Microbiome exploration of deep-sea carnivorous Cladorhizidae sponges

by Joost Theo Petra Verhoeven

A Thesis submitted to the School of Graduate Studies

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biology

Memorial University of Newfoundland

March 2019

St. John's, Newfoundland and Labrador

ABSTRACT

Members of the sponge family Cladorhizidae are unique in having replaced the typical filter-feeding strategy of sponges by a predatory lifestyle, capturing and digesting small prey. These carnivorous sponges are found in many different environments, but are particularly abundant in deep waters, where they constitute a substantial component of the benthos. Sponges are known to host a wide range of microbial associates (microbiome) important for host health, but the extent of the microbiome in carnivorous sponges has never been extensively investigated and their importance is poorly understood. In this thesis, the microbiome of two deep-sea carnivorous sponge species (*Chondrocladia grandis* and *Cladorhiza oxeata*) is investigated for the first time, leveraging recent advances in high-throughput sequencing and through custom developed bioinformatic and molecular methods. Microbiome analyses showed that the carnivorous sponges co-occur with microorganisms and large differences in the composition and type of associations were observed between sponge species. Tissues of *C. grandis* hosted diverse bacterial communities, similar in composition between individuals, in stark contrast to *C. oxeata* where low microbial diversity was found with a high host-to-host variability. In *C. grandis* the microbiome was not homogeneous throughout the host tissue, and significant shifts occured within community members across anatomical regions, with the enrichment of specific bacterial taxa in particular anatomical niches, indicating a potential symbiotic role of such taxa within processes like prey digestion and chemolithoautotrophy. The potential for bacteria-mediated chemolithoautotrophy was further supported by stable-isotope analysis in *C. grandis* but not in *C. oxeata*. Metagenome analysis further showed a potential

role for archaea in nitrogenous-waste detoxification, as well as an assemblage of bacteriophages potentially influencing the bacterial community composition and functionality. Lastly, metabolic pathways of associated microbes were diverse and included metabolism of elemental substrates (nitrogen, sulfur) and methane. This thesis provides novel insights into the role of microorganisms in carnivorous sponges and their relation to carnivory; the microbiome of *C. grandis* is linked to many metabolic functions, including the facilitation of carnivory, whereas heterogeneity and absence of apparent symbiotic potential in *C. oxeata* suggest that in some sponges, other processes govern carnivory.

ACKNOWLEDGMENTS

Looking back, my journey towards this thesis has been quite an extensive one, starting as a student to become a laboratory technician all the way back in 2003 in The Netherlands. Back then, I would have never thought that one day I would be pursuing a PhD degree abroad, and many people along the way have been so supportive and helped me greatly along the path to get there.

I would like to take this opportunity to thank all my previous supervisors and places that offered me internships over the years, including Johnny and Gonda Rottier and all the fine folk at Micro-Analyse Zeeland for putting up with me during my very first internship when I had no idea how a lab worked, Lucy Buyck at the Medical Microbiology Lab Terneuzen, as well as Douglass Turnbull, Robert Taylor and Laura Greaves at the Wellcome Trust Centre for Mitochondrial Research in Newcastle, and Kimberly Benschop at the National Institute for Public Health and the Environment in Bilthoven. A special mention goes out to Lia van der Hoek at the Academic Medical Center for a very formative internship in a great environment (which also included many pubs in Amsterdam!) in which your support and trust in me to manage my own little research project most certainly kindled my confidence to keep learning.

Of course, this thesis could not have come to be if not for Suzanne Dufour, who graciously agreed to be my supervisor. Thank you for giving me a chance to pursue an incredibly interesting topic and learn so many new things, I greatly appreciate all the guidance and opportunities you have given me over the course of my PhD program. This

gratitude also extends to my committee members: Annie Mercier, Lourdes Peña-Castillo and Kapil Tahlan, thank you for all your help and patience over the last few years. I would also like to thank my lab mates for all the nice times in and out of the lab!

A source of never-failing help that needs to be acknowledged is my parents, brother and sister. Moving away from home to Canada has not always been easy, and this PhD would not have been possible if it weren't for your support, continued interest in my work, your visits to Canada, and kindness whenever I had to chance to visit home, thank you so much.

Finally, I would like to thank Marta, you have been at the center of this thesis adventure together with me, and I could not have done it without you. Your continued patience, eagerness to teach me things, discuss science (usually over a pint of beer) and all-around encouragement has been an enormous help, and I am looking forward to many more adventures together.

CO-AUTHORSHIP STATEMENT

I am the primary author of all chapters in this thesis, for which I performed the associated laboratory work, development of bioinformatic scripts and data analysis. With the exception of Chapter 5 (see below) I authored, in completeness, all the text for the presented manuscripts, which then received editorial inputs from my co-authors.

Chapter 2: I wrote the SPONS pipeline software, performed the validation experiments, wrote the thesis chapter and created the supporting material such as figures and tables. Dr. Suzanne Dufour and Dr. Lourdes Peña-Castillo provided editorial inputs which were incorporated into the thesis chapter.

Chapter 3: The samples used in this study were collected by Dr. Dufour. All laboratory experiments were performed by me, with the exception of spicule analysis through electron microscopy, which was performed by Dr. Dufour and Alana Kavanagh. I performed the bioinformatic analyses, wrote the manuscript and created the supporting material such as figures and tables. Editorial inputs from my co-authors, and peer reviewers upon publishing in *FEMS* [1] were incorporated into the chapter.

Chapter 4: I collected the samples, performed laboratory experiments and conducted the bioinformatic analyses. Spicule analysis through electron microscopy was performed by Dr. Dufour and Alana Kavanagh. The manuscript, including supporting materials, was written by me and received editorial inputs, incorporated into the thesis

[1] Verhoeven JTP, Kavanagh AN, Dufour SC. Microbiome analysis shows enrichment for specific bacteria in separate anatomical regions of the deep-sea carnivorous sponge *Chondrocladia grandis*. *FEMS Microbiol Ecol* (2017) 93: doi:10.1093/femsec/fiw214

chapter, from co-authors as well as peer reviewers upon submitting the manuscript to *Arctic Science*[2].

Chapter 5: This manuscript was the result of a combined effort with Dr. Marta Canuti. Dr. Canuti contributed to the conceptualization and design of the molecular methods, which were validated in a series of experiments performed by both Dr. Canuti and myself. I developed the bioinformatics pipeline. The manuscript was co-written; Dr. Canuti wrote the sections related to virus discovery and influenza virus genome sequencing, as well as the molecular methods, while I wrote the introduction, discussion, bioinformatic methods and the sponge microbiome results. The manuscript further received minor editorial inputs from other co-authors, and peer reviewers after submission to the *Canadian Journal of Microbiology*[3].

Chapter 6: All experiments, analysis and manuscript preparation were performed by me. Dr. Dufour provided editorial inputs which were incorporated into the thesis chapter.

---

[2] Verhoeven JTP, Dufour SC. Microbiomes of the Arctic carnivorous sponges *Chondrocladia grandis* and *Cladorhiza oxeata* suggest a specific, but differential involvement of bacterial associates. *Arct Sci* (2017) doi:10.1139/AS-2017-0015.

[3] Verhoeven JTP, Canuti M, Munro HJ, Dufour SC, Lang AS. ViDiT-CACTUS: an inexpensive and versatile library preparation and sequence analysis method for virus discovery and other microbiology applications. *Can J Microbiol* (2018)1–13. doi:10.1139/cjm-2018-0097

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

1/D	Reciprocal Simpson index
16S rRNA	16S (Svedburg) ribosomal RNA
28S rDNA	28S (Svedburg) ribosomal RNA
AIV	Avian influenza A virus
ALG11	Asparagine-linked glycosylation 11 homolog gene
ANOVA	Analysis of variance
Bg	Baffin Bay *Chondrocladia grandis*
BIOM	Biological observation matrix (format)
BLAST	Basic local alignment search tool
Bo	Baffin Bay *Cladorhiza oxeata*
C1-D2	28S ribosomal RNA gene C1-D2 partition
cDNA	Copy DNA
COG	Cluster of orthologous groups
COI	Cytochrome c oxidase subunit I
CPU	Central processing unit
DIAMOND	Double index alignment of next-generation sequencing data
dsDNA	Double-stranded DNA
E	Simpson's equitability measure
EC2	Elastic compute cloud
eggNOG	Evolutionary genealogy of genes: non-supervised orthologous group
FASTA	Fast-all (file containing nucleotide or protein sequences)

FASTQ	Fast-all (file containing nucleotide or protein sequences), with quality information included
FFT-NS-2	Progressive method heuristic in MAFFT
GTR	Generalised time reversible model
HTS	High throughput sequencing
KEGG	Kyoto encyclopedia of genes and genomes
LCA	Lowest common ancestor
LDA	Linear discriminant analysis
LEfSe	Linear discriminant analysis effect size
MAFFT	Multiple sequence alignment fast Fourier transform
MEGAN	Metagenome analyzer
Mg	(Gulf of) Maine *Chondrocladia grandis*
MinPath	Minimal set of pathways
NA	Nucleic acid
NCBI	National center for biotechnology information
NMDS	Non-metric multidimensional scaling
OG	Orthologous group
OTU	Operational taxonomic unit
PAST	Paleontological statistics
PBS	Phosphate-buffered saline
PCoA	Principal coordinates analysis
PCR	Polymerase chain reaction
PD	Phylogenetic distance
PEAR	Paired-end read merger

PGM	Personal genome machine
PHRAP	*Phragment* assembly program
QIIME	Quantitative insights into microbial ecology
R	R programming language
RAXML	Randomized *axelerated* maximum likelihood
RDP	Ribosomal database project
RNA	Ribonucleic acid
ROV	Remote operate vehicle
RT-PCR	Reverse-transcriptase polymerase chain reaction
SEM	Scanning electron microscopy
SPONS	Streamline processor of next-generation sequences
ssDNA	Single stranded DNA
SSU	Small subunit ribosomal ribonucleic acid
SUMO	Super mohawk II ROV
TAP	Taxonomy and phylogeny
UniVec	Database of sequences of potential vector origin
UP	Unified processor
V4	Hypervariable region 4 of the 16S rRNA gene
V6	Hypervariable region 6 of the 16S rRNA gene
V7	Hypervariable region 7 of the 16S rRNA gene
V8	Hypervariable region 8 of the 16S rRNA gene
ViDiT	Virus discovery Ion Torrent
VSEARCH	Vectorized search
VST	Variance stabilizing transformation

CHAPTER 1: INTRODUCTION AND OVERVIEW

1.1 Introduction

1.1.1 Overview of the phylum Porifera

Sponges (phylum Porifera) are sessile, multicellular, aquatic organisms (1). Members of the Porifera are among the oldest multicellular organisms, thought to have arisen as early as 600 million years ago (1,2). Today, sponges are found as benthic dwellers of both marine and fresh water environments, across a variety of temperate, tropical and polar habitats, a testament to their ecological and evolutionary success (3).

The majority of sponges are filter-feeders (3), fulfilling their energetic requirements by filtering out particles and ingesting nutrients from the surrounding water column through a highly efficient aquiferous system (4). Briefly, pores (ostia) on the outer surface of the sponge (pinacoderm) lead to a complex network of interior canals and chambers lined with flagellated cells (choanocytes) that create a current, which draws water in through the ostia (4). In addition to generating currents, choanocytes generally also mediate the transport of food particles and dissolved organic matter from the inhalant water to the inner connective tissue (mesohyl) of the sponge, where phagocytosis and downstream metabolism occurs (4,5). After transportation through the aquiferous system, filtered water is eventually expelled through an exhalant opening (osculum).

Sponges are a significant component of benthic communities in many marine environments, often present in high numbers and biomass, and are known to play important roles in ecosystem functions (6). These roles can be niche specific; for example, in coral

reef systems, sponges can directly influence the consolidated minerals of lithic substrates through bioerosion, and play a key role in substrate stabilization and consolidation leading to increased coralline survival rates (7,8). On a larger scale, sponges play a key role in benthic-pelagic coupling, i.e.: the exchange of energy, mass, or nutrients between benthic and pelagic habitats (9). By filtering vast amounts of water, sponges effectively remove carbon and other nutrients (such as nitrogen, oxygen and silicon) from the water column, directly influencing the pelagic environment, and driving nutrient flow (10). Other ecosystem roles played by sponges are more intimate and consist of direct interactions with other community members (6). For example, individual sponge morphology as well as aggregates of sponges (sponge grounds) provide (micro) habitats utilized by other organisms (6). This can include the provision of a substrate to physically attach to (6), as well as shelter (11), with some organisms even residing within the aquiferous system of sponges to evade predation (12,13).

1.1.2 Sponge-associated microbial communities

The evolutionary success of sponges, evident by their global distribution, entrenchment in many ecosystems, and general survival since ancient times, is likely, in part, due to their characteristic of associating with other organisms (14). As described previously, such associations can be on a macro-organismal scale, but sponges are especially known for hosting dense, diverse and highly specific microbial communities (15–17). This collection of bacteria, archaea, microbial eukaryotes and viruses can be referred to as the sponge microbiome (18).

Sponge-associated microbial communities can comprise as much as 40% of the total sponge biomass and microbial cells are often found in high densities, up to 10^9 cells/cm^3 within the mesohyl, residing within specialized bacteriocyte cells in some species, and aggregated below the pinacoderm (19,20). Describing the composition and diversity of the sponge microbiome has been the focal point of many studies, as it is thought to play crucial roles in sponge physiology and ecology (21).

Microbiome analyses of different sponge species has revealed an incredible biological diversity, comprised of microorganisms from over 50 distinct phyla and candidate phyla (20). While the phylum level biodiversity found in sponges is extremely high, recent reviews have shown that, overall, the most abundant taxa represented in sponge microbiome encompass the Gamma and Alphaproteobacteria*,* Actinobacteria*,* Chloroflexi*,* Cyanobacteria and Crenarchaeota (16).

While the composition and richness of sponge-associated microbial communities can vary widely between species, intra-species variation is generally low, with microbial communities being remarkably specific and stable over large geographical distances and environmental conditions (20). Relatively little is known about the evolutionary, ecological and physiological processes that govern this specificity of microbiome diversity and composition, although recent work has suggested that most sponges acquire their microbiomes through a mixed mode of transmission which convergently evolved independently across the Porifera phylum (16,22). In the mixed mode of transmission model, it is thought that part of the microbial associates are passed on from parent to

offspring through vertical transmission (19). This vertically transmitted component is most likely supplemented through horizontal transmission, where additional microorganisms are acquired from the environment (19). This ratio between vertical and horizontal transmission is likely to vary across host species, with some hosts exerting a greater control over symbiont communities by favoring vertical transmission (19).

It is expected that many sponges are mixotrophs (20) and, while the retention of specific microorganisms in sponges has often been interpreted to reflect symbiotic relationships involved in supporting mixotrophy, careful consideration is needed to label them as such. Since relatively little is known on the physiological interactions between microbiome members and the host sponge, the type of relationship is often unclear and can range from mutualism and commensalism to parasitism and pathogenicity (14). However, numerous associations of apparent benefit for the host sponge have been described, including nutritional interactions involving the acquisition of major, biologically relevant chemical elements such as carbon and nitrogen (20). For example, in many tropical sponges, carbon metabolism by sponges is heavily dependent on their associated photosynthetic bacterial community, as the translocation of cyanobacterial photosynthates to host cells provides the majority of the carbon required (14,23). Likewise, a diverse set of bacteria and archaea have been linked to nitrogen fixation as well as the breakdown of nitrogenous waste. Metabolic functions conferred by these microorganisms are providing the host sponge with a source of bioavailable nitrogen critical for host health, as well as an effective pathway to handle the build-up of toxic nitrogenous waste metabolites (4,24,25).

Similar associations have been observed for other processes, including methane oxidation, sulfate reduction, dehalogenation and the production of vitamins (14,20).

Sponges are also well known for the production of a wide range of chemical compounds, which are not directly involved in primary metabolic pathways, referred to as natural products or secondary metabolites (26). Marine sponges are the richest known sources of secondary metabolites with strong biological effects (27), and have been the source of many pharmaceutical compounds showing antitumor, antibacterial as well as antiviral properties (28). Within the native sponge environment, these compounds have been implicated in a variety of ecological functions, including the deterrence of benthic predators, prevention of overgrowth, protection against biofouling and antimicrobial activity (22,26,29). It has long been speculated that a large proportion of these compounds are likely produced by sponge-associated microorganisms (14,20), and recent work has indeed shown a rich variety of specialized, secondary metabolism biosynthetic gene clusters in sponge-associated bacteria (30,31).

In addition to bacteria and archaea, the filter-feeding activity of sponges is likely to bring them into contact with a wide variety of viruses, highly abundant in aquatic ecosystems. While virus-like particles have been observed in sponges as early as 1977 (32), to date almost nothing is known about their infectious potential towards the host sponge or their functional roles within the microbiome (20). However, within the context of the sponge microbiome, it has been speculated that phage-mediated horizontal gene transfer might play a role in the distribution of core functions across associated microorganisms

(20). In addition, phages could potentially shape the composition of the sponge microbiome by regulating the abundance of specific microbial members through infection and subsequent lytic action (14,33).

It is clear that a wide variety of interactions exists within the sponge microbiome. Bacteria, archaea and viruses all appear to play important roles, critical in shaping the microbiome that the host sponge appears to need for survival. Indeed, recent studies showing a decline in sponge health upon microbiome disruption further support this central role of the microbiome in sponge physiology and function (34,35).

1.1.3 The carnivorous sponge family Cladorhizidae

Historically, filter-feeding was considered a defining feature of all members within the phylum Porifera (36). More recently however, this definition has become obsolete after the discovery of the carnivorous sponge family Cladorhizidae (37). Carnivorous sponges are unique among the Porifera as they consume small invertebrates (presumed to include the crustacean taxa Copepoda, Isopoda and Ostracod, as well as small polychaete worms) as their main source of nutrients (38).

Carnivory was suspected as early as 1872, when G.O. Sars described a sponge presumed to be non-filter feeding: "*By means of these innumerable microscopic 'claws' which project everywhere from the surface of the sponge, all the more minute animals and the light floating particles which come into immediate contact with the sponge, become attached to it, and thus probably fulfill an essential condition for its nourishment"* and *"As there appears here no trace of any pores interior cavity or oscula, the alimentation must*

be supposed to be effected by a process different from that which takes place where these apparatus exist." (39). This assumption was only verified in 1995 by the direct observation of carnivory in *Lycopodina hypogea* (37).

The carnivorous lifestyle is often characterized as an evolutionary adaption to overcome the inefficient nature of filter-feeding in deep and oligotrophic environments where cladorhizids are often found (38,40). However, whereas other sponges that reside in oligotrophic habitats generally evolved highly optimized current-generating and filtering capabilities to offset a lack of nutrients, the carnivorous sponges have no or severely diminished aquiferous systems, and perform no filter feeding (38,41). A comprehensive study of the molecular phylogeny of the carnivorous sponges has shown that the family Cladorhizidae is monophyletic, most closely related to the filter-feeding families Mycalidae and Guitarridae (41). This monophyly suggests that the radical change to carnivory has evolved only once within the phylum Porifera. The carnivorous genus *Chondrocladia* retains a partial aquiferous system, which led to the speculation that these sponges could represent an intermediate between filter-feeding and carnivory. However, phylogenetic evidence indicates that this is not the case and *Chondrocladia* species are nested within the Cladorhizidae, suggestive of an independent loss of the aquiferous system in other know cladorhizid lineages (41).

In general, cladorhizid species are considered deep-sea sponges. While certain species have been observed in shallow waters, such individuals usually live in cryptic environments like oligotrophic marine caves (40). At depths of 400 meters or more, the

Cladorhizidae become more abundant and more dominant members of the total sponge fauna, increasing in abundance to bathyal, abyssal and hadal depths where they represent one of the dominant sponge groups (40,42). Our current understanding of the global biogeography of the Cladorhizidae is still rudimentary, although a largely bipolar cold-water distribution has been suggested, and a high number of species have been reported within northeast Atlantic and Arctic regions (42). Although these sponges are highly abundant in deep waters, across a large geographical area, concrete knowledge on the role of carnivorous sponges within their respective ecosystems is not known.

In the absence of filter-feeding capabilities, the cladorhizids rely on specialized biological adaptation to allow for the capture and digestion of their prey (41). While there is a paucity of knowledge regarding the feeding behavior of carnivorous sponges, with (to my knowledge) only *L. hypogea* studied *in vitro*, several intriguing biological processes have been documented, which are likely conserved across the Cladorhizidae (38,43). For example, in *L. hypogea,* outward facing filaments on the pinacoderm are covered by microsclere spicules and are thought to be pivotal in prey capture, as the highly organized arrangement of hook-like spicules creates a surface in which bristle-like setae of crustacean prey are efficiently trapped (38). Upon capturing prey, filaments widen and cover the prey, sometimes with the recruitment of additional filaments, and contract. Migrating sponge cells then create a veil over the closely associated prey, forming a cavity (38,43,44). Following capture and envelopment, the original pinacoderm cell layer quickly dedifferentiates (within several hours), leaving the prey in direct contact with the internal sponge mesohyl (38). After ingestion, motile sponge cells (totipotent archeocytes and

bacteriocytes) aggregate in large number around the ingested prey, creating a cyst-like structure in what essentially represents a temporary digestive cavity. Within this cavity, prey degradation occurs over the course of several days with sponge cells disrupting and infiltrating the prey, followed by phagocytosis of prey debris by archeocytes and bacteriocytes (38).

Like their filter-feeding contemporaries, carnivorous sponges can have high numbers of bacterial cells within their tissues, suggesting a close association between sponge and microbial partners (45). It is likely that the microbiome of carnivorous sponges plays symbiotic roles in processes important for host function, possibly extending to involvement in prey digestion and nutrient acquisition. Indeed, feeding experiments in *Lycopodina hypogea* show that upon prey ingestion, an unusually large number of mesohyl-associated bacteria aggregate in and around the temporary digestive cyst (38).

In addition, prey debris phagocytized by sponge bacteriocytes are often intracellularly combined with phagosomes containing bacterial cells which, over the course of digestion, increase in density and are more frequently dividing (38). Another example of microbial interdependence was documented within *Cladorhiza methanophila*, a methanotrophic carnivorous sponge that relies on chemolithoautotrophic symbionts for part of its nutritional demand (46,47).

Overall, it is likely that carnivorous sponges depend on the microbial community for host health and survival, however, relatively little is known about their microbiome compared to filter-feeding sponges. At the start of this PhD project, a single study by

Dupont *et al*. on the shallow-water, cave-dwelling *Lycopodina hypogea* was available, offering the first insight into the carnivorous sponge microbiome (45). In *L. hypogea*, a microbiome similar to filter-feeders was observed, and high relative abundances of ammonium-oxidizing archaea and sulfate-reducing bacteria were noted (45). In a follow-up study by Dupont *et al.*, in which 67 *L. hypogea*-associated bacteria were cultured and a number of strains were identified with antimicrobial and antioxidant activities, it was suggested that microbiome-based secondary metabolite production is common in carnivorous sponges (48).

1.1.4 Advances and challenges in molecular microbiology

Advances in our understanding of the sponge microbiome, and indeed microbial ecology as a whole, have in large part been due to progress in high-throughput sequencing (HTS) technologies and associated analysis software (49). In short, HTS methods allow for fast and parallel sequencing, in contrast to traditional Sanger sequencing, in which the nucleotide sequence of a single nucleic-acid fragment is analyzed (50).

HTS is used for many applications across different fields, including complete genome characterization, mapping of regulatory information, characterization of the cell transcriptome, disease surveillance and cancer typing (50). However, in microbial ecology, HTS has paved the way for transitioning from traditional culture based methods to molecular investigations, allowing extensive high-resolution cataloguing of microbiomes in a wide variety of niches (49). Such studies are often performed through the analyses of conserved marker-genes, such as the ribosomal 16S rRNA gene, which can be used to infer

the composition, diversity and the relationship between microbiomes. In addition, more comprehensive "shotgun sequencing" is often used to indiscriminately amplify all genomic material found in a microbial community. Data from such experiments is extremely valuable as, in addition to taxonomy, it allows for the reconstruction of complete microbial genomes from which organism function and lifestyle can be inferred (50).

More recently, advances in HTS chemistry and hardware have driven a significant drop in cost while simultaneously increasing the throughput capabilities of HTS machines, with some devices sequencing up to hundreds of gigabases per run (51). However, this decrease in cost, coupled with the speed, affordability and volume in which data can now be generated has created new bioinformatic challenges (52). The computational infrastructure needed to effectively analyze high-dimensional datasets has not seen as dramatic of a decrease in cost as HTS experiments, often creating a major bottleneck in HTS experiments, whereby data is generated faster then we are able to analyze it (51,52). The development and implementation of novel computational algorithms, efficient data-analysis pipelines, and the use of de-centralized cloud-based computing platforms have been suggested as potential ways to mitigate the "big-data" problem (52,53).

1.2 Thesis objectives and overview

In oligotrophic and deep waters, carnivorous sponges often make large contributions to the benthic biomass. Little is known on the physiology and environmental role of these sponges, however, extrapolating from knowledge on filter-feeding sponges and the multitude of crucial roles they play in their habitats, it seems likely that the

cladorhizids hold a similar importance within their respective ecosystem. However, studying the Cladorhizidae is challenging due to their tendency to live in extremely deep waters, which can make collection and *in vivo* observation of the host sponge difficult.

In this thesis, the deep-sea carnivorous sponge species *Chondrocladia grandis* and *Cladorhiza oxeata* were investigated. The microbiome of both sponges was characterized in-depth by leveraging recent advances in sequencing technology as well as through the development and use of novel molecular and bioinformatic tools. The main research objectives set out in this thesis were 1) to describe the overall diversity and stability of the *C. grandis* and *C. oxeata* microbiomes, 2) detect and investigate potential symbiotic partners that may be important for carnivory and overall host health, and 3) predict ecosystem roles of carnivorous sponges through the examination of functional roles found in sponge associated microorganisms. An additional objective was to develop molecular and bioinformatic tools to efficiently process and mine the data obtained while addressing the research objectives.

In Chapter 2 the SPONS pipeline is described. This bioinformatic tool was designed to analyze data from 16S rRNA gene sequencing experiments, with a specific focus on throughput. This pipeline was subsequently used to process and analyze data presented in Chapter 3 and Chapter 4.

Chapter 3 describes the microbiome of *Chondrocladia grandis* sampled in the Gulf of Maine. Four specimens are examined, and the overall bacterial community composition and diversity is characterized in three distinct anatomical regions with predicted

involvement in prey capture (sphere), support (axis) and benthic substrate attachment (root).

Building upon the data presented in Chapter 3, an in-depth analysis of bacterial associates of *Chondrocladia grandis* and *Cladorhiza oxeata* collected in the Canadian Arctic is presented in Chapter 4. The overall microbiome as well as potential symbionts involved in carnivory are described. Furthermore, oligotype characterization and stable isotope analysis were performed to elucidate the geographical and host specificity of the microbiome and to investigate the potential for chemolithoautotrophic symbiosis.

In Chapter 5, the ViDiT method is described. ViDiT is a customizable, PCR-based and versatile library preparation method, designed to be extremely cost efficient to facilitate metagenomic experiments. In addition, the "CACTUS" software package is also presented, a complete analysis pipeline designed to rely on cloud computing power to generate high-quality data from ViDiT-based experiments without the need of expensive servers.

Chapter 6 shows the versatility of the ViDiT-CACTUS method, where it is applied together with HTS to describe the metagenome of *C. grandis*. The bacterial, archaeal and viral communities are described and compared to the marker-gene studies performed in Chapter 3 and Chapter 4. Furthermore, through metabolic pathway prediction, a putative functional profile of the microbiome of *C. grandis* was created, granting a first insight into the potential ecological roles of carnivorous sponges.

1.3 References

1. Wörheide G, Dohrmann M, Erpenbeck D, Larroux C, Maldonado M, Voigt O, Borchiellini C, Lavrov DV. "Deep phylogeny and evolution of sponges (Phylum Porifera)," in *Advances in marine biology* Advances in sponge science: phylogeny, systematics, ecology., eds. M. A. Becerro, M. J. Uriz, M. Maldonado, X. Turon (Academic Press), 1–78. doi:10.1016/B978-0-12-387787-1.00007-6

2. Li C-W, Chen J-Y, Hua T-E. Precambrian sponges with cellular structures. *Science* (1998) **279**:879–882. doi:10.1126/science.279.5352.879

3. Van Soest RWM, Boury-Esnault N, Vacelet J, Dohrmann M, Erpenbeck D, De Voogd NJ, Santodomingo N, Vanhoorne B, Kelly M, Hooper JNA. Global diversity of sponges (Porifera). *PLoS One* (2012) **7**: doi:10.1371/journal.pone.0035105

4. Leys SP, Hill A. The physiology and molecular biology of sponge tissues. *Adv Mar Biol* (2012) **62**:1–56. doi:10.1016/B978-0-12-394283-8.00001-1

5. Goeij JM de, Berg H van den, Oostveen MM van, Epping EHG, Duyl FC van. Major bulk dissolved organic carbon (DOC) removal by encrusting coral reef cavity sponges. *Marine Ecology Progress Series* (2008) **357**:139–151. doi:10.3354/meps07403

6. Bell JJ. The functional roles of marine sponges. *Estuarine, Coastal and Shelf Science* (2008) **79**:341–353. doi:10.1016/j.ecss.2008.05.002

7. Wulff JL. Sponge-mediated coral reef growth and rejuvenation. *Coral Reefs* (1984) **3**:157–163. doi:10.1007/BF00301960

8. Zundelevich A, Lazar B, Ilan M. Chemical versus mechanical bioerosion of coral reefs by boring sponges - lessons from Pione cf. vastifica. *Journal of Experimental Biology* (2007) **210**:91–96. doi:10.1242/jeb.02627

9. Griffiths JR, Kadin M, Nascimento FJA, Tamelander T, Törnroos A, Bonaglia S, Bonsdorff E, Brüchert V, Gårdmark A, Järnström M, et al. The importance of benthic-pelagic coupling for marine ecosystem functioning in a changing world. *Glob Chang Biol* (2017) **23**:2179–2196. doi:10.1111/gcb.13642

10. Goeij JM de, Oevelen D van, Vermeij MJA, Osinga R, Middelburg JJ, Goeij AFPM de, Admiraal W. Surviving in a Marine Desert: The sponge loop retains resources within coral reefs. *Science* (2013) **342**:108–110. doi:10.1126/science.1241981

11. Beazley LI, Kenchington EL, Murillo FJ, Sacau M del M. Deep-sea sponge grounds enhance diversity and abundance of epibenthic megafauna in the Northwest Atlantic. *ICES J Mar Sci* (2013) **70**:1471–1490. doi:10.1093/icesjms/fst124

12. Fiore CL, Jutte PC. Characterization of macrofaunal assemblages associated with sponges and tunicates collected off the southeastern United States. *Invertebrate Biology* (2010) **129**:105–120.

13. Pavloudi C, Christodoulou M, Mavidis M. Macrofaunal assemblages associated with the sponge *Sarcotragus foetidus* Schmidt, 1862 (Porifera: Demospongiae) at the coasts of Cyprus and Greece. *Biodivers Data J* (2016) doi:10.3897/BDJ.4.e8210

14. Taylor MW, Radax R, Steger D, Wagner M. Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiol Mol Biol Rev* (2007) **71**:295–347. doi:10.1128/MMBR.00040-06

15. Schmitt S, Tsai P, Bell J, Fromont J, Ilan M, Lindquist N, Perez T, Rodrigo A, Schupp PJ, Vacelet J, et al. Assessing the complex sponge microbiota: core, variable and species-specific bacterial communities in marine sponges. *ISME J* (2012) **6**:564–576. doi:10.1038/ismej.2011.116

16. Thomas T, Moitinho-Silva L, Lurgi M, Björk JR, Easson C, Astudillo-García C, Olson JB, Erwin PM, López-Legentil S, Luter H, et al. Diversity, structure and convergent evolution of the global sponge microbiome. *Nat Commun* (2016) **7**: doi:10.1038/ncomms11870

17. Webster NS, Taylor MW, Behnam F, Lücker S, Rattei T, Whalan S, Horn M, Wagner M. Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts. *Environ Microbiol* (2010) **12**:2070–2082. doi:10.1111/j.1462-2920.2009.02065.x

18. Marchesi JR, Ravel J. The vocabulary of microbiome research: a proposal. *Microbiome* (2015) **3**: doi:10.1186/s40168-015-0094-5

19. Thacker RW, Freeman CJ. "Sponge–microbe symbioses: recent advances and new directions," in *Advances in marine biology* Advances in sponge science: physiology, chemical and microbial diversity, Biotechnology., eds. M. A. Becerro, M. J. Uriz, M. Maldonado, X. Turon (Academic Press), 57–111. doi:10.1016/B978-0-12-394283-8.00002-3

20. Webster NS, Thomas T. The sponge hologenome. *mBio* (2016) **7**:e00135-16. doi:10.1128/mBio.00135-16

21. Moitinho-Silva L, Nielsen S, Amir A, Gonzalez A, Ackermann GL, Cerrano C, Astudillo-Garcia C, Easson C, Sipkema D, Liu F, et al. The sponge microbiome project. *Gigascience* (2017) **6**:1–7. doi:10.1093/gigascience/gix077

22. Fan L, Reynolds D, Liu M, Stark M, Kjelleberg S, Webster NS, Thomas T. Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. *Proc Natl Acad Sci U S A* (2012) **109**:E1878–E1887. doi:10.1073/pnas.1203287109

23. Wilkinson CR. Net primary productivity in coral reef sponges. *Science* (1983) **219**:410–412. doi:10.1126/science.219.4583.410

24. Steger D, Ettinger-Epstein P, Whalan S, Hentschel U, de Nys R, Wagner M, Taylor MW. Diversity and mode of transmission of ammonia-oxidizing archaea in marine sponges. *Environ Microbiol* (2008) **10**:1087–1094. doi:10.1111/j.1462-2920.2007.01515.x

25. Wilkinson CR, Fay P. Nitrogen fixation in coral reef sponges with symbiotic cyanobacteria. *Nature* (1979) **279**:527–529. doi:10.1038/279527a0

26. Pawlik JR. The Chemical Ecology of sponges on caribbean reefs: natural products shape natural systems. *BioScience* (2011) **61**:888–898. doi:10.1525/bio.2011.61.11.8

27. Blunt JW, Copp BR, Munro MHG, Northcote PT, Prinsep MR. Marine natural products. *Nat Prod Rep* (2011) **28**:196–268. doi:10.1039/c005001f

28. Müller WE, Böhm M, Batel R, De Rosa S, Tommonaro G, Müller IM, Schröder HC. Application of cell culture for the production of bioactive compounds from sponges: synthesis of avarol by primmorphs from *Dysidea avara*. *J Nat Prod* (2000) **63**:1077–1081.

29. Thomas TRA, Kavlekar DP, LokaBharathi PA. Marine drugs from sponge-microbe association—A Review. *Marine Drugs* (2010) **8**:1417–1468. doi:10.3390/md8041417

30. Jackson SA, Crossman L, Almeida EL, Margassery LM, Kennedy J, Dobson ADW. Diverse and abundant secondary metabolism biosynthetic gene clusters in the genomes of marine sponge derived *Streptomyces* spp. isolates. *Marine Drugs* (2018) **16**:67. doi:10.3390/md16020067

31. Mori T, Cahn JKB, Wilson MC, Meoded RA, Wiebach V, Martinez AFC, Helfrich EJN, Albersmeier A, Wibberg D, Dätwyler S, et al. Single-bacterial genomics validates rich and varied specialized metabolism of uncultivated *Entotheonella* sponge symbionts. *PNAS* (2018) **115**:1718–1723. doi:10.1073/pnas.1715496115

32. Vacelet J, Donadey C. Electron microscope study of the association between some sponges and bacteria. *Journal of Experimental Marine Biology and Ecology* (1977) **30**:301–314. doi:10.1016/0022-0981(77)90038-7

33. Lohr JE, Chen F, Hill RT. Genomic analysis of bacteriophage ΦJL001: insights into its interaction with a sponge-associated alpha-proteobacterium. *Appl Environ Microbiol* (2005) **71**:1598–1609. doi:10.1128/AEM.71.3.1598-1609.2005

34. Fan L, Liu M, Simister R, Webster NS, Thomas T. Marine microbial symbiosis heats up: the phylogenetic and functional response of a sponge holobiont to thermal stress. *ISME J* (2013) **7**:991–1002. doi:10.1038/ismej.2012.165

35. Ramsby BD, Hoogenboom MO, Whalan S, Webster NS. Elevated seawater temperature disrupts the microbiome of an ecologically important bioeroding sponge. *Molecular Ecology* (2018) **27**:2124–2137. doi:10.1111/mec.14544

36. Patricia R. Bergquist. *Sponges*. Berkeley: University of California Press (1978).

37. Vacelet J, Boury-Esnault N. Carnivorous sponges. *Nature; London* (1995) **373**:333.

38. Vacelet J, Duport E. Prey capture and digestion in the carnivorous sponge *Asbestopluma hypogea* (Porifera: Demospongiae). *Zoomorphology* (2004) **123**:179–190. doi:10.1007/s00435-004-0100-0

39. Sars GO (Georg O, Sars M (Michael). *On some remarkable forms of animal life from the great deeps off the Norwegian coast.*. Christiania, Printed by Brøgger & Christie [etc.] (1872).

40. Hestetun JT, Tompkins-Macdonald G, Rapp HT. A review of carnivorous sponges (Porifera: Cladorhizidae) from the Boreal North Atlantic and Arctic. *Zool J Linn Soc* (2017) **181**:1–69. doi:10.1093/zoolinnean/zlw022

41. Hestetun JT, Vacelet J, Boury-Esnault N, Borchiellini C, Kelly M, Ríos P, Cristobo J, Rapp HT. The systematics of carnivorous sponges. *Molecular Phylogenetics and Evolution* (2016) **94**:327–345. doi:10.1016/j.ympev.2015.08.022

42. Hestetun JT, Fourt M, Vacelet J, Boury-Esnault N, Rapp HT. Cladorhizidae (Porifera, Demospongiae, Poecilosclerida) of the deep Atlantic collected during Ifremer cruises, with a biogeographic overview of the Atlantic species. *Journal of the Marine Biological Association of the United Kingdom* (2015) **95**:1311–1342. doi:10.1017/S0025315413001100

43. Lee WL, Reiswig HM, Austin WC, Lundsten L. An extraordinary new carnivorous sponge, *Chondrocladia lyra*, in the new subgenus Symmetrocladia (Demospongiae, Cladorhizidae), from off of northern California, USA. *Invertebrate Biology* (2012) **131**:259–284.

44. Vacelet J. New carnivorous sponges (Porifera, Poecilosclerida) collected from manned submersibles in the deep Pacific. *Zool J Linn Soc* (2006) **148**:553–584. doi:10.1111/j.1096-3642.2006.00234.x

45. Dupont S, Corre E, Li Y, Vacelet J, Bourguet-Kondracki M-L. First insights into the microbiome of a carnivorous sponge. *FEMS Microbiol Ecol* (2013) **86**:520–531. doi:10.1111/1574-6941.12178

46. Vacelet J, Boury-Esnault N, Fiala-Medioni A, Fisher CR. A methanotrophic carnivorous sponge. *Nature* (1995) **377**:296. doi:10.1038/377296a0

47. Hestetun JT, Dahle H, Jørgensen SL, Olsen BR, Rapp HT. The microbiome and occurrence of methanotrophy in carnivorous sponges. *Front Microbiol* (2016) **7**: doi:10.3389/fmicb.2016.01781

48. Dupont S, Carre-Mlouka A, Domart-Coulon I, Vacelet J, Bourguet-Kondracki M-L. Exploring cultivable bacteria from the prokaryotic community associated with the carnivorous sponge *Asbestopluma hypogea*. *FEMS Microbiol Ecol* (2014) **88**:160–174. doi:10.1111/1574-6941.12279

49. Boughner LA, Singh P. Microbial ecology: where are we now? *Postdoc J* (2016) **4**:3–17. doi:10.14304/SURYA.JPR.V4N11.2

50. Reuter JA, Spacek D, Snyder MP. High-throughput sequencing technologies. *Mol Cell* (2015) **58**:586–597. doi:10.1016/j.molcel.2015.05.004

51. D'Argenio V. The high-throughput analyses era: are we ready for the data struggle? *High-Throughput* (2018) **7**:8. doi:10.3390/ht7010008

52. Sboner A, Mu XJ, Greenbaum D, Auerbach RK, Gerstein MB. The real cost of sequencing: higher than you think! *Genome Biology* (2011) **12**:125. doi:10.1186/gb-2011-12-8-125

53. Merelli I, Pérez-Sánchez H, Gesing S, D'Agostino D. Managing, analysing, and integrating big data in medical bioinformatics: open problems and future perspectives. *BioMed Research International* (2014) doi:10.1155/2014/134023

CHAPTER 2: SPONS, A SLIM PYTHON-BASED PIPELINE FOR ULTRA-FAST ANALYSIS OF 16S RRNA GENE AMPLICONS.

2.1 Abstract

High-throughput sequencing of 16S ribosomal RNA (rRNA) genes is used in many studies to analyze bacterial community composition from a wide variety of niches. As sequencing methods continue to produce larger datasets and as the size of studies increases, computational requirements remain a significant bottleneck. Here, we present the SPONS (Streamlined Processor of Next-gen Sequences) pipeline, a Python based wrapper script that provides a complete workflow for the creation of quality-checked, chimera free OTUs with associated taxonomic assignment and phylogeny from raw high-throughput sequences. SPONS is designed to be heavily optimized, using a variety of technologies that leverage modern computer-processor instructions and multi-threading, significantly reducing the execution time of analyses. We analyzed four mock communities to validate the accuracy and speed of the presented pipeline. Results from these analyses show that SPONS is capable of accurately characterizing bacterial communities, recovering near complete diversity estimates, in most cases up to genus level, with minimal deviation from known abundances. As such, the SPONS pipeline makes a useful addition to the 16S rRNA analysis toolkit, especially for users wishing to analyze large datasets efficiently.

2.2 Introduction

Over the last decade, high-throughput sequencing (HTS) has seen massive incremental improvements in both chemistry and hardware (1). Newer HTS platforms

(such as Illumina HiSeq) have resulted in a significant increase in the amount of sequence data (or throughput) provided by a single run, accompanied by a decrease in the per-base cost (1). These improvements and cost decreases have in turn brought HTS experiments within the reach of many laboratories (2), including those studying microbial ecology. Indeed, HTS technologies have been crucial in revealing many aspects of microbial diversity and symbiotic interactions within ecosystems (3).

On a gross scale, HTS experiments can be divided into two categories: marker-gene analysis and shotgun metagenomics. Marker-gene analysis typically uses specific genomic regions (or markers) as a target for amplification, with generated sequences then compared to a known set of reference sequences *in-silico* to infer a taxonomy (4). One of the most commonly used markers has been the 16S rRNA gene for the identification of bacteria and archaea (5), used in countless studies. Shotgun metagenomic approaches aim to sequence the full genomic content (i.e.: complete or near complete genomes) of all organisms present within a given sample (6). In microbiologic studies, such metagenomic approaches offer a more complete view of the microbial diversity since other organisms, such as viruses, fungi, and protists are also detected and sequenced. However, metagenomic approaches are less cost-effective (7), especially when compared to marker-gene analysis, due to the inherent sequencing depth required to fully characterize complete genomes. As such, marker-gene analyses like 16S rRNA gene sequencing are still highly popular as a valid, cost-effective way of characterizing microbial communities.

Processing and analyzing sequence data stemming from HTS experiments can be difficult and has a steep learning curve, especially for scientists without an in-depth knowledge of bioinformatics. Multiple processes including trimming, read-merging, clustering of sequences into operational taxonomic units (OTU) and taxonomy assignment are needed to transform raw sequences into insightful data, which can be overwhelming. Several platforms have been developed to partly mitigate this problem and make analysis tools more accessible for scientists. Most notable are the QIIME pipeline (8) and the Mothur (9) program which, together with their respective user communities, have brought microbial community analysis to the masses.

However, such frameworks are designed for a broad user group with diverse needs and research questions. Many of the applications, workflows and settings included may not be needed or even appropriate for all users. Furthermore, the centralized development can hinder the timely roll-out of improvements to standard settings, in some cases making it impossible to update existing applications to newer versions and delaying the implementation of novel applications. For intermediate to experienced users who expect to process large amounts of data, there can be a significant benefit in developing custom pipelines or wrapper scripts. Such benefits include a greater control over installed dependencies, the possibility of incorporating the newest applications and application versions, and more importantly the ability to maximize speed, efficiency and accuracy in a way that is tailored to the type of samples being analyzed.

Here, we present SPONS (Streamlined Processor of Next-gen Sequences), a 16S rRNA gene analysis pipeline and wrapper script written in Python that is heavily optimized for speed, but still retains high levels of accuracy.

2.3 Implementation

2.3.1 SPONS architecture overview

The SPONS pipeline (source available on request), written in Python (version 2.7), provides a complete workflow for 16S rRNA processing including read-merging, quality control, primer trimming, OTU creation, taxonomy assignment and full dataset phylogenetic analysis. Two main modules make up the SPONS pipeline; the Unified Processing (UP) module and the Taxonomy And Phylogeny (TAP) module (Figure 2.1). The UP-module acts as the entry point into the pipeline and handles all data preprocessing, quality control, OTU construction and generation of output files, while the TAP module is responsible for OTU taxonomic assignment and phylogeny. While the operational behavior of SPONS is adjustable and expandable by the user, the typical workflow depicted in Figure 2.1, and described in detail below.

2.3.2 Input files

As SPONS was specifically designed for paired-end Illumina based HTS data, the workflow requires unmerged paired-end sequences as input. Sequences are required to be in FASTQ format in order for quality information to be available for downstream quality control steps. In addition to sequence files, the user also supplies a meta-data file, a tab-delimited plain-text file which contains any number of variables (i.e.: sample name, type,

location, source, etc.) associated with each sample. Lastly, a SPONS configuration file has to be supplied in which the user can set various parameters (Table 2.1) affecting the behavior of pipeline components, such as quality control stringency, OTU creation parameters and settings governing taxonomy assignment.

2.3.3 Reference files

SPONS was designed to use the SILVA database as the main taxonomic reference (10). SPONS expects the reference database to be in FASTA format, containing the SILVA taxonomy within the sequence header. While designed to use the SILVA database, other reference databases can be used if they are converted into the SILVA taxonomy format.

2.3.4 Sequence quality control and paired-end assembly

When initiating the SPONS pipeline, unmerged sequences are first checked for the presence of primers defined by the user in the configuration file. CutAdapt version 1.16 (11) is used for the detection and removal of primers, and trimmed FASTQ files are stored for quality control. Optionally, the user can select to discard reads in which no forward or reverse primer were found in order to reduce erroneous reads with a high error-rate probability. Subsequently, primer-free reads are quality checked using Trimmomatic version 0.36 (12), which employs a sliding window and a minimum length filter to remove low quality regions from reads. Read pairs in which either forward or reverse primer do not pass the minimum length filter after trimming are discarded. After quality control, reads that pass the quality checks are merged using PEAR version 0.9.5 (13). To reduce false-positive assemblies, a minimum overlap of 25 bases is required between the sequences of

each pair, and pairs that do not successfully assemble are discarded. Following read merging, paired-end assemblies are quality checked again using a native SPONS filtering algorithm, in which assembled reads with an overall average Phred score below the minimum defined in the configuration file are removed. Furthermore, assemblies containing more than a user-specified number of bases which have a Phred score below a user-defined cutoff are also removed, as random drops in quality are often indicative of low-quality sequences.

2.3.5 OTU creation and chimera detection

Prior to OTU creation, VSEARCH version 2.7.1 (14) is used to transform FASTQ files (containing assembled and quality-controlled sequences) to FASTA format, as well as to calculate the expected number of errors in each sequence and append this information to the FASTA sequence header. Sequences containing ambiguous nucleotides are also removed during this step, as the downstream OTU creation process does not accept sequences containing ambiguities. Subsequently, all FASTA files are dereplicated, a process in which all identical sequences are combined to reduce the total amount of sequences to be processed in downstream steps. The original abundance of each collapsed sequence is recorded to allow for the reconstruction of the true abundance of a given amplicon during result generation. This dereplication is followed by a combination of all individual dereplicated sample assemblies into one repository ("pile-file"), which is then dereplicated again using VSEARCH.

The dereplicated pile-file is then used as input for the SWARM version 2.2.2 (15) based OTU creation process. SWARM is a single-linkage clustering method that creates OTUs by iteratively extending a OTU seed sequence by detecting local links to other amplicons within a pre-defined distance threshold (representing the maximum number of differences between two amplicons). This creates stable clusters reaching their natural limits, with centroids based on local peaks of abundance, and avoids problems associated with using arbitrary global clustering thresholds utilized in traditional clustering methods and prevents input-order dependencies (15,16). Subsequently, any potential chimeric OTU sequences are detected and removed using the VSEARCH implementation of UCHIME *de novo* (17).

2.3.6 Taxonomy and phylogeny

Chimera-filtered, representative OTU sequences are used by the TAP module for taxonomic inference and phylogenetic analysis. For taxonomy classification, SPONS supports two approaches, an alignment-based classification using VSEARCH and an inference-based classification using the Ribosomal Database Project (RDP) naïve Bayesian classifier, release 2.12 (18). The alignment-based approach uses VSEARCH to detect similarities between OTU query sequences and known reference sequences in the user-supplied reference database. When a query sequence shows similarity with more than one reference sequence, the hits are parsed to determine the lowest common ancestor (LCA) within the potential taxonomic assignments. The RDP classification algorithm, which is re-trained using the reference database, can assign taxonomy down to genus level based on bootstrap-like confidence values. A full dataset alignment of OTU

sequences is generated using MAFFT version 7.397 (19), from which poorly aligned or divergent positions are subsequently detected and deleted using GBlocks version 0.91b (20). This filtered alignment is then used as input to generate an approximate maximum-likelihood phylogeny using FastTreeMP version 2.1.10 with the General Time Reversible model and gamma options (21).

2.3.7 Output files

After taxonomy assignment and phylogeny creation, SPONS produces several output files for downstream analyses. This includes a tab-delimited OTU table that can be opened by most spreadsheet software, containing information on all found OTUs, such as the nucleotide sequence, spread (the number of samples a given OTU occurs in), overall abundance and abundance in each individual sample. Similarly, a Biological Observation Matrix (BIOM) file is generated (22), containing the same information, but designed to be machine readable by other applications to facilitate the transfer of data between SPONS and other downstream analysis packages such, as R. Furthermore, the full alignment of all OTU sequences in the dataset, as well as the phylogenetic tree are stored, along with a FASTA file containing all OTU nucleotide sequences. Finally, the configuration file of SPONS can be set to automatically invoke an external R script file for further custom analyses.

2.4 Material and methods

2.4.1 Mock datasets

Validation of SPONS was performed by analyzing four datasets obtained from the mockrobiota resource (23) including mock-12 (24), mock-18 (25), mock-20 (26) and mock-23 (26). The sequencing target for all mock communities was the V4 region of the 16S rRNA gene. For all mock sets, the biological composition and distribution of abundance is known, and this "ground truth" can be used to verify the results obtained by the SPONS pipeline. Of note, mock community 12 is an "extreme" dataset, which includes highly related taxa, among which several are only distinguishable by one nucleotide. Mock 18, 20 and 23 were arbitrarily chosen. Mock-18 consists of an even mixture of cloned 16S rRNA genes in pUC19 plasmid vector from 15 bacterial strains. Mock-20 and 23 are composed of purified genomic DNA from 20 bacterial strains, in even, and uneven (staggered) amounts, respectively. Additional details on the composition of mock datasets is available on the mockrobiota resource (23).

2.4.2 Pipeline settings and validation methods

Validation of the SPONS pipeline was performed by an automation script that, for each mock community, executed the SPONS pipeline using both the RDP and VSEARCH classification methods. Settings for preprocessing and OTU creation were based on previous tests (data not shown) and common to both the RDP and VSEARCH runs. In summary, primer trimming was set such that "discard untrimmed" was disabled, and Trimmomatic based quality control on primer trimmed reads was performed using a sliding window of 4 nucleotides, initiating the read trimming when the Phred quality score of the

window dropped below 10. Read merging with PEAR required a minimal overlap of 25 nucleotides, with a p-value cutoff of < 0.01. The SPONS quality control filter on assembled reads was set to reject any read with an overall average Phred score below 27, and further checked for a maximum allowable number of drops below a Phred score of 15 set to 10, rejecting the read if this threshold was passed. Detection of OTUs with SWARM was performed with the "fastidious" mode in SWARM (maximum distance of reads (*d*) of 1) and the ceiling for fastidious mode set to 100 gigabytes. Chimera removal with VSEARCH was configured with a minimum score setting of 0.28, minimum number of differences in a segment of 10, and minimum divergence from the closest parent at 1.5. For both VSEARCH and RDP based taxonomy assignment, the SILVA small subunit 16S rRNA reference database (release 132) was used (24). Comparison of OTU sequences to the SILVA reference database with VSEARCH was set with a minimum similarity threshold of 97%, and with no limit set on the amount of hits reported. RDP classification was done by retraining the RDP classifier with the SILVA reference database and performing taxonomic inference on the OTU sequences, assigning taxonomy when the minimum posterior likelihood was higher than 0.51. Validation of the MAFFT aligner and FastTree phylogenetic analysis was outside the scope of this study and was not performed.

To evaluate the performance of the SPONS pipeline, we calculated the recall, deviance, unclassified, misclassified and misclassified abundance for each taxonomic level in each mock community for both the VSEARCH and RDP modes. Recall was defined as the percentage of total taxa known to be present in a mock community identified by SPONS. Deviance was based on the average relative abundance deviance when comparing

the known abundance of a taxon within the mock community and the SPONS reported abundance of the same taxon. Unclassified represents the total abundance of reads within a mock community that was not given a taxonomic classification. Misclassified and abundance of misclassified are defined as the number of reported taxa that were not within the tested mock community and the relative abundance of such erroneously reported taxa.

2.4.3 Test environment and hardware

All tests were performed on a Dell PowerEdge T430 server with two Intel Xeon processors (E5-2630 v3 at 2.4GHz), 128 Gb of RDIMM RAM (2133 MT/s), and 7.2K RPM NLSAS (6 Gbps) hard drives in a RAID-0 configuration, running Ubuntu Server 16.04 (Xenial Xerus). All SPONS runs were performed using the maximum 32 threads available on the server.

2.5 Results

2.5.1 Recall and abundance deviance

Recall, or the percentage of known phyla present in the mock communities detected, indicated that, overall, the RDP mode of SPONS provided a better taxonomic classification (Table 2.2, Figure 2.2), especially at lower taxonomic ranks (90% recall in contrast to 87% recall at genus level). At species level, while RDP performed better overall, recall was generally low for both methods (61% and 7% for RDP and VSEARCH respectively). The performance of SPONS analyzing mock-12, containing closely related taxa, was overall lower, with recall percentages diminishing quickly for taxonomic levels lower than order (Table 2.2). Likewise, deviance from known abundances was low among

datasets mock-18, mock-20 and mock-23 for both RDP and VSEARCH processing modes, across all taxonomic levels. However, significantly higher deviances were detected for the mock-12 community.

2.5.2 Unclassified and misclassified taxa

Evaluation of the percentage of unclassified reads in each mock community showed that SPONS in RDP mode improved classification rates compared to the VSEARCH mode (Figure 2.2), which generally showed a higher percentage of unclassified reads, especially at lower taxonomic levels and the VSEARCH mode failed to classify a large number of reads (genus: 0.16% - 35.72%, species: 29.58% - 98.44%, for RDP and VSEARCH, respectively). High unclassified percentages were observed at species level across all communities.

The number of misclassified taxa was generally low (Table 2.2), however, an increase in misclassifications was observed at lower taxonomic levels. This effect was more pronounced in the RDP classification mode. The total abundance of misclassified taxa was found to be low, less than 0.1% up to genus level. At species level, misclassification abundances were higher, specifically for the RDP classification mode (4.31% and 0.14% for RDP and VSEARCH modes respectively).

2.5.3 Performance

No significant time difference was observed between the RDP and VSEARCH modes. For small datasets like mock-19, mock-20 and mock-23, SPONS completed analysis in under three minutes (Table 2.3), processing an average of 1812 reads per

second. The larger mock-12 dataset took approximately 9 minutes to complete, indicating a throughput of 3683 reads per second.

2.6 Discussion

The SPONS pipeline presented here is an automated 16S rRNA gene sequencing analysis workflow, including read trimming and quality control, OTU creation, and taxonomic classification. SPONS was designed to be centered around applications and scripts that are throughput optimized, central to which are the VSEARCH sequence processing tool and SWARM OTU clustering software. Both tools leverage modern computer processor capabilities, such as efficient multi-threading and the Streaming SIMD Extensions (SSE2), potentially enabling a reduction in computational time over other comparable tools.

To validate the performance of SPONS, four mock communities with known taxonomic composition were processed to test the ability of SPONS to correctly construct OTUs and classify taxa from phylum to species level. These mock datasets, generated through sequencing of samples of true biological origin, as opposed to being generated *in-silico*, provide test data which is technologically relevant as they represent true experimental observations and enable the validation of bioinformatics methods including experimental conditions and variations (23).

Analysis of mock communities showed that, overall, SPONS achieves a high rate of recall, recovering up to 100% of the taxonomic diversity, especially at higher taxonomic levels including phylum, class and order (Table 2.2, Figure 2.2). Likewise, deviance from

known abundance values and misclassification of taxa was generally low. Similar to other 16S rRNA analysis methods such as QIIME and mothur, recall rates were lower at the family and genus level (27), especially for classifications performed using VSEARCH (87% and 78% for family and genus respectively); a recall of over 90% was still observed using RDP (90% and 91%, family and genus respectively), suggesting that for experiments where a lower classification is required, RDP based taxonomic assignment is recommended.

Species-level classification was poor (Table 2.2), and both VSEARCH or RDP performed sub optimally. While included in this study for completeness, it is generally not recommended to perform species level classification using partial 16S rRNA gene data, as hyper-variable regions can be highly similar between species within the same genus, leading to potential misclassifications (4).

The challenging nature of highly similar taxa was also evident in the analysis of the "extreme" mock community. In comparison to the other analyzed mock communities, the "extreme" dataset showed a significant decrease in taxonomic recall and increase in abundance deviance, especially at lower taxonomic levels. Furthermore, a slight increase in mis- and unclassified taxa was also observed. Taxa within this dataset are differentiable by as few as a single nucleotide over the length of the sequenced region, and it is likely that such taxa were merged together during the OTU formation process. As such, erroneous mergers could group two or more distinct taxa together causing an incorrect inflation of relative abundance values. However, in experiments where highly related taxa are

expected, this could in part be overcome by setting the OTU distance value to zero, effectively disabling OTU extension and performing only a dereplication step instead.

Furthermore, the existence of highly related taxa inherently reduces the available resolution for taxonomic assignment within reference databases. Especially at low taxonomic levels, it is possible that a given query sequence could match to multiple related references sequences for such closely related taxa, subsequently introducing ambiguity in both the alignment-based LCA determination as well as the probability calculation within the RDP classification method. As seen in the "extreme" mock community, the default behavior of SPONS is to be conservative, and report the classification at a higher taxonomic level, instead of producing a potential misclassification.

Overall, we show that SPONS can be used to accurately reconstruct a varied type of mock communities. The ability of SPONS to quickly analyze large datasets, including highly complex communities, makes it a powerful addition to a 16S rRNA gene analysis toolkit, and could be of benefit to researchers aiming to implement an expandable, fast and customizable pipeline.

2.7 References

1. Reuter JA, Spacek D, Snyder MP. High-throughput sequencing technologies. *Mol Cell* (2015) **58**:586–597. doi:10.1016/j.molcel.2015.05.004

2. Lightbody G, Haberland V, Browne F, Taggart L, Zheng H, Parkes E, Blayney JK. Review of applications of high-throughput sequencing in personalized medicine: barriers and facilitators of future progress in research and clinical application. *Brief Bioinform* (2018) doi:10.1093/bib/bby051

3. The Human Microbiome Project Consortium, Methé BA, Nelson KE, Pop M, Creasy HH, Giglio MG, Huttenhower C, Gevers D, Petrosino JF, Abubucker S, et al. A framework for human microbiome research. *Nature* (2012) **486**:215–221. doi:10.1038/nature11209

4. Case RJ, Boucher Y, Dahllöf I, Holmström C, Doolittle WF, Kjelleberg S. Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. *Appl Environ Microbiol* (2007) **73**:278–288. doi:10.1128/AEM.01177-06

5. Pace NR, Stahl DA, Lane DJ, Olsen GJ. "The analysis of natural microbial populations by ribosomal RNA sequences," in *Advances in Microbial Ecology* Advances in Microbial Ecology., ed. K. C. Marshall (Boston, MA: Springer US), 1–55. doi:10.1007/978-1-4757-0611-6_1

6. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology* (2017) **35**:833–844. doi:10.1038/nbt.3935

7. Sboner A, Mu XJ, Greenbaum D, Auerbach RK, Gerstein MB. The real cost of sequencing: higher than you think! *Genome Biology* (2011) **12**:125. doi:10.1186/gb-2011-12-8-125

8. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, et al. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* (2010) **7**:335–336. doi:10.1038/nmeth.f.303

9. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, et al. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl Environ Microbiol* (2009) **75**:7537–7541. doi:10.1128/AEM.01541-09

10. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* (2013) **41**:D590–D596. doi:10.1093/nar/gks1219

11. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* (2011) **17**:10–12. doi:10.14806/ej.17.1.200

12. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* (2014) **30**:2114–2120. doi:10.1093/bioinformatics/btu170

13. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* (2014) **30**:614–620. doi:10.1093/bioinformatics/btt593

14. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* (2016) **4**: doi:10.7717/peerj.2584

15. Mahé F, Rognes T, Quince C, Vargas C de, Dunthorn M. Swarm v2: highly-scalable and high-resolution amplicon clustering. *PeerJ* (2015) **3**:e1420. doi:10.7717/peerj.1420

16. Mahé F, Rognes T, Quince C, Vargas C de, Dunthorn M. Swarm: robust and fast clustering method for amplicon-based studies. *PeerJ* (2014) **2**:e593. doi:10.7717/peerj.593

17. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* (2011) **27**:2194–2200. doi:10.1093/bioinformatics/btr381

18. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* (2007) **73**:5261–5267. doi:10.1128/AEM.00062-07

19. Katoh K, Standley DM. MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* (2013) **30**:772–780. doi:10.1093/molbev/mst010

20. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* (2000) **17**:540–552. doi:10.1093/oxfordjournals.molbev.a026334

21. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately maximum-likelihood trees for large alignments. *PLOS ONE* (2010) **5**:e9490. doi:10.1371/journal.pone.0009490

22. McDonald D, Clemente JC, Kuczynski J, Rideout JR, Stombaugh J, Wendel D, Wilke A, Huse S, Hufnagle J, Meyer F, et al. The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome. *Gigascience* (2012) **1**:7. doi:10.1186/2047-217X-1-7

23. Bokulich NA, Rideout JR, Mercurio WG, Shiffer A, Wolfe B, Maurice CF, Dutton RJ, Turnbaugh PJ, Knight R, Caporaso JG. mockrobiota: a public resource for microbiome bioinformatics benchmarking. *mSystems* (2016) **1**:e00062-16. doi:10.1128/mSystems.00062-16

24. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods* (2016) **13**:581–583. doi:10.1038/nmeth.3869

25. Tourlousse DM, Yoshiike S, Ohashi A, Matsukura S, Noda N, Sekiguchi Y. Synthetic spike-in standards for high-throughput 16S rRNA gene amplicon sequencing. *Nucleic Acids Res* (2017) **45**:e23. doi:10.1093/nar/gkw984

26. Gohl DM, Vangay P, Garbe J, MacLean A, Hauge A, Becker A, Gould TJ, Clayton JB, Johnson TJ, Hunter R, et al. Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. *Nat Biotechnol* (2016) **34**:942–949. doi:10.1038/nbt.3601

27. Almeida A, Mitchell AL, Tarkowska A, Finn RD. Benchmarking taxonomic assignments based on 16S rRNA gene profiling of the microbiota from commonly sampled environments. *Gigascience* (2018) **7**: doi:10.1093/gigascience/giy054

Table 2.1: Configuration settings available for the SPONS pipeline.

Section	Variable	Type	Description
Quality control	minAverageQuality	Integer	Minimum average Phred score threshold to retain a sequence
	maxNumberOfDips	Integer	Maximum number bases with a “dip" in Phred quality score allowed in a sequence
	dipCheckTreshold	Integer	Phred score cutoff to consider a base as a “dip” in quality
CutAdapt	aPrimer	String	Location of FASTA file containing the forward primers that should be trimmed
	gPrimer	String	Location of FASTA file containing the reverse primers that should be trimmed
Rscript	doRanalysis	Boolean	Should SPONS automatically run an R script after completion?
	Rscript	String	Location of the R script to run
	Design	String	This string is passed to the R script and can be used to set an experimental design.
SWARM	Resolution	Integer	SWARM local-linking distance threshold
Binaries	Pear	String	Location of the PEAR binary on the system
	Trimmomatic	String	Location of the Trimmomatic binary on the system
	VSEARCH	String	Location of the VSEARCH binary on the system
	SWARM	String	Location of the SWARM binary on the system
	MAFFT	String	Location of the MAFFT binary on the system
	Gblocks	String	Location of the Gblocks binary on the system
	Fasttreemp	String	Location of the FastTreeMP binary on the system
	Cutadapt	String	Location of the CutAdapt binary on the system
	Rdpjarpath	String	Location of the RDP Java JAR on the system
Databases	Silvavsearchdb	String	Location of the SILVA taxonomy database
RDP	Threads	Integer	Number of threads to use when running the RDP classifier
	Maxmem	Integer	Maximum amount of memory to use (in Mb) when running the RDP classifier

	Posteriorconfidence	Float	Minimum confidence a taxonomic hit must obtain to be accepted (0 – 1)
VSEARCH	Threads	Integer	Number of threads to use when running the VSEARCH classifier
	Maxaccepts	Integer	Maximum number of hits to accept before stopping the search
	Maxrejects	Integer	Maximum number of non-matching target sequences to consider before stopping search
	Id	Float	Minimum pair-wise identity needed to accept a hit
	Iddef	Integer	Pairwise identity definition used by VSEARCH

Table 2.2: Results of SPONS validation test on four mock datasets, using both the RDP and VSEARCH assignment methods.

Test	Level	Mock-12		Mock-18		Mock-20		Mock-23		Overall			
		R	V	R	V	R	V	R	V	R	Rsd	V	Vsd
Recall (%)	Phylum	100	100	88	88	100	100	100	100	97.0	6.00	97.0	6.00
	Class	100	100	92	92	100	100	100	100	98.0	4.00	98.0	4.00
	Order	100	100	93	93	100	100	100	100	98.3	3.50	98.3	3.50
	Family	67	67	93	93	100	100	100	88	90.0	15.7	87.0	14.2
	Genus	69	50	93	80	100	100	100	82	90.5	14.7	78.0	20.7
	Species	50	4	60	7	80	10	55	5	61.3	13.2	6.50	2.65
Deviance (%)	Phylum	36	19	3	3	6	7	3	3	12.0	16.1	8.00	7.57
	Class	24	13	2	2	4	5	2	2	8.0	10.7	5.50	5.20
	Order	24	13	2	2	4	4	2	3	8.0	10.7	5.50	5.07
	Family	19	12	2	2	3	4	1	3	6.3	8.54	5.25	4.57
	Genus	9	6	2	2	3	4	1	3	3.8	3.59	3.75	1.71
	Species	8	3	3	1	3	5	1	22	3.8	2.99	7.75	9.64
Un-classified (%)	Phylum	0	38.2	0.06	0.23	0	10.6	0.07	0.09	0.03	0.04	12.3	18.0
	Class	0	38.2	0.17	0.23	0	11.9	0.07	0.10	0.06	0.08	12.6	17.9
	Order	0	38.2	0.21	0.23	0.01	25.2	0.07	30.5	0.07	0.10	23.5	16.4
	Family	0	38.5	0.21	0.23	0.26	36.2	0.07	33.6	0.14	0.12	27.1	18.0
	Genus	0.1	38.6	0.21	14.0	0.28	36.2	0.07	54.0	0.16	0.10	35.7	16.5
	Species	9.9	99.7	42.2	94.4	5.38	99.8	60.8	99.9	29.6	26.5	98.4	2.72
Mis-classified (N)	Phylum	2	2	0	0	0	0	1	1	0.75	0.96	0.75	0.96
	Class	3	3	0	0	0	0	2	2	1.25	1.50	1.25	1.50
	Order	3	3	0	0	0	0	2	2	1.25	1.50	1.25	1.50
	Family	4	3	0	0	0	1	3	3	1.75	2.06	1.75	1.50
	Genus	7	3	0	0	1	1	3	3	2.75	3.10	1.75	1.50
	Species	5	1	1	0	4	6	2	4	3.00	1.83	2.75	2.75
Abundance of mis-classified (%)	Phylum	0.01	0.01	0	0	0	0	0.01	0.01	0	0.01	0	0.01
	Class	0.02	0.02	0	0	0	0	0.02	0.02	0.01	0.01	0.01	0.01
	Order	0.02	0.02	0	0	0	0	0.02	0.02	0.01	0.01	0.01	0.01
	Family	0.10	0.08	0	0	0	0.01	0.02	0.02	0.03	0.05	0.03	0.04
	Genus	0.03	0.01	0	0	0.01	0.01	0.02	0.02	0.02	0.01	0.01	0.01
	Species	0.03	0	0.03	0	17.2	0.52	0.02	0.05	4.31	8.58	0.14	0.25

Recall: Percentage of known taxa recovered, deviance: average percentage of deviance from theoretical known abundance, unclassified: percentage of reads with no taxonomic classification, misclassified: number of taxa misclassified, abundance of misclassified: relative abundance of misclassified abundance, R: results for RDP classifier, V: results for VSEARCH classifier, Rsd: standard deviation for RDP classifier, Vsd: standard deviation for VSEARCH classifier.

Table 2.3: Processing time of SPONS in RDP and VSEARCH mode on mock communities.

Community	Number of reads	Run time	
		RDP	VSEARCH
Mock-12	2,040,485	9m 14s	9m 11s
Mock-18	169,516	1m 44s	1m 46s
Mock-20	207,197	2m 9s	2m 7s
Mock-23	336,667	2m 33s	2m 35s

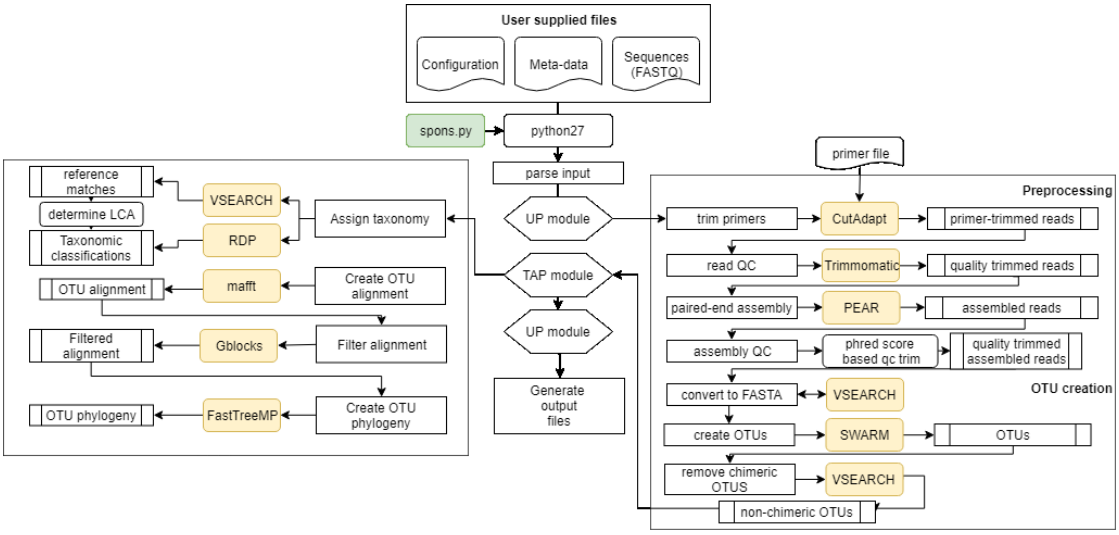

Figure 2.1: Schematic overview of the SPONS workflow. Shown are the interactions between the steps (indicated in white rectangles) within the universal processing module (UP) and Taxonomy and Phylogeny (TAP) modules (indicated with diamond shapes). The green rectangle indicates the starting point where the SPONS script is loaded, whereas serrated rectangles indicate input files, rounded rectangles denote processes, with externally called applications shaded in orange, double-edge rectangle indicate output of processes.

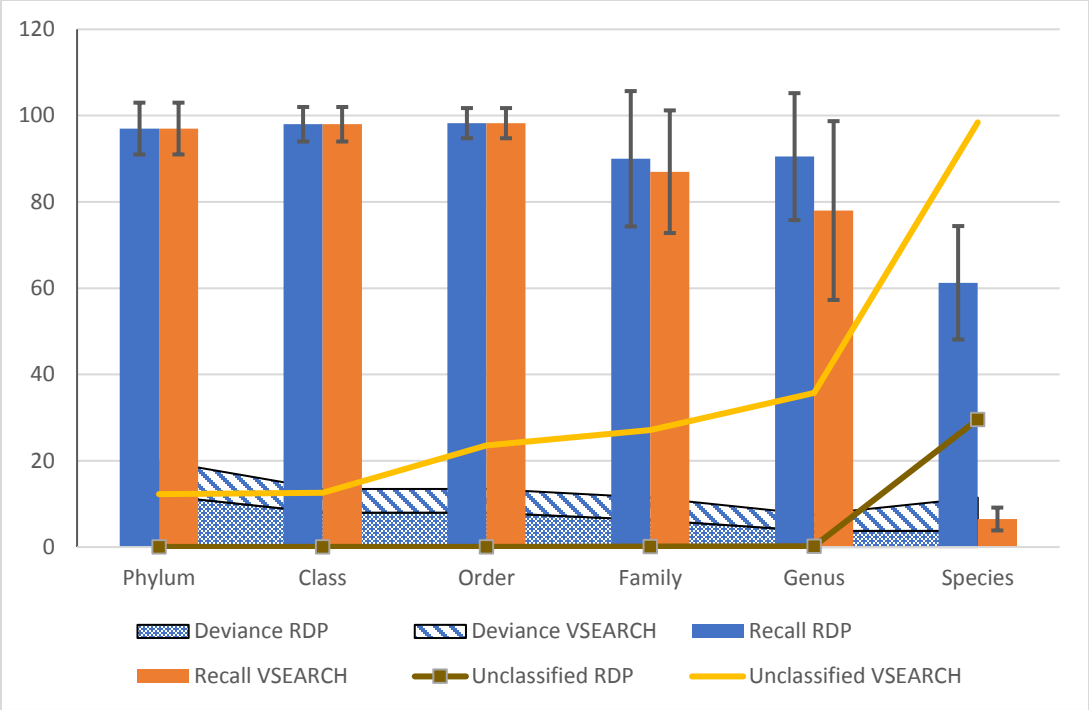

Figure 2.2: Average analysis performance of SPONS using four distinct mock communities. At each taxonomic level (Phylum to Species), the percentage recall, abundance deviance, and relative abundance of unclassified reads are shown as bars (whiskers on bars indicate standard deviation), area and lines, respectively.

CHAPTER 3: MICROBIOME ANALYSIS SHOWS ENRICHMENT FOR SPECIFIC BACTERIA IN SEPARATE ANATOMICAL REGIONS OF THE DEEP-SEA CARNIVOROUS SPONGE *CHONDRCOLADIA GRANDIS*

Adapted from the published version available in FEMS Microbiology Ecology[4]
https://doi.org/10.1139/cjm-2018-0097

3.1 Abstract

The Cladorhizidae is a unique family of carnivorous marine sponges characterized by either the absence or reduction of the aquiferous system and by the presence of specialized structures to trap and digest mesoplanktonic prey. Previous studies have postulated a key role of host-associated bacteria in enabling carnivory in this family of sponges. In this study, we employed high-throughput Illumina-based sequencing to identify the bacterial community associated with four individuals of the deep-sea sponge *Chondrocladia grandis* sampled in the Gulf of Maine. By characterizing the V6 through V8 region of the 16S rRNA gene, we compared the bacterial community composition and diversity in three distinct anatomical regions with predicted involvement in prey capture (sphere), support (axis) and benthic substrate attachment (root). A high abundance of *Tenacibaculum*, a known siderophore producing bacterial genus, was present in all anatomical regions and specimens. The abundance of *Colwellia* and *Roseobacter* was greater in sphere and axis samples, and bacteria from the hydrocarbon-degrading *Robiginitomaculum* genus were most abundant in the root. This first description of the bacterial community associated with *C. grandis* provides novel insights into the

[4] This version was modified to include supplementary figures and tables in the main text

contribution of bacteria to the carnivorous lifestyle while laying foundations for future cladorhizid symbiosis studies.

3.2 Introduction

Marine sponges (phylum Porifera) are ubiquitous, widespread and ecologically important members of many marine environments (1). Sponges harbour phylogenetically diverse, yet specific and stable, microbial communities; bacteria, archaea, single-celled eukaryotes and even viruses are thought to play central roles in maintaining host biology and function (2–5). Microbial residents can comprise up to 47% of the overall sponge biomass, are proposed to contribute to host metabolism and produce biologically active compounds beneficial to the sponge (2,3,6).

Most sponges are filter feeders that circulate large volumes of water through a specialized aquiferous system and obtain nutrients such as carbon, oxygen, silica and nitrogen from the water column (7,8). As the majority of sponges follow this lifestyle, the aquiferous system has been regarded as a defining feature of the phylum (9,10). However, an exception to this rule is the carnivorous feeding strategy observed in the sponge family Cladorhizidae (10,11). Cladorhizids display unique morphological adaptions which enable carnivory, including an erect body morphology, a diminished or absent aquiferous system, and the production of filaments or inflatable spheres serving as adhesive traps to ensnare live prey (10,11). This carnivorous feeding strategy is often regarded as an evolutionary adaptation to overcome the limitations of filter feeding in deep-sea oligotrophic

environments, but has also been observed in species dwelling in food-rich environments (10,12).

Laboratory feeding experiments on *Lycopodina hypogea* (13) and *in situ* observations of *Chondrocladia* (*Chondrocladia*) *gigantea* (14) suggest a feeding sequence that begins when captured prey are quickly engulfed by a veil of migrating sponge pinacocyte cells (11,12,15). The engulfed prior pinacocyte layer then rapidly dedifferentiates and diminishes, leaving the prey in direct contact with the intercellular matrix (mesohyl) of the sponge. Archaeocytes and bacteriocytes then concentrate around the prey body, forming a 'cyst' in which prey degradation occurs. Interestingly, bacteriocytes contain large vacuoles harboring numerous symbiotic bacteria similar to those observed in the mesohyl (11). The archaeocytes, bacteriocytes and associated symbiotic bacteria are thought to facilitate prey degradation, presumably through enzymatic action. After degradation and disruption of the prey body, small elements are phagocytosed by the archaeocytes and bacteriocytes, and in the latter, the phagosomes can subsequently fuse with vacuoles containing symbiotic bacteria. Early stages of intracellular digestion might be dependent on bacteria-mediated actions or secreted proteins, given that the symbiotic bacteria within vacuoles increase in abundance and division rate at this stage (11).

The functionality and importance of the carnivorous sponge microbiome likely extends beyond prey digestion. Recent studies have revealed the high abundance of ammonium-oxidizing archaea and sulfate-reducing bacteria in *L. hypogea*, suggesting the

involvement of the microbiome in detoxification and both nitrogen and sulfur cycles (16,17). In filter-feeding sponges, secondary metabolites can act as protective agents against predators and epibionts (18,19). Part of these secondary metabolites are thought to be of bacterial origin, constituting a mutually beneficial relationship between bacteria and host, which could explain the evolutionary and ecological success of sponges (1,18–20). We expect similar relationships between carnivorous sponge species and their microbial residents.

Carnivorous sponges are mostly confined to deep-sea environments (21) and have not been studied to the same extent as filter-feeding sponges. To the best of our knowledge, the only investigation of a carnivorous sponge microbiome focuses on *L. hypogea*, a species found in shallow-water caves (16). Here, we investigate the bacterial community (from here on colloquially referred to as the microbiome) associated with a deep-sea cladorhizid from the Gulf of Maine, identified as *C.* (*Chondrocladia*) *grandis* (22), using high-throughput sequencing (HTS) of the V6/V7/V8 region of the 16S rRNA gene. As the specimens have a unique morphology, with anatomical regions predicted to have different functions (10,11), we compared three distinct anatomical regions within four individual sponge specimens to investigate the variability of the microbiome within and among specimens.

3.3 Materials and methods

3.3.1 Sample collection

Chondrocladia specimens (N = 4) were collected from the base of the Northeast Channel of the Gulf of Maine (N41°48.2302′, W65° 41.0412′) on 27 July 2014. From the National Oceanic and Atmospheric Administration (NOAA) vessel R/V Henry B. Bigelow, the CSSF-ROPOS (Canadian Scientific Submersible Facility, Remotely Operated Platform for Ocean Sciences) was used to locate and extract specimens from the seafloor at a depth of ~852 m. Sponges displayed an erect body morphology, with a central cylindrical body (primary axis) perched on a stalk or 'root' embedded within the sediment (Figure 3.1). From the central body axis, various lateral appendages (secondary axes) radiated outwards and terminated in translucent spheres. Upon surfacing, specimens were immediately dissected, and a single sphere, secondary axis and root sample were retained from each specimen for DNA extraction. These samples were frozen at −20°C in sterile containers until laboratory arrival, where they were transferred to a −80°C freezer until further analysis. Additional samples were fixed in buffered formaldehyde.

3.3.2 Spicule preparation and examination

Spicules were isolated from sphere, secondary axis and root regions of a single specimen by immersing sponge fragments in 10% sodium hypochlorite. After carefully rinsing spicules in distilled water and 100% ethanol, they were transferred to stubs and imaged using a Quanta 650 Scanning Electron Microscope (FEI Company, Hillsboro, OR, USA). Between 2 and 23 spicules of each type were measured, and sizes reported as average ± standard deviation (range).

3.3.3 DNA extraction and HTS

Sponge sections were thawed on ice and rinsed three times with sterile nucleic-acid-free PBS to remove loosely associated bacteria and other contaminants. Samples were then aseptically cut into multiple thin sections and transferred to bead beating tubes containing 0.1 mm glass beads (MO BIO Laboratories, Inc., Carlsbad, CA, USA). DNA was subsequently isolated from cut sections using the MO Bio PowerViral DNA/RNA isolation kit (MO BIO Laboratories) according to the manufacturer's specifications. Extracted nucleic acids were checked for quality and concentration by using a NanoDrop 1000 (Thermo Scientific, Waltham, MA, USA) and outsourced for Illumina MiSeq 2 × 300-bp paired-end sequencing of the 16S rRNA (V6/V7/V8 region) gene at the Integrated Microbiome Resource of the Center for Comparative Genomics and Evolutionary Bioinformatics (Dalhousie University, Halifax, Canada), using previously published primers B969F and B1406R (23).

3.3.4 Molecular identification of sponge specimens

Collected sponge specimens were characterized by sequencing the 18S rDNA, 28S rDNA C1-D2 partition, the overlapping 'Folmer' and 'Erpenbeck' mitochondrial COI gene fragments and part of the asparagine-linked glycosylation 11 homolog (ALG11) gene, using primers and methods as in Hestetun et al. (10). Amplified PCR products were purified using Agencourt AMPure XP beads (Beckman Coulter, Brea, CA, USA) and submitted for Sanger sequencing to the Centre for Applied Genomics (Hospital for Sick Children, Toronto, Canada). Obtained sequences were deposited in NCBI GenBank database under accession numbers KX950006–KX950009.

Reference sequences of other carnivorous sponges (N = 8) and of non-carnivorous sponges (N = 2) as outgroups were downloaded from GenBank and used for the phylogenetic analysis, which was performed in RAxML 8.1.21 invoked from RAxMLGUI 1.5b. To determine sponge taxonomy, 18S rDNA (1737 nt) alignments and concatenated alignments of the 28S rRNA (841 nt), COI (572 nt) and ALG11 (723 nt) genes were generated with MAFFT (version 7.221). Concatenated alignments were divided into three unlinked partitions based on included genes in RAxMLGUI. Phylogenetic trees were constructed using the maximum-likelihood method (24) and the GTR+GAMMA model (24,25). The robustness of the analysis was tested with the bootstrap method using 10 000 rapid bootstrap replicates.

3.3.5 Processing and taxonomic profiling of HTS data

The in-house developed SPONS (Streamlined Processor Of Next-gen Sequences) pipeline was used to process Illumina MiSeq sequence data. An initial quality check was performed using Trimmomatic (version 0.33) to remove remaining primer and adapter sequences, check reads for quality (leading and trailing quality threshold = 25) and ensure that the remaining sequence information was of sufficient (>100 bp) length (26). Paired-end reads passing the initial quality check were subsequently merged using PEAR version 0.9 (27), requiring a minimal overlap of 105 bp and a P-value of <0.001 for read merging to occur. Operational taxonomic units (OTUs) were constructed by dereplicating the total sequence dataset using VSEARCH (merging strictly identical sequences, removing sequences with any ambiguities and adding abundance information) and using the dereplicated weighted dataset in conjunction with the *de novo* clustering program SWARM

to create OTUs, retaining the seed of each OTU as representative sequence (28,29). The SWARM option 'fastidious' was used to promote merging of low-abundance swarms to closely related high abundance swarms, as to generate fewer OTUs with a higher rate of inclusion. Subsequently, representative OTU sequences were checked for chimeric sequences using VSEARCH (UCHIME *de novo* implementation, minimum score (h) = 0.30).

An internal SPONS wrapper function was used to infer taxonomy for all detected OTUs by invoking four distinct taxonomic classifiers through the assign_taxonomy.py command within the QIIME framework: UCLUST (maximum number of hits to consider = 1, minimum percent similarity = 0.8, minimum consensus fraction = 0.76), RDP (minimum confidence = 0.5), SortMeRNA (minimum consensus fraction = 0.51, similarity = 0.8, best alignments per read = 1, coverage = 0.8, e-value = 1.0) and BLAST (e-value = 0.001) (30–34). Subsequently, a consensus taxonomy was extracted for each OTU at the lowest taxonomic level at which a minimum of three classifier methods reported an identical result. For all classification methods, the full redundant SILVA SSU database and taxonomy (release 123) was used as reference (35,36). OTUs with taxonomic assignments other than bacteria were removed from the dataset.

Full dataset phylogeny was performed by creating alignments of OTU sequences in MAFFT7 using the FFT-NS-2 algorithm, which were subsequently trimmed using gBlocks version 0.91b (minimum block length = 3, all gap positions allowed). FastTree (version 2.18, fastest option enabled in combination with the generalized time reversible

model) was used to infer approximately maximum-likelihood phylogeny (37). The source code of the SPONS pipeline is available upon request.

3.3.6 Bacterial diversity analyses, visualization and statistics

Analysis of alpha diversity was performed with QIIME's *alpha_rarefaction.py* workflow utilizing the phylogenetic diversity measure (PD_whole_tree) and Simpson's Equitability measure E (simpson_e) (38). Statistical comparison of alpha diversity measurements was accomplished using *compare_alpha_diversity.py* in QIIME using the non-parametric test option with 999 permutations. Beta diversity analysis was performed using R with the PhyloSeq package (39,40). In short, OTU abundance data were converted to relative abundance, which was subsequently used for principal coordinate analysis (PCoA) using the Bray–Curtis dissimilarity measure. PCoA analysis was also performed on variance stabilized OTU count data, generated by applying the variance stabilizing transformation (VST) function available in the DeSeq2 R package on original count data (41). Community structure was visualized by collating OTU count data of samples according to their anatomical region and subsequently plotting relative abundance at phylum, family and genus levels as heatmaps using the enhanced heatmap function within the gplots R package (42). A supporting dendrogram was constructed using hierarchical clustering of the Bray–Curtis dissimilarity generated from DeSeq2 stabilized OTU count data with the 'vegan' community ecology R package (43). Analysis of differentially abundant phyla, families and genera among anatomical regions was performed on non-collated data using the PhyloSeq implementation of DESeq2 (using the Wald significance

test, local dispersion fit type and $P < 0.05$) to perform statistical testing. Interactive taxonomy visualizations were created using Krona (44).

3.4 Results

3.4.1 Genetic and spicule-based identification of sponge specimens

Five types of spicules were identified through electron microscopy (Figure 3.2): megascleres (mycalostyles and, in the root, microspinose strongyles), two unguiferous anchorate isochelae (large and small), and sigmas. Mycalostyles showed an average length of 1190.52 ± 370.68 μm (817.94–1994.71 μm) and width of 21.21 ± 4.81 μm (12.58–30.86 μm) with shorter mycalostyles occurring in the spheres and longer ones in the axis. The shorter microspinose strongyles were on average 207.34 ± 52.73 μm (152–257 μm) long and 5.24 ± 2.03 μm (3.01–7 μm) wide. Sigmas were consistent in size with an average length of 42.44 ± 2.56 μm (39.79–47.36 μm) and width of 2.06 ± 0.20 μm (1.78–2.34 μm). Two varieties of unguiferous anchorate isochelae were observed: large and small. The large isochelae were 65.77 ± 2.30 μm (63.29–68.98 μm) in length × 4.49 ± 0.32 μm (3.96–4.98 μm) in width, and displayed six teeth at each extremity with an average length of 10.96 ± 1.56 μm (8.39–13.12 μm). Small isochelae measured 22.78 ± 3.28 μm (19.17–25.57 μm) in length and 2.30 ± 0.20 μm (2.08–2.48 μm) in width, and had between five and nine observed teeth at each end.

Phylogenetic maximum-likelihood analysis of concatenated 28S, COI and ALG11 alignments (Figure 3.3A) placed the studied specimens within the carnivorous family Cladorhizidae and specifically within the *Chondrocladia* clade. The studied sponge had a

distinct concatenated gene sequence, but clustered with the *Chondrocladia gigantea* reference sequence (100% bootstrap) with high-sequence identity (28S rDNA: 99.8%, COI: 100%, ALG11: 99%). Sequence analysis of the 18S rDNA showed a low-sequence identity to the 18S reference sequence of *C. gigantea* (97.8%) and the phylogenetic analysis placed the studied specimen as an outgroup of the *C. gigantea* and *Cladorhiza abyssicola* clade (Figure 3.3B). We tentatively identified our specimens as *C. grandis* based on the genetic distance observed within the 18S rDNA, 28S rDNA and ALG11 genes, the presence of microspinose strongyles (not reported to be present in *C. gigantea*) and the proximity of our collection site to the type location for *C. grandis* (Nova Scotia, Canada (22)).

3.4.2 Inspection of HTS data

Sequencing of the V6/V7/V8 16S rRNA gene resulted in over 740000 merged read-pairs passing quality checks, with an average of ~47000 sequences per sample. Dereplication, chimera removal and clustering resulted in the formation of 3725 high-quality, non-singleton bacterial OTUs distributed among samples (Table 3.1). Furthermore, rarefaction curve analysis of observed OTUs indicated that all samples reached their diversity plateau, suggesting sufficient sequencing depth (data not shown). A single specimen (root of sponge 1) displayed both an extremely high OTU count and phylogenetic diversity compared to the other root samples analyzed. As this was most likely the result of contamination with sediment bacteria, we excluded this sample from downstream analyses.

3.4.3 Bacterial community diversity and structure

Comparison of the Simpson Equitability (*E*) indices showed the evenness of species within bacterial communities to be low in all anatomical regions, with roots (divergent root sample excluded from all metrics) presenting the most uneven community (mean E = 0.006), significantly different from axis (mean E = 0.018, $P < 0.05$) and sphere (mean E = 0.021, $P < 0.05$) samples (Table 3.1). Phylogenetic diversity (PD) analysis revealed a significantly higher diversity in the root samples (mean PD = 41.48) compared to the axis (mean PD = 19.87, $P < 0.05$) and sphere (mean PD = 20.22, $P < 0.05$) samples (Table 3.1). No significant differences were observed for either *E* or PD between the sphere and axis samples.

A similar pattern was noted in results of the beta-diversity analyses (Bray–Curtis based PCoA). Using proportional count data, axis and sphere samples formed a distinct cluster, distant from root samples along the first principal coordinate, which represented 47.6% of the total variation inherent to the dataset and suggested distinct communities (Figure 3.4A). However, the axis and sphere samples did not form a tight cluster along the secondary principal coordinate axis. PCoA analysis on VST transformed data (reducing the importance of highly abundant taxa) showed a similar distinction between root and sphere/axis samples along the first principal coordinate (28% of the variability within the dataset). The sphere and axis cluster was more refined using transformed data showing the close association of sphere and axis samples and an additional clustering per individual (Figure 3.4B).

We examined the bacterial composition of samples at phylum, family and genus levels. The observed OTUs were assigned to 26 phyla, 277 families and 429 unique bacterial genera. For each anatomical region, a fraction of the total generic diversity found was shared among all sponge individuals (Figure 3.5). However, when considering sequence abundance, these shared genera encompassed the vast majority (on average 98%) of all sequences recovered from a given anatomical region.

Proteobacteria and Bacteroidetes were the most abundant phyla observed in all anatomical regions, ranging between 33%–65% and 28%–42% in relative abundance, respectively (Figure 3.6A). The abundance of these two phyla in the root samples was significantly lower ($P < 0.001$) than in sphere and axis samples. This lower abundance was offset by a significantly higher abundance of Actinobacteria in roots in comparison to axis and sphere samples ($P < 0.001$, 34% versus 4% and 3%, respectively).

The Flavobacteriaceae was the only highly abundant family observed in all sample groups (root: 27%, sphere: 29%, axis: 41%, Figure 3.6B); this family was present at a significantly higher abundance in the axis group ($P < 0.001$). Anatomical region-specific enrichment was observed, including a significantly higher abundance of the 'Sva0996 marine group' in the root group ($P < 0.001$, root: 34%, sphere: 3%, axis: 4.3%). Axis and sphere groups displayed a distinctly higher abundance of Rhodobacteraceae and a significant increase ($P < 0.001$) of an uncultured bacterium belonging to the 'E01-9C-26' unclassified Gammaproteobacteria lineage (both taxa at average abundances of 16% in axis and sphere groups compared to 5.5% Rhodobacteraceae and 2% E01-9C-26 in the root

group). We also observed higher relative abundances of both Alteromonadaceae (not significant) and Colwelliaceae ($P = 0.024$) in the sphere group (11.4% and 6.8% for each family respectively) compared to root (7%, 0.2%) and axis (6.6%, 2.3%) groups. Furthermore, a non-statistically significant increase of Hyphomonadaceae was observed in the root group with an average relative abundance of 8.4%, higher than the values observed in sphere (4.8%) and axis (2.9%) groups.

Genus-level comparisons (Figure 3.6C) indicated that the patterns observed at the family level are mostly driven by single genera. For example, the high abundance of the Flavobacteriaceae family translated into a high abundance of a single genus: *Tenacibaculum*. This genus was observed in all sample groups (root: 27%, sphere: 29%) and was more abundant in the axis group (41%). A significantly higher ($P < 0.01$) abundance of an uncultured Actinobacterium ('Sva0996 marine group') was observed in the root (34%) compared to the axis and sphere groups (4% and 3%, respectively). We also noted a significantly ($P < 0.01$) higher abundance of *Robiginitomaculum* in the root group at 8% than at 5% in the sphere and 3% in the axis group. Furthermore, abundance of both the NAC11-7 *Roseobacter* subclade and the uncultured 'E01-9C-26' Gammaproteobacterium was elevated in both the axis (16%) and sphere (15%) groups compared to the root group (5%, 2%), significantly so for the uncultured Gammaproteobacterium ($P < 0.01$). The OM60/NOR5 clade showed a higher abundance in the sphere group (11%, root: 7%, axis: 6%) but this was not statistically significant. Finally, the *Colwellia* genus comprised 7.9% of the sphere community and was

significantly (P = 0.02) more abundant in contrast to 0.2% and 2% in the root and axis regions, respectively.

3.5 Discussion

Despite substantial scientific interest, our understanding of bacteria–sponge interactions and symbiosis is incomplete and significant gaps remain in our knowledge of the composition, maintenance and function of these symbiotic relationships. In this study, we report on the microbiome of four carnivorous sponge specimens collected from the base of the continental shelf in the Gulf of Maine. Our collection site is very close to the type locality of the species *Chondrocladia grandis* (22). As the gross morphology of the latter species is highly similar to that of *C. gigantea*, as described by Lundbeck (45), some authors have proposed that *C. grandis* may be a synonym of *C. gigantea*. Although no spicule information or genetic sequences are available for *C. grandis*, phylogenetic analysis of the 18S rDNA, 28S rDNA, ALG11 and mitochondrial COI genes showed that the sponge specimens collected in this study were distinct from *C. gigantea* (Figure 3.3). Spicule types mainly resembled those previously described for *C. gigantea*, except for the identified microspinose strongyles. Based on spicule data, collection location and genetic differences with *C. gigantea*, we identified the collected specimens as *C. grandis* and deposited the first sequences for this species in GenBank. However, additional analyses are needed to fully confirm the taxonomy of this species.

To our knowledge, the bacterial community composition of *Chondrocladia* has never been investigated. The morphological structures used for prey trapping (sphere),

attachment to the benthic substrate (root) and sphere support (axis) are unique to carnivorous sponges and we have placed a large emphasis on detecting common (inter-individual) patterns in bacterial community composition that are driven by the different anatomical regions.

Our results revealed a stark difference in overall bacterial community structure among the three anatomical regions. Analysis of both phylogenetic diversity and Simpson Equitability showed the bacterial communities present in the sphere and axis to be of similar diversity and evenness, while those in the root were significantly more diverse ($P < 0.05$) and less even ($P < 0.05$). The divergent nature of the bacterial community associated with the root is also evident from the PCoA analyses, where root samples were relatively distant from a tight cluster composed of axis and sphere samples along the first principal coordinate axis. This apparent difference in bacterial community could indicate that root and sphere/axis sections retain specific bacteria to support distinct biological functions. The increase in diversity within root sections could be due to their physical location within the marine sediments, where associations with benthic bacteria could potentially occur. Although samples were pretreated to remove sediment particles, the co-isolation of attached sedimentary bacteria cannot be excluded.

Analysis of bacterial community structure revealed the presence of 429 distinct bacterial genera, of which the 10 most abundant constituted the vast majority of the entire microbiome, in line with the low Simpson Equitability indices we observed in all samples. For each anatomical region, an examination of the genera shared by all individuals showed

that, although only a select subset (on average 22%) were ubiquitous, this subset composed the majority (~98%) of sequences found. Ubiquitous genera more likely represent true symbionts, while the diverse but rare non-ubiquitous genera could be prey-associated bacteria used as a food source, bacteria introduced through the aquiferous system that *C. grandis* uses to inflate trapping spheres, or surface-associated bacteria.

Remarkably, the microbiome showed a high abundance (up to 41%) of *Tenacibaculum* (family: Flavobacteriaceae) in all sample groups. Although *Tenacibaculum* have been reported in algae, corals, ascidians and sponges (16,46–48), they were not detected in such high abundance. Certain members of the genus *Tenacibaculum* can produce hydroxamate siderophores capable of chelating dissolved inorganic iron, an essential nutrient (47). Although iron does not always constitute a limited nutrient in deeper waters, due to iron remineralization and dissolved iron fluxes from benthic sediment (49,50), the high abundance of *Tenacibaculum* could indicate a symbiotic relationship where bacterial siderophores are utilised to chelate and scavenge ferric iron from the water column, although this has not been verified in this study. The previously reported production of (non-siderophore) metal-binding peptides by filter-feeding sponges does, however, support a crucial role for iron within sponge biology (51).

Other ubiquitous microbiome members were the nitrogen cycle related *Nitrospina* (family: Nitrospinaceae) and *Nitrosomonas* (family: Nitrosomonadaceae, not visible on heatmap), respectively, nitrite- and ammonia-oxidising bacteria (52,53). Microorganisms may contribute towards the nitrogen budget of sponges, and in carnivorous sponges the

presence of both *Nitrospina* and *Nitrosomonas* might indicate a coupled nitrification system facilitating the uptake of nitrogen during prey digestion (2). The presence of *Nitrospina* and *Nitrosomonas* should be further investigated to clarify the importance of these bacteria for nitrogen acquisition and remediation of toxic ammonia intermediates during prey digestion.

Most bacteria were not evenly distributed among anatomical regions. The most notable differences in bacterial abundance were observed within the sphere, where *Colwellia* (family: Colwelliaceae) and *Pseudoalteromonas* (family: Pseudoalteromonadaceae), both within the Alteromonadales order, were more abundant. Multiple species of *Colwellia* and *Pseudoalteromonas* can hydrolyse chitin, a glucose-derived polymer found in the crustacean exoskeleton (54,55). The selective retention of Alteromonadales could indicate the need for chitin hydrolysis in order to facilitate prey breakdown and digestion.

The *Roseobacter* NAC11-7 lineage (family: Rhodobacteraceae), enriched in the sphere and axis, is ecologically important in marine environments and has been found associated with diverse eukaryotic organisms, including sponges (56,57). They display a wide range of metabolic capabilities, including sulfur transformations, carbon monoxide oxidation and aromatic compound degradation. In addition, many *Roseobacter* species produce secondary metabolites, including antibiotic compounds effective against other marine bacteria and algae, and various toxic substances such as okadaic acid and paralytic toxins (56).

Finally, the root samples showed affinity for an uncultured bacterium of the Sva0996 marine group and *Robiginitomaculum* (family: Hyphomonadaceae). The presence of Sva0996 marine group has been reported in sponges; however, its functional and ecological aspects remain poorly understood. The presence of *Robiginitomaculum* is intriguing as this genus is thought to be a member of the hydrocarbon-degrading bacterial community (58). Besides the known function of marine sediments as natural hydrocarbon sinks, an increase in abundance and variation of cladorhizid species in the vicinity of hydrocarbon-enriched habitats, such as seeps, has previously been reported (59). Stable isotope analysis and detection of methanol dehydrogenase pathways in *Cladorhiza* sp. indicate that methanotrophic bacteria can play an important role for nutritional carbon acquisition in carnivorous sponges (60). Although the sponges sampled in this study were not situated near an evident seep, pockmarks of natural gas and methane hydrates have been observed in the Gulf of Maine (61). The hypothesis of utilizing hydrocarbon as nutrient source and the potential contribution of *Robiginitomaculum* to this process is intriguing and needs to be further explored.

While the sole presence of bacteria in a host is not sufficient to conclusively establish whether relationships are truly symbiotic, identifying bacteria that are shared between multiple host specimens can reduce the bias due to detection of transient bacterial species and uncover valuable information on bacterial residents. Here we show that the bacterial community composition of *C. grandis* includes several highly abundant bacteria present in the whole sponge and shared among sponge specimens, and that certain genera are enriched in specific anatomical regions. The anatomical compartmentalization of

biological functions, including cell proliferation and differentiation, has been suggested for the carnivorous sponge *Lycopodina hypogea* (15); our results could indicate a similar functional compartmentalization in *C. grandis* with bacteria likely making important contributions to host biology.

Based on data presented herein, we hypothesise that in *C. grandis* the sphere and axis communities facilitate prey capture and digestion whereas the root community might be involved in hydrocarbon breakdown. The diminished aquiferous system of *C. grandis*, in combination with the temporal/seasonal variability in copepod availability, could explain the need for this sponge to exploit all available food sources. Supplementing predation with nutrients from the water column and sediment and using bacterial symbionts as intermediates for these nutritional transactions could be an advantageous strategy for carnivorous sponges.

In conclusion, we report here the bacterial community associated with *C. grandis* and show that specific bacteria are retained in distinct anatomical regions. Multiple sponge individuals shared bacteria in sphere (*Colwellia*), axis (*Roseobacter*) and root (*Actinobacteria*, *Robiginitomaculum*) regions, suggesting biologically important symbiotic relationships in *C. grandis*. As many deep-sea species are difficult to maintain in vitro, molecular methods remain crucial to developing a basic understanding of the workings of these unique organisms.

3.6 Acknowledgments

The authors thank chief scientists A. Metaxas (Dalhousie University) and M. Nizinski (National Systematics Laboratory, National Museum of Natural History, Smithsonian Institution), the crew of the National Oceanographic and Atmospheric Administration (NOAA) ship Henry B. Bigelow, and the ROPOS team from the Canadian Scientific Submersible Facility. D. Goudie and S. Hill assisted with SEM work.

3.7 Funding

This work was supported by the Natural Sciences and Engineering Research Council of Canada Ship Time Grant [436808-2013] and Discovery Grant [RGPIN-2015-06548]. The research cruise was part of the Northeast Regional Deep-Sea Coral Initiative, funded in part by NOAA National Marine Fisheries Service, NOAA Office of Marine and Aviation Operations, and NOAA's Deep-Sea Coral Research and Technology Program.

3.8 References

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

2. Taylor MW, Radax R, Steger D, Wagner M. Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiol Mol Biol Rev* (2007) **71**:295–347. doi:10.1128/MMBR.00040-06

3. Webster NS, Blackall LL. What do we really know about sponge-microbial symbioses? *ISME J* (2009) **3**:1–3. doi:10.1038/ismej.2008.102

4. Harrington C, Del Casale A, Kennedy J, Neve H, Picton BE, Mooij MJ, O'Gara F, Kulakov LA, Larkin MJ, Dobson ADW. Evidence of bacteriophage-mediated horizontal transfer of bacterial 16S rRNA genes in the viral metagenome of the marine sponge *Hymeniacidon perlevis*. *Microbiology (Reading, Engl)* (2012) **158**:2789–2795. doi:10.1099/mic.0.057943-0

5. Rodríguez-Marconi S, De la Iglesia R, Díez B, Fonseca CA, Hajdu E, Trefault N. Characterization of bacterial, archaeal and eukaryote symbionts from Antarctic sponges reveals a high diversity at a three-domain level and a particular signature for this ecosystem. *PLoS ONE* (2015) **10**:e0138837. doi:10.1371/journal.pone.0138837

6. Taylor MW, Hill RT, Piel J, Thacker RW, Hentschel U. Soaking it up: the complex lives of marine sponges and their microbial associates. *The ISME Journal* (2007) **1**:187–190. doi:10.1038/ismej.2007.32

7. Bell JJ. The functional roles of marine sponges. *Estuarine, Coastal and Shelf Science* (2008) **79**:341–353. doi:10.1016/j.ecss.2008.05.002

8. Goeij JM de, Oevelen D van, Vermeij MJA, Osinga R, Middelburg JJ, Goeij AFPM de, Admiraal W. Surviving in a marine desert: the sponge loop retains resources within coral reefs. *Science* (2013) **342**:108–110. doi:10.1126/science.1241981

9. Lee WL, Reiswig HM, Austin WC, Lundsten L. An extraordinary new carnivorous sponge, *Chondrocladia lyra*, in the new subgenus *Symmetrocladia* (Demospongiae, Cladorhizidae), from off of northern California, USA. *Invertebrate Biology* (2012) **131**:259–284.

10. Hestetun JT, Vacelet J, Boury-Esnault N, Borchiellini C, Kelly M, Ríos P, Cristobo J, Rapp HT. The systematics of carnivorous sponges. *Molecular Phylogenetics and Evolution* (2016) **94**:327–345. doi:10.1016/j.ympev.2015.08.022

11. Vacelet J, Duport E. Prey capture and digestion in the carnivorous sponge *Asbestopluma hypogea* (Porifera: Demospongiae). *Zoomorphology* (2004) **123**:179–190. doi:10.1007/s00435-004-0100-0

12. Watling L. Predation on copepods by an Alaskan cladorhizid sponge. *Journal of the Marine Biological Association of the United Kingdom* (2007) **87**:1721–1726. doi:10.1017/S0025315407058560

13. Vacelet J, Boury-Esnault N (Universite de la M. A new species of carnivorous sponge (Demospongiae: Cladorhizidae) from a Mediterranean cave. in *Bulletin van het Koninlijk Belgisch Instituut voor Natuurwetenschappen - Biologie* Available at: http://agris.fao.org/agris-search/search.do?recordID=BE1997001471 [Accessed December 4, 2018]

14. Hansen GA. Spongiadae. The Norwegian North-Atlantic Expedition 1876-1878. *Zoology* (1885) Available at: http://www.marinespecies.org/aphia.php?p=sourcedetails&id=7585

15. Martinand-Mari C, Vacelet J, Nickel M, Wörheide G, Mangeat P, Baghdiguian S. Cell death and renewal during prey capture and digestion in the carnivorous sponge *Asbestopluma hypogea* (Porifera: Poecilosclerida). *J Exp Biol* (2012) **215**:3937–3943. doi:10.1242/jeb.072371

16. Dupont S, Corre E, Li Y, Vacelet J, Bourguet-Kondracki M-L. First insights into the microbiome of a carnivorous sponge. *FEMS Microbiol Ecol* (2013) **86**:520–531. doi:10.1111/1574-6941.12178

17. Dupont S, Carre-Mlouka A, Domart-Coulon I, Vacelet J, Bourguet-Kondracki M-L. Exploring cultivable Bacteria from the prokaryotic community associated with the carnivorous sponge *Asbestopluma hypogea*. *FEMS Microbiol Ecol* (2014) **88**:160–174. doi:10.1111/1574-6941.12279

18. Pawlik JR. The chemical ecology of sponges on Caribbean reefs: natural products shape natural systems. *BioScience* (2011) **61**:888–898. doi:10.1525/bio.2011.61.11.8

19. Núñez-Pons L, Avila C. Defensive metabolites from Antarctic invertebrates: does energetic content interfere with feeding repellence? *Mar Drugs* (2014) **12**:3770–3791. doi:10.3390/md12063770

20. Blunt JW, Copp BR, Munro MHG, Northcote PT, Prinsep MR. Marine natural products. *Nat Prod Rep* (2011) **28**:196–268. doi:10.1039/c005001f

21. Van Soest RWM, Boury-Esnault N, Vacelet J, Dohrmann M, Erpenbeck D, De Voogd NJ, Santodomingo N, Vanhoorne B, Kelly M, Hooper JNA. Global diversity of sponges (Porifera). *PLoS One* (2012) **7**: doi:10.1371/journal.pone.0035105

22. Verrill AE. Notice of recent additions to the marine Invertebrata of the northeastern coast of America, with descriptions of new genera and species and critical remarks on others. Part 1 Annelida, Gephyraea, Nemertina, Nematoda, Polyzoa, Tunicata, Mollusca, Anthozoa, Echinodermata, Porifera. (1879) Available at: http://repository.si.edu//handle/10088/12271 [Accessed December 4, 2018]

23. Comeau AM, Li WKW, Tremblay J-É, Carmack EC, Lovejoy C. Arctic ocean microbial community structure before and after the 2007 record sea ice minimum. *PLoS ONE* (2011) **6**:e27492. doi:10.1371/journal.pone.0027492

24. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* (2014) **30**:1312–1313. doi:10.1093/bioinformatics/btu033

25. Lanave C, Preparata G, Saccone C, Serio G. A new method for calculating evolutionary substitution rates. *J Mol Evol* (1984) **20**:86–93.

26. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* (2014) **30**:2114–2120. doi:10.1093/bioinformatics/btu170

27. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* (2014) **30**:614–620. doi:10.1093/bioinformatics/btt593

28. Mahé F, Rognes T, Quince C, Vargas C de, Dunthorn M. Swarm: robust and fast clustering method for amplicon-based studies. *PeerJ* (2014) **2**:e593. doi:10.7717/peerj.593

29. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* (2016) **4**: doi:10.7717/peerj.2584

30. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* (1990) **215**:403–410. doi:10.1016/S0022-2836(05)80360-2

31. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* (2007) **73**:5261–5267. doi:10.1128/AEM.00062-07

32. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, et al. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* (2010) **7**:335–336. doi:10.1038/nmeth.f.303

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

34. Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* (2012) **28**:3211–3217. doi:10.1093/bioinformatics/bts611

35. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* (2013) **41**:D590–D596. doi:10.1093/nar/gks1219

36. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glöckner FO. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res* (2014) **42**:D643-648. doi:10.1093/nar/gkt1209

37. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately maximum-likelihood trees for large alignments. *PLOS ONE* (2010) **5**:e9490. doi:10.1371/journal.pone.0009490

38. Faith DP, Baker AM. Phylogenetic diversity (PD) and biodiversity conservation: some bioinformatics challenges. *Evol Bioinform Online* (2007) **2**:121–128.

39. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* (2013) **8**:e61217. doi:10.1371/journal.pone.0061217

40. R Core Team. R: A language and environment for statistical computing. (2015) Available at: http://www.R-project.org/

41. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* (2014) **15**:550. doi:10.1186/s13059-014-0550-8

42. Warnes GR, Bolker B, Bonebakker L, Gentleman R, Liaw WHA, Lumley T, Maechler M, Magnusson A, Moeller S, Schwartz M, et al. *gplots: Various R programming tools for plotting Data*. (2016). Available at: https://CRAN.R-project.org/package=gplots [Accessed December 4, 2018]

43. Oksanen J, Guillaume Blanchet F, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin P, O'Hara RB, Simpson GL, Solymos P, et al. vegan: community ecology package. R package version 2.4-6. (2018) Available at: https://CRAN.R-project.org/package=vegan

44. Ondov BD, Bergman NH, Phillippy AM. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* (2011) **12**:385. doi:10.1186/1471-2105-12-385

45. Lundbeck W. "Porifera. (Part II.) Desmacidonidae (pars.)," in *The Danish Ingolf-Expedition* (Copenhagen: BiancoLuno), 1–219.

46. Carlos C, Torres TT, Ottoboni LMM. Bacterial communities and species-specific associations with the mucus of Brazilian coral species. *Sci Rep* (2013) **3**:1624. doi:10.1038/srep01624

47. Fujita MJ, Nakano K, Sakai R. Bisucaberin B, a linear hydroxamate class siderophore from the marine bacterium *Tenacibaculum mesophilum*. *Molecules* (2013) **18**:3917–3926. doi:10.3390/molecules18043917

48. Tianero MDB, Kwan JC, Wyche TP, Presson AP, Koch M, Barrows LR, Bugni TS, Schmidt EW. Species specificity of symbiosis and secondary metabolism in ascidians. *ISME J* (2015) **9**:615–628. doi:10.1038/ismej.2014.152

49. Boyd PW, Ellwood MJ. The biogeochemical cycle of iron in the ocean. *Nature Geoscience* (2010) **3**:675–682. doi:10.1038/ngeo964

50. Dale AW, Nickelsen L, Scholz F, Hensen C, Oschlies A, Wallmann K. A revised global estimate of dissolved iron fluxes from marine sediments. *Global Biogeochemical Cycles* (2015) **29**:691–707. doi:10.1002/2014GB005017

51. Guan LL, Sera Y, Adachi K, Nishida F, Shizuri Y. Isolation and evaluation of nonsiderophore cyclic peptides from marine sponges. *Biochemical and Biophysical Research Communications* (2001) **283**:976–981. doi:10.1006/bbrc.2001.4890

52. Kennedy J, Flemer B, Jackson SA, Morrissey JP, O'Gara F, O'Gara F, Dobson ADW. Evidence of a putative deep sea specific microbiome in marine sponges. *PLoS ONE* (2014) **9**:e91092. doi:10.1371/journal.pone.0091092

53. Ngugi DK, Blom J, Stepanauskas R, Stingl U. Diversification and niche adaptations of *Nitrospina*-like bacteria in the polyextreme interfaces of Red Sea brines. *The ISME Journal* (2016) **10**:1383–1399. doi:10.1038/ismej.2015.214

54. Garrity G, Brenner DJ, Krieg NR, Staley JR eds. *Bergey's Manual® of Systematic Bacteriology: Volume 2: The Proteobacteria, Part B: The Gammaproteobacteria*. 2nd ed. Springer US (2005). Available at: //www.springer.com/gp/book/9780387241449 [Accessed December 4, 2018]

55. Thomas T, Evans FF, Schleheck D, Mai-Prochnow A, Burke C, Penesyan A, Dalisay DS, Stelzer-Braid S, Saunders N, Johnson J, et al. Analysis of the *Pseudoalteromonas tunicata* genome reveals properties of a surface-associated life style in the marine environment. *PLoS ONE* (2008) **3**:e3252. doi:10.1371/journal.pone.0003252

56. Buchan A, González JM, Moran MA. Overview of the Marine *Roseobacter* Lineage. *Appl Environ Microbiol* (2005) **71**:5665–5677. doi:10.1128/AEM.71.10.5665-5677.2005

57. Wichels A, Würtz S, Döpke H, Schütt C, Gerdts G. Bacterial diversity in the breadcrumb sponge *Halichondria panicea* (Pallas). *FEMS Microbiol Ecol* (2006) **56**:102–118. doi:10.1111/j.1574-6941.2006.00067.x

58. Guibert LM, Loviso CL, Marcos MS, Commendatore MG, Dionisi HM, Lozada M. Alkane biodegradation genes from chronically polluted subantarctic coastal sediments and their shifts in response to oil exposure. *Microb Ecol* (2012) **64**:605–616. doi:10.1007/s00248-012-0051-9

59. Hestetun JT, Fourt M, Vacelet J, Boury-Esnault N, Rapp HT. Cladorhizidae (Porifera, Demospongiae, Poecilosclerida) of the deep Atlantic collected during Ifremer cruises, with a biogeographic overview of the Atlantic species. *Journal of the Marine Biological Association of the United Kingdom* (2015) **95**:1311–1342. doi:10.1017/S0025315413001100

60. Vacelet J, Fiala-Médioni A, Fisher CR, Boury-Esnault N. Symbiosis between methane-oxidizing bacteria and a deep-sea carnivorous cladorhizid sponge. *Marine Ecology Progress Series* (1996) **145**:77–85. doi:10.3354/meps145077

61. Rogers JN, Kelley JT, Belknap DF, Gontz A, Barnhardt WA. Shallow-water pockmark formation in temperate estuaries: A consideration of origins in the western gulf of Maine with special focus on Belfast Bay. *Marine Geology* (2006) **225**:45–62. doi:10.1016/j.margeo.2005.07.011

Table 3.1: Summary of high-throughput sequencing results, OTU generation and associated biological alpha-diversity parameters for bacterial OTUs extracted from distinct *Chondrocladia grandis* anatomical regions.

Sponge specimen	Region/ Sample Group	Paired-sequences	Obser-ved OTUs	Simpson Equitability (*E*)	Phylogenetic diversity (PD)	*E* group mean[a]	PD group mean[a]
1	Axis	63832	354	0.017	19.44	0.018	19.87
2	Axis	59822	253	0.022	21.25		
3	Axis	95156	394	0.017	21.73		
4	Axis	70966	272	0.015	17.06		
1	Root[b]	49218	5183	0.015	160.40	0.006*	41.48*
2	Root	68332	876	0.006	38.38		
3	Root	36577	920	0.007	42.89		
4	Root	93840	984	0.006	43.17		
1	Sphere	47634	367	0.026	20.45	0.021	20.22
2	Sphere	76956	364	0.022	15.55		
3	Sphere	76894	362	0.020	19.65		
4	Sphere	52082	265	0.015	25.24		

[a] Metric calculated after exclusion of root sample 1, at even rarefaction depth 36577.

[b] Sample excluded from downstream analysis due to uncharacteristically high OTU count and diversity.

* Value significantly different from other sample groups ($P < 0.05$).

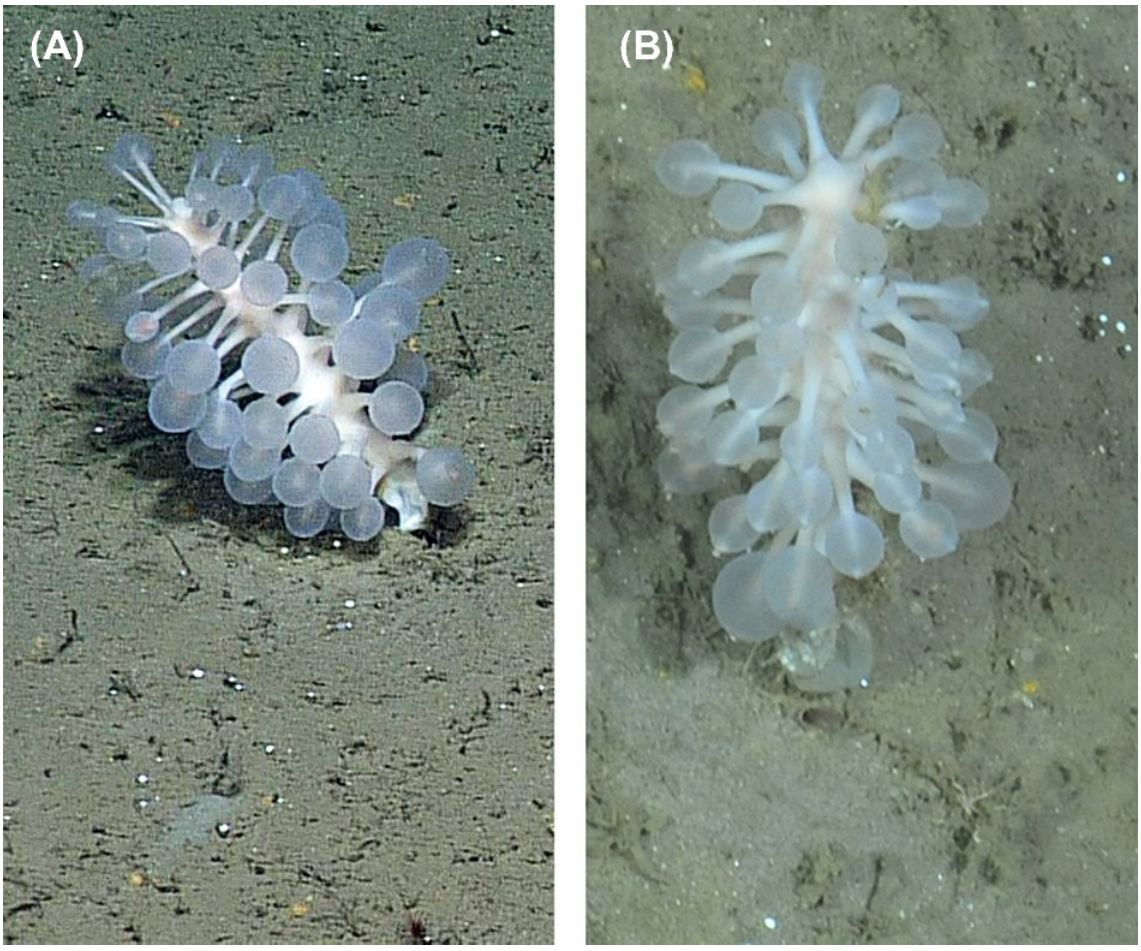

Figure 3.1: *Chondrocladia grandis* as observed in situ. (A) Side view, (B) top-down view.

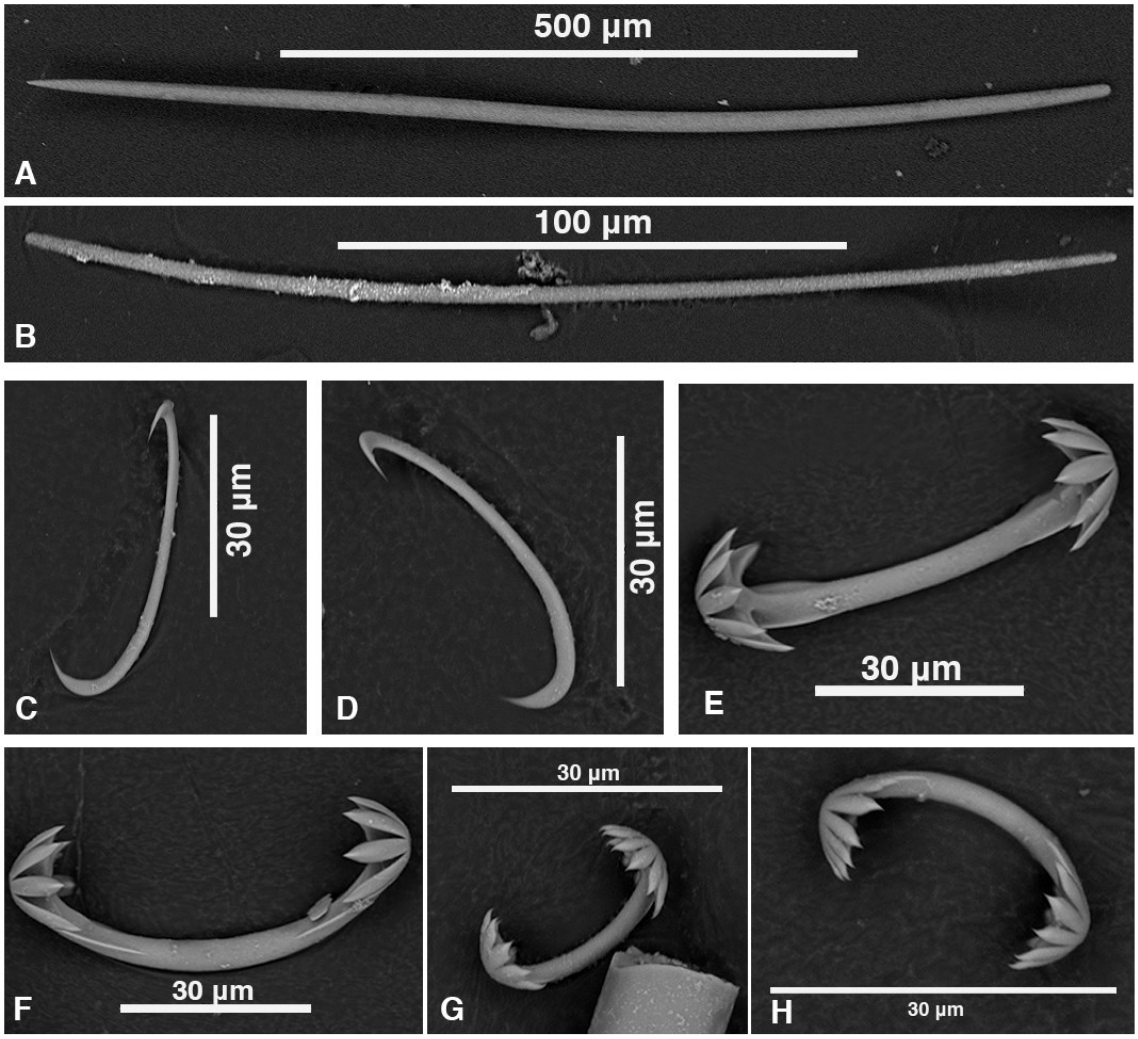

Figure 3.2: Scanning Electron Micrographs (SEM) of *Chondrocladia grandis* spicules. (A) Mycalostyles, (B) microspinose strongyles (C,D) sigmas (E,F) large and (G,H) small unguiferous anchorate isochelae.

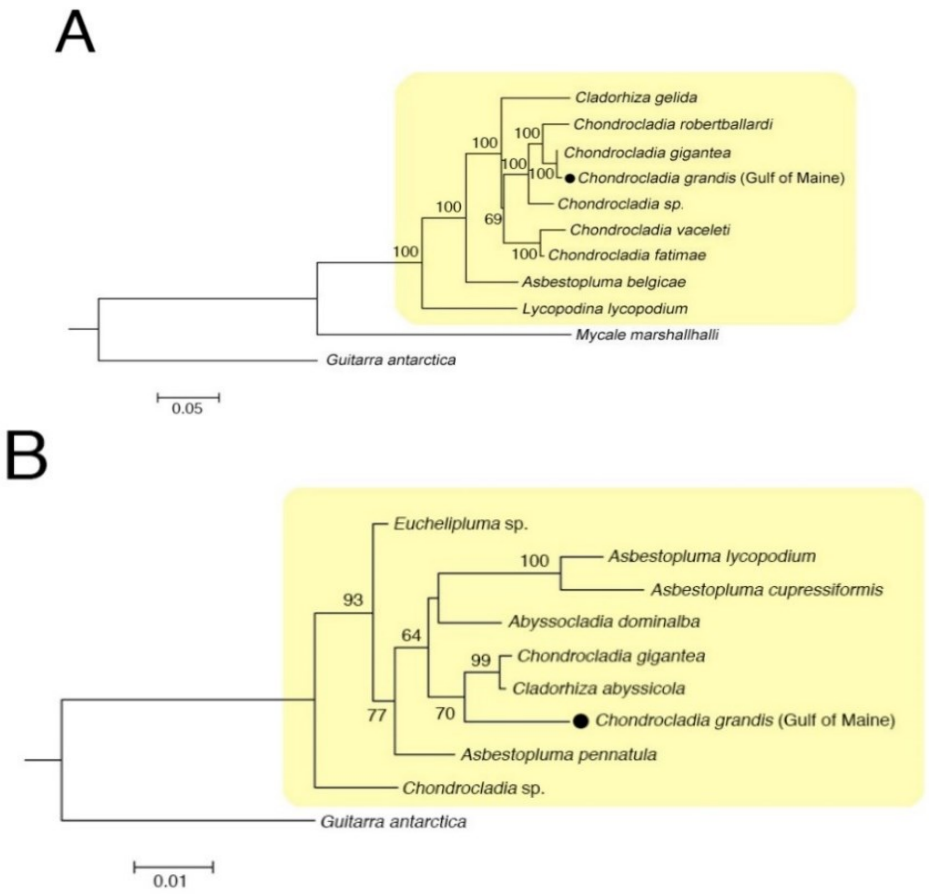

Figure 3.3: Maximum Likelihood analysis of (A) concatenated 28S rDNA, COI and ALG11 and (B) 18S rDNA alignments of *Chondrocladia grandis* with reference sequences. The evolutionary history was inferred with the GTR models and a discrete gamma distribution was used to model evolutionary rate differences among sites. The outcome of the bootstrap analysis is shown next to the nodes and only values higher than 75 are shown. Branch lengths are proportional to genetic distances as indicated by the scale bar. Sample collected in this study is indicated as Gulf of Maine and a black circle (KX950006- KX950009), all other species represent sequences obtained from GenBank (Accession numbers for concatenated alignments: LN870588.1, LN870518.1, LN870450.1, LN870535.1, LN870608.1, LN870469.1, LN870617.1, LN870558.1, LN870481.1, LN870559.1, LN870483.1, LN870619.1, LN870627.1, LN870567.1, LN870493.1, LN870490.1, LN870566.1, LN870626.1, LN870568.1, LN870628.1, LN870494.1, LN870635.1, LN870554.1, LN870503.1, LN870573.1, LN870642.1, LN870510.1, LN870647.1, LN870515.1, LN870575.1, Accession numbers for 18S rDNA alignments: LN870434.1, LN870435.1, LN870429.1, LN870430.1, LN870431.1, LN870432.1, LN870436.1, LN870433.1, LN870437.1). The clade including carnivorous sponges is in yellow.

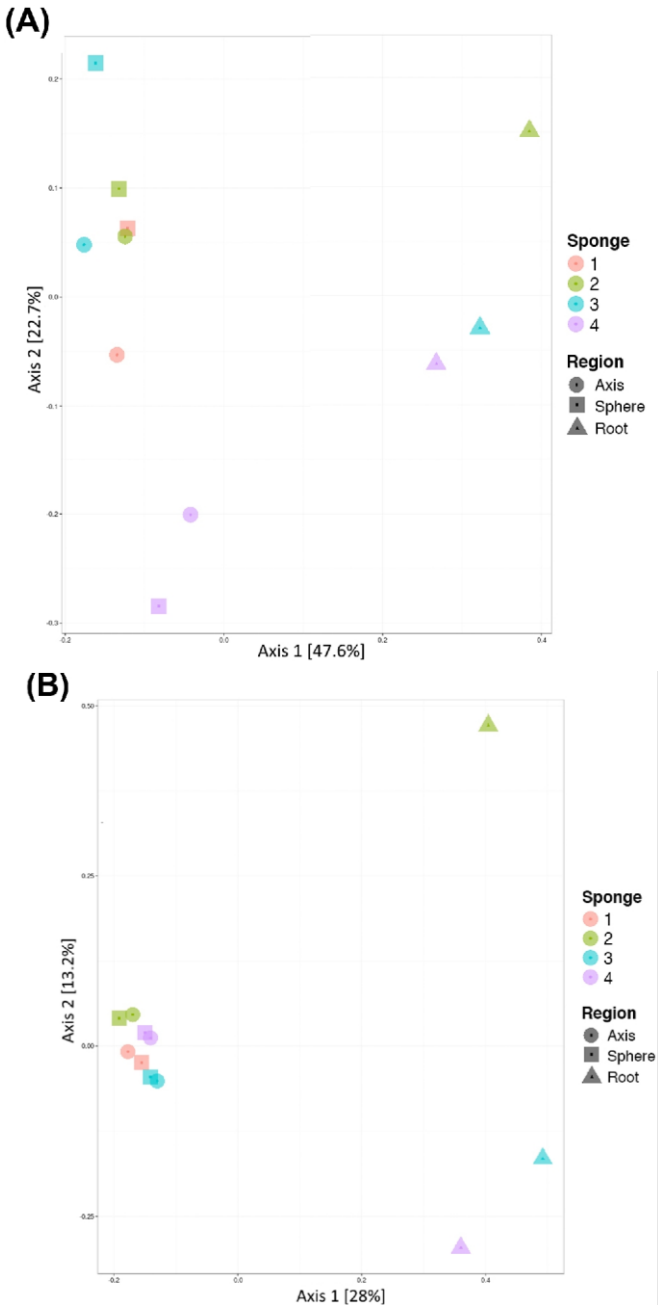

Figure 3.4: PCoA of bacterial community composition in carnivorous sponge samples, based on the Bray–Curtis dissimilarity. Analyses were performed on (A) proportional count data and (B) variance stabilized transformation pretreated count data. The percentage of dataset variation that each principal coordinate represents is shown on each axis. Shapes indicate type of sample (anatomical region), while colors represent the sponge specimen from which the sample originated.

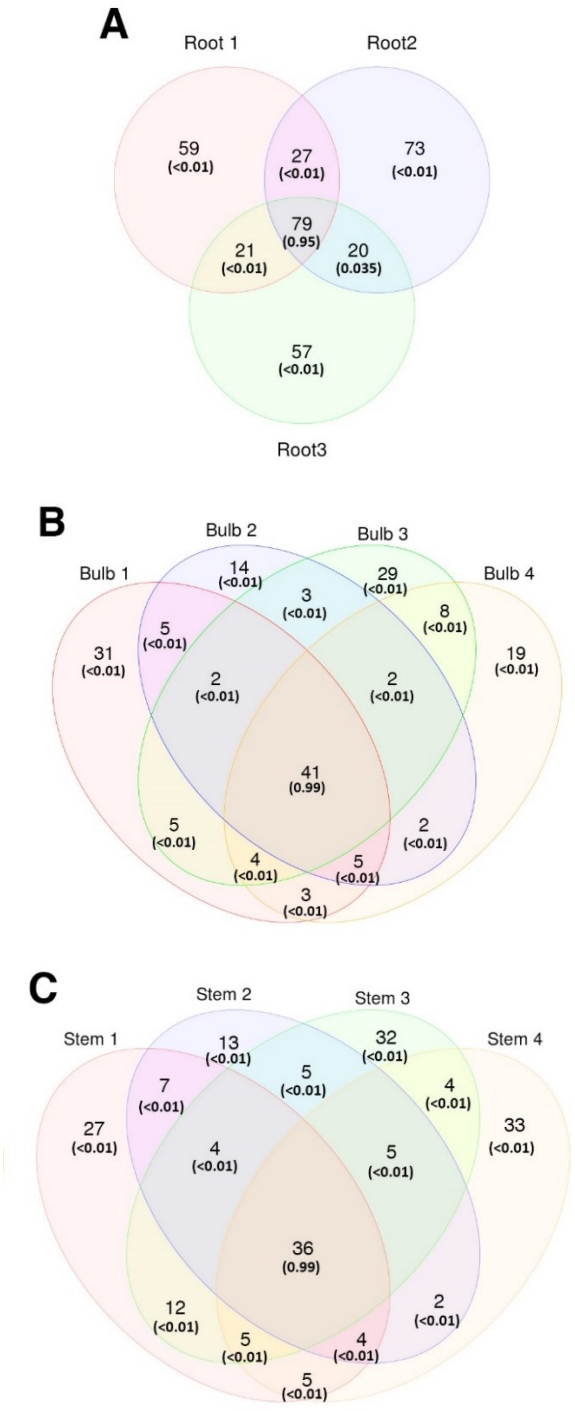

Figure 3.5: Relatedness of bacterial genera present in different anatomical regions in *Chondrocladia grandis*. Indicated within Venn diagrams are the number of bacterial genera present and the total proportion of sequences attributed to these genera (in parenthesis) shared amongst individual samples (from different sponge specimens) of the root (A), sphere (B) and axis (C).

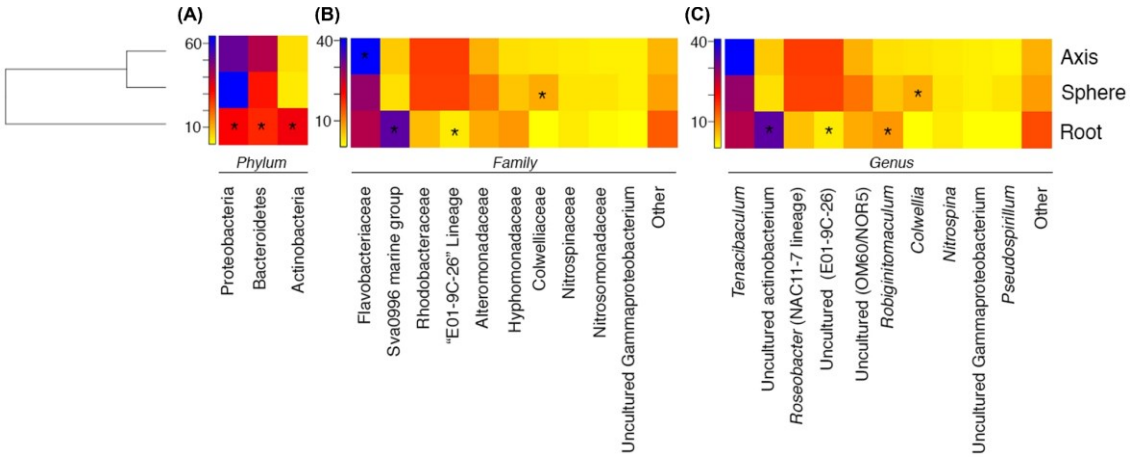

Figure 3.6: Heat map visualization of abundance of the most prevalent bacterial taxa found in distinct anatomical regions of *C. grandis* at the phylum (A), family (B) and genus (C) levels. Count data from individual samples were collated based on anatomical region. A dendrogram constructed from variance stabilized OTU count data using hierarchical clustering of the Bray–Curtis dissimilarity is displayed on the left. Coloring of squares indicates OTU abundance (in relative percentage of sequence abundance) of each taxon within each anatomical region. Stars indicate, for a given taxon, the anatomical region in which the abundance of that taxon differs significantly from that in the other two regions ($P < 0.05$).

CHAPTER 4: MICROBIOMES OF THE ARCTIC CARNIVOROUS SPONGES *CHONDROCLADIA GRANDIS* AND *CLADORHIZA OXEATA* SUGGEST A SPECIFIC, BUT DIFFERENTIAL INVOLVEMENT OF BACTERIAL ASSOCIATES

Adapted from the published version available in Arctic Science[5]
https://doi.org/10.1139/as-2017-0015

4.1 Abstract

While sponges are generally known to host a wide range of microbial associates, the composition and specificity of the microbial communities in carnivorous sponges are poorly understood. We used 16S rRNA gene data to examine and compare the bacterial communities associated with distinct anatomical regions of two carnivorous sponge species, *Chondrocladia grandis* and *Cladorhiza oxeata*, sampled from Baffin Bay and the Gulf of Maine (*C. grandis* only). The two sponge species hosted distinct bacterial communities, with taxonomic diversity being greater in *C. grandis*. Some bacterial taxa (including particular oligotypes) were consistently recovered in multiple host individuals from geographically distant sites, suggesting specificity. Within *C. grandis*, certain bacterial taxa were enriched in particular anatomical regions, suggesting functional roles in carnivorous sponge metabolism or other biological processes. Stable isotope analysis provided no evidence for methanotrophy in the sponges examined, but Gulf of Maine *C. grandis* might incorporate ^{13}C-depleted carbon via the bacteria-mediated heterotrophic degradation of other hydrocarbons. Overall, our results demonstrate that the carnivorous

[5] This version was modified to include supplementary figures and tables in the main text

sponge microbiome appears host species specific and stable, even over large geographical areas. The observed differences in bacterial community composition and diversity between *C. grandis* and *C. oxeata* may reflect differences in trophic adaptability, specialization, and overall reliance on associated bacteria.

4.2 Introduction

The Porifera (commonly known as sponges) is a unique phylum of ancient metazoans, often defined by their ability to efficiently obtain nutrients by filtering water through an aquiferous system. They are a key macrobenthic component in many environments and perform important ecological functions, including the enhancement of biochemical fluxes between benthic and pelagic zones and the conversion of dissolved organic matter into particulate organic matter (1). The ecological importance of marine sponges is thought to be driven in part by their ability to readily associate with microorganisms. Indeed, many studies have uncovered the vast diversity of microbes associated with different sponge species. Such associations can constitute symbiotic relationships and are thought to contribute to filter-feeding efficiency. Metabolic exchanges between sponges and their symbionts have been observed in processes such as nitrogen cycling, CO_2 fixation, and secondary metabolite production (2).

Although most sponges rely on filter-feeding, the carnivorous sponges (family Cladorhizidae) show a diminished or absent aquiferous system and are uniquely capable of capturing and engulfing prey, which range from planktonic organisms to small crustaceans. Carnivorous sponges show several morphological adaptations that enable this feeding

strategy, including an erect body morphology and structures with adhesive surfaces and filaments that can ensnare live prey. Following prey capture, motile sponge cells migrate to the capture site and quickly engulf the prey, creating a temporary cavity in which digestion can occur (3). The associated microbial community of the sponge is presumed to play a large role in this process, as the digestion of prey and production of biologically available nutrients are thought to be facilitated by symbiotic bacteria (4), which are actively shuttled to these digestive cavities (5). Interestingly, methanotrophic *Methylococcales* were recently identified within *Cladorhiza methanophila* (3), supporting previous reports of this Cladorhizidae species being capable of supplementing a carnivorous diet through chemolithoautotrophic symbiosis (6,7).

Carnivorous sponges have been observed in different environments, ranging from littoral caves to the deep sea, but are thought to have originally evolved to overcome the limitations of filter-feeding in deep-sea oligotrophic niches (5,8). However, our current understanding of the global biogeography of the Cladorhizidae is still rudimentary, although a bipolar cold-water distribution pattern has been suggested, and a high number of species have been reported within northeast Atlantic and Arctic regions (9). Cladorhizids are commonly encountered at shallow to mid-bathyal depths at high latitudes (10) and may be dominant members of certain benthic communities. For example, Chondrocladia was a defining taxon for one of 10 biotopes identified in the Norwegian MAREANO seabed mapping program; this biotope was associated with mixed sediments along canyons and steep slopes at a mean depth of 1390 m (11). The ecological importance of carnivorous

sponges in Arctic food webs remains poorly understood given our limited knowledge of their trophic strategies and microbial symbiotic relationships.

The overarching goal of this work was to describe similarities and differences in the bacterial communities (from here on colloquially referred to as the microbiomes) associated with multiple individuals of *Chondrocladia grandis* and *Cladorhiza oxeata* sampled from Baffin Bay and from the Gulf of Maine (*C. grandis* only). First, we compare microbiome diversity across all samples and determine which factors best explain assemblage composition. We then identify bacterial biomarkers for each anatomical region of these carnivorous sponges (i.e., taxa that are consistently and significantly enriched within distinct body regions) and highlight certain bacterial community members that may have particular functional importance. We also investigate the stability of selected common and biomarker taxa at the oligotype level, across host specimens from various geographical regions and depths. Finally, we employ stable isotope analysis of $\delta^{13}C$ and $\delta^{15}N$ to investigate the potential for chemoautotrophic symbiosis within these sponges.

4.3 Materials and methods

4.3.1 Sponge description and sample collection

Arctic carnivorous sponges (Table 4.1) were collected off the coast of Baffin Island (Qikiqtaaluk Region (Inuktitut: ᖃᑭᖅᑕᐃᒪᐅᖅ), Nunavut, Canada) using the Super Mohawk II (SUMO) ROV launched from the Canadian Coast Guard ship Amundsen during the ArcticNet 2015 multidisciplinary scientific cruise. Upon successful collection from the seafloor, sponges were immediately placed on ice before being dissected for subsample

collection. Dissection was performed using sterile scalpels and anatomically distinct regions were excised. *Chondrocladia* sponges (one from Scott Inlet and two from Davis Strait, southeast of Broughton Island) displayed a single central stem morphology (referred to here as axis) with numerous side branches (stems) terminating in inflated spheres (Figure 4.1A). From its terminal end, the axis changed in appearance roughly 5 cm from the seafloor before protruding into the sediment; we termed this structure (both above and below the sediment) root. Within the sediment, branching tendrils from the root (root tips) were observed and sampled. Furthermore, multiple white, oval inclusions were found embedded within the axis of two specimens and dissected for further histological and molecular investigation. Replicate subsamples of each anatomical region, axis, inclusions, stems, inflatable spheres, root, and root tips, were collected where possible (Table 4.2). *Cladorhiza* specimens (one from Scott Inlet and one from Davis Strait) displayed a "bush-like" morphology, with a larger base axis branching into thinner secondary axes; unlike *Chondrocladia*, *Cladorhiza* were attached to rocks, together with larger epifauna such as anemones, and had no roots. Multiple, evenly spaced shorter branching stems terminating in a slight swelling ("sphere") were seen on the secondary axes (Figure 4.1B). As in *Chondrocladia*, oval inclusions were seen within the base axis and sampled. Replicate subsamples of each anatomical region, base axis, secondary axis, the tip or "sphere" of secondary axis, and inclusions, were collected where possible (Table 4.2). All subsamples were immediately stored in cryovials at −80 °C until further processing. Furthermore, four specimens of *C. grandis* from the Northeast Channel of the Gulf of Maine (Canada) described in a previous study (4) were included in our analysis (Table 4.1).

4.3.2 Identification of sponge specimens

Sponge subsamples were pretreated and nucleic acids extracted as previously described (4). Taxonomy of the collected carnivorous sponges was established by sequencing the host 18S rRNA gene, the overlapping "Folmer" and "Erpenbeck" mitochondrial COI gene fragments, and part of the asparagine-linked glycosylation 11 homolog (ALG11) gene, amplified using primers described in Hestetun et al. (10). PCR products were obtained using Dream Taq polymerase (Thermo Fisher Scientific, Waltham, Massachusetts, USA), purified using Agencourt AMPure XP beads (Beckman Coulter, Brea, California, USA), and outsourced for Sanger sequencing to the Centre for Applied Genomics (Hospital for Sick Children, Toronto, Ontario, Canada). Phylogenetic analysis was performed as described earlier (4). Sequences were deposited in the NCBI GenBank database with accession numbers MG193128–MG193135.

Spicules were isolated from specimens by immersing sponge tissues in 10% sodium hypochlorite. Following tissue dissolution, the remaining spicules were carefully rinsed with distilled water and 100% ethanol before being transferred to stubs and imaged using a Quanta 650 scanning electron microscope (FEI Company, Hillsboro, Oregon, USA).

4.3.3 High-throughput 16S rRNA gene sequencing

Bacterial communities were sequenced by Illumina MiSeq 2 × 300 bp paired-end sequencing of the V6–V8 region of the 16S rRNA gene, amplified using universal bacterial primers B969F and B1406R (12,13), at the Integrated Microbiome Resource of the Centre

for Comparative Genomics and Evolutionary Bioinformatics (Dalhousie University, Halifax, Nova Scotia, Canada).

4.3.4 Sequence processing and analysis

Illumina reads from Baffin Bay specimens obtained in this study and from *C. grandis* obtained previously from the Gulf of Maine (4) were analyzed with the in-house developed SPONS-2 (Streamlined Processor Of Next-gen Sequences, version 2) pipeline. This includes the creation of operational taxonomic units (OTUs), taxonomic assignment, and basic filtering. Briefly, sequences were quality checked using Trimmomatic version 0.33 to remove low-quality bases (sliding window of 20 bases, minimum average quality per base of 15) and low length reads (<100 bases) were filtered out (14). Remaining reads were subsequently merged using PEAR version 0.9 (using modified settings of minimal overlap required of 105 bp, p value required <0.01, and minimum and maximum fragment length of 300 and 500 bases, respectively) (15). CutAdapt was then used to trim primers from merged reads (maximum error rate 0.2) while simultaneously dropping reads in which forward or reverse primers were not detected (16). A final quality check was performed and reads with an average Phred quality below 30 were removed as well as reads containing more than 10 bases at any position dropping below a Phred score of 15. Subsequently, OTUs were created as described previously (4). Taxonomy was assigned to OTUs by comparing sequences against the SILVA SSU database (release 128) using VSEARCH (version 2.0.3, using the USEARCH global implementation, minimum pairwise identity of 0.92, with an unlimited number of maximum number of hits to accept and a maximum for nonmatching sequences of 32, and edit distance was set to pairwise identity definition) and

subsequently collecting the lowest common ancestor amongst database hits (17). Full dataset phylogeny was performed as previously described (4).

Count data, taxonomic classification, and full dataset phylogeny were exported to R for further analysis using the PhyloSeq R package (18). Additional filtering included the removal of spurious phyla (phyla with <10 reads across the full dataset) and removing OTUs presumed to be sequencing "noise" (retaining OTUs observed at least twice across the dataset). Rarefaction curve analysis was performed using the "observed" and "Chao1" alpha diversity metrics, measuring 10 replicates at sequence depths 1, 10, 100, and 1000 and continuing at 1000 sequence intervals up to a sequence depth of 42000. Alpha diversity analysis of bacterial communities was performed using the PhyloSeq estimate richness command, using the metric "observed" to visualize the amount of OTUs observed, the "Chao1" index to estimate total community richness, and the inverse of the Simpson's index "invSimpson" to estimate community evenness. An unpaired t test and an ANOVA followed by a Tukey's test were used to determine any significant (defined as a $p < 0.05$) differences of alpha diversity measurements between sponge species and anatomical regions, respectively. To compare bacterial communities among individual sponge specimens and anatomical regions, ordination analysis was performed on log-transformed datasets using the PhyloSeq implementation of the Jensen–Shannon divergence to generate distance matrices and ordinated using nonmetric multidimensional scaling available in the R package "vegan" (19). Permutational multivariate ANOVA was performed using the "Adonis" command, also available in the "vegan" package (19); this approach allowed us to examine how the variation among microbiomes could be attributed to different meta-

data factors, defined as host species, host individual, anatomical region, and sampling site (Davis Strait, Scott Inlet (Arctic) or Gulf of Maine). Factors explaining variation (their importance reported herein as effect-size, also known as R^2) with a $p < 0.05$ were considered statistically significant.

4.3.5 Biomarker analysis of anatomical regions

LEfSe analysis (20) was used to detect any biomarkers (taxa that are differentially abundant and explain, with statistical significance, the differences among multiple microbial communities) specific to unique anatomical regions of sponges. OTU count data were filtered to remove OTUs with low abundance (<100 reads in total) and subsequently converted to relative abundance. Biomarkers were considered valid when significant ($p < 0.01$) and when the linear discriminant analysis score was >3.

4.3.6 Oligotype analysis

To examine host specificity in bacterial genera that are common and/or biomarkers for particular anatomical regions, oligotyping was performed using the Oligotyping-pipeline version 2.1 (21). Original Illumina reads belonging to genera of interest were extracted and used for entropy analysis followed by oligotyping analysis. To reduce the effect of sequencing noise, only oligotypes with a minimum abundance of 100 were retained.

4.3.7 Histology of sponge inclusions

Excised sponge inclusions were fixed in 2.5% glutaraldehyde in a 0.1 mol/L sodium cacodylate – 1% HCl buffer for 48 h at 4 °C. Once fixed, tissues were dehydrated using an

ascending ethanol series and cleared using xylene substitute (VWR, Radnor, Pennsylvania, USA). Cleared tissues were paraffin embedded and sections (6 μm) were cleared in xylene substitute, rehydrated, and subsequently stained using hematoxylin with eosin counterstain.

4.3.8 Stable isotope analysis

Stable isotope ratio analysis on $\delta^{13}C$ and $\delta^{15}N$ was performed on subsamples of all collected sponge specimens from Baffin Bay (frozen at −80 ˚C) as well as ethanol-preserved specimens from the Gulf of Maine. Sponge tissues were dried for 24 h at 80 °C in acid-washed (0.5 mol/L HCl) glass vials. Once dry, samples were grounded to powder using a ceramic mortar and pestle. Lipid removal was performed by adding a mixture of methanol, chloroform, and distilled water (5:3:2) and incubating for 2 h at room temperature, after which the organic phase was discarded, samples rinsed with distilled water, and dried again at 80 °C. Once dry, 0.5 mol/L HCl was added to samples to remove inorganic carbon and samples were incubated for 5 min at room temperature. Finally, acid waste was removed by washing the samples with distilled water, and samples were dried for 24 h at 80 °C. Treated and dried samples were weighed in tin capsules and measured using a Delta V isotope ratio mass spectrometer (Thermo Fisher Scientific). After calibrating against internal standards, isotope ratios were determined relative to Vienna Pee Dee Belemnite for $\delta^{13}C$ and atmospheric air for $\delta^{15}N$. The variation and analysis uncertainty based upon standards was <0.1‰.

4.4 Results

4.4.1 Verification of sponge taxonomy

Sponge identification based on gross morphology was confirmed by comparing host gene sequences to reference sequences in NCBI GenBank. Host genes for Baffin Bay *Chondrocladia* where identical to reference *C. grandis* sequences (18S: KX950006, 28S: KX266202, COI: KX266219, ALG11: KX266228). Identical sequences were also found for *Cladorhiza*, verifying the taxonomy as *C. oxeata* (28S: KX266195, COI: KX266209, ALG11: KX26625). No 18S sequence from *C. oxeata* was present in the NCBI database and the closest relative for this gene was *Chondrocladia gigantea* (99%: LN870434). Phylogenetic analysis further supported the taxonomic assignment (Figure 4.2). A tentative analysis of spicules in Baffin Bay *C. grandis* revealed megascleres (mycalostyles), s-shaped sigmas, and unguiferous anchorate isochelae, which displayed six teeth at each extremity. A larger diversity of spicules was observed in *C. oxeata*, including a combination of megascleres (mycalostyles, strongyles, and sybtylostyles), c- and s-shaped sigmas, centrotylote oxea, unguiferous anchorate isochelae (six teeth on each extremity), and multidentate anisochelas (five teeth on each extremity).

4.4.2 Histology of sponge inclusions

Sponge inclusions measured approximately 4.76 ± 0.72 and 3.81 ± 0.19 mm in diameter for *C. grandis* and *C. oxeata*, respectively. Gross morphology of sponge inclusions was determined for *C. grandis* using light microscopy and revealed a central core of various cell types, including amoeboid cells suspended in apparent collagen fibers. Surrounding the interior was a circumferential layer of elongated, squamous-like cells

surrounded by an outer layer of columnar, stratified appearance. The outer layers were further lined with numerous anchorate isochelae. These structures resemble previously identified embryos (22–24) and are here identified as such.

4.4.3 High-throughput sequencing data inspection

High-throughput 16S rRNA gene reads included in the analyses were obtained from three *C. grandis* and two *C. oxeata* specimens collected in Baffin Bay and compared to reads obtained during a previous study on four *C. grandis* specimens collected in the Gulf of Maine (4). In total, these included 53 subsamples from sectioned anatomical regions. Approximately 3 million read-pairs were considered, 2.5 million of which passed assembly and subsequent quality controls. On average, about 47600 sequences were recovered per sample (Table 4.2). Dereplication, clustering, and chimera removal resulted in the creation of 5736 high-quality nonsingleton bacterial OTUs distributed amongst analyzed samples. Rarefaction curve analysis indicated that all samples reached or tended towards their respective alpha diversity plateaus, suggesting adequate sequencing depth (Figure 4.3).

4.4.4 Bacterial community diversity and similarity

Analysis of bacterial community richness (Figure 4.4), comparing the number of observed OTUs, Chao1, and mean Simpson's reciprocal index (1/*D*), showed considerable variation amongst samples with observed OTUs ranging from 41 to 1611, Chao1 values between 55 and 1771, and 1/*D* in the range of 1.28 to 17.1. Overall, *C. grandis* (Figure 4.4A) showed a significantly higher mean value for the number of observed OTUs (335, $p = 0.003$) and Chao1 (392, $p = 0.002$) compared to *C. oxeata* (114 and 142, respectively)

(Figure 4.4B) in addition to a highly significant difference ($p < 0.001$) in mean $1/D$ (5.37 as compared to 2.14 in *C. oxeata*).

Considering anatomical regions of *C. grandis* separately, significantly ($p < 0.05$) greater observed diversity and evenness were found within root and root tip samples in contrast to axis, sphere, and stem sections (Figure 4.4A; Table 4.3). Embryonic samples had the lowest observed diversity and Chao1 values and the highest mean inverse Simpson's index (Figure 4.4A; Table 4.3), indicating a more evenly distributed bacterial community with fewer rare taxa.

In *C. oxeata*, both the observed number of taxa and the Chao1 metric varied widely among anatomical regions (Figure 4.4B; Table 4.3). However, both base axis and secondary axes had a higher estimated evenness according to the inverse Simpson's metric, likely due to a large proportion of rare taxa within spheres, embryos, and stems (Figure 4.4B; Table 4.3).

Nonmetric multidimensional scaling ordination analysis revealed a systematic difference in bacterial community composition as a function of sponge species, clearly visible as two distinct clusters (Figure 4.5A). ANOVA further indicated host species as the highest scoring significant variable explaining variation within the dataset (effect size of 37%, $p < 0.001$). After repeating the ordination analysis restricted to the *C. grandis* sample cluster (Figure 4.5B), further grouping was observed based on anatomical region, with root and root tip samples separate from axis, stem, sphere, and embryo samples. Within these clusters, grouping based on sampling location was also observed. ANOVA within *C.*

grandis samples revealed a highly significant effect size for anatomical region and sampling location (effect size of 40% and 22%, respectively, $p < 0.001$), while host individual had a more moderate impact (effect size 8%, $p = 0.012$). In contrast, within *C. oxeata* (Figure 4.5C), no clear pattern was observed. Indeed, the largest significant explanatory variable for variability in the dataset was the host individual (effect size of 22%, $p = 0.003$).

4.4.5 Taxonomic composition of bacterial communities

The taxonomic assignment of created OTUs was distributed amongst 35 unique bacterial phyla, with the vast majority (93%) being assigned to either the Proteobacteria or the Bacteroidetes (Figure 4.6A). Variations amongst the host species, anatomical regions, and geographical regions were observed, including a higher relative abundance of Bacteroidetes in *C. oxeata*, an increase in Actinobacteria within root samples of *C. grandis* from the Gulf of Maine, and a higher relative abundance of Proteobacteria within Arctic *C. grandis* in contrast with the Gulf of Maine specimens.

On the class level, a clear distinction between *C. oxeata* and *C. grandis* was observed (Figure 4.6B). Within *C. oxeata* specimens, the Flavobacteriia were highly abundant, while other classes exhibited a more variable and less abundant profile, in line with the low alpha diversity measurements. Commonly encountered OTUs were distributed amongst a more diverse set of classes within *C. grandis*, with Flavobacteriia, Gammaproteobacteria, and Alphaproteobacteria being particularly abundant.

Gammaproteobacteria were highly abundant in all Arctic *C. grandis*, while this class was less represented within the Gulf of Maine specimens.

Taxonomic distribution analysis on the family level revealed a diverse bacterial community within *C. grandis* samples (Figure 4.6C). High abundances of Flavobacteriaceae, especially within the Gulf of Maine specimens, were detected. Other families were present at fluctuating rates, including the Rhodobacteraceae, Halieaceae, the "Sva0996 marine group" of the Acidimicrobiia, Colwelliaceae, and Hyphomonadaceae. Within *C. oxeata*, a bacterial community mainly dominated by Flavobacteriaceae was observed. The Rhodobacteraceae, Nitrosomonadaceae, Flammeovirgaceae, and Phyllobacteriaceae were abundant in some samples but not ubiquitous amongst all *C. oxeata* samples.

4.4.6 Biomarkers of anatomical regions

We investigated the potential for enrichment of specific bacterial taxa (biomarkers) within anatomical regions for both *C. grandis* and *C. oxeata*. Although no statistically significant compartmentalization of taxa was observed in *C. oxeata*, 45 taxa with significant differences across anatomical regions were detected within *C. grandis* (Figure 4.7). Of particular interest were the *Colwellia* (Colwelliaceae) and *Reichenbachiella* (Flammeovirgaceae) observed within the trapping spheres. However, the majority of biomarker taxa (N = 39) were found in the root and root tip. Assignments at the genus level for these taxa included the *Robiginitomaculum*, *Maritimimonas*, *Granulosicoccus*, *Mycobacterium*, *Anderseniella*, and *Nitrospira*.

4.4.7 Oligotyping of common bacterial associates

Oligotyping was performed on selected bacterial genera common to both species (*Tenacibaculum*, *Candidatus* Branchiomonas, *Fulvivirga*, and *Pseudahrensia*) as well as all identified biomarkers in all anatomical regions for which taxonomy was resolved to genus level (*Robiginitomaculum*, *Colwellia*, *Maritimimonas*, *Granulosicoccus*, *Mycobacterium*, *Anderseniella*, *Reichenbachiella*, and *Nitrospira*).

Two oligotypes were detected for *Tenacibaculum* (Figure 4.8A). Oligotype G was almost exclusively found in *C. oxeata*, while oligotype A was highly abundant in *C. grandis* (in both Gulf of Maine specimens and Arctic specimens) but rare in *C. oxeata*. A similar distribution was observed for *Candidatus* Branchiomonas (Figure 4.8B), where oligotype G was found in high abundance among *C. oxeata* specimens whereas oligotype A was only abundant in *C. grandis*. Seven *Fulvivirga* oligotypes were detected. Interestingly, within *C. grandis* specimens, only one of the seven oligotypes (CGTCA) was abundant whereas in *C. oxeata*, this type was not present and a diverse assemblage of other oligotypes was found instead (Figure 4.8C). Finally, a high noise level was consistently observed in the entropy analysis of Pseudahrensia and the oligotype composition could not be successfully resolved for this genus.

Within *C. grandis*, we detected five oligotypes of the sphere associated biomarker genus *Colwellia* (Figure 4.8D), two of which (oligotypes CGCG and TGCG) accounted for the vast majority of *Colwellia* sequences. Three other oligotypes (CACG, CGCA, and CATG) were recovered at lower abundances and only from *C. grandis* sampled in Davis

Strait. Two oligotypes of *Reichenbachiella* (Figure 4.8E) were detected with an apparent mixed abundance across *C. grandis* spheres, independent of sampling location. The majority of the root-associated *Maritimimonas* sequences were found in arctic *C. grandis* and composed of oligotypes TA and CA (Figure 4.8F), the latter of which was also found to be the sole oligotype found in Gulf of Maine *C. grandis*. Finally, only one ubiquitous *Robiginitomaculum* oligotype was detected. Oligotyping of *Granulosicoccus*, *Mycobacterium*, *Anderseniella*, and *Nitrospira* did not produce satisfactory results due to a low number of reads (<1000 reads total).

4.4.8 Stable isotope analysis

Stable isotope analysis showed a comparable signature for *C. grandis* and *C. oxeata* specimens collected at Arctic sampling sites, with mean $\delta^{13}C$ values of −24.78‰ ± 0.68‰ (N = 5) and −24.88‰ ± 0.68‰ (N = 5), respectively. However, *C. grandis* specimens collected within the Gulf of Maine showed a significantly lighter (unpaired t test, $p < 0.001$) $\delta^{13}C$ mean value of −33.27‰ ± 2.53‰ (N = 5), separating them from other sponge specimens. Mean $\delta^{15}N$ values were similar across all groups with no large-scale differences observed between *C. grandis* (10.6‰ and 11.22‰ for Gulf of Maine and Arctic samples, respectively) and *C. oxeata* (12.83‰).

4.5 Discussion

In deep-sea and Arctic environments, carnivorous sponges are often seen instead of filter-feeding sponges, but their ecological functions within these environmental niches are largely unknown. Previous microbiological studies on both filter-feeding and carnivorous

sponges elucidated that several biological functions of sponges are, to some extent, mediated by associated (symbiotic) microbial communities (2). In this study, we aimed to describe the microbial communities associated with two carnivorous sponge species, *Chondrocladia grandis* and *Cladorhiza oxeata*. Furthermore, we investigated the variability of these communities across two geographical regions and the possibility of host, species, and anatomical region specificity among associated bacteria.

4.5.1 Bacterial community diversity

Significant, systematic differences between the bacterial communities of the two sponge species were observed, highlighted by the formation of species-specific clusters within our ordination analysis (Figure 4.5). As the investigated sponges are part of different genera, differences in microbiome composition are not unexpected (2). However, for *C. grandis*, a significant difference in bacterial composition among anatomical regions was also seen, with unique assemblages of bacteria within the root and root tip samples, setting these regions apart from the remainder of the sponge body. This trend was consistently observed within samples from both the Gulf of Maine and Davis Strait and Scott Inlet, suggesting a stable, species-specific core microbial assemblage in these two clusters. ANOVA showed that geographical location and host individual also influenced the composition of the microbiome (Figure 4.5B), albeit with a smaller effect size. Recent studies have discussed the flexibility and adaptive capabilities of sponge microbiomes (25), and the geographical differences in *C. grandis* microbiomes could indicate adaptations to local conditions and the acquisition, or transient presence, of divergent environmental bacterial taxa. The clustering of *C. oxeata* samples within the ordination analysis (Figure

4.5C) could not be explained by available variables within our dataset and appeared more diffuse, although the lower number of samples for *C. oxeata* could potentially have obscured present trends.

Alpha diversity analysis indicated that the microbial community is not evenly distributed amongst anatomical regions. Specifically, within *C. grandis*, the root and root tip samples displayed the most diverse communities, while axis, sphere, and stem were markedly less diverse. Although a large variation was seen within some anatomical regions of *C. oxeata*, the base axis and embryos presented the highest mean diversity. The increase in bacterial diversity in specific anatomical regions (such as the root sections of *C. grandis*) might point towards a particular functional importance of these regions. In fact, studies have suggested that in *Asbestopluma hypogea*, a carnivorous sponge also displaying a distinctive heterogeneous body plan, a single anatomical region might drive the lifecycle of the host sponge, localizing processes such as cell renewal and differentiation (26). In *A. hypogea*, dynamic cell motility was observed, with cells migrating from the base of the sponge peduncle to prey capture sites during starvation and shortly after prey capture and towards the peduncle during later stages of digestion, supporting the possibility that one specific region might represent the "main body" of the host, working in tandem with other auxiliary, function-oriented regions.

4.5.2 Bacterial community composition

Bacterial communities in both *C. oxeata* and *C. grandis* included a high abundance of Flavobacteriia, Gammaproteobacteria, and Alphaproteobacteria. While the

Alphaproteobacteria and Gammaproteobacteria have been commonly observed in both filter-feeding and carnivorous sponge microbiomes (2), the relatively high abundance of Flavobacteriia is less common and appears restricted to deep-sea and carnivorous sponges (3,4,27). At the family level, striking differences between *C. grandis* and *C. oxeata* were observed in terms of bacterial abundance and richness. Specifically, within *C. oxeata*, the only ubiquitous family was the Flavobacteriaceae, whose members represented the majority of the community, in line with the detected low community diversity and evenness metrics. Some of the observed traits of *C. oxeata*, including a single clade dominance and little overlap between intrasample microbiomes, could point towards this species being a potential "low microbial abundance" sponge (28). In contrast, *C. grandis* showed a larger diversity, with abundance values distributed amongst the most common families including the Flavobacteriaceae, Rhodobacteraceae, Halieaceae, Colwelliaceae, and Hyphomonadaceae.

4.5.3 Enrichment of bacterial taxa within different anatomic regions

Due to their unique morphology, we previously hypothesized that the anatomical regions of the carnivorous sponges might have different functional roles and reported on the association of specific bacterial genera within anatomical regions of *C. grandis* (4). Here, we further investigated this hypothesis using biomarker analysis (LEfSe). While no biomarker taxa were found for *C. oxeata*, the association between specific bacterial taxa and anatomical regions does occur within Arctic *C. grandis*, in agreement with what has been previously observed in the Gulf of Maine (4). The majority of enriched bacteria were

detected within the root and root tips, further supporting the hypothesis that this could represent an area of particular importance within *C. grandis*.

Biomarkers with genus-level taxonomy assignment included the *Mycobacterium*, *Maritimimonas*, *Nitrospira*, *Robiginitomaculum*, *Anderseniella*, and *Granulosicoccus* in the root and root tip. In addition, two biomarkers were found within the trapping spheres: *Colwellia* and *Reichenbachiella*. The ecological role and importance of many of these genera are not fully understood and inferring functional importance can be highly speculative, as bacteria with identical 16S rRNA gene sequences can have highly divergent genomes (29). However, the biomarker status of *Colwellia* and *Reichenbachiella* within trapping spheres is intriguing, as members within both genera have proven to have polysaccharide degradation activities. Previous studies indicated that growth of *Reichenbachiella* is highly stimulated by the presence of chitin in cold-water environments (30), while several species within the *Colwellia* genus are known chitinase producers (31). Mesoplanktonic organisms caught by carnivorous sponges likely have a chitin-rich exoskeleton that inhibits phagocytosis and restricts access to nutrients stored within the prey. The increase in the abundance of both of these chitin-degradation-associated genera is suggestive of an active role in prey-associated chitin hydrolysis, facilitating prey breakdown and subsequent phagocytosis processes, although further functional studies are needed to fully confirm this hypothesis.

4.5.4 Host and species specificity of common and biomarker bacterial genera

Oligotyping of genera common to both sponge species (*Tenacibaculum*, Candidatus *Branchiomonas*, and *Fulvivirga*) showed a host-species-specific pattern of bacterial oligotypes, consistent throughout the entire sponge body. Within *C. grandis*, this specificity remains unchanged even across a wide geographical distance (Arctic waters to the Gulf of Maine). This host species specificity occurs even in sponges originating from the same geographical location (i.e., Bo1 and Bg3, Figure 4.8). This suggests that both *C. grandis* and *C. oxeata* are capable of selectively retaining specific oligotypes of bacterial taxa from the water column. Similar host specificity was previously reported for filter-feeding sponges (32).

Multiple oligotypes were found for the sphere biomarkers *Colwellia* and *Reichenbachiella*. The distribution of sequences indicated the presence of two dominant oligotypes of *Colwellia* and *Reichenbachiella* whereas additional, rare *Colwellia* oligotypes were only seen in host samples from one sponge in one specific geographical location. Although inconclusive, these data seem to suggest that, although multiple *Colwellia* oligotypes can be present in the environment, there are two specific oligotypes that are generally preferentially retained. Unfortunately, the lack of seawater samples prevented us from comparing sponge-associated oligotypes to environmental oligotypes, which would have given us a more complete understanding on this potential selection.

4.5.5 Potential for vertical transmission of bacterial communities

The biology of reproduction within deep-sea carnivorous sponges is still poorly understood. The Cladorhizidae are presumed to be hermaphroditic and viviparous, with an internal developmental stage where large embryos develop into larvae (8,33). Preliminary gross morphology and histological analysis of the oval inclusions found within *C. grandis* and *C. oxeata* indicated several features, such as their localization on the outermost edge of the adult, a yolky interior upon sectioning, a double cell layer exterior including a peripheral "sheath", and the presence of multiple cell types in an apparent collagen-like matrix in the center of the inclusion. All of these features have been described within embryos of both haplosclerid demosponges and the carnivorous sponge *Lycopodina occidentalis* (22–24).

The presence of embryos gave us the unique opportunity to investigate their associated microbiome and speculate about the vertical transmission of symbionts and other bacteria within carnivorous sponges. In *C. grandis*, the microbial diversity within embryos was the lowest amongst all anatomical regions; however, the evenness of the associated communities was the highest. No statistically significant enriched bacteria were detected within the embryos. In fact, the bacterial composition of embryos mostly resembled an "average" of the total sponge microbiome, that is, the combined taxa found preferentially in particular anatomical regions. For example, the Hyphomonadaceae and Nitrospinaceae, most abundant in the root, were observed within embryos at higher levels than in the surrounding tissue (axis). This seems to suggest that a different microbial assemblage is included within the embryos, possibly constituting a vertically transmitted

toolkit needed for the embryo to successfully thrive after its release from the parent. Such vertical transmission has previously been described in detail within filter-feeding sponges (25).

In contrast, the embryo-associated microbial diversity in *C. oxeata* was the highest amongst anatomical regions, with a low community evenness. Furthermore, the composition of two embryos collected from the same host sponge showed a high variability, while this was not the case in *C. grandis*. As such, results from this species are harder to interpret. However, the low community evenness and high variability suggest a different involvement for bacterial associates in embryos of *C. oxeata* as opposed to *C. grandis*. Further studies are needed to fully explore potential vertical transmission in these sponges.

4.5.6 Environmental hydrocarbons as a driver for an alternative source of nutrients

Previous studies have either shown or hypothesized that at least some carnivorous sponges can gain nutrition from chemosynthetic symbionts utilizing environmental hydrocarbon, such as methane (3,7). Indeed, carnivorous sponges are often found in higher abundances in areas of general enrichment such as vent and seep sites (3). While this could in part be due to a general increase in prey availability in these regions, some carnivorous sponge species are thought to settle in these areas to utilize these environmental hydrocarbons, notably through methanotrophy (3).

To examine whether methanotrophy might occur within the carnivorous sponges studied herein, stable isotope analysis was performed. This revealed a lighter $\delta^{13}C$ ratio for

Gulf of Maine (−33.27‰) *C. grandis* specimens in comparison to both species collected within the Arctic regions (−24.78‰ to −24.88‰). However, $\delta^{13}C$ values were notably higher than those reported for the methanotrophic carnivorous sponge *C. methanophila* (3).

Together with $\delta^{15}N$ values suggestive of carnivory, carbon isotope values suggest that methanotrophy is unlikely to occur (at least to a significant degree) within *C. grandis* from our study. Indeed, we noted no particular methanotrophic bacteria in the microbial communities of the sponges investigated here. Nevertheless, the relatively lower $\delta^{13}C$ values for Gulf of Maine *C. grandis* could indicate the incorporation of ^{13}C-depleted carbon via the heterotrophic degradation of other hydrocarbons (e.g., oils), which typically have $\delta^{13}C$ values that are slightly more depleted than carbon in source substrates (34). However, hydrocarbon degradation is unlikely to be the main source of nutrient acquisition in this sponge. Site-specific differences in sediment hydrocarbon availability could partly explain the observed geographic difference in carbon isotope ratios. Furthermore, as sampling efforts were performed during different months (Gulf of Maine: June, Arctic: October), a seasonal factor could be influencing the availability of prey as seasonal vertical migrations of mesoplanktonic prey occurs, with many species departing from deeper waters in the month of May, not to return until July (35). Therefore, one hypothesis could be that in times of high prey abundance, hydrocarbon degradation might not be necessary and energetically unfavorable whereas in low prey abundance situations, nutritional strategies alternative to carnivory may have to be employed. A prolonged sampling effort of these species over a larger period of time might elucidate if such fluctuations indeed occur or if hydrocarbon degradation is only correlated with the presence of hydrocarbon-rich sources.

As in our previous study (4), we observed an association of *Robiginitomaculum* (Hyphomonadaceae) with root tips of *C. grandis*. A potential role for *Robiginitomaculum* in hydrocarbon degradation has been suggested elsewhere (36) and genera within the Hyphomonadaceae are often found in oligotrophic niches, such as deep-sea vents (37). Therefore, *Robiginitomaculum*, or similar uncharacterized bacteria, could have a role in hydrocarbon degradation within *C. grandis*. In contrast, no specific enrichments of suspected hydrocarbon-associated bacteria were detected within *C. oxeata*.

4.5.7 Overall conclusions

In conclusion, our results shed light on the bacterial communities associated with *C. grandis* and *C. oxeata* and revealed a highly conserved species-specific microbiome among sponges sampled in different geographical locations. Large-scale differences in microbiome composition and diversity clearly set the two sponge species apart. Furthermore, we demonstrate the potential ability for active selection of specific bacterial oligotypes in both *C. grandis* and *C. oxeata* and the association of distinct bacterial taxa with specific anatomical regions in *C. grandis*. Lastly, the finding that *C. grandis* might perform symbiotic hydrocarbon degradation could indicate an alternative avenue through which carbon is transferred to carnivorous sponges in deep-sea food webs, including in select Arctic environments.

The exact functional roles of bacteria within carnivorous sponges and the range of nutritional strategies that the host sponge itself might employ require further investigation,

possibly through the use of lipid or fatty acid markers and transcriptomic experiments, ideally from samples collected at different locations and times.

4.6 Acknowledgments

The authors would like to acknowledge the crew of the Canadian Coast Guard ship Amundsen and the SUMO ROV pilots for making sponge collection possible. We would further like to thank Bárbara de Moura Neves and Vonda Wareham for their indispensable help, including sponge identification during ROV dives. Alana Kavanagh helped with spicule preparation and electron microscopy.

4.7 Funding

This work was funded by the ArcticNet Grant “Hidden Biodiversity and Vulnerability of Hard-Bottom and Surrounding Environments in the Canadian Arctic” and NSERC Discovery Grant 2015-06548 to S.C.D.

4.8 References

1. Maldonado M. Sponge waste that fuels marine oligotrophic food webs: a re-assessment of its origin and nature. *Marine Ecology* (2015) **37**:477–491. doi:10.1111/maec.12256

2. Thomas T, Moitinho-Silva L, Lurgi M, Björk JR, Easson C, Astudillo-García C, Olson JB, Erwin PM, López-Legentil S, Luter H, et al. Diversity, structure and convergent evolution of the global sponge microbiome. *Nature Communications* (2016) **7**:11870.

3. Hestetun JT, Dahle H, Jørgensen SL, Olsen BR, Rapp HT. The microbiome and occurrence of methanotrophy in carnivorous sponges. *Front Microbiol* (2016) **7**: doi:10.3389/fmicb.2016.01781

4. Verhoeven JTP, Kavanagh AN, Dufour SC. Microbiome analysis shows enrichment for specific bacteria in separate anatomical regions of the deep-sea carnivorous sponge *Chondrocladia grandis*. *FEMS Microbiol Ecol* (2017) **93**: doi:10.1093/femsec/fiw214

5. Vacelet J, Duport E. Prey capture and digestion in the carnivorous sponge *Asbestopluma hypogea* (Porifera: Demospongiae). *Zoomorphology* (2004) **123**:179–190. doi:10.1007/s00435-004-0100-0

6. Vacelet J, Boury-Esnault N, Fiala-Medioni A, Fisher CR. A methanotrophic carnivorous sponge. *Nature* (1995) **377**:296. doi:10.1038/377296a0

7. Vacelet J, Fiala-Médioni A, Fisher CR, Boury-Esnault N. Symbiosis between methane-oxidizing bacteria and a deep-sea carnivorous cladorhizid sponge. *Marine Ecology Progress Series* (1996) **145**:77–85. doi:10.3354/meps145077

8. Vacelet J. Diversity and evolution of deep-sea carnivorous sponges. *Porifera research: biodiversity, innovation and sustainability Série Livros* (2007) **28**:107–115.

9. Hestetun JT, Fourt M, Vacelet J, Boury-Esnault N, Rapp HT. Cladorhizidae (Porifera, Demospongiae, Poecilosclerida) of the deep Atlantic collected during Ifremer cruises, with a biogeographic overview of the Atlantic species. *Journal of the Marine Biological Association of the United Kingdom* (2015) **95**:1311–1342. doi:10.1017/S0025315413001100

10. Hestetun JT, Vacelet J, Boury-Esnault N, Borchiellini C, Kelly M, Ríos P, Cristobo J, Rapp HT. The systematics of carnivorous sponges. *Molecular Phylogenetics and Evolution* (2016) **94**:327–345. doi:10.1016/j.ympev.2015.08.022

11. Buhl-Mortensen L, Buhl-Mortensen P, Dolan MFJ, Holte B. The MAREANO programme – A full coverage mapping of the Norwegian off-shore benthic environment and fauna. *Marine Biology Research* (2015) **11**:4–17. doi:10.1080/17451000.2014.952312

12. Comeau AM, Li WKW, Tremblay J-É, Carmack EC, Lovejoy C. Arctic Ocean microbial community structure before and after the 2007 record sea ice minimum. *PLoS ONE* (2011) **6**:e27492. doi:10.1371/journal.pone.0027492

13. Comeau AM, Douglas GM, Langille MGI. Microbiome Helper: a custom and streamlined workflow for microbiome research. *mSystems* (2017) **2**:e00127-16. doi:10.1128/mSystems.00127-16

14. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* (2014) **30**:2114–2120. doi:10.1093/bioinformatics/btu170

15. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* (2014) **30**:614–620. doi:10.1093/bioinformatics/btt593

16. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* (2011) **17**:10–12. doi:10.14806/ej.17.1.200

17. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* (2016) **4**: doi:10.7717/peerj.2584

18. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* (2013) **8**:e61217. doi:10.1371/journal.pone.0061217

19. Oksanen J, Guillaume Blanchet F, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin P, O'Hara RB, Simpson GL, Solymos P, et al. vegan: community ecology package. R package version 2.4-6. (2018) Available at: https://CRAN.R-project.org/package=vegan

20. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C. Metagenomic biomarker discovery and explanation. *Genome Biology* (2011) **12**:R60. doi:10.1186/gb-2011-12-6-r60

21. Eren AM, Maignien L, Sul WJ, Murphy LG, Grim SL, Morrison HG, Sogin ML. Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods in Ecology and Evolution* (2013) **4**:1111–1119. doi:10.1111/2041-210X.12114

22. Leys SP, Degnan BM. Embryogenesis and metamorphosis in a haplosclerid demosponge: gastrulation and transdifferentiation of larval ciliated cells to choanocytes. *Invertebrate Biology* (2005) **121**:171–189. doi:10.1111/j.1744-7410.2002.tb00058.x

23. Riesgo A, Taylor C, Leys SP. Reproduction in a carnivorous sponge: the significance of the absence of an aquiferous system to the sponge body plan. *Evolution & Development* (2007) **9**:618–631. doi:10.1111/j.1525-142X.2007.00200.x

24. Vacelet J, Boury-Esnault N. A new species of carnivorous deep-sea sponge (Demospongiae: Cladorhizidae) associated with methanotrophic bacteria. *Cahiers de Biologie Marine* (2002)

25. Webster NS, Thomas T. The sponge hologenome. *mBio* (2016) **7**:e00135-16. doi:10.1128/mBio.00135-16

26. Martinand-Mari C, Vacelet J, Nickel M, Wörheide G, Mangeat P, Baghdiguian S. Cell death and renewal during prey capture and digestion in the carnivorous sponge *Asbestopluma hypogea* (Porifera: Poecilosclerida). *J Exp Biol* (2012) **215**:3937–3943. doi:10.1242/jeb.072371

27. Dupont S, Corre E, Li Y, Vacelet J, Bourguet-Kondracki M-L. First insights into the microbiome of a carnivorous sponge. *FEMS Microbiol Ecol* (2013) **86**:520–531. doi:10.1111/1574-6941.12178

28. Gloeckner V, Wehrl M, Moitinho-Silva L, Gernert C, Schupp P, Pawlik JR, Lindquist NL, Erpenbeck D, Wörheide G, Hentschel U. The HMA-LMA dichotomy revisited: an electron microscopical survey of 56 sponge species. *The Biological Bulletin* (2014) **227**:78–88. doi:10.1086/BBLv227n1p78

29. Jaspers E, Overmann J. Ecological significance of microdiversity: identical 16S rRNA gene sequences can be found in bacteria with highly divergent genomes and ecophysiologies. *Appl Environ Microbiol* (2004) **70**:4831–4839. doi:10.1128/AEM.70.8.4831-4839.2004

30. Wietz M, Wemheuer B, Simon H, Giebel H-A, Seibt MA, Daniel R, Brinkhoff T, Simon M. Bacterial community dynamics during polysaccharide degradation at contrasting sites in the Southern and Atlantic Oceans. *Environ Microbiol* (2015) **17**:3822–3831. doi:10.1111/1462-2920.12842

31. Bowman J, Gosink JJ, McCammon SA, Lewis TE, Nichols DS, Nichols PD, Skerratt JH, Staley JT, McMeekin TA. *Colwellia demingiae* sp. nov., *Colwellia hornerae* sp. nov., *Colwellia rossensis* sp. nov. and *Colwellia psychrotropica* sp. nov.: psychrophilic Antarctic species with the ability to synthesize docosahexaenoic acid (22:ω63). *International Journal of Systematic and Evolutionary Microbiology* (1998) **48**:1171–1180. doi:10.1099/00207713-48-4-1171

32. Reveillaud J, Maignien L, Eren AM, Huber JA, Apprill A, Sogin ML, Vanreusel A. Host-specificity among abundant and rare taxa in the sponge microbiome. *The ISME Journal* (2014) **8**:1198–1209. doi:10.1038/ismej.2013.227

33. Maldonado M, Riesgo A. Reproduction in Porifera: a synoptic overview. *Societat Catalana de Biologia* (2008)

34. Hoefs J. *Stable Isotope Geochemistry*. 7th ed. Springer International Publishing (2015). Available at: //www.springer.com/gp/book/9783319307527 [Accessed December 4, 2018]

35. Darnis G, Fortier L. Temperature, food and the seasonal vertical migration of key arctic copepods in the thermally stratified Amundsen Gulf (Beaufort Sea, Arctic Ocean). *J Plankton Res* (2014) **36**:1092–1108. doi:10.1093/plankt/fbu035

36. Guibert LM, Loviso CL, Marcos MS, Commendatore MG, Dionisi HM, Lozada M. Alkane biodegradation genes from chronically polluted subantarctic coastal sediments and their shifts in response to oil exposure. *Microb Ecol* (2012) **64**:605–616. doi:10.1007/s00248-012-0051-9

37. Abraham W-R, Rohde M. "The Family Hyphomonadaceae," in *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria*, eds. E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, F. Thompson (Berlin, Heidelberg: Springer Berlin Heidelberg), 283–299. doi:10.1007/978-3-642-30197-1_260

Table 4.1: Sponge specimens considered in this study.

Specimen ID	Species	Location Date	Coordinates	Depth (m)
Bg1	*Chondrocladia grandis*	Scott Inlet 17-10-2015	71° 30.9911′ N 70° 01.0464′ W	539
Bg2	*Chondrocladia grandis*	Scott Inlet 17-10-2015	71° 30.9627′ N 70° 16.6843′ W	537
Bg3	*Chondrocladia grandis*	Davis Strait 20-10-2015	67° 28.6157′ N 63° 41.0751′ W	702
Bo1	*Cladorhiza oxeata*	Davis Strait 20-10-2015	67° 28.6491′ N 63° 40.8085′ W	680
Bo2	*Cladorhiza oxeata*	Scott Inlet 17-10-2015	71° 30.9889′ N 70° 16.6472′ W	538
Mg1	*Chondrocladia grandis*[1]	Gulf of Maine 27-06-2014	41° 48.2302′ N 65° 41.0412′ W	852
Mg2	*Chondrocladia grandis*[1]	Gulf of Maine 27-06-2014	41° 48.2302′ N 65° 41.0412′ W	852
Mg3	*Chondrocladia grandis*[1]	Gulf of Maine 27-06-2014	41° 48.2302′ N 65° 41.0412′ W	852
Mg4	*Chondrocladia grandis*[1]	Gulf of Maine 27-06-2014	41° 48.2302′ N 65° 41.0412′ W	852

1. Samples from previous study (4)

Table 4.2: Run statistics for individual samples.

Sponge	*Sample*	*Region*	*Raw Reads*	*PASSED QC*	*Unique OTUs*
Bg1					
	AS1	Sphere	58597	51198	6652
	AS2	Sphere	48517	43384	7147
	AS3	Stem	46156	41446	5057
	AS4	Stem	57498	51390	6085
	AS5	Root	68983	61646	12786
	AS6	Root	70759	62505	14643
	AS7	Root tip	61231	53997	9042
	AS8	Root tip	41978	37534	6650
	AS9	Axis	41747	37434	6902
	AS10	Axis	9566	8484	1460
	AS20	Embryo	52258	46668	7384
	AS21	Embryo	49838	44717	6138
Bo2					
	AS31	Sphere	46015	40788	4364
	AS32	Stem	59288	52678	4796
	AS33	Stem	73390	64500	5126
	AS34	Secondary axis	36674	32484	2920
	AS35	Secondary axis	51700	45890	4142
	AS36	Basal axis	48174	42917	7172
	AS37	Basal Axis	74212	66024	7749
	AS43	Embryo	61204	54712	4897
	AS44	Embryo	61095	51776	6401
Bg2					
	AS11	Sphere	29809	26795	4178
	AS12	Sphere	46498	41567	6926
	AS13	Stem	48801	43535	5110
	AS14	Stem	51869	45773	6017
	AS15	Axis	66397	59798	8087
	AS16	Axis	30541	27313	4057
	AS23	Embryo	74552	66075	8166
Bg3					
	AS17	Sphere	51431	45916	7102
	AS18	Sphere	44954	40164	5468
	AS19	Stem	56017	50639	5370
	AS25	Root	56134	49805	15426
	AS26	Root	38963	34638	8857
	AS27	Root tip	62431	55678	14011
	AS29	Axis	47995	43164	6955
	AS30	Axis	44342	40070	5626
Bo1					

	AS38	Sphere	45696	40241	2957
	AS39	Stem	38356	34236	3458
	AS40	Secondary axis	49342	44002	5455
	AS41	Secondary axis	74108	65042	6199
	AS42	Basal axis	33937	29928	3577
	AS45	Basal axis	20138	16354	1825
Mg1					
	1B	Sphere	61227	41006	13063
	1S	Stem	78867	55512	12938
Mg2					
	5B	Sphere	97270	66531	18532
	5S	Stem	72561	51555	1970
	7R	Root	86405	56750	16418
Mg3					
	9B	Sphere	95459	65799	16619
	9S	Stem	118153	81363	15950
	11R	Root	47238	30843	9416
Mg4					
	13B	Sphere	64728	45043	9290
	13S	Stem	88909	60143	12782
	15RM	Root	118791	79195	20457
Total			**3060799**	**2526645**	**428775**

Table 4.3: Alpha-diversity measurements for distinct anatomical regions within the carnivorous sponges *Chondrocladia grandis* and *Cladorhiza oxeata*.

Sample	**Observed**		**Chao1**		**invSimpson (1/D)**	
	Min	**Max**	**Min**	**Max**	**Min**	**Max**
Chondrocladia grandis						
Root	427	1611	433	1771	3.6	17.1
Root tip	388	1032	526	1158	4.3	12.1
Axis	41	172	83	223	2.9	5.2
Stem	86	246	109	341	3.1	5.0
Sphere	87	255	107	358	3.6	8.5
Embryo	81	107	105	210	4.6	6.6
Cladorhiza oxeata						
Base axis	53	297	55	345	2.3	2.9
Axis	62	160	83	185	1.7	4.5
Sphere	64	89	77	109	1.3	1.6
Stem	60	67	77	115	1.5	1.6
Embryo	113	242	150	275	1.6	1.8

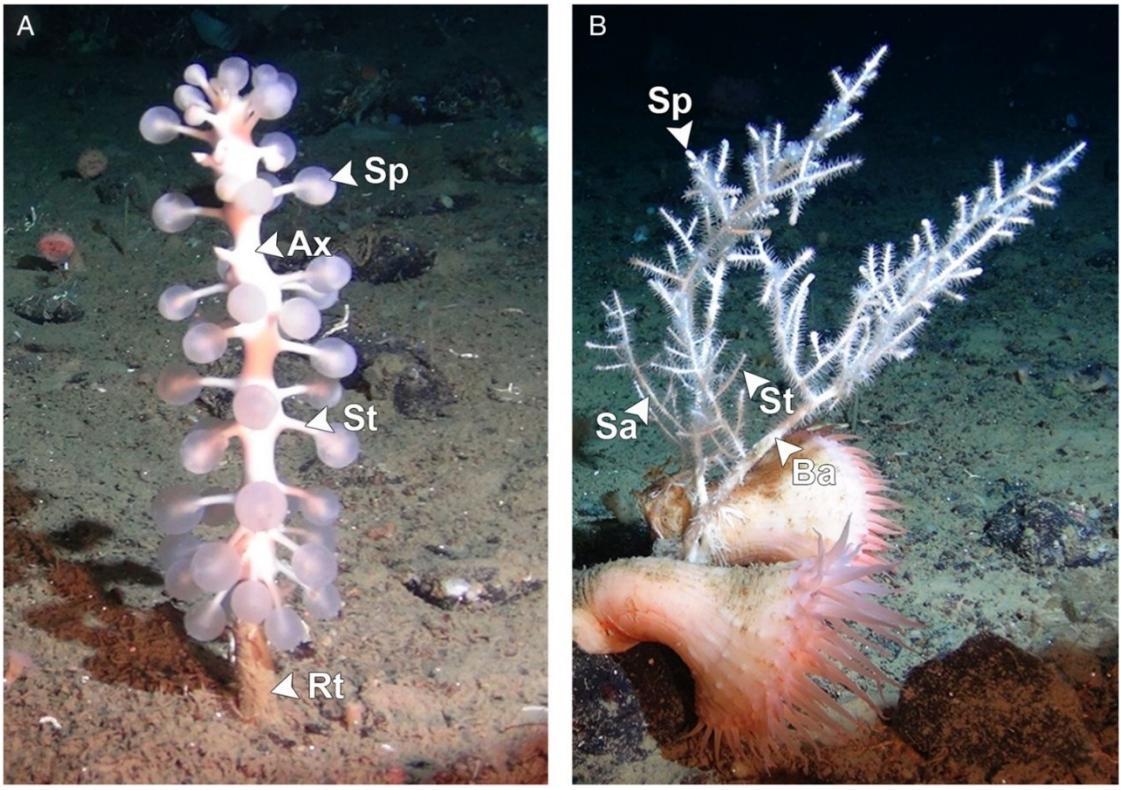

Figure 4.1: *Chondrocladia grandis* and *Cladorhiza oxeata* as observed in situ. (A) *Chondrocladia grandis*: arrowheads indicate different anatomical regions. Ax, axis; St, stem; Rt, root; Sp, sphere. (B) *Cladorhiza oxeata*: arrowheads indicate different anatomical regions. Ba, base axis; Sa, secondary axis; St, stem; Sp, sphere.

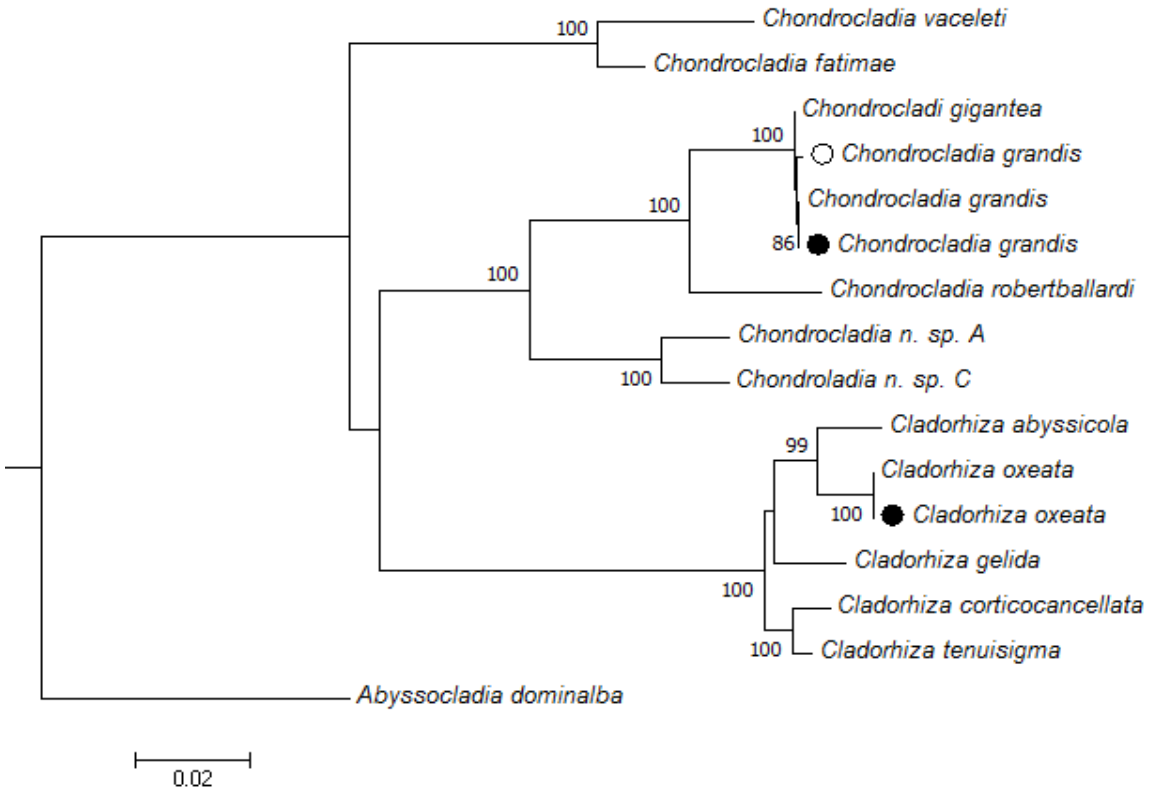

Figure 4.2: Maximum Likelihood analysis of concatenated 28S rDNA, COI and ALG11 alignments of *Chondrocladia grandis and Cladorhiza oxeata* with reference sequences. The evolutionary history was inferred with the GTR model and a discrete gamma distribution was used to model evolutionary rate differences among sites. The outcome of the bootstrap analysis is shown next to the nodes and only values higher than 65 are shown. Branch lengths are proportional to genetic distances as indicated by the scale bar. *Abyssocladia dominalba* was used as outgroup. Arctic *Chondrocladia grandis* and *Cladorhiza oxeata* samples collected in this study are indicated by a filled circle, *Chondrocladia* grandis collected in the Gulf of Maine is indicated by an empty circle (KX950006-KX950009). Accession numbers of sequences used to build the tree included *C. fatimae*: LN870616, LN870480, LN870557; *C. gigantea*: LN870619, LN870483, LN870559; *C. robertballardi*: LN870627, LN870493, LN870567; *C. vaceleti*: LN870628, LN870494, LN870568; *C. n. sp. A*: LN870622, LN870486, LN870562; *C. n. sp. C*: LN870626, LN870490, LN870566; *C. grandis*: KX266196, KX266210, KX266226; *C. oxeata*: KX266192,KX266209, KX266225; *C. abyssicola*: LN870631, LN870499, LN870550; *C. corticocancellata*: LN870632, LN870500, LN870551; *C. gelida*: LN870633, LN870501, LN870552; *C. tenuisigma*: LN870636, LN870504, LN870555; *A. dominalba*: LN870578, LN870440, LN870544.

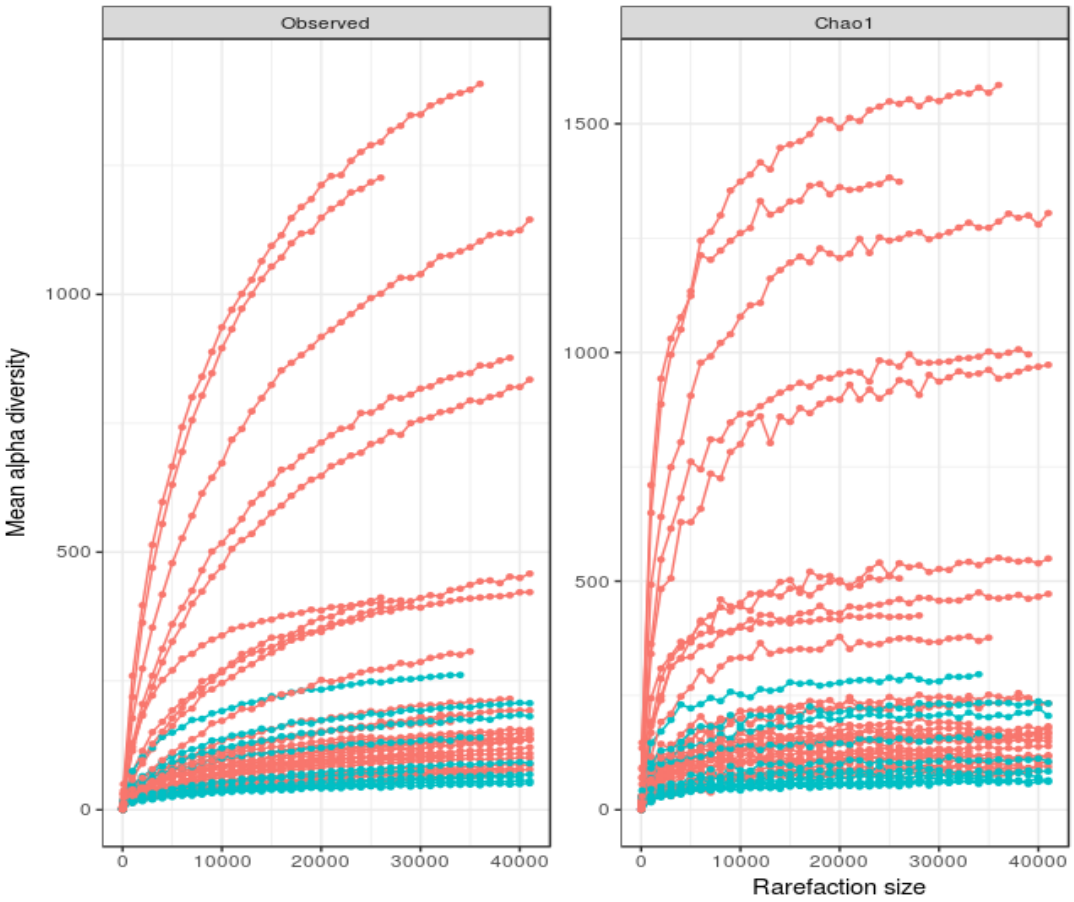

Figure 4.3: Rarefaction curve analysis of high-throughput sequencing data for Arctic carnivorous sponges. Data points indicate mean alpha diversity at specified rarefaction sizes for "observed" (number of operational taxonomic units observed) and Chao1 metrics, left and right panel, respectively. Data series displayed represent, grouped subsamples of distinct anatomical regions within separate sponge individuals. Red and turquoise lines indicate *Chondrocladia grandis* and *Cladorhiza oxeata* specimens, respectively.

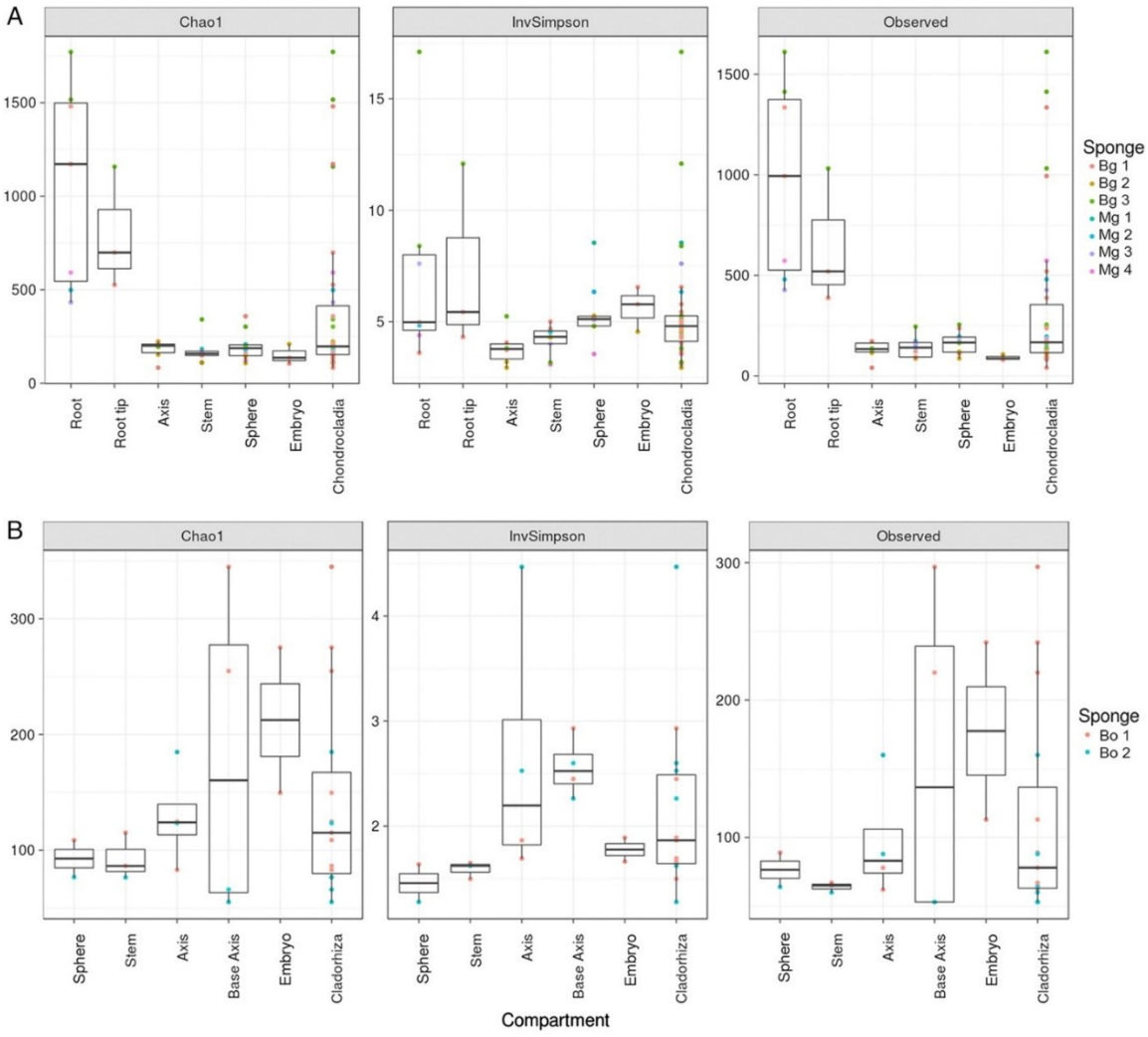

Figure 4.4: Estimated bacterial community richness and evenness within distinct anatomical regions. (A) *Chondrocladia grandis*; (B) *Cladorhiza oxeata*. Anatomical regions for each sponge specimen and compiled data for each species are displayed on the x-axis, while boxes indicate metrics: Chao1 estimated community richness (Chao1), the reciprocal Simpson index (invSimpson), and number of OTUs observed (Observed). Boxplot rectangles indicate the interquartile range with an internal line denoting the median; whiskers extend to minimum and maximum values. Whiskers on individual Chao1 values indicate standard deviation. Sponge individuals are represented by different numbers and colors and are coded by broader geographic region (B, Baffin Bay; M, Gulf of Maine) and species (g, *C. grandis*; o, *C. oxeata*).

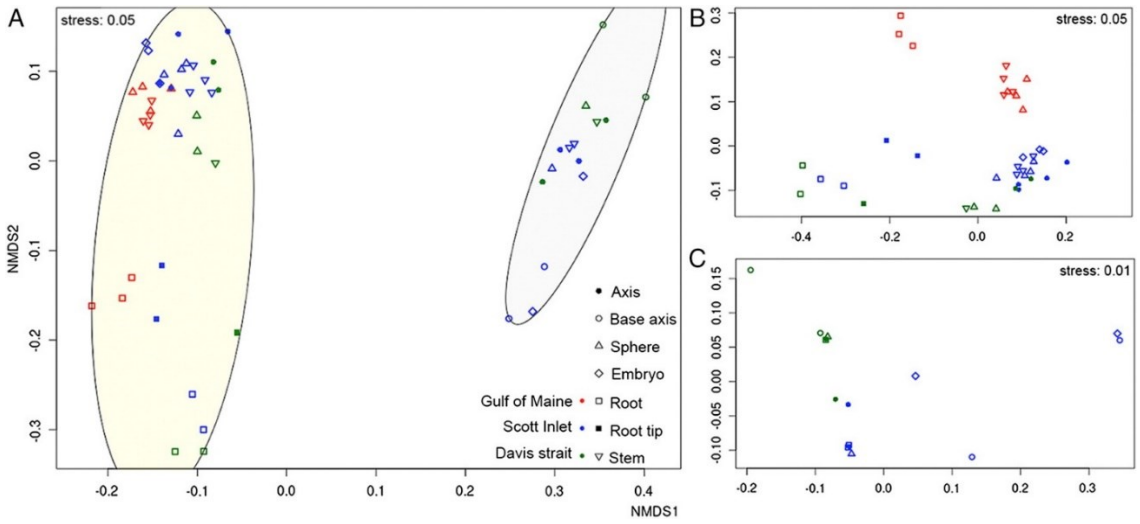

Figure 4.5: Bacterial community similarity. Ordination analysis (nonmetric multidimensional scaling (NMDS) based on Jensen–Shannon divergence of bacterial communities in sponge samples from different geographical locations and host species. Displayed are (A) combined *Chondrocladia grandis* (left cluster) and *Cladorhiza oxeata* (right cluster) samples from Davis Strait (green), Gulf of Maine (red), and Scott Inlet (blue), (B) *C. grandis* samples, and (C) *C. oxeata* samples. Symbols indicate anatomical regions. Each plot has the NMDS stress value reported in the top right or left corner.

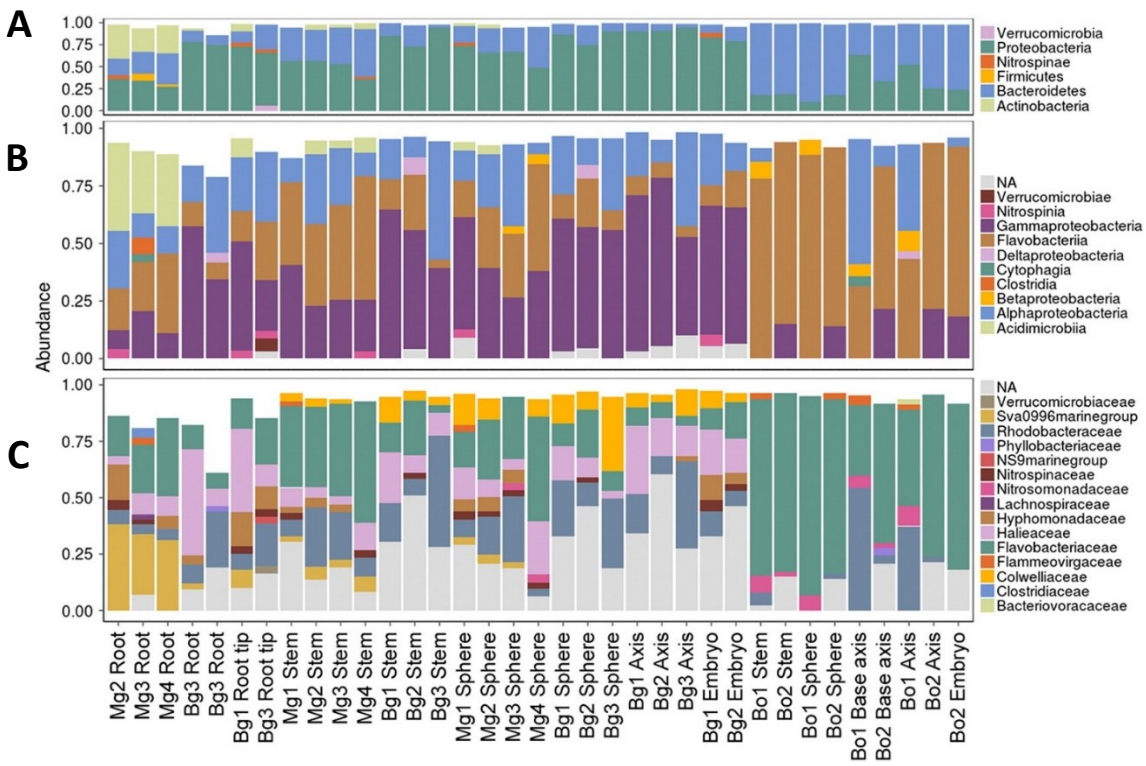

Figure 4.6: Bar graph of most common bacterial taxa found in anatomical regions of *Chondrocladia grandis* and *Cladorhiza oxeata*. The x-axis represents anatomical regions (grouped per sponge individual), while color-coded segments in each bar indicate taxa with a relative abundance higher than 2%, using count data agglomerated at the Phylum (top panel, A), Class (middle panel, B), and Family (bottom panel, C) levels. Sponge individuals are represented by different numbers and are coded by broader geographic region (B, Baffin Bay; M, Gulf of Maine) and species (g, *C. grandis*; o, *C. oxeata*).

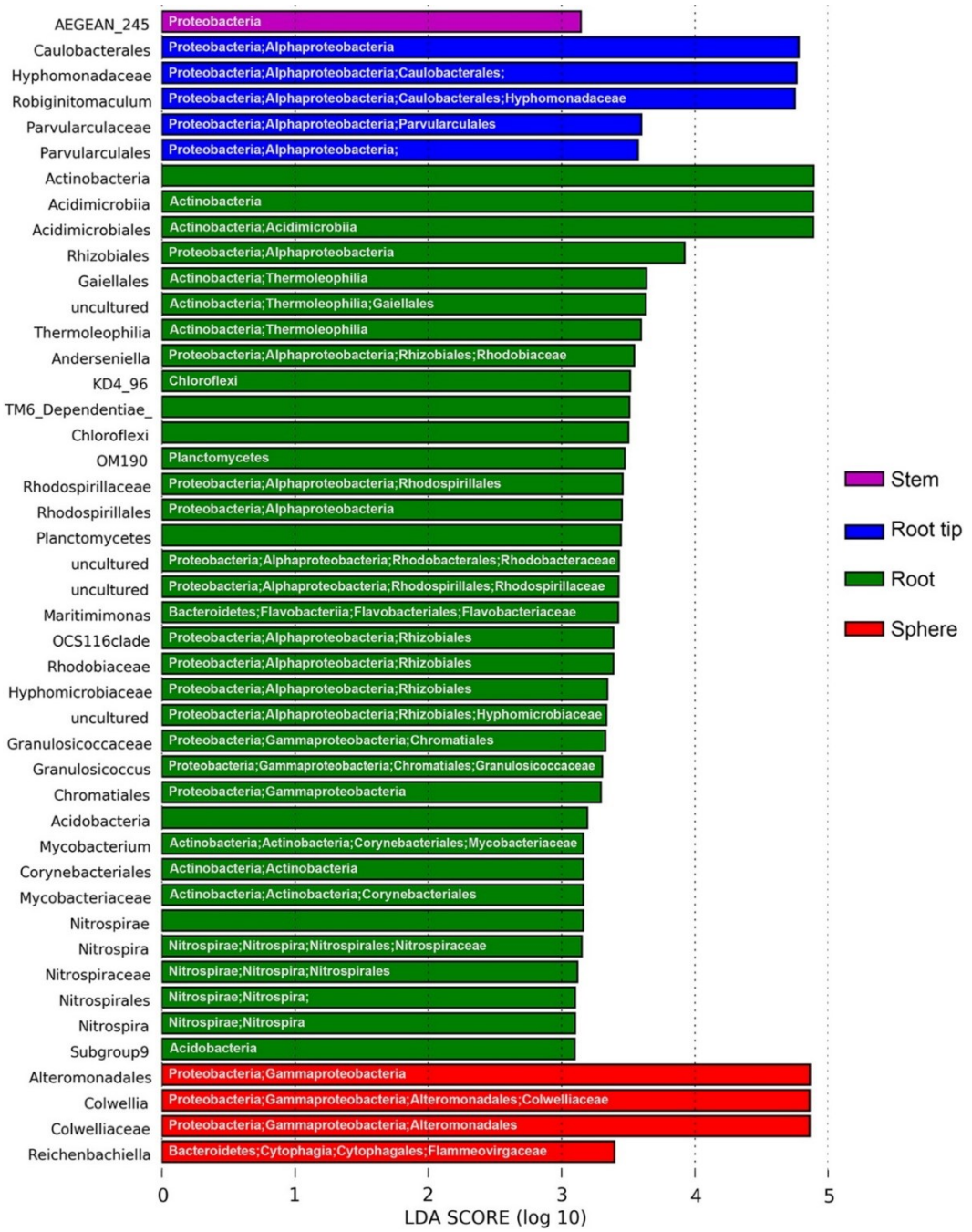

Figure 4.7: Differentially abundant bacterial genera (biomarkers) in *Chondrocladia grandis*. Each bar indicates the linear discriminant analysis (LDA) score for a specific biomarker that is significantly ($p < 0.01$) more abundant within the respective anatomical region (indicated by bar color). The most specific taxonomy is given on the y-axis, while additional taxonomic information is displayed within each bar (if available).

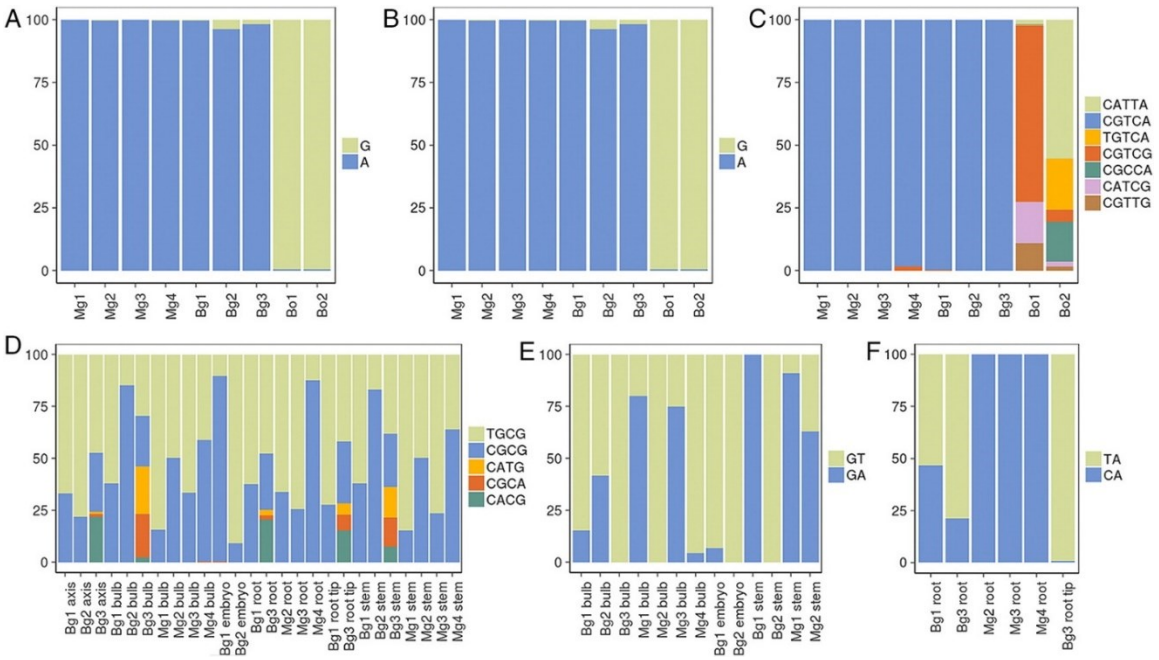

Figure 4.8: Oligotype results of both common and biomarker genera in *Chondrocladia grandis* and *Cladorhiza oxeata*. Bars represent sponge individuals or anatomical regions per sponge, while stacked bars show the relative abundance of each oligotype within the common genera (A) *Tenacibaculum*, (B) *Candidatus* Branchiomonas, (C) *Fulvivirga*, as well as the *Chondrocladia grandis* specific biomarkers, (D) *Colwellia*, (E) *Reichenbachiella*, and (F) *Maritimimonas*. Comparison of common genera oligotypes is presented grouped on sponge individual, while biomarker oligotypes are displayed per anatomical region. Samples with fewer than 10 reads for a particular genus were not displayed in the graph. Sponge individuals are represented by different numbers and are coded by broader geographic region (B, Baffin Bay; M, Gulf of Maine) and species (g, *C. grandis*; o, *C. oxeata*).

CHAPTER 5: VIDIT-CACTUS: AN INEXPENSIVE AND VERSATILE LIBRARY PREPARATION AND SEQUENCE ANALYSIS METHOD FOR VIRUS DISCOVERY AND OTHER MICROBIOLOGY APPLICATIONS

Adapted from the published version available in The Canadian Journal of Microbiology[6]
https://doi.org/10.1139/cjm-2018-0097

5.1 Abstract

High-throughput sequencing (HTS) technologies are becoming increasingly important within microbiology research, but aspects of library preparation, such as high cost per sample or strict input requirements, make HTS difficult to implement in some niche applications and for research groups on a budget. To answer these necessities, we developed ViDiT, a customizable, PCR-based, extremely low-cost (<5 US dollars per sample) and versatile library preparation method, and CACTUS, an analysis pipeline designed to rely on cloud computing power to generate high-quality data from ViDiT-based experiments without the need of expensive servers. We demonstrate here the versatility and utility of these methods within three fields of microbiology: virus discovery, amplicon-based viral genome sequencing and microbiome profiling. ViDiT-CACTUS allowed the identification of viral fragments from 25 different viral families from 36 oropharyngeal-cloacal swabs collected from wild birds, the sequencing of three almost complete genomes of avian influenza A viruses (>90% coverage), and the characterization and functional profiling of the complete microbial diversity (bacteria, archaea, viruses) within a deep-sea carnivorous sponge. ViDiT-CACTUS demonstrated its validity in a wide

[6] This version was modified to include supplementary figures and tables in the main text

range of microbiology applications and its simplicity and modularity make it easily implementable in any molecular biology laboratory, towards various research goals.

5.2 Introduction

Over the last decade, high-throughput sequencing (HTS) technologies have seen substantial improvements to both chemistry and hardware (1). Combined with low per-base costs, HTS has become routinely used in many research laboratories. This has enabled advances in diverse areas of microbiology including pathogen discovery, disease surveillance, full-genome analyses and large-scale ecological investigations, such as charting microbial diversity and interactions within ecological niches (2).

HTS methods require alteration of original sample-derived nucleic acids (NA) to create standardized libraries compatible with the specifications of the sequencing instrument being used (i.e., the sequencing libraries). While a large array of commercial library preparation methods exists (3), the varying input requirements, including adequate NA quality and concentration, can make selecting an appropriate library preparation method problematic. This complexity is exacerbated when working with samples that do not fit the requirements of commercially available kits. For example, the low amount of NA that can be recovered from specific sample types, such as clinical samples, may lead to failed or low-quality libraries. Furthermore, the associated cost of library preparation can be challenging for research groups on a budget (4). The procurement of specialized equipment can also require a substantial initial investment and recurring per-sample costs with any given library preparation kit can influence the quantity of samples that can be

processed. To overcome such obstacles, some research groups have created custom library preparation methods that are, typically, aimed at one specific application (5–8).

This left a niche for an easily applicable library preparation method that overcomes both budget and sample prerequisite issues, can be used in multiple applications, and can be readily implemented into existing laboratory workflows while still providing satisfactory performance. To fill this niche, we developed the ViDiT (Virus Discovery Ion Torrent) method, a simple three-step random-PCR-based library preparation technique that is a fast, inexpensive and versatile in-house system. Although, as the name implies, it was originally conceptualized for the purpose of virus discovery with one sequencing platform, it is equally suitable to study different aspects of microbial diversity in a broad range of samples using any HTS platform.

The immense volume of HTS-derived data creates further challenges. Specifically, configuring bioinformatic applications to handle subsequent steps, such as sequence quality control, contig assembly and alignment of data to reference databases, can be a daunting task for some laboratories (9). These applications typically require powerful servers that might not be available in all laboratories. As such, complementary to ViDiT, we designed the in-house bioinformatic pipeline CACTUS (not an acronym), which was used to generate the results in this study. Once properly configured, a single workflow handles the majority of operations needed for quality control, assembly and alignments. CACTUS is easily implementable on local systems or cloud computing environments, allowing for virtually any laboratory to perform data analysis at a low cost.

To evaluate the validity of our methods within various fields of microbiology, we assessed three possible applications of the ViDiT-CACTUS system: amplicon-based viral genome sequencing, virus discovery, and microbiome community profiling.

5.3 Material and Methods

Depending on the application, each set of samples was subjected to different pre-processing steps and procedures for NA isolation and purification. Obtained NA (amplicons, genomic DNA or cDNA) were then used as input for the ViDiT method and the reads obtained were fed into the CACTUS pipeline. Figure 5.1 summarizes the procedural steps utilized in this study. We provide detailed protocols as supplementary material (Submitted with this thesis as supplementary file 1) and the CACTUS pipeline is available on GitHub (https://github.com/jtpverhoeven/cactus).

5.3.1 Materials

Samples from birds were collected in 2014 and 2015 at various locations on the Avalon Peninsula (Newfoundland, Canada) during avian influenza A virus (AIV) surveillance studies. This sampling was carried out under the guidelines specified by the Canadian Council on Animal Care with approved protocols 13-01-AL and 14-01-AL from the Memorial University Institutional Animal Care Committee. Paired oropharyngeal and cloacal swabs were collected from each bird (polyester swabs, Starplex Scientific Inc.), submerged together into 3 mL viral transport medium (Starswab Multitrans System, Starplex Scientific Inc.), and frozen until processing. Two swab samples (14-AIV_1475 and 15-AIV_1590) collected from American Herring Gulls (*Larus smithsonianus*) and one

sample (14-AIV_1623) from an American Black Duck (*Anas rubripes*) that tested-positive for AIV (by real-time reverse transcriptase (RT)-PCR (10) with Ct values ranging from 23 to 40) were selected for AIV complete genome sequencing. Thirty-six AIV-negative swab samples collected from 6 different wild bird species (*Anas rubripes, Anas platyrhynchos, Fratercula arctica, Uria aalge*, *Larus marinus* and *Larus smithsonianus*) were used as input for virus discovery.

Sponge tissues used in this study consisted of fragments of a *Chondrocladia grandis* specimen collected in the Canadian Arctic that were preserved at -80˚C until processing (11). Four tissue replicates from different anatomical regions from the sponge specimen were used in this study. Each tissue section was rinsed in PBS three times prior to NA isolation to remove loosely associated microorganisms and rinse off seawater.

5.3.2 Influenza A virus complete genome amplification and product purification

Two hundred and twenty µL of each swab sample were centrifuged at 10000 x g for 10 min to remove material that could cause column clogging and RNA was isolated from 200 µL of the supernatant using the ZR Viral RNA Kit™ (Zymo Research) according to the manufacturer's specifications. An RT-PCR (modified from (12)) was then performed that could simultaneously amplify all 8 viral segments using the SuperScript® III One-Step RT-PCR System with Platinum® Taq DNA Polymerase (ThermoFisher Scientific) in a reaction mix of 25 µL final volume. Tailed (tail sequences underlined) universal primers MBT-Uni13 (ACGCGTGATCAGTAGAAACAAGG), MBT-Uni12 (ACGCGTGATCAGCAAAAGCAGG) and MBT-Uni12G

(ACGCGTGATCAGCGAAAGCAGG) were used at a final concentration of 0.2 µM, 0.12 µM and 0.08 µM, respectively. The reaction was performed using the following incubation conditions: 5 min at 40˚C, 60 min at 45˚C, 2 min at 95˚C, followed by 5 cycles of 30 sec at 95˚C, 30 sec at 45˚C, and 3 min at 68˚C and 35 cycles of 30 sec at 95˚C, 30 sec at 57˚C, and 3 min at 68˚C and a final 7 min at 68˚C. Amplified products were purified with Agencourt AMPure XP beads (Beckman Coulter) with a ratio of 0.6:1 (bead solution:sample, v:v). Purified products were diluted 1:10 and used as input for the ViDiT library preparation.

5.3.3 Pre-treatment and nucleic acid isolation for microbial community profiling

NA were isolated from sponge tissue samples with the MO BIO PowerViral environmental RNA/DNA isolation kit (MO BIO Laboratories) according to the manufacturer's specifications. For microbiome determination, 250 ng of isolated NA were first processed using the NEBNext Microbiome DNA Enrichment Kit (New England Biolabs) to remove any methylated DNA. The enriched DNA was used as input for ViDiT library preparation.

5.3.4 Pre-treatment and nucleic acid isolation for virus discovery

A 210 µL volume of each avian swab sample was centrifuged at 10000 x g for 10 min to remove debris and intact cells and 190 µL of the supernatant were transferred to a new tube and treated with DNase I (New England Biolabs) using 10 µL enzyme (2 U/µL) and 22 µL 10X reaction buffer at 37˚C for 30 min, followed by enzyme inactivation by the addition of 2.2 µL of 0.5 M EDTA (pH 8.0) and incubation at 75˚C for 10 min. NA were

then isolated using the QIAamp® DNA Mini Kit (Qiagen) according to the manufacturer's specifications (bodily fluid spin protocol) and NA were eluted in 40 µL elution buffer. Purified NA were used as input for the RT reaction prior to library preparation.

5.3.5 Retro-transcription of RNA into cDNA

Ten µL of purified NA solution were used to perform an RT reaction using the ProtoScript® II First Strand cDNA Synthesis Kit (New England Biolabs). NA were mixed with 2 µL random primer mix (mixture of hexamers and oligo-dT primer (dT23)) and incubated at 65˚C for 5 min and then immediately chilled on ice. A solution containing 13 µL 2X reaction mix and 2 µL enzyme mix (Protoscript II RT 25 U/µL, RNase H inhibitor 10 U/µL) was added and the solution was incubated at 25˚C for 5 min, 42˚C for 60 min and 80˚C for 5 min. Finally, residual primers were removed by adding 2 µL of ExoSAP-IT reagent (Affymetrix) and incubating the mixture for 15 min at 37˚C, followed by enzyme inactivation by incubation at 80˚C for 15 min.

5.3.6 Library preparation performed with the ViDiT method

ViDiT is an in-house library preparation method originally optimized for Ion Torrent PGM sequencing that includes 3 fundamental steps (Figure 5.2): i) random amplification, where random octamers with a tail (containing either the sequence for one Ion Xpress™ Barcode and part of the A primer or part of the P primer) are used to randomly amplify DNA; ii) sequencing library generation, where primers containing the whole Ion Torrent PGM adaptor sequences (either one Ion Xpress™ Barcode and the whole A primer or the whole P primer) are used to produce amplicons ready for sequencing; and iii) library

enrichment, where Ion Torrent PGM A and P primers are used to increase the library concentration.

Specifically, 10 µL of purified amplicons (for AIV complete genome sequencing), 21 µL of DNA/cDNA (for virus discovery), or 21 µL microbiome-enriched DNA (for microbial community profiling) were used as input in a 50 µL reaction mix containing 1X Dream-Taq Green PCR Master Mix (ThermoFisher Scientific) and 0.4 µM primers ViDiT_AXXN (GXXXXXXXXXXXXGATNNNNNNNN), where Xs represent the barcode sequence, and ViDiT_PN (TATGGGCAGTCGGTGATNNNNNNNN). The random amplification was performed according to the following cycling conditions: 95˚C for 5 min, followed by 15-25 cycles of 30 sec at 95˚C, 2 min at 26˚C and 20 sec at 72˚C, and a final extension step of 7 min at 72˚C. This was followed by a PCR clean-up with Agencourt AMPure XP beads with a ratio of 0.85:1 and elution in 15 µL 0.1X Tris-EDTA (TE) buffer (pH 8.0). The sequencing library generation was performed using 10 µL of product from the previous step in a 25 µL reaction mix containing 1X Dream-Taq Green PCR Master Mix and 0.2 µM primers ViDIT_AXX (CCATCTCATCCCTGCGTGTCTCCGACTCAGXXXXXXXXXXXXGAT) and ViDiT_P (CCTCTCTATGGGCAGTCGGTGAT). The reaction was performed using the following incubation conditions: 5 min at 95˚C, followed by 3 cycles of 30 sec at 95˚C, 30 sec at 48˚C, and 30 sec at 72˚C and 17-32 cycles of 30 sec at 95˚C, 30 sec at 62˚C, and 30 sec at 72˚C and a final 7 min at 72˚C. Amplified products were subjected to a second PCR clean-up as described above. Finally, the library enrichment was performed using 4-6 µL of product from the previous step in a 25 µL reaction mix containing 1X Dream-Taq Green

PCR Master Mix and 0.2 μM primers A (CCATCTCATCCCTGCGTGTCTCCGACT) and ViDiT_P. The reaction was performed using the following incubation conditions: 5 min at 95˚C, followed by 10-30 cycles of 30 sec at 95˚C, 30 sec at 58˚C, and 30 sec at 72˚C and a final 7 min at 72˚C. A size-selection followed by a final clean-up were then performed with Agencourt AMPure XP beads with a ratio of 0.6-0.8:1 for the size selection and 0.85:1 for the final clean-up, and elution was done in 15 μL 0.1X TE buffer. DNA concentration was measured for each sample with the Qubit® dsDNA HS Assay Kit (ThermoFisher Scientific), concentrations were equalized, and samples were pooled according to the barcode and the quality of the libraries (for each pool) was assessed on a Bioanalyzer 2100-expert with the High Sensitivity DNA Kit (Agilent Technologies). Finally, libraries were equimolarly pooled and the final pool was diluted to 2.6 pM.

5.3.7 Sequencing

Twenty-five μL of the 2.6 pM library were used as input for the emulsion PCR, performed with the Ion PGM™ Hi-Q™ View OT2 Kit on the Ion OneTouch™ 2 System (Life Technologies). Sequencing was then performed using the Ion PGM™ Hi-Q™ View Sequencing Kit and an Ion 316™ Chip run on the Ion PGM System (Life Technologies).

5.3.8 Read processing and sequence analyses using CACTUS

Reads from sequencing experiments were processed using the CACTUS pipeline (Figure 5.1), a Python-based (version 3) wrapper script handling quality checks, filtering, assembly and alignment of sequences to reference databases using both custom scripts and third-party dependencies. Details of settings for all the aspects of CACTUS used for

experiments presented in this study are available in Table 5.1. In short, demultiplexed FASTQ reads from the Ion PGM system were processed by CACTUS to identify suboptimal sequences not suitable for assembly. Each read was analyzed for overall average Phred score, number of bases with a lower than average Phred score, and both general sequence length and complexity. Reads were rejected if any of the detected values for a given sequence was outside of the parameters supplied (settings used in this study available in Table 5.1). Reads passing quality checks were subsequently analyzed with CutAdapt version 1.16 (13) to trim any remaining primer and barcode sequences (maximum error rate 0.2). The detection of common contaminant sequences in the form of vector, linker, adapter or primer sequences was performed with BLASTn version 2.7.1+ (14) using the Univec database (build 10.0, http://www.ncbi.nlm.nih.gov/tools/vecscreen/univec/) as reference. Any sequence with a high identity to known contaminants was removed from the dataset. Finally, to reduce computational time during assembly, an optional dereplication step, using VSEARCH version 2.7.1 (15), was performed to merge identical sequences.

Assembly of sequences was performed using a custom "staggered" multi-threaded implementation of the PHRAP (16) assembly application (version 1.090518). Sequences were distributed evenly into pools, with the amount of pools reflecting the amount of processor (CPU) cores available on the analysis server. All pools were then assembled with strict settings (only allowing the joining of high-scoring matches, see Table 5.1) by a dedicated PHRAP process for each pool, each running on a dedicated core. Contig and unassembled sequences for each pool were collected, combined and redistributed over a

lower number of pools and once again assembled with the same settings. This process was performed with a pool size of 32, 16, 8, 4 and 1. After the strict assembly, two more rounds of pooled assembly (pools sizes 4 and 1) were performed using less strict settings (Table 5.1) to facilitate joins and creation of longer contigs.

To improve speed, comparison of contig sequences to reference databases was performed using the built-in CACTUS cloud-blast option. Sequences were split into 15 equal groups, and each group was processed by a dedicated Amazon EC2 instance (c4.4xlarge), using the standalone BLASTn application from the NCBI BLAST+ suite (14) and the NCBI non-redundant nucleotide database as reference (retrieved 13 November 2017). Once complete, BLASTn results were automatically retrieved by CACTUS from each EC2 instance and combined. Translated DNA searches were performed using DIAMOND version 0.9.18.119 (17) through CACTUS, on a single, memory-optimized EC2 instance (x1.16xlarge), using the NCBI non-redundant protein database as reference (retrieved 23 November 2017).

5.3.9 Data visualization and interpretation

MEGAN 6 Community Edition (18) was used to identify viral sequences and for analyzing both taxonomic and predicted functional profiles of metagenomes (see Table 5.1 for specific settings and LCA parameters used for each application). Functional profiles of microbiomes were generated by comparing hits from translated nucleotide searches to known Orthologous Groups (OGs) of proteins by linking query matches to the eggNOG (evolutionary genealogy of genes: Non-supervised Orthologous Groups) database (19).

Comparison of ViDiT-based microbiome data to previously published marker gene data (11) was accomplished by importing 16S rRNA taxa, including abundance data and taxonomy, into MEGAN and using the built in "compare" function to compare samples. Linear Pearson correlation of taxon abundance between microbiome and marker-gene analysis results was calculated using PAST version 3.20 (20).

The visualization of sequencing coverage for AIV genomes was performed by mapping single reads to reference sequences with Geneious version 8.1.6 (Biomatters).

5.4 Results

The validity and versatility of the developed ViDiT library preparation method and CACTUS sequence analysis pipeline were tested for three different microbiological applications: virus discovery, amplicon-based sequencing of influenza A virus genomes and microbial community profiling.

5.4.1 Virus Discovery

Virus discovery was performed on 36 paired oropharyngeal and cloacal swab samples collected from wild birds belonging to 6 different species. Libraries were successfully obtained from NA solutions with low to unmeasurable concentrations (Table 5.2). Samples were sequenced in pools (4 pools of 7 samples and one pool of 6 samples) and a total of 898740 reads were obtained (24965 average per sample). Reads were processed by CACTUS and obtained contigs and unassembled sequences were compared to the NCBI non-redundant database using BLASTn and DIAMOND. Results were visualized in MEGAN and each potential match (in root, virus and not-assigned branches)

was manually inspected. A sequence was considered as potentially viral if a valid match (encompassing at least two-thirds of the sequence if the sequence was shorter than 200 nt or being at least 200 nt for longer sequences, and excluding hits involving low-complexity regions) was identified at either the nucleotide or amino acid level. In total, sequences potentially derived from 25 viral families (including both DNA and RNA viruses) were identified, and these are summarized in Table 5.3 with a complete overview of each match provided as supplementary material to this thesis (Supplementary File 2). The vast majority of the sequences identified as potentially viral were classified as archaeal viruses or bacteriophages (families *Cystoviridae*, *Inoviridae*, *Microviridae*, *Myoviridae*, *Podoviridae*, *Siphoviridae*, and *Tectiviridae*), but viruses infecting plants (*Alphaflexiviridae*, *Geminiviridae*, and viruses without family assignment), insects (*Dicistroviridae*, *Iflaviridae Nodaviridae* and *Parvoviridae*) and vertebrates (*Adenoviridae*, *Caliciviridae*, *Circoviridae*, *Coronaviridae*, *Hepeviridae*, *Papillomaviridae*, *Parvoviridae*, *Picornaviridae*, *Pneumoviridae*, and *Rhabdoviridae*) were also identified. Finally, a small proportion of viruses whose primary host has not yet been identified were also detected (Circo/Geminiviridae and *Genomoviridae*). Some of the identified viral fragments showed a low identity with reference sequences, indicating the potential presence of "novel" viruses within our samples. The presence of several of these viruses was confirmed by specific PCRs in a subset of samples (indicated as underlined and bold in Table 5.3) and both complete genome sequencing and epidemiological investigations on these viruses are currently ongoing.

5.4.2 Amplicon-based sequencing of complete influenza A virus genomes

The complete AIV genomes were amplified by an RT-PCR that can simultaneously amplify all 8 viral genomic segments. Although not all bands corresponding to all the 8 segments could be observed using gel electrophoresis (data not shown), PCR products were purified and used for ViDiT library preparation and sequencing. On average, approximately 18000 reads were obtained for each sample (Figure 5.3) and processed through the CACTUS pipeline. Obtained contigs were aligned to the non-redundant database of the NCBI using BLASTn.

In all cases in which an entire segment was amplified and sequenced, the corresponding reads successfully assembled into one single contig covering the segment. Figure 5.3 shows the sequence coverage obtained for each genome and how, in each case, almost the entire viral genome could be sequenced (genome coverage of 99.0%, 90.4% and 99.6% for 14-AIV_1475, 15-AIV_1590 and 14-AIV_1623, respectively). The lower coverage obtained for the 3 segments coding for the polymerase subunits (PB2, PB1 and PA) was likely due to a lower PCR efficiency for these longest segments. Indeed, bands corresponding to amplicons longer than 1.8 Kb were never observed on agarose gel. Mispriming by primer MBT-12 during the RT reaction is the likely explanation for the partial amplification, and therefore the lower sequencing coverage, obtained for segment 4 of virus 15-AIV_1590. In fact, the sequence GCAGG, which allowed the mispriming, is found in several AIV strains at position 1,045-1,094 of the hemagglutinin gene. Nonetheless, viruses could be successfully typed and the obtained genomes are currently being utilized for AIV epidemiological studies.

5.4.3 Microbial community profiling

Microbial community profiling was performed on four microbiome-enriched DNA samples isolated from a deep-sea carnivorous sponge (*C. grandis*). A total of 2138671 reads (534667 average per sample) were recovered and processed with CACTUS, generating 63041 contigs and 438836 single unassembled sequences larger than 75 nt.

To determine bacterial composition, taxonomic profiles of BLASTn results were computed and analyzed in MEGAN. The majority of the microbe-derived sequences were assigned to taxa within the domain *Bacteria*, however a smaller number of reads were also assigned to the domain *Archaea* and to viruses. Bacterial phylum-level composition showed a diverse assemblage including Proteobacteria, Bacteroidetes, Cyanobacteria, Actinobacteria and Firmicutes, while the detected archaeal phyla were Euryarchaeota and Thaumarchaeota. Detected viral diversity comprised double-stranded DNA viruses (*Caudovirales*, *Mimiviridae* and *Phycodnaviridae*) and single-stranded DNA viruses (*Microviridae*). To test the performance and coverage of the ViDiT library preparation method in microbiome applications, taxonomic profiles (phylum, class, order and family level) of the bacterial community were created for each sample and compared to results obtained previously with HTS of the 16S rRNA marker gene (11). The obtained results are summarized in Table 5.4. ViDiT-based taxonomic profiles showed recovery of the majority of taxa found using 16S rRNA gene data. Relative abundance of taxa not recovered with the ViDiT method (2 for both phylum and class levels, 3 and 6 for order and family levels, respectively) were investigated and found to be rare taxa with an average relative abundance of 0.15% (0.01-0.75%) according to the 16S data. Furthermore, ViDiT-based

taxonomic profiles consistently revealed additional taxa not observed using 16S rDNA analysis and achieved a higher number of family-level assignments. In addition, a high correlation was found for all samples when comparing the number of reads assigned to each taxonomic node between 16S- and ViDiT-based analyses (average Pearson correlation coefficient 0.92 ± 0.01), indicating that taxa were detected with similar abundances with both methods.

Predicted metagenomic functional compositions, which compared sequence matches to Clusters of Orthologous Groups (COGs) and associated functional annotations, showed a wide variety of functional categories (Figure 5.4). More specifically, COGs high in relative abundance included amino acid transport and metabolism, energy production and conversion, as well as replication, recombination and repair, similar to COG distributions described from filter feeding sponges in other studies (21,22). On a more granular level, a total of 1961 unique orthologous groups (OGs) were found. An in-depth study on the identified microorganisms and their predicted functional and ecological roles within this carnivorous sponge is currently ongoing.

5.5 Discussion

High-throughput sequencing has proven to be one of the most valuable instruments in the molecular toolbox. With a wide variety of applications, including full genome sequencing, pathogen discovery, transcriptome characterization and the description of microbiomes, HTS is now a major component of many scientific studies across a wide variety of fields (2).

Different HTS systems utilize unique protocols to prepare libraries appropriate for sequencing. In addition, the wide variety of applications HTS technologies provide creates further divergent workflows for library preparation, most with unique prerequisites for source material, reagents and dedicated equipment. As such, the creation of sequence libraries compatible with HTS can remain challenging. Furthermore, budgetary constraints, niche applications or sub-optimal source materials can all be limiting factors in making HTS-based experiments available to all research groups (4).

Likewise, the complexity of data analyses for HTS methods remains a challenge. The high volumes of data generated by HTS technologies are generally incompatible with traditional desktop computer analyses, and optimized workflows in conjunction with high-end computers are typically needed to process data in a workable timeframe. Although for some specific applications optimized data analysis pipelines exist and are readily available (23,24), for other types of studies, like virus discovery, data analyses can remain a bottleneck.

In this study, we demonstrated that some of the methodological issues can be overcome by using common, low-cost laboratory reagents. We were able to create ViDiT, a cost-efficient library preparation method, which can easily be performed in any laboratory equipped for basic molecular biology analyses, adaptable to most HTS platforms and used for various applications within the field of microbiology starting from diverse types of samples. Coupled to the ViDiT method, we developed CACTUS, a semi-automated, easily implementable pipeline that can handle all ViDiT-generated reads for

various applications. Furthermore, to reduce the costs associated with data analyses, we have optimized the system to run on cloud computers and provided all scripts for its implementation.

ViDiT and CACTUS were successfully used to generate libraries from NA obtained from various types of samples (both liquid and tissue) and identify sequences potentially derived from both DNA (ssDNA and dsDNA) and RNA viruses. Virus discovery in oropharyngeal-cloacal swabs collected from birds allowed us to identify sequences potentially derived from viruses, some of which never described before, belonging to 25 viral families. The vast majority of viral sequences showed identity with reference sequences of bacteriophages and archaeal viruses. Many bacterial and archaeal viruses are often found integrated within the DNA of their host (25,26) and, although the sample pre-processing steps were designed to enrich for viral particles and remove cellular components, we could not determine with certainty whether a proportion of the detected fragments originated from "contaminating" host NA. Furthermore, based on previous observations made by others (27) and in our laboratory concerning certain gemycircularvirus-like sequences (unpublished), we have reasons to believe that some of the detected viral sequences might have originated from DNA contained in kits and reagents used during sample processing. Therefore, further investigations are crucial to establish with certainty the origin of the detected viruses (28)**.** Nevertheless, the presence of several of the identified animal viruses was PCR-confirmed in select samples, excluding the possibly that these derived from reagents and proving the utility of the ViDiT-CACTUS method for virus discovery applications.

The positive characterization of influenza A virus genomes showed that ViDiT can also be used in conjunction with amplicon-based input material and recover full or near-complete genomes. Although specific experiments on RT-PCR optimization prior to the ViDiT library preparation were not performed, we postulate that the main limiting factor in obtaining a homogeneous sequence coverage is RT-PCR efficiency prior to library preparation. Lower coverage was observed for longer segments, presumably less efficiently amplified, and viral culturing prior to genome amplification greatly improves amplification (12). Nevertheless, we were able to obtain almost-complete genomes from NA isolated directly from swab samples and showed how the ViDiT approach can be used to easily characterize amplified genomic targets in a streamlined and multiplex-capable setting. Our results also demonstrate the good performance of CACTUS in handling viral reads and complete genome assembly. We demonstrated that, when sequencing coverage is sufficient, all reads can consistently be assembled in long contigs spanning the complete viral genome.

Lastly, ViDiT can also be a valuable tool to create shotgun libraries for metagenomic studies. ViDiT-based sequencing was successful in characterizing the bacterial, archaeal and viral diversity in a deep-sea carnivorous sponge. By comparing this microbial community data to previously published bacterial 16S rDNA-based community data of the same sponge (11), we showed that ViDiT-CACTUS performs well and can capture all but the least abundant bacterial taxa. While PCR-based library preparation methods can be affected by amplification bias (29), we observed a high correlation and no disproportional shifts in the relative abundance of taxa comparing 16S- and ViDiT-based

data. Besides taxonomy assignment and bacterial community composition determination, methods alternative to partial 16S sequencing, such as ViDiT-CACTUS, have additional benefits, including the creation of longer, high-quality contigs that can provide further information on the role of bacterial communities in specific ecological niches. In this study, these contigs allowed for a finer-scale taxonomic classification of divergent microorganisms that were not resolvable by marker-gene analysis. Furthermore, ViDiT-CACTUS data were suitable to predict the functional metagenome of samples by annotating contig matches to eggNOG. A wide variety of proteins with associated OGs and biological processes were detected, with distributions similar to those found in sponges in other studies (21,22). This analysis allowed us to identify predicted functional characteristics of the metagenome and further studies will investigate whether such functions may be relevant in explaining symbiotic processes between the microbial community and the sponge host.

The data presented here serve as a proof-of-concept that ViDiT can be a valuable library preparation method in a variety of applications. ViDiT fills a unique niche and provides several advantages over commercially available kits. First and foremost, the low cost of this method (at time of writing under 5 USD per sample for ViDiT and under 14 USD for the entire virus discovery process, including pre-treatment, NA isolation, RT and ViDiT) allows for large-volume sample processing at a low cost. In addition, the high sensitivity of ViDiT, as shown by the successful detection of viruses in samples where NA were below detection limit, opens opportunities for the characterization of samples that cannot be handled by other less sensitive methods, including some commercially available

ones. The PCR backbone of ViDiT makes it a flexible and robust method and as such it can be used in a wide variety of applications, demonstrated herein as virus discovery, amplicon sequencing and metagenomic profiling, but could be easily extended to encompass additional tasks. Finally, the ViDiT method can be customized for any sequencing platform. Here, we choose to generate sequence data by using Ion Torrent PGM sequencing, however, adapting the primer sequences to those used by other sequencing platforms would be trivial.

The CACTUS pipeline, as used in this study, was created to facilitate and streamline the assembly of ViDiT data. By wrapping most tasks in one workflow, accessible with a simple configuration file, running assembly and analyses on raw sequence data becomes a straightforward task. Furthermore, users with scripting experience can easily extend CACTUS to accommodate their bioinformatic needs. The custom implementation of the PHRAP (Smith-Waterman based) aligner facilitates the creation of long, high-quality contigs from low-coverage datasets at adequate speeds, while the "out-of-the-box" support for off-loading computationally intensive workloads to the Amazon Elastic Cloud platform allows research groups without local bioinformatics-rated hardware to start generating results quickly and at a low cost, without the need for hardware investment.

In conclusion, we present here ViDiT, a low-cost, robust, sensitive and versatile library preparation method, and the coupled sequence analysis method CACTUS, a customizable pipeline designed to generate high-quality data within a reasonably short time

and without the need of expensive computers. The ViDiT-CACTUS method proved to be useful in a wide range of applications within the field of microbiology and its low complexity for implementation and low reliance on expensive material and equipment makes it highly beneficial for laboratories with limited budgets engaging in HTS experiments with various research goals.

5.6 Acknowledgments

We would like to thank Ashley Kroyer for her careful selection of samples for virus discovery and Dr. Hugh Whitney and Dr. Davor Ojkic for providing influenza virus diagnostic results.

5.7 Funding

This research was supported by funding from the Joint Mink Research Committee, the Newfoundland and Labrador Research & Development Corporation (5404.1115.104), the Natural Sciences and Engineering Research Council of Canada, and the ArcticNet Grant "Hidden Biodiversity and Vulnerability of Hard Bottom and Surrounding Environments in the Canadian Arctic (HiBio)".

5.8 Supplementary materials

This chapter includes two supplementary files available in digital format. A full overview of each viral match found in analyzed samples is available as a Microsoft Excel file. In addition, a detailed step by step protocol to perform ViDiT based experiments is available in PDF format.

5.9 References

1. Morey M, Fernández-Marmiesse A, Castiñeiras D, Fraga JM, Couce ML, Cocho JA. A glimpse into past, present, and future DNA sequencing. *Mol Genet Metab* (2013) **110**:3–24. doi:10.1016/j.ymgme.2013.04.024

2. Reuter JA, Spacek D, Snyder MP. High-throughput sequencing technologies. *Mol Cell* (2015) **58**:586–597. doi:10.1016/j.molcel.2015.05.004

3. Aigrain L, Gu Y, Quail MA. Quantitation of next generation sequencing library preparation protocol efficiencies using droplet digital PCR assays - a systematic comparison of DNA library preparation kits for Illumina sequencing. *BMC Genomics* (2016) **17**: doi:10.1186/s12864-016-2757-4

4. Helmy M, Awad M, Mosa KA. Limited resources of genome sequencing in developing countries: Challenges and solutions. *Appl Transl Genomics* (2016) **9**:15–19. doi:10.1016/j.atg.2016.03.003

5. Djikeng A, Halpin R, Kuzmickas R, DePasse J, Feldblyum J, Sengamalay N, Afonso C, Zhang X, Anderson NG, Ghedin E, et al. Viral genome sequencing by random priming methods. *BMC Genomics* (2008) **9**:5. doi:10.1186/1471-2164-9-5

6. Gansauge M-T, Meyer M. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat Protoc* (2013) **8**:737. doi:10.1038/nprot.2013.038

7. Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M, Turner DJ. Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nat Methods* (2009) **6**:291. doi:10.1038/nmeth.1311

8. de Vries M, Deijs M, Canuti M, Schaik BDC van, Faria NR, Garde MDB van de, Jachimowski LCM, Jebbink MF, Jakobs M, Luyf ACM, et al. A sensitive assay for virus discovery in respiratory clinical samples. *PLOS ONE* (2011) **6**:e16118. doi:10.1371/journal.pone.0016118

9. Sboner A, Mu XJ, Greenbaum D, Auerbach RK, Gerstein MB. The real cost of sequencing: higher than you think! *Genome Biol* (2011) **12**:125. doi:10.1186/gb-2011-12-8-125

10. Spackman E, Senne DA, Myers TJ, Bulaga LL, Garber LP, Perdue ML, Lohman K, Daum LT, Suarez DL. Development of a real-time reverse transcriptase PCR assay for type A influenza virus and the avian H5 and H7 hemagglutinin subtypes. *J Clin Microbiol* (2002) **40**:3256–3260.

11. Verhoeven JTP, Dufour SC. Microbiomes of the Arctic carnivorous sponges *Chondrocladia grandis* and *Cladorhiza oxeata* suggest a specific, but differential involvement of bacterial associates. *Arct Sci* (2017) doi:10.1139/AS-2017-0015

12. Zhou B, Donnelly ME, Scholes DT, St George K, Hatta M, Kawaoka Y, Wentworth DE. Single-reaction genomic amplification accelerates sequencing and vaccine production for classical and Swine origin human influenza a viruses. *J Virol* (2009) **83**:10309–10313. doi:10.1128/JVI.01109-09

13. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* (2011) **17**:10–12. doi:10.14806/ej.17.1.200

14. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinformatics* (2009) **10**:421. doi:10.1186/1471-2105-10-421

15. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* (2016) **4**:e2584. doi:10.7717/peerj.2584

16. Green P. Phrap documentation. (1996) Available at: http://www.phrap.org/phredphrap/phrap.html [Accessed January 25, 2018]

17. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* (2015) **12**:59–60. doi:10.1038/nmeth.3176

18. Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. *Genome Res* (2007) **17**:377–386. doi:10.1101/gr.5969107

19. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, et al. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res* (2016) **44**:D286–D293. doi:10.1093/nar/gkv1248

20. Hammer O, Harper D, Ryan P. PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica* (2001) **4**:

21. Liu M, Fan L, Zhong L, Kjelleberg S, Thomas T. Metaproteogenomic analysis of a community of sponge symbionts. *ISME J* (2012) **6**:1515–1525. doi:10.1038/ismej.2012.1

22. Thomas T, Rusch D, DeMaere MZ, Yung PY, Lewis M, Halpern A, Heidelberg KB, Egan S, Steinberg PD, Kjelleberg S. Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. *ISME J* (2010) **4**:1557–1567. doi:10.1038/ismej.2010.74

23. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* (2010) **7**:335–336. doi:10.1038/nmeth.f.303

24. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, et al. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl Environ Microbiol* (2009) **75**:7537–7541. doi:10.1128/AEM.01541-09

25. Fortier L-C, Sekulovic O. Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence* (2013) **4**:354–365. doi:10.4161/viru.24498

26. Krupovic M, Prangishvili D, Hendrix RW, Bamford DH. Genomics of Bacterial and Archaeal viruses: dynamics within the prokaryotic virosphere. *Microbiol Mol Biol Rev* (2011) **75**:610–635. doi:10.1128/MMBR.00011-11

27. Naccache SN, Greninger AL, Lee D, Coffey LL, Phan T, Rein-Weston A, Aronsohn A, Hackett J, Delwart EL, Chiu CY. The perils of pathogen discovery: origin of a novel parvovirus-like hybrid genome traced to nucleic acid extraction spin columns. *J Virol* (2013) **87**:11966–11977. doi:10.1128/JVI.02323-13

28. Canuti M, van der Hoek L. Virus discovery: are we scientists or genome collectors? *Trends Microbiol* (2014) **22**:229–231. doi:10.1016/j.tim.2014.02.004

29. van Dijk EL, Jaszczyszyn Y, Thermes C. Library preparation methods for next-generation sequencing: Tone down the bias. *Exp Cell Res* (2014) **322**:12–20. doi:10.1016/j.yexcr.2014.01.008

Table 5.1: Configuration and settings used within the CACTUS pipeline

Module and commands	Application or application step		Description
Trimming	**VD & AS**	**M**	
enabled	Yes	Yes	Enable or disable all quality control steps
enableCompression	No	Yes	Merge duplicate reads into a single read, if identity of those sequences is above threshold set in *compressionId*
compressionId	1.00	1.00	Identity sequences must match in order to be merged
enableTrimmer	Yes	Yes	Enable or disable trimming of raw reads based on quality parameters
trimOnlyLength	No	No	Enable or disable trimming beyond the length based dataset trimming
minLength	50	75	Reads under this length are removed and not considered for assembly
minAverageQuality	15	20	Minimum average quality (PHRED, sum of individual base qualities divided by number of bases) a read must obtain to be considered for assembly
maxNumberOfDips	60	20	Maximum number of low quality bases (cut off defined in *dipCheckTreshold*) a read is allowed to have before being discarded
dipCheckTreshold	10	12	Quality threshold under which a base will be considered a low-quality dip
salvageEnabled	Yes	Yes	Enable the recovery of reads which fail quality or compression checks, but are over a certain length (as defined in *salvageLength*)
salvageLength	70	150	Length threshold, above which reads failing quality checks should be salvaged (if enabled through *salvageEnabled*)
checkCompressionRatio	Yes	Yes	Enable or disable the check of compression ratios of reads
compressionRatioCutOff	0.15	0.15	Compression ratio below which a read should be considered low complexity and removed if enabled via *checkCompressionRatio*
enableCutAdapt	Yes	Yes	Enable or disable CutAdapt to remove primers
Univec screening	**VD & AS**	**M**	
enableUnivec	Yes		Enable or disable BLAST + Univec based screening on sequences to remove contaminants
minLengthForUnivecTrim	50		Minimum length a read must have after Univec trimming, otherwise discard
reward	1		See BLAST documentation[1]
penalty	-4		See BLAST documentation[1]
gapopen	3		See BLAST documentation[1]
gapextend	3		See BLAST documentation[1]
dust	Yes		See BLAST documentation[1]
soft_masking	Yes		See BLAST documentation[1]
evalue	700		See BLAST documentation[1]

searchsp	1750000000000		See BLAST documentation[1]
max_target_seqs	1		See BLAST documentation[1]

Assembly with PHRAP	**Strict**	**Broad**	
penalty	-3	-3	See PHRAP documentation[2]
bandwidth	4	4	See PHRAP documentation[2]
minscore	36	30	See PHRAP documentation[2]
maxgap	35	50	See PHRAP documentation[2]
repeat_stringency	0.92	0.90	See PHRAP documentation[2]
forcelevel	1	1	See PHRAP documentation[2]
gap_init	-3	-3	See PHRAP documentation[2]
gap_ext	-2	-2	See PHRAP documentation[2]
Node_seg	16	16	See PHRAP documentation[2]
node_space	12	12	See PHRAP documentation[2]
minmatch	20	18	See PHRAP documentation[2]
maxmatch	28	26	See PHRAP documentation[2]
indexwordsize	10	10	See PHRAP documentation[2]
trim_qual	2	2	See PHRAP documentation[2]
trim_score	2	2	See PHRAP documentation[2]

Assembly check	**VD & AS**	**M**	
trimOutHighCompression	Yes	Yes	Enable or disable the check of compression ratios of contigs
compressionRatioCutOff	0.18	0.18	Compression ratio below which a contig should be considered low complexity and removed if enabled via *trimOutHighCompression*

Reference alignment BLASTn	**VD & AS**	**M**	
Penalty	-1	-1	See BLAST documentation[1]
Reward	1	1	See BLAST documentation[1]
Gapopen	2	2	See BLAST documentation[1]
Gapextend	1	2	See BLAST documentation[1]
Evalue	10	10	See BLAST documentation[1]
max_hsps	10	10	See BLAST documentation[1]
culling_limit	1	1	See BLAST documentation[1]
num_alignments	10	10	See BLAST documentation[1]
num_descriptions	10	10	See BLAST documentation[1]

Reference alignment DIAMOND	VD & AS	M	
sensitive	Yes	Yes	See DIAMOND documentation[3]
evalue	10000	10000	See DIAMOND documentation[3]
maxtargetseqs	5	5	See DIAMOND documentation[3]

MEGAN LCA settings	**Virus discovery**	**Microbiome taxonomy**	**Microbiome functional profile**	
Min score	10	80	50	See MEGAN documentation[4]
Max expected	10000	10	0.1	See MEGAN documentation[4]
Min percent identity	0	60	0	See MEGAN documentation[4]
Top percent	100	100	10	See MEGAN documentation[4]
Min support percent	1	0.01	0	See MEGAN documentation[4]
Min support	0	1	1	See MEGAN documentation[4]

Min complexity	0	0.3	0	See MEGAN documentation[4]
LCA algorithm	Naïve	Naïve	longReads	See MEGAN documentation[4]
Percent to cover	100	100	80	See MEGAN documentation[4]
Read assignment mode	Read count	Read count	readMagnitude	See MEGAN documentation[4]

VD: virus discovery, AS: amplicon sequencing, M: microbiome profiling

[1] https://www.ncbi.nlm.nih.gov/books/NBK279690/
[2] http://www.phrap.org/phredphrap/phrap.html
[3] https://github.com/bbuchfink/diamond/blob/master/diamond_manual.pdf
[4] http://ab.inf.uni-tuebingen.de/data/software/megan6/download/manual.pdf

Table 5.2: Nucleic acid concentrations in extractions used for virus discovery

Sample	Concentration*	PCR-confirmed viral presence
B1	4.3	
B2	<2	
B3	<2	
B4	<2	
B5	<2	
B6	<2	+
B7	8.7	
B8	36.7	+
B9	2.2	
B10	<2	
B11	<2	
B12	<2	
B13	<2	+
B14	3.5	
B15	<2	
B16	<2	
B17	<2	
B18	<2	
B19	4.3	
B20	3.8	
B21	<2	
B22	3.1	
B23	<2	
B24	<2	
B25	<2	
B26	<2	
B27	8.2	
B28	12.6	
B29	1.8	+
B30	NA	+
B31	6.8	
B32	<2	
B33	7.4	
B34	<2	+
B35	9.5	
B36	5.3	

* Concentration (ng/µl) measured with the Nanodrop N-1000 (detection limit: 2 ng/µl).
NA=not available

Pool (N reads)	Potential viral fragments							
	Viral family	Putative host	BLASTn & DIAMOND results			Fragments (N)	Reads (N)	Fragment length(s)/ range
			Closest relative*	Identity nt (%)†	Identity aa (%)†			
B1-7 (64892)	*Alphaflexiviridae*	Plant	White clover mosaic virus	94-98	94-100	7	12	71-372
	***Caliciviridae*‡**	Vertebrates	Goose calicivirus	U-80	63-95	10	17	58-500
	Circo/Geminiviridae	Unknown	Wastewater_Circular_virus_FL16	69-74	60-74	2	2	139-254
	Circoviridae	Vertebrates	Various	78	77	1	1	171
	Dicistroviridae	Insect	Hubei picorna-like virus 15	U-79	53-92	4	4	260-459
	Genomoviridae	Unknown	Various	U-87	40-95	5	17	267-462
	Iflaviridae	Insect	Various	U-61	40-58	3	3	161-386
	Microviridae	Bacteria	Various	U-78	30-79	28	34	170-442
	Myoviridae	Bacteria/Archaea	*Pseudomonas* phage (various)	U-71	41-77	6	7	202-406
	***Parvoviridae*‡**	Animal	Various	U	50-55	3	13	216-416
	Podoviridae	Bacteria	Various	63-75	U-78	4	4	248-589
	Siphoviridae	Bacteria/Archaea	*Enterococcus* phage IME-EFm5/IME-EFm1	U-91	U-100	44	112	65-920
	Unknown	Unknown	Various	U-64	38-63	5	5	184-365
B8-14 (97262)	*Circoviridae*	Vertebrates	*Acartia tonsa* copepod circovirus	U	50	1	1	305
	Genomoviridae	Unknown	Sewage-associated gemycircularvirus-10b	79	83	1	1	279
	Microviridae	Bacteria	Gokushovirus WZ-2015a	68	53	1	1	270
	Myoviridae	Bacteria/Archaea	*Pseudomonas* phage (various)	U-82	U-87	102	189	85-854
	***Papillomaviridae*‡**	Vertebrates	*Pygoscelis adeliae* papillomavirus	U-68	38-67	3	3	171-249
	Podoviridae	Bacteria	*Pseudomonas* phage (various)	44-85	U-84	14	24	135-771
	Siphoviridae	Bacteria/Archaea	Various	U-70	48-79	6	9	54-566
	Unknown	Unknown	Various	U	41-57	3	10	222-506
B15-21 (27909)	*Myoviridae*	Bacteria/Archaea	Various	U-66	41-95	3	4	201-338
	Siphoviridae	Bacteria/Archaea	*Gordonia* phage (various)	U-63	U-71	2	2	64-210

	Unknown	Unknown	*Bradyrhizobium* phage BDU-MI-1	80	U	1	1	51
B22-28 (103007)	*Microviridae*	Bacteria	Gokushovirus WZ-2015a	71	80	1	1	244
	Myoviridae	Bacteria/Archaea	*Pseudomonas* phage (various)	U-99	U-100	275	3173	96-2760
	Podoviridae	Bacteria	Various	U-99	U-100	94	206	86-1023
	Siphoviridae	Bacteria/Archaea	Various	U-98	U-100	262	5664	62-5724
	Unknown	Unknown	Various	U-89	28-86	56	1701	172-1961
B29-36 (605670)	*Adenoviridae*	Vertebrates	Various	U-73	56-77	2	2	158-436
	Alphaflexiviridae	Plant	Various	95-98	97	2	2	119-244
	Circoviridae	Animal	Bat circovirus	67-82	46-82	2	5	133-210
	***Coronaviridae*‡**	Animal	Various	U-83	U-90	7	9	78-328
	Cystoviridae	Bacteria	*Pseudomonas* phage phi13	96	100	1	1	223
	Geminiviridae	Plant	Sugarcane streak virus	U-65	40-61	4	5	62-672
	Genomoviridae	Unknown	Pacific flying fox faeces associated gemycircularvirus	U	63	1	1	77
	Hepeviridae	Animal	Barns Ness breadcrumb sponge hepe-like virus	78-98	U-99	2	4	353-379
	Inoviridae	Bacteria	Enterobacteria phage f1	59-97	100	3	46	87-459
	Microviridae	Bacteria	Various	U-82	U-88	11	15	105-309
	Myoviridae	Bacteria/Archaea	Various	U-100	U-100	2092	15389	51-2568
	Nodaviridae	Animal	Various	U-98	37-95	4	9	95-349
	***Papillomaviridae*‡**	Vertebrates	Bird papillomaviruses	U-89	U-76	17	29	181-610
	Parvoviridae	Animal	*Blattella germanica* densovirus	U-87	38-66	2	2	163-271
	Picornaviridae	Vertebrates	Sicinivirus Pf-CHK1/SiV/Sicinivirus A	73	46	1	2	354
	***Pneumoviridae*‡**	Vertebrates	Human metapneumovirus	U-82	67-96	3	4	63-209
	Podoviridae	Bacteria	Various	U-100	U-100	1192	6490	52-1947
	Rhabdoviridae	Vertebrates	*Tupaia* virus	U	43-45	2	11	321-586
	Siphoviridae	Bacteria/Archaea	Various	U-100	U-100	1163	247327	51-16869
	Tectiviridae	Bacteria	*Bacillus* phage Wip1	66-71	39-48	2	4	139-319
	Unassigned	Plant	Caledonia beadlet anemone sobemo-like virus 1	U-93	64-97	6	25	174-1129
	Unknown	Unknown	Various	U-99	U-100	807	15253	52-4354

*Majority (>65%) of the hits; † The label "undetermined" (U) was assigned when t
‡Viral presence PCR-confirmed and complete genome sequencing ongoing

taxonomic levels, identified wit 6S rRNA gene sequencing
ion) from a carnivorous sponge (

Sample	Phylum				Class				Order				Family			
	AS 2	AS 3	AS 6	AS 9	AS 2	AS 3	AS 6	AS 9	AS 2	AS 3	AS 6	AS 9	AS 2	AS 3	AS 6	AS 9
ViDiT	28	23	35	28	51	40	63	47	110	93	139	105	214	179	269	204
16S rRNA reference method*	14	9	21	9	20	17	28	13	38	22	56	19	40	18	72	18
Not recovered by ViDiT-CACTUS	2	1	5	1	4	1	1	1	8	1	3	1	12	3	7	3
Cumulative abundance of taxa missed by ViDiT-CACTUS(%)	0.1	0.0	0.1	0.1	0.2	0.0	0.1	0.1	0.3	0.0	0.1	0.1	0.4	0.1	0.8	0.2

* Data from (11)

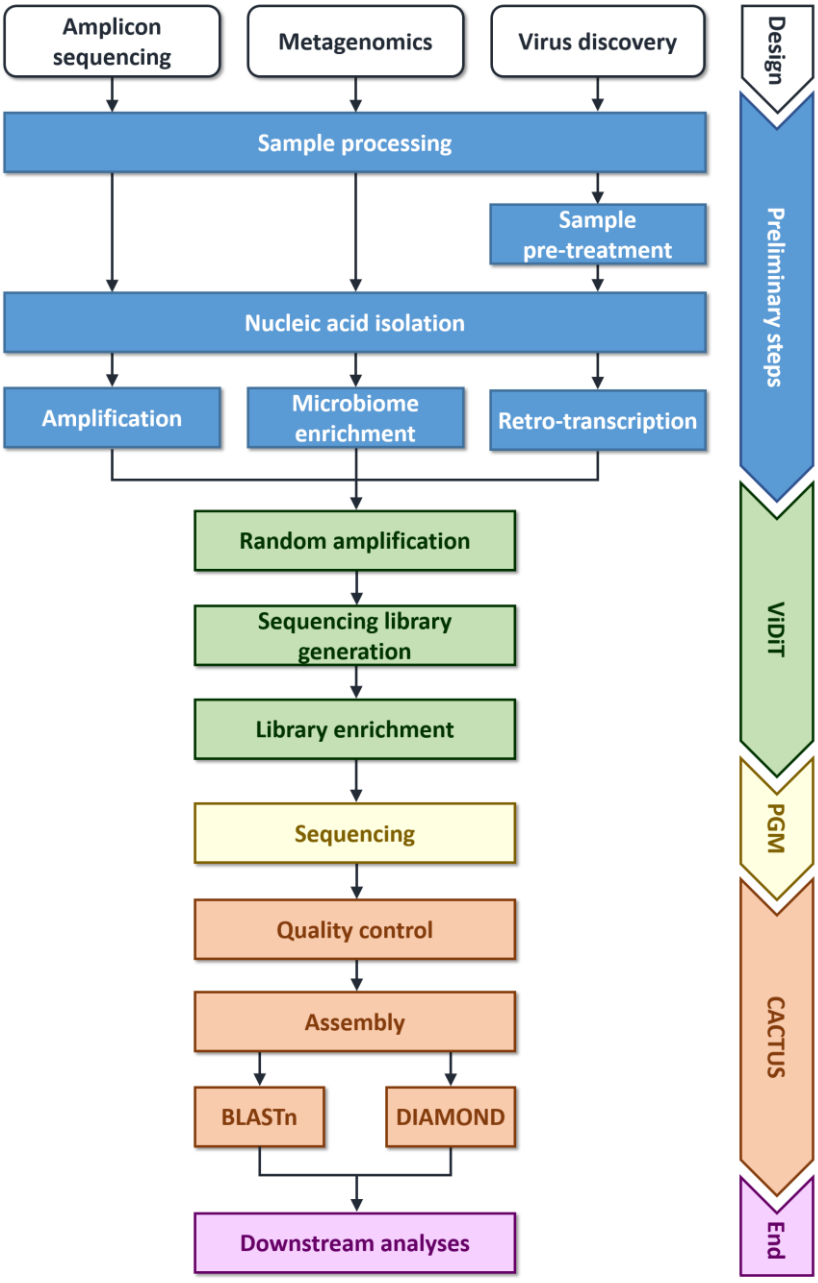

Figure 5.1: Overview of the different procedural steps used in this study. Depending on the type of study, different sample pre-processing methods and steps for nucleic acid isolation and purification were used. Library preparation of the obtained nucleic acids (amplicons, DNA or cDNA) was then performed using the ViDiT method, a three-step random-PCR-based library preparation technique. Libraries were sequenced with the Ion Torrent PGM and the obtained reads were analyzed with the CACTUS pipeline.

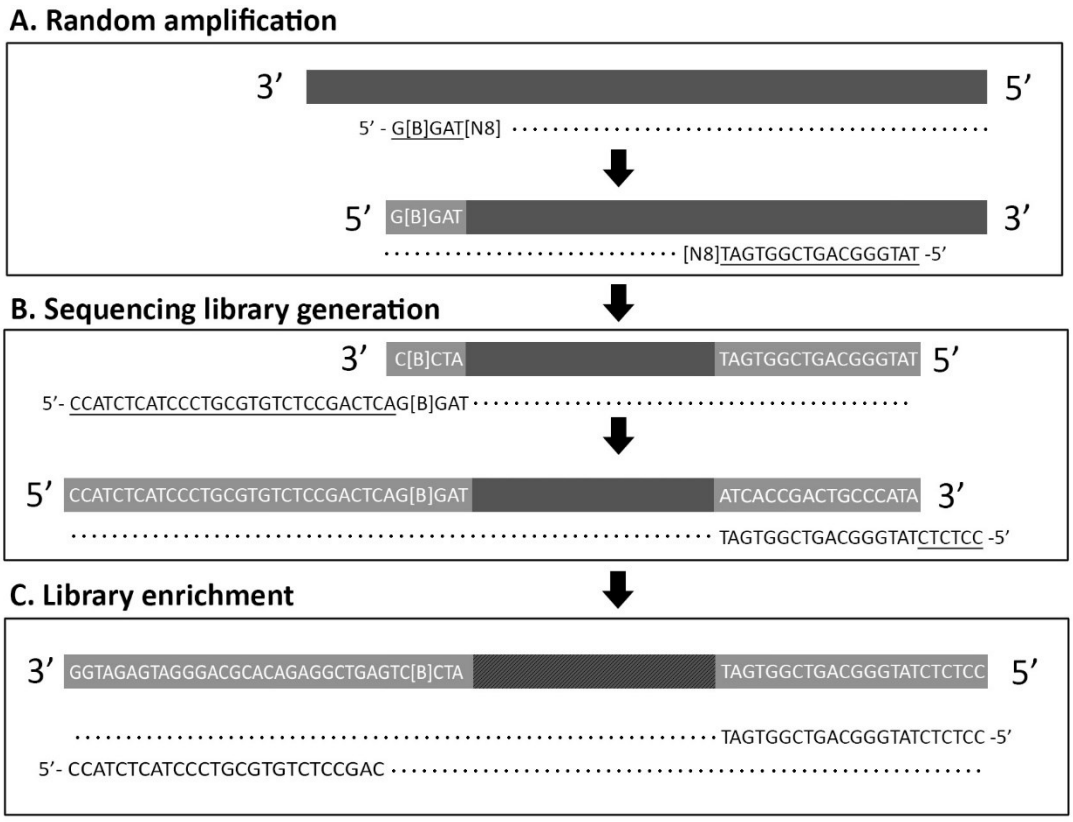

Figure 5.2: Schematic representation of the ViDiT library preparation method. The in-house library preparation method ViDiT optimized for Ion Torrent PGM sequencing includes three fundamental steps, marked here by square boxes. In this schematic, unknown DNA is represented by a dark grey rectangle, the incorporated adaptor sequence by a light grey rectangle, primer sequences are indicated below the DNA molecules with the primer tails underlined, and the dotted line indicates complementary DNA being synthetized. A. During the random amplification, random octamers [N8] with a tail containing either the sequence for the Ion Torrent linker sequence (GAT), the 12-nt long Ion Xpress™ Barcode [B] and part of the A primer (on the left) or part of the P primer (on the right) are used to randomly amplify portions of DNA. B. During the sequencing library generation primers containing the whole Ion Torrent PGM adaptor sequences are used to produce amplicons ready for sequencing. C. Finally, the library enrichment step uses Ion Torrent PGM A and P primers to increase the library concentration.

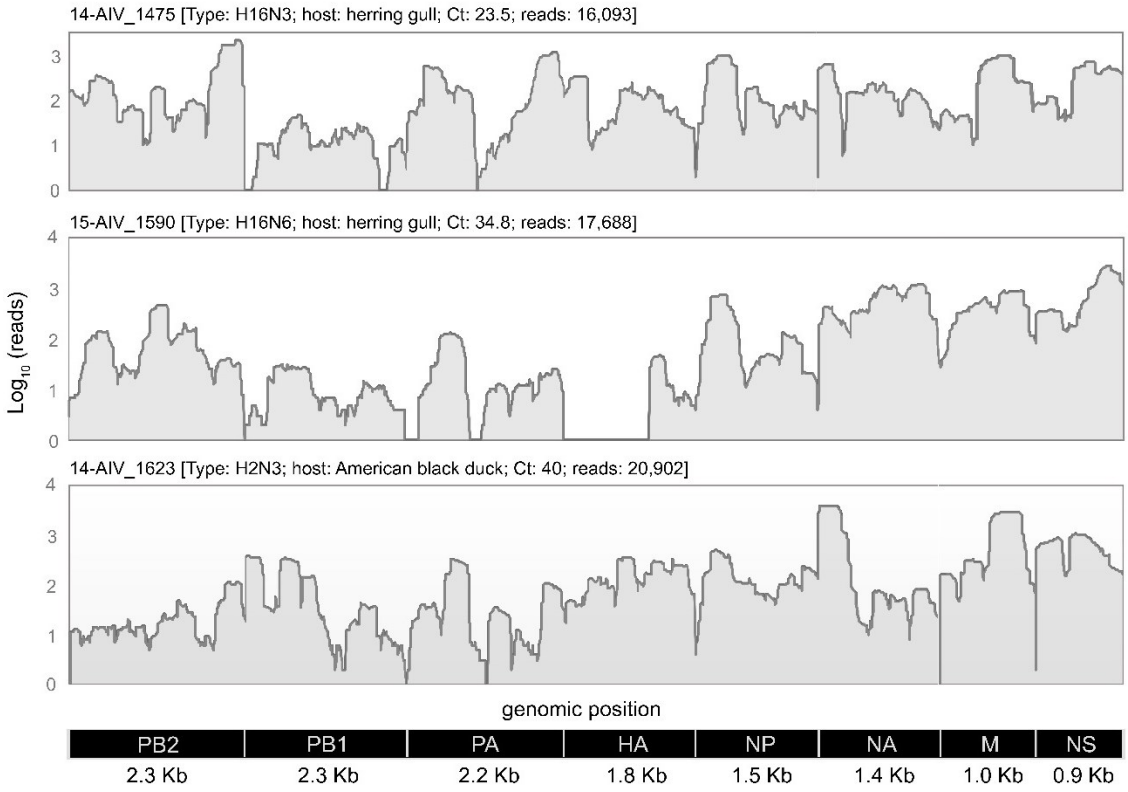

Figure 5.3: Depth and coverage of AIV whole-genome sequencing. Sequencing depth is shown for each virus as Log_{10}-transformed number of reads (Y axis) for each nucleotide of the AIV complete genome (X axis). Genomic segments have been concatenated and include: segments coding for the viral polymerases (segment 1: PB2; segment 2: PB1; segment 3: PA), those coding for the structural proteins hemagglutinin (segment 4: HA), nucleoprotein (segment 5: NP), neuraminidase (segment 6: NA) and matrix 1 and 2 (segment 7: M), and the segment coding for the non-structural proteins NS1 and NS2 (segment 8: NS). The approximate segment size (in kb) is indicated for each segment. For each strain the viral type, the host, the Ct-value obtained during the RT-PCR screening, and the number of obtained reads are indicated in square parenthesis.

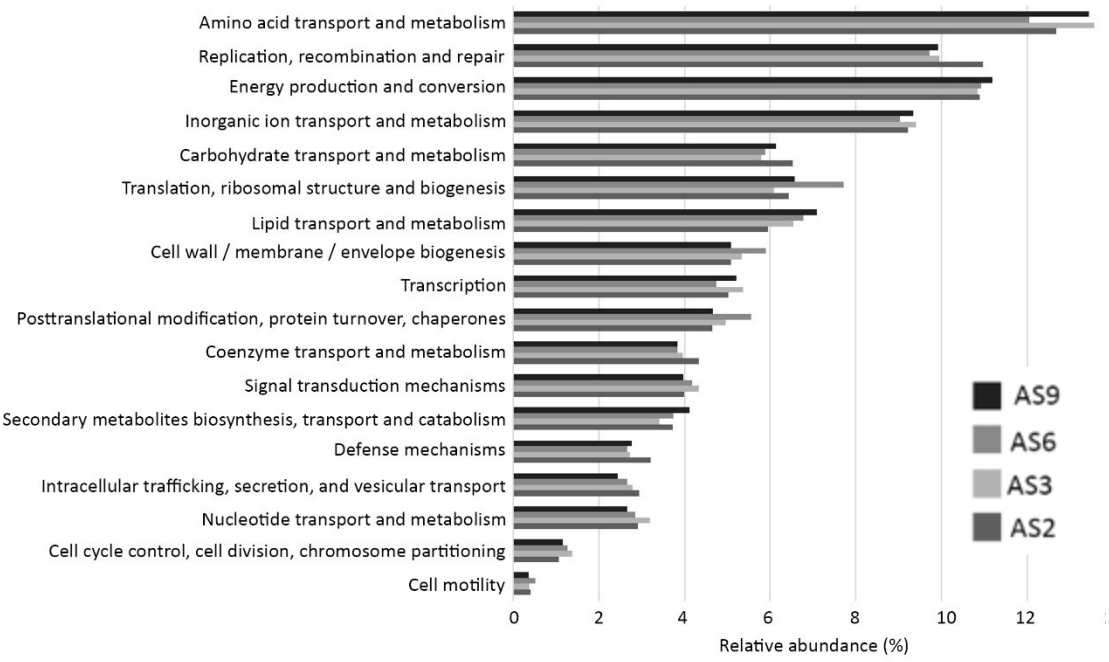

Figure 5.4: Cluster of orthologue groups (COG) categories detected in the bacterial communities associated with the carnivorous sponge *Chondrocladia grandis*. Presented are all COG categories detected, with count data displayed as percentage, calculated as the number of reads in each category divided by the sum of counts from all categories. AS2-9 indicate replicates of *C. grandis* samples.

CHAPTER 6: METAGENOMIC ANALYSIS OF THE CARNIVOROUS SPONGE *CHONDROCLADIA GRANDIS*

6.1 Abstract

Carnivorous sponges (family Cladorhizidae) are distinct from other filter feeding sponges because of their ability to trap and digest mesoplanktonic prey. Previous studies have suggested that host health and carnivory are supported by the associated microbial community. Here, we used ViDiT (a random shotgun metagenomic approach) sequencing to characterize the bacterial, archaeal and viral diversity and detect potential symbionts within the microbiome of the carnivorous sponge *Chondrocladia grandis*. Our results reveal a bacterial community dominated by the phyla Proteobacteria, Bacteroidetes and Firmicutes. In addition, while not abundant, a large diversity of Actinobacteria was also found. The archaeal community was composed mainly of Thaumarchaeota and Euryarchaeota, with many taxa identified as ammonia-oxidizing archaea from the Nitrosopumilales. We identified several potential symbiotic bacteria, including *Colwellia* species which we hypothesize is involved in prey degradation, as well as a significant number of methanotrophic microorganisms (*Methylophaga*, *Methylomonas*, *Methylomicrobium*, *Methylomarinum* and *Methylobacter*), suggesting bacterial symbiont mediated chemolithoautotrophy, as well as hydrocarbonoclastic (*Alcanivorax*, *Thalassolituus*, *Oleiphilus* and *Oleispira*) genera. Viral diversity was dominated by bacteriophages, which are most likely targeting bacteria within the microbiome, potentially playing key roles in horizontal gene transfer. A wide variety of predicted metabolic pathways further suggests a wide repertoire of metabolic capabilities within the microbiome of *C. grandis*. In all, this study reveals many novel members of the *C. grandis* microbiome and shows the benefit of using in-depth metagenomic sequencing to increase our understanding of the true biological diversity within (carnivorous) sponges

6.2 Introduction

The Cladorhizidae (1) is a unique family of sponges (Porifera) observed in many different environments, ranging from littoral caves to the deep sea (2). Unlike the majority of sponges, cladorhizids do not perform filter feeding but are distinctly capable of capturing, engulfing and digesting live prey ranging from planktonic organisms to small crustaceans (3). Several morphological adaptations enable this carnivory, including an erect morphology and structures with adhesive surfaces and filaments that can ensnare live prey (3,4), which are thought to have evolved to overcome the limitations of filter-feeding in deep-sea oligotrophic niches (5). When prey is successfully captured, motile sponge cells migrate to the capture site and quickly encapsulate the prey, creating a temporary cavity in which digestion can occur (5).

In filter-feeding sponges, the mesohyl-associated microbial community provides various functions relating to the metabolism of the host sponge, including nitrogen cycling, carbon fixation, and the production of biologically active secondary metabolites (6). Likewise, it is suspected that the microbial associates of carnivorous sponges play an equally central role in maintaining host health, potentially extending to functions that include the degradation of prey, such as the breakdown of complex molecules (7). Indeed, previous studies describing feeding processes in the carnivorous sponge *Asbestopluma hypogea* observed the presence of unusually numerous sponge-associated bacteria around prey capture sites, supportive of a central bacterial role in prey degradation (5).

Previously, we reported on the bacterial community composition associated with the carnivorous sponge *Chondrocladia grandis*, collected within Baffin Bay and the Gulf of Maine (7,8). Through the use of 16s rRNA gene analysis, we investigated both the overall sponge associated bacterial community and specific bacterial taxa that were differentially abundant within

distinct anatomical regions of the sponge. Using this approach, we identified several potentially symbiotic bacterial genera (7,8), with suspected roles in prey-associated polysaccharide degradation (*Colwellia* and *Reichenbachiella*) and potentially hydrocarbon degradation (*Robiginitomaculum*). However, 16S rRNA gene analysis offers limited resolution when reconstructing the taxonomic composition of microbiomes, as classification can generally only be reliably attained to the genus level (9). Furthermore, elucidating microbial functions and detecting other microbiome members such as archaea (depending on which 16S rRNA region is sequenced), and viruses is not feasible with 16S rRNA gene analysis, and as such, important components of the microbiome may remain unanalyzed or misinterpreted.

In this study, we performed shotgun metagenomics using ViDiT, a random-amplification method coupled with Ion Torrent High-Throughput Sequencing (HTS) to investigate the metagenomic composition of *Chondrocladia grandis*. We compare the bacterial diversity to the previously published 16S rRNA gene data, investigate potential symbiont genera, describe the archaeal and viral communities and predict functional and metabolic pathways present in *C. grandis.* This is, to our knowledge, the first metagenomic description of *Chondrocladia grandis* and of deep-sea carnivorous sponges in general.

6.3 Material and methods

6.3.1 Sample collection

The *Chondrocladia grandis* specimen used in this study was collected off the coast of Baffin Island (Qikiqtaaluk Region) in Scott Inlet (71° 30.9911′ N, 70° 01.0464′ W), from a depth of 539 meters on the 17th of October 2015. The Super Mohawk II (SUMO) ROV was used to retrieve the sponge specimen, which after collection was immediately placed on ice before being dissected for sub-sampling, as previously described in Verhoeven and Dufour, 2017 (7) .

6.3.2 DNA extraction and sequencing

Nucleic acids (NA) were isolated from four distinct anatomical sponge sections: root, sphere, axis and stem. Thawed sections were kept on ice and rinsed three times with PBS to remove loosely associated bacteria and other contaminants, and NAs were subsequently isolated using the MO BIO PowerViral DNA/RNA isolation kit (MO Bio Laboratories) according to manufacturer's specifications. Microbiome-enriched NA solutions were created by using the NEBNext Microbiome DNA Enrichment Kit (New England Biolabs) to remove any methylated DNA.

Random DNA shotgun sequencing libraries were created using the ViDiT library preparation method (10) and sequenced on the Ion Torrent PGM using the Ion PGM™ Hi-Q™ View OT2 Kit on the Ion OneTouch™ 2 System (Life Technologies). Raw sequences were trimmed for primer contaminants and low quality regions using CACTUS (10).

6.3.3 Taxonomic classification and functional annotation

Taxonomic classification of sequencing reads was obtained with Kaiju (11), using the NCBI BLAST (*nr*+euk) reference database and the "greedy" run mode (minimum length match: 11, minimum match score: 75, allowed mismatches: 5). Kaiju classifies reads based on protein-level matches to the reference database, and in the case of multiple similar scoring matches generates a lowest common ancestor assignment. The web-version of mi-faser (12) was used for functional annotation of sequencing reads using the GS+ reference database. Metabolic and functional pathways were analyzed by exporting mi-faser detected enzyme commission numbers and using MinPath (13) to parsimoniously estimate potential biological pathways within our dataset. The Interactive Pathways Explorer was used to visualize and map functional annotations to KEGG pathways (14). Metagenome Krona plots were made using the export Krona export function in Kaiju and KronaTools (15).

6.4 Results and discussion

6.4.1 Overall sequencing results and metagenomic composition

ViDiT based metagenomic shotgun sequencing resulted in a total of 2138671 reads of which 1989987 passed quality control. Out of all sequences, Kaiju was able to classify 237,068 (12%) sequences, which were divided among bacterial (~94%), microbial eukaryote (~3%), archaeal (~3%) and viral (~0.05%) sequences. The relatively low number of classified reads might be attributable to the largely unexplored nature of carnivorous sponge associated microorganisms, and therefore the lack of close relatives within reference databases could have hindered the correct classification of reads (16,17). Furthermore, the Kaiju *nr+euk* database does not contain non-microbial eukaryote reads and, while microbiome enrichment through the removal of methylated host DNA was performed, the success of this method is largely dependent on the density of methylated sites within target genomes. While sponges have varying levels of methylation (18), it is possible that a low methylation rate in *C. grandis* was not conducive to successfully remove host DNA and many of the unclassified read could be *C. grandis* genome fragments. Indeed, this was partially confirmed by classification of sequences using NCBI BLASTn (data not shown) showing a large number of sequences belonging to the phylum Porifera.

6.4.2 Bacterial community composition

The previously published 16S rRNA marker-gene analysis of the *C. grandis* microbiome indicated a dominant role for the phyla Proteobacteria and Bacteroidetes, with the classes Flavobacteriia, Gammaproteobacteria and Alphaproteobacteria being particularly abundant (7). The results in this study show a similar composition (Figure 6.1), with the majority of the reads classified within the phyla Proteobacteria (80%), Bacteroidetes (4%), Firmicutes (1.3%), and classes Gammaproteobacteria (52%) and Alphaproteobacteria (19%). While similar, it is important

to note that the calculated relative abundance represents a simple percentage of the amount of reads from the total dataset mapped to a designated taxon, which may be biased due to differences in bacterial genome sizes and preferential amplification of specific genomic regions. Finally, abundance of some phyla may be under- or over- represented due to distantly related reads not being correctly classified (19,20). Taxonomic classification was generally restricted to genus level, as many reads showed low identity to known sequences in the Kaiju *nr*+*euk* database.

The most interesting discrepancy between shotgun metagenomic and 16S rRNA gene data was the detection of a high abundance of Actinobacteria with the former. In fact, 2% of all bacterial reads obtained with the ViDiT method were classified as Actinobacteria, a phylum that was almost absent from the 16S rRNA sequence pool. Among these reads, the majority was assigned to the orders Streptomycetales, Corynebacteriales, Micrococcales, Pseudonocardiales and Proprionibacteriales, each one being represented by a diverse assemblage of species. Members of the Actinobacteria phylum are extremely diverse, both in physiology and metabolic properties, are well known for their ability to produce a wide variety of bioactive secondary metabolites (21), and are thought to be true symbionts in several filter feeding sponges (22). Secondary metabolites such as those produced by Actinobacteria have important roles for maintaining sponge health, ranging from chemical defenses against predation to anti-overgrowth, antifouling, and antimicrobial defenses (23). The rich diversity of actinobacteria found in this study could indicate an equally important role within carnivorous sponges. Strikingly, we did not detect a significant amount of reads from the Actinobacteria orders – and underlying diversity – during our previous studies using 16S rRNA sequencing. It is known that marker-gene analyses using universal primers can lead to the under-representation of specific bacterial taxa (24) and shows the added value of performing full metagenomic analysis using random amplification.

6.4.3 Diversity of potential bacterial symbionts

Based on 16s rRNA gene sequencing, we previously tentatively identified several genera of potential bacterial symbionts with predicted roles in prey degradation and metabolism, as well as hydrocarbon degradation (7). In this study, we used obtained sequences to further investigate these genera, in an attempt to obtain a better taxonomic resolution and a more reliable prediction of their functional role within the sponge microbiome.

The genera *Colwellia* and *Reichenbachiella* were thought to constitute carnivorous sponge symbionts, acting in prey digestion through degradation of marine polysaccharides, based on the reported chitinase activity of members within these genera and their increased abundance within the trapping spheres of *Chondrocladia grandis* (7,8). Within the *Colwellia* genus, the closest relatives to the identified sequences were *Colwellia psychreythraea, Colwellia sediminilitoris and Colwellia piezophilia,* comprising 23%, 18% and 6% of the *Colwellia* reads, respectively. The high relative abundance of sequences related to *C. psychreythraea* supports the hypothesis that *Colwellia* might be involved in prey breakdown as this species is known hydrolyze chitin (25). In addition, *C. psychreythraea* thrives at low temperatures (-1 °C to 10 °C), temperatures typically observed at *C. grandis* habitats, and produces a repertoire of cold-active enzymes with functionalities in carbon and nutrient cycling, as well as the degradation of high-molecular-weight organic compounds, which in turn might provide additional nutrient conversions beyond chitin (26). Two taxa closely related to *Reichenbachiella faecimaris* and *Reichenbachiella agariperforans* were also detected. Although the genus *Reichenbachiella* has been reported to be growth-stimulated by chitin (27), we did not find any evidence in literature that either *R. faecimaris* or *R. agariperforans* is chitinolytic (28). While the absence of chitinase is unsupportive of the *Reichenbachiella* genus constituting a chitin degrading symbiont, it is interesting to note that *R. agariperforans* is hydrolytic to a wide

variety of other substrates, including agar, gelatin, alginate and DNA (28), which could explain the increased abundance of species within this genera in the trapping spheres.

Chemolithoautotrophic symbiosis has previously been observed within the carnivorous sponge *Cladorhiza methanophila,* where methanotrophic bacteria play a crucial role in converting environmental methane to bioavailable substrates (3). We reported a lack of such bacterial taxa in *C. grandis* (7), however did observe lighter $\delta^{13}C$ stable isotope values in some *C. grandis* samples. This observation lead to the hypothesis that *C. grandis* might obtain additional nutrients through chemosynthetic symbionts utilizing other environmental hydrocarbons as a source of carbon besides methane, potentially through a bacterial symbiont *Robiginitomaculum,* tentatively associated with hydrocarbon degradation (29), which was found at higher abundances in sediment-associated anatomical regions of *C. grandis* (7,8). Unfortunately, the classification of obtained sequences from the *Robiginitomaculum* genus had a low resolution and no specific close relative species could be identified, making detailed analysis of the role of this genus in *C. grandis* impossible.

However, interestingly, we identified several taxa with predicted potential for hydrocarbon degradation as well as several methanotrophic taxa, not observed using our 16S rRNA gene survey. Methanotrophic taxa included, among others, the *Methylophaga*, *Methylomonas*, *Methylomicrobium*, *Methylomarinum* and *Methylobacter* (Table 6.1). In addition, a variety of obligate hydrocarbonoclastic bacterial genera were detected, including *Alcanivorax*, *Thalassolituus*, *Oleiphilus* and *Oleispira* (Table 6.1). Due to their hydrocarbonoclastic nature, these bacteria are highly dependent on utilizing a wide variety of alkanes to fulfill their energy and carbon needs. The detection of these bacteria within sponge tissue - specifically the hydrocarbonoclastic genera, which cannot thrive in the absence alkanes - provides additional evidence that hydrocarbon degradation through symbiotic microbial action might occur within *C. grandis*.

While the described tentative symbiont taxa represent interesting targets for future research, further work is needed to fully elucidate the role of these taxa within *C. grandis*.

6.4.4 Carnivorous sponges harbor a distinct Archaeal community

In addition to the bacterial community, a distinct assemblage of archaeal taxa was found to be associated with *C. grandis*. The majority of archaeal sequences were classified as the phylum Thaumarchaeota (88% of Archaea) while the Euryarchaeota phyla was found at a lower abundance (9%).

While a large portion of Thaumarchaeota reads (62%) were not classified beyond phylum level a diverse group of ammonia-oxidizing archaea from the Nitrosopumilales order were found (30), including the genera *Nitrosopumilus*, *Nitrosoarchaeum* and *Nitrosomarinus*. Most sponges excrete ammonium as a metabolic waste product, and as such the ability to metabolize nitrogenous waste through ammonia- and nitrite-oxidation is critical for survival (31). It has previously been suggested that the co-occurrence of archaeal ammonium oxidizers and filter-feeding sponges could constitute a mutualistic relationship, facilitating this nitrification process needed for host health (31,32). The detection of numerous ammonia-oxidizing archaea in *C. grandis* indicates a similar relationship may be present within carnivorous sponges.

Similar to the Thaumarchaeota, a large portion (63%) of Euryarchaeota were unclassified and could not be further investigated. Interestingly, a diverse assemblage of a methanogenic archaea classes was detected, including Methanomicrobia (17%), Methanobacteria (4%) and Methanococci (2%). Although methanogenic archaea have previously been detected in sponges (33), it is unknown if their presence could constitute a symbiotic relationship, and what exact role methanogens play within sponges. As previously suggested, the often-anaerobic methanogenic

archaea could survive within anaerobic micro niches (34), where metabolic byproducts of other bacterial symbionts are utilized for their metabolism (33).

6.4.5 Viral diversity

Relatively little is known on the viral diversity associated with marine sponges. In addition to the potential for sponge host cell infection, viruses can affect the sponge-associated microbial population either directly through bacterial cell lysis, or by contributing to the exchange of genomic information through horizontal gene transfer (35). Similar to what has been described in other sponge viromes (35,36), the majority of viral fragments in *C. grandis* were assigned to the *Microviridae*, *Phycodnaviriadae*, *Caudovirales* (including the *Podoviridae*, *Siphoviridae* and *Myoviridae*), *Mimiviridae* and *Iridoviridae*.

Microviridae representatives included unclassified marine Gokushovirus and Pequeñoviruses. Gokushoviruses are widespread in aquatic ecosystems (37), but their possible hosts and ecological roles are not fully understood (38). Likewise, it is unclear if the Gokushoviruses detected here are lytic towards bacteria within the sponge microbiome or are transiently present from the water column. The identification of a Pequeñovirus is intriguing as these phages were shown to be highly abundant at deep-sea methane seep habitats (39), where it is speculated they might be involved in providing key metabolism genes and physiological enhancements to their hosts through the expression of virally encoded genes and horizontal gene transfer, respectively, with Methylococcaceae as one of the virus's suspected hosts (39). As discussed earlier, we suggest the multiple Methylococcaceae taxa found in *C. grandis* could constitute symbiotic methanotrophic relationships, and a similar interplay between these bacteria and the pequeñoviruses could be hypothesized.

Sequences identified to belong to members of the viral families *Podoviridae*, *Siphoviridae* and *Myoviridae* most likely represent an assemblage of viruses that can infect bacteria which are associated with the *C. grandis* microbiome. Furthermore, as the aforementioned viral families are all known to contain integrating species (40,41), it is possible that a selection of such viruses could play a role in horizontal gene transfer within the sponge microbiome. Likewise, since no cell-removal pre-treatment was performed before library preparation, it is likely that some of the detected fragments originate from integrated prophages, rather then from infectious viral particles. In any case, as prophages are dormant viruses, the significant of our finding is independent from the source (host or virus) of the sequenced DNA.

The presence of algae infecting *Phycodnaviriadae* is unusual, as the host for these viruses are phototrophic algae (42) that are not expected to occur within the lightless deep depths of the *C. grandis* habitat. Although highly speculative, a potential explanation for the presence of viral fragments classifiable within this viral family in *C. grandis* could be the ingestion of prey, such as copepods, which during detritus feeding may have taken up remnants of Phycodnaviriadae infected phototrophic microorganisms. Likewise, the amoeba infecting Mimiviridae (43), could have entered through a similar pathway. However, it is also possible that viruses detected here infect hosts that are different to those of their closest relatives and are native to the depths of the *C. grandis* habitat.

Finally, several identified sequences were classified as *Iridovirus*. Interestingly, as arthropods are a natural reservoir for these viruses (44), and the closest relatives to identified viruses were arthropod viruses, it is conceivable that the presence of this viral family in *C. grandis* is a direct consequence of the carnivorous lifestyle. In fact, it is possible that, during feeding on small arthropods prey, such as copepods, *C. grandis* passively acquires iridoviruses. However,

iridoviruses are also capable of infecting invertebrates hosts and we cannot exclude that the detected viruses are sponge pathogens and further studies to determine the prevalence, abundance and distribution of these viruses are required to answer this question.

6.4.6 Functional pathways

Functional annotation of sequencing reads using mi-faser uncovered 538 unique protein families that, through pathway inference, were placed as members of 122 predicted biological pathways (Table 6.2). The majority of identified pathways were of metabolic origin with functions within carbohydrate metabolism, xenobiotics biodegradation and metabolism, amino acid metabolism, lipid metabolism, metabolism of cofactors and vitamins and the biosynthesis of secondary metabolites.

A wide variety of pathways specific for elemental substrates (Table 6.2) were predicted, including nitrogen, sulfur and methane metabolism, in line with the metabolic predictions based on the taxonomic composition of the *C. grandis* microbiome. In addition, a large number of other degradation and metabolism pathways were found, such as pathways specific for a diversity of lipids as well as xenobiotics including aromatic organic compounds, hydrocarbons and steroids. The prediction of these pathways seems to suggest a wide repertoire of metabolic capabilities within the microbiome of *C. grandis*.

Investigating the secondary metabolite associated pathways showed a potential for the production of several antibiotic compounds, such as monobactam, carbapenem, novobiocin, streptomycin and validamycin, and we also identified pathways for aflatoxin and phenazine biosynthesis. Production of secondary metabolites is not unexpected as sponges and associated microbiomes are known to be the most prolific producers of bioactive compounds and the prediction of biosynthesis pathways for a variety of compounds could suggest a similar productivity

within *C. grandis*. Although further studies are necessary to evaluate if this potential for secondary metabolite production can be corroborated, these results are interesting as they might encourage future research in natural product discovery within deep-sea carnivorous sponges.

6.5 Conclusions

The results presented here show that *C. grandis* has a wide variety of microbial associates, including a diverse set of bacteria, viruses and archaea. The bacterial community composition found with metagenomic sequencing was comparable to results obtained with 16S rRNA gene analysis (7), but, the detection of additional taxa and a more refined taxonomic assignment shows the benefit of using in-depth metagenomic sequencing to study the composition and functional potential of microbial communities, including those of (carnivorous) sponges. While many microbial taxa with a potential chemoautotrophic symbiotic role were found, additional research is needed to establish if the presence of these taxa does indeed constitute symbiotic relationship and what the exact function of these taxa within the *C. grandis* microbiome.

6.6 References

1. Dendy A. No. I.—Report on the Sigmatotetraxonida collected by H.M.S. "Sealark" in the Indian Ocean. *Transactions of the Linnean Society of London 2nd Series Zoology* (1922) **18**:1–164. doi:10.1111/j.1096-3642.1922.tb00547.x

2. Hestetun JT, Vacelet J, Boury-Esnault N, Borchiellini C, Kelly M, Ríos P, Cristobo J, Rapp HT. The systematics of carnivorous sponges. *Molecular Phylogenetics and Evolution* (2016) **94**:327–345. doi:10.1016/j.ympev.2015.08.022

3. Hestetun JT, Dahle H, Jørgensen SL, Olsen BR, Rapp HT. The microbiome and occurrence of methanotrophy in carnivorous sponges. *Front Microbiol* (2016) **7**: doi:10.3389/fmicb.2016.01781

4. Dupont S, Corre E, Li Y, Vacelet J, Bourguet-Kondracki M-L. First insights into the microbiome of a carnivorous sponge. *FEMS Microbiol Ecol* (2013) **86**:520–531. doi:10.1111/1574-6941.12178

5. Vacelet J, Duport E. Prey capture and digestion in the carnivorous sponge *Asbestopluma hypogea* (Porifera: Demospongiae). *Zoomorphology* (2004) **123**:179–190. doi:10.1007/s00435-004-0100-0

6. Taylor MW, Radax R, Steger D, Wagner M. Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiol Mol Biol Rev* (2007) **71**:295–347. doi:10.1128/MMBR.00040-06

7. Verhoeven JTP, Dufour SC. Microbiomes of the Arctic carnivorous sponges *Chondrocladia grandis* and *Cladorhiza oxeata* suggest a specific, but differential involvement of bacterial associates. *Arctic Science* (2017)1–19. doi:10.1139/as-2017-0015

8. Verhoeven JTP, Kavanagh AN, Dufour SC. Microbiome analysis shows enrichment for specific bacteria in separate anatomical regions of the deep-sea carnivorous sponge *Chondrocladia grandis*. *FEMS Microbiol Ecol* (2017) **93**: doi:10.1093/femsec/fiw214

9. Jovel J, Patterson J, Wang W, Hotte N, O'Keefe S, Mitchel T, Perry T, Kao D, Mason AL, Madsen KL, et al. Characterization of the gut microbiome using 16S or shotgun metagenomics. *Front Microbiol* (2016) **7**: doi:10.3389/fmicb.2016.00459

10. Verhoeven JTP, Canuti M, Munro HJ, Dufour SC, Lang AS. ViDiT-CACTUS: an inexpensive and versatile library preparation and sequence analysis method for virus discovery and other microbiology applications. *Can J Microbiol* (2018)1–13. doi:10.1139/cjm-2018-0097

11. Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nature Communications* (2016) **7**:11257. doi:10.1038/ncomms11257

12. Zhu C, Miller M, Marpaka S, Vaysberg P, Rühlemann MC, Wu G, Heinsen F-A, Tempel M, Zhao L, Lieb W, et al. Functional sequencing read annotation for high precision microbiome analysis. *Nucleic Acids Res* (2018) **46**:e23–e23. doi:10.1093/nar/gkx1209

13. Ye Y, Doak TG. A Parsimony Approach to Biological Pathway Reconstruction/inference for genomes and metagenomes. *PLOS Computational Biology* (2009) **5**:e1000465. doi:10.1371/journal.pcbi.1000465

14. Darzi Y, Letunic I, Bork P, Yamada T. iPath3.0: interactive pathways explorer v3. *Nucleic Acids Res* (2018) **46**:W510–W513. doi:10.1093/nar/gky299

15. Ondov BD, Bergman NH, Phillippy AM. Interactive metagenomic visualization in a web browser. *BMC Bioinformatics* (2011) **12**:385. doi:10.1186/1471-2105-12-385

16. Malmstrom RR, Rodrigue S, Huang KH, Kelly L, Kern SE, Thompson A, Roggensack S, Berube PM, Henn MR, Chisholm SW. Ecology of uncultured *Prochlorococcus* clades revealed through single-cell genomics and biogeographic analysis. *ISME J* (2013) **7**:184–198. doi:10.1038/ismej.2012.89

17. Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, Kunin V, Goodwin L, Wu M, Tindall BJ, et al. A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* (2009) **462**:1056–1060. doi:10.1038/nature08656

18. Regev A, Lamb MJ, Jablonka E. The role of DNA methylation in invertebrates: developmental regulation or genome defense? *Mol Biol Evol* (1998) **15**:880–880. doi:10.1093/oxfordjournals.molbev.a025992

19. Jones MB, Highlander SK, Anderson EL, Li W, Dayrit M, Klitgord N, Fabani MM, Seguritan V, Green J, Pride DT, et al. Library preparation methodology can influence genomic and functional predictions in human microbiome research. *Proc Natl Acad Sci USA* (2015) **112**:14024–14029. doi:10.1073/pnas.1519288112

20. Nayfach S, Pollard KS. Toward accurate and quantitative comparative metagenomics. *Cell* (2016) **166**:1103–1116. doi:10.1016/j.cell.2016.08.007

21. Manivasagan P, Venkatesan J, Sivakumar K, Kim S-K. Pharmaceutically active secondary metabolites of marine actinobacteria. *Microbiological Research* (2014) **169**:262–278. doi:10.1016/j.micres.2013.07.014

22. Montalvo NF, Mohamed NM, Enticknap JJ, Hill RT. Novel actinobacteria from marine sponges. *Antonie Van Leeuwenhoek* (2005) **87**:29–36. doi:10.1007/s10482-004-6536-x

23. Pawlik JR. The chemical ecology of sponges on caribbean reefs: natural products shape natural systems. *BioScience* (2011) **61**:888–898. doi:10.1525/bio.2011.61.11.8

24. Farris MH, Olson JB. Detection of Actinobacteria cultivated from environmental samples reveals bias in universal primers. *Lett Appl Microbiol* (2007) **45**:376–381. doi:10.1111/j.1472-765X.2007.02198.x

25. Bowman J, Gosink JJ, McCammon SA, Lewis TE, Nichols DS, Nichols PD, Skerratt JH, Staley JT, McMeekin TA. *Colwellia demingiae* sp. nov., *Colwellia hornerae* sp. nov., *Colwellia rossensis* sp. nov. and *Colwellia psychrotropica* sp. nov.: psychrophilic Antarctic species with the ability to synthesize docosahexaenoic acid (22:ω63). *International Journal of Systematic and Evolutionary Microbiology* (1998) **48**:1171–1180. doi:10.1099/00207713-48-4-1171

26. Techtmann SM, Fitzgerald KS, Stelling SC, Joyner DC, Uttukar SM, Harris AP, Alshibli NK, Brown SD, Hazen TC. *Colwellia psychrerythraea* Strains from Distant Deep Sea Basins Show Adaptation to Local Conditions. *Front Environ Sci* (2016) **4**: doi:10.3389/fenvs.2016.00033

27. Wietz M, Wemheuer B, Simon H, Giebel H-A, Seibt MA, Daniel R, Brinkhoff T, Simon M. Bacterial community dynamics during polysaccharide degradation at contrasting sites in the Southern and Atlantic Oceans. *Environ Microbiol* (2015) **17**:3822–3831. doi:10.1111/1462-2920.12842

28. Cha I-T, Oh Y-S, Park S-J, Park B-J, Lee J-K, Lim C-S, Park A-R, Yoo J-S, Lee D-H, Rhee S-K, et al. *Reichenbachiella faecimaris* sp. nov., isolated from a tidal flat, and emended descriptions of the genus *Reichenbachiella* and *Reichenbachiella agariperforans*. *Int J Syst Evol Microbiol* (2011) **61**:1994–1999. doi:10.1099/ijs.0.026849-0

29. Guibert LM, Loviso CL, Marcos MS, Commendatore MG, Dionisi HM, Lozada M. Alkane biodegradation genes from chronically polluted subantarctic coastal sediments and their shifts in response to oil exposure. *Microb Ecol* (2012) **64**:605–616. doi:10.1007/s00248-012-0051-9

30. Bayer B, Vojvoda J, Offre P, Alves RJE, Elisabeth NH, Garcia JA, Volland J-M, Srivastava A, Schleper C, Herndl GJ. Physiological and genomic characterization of two novel marine thaumarchaeal strains indicates niche differentiation. *The ISME Journal* (2016) **10**:1051–1063. doi:10.1038/ismej.2015.200

31. Zhang F, Pita L, Erwin PM, Abaid S, López-Legentil S, Hill RT. Symbiotic archaea in marine sponges show stability and host specificity in community structure and ammonia oxidation functionality. *FEMS Microbiol Ecol* (2014) **90**:699–707. doi:10.1111/1574-6941.12427

32. Turque AS, Batista D, Silveira CB, Cardoso AM, Vieira RP, Moraes FC, Clementino MM, Albano RM, Paranhos R, Martins OB, et al. Environmental shaping of sponge associated archaeal communities. *PLOS ONE* (2010) **5**:e15774. doi:10.1371/journal.pone.0015774

33. Webster NS, Watts JEM, Hill RT. Detection and phylogenetic analysis of novel Crenarchaeote and Euryarchaeote 16S ribosomal RNA gene sequences from a Great Barrier Reef sponge. *Mar Biotechnol* (2001) **3**:600–608. doi:10.1007/s10126-001-0065-7

34. Hoffmann F, Larsen O, Thiel V, Rapp HT, Pape T, Michaelis W, Reitner J. An anaerobic world in sponges. *Geomicrobiology Journal* (2005) **22**:1–10. doi:10.1080/01490450590922505

35. Laffy PW, Wood-Charlson EM, Turaev D, Jutz S, Pascelli C, Botté ES, Bell SC, Peirce TE, Weynberg KD, van Oppen MJH, et al. Reef invertebrate viromics: diversity, host specificity and functional capacity. *Environmental Microbiology* (2018) **20**:2125–2141. doi:10.1111/1462-2920.14110

36. Laffy PW, Wood-Charlson EM, Turaev D, Weynberg KD, Botté ES, van Oppen MJH, Webster NS, Rattei T. HoloVir: A workflow for investigating the diversity and function of viruses in invertebrate holobionts. *Front Microbiol* (2016) **7**: doi:10.3389/fmicb.2016.00822

37. Roux S, Enault F, Robin A, Ravet V, Personnic S, Theil S, Colombet J, Sime-Ngando T, Debroas D. Assessing the Diversity and Specificity of Two Freshwater Viral Communities through Metagenomics. *PLOS ONE* (2012) **7**:e33641. doi:10.1371/journal.pone.0033641

38. Labonté JM, Hallam SJ, Suttle CA. Previously unknown evolutionary groups dominate the ssDNA gokushoviruses in oxic and anoxic waters of a coastal marine environment. *Front Microbiol* (2015) **6**: doi:10.3389/fmicb.2015.00315

39. Bryson SJ, Thurber AR, Correa AMS, Orphan VJ, Vega Thurber R. A novel sister clade to the enterobacteria microviruses (family Microviridae) identified in methane seep sediments. *Environ Microbiol* (2015) **17**:3708–3721. doi:10.1111/1462-2920.12758

40. Fortier L-C, Sekulovic O. Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence* (2013) **4**:354–365. doi:10.4161/viru.24498

41. Krupovic M, Prangishvili D, Hendrix RW, Bamford DH. Genomics of bacterial and archaeal Viruses: dynamics within the prokaryotic virosphere. *Microbiol Mol Biol Rev* (2011) **75**:610–635. doi:10.1128/MMBR.00011-11

42. Wilson WH, Van Etten JL, Allen MJ. The Phycodnaviridae: The story of how tiny giants rule the world. *Curr Top Microbiol Immunol* (2009) **328**:1–42.

43. Claverie J-M, Grzela R, Lartigue A, Bernadac A, Nitsche S, Vacelet J, Ogata H, Abergel C. Mimivirus and Mimiviridae: Giant viruses with an increasing number of potential hosts, including corals and sponges. *Journal of Invertebrate Pathology* (2009) **101**:172–180. doi:10.1016/j.jip.2009.03.011

44. Williams T. Natural invertebrate hosts of iridoviruses (Iridoviridae). *Neotrop Entomol* (2008) **37**:615–632.

45. Boden R, Cunliffe M, Scanlan J, Moussard H, Kits KD, Klotz MG, Jetten MSM, Vuilleumier S, Han J, Peters L, et al. Complete genome sequence of the aerobic marine methanotroph *Methylomonas methanica* MC09. *Journal of Bacteriology* (2011) **193**:7001–7002. doi:10.1128/JB.06267-11

46. Akberdin IR, Thompson M, Hamilton R, Desai N, Alexander D, Henard CA, Guarnieri MT, Kalyuzhnaya MG. Methane utilization in *Methylomicrobium alcaliphilum* 20Z R : a systems approach. *Scientific Reports* (2018) **8**:2512. doi:10.1038/s41598-018-20574-z

47. Hirayama H, Fuse H, Abe M, Miyazaki M, Nakamura T, Nunoura T, Furushima Y, Yamamoto H, Takai K. *Methylomarinum vadi* gen. nov., sp. nov., a methanotroph isolated from two distinct marine environments. *Int J Syst Evol Microbiol* (2013) **63**:1073–1082. doi:10.1099/ijs.0.040568-0

48. Svenning MM, Hestnes AG, Wartiainen I, Stein LY, Klotz MG, Kalyuzhnaya MG, Spang A, Bringel F, Vuilleumier S, Lajus A, et al. Genome Sequence of the Arctic Methanotroph *Methylobacter tundripaludum* SV96. *J Bacteriol* (2011) **193**:6418–6419. doi:10.1128/JB.05380-11

49. Barbato M, Scoma A, Mapelli F, De Smet R, Banat IM, Daffonchio D, Boon N, Borin S. Hydrocarbonoclastic *Alcanivorax* Isolates Exhibit Different Physiological and Expression Responses to n-dodecane. *Front Microbiol* (2016) **7**: doi:10.3389/fmicb.2016.02056

50. Dong C, Chen X, Xie Y, Lai Q, Shao Z. Complete genome sequence of *Thalassolituus oleivorans* R6-15, an obligate hydrocarbonoclastic marine bacterium from the Arctic Ocean. *Stand Genomic Sci* (2014) **9**:893–901. doi:10.4056/sigs.5229330

51. Golyshin PN, Chernikova TN, Abraham W-R, Lünsdorf H, Timmis KN, Yakimov MM. Oleiphilaceae fam. nov., to include *Oleiphilus messinensis* gen. nov., sp. nov., a novel marine bacterium that obligately utilizes hydrocarbons. *International Journal of Systematic and Evolutionary Microbiology* (2002) **52**:901–911. doi:10.1099/00207713-52-3-901

52. Yakimov MM, Giuliano L, Gentile G, Crisafi E, Chernikova TN, Abraham W-R, Lünsdorf H, Timmis KN, Golyshin PN. *Oleispira antarctica* gen. nov., sp. nov., a novel hydrocarbonoclastic marine bacterium isolated from Antarctic coastal sea water. *Int J Syst Evol Microbiol* (2003) **53**:779–785. doi:10.1099/ijs.0.02366-0

Table 6.1: Methanotrophic and hydrocarbonoclastic genera detected in tissue of *Chondrocladia grandis*

Type	Genus	Closest relatives	Reference
Methanotrophic	*Methylophaga*	*M. thiooxydans* *M. lonarensis* *M. frappieri* *M. sulfidovorans* *M. aminisulfidivorans* *M. muralis*	(18)
	Methylomonas	*M. methanica* *M. lenta* *M. koyamae* *M. denitrificans*	(45)
	Methylomicrobium	*M. alcaliphilum* *M. buryatense*	(46)
	Methylomarinum	*M. vadi*	(47)
	Methylobacter	*M. pseudosasicola* *M. aquaticum* *M. extorquens* *M. nodulans* *M. populi* *M. salsuginis* *M. platani* *M. tarhaniae*	(48)
Hydrocarbonoclastic	*Alcanivorax*	*A. pacificus* *A. hongdengensis* *A. nanhaiticus* *A. jadensis* *A. dieselolei*	(49)
	Thalassolituus	*T. oleivorans*	(50)
	Oleiphilus	*O. messinensis*	(51)
	Oleispira	*O. antarctica*	(52)

Table 6.2: Predicted biological pathways present within the microbiome of *C. grandis*.

Category	**Type**	**Modules**
Metabolism	Amino acid metabolism	Arginine biosynthesis Alanine, aspartate and glutamate metabolism Glycine, serine and threonine metabolism Cysteine and methionine metabolism Valine, leucine and isoleucine degradation Valine, leucine and isoleucine biosynthesis Lysine biosynthesis Lysine degradation Arginine and proline metabolism Histidine metabolism Tyrosine metabolism Phenylalanine metabolism Tryptophan metabolism Phenylalanine, tyrosine and tryptophan biosynthesis
	Biosynthesis of other secondary metabolites	Aflatoxin biosynthesis Monobactam biosynthesis Carbapenem biosynthesis Novobiocin biosynthesis Phenazine biosynthesis Streptomycin biosynthesis Acarbose and validamycin biosynthesis Phenylpropanoid biosynthesis Isoquinoline alkaloid biosynthesis Tropane, piperidine and pyridine alkaloid biosynthesis Glucosinolate biosynthesis Biosynthesis of secondary metabolites - unclassified
	Carbohydrate metabolism	Glycolysis / Gluconeogenesis Citrate cycle (TCA cycle) Pentose phosphate pathway Pentose and glucuronate interconversions Fructose and mannose metabolism Galactose metabolism Ascorbate and aldarate metabolism Starch and sucrose metabolism Amino sugar and nucleotide sugar metabolism Inositol phosphate metabolism Pyruvate metabolism Glyoxylate and dicarboxylate metabolism Propanoate metabolism

		Butanoate metabolism C5-Branched dibasic acid metabolism
	Energy metabolism	Oxidative phosphorylation Photosynthesis Methane metabolism Carbon fixation in photosynthetic organisms Carbon fixation pathways in prokaryotes Nitrogen metabolism Sulfur metabolism
	Glycan biosynthesis and metabolism	Other glycan degradation Various types of N-glycan biosynthesis Glycosaminoglycan degradation Lipopolysaccharide biosynthesis Peptidoglycan biosynthesis Glycosphingolipid biosynthesis (globo/isoglobo) Glycosphingolipid biosynthesis (ganglio)
	Lipid metabolism	Fatty acid biosynthesis Fatty acid elongation Fatty acid degradation Synthesis and degradation of ketone bodies Cutin, suberine and wax biosynthesis Secondary bile acid biosynthesis Steroid hormone biosynthesis Glycerolipid metabolism Glycerophospholipid metabolism Arachidonic acid metabolism alpha-Linolenic acid metabolism Sphingolipid metabolism Biosynthesis of unsaturated fatty acids
	Metabolism of cofactors and vitamins	Ubiquinone and other terpenoid-quinone biosynthesis One carbon pool by folate Thiamine metabolism Riboflavin metabolism Vitamin B6 metabolism Nicotinate and nicotinamide metabolism Pantothenate and CoA biosynthesis Biotin metabolism Lipoic acid metabolism Folate biosynthesis Retinol metabolism Porphyrin and chlorophyll metabolism

	Metabolism of other amino acids	beta-Alanine metabolism Taurine and hypotaurine metabolism Phosphonate and phosphinate metabolism Selenocompound metabolism Cyanoamino acid metabolism D-Glutamine and D-glutamate metabolism D-Alanine metabolism Glutathione metabolism
	Metabolism of terpenoids and polyketides	Geraniol degradation Polyketide sugar unit biosynthesis Terpenoid backbone biosynthesis Limonene and pinene degradation Zeatin biosynthesis Insect hormone biosynthesis Biosynthesis of ansamycins Biosynthesis of vancomycin group antibiotics
	Nucleotide metabolism	Purine metabolism Pyrimidine metabolism
	Xenobiotics biodegradation and metabolism	Chlorocyclohexane and chlorobenzene degradation Benzoate degradation Dioxin degradation Xylene degradation Chloroalkane and chloroalkene degradation Naphthalene degradation Aminobenzoate degradation Ethylbenzene degradation Styrene degradation Atrazine degradation Caprolactam degradation Metabolism of xenobiotics by cytochrome P450 Drug metabolism - cytochrome P450 Drug metabolism - other enzymes Steroid degradation
Organismal Systems	Environmental adaptation	Thermogenesis
	Endocrine system	Relaxin signaling pathway
Genetic Information Processing	Translation	Aminoacyl-tRNA biosynthesis

Environmental Information Processing	Signal transduction	Phosphatidylinositol signaling system mTOR signaling pathway PI3K-Akt signaling pathway
Human Diseases	Infectious diseases: Viral	Human cytomegalovirus infection Human papillomavirus infection Human immunodeficiency virus 1 infection

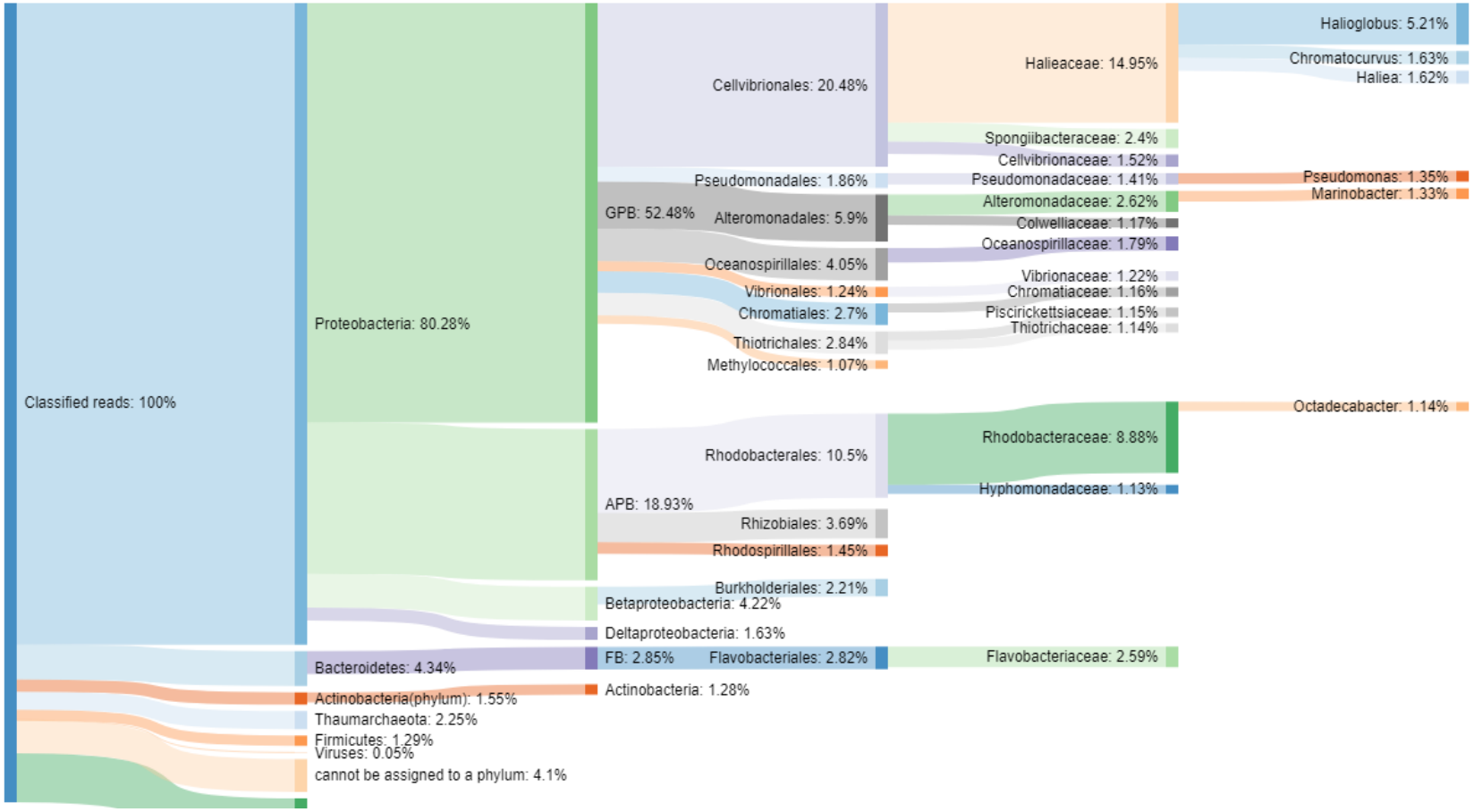

Figure 6.1: Sankey diagram showing the taxonomic composition of the microbiome of *Chondrocladia grandis*. Flows depict the contribution of different taxonomic groups, flowing from phylum (left) to genus (right); the width of each flow corresponds to the relative abundance of a given taxa. Only taxa with a relative abundance of at least 1% are shown. GBP: Gammaproteobacteria, APB: Alphaproteobacteria, FB: Flavobacteriia.

CHAPTER 7: SUMMARY

Over the last decade, improvements in high-throughput sequencing technologies and bioinformatic methods have been at the forefront of revitalizing the field of microbial ecology (1). Tools to explore and accurately characterize microbial communities are now readily available and, with sequencing costs decreasing every year, more easily accessible to scientists from every area of biology (2). These technological advances make it possible to obtain millions of sequences from a single sample, providing a resolution high enough to give an answer to many biological questions at an unprecedented speed (3).

Within the field of microbiology, these methods have already shown their potential. In fact, next-generation sequencing methods are now widely used in microbiology to reveal the diversity of microbial communities in ecological niches, to discover pathogens, evaluate the role of certain communities or study the functioning of organisms (1,3). However, sequence analysis remains a significant bottleneck, and since the science of next-generation sequencing is still evolving rapidly, sequence analysis software packages are often only available for mainstream applications and advanced informatics skills are regularly required to customize available pipelines to personal research needs.

During the course of this PhD program, a key objective was to develop customizable laboratory methods with a broad application in conjunction with more streamlined sequence analysis bioinformatic tools. Development of this toolkit was a crucial step towards the achievement of the second objective, investigating the microbiome of two deep-sea carnivorous sponges, *Chondrocladia grandis* and *Cladorhiza oxeata.*

These sponges, together with the other members of the Cladorhizidae, evolved a carnivorous feeding strategy that is unique to this family within the phylum Porifera. While much has been studied about the relationships between filter-feeding sponges and their microbiome, extremely little is known about the microorganisms inhabiting carnivorous sponges. Indeed, at the start of this PhD program, no microbiome data was available for deep-sea carnivorous sponges. Uncovering the microbiome of carnivorous sponges presented an opportunity to gain valuable insight on relationships between the host organism and their associated microorganisms, understanding the potential role of the microbiome in facilitating carnivory, characterizing the evolution and maintenance of the carnivorous sponge microbiome, and, finally, gaining an understanding of the potential role of carnivorous sponges within their environments.

7.1 Microbiome and metagenomics method development

In microbial ecology, high-throughput sequencing (HTS) has proven to be one of the most valuable instruments in the molecular toolbox (4). Molecular approaches have gained significant preference over more traditional methods, such as culturing, as they are better suited to provide unbiased, fast and comprehensive information on complex microbial communities (1).

7.1.1 SPONS pipeline

One of the most prolific methods used in microbial ecology is community analysis using targeted 16S rRNA gene sequencing, which provides rapid, reliable and comprehensive characterization of microbial niches (1). While the sequencing

methodology is well established, analysis tools for 16S rRNA sequencing experiments are still evolving and maturing. Improved sequence analysis tools are continually released and "best practices" on how to analyze data are still shifting as new algorithms and methods are being established (i.e.: (5)). This quick turn-over can sometimes lead to lagging implementation of novel tools and methods in larger, broad user-group, analysis frameworks such as Mothur and QIIME.

The SPONS pipeline presented in this thesis (Chapter 2) was designed to overcome these issues by being a highly flexible and adaptable workflow that users with bioinformatic experience can tailor to their unique preferences. In the default configuration, SPONS is a fully automated workflow for analyzing 16S rRNA gene sequences, that performs all steps necessary (for example: read trimming, quality control, OTU creation, taxonomic classification) to transform raw datasets into useful, interpretable data. By integrating leading-edge applications, SPONS is highly efficient and can rapidly analyze large datasets with excellent performance.

Within this thesis the SPONS pipeline was a crucial component of the data workflow and was used for analyzing the bacterial diversity within *Chondrocladia grandis* (Chapter 3 and 4) and *Cladorhiza oxeata* (Chapter 4). However, it has also proven to be useful in a number of other studies, across a variety of ecological niches. For example, SPONS was successfully used to characterize the impact of aquaculture operations on bacterial communities in the environment (6,7), describe the bacterial associates of cold-water corals (8), report shifts in the bacterial composition of rat intestines under different

diets (9), detect bacterial associates of ticks (10), and to study potential causative bacterial pathogens in mink pododermatitis (unpublished).

7.1.2 The ViDiT-CACTUS metagenomic tools

Evidently, 16S rRNA gene HTS experiments are extremely useful and widely used. However, the target specific nature of this approach means no additional information outside of the 16S rRNA gene sequence is obtained and, as such, other microbial organisms lacking the 16S rRNA gene (such as viruses, fungi and protozoa) evade detection and cannot be characterized (1). For a broader characterization, metagenomics (or shotgun sequencing) is often used; a method by which all the DNA extracted from a sample is indiscriminately sequenced. In addition to detecting all organisms present in a sample, this method further allows for the reconstruction of complete genomes, which can be valuable to characterize the predicted functional profile of a microbial community.

Unfortunately, metagenomic experiments are a magnitude more expensive than marker-gene analysis, such as 16S rRNA sequencing (4). Often metagenomic experiments require alteration of original sample-derived nucleic acids to create standardized libraries compatible with the intended sequencing instrument (11). Such alterations are usually high in price and, as such, these types of experiments are often outside budgetary considerations or severely limit the number of samples that can be processed (11). Additionally, isolated NA intended for metagenomic sequencing typically needs to conform to strict requirements in terms of quality and quantity, making it difficult to analyze sub-optimal samples (11).

The ViDiT method, as presented in this thesis (Chapter 5), was created to address some of these issues and make metagenomic experiments more streamlined and accessible. ViDiT is a customizable, PCR-based, extremely low-cost and versatile library preparation method, which was coupled with a bioinformatics analysis workflow named CACTUS, which automates and leverages cloud-computing to handle many of the tasks needed to transform shotgun metagenomics reads into insightful data.

In the scope of this work, ViDiT-CACTUS was used to characterize the metagenome of *Chondrocladia grandis* (Chapter 6), where it provided essential insight into the bacterial, archaeal and viral diversity within this carnivorous sponge. In addition, ViDiT-CACTUS has been successfully used in a virus discovery setting, shedding light on novel papillomaviruses in wild birds (12), as well as for epidemiological investigations, being used for influenza genome sequencing and characterization(13,14), showing the modular versatility of the method.

In all, the power of molecular tools is indisputable, and they have been instrumental in shaping microbial ecology, providing a way to take a broader and more thorough look into many diverse ecological niches (1,4). The continued usage, evolution and improvement of sequencing methods and bioinformatic tools for analysis, such as SPONS and ViDiT-CACTUS, will undoubtedly lead to many novel insights into the microbial landscape.

7.2 The carnivorous sponge microbiome

The majority of the work in this thesis was aimed at describing and characterizing the similarities and differences between the microbial communities associated with two carnivorous sponge species, *Chondrocladia grandis* and *Cladorhiza oxeata*. Furthermore, the variability of these communities across two geographical regions and the possibility of host, species, and anatomical region specificity among associated bacteria was investigated.

7.2.1 Overall bacterial diversity and composition in carnivorous sponges

While *Chondrocladia grandis* and *Cladorhiza oxeata* reside in similar habitats, significant systematic differences between the bacterial communities of the two sponge species were consistently observed (Chapters 3, 4). On the most basal level, the largest difference observed was the overall biodiversity of the bacterial communities, which appear significantly richer in *C. grandis* when compared to *C. oxeata*. This difference is evident within the composition of their respective microbiome communities. In fact, *C. grandis* specimens often contain bacteria distributed across several families, with the Flavobacteriaceae being especially prominent, and other families, such as the Rhodobacteraceae, Halieaceae, Colwelliaceae, and Hyphomonadaceae also observed. In contrast, *C. oxeata* specimens displayed a microbiome community completely dominated by Flavobacteriaceae. This single clade dominance could point towards this species being a potential "low microbial abundance" sponge (15). Overall, the bacterial community members observed are not unlike those observed in filter-feeding sponges (16), except for the high abundance of Flavobacteriaceae, which appears to be a trait constrained to deep-water carnivorous sponges.

As *C. grandis* and *C. oxeata* are part of different genera, differences in microbiome composition were not unexpected. Perhaps more interesting was the discovery of large scale changes in bacterial communities among anatomical regions within each *C. grandis* specimen. The anatomy of *C. grandis* is quite unique; the sponge seems to burrow into the benthic sediment with a unique structure (root), while above the sediment a single central axis diverges in numerous side branches (stems) terminating in inflated spheres. These unique anatomical characteristics are presumed to have specific biological functions, with the spheres specifically being implicated in prey capture. Comparing the bacterial communities found in distinct anatomical regions revealed that the root tissues have a uniquely rich and diverse bacterial composition, setting this region apart from the remainder of the sponge body. We hypothesized that this increase in community richness and diversity indicates a particular functional importance of the root tissue in *C. grandis*, perhaps constituting a "main body" of the host, working in tandem with other auxiliary, function-oriented regions, in a manner similar to what has been reported for *Lycopodina hypogea* (17).

In contrast, *C. oxeata* lacks a root and presents a "bush-like" morphology, with little variety in anatomical regions. Within the investigated sponges, bacterial community richness and composition appeared random, with no clear pattern associating these changes to the type of sample or other available meta-data. This is in stark contrast to *C. grandis*, where highly conserved communities were seen across anatomical regions.

Another significant difference was observed in the stability of the microbial community between host individuals of *C. grandis* and *C. oxeata*. The bacterial communities in *C. grandis* seem to be robust when comparing host to host, as samples collected in the Canadian Arctic are highly similar to those collected in the Gulf of Maine. This suggests that the microbiome within *C. grandis* is highly curated, and an affinity for certain bacterial taxa is present, similar to the preferential retention seen in filter-feeding sponges (18). This retention appears to be highly specific, as oligotyping of common bacterial community members showed that *C. grandis* seems to select bacteria to a strain specific level. In *C. oxeata,* host to host variability was high, and the microbial community composition did not appear conserved, nor stable, as even replicates from anatomical regions within the same sponge showed high variability amongst each other.

7.2.2 Localized enrichment of potential bacterial symbionts

As the overall bacterial community composition analysis of *C. grandis* and *C. oxeata* revealed that the microbiome is not homogeneous and intra-specimen variations within the structure of the microbiome exist, the potential for enrichment of specific bacterial taxa (or biomarkers) was investigated (Chapter 3, Chapter 4). Specific associations between biomarkers and anatomical regions can potentially signify symbiotic partnerships, which in turn can be useful in extrapolating the potential function of each anatomical region.

While no biomarkers were found for *C. oxeata*, over 40 taxa with significant affinity for specific anatomical regions were consistently detected within *C. grandis*

samples. The majority of these enriched bacteria were detected within the root structure, further supporting the hypothesis that this could represent an area of particular importance within *C. grandis*, distinct from the rest of the sponge. Several biomarkers were also detected in the trapping spheres and their supporting stems. Unfortunately, the ecological role and importance of many of the identified biomarkers are not fully understood and inferring functional importance was challenging and, admittedly, speculative. Of particular note, however, was the presence of two chitin degradation associated bacteria (*Colwellia* and *Reichenbachiella*) within the trapping spheres of *C. grandis*, which is highly suggestive of a bacterial role in prey digestion (19,20). Mesoplanktonic organisms caught by carnivorous sponges likely have a chitin-rich exoskeleton that inhibits phagocytosis, and the affinity of chitin-degradation-associated bacterial genera could point towards a bacterially-mediated chitin hydrolysis that facilitates prey breakdown and subsequent phagocytosis processes. In addition, a potential for bacteria-mediated heterotrophic degradation was observed within sediment associated root tissues, to which the hydrocarbon-degradation linked *Robiginitomaculum* genus showed particular affinity (21). The carnivorous sponge *Cladorhiza methanophila* has been shown to perform bacterial symbiont mediated chemolithoautotrophy (22), utilizing the affinity for environmental hydrocarbon of chemosynthetic symbionts as an additional source of nutrients, and the same relationship could exist between *Robignitomaculum* and *C. grandis*.

7.2.3 Metagenomic investigation of C. grandis

ViDiT based metagenome sequencing of *C. grandis* (Chapters 5 and 6) showed the benefit of performing additional analyses beyond 16S rRNA gene sequencing, as many

additional characteristics of the microbiome were revealed through random shotgun sequencing. Remarkably, a hidden biodiversity of the Actinobacteria phylum was uncovered, which was not detected in the performed marker-gene analysis. Actinobacteria are thought to be true symbionts in several filter-feeding sponges and are well known for their ability to produce a wide variety of bioactive secondary metabolites, where these metabolites play important roles in maintaining sponge health (23). The successful characterization of this phylum in carnivorous sponges is a first step to assess if a similar symbiotic relationship exists in these organisms. Similarly, previously undetected methanotrophic and obligate hydrocarbonoclastic taxa were observed, supporting the hypothesis that *C. grandis* may be partly reliant on microbial mediated chemolithoautotrophy and degradation of environmental hydrocarbons.

Furthermore, a distinct assemblage of archaeal taxa from the phyla Thaumarchaeota and Euryarchaeota was found to be associated with *C. grandis*. The majority of detected Archaea could not be classified beyond the phylum level, suggesting the presence of distant, yet undescribed taxa. However, the observed presence of a diverse set of ammonia-oxidizing archaea from the Nitrosopumilales order could potentially indicate them as important symbionts, contributing to host health by metabolizing nitrogenous waste, secreted by most sponges as a metabolic waste product (24).

While little is known on the viral diversity in sponges, the virome of *C. grandis* was similar to what has been described in the limited work performed on filter-feeding sponges. The majority of detected viruses constituted bacteriophages (including the

Podoviridae, *Siphoviridae* and *Myoviridae*), and could play a role in influencing the composition of the *C. grandis* associated microbial population directly through mortality, in addition to contributing to the exchange of genomic information through horizontal gene transfer (25,26).

Lastly, the prediction of functional capabilities through the detection of protein coding genes suggests that the microbiome of *C. grandis* contains a wide repertoire of metabolic capabilities. Pathways specific for processing elemental substrates were predicted, including nitrogen, sulfur and methane metabolism, in line with the metabolic predictions based on the taxonomic composition of the microbiome. In addition, a large number of other degradation and metabolic pathways were found, such as pathways specific for a diversity of lipids as well as xenobiotics, including aromatic organic compounds, hydrocarbons and steroids.

7.2.4 Overall conclusions

Overall, the analysis performed of *C. grandis* show that carnivorous sponges, like their filter-feeding contemporaries, can be host to a diverse and extensive microbiome, which is seemingly intrinsically linked to many metabolic functions, including the facilitation of carnivory and chemolithoautotrophy. However, the microbiome heterogeneity observed and absence of apparent symbiotic potential in *C. oxeata* suggest that not all carnivorous sponges may be reliant on their microbiome to the same extent, and perhaps use innate processes to enable carnivory.

Microbiome analysis further revealed that in carnivorous sponges like *C. grandis*, unique micro-niches exist across the sponge body, in which specific microbial communities are found, which are most likely hot-spots for symbiotic processes to occur.

A prominent role for the carnivorous sponges in the deep-sea habitat is probable as carnivory, chemolithoautotrophy, and the microbial associates with a wide repertoire of metabolic capabilities presumed to drive these processes, are likely to influence nutrient fluxes and benthic–pelagic coupling in the immediate environment. In addition to the microbiological aspects, the common occurrence of carnivorous sponges in a wide range of habitats suggest a large-scale functional importance that has yet to be fully elucidated.

7.3 Future work and perspectives

While the work presented here has revealed significant insights into the composition, structure and stability of the sponge microbiome, fundamental questions remain, and results presented here have in turn spawned new hypotheses that require follow-up studies.

Overall, the Cladorhizidae family is extremely diverse, with roughly 300 species described and, while the work presented here performs in-depth analyses, it did not have an extensive taxonomic reach across the Cladorhizidae. Our understanding of the carnivorous sponge microbiome structure, diversity and stability could significantly be improved by the sampling and microbiome characterization of additional Cladorhizidae species, preferably collected globally in different habitats. Such work could help facilitate studies similar to a recent excellent investigation in which diversity, structure and

convergent evolution of the global sponge microbiome was evaluated (16). Sadly, this study did not include any carnivorous sponge taxa, however, combining such a study with a "global carnivorous sponge microbiome" dataset would offer a unique opportunity to better understand the relation of carnivorous sponge microbiomes to those found in filter-feeder sponges.

While the metagenomic study performed in this thesis allowed us to glance into the true biodiversity residing within *C. grandis*, the overall recovery of microbial fragments was low, and no complete metagenomes were recovered. Further studies, perhaps combining shotgun metagenomics with the newly emerging long-read sequencing technologies, such as the Oxford Nanopore MinION, should be performed to improve this coverage. This will undoubtably lead to the detection of additional (rare) species, while more complete genomes could improve taxonomic classification and functional prediction for distant, undiscovered microbial associates.

On a finer scale, the potential symbiotic partners described in this thesis should be evaluated in more depth to confirm their suggested roles. The application of transcriptomics, in which messenger / protein-coding RNA molecules are sequenced as opposed to DNA, could elucidate which genes are actively being transcribed and thus give a better understanding on the functional repertoire of sponge associated microbial communities. Such data could irrefutably show specific metabolic activities, such as the suggested chitinase activity in *C. grandis*, especially if tissue samples from sponges that

recently fed, perhaps through the visual identification of prey inclusions, could be compared to those that do not show recent feeding.

Additionally, cytogenic studies, such as electron microscopy and visualizing suspected symbionts in sponge tissue using fluorescent in-situ hybridization, could aid in creating a better understanding on how symbionts aggregate in different anatomical regions, if these symbionts appear intra or inter-cellularly and show potential interactions with captured prey. Such approaches would be extremely useful if the host sponge could be maintained in an *in vitro* setting so feeding trials could be conducted, as previously performed in *Lycopodina hypogea* (17). However, the expertise and equipment required to collect, transfer and maintain a deep-sea sponge in a condition mimicking its natural environment, will almost certainly make such studies extremely challenging.

7.4 References

1. Boughner LA, Singh P. Microbial Ecology: Where are we now? *Postdoc J* (2016) **4**:3–17. doi:10.14304/SURYA.JPR.V4N11.2

2. Lightbody G, Haberland V, Browne F, Taggart L, Zheng H, Parkes E, Blayney JK. Review of applications of high-throughput sequencing in personalized medicine: barriers and facilitators of future progress in research and clinical application. *Brief Bioinform* (2018) doi:10.1093/bib/bby051

3. Reuter JA, Spacek D, Snyder MP. High-throughput sequencing technologies. *Mol Cell* (2015) **58**:586–597. doi:10.1016/j.molcel.2015.05.004

4. Escobar-Zepeda A, Vera-Ponce de León A, Sanchez-Flores A. The road to metagenomics: from microbiology to DNA sequencing technologies and bioinformatics. *Front Genet* (2015) **6**: doi:10.3389/fgene.2015.00348

5. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: and this is not optional. *Front Microbiol* (2017) **8**: doi:10.3389/fmicb.2017.02224

6. Verhoeven JTP, Salvo F, Hamoutene D, Dufour SC. Bacterial community composition of flocculent matter under a salmonid aquaculture site in Newfoundland, Canada. *Aquaculture Environment Interactions* (2016) **8**:637–646. doi:10.3354/aei00204

7. Verhoeven JTP, Salvo F, Knight R, Hamoutene D, Dufour S. Temporal bacterial surveillance of salmon aquaculture sites indicates a long lasting benthic impact with minimal recovery. *Front Microbiol* (2018) **9**: doi:10.3389/fmicb.2018.03054

8. Weiler BA, Verhoeven JTP, Dufour SC. Bacterial Communities in Tissues and Surficial Mucus of the Cold-Water Coral *Paragorgia arborea*. *Front Mar Sci* (2018) **5**: doi:10.3389/fmars.2018.00378

9. Brosnan ME, Clow KA, Pongnopparat T, Verhoeven JTP, Canuti M, Brosnan JT. An Apple a Day: the Role of the Intestinal Microbiota in Formate Metabolism. (2018)

10. Munro HJ. Circulation of pathogenic spirochetes in the genus *Borrelia* within ticks and seabirds in breeding colonies of Newfoundland and Labrador. (2018)

11. Verhoeven JTP, Canuti M, Munro HJ, Dufour SC, Lang AS. ViDiT-CACTUS: an inexpensive and versatile library preparation and sequence analysis method for virus discovery and other microbiology applications. *Can J Microbiol* (2018)1–13. doi:10.1139/cjm-2018-0097

12. Canuti M, Munro HJ, Robertson G, Kroyer A, Roul S, Ojkic D, Whitney HG, Lang AS. The discovery of novel papillomaviruses in wild birds sheds light over avian papillomavirus ecology and evolution. *Front Microbiol (submitted)* (2018)

13. Roul S. Assessing the impact of low-pathogenicity avian influenza virus on the health of American Black Ducks (Anas rubripes). (2018)

14. Benkaroun J. Genetic and evolutionary dynamics of avian influenza A virus in wild birds. (2018)

15. Gloeckner V, Wehrl M, Moitinho-Silva L, Gernert C, Schupp P, Pawlik JR, Lindquist NL, Erpenbeck D, Wörheide G, Hentschel U. The HMA-LMA dichotomy revisited: an electron microscopical survey of 56 sponge species. *The Biological Bulletin* (2014) **227**:78–88. doi:10.1086/BBLv227n1p78

16. Thomas T, Moitinho-Silva L, Lurgi M, Björk JR, Easson C, Astudillo-García C, Olson JB, Erwin PM, López-Legentil S, Luter H, et al. Diversity, structure and convergent evolution of the global sponge microbiome. *Nature Communications* (2016) **7**:11870.

17. Martinand-Mari C, Vacelet J, Nickel M, Wörheide G, Mangeat P, Baghdiguian S. Cell death and renewal during prey capture and digestion in the carnivorous sponge *Asbestopluma hypogea* (Porifera: Poecilosclerida). *J Exp Biol* (2012) **215**:3937–3943. doi:10.1242/jeb.072371

18. Glasl B, Smith CE, Bourne DG, Webster NS. Exploring the diversity-stability paradigm using sponge microbial communities. *Scientific Reports* (2018) **8**:8425. doi:10.1038/s41598-018-26641-9

19. Wietz M, Wemheuer B, Simon H, Giebel H-A, Seibt MA, Daniel R, Brinkhoff T, Simon M. Bacterial community dynamics during polysaccharide degradation at contrasting sites in the Southern and Atlantic Oceans. *Environ Microbiol* (2015) **17**:3822–3831. doi:10.1111/1462-2920.12842

20. Bowman J, Gosink JJ, McCammon SA, Lewis TE, Nichols DS, Nichols PD, Skerratt JH, Staley JT, McMeekin TA. *Colwellia demingiae* sp. nov., *Colwellia hornerae* sp. nov., *Colwellia rossensis* sp. nov. and *Colwellia psychrotropica* sp. nov.: psychrophilic Antarctic species with the ability to synthesize docosahexaenoic acid (22:ω63). *International Journal of Systematic and Evolutionary Microbiology* (1998) **48**:1171–1180. doi:10.1099/00207713-48-4-1171

21. Guibert LM, Loviso CL, Marcos MS, Commendatore MG, Dionisi HM, Lozada M. Alkane biodegradation genes from chronically polluted subantarctic coastal sediments and their shifts in response to oil exposure. *Microb Ecol* (2012) **64**:605–616. doi:10.1007/s00248-012-0051-9

22. Hestetun JT, Dahle H, Jørgensen SL, Olsen BR, Rapp HT. The microbiome and occurrence of methanotrophy in carnivorous Sponges. *Front Microbiol* (2016) **7**: doi:10.3389/fmicb.2016.01781

23. Pawlik JR. The chemical ecology of sponges on Caribbean reefs: natural products shape natural systems. *BioScience* (2011) **61**:888–898. doi:10.1525/bio.2011.61.11.8

24. Zhang F, Pita L, Erwin PM, Abaid S, López-Legentil S, Hill RT. Symbiotic archaea in marine sponges show stability and host specificity in community structure and ammonia oxidation functionality. *FEMS Microbiol Ecol* (2014) **90**:699–707. doi:10.1111/1574-6941.12427

25. Laffy PW, Wood-Charlson EM, Turaev D, Jutz S, Pascelli C, Botté ES, Bell SC, Peirce TE, Weynberg KD, van Oppen MJH, et al. Reef invertebrate viromics: diversity, host specificity and functional capacity. *Environmental Microbiology* (2018) **20**:2125–2141. doi:10.1111/1462-2920.14110

26. Bryson SJ, Thurber AR, Correa AMS, Orphan VJ, Vega Thurber R. A novel sister clade to the enterobacteria microviruses (family Microviridae) identified in methane seep sediments. *Environ Microbiol* (2015) **17**:3708–3721. doi:10.1111/1462-2920.12758