SUBSEAFLOOR PROKARYOTIC AND VIRAL COMMUNITIES IN SEDIMENT FROM AN ANOXIC MARINE BASIN

A Thesis

by

CATHERINE ANN RISLEY

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Jessica M. Labonté
Committee Members,	Peter van Hengstum
	Brandi Kiel Reese
	Ron I. Eytan
Head of Department,	Daniel Roelke

August 2020

Major Subject: Marine Biology

ABSTRACT

Prokaryotes make up the majority of the biomass in sediment, where they play a role in cycling organic carbon and regulate the fluctuation of organic matter. Studies of abundance, genomic composition, and ecological roles of prokaryotes and viruses indicate that viruses are significant contributors to subseafloor ecosystems. In anoxic coastal sediment, the availability of electron donors and acceptors could be an important driver of the microbial community composition and functionality. The main goal of this study was to define the relationships between prokaryotic diversity and activity, nutrient availability, and organic matter source within sediment of an anoxic basin (Blackwood Sinkhole, Bahamas), to determine their roles in biogeochemical processes. The main goal was achieved by answering the following specific objectives: (1) determine the source of carbon in a 0.9 m long core dated to ~2,500 years, (2) define the relationship between the origin of carbon and microbial community composition, (3) characterize the diversity and analyze the functionality of prokaryotes from an anoxic sinkhole basin, and (4) determine the availability of electron donors and acceptors and their relation with microbial genomic potential and activity, using viral production and virus-induced microbial mortality as a proxy for microbial activity. The source of carbon, whether terrestrial or planktonic, impacted the microbial community composition as there was a significant correlation between the microbial community abundance and the C:N ratio. Genes involved in sulfate reduction, sulfur oxidation and denitrification were among the main metabolisms found. The rate of viral production and virus-induced microbial

mortality was linked to nutrient concentration. These results show that the microbial community present (Bacteria, Archaea, and viruses) and their interactions with nutrients available in anoxic sinkhole sediment may play an important role in the global sulfur, nitrogen, and carbon cycles.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Jessica Labonté, and my committee members, Dr. van Hengstum, Dr. Reese, and Dr. Eytan, for their guidance and support throughout the course of this research.

Thanks also go to my lab mates (Labonté Viral Ecology Laboratory) for assisting me in my many extractions, sequencing preparation, and data analysis. Specifically, many thanks to Cameron Jackson and Jordan Walker for staying late in the lab working on sampling my incubations with me and to Madeleine Thompson for making over 200 slides for me. Thank you to Annie Tamalavage for all the assistance with my carbon and nitrogen measurements and analysis.

Finally, thanks to my parents for their encouragement, to my partner for his patience and love, and thanks to my two rescue puppies for keeping me sane.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professor Jessica Labonté of the Department of Marine Biology, Pete van Hengstum of the Department of Marine Science, Brandi Reese of the Department of Life Sciences at TAMU-CC, and Ron Eytan of the Department of Marine Biology.

The data analysis for some of Chapter II was performed with the assistance of Annie Tamalavage. Portions of this research were conducted using the advanced computing resources and consultation provided by Texas A&M High Performance Research Computing. All other work conducted for the dissertation was completed by the student independently.

Funding Sources

The study was mainly supported by a Research Grant from the National Science Foundation Technology Center for Dark Energy Biosphere Investigations to Drs. Jessica Labonté and Pete van Hengstum. Catherine Risley was also the recipient of a Texas Sea Grant Grants-in-Aid. Summer funding support was provided by the Research and Graduate Studies office at Texas A&M University Galveston.

NOMENCLATURE

˚C	Degrees Celsius
BLWD	Blackwood Sinkhole
C:N	Carbon to Nitrogen atomic ratio
Cal yrs BP	Calibrated years Before Present
DNA	Deoxyribose Nucleic Acid
dsDNA	Double Stranded DNA
OC	Organic Carbon (%)
OTU	Operational Taxonomic Unit
POM	Particulate Organic Matter
rRNA	Ribosomal Ribonucleic Acid
TC	Total Carbon
TOC	Total Organic Carbon
TN	Total Nitrogen (%)

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. INTRODUCTION

Microbes, including protists, prokaryotes, and viruses, play an important role in cycling organic carbon and regulating the fluctuation of organic matter in sediment. Within coastal sediment the microbial community interactions with geochemical pathways effect ecosystem function, which may be different to deep-sea sediment communities (Lachnit *et al.*, 2019). In the deep sea, the rate of biogeochemical cycles are much slower in the sediment than in the water column; and in general, microbial life buried in subsurface sediment have an extremely slow mean metabolic rate and growth rate where generation times can be up to thousands of years (Jorgensen *et al.*, 2016). A microbial community may only go through 10,000 generations from the burial time until it reaches a depth of tens of meters in the time span of several millions of years (Meister *et al.*, 2005). Slow growth rates could be due to the low energy flux available to individual cells, labile carbon unavailability, or regulated by virus-induced mortality (Jorgensen *et al.*, 2007; Jorgensen *et al.*, 2016).

1.1. Microbial life in an anoxic sediment

The diversity of prokaryotes is driven by their use of a variety of energy sources and electron acceptors (Nealson, 1997; Walsh *et al.*, 2016). Both Bacteria and Archaea stimulate essential processes within sediment, such as the production of methane and other hydrocarbons, removal of sulfate, and the oxidation of organic matter (Kallmeyer *et al.*, 2012).

Microbial communities in sediment play an important role in cycling organic carbon and regulate the fluctuation of organic matter. In deep-sea sediment, generation times of the microbial communities can be up to thousands of years and the regulation of slow mean metabolic microbial growth rate is believed to be caused by low energy flux availability and carbon unavailability (Jorgensen *et al.*, 2007; Jorgensen *et al.*, 2016). These slow growth rates and difficulty in recreating these environments in the laboratory make cultivation in the lab challenging (Harwani, 2013), therefore cultivation-independent techniques such as 16S ribosomal RNA (rRNA) survey and metagenomics are some of the current approaches used to study these microbial communities.

Prokaryotic diversity can be studied through the sequencing of the 16S rRNA gene. This gene is present in all species, is highly conserved within prokaryotic species and its function has not changed over time (Janda *et al.*, 2007). The Earth Microbiome Project has been able to use 16S rRNA gene sequencing to characterize, at a taxonomic level (Thompson *et al.*, 2017), the microbial communities from crowd-sourced environmental samples (Gilbert *et al.*, 2014). The Census of Deep Life, a Deep Carbon Observatory initiative, aims to explore the microbial communities living below the seafloor and in the continental crust using 16S rRNA gene deep sequencing. As a part of the Census for Deep Life, Labonté et al. (2017) characterized the microbial diversity within the subsurface basalt-sediment interface of the Juan de Fuca Ridge flank. The sediment samples closest to the sediment-basement interface were dominated by three phyla, Chloroflexi, Aerophobetes, and Nitrospirae, and differed greatly from the basalt samples (Labonté *et al.*, 2017); the main influence affecting sediment microbial

communities along the subsurface was based on the depth and the type of sediment sampled (*i.e.*, clays or carbonate-rich) rather than the geochemistry of the pore water.

In contrast to 16S rRNA gene surveys, metagenomic analysis (the study of all the genetic material recovered from an environmental sample) provides insight into the functionality of the entire community. Metagenomic sequencing was used to identify the percentage of genes for specific metabolic pathways between sample depths in a sediment core from the Peru Margin as well as to determine the taxonomic assignment based on 16S rRNA gene conservation (Biddle *et al.*, 2008). The major metabolic pathways found were sulfate reduction and methanogenesis (Biddle *et al.*, 2008). Given the difficulty of cultivating microbes in the laboratory, metagenomic data served as a pivotal tool in the discovery and exploration of ecologically important microorganisms.

1.2. Nutrient geochemistry and availability within coastal, freshwater, and marine sediment

In deep-subsurface environments, pore water geochemistry has been linked to the residential microbial communities with their metabolism serving as a reflection of the main element found within the sediment (Konhauser, 2009), which contains mineral particles, organic matter, and organic mineral particles (Parmar *et al.*, 2013). Microbes play major roles in several fundamental processes in ocean subsurface sediment, including the removal of sulfate from the oceans, the oxidation of organic matter, and the production of hydrocarbons (Kallmeyer *et al.*, 2012). They also catalyze the transformation of sulfur for their metabolism and biomass (Ehrlich *et al.*, 2015).

Microbes inhabiting sediment depend on the availability of electron donors and acceptors, as well as other environmental conditions, such as depth and temperature (Konhauser, 2009). Different types of microbes, such as chemolithoautotrophic and chemoheterotrophic bacteria, perform different functions based on their metabolic properties (*e.g.*, reduction, oxidation of organic compounds). In sediment, oxygen gets quickly depleted, while other electron acceptors such as sulfate, nitrate, and ferric iron increase in concentration with depth. This results in a sequence of oxidation-reduction reactions (redox tower) that changes the pH, and influences metal cation solubility, valence of ions, and nature of the molecules dissolved in pore water (Konhauser, 2009). Without the use of oxygen, the availability of alternative electron donors and acceptors in anoxic sediment is highly dependent on concentrations of organic carbon (Orsi, 2018). Due to these gradients, microbial communities can be vertically stratified and composed of interdependent layers of chemotrophic and heterotrophic microorganisms (Biddanda *et al.*, 2012).

In anaerobic environments, acetate, a compound produced via acetogenesis that can be anaerobically oxidized to CO_2, is an important intermediate of the organic carbon cycle. Acetogenesis, the process through which acetate is produced either by the reduction of CO_2 or by the reduction of organic acids via the acetyl-CoA pathway, is completed in anaerobic environments mostly by microbes in the phylum Firmicutes (Schuchmann *et al.*, 2016). Acetogenic microbes play an important role in carbon cycling due to the flux of organic carbon into the acetate pool (Liou *et al.*, 2008).

Proteogenomics on acetate-amended sediment samples from a freshwater aquifer in Colorado showed that *Desulfuromonadales* contribute to acetate fueled sulfate reduction, Bacteroidetes reduce nitrogen, and that Epsilonproteobacteria facilitate sulfide oxidation and carbon fixation (Handley *et al.*, 2013). Microbial communities influenced by acetate facilitate the sulfur, carbon and nitrogen cycles (Handley *et al.*, 2013).

Porewater geochemistry and the sediment microbial community structure were compared at intertidal areas of a saltmarsh (Koretsky *et al.*, 2005). Saltmarshes have high primary productivity, organic carbon rates, and redox stratification (Koretsky *et al.*, 2005). Organic carbon can be oxidized via aerobic respiration, ferric iron reduction, and methanogenesis and the oxidation of organic carbon is usually coupled with microbe facilitated sulfate reduction. In addition to porewater measurements of alkalinity, pH, conductivity sediment cores were also taken in three saltmarsh areas with different vegetation levels. Researchers identified a large number of anaerobic microbes (*e.g.*, manganese-reducing bacteria, iron-reducing bacteria, and sulfate-reducing bacteria) in the shallower depth of the cores. The porewater profiles showed a vertical redox stratification, with no clear trends to suggest microbial community structure was impacted by the porewater geochemistry (Koretsky *et al.*, 2005). These results demonstrate only a small distinction between the redox potential, carbon content and vegetation impact on the microbial community structure in saltmarshes versus deep subsurface environments.

1.3. Importance of the nitrogen cycle in sediment

The nitrogen cycle provides an essential component of life to all organisms and is limited in marine environments (**Figure 1**). A study by Kuypers et al. (2003) found an abundance of denitrifying bacteria (able to convert nitrate/nitrite to nitrogen gas), within the Black Sea, which is naturally anoxic and harbors high concentrations of ammonium. Prior to that discovery, it was thought that denitrification only occurred aerobically (Kuypers *et al.*, 2003). Anaerobic ammonium oxidation (anammox) bacteria belong to the order Planctomycetales and directly oxidize ammonia to nitrogen gas with nitrite as the electron acceptor. Biomarkers (*e.g.*, ladderane lipids (Damste *et al.*, 2002)) were used to trace anammox bacteria in the particulate organic matter within the anoxic water column samples. Results indicated that anammox bacteria are important players in the nitrogen cycle of the Black Sea. Both anammox and denitrification reactions are readily occurring in anoxic sediment (Ehrlich *et al.*, 2015). Anammox bacteria could have an important role in the oceanic loss of fixed nitrogen due to ammonia within the water column (Kuypers *et al.*, 2003).

Nitrification occurs when ammonia is oxidized to nitrate (Bernhard, 2010) (**Figure 1**) and is usually carried out by heterotrophic microbes, either bacteria or fungi (Mouton *et al.*, 2012). Dissimilatory nitrate reduction (when nitrate transforms into nitrite) further reduces nitrite to ammonium (**Figure 1**). Nitrogen cycling reactions are likely governed by microbial metabolism (Ehrlich *et al.*, 2015).

A key process of the nitrogen cycle is the bacterial oxidation of nitrite to nitrate (nitrification) (**Figure 1**). Ammonia-oxidizing bacteria and nitrite-oxidizing bacteria

facilitate the process of nitrification, where the oxidation of ammonia or ammonium to nitrite is followed by the oxidation of the nitrite to nitrate (Koch *et al.*, 2014). The overall process of oxidation of ammonia to nitrate (comammox) (**Figure 1**) is energetically favorable and can be potentially completed by *Nitrospira* bacteria (Santoro, 2016). Sequencing of two isolated strains of *Nitrospira* sp. (*Candidatus* Nitrospira nitrosa and *Candidatus* Nitrospira nitrificans) introduced to low concentrations of ammonium in a stable bioreactor revealed that they were able to complete the entire nitrification process. The genome of both species were found to possess the genes that encode for the ammonia monooxygenase (*amo*) and hydroxylamine dehydrogenase (*hao*) for ammonia oxidation, as well as the nitrite oxidoreductase (*nxr*) subunits needed for nitrite oxidation in *Nitrospira* (Daims *et al.*, 2015; van Kessel *et al.*, 2015).

Metagenomic data can be used to identify metabolic pathways if genes associated with certain pathways are present. Ammonia-oxidizing bacteria and archaea were found in sedimentary environments from Shanghai, China (Xia *et al.*, 2018). The ammonia-oxidizing bacteria, related to *Nitrospira* sublineage II, encoded for nitrification genes (*amoA*, *nxr*, etc.) (Poghosyan *et al.*, 2019; Xia *et al.*, 2018).

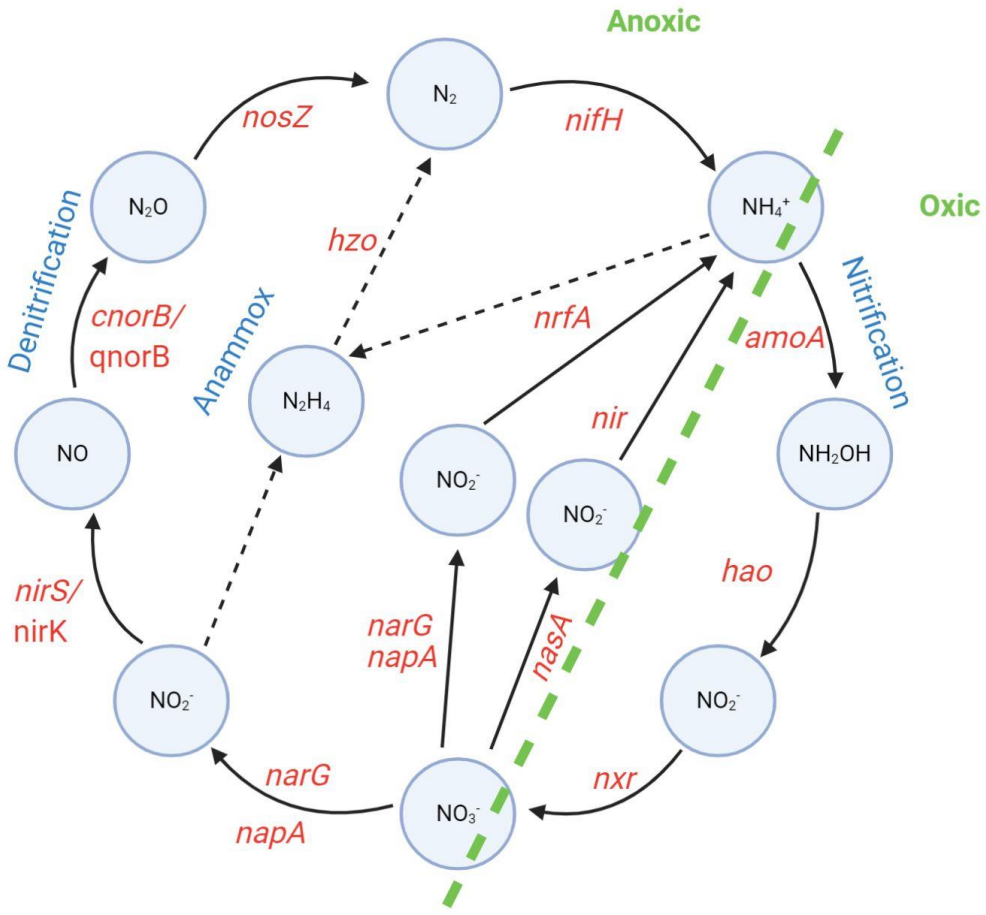

Figure 1: Microbial facilitated nitrogen cycle with associated genes (indicated in red): Anoxic and oxic pathways are separated by the green dashed line. Black dashed arrows show steps of the anaerobic nitrogen cycle specifically associated with the anammox pathway. Adapted from (Smith *et al.*, 2015).

1.4. Importance of the sulfur cycle in sediment

The sulfur cycle involves all processes influencing the transfer of sulfur compounds to and from the various Earth sulfur reservoirs (**Figure 2**). Biogeochemical processes, such as sulfate reduction and sulfur oxidation, depend on microbial catalysts and sediment depths where nutrients are more prevalent (Holmkvist *et al.*, 2011). Sulfate reduction, methanogenesis, and fermentation are the primary metabolic processes in

subsurface sediment (D'Hondt *et al.*, 2002). A study by Leloup et al. (2007) using sediment from the Black Sea, showed that dissimilatory (bi)sulfite reductase (*dsrAB*) genes can be used as functional markers instead of 16S rRNA genes to investigate the sulfate-reducing community and the metabolic processes. Moreover, the sulfate-reducing bacterial population in water samples off the Chilean Coast was found to be diverse at oxygen minimum zones (OMZ) and included *Desulfatibacillum*, *Desulfobacterium*, *Desulfococcus*, *Syntrophobacter*, and *Desulfovibrio* species. Although the taxa were genetically diverse, the sulfite reductase enzyme (*dsr*) gene was present in all groups (Canfield *et al.*, 2010).

Microbial sulfate reduction (**Figure 2**) is catalyzed by Archaea and Bacteria. Sulfate-reducing bacteria are both organotrophs (*i.e.*, use organic compounds as an electron donors) and lithotrophs (*i.e.*, use hydrogen as an electron donor), belonging to the genera *Desulfovibrio*, *Desulfotomaculum*, and *Desulfomonas* (Ehrlich *et al.*, 2015). The sulfate reducing Archaea are the Euryarchaeota and Crenarchaeota (*i.e.*, phyla). Some sulfate reducers are also involved in methane oxidation. Some sulfate reducers are able to use oxygen as an electron acceptor, but this is rare due to the anaerobic environments these reducers are known to occupy. There is little known about coastal anoxic environments as most studies have been focused on water column or deep-sea communities.

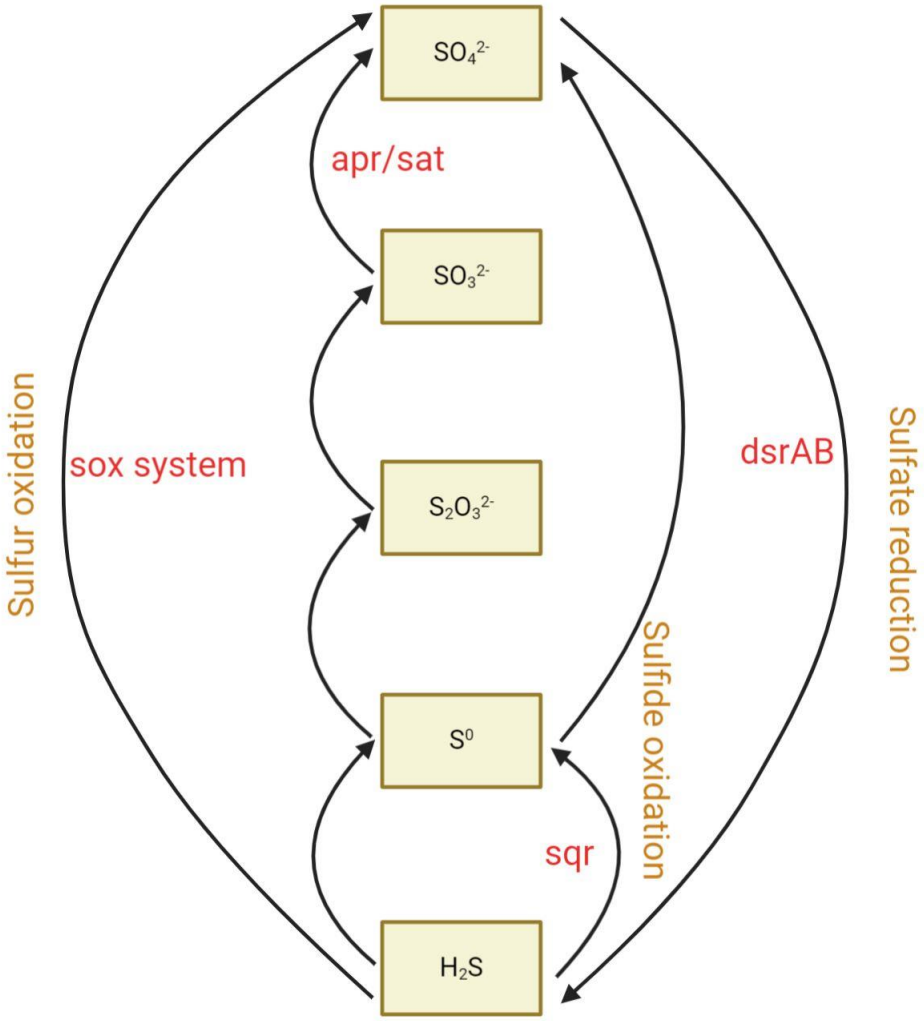

Figure 2: Major microbial sulfur cycle metabolisms and associated genes (indicated in red): sulfate reduction transforms sulfate to hydrogen sulfide (dsrAB), sulfide oxidation transforms sulfide to a more oxidized state between elemental sulfur and sulfate (sqr), and sulfur transforms these compounds to both H_2S and SO_4^{2-} Adapted from (Ehrlich *et al.*, 2015).

1.5. Abundant and important viruses in coastal, deep-sea, and freshwater sediment

With the oceans estimated to contain ~ $4x10^{30}$ viruses, viruses are the most abundant life forms on Earth. It is estimated that there is an average of ~$5x10^8$ viruses

per gram of sediment (Corinaldesi *et al.*, 2010; Suttle, 2005). Viruses are obligate intracellular parasites, as they rely on a host cell for energy, metabolic processes, and protein synthesis. They infect organisms from all three domains of life, and are extremely diverse in regards to their morphology, genetic material, and the type of hosts they infect (Anderson *et al.*, 2013). Bacterial viruses are known as bacteriophages, or phages, and they follow two major life cycles: lysis and lysogeny. The lytic cycle occurs when the virus injects its genetic material into the host cell and highjacks the cell to multiply, it ends with the lysis of the cell and release of multiple viral particles (Anderson *et al.*, 2013; Labonté *et al.*, 2015). In the environment, the process known as the "viral shunt" occurs when viruses lyse cells, converting biomass into dissolved and particulate organic matter (DOM), which then stimulates bacterial production by providing a source of DOM and contributes to respiration (Suttle, 2007). The lytic cycle involves therefore the release of organic carbon and other nutrients back into the environment (Weitz *et al.*, 2012). Some bacteriophages can go into the lysogenic cycle. Lysogeny occurs when the phage genome gets integrated into the host's genome, where it will replicate in conjunction with the host chromosome (Suttle, 2005). The lysogenic cycle facilitates phage-mediated horizontal gene transfer through transduction. Lysogeny changes the composition of the host DNA and most commonly occurs when the environmental conditions are unfavorable (Brüssow *et al.*, 2004; Finke *et al.*, 2017). The phages may later be induced due to environmental triggers and go into the lytic cycle (Brüssow *et al.*, 2004; Racine, 2014).

Viruses are significant contributors to subseafloor ecosystems because of their abundance, genomic composition, and ecological roles (Breitbart *et al.*, 2004). Currently, it is hypothesized that lysogeny is more prevalent under suboptimal conditions for host survival and growth in environments such as the subsurface (Williamson *et al.*, 2008). Viruses can also be a driving factor of nutrients recycling through mortality via lysis of prokaryotes in deep-sea sediment. Sediment properties, such as iron rich areas, affect the viral population and could provide a greater number of viruses due to iron oxyhydroxide minerals absorbing viral particles (Pan, Morono, *et al.*, 2019). Cell abundance greatly depends on the slow microbial death rate in sediment (Orsi, 2018).

In marine sediment, viruses are highly abundant and persistently outnumber prokaryotic cells by one to two orders of magnitude (Engelhardt *et al.*, 2014). A high virus-to-bacteria ratio, as observed in sediment from the Peru margin and equatorial Pacific (Engelhardt *et al.*, 2014), suggests that the production of viruses through lysis is ongoing; indicating that viruses in sediment may play a role in the viral shunt. A low virus-to-bacteria ratio suggests a low production of viruses, and possibly a higher incidence of lysogeny. With an increase of depth below seafloor and organic carbon, viral abundance also increased (Cai *et al.*, 2019). Within deep-sea sediment, extrapolated estimates of viral abundances are between 0.25 and 3.5 x 10^{30} particles (Parikka *et al.*, 2017). Coastal sediment carries up to 33% more cells than deep-sea sediment potentially due to carbon rich areas (Pan, Morono, *et al.*, 2019). Although cell abundance is increased in coastal areas, compared to the highly oxygenated sediment,

the virus-to-prokaryote ratio is known to be lower in anoxic sediment compared to oxic sediment (Cassman *et al.*, 2012; Orsi, 2018). Exploring the dynamics between microbes and the associated viral community will further our understanding of the role of bacteria and viruses in biogeochemical cycles.

Metagenomic sequencing has provided the opportunity to capture the genetic richness of viruses in marine ecosystems through collection of genetic material directly from the environment. Metagenomic analysis was used to evaluate the single-stranded DNA (ssDNA) and double-stranded DNA (dsDNA) viruses in offshore Japan sediment (Yoshida *et al.*, 2018). The dsDNA sequences were dominated by four main viral families, the tailed bacteriophage families *Siphoviridae*, *Podoviridae*, *Myoviridae*, and the eukaryotic algae viral family *Phycodnaviridae* (Yoshida *et al.*, 2018). Shotgun metagenomic sequencing conducted on water samples from five different hydrothermal vent plumes identified sulfur-oxidizing bacteria, along with 18 double-stranded DNA viral sequences. Some of these viruses belonged to three marine viral families of the orders Caudovirales: *Podoviridae*, *Siphoviridae*, and *Myoviridae* (all bacteriophages). Further exploration into the amino acid sequences revealed certain genes encoded for sulfur-oxidizing enzymes, suggesting that viruses carrying metabolic genes may play a role in the biogeochemical cycles (Anantharaman *et al.*, 2014).

Metagenomic sequencing was conducted to evaluate viral populations within microbial communities and to predict the environmental drivers of community diversity. The Global Ocean Viromes 2.0 (GOV 2.0) dataset, consisting of marine viral metagenomic samples, was compiled from the original Tara Oceans and Malaspina

expeditions to provide a more comprehensive view of viral populations. The data mostly consists of viral dsDNA and unveiled how the physiochemical structure of microbial communities play an important role in the global host range of viruses (Gregory *et al.*, 2019). Using metagenomic sequencing and methods to specifically extract viral DNA, the viral diversity within microbial communities and virus–host interactions is becoming better understood (Corinaldesi *et al.*, 2017).

Virus–host interactions are impacted by host availability, productivity, and nutrients within their environment. Viruses are abundant in many ecosystems and impact a variety of hosts. Particulate organic matter (POM) and eutrophic conditions can affect the microbial community abundance and nutrient availability, which in turn affects the virus–host interactions. After comparing contaminated and reference coastal surface sediment, viral community composition was found to be altered due to the carbon and nitrogen content, as well as the anthropogenic contaminants, such as pesticides, heavy metals and waste water (Lachnit *et al.*, 2019). Further exploration into the effects on the microbial community structure prove vital to understanding the abundance of these communities within surface sediment.

1.6. Objectives

The main goal of this study is to define the relationships between prokaryotic diversity and activity, nutrient availability, and organic matter source within sediment of an anoxic basin (Blackwood Sinkhole, Bahamas), to determine their roles in biogeochemical processes.

Objective 1: Determine the origin of carbon (terrestrial or planktonic) in a 0.9 m long core dated to ~2,500 years. Hypothesis: There will be variations in the measured carbon-to-nitrogen ratio in each stratigraphic layer, based on the potential source of organic matter, whether it is >20 (terrestrial input) or <20 (phytoplankton input). For example, if leaves can be seen in any layers the C:N ratio will be much higher than 20.

Objective 2: Define the relationship between the origin of carbon and the microbial community composition. Hypothesis: The differences in subseafloor microbial communities will be due to the organic carbon originally available at time of burial.

Objective 3: Analyze the prokaryotic diversity and metabolic potential from an anoxic sinkhole basin. Hypothesis: Microbial communities will be mainly influenced by geochemical gradients and nutrients, therefore it is expected that there will be a succession of metabolic genes involved in sulfur oxidation (*sox*), sulfate reduction (*dsr*), denitrification (*nosZ*, *nifH/nifS*), as we go down core, following a redox gradient.

Objective 4: Determine the effect of electron donors and acceptors availability on microbial activity, using viral productions as a proxy for microbial activity. Hypothesis: Due to the anoxic nature of the sediment core sample, nitrates, sulfates, and phosphates will be the main electron acceptors. Within each sediment layer the microbial mortality rate will be higher when the source of carbon is terrestrial and will be lower when the source of carbon is phytoplankton. Viral activity will be low when there is a decrease in nutrients within anoxic sediment.

2. IMPACT OF CARBON CONTENT AND SOURCE ON THE MICROBIAL COMMUNITIES INHABITING SEDIMENT IN AN ANOXIC SINKHOLE

2.1. Introduction

Sinkholes are developed from the subsurface dissolution and subsequent collapse of limestone bedrock (Mylroie *et al.*, 1995). In some cases when the sinkhole is close to the ocean, there is a vertical migration of the water table, *i.e.*, there is a saltwater input from the bottom and freshwater input on the surface, causing salinity stratification (Collins *et al.*, 2015; van Hengstum *et al.*, 2011). The Blackwood Sinkhole has a sedimentation rate of ~0.4–1.2 mm yr^{-1} (~1,600 Cal yrs BP), which is caused by organic matter deposition from primary productivity and erosion (Tamalavage *et al.*, 2018). The relatively high sedimentation rates provide highly stratified sediment with great preservation potential that have been used to identify environmental changes and hydroclimate records over long timescales (Tamalavage *et al.*, 2018). Anoxic coastal basins provide ideal, easy to access, settings to provide greater understanding of microbial communities living under anaerobic conditions, and could potentially serve as an environmental analog to anoxic marine basins.

Sequencing of the 16S ribosomal RNA (rRNA) gene has been used to identify the members of the microbial community inhabiting sediment from coastal marine environments (Koretsky *et al.*, 2005; Lenk *et al.*, 2012; Ul-Hasan *et al.*, 2018), freshwater lakes (Biddanda *et al.*, 2012; Kojima *et al.*, 2012), and the deep-sea (Engelen *et al.*, 2008; Inagaki, 2015; Labonté *et al.*, 2017; Reed *et al.*, 2002). The microbes found

within these different environments below the subsurface are physiologically versatile, and extremely diverse in terms of their functionality (Orcutt *et al.*, 2011).

Although there is a significant contribution from primary productivity within sinkholes, organic matter found buried in the sediment can also come from the terrestrial environment, *i.e.* leaves and tree debris (Lamb *et al.*, 2006). Similarly to sinkholes, in coastal marine environments, organic carbon content can be deposited from the terrestrial landscape and phytoplankton productivity in the overlying water column (Glud, 2008; Lamb *et al.*, 2006). The origin of organic matter, terrestrial or phytoplanktonic, can influence the microbial community due to the functionality of the microbes (Fagervold *et al.*, 2014).

There are many hypotheses as to what the drivers of the microbial communities inhabiting sediments are. It was showed that variations in microbial communities may be caused by changes in sediment chemistry and texture, which includes clays, carbonate-content, and organic carbon rather than by the pore water geochemistry (Labonté *et al.*, 2017). In various sedimentary environments that include the North Frisian Wadden Sea sand flat and the Belgian Continental Shelf, the microbial community has been linked to enzymatic activity (Böer *et al.*, 2009) and porewater nutrients (Franco *et al.*, 2007). Microphytobenthic algae productivity also affects the distribution of the microbial community within sediment (Böer *et al.*, 2009).

The **goal of this study** was to define the relationship between prokaryotic diversity and activity, and organic matter source within sediment of an anoxic basin. We measured the carbon-to-nitrogen ratio to identify the potential source of organic matter,

whether it is from terrestrial plants or aquatic primary productivity. We used 16S rRNA gene sequencing to characterize the microbial community. I hypothesized that the differences in subseafloor microbial communities would be due to the organic carbon originally available at time of burial.

2.2. Study site

Blackwood Sinkhole is located ~220 m from the shoreline on the northeastern coast of the Great Abaco Island in the northern Bahamas (**Figure 3**) (van Hengstum *et al.*, 2016). The regional terms sinkhole (North America), blue hole (Bahamas mostly), and cenote (Mexico) all describe similar geomorphologic features on carbonate landscapes that are the surface expression of subsurface dissolution and subsequent overburden collapse (Mylroie *et al.*, 1995). On Quaternary timescales, water level in blue holes is concomitantly linked to eustatic sea-level change, whereby sinkholes are repetitively flooded and drained by eustatic sea-level rise and fall, respectively (van Hengstum *et al.*, 2016; van Hengstum *et al.*, 2011). Currently flooded blue holes are groundwater-fed aquatic environments whereby, in general, the modern hydrography is characterized by an upper meteoric water mass of fresh to brackish water that overlies anoxic, saline groundwater intruding from the ocean (van Hengstum *et al.*, 2011). Blackwood Sinkhole is similarly a stratified and groundwater-fed system, characterized by stratified water column of 15 m of meteoric water, 40 m of anoxic saline groundwater (39.9 psu).

The groundwater stratification and geomorphology of Blackwood Sinkhole has promoted great sediment preservation from a lack of vertical mixing from either

invertebrate bioturbation, or physical mixing from wave action or winds. Basal sedimentary deposits in Blackwood Sinkhole are carbonate gravels, which abruptly transition to laminated sapropel interbedded with carbonate horizons that were deposited over the last 3000 years (van Hengstum *et al.*, 2016). The laminated sapropels have been previously determined to have a high organic total carbon content (TOC of ~10%, (Tamalavage *et al.*, 2018). However, sedimentation rate and location in the sinkhole has not been laterally or temporally uniform over time. Constant sedimentation (~0.4 mm yr^{-1}) was initiated 3,000 years ago on the sinkhole periphery (BLWD-C2), with evidence of coarse-grained particle deposition during intense hurricane strikes on Abaco Island (van Hengstum *et al.*, 2016). In contrast, the onset of sedimentation was delayed in the center of the sinkhole until ~1,600 Cal yrs BP, and sedimentation rates were elevated (BLWD-C3, 1.2 mm yr^{-1}), which was caused by increased organic matter delivery to the benthos from (i) primary productivity and (ii) erosion of sediment from the adjacent wetlands at ~1,000 years Cal yrs BP (Tamalavage *et al.*, 2018).

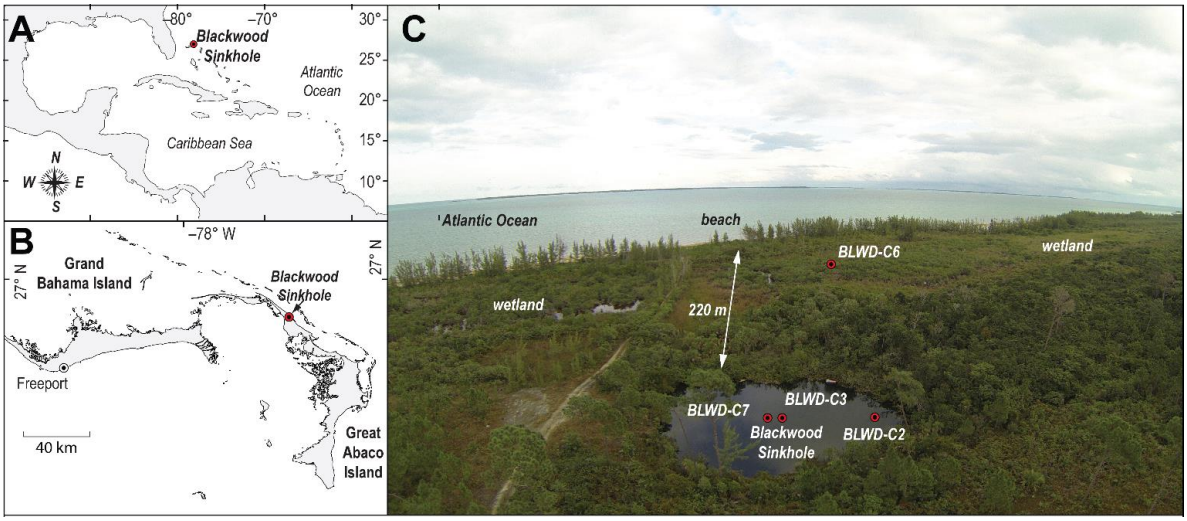

Figure 3: Map of BLWD-C7 and the Bahamas region. (A) The western North Atlantic region noting the location of the Blackwood Sinkhole on the Little Bahama Bank. (B) Islands on the Little Bahama Bank, and the position of the Blackwood Sinkhole on the Great Abaco Island. (C) Aerial photograph of the Blackwood Sinkhole and the relation (220 m) to the Atlantic Ocean. Figure adapted from Tamalavage et al. 2018.

2.3. Methodology

2.3.1. Sample collection and handling

A push core (BLWD-C7, 26.79°N, 77.42°W) was taken on 07/29/2018 by using advanced scuba techniques. The periphery of the sinkhole bottom (not center) was targeted for core sampling in an attempt to re-collect the last 3000 years of sedimentary deposition, as has been previously documented (van Hengstum *et al.*, 2016). In the lab, the core was sectioned lengthwise, photographed and x-radiographed, and subsequently stored at 4°C until further analysis. Twenty-four stratigraphic layers were identified based on coloring and soil type. The 24 layers were extracted and stored at -80°C for further DNA extraction. The core was draw in color based on distinct light and dark

sediment layers, along with notes of the location of leaves, twigs, and coarse-grained particles. For five samples (20.1−22.2 cm, 32.9−35 cm, 52.1−54.2 cm, 64.5−67.7 cm, and 79.5−86.9 cm), age constraint for the core (BLWD-C7) was determined by radiocarbon dating terrestrial plant macrofossils (*e.g.*, leaves, twigs) by accelerator mass spectrometry at the National Ocean Sciences Accelerator Mass Spectrometry facility at Woods Hole Oceanographic Institution. Conventional radiocarbon results were calibrated into calendar years before present (Cal yrs BP) with IntCAL13, using 1950 CE as present (Reimer *et al.*, 2013). Final downcore Bayesian age models for BLWD-C7 were computed using the R program Bacon v2.2 (Blaauw *et al.*, 2011), which provides probability estimates at each core depth.

2.3.2. Measurement of carbon and nitrogen content

Total carbon and nitrogen content measurements were performed for 24 sediment subsamples down core (**Table 1**). First, the samples were freeze dried overnight, homogenized, and 2–6 mg of ground sample were placed into tin capsules and measured on a CHN analyzer (Costech instruments ECS 4010 CHNSO Analyzer) to measure total carbon (TC) and total nitrogen (TN). Data calibration was determined relative to acetanilide and standard reference material for marine sediment according to the National Institute of Standards and Technology (NIST). To measure organic carbon (with mass correction applied) the samples were acidified using 8 mL of 1M hydrochloric acid for 24 hr or until effervescence ceased, then desiccated at 60˚C, and re-homogenized. The grounded, acidified samples were weighed (0.5 to 1.8 mg) into silver capsules then analyzed on the CHN analyzer. The potential loss of carbon from the

direct acidification process was corrected by multiplying the percent of sample remaining (post acidification weight subtracted from pre acidification weight/pre acidification weight) (Tamalavage *et al.*, 2018). The atomic C:N ratio was determined using the organic carbon (OC) acidified values divided by the TN values unacidified and multiplied by the molecular weight ratio (14.01/12.01) (Tamalavage *et al.*, 2018). For measurements of TC and TN, replicates yield a mean precision of 2%.

Table 1: Downcore C:N atomic ratios, OC% (organic carbon corrected), TN% (total nitrogen), and sample loss % for core BLWD C7.

Core	Depth Interval (cm)	C:N atomic	OC %	TN %
BLWD C7	2-3	14.35	9.60	0.78
BLWD C7	4-6	17.26	9.32	0.63
BLWD C7	6-13.5	18.47	11.40	0.72
BLWD C7	14-15.6	21.64	10.02	0.54
BLWD C7	16.5-19.7	18.33	11.63	0.74
BLWD C7	20.1-22.2	14.05	6.99	0.58
BLWD C7	22.4-23.7	19.56	15.09	0.90
BLWD C7	24.3-27.5	14.43	5.32	0.43
BLWD C7	27.6-30.2	16.04	8.66	0.63
BLWD C7	30.2-32.8	25.43	16.57	0.76
BLWD C7	32.9-35	21.48	14.91	0.81
BLWD C7	35.2-39.4	19.19	9.54	0.58
BLWD C7	39.6-42.3	23.77	16.10	0.79
BLWD C7	43-45	25.01	12.87	0.60
BLWD C7	46.5-52	32.20	22.64	0.82
BLWD C7	52.1-54.2	19.36	10.95	0.66
BLWD C7	54.5-58.2	23.98	14.59	0.71
BLWD C7	58.5-61	21.32	7.49	0.41
BLWD C7	61.5-64	17.98	7.09	0.46
BLWD C7	64.5-67.7	22.21	13.90	0.73
BLWD C7	68-72.3	27.53	12.51	0.53
BLWD C7	72.5-76.9	39.51	17.27	0.51
BLWD C7	77.1-79	21.55	3.88	0.21
BLWD C7	79.5-86.9	19.45	8.50	0.51

2.3.3. 16S rRNA gene and analysis

Total DNA was extracted from all 24 samples using the DNeasy PowerSoil kit (Qiagen, USA) and stored at -20˚C until PCR amplification (**Table 2**). A blank control was also extracted to identify possible contamination from the ambient lab and kit reagents (Glassing *et al.*, 2016). The PCR primers 515F (5'-GTGYCAGCMGCCGCGGTAA-3') and 806R (5'-CCGYCAATTYMTTTRAGTTT-3') were used to target the V4 region of the 16S rRNA gene (Parada *et al.*, 2016). Thermal cycling was performed under the following conditions: initial preheating for 3 min at 94˚C; 35 cycles of denaturation at 94˚C for 1 min, annealing at 50˚C for 1 min, and extension at 72˚C for 1 min 45 sec; final extension at 72˚C for 10 min. PCR was completed in triplicates and the products were pooled and cleaned using the MinElute PCR Purification Kit (Qiagen, USA). Amplicons (100-200 ng) were sequenced with Illumina MiSeq with 250 bp paired-ends at the Texas A&M AgriLife Bioinformatics and Genomics facility. The library was prepped using the Amplicon library preparation kit.

PCR analyses of the 16S rRNA gene amplicons were completed using *mothur* as presented in the MiSeq Standard Operating Procedure (SOP) example analysis (Kozich *et al.*, 2013), which included reducing sequencing and PCR errors, processing the improved sequences, running an alignment using the reference SILVA v132 alignment, removing poorly aligned sequences (sequences that start and end at a different position due to non-specific amplification) and undesirables, pre-clustering the sequences into amplicon sequence variants (ASVs), clustering the sequences using the split method argument to classify (splitting the sequences into distinct taxonomy groups, and splitting

the distance file based on those groups), running ASV-based analyses, running Operational Taxonomic Unit (OTU)-based analysis (rarefaction and heatmap), and generating visualizations. Using Bray-Curtis dissimilarity matrix, beta-diversity between samples was examined and ordinated by non-metric multidimensional scaling (NMDS) in R (RCore Team, 2013), with overlaying the carbon content parameters applying the envfit function from the vegan package (Dixon, 2003; Franco *et al.*, 2017). Analysis of stratigraphically-constrained hierarchical clustering according to the age model of the core was completed using the package rioja in R (Juggins *et al.*, 2019).

Table 2: Sediment sample initial weight (g), DNA extractions weight taken, and weight leftover for porewater extraction. Highlighted in blue are samples greater than 50g.

Sample	Lamination Thickness	Initial Weight (g)	Weight (g) taken for Incubation Experiments	Weight (g) taken for 16s rRNA
BLWD C7 2–3 cm	1 cm	21.345	1.5g	2g
BLWD C7 4–6 cm	2 cm	20.258	1.5g	2g
BLWD C7 6–13.5 cm	7.5 cm	100.363	1.5g	2g
BLWD C7 14–15.6 cm	1.6 cm	30.265	1.5g	2g
BLWD C7 16.2–19.7 cm	3.5 cm	49.335	1.5g	2g
BLWD C7 20.1–22.2 cm	2.1 cm	29.152	1.5g	2g
BLWD C7 22.4–23.7 cm	1.3 cm	29.542	1.5g	2g
BLWD C7 24.3–27.6 cm	3.3 cm	51.087	1.5g	1g
BLWD C7 27.5–30.2 cm	2.7 cm	36.495	1.5g	2g
BLWD C7 30.2–32.8 cm	2.6 cm	38.976	1.5g	2g
BLWD C7 32.9–35 cm	2.1 cm	38.482	1.5g	2g
BLWD C7 35.2–39.4 cm	4.2 cm	60.493	1.5g	1g
BLWD C7 39.6–42.3 cm	2.7 cm	41.896	1.5g	2g
BLWD C7 43–45 cm	2 cm	33.302	1.5g	2g
BLWD C7 46.5–52 cm	5.5 cm	43.917	1.5g	2g
BLWD C7 52.1–54.2 cm	2.1 cm	31.159	1.5g	2g
BLWD C7 54.5–58.2 cm	3.7 cm	34.818	1.5g	2g
BLWD C7 58.5–61 cm	2.5 cm	39.088	1.5g	2g
BLWD C7 61.5–64 cm	2.5 cm	38.798	1.5g	2g
BLWD C7 64.5–67.7 cm	3.2 cm	38.942	1.5g	2g
BLWD C7 68–72.3 cm	4.3 cm	54.696	1.5g	1g
BLWD C7 72.5–76.9 cm	4.4 cm	69.386	1.5g	2g
BLWD C7 77.1–79 cm	1.9 cm	31.047	1.5g	2g
BLWD C7 79.5–86.9 cm	7.4 cm	90.032	1.5g	2g

2.4. Results

2.4.1. Age model and sediment organic geochemistry

Basal sediment in BLWD-C7 (this study) is carbonate gravel, similar to previous results (Tamalavage *et al.*, 2018; van Hengstum *et al.*, 2016), above which is laminated algal sapropel interbedded with layers of increased calcium carbonate. Sedimentation rate within BLWD-C7 was 0.3 to 0.6 mm yr^{-1}, which is similar to what has been previously documented on the sinkhole periphery (BLWD-C2: ~0.4 mm yr^{-1}). However, the onset of sedimentation was delayed at BLWD-C7 until ~1900 Cal yrs BP relative to BLWD-C2 (onset ~3000 Cal yrs BP) (**Figure 4**). A layer of non-laminated can be observed at 23–25 cm depth in BLWD-C7 (2σ median age range: 298-373 Cal yrs BP), which was previously observed in other cores (BLWD-C2: 10–16 cm median age 371–158 Cal yrs BP, BLWD-C3: 25–31 cm median age 291–223 Cal yrs BP) occurs at a depth of in 23–25 cm in BLWD-C7. Only three of the five samples were kept to create the age model because the other two samples had older materials that had become incorporated into the sediment record, therefore they were removed to provide a more realistic sedimentary profile (**Figure 4**).

Carbon and nitrogen content show variation in concentrations according to the stratigraphic layers along the depth of the core (**Table 1**). The variations in C:N ratios and organic carbon content in the different sediment core layers are due to the distribution of organic matter (*i.e.*, the ratio will be high if there are more leaves and twigs) trapped within the sediment. Particulate organic matter in Blackwood sinkhole comes from different sources, *i.e.*, freshwater or marine phytoplankton and vegetation

(Tamalavage *et al.*, 2018). In general, sedimentary layers with high TOC content (%) also have a corresponding high C:N ratio (**Table 1 and Figure 5**). A high C:N ratio (>20) indicates that the source of carbon fluxed to the sediment was likely dominated by terrestrial organic carbon input, whereas a lower C:N ratio (<20) indicates a higher proportion of carbon from primary productivity (*e.g.*, phytoplankton, (Stenuite *et al.*, 2007)). In Blackwood sinkhole, the C:N ratio fluctuates between the 24 layers of the sediment core, with the lowest ratio at 20.1–22.2 cm (C:N ratio of 14) and the highest at 72.5–76.9 cm (C:N ratio of 39.5) (**Figure 5**). The algal sapropel layer seen at approximately 23–25cm contains 15% carbon and has a C:N ratio of 19.

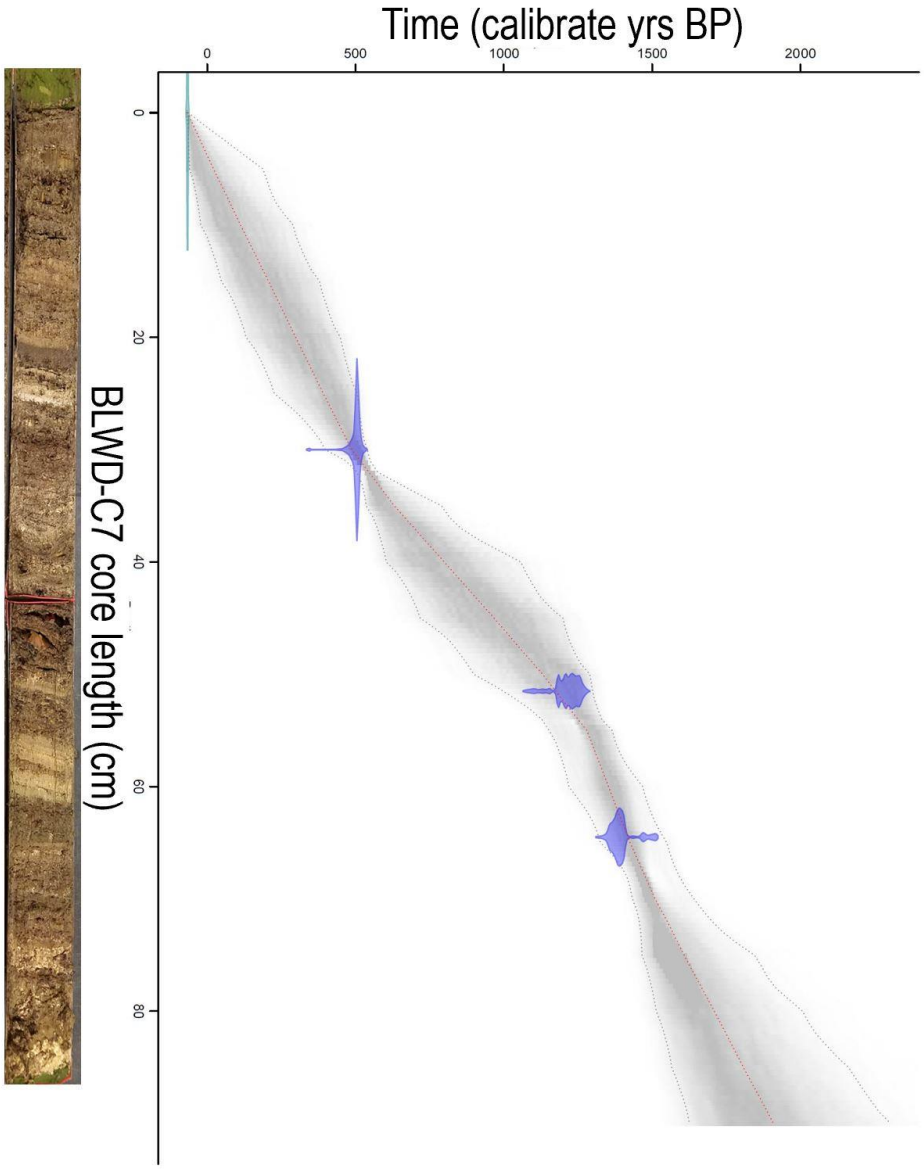

Figure 4: Downcore age model for BLWD-C7 showing core to be up to ~1,900 years old. Radiocarbon results were calibrated into calendar years before present (Cal yrs BP) with IntCAL13 and visualized using Bacon v2.2. The calibrated ^{14}C dates are shown in the transparent blue regions, the grey coloring indicates more likely calendar ages, and the red line shows single 'best' model based on the weighted mean age for each depth.

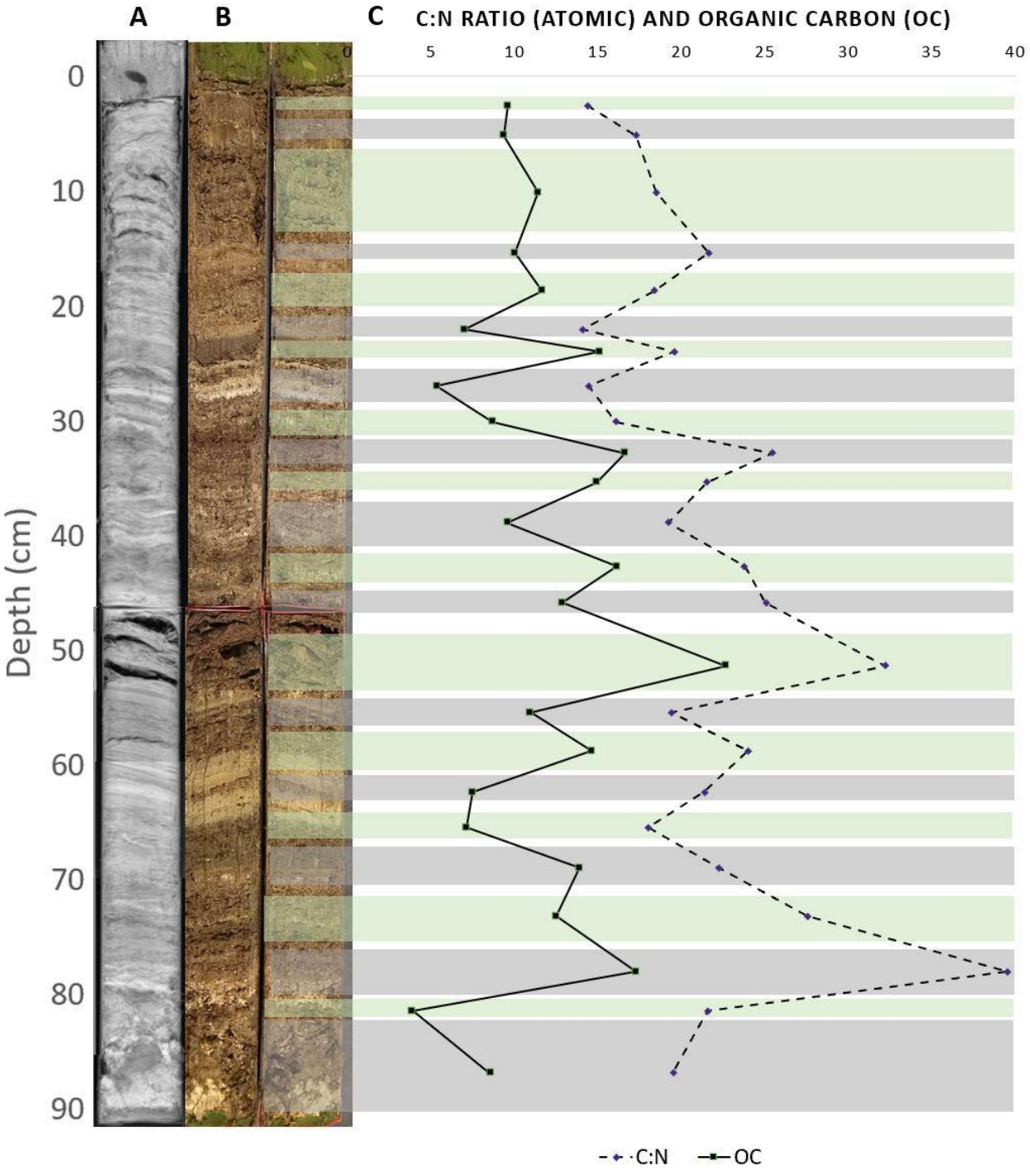

Figure 5: C:N ratio and OC (%) according to core depth. Sediment core **(A)** x-radiograph and **(B)** photograph with clear stratigraphic layers, **(C)** C:N ratio (dashed line), and OC (%) (solid line) in accordance to depth in centimeters. The green and grey bars over the lines indicate the sediment layer interval.

2.4.2. 16S rRNA gene analysis

We sequenced on average 749,985 reads for each of the 24 samples, for a total of 20,669,626 raw reads. After trimming to a maximum length of 450 bp to keep any longer strands (the amplicon product is 254 bp) and deduplication, between 277,000-551,107 reads per sample, or a total of 13,907,092 reads, remained (**Table 3**). After removal of poorly aligned and chimeric sequences, we were left with a total of 4,385,399 ASVs to assign taxonomy to.

Table 3: Reads per sample after each step of the *mothur* analysis.

Sample	Trim and deduplication Read #	Removing poorly aligned sequences Read #	Filtering and omitting redundancy ASV #	Chimera and contaminant removal ASV #
BLWD C7 2–3 cm	494,377	421,180	190,007	136,993
BLWD C7 4–6 cm	401,304	347,866	168,279	118,684
BLWD C7 6–13.5 cm	390,257	332,427	154,288	109,889
BLWD C7 14–15.6 cm	388,637	330,896	155,537	111,279
BLWD C7 16.2–19.7 cm	416,007	356,997	168,725	122,324
BLWD C7 20.1–22.2 cm	388,159	326,509	152,042	113,380
BLWD C7 22.4–23.7 cm	372,830	317,607	160,981	126,463
BLWD C7 24.3–27.6 cm	401,594	321,154	158,787	120,246
BLWD C7 27.5–30.2 cm	534,975	439,929	216,426	167,291
BLWD C7 30.2–32.8 cm	393,735	334,419	154,309	118,188
BLWD C7 32.9–35 cm	361,984	309,549	148,369	110,567
BLWD C7 35.2–39.4 cm	321,766	269,173	121,287	84,371
BLWD C7 39.6–42.3 cm	277,000	247,207	116,018	81,854
BLWD C7 43–45 cm	374,845	295,739	133,760	94,305
BLWD C7 46.5–52 cm	337,855	303,124	142,325	107,701
BLWD C7 52.1–54.2 cm	414,757	359,680	165,007	119,696
BLWD C7 54.5–58.2 cm	551,107	500,340	250,488	202,627
BLWD C7 58.5–61 cm	452,601	402,125	177,635	123,174

Table 3: Continued

Sample	Trim and deduplication Read #	Removing poorly aligned sequences Read #	Filtering and omitting redundancy ASV #	Chimera and contaminant removal ASV #
BLWD C7 61.5–64 cm	434,724	389,704	220,377	175,318
BLWD C7 64.5–67.7 cm	390,283	342,919	165,597	123,841
BLWD C7 68–72.3 cm	385,038	344,842	172,055	133,112
BLWD C7 72.5–76.9 cm	385,584	347,816	179,269	135,198
BLWD C7 77.1–79 cm	467,692	410,833	205,628	157,755
BLWD C7 79.5–86.9cm	469,387	400,612	204,547	155,733
Negative control	244,543	242,032	237,489	235,545

The main taxa found in all the layers belong to the classes Dehalococcoidia (21%), Gammaproteobacteria (12%), Bathyarchaeota (12%), and Campylobacteria (13%), all commonly found in sediment (Wang *et al.*, 2012). A number of Archaea phyla, such as Euryarchaeota (5%), Nanoarchaeota (3%), and Crenarcheaota (13%), were also found in each layer of the anoxic sediment (Huber *et al.*, 2003; Iverson *et al.*, 2012; Kubo *et al.*, 2012). The minor fluctuation in class relative abundance can be seen by changes in the community in samples 6–13.5 cm, 22.4–23.7 cm, 54.5–58.2 cm, and 61.5–64 cm according to depth (**Figure 6**). We ran a negative control to identify possible extraction kit contamination. The negative control made up ~5% of the entire dataset, therefore removing contaminants would not change the overall microbial community composition.

To determine the distribution of the microbial community (beta diversity), the relative abundance of the different classes was analyzed in relation to the carbon content variables (organic carbon content (%), total nitrogen (%), and C:N ratio) using non-

metric multidimensional scaling (NMDS). Organic carbon ($p = 0.752$) and total nitrogen ($p = 0.152$) are not statistically correlated to the relative abundance of the microbial community. The C:N ratio is however statistically correlated to the relative abundance for the microbial community with a $p < 0.05$ ($p = 0.012$) (**Figure 7**).

Diversity includes both taxon richness and evenness and was observed to vary going deeper in the core. Sample BLWD 35.2–39.4 cm displayed the highest diversity (inverse Simpson index value of 495.6) while BLWD 27.5–30.2 cm displayed the lowest diversity (10.9) (**Table 4**). The samples with the top five highest inverse Simpson index values are BLWD 2–3 cm (126.4), 6–13.5 cm (92.6), 35.2–39.4 cm (495.6), 39.6–42.3 cm (326.8), and 43–45 cm (152.5). None of the rarefaction curves reached a plateau, indicating that our sequencing efforts did not recover the complete richness of our samples (**Figure 9**). The distribution of taxa across the samples indicate a similar microbial community within the top ten layers (2–32.8 cm) (**Figure 10**). A few outliers on the heatmap of the top 150 shared by all OTUs in the samples (**Figure 10**) that can be seen are sample BLWD 72.5–76.9 cm, 77.1–79 cm, and 79.5–86.9cm, which could indicate similar microbes towards the lower depths of the sediment core. The microbial community were clustered with distinct age groups (**Figure 8**). The stratigraphically-constrained hierarchical clustering conformed to three groups previously identified; Group 1: 0-1008 Cal yrs BP, Group 2: 1009-1502 Cal yrs BP , and Group 3: 1520-2990 Cal yrs BP (Tamalavage, 2016). The main taxa present in all three groups were Dehalococcoidia and Bathyarchaeota, compared to Gammaproteobacteria mainly present in Group 1 and Group 3, as well as Campylobacteria mainly present in Group 3.

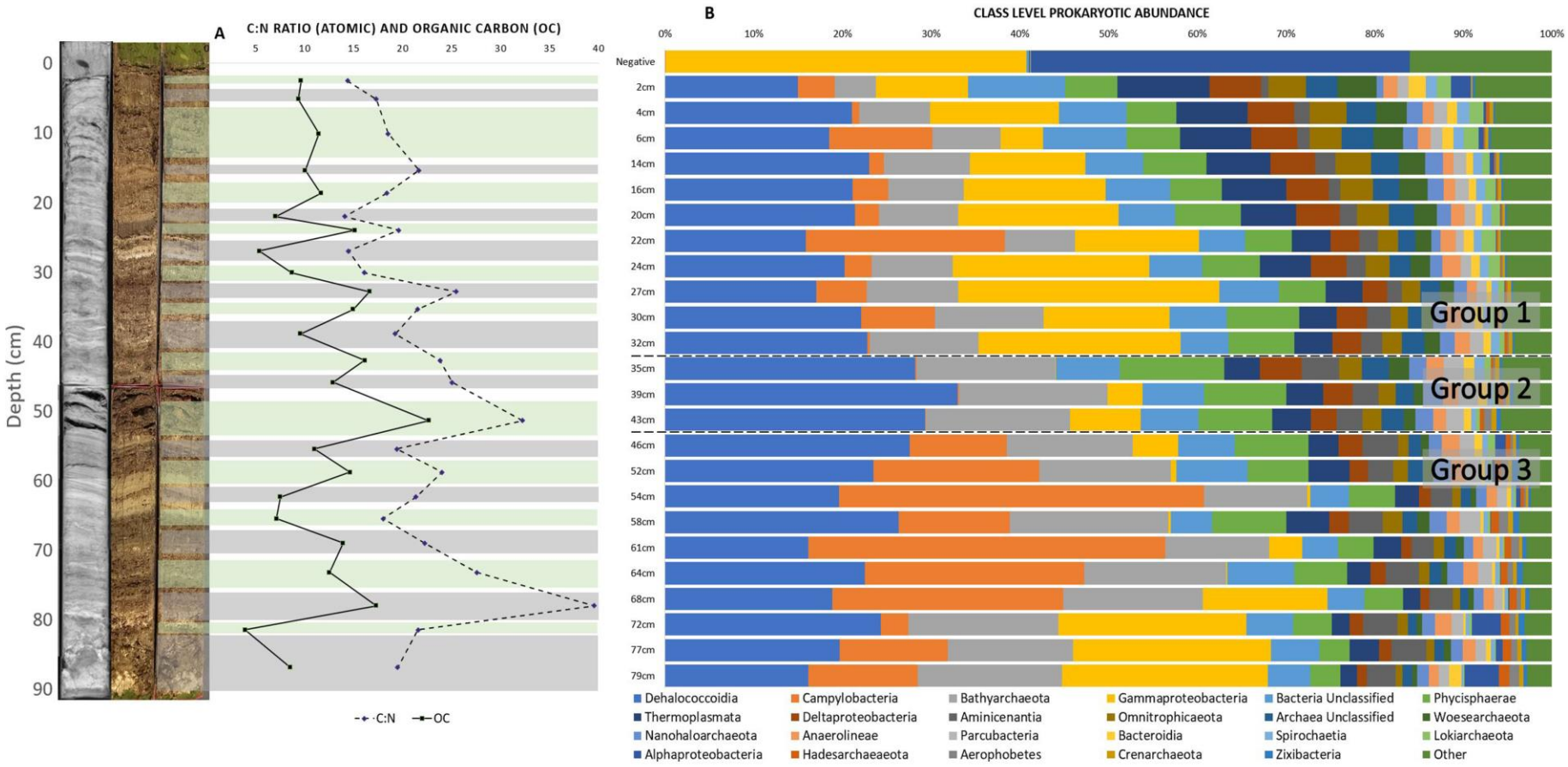

Figure 6: (A) Core image, C:N, and Organic Carbon (OC), and **(B)** class level prokaryotic relative abundance according to depth of sediment core. C:N Ratio (purple dashed line) and OC (black line). Groups 1, 2, and 3 are denoted by the dashed line.

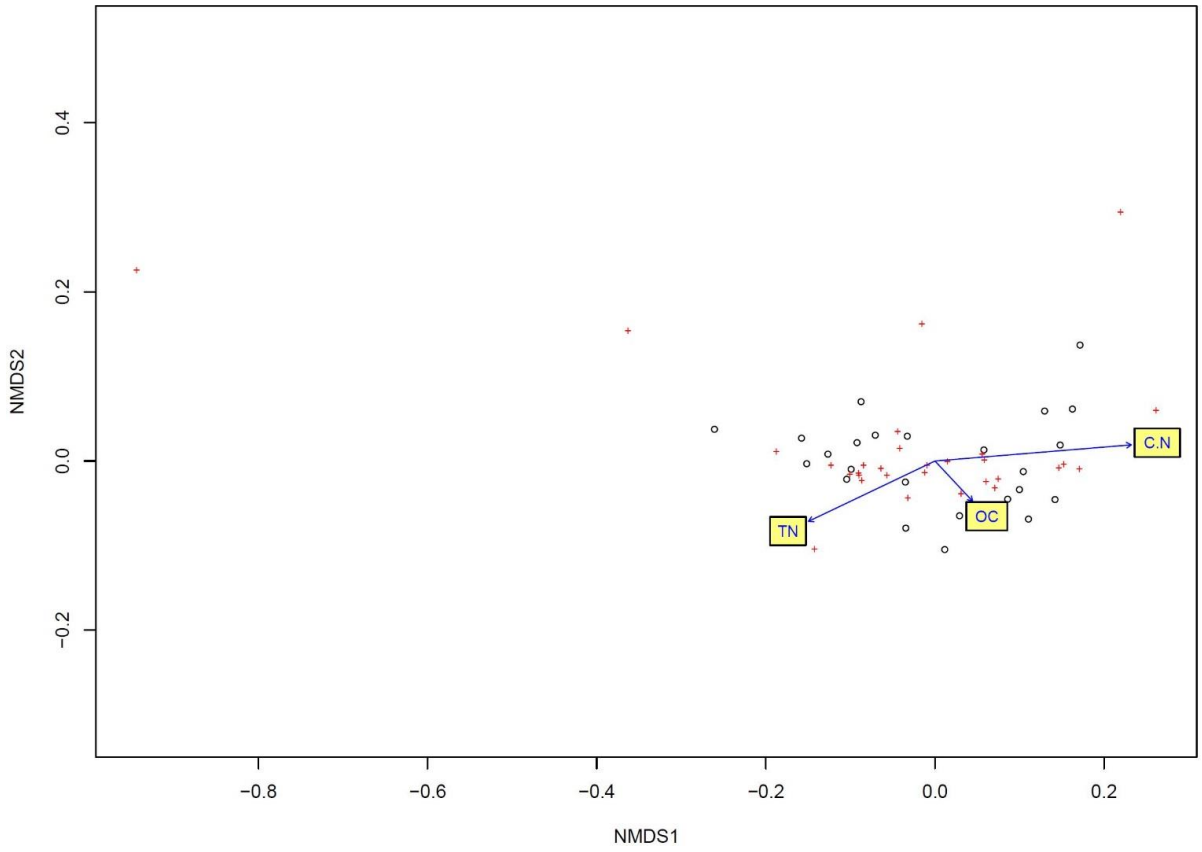

Figure 7: Non-metric multidimensional scaling (NMDS) ordination. Black dots represent sample depths, and red crosses are microbial community relative abundance taxon (class level). Each arrow is correlated to the ordination and represents the direction and strength of the carbon influence (C:N ratio, OC, and TN). C:N ratio is significantly correlated ($p = 0.012$).

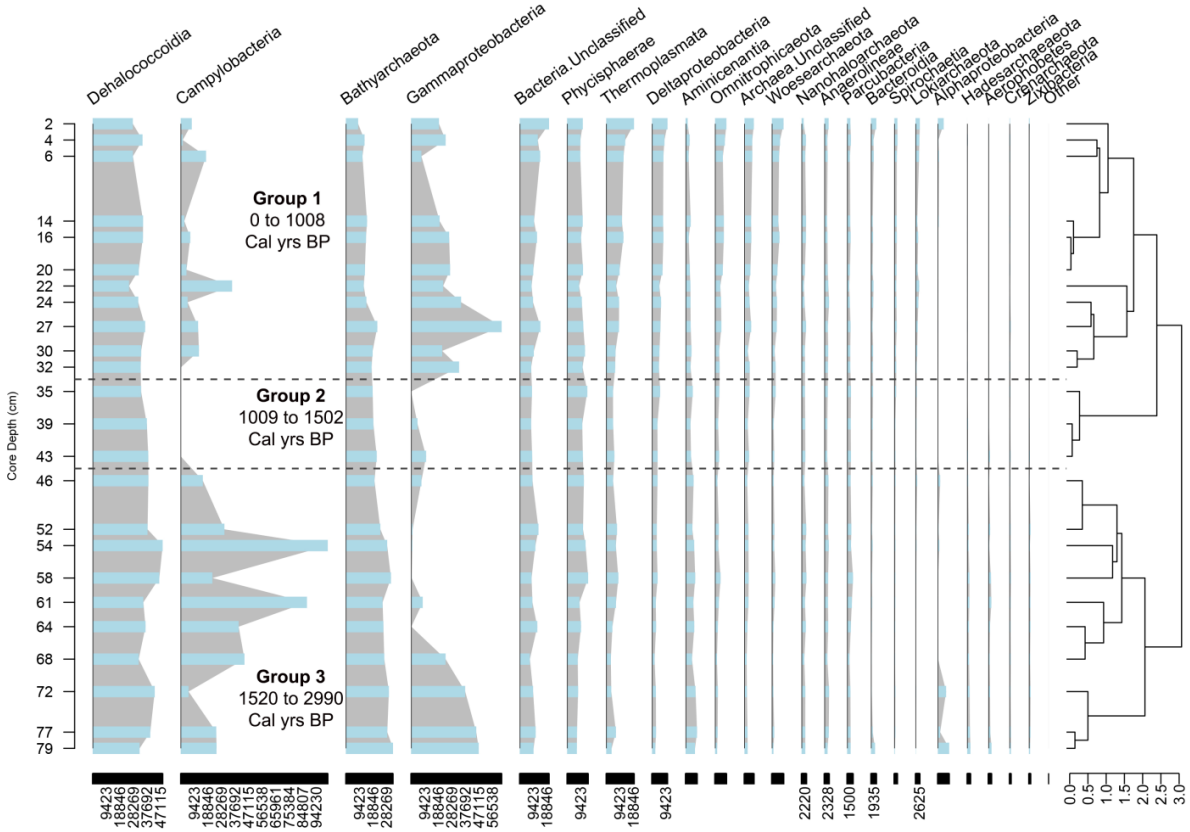

Figure 8: Stratigraphically-constrained hierarchical cluster analysis of microbial community downcore based on previously determined groups of organic matter sources.

Table 4: Number of sequences (N_{seq}), sequence coverage, observed richness (S_{OBS}), and inverse Simpson index for each sediment layer.

Group	N_{seqs}	Coverage	S_{OBS}	Invsimpson	C:N ratio
BLWD 2–3 cm	136,993	60.10%	63,027	126.4	14.35
BLWD 4–6 cm	118,684	61.70%	52,126	38.2	17.26
BLWD 6–13.5 cm	109,889	62.10%	47,754	92.6	18.47
BLWD 14–15.6 cm	111,279	64.10%	45,426	47.9	21.64
BLWD 16.2–19.7 cm	122,324	64.80%	49,091	34.0	18.33
BLWD 20.1–22.2 cm	113,380	66.70%	43,363	28.4	14.05
BLWD 22.4–23.7 cm	126,463	75.30%	35,915	12.9	19.56
BLWD 24.3–27.6 cm	120,246	69.50%	41,905	20.3	14.43
BLWD 27.5–30.2 cm	167,291	73.10%	50,841	10.9	16.04
BLWD 30.2–32.8 cm	118,188	70.40%	39,947	40.5	25.43
BLWD 32.9–35 cm	110,567	68.90%	39,209	13.8	21.48
BLWD 35.2–39.4 cm	84,371	58.20%	39,879	495.6	19.19
BLWD 39.6–42.3 cm	81,854	60.00%	37,013	326.8	23.77
BLWD 43–45 cm	94,305	60.80%	41,483	152.5	25.01
BLWD 46.5–52 cm	107,701	68.30%	38,564	54.9	32.20
BLWD 52.1–54.2 cm	119,696	70.70%	39,976	25.3	19.36
BLWD 54.5–58.2 cm	202,627	79.70%	46,694	12.6	23.98
BLWD 58.5–61 cm	123,174	68.20%	44,616	62.6	21.32
BLWD 61.5–64 cm	175,318	81.70%	36,890	6.3	17.98
BLWD 64.5–67.7 cm	123,841	74.10%	36,533	19.7	22.21
BLWD 68–72.3 cm	133,112	76.90%	34,785	16.0	27.53
BLWD 72.5–76.9 cm	135,198	73.90%	40,232	25.3	39.51
BLWD 77.1–79 cm	157,755	77.10%	41,277	29.0	21.55
BLWD 79.5–86.9cm	155,733	79.60%	36,759	29.0	19.45

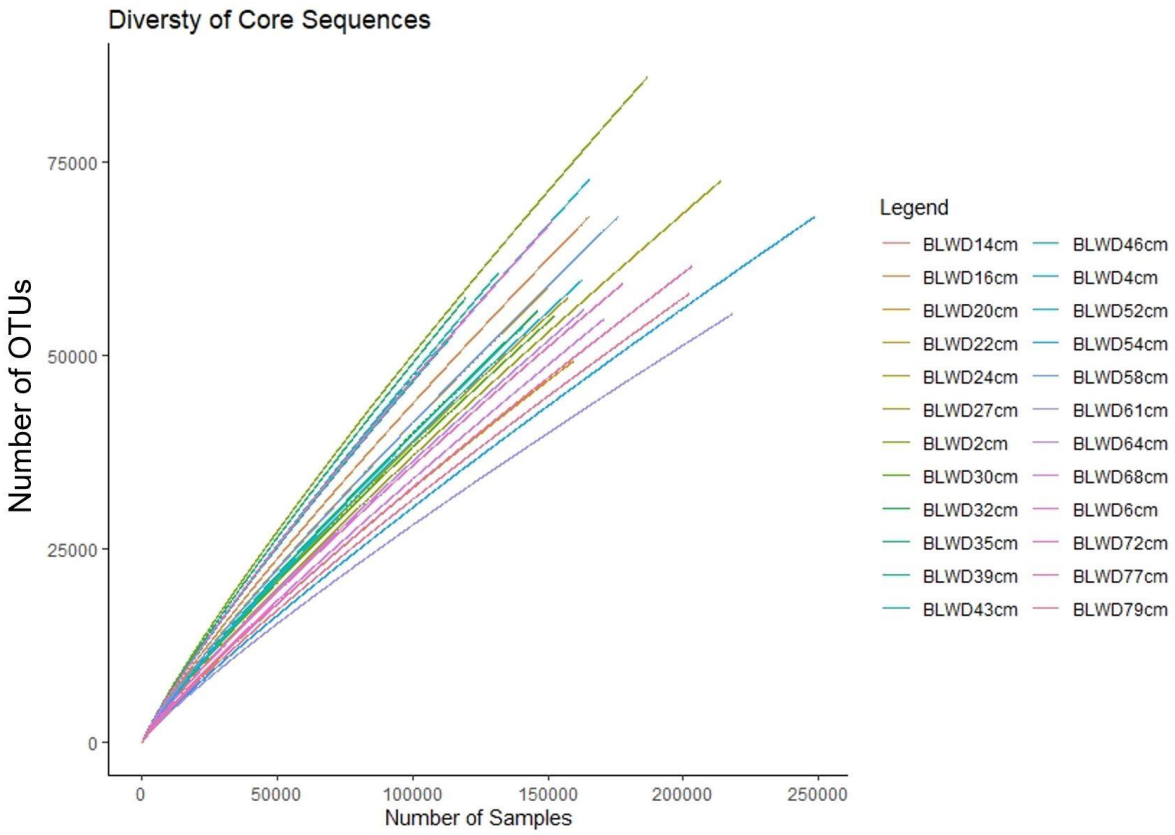

Figure 9: Rarefaction curve to show richness for sediment depth interval samples.

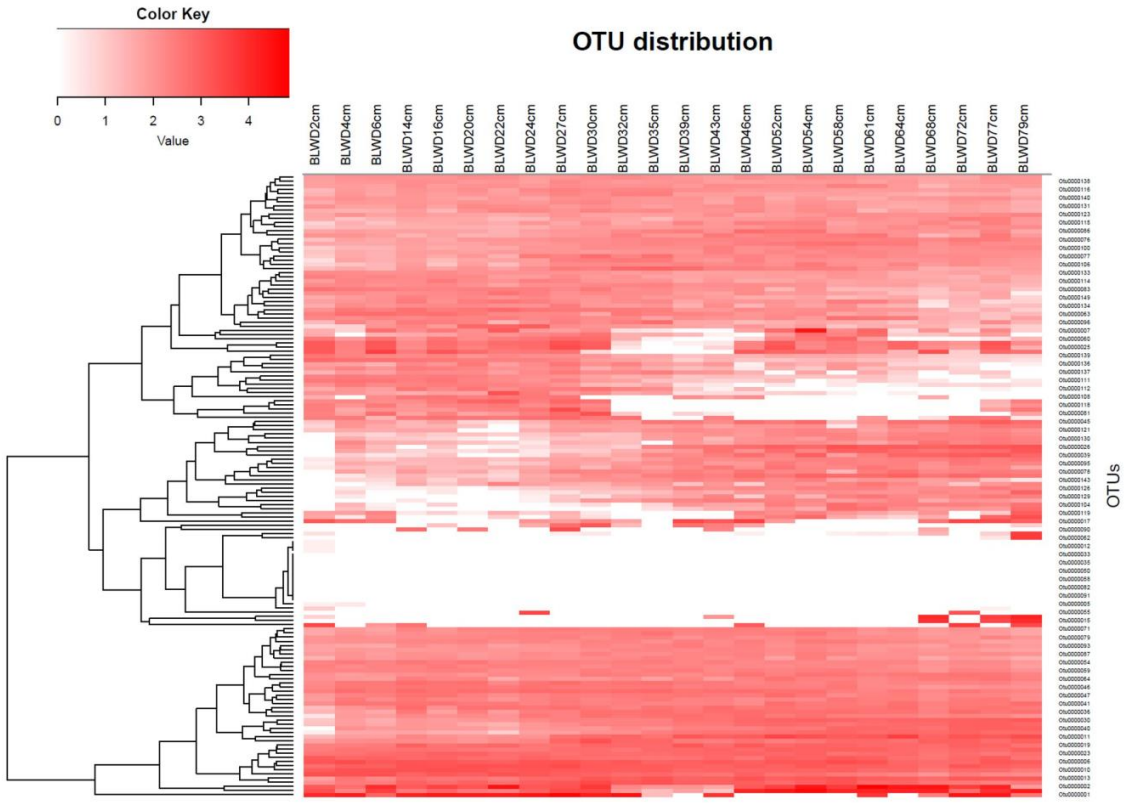

Figure 10: Heatmap of the top 150 shared by all OTU of the sediment core samples. The darker red colors indicate that there were more sequences of the particular OTU and the white means there were none of that OTU.

2.5. Discussion

Sinkholes provide easy-to-access samples, allowing the study of microbial communities buried in sediment. Here, we looked at the influence of the source of organic matter on microbial community composition. In sinkholes, carbon is present in excess, although the lowest TOC was 3.8% at layer 77.1–79 cm, with a corresponding C:N ratio of 21.55. In Blackwood sinkhole, we found that the C:N ratio is statistically correlated to the relative abundance of the microbial community with a $p < 0.05$,

strongly suggesting that the C:N ratio significantly affects the microbial community composition.

Within carbon-rich anoxic sediment, the main electron donor is organic carbon. However, the source of carbon (terrestrial or planktonic) can impact the microbial community. Bacteria seen in sediment with a C:N ratio of <20 had an organic matter source of phytoplankton and bacteria seen in sediment with a C:N ratio>20 had an organic matter source of terrestrial material (Fuentes *et al.*, 2016). The large increase of organic carbon (as shown at 72.5–76.9 cm), as well as the source of particulate organic matter (POM) changes the microbial community.

The stratigraphically-constrained hierarchical cluster analysis showed the microbial community clustered within the three distinct organic matter source groups previously analyzed (Tamalavage, 2016). Group 1 (0-1008 Cal yrs BP) POM input was mainly marine and terrestrial, Group 2 (1009-1502 Cal yrs BP) was identified as mainly freshwater productivity and terrestrial POM input, and Group 3 (1520-2990 Cal yrs BP) was identified as mainly terrestrial POM (Tamalavage, 2016). The source of organic matter for the three groups influenced the taxa present. The main taxa present in all three groups were Dehalococcoidia and Bathyarchaeota, therefore the source of organic matter was both terrestrial and phytoplanktonic. The source of organic matter for Gammaproteobacteria, the main taxa present in Group 1 and Group 3, was marine and terrestrial. The source of organic matter for Campylobacteria, the main taxa present in Group 3, was terrestrial. The Blackwood Sinkhole sedimentation rate of ~0.4–1.2 mm yr^{-1} (~1,600 Cal yrs BP) was caused by organic matter deposition from primary

productivity and erosion (Tamalavage *et al.*, 2018). This sedimentation rate is relatively high compared to sinkholes from other environments, such as in the Tennessee watershed had a sedimentation rate of 36,000 mm yr^{-1} between 1997 and 2003 (Hart *et al.*, 2005) and in the Nesher Quarry on the Mediterranean coast had a sedimentation rate of ~ 61 mm ka^{-1} (ka= 1,000 yrs) from ~ 200 to 78 ka (Frumkin *et al.*, 2015).

According to the 16S rRNA gene data, the most prevalent classes within the entire sediment core are Alphaproteobacteria, Dehalococcoidia, Gammaproteobacteria, Bathyarchaeota, and Campylobacter. Campylobacteria are nitrogen fixers commonly found in the Spartina root system (McClung *et al.*, 1980; Miller *et al.*, 2007). In North America common salt marsh plants and C_4 carbon fixators are Spartina (Lamb *et al.*, 2006). Blackwood Sinkhole is mainly dominated by pine and palm trees (Tamalavage *et al.*, 2018), although as a coastal environment it is inhabited by wetland plants. Alphaproteobacteria is known for a variety of metabolic abilities within anoxic sediment, including nitrogen fixation, ammonia oxidation, and methylotrophy (Acosta-González *et al.*, 2016; Williams *et al.*, 2007). Studies completed in similar anoxic environments have often discovered an abundance of the class Dehalococcoidia, which was also very prevalent in this study (Biderre-Petit *et al.*, 2016; Cetecioğlu *et al.*, 2009). Gammaproteobacteria facilitating carbon fixation, have also been found in high abundance from coastal sediment sites in Western Europe, Germany, and Australia (Dyksma *et al.*, 2016; Lenk *et al.*, 2011). Bathyarchaeota, prevalently found in estuary sediment in North Carolina, were able to break down carbohydrates of terrestrial origin (Lazar *et al.*, 2016). Members of the phylum Chloroflexi are commonly found in anoxic

environments, within the water column (Cetecioğlu *et al.*, 2009). Specifically, members of the class Dehalococcoidia are known to only inhabit environments with strict anoxic conditions, such as river sediments, aquifers, and sludge (Löffler *et al.*, 2013). Due to the coastal anoxic nature of the Blackwood Sinkhole sediment, all the most prevalent classes of microbes present in this system may use alternative forms of nutrients.

It was suggested that sinkholes could be considered analogs to deep-sea sedimentary environments, due to the carbon influence on the residing community. The very high C:N ratio observed in some layers is a clear indication that the organic material found in the sediment is of terrestrial origin (plants and trees) and is very different than what would be found in the deep-sea, which is remote from terrestrial input. Sinkholes also have a much higher sedimentation rate with much higher concentrations of organic material (van Hengstum *et al.*, 2016). Therefore, coastal anoxic sediment, like the Blackwood Sinkhole, microbial communities should be different, compared to deep-sea sediment and cannot be used as analogs.

Overall, the entire core displayed an excess of carbon, allowing for a large number of prokaryotes and a diverse subsurface microbial community. The microbial communities were dominated by members of the Chloroflexi, Proteobacteria, and Epsilonproteobacteria phyla, all phyla commonly found in coastal anoxic sediment. Since organic carbon, the main electron donor, is always in excess, it seems that the limiting factor for microbial growth in the studied core was the availability of electron acceptors, which will be the subject of the next chapter. The C:N ratio significantly correlates to the microbial community, and further investigation should include

alternative nutrient and energy sources. Metagenomic analysis and nutrient data could provide more information as to the major impacts on the microbial communities.

3. GENOMIC POTENTIAL AND VIRAL ACTIVITY ARE DICTATED BY NUTRIENT AVAILABILITY IN SEDIMENT IN AN ANOXIC SINKHOLE

3.1. Introduction

With an estimated 10^{29} microbial cells (mainly prokaryotes), marine sediment encompasses a large portion of the Earth's biomass (Orsi, 2018). Prokaryotic diversity is driven by the diversity of energy sources and electron acceptors (Nealson, 1997; Walsh *et al.*, 2016). Because of redox potential gradients, microbial communities can be vertically stratified and are composed of interdependent layers of chemotrophic and heterotrophic microorganisms (Biddanda *et al.*, 2012). Both Bacteria and Archaea stimulate essential processes within sediment, such as the removal of sulfate and the oxidation of organic matter (Kallmeyer *et al.*, 2012).

Prokaryotic diversity can be greatly impacted by viruses. Viruses are the most abundant life forms on Earth, with estimates of ~5×10^{8} viral particles (VLP) per gram of sediment in the Mediterranean Sea (Corinaldesi *et al.*, 2017), 1×10^{4} to 1×10^{9} VLP/g in the South Pacific Gyre (Engelhardt *et al.*, 2014), and 10^{6} to 10^{7} VLP/g in anoxic subsurface sediment viral particles (Orsi, 2018). Viral abundance is highly dependent on their host abundance, which relies on total organic carbon and energy flux within the sediment (Cai *et al.*, 2019), biomass production (Lachnit *et al.*, 2019), as well as organic matter (Lachnit *et al.*, 2019). Moreover, viral production changes with the conditions of the sediment environment (*e.g.*, oxygen levels and carbon content) (Glud *et al.*, 2004). Studies of abundance, genomic composition, and ecological roles of viruses indicate that

viruses are significant contributors to subseafloor ecosystems. Currently, it is hypothesized that lysogeny is more prevalent under suboptimal conditions for host survival and growth such as the subsurface environments (Williamson *et al.*, 2008). Viruses can also be a driving factor of nutrients recycling through mortality via lysis of their prokaryotic hosts. In sediment, viral abundances greatly depend on the concentration and the age of the organic matter (Jorgensen *et al.*, 2007). Viruses can also be used as a carbon source. Indeed, virus decomposition rates, which are higher in deep-sea sediment than in coastal sediment, are controlled by the extracellular enzymatic activities that break down the viral capsid proteins to provide an important source of labile organic compounds (Dell'Anno *et al.*, 2015). Therefore, viruses play an important role in biogeochemical cycles (Dell'Anno *et al.*, 2015).

Metagenomic analysis provides a snapshot of the metabolic potential and biochemical functions to provide insight into the workings of the entire community. Major microbial groups such as Aquificae, Chloroflexi, Bacteroidetes, Verrucomicrobia, and Crenarchaeota, were found to encode denitrification genes (*nosZ*) within a metagenomes from Ushuaia Bay (Tierra del Fuego Island, Argentina) (Calderoli *et al.*, 2018). Nitrogen concentrations influenced the communities, specifically nitrogen fixation genes (*nif*) were found within Alphaproteobacteria (Reese *et al.*, 2018). Sulfate reduction (*dsr*) genes, found in Desulfobacterales and Desulfovibrionales families, have been used as functional markers instead of 16S rRNA gene analysis (Leloup *et al.*, 2007).

The **main goal of this study** was to determine the effect of nutrient availability on the genomic potential and activity of the prokaryotic and viral communities within the Blackwood Sinkhole. We conducted metagenomics on the sediment samples to assign genomic potential, measured the nutrient content of the sediment to identify possible electron acceptors, and calculated the viral production rate, microbial mortality rate, and virus-to-prokaryote ratio to identify the impact of viruses on the microbial community. I hypothesized that metabolic genes of microbial communities would be driven by nutrient availability and redox potential and that denitrification (*nosZ*), sulfur oxidation (*sox*), and sulfate reduction (*dsr*) will be the main pathways. Viral activity would be low when there is a decrease in nutrients within anoxic sediment.

3.2. Methodology

3.2.1. Sample collection and handling

A push core (BLWD-C7, 26.79°N, 77.42°W) was sampled on 07/29/2018 using advanced scuba techniques. The periphery of the sinkhole bottom (not center) was targeted for core sampling (**Figure 3**) in an attempt to re-collect the last 3000 years of sedimentary deposition, as has been previously documented (van Hengstum *et al.*, 2016). In the lab, the core was sectioned lengthwise, photographed and x-radiographed, and subsequently stored at 4˚C until further analysis. Twenty-four stratigraphic layers of varying widths were identified based on coloring and texture. The 24 layers were extracted and stored at -80˚C for further DNA extraction (see Chapter 2). Of the 24 layers, only six samples (6–13.5 cm, 24.3–27.6 cm, 35.2–39.4 cm, 68–72.3 cm, 72.5–76.9 cm, and 79.5–86.9 cm) had enough material for nutrient chemistry analysis,

prokaryotic metagenomics, viral metagenomics, and virus-induced microbial mortality experiments.

3.2.2. Porewater nutrient geochemistry

Porewater was extracted from the six core samples using a centrifuge to spin down the sediment for 15 min at full speed (3220xg) and the supernatant (~10 ml) was pipetted out for analysis. Nutrient (silicate, phosphate, ammonium, nitrate/nitrite) analysis of pore water was determined at the University of Texas at Austin – Marine Science Institute. N as ammonia was measured by using the indicator indophenol blue, which, when measured at 630 nm, was proportional to the original ammonia concentration (Liao, 2002). P as orthophosphate was measured using an antimony-phospho-molybdate complex. This complex was reduced to an intensely blue-colored complex by ascorbic acid. The color produced by the reduction of the complex (using ascorbic acid) was then proportional to the phosphate concentration in the sample (Liao, 2002). SiO_2 in silicate was measured using the absorbance of a yellow silicamolybdate complex. This complex was subsequently reduced with stannous chloride to form a heteropoly blue complex which has an absorbance maximum at 820 nm, which was proportional to the concentration of "molybdate reactive" silica (Liao, 2002). To measure nitrate/nitrite the sample was passed through a column containing granulated copper-cadmium to reduce nitrate to nitrite. N as nitrite was determined by diazotizing with sulfanilamide and coupling with N-(1-napthyl)-ethylenediamine dihydrochloride to form a highly colored azo dye which was measured calorimetrically (Liao, 2002).

3.2.3. Prokaryotic and viral DNA extraction and metagenomic sequencing

Total DNA was extracted using the DNeasy PowerMax Soil kit (Qiagen, USA) from 10 g of sediment (**Table 5**), then quantified using the Qubit dsDNA HS Assay kit (Invitrogen, MA) and stored at -20˚C until further used. A blank control was also extracted to identify possible contamination from the ambient lab and kit reagents, although due to low yield was not sequenced.

Prior to viral DNA extraction, a sediment slurry was made from 50 g sediment and 50 mL autoclaved virus-free seawater (Table 5). Viruses were extracted from the sediment slurry by adding tetrasodium pyrophosphate to a final concentration of 5 mM, followed by three one-minute ultrasound treatments at 47 KHz in a Branson 2800 Ultrasonicator. Samples were centrifuged for 10 min at 800xg, the supernatant was collected, and 4 mL of virus-free seawater were added (Corinaldesi *et al.*, 2017). Bacteria and larger particles were sequentially filtered through a GF/D (2.7 µm), 0.45 µm, 0.22 µm and 0.1 µm filters. Viruses were concentrated using a Pellicon XL 50 Ultrafiltration Cassette (tangential flow filter) (Millipore, USA). Further concentration was completed using the Amicon Ultra Centrifugal Filters (Millipore, USA) with centrifugation at 2087xg for ~5 min). Extracellular DNA was degraded using 2 U/mL DNase I (Invitrogen) and incubated for 15 min at room temperature. DNase I was inactivated with the addition of EDTA to a final concentration 0.02 mM to the reaction mixture followed by heating for 10 min at 65°C (Corinaldesi *et al.*, 2017). Proteins and non-nucleic acid components were broken down using proteinase K (50 ug/mL final concentration) and SDS (0.5% final concentration) and the viral DNA was extracted

using phenol-chloroform followed by ethanol precipitation (Sambrook *et al.*, 1989). The extracted DNA was purified one more time using the QiaAmp DNA purification kit (Qiagen, USA) to remove residual phenol and chloroform before quantification using Qubit dsDNA HS Assay kit (Invitrogen, MA). To get enough DNA for metagenomic sequencing, viral DNA was amplified using the Qiagen REPLI-g WGA kit (Qiagen, USA) following manufacturer's instructions. The amplified DNA was not furthered cleaned, as per the manufacturer's instructions. The methodology used for extraction was designed to specifically look at DNA viruses.

Metagenomic samples were sent for sequencing using Illumina NovaSeq 6000 chemistry with 150 bp paired-end at the Texas A&M AgriLife Bioinformatics and Genomics facility (**Table 6**). The library was prepped using the Bioo DNA library preparation kit. Due to low yields, sequencing was not completed for the negative controls.

Table 5: Lamination thickness, sediment sample initial weight (g), DNA extractions weight taken (g) (PowerMax kit and viral).

Sample	Lamination Thickness	Initial Weight (g)	Weight (g) taken for PowerMax kit	Weight (g) taken for viral extraction
BLWD C7 6–13.5 cm	7.5 cm	100.363	10g	50g
BLWD C7 24.3–27.6 cm	3.3 cm	51.087	6g	44g
BLWD C7 35.2–39.4 cm	4.2 cm	60.493	10g	50g
BLWD C7 68–72.3 cm	4.3 cm	54.696	8g	45g
BLWD C7 72.5–76.9 cm	4.4 cm	69.386	10g	50g
BLWD C7 79.5–86.9 cm	7.4 cm	90.032	10g	50g

Table 6: Total concentration (µg/µL), volume (µL) of sample sent, and total DNA sent (µg) for metagenomic sequencing. PSM= Power Soil Max kit for prokaryotic samples. Viral= Amplified viral samples.

Sample	Total DNA Concentration (µg/µL)	Volume (µL)	Total DNA sent (µg)
PSM BLWD C7 6–13.5 cm	0.1151	90	10.35
PSM BLWD C7 24.3–27.6 cm	0.085	90	7.65
PSM BLWD C7 35.2–39.4 cm	0.1272	90	11.44
PSM BLWD C7 68–72.3 cm	0.0789	90	7.10
PSM BLWD C7 72.5–76.9 cm	0.0263	90	2.36
PSM BLWD C7 79.5–86.9 cm	0.0957	90	8.55
Viral BLWD C7 24.3–27.6 cm	0.025	54	1.35
Viral BLWD C7 68–72.3 cm	0.6	54	32.40

3.2.4. Metagenomic analysis

After screening for decontamination with FastQC (Andrews, 2010), matching reads were paired with BBMerge from BBTools (Bushnell *et al.*, 2017). Adapters were trimmed using the Illumina Truseq and Nextera adapter sequences reference file and filtering of low-quality regions (quality-trim to Q10 using the Phred algorithm) with BBDuk from BBTools (Bushnell *et al.*, 2017). *De novo* assembly was performed with MEGAHIT using --12 (comma-separated list of interleaved fastq paired-end files) and -r (comma-separated list of fastq) single-end files and no advanced parameters (Tamura *et al.*, 2011). Gene prediction was completed using the -meta parameter in Prodigal (Hyatt *et al.*, 2010). For functional and taxonomic assignment, the proteins were compared to the NCBI non redundant (nr) database for the prokaryotes and the vContact2 database

(ProkaryoticViralRefSeq94-Merged) (Jang *et al.*, 2019) for the viruses using the BLASTp algorithm within DIAMOND (Buchfink *et al.*, 2014). Results were viewed and analyzed in MEGAN (Huson *et al.*, 2016) (**Figure 11**).

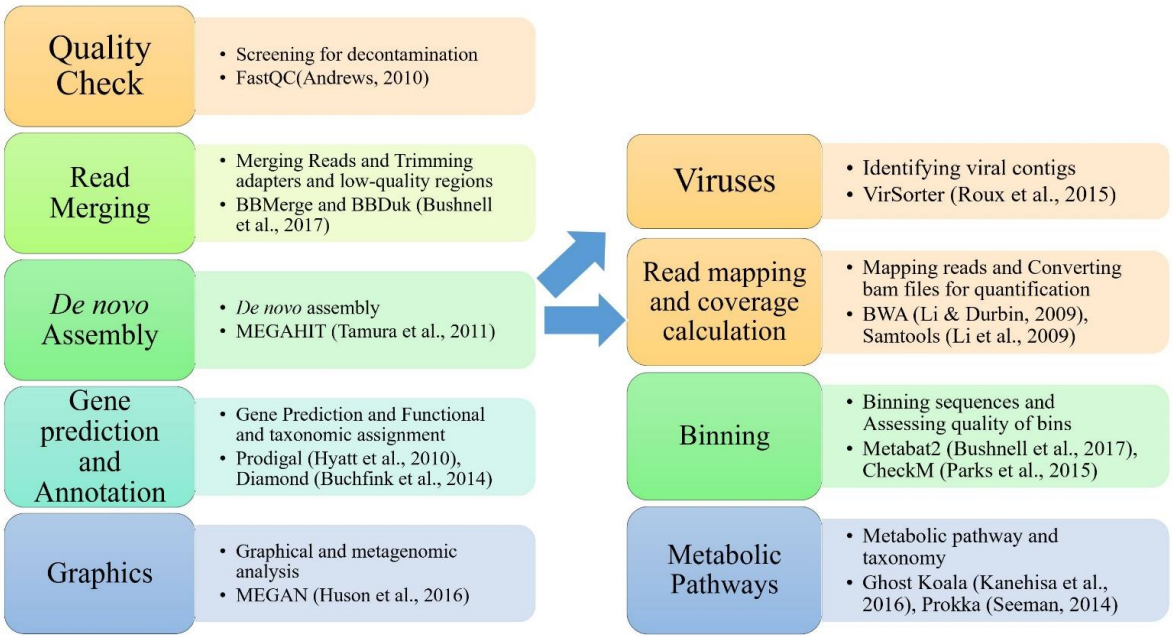

Figure 11: Metagenomic analysis bioinformatics pipeline. Quality checking and assembly was performed using FastQC, BBMerge, BBDuk, and MEGAHIT. Prediction of the genes, functionality, taxonomy, and graphical analysis was completed using Prodigal, DIAMOND, and MEGAN. Identification of viral contigs was performed using VirSorter. Metagenome assembled genomes and metabolic genes were discovered using BWA, Samtools, Metabat2, CheckM, Prokka, and Ghost Koala.

To identify viral contigs within the prokaryotic metagenomes, all contigs were ran through VirSorter (Roux *et al.*, 2015) using both the RefSeq and Virome databases. The resulting predicted viral sequences (at category 1 and category 2, which are the "most confident" and "likely" predictions (Roux *et al.*, 2015) were run through Prodigal and then DIAMOND (using the NCBI nr database) for annotations.

3.2.5. Metabolic pathways and metagenome assembled metagenome reconstruction

Binning of the contigs was performed to identify near-complete genomes and determine the metabolic pathways of the main microbial members of the community. Coverage was calculated using BWA (Li & Durbin, 2009) with the BWA-MEM algorithm to map the reads and Samtools (Li, Handsaker, *et al.*, 2009) was used to convert the read coordinates into coverage depth values. To complete binning of sequences using Metabat2 (Kang *et al.*, 2019), a depth file assessing base coverage depth from BAM file was created and then used to create bins. The quality of the bins were assessed using CheckM (Parks *et al.*, 2015) and bins with less than 5% contamination and greater than 90% completion (the percentage of present genes derived from isolates, single cells, and metagenomes using lineage-specific marker genes) were kept for further analysis. For the selected bins, a KEGG metabolic pathway was reconstructed with Ghost Koala (Kanehisa *et al.*, 2016) using protein sequences as input.

The contig files for the binned genomes, or metagenome assembled genomes (MAGs), were annotated with Prokka (Seemann, 2014) using the metagenome option to improve gene predictions for highly fragmented genomes. The 16S rRNA gene region was extracted using Geneious (Kearse *et al.*, 2012) and a search was conducted for approximate identification in RDP-Classifier (Wang *et al.*, 2007). The genes *nif, sox, dsr, nosZ, napA, nirS, sat/cysC, sqr, hzo, nrfA, nir, nasA, cnorB* were manually found in the annotated MAGs using Geneious.

3.2.6. Virus-induced microbial mortality experiments

Triplicate sediment slurries were made by mixing 0.5 g of sediment core sample and 4.5 mL of virus-free water (final volume of 5 mL). Samples were diluted with 21 mL of virus-free seawater and incubated at 25˚C (Danovaro *et al.*, 2010). Samples of 1 mL were taken in triplicate from each sediment replicate after 0(fresh), 1, 3, 6, 12, and 24 hours of incubation (Corinaldesi *et al.*, 2010). The samples were fixed with formaldehyde to a final concentration of 37% and stored at -80ºC until quantification (Patel *et al.*, 2007). For quantification, samples were sonicated 3x1 min to release viral particles, then sequentially filtered through a 0.2 µm and 0.02 µm Anodisc filters to retain bacteria and viruses, respectively, before staining with SYBR Green for epifluorescence microscopy (Patel *et al.*, 2007). At least 20 fields were photographed for each filter and processed using ImageJ using 10-30 (viruses) and 80-200 (bacteria) pixel particle size (Abràmoff *et al.*, 2004). The equation to measure the number of particles was adapted from (Patel *et al.*, 2007): RSF × (100/n) /V, where RSF is the grid-reticle scaling factor, n is the number of counted photos, and V is the volume of sample filtered. The microbial mortality rate was calculated as the decline (slope) of the prokaryotic particles divided by the incubation time. Viral production was calculated using the equation: $VP=(V_{max} - V_{min})/(t_{max} - t_{min})$ (Luef *et al.*, 2009), where VP stands for viral production, V_{max} for the maximum viral abundance, V_{min} for the minimum viral abundance, and t for the number of hours of incubation. To identify statistically significant relationships between VP and the nutrient data, a Spearman's rank correlation analysis and corresponding t-test was completed using R.

3.3. Results

3.3.1. Nutrient geochemistry

Nutrient chemical analysis (SiO_2 as Silicate, P as Orthophosphate, and N as ammonia) was completed for samples 6–13.5 cm, 24.3–27.6 cm, 35.2–39.4 cm, 68–72.3 cm, 72.5–76.9 cm, and 79.5–86.9 cm (**Figure 12**). The concentrations varied down core, with a similar trend for each tested nutrient, i.e., if one nutrient was present in high concentration, all measured nutrients were in high concentration. The highest concentration of silicate (1.0247 ug/mL) was at 6–13.5 cm. The highest concentration of phosphate (0.2796 ug/mL) was at 35.2–39.4 cm. The highest concentrations of N as ammonia were observed at 6–13.5 cm (2.278 ug/mL) and 79.5–86.9 cm (2.2051 ug/mL). The lowest concentration of silicate, phosphate and N as ammonia were all observed at 68–72.3 cm (0.0183, 0.0316, and 0.0418 ug/mL respectively). The concentration of nitrate and nitrite in all samples was below standard limit of detection of 0.025 mg.

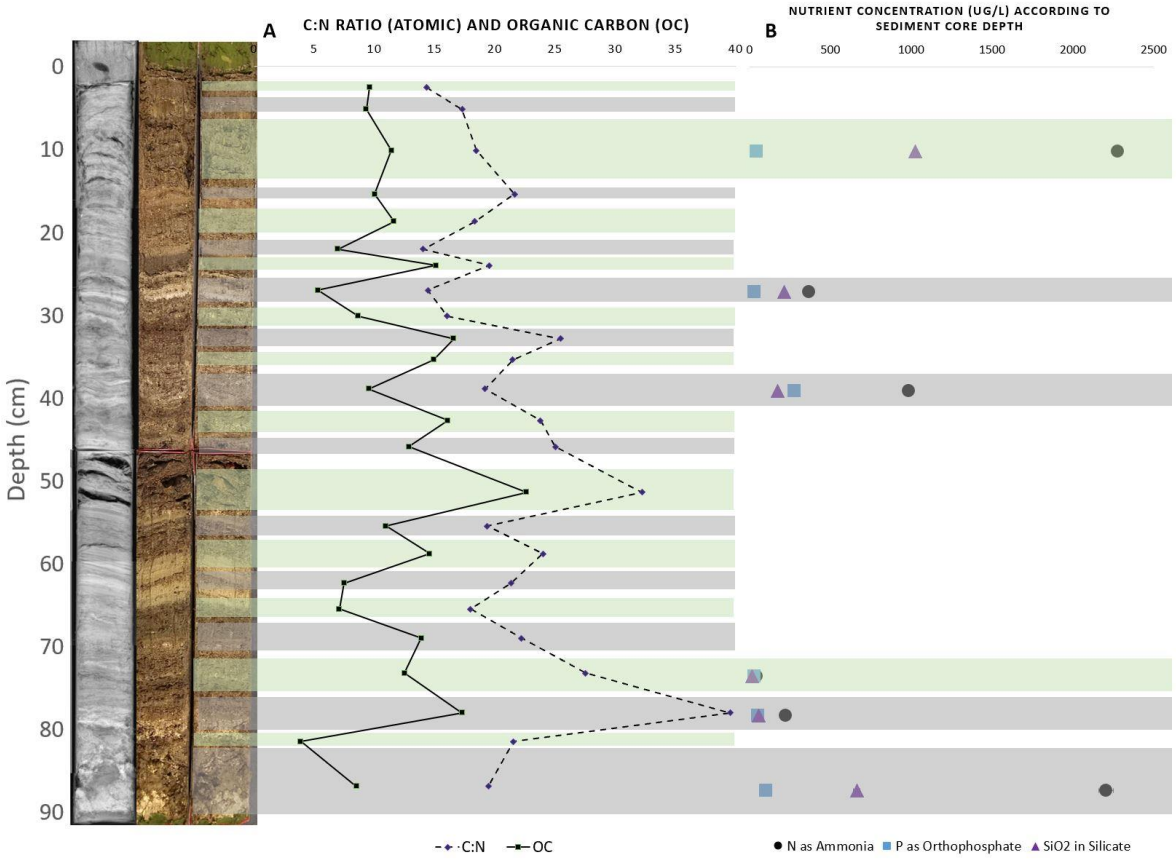

Figure 12: The carbon to nitrogen ratio and percentage of organic carbon in 24 samples of sediment compared to the corresponding nutrient concentrations at six different sediment depths. **(A)** C:N, and Organic Carbon (OC) and **(B)** Nutrients (Ammonia, Phosphate, Silicate) according to depth of sediment core. C:N Ratio (blue dashed line), OC (black line), N as Ammonia (green circles dashed line), P as Orthophosphate (blue squares), SiO_2 as Silicate (purple triangles dashed line).

3.3.2. Taxonomic assignment and metabolic potential

Metagenomic analysis (**Table 7**) revealed that the phyla dominating the samples are Chloroflexi, Proteobacteria, and Planctomycetes (**Figure 13**), which were the same dominating phyla in the 16S rRNA gene analysis (**Figure 6**). There was a shift in the

phyla distribution going deeper in the core, from 6–13.5 cm and 24.3–27.6 cm, to 72.5–76.9 cm and 79.5–86.9 cm (**Figure 14**). For example, the *Anaerolineaceae* family is prevalent in all samples except at 72.5–76.9 cm, when *Rhodobacteraceae* and *Desulfuromonadaceae* increases. The *Desulfobacteraceae* family also decreased down the core, although increased again at 79.5–86.9 cm. Archaea phyla (*e.g.*, Crenarchaeota and Bathyarchaeota) are more prevalent in the deeper layers of the core (68–72.3 cm, 72.5–76.9 cm and 79.5–86.9 cm). Metagenomic taxonomic assignment indicated the 72.5–76.9 cm layer was slightly different, which could be due to the drastic increase in carbon, as described in section 2.4.1 (**Figure 14**). Due to the negative controls not being sequenced we are unable to report possible contamination.

A SEED analysis (annotation of genomic data) was completed on the metagenomic data to determine functionality. Proteins involved in nitrogen, phosphorus and sulfur metabolisms were found in a higher mean abundance in the shallower layers (6–13.5 cm and 24.3–27.6 cm) of the core compared to the deeper layers (**Figure 15**). Carbohydrate degradation (432,131 proteins), iron acquisition (18,673 proteins), aromatic compound degradation (62,347 proteins), nitrogen cycling (39,891 proteins), photosynthesis (1,398 proteins), polyamines (615 proteins), phosphorus (40,609 proteins), and sulfur (71,432 reads) metabolisms were found in all samples (**Figure 15**).

Table 7: Metagenomic data counts for all prokaryotic samples after each step. '_ambig', short for ambiguous, were the unmerged ambiguous reads then assembled together using MEGAHIT.

Sample	# reads BBmerge	#reads BBduk (adapters and contaminants)	# reads BBduk (trimming)	total bp	#contigs	Avg length of contigs	N50
6–13.5 cm	4673572	4673572	4673555	1253201127	2026860	618 bp	641 bp
6–13.5 cm_ambig	100373348	100208776	100208640				
24.3–27.6 cm	4286891	4286884	4286884	1182309635	1756573	673 bp	708 bp
24.3–27.6 cm_ambig	87934402	87768280	87768184				
35.2–39.4 cm	3385845	3385838	3385838	1229820647	1684007	730 bp	805 bp
35.2–39.4 cm_ambig	85331604	85163526	85163412				
68–72.3 cm	3748499	3748494	3748494	1041001015	1351916	770 bp	861 bp
68–72.3 cm_ambig	80870786	80745112	80744990				
72.5–76.9 cm	5793742	5793725	5793725	784043103	1023073	766 bp	858 bp
72.5–76.9 cm_ambig	79653130	79308474	79308370				
79.5–86.9 cm	4305154	4305124	4305124	936290271	1201610	779 bp	876 bp
79.5–86.9 cm_ambig	78085858	77954968	77954880				

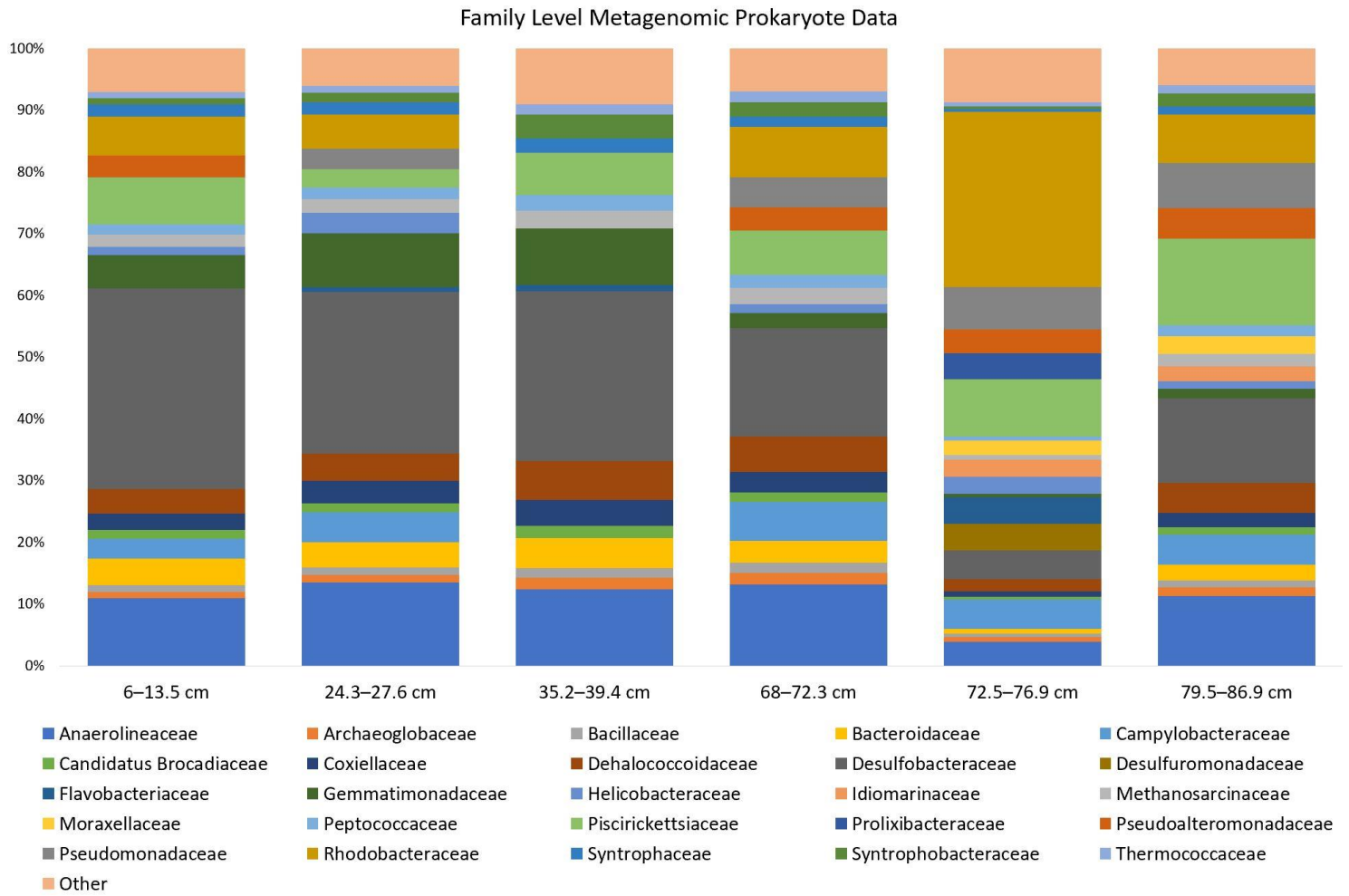

Figure 13: Comparison of six prokaryotic metagenomic samples at the family level. "Other" consist of the families that are present at < 0.97%.

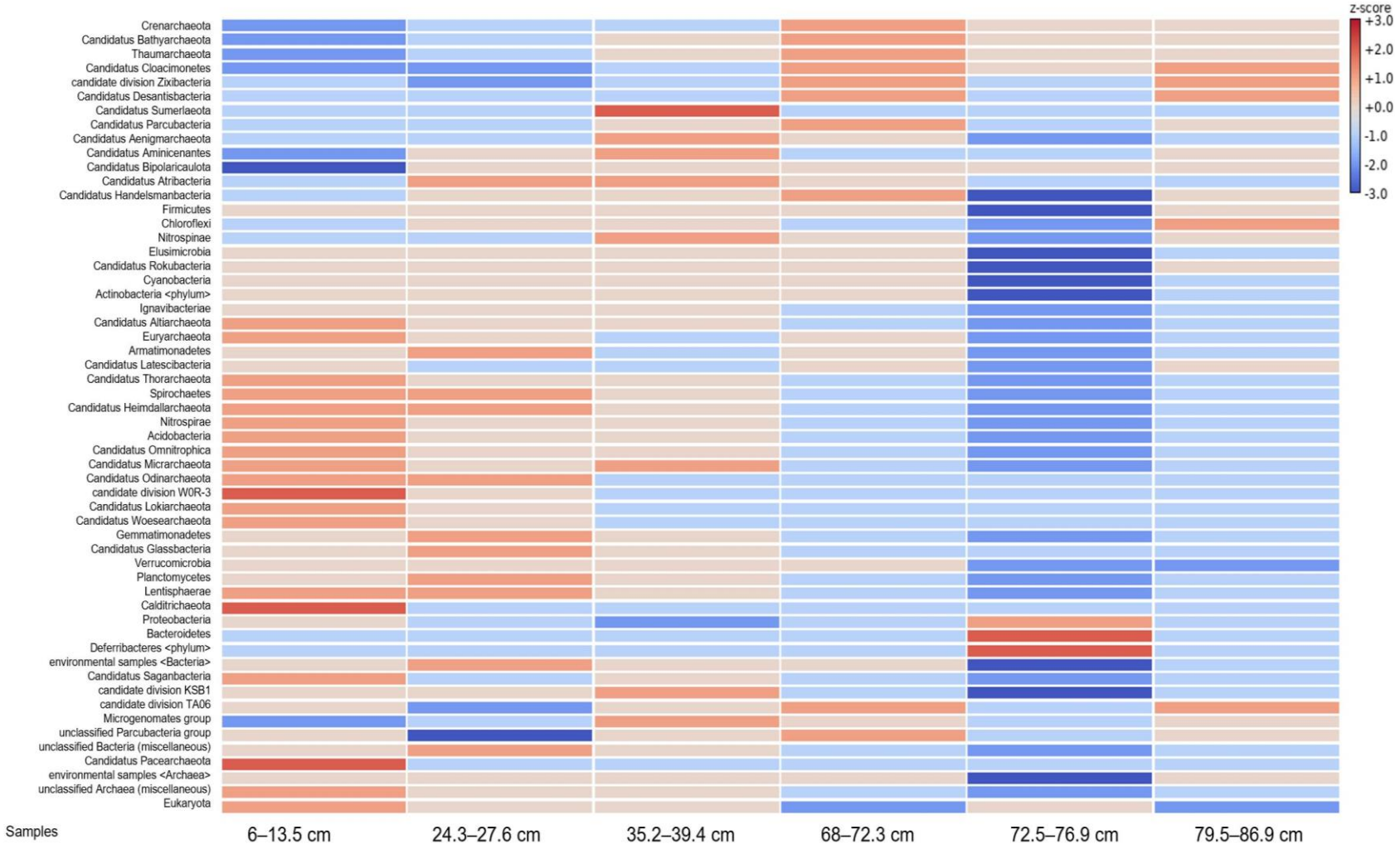

Figure 14: Phylum level heatmap of six prokaryotic metagenomic samples based on Z-score (number of standard deviations from the mean).

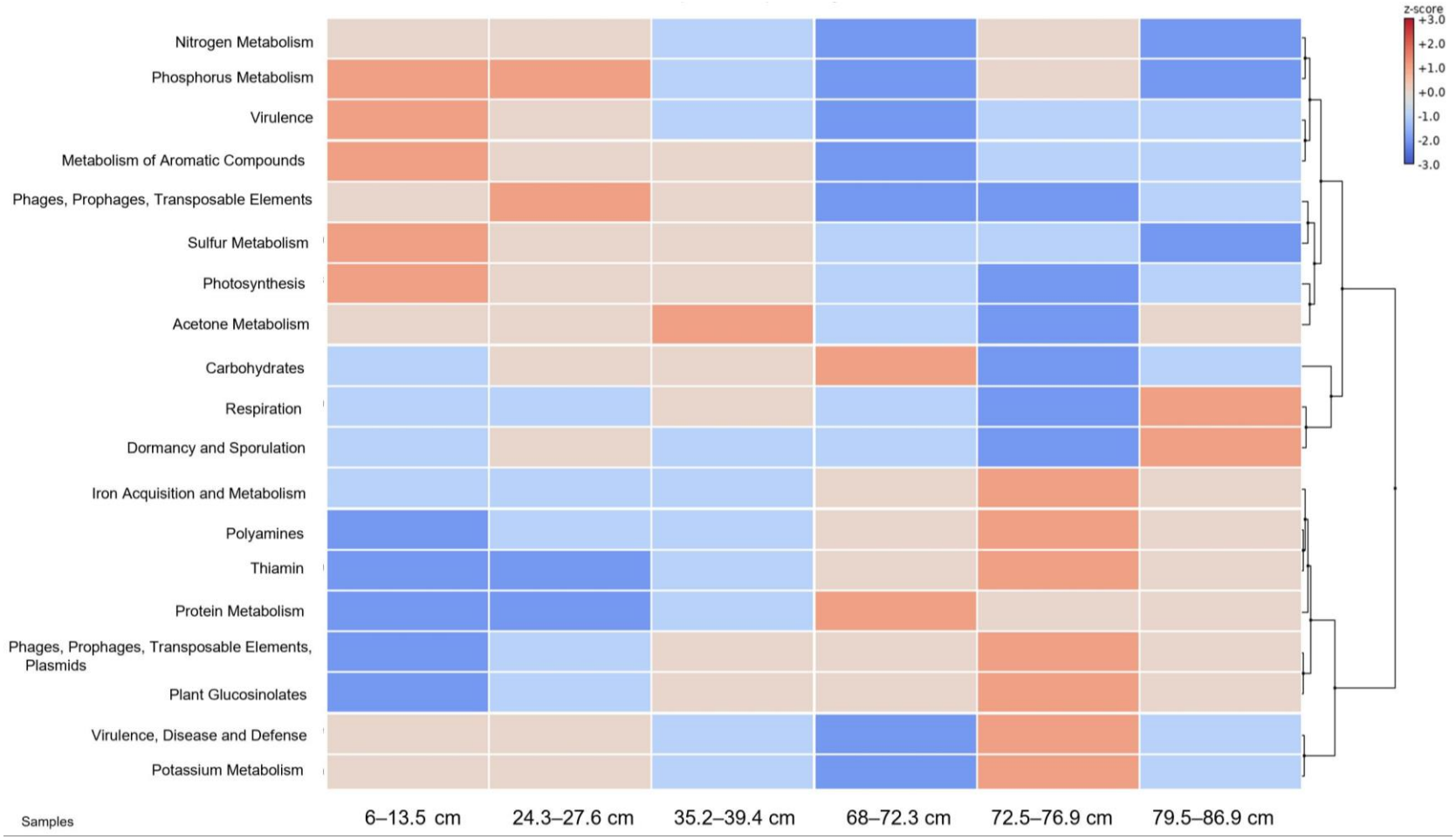

Figure 15: Distribution of proteins in SEED metabolic pathways of interest (carbohydrates, iron acquisition, aromatic compounds, nitrogen, phosphorus, photosynthesis, polyamines, and sulfur).

3.3.3. Binning and identification of nearly-complete genomes

A total of 116 bins were generated (**Appendix A**), with 23 used for further analysis as they were less than 5% contaminated and had greater than 90% genome completeness (**Appendix A**). Out of 23 bins, seven had a 16S rRNA gene. The bins, from now on called metagenome assembled genomes (MAGs), were searched for identifying geochemical pathway (**Table 8**). The strain heterogeneity indicates the proportion of contamination that appears to be from the same or similar strains. The bins were labeled with the corresponding depth *(e.g*. 68 cm) and bin number as denoted by CheckM. Based on the 16S rRNA sequence, 68bin7 (from sample 68–72.3 cm), 79bin17 (from sample 79.5–86.9 cm), and 72bin19 (from sample 72.5–76.9 cm) were closely related to *Desulfurobacteriales* (**Table 9**). 72bin14 and 72bin28 (from sample 72.5–76.9 cm) are identified as *Methanopyrales*. 72bin1 (from sample 72.5–76.9 cm) is identified as *Kallotenuales*. 72bin7 (from sample 72.5–76.9 cm) is identified as *Caldilineale*. The sulfur oxidizing (*sox*) gene was found in MAGs 79bin17, 68bin7, 72bin1, 72bin19, 72bin14, and 72bin28 (**Table 9**). The *dsr* (dissimilatory sulfite reductase) gene was found in MAGs 68bin7, 72bin7, 72bin19, and 72bin28. The *nif* (nitrogen fixation) gene was found in samples 68bin7, 72bin1, and 72bin14, which are members of the Gammaproteobacteria, Rhodobacteraceae, and Rhodobacteraceae class, respectively. The *nosZ* (nitrous oxide reductase) gene utilized by denitrifying bacteria was found in MAGs 79bin17, 68bin7, and 72bin19.

Parts of the carbon fixation pathway found in MAGs 68bin7, 79bin17, 72bin1, 72bin7, 72bin14, 72bin19, and 72bin28 included the Crassulacean acid metabolism,

reductive citrate cycle, and the reductive acetyl-CoA pathway. The nitrogen metabolism found included nitrogen fixation, dissimilatory nitrate reduction, denitrification, and nitrification (**Figure 17**). The sulfur metabolism found included assimilatory sulfate reduction, dissimilatory sulfate reduction, and thiosulfate oxidation by *sox* complex (**Figure 16**). The assimilatory nitrate reduction metabolism was found in 79bin17, 72bin1, 72bin7, 72bin14, 72bin19, and 72bin28 (**Figure 17**).

Table 8: CheckM output to assess the quality of the metagenomes. Marker lineage identifies the possible taxonomic level (*e.g.*, C for class and F for family). Strain heterogeneity indicates the proportion of the contamination that appears to be from the same or similar strains.

Bin ID	Marker Lineage	# genomes	# markers	# marker sets	Completeness	Contamination	Strain heterogeneity
68bin7	C_Gammaproteobacteria	228	583	244	99.18	1.09	20.00
72bin1	F__Rhodobacteraceae	84	568	330	97.73	1.36	40.00
72bin7	C_Gammaproteobacteria	263	506	232	99.52	1.08	0.00
72bin14	F_Rhodobacteraceae	67	616	330	90.71	2.94	39.13
72bin19	C__Gammaproteobacteria	228	583	244	97.53	1.72	50.00
72bin28	C__Gammaproteobacteria	119	544	284	95.51	4.14	41.18
79bin17	C_Gammaproteobacteria	228	583	244	97.92	1.09	20.00

Table 9: Classification of the MAGs based on the 16S rRNA gene comparison using the RDP-Classifier, confidence interval percentage, and genes present

Bin/Sample	Identification Order level	Confidence (%)	Genes Present
79bin17	*Desulfurobacteriales*	4%	*soxABDG, nosZ, dsrE, dsrF, napA, nirS*
68bin7	*Desulfurobacteriales*	4%	*nifH, soxABDG, nosZ, dsrE, dsrF, napA, nirS*
72bin1	*Kallotenuales*	12%	*nifS, soxABD, sat/cysC*
72bin7	*Caldilineales*	9%	*dsrE, napA*
72bin19	*Desulfurobacteriales*	4%	*nosZ, soxABDG, dsrE, dsrF, napA*
72bin14	*Methanopyrales*	17%	*nifS, soxABD, napA, narG, sat/cycC*
72bin28	*Methanopyrales*	20%	*dsrE2, soxA*

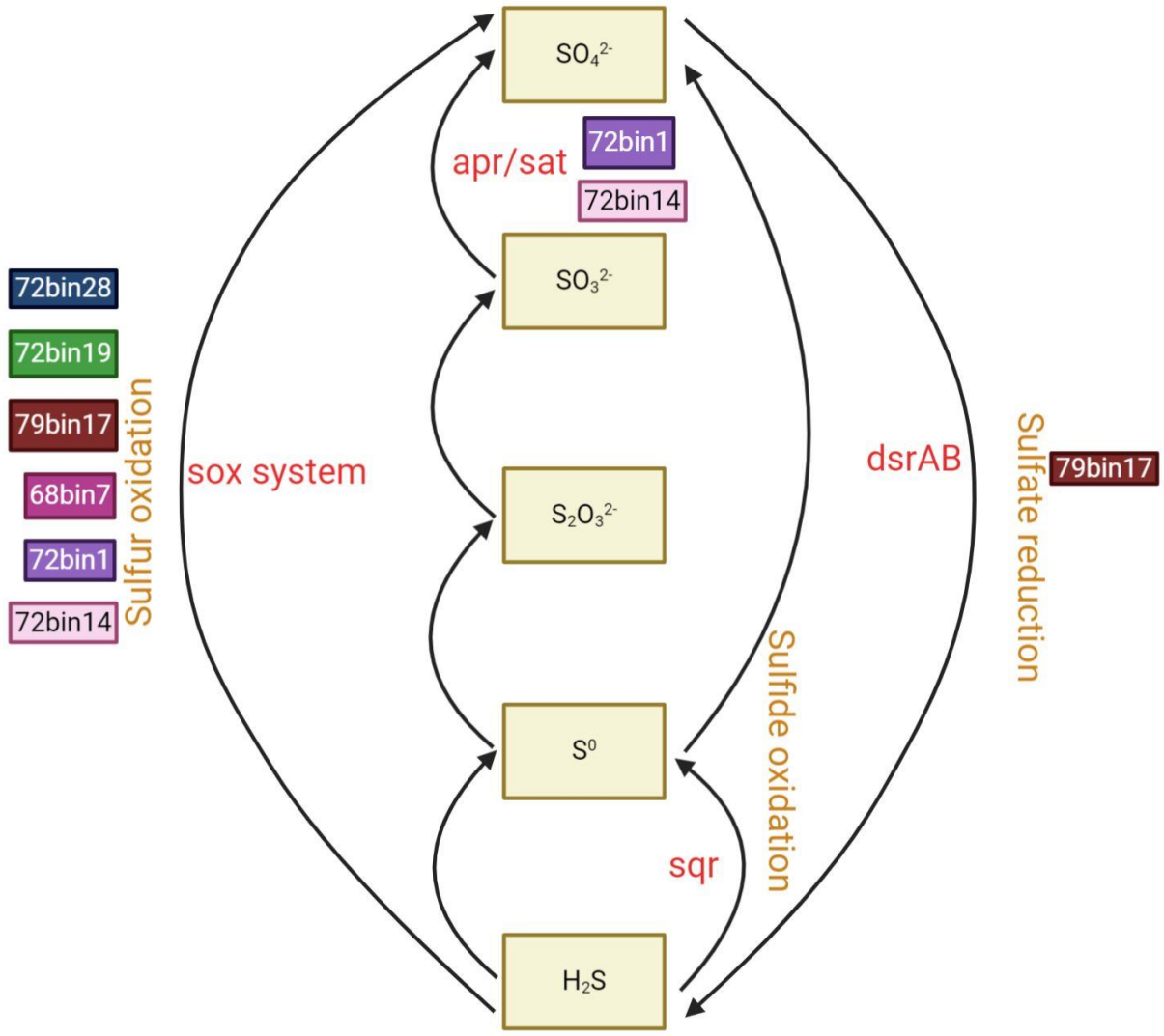

Figure 16: Major microbial sulfur cycle metabolisms and associated genes (indicated in red): sulfate reduction transforms sulfate to hydrogen sulfide (dsrAB), sulfide oxidation transforms sulfide to a more oxidized state between elemental sulfur and sulfate (sqr), and sulfur transforms these compounds to both H_2S and SO_4^{2-}. MAGs capable of these functions are labeled. Adapted from (Ehrlich *et al.*, 2015).

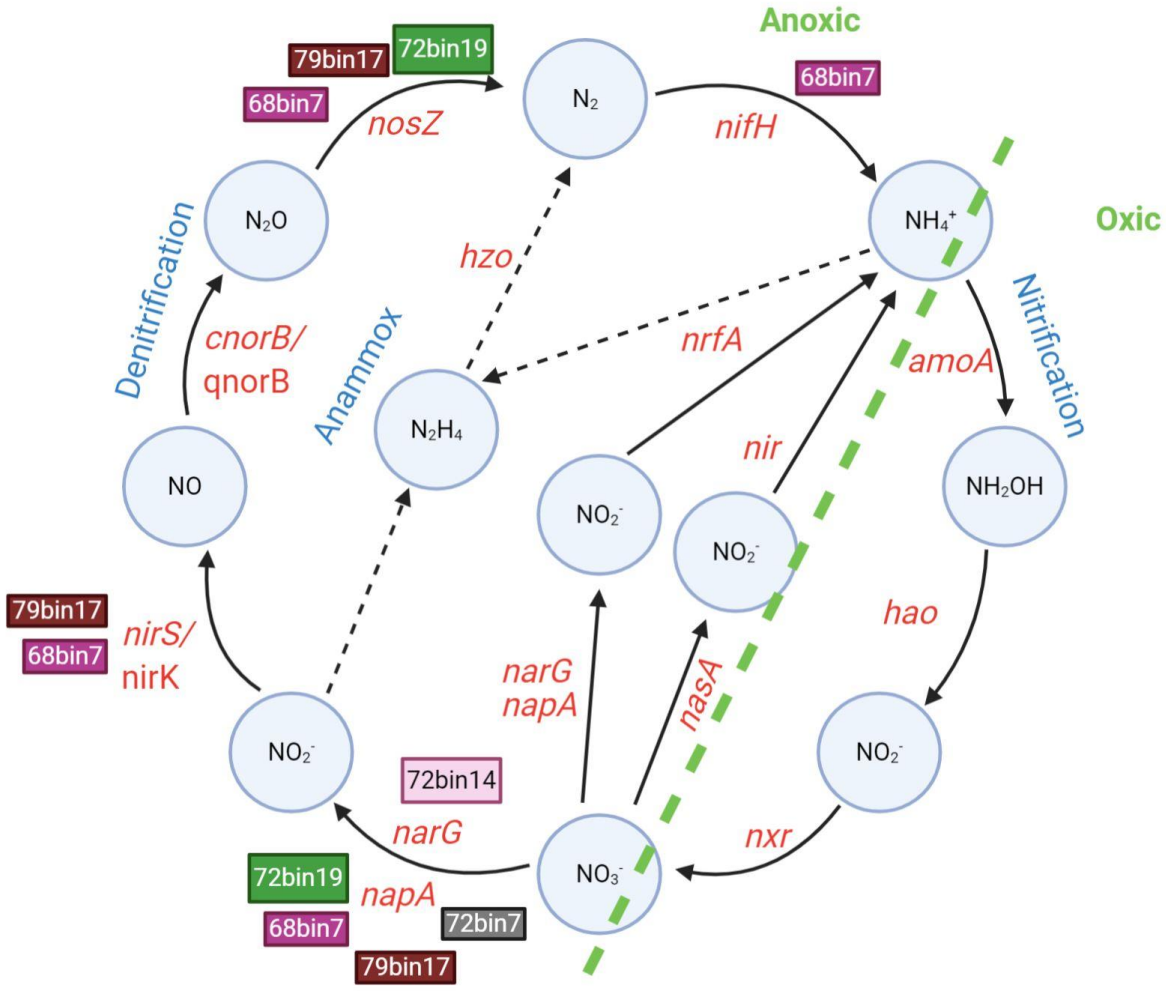

Figure 17: Microbial facilitated nitrogen cycle with associated genes (indicated in red): Anoxic and oxic pathways are indicated with the green dashed line. Black dashed arrows show steps of the anaerobic nitrogen cycle specifically associated with the anammox pathway. MAGs capable of these functions are labeled. Adapted from (Smith *et al.*, 2015).

3.3.4. Viral metagenomics

We sequenced the virome of two samples, 24.3–27.6 cm and 68–72.3 cm. Viral community composition in the sinkhole sediment samples show both Bacteria- and Archaea-associated viruses. The main families found at 24.3–27.6 cm were *Myoviridae*

(48%), *Bicaudaviridae* (Archaeal virus) (2%), and unclassified dsDNA Archaeal viruses (4%). The main families found at 68–72.3 cm were *Siphoviridae* (38%), *Podoviridae* (10%), and *Herelleviridae* (3%), (**Figure 18**). After the steps to merge, remove adapters and contaminants, and trimming the reads, the viral metagenomic samples (24.3–27.6 cm and 68–72.3 cm) had a rage of length 200-7,575 bp and 200-52,946 bp (**Table 10**). With a maximum length of 7 kb and 52 kb the samples may contain complete viral genomes. The number of reads removed after BBmerge and BBduk were 15 for 24.3–27.6 cm and 119,146 24.3–27.6 cm_ambig. The number of reads removed after BBmerge and BBduk were 23 for 24.3–27.6 cm and 3,751,6396 24.3–27.6 cm_ambig. The number of viral contigs assembled were 728,526 (24.3–27.6 cm) and 526,570 (68–72.3 cm). The average length of the 24.3–27.6 cm sequences (N50) was 578 bp and of the 68–72.3 cm sequences was 733 bp.

We mined the prokaryotic metagenomes for the presence of viral sequences. Viral contigs obtained using VirSorter had a range of length 104-698 bp. Multiple viral "hallmark" genes were found; major capsid protein, tail protein, spike protein, protease, and chaperone. A total of 641 lytic viral genes using both databases (virome and refseq) were found in all metagenomic samples (**Table 11**). *Citrobacter* and *Pseudomonas* sp. prophage viral sequences were found in samples at 72.5–76.9 cm and 6–13.5 cm (**Table 11**). An increase in the number of viral contigs within the prokaryotic metagenomes did not align with increased viral production rates (**Table 12**) (**Figure 19**).

Table 10: Metagenomic data read counts for viral samples (24.3–27.6 cm and 68–72.3 cm) after each step. '_ambig', short for ambiguous, were the unmerged ambiguous reads then assembled together using MEGAHIT.

Sample	# reads BBmerge	#reads BBduk (adapters and contaminants)	# reads BBduk (trimming)	total bp	#contigs	Avg length of contigs	N50
24.3–27.6 cm	14794535	14794520	14794520	396965184 bp	728526	545 bp	578 bp
24.3–27.6 cm_ambig	48775580	48656434	48656324				
68–72.3 cm	12745399	12745376	12745376	351152453 bp	526570	667 bp	733 bp
68–72.3 cm_ambig	50261772	49991240	12745376				

Table 11: VirSorter number of genes found and sequences of interest, using both databases (virome and refseq).

Sample	Total viral genes (virome and refseq)	Prophages	Viral sequences of interest
6–13.5 cm	1,250	*Citrobacter* and *Pseudomonas* sp. prophage	Major capsid protein, tail protein
24.3–27.6 cm	290		Major capsid protein, tail protein
35.2–39.4 cm	2,405		Major capsid protein, tail protein, spike protein
68–72.3 cm	1,868		Major capsid protein, tail protein, spike protein, protease
72.5–76.9 cm	9,392	*Citrobacter* and *Pseudomonas* sp. prophage	Major capsid protein, tail protein, spike protein, protease
79.5–86.9 cm	4,760		Major capsid protein, tail protein, protease, chaperone

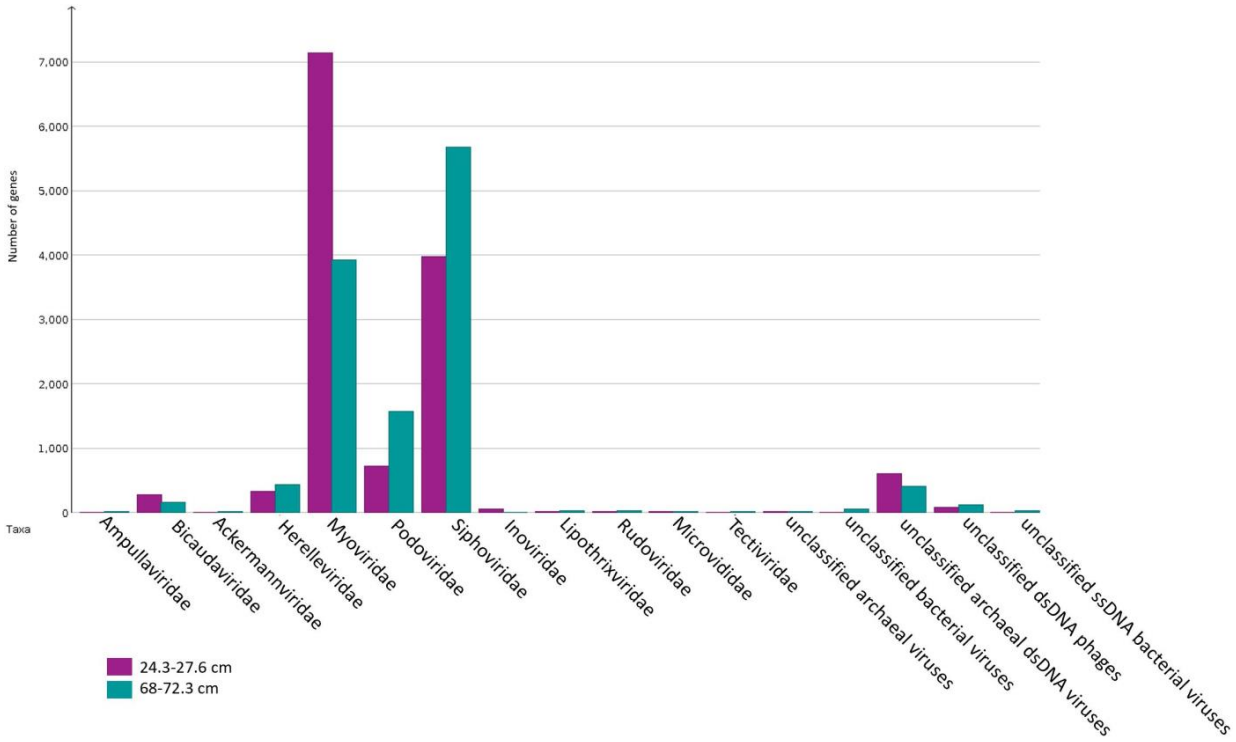

Figure 18: Viral Metagenomic samples compared at the Family level. The purple is sample 24.3–27.6 cm and the teal bar is sample 68–72.3 cm.

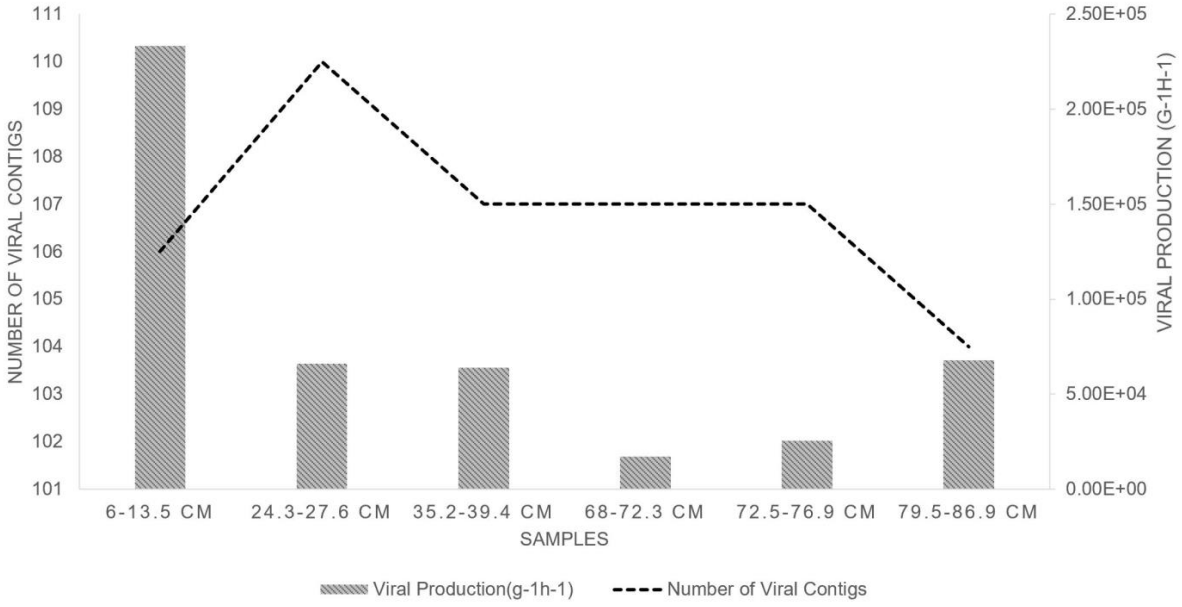

Figure 19: Total number of viral contigs (dashed line) found in VirSorter compared to viral production ($g^{-1}h^{-1}$) (**Table 12**) (gray columns).

3.3.5. Viral activity measurements

As a proxy for microbial activity, we measured viral production and virus-induced microbial mortality. The virus induced microbial mortality rate was different in each layer (**Figure 20**). The rate of viral production was higher in samples 24.3–27.6 cm (6.60×10^4 $g^{-1}h^{-1}$), 35.2–39.4 cm (6.41×10^4 $g^{-1}h^{-1}$), and 79.5–86.9 cm (6.79×10^4 $g^{-1}h^{-1}$) (**Table 12**), which corresponds to low concentrations of phosphate (**Figure 20A**) at 0.0285 ug/mL, 0.2796 ug/mL, and 0.1017 ug/mL, respectively, and high concentrations of nitrogen as ammonia at 0.3669 ug/mL, 0.9856 ug/mL, and 2.2051 ug/mL, respectively. When the rate of viral production increased, so did the N as ammonia, suggesting ammonia was a factor influencing microbial activity. The rate of viral production was the lowest at 68–72.3 cm (1.72×10^4 $g^{-1}h^{-1}$). The viral induced microbial

mortality rate started out high at the surface sample 6–13.5 cm (-4.83 x 10^3 $g^{-1}h^{-1}$), dropped and then increased again at deeper depths 72.5–76.9 cm (-5.25 x 10^4 $g^{-1}h^{-1}$) and 79.5–86.9 cm (-5.17 x 10^4 $g^{-1}h^{-1}$). The virus-to-prokaryote started out high at the shallower depth of 6–13.5 cm (0.63 v/b) and then decreased with depth, 24.3–27.6 cm (0.15 v/b), 35.2–39.4 cm (0.38 v/b), 68–72.3 cm (0.23 v/b), 72.5–76.9 cm (0.26 v/b). At the lowest depth (79.5–86.9 cm) the virus-to-prokaryote ratio increased to 2.57 v/b.

A Spearman rank correlation analysis was carried out to test the relationships among virus activity and the nutrient data. The rate of viral production was significantly positively correlated with the N as ammonia ($r_s = 0.942$, $p < 0.05$) (**Figure 21**). The rate of viral production was not significantly correlated with phosphate ($r_s = 0.085$) and silicate ($r_s = 1.00$).

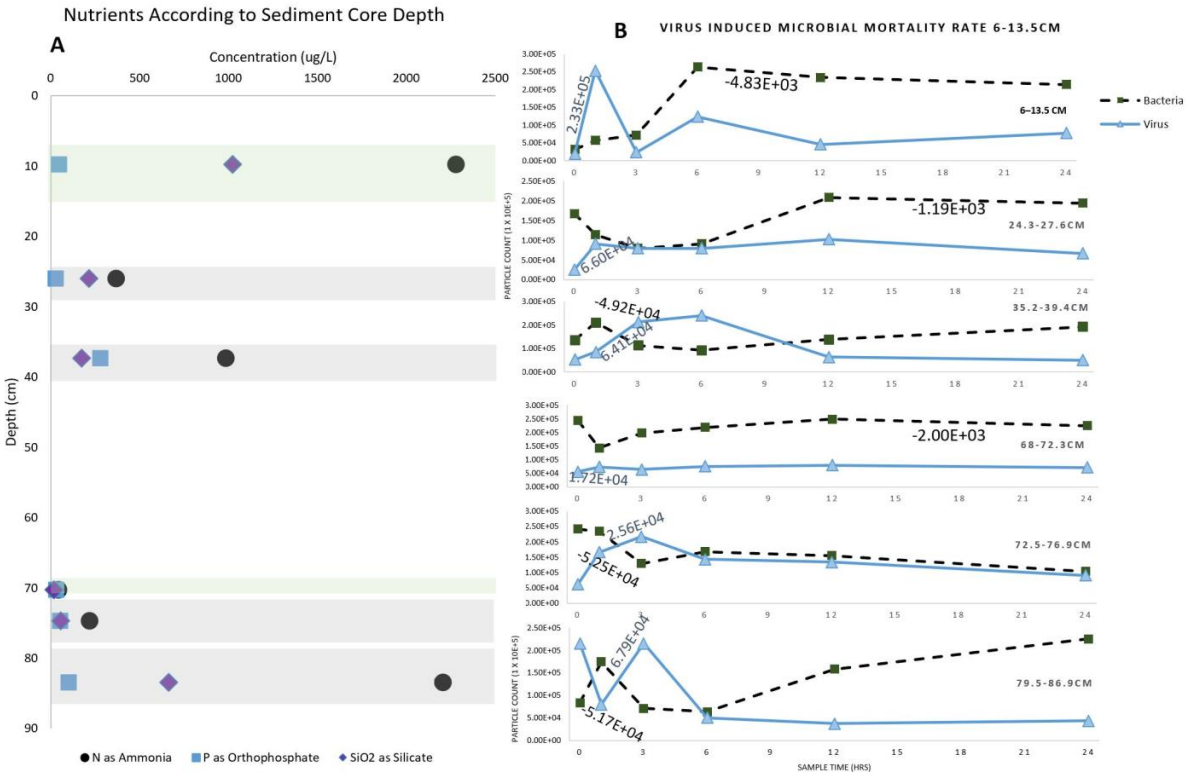

Figure 20: Nutrients in sediment sampled at six different depths compared to the corresponding virus induced microbial mortality rates. **(A)** Nutrients (Ammonia, Phosphate, Silicate) according to depth of sediment core) and **(B)** Virus Induced Microbial Mortality Rate for sample 6–13.5 cm, 24.3–27.6 cm, 35.2–39.4 cm, 68–72.3 cm, 72.5–76.9 cm, and 79.5–86.9 cm at six time points (0, 1, 3, 6, 12, 24 hrs). The green line is bacteria counted and the blue line is viruses counted. N in Ammonia (green circles dashed line), P in Orthophosphate (blue squares), and SiO_2 in Silicate (purple triangles dashed line).

Table 12: Viral production, microbial mortality, and virus-to-prokaryote ratio rates.

Sample	Viral Production	Microbial Mortality	Virus-to-prokaryote Ratio
6–13.5 cm	2.33×10^5 $g^{-1}h^{-1}$	-4.83×10^3 $g^{-1}h^{-1}$	0.63 v/b
24.3–27.6 cm	6.60×10^4 $g^{-1}h^{-1}$	-1.19×10^3 $g^{-1}h^{-1}$	0.15 v/b
35.2–39.4 cm	6.41×10^4 $g^{-1}h^{-1}$	-4.92×10^4 $g^{-1}h^{-1}$	0.38 v/b
68–72.3 cm	1.72×10^4 $g^{-1}h^{-1}$	-2.00×10^3 $g^{-1}h^{-1}$	0.23 v/b
72.5–76.9 cm	2.56×10^4 $g^{-1}h^{-1}$	-5.25×10^4 $g^{-1}h^{-1}$	0.26 v/b
79.5–86.9 cm	6.79×10^4 $g^{-1}h^{-1}$	-5.17×10^4 $g^{-1}h^{-1}$	2.57 v/b

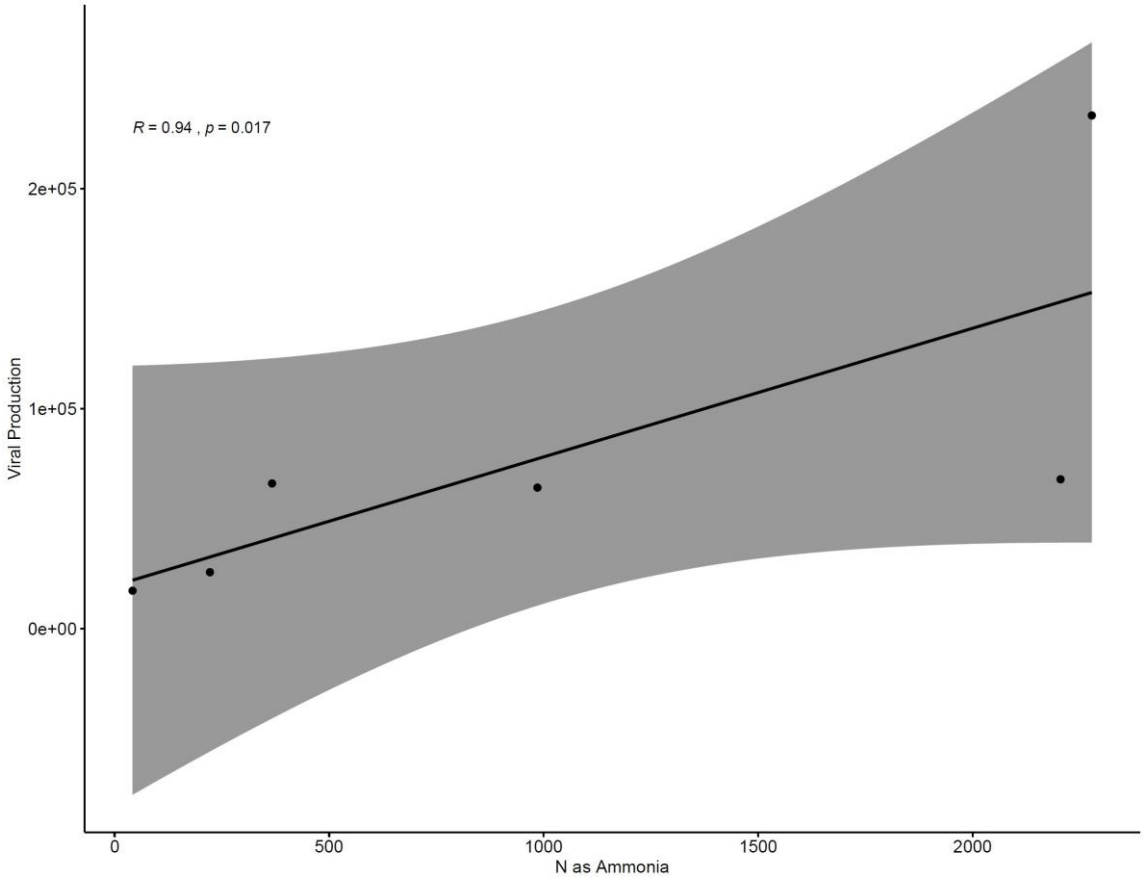

Figure 21: Spearman rank correlation analysis of the significant relationship between the rate of viral production and the N as ammonia. The rate of viral production was significantly positively correlated with the N as ammonia ($rs = 0.942$, $p < 0.05$).

3.4. Discussion

3.4.1. Metagenomic diversity and the potential nutrients available in sediment

Coastal anoxic environments, such as sinkholes, provide easy-to-access sediment samples and allow the study of microbial communities and the alternative nutrients available potentially utilized by the community. Here, we looked at the microbial community composition, the genes associated with metabolisms corresponding to measured nutrients, and the viral impact on the community. According to the

metagenomic data, there is a large number of various taxonomic classes found within each anoxic sediment layer. The phyla dominating the metagenomic samples are the same as the 16S rRNA gene data; Chloroflexi, Proteobacteria, and Planctomycetes. The phyla Crenarchaeota and Bathyarchaeota are more prevalent in the deeper layers of the core (68–72.3 cm, 72.5–76.9 cm and 79.5–86.9 cm), which could be due to the increase in nutrients and carbon. Crenarchaeota are mainly located in deep-sea sediment and have been associated with sulfur reduction (Behera *et al.*, 2020; Kubo *et al.*, 2012). Although Bathyarchaeota is a diverse and abundant phylum, total organic carbon was found to be impacting the mangrove sediment abundance of Bathyarchaeota in Shenzhen, China (Pan, Chen, *et al.*, 2019). The highest C:N ratio (39.5) and the drastic increase in carbon showed a different metagenomic taxonomy at the 72.5–76.9 cm layer. The family Rhodobacteraceae compiled 13% of the microbial abundance in the layer. Rhodobacteraceae has been found in Brazilian mangrove sediment and tied to functional abilities, such as methane and nitrogen metabolism (Andreote *et al.*, 2012). The family Pseudomonadaceae was 5% of total layer, which is a denitrifying bacterium. The bins found with the *nosZ* gene encoding denitrification, were located in the same layers as Pseudomonadaceae; 68–72.3 cm, 72.5–76.9 cm and 79.5–86.9 cm.

Bacteria and Archaea play an important role in the nitrogen cycle. Due to the anaerobic nature of the studied sediment, denitrification and anaerobic ammonium oxidation (anammox) are the major pathways for nitrogen (Engström *et al.*, 2005). Microbial communities are impacted by nutrient availability and chemical gradients, such as nitrogen (Reese *et al.*, 2018), phosphorus (März *et al.*, 2018) and sulfur

metabolisms (Berben *et al.*, 2019). Sulfur oxidation and denitrification occur in Chloroflexi and Crenarchaeota, known to use nutrients, such as sulfur and nitrogen (Kubo *et al.*, 2012; Meyer *et al.*, 2007). Both Chloroflexi and Crenarchaeota were found in all the metagenomic samples. Chloroflexi make up a large portion of the microbial community in other sinkhole environments, including freshwater lakes (Nold *et al.*, 2010). Carbohydrate metabolism, found within the metagenomes, is a biochemical process responsible for the metabolic formation, breakdown, and interconversion of carbohydrates. Within sediment, Bathyarchaeota are capable of extracellular carbohydrate degradation (Orsi *et al.*, 2018). Iron acquisition is utilized by microbes (*e.g.* the orders *Actinomycetales* and *Chromatiales*) for biochemical processes under anaerobic conditions (Edwards *et al.*, 2006; Konhauser *et al.*, 2011). Bathyarchaeota was one of the more abundant phyla found within the metagenomic samples. Aromatic compound metabolism proteins encoded in microbial genomes (*e.g.* the phyla Bathyarchaeota and Aminicenantes) include activation of the benzene ring and are thermodynamically favorable in conditions of deep-sea sediment, such as high pressure and low temperature (Dong *et al.*, 2019). Archaea, specifically Euryarchaeota and Crenarchaeota, were found in great abundance in similar a sinkhole environment and were potentially participating in the sulfur cycle (Ferrer *et al.*, 2011).

3.4.2. Metabolic genes found indicate possible nutrient cycling occurring in sediment

Metabolic genes of microbial communities are driven by nutrient availability and redox potential (Orsi, 2018). The genes found from the compiled MAGs that correspond

to sulfur and nitrogen metabolism were *soxABDG* subunits, *nifH, nifS, dsrE,* and *nosZ.* Sulfur oxidizing bacteria utilize the sox gene, which is comprised of multiple subunits (*soxA, soxB, soxD,* and *soxG*) (Friedrich *et al.*, 2001). The *dsrE* is part of a larger cluster of *dsr* (dissimilatory sulfite reductase) genes, which are used for sulfur oxidation to sulfite (Berben *et al.*, 2019). The *nif* genes (*nifH* and *nifS*) are found in nitrogen-fixing bacteria and correspond to a wide range of taxa, encompassing members of the Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria, and Firmicutes (Dias *et al.*, 2012; Reese *et al.*, 2018). The nifH protein (iron protein or dinitrogenase reductase) functions as an ATP-dependent electron donor to a larger component, which is composed of the nifD and nifK proteins (Mus *et al.*, 2018). The *nosZ* (nitrous oxide reductase) gene is utilized by denitrifying bacteria (Horn *et al.*, 2006; Orellana *et al.*, 2014). All genes found within the metagenomes can be potentially expressed by microbes to facilitate biogeochemical pathways, such as the nitrogen and sulfur cycle.

MAG 79bin17 was identified as *Desulfurobacteriales*, an autotrophic sulfur reducing bacteria (*dsr* gene), encoded for the *sox* and *nosZ* genes. MAG 68bin7 was identified as *Desulfurobacteriales*, and contained *sox*, *dsr, nif*, and *nosZ* genes. *Desulfurobacteriales* are usually found in relatively high abundance in areas of high sulfide concentrations within a low-oxygen sinkhole environment (Kinsman-Costello *et al.*, 2017). MAG 72bin1 was identified as *Kallotenuales*, a known thermophile (Cole *et al.*, 2013), and contained *sox* and *nif* genes. MAG 72bin7 was identified as *Caldilineales*, a natural dechlorinator (D'Angelo *et al.*, 2010), and contained *dsr* genes. MAG 72bin19 from was identified as *Desulfurobacteriales*, and contained *sox*, *dsr*, and

nosZ genes, which are autotrophic sulfur reducing bacteria (Trembath-Reichert *et al.*, 2019). 72bin14 was identified as *Methanopyrales*, known methanogens that are associated with denitrifying bacteria, contained *sox* and *nif* genes (Andalib *et al.*, 2011; Luo *et al.*, 2009). MAG 72bin28 was identified as *Methanopyrales*, methanogens, contained *sox* and *dsr* genes (Andalib *et al.*, 2011; Luo *et al.*, 2009). The Roseobacter clade genomes, a member of the family Rhodobacteraceae, was predominantly found within the 72.5–76.9 cm layer of Blackwood sediment samples, were found to encode the *soxB* gene for thiosulfate oxidation in a similar coastal environment (Lenk *et al.*, 2012). Although, without geochemical data for sulfur compounds, we can only assume the genes present within the MAGs indicated the microbial community present could have potentially played a role in sulfur cycling.

In sediment, nitrogen exists in both inorganic and organic forms. Inorganic forms include, dissolved ammonia, ammonium, nitrate, and nitrite. The nitrogen as ammonia showed high levels of concentration (0.0418 ug/mL to 2.278 ug/mL) in all six sediment layers measured. The concentration levels are similar to sediment samples taken from Liangzi Lake, China (0.3 mg/L to 1.0 mg/L) (Meng *et al.*, 2019). Most denitrification occurs in the sediment and can be combined with ammonification (Thamdrup *et al.*, 2002). The appearance of the *nosZ* gene within metabolic genes found at 72.5–76.9 cm and 79.5–86.9 cm indicates the Blackwood Sinkhole sediment could potentially contribute to denitrification (Na *et al.*, 2018). Organic nitrogen is commonly a limiting element within the ocean. A limiting factor to primary productivity is the availability of fixed nitrogen (Ehrlich *et al.*, 2015). In our study, nitrate/nitrite concentrations were too

low to be measured, indicating that the nutrient was completely consumed by the microbial community or not present in the sediment at all. Viral activity was lower when the ammonia concentration was low, suggesting that nitrogen is a limiting factor of microbial and viral activity in the anoxic sediment of Blackwood Sinkhole. With significant correlation of N in ammonia to viral production, and denitrification genes present, the Blackwood Sinkhole microbial community is potentially influenced by nitrogen.

High rates of silicon dioxide (silica) found in this sediment core are due to the phytoplankton particulate matter in the water column of the Blackwood Sinkhole (Lamb *et al.*, 2006). The concentration of silicate measured in one of the top sediment layers (6–13.5 cm) indicates the coastal organic matter (*e.g.* diatoms) within the core. The biological inputs and outputs of silica in the ocean is important in various biogeochemical processes. Dissolved silicon is an important part of the organic matter in coastal sediment. Organisms such as diatoms, use silicic acid (silica) for their shells. Input of silica into the ocean can be via sediment weathering, rivers, and basalt weathering (Treguer *et al.*, 1995). The C:N ratio at 6–13.5 cm was low (18.47), indicating a phytoplankton organic matter source.

3.4.3. Viral metagenomes present linked to nutrient sources

The main families found in samples 24.3–27.6 cm and 68–72.3 cm was *Myoviridae*, *Siphoviridae*, *Podoviridae*, *Herelleviridae*, *Bicaudaviridae*, and unclassified dsDNA Archaeal viruses. The families *Myoviridae*, *Siphoviridae*, *Podoviridae*, and *Herelleviridae* are tailed bacteriophages dsDNA viruses belonging to the order

Caudovirales and commonly occur in various marine and freshwater environments (Garin-Fernandez, 2019; Roux *et al.*, 2012; Suttle, 2005; Yoshida *et al.*, 2018). The families *Ampullaviridae* and *Bicaudaviridae* are associated with hyperthermophilic Archaea from the family *Sulfolobaceae*, usually located in extreme environments dominated by Archaea (Häring *et al.*, 2005; Prangishvili *et al.*, 2017; Prangishvili *et al.*, 2018). Citrobacter and Pseudomonas sp. prophage viral sequences were found in samples at 72.5–76.9 cm and 6–13.5 cm, which correspond to the higher mean gene expression of virulence seen in those layers. All viruses found occur in marine environments and prophage sequences found could be associated with Blackwood Sinkhole sediment microbial community virulence genes. The viral community could be studied more in depth to proactively link viruses with their hosts using various bioinformatics methods that include identifying CRISPR regions using metaCRT (Sanguino *et al.*, 2015) and assembling CRISPR arrays to compare viral contigs and prokaryotic metagenomic data using the Crass assembler (Skennerton *et al.*, 2013). RNA viruses could be extracted to provide actual activity present to further link viruses to their hosts.

The highest viral production rate in sample 79.5–86.9 cm (6.79×10^4 $g^{-1}h^{-1}$) is slow compared to studies completed in deep-sea sediment samples in the Mediterranean Sea ($61.2 \pm 10.8 \times 10^7$ $g^{-1}h^{-1}$) (Corinaldesi *et al.*, 2010). The rate of viral production was significantly positively correlated with the N in ammonia ($r_s = 0.942$, $p < 0.05$), suggesting that viral production is related to the concentration of N in ammonia throughout the core. The virus-to-prokaryotes ratios observed in Blackwood Sinkhole

are different than what has been observed within the deep-sea. This relationship is difficult to compare due to the lack of robust studies of coastal anoxic sediment. In our samples, the virus-to-prokaryote ratio increased with depth (2.57 v/b at 79.5–86.9 cm), which is consistent with other observations in deep-sea sediments (Corinaldesi *et al.*, 2010). The ratio also increased at 6–13.5 cm (0.63 v/b) and 79.5–86.9 cm (2.57 v/b) when nutrient levels, specifically nitrogen as ammonia increased at 6–13.5 cm (2.278 ug/mL) and 79.5–86.9 cm (2.2051 ug/mL). Within each sediment layer the microbial mortality rate depended on organic carbon and other nutrients. There is a high virus-to-bacteria ratio in samples with high carbon content and high nutrient concentrations.

Overall, the microbial community was dominated by members of the Chloroflexi, Proteobacteria, Planctomycetes, Crenarchaeota and Bathyarchaeota phyla in coastal anoxic sediment. The microbial community found potentially utilize both sulfur and nitrogen metabolisms. Metagenomic analysis shows the range of metabolic genes present for sulfur oxidation, sulfate reduction, denitrification, and nitrogen fixation found within the community. These metabolic genes present within the community provide a snapshot of potential metabolisms, which are linked to the alternative nutrient sources to oxygen, which were silicate, phosphate, and N in ammonia. As the more available form of nitrogen, ammonia increases are significantly correlated ($r_s = 0.942$, $p < 0.05$) to the rate of viral production increases. The Blackwood Sinkhole microbial community functions on alternative electron acceptors and viruses have an active presence within the community.

4. CONCLUSIONS

4.1. The Anoxic Blackwood Sinkhole is home to a diverse microbial community

With a range of 3.8% to 22.6% of TOC in each analyzed stratigraphic layer, the carbon content is significantly higher than deep-sea sediment. Indeed, TOC in the Mediterranean Sea sediment ranges from 0.5% to 2.8% (Corinaldesi *et al.*, 2010) while it ranges from 0.67% to 0.86% in the Mariana Trench (Yoshida *et al.*, 2013). Carbon is therefore not thought to be a limiting growth factor for the microbial communities inhabiting the Blackwood Sinkhole sediment. In general, the source of carbon is phytoplanktonic when the C:N ratio was <20 and is terrestrial (leaves and twigs) when the C:N ratio was >20. Within approximately 2–30 cm of the core, the C:N ratio was <20 indicating the shallower depths have an influx of phytoplankton POM. From approximately 30–79 cm the C:N ratio was >20, showing the deeper layers of the core were built up over time with terrestrial POM upon fluctuation of the basin. Although carbon is not a limiting nutrient, the source of carbon also plays a role in shaping the microbial communities as we found that the C:N ratio significantly correlated with microbial community composition and functionality. Heterotrophic Bacteria were commonly found in sediment where the source of carbon is phytoplankton and autotrophic Bacteria are commonly found in sediment where the source of carbon is terrestrial (Fuentes *et al.*, 2016).

According to the 16S rRNA gene and metagenomic data, the main taxa found in all the layers belong to the classes Dehalococcoidia, Gammaproteobacteria,

Bathyarchaeota, and Campylobacter, all commonly found in sediment (Wang *et al.*, 2012). A number of Archaea phyla, such as Euryarchaeota, Nanoarchaeota, and Crenarcheaota, were also found in each layer of the anoxic sediment (Huber *et al.*, 2003; Iverson *et al.*, 2012; Kubo *et al.*, 2012). There is a notable shift in the bacterial families with more Desulfobacteraceae, sulfate reducers (Kümmel *et al.*, 2015), in the shallower depths and Rhodobacteraceae, sulfur oxidizers (Pohlner *et al.*, 2019), deeper in the core. At 72.5–76.9 cm the metagenomic data shows a different microbial community, which could be due to the drastic increase in the C:N ratio. The texture of this layer was also very different as it was very leafy.

The stratigraphically-constrained hierarchal cluster analysis showed the microbial community clustered within three groups; Group 1 (0-1008 Cal yrs BP), Group 2 (1009-1502 Cal yrs BP), and Group 3 (1520-2990 Cal yrs BP). The source of organic matter, *e.g.*, Group 1 POM input was mainly marine and terrestrial, Group 2 was mainly freshwater productivity and terrestrial POM input, and Group 3 was mainly terrestrial POM, for the three groups influenced the taxa present at time of deposition.

4.2. Alternative to oxygen electron acceptors provide a range of redox reactions within the Blackwood Sinkhole

Another factor influencing the microbial community activity is the availability of electron donors and acceptors. Given the excess of TOC, organic carbon is most likely the main electron donor. The availability of different nutrients provides a range of redox reactions, capable of facilitating a diverse microbial community. Here, we looked at the concentration of different nutrients (silicate, phosphate, and N in ammonia). The highest

concentration of silicate was at 6–13.5 cm, when the C:N ratio was less than 20. This could be due to the most common type of phytoplankton, diatoms, whose cell walls consist of silica.

We searched for genes associated to the main metabolic pathways commonly found in anoxic sediment. Genes found within MAGS are involved in nitrogen fixation (*nifH* and *nifS*), denitrification (*nosZ*), and sulfur oxidation (*sox*). The MAGs found were nearly complete genomes, therefore they are likely from the most abundant microbes found within the sediment. The *nif* genes correspond to a wide range of taxa, encompassing members of the Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria, and Firmicutes (Dias *et al.*, 2012; Reese *et al.*, 2018). The genes found within the most abundant microbes, indicate the metabolisms, such as sulfate reduction (Kubo *et al.*, 2012) and nitrite reduction (Xiang *et al.*, 2017), are important, and that microorganisms play substantial roles in the nitrogen, sulfur and carbon cycles.

4.3. Viruses are prevalent and important members of the microbial community

We observed a lower viral activity in samples where the nutrient concentrations were at their lowest, suggesting that viral activity relies upon their host activity, which depends on nutrient availability. Viral metagenomics from two layers (24.3–27.6 cm and 68–72.3 cm) revealed viruses associated with both Bacteria and Archaea. Moreover, all prokaryotic metagenomic samples contained viral contigs, suggesting that there were ongoing infections at time of sampling. The rate of viral production was significantly correlated with the ammonia concentration ($r_s = 0.942$, $p < 0.05$), proving a relationship

between nutrients (specifically, ammonia) and viruses. Viral activity does not necessarily depend on depth, but the nutrients available in the sediment.

4.4. Conclusions

Overall, Blackwood Sinkhole is a great study site with potential for further research into anoxic coastal sediment microbial and viral community interactions. Specifically, further research could be conducted to determine impact of other nutrients and connections to genes present, such as methane and sulfur, the viral community could be studied more in depth to proactively link viruses with their hosts, and an in depth analysis of the microbial community within the water column to see if there is any correlation to what's going on in the sediment.

Through the genomic characterization of the resident microbial communities inhabiting sediments, we learned about the mechanisms that are driving their survival and adaptation to the varying nutrient concentrations available in sediments. This study was unique as it is one of the first that looked at the genomic composition of both prokaryotes and viruses. We were also able to estimate microbial activity, using viral production as a proxy, as well as identify the source of organic matter.

REFERENCES

Abràmoff, M. D., Magalhães, P. J., & Ram, S. J. (2004). Image processing with ImageJ. *Biophotonics international, 11*(7), 36-42.

Acosta-González, A., & Marqués, S. (2016). Bacterial diversity in oil-polluted marine coastal sediments. *Current opinion in biotechnology, 38*, 24-32.

Anantharaman, K., Duhaime, M. B., Breier, J. A., Wendt, K., Toner, B. M., & Dick, G. J. (2014). Sulfur oxidation genes in diverse deep-sea viruses. *Science*, 1252229.

Andalib, M., Nakhla, G., McIntee, E., & Zhu, J. (2011). Simultaneous denitrification and methanogenesis (SDM): Review of two decades of research. *Desalination, 279*(1-3), 1-14.

Anderson, R. E., Brazelton, W. J., & Baross, J. A. (2013). The deep viriosphere: assessing the viral impact on microbial community dynamics in the deep subsurface. *Reviews in Mineralogy and Geochemistry, 75*(1), 649-675.

Andreote, F. D., Jiménez, D. J., Chaves, D., Dias, A. C. F., Luvizotto, D. M., Dini-Andreote, F., . . . Taketani, R. G. (2012). The microbiome of Brazilian mangrove sediments as revealed by metagenomics. *PLoS One, 7*(6).

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

Behera, P., Mohapatra, M., Kim, J. Y., & Rastogi, G. (2020). Benthic archaeal community structure and carbon metabolic profiling of heterotrophic microbial communities in brackish sediments. *Science of The Total Environment, 706*, 135709.

Berben, T., Overmars, L., Sorokin, D. Y., & Muyzer, G. (2019). Diversity and distribution of sulfur oxidation-related genes in thioalkalivibrio, a genus of chemolithoautotrophic and haloalkaliphilic sulfur-oxidizing bacteria. *Frontiers in Microbiology, 10*, 160.

Bernhard, A. (2010). The nitrogen cycle: Processes. *Players, and Human.*

Biddanda, B. A., Nold, S. C., Dick, G. J., Kendall, S., Vail, J., Ruberg, S., & Green, C. (2012). Rock, water, microbes: underwater sinkholes in Lake Huron are habitats for ancient microbial life. *Nature Education Knowledge, 3*, 13.

Biddle, J. F., Fitz-Gibbon, S., Schuster, S. C., Brenchley, J. E., & House, C. H. (2008). Metagenomic signatures of the Peru Margin subseafloor biosphere show a genetically distinct environment. *Proceedings of the National Academy of Sciences, 105*(30), 10583-10588.

Biderre-Petit, C., Dugat-Bony, E., Mege, M., Parisot, N., Adrian, L., Moné, A., . . . Boucher, D. (2016). Distribution of Dehalococcoidia in the anaerobic deep water of a remote meromictic crater lake and detection of Dehalococcoidia-derived reductive dehalogenase homologous genes. *PLoS One, 11*(1), e0145558.

Blaauw, M., & Christen, J. A. (2011). Flexible paleoclimate age-depth models using an autoregressive gamma process. *Bayesian analysis, 6*(3), 457-474.

Böer, S. I., Hedtkamp, S. I., Van Beusekom, J. E., Fuhrman, J. A., Boetius, A., & Ramette, A. (2009). Time-and sediment depth-related variations in bacterial diversity and community structure in subtidal sands. *The ISME journal, 3*(7), 780.

Breitbart, M., Miyake, J. H., & Rohwer, F. (2004). Global distribution of nearly identical phage-encoded DNA sequences. *FEMS Microbiol Lett, 236*(2), 249-256. doi:10.1016/j.femsle.2004.05.042

Brüssow, H., Canchaya, C., & Hardt, W.-D. (2004). Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiology and molecular biology reviews, 68*(3), 560-602.

Buchfink, B., Xie, C., & Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. *Nature methods, 12*(1), 59.

Bushnell, B., Rood, J., & Singer, E. (2017). BBMerge–accurate paired shotgun read merging via overlap. *PLoS One, 12*(10), e0185056.

Cai, L., Jørgensen, B. B., Suttle, C. A., He, M., Cragg, B. A., Jiao, N., & Zhang, R. (2019). Active and diverse viruses persist in the deep sub-seafloor sediments over thousands of years. *The ISME journal*, 1.

Calderoli, P. A., Espinola, F. J., Dionisi, H. M., Gil, M. N., Jansson, J. K., & Lozada, M. (2018). Predominance and high diversity of genes associated to denitrification in metagenomes of subantarctic coastal sediments exposed to urban pollution. *PLoS One*, *13*(11).

Canfield, D. E., Stewart, F. J., Thamdrup, B., De Brabandere, L., Dalsgaard, T., Delong, E. F., . . . Ulloa, O. (2010). A cryptic sulfur cycle in oxygen-minimum–zone waters off the Chilean coast. *Science*, 1196889.

Cassman, N., Prieto-Davó, A., Walsh, K., Silva, G. G., Angly, F., Akhter, S., . . . Haggerty, J. M. (2012). Oxygen minimum zones harbour novel viral communities with low diversity. *Environmental Microbiology, 14*(11), 3043-3065.

Cetecioğlu, Z., Ince, B. K., Kolukirik, M., & Ince, O. (2009). Biogeographical distribution and diversity of bacterial and archaeal communities within highly polluted anoxic marine sediments from the Marmara Sea. *Marine Pollution Bulletin, 58*(3), 384-395.

Cole, J. K., Gieler, B. A., Heisler, D. L., Palisoc, M. M., Williams, A. J., Dohnalkova, A. C., . . . Li, W.-J. (2013). Kallotenue papyrolyticum gen. nov., sp. nov., a cellulolytic and filamentous thermophile that represents a novel lineage (Kallotenuales ord. nov., Kallotenuaceae fam. nov.) within the class Chloroflexia. *International journal of systematic and evolutionary microbiology, 63*(12), 4675-4682.

Collins, S., Reinhardt, E., Werner, C., Le Maillot, C., Devos, F., & Meacham, S. (2015). Regional response of the coastal aquifer to Hurricane Ingrid and sedimentation flux in the Yax Chen cave system (Ox Bel Ha) Yucatan, Mexico. *Palaeogeography, Palaeoclimatology, Palaeoecology, 438*, 226-238.

Corinaldesi, C., Dell'Anno, A., Magagnini, M., & Danovaro, R. (2010). Viral decay and viral production rates in continental-shelf and deep-sea sediments of the Mediterranean Sea. *FEMS Microbiology Ecology, 72*(2), 208-218.

Corinaldesi, C., Tangherlini, M., & Dell'Anno, A. (2017). From virus isolation to metagenome generation for investigating viral diversity in deep-sea sediments. *Scientific Reports, 7*(1), 8355.

D'Hondt, S., Rutherford, S., & Spivack, A. J. (2002). Metabolic activity of subsurface life in deep-sea sediments. *Science, 295*(5562), 2067-2070.

D'Angelo, E., & Nunez, A. (2010). Effect of environmental conditions on polychlorinated biphenyl transformations and bacterial communities in a river sediment. *Journal of soils and sediments, 10*(6), 1186-1199.

Daims, H., Lebedeva, E. V., Pjevac, P., Han, P., Herbold, C., Albertsen, M., . . . Bulaev, A. (2015). Complete nitrification by Nitrospira bacteria. *Nature, 528*(7583), 504.

Damste, J. S. S., Strous, M., Rijpstra, W. I. C., Hopmans, E. C., Geenevasen, J. A., Van Duin, A. C., . . . Jetten, M. S. (2002). Linearly concatenated cyclobutane lipids form a dense bacterial membrane. *Nature, 419*(6908), 708.

Danovaro, R., & Middelboe, M. (2010). Separation of free virus particles from sediments in aquatic systems. 74-81. doi:10.4319/mave.2010.978-0-9845591-0-7.74

Dell'Anno, A., Corinaldesi, C., & Danovaro, R. (2015). Virus decomposition provides an important contribution to benthic deep-sea ecosystem functioning. *Proceedings of the National Academy of Sciences, 112*(16), E2014-E2019.

Dias, A. C. F., e Silva, M. d. C. P., Cotta, S. R., Dini-Andreote, F., Soares, F. L., Salles, J. F., . . . Andreote, F. D. (2012). Abundance and genetic diversity of nifH gene sequences in anthropogenically affected Brazilian mangrove sediments. *Appl. Environ. Microbiol., 78*(22), 7960-7967.

Dixon, P. (2003). VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science, 14*(6), 927-930.

Dong, X., Greening, C., Rattray, J. E., Chakraborty, A., Chuvochina, M., Mayumi, D., . . . Bernard, B. B. (2019). Metabolic potential of uncultured bacteria and archaea

associated with petroleum seepage in deep-sea sediments. *Nature communications, 10*(1), 1-12.

Dyksma, S., Bischof, K., Fuchs, B. M., Hoffmann, K., Meier, D., Meyerdierks, A., . . . Stepanauskas, R. (2016). Ubiquitous Gammaproteobacteria dominate dark carbon fixation in coastal sediments. *The ISME journal, 10*(8), 1939-1953.

Edwards, R. A., Rodriguez-Brito, B., Wegley, L., Haynes, M., Breitbart, M., Peterson, D. M., . . . Rohwer, F. (2006). Using pyrosequencing to shed light on deep mine microbial ecology. *BMC genomics, 7*(1), 57.

Ehrlich, H. L., Newman, D. K., & Kappler, A. (2015). *Ehrlich's geomicrobiology*: CRC press.

Engelen, B., Ziegelmüller, K., Wolf, L., Köpke, B., Gittel, A., Cypionka, H., . . . Steinsbu, B. r. O. (2008). Fluids from the Oceanic Crust Support Microbial Activities within the Deep Biosphere. *Geomicrobiology Journal, 25*(1), 56-66. doi:10.1080/01490450701829006

Engelhardt, T., Kallmeyer, J., Cypionka, H., & Engelen, B. (2014). High virus-to-cell ratios indicate ongoing production of viruses in deep subsurface sediments. *ISME J, 8*(7), 1503-1509. doi:10.1038/ismej.2013.245

Engström, P., Dalsgaard, T., Hulth, S., & Aller, R. C. (2005). Anaerobic ammonium oxidation by nitrite (anammox): implications for N2 production in coastal marine sediments. *Geochimica et Cosmochimica Acta, 69*(8), 2057-2065.

Fagervold, S. K., Bourgeois, S., Pruski, A. M., Charles, F., Kerherve, P., Vetion, G., & Galand, P. E. (2014). River organic matter shapes microbial communities in the sediment of the Rhone prodelta. *The ISME journal, 8*(11), 2327-2338.

Ferrer, M., Guazzaroni, M.-E., Richter, M., García-Salamanca, A., Yarza, P., Suárez-Suárez, A., . . . Molina-Henares, M. A. (2011). Taxonomic and functional metagenomic profiling of the microbial community in the anoxic sediment of a sub-saline shallow lake (Laguna de Carrizo, Central Spain). *Microbial Ecology, 62*(4), 824-837.

Finke, J. F., Hunt, B. P., Winter, C., Carmack, E. C., & Suttle, C. A. (2017). Nutrients and other environmental factors influence virus abundances across oxic and hypoxic marine environments. *Viruses*, *9*(6), 152.

Franco, D. C., Signori, C. N., Duarte, R. T., Nakayama, C. R., Campos, L. S., & Pellizari, V. H. (2017). High prevalence of gammaproteobacteria in the sediments of admiralty bay and north bransfield Basin, Northwestern Antarctic Peninsula. *Frontiers in Microbiology*, *8*, 153.

Franco, M., De Mesel, I., Diallo, M. D., Van Der Gucht, K., Van Gansbeke, D., Van Rijswijk, P., . . . Vanaverbeke, J. (2007). Effect of phytoplankton bloom deposition on benthic bacterial communities in two contrasting sediments in the southern North Sea. *Aquatic Microbial Ecology*, *48*(3), 241-254.

Friedrich, C. G., Rother, D., Bardischewsky, F., Quentmeier, A., & Fischer, J. (2001). Oxidation of reduced inorganic sulfur compounds by bacteria: emergence of a common mechanism? *Appl. Environ. Microbiol.*, *67*(7), 2873-2882.

Frumkin, A., Zaidner, Y., Na'aman, I., Tsatskin, A., Porat, N., & Vulfson, L. (2015). Sagging and collapse sinkholes over hypogenic hydrothermal karst in a carbonate terrain. *Geomorphology*, *229*, 45-57.

Fuentes, J. L., Garbayo, I., Cuaresma, M., Montero, Z., González-del-Valle, M., & Vílchez, C. (2016). Impact of microalgae-bacteria interactions on the production of algal biomass and associated compounds. *Marine drugs*, *14*(5), 100.

Garin-Fernandez, A. (2019). *VIRUSES IN THE NORTH SEA.* Department of Life Sciences and Chemistry, Jacobs University Bremen,

Gilbert, J. A., Jansson, J. K., & Knight, R. (2014). The Earth Microbiome project: successes and aspirations. *BMC biology*, *12*(1), 69.

Glassing, A., Dowd, S. E., Galandiuk, S., Davis, B., & Chiodini, R. J. (2016). Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. *Gut pathogens*, *8*(1), 24.

Glud, R. N. (2008). Oxygen dynamics of marine sediments. *Marine Biology Research, 4*(4), 243-289.

Glud, R. N., & Mathias, M. (2004). Virus and bacteria dynamics of a coastal sediment: implication for benthic carbon cycling. *Limnology and Oceanography, 49*(6), 2073-2081.

Gregory, A. C., Zayed, A. A., Conceição-Neto, N., Temperton, B., Bolduc, B., Alberti, A., . . . Cruaud, C. (2019). Marine DNA viral macro-and microdiversity from pole to pole. *Cell, 177*(5), 1109-1123. e1114.

Handley, K. M., VerBerkmoes, N. C., Steefel, C. I., Williams, K. H., Sharon, I., Miller, C. S., . . . Shah, M. B. (2013). Biostimulation induces syntrophic interactions that impact C, S and N cycling in a sediment microbial community. *The ISME journal, 7*(4), 800.

Häring, M., Rachel, R., Peng, X., Garrett, R. A., & Prangishvili, D. (2005). Viral diversity in hot springs of Pozzuoli, Italy, and characterization of a unique archaeal virus, Acidianus bottle-shaped virus, from a new family, the Ampullaviridae. *Journal of virology, 79*(15), 9904-9911.

Hart, E. A., & Schurger, S. G. (2005). Sediment storage and yield in an urbanized karst watershed. *Geomorphology, 70*(1-2), 85-96.

Harwani, D. (2013). The great plate count anomaly and the unculturable bacteria. *Microbiology, 2*(9), 350-351.

Holmkvist, L., Ferdelman, T. G., & Jørgensen, B. B. (2011). A cryptic sulfur cycle driven by iron in the methane zone of marine sediment (Aarhus Bay, Denmark). *Geochimica et Cosmochimica Acta, 75*(12), 3581-3599.

Horn, M. A., Drake, H. L., & Schramm, A. (2006). Nitrous oxide reductase genes (nosZ) of denitrifying microbial populations in soil and the earthworm gut are phylogenetically similar. *Appl. Environ. Microbiol., 72*(2), 1019-1026.

Huber, H., Hohn, M. J., Stetter, K. O., & Rachel, R. (2003). The phylum Nanoarchaeota: present knowledge and future perspectives of a unique form of life. *Research in Microbiology, 154*(3), 165-171.

Huson, D. H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., . . . Tappu, R. (2016). MEGAN community edition-interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS computational biology, 12*(6), e1004957.

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*(1), 119.

Inagaki, F. (2015). Exploring deep microbial life in coal-bearing sediment down to ~2.5km below the ocean floor

Iverson, V., Morris, R. M., Frazar, C. D., Berthiaume, C. T., Morales, R. L., & Armbrust, E. V. (2012). Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. *Science, 335*(6068), 587-590.

Janda, J. M., & Abbott, S. L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *Journal of clinical microbiology, 45*(9), 2761-2764.

Jang, H. B., Bolduc, B., Zablocki, O., Kuhn, J. H., Roux, S., Adriaenssens, E. M., . . . Lavigne, R. (2019). Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nature biotechnology, 37*(6), 632.

Jorgensen, B. B., & Boetius, A. (2007). Feast and famine--microbial life in the deep-sea bed. *Nat Rev Microbiol, 5*(10), 770-781. doi:10.1038/nrmicro1745

Jorgensen, B. B., & Marshall, I. P. (2016). Slow Microbial Life in the Seabed. *Ann Rev Mar Sci, 8*, 311-332. doi:10.1146/annurev-marine-010814-015535

Juggins, S., & Juggins, M. S. (2019). Package 'rioja'.

Kallmeyer, J., Pockalny, R., Adhikari, R. R., Smith, D. C., & D'Hondt, S. (2012). Global distribution of microbial abundance and biomass in subseafloor sediment. *Proc Natl Acad Sci U S A, 109*(40), 16213-16216. doi:10.1073/pnas.1203849109

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.

Kang, D., Li, F., Kirton, E. S., Thomas, A., Egan, R. S., An, H., & Wang, Z. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ Preprints, 7*, e27522v27521.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., . . . Duran, C. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics, 28*(12), 1647-1649.

Kinsman-Costello, L. E., Sheik, C. S., Sheldon, N., Allen Burton, G., Costello, D., Marcus, D., . . . Dick, G. (2017). Groundwater shapes sediment biogeochemistry and microbial diversity in a submerged Great Lake sinkhole. *Geobiology, 15*(2), 225-239.

Koch, H., Galushko, A., Albertsen, M., Schintlmeister, A., Gruber-Dorninger, C., Lücker, S., . . . Richter, A. (2014). Growth of nitrite-oxidizing bacteria by aerobic hydrogen oxidation. *Science, 345*(6200), 1052-1054.

Kojima, H., Tsutsumi, M., Ishikawa, K., Iwata, T., Mußmann, M., & Fukui, M. (2012). Distribution of putative denitrifying methane oxidizing bacteria in sediment of a freshwater lake, Lake Biwa. *Systematic and Applied Microbiology, 35*(4), 233-238.

Konhauser, K. O. (2009). *Introduction to geomicrobiology*: John Wiley & Sons.

Konhauser, K. O., Kappler, A., & Roden, E. E. (2011). Iron in microbial metabolisms. *Elements, 7*(2), 89-93.

Koretsky, C. M., Van Cappellen, P., DiChristina, T. J., Kostka, J. E., Lowe, K. L., Moore, C. M., . . . Viollier, E. (2005). Salt marsh pore water geochemistry does not correlate with microbial community structure. *Estuarine, Coastal and Shelf Science, 62*(1-2), 233-251.

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., & 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*(17), 5112-5120.

Kubo, K., Lloyd, K. G., Biddle, J. F., Amann, R., Teske, A., & Knittel, K. (2012). Archaea of the Miscellaneous Crenarchaeotal Group are abundant, diverse and widespread in marine sediments. *The ISME journal, 6*(10), 1949.

Kümmel, S., Herbst, F.-A., Bahr, A., Duarte, M., Pieper, D. H., Jehmlich, N., . . . Richnow, H. H. (2015). Anaerobic naphthalene degradation by sulfate-reducing Desulfobacteraceae from various anoxic aquifers. *FEMS Microbiology Ecology, 91*(3).

Kuypers, M. M., Sliekers, A. O., Lavik, G., Schmid, M., Jørgensen, B. B., Kuenen, J. G., . . . Jetten, M. S. (2003). Anaerobic ammonium oxidation by anammox bacteria in the Black Sea. *Nature, 422*(6932), 608.

Labonté, J. M., Field, E. K., Lau, M., Chivian, D., Van Heerden, E., Wommack, K. E., . . . Stepanauskas, R. (2015). Single cell genomics indicates horizontal gene transfer and viral infections in a deep subsurface Firmicutes population. *Frontiers in Microbiology, 6*, 349.

Labonté, J. M., Lever, M. A., Edwards, K. J., & Orcutt, B. N. (2017). Influence of Igneous Basement on Deep Sediment Microbial Diversity on the Eastern Juan de Fuca Ridge Flank. *Frontiers in Microbiology, 8*, 1434.

Lachnit, T., Dafforn, K. A., Johnston, E. L., & Steinberg, P. (2019). Contrasting distributions of bacteriophages and eukaryotic viruses from contaminated coastal sediments. *Environmental Microbiology, 21*(6), 1929-1941.

Lamb, A. L., Wilson, G. P., & Leng, M. J. (2006). A review of coastal palaeoclimate and relative sea-level reconstructions using δ13C and C/N ratios in organic material. *Earth-Science Reviews, 75*(1-4), 29-57.

Lazar, C. S., Baker, B. J., Seitz, K., Hyde, A. S., Dick, G. J., Hinrichs, K. U., & Teske, A. P. (2016). Genomic evidence for distinct carbon substrate preferences and ecological niches of B athyarchaeota in estuarine sediments. *Environmental Microbiology, 18*(4), 1200-1211.

Leloup, J., Loy, A., Knab, N. J., Borowski, C., Wagner, M., & Jørgensen, B. B. (2007). Diversity and abundance of sulfate-reducing microorganisms in the sulfate and methane zones of a marine sediment, Black Sea. *Environmental Microbiology, 9*(1), 131-142.

Lenk, S., Arnds, J., Zerjatke, K., Musat, N., Amann, R., & Mußmann, M. (2011). Novel groups of Gammaproteobacteria catalyse sulfur oxidation and carbon fixation in a coastal, intertidal sediment. *Environmental Microbiology, 13*(3), 758-774.

Lenk, S., Moraru, C., Hahnke, S., Arnds, J., Richter, M., Kube, M., . . . Amann, R. (2012). Roseobacter clade bacteria are abundant in coastal sediments and encode a novel combination of sulfur oxidation genes. *The ISME journal, 6*(12), 2178-2187.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics, 25*(14), 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079.

Liao, N. (2002). Determination of ammonia in brackish or seawater by flow injection analysis. QuikChem method 31-107-06-1-B. In: Milwaukee: Lachat Instruments. Revised.

Liou, J.-C., & Madsen, E. (2008). Microbial Ecological Processes: Aerobic/Anaerobic.

Löffler, F. E., Yan, J., Ritalahti, K. M., Adrian, L., Edwards, E. A., Konstantinidis, K. T., . . . Spormann, A. M. (2013). Dehalococcoides mccartyi gen. nov., sp. nov., obligately organohalide-respiring anaerobic bacteria relevant to halogen cycling and bioremediation, belong to a novel bacterial class, Dehalococcoidia classis nov., order Dehalococcoidales ord. nov. and family Dehalococcoidaceae fam. nov., within the phylum Chloroflexi. *International journal of systematic and evolutionary microbiology, 63*(2), 625-635.

Luef, B., Luef, F., & Peduzzi, P. (2009). Online program 'vipcal'for calculating lytic viral production and lysogenic cells based on a viral reduction approach. *Environmental microbiology reports, 1*(1), 78-85.

Luo, H., Sun, Z., Arndt, W., Shi, J., Friedman, R., & Tang, J. (2009). Gene order phylogeny and the evolution of methanogens. *PLoS One, 4*(6).

März, C., Riedinger, N., Sena, C., & Kasten, S. (2018). Phosphorus dynamics around the sulphate-methane transition in continental margin sediments: Authigenic apatite and Fe (II) phosphates. *Marine Geology, 404*, 84-96.

McClung, C., & Patriquin, D. (1980). Isolation of a nitrogen-fixing Campylobacter species from the roots of Spartina alterniflora Loisel. *Canadian Journal of Microbiology, 26*(8), 881-886.

Meister, P., Prokopenko, M., Skilbeck, G., & Watson, M. (2005). *Compilation of total organic and inorganic carbon data from Peru margin and eastern equatorial Pacific drill sites.* Paper presented at the Proceedings of the Ocean Drilling Program.

Meng, X., Zhang, W., & Shan, B. (2019). Distribution of nitrogen and phosphorus and estimation of nutrient fluxes in the water and sediments of Liangzi Lake, China. *Environmental Science and Pollution Research*, 1-9.

Meyer, B., Imhoff, J. F., & Kuever, J. (2007). Molecular analysis of the distribution and phylogeny of the soxB gene among sulfur-oxidizing bacteria–evolution of the Sox sulfur oxidation enzyme system. *Environmental Microbiology, 9*(12), 2957-2977.

Miller, W. G., Parker, C. T., Rubenfield, M., Mendz, G. L., Wösten, M. M., Ussery, D. W., . . . Wang, G. (2007). The complete genome sequence and analysis of the epsilonproteobacterium Arcobacter butzleri. *PLoS One, 2*(12).

Mouton, M., Postma, F., Wilsenach, J., & Botha, A. (2012). Diversity and characterization of culturable fungi from marine sediment collected from St. Helena Bay, South Africa. *Microbial Ecology, 64*(2), 311-319.

Mus, F., Alleman, A. B., Pence, N., Seefeldt, L. C., & Peters, J. W. (2018). Exploring the alternatives of biological nitrogen fixation. *Metallomics, 10*(4), 523-538.

Mylroie, J. E., Carew, J. L., & Vacher, H. (1995). Karst development in the Bahamas and Bermuda. *Geological Society of America Special Papers, 300*, 251-267.

Na, T., Thamdrup, B., Kim, B., Kim, S. H., Vandieken, V., Kang, D. J., & Hyun, J. H. (2018). N2 production through denitrification and anammox across the continental margin (shelf–slope–rise) of the Ulleung Basin, East Sea. *Limnology and Oceanography, 63*(S1).

Nealson, K. H. (1997). Sediment bacteria: who's there, what are they doing, and what's new? *Annual Review of Earth and Planetary Sciences, 25*(1), 403-434.

Nold, S. C., Pangborn, J. B., Zajack, H. A., Kendall, S. T., Rediske, R. R., & Biddanda, B. A. (2010). Benthic bacterial diversity in submerged sinkhole ecosystems. *Appl. Environ. Microbiol., 76*(1), 347-351.

Orcutt, B. N., Sylvan, J. B., Knab, N. J., & Edwards, K. J. (2011). Microbial ecology of the dark ocean above, at, and below the seafloor. *Microbiol. Mol. Biol. Rev., 75*(2), 361-422.

Orellana, L., Rodriguez-R, L., Higgins, S., Chee-Sanford, J., Sanford, R. A., Ritalahti, K., . . . Konstantinidis, K. (2014). Detecting nitrous oxide reductase (nosZ) genes in soil metagenomes: method development and implications for the nitrogen cycle. *MBio, 5*(3), e01193-01114.

Orsi, W. D. (2018). Ecology and evolution of seafloor and subseafloor microbial communities. *Nature Reviews Microbiology*, 1.

Orsi, W. D., Richards, T. A., & Francis, W. R. (2018). Predicted microbial secretomes and their target substrates in marine sediment. *Nature microbiology, 3*(1), 32.

Pan, D., Morono, Y., Inagaki, F., & Takai, K. (2019). An improved method for extracting viruses from sediment: detection of far more viruses in the subseafloor than previously reported. *Frontiers in Microbiology, 10*, 878.

Pan, J., Chen, Y., Wang, Y., Zhou, Z., & Li, M. (2019). Vertical distribution of bathyarchaeotal communities in mangrove wetlands suggests distinct niche preference of Bathyarchaeota subgroup 6. *Microbial Ecology, 77*(2), 417-428.

Parada, A. E., Needham, D. M., & Fuhrman, J. A. (2016). Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology, 18*(5), 1403-1414.

Parikka, K. J., Le Romancer, M., Wauters, N., & Jacquet, S. (2017). Deciphering the virus-to-prokaryote ratio (VPR): insights into virus–host relationships in a variety of ecosystems. *Biological Reviews, 92*(2), 1081-1100.

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research, 25*(7), 1043-1055.

Parmar, N., & Singh, A. (2013). *Geomicrobiology and Biogeochemistry* (Vol. 39): Springer.

Patel, A., Noble, R. T., Steele, J. A., Schwalbach, M. S., Hewson, I., & Fuhrman, J. A. (2007). Virus and prokaryote enumeration from planktonic aquatic environments by epifluorescence microscopy with SYBR Green I. *Nat Protoc, 2*(2), 269-276. doi:10.1038/nprot.2007.6

Poghosyan, L., Koch, H., Lavy, A., Frank, J., van Kessel, M. A., Jetten, M. S., . . . Lücker, S. (2019). Metagenomic recovery of two distinct comammox Nitrospira from the terrestrial subsurface. *Environmental Microbiology, 21*(10), 3627-3637.

Pohlner, M., Dlugosch, L., Wemheuer, B., Mills, H., Engelen, B., & Reese, B. K. (2019). The majority of active Rhodobacteraceae in marine sediments belong to uncultured genera: a molecular approach to link their distribution to environmental conditions. *Frontiers in Microbiology, 10*, 659.

Prangishvili, D., Bamford, D. H., Forterre, P., Iranzo, J., Koonin, E. V., & Krupovic, M. (2017). The enigmatic archaeal virosphere. *Nature Reviews Microbiology, 15*(12), 724.

Prangishvili, D., & Krupovic, M. (2018). ICTV Virus Taxonomy Profile: Bicaudaviridae.

Racine, V. (2014). Lysogenic Bacteria as an Experimental Model at the Pasteur Institute (1915-1965). *Embryo Project Encyclopedia*.

RCore Team. (2013). R: A language and environment for statistical computing.

Reed, D. W., Fujita, Y., Delwiche, M. E., Blackwelder, D. B., Sheridan, P. P., Uchida, T., & Colwell, F. S. (2002). Microbial communities from methane hydrate-bearing deep marine sediments in a forearc basin. *Applied and Environmental Microbiology, 68*(8), 3759-3770.

Reese, B. K., Zinke, L. A., Sobol, M. S., LaRowe, D. E., Orcutt, B. N., Zhang, X., . . . Defforey, D. (2018). Nitrogen cycling of active bacteria within oligotrophic sediment of the mid-atlantic ridge flank. *Geomicrobiology Journal, 35*(6), 468-483.

Reimer, P. J., Bard, E., Bayliss, A., Beck, J. W., Blackwell, P. G., Ramsey, C. B., . . . Friedrich, M. (2013). IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon, 55*(4), 1869-1887.

Roux, S., Enault, F., Hurwitz, B. L., & Sullivan, M. B. (2015). VirSorter: mining viral signal from microbial genomic data. *PeerJ, 3*, e985.

Roux, S., Enault, F., Robin, A., Ravet, V., Personnic, S., Theil, S., . . . Debroas, D. (2012). Assessing the diversity and specificity of two freshwater viral communities through metagenomics. *PLoS One, 7*(3).

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular cloning: a laboratory manual*: Cold spring harbor laboratory press.

Sanguino, L., Franqueville, L., Vogel, T. M., & Larose, C. (2015). Linking environmental prokaryotic viruses and their host through CRISPRs. *FEMS Microbiology Ecology, 91*(5).

Santoro, A. E. (2016). The do-it-all nitrifier. *Science, 351*(6271), 342-343.

Schuchmann, K., & Müller, V. (2016). Energetics and application of heterotrophy in acetogenic bacteria. *Appl. Environ. Microbiol., 82*(14), 4056-4069.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics, 30*(14), 2068-2069.

Skennerton, C. T., Imelfort, M., & Tyson, G. W. (2013). Crass: identification and reconstruction of CRISPR from unassembled metagenomic data. *Nucleic Acids Research, 41*(10), e105-e105.

Smith, C. J., McKew, B. A., Coggan, A., & Whitby, C. (2015). Primers: functional genes for nitrogen-cycling microbes in oil reservoirs. In *Hydrocarbon and Lipid Microbiology Protocols* (pp. 207-241): Springer.

Stenuite, S., Pirlot, S., HARDY, M. A., Sarmento, H., TARBE, A. L., Leporcq, B., & DESCY, J. P. (2007). Phytoplankton production and growth rate in Lake Tanganyika: evidence of a decline in primary productivity in recent decades. *Freshwater Biology, 52*(11), 2226-2239.

Suttle, C. A. (2005). Viruses in the sea. *Nature, 437*(7057), 356-361. doi:10.1038/nature04160

Suttle, C. A. (2007). Marine viruses—major players in the global ecosystem. *Nature Reviews Microbiology, 5*(10), 801.

Tamalavage, A. E. (2016). *Stable Organic Isotopic (δ^ 15Norg, δ^ 13Corg) Evidence for Climate-Forced Landscape and Aquatic Change During the Late Holocene in Abaco Island, The Bahamas.*

Tamalavage, A. E., van Hengstum, P. J., Louchouarn, P., Molodtsov, S., Kaiser, K., Donnelly, J. P., . . . Fall, P. L. (2018). Organic matter sources and lateral sedimentation in a Bahamian karst basin (sinkhole) over the late Holocene: Influence of local vegetation and climate. *Palaeogeography, Palaeoclimatology, Palaeoecology*.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution, 28*(10), 2731-2739.

Thamdrup, B., & Dalsgaard, T. (2002). Production of N2 through anaerobic ammonium oxidation coupled to nitrate reduction in marine sediments. *Applied and Environmental Microbiology, 68*(3), 1312-1318.

Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., Ladau, J., Locey, K. J., . . . Ackermann, G. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. *Nature, 551*(7681).

Treguer, P., Nelson, D. M., Van Bennekom, A. J., DeMaster, D. J., Leynaert, A., & Quéguiner, B. (1995). The silica balance in the world ocean: a reestimate. *Science, 268*(5209), 375-379.

Trembath-Reichert, E., Butterfield, D. A., & Huber, J. A. (2019). Active subseafloor microbial communities from Mariana back-arc venting fluids share metabolic strategies across different thermal niches and taxa. *The ISME journal, 13*(9), 2264-2279.

Ul-Hasan, S., Bowers, R. M., Figueroa-Montiel, A., Licea, A. F., Beman, J. M., Woyke, T., & Nobile, C. J. (2018). Archaeal, bacterial, and eukaryal microbial community structure of sediment and seawater in a coastal region near Puerto Nuevo, Baja California. *BioRxiv*, 324442.

van Hengstum, P., Donnelly, J. P., Fall, P. L., Toomey, M. R., Albury, N. A., & Kakuk, B. (2016). The intertropical convergence zone modulates intense hurricane strikes on the western North Atlantic margin. *Sci Rep, 6*, 21728. doi:10.1038/srep21728

van Hengstum, P., Scott, D. B., Gröcke, D. R., & Charette, M. A. (2011). Sea level controls sedimentation and environments in coastal caves and sinkholes. *Marine Geology, 286*(1-4), 35-50. doi:10.1016/j.margeo.2011.05.004

van Kessel, M. A., Speth, D. R., Albertsen, M., Nielsen, P. H., den Camp, H. J. O., Kartal, B., . . . Lücker, S. (2015). Complete nitrification by a single microorganism. *Nature, 528*(7583), 555.

Walsh, E. A., Kirkpatrick, J. B., Pockalny, R., Sauvage, J., Spivack, A. J., Murray, R. W., . . . D'Hondt, S. (2016). Relationship of bacterial richness to organic degradation rate and sediment age in subseafloor sediment. *Applied and Environmental Microbiology*, AEM. 00809-00816.

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol., 73*(16), 5261-5267.

Wang, Y., Sheng, H.-F., He, Y., Wu, J.-Y., Jiang, Y.-X., Tam, N. F.-Y., & Zhou, H.-W. (2012). Comparison of the levels of bacterial diversity in freshwater, intertidal wetland, and marine sediments by using millions of illumina tags. *Appl. Environ. Microbiol., 78*(23), 8264-8271.

Weitz, J. S., & Wilhelm, S. W. (2012). Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 biology reports, 4*.

Williams, K. P., Sobral, B. W., & Dickerman, A. W. (2007). A robust species tree for the alphaproteobacteria. *Journal of bacteriology, 189*(13), 4578-4586.

Williamson, S. J., Cary, S. C., Williamson, K. E., Helton, R. R., Bench, S. R., Winget, D., & Wommack, K. E. (2008). Lysogenic virus-host interactions predominate at deep-sea diffuse-flow hydrothermal vents. *ISME J, 2*(11), 1112-1121. doi:10.1038/ismej.2008.73

Xia, F., Wang, J.-G., Zhu, T., Zou, B., Rhee, S.-K., & Quan, Z.-X. (2018). Ubiquity and diversity of complete ammonia oxidizers (comammox). *Appl. Environ. Microbiol., 84*(24), e01390-01318.

Xiang, X., Wang, R., Wang, H., Gong, L., Man, B., & Xu, Y. (2017). Distribution of Bathyarchaeota communities across different terrestrial settings and their potential ecological functions. *Scientific Reports, 7*, 45028.

Yoshida, M., Mochizuki, T., Urayama, S.-I., Yoshida-Takashima, Y., Nishi, S., Hirai, M., . . . Takai, K. (2018). Quantitative Viral Community DNA Analysis Reveals the Dominance of Single-Stranded DNA Viruses in Offshore Upper Bathyal Sediment from Tohoku, Japan. *Frontiers in Microbiology, 9*, 75.

Yoshida, M., Takaki, Y., Eitoku, M., Nunoura, T., & Takai, K. (2013). Metagenomic analysis of viral communities in (hado)pelagic sediments. *PLoS One, 8*(2), e57271. doi:10.1371/journal.pone.0057271

APPENDIX

Appendix A: CheckM output to assess the quality of the metagenomes. Marker lineage identifies the possible taxonomic level (*e.g.*, C for class and F for family). Strain heterogeneity indicates the proportion of the contamination that appears to be from the same or similar strains.

Bin ID	Marker Lineage	# genomes	# markers	# marker sets	Completeness	Contamination	Strain heterogeneity
6bin2	C_Gammaproteobacteria	119	544	284	98.61	3.70	18.75
6bin9	F__Rhodobacteraceae	67	615	329	87.46	2.33	84.62
6bin8	K_Bacteria	924	151	101	85.54	20.11	75.76
6bin6	K_Bacteria	5449	104	58	43.97	3.45	100.00
6bin4	K_Bacteria	138	338	246	26.99	0.41	100.00
6bin7	K_Archaea	207	149	107	25.84	6.61	11.11
6bin10	K_Bacteria	924	151	101	18.14	0.00	0.00
6bin5	C__Deltaproteobacteria	83	247	155	14.89	0.77	5.00
6bin3	K_Bacteria	5449	104	58	12.07	0.00	0.00
6bin1	Root	5656	56	24	0.00	0.00	0.00
24bin4	P__Proteobacteria	1495	261	164	98.78	1.22	50.00
24bin15	K_Bacteria	5449	104	58	97.41	204.33	11.55
24bin12	K_Bacteria	2993	147	91	95.10	4.40	0.00
24bin7	K_Archaea	207	149	107	90.65	132.75	17.05
24bin16	K_Bacteria	924	163	110	89.09	3.36	20.00
24bin10	F__Rhodobacteraceae	84	568	330	80.66	0.00	0.00

24bin3	K_Bacteria	924	151	101	77.45	0.99	100.00
24bin6	K_Bacteria	2921	143	88	65.50	10.15	53.85
24bin9	K_Bacteria	5449	104	58	51.72	1.72	0.00
24bin2	K_Bacteria	924	151	101	24.69	2.97	0.00
24bin11	K_Archaea	207	145	103	16.50	0.00	0.00
24bin14	K_Bacteria	2258	188	117	10.26	0.00	0.00
24bin1	K_Archaea	207	149	107	4.98	0.00	0.00
24bin5	K_Bacteria	5449	104	58	4.23	0.00	0.00
24bin8	Root	5656	56	24	0.00	0.00	0.00
24bin13	Root	5656	56	24	0.00	0.00	0.00
35bin8	Root	5656	56	24	95.83	489.76	12.54
35bin4	K_Bacteria	5449	103	57	95.61	205.89	12.01
35bin3	K_Bacteria	2921	143	88	91.41	4.55	0.00
35bin10	K_Bacteria	924	151	101	74.92	0.99	0.00
35bin6	K_Archaea	207	149	107	60.51	9.27	23.33
35bin9	K_Archaea	207	145	103	59.26	11.33	6.67
35bin7	K_Bacteria	5449	104	58	43.97	1.72	100.00
35bin1	P__Euryarchaeota	148	188	125	18.53	0.00	0.00
35bin2	K_Bacteria	5449	104	58	4.62	0.00	0.00
35bin5	Root	5656	56	24	0.00	0.00	0.00
68bin13	K_Archaea	207	149	107	99.69	2.41	0.00

68bin7	C_Gammaproteobacteria	228	583	244	99.18	1.09	20.00
68bin9	K_Bacteria	5449	104	58	99.14	21.47	36.67
68bin12	Root	5656	56	24	95.83	1283.12	13.10
68bin22	C__Gammaproteobacteria	119	544	284	94.77	2.82	9.09
68bin18	K_Bacteria	5449	103	57	87.64	123.12	7.00
68bin3	K_Archaea	207	149	107	86.84	5.46	31.82
68bin20	K_Bacteria	138	335	244	81.63	1.37	0.0
68bin2	K_Bacteria	924	151	101	72.09	4.07	83.33
68bin16	K_Bacteria	5449	103	57	58.44	33.33	32.00
68bin11	K_Archaea	207	149	107	57.01	12.31	0.00
68bin10	K_Bacteria	5449	104	58	49.48	0.00	0.00
68bin4	K_Bacteria	2921	143	88	40.91	0.67	100.00
68bin15	K_Bacteria	5449	104	58	35.34	5.17	100.00
68bin24	K_Archaea	207	145	103	27.30	3.88	0.00
68bin19	K_Bacteria	5449	104	58	19.75	0.00	0.00
68bin1	K_Archaea	207	149	107	17.45	0.00	0.00
68bin6	K_Bacteria	5449	104	28	17.24	0.00	0.00
68bin5	K_Archaea	207	145	103	9.52	0.00	0.00
68bin17	K_Bacteria	5449	104	58	9.48	5.17	100.00
68bin8	K_Archaea	207	145	103	8.74	0.00	0.00
68bin23	Root	5656	56	24	0.00	0.00	0.00

68bin21	Root	5656	56	24	0.00	0.00	0.00
68bin14	Root	5656	56	24	0.00	0.00	0.00
72bin7	C_Gammaproteobacteria	263	506	232	99.52	1.08	0.00
72bin22	K_Bacteria	138	338	246	98.25	4.07	0.00
72bin24	P__Proteobacteria	1495	261	164	98.17	0.81	0.00
72bin20	C__Gammaproteobacteria	228	583	244	97.94	0.22	50.00
72bin1	F__Rhodobacteraceae	84	568	330	97.73	1.36	40.00
72bin5	C__Gammaproteobacteria	52	693	297	97.64	0.73	0.00
72bin19	C__Gammaproteobacteria	228	583	244	97.53	1.72	50.00
72bin26	C__Gammaproteobacteria	119	544	284	96.67	2.94	7.69
72bin3	Root	5656	56	24	95.83	985.92	11.64
72bin28	C__Gammaproteobacteria	119	544	284	95.51	4.14	41.18
72bin27	P_Proteobacteria	1495	261	164	93.51	4.57	7.69
72bin13	C__Gammaproteobacteria	26	752	296	93.30	3.30	62.50
72bin10	F__Rhodobacteraceae	67	615	329	92.59	1.45	44.44
72bin21	K_Bacteria	138	338	246	90.89	2.44	42.86
72bin14	F_Rhodobacteraceae	67	616	330	90.71	2.94	39.13
72bin6	K_Bacteria	5449	103	57	89.55	136.09	17.70
72bin23	K_Bacteria	5449	104	58	85.47	56.56	56.60
72bin12	K_Bacteria	5449	104	58	56.50	1.72	0.00
72bin18	K_Bacteria	5443	105	59	53.39	0.85	0.00

72bin17	K_Bacteria	924	151	101	49.17	6.71	30.77
72bin4	K_Archaea	207	145	103	45.25	1.94	0.00
72bin32	K_Bacteria	138	338	246	41.84	0.41	50.00
72bin15	K_Bacteria	2921	152	93	32.10	0.10	100.00
72bin31	K_Archaea	207	149	107	26.99	0.93	0.00
72bin30	K_Bacteria	2921	152	93	23.17	0.00	0.00
72bin9	K_Bacteria	5449	104	58	20.69	0.00	0.00
72bin25	K_Archaea	207	149	107	19.37	1.87	0.00
72bin11	Root	5656	56	24	12.50	0.00	0.00
72bin2	K_Bacteria	5449	104	58	10.85	0.00	0.00
72bin16	K_Bacteria	5449	104	58	4.23	0.00	0.00
72bin33	K_Archaea	207	149	107	0.93	0.00	0.00
72bin8	Root	5656	56	24	0.00	0.00	0.00
72bin29	Root	5656	56	24	0.00	0.00	0.00
79bin18	K_Bacteria	5449	103	57	99.12	50.24	9.09
79bin17	C_Gammaproteobacteria	228	583	244	97.92	1.09	20.00
79bin20	K_Bacteria	5449	103	57	97.05	311.82	4.51
79bin6	Root	5656	56	24	95.83	1352.74	11.63
79bin22	K_Bacteria	2921	143	88	91.48	7.10	28.57
79bin8	K_Archaea	207	149	107	90.34	1.94	25.00
79bin16	K_Bacteria	924	163	110	87.73	12.76	34.78

79bin9	K_Bacteria	5449	103	57	86.84	135.37	36.24
79bin12	F__Rhodobacteraceae	67	615	329	86.49	0.53	50.00
79bin11	K_Bacteria	138	338	246	76.00	1.22	0.00
79bin24	F__Rhodobacteraceae	84	568	330	59.40	0.36	66.67
79bin19	P__Euryarchaeota	148	188	125	57.06	2.80	0.00
79bin14	K_Archaea	207	149	107	56.47	6.54	20.00
79bin13	K_Bacteria	138	338	246	52.58	0.41	100.00
79bin2	K_Archaea	207	149	107	33.07	1.87	0.00
79bin21	K_Bacteria	2258	188	117	24.36	1.28	50.00
79bin10	K_Bacteria	5449	104	58	21.55	0.00	0.00
79bin4	K_Archaea	207	145	103	16.50	1.94	0.00
79bin23	K_Bacteria	5449	104	28	12.07	0.00	0.00
79bin1	K_Bacteria	5449	104	28	10.34	0.00	0.00
79bin5	K_Bacteria	5449	104	28	8.62	0.00	0.00
79bin7	K_Bacteria	5449	104	28	5.17	0.00	0.00
79bin3	Root	5656	56	24	0.00	0.00	0.00
79bin15	Root	5656	56	24	0.00	0.00	0.00