UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE QUÍMICA – DEPARTAMENTO DE BIOQUÍMICA
PROGRAMA INTERUNIDADES DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA

ANA CAROLINA DOS SANTOS SOARES

Sponge Microbiomes from Great Amazon Reef System

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

São Paulo
2023

ANA CAROLINA DOS SANTOS SOARES

Sponge Microbiomes from Great Amazon Reef System

Final Version

Ph. D. Thesis presented to the Interunit Bioinformatics Graduate Program at the Institute of Chemistry in University of São Paulo, Brazil to obtain the degree of Doctor of Science.

Concentration area: Bioinformatics

Supervisor: Prof. Dr. João Carlos Setubal (USP)

Co-supervisor: Prof. Dr. Fabiano Lopes Thompson (UFRJ)

São Paulo

2023

Ficha catalográfica elaborada com dados inseridos pelo(a) autor(a)
Biblioteca Carlos Benjamin de Lyra
Instituto de Matemática e Estatística
Universidade de São Paulo

Soares, Ana Carolina dos Santos
Sponge Microbiomes from Great Amazon Reef System / Ana Carolina dos Santos Soares; orientador, João Carlos Setubal; coorientador, Fabiano Lopes Thompson. - São Paulo, 2023.
82 p.: il.

Tese (Doutorado) - Programa Interunidades de Pós-Graduação em Bioinformática / Instituto de Matemática e Estatística / Universidade de São Paulo.
Bibliografia
Versão corrigida

1. Microbioma. 2. Metagenômica. 3. Esponjas marinhas. 4. Diversidade microbiana. I. Setubal, João Carlos. II. Título.

Bibliotecárias do Serviço de Informação e Biblioteca Carlos Benjamin de Lyra do IME-USP, responsáveis pela estrutura de catalogação da publicação de acordo com a AACR2: Maria Lúcia Ribeiro CRB-8/2766; Stela do Nascimento Madruga CRB 8/7534.

Dedico a minha mãe Maria Estela, ao meu pai Pedro Soares e ao meu irmão Pedro Alexandre com amor e gratidão.

AGRADECIMENTOS

Para meus pais Maria Estela e Pedro pelo amor incondicional e especialmente por sempre me apoiarem na minha trajetória profissional, dando todo o suporte emocional e financeiro.

Para meu irmão Pedro Alexandre pela parceria e momentos de descontração.

Para meu orientador Professor Dr. João Carlos Setubal pela oportunidade de integrar seu grupo de pesquisa no Laboratório de Bioinformática durante os últimos 6 anos e por proporcionar um ambiente de muito aprendizado e crescimento profissional, me abrindo diversas portas.

Para meu coorientador Professor Dr. Fabiano Lopes Thompson por todo o suporte fornecido desde a graduação e iniciação científica no laboratório de Microbiologia da UFRJ, e por me proporcionar diversas oportunidades que colaboraram na formação da profissional que sou hoje.

Para a Professora Cristiane Thompson do Laboratório de Microbiologia por todo apoio na minha trajetória acadêmica, e principalmente por sempre trazer palavras doces e leves nos momentos difíceis.

Para meus colegas do Laboratório de Bioinformática, especialmente Raquel, Deyvid, Jhonatas, Livia, Meline, Carlos, Suzana, Lucas, Fernando e Bruno pela amizade e parceria nas disciplinas e momentos difíceis do doutorado.

Para meus colegas do Laboratório de Microbiologia, especialmente Gustavo Gregoracci, Adriana, Gisele, Gizele, Graciela, Cintia Rua, Cynthia Silveira, Luciana Reis, Felipe Coutinho, Bruno, Louise, Pedro Meirelles, Diogo, Tainá, Maria Soares, Nelson, Cecília, Juline, Ana Paula, Mayanne, Ana Carolina Gomes, Tatiane, Itamar e Angélica pela parceria e amizade.

Para minhas amigas Marcele, Isabela, Jessica e Ana Carolina Carvalho pela amizade de muitos anos.

Para Andreza, Melissa, Rayanne, Camila, Gabriela, Francine, Rafaela e Arielly por terem feito a minha temporada em São Paulo muito mais leve, divertida e menos solitária.

A CAPES por me fornecer uma bolsa de estudos para executar o projeto do meu doutorado.

ABSTRACT

The Great Amazon Reef System is located on the Brazilian continental shelf underneath the Amazon River plume. This environment has a benthic community mainly formed by sponges and calcareous algae. Sponges are the main members of the marine benthic community. Through filtration of large amounts of water, sponges are important links between the water column and benthic compartments. They are able to influence nutrient cycling by removal, processing and release of the filtered materials. Sponges are known to harbor a large abundance of microbes. The microbial abundance can be 3 to 4 orders of magnitude higher than the density in the surrounding water. Most of microorganisms associated with sponges are species specific. Among the functions of these microbial communities are those related to the symbiotic lifestyle, such as nutrient supply, degradation of complex carbohydrates, sponge skeletal stabilization, waste processing and production of secondary metabolites involved in sponge defense. Using a metagenomics approach, the main goal of this study is to evaluate the microbial diversity associated with sponges from the Amazon Reef System in terms of taxonomy, genomic and functional metabolism. A total of 236131258 good quality paired-end sequences were generated for 37 metagenomes from 20 sponge species. Taxonomic assignments showed a higher contribution of Actinobacteria, Euryarchaeota, Chloroflexi and Planctomycetes, besides the classes Betaproteobacteria and Deltaproteobacteria. This taxonomic profile was distinct between the sponges and surrounding water, showing that sponges harbour their own unique microbial community. From the metagenome data we obtained 115 metagenome-assembled genomes, revealing new species of bacteria and archaea. The sponge microbiome also showed an enrichment of genes related to carbon, nitrogen and sulfur metabolism. These results bring new knowledge about marine sponge microbial symbionts.

Keywords: marine sponge, microbiome, metagenomics, microbial diversity.

RESUMO

O Sistema Recifal Amazônico está localizado na Plataforma Continental Amazônica, embaixo da pluma do rio Amazonas. Este ambiente é formado principalmente por esponjas e algas calcárias. Esponjas são os principais membros da comunidade bentônica marinha. Através da filtração de grandes quantidades de água, exercem um papel importante na conexão entre o compartimento pelágico e bêntico. Além disso, são capazes de influenciar na ciclagem de nutrientes pela remoção, processamento e liberação do filtrado. Esponjas também são conhecidas por abrigar uma grande quantidade de microrganismos, no qual a abundância pode ser até 3 a 4 ordens de grandeza maior que a da água do entorno. Muitas funções estão relacionadas a estes simbiontes, como suprimento de nutrientes, estabilização do esqueleto da esponja, síntese e processamento de metabólitos secundários, que estão envolvidos na defesa das esponjas. Através de uma abordagem “ômicas”, o principal objetivo deste estudo é analisar o microbioma de esponjas do Sistema Recifal Amazônico, em termos taxonômicos, genômicos, funcionais e comparativos. Um total de 236.131.258 sequências de boa qualidade foram geradas para 37 metagenomas de 20 espécies de esponjas. A análise taxonômica mostrou que a comunidade microbiana apresenta uma alta abundância dos filos Actinobacteria, Euryarchaeota, Chloroflexi e Planctomycetes, além das classes Betaproteobacteria and Deltaproteobacteria. O perfil do microbioma das esponjas também foi distinto do microbioma da água do entorno, demonstrando que esponjas possuem uma microbiota própria, corroborando a literatura. Além disso, foram recuperados 115 genomas dos metagenomas, revelando novas espécies de bactérias e arquéias. A análise funcional indicou um enriquecimento dos genes relacionados aos metabolismos de carbono, nitrogênio e enxofre. Estes resultados revelam um novo conhecimento sobre os microrganismos simbiontes de esponjas.

Palavras-chave: esponjas marinhas, microbioma, metagenômica, diversidade microbiana.

Table of Contents

Figures

Tables

1 Introduction

1.1 Marine Sponges and its ecological functions

Marine sponges are the main members of marine benthic communities throughout the world. These animals have been around for at least 600 million years and are among the first multicellular organisms (Metazoa) on Earth (YIN et al., 2015). In reef ecosystems play many functional related to bioerosion, reef creation and manutention, promoting substrate consolidation, stabilization and regeneration (BELL, 2008; YIN et al., 2015). This impacts on substrate promotes a habitat formation for many aquatic animals, such as fishes and crustaceans. Through a complex filtering system, sponges can filter up to 24.000 liters of water in one day (VOGEL, 1977) (**Figure 1**). Thus they establish an important connection between the water column and benthic compartments (GOEIJ et al., 2013; SILVEIRA et al., 2015), influencing nutrient cycling by removal, processing and release of the filtered materials.

Through filtration, these animals remove organic matter, particles and nutrients from the water column (e.g. carbon, nitrogen and phosphorus), promoting significant changes in the pelagic environment (LESSER et al., 2006; COPPARI et al., 2016, GOEIJ et al., 2008). As a result of the respiration of organic matter, dissolved inorganic compounds are released into the water column, such as nitrate, nitrite, ammonia, carbon dioxide and phosphate (HADAS; SHPIGEL; ILAN, 2009; MALDONADO; RIBES; VAN DUYL, 2012; ZHANG et al., 2015). The release of excreta from the sponge, resulting from metabolic processes, also contributes to the cycling of nutrients in the pelagic environment, since the material can be rapidly consumed by detritivores (GOEIJ et al., 2013).

Sponges benefit from increased concentrations of particulate and dissolved organic carbon in water by consuming organic matter (GOEIJ et al., 2008; HADAS et al., 2006; HADAS et al., 2009) and enrich the surrounding environment with metabolic organic matter.

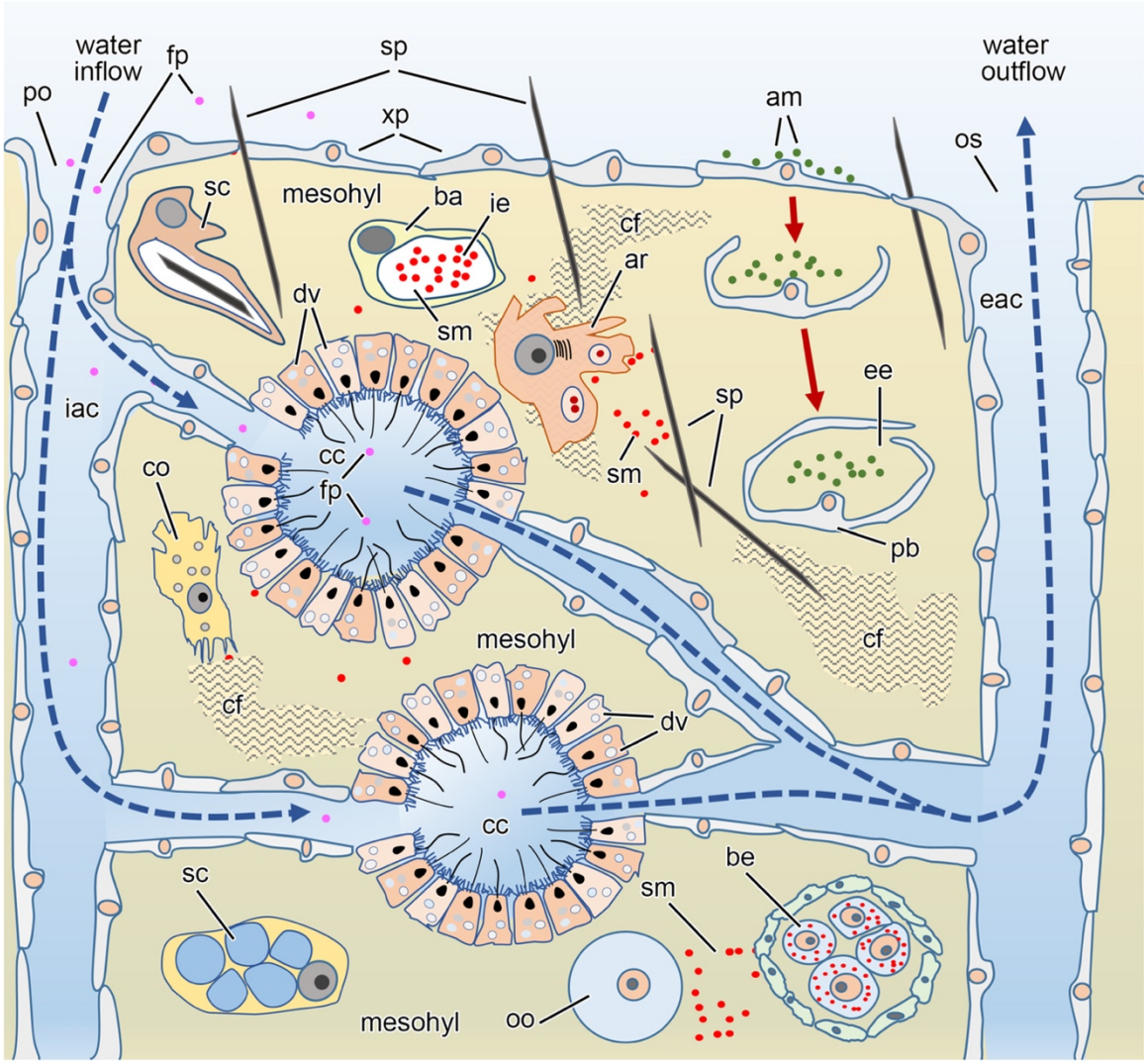

Figure 1 Schematic diagram summarizing the essential cytology known for demosponges, as reported previously. The external epithelium consists of pavement-like cells (exopinacocytes, xp). Figure shows the inhalant aquiferous canals (iac) through the pores (po) at the sponge "skin" carrying particles (fp) in suspension to the choanocyte chambers (cc). Bacteriocytes (ba) are cells that host symbiotic microbes (sm) in the intracellular environment (ie) of a large intracytoplasmic vesicle. Digesting vesicles (dv), exhalant aquiferous canals (eac), oscule (os), collagen fibrils (cf), pocket bacteriocyte (pb), free-living microbes (am), oocytes (oo) and brooded embryos (be). Figure from Carrier, et al BMC Biol (2022): https://doi.org/10.1186/s12915-022-01291-6.

1.2 Marine Sponge Microbiomes

Sponges are known to harbor a high abundance of microorganisms (WEBSTER; TAYLOR, 2012). These microorganisms can represent up to approximately 35% of the total biomass of a sponge and a density of 10^9 microbial cells per centimeter of tissue, which is 3 to 4 orders of magnitude more than the density in the surrounding water (WEBSTER; HILL, 2001). Most symbiotic microorganisms are found in the mesohyl, an extracellular matrix that makes up

most of the sponge body, but some are intracellular symbionts (WEBSTER; TAYLOR, 2012).

Based on the abundance and diversity of microorganisms, sponges are commonly divided into two functional groups: Low Microbial Abundance (LMA) and High Microbial Abundance (HMA) (GLOECKNER et al., 2014; MOITINHO-SILVA et al., 2017b) **(Figure 2)**. LMA sponge species are generally small and delicate, have high filtration rates, well-irrigated tissues, and a sparse and diverse microbial community. In contrast, the HMA sponges are often large, massive, and have a dense tissue with low filtration rates, and an abundant, diverse and specific microbiota, which is very different from the microbiota of the water column (GLOECKNER et al., 2014).

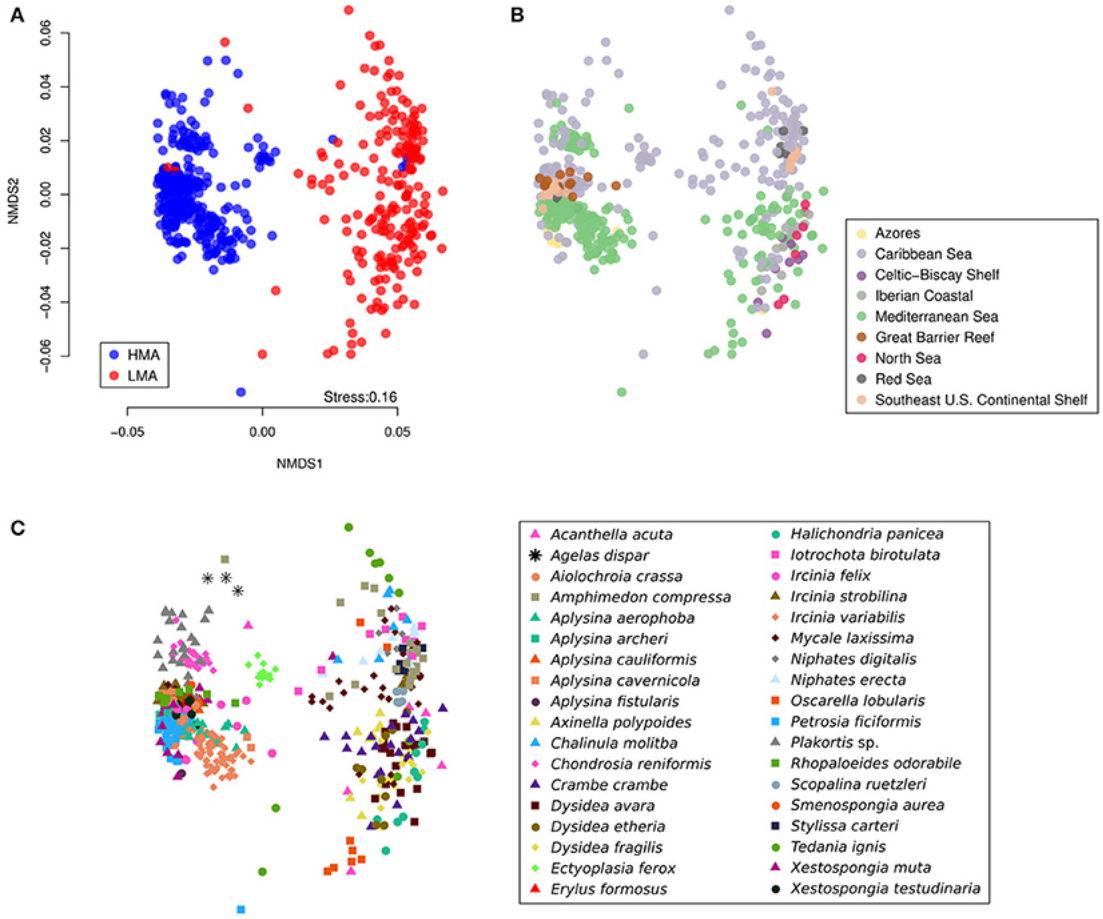

Figure 2 Structure of microbial communities associated with 575 HMA and LMA sponge samples from Sponge Microbiome Project dataset. Non-metric Multidimensional Scaling plots represent the same analysis, which sample symbols and colors stand for (A) HMA-LMA status, (B) geographic region, and (C) host identity. Figure from Moitinho-Silva et al. 2017: https://doi.org/10.3389/fmicb.2017.00752.

The two groups have different feeding strategies (WEISZ; LINDQUIST; MARTENS, 2008). HMA sponges, whose filtration system is less developed,

supply their nutritional needs preferentially by ingesting organic matter produced by the associated microbial community, through autotrophic metabolisms such as photosynthesis and chemosynthesis (RADAX et al., 2012; SCHLÄPPY et al., 2010; WEISZ; LINDQUIST; MARTENS, 2008). The LMA type acquire most of the nutrients through the process of filtration and ingestion of organic matter from the water column, depending less on the microbiota (WEISZ et al., 2007). Both groups depend on both filtration and the microbiota for their nutrition but invest in different strategies.

Most microorganisms associated with sponges are specific to a particular host species (REVEILLAUD et al., 2014; SCHMITT et al., 2012). This pattern is independent of the geographic region and physicochemical conditions in which the species are found (HENTSCHEL et al., 2002). Analyzes of sponge microbiome diversity also revealed that part of the microbiota is shared between different species, even though they are phylogenetically distant and found in different locations (SIMISTER et al., 2012; TAYLOR et al., 2007). Studies using 16S ribosomal RNA gene sequencing revealed that certain microorganisms occurred exclusively in sponges (TAYLOR et al., 2007), but more recent studies, with greater sequencing depth, suggest that these taxa are actually rare or not detectable in other environments (TAYLOR et al., 2013), as they occur in very low abundance.

Although sponge species have a taxonomically distinct microbial community, it has been demonstrated that there is a sharing of functional categories between different species (FAN et al., 2012; RIBES et al., 2012). The main functions are important for the establishment of symbiosis and are related to the adaptation of microorganisms to the environment in the host (THOMAS et al., 2016). For species that occupy similar niches, the symbiotic microorganisms converge in relation to the metabolic profile. It was observed distinct symbiotic microorganisms from different sponges, capable of carrying out the processes of denitrification and nitrification, use different analogous enzymes to perform equivalent functions (FAN et al., 2012; RIBES et al., 2012). Thus, it is suggested that the association of symbionts with sponges evolved independently in the phylum Porifera and that converging forces resulted in a similar functional profile (THOMAS et al., 2016).

1.3 Sponge Symbionts Functional Roles

Sponge symbionts carry out a wide range of functions, such as supplying nutrients, stabilizing the sponge skeleton, processing sponge excreta (TAYLOR et al., 2007; WEBSTER; TAYLOR, 2012) and producing secondary metabolites that are involved in sponge defense (INDRANINGRAT; SMIDT; SIPKEMA, 2016; THOMAS; KAVLEKAR; LOKABHARATHI, 2010). On the other hand, the symbiotic microorganisms are benefited by the supply of nutrients, through the filter feeding mechanism of the sponge and nitrogen from the ammonia excreted by the sponge itself, as a final product of metabolism.

Symbiotic microorganisms can perform various metabolic processes, such as photosynthesis, nitrification, denitrification, anammox (anaerobic ammonia oxidation), sulfur oxidation and reduction, and nitrogen fixation (HOFFMANN et al., 2009; MOHAMED et al., 2010; RIBES et al., 2015; WEISZ et al., 2007), providing nutrients such as carbon and nitrogen to the sponge. These processes allow the reuse and recirculation of nutrients within the host-microbiota complex since denitrification and anammox remove inorganic nitrogen from the environment. Sponges that harbor microorganisms capable of carrying out these processes can function as a nitrogen sink in places where they are abundant (RADAX et al., 2012). Additionally, heterotrophic bacteria assist in the degradation of complex carbohydrates removed from the water column by filtration. Certain associated microorganisms revealed a potential for the synthesis of essential vitamins for the sponge, such as vitamin B1 (thiamine), B2 (riboflavin), B7 (biotin) and B12 (cobalamin), which must be obtained through the diet (HENTSCHEL et al., 2012).

Another contribution of symbiotic microorganisms is in relation to host defense. Many studies have evaluated the ability to synthesize secondary metabolites that can protect sponges from predators and epibionts (MEHBUB et al., 2014; THOMAS; KAVLEKAR; LOKABHARATHI, 2010). This chemical defense has contributed to the evolutionary and ecological success of sponges.

Due to the sponge filtration mechanism, the symbiotic microorganisms are exposed to viral infection, given that the abundance of viruses in seawater is approximately 10^7 particles/ml (SUTTLE, 2007). Thus, the sponge-associated

microbial community has mechanisms to prevent viral lysis and exogenous DNA insertion, such as restriction modification system (R-M), toxin-antitoxin system (T-A) and the CRISPRs system (clustered regularly interspaced short palindromic repeats) (FAN et al., 2012). These mechanisms of resistance to exogenous DNA insertion are enriched in sponge symbiotic microorganisms (HORN et al., 2016; THOMAS et al., 2010), playing a key role in maintaining the hologenome (WEBSTER; THOMAS, 2016).

1.4 Amazon River Plume

The Amazon River is the largest river in terms of water discharge on the planet. Annually, 63 x 10^{11} m^3 of fresh water are discharged into the Tropical North Atlantic Ocean (DAGG et al., 2004). This volume corresponds to approximately 20% of all global freshwater discharge, promoting the formation of an immense plume that extends towards the north of the Atlantic Ocean (COLES et al., 2013). Between January and April, the Amazon River plume extends towards the northwest Atlantic Ocean, closely adjacent to the continental shelf. From April to July, the river maximum discharge period, the plume reaches its greatest extent, reaching the Caribbean Sea. During the retroflexion of the North Brazilian current, between the months of August and December, the plume flows towards the east of the Atlantic Ocean (MOLLERI; DE M. NOVO; KAMPEL, 2010). The Amazon River plume is mainly characterized by a surface layer of water with salinity below average for seawater (~35 psu) and a higher concentration of dissolved oxygen. In a vertical section of the water column up to a depth of at least 60 meters, three distinct layers are observed: 1. Plume (can reach up to approximately 25 meters in depth) promoting stratification in the water column, due to the presence of suspended material, which drastically reduces the incidence of light and the concentration of oxygen in the layer below the plume; 2. Sub-plume layer with low light levels, dissolved oxygen concentration < 3 mg/L, and predominance of chemosynthesis, and 3. Benthic compartment (e.g. marine sponges).

The Amazon River plume exports to the ocean, on average, 11 to 13 x 10^8 tons of sediment per year and 19.1 x 10^6 tons of dissolved organic carbon (DOC) per year (KINEKE et al., 1996), classifying the Amazon River as the river that has the largest DOC release into the ocean in the world (DAGG et al., 2004). In addition

to the high concentration of sediment and DOC, the plume is also rich in silicate, and several nutrients, such as dissolved inorganic nitrogen (DIN: ammonia, nitrate and nitrite) and inorganic phosphorus.

1.5 Amazon River Plume influence on organisms

Due to its physicochemical characteristics and high abundance of nutrients, the Amazon River plume strongly influences the biological communities of the Amazonian Continental Shelf region. However, most of the studies carried out so far have focused on understanding planktonic communities along the salinity gradient. It has been demonstrated that phytoplankton biomass and primary production is higher in regions where the nutrient concentration remains high, the plume salinity is intermediate and the sediment concentration lower, allowing the penetration of light into the water column (GOES et al., 2014; SUBRAMANIAM et al., 2008). This same pattern was also observed for diazotrophic microorganisms, that is, bacteria and archaea capable of fixing nitrogen (GOEBEL et al., 2010; HILTON et al., 2015; SUBRAMANIAM et al., 2008).

Several studies have also evaluated the diversity and activity of total microbial communities in the plume in different salinity regions (SATINSKY et al., 2014, 2015). However, so far, there are not integrated studies involving biodiversity analyzes of the three layers (plume, sub-plume and benthos) of the new reef system (LEAL et al., 2017; MOURA et al., 2016).

1.6 Great Amazon Reef System

A recent study revealed the presence of a new Reef System, located at depths of up to approximately 220 m in the vicinity of the mouth of the Amazon River, under the river plume (**Figure 3**) (MOURA et al., 2016) (FRANCINI-FILHO et al., 2018). This reef called the Great Amazon Reef System (GARS) presents several peculiarities in relation to other reefs in the world. It occurs in regions of high turbidity and concentration of organic matter and low incidence of light. These conditions are unfavorable for reef formation. Until 2016, it was considered that the reefs would occur at most up to the Parcel de Manoel Luís in the north of the state of Maranhão, although there were studies pointing to the presence of reef fish

associated with sponge banks and vast covers of calcium carbonate on the platform (Projeto PIATAM).

The Great Amazon Reef System is approximately 56,000 km^2 and has three sectors differentiated mainly by the residence time of the Amazon River plume: North, Central and South (FRANCINI-FILHO et al., 2018; MOURA et al., 2016). The North sector is located close to the continental shelf break, so that at certain points the depth can reach approximately 220 meters. In this sector, the plume occurs throughout the year, so the incidence of light in the water column is reduced, characterizing this region as mesophotic and suboxic. As a result, an increase in the proportion of chemolithoautotrophic and anaerobic metabolisms in the subplume layer has been observed (FRANCINI-FILHO et al., 2018; MOURA et al., 2016). The Central and South sectors are shallower, with the depth varying from 50 m to 100 m in the Central, and in the South reaching a maximum of approximately 25 m. The occurrence of the plume is seasonal and intermittent, respectively, so in these sectors photosynthetic metabolism was more abundant when the plume was absent (FRANCINI-FILHO et al., 2018; MOURA et al., 2016).

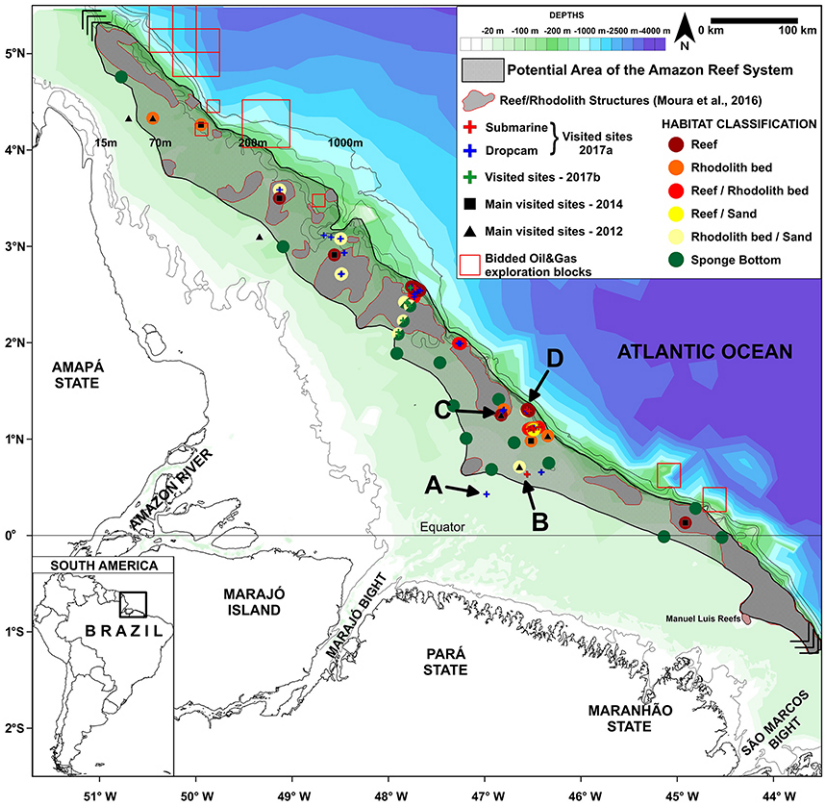

Figure 3 Map of Great Amazon Reef System (GARS). Figure from (FRANCINI-FILHO et al., 2018).

Regarding the composition of the benthic community, the study by Moura et al. 2016 showed that sponges and calcareous algae (**Figure 4**), mainly rhodoliths, are abundant in the three sectors of the GARS, however the greatest diversity was evaluated in the Central and Southern sectors. In the North sector, an environment with lower luminosity and dissolved oxygen, a high richness of filtering organisms was recorded, mainly sponges. In the Central and Southern sectors, where the plume was absent during the study, in addition to the high diversity of sponges, it was possible to observe a high abundance of light-dependent reef organisms, such as corals. The presence of reef-building corals was restricted to areas of the RAS where the plume was absent (Central and Southern), however sponges were observed from North to South of the reefs, within the mesophotic and photic zones (FRANCINI-FILHO et al., 2018; LEAL et al., 2017; MOURA et al., 2016).

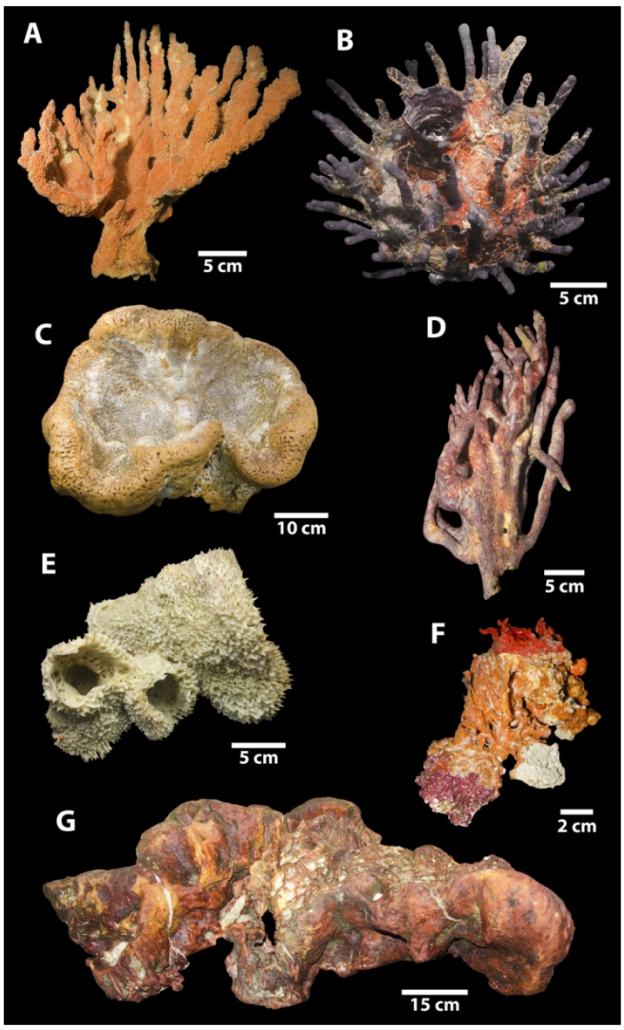

Figure 4 Representative species of sponges collected off the Amazon River mouth. Figure from Moura et al., 2016. (A) *Clathria nicoleae; (B) Coelocarteria bartschi; (C) Agelas clathrodes; (D) Aplysina fulva; (E) Callyspongia aculeata; (F) Monanchora arbuscula (attached to live and dead rhodoliths); (G) Geodia neptuni.*

1.7 Motivation

Studies on the taxonomic and microbial functional diversity of sponges are very limited in Brazil, being mostly concentrated in the southeastern region of our country (TRINDADE-SILVA et al., 2012). The microbial diversity associated with sponges has been intensely studied in the Caribbean, but the region of the mouth of the Amazon River is still unexplored, forming a gap of studies between the Caribbean and the reefs of Northern Maranhão. This thesis extends the first analyzes of microbial communities associated with the benthic components that form the GARS (MOURA et al., 2016), expanding the still scarce biological knowledge of the Brazilian Amazon Continental Shelf. This region is under strong

pressure from oil and gas companies, which are very interested in exploring for oil in the region, subjecting this distinct biome to high risks (FRANCINI-FILHO et al., 2018).

The high abundance of sponges in the Great Amazon Reef System suggests that these organisms play a relevant ecological role in this system. Most sponge-associated microorganisms are species-specific and of restricted distribution (endemic), as the microbiome is transmitted vertically by the parent sponge. The smallest part is acquired from the surrounding water. Thus, species of sponges from differentiated and not yet studied biomes, such as the Amazonian Reefs, may present different microbiomes from those previously analyzed, favoring the discovery of new microbial communities and new species of microorganisms. Among the 20 sponge species to be studied (**Table 1**), 4 are new species, *Arenosclera amazonensis, Arenosclera klausi, Coelocarteria alcoladoi sp. nov. and Coelocarteria amadoi sp. nov.* Therefore, the project can make important discoveries in the field of microbial diversity and ecology. This thesis is related to the research project “Sistema Recifal Amazônico” (projeto Novo Bioma Recifal) led by professor Dr. Fabiano L. Thompson from Federal University of Rio de Janeiro (UFRJ).

1.8 Aims

The main goal of this study is to characterize the sponge microbiomes structure composition and diversity from Great Amazon Reef System using a metagenomic approaches. In this thesis we evaluate these microbiomes based on reads analysis and metagenomes-assembled genomes.

In the chapter 1 we present the Introduction. In the chapter 2 the Materials and Methods applied for chapter 3 and 4. In the chapter 3 entitled “Metagenomic Profiling of Sponge Microbiomes from Great Amazon Reef System using read-based analysis” we study the composition, diversity and structure of marine sponge microbiomes through shotgun metagenomics using read-based analysis. In addition, we compared the microbial communities among the different sponge species and assessed the drivers of microbiome composition. In the chapter 4 entitled “Ecogenomic resource partitioning in the Great Amazon Reef Sponge Symbionts” we evaluate the genomes recovered from metagenomes of marine

sponge from Great Amazon Reef System (GARS) in terms of taxonomic diversity and ecological functional features. Finally, in the chapter 5 we present the conclusions of this study.

2 Materials and Methods

2.1 Study Area

The Great Amazon Reef System (GARS) is a mesophotic reef ecosystems off Amazon river mouth in the Brazilian Amazon Continental Shelf (BANHA et al., 2022; DE MAHIQUES et al., 2019; FRANCINI-FILHO et al., 2018; MOURA et al., 2016; OMACHI et al., 2019). This extensive reef encompasses an area of approximately 56,000 km^2 and is composed by a diverse and complex habitats, such as reef platforms, reef walls, rhodolith beds and sponge bottoms (DE MAHIQUES et al., 2019; FRANCINI-FILHO et al., 2018; MOURA et al., 2016; OMACHI et al., 2019). Under the influence of the Amazon River Plume, the light penetration is limited due to high sediment concentration, particularly in the northern sector of the reef (**Figure 5**). Although, it was demonstrated that there is enough light levels for photosynthetic organisms across the reef (OMACHI et al., 2019).

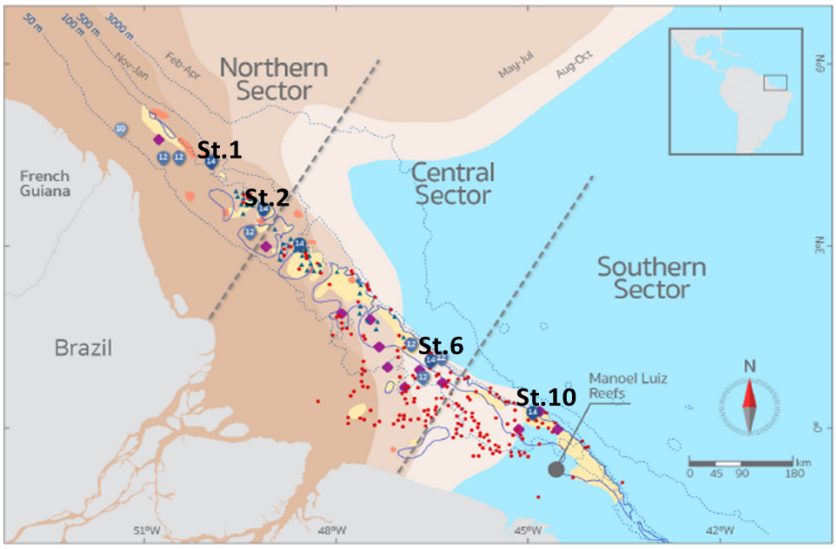

Figure 5 Map of Great Amazon Reef System area in Brazil showing sampling sites. Station 1 and station 2 are located in the Northern Sector where the plume is permanent, whereas station 6 and station 10 in the Central Sector and Southern sector, respectively, lacked the Amazon River Plume on the sampling season. Figure edited from (MOURA et al., 2016), Science Advances.

2.2 Sampling

During the oceanographic expedition onboard the Brazilian Navy ship NHo Cruzeiro do Sul (H-38) between September 24th to 29th in 2014, we collected 30 specimens of 20 marine sponge species **(Table 1) (Figure 5)** by bottom trawls and dredges. A total of 7 species were collected in the North sector, 2 at station 1

(N4°23.492', W50°42.575') and 5 at station 3 (N03°35.4267, W049°07.6028) **(Table 1, Table 2)**. In the Central and South sectors were collected 15 sponge species. Of this total 13 at station 6 (N1°18.329', W46°47.840'), 1 at station 8 (N00°45.359', W046°38.49') and 1 at station 10 (S0°15.877', W44 °52.416') **(Table 1, Table 2)**. The samples were stored and frozen in liquid nitrogen until processing. The water samples from the sites where the sponges were collected were previously evaluated in the student's master's thesis, regarding the physical-chemical profile and diversity of the microbial community through metagenomes analysis. These samples will be the controls for the analyzes of the sponge microbiomes (**Table 3**).

Table 1 Marine sponge species collected in the Great Amazon Reef System (GARS). Column "Specie.Ab" stands for sponge species name abbreviation. 1.: (GLOECKNER et al., 2014), 2.: (MOITINHO-SILVA et al., 2017b) 3.: (DE MENEZES et al., 2022), (*): inference from this study.

Sponge Species	Microbial Abundance	Order	Family	Specie.Ab	Station
Aplysina cauliformis	HMA (1)	Verongiida	Aplysinidae	Aply.caul	St.6
Aplysina fistularis	HMA (1)	Verongiida	Aplysinidae	Aply.fist	St.6
Arenosclera amazonensis	LMA (3)	Haplosclerida	Callyspongiidae	Aren.amaz	St.6, St.8
Arenosclera klausis	LMA*	Haplosclerida	Callyspongiidae	Aren.klau	St. 10
Callyspongia aculeata	LMA (2,3)	Haplosclerida	Callyspongiidae	Callys.acul	St.6
Cinachyrella kuekenthali	HMA (3)	Tetractinellida	Tetillidae	Cin.kuek	St.3
Clathria nicoleae	LMA (2,3)	Poecilosclerida	Microcionidae	Clat.nicol	St.6
Coelocarteria alcoladoi sp. nov.	HMA (3)*	Poecilosclerida	Isodictyidae	Coe.alcol	St.3
Coelocarteria amadoi sp. nov.	HMA (3)	Poecilosclerida	Isodictyidae	Coe.amad	St.3, St.6
Coelocarteria bartschi	HMA (3)	Poecilosclerida	Isodictyidae	Coe.bart	St.6
Geodia cf. corticostylifera	HMA (2)	Tetractinellida	Geodiidae	Geo.cort	St.6
Geodia neptuni	HMA (2)	Tetractinellida	Geodiidae	Geo.nep	St.6
Geodia sp.	HMA (2)	Tetractinellida	Geodiidae	Geo	St.6
Hyattella cavernosa	HMA*	Dictyoceratida	Spongiidae	Hyat.cav	St.1
Monanchora arbuscula	LMA (1)	Poecilosclerida	Crambeidae	Mon.arb	St.6
Neopetrosia proxima	HMA (2)	Haplosclerida	Petrosiidae	Neo.prox	St.3
Perissinella fosteri	LMA (3)	Bubarida	Dictyonellidae	Per.fost	St.6
Petromica citrina	HMA*	Bubarida	Desmanthidae	Petro.cit	St.3

Topsentia ophiraphidites	HMA (3)	Suberitida	Halichondriidae	Top.ophi	St.6
Tribrachium schmidtii	HMA (3)	Tetractinellida	Ancorinidae	Trib.schm	St.1

2.3 Sponge Species Classification

The taxonomic identification of sponge specimens was based on microscopic preparations of dissociated spicules and thick anatomical sections of samples fragments, as previously described (HADJU; PEIXINHO; FERNANDEZ, 2011).

Table 2 Data of water physical-chemical parameters from sites where sponges were collected in GARS. St.: Station. DO: dissolved oxygen. DOC: dissolved organic carbon. POC: particulate organic carbon. NT: nitrogen total. NTP: nitrogen total particulated.

Station	**St.1**	**St.3**	**St.6**	**St.8**	**St. 10**
Sector	Northern	Northern	Central	Southern	Southern
Area	Plume	Plume	Non-plume	Non-plume	Non-plume
Depth (m)	64	91	53	51	23
Temperature (°C)	26.45	25.74	27.14	27.44	27.13
Salinity (psu)	36.34	36.3	36.15	36.31	36.35
DO (ml/L)	4.79	4.62	2.87	4.64	5.08
DOC (µM)	144	99	138	103	134
POC(µM)	12.32	4.32	4.71	7.04	2.95
NT(µM)	14	12	16	10	12
NTP(µM)	1.66	0.49	0.51	0.71	0.28

2.4 Sponge Samples DNA Extraction

Fragments of sponge samples were crushed with sterile mortar and pestle in the presence of liquid nitrogen. The DNA extraction of the material crushed was carried out according to the protocol in (GARCIA et al., 2013; HADJU; PEIXINHO; FERNANDEZ, 2011).

2.5 Metagenomic DNA Quality and Sequencing

Metagenomic DNA purity and quality was evaluated through Nanodrop absorbance (Nanodrop 2000 Spectrophotometer) and quantified using Qubit (High Sensitivity DNA Kit - Agilent, Santa Clara, CA, USA). In addition, DNA integrity was accessed through 1% agarose gel analysis. Metagenomic DNA libraries were sequenced by Illumina HiSeq 2500 (paired-end sequencing, 2 × 150 base pairs) according to manufacturer's protocol.

2.6 Data Quality Control

Quality control of sequences was performed with Trimmomatic-0.36 (BOLGER; LOHSE; USADEL, 2014) by removing reads with phred score lower than 25. Host sequences were removed in silico using BWA-mem (LI; DURBIN, 2009) with default parameters by mapping reads against the only two genomes of sponges deposited in the NCBI, *Amphimedon queenslandica* (ID: 2698) and *Aplysina aerophoba* (ID: 67299). The reads not mapped were submitted to specific analyses described in the following chapters of results: chapter 3 (Metagenomic Profiling of Sponge Microbiomes from Great Amazon Reef System using read-based analysis) and chapter 4 (Ecogenomic resource partitioning in the Great Amazon Reef Sponge Symbionts).

Table 3 Metagenomes of column water from GARS.

Sampling Sites	Sequences (N)
St.1 (2 m) 1	1206798
St.1 (2 m) 2	1396143
St.2 (2 m) 1	1662798
St.2 (2 m) 2	1284472
St.1 (54 m) 1	1433696
St.1 (54 m) 2	1284412
St.2 (110 m) 1	1982981
St.2 (110 m) 2	477306
St.6 (2 m) 1	1376984
St.6 (2 m) 2	1399868
St.10 (2 m) 1	1378756
St.10 (2 m) 2	1473819
St.6 (43 m) 1	1318101
St.6 (43 m) 2	1370973
St.10 (15 m) 1	1088977
St.10 (15 m) 2	2479967

Table 4 Sequences counts of Sponge metagenomes from GARS. N: number of sequences. QC: quality control. Column “Samples.Ab” shows name abbreviation for the metagenomic samples per sponge species.

Sponge Species	Samples.Ab	Sequences (N)	Post-QC (N)(%)	Host Unmapped (N)(%)
Aplysina cauliformis	Aply.caul-25	2750602	2644594 (96.15)	2402580 (90.85)
Aplysina cauliformis	Aply.caul-27	4867699	4678337 (96.11)	4094457 (87.52)
Aplysina fistularis	Aply.fist-26	14971707	12555439 (83.86)	11959169 (95.25)
Aplysina fistularis	Aply.fist-28	3459151	3340113 (96.56)	2846342 (85.22)
Arenosclera amazonensis	Aren.amaz-58	14373435	12362262 (86.01)	11734118 (94.92)
Arenosclera amazonensis	Aren.amaz-778	1149731	766692 (66.68)	605546 (78.98)

Arenosclera klausis	Aren.klau-757	433937	302695 (69.76)	218816 (72.29)
Callyspongia aculeata	Callys.acul-43	3042796	2816064 (92.55)	1847836 (65.62)
Cinachyrella kuekenthali	Cin.kuek-62	12739526	7813853 (61.34)	7363870 (94.24)
Cinachyrella kuekenthali	Cin.kuek-62-2	2491782	2280599 (91.52)	1730736 (75.89)
Clathria nicoleae	Clat.nicol-42-2	2961545	2856917 (96.47)	2085281 (72.99)
Coelocarteria alcoladoi sp. nov.	Coe.alcol-61-1	2323801	2245548 (96.63)	1805545 (80.41)
Coelocarteria alcoladoi sp. nov.	Coe.alcol-61-2	3119498	3005578 (96.35)	2402656 (79.94)
Coelocarteria amadoi sp. nov.	Coe.amad-44	14737005	9777533 (66.35)	9222668 (94.33)
Coelocarteria amadoi sp. nov.	Coe.amad-44-2	1996317	1928687 (96.61)	1460645 (75.73)
Coelocarteria amadoi sp. nov.	Coe.amad-63	31398977	20682438 (65.87)	19499824 (94.28)
Coelocarteria amadoi sp. nov.	Coe.amad-63-2	5159308	4981726 (96.56)	3686911 (74.01)
Coelocarteria bartschi	Coe.bart-17	12418751	7260122 (58.46)	6985720 (96.22)
Geodia cf. corticostylifera	Geo.cort-39	17187284	14212604 (82.69)	13669259 (96.18)
Geodia cf. corticostylifera	Geo.cort-40	11280196	9317116 (82.6)	8953560 (96.1)
Geodia neptuni	Geo.nep-45	14038441	8046549 (57.32)	7742127 (96.22)
Geodia sp.	Geo-48	11178642	8521542 (76.23)	8222383 (96.49)
Hyattella cavernosa	Hyat.cav-56-1	2708583	2403363 (88.73)	2127862 (88.54)
Hyattella cavernosa	Hyat.cav-56-2	3339201	3115098 (93.29)	2713107 (87.1)
Monanchora arbuscula	Mon.arb-22	15622090	14866511 (95.16)	9642311 (64.86)
Monanchora arbuscula	Mon.arb-47	15591586	12588669 (80.74)	11597357 (92.13)
Neopetrosia proxima	Neo.prox-59	17744749	11730795 (66.11)	10624958 (90.57)
Neopetrosia proxima	Neo.prox-60	3228027	3111934 (96.4)	2302700 (74)
Perissinella fosteri	Per.fost-18	14201585	9672962 (68.11)	9010108 (93.15)
Perissinella fosteri	Per.fost-19	14962419	9830928 (65.7)	9156597 (93.14)
Petromica citrina	Petro.cit-53-1	3622358	3301926 (91.15)	2552690 (77.31)
Petromica citrina	Petro.cit-53-2	3810516	3630432 (95.27)	3208997 (88.39)
Topsentia ophiraphidites	Top.ophi-12	2444807	2317381 (94.79)	1963815 (84.74)
Topsentia ophiraphidites	Top.ophi-20	5016953	4671951 (93.12)	4112745 (88.03)
Topsentia ophiraphidites	Top.ophi-24	40695486	23257169 (57.15)	22240558 (95.63)
Tribrachium schmidtii	Trib.schm-57	16076883	13138613 (81.72)	12407738 (94.44)
Tribrachium schmidtii	Trib.schm-57-2	2580712	2409945 (93.38)	1929666 (80.07)

2.7 Credits

The steps of sponge samples sampling were executed by members of the oceanographic expedition onboard the Brazilian Navy ship NHo Cruzeiro do Sul (H-38) in 2014 (MOURA et al., 2016), in which I participated. DNA Extraction and sequencing were performed by members of Laboratory of Microbiology in Federal University of Rio de Janeiro, including Louise Oliveira, Ana Paula Moreira and Tatiane de Menezes.

3 Metagenomic Profiling of Sponge Microbiomes from Great Amazon Reef System using read-based analysis

3.1 Introduction

In this chapter, we studied the marine sponge microbiomes composition, diversity and structure of 20 species from Great Amazon Reef System through shotgun metagenomics using read-based analysis. We also compared the microbial communities among the different sponge species and assessed the drivers of microbiome composition.

3.2 Materials and Methods

3.2.1 Taxonomic Assignment of Reads

The metagenomic sequences which passed the previous quality control step and removal of host-derived DNA were submitted to taxonomic classification step with software Kraken 2 (WOOD; LU; LANGMEAD, 2019) using the database nr from NCBI. The metagenomes of column water surrounding sponges from GARS were analyzed by the student in the Master's Dissertation. These water samples were evaluated as control samples. Microbial taxa count generated by kraken report were normalized by total reads counts per sample to obtain the relative abundance of each taxon in the sponge metagenomes.

3.2.2 Alpha and Beta Diversity

Alpha diversity of sponge metagenomic samples was evaluated considering diversity indices and richness metrics. Using function diversity of vegan package (DIXON, 2003), we estimated the alpha diversity of sponge microbiomes at taxonomic levels of Phylum, Order, Genus and Species by the indices Shannon's H', Inverse Simpson, Simpson and Pielou's Evenness. Richness was predicted by counts of taxa (i.e. Richness) and Chao1 estimator. Boxplots graphs of alpha diversity were generated with ggplot2 (WICKHAM, 2016) in R (DE MICHEAUX; DROUILHET; LIQUET, 2014).

Beta diversity analysis was performed to establish the correlation between microbiomes from different sponge species from GARS. Non-metric multidimensional scaling (nMDS) based on Bray-Curtis dissimilarity matrix was

calculated from relative abundance data using function metaMDS of vegan (DIXON, 2003) package in R (DE MICHEAUX; DROUILHET; LIQUET, 2014). The plots to visualize the nMDS ordination was generated with ggplot2 (WICKHAM, 2016).

3.2.3 Statistical Analysis

To analyze the contribution of microorganisms associated to sponge metagenomes samples by phylogenetic group or ecological group, the abundance results were shown in the format of mean and standard deviation (mean ± sd).

PERMANOVA (permutational multivariate analysis of variance) test was performed using function *adonis2* of vegan (DIXON, 2003) package in R (DE MICHEAUX; DROUILHET; LIQUET, 2014) to assess factors which influence in the sponge microbiome structure. The factors tested were host phylogeny, microbial abundance status (HMA or LMA) and environmental variables (**Table 2**). Results with p-value < 0.05 were considered significant (*).

3.3 Results

A total of 262,444,685 high quality *paired-end* sequences were generated for the 37 sponge metagenomes of 20 sponge species from GARS (**Table 1** and **Table 4**). Of this total, 236,131,258 reads were not mapped against host sponge genomes (**Table 4**). The percent of reads after this filter that were assigned taxonomically to *Archaea* or *Bacteria* domain was 55.77% (N = 131,699,180).

3.3.1 Taxonomic Composition

The taxonomic profile of microorganisms associated to the 20 sponge species from GARS were evaluated in taxonomic levels of phylum, order, genus and species. Due to the high diversity of phylum *Proteobacteria*, the members of this group were shown in class taxonomic level (we refer to non proteobacteria phyla and proteobacteria classes generically as high-level taxa). The composition and diversity analysis were also carried out comparing HMA and LMA sponge species because of the importance of this feature in the structure of sponge microbiomes.

The taxonomic composition analysis revealed 167 microbial high-level taxa for the 37 sponge metagenomes. The predominant bacterial high-level taxa were *Gammaproteobacteria* (18.86% ± 4.54%), followed by Firmicutes (16.24% ± 4.40%), *Alphaproteobacteria* (14.69% ± 2.17%), *Actinobacteria* (13.63% ± 3.76%), *Bacteroidetes* (8.5% ± 2.47%), *Betaproteobacteria* (5.24% ± 0.82%), *Chloroflexi* (3.52% ± 2.04%), *Deltaproteobacteria* (2.76% ± 0.39%), *Cyanobacteria* (2.64% ± 0.45%) and *Acidobacteria* (2.61% ± 1.28%) (**Figure 6**). The two predominant archaeal phyla assigned in sponge metagenomes were *Thaumarchaeota* (0.75% ± 0.91) and *Euryarchaeota* (0.68% ± 0.17) (**Figure 6**). Unlike *Euryarchaeota*, *Thaumarchaeota* proportion presented a high variability, ranging from 0.05% (*Arenosclera amazonensis* - LMA) to 4.02% (*Cinachyrella kuekenthali* - HMA) (**Figure 7** and **Figure 8**).

HMA and LMA sponge species could be distinguished by the fact that specific phyla were more abundant in HMA species than in LMA species. Examples are: *Chloroflexi* (HMA = 4.37% ± 1.58%, LMA = 0.90% ± 0.32%), *Poribacteria* (HMA = 0.17% ± 0.118%, LMA = 0.01% ± 0.040%), *Dadabacteria* (HMA = 0.097 % ± 0.072%, LMA = 0.007% ± 0.003%), *Nitrospirae* (HMA = 0.525% ± 0.187%, LMA = 0.194% ± 0.039%) and *Thaumarchaeota* (HMA = 0.91% ± 0.973%, LMA = 0.28% ± 0.462) (**Figure 7** and **Figure 8**).

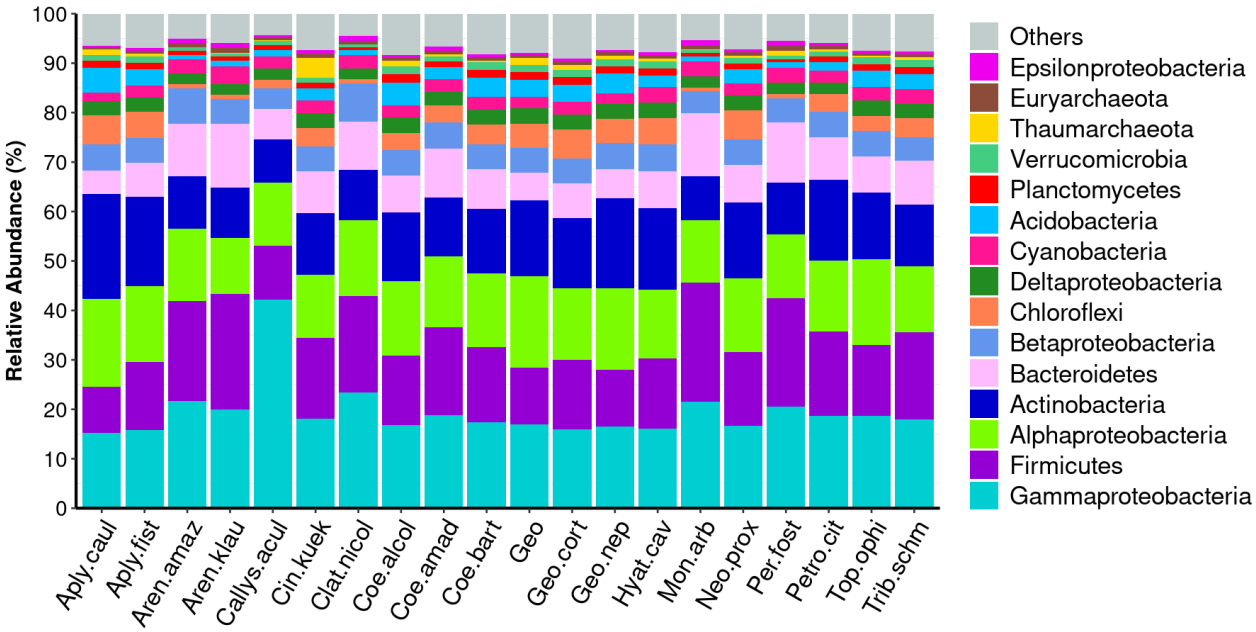

Figure 6 Bar graph of relative abundance of sponge species microbiome from GARS in phylum level. The phylum proteobacteria is shown in classes.

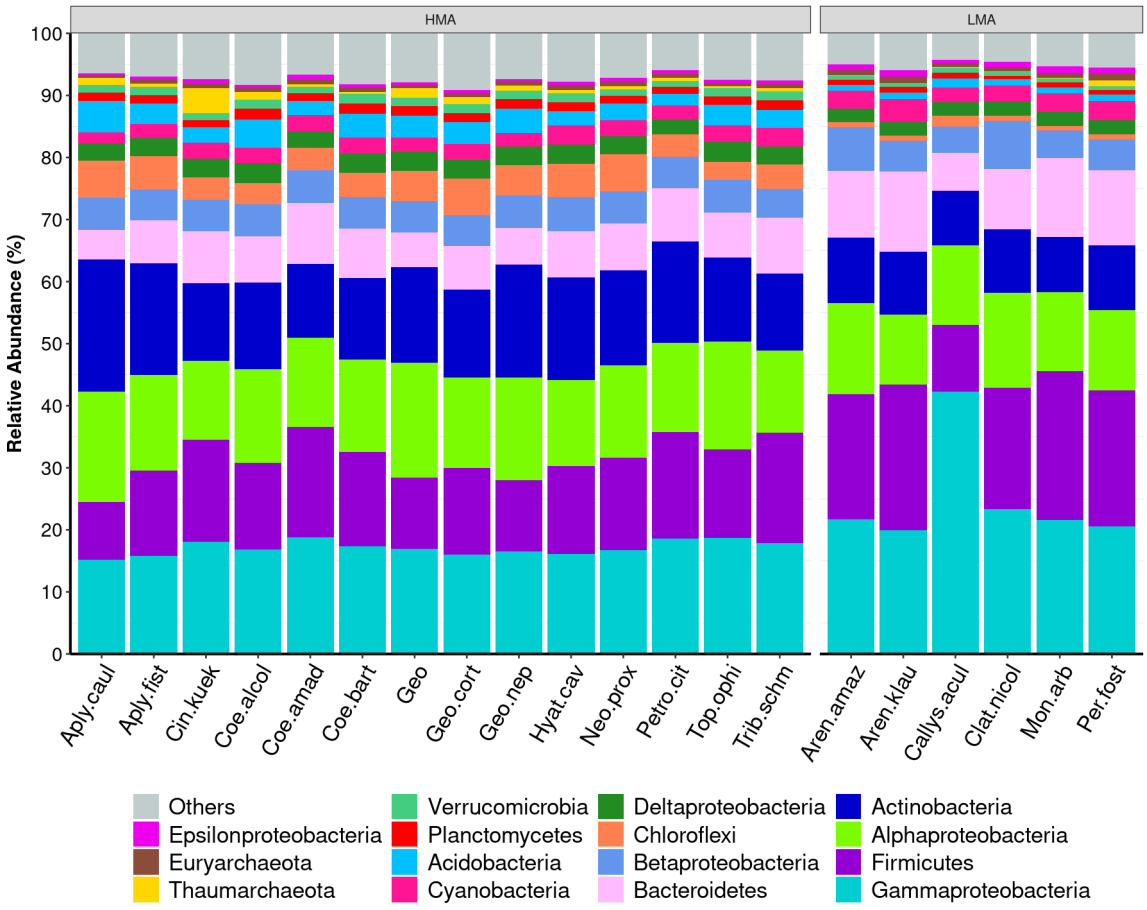

Figure 7 Bar graph of relative abundance of sponge species microbiome from GARS in phylum level. The sponge species in axis-x were divided in HMA and LMA. The phylum proteobacteria is shown in classes.

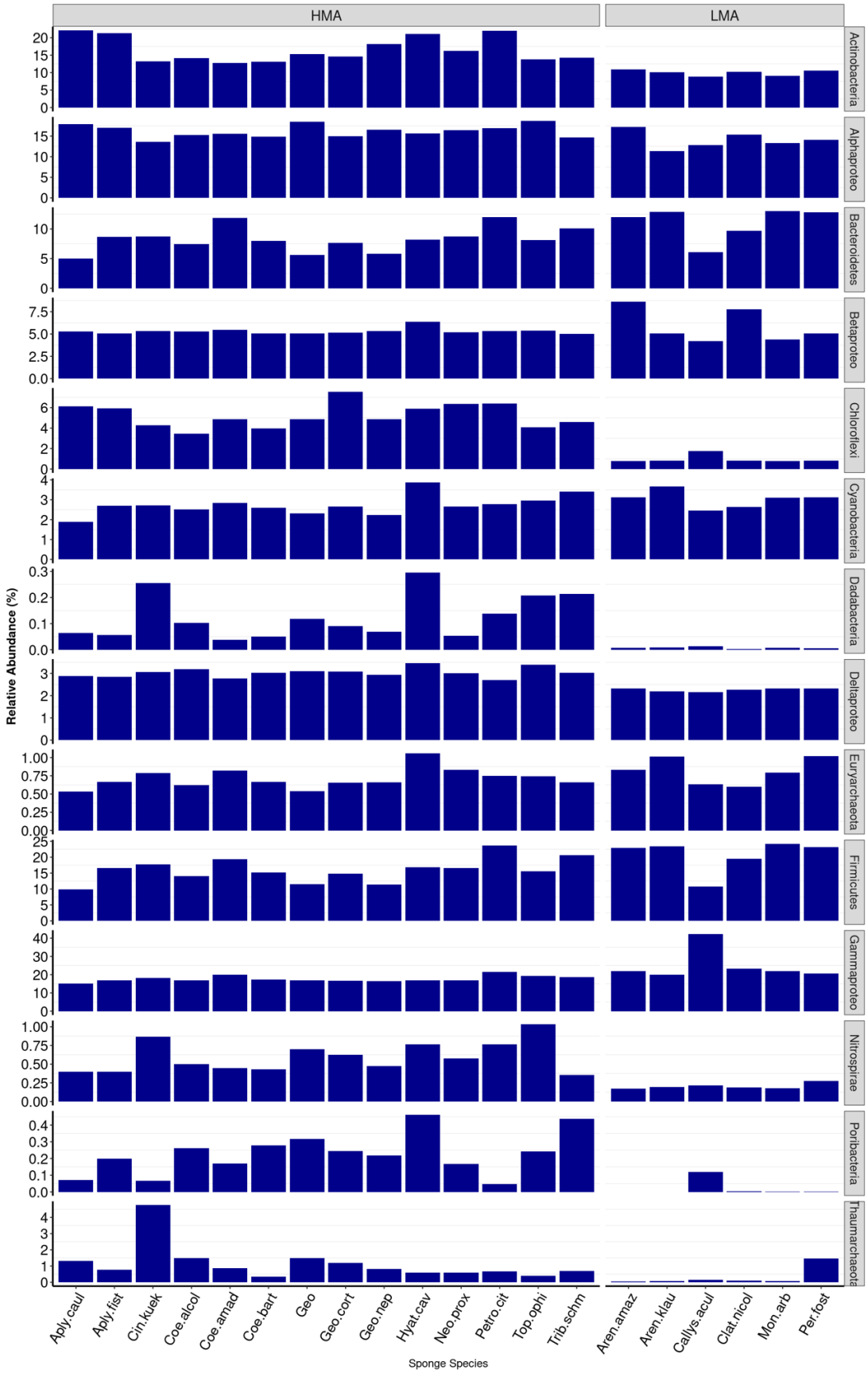

Figure 8 Bar graph of relative abundance of specific phyla of sponge species microbiome from GARS: 9 most abundant bacterial phyla, the two archaeal phyla (Thaumarchaeota and Euryarchaeota) plus phyla Poribacteria, Nitrospirae and Dadabacteria. The sponge species in the x-axis were divided in HMA and LMA.

A total of 31,611 microbial species were assigned for the 37 sponge metagenomes. The twenty most abundant microbial species in sponge microbiomes were *Acidobacteria bacterium* (2.35% ± 1.2%), followed by *bacterium TMED15* (1.54% ± 1.24%), *Thioalkalivibrio sp. HK1* (0.92% ± 4.74%), *Chloroflexi bacterium* (0.72% ± 0.48%), *Rhizobiales bacterium* (0.55% ± 0.04%), *Candidatus Handelsmanbacteria bacterium* (0.54% ± 0.37%), *Candidatus Rokubacteria bacterium* (0.41% ± 0.21%), *Verrucomicrobia bacterium* (0.39% ± 0.07%), *Rhodospirillaceae bacterium Spongia-Bin9* (0.38% ± 0.35%), *Opitutae bacterium TARA_B100001123* (0.35% ± 0.24%), *SAR202 cluster bacterium Io17-Chloro-G9* (0.3% ± 0.25%), *Dehalococcoidia bacterium* (0.24% ± 0.2%), *uncultured bacterium* (0.24% ± 0.02%), *Escherichia coli* (0.24% ± 0.07%), *Litorilinea aerophila* (0.21% ± 0.14%), *uncultured Poribacteria bacterium 64K2* (0.18% ± 0.16%), *SAR202 cluster bacterium Casp-Chloro-G4* (0.18% ± 0.17%), *SAR202 cluster bacterium Casp-Chloro-G3* (0.18% ± 0.15%), *Bacteroidetes bacterium* (0.18% ± 0.12%) and *Bacillus cereus* (0.17% ± 0.07%) (**Figure 9**).

Thioalkalivibrio sp. HK1 displayed the third highest mean for all sponges, mainly because its abundance in the LMA sponge *Callyspongia aculeata* (28.94%) (**Figure 10** and **Table 5**).

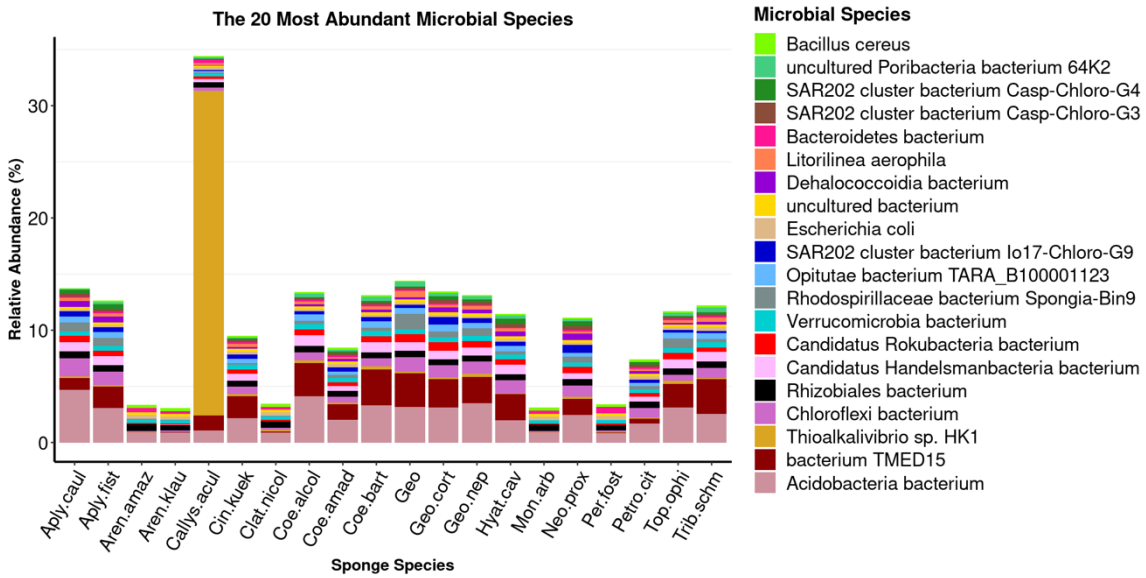

Figure 9 Bar graph of the 20 most abundant microbial species in sponge microbiomes from GARS. The x-axis shows the sponge species and y-axis the relative abundance (%).

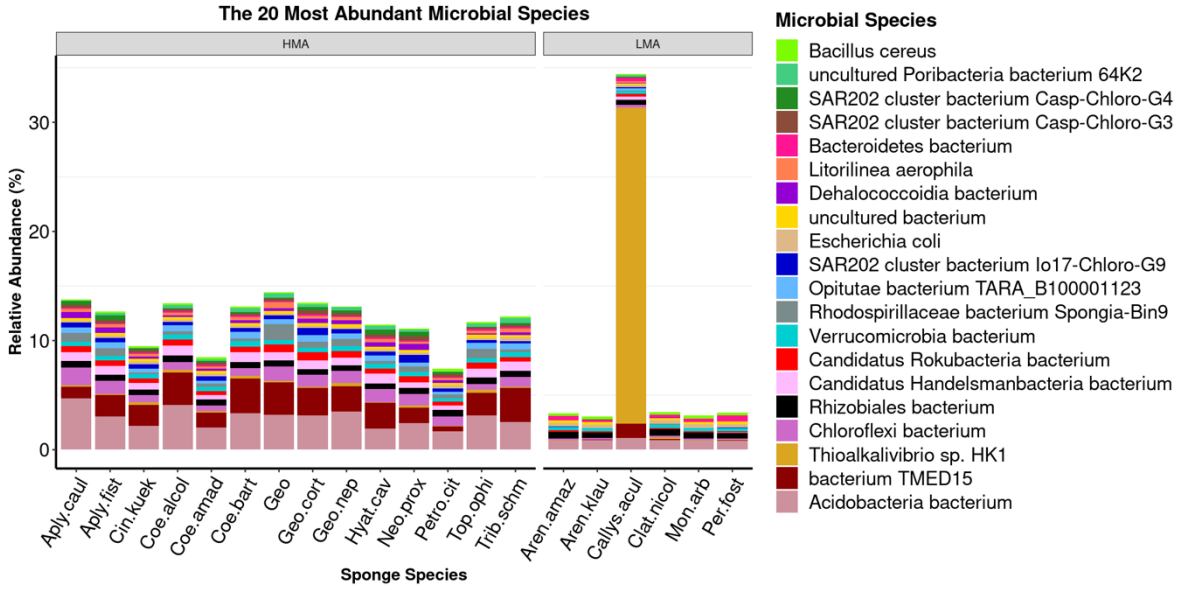

Figure 10 Bar graph of the 20 most abundant microbial species in sponge microbiomes from GARS. The sponge species in x-axis were divided in HMA and LMA.

Table 5 Relative Abundance of the 20 most abundant microbial species in sponge microbiomes divided by HMA and LMA sponges. Results is shown by Mean ± standard deviation.

Species	HMA	LMA
Acidobacteria bacterium	2.83% ± 0.98%	0.88% ± 0.08%
bacterium TMED15	1.98% ± 1.08%	0.17% ± 0.43%
Thioalkalivibrio sp. HK1	0.17% ± 0.1%	3.26% ± 9.63%
Chloroflexi bacterium	0.9% ± 0.41%	0.17% ± 0.05%
Rhizobiales bacterium	0.56% ± 0.04%	0.54% ± 0.05%
Candidatus Handelsmanbacteria bacterium	0.7% ± 0.27%	0.04% ± 0.1%
Candidatus Rokubacteria bacterium	0.5% ± 0.15%	0.13% ± 0.04%
Verrucomicrobia bacterium	0.42% ± 0.04%	0.29% ± 0.02%
Rhodospirillaceae bacterium Spongia-Bin9	0.49% ± 0.33%	0.03% ± 0.02%
Opitutae bacterium TARA_B100001123	0.46% ± 0.17%	0.02% ± 0.06%
SAR202 cluster bacterium Io17-Chloro-G9	0.4% ± 0.21%	0.02% ± 0.03%
Escherichia coli	0.22% ± 0.06%	0.31% ± 0.04%
uncultured bacterium	0.25% ± 0.02%	0.22% ± 0.03%
Dehalococcoidia bacterium	0.31% ± 0.17%	0.02% ± 0.02%
Litorilinea aerophila	0.26% ± 0.11%	0.02% ± 0.05%
Bacteroidetes bacterium	0.13% ± 0.03%	0.34% ± 0.14%
SAR202 cluster bacterium Casp-Chloro-G3	0.23% ± 0.12%	0.01% ± 0.02%
SAR202 cluster bacterium Casp-Chloro-G4	0.23% ± 0.16%	0.01% ± 0.01%
uncultured Poribacteria bacterium 64K2	0.23% ± 0.16%	0.02% ± 0.05%
Bacillus cereus	0.15% ± 0.06%	0.25% ± 0.07%

3.3.2 Alpha Diversity of Sponges Microbiomes from GARS

The microbial community diversity of sponges was investigated at taxonomic levels of Phylum, Order, Genus and Species using the indices Shannon's H', Inverse Simpson index, Simpson and Pielou's Evenness, together with richness metrics.

The alpha diversity analysis indicated that HMA sponge microbiomes presented a higher diversity than LMA sponges through Shannon's H' index, Inverse Simpson index, Simpson index and Pielou's Evenness index for all evaluated taxonomic levels (**Figure 11**A-D, **Figure 12**A-D, **Figure 13**A-D), except species level (**Figure 14**A-D). The same pattern was observed in species level analysis only for richness estimator metrics, such as Chao1 (**Figure 14**E) and taxon counts (i.e., Richness) (**Figure 14**F).

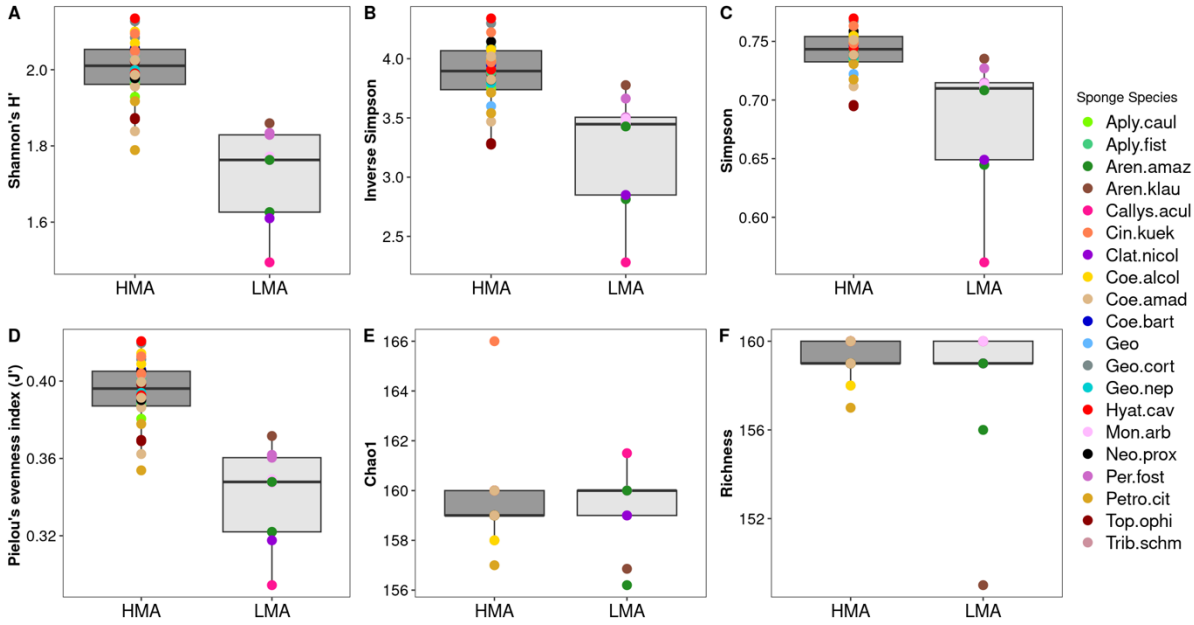

Figure 11 Phylum Alpha diversity of HMA and LMA sponge species samples (N=37). Diversity indexes **(A, B, C, D)** and Richness metrics **(E, F)** were calculated using phyla abundances. Boxplots were generated with ggplot2 in R. The colored dots represent each sponge species.

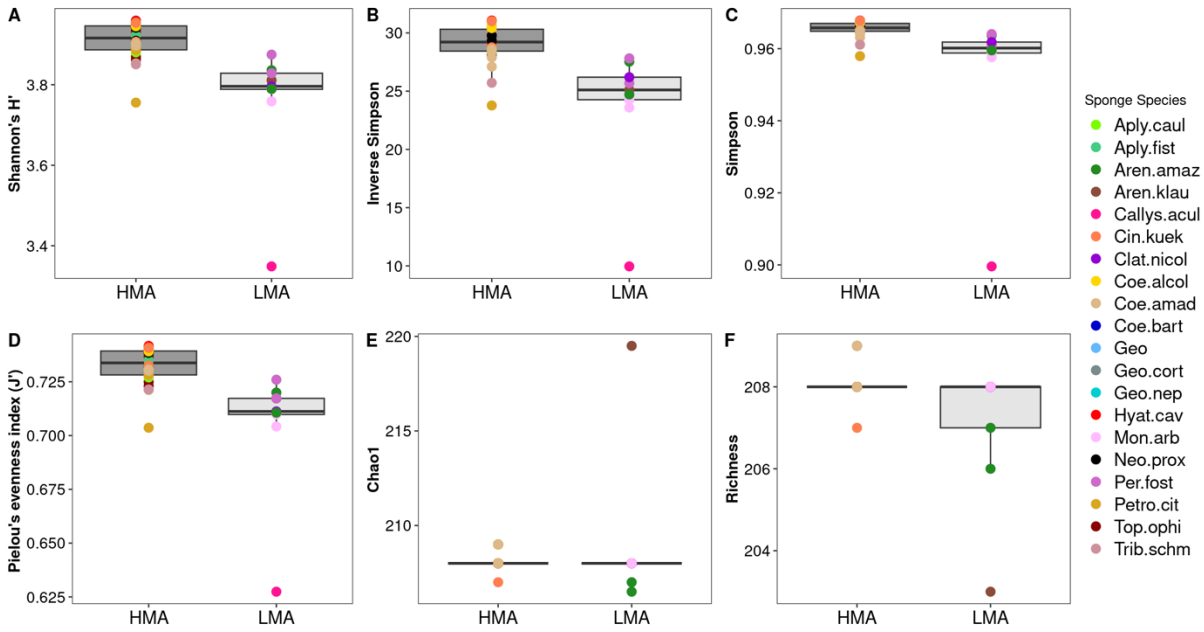

Figure 12 Order Alpha diversity of HMA and LMA sponge species samples (N=37). Diversity indexes **(A, B, C, D)** and Richness metrics **(E, F)** were calculated using order abundances. Boxplots were generated with ggplot2 in R. The colored dots represent each sponge species.

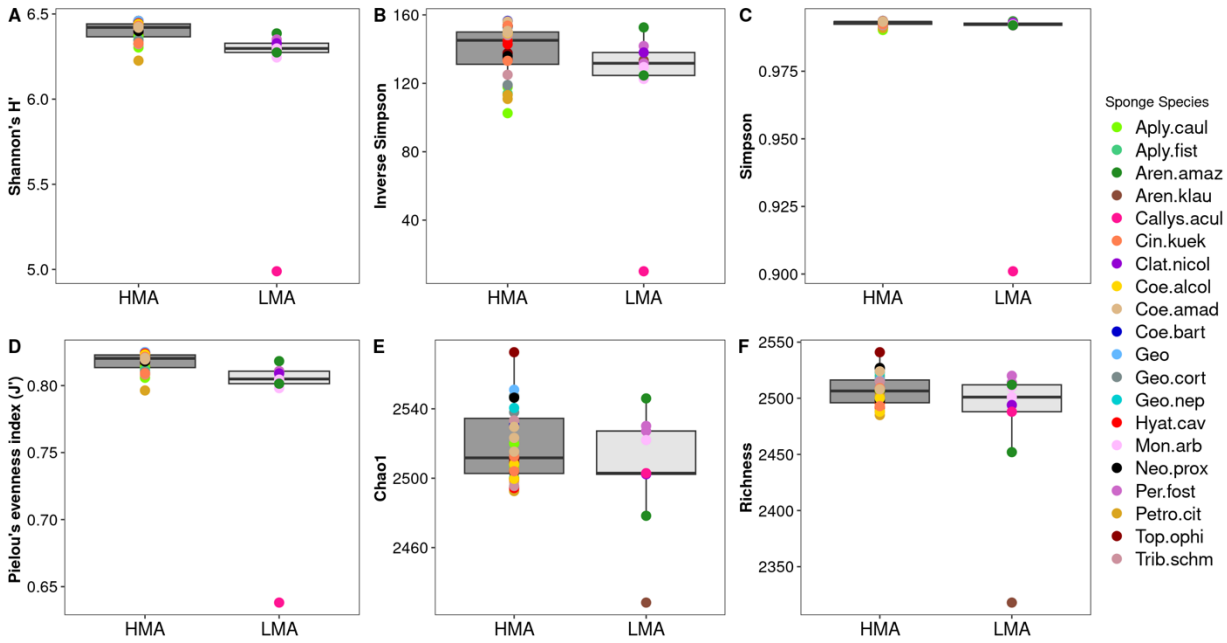

Figure 13 Genus Alpha diversity of HMA and LMA sponge species samples (N=37). Diversity indexes **(A, B, C, D)** and Richness metrics **(E, F)** were calculated using genus abundances. Boxplots were generated with ggplot2 in R. The colored dots represent each sponge species.

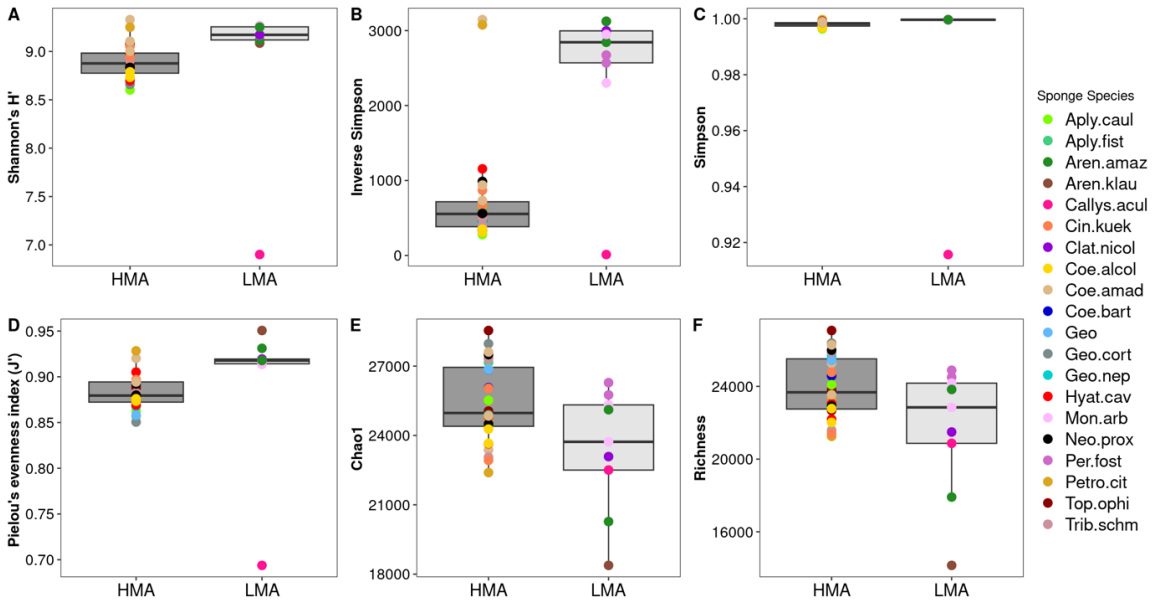

Figure 14 Species Alpha diversity of HMA and LMA sponge species samples (N=37). Diversity indexes **(A, B, C, D)** and Richness metrics **(E, F)** were calculated using species abundances. Boxplots were generated with ggplot2 in R. The colored dots represent each sponge species.

3.3.3 Host Identity and Environmental factors Influence on the composition of sponge microbiomes from GARS

Beta diversity analysis were performed to compare the microbial community between microbiomes of different sponge species from GARS.

The NMDS plot based on microbial species profile showed that sponge microbiomes from the same host species displayed a specific microbial community (**Figure 15**A). The PERMANOVA test corroborated the grouping pattern by sponge species observed in NMDS ordination (p-value 0.011, R2 = 0.64873) (**Table 6**). This result is not statistically significant for higher taxonomic levels of host phylogeny, such as order and genera (**Table 6**) but the NMDS exposed a similarity between sponge microbiomes of the same order (**Figure 15**C). Likewise, it was not possible statistically to distinguish the HMA and LMA sponges from GARS based on microbial species profile (**Table 6**), even though the NMDS ordination pointed to a similarity among sponges from same microbial abundance status (**Figure 15**B).

There were not statistically significant results to reveal the influence of environmental factors (**Table 2**) on sponge microbiome composition from GARS

through the following factors: geographical localization, depth, temperature, dissolved oxygen, carbon and nitrogen concentration.

Table 6 PERMANOVA (permutational multivariate analysis of variance) test based on Bray–Curtis dissimilarity matrix of total microbial species abundance. Column 1 stands for factors. (*) means p-value < 0.05.

Microbial Species Profile	Df	SumOfSqs	R2	F	Pr(>F)
Host Sponge Species	19	0.98316	0.64873	1.6524	0.011*
Host Sponge Genera	13	0.65968	0.43529	1.3638	0.083
Host Sponge Order	8	0.41307	0.27256	1.3114	0.126
Sponge Status HMA-LMA	1	0.03066	0.02023	0.7228	0.631

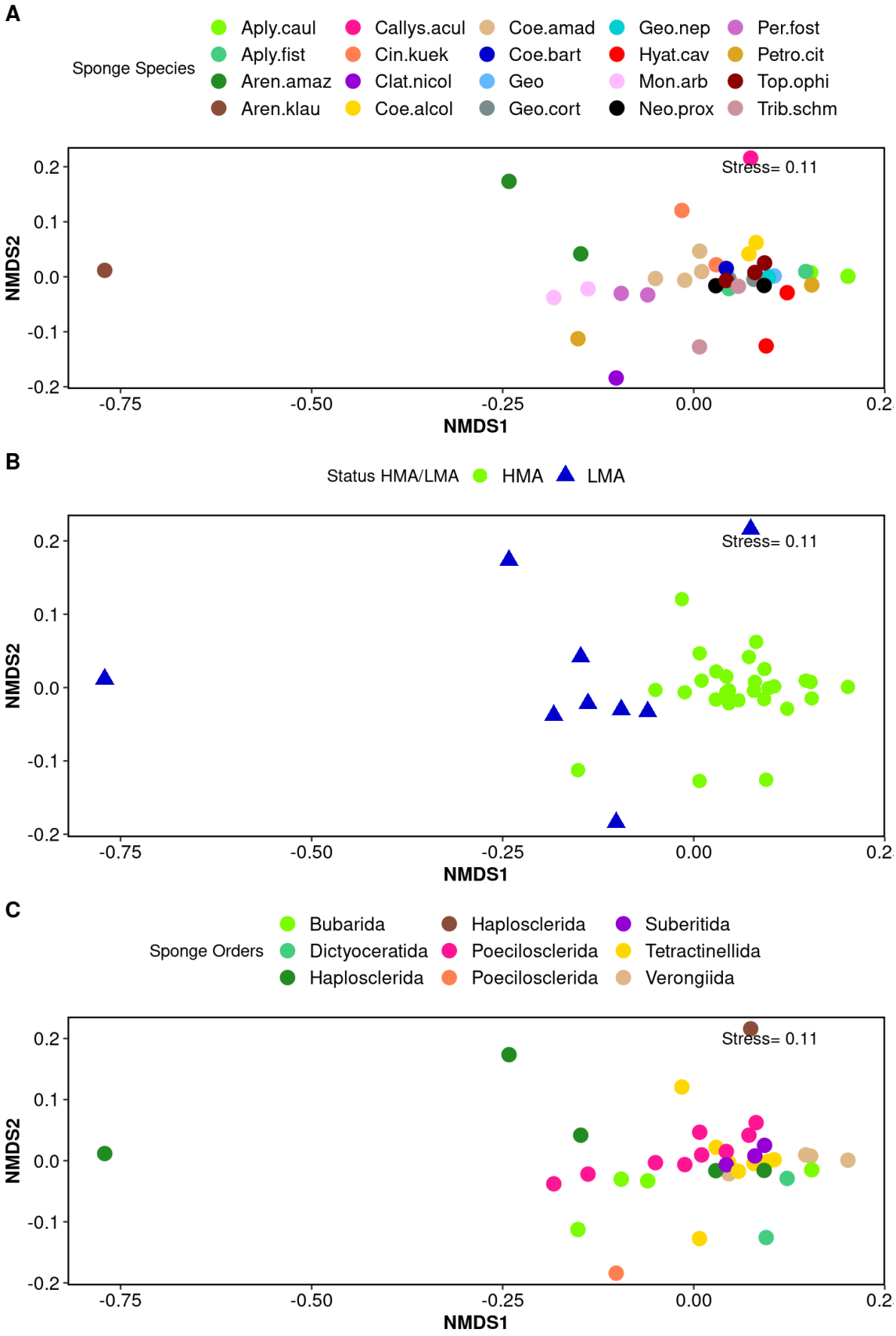

Figure 15 Non-metric multidimensional scaling (NMDS) of the 37 sponge metagenomes based on Bray–Curtis dissimilarity matrix of total microbial species abundance. **A.** Colors and shapes classifies the sponge host

species. **B.** Colors shows the Microbial Abundance type of sponge species. **C.** Colors classifies the sponge host orders.

3.3.4 Comparative Analysis of Sponges Microbiome and Surrounding Water

The taxonomic microbial composition of samples from water column collected at surface depth (2 m) and above seabed in the three sectors of GARS was compared with sponge microbiomes from GARS in order to evaluate the specificity of sponge associated microorganisms. The analysis evidenced that taxonomic profile of sponge microbiomes is distinguishable from the surrounding water from GARS (PERMANOVA: P = 0.001, R2 = 0.4681; **Figure 16**).

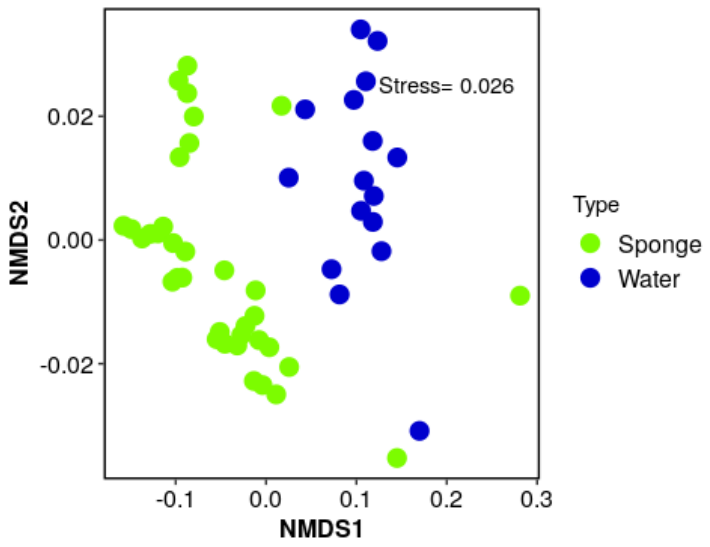

Figure 16 Non-metric multidimensional scaling (NMDS) of the 35 sponge metagenomes (lacks *Callyspongia aculeata* and *Clathria nicoleae*) based on Bray-Curtis distance matrix of microbial community abundance of phyla (log transformed) including metagenomes of sponges and surrounding water from GARS. The colors in NMDS stands for sample type: sponges in green and water samples in blue.

The alpha diversity analysis using Shannon index indicated that sponge microbiomes depicted a higher diversity than the surrounding water samples from the 3 sectors of GARS in different depths (**Figure 17**). The Shannon index value of sponges is approximately 9, whereas the Shannon water index range from about 6 in surface water at North sector (Plume (2 m)) to 7.5 (Subplume 110 m) also in North sector.

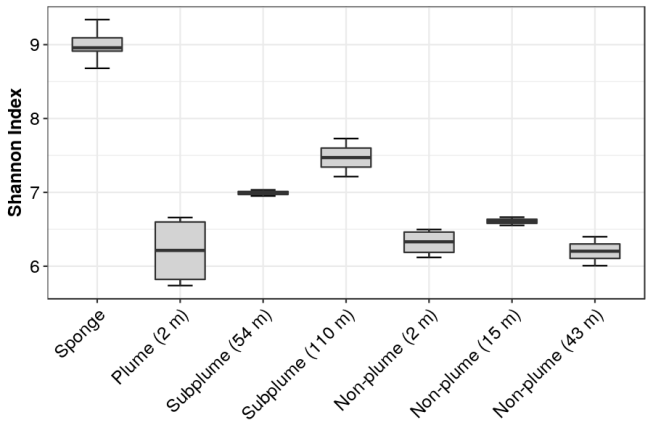

Figure 17 Alpha diversity boxplot of Shannon Index comparing microbial diversity of sponge samples with surrounding water in GARS based on species abundance. The axis-x stands for water samples sites and its relative sampling depth. Plume and Subplume comprise in Stations 1 and 3. Non-plume area comprises Stations 6, 8 and 10.

3.4 Discussion

Our results provide the first insight of sponge microbiomes composition and diversity from Great Amazon Reef System.

Generalists Microbial Symbionts Taxa in Sponges from Great Amazon Reef System

The predominant microbial taxa in the sponge microbiomes from GARS were Proteobacteria, specifically *Alphaproteobacteria, Betaproteobacteria and Deltaproteobacteria,* also *Firmicutes, Actinobacteria, Bacteroidetes, Chloroflexi, Cyanobacteria* and *Acidobacteria.* These bacterial phyla are widely reported to inhabit many sponges species (CLEARY et al., 2019; GILES et al., 2013; THOMAS et al., 2016; TRINDADE-SILVA et al., 2012; WEBSTER; TAYLOR, 2012). Together with the archaeal phyla *Thaumarchaeota* and *Euryarchaeota* they contribute to the core sponge microbiome from GARS (PITA et al., 2018).

Dominant Microbial Taxa of HMA and LMA Sponges

We identified some specific microbial phyla more abundant in sponges from GARS classified as HMA, such as *Chloroflexi, Actinobacteria, Nitrospirae* and *Poribacteria*. These phyla were considered HMA indicators in different HMA sponge species (MOITINHO-SILVA et al., 2017b). Similarly, in LMA sponges from

GARS the taxa *Bacteroidetes, Firmicutes* and *Betaproteobacteria*, which are microbial indicators taxa of LMA sponges (MOITINHO-SILVA et al., 2017b) have been found in higher abundance.

Predominance of *Thioalkalivibrio sp. HK1* in LMA sponge *Callyspongia aculeata*

Thioalkalivibrio sp. HK1 make up about 28% of *Callyspongia aculeata* microbial community, a LMA sponge (FREEMAN et al., 2021). This bacterial species belongs to family *Ectothiorhodospiraceae* (class *Gammaproteobacteria*, order *Chromatiales*), a phylogenetic lineage of purple sulfur bacteria (PSB) (OREN, 2014). The genus *Thioalkalivibrio* comprises obligate chemolithoautotrophic haloalkaliphilic sulfur-oxidizing bacteria (SOB) (AHN et al., 2017). The genome of species *Thioalkalivibrio sp. HK1* was recovered from *Haliclona cymaeformis* sponge metagenome*, which the study* showed a dominance of symbiotic sulfur-oxidizing *Ectothiorhodospiraceae* in its microbiome (TIAN et al., 2014). The genome shows bacterial symbionts features, such as an abundance of ankyrin repeat protein domains and lacks of transposases (LAVY et al., 2018). *Thioalkalivibrio sp. HK1* is mixotrophic, so is capable to oxidizes sulfide or sulfite through the reverse sulfate reduction pathway (FRIEDRICH *et al*., 2001) and present heterotrophic capability (LAVY *et al*., 2016). These characteristics suggest this bacteria is the main responsible for sulfur cycling within *Callyspongia aculeata* and also relevant in the oxidized environments.

Symbiotic Prokaryotic Species Hosted by Marine Sponges from GARS

The most abundant prokaryotic species identified in sponge microbiomes have been reported previously in sponge microbiomes. Among them, *bacterium TMED15 Chloroflexi bacterium*, *SAR202 cluster bacterium Io17-Chloro-G9, SAR202 cluster bacterium Casp-Chloro-G4*, *SAR202 cluster bacterium Casp-Chloro-G3 and Dehalococcoidia bacterium* are metagenome bins classified into clades within the phylum *Chloroflexi* (BAYER et al., 2018). *Dehalococcoidia members* are known as organohalide-respiring bacteria, playing relevant roles in carbon cycling in anoxic ecosystems (HUG et al., 2013; YANG et al., 2020). *Candidatus Handelsmanbacteria bacterium* belongs to phylum *Latescibacterota*

isolated from a soil metagenomic sample. *Rhodospirillaceae bacterium Spongia-Bin9* (KARIMI et al., 2018) *and Poribacteria bacterium 64K2* (SIEGL et al., 2011) were both isolated from sponge metagenomes.

Host Specificity in Sponge Microbiomes from GARS

We showed that sponge microbiomes from GARS are distinct from surrounding water, demonstrating that these sponges harbor their own microbial community, corroborating previous studies (THOMAS et al., 2016). Their microbiome is also species-specific for both LMA and HMA, but LMA seems to have an extended variability in core microbiome (BUSCH et al., 2022).

Drivers of Sponge Microbiomes Structure

The host phylogeny seems to be the main driver of sponge microbiomes structure and composition from GARS. This variable is usually the most relevant driver of microbiome composition in sponge studies (MOITINHO-SILVA et al., 2017b; SCHMITT et al., 2012).

Despite difference on microbial abundance and diversity in the sponge microbiomes, it was not possible statistically to distinguish the HMA and LMA sponges from GARS observing the species abundance profile, preventing us to confirm the influence of this variable. One reason which could explain this result is the smaller number of LMA sponge specimens that may disturb the analysis.

Although the environmental factors and water physical-chemical parameters evaluated in this study does not seem to drive microbiome composition, in (DE MENEZES et al., 2022) the analysis of microbial composition associated with sponges from GARS, in addition to lipidomic and isotopic analysis indicated that LMA sponges rely on the Amazon River Plume for nutrition. This result point to the importance of environment in sponge microbiomes structure from Great Amazon Reef System.

3.5 Final Remarks

The present study explores the first results of marine sponge microbiomes from Great Amazon Reef System. We showed that there is a group of predominant

microbial taxa that are consistently present in all sponge microbiomes. This suggests that these core microbial taxa play important roles in the sponge health. Further studies with functional analysis may be conducted to better understand the interactions and functions of these taxa. Specific taxa indicators of LMA and HMA sponges could be used as markers for different types of sponge from GARS. Although, the host phylogeny seems to be a major factor in shaping the composition and diversity of sponge microbiome, extra analysis with functional profile is necessary to investigate other factors, such as environmental conditions influence, mainly LMA sponges. These sponges are able to benefit from the nutrients present in the plume of the Amazon River (DE MENEZES et al., 2022).

4 Ecogenomic resource partitioning in the Great Amazon Reef Sponge Symbionts

4.1 Introduction

In this chapter, we studied the genomes recovered from metagenomes of marine sponge from Great Amazon Reef System (GARS). The metagenomes-assembled genomes (MAGs) were evaluated in terms of quality, taxonomic affiliation and functional features profile to reveal the ecological functions of sponge-associated microorganisms and their role in the sponge ecology and physiology.

4.2 Materials and Methods

4.2.1 Samples

A total of 35 metagenomes of 18 marine sponge species of class Demospongiae from Great Amazon Reef System (BANHA et al., 2022) were analyzed in this chapter (**Table *7***, **Table 4**). The samples processing was described in chapter 2.

Table 7 Sponge species used in this study with information on collection sites and their taxonomic assignments (Samples, Station, Area, Class, Order, Family, Species).

Samples	Station	Area	Order	Family	Species
Aply.caul-25	St.6	Non plume	Verongiida	Aplysinidae	*Aplysina cauliformis*
Aply.caul-27	St.6	Non plume	Verongiida	Aplysinidae	*Aplysina cauliformis*
Aply.fist-26	St.6	Non plume	Verongiida	Aplysinidae	*Aplysina fistularis*
Aply.fist-28	St.6	Non plume	Verongiida	Aplysinidae	*Aplysina fistularis*
Aren.amaz-778	St.6	Non plume	Haplosclerida	Callyspongiidae	*Arenosclera amazonensis*
Aren.amaz-58	St.8	Non plume	Haplosclerida	Callyspongiidae	*Arenosclera amazonensis*
Aren.klau-757	St. 10	Non plume	Haplosclerida	Callyspongiidae	*Arenosclera klausis*
Cin.kuek-62	St.3	Plume	Tetractinellida	Tetillidae	*Cinachyrella kuekenthali*
Cin.kuek-62-2	St.3	Plume	Tetractinellida	Tetillidae	*Cinachyrella kuekenthali*
Coe.alcol-61-2	St.3	Plume	Poecilosclerida	Isodictyidae	*Coelocarteria alcoladoi sp. nov.*
Coe.alcol-61-1	St.3	Plume	Poecilosclerida	Isodictyidae	*Coelocarteria alcoladoi sp. nov.*
Coe.amad-63-2	St.3	Plume	Poecilosclerida	Isodictyidae	*Coelocarteria amadoi sp. nov.*
Coe.amad-63	St.3	Plume	Poecilosclerida	Isodictyidae	*Coelocarteria amadoi sp. nov.*
Coe.amad-44	St.6	Non plume	Poecilosclerida	Isodictyidae	*Coelocarteria amadoi sp. nov.*
Coe.amad-44-2	St.6	Non plume	Poecilosclerida	Isodictyidae	*Coelocarteria amadoi sp. nov.*
Coe.bart-17	St.6	Non plume	Poecilosclerida	Isodictyidae	*Coelocarteria bartschi*
Geo.cort-39	St.6	Non plume	Tetractinellida	Geodiidae	*Geodia cf. corticostylifera*
Geo.cort-40	St.6	Non plume	Tetractinellida	Geodiidae	*Geodia cf. corticostylifera*
Geo.nep-45	St.6	Non plume	Tetractinellida	Geodiidae	*Geodia neptuni*

Geo-48	St.6	Non plume	Tetractinellida	Geodiidae	*Geodia sp.*
Hyat.cav-56-1	St.1	Plume	Dictyoceratida	Spongiidae	*Hyattella cavernosa*
Hyat.cav-56-2	St.1	Plume	Dictyoceratida	Spongiidae	*Hyattella cavernosa*
Mon.arb-22	St.6	Non plume	Poecilosclerida	Crambeidae	*Monanchora arbuscula*
Mon.arb-47	St.6	Non plume	Poecilosclerida	Crambeidae	*Monanchora arbuscula*
Neo.prox-59	St.3	Plume	Haplosclerida	Petrosiidae	*Neopetrosia proxima*
Neo.prox-60	St.3	Plume	Haplosclerida	Petrosiidae	*Neopetrosia proxima*
Per.fost-18	St.6	Non plume	Bubarida	Dictyonellidae	*Perissinella fosteri*
Per.fost-19	St.6	Non plume	Bubarida	Dictyonellidae	*Perissinella fosteri*
Petro.cit-53-1	St.3	Plume	Bubarida	Desmanthidae	*Petromica citrina*
Petro.cit-53-2	St.3	Plume	Bubarida	Desmanthidae	*Petromica citrina*
Top.ophi-12	St.6	Non plume	Suberitida	Halichondriidae	*Topsentia ophiraphidites*
Top.ophi-20	St.6	Non plume	Suberitida	Halichondriidae	*Topsentia ophiraphidites*
Top.ophi-24	St.6	Non plume	Suberitida	Halichondriidae	*Topsentia ophiraphidites*
Trib.schm-57	St.1	Plume	Tetractinellida	Ancorinidae	*Tribrachium schmidtii*
Trib.schm-57-2	St.1	Plume	Tetractinellida	Ancorinidae	*Tribrachium schmidtii*

4.2.2 DNA Extraction and Metagenomic Sequencing

These steps were described in chapter 2 entitled “Materials and Methods”.

4.2.3 Assembly and Binning of Sequences

The reads from the previous step were submitted to the binning process. The recovery of genomes from metagenomes was performed using metaWRAP v1.0 (URITSKIY; DIRUGGIERO; TAYLOR, 2018) pipeline. First, all metagenomes were co-assembled with MEGAHIT v1.1.3 (LI et al., 2016), then contigs were binned with Metabat2 (KANG et al., 2019), MaxBin2 (WU; SIMMONS; SINGER, 2016) and CONCOCT (ALNEBERG et al., 2014) with default parameters of metaWRAP. The three sets of genomic bins were consolidated into a single bin set by metaWRAP's Bin_refinement module with minimum completion and maximum contamination equal to 50% and 5%, respectively. The completion and contamination were evaluated using CheckM (PARKS et al., 2015). Assembly quality and genomic features of the metagenome-assembled genomes (MAGs) were assessed using softwares QUAST v4.6.3 (GUREVICH et al., 2013) and Prokka v1.14.5 (SEEMANN, 2014).

4.2.4 Taxonomic Assignment and Functional Genomics of MAGs

Taxonomic assignments of MAGs were performed using GTDB-tk (classify_wf module) (CHAUMEIL et al., 2019). Prokaryotic protein coding

sequences (CDSs) in the contigs were called by Prodigal v2.6.3 (TATUSOV et al., 2003). COG (Clusters of Orthologous Groups) metabolic functional categories of the CDSs were predicted using rpsblast from BLAST+ version 2.9.0+ against NCBI's Conserved Domain Database (CDD) (TATUSOV et al., 2003). CDSs annotations with best blast hit and e-value ≤ $1x10^{-5}$ were kept for following analyses. COG classification was also used to predict genes of vitamin B12 synthesis, CRISPR/Cas proteins and Eukaryotic-like proteins (ELPs).

Specific metabolic pathways related to energy production (Nitrogen metabolism and Sulfur metabolism) were predicted using a set of HMMs from https://github.com/banfieldlab/metabolic-hmms, PFAM (MISTRY et al., 2021), TIGRFAM (HAFT et al., 2001) and Fungene (FISH et al., 2013) through software ggHMM v1.1 (https://github.com/banfieldlab/ggHMM). Results were filtered based on an e-value cutoff of $1x10^{-4}$ and bit-score threshold according to the supplementary table (**Appendix 1**). Potential to degrade and transform complex carbohydrates was assessed with dbCAN (YIN et al., 2012) and classified according to the carbohydrate-active enzymes (CAZymes) database.

In addition, all genomes were submitted to IMG/MER (CHEN et al., 2019) for complementary analyses to identify maker genes of carbon fixation pathways (the Wood-Ljungdahl (WL) pathway, reductive citric acid (rTCA) cycle, 3-hydroxypropionate/4-hydroxybutyrate (HP-HB) cycle, 3-hydroxypropionate (3-HP) bicycle, dicarboxylate/4-hydroxybutyrate (DC-HB) cycle, and the Calvin–Benson–Bassham (CBB) cycle) (HÜGLER; SIEVERT, 2011) against the KEGG Orthology (KO) database (KANEHISA et al., 2016).

4.2.5 Statistical Analysis

Non-metric multidimensional scaling (NMDS) based on Bray-Curtis dissimilarity matrix of reads counts of CDSs annotated for each COG functional category in the genomes were conducted through *metaMDS* function through vegan package (DIXON, 2003) in software R (http://www.r-project.org). Clustered heatmaps were generated using seaborn and matplotlib libraries (CASWELL et al., 2020; WASKOM et al., 2017) in Python.

4.3 Results

4.3.1 Genomic characteristics

Metagenomes were analyzed covering 18 GARS sponge species (n=35; 232198141 metagenomic sequences). A total of 115 metagenome assembled genomes were obtained (**Table 8**, **Figure 18**) (completeness: >90%, 107 MAGs; >50%, 08 MAGs. Contamination: <5%) (**Appendix 2**). The genomes varied in size from 0.95 Mbp (GARS044 - *Chloroflexota*) to 5.66 Mbp (GARS017 - *Acidobacteriota*). A large variability of the number of contigs was observed, with values ranging between 89 (GARS106 - *Gammaproteobacteria*) and 1973 (GARS017 - *Acidobacteriota*). The average number of predicted coding sequences (CDSs) in the genomes was 2608.1 ± 967.78 (N=115), of which 1717.17 ± 590.0 (66.68% ± 7.82%) were assigned to COGs. (**Appendix 2**).

Table 8 Taxonomic Affiliation of MAGs. Columns refers to MAGs code and Lineage (Phylum, Class, Order, Family, Genus, Specie.

MAGs	Phylum	Class	Order	Family	Genus	Specie
GARS001	Crenarchaeota	Nitrososphaeria	Nitrososphaerales	Nitrosopumilaceae	Cenarchaeum	Unclassified
GARS002	Crenarchaeota	Nitrososphaeria	Nitrososphaerales	Nitrosopumilaceae	Cenarchaeum	Unclassified
GARS003	Crenarchaeota	Nitrososphaeria	Nitrososphaerales	Nitrosopumilaceae	Cenarchaeum	Unclassified
GARS004	Crenarchaeota	Nitrososphaeria	Nitrososphaerales	Nitrosopumilaceae	Cenarchaeum	Unclassified
GARS005	Crenarchaeota	Nitrososphaeria	Nitrososphaerales	Nitrosopumilaceae	Cenarchaeum	Unclassified
GARS006	Crenarchaeota	Nitrososphaeria	Nitrososphaerales	Nitrosopumilaceae	Cenarchaeum	Unclassified
GARS007	Crenarchaeota	Nitrososphaeria	Nitrososphaerales	Nitrosopumilaceae	Cenarchaeum	Unclassified
GARS008	Crenarchaeota	Nitrososphaeria	Nitrososphaerales	Nitrosopumilaceae	Nitrosopumilus	Unclassified
GARS009	Crenarchaeota	Nitrososphaeria	Nitrososphaerales	Nitrosopumilaceae	Unclassified	Unclassified
GARS010	Crenarchaeota	Nitrososphaeria	Nitrososphaerales	Nitrosopumilaceae	Unclassified	Unclassified
GARS011	Acidobacteriota	Acidobacteriae	Solibacterales	UBA6623	Unclassified	Unclassified
GARS012	Acidobacteriota	Acidobacteriae	Solibacterales	UBA6623	Unclassified	Unclassified
GARS013	Acidobacteriota	Acidobacteriae	Unclassified	Unclassified	Unclassified	Unclassified
GARS014	Acidobacteriota	Acidobacteriae	Unclassified	Unclassified	Unclassified	Unclassified
GARS015	Acidobacteriota	Acidobacteriae	Unclassified	Unclassified	Unclassified	Unclassified
GARS016	Acidobacteriota	Blastocatellia	Unclassified	Unclassified	Unclassified	Unclassified
GARS017	Acidobacteriota	Luteitaleia	Luteitaleales	UBA8438	Unclassified	Unclassified
GARS018	Acidobacteriota	bin61	bin61	bin61	Unclassified	Unclassified
GARS019	Acidobacteriota	bin61	bin61	bin61	bin61	Unclassified
GARS020	Acidobacteriota	bin61	bin61	bin61	bin61	Unclassified
GARS021	Acidobacteriota	bin61	bin61	bin61	bin61	Unclassified
GARS022	Acidobacteriota	bin61	bin61	bin61	bin61	Unclassified
GARS023	Actinobacteriota	Acidimicrobiia	Microtrichales	Bin134	Bin134	Unclassified

GARS024	Actinobacteriota	Acidimicrobiia	Microtrichales	Bin134	Unclassified	Unclassified
GARS025	Actinobacteriota	Acidimicrobiia	Microtrichales	Bin134	Unclassified	Unclassified
GARS026	Actinobacteriota	Acidimicrobiia	Microtrichales	TK06	Unclassified	Unclassified
GARS027	Actinobacteriota	Acidimicrobiia	Microtrichales	UBA11606	Unclassified	Unclassified
GARS028	Actinobacteriota	Acidimicrobiia	bin76	bin76	bin76	Unclassified
GARS029	Bdellovibrionota	Bdellovibrionia	Bdellovibrionales	Unclassified	Unclassified	Unclassified
GARS030	Chloroflexota	Anaerolineae	Caldilineales	Caldilineaceae	bin5	Unclassified
GARS031	Chloroflexota	Anaerolineae	Caldilineales	bin34	Unclassified	Unclassified
GARS032	Chloroflexota	Anaerolineae	Caldilineales	bin34	Unclassified	Unclassified
GARS033	Chloroflexota	Anaerolineae	Caldilineales	bin34	Unclassified	Unclassified
GARS034	Chloroflexota	Anaerolineae	SBR1031	A4b	UBA6055	Unclassified
GARS035	Chloroflexota	Dehalococcoidia	SAR202	UBA11138	Bin90	Unclassified
GARS036	Chloroflexota	Dehalococcoidia	SAR202	UBA11138	Bin90	Unclassified
GARS037	Chloroflexota	Dehalococcoidia	SAR202	UBA11138	Bin90	Unclassified
GARS038	Chloroflexota	Dehalococcoidia	SAR202	UBA11138	Unclassified	Unclassified
GARS039	Chloroflexota	Dehalococcoidia	UBA1151	Unclassified	Unclassified	Unclassified
GARS040	Chloroflexota	Dehalococcoidia	UBA1151	bin127	Unclassified	Unclassified
GARS041	Chloroflexota	Dehalococcoidia	UBA1151	bin127	bin127	Unclassified
GARS042	Chloroflexota	Dehalococcoidia	UBA1151	bin127	bin127	Unclassified
GARS043	Chloroflexota	Dehalococcoidia	UBA1151	bin127	bin127	Unclassified
GARS044	Chloroflexota	Dehalococcoidia	UBA2963	Unclassified	Unclassified	Unclassified
GARS045	Chloroflexota	Dehalococcoidia	UBA2991	UBA2991	Unclassified	Unclassified
GARS046	Chloroflexota	Dehalococcoidia	UBA3495	UBA3495	Bin22	Unclassified
GARS047	Chloroflexota	Dehalococcoidia	UBA3495	UBA3495	Bin87	Unclassified
GARS048	Chloroflexota	Dehalococcoidia	UBA3495	UBA3495	Bin87	Unclassified
GARS049	Chloroflexota	Dehalococcoidia	UBA3495	UBA3495	Bin87	Unclassified
GARS050	Chloroflexota	Dehalococcoidia	UBA3495	UBA3495	Unclassified	Unclassified
GARS051	Chloroflexota	Dehalococcoidia	UBA3495	UBA3495	Unclassified	Unclassified
GARS052	Chloroflexota	Dehalococcoidia	UBA3495	UBA3495	Unclassified	Unclassified
GARS053	Chloroflexota	Dehalococcoidia	bin125	Unclassified	Unclassified	Unclassified
GARS054	Chloroflexota	Dehalococcoidia	bin125	bin125	bin125	Unclassified
GARS055	Chloroflexota	Dehalococcoidia	bin125	bin125	bin125	Unclassified
GARS056	Chloroflexota	Dehalococcoidia	bin125	bin125	bin125	Unclassified
GARS057	Dadabacteria	UBA1144	UBA2774	Unclassified	Unclassified	Unclassified
GARS058	Dadabacteria	UBA1144	UBA2774	Unclassified	Unclassified	Unclassified
GARS059	Dadabacteria	UBA1144	UBA2774	Unclassified	Unclassified	Unclassified
GARS060	Dadabacteria	UBA1144	UBA2774	Unclassified	Unclassified	Unclassified
GARS061	Dadabacteria	UBA1144	UBA2774	Unclassified	Unclassified	Unclassified
GARS062	Dadabacteria	UBA1144	UBA2774	Unclassified	Unclassified	Unclassified
GARS063	Dadabacteria	UBA1144	Unclassified	Unclassified	Unclassified	Unclassified
GARS064	Deinococcota	Deinococci	Deinococcales	Trueperaceae	Truepera	GCA_002239005.1
GARS065	Entotheonellota	Entotheonellia	Entotheonellales	Entotheonellaceae	Unclassified	Unclassified

GARS066	Entotheonellota	Entotheonellia	Entotheonellales	Unclassified	Unclassified	Unclassified
GARS067	Gemmatimonadota	Gemmatimonadetes	SG8-23	BD2-11	BD2-11	Unclassified
GARS068	Gemmatimonadota	Gemmatimonadetes	SG8-23	BD2-11	bin94	Unclassified
GARS069	Gemmatimonadota	Gemmatimonadetes	SG8-23	UBA6960	Unclassified	Unclassified
GARS070	Latescibacterota	UBA2968	UBA2968	GCA-2709665	Unclassified	Unclassified
GARS071	Latescibacterota	UBA2968	UBA2968	GCA-2709665	Unclassified	Unclassified
GARS072	Latescibacterota	UBA2968	UBA2968	GCA-2709665	Unclassified	Unclassified
GARS073	Latescibacterota	UBA2968	UBA2968	UBA2968	Unclassified	Unclassified
GARS074	Latescibacterota	UBA2968	UBA2968	UBA2968	Unclassified	Unclassified
GARS075	Latescibacterota	UBA2968	UBA2968	UBA2968	Unclassified	Unclassified
GARS076	Latescibacterota	UBA2968	UBA8231	GCA-002724215	GCA-2724215	Unclassified
GARS077	Latescibacterota	Unclassified	Unclassified	Unclassified	Unclassified	Unclassified
GARS078	Latescibacterota	Unclassified	Unclassified	Unclassified	Unclassified	Unclassified
GARS079	Latescibacterota	Unclassified	Unclassified	Unclassified	Unclassified	Unclassified
GARS080	Latescibacterota	Unclassified	Unclassified	Unclassified	Unclassified	Unclassified
GARS081	Nitrospirota	Nitrospiria	Nitrospirales	UBA8639	bin75	Unclassified
GARS082	Nitrospirota	Nitrospiria	Nitrospirales	UBA8639	bin75	Unclassified
GARS083	Poribacteria	WGA-4E	WGA-4E	Unclassified	Unclassified	Unclassified
GARS084	Poribacteria	WGA-4E	WGA-4E	Unclassified	Unclassified	Unclassified
GARS085	Poribacteria	WGA-4E	WGA-4E	WGA-3G	Unclassified	Unclassified
GARS086	Poribacteria	WGA-4E	WGA-4E	WGA-3G	Unclassified	Unclassified
GARS087	Poribacteria	WGA-4E	WGA-4E	WGA-3G	Unclassified	Unclassified
GARS088	Poribacteria	WGA-4E	WGA-4E	WGA-3G	Unclassified	Unclassified
GARS089	Poribacteria	WGA-4E	WGA-4E	WGA-3G	Unclassified	Unclassified
GARS090	Poribacteria	WGA-4E	WGA-4E	WGA-3G	WGA-3G	Unclassified
GARS091	Poribacteria	WGA-4E	WGA-4E	WGA-3G	WGA-3G	Unclassified
GARS092	Poribacteria	WGA-4E	WGA-4E	WGA-3G	WGA-3G	Unclassified
GARS093	Poribacteria	WGA-4E	WGA-4E	WGA-3G	WGA-3G	Unclassified
GARS094	Poribacteria	WGA-4E	WGA-4E	WGA-3G	WGA-3G	Unclassified
GARS095	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Unclassified	Unclassified
GARS096	Proteobacteria	Alphaproteobacteria	Sneathiellales	UBA2966	Unclassified	Unclassified
GARS097	Proteobacteria	Alphaproteobacteria	UBA7887	Unclassified	Unclassified	Unclassified
GARS098	Proteobacteria	Gammaproteobacteria	AqS2	AqS2	Unclassified	Unclassified
GARS099	Proteobacteria	Gammaproteobacteria	AqS2	AqS2	Unclassified	Unclassified
GARS100	Proteobacteria	Gammaproteobacteria	AqS2	Unclassified	Unclassified	Unclassified
GARS101	Proteobacteria	Gammaproteobacteria	AqS2	Unclassified	Unclassified	Unclassified
GARS102	Proteobacteria	Gammaproteobacteria	Pseudomonadales	HTCC2089	bin55	Unclassified
GARS103	Proteobacteria	Gammaproteobacteria	Pseudomonadales	HTCC2089	bin55	Unclassified
GARS104	Proteobacteria	Gammaproteobacteria	Pseudomonadales	HTCC2089	bin55	Unclassified
GARS105	Proteobacteria	Gammaproteobacteria	Thiotrichales	Unclassified	Unclassified	Unclassified
GARS106	Proteobacteria	Gammaproteobacteria	TsSOB	Unclassified	Unclassified	Unclassified
GARS107	Proteobacteria	Gammaproteobacteria	TsSOB	Unclassified	Unclassified	Unclassified
GARS108	Proteobacteria	Gammaproteobacteria	TsSOB	Unclassified	Unclassified	Unclassified

GARS109	Proteobacteria	Gammaproteobacteria	UBA10353	Unclassified	Unclassified	Unclassified
GARS110	Proteobacteria	Gammaproteobacteria	UBA10353	Unclassified	Unclassified	Unclassified
GARS111	Proteobacteria	Gammaproteobacteria	UBA10353	Unclassified	Unclassified	Unclassified
GARS112	Proteobacteria	Gammaproteobacteria	UBA11654	UBA11654	Unclassified	Unclassified
GARS113	Spirochaetota	Spirochaetia	Spirochaetales	RBG-16-67-19	Unclassified	Unclassified
GARS114	UBP10	GR-WP33-30	bin18	bin18	Unclassified	Unclassified
GARS115	UBP10	GR-WP33-30	bin18	bin18	Bin 18	GCA_002238415.1

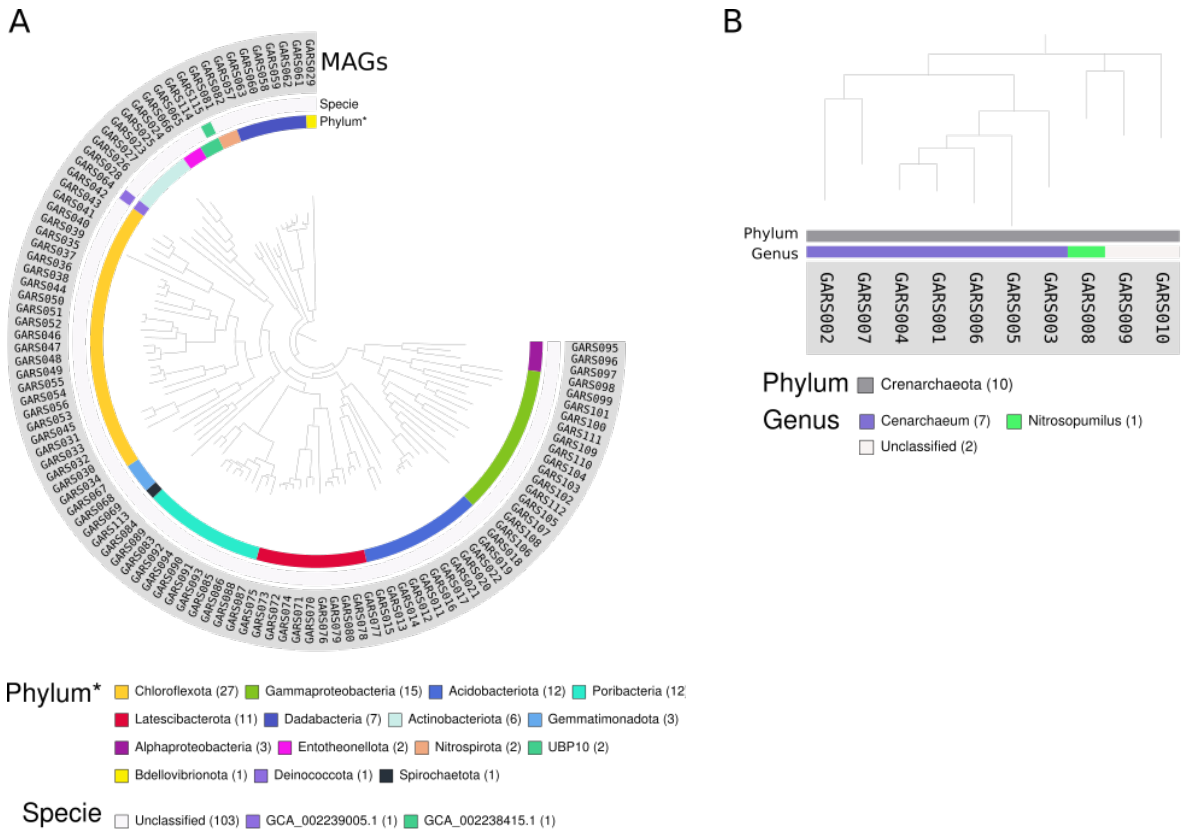

Figure 18 Assessment of phylogenetic and taxonomic diversity of MAGs. A: 105 Bacterial MAGs, B: 10 Archaeal MAGs.

4.3.2 Taxonomic affiliation of sponge symbionts

The 115 obtained genomes belong to *Chloroflexota* (*Dehalococcoidia*, N=22; *Anaerolineae*, N=5), *Proteobacteria* (*Gammaproteobacteria*, N=15; *Alphaproteobacteria*, N=3), *Acidobacteriota* (N=12), *Poribacteria* (N=12), *Latescibacterota* (N=11), *Dadabacteria* (N=7), *Actinobacteriota* (*Acidimicrobiia* , N=6), *Gemmatimonadota* (N=3), *UBP10* (N=2), *Nitrospirota* (N=2), *Entotheonellota* (N=2), *Spirochaetota* (N=1), *Deinococcota* (N=1), *Bdellovibrionota*

(N=1), and *Crenarchaeota* (N=10), family *Nitrosopumilaceae* (generally included in the phylum *Thaumarchaeota*) (**Figure 18**, **Figure 19**).

Of the 11 MAGs were assigned as *Latescibacteriota*, 7 are part of order *UBA2968* and 4 of them probably form a new class.

Two MAGs had GTDB-tk matches to MAGs recovered from other projects: GARS064 matched MAG GCA_002239005.1, which appears in the GenBank database classified as *Truepera* sp. and GARS115, which matched GCA_002238415.1 appears in the GenBank database classified as a member of the *Desulfurellaceae* family. Both of these other MAGs were recovered from metagenome data from sponge species *Aplysina aerophoba*, obtained in Marine Biology Station Piran, Slovenia. Following the nomenclature suggested by (SETUBAL, 2021), these results mean that two of the 115 MAGs are conserved hypothetical MAGs, and the rest are hypothetical MAGs; and none is a SMAG (a MAG for which a species can be assigned).

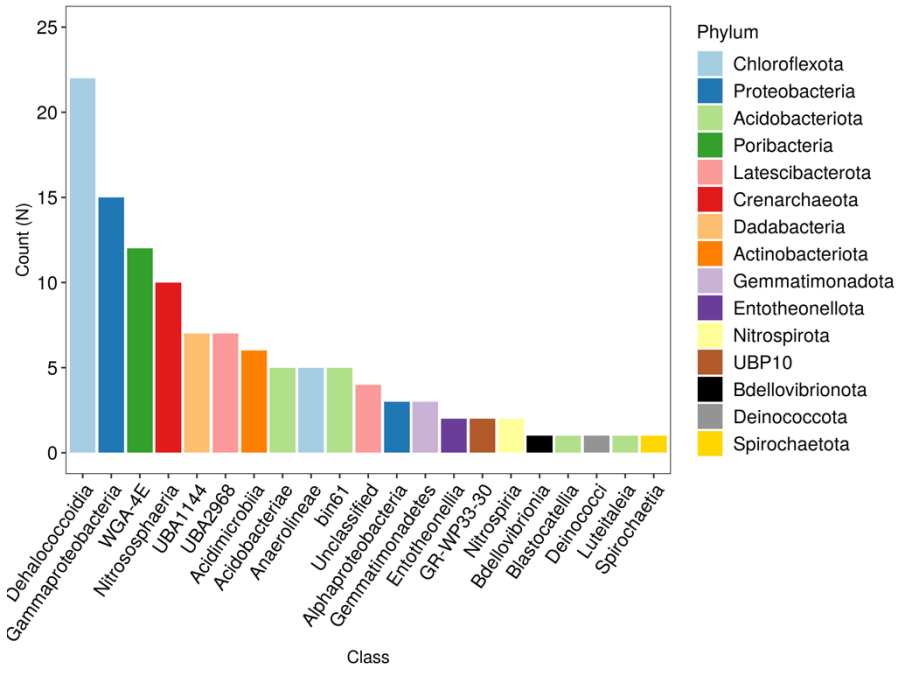

Figure 19 Bar plot of MAGs counts by taxonomic class. The bar colors represent the phylum of the class.

4.3.3 Major metabolic features of sponge symbionts

In order to evaluate the functional diversity, we analyzed the metabolic profile of sponge-associated microbial genomes using Cluster of Orthologous Groups (COGs). The relative abundance of functional categories of COGs showed

that genomes were grouped by taxonomic affiliation (**Figure 20A).** The same distribution pattern was observed for the COGs category Energy production and Conversion [C], mainly because archaeal genomes (Crenarchaeota:Nitrosopumilaceae) displayed a distinct metabolic profile (**Figure 20B**).

Carbon Fixation

To identify potential autotroph microbial genomes, we searched for gene markers of six carbon fixation pathways: Wood-Ljungdahl (WL) pathway, reductive citric acid cycle (rTCA), 3-hydroxypropionate bicycle (3-HB), 3-hydroxypropionate/4-hydroxybutyrate cycle (3HP-4HB), dicarboxylate/4-hydroxybutyrate cycle (DC-HB) and Calvin–Benson–Bassham cycle (CBB). Out of the six, three pathways were identified in 19 putative autotrophic genomes (**Figure 21**): reverse TCA (rTCA), 3-Hydroxypropionate bicycle (3HB) and 3-hydroxypropionate/4-hydroxybutyrate (3-HP/4HB). The 3HB genes markers malyl-CoA/(S)-citramalyl-CoA lyase (*mcl*), 2-methylfumaryl-CoA isomerase (*mct*) and 3-methylfumaryl-CoA hydratase (*meh*) were identified in two genomes of *Alphaproteobacteria* (GARS95, GARS96) and one *Chloroflexota* genome (GARS054). In addition, the genes encoding CO2-fixing enzymes acetyl-CoA/propionyl-CoA carboxylases (*accC* and *accA*), required for 3HB cycle and 3-HP/4HB cycle were also predicted in these three genomes (**Figure 21**). A total of 13 genomes showed capability of fixing carbon through 3-HP/4HB cycle: two *Acidobacteriota* (GARS018 and GARS019), two *Gammaproteobacteria* (GARS109 and GARS110), and all archaeal genomes, except one (GARS007). All of them contained the acetyl-CoA/propionyl-CoA carboxylases genes (archaeal: *accC* and *accB*, bacterial: *accC* and *accA*), and other key genes involved in the 3-HP/4HB cycle, such as 4-hydroxybutyryl-CoA dehydratase/vinylacetyl-CoA-Delta-isomerase (*abfD*), enoyl-CoA hydratase/ 3-hydroxyacyl-CoA dehydrogenase (*hcdh*) and acetyl-CoA C-acetyltransferase (*atoB*). The two *Nitrospirota* genomes (GARS81, GARS82) and one *Acidobacteriota* genome (GARS016) encoded key genes for the rTCA cycle (*aclA*, *aclB*) (**Figure 21**).

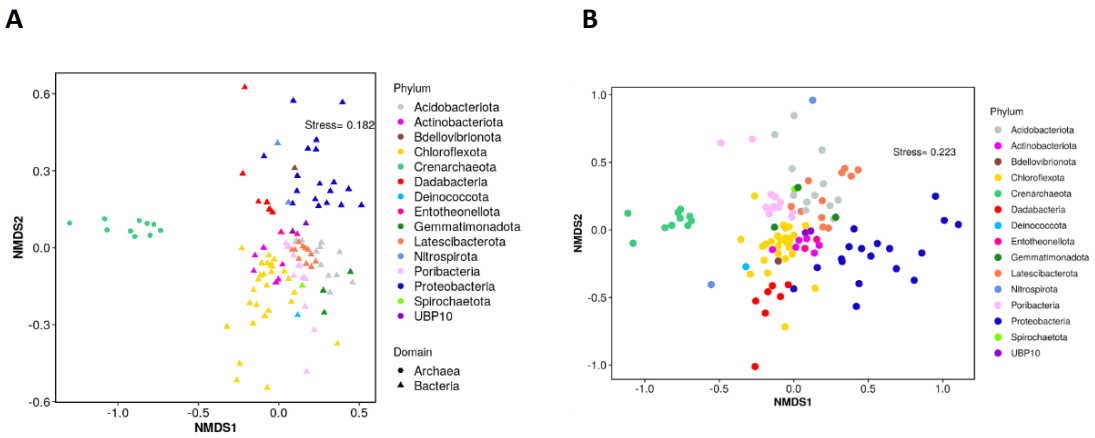

Figure 20 Non-metric multidimensional scaling (NMDS) based on Bray-Curtis dissimilarity matrix of CDSs abundance annotated for each COG functional category in the 115 MAGs. A: nMDS based on abundance of COGs Functional Categories. B: nMDS based on COGs of category Energy_production_Conversion (C). The colors represent the phylum.

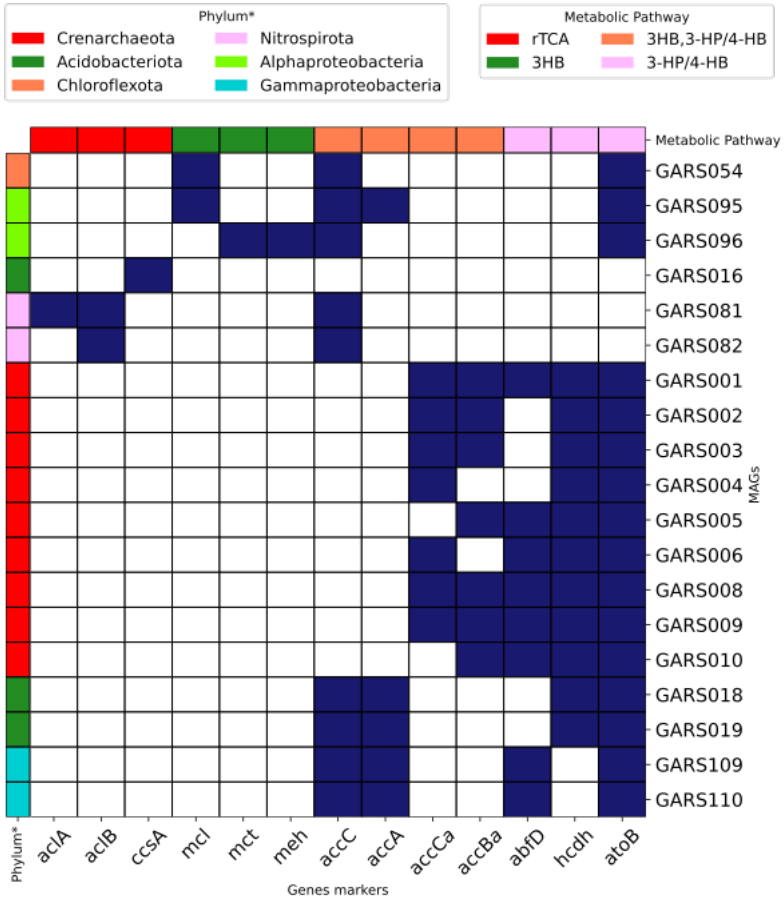

Figure 21 Graph of presence (dark blue) and absence (white) of genes (KEGG-KO) related to Carbon Fixation Pathways predicted in MAGs. The side color bar corresponds to the phyla of each MAG, and the upper color bar to the metabolic pathway of the of the corresponding gene.

Nitrification

A total of 15 MAGs have potential for nitrification (**Figure 22**). This autotrophic process requires significant ammounts of ammonia to efficiently produce nitrite and nitrate. For the first step of nitrification, 11 MAGs encodes ammonia oxidation key marker genes (*amoA* and *amoC*): 9 *Crenarchaeota* (GARS001, GARS002, GARS003, GARS004, GARS005, GARS007, GARS008, GARS009, GARS010) presented *amoA* gene and 2 *UBP10* phylum (GARS114, GARS115) both genes (**Figure 22**). For the second step of nitrification, 4 MAGs presented nitrite oxidation genes *nxrA* and *nxrB*: 1 *Alphaproteobacteria* (GARS095), 1 *Gammaproteobacteria* (GARS111) and 2 *Nitrospirota* (GARS081, GARS082) (**Figure 22**). All these 4 putative nitrite-oxidizing bacterial genomes also showed denitrification genes (**Figure 22**).

Denitrification

Denitrification is a metabolic pathway for nitrogen removal carried out in four steps, in which nitrate or nitrite is reduced to nitrous oxide or dinitrogen gases. No genome presented the genes markers required for the complete process. The nitrate reductase genes *narGH* and *napA* necessary for the first step (e.g. reduction of nitrate to nitrite) were predicted in 5 bacterial genomes: 2 *Gammaproteobacteria* (GARS099, GARS111), 1 *Alphaproteobacteria* (GARS095) and 1 *Chloroflexota* (GARS043) (**Figure 22**). The genes *narG* and *narH* were identified in these 3 proteobacterial genomes and GARS043 only had predicted gene *narG*. In addition, nitrate reductase gene *napA* were predicted in one *Gammaproteobacteria* genome (GARS100) (**Figure 22**). Nitrite reductase genes (*nirK* and *nirS*) required for the second step (e.g. reduction of nitrite to nitric oxide) were searched in all genomes. The gene *nirK* was identified in 9 genomes: 4 *Chloroflexota* (GARS043, GARS047, GARS050, GARS051), 1 *Dadabacteria* (GARS057), 2 nitrite-oxidizing bacterial genomes from *Nitrospirota* (GARS081, GARS082), 1 *Alphaproteobacteria* (GARS097) and 1 *Gammaproteobacteria* (GARS110) (**Figure 22**). The *Chloroflexota* genome GARS043 was the only that presented the gene key markers for the two early denitrification steps. The gene *nirS* was lacking in all genomes.

Nineteen MAGs have key marker genes of the third denitrification step (e.g nitric oxide into nitrous oxide). The nitric oxide reductase genes *norBC* were both encoded by 3 genomes: 1 *Poribacteria* (GARS083), 1 *Gammaproteobacteria* (GARS103) and 1 *UBP10* (GARS115) that is also an AOB (ammonia oxidizing bacteria) (**Figure 22**). Others present one gene: 7 *Poribacteria* (GARS084, GARS085, GARS086, GARS087, GARS088, GARS089, GARS090), 4 *Acidobacteriota* (GARS011, GARS012, GARS014, GARS019), 2 *Gammaproteobacteria* (GARS104, GARS105), 2 *Entotheonellota* (GARS065, GARS066) and 1 *Spirochaetota* (GARS113) (**Figure 22**). For the final denitrification step (e.g. nitrous oxide to dinitrogen) 2 *Poribacteria* genomes (GARS083, GARS085) presented the marker genes *nosZ* and *nosD*. These 2 *Poribacteria* genomes have the coding genes enzymes for the two final steps of denitrification (**Figure 22**). All symbiotic genomes lacked nitrogen fixation encoding genes.

Sulfur metabolism

A total of 8 putative sulfur-oxidizing bacteria (SOB) genomes were predicted by identification of two enzymatic pathways: thiosulfate oxidation (*sox* enzymes complex) and *dsrAB* enzymes (dissimilatory (bi)sulfite reductase) (**Figure 22**). The first one was searched by identification of key marker genes *soxBCY* from the *sox* enzyme complex and the second one by *dsrAB* genes. One *Actinobacteria* genome (GARS026) and 3 *Gammaproteobacteria* genomes (GARS109, GARS110, GARS111) encode the *soxB* key gene, of which GARS109, GARS110 also present the *soxY* gene (**Figure 22**). In addition, the *dsrAB* key genes were detected in 4 other genomes of *Gammaproteobacteria*: GARS105, GARS106, GARS107 and GARS108. All these 8 putative SOB genomes, except for GARS108 also had predicted the *tauD* gene (taurine dioxygenase) required for the utilization of taurine as a sulfur source (**Figure 22**).

Metabolism of Urea, Carnitine and Taurine

A prediction of the potential sources of carbon, nitrogen and sulfur that microbial symbionts can make use of for metabolism and energy were evaluated. An enrichment of genes related to utilization of Urea and Carnitine as a source of nitrogen was observed in 23 (20%) and 27 (23.48%) microbial genomes,

respectively (**Figure 22**). The urease encoding genes *ureABC* were predicted in 8 bacterial phyla (15 MAGs) and archaeal phylum *Crenarchaeota* (8 MAGs). These include part of the putative genomes predicted as AOA (7 *Crenarchaeota*) and all AOB (2 *UBP10*). Carnitine transporter gene *bcct* was also identified in 8 bacterial phyla (27 MAGs), but not in archaeal genomes (**Figure 22**). Moreover 51 (44.35%) genomes encodes taurine dioxygenase gene (*tauD*), suggesting that taurine is an important source of sulfur (**Figure 22**).

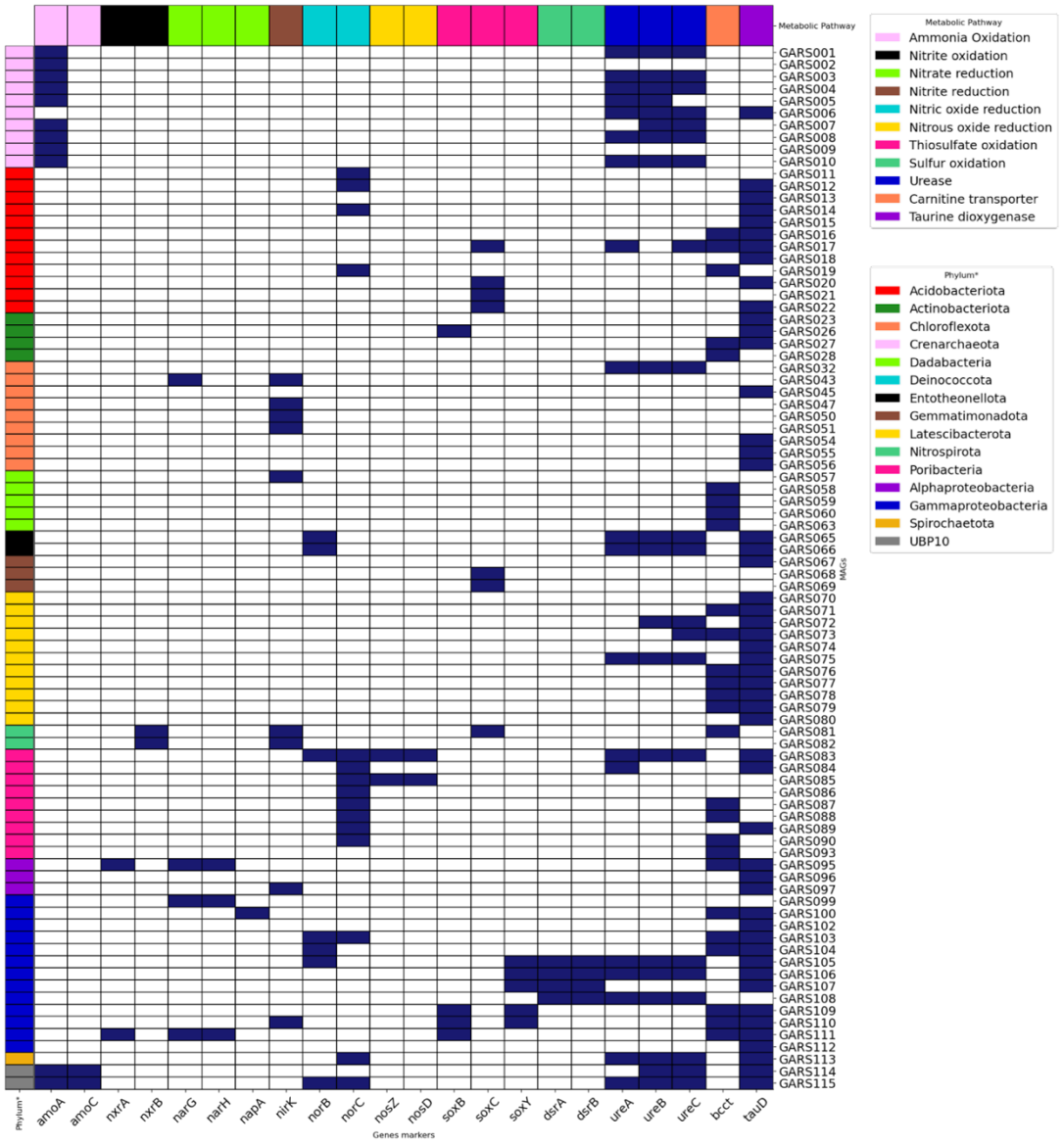

Figure 22 Graph of presence (dark blue) and absence (white) of genes (HMM) related to metabolism of Nitrogen and Sulfur in MAGs. The side color bar corresponds to the phyla of each MAG, and the upper color bar to the metabolic pathway of the of the corresponding gene.

Sponge Microbial Symbiont Features

Vitamins B

A total of 67 (58.26%) microbial genomes of 14 phyla showed a potential to synthesize vitamin B12 (**Figure 23**). The two key enzymes responsible for vitamin B12 synthesis uroporphyrinogen-III synthase (COG1587) and ATP:corrinoid adenosyltransferase/Cob(I)alamin adenosyltransferase (COG2109) were detected in 51 and 40 microbial genomes, respectively. Most of the genomes that encode these genes belong to *Chloroflexota* (N=22), which corresponds to 81.48% of genomes from this phylum. In addition, all 10 genomes of *Crenarchaeota* also presented this potential, as well as all *Actinobacteriota* (N=6, 100%), *Entotheonellota* (N=2,100%), *UBP10* (N=2, 100%), *Deinococcota* (GARS064), *Spirochaetota* (GARS113), and also 8 out of 15 *Gammaproteobacteria* genomes (53.33%), *Acidobacteriota* (N=5), *Latescibacterota* (N=5), *Dadabacteria* (N=2), *Alphaproteobacteria* (GARS095), *Nitrospirota* (GARS082) and 1 genome of *Poribacteria* out of 12 (GARS087). Only genomes of phyla *Gemmatimonadota* and *Bdellovibrionota* lacked genes encoding for vitamin B12 synthesis (**Figure 23**).

Eukaryotic-like proteins (ELPs)

Eukaryotic-like repeat motifs proteins allow the protein-protein interactions between host and symbiotic microbes, playing a crucial role in host symbiont recognition. An enrichment of genes encoding for eukaryotic-like repeated domain proteins were observed in 110 genomes (95.65%) (**Figure 23**). The Tetratricopeptide (TPR) repeat (COG0457) was the most abundant repeated domains, detected in 95 (82.61%) microbial genomes. Moreover, the Ankyrin repeat (COG0666), Leucine-rich repeat (LRR) (COG4886), WD40 repeat (COG2319) and also TPR repeat (COG0790) were detected in 75 (65.22%), 69 (60.00%), 47 (40.87%) and 38 (33.04%) microbial genomes, respectively. Only 5 genomes lacked genes that encode for eukaryotic-like repeat motifs proteins: 2 *Actinobacteriota* (GARS023, GARS024), 2 *Chloroflexota* of class *Dehalococcoidia* (GARS053, GARS054) and 1 *Dadabacteria* (GARS061) (**Figure 23**).

Utilization of Amazon Forest nutrients by sponge symbionts

Carbohydrate-active enzymes (CAZymes) are responsible for the synthesis, degradation and modification of all carbohydrates molecules. The

CAZyme classes glycoside hydrolases (GHs) and Carbohydrate esterases (CEs) were identified in all microbial phyla and vast majority of MAGs (N=108 (93.91%)).

The Amazon Forest exports great amounts of lignocellulolytic materials to the ocean. Genes responsible for the degradation of lignocellulose are more abundant in the genomes of *Chloroflexota* (gene families CE7 and CE15). CE7 family was detected in 25 MAGs with predominance in the phylum *Chloroflexota* (18 of 27 MAGs), followed by *Latescibacterota* (N=3), *Acidobacteriota* (N=1), *Entotheonellota* (N=1), *Gammaproteobacteria* (N=1) and *Poribacteria* (N=1). In addition, the CE15 family was found in 4 MAGs: *Latescibacterota* (N=2), *Acidobacteriota* (N=1) and *Spirochaetota* (N=1).

CAZymes families GH33 and GH88 related to carbohydrates found within sponge tissue were detected in 39 and 23 MAGs, respectively. GH33 family were found in all *Poribacteria* genomes (N=12) and also in *Latescibacterota* (N=10), *Acidobacteriota* (N=7), *Chloroflexota* (N=6), *Actinobacteriota* (N=2), *Alphaproteobacteria* (N=1) and *Spirochaetota* (N=1). In addition GH88 family, that acts on glycosaminoglycans were enriched in *Poribacteria* (N=8), *Latescibacterota* (N=7), *Chloroflexota* (N=4), *Acidobacteriota* (N=2), *Alphaproteobacteria* (N=1) and *Spirochaetota* (N=1).

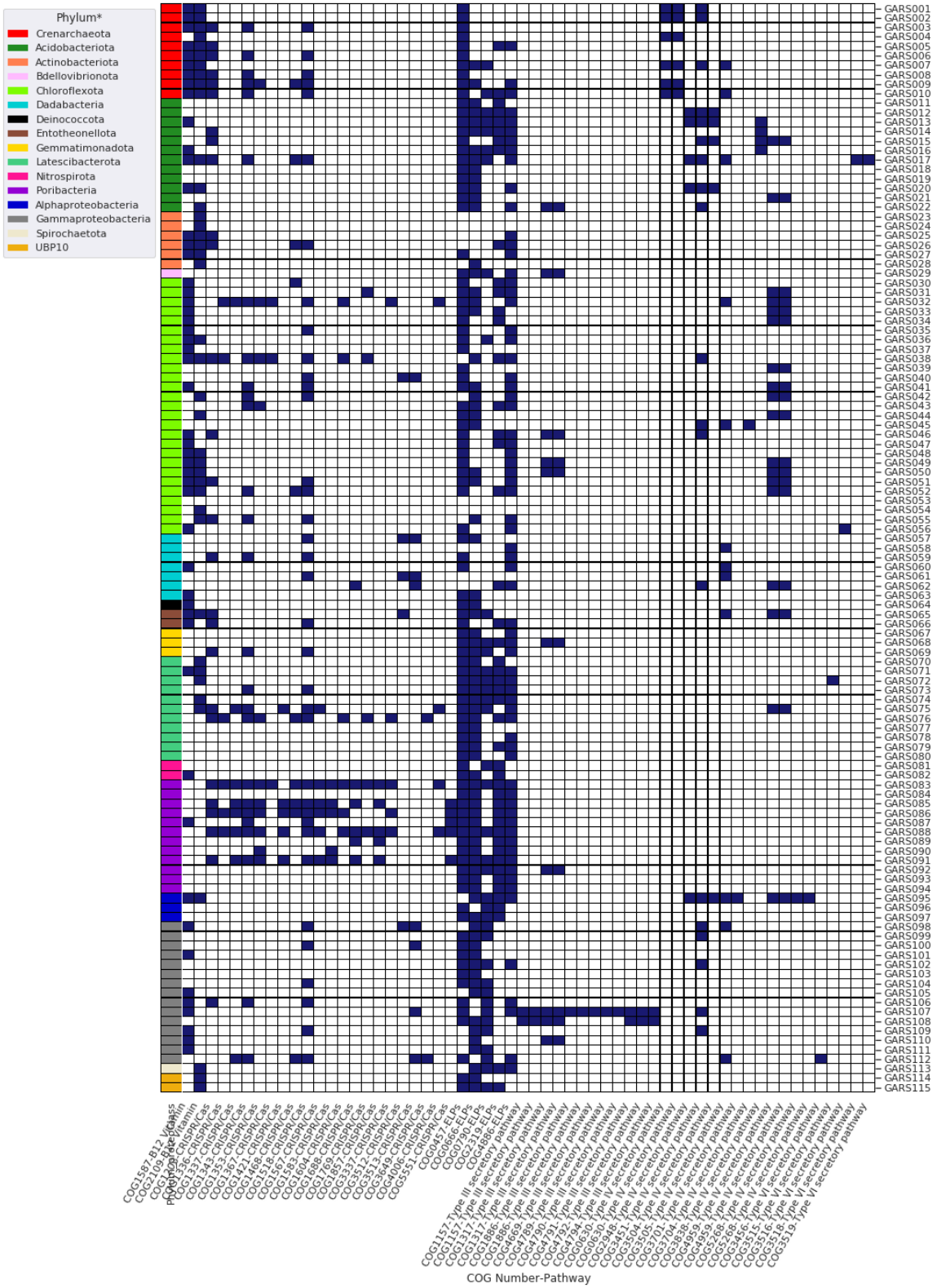

Figure 23 Graph of presence (dark blue) and absence (white) of COGs related to Symbiotic features in MAGs. The side color bar corresponds to the phyla of each MAG.

4.4 Discussion

In this study we showed the metabolic diversity of microbial genomes recovered from sponge metagenomes from Great Amazon Reef System.

Major metabolic features of sponge symbionts

Archaeal MAGs

The Archaea MAGs have genes for vitamin B12 synthesis and CRISPR/Cas 3 category and Eukaryotic-like proteins (ELPs). Taken together, the data indicate that the archaea present in GARS sponges play a key role in the sponge holobionts in different ways, through nutrition, defense, and cell-host interactions.

Bacterial MAGs

MAGs assigned to *Actinobacteriota, Chloroflexota* and *Entotheonellota, Acidobacteriota, Latescibacteriota, Gammaproteobacteria,* and *UBP10* also have the potential for B12 vitamin production. All bacterial genomes have eukaryotic like protein (ELPs), but *Poribacteria* stand out with an enrichment of ELPs (Ankyrin repeat, WD40 , Leucine-rich repeat (LRR) protein, Tetratricopeptide TPR repeat, TPR repeat) as a possible consequence of host-microbe co-evolution (MOITINHO-SILVA et al., 2017a). *Poribacteria* also stand out for the abundance of genes related to the CRISPR/Cas system, another sign of co-evolution for host protection (HORN et al., 2016). These genetic elements may confer protection against a wide variety of mobile genetic elements present in the seawater (e.g. viruses, transposable elements, and plasmids) (HORN et al., 2016; SLABY et al., 2017). *Latescibacteriota* and *Chloroflexota* also contribute to the CRISP/Cas pool in the sponges.

Guilds of Nutrition Specialization

Genomes assigned to the phylum *Latescibacterota* and *Chloroflexota*, showed metabolisms linked to the degradation of xylan polymers (CE15 - glucuronyl esterases and CE7 - acetyl-xylan esterases). These enzymes are involved in lignocellulose and hemicellulose degradation and were initially characterized in terrestrial plants (HETTIARACHCHI et al., 2019) and have already been found in sponge-associated bacteria (ROBBINS et al., 2021). *Latescibacteriota* and *Chloroflexota* stood out in terms of abundance of pathways

related to xylan degradation in these bacteria may indicate that the dissolved organic matter present in the Amazon River plume, especially that of plant origin from terrestrial plants, plays an important role in the nutrition of sponges (BAYER et al., 2018; DE MENEZES et al., 2022).

MAGs with potential to chemoautotrophy, such as Ammonia oxidizing archaea, Ammonia Oxidizing Bacteria (AOB) and Sulfur oxidizing Bacteria (SOB) were recovered from sponge metagenomes. A relevant contribution to the nutrition of the sponge hosts appears to come these autotrophic microorganisms, as they contribute to the nutrition of the sponge host through primary production (THOMAS et al., 2016).

All 10 recovered archaeal genomes are nitrifying, and 8 showed urea degradation genes. Urea is an excretion product of sponge metabolism, which serves as an important source of ammonia (ZHANG et al., 2019). The use of urea is also important, as a source of nitrogen and carbon, for other symbiotic microorganisms, as the breakdown of urea generates CO_2 and ammonia (SU et al., 2013).

Many genomes of Gammaproteobacterial seems to be SOB (sulfur oxidizing bacteria). This group of symbionts oxidize reduced sulfur compounds produced by sulfate-reducing bacteria (SRB) (HOFFMANN et al., 2005). Studies with genomes retrieved from sponge metagenomes described new taxa of SOBs and showed that these genomes are present in several orders of sponges (LAVY et al., 2018).

4.5 Final Remarks

In this chapter we showed the taxonomic and functional diversity of microbial genomes recovered from metagenomes of marine sponge from Great Amazon Reef System (GARS). These genomes displayed a wide functional repertoire and seem to be vital to survive of sponge host in the environment of Amazon Reefs.

4.6 Publication

The results of this chapter are in an ongoing manuscript with Mayanne A.M. de Freitas, Fabiano L. Thompson and João Carlos Setubal.

5 Conclusions

In this thesis we studied the diversity of microbial community associated with sponges from Great Amazon Reef System.

Using shotgun metagenomics data, we explored the sponge microbiomes from Great Amazon Reef System and showed that these sponges possess a core microbiome, sharing a predominant group of microbial taxa that are consistently present in all sponge microbiomes. Besides the core microbiome, specific taxa were recognized as indicators of LMA and HMA sponges and results indicated that microbiomes of sponge from GARS is species-specific.

In chapter 4, the analyses indicated that microbes associated with these sponges are able to benefit from the nutrients present in the plume of the Amazon River also demonstrated in (DE MENEZES et al., 2022). We also found that chemolithotrophs microbes seem to be relevant to sponge host survival.

Further studies with the total microbial community based on functional analysis may be conducted to better understand the interactions symbiotic microbes and sponge host. Together with more analyses of genomes recovered from sponge metagenomes are required to understand the influence of host identity and environmental specificity on nutrition specialization of sponge microbial symbionts from GARS.

Contributions of this thesis:

1. Analyzes of the microbial community associated with sponges from GARS extended the scarce biological knowledge of the Brazilian Amazon Continental Shelf.

2. Through the prospecting of genomes from sponge, we could recover new microbial genomes. These genomes revealed a wide functional repertoire displaying the metabolic diversity of sponge microbiomes from GARS.

3. The sponge microbiomes from GARS seem to be vital to survive of sponge host in the environment of Amazon Reefs.

6 References

AHN, A.-C. et al. Genomic diversity within the haloalkaliphilic genus Thioalkalivibrio. **PloS one**, v. 12, n. 3, p. e0173517, 10 Mar. 2017.

ALNEBERG, J. et al. Binning metagenomic contigs by coverage and composition. **Nature methods**, v. 11, n. 11, p. 1144–1146, Nov. 2014.

ANANTHARAMAN, K. et al. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. **Nature communications**, v. 7, n. 1, p. 13219, 24 Oct. 2016.

BANHA, T. N. S. et al. **The Great Amazon Reef System: A fact**. **Frontiers in Marine Science**, 2022. Disponível em: <http://dx.doi.org/10.3389/fmars.2022.1088956>

BAYER, K. et al. Marine Sponges as Chloroflexi Hot Spots: Genomic Insights and High-Resolution Visualization of an Abundant and Diverse Symbiotic Clade. **mSystems**, v. 3, n. 6, 26 Dec. 2018.

BELL, J. J. **The functional roles of marine sponges**. **Estuarine, Coastal and Shelf Science**, 2008. Disponível em: <http://dx.doi.org/10.1016/j.ecss.2008.05.002>

BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics** , v. 30, n. 15, p. 2114–2120, 1 Aug. 2014.

BUSCH, K. et al. Biodiversity, environmental drivers, and sustainability of the global deep-sea sponge microbiome. **Nature communications**, v. 13, n. 1, p. 5160, 2 Sep. 2022.

CASWELL, T. A. et al. **matplotlib/matplotlib v3.1.3**. [s.l: s.n.].

CHAUMEIL, P.-A. et al. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. **Bioinformatics** , 15 Nov. 2019.

CHEN, I.-M. A. et al. IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. **Nucleic acids research**, v. 47, n. D1, p. D666–D677, 8 Jan. 2019.

CLEARY, D. F. R. et al. The sponge microbiome within the greater coral reef microbial metacommunity. **Nature communications**, v. 10, n. 1, p. 1644, 9 Apr. 2019.

COLES, V. J. et al. **The pathways and properties of the Amazon River Plume in the tropical North Atlantic Ocean**. **Journal of Geophysical Research: Oceans**, 2013. Disponível em: <http://dx.doi.org/10.1002/2013jc008981>

DAGG, M. et al. **Transformation of dissolved and particulate materials on continental shelves influenced by large rivers: plume processes**. **Continental Shelf Research**, 2004. Disponível em:

<http://dx.doi.org/10.1016/j.csr.2004.02.003>

DE MAHIQUES, M. M. et al. Insights on the evolution of the living Great Amazon Reef System, equatorial West Atlantic. **Scientific reports**, v. 9, n. 1, p. 1–8, 23 Sep. 2019.

DE MENEZES, T. A. et al. Fluxes of the Amazon River plume nutrients and microbes into marine sponges. **The Science of the total environment**, v. 847, n. 157474, p. 157474, 15 Nov. 2022.

DE MICHEAUX, P. L.; DROUILHET, R.; LIQUET, B. **The R Software: Fundamentals of Programming and Statistical Analysis**. [s.l.] Springer Science & Business, 2014.

DIXON, P. **VEGAN, a package of R functions for community ecology**. **Journal of Vegetation Science**, 2003. Disponível em: <http://dx.doi.org/10.1111/j.1654-1103.2003.tb02228.x>

FAN, L. et al. Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. **Proceedings of the National Academy of Sciences of the United States of America**, v. 109, n. 27, p. E1878-87, 3 Jul. 2012.

FISH, J. A. et al. **FunGene: the functional gene pipeline and repository**. **Frontiers in Microbiology**, 2013. Disponível em: <http://dx.doi.org/10.3389/fmicb.2013.00291>

FRANCINI-FILHO, R. B. et al. Perspectives on the Great Amazon Reef: Extension, Biodiversity, and Threats. **Frontiers in Marine Science**, v. 5, 2018.

FREEMAN, C. J. et al. Sponge–microbe interactions on coral reefs: Multiple evolutionary solutions to a complex environment. **Frontiers in marine science**, v. 8, 20 Jul. 2021.

GARCIA, G. D. et al. Metagenomic analysis of healthy and white plague-affected Mussismilia braziliensis corals. **Microbial ecology**, v. 65, n. 4, p. 1076–1086, May 2013.

GILES, E. C. et al. Bacterial community profiles in low microbial abundance sponges. **FEMS microbiology ecology**, v. 83, n. 1, p. 232–241, Jan. 2013.

GLOECKNER, V. et al. The HMA-LMA dichotomy revisited: an electron microscopical survey of 56 sponge species. **The Biological bulletin**, v. 227, n. 1, p. 78–88, Aug. 2014.

GOEBEL, N. L. et al. **Abundance and distribution of major groups of diazotrophic cyanobacteria and their potential contribution to N2 fixation in the tropical Atlantic Ocean**. **Environmental Microbiology**, 2010. Disponível em: <http://dx.doi.org/10.1111/j.1462-2920.2010.02303.x>

GOEIJ, J. M. DE et al. **Surviving in a Marine Desert: The Sponge Loop Retains Resources Within Coral Reefs**. **Science**, 2013. Disponível em:

<http://dx.doi.org/10.1126/science.1241981>

GOES, J. I. et al. **Influence of the Amazon River discharge on the biogeography of phytoplankton communities in the western tropical north Atlantic**. **Progress in Oceanography**, 2014. Disponível em: <http://dx.doi.org/10.1016/j.pocean.2013.07.010>

GUREVICH, A. et al. QUAST: quality assessment tool for genome assemblies. **Bioinformatics** , v. 29, n. 8, p. 1072–1075, 15 Apr. 2013.

HADAS, E.; SHPIGEL, M.; ILAN, M. Particulate organic matter as a food source for a coral reef sponge. **The Journal of experimental biology**, v. 212, n. Pt 22, p. 3643–3650, Nov. 2009.

HADJU, E.; PEIXINHO, S.; FERNANDEZ, J. C. C. **Esponjas marinhas da Bahia: guia de campo e laboratório**. [s.l: s.n.].

HAFT, D. H. et al. TIGRFAMs: a protein family resource for the functional identification of proteins. **Nucleic acids research**, v. 29, n. 1, p. 41–43, 1 Jan. 2001.

HENTSCHEL, U. et al. Molecular evidence for a uniform microbial community in sponges from different oceans. **Applied and environmental microbiology**, v. 68, n. 9, p. 4431–4440, Sep. 2002.

HENTSCHEL, U. et al. Genomic insights into the marine sponge microbiome. **Nature reviews. Microbiology**, v. 10, n. 9, p. 641–654, Sep. 2012.

HETTIARACHCHI, S. A. et al. Characterization of an acetyl xylan esterase from the marine bacterium Ochrovirga pacifica and its synergism with xylanase on beechwood xylan. **Microbial cell factories**, v. 18, n. 1, p. 122, 8 Jul. 2019.

HILTON, J. A. et al. Metatranscriptomics of N2-fixing cyanobacteria in the Amazon River plume. **The ISME journal**, v. 9, n. 7, p. 1557–1569, Jul. 2015.

HOFFMANN, F. et al. **An Anaerobic World in Sponges**. **Geomicrobiology Journal**, 2005. Disponível em: <http://dx.doi.org/10.1080/01490450590922505>

HOFFMANN, F. et al. Complex nitrogen cycling in the sponge Geodia barretti. **Environmental microbiology**, v. 11, n. 9, p. 2228–2243, Sep. 2009.

HORN, H. et al. An Enrichment of CRISPR and Other Defense-Related Features in Marine Sponge-Associated Microbial Metagenomes. **Frontiers in microbiology**, v. 7, p. 1751, 8 Nov. 2016.

HUG, L. A. et al. Overview of organohalide-respiring bacteria and a proposal for a classification system for reductive dehalogenases. **Philosophical transactions of the Royal Society of London. Series B, Biological sciences**, v. 368, n. 1616, p. 20120322, 19 Apr. 2013.

HÜGLER, M.; SIEVERT, S. M. Beyond the Calvin cycle: autotrophic carbon fixation in the ocean. **Annual review of marine science**, v. 3, p. 261–289, 2011.

INDRANINGRAT, A. A. G.; SMIDT, H.; SIPKEMA, D. Bioprospecting Sponge-Associated Microbes for Antimicrobial Compounds. **Marine drugs**, v. 14, n. 5, 2 May 2016.

KANEHISA, M. et al. KEGG as a reference resource for gene and protein annotation. **Nucleic acids research**, v. 44, n. D1, p. D457-62, 4 Jan. 2016.

KANG, D. D. et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. **PeerJ**, v. 7, p. e7359, 26 Jul. 2019.

KARIMI, E. et al. Metagenomic binning reveals versatile nutrient cycling and distinct adaptive features in alphaproteobacterial symbionts of marine sponges. **FEMS microbiology ecology**, v. 94, n. 6, 1 Jun. 2018.

KINEKE, G. C. et al. **Fluid-mud processes on the Amazon continental shelf**. **Continental Shelf Research**, 1996. Disponível em: <http://dx.doi.org/10.1016/0278-4343(95)00050-x>

LAVY, A. et al. A novel Chromatiales bacterium is a potential sulfide oxidizer in multiple orders of marine sponges. **Environmental microbiology**, v. 20, n. 2, p. 800–814, Feb. 2018.

LEAL, C. V. et al. **Integrative Taxonomy of Amazon Reefs' Arenosclera spp.: A New Clade in the Haplosclerida (Demospongiae)**. **Frontiers in Marine Science**, 2017. Disponível em: <http://dx.doi.org/10.3389/fmars.2017.00291>

LI, D. et al. MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. **Methods** , v. 102, p. 3–11, 1 Jun. 2016.

LI, H.; DURBIN, R. Fast and accurate short read alignment with Burrows-Wheeler transform. **Bioinformatics** , v. 25, n. 14, p. 1754–1760, 15 Jul. 2009.

MALDONADO, M.; RIBES, M.; VAN DUYL, F. C. Nutrient fluxes through sponges: biology, budgets, and ecological implications. **Advances in marine biology**, v. 62, p. 113–182, 2012.

MEHBUB, M. F. et al. Marine sponge derived natural products between 2001 and 2010: trends and opportunities for discovery of bioactives. **Marine drugs**, v. 12, n. 8, p. 4539–4577, 19 Aug. 2014.

MISTRY, J. et al. Pfam: The protein families database in 2021. **Nucleic acids research**, v. 49, n. D1, p. D412–D419, 8 Jan. 2021.

MOHAMED, N. M. et al. Diversity of aerobic and anaerobic ammonia-oxidizing bacteria in marine sponges. **The ISME journal**, v. 4, n. 1, p. 38–48, Jan. 2010.

MOITINHO-SILVA, L. et al. Integrated metabolism in sponge–microbe symbiosis revealed by genome-centered metatranscriptomics. **The ISME journal**, v. 11, n. 7, p. 1651–1666, 24 Mar. 2017a.

MOITINHO-SILVA, L. et al. Predicting the HMA-LMA Status in Marine Sponges by Machine Learning. **Frontiers in microbiology**, v. 8, p. 752, 8 May 2017b.

MOLLERI, G. S. F.; DE M. NOVO, E. M. L.; KAMPEL, M. **Space-time variability of the Amazon River plume based on satellite ocean color**. **Continental Shelf Research**, 2010. Disponível em: <http://dx.doi.org/10.1016/j.csr.2009.11.015>

MOURA, R. L. et al. An extensive reef system at the Amazon River mouth. **Science advances**, v. 2, n. 4, p. e1501252, Apr. 2016.

OMACHI, C. Y. et al. **Light availability for reef-building organisms in a plume-influenced shelf**. **Continental Shelf Research**, 2019. Disponível em: <http://dx.doi.org/10.1016/j.csr.2019.05.005>

OREN, A. The Family Ectothiorhodospiraceae. In: **The Prokaryotes**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014. p. 199–222.

PARKS, D. H. et al. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. **Genome research**, v. 25, n. 7, p. 1043–1055, Jul. 2015.

PITA, L. et al. The sponge holobiont in a changing ocean: from microbes to ecosystems. **Microbiome**, v. 6, n. 1, Dec. 2018.

RADAX, R. et al. Ammonia-oxidizing archaea as main drivers of nitrification in cold-water sponges. **Environmental microbiology**, v. 14, n. 4, p. 909–923, Apr. 2012.

REVEILLAUD, J. et al. Host-specificity among abundant and rare taxa in the sponge microbiome. **The ISME journal**, v. 8, n. 6, p. 1198–1209, Jun. 2014.

RIBES, M. et al. **Functional convergence of microbes associated with temperate marine sponges**. **Environmental Microbiology**, 2012. Disponível em: <http://dx.doi.org/10.1111/j.1462-2920.2012.02701.x>

RIBES, M. et al. Microbial Diversity and Putative Diazotrophy in High- and Low-Microbial-Abundance Mediterranean Sponges. **Applied and environmental microbiology**, v. 81, n. 17, p. 5683–5693, 1 Sep. 2015.

ROBBINS, S. J. et al. A genomic view of the microbiome of coral reef demosponges. **The ISME journal**, v. 15, n. 6, p. 1641–1654, Jun. 2021.

SATINSKY, B. M. et al. Microspatial gene expression patterns in the Amazon River Plume. **Proceedings of the National Academy of Sciences of the United States of America**, v. 111, n. 30, p. 11085–11090, 29 Jul. 2014.

SATINSKY, B. M. et al. Metagenomic and metatranscriptomic inventories of the lower Amazon River, May 2011. **Microbiome**, v. 3, p. 39, 10 Sep. 2015.

SCHLÄPPY, M.-L. et al. Evidence of nitrification and denitrification in high and low microbial abundance sponges. **Marine biology**, v. 157, n. 3, p. 593–602,

2010.

SCHMITT, S. et al. Assessing the complex sponge microbiota: core, variable and species-specific bacterial communities in marine sponges. **The ISME journal**, v. 6, n. 3, p. 564–576, Mar. 2012.

SEEMANN, T. Prokka: rapid prokaryotic genome annotation. **Bioinformatics** , v. 30, n. 14, p. 2068–2069, 15 Jul. 2014.

SETUBAL, J. C. Metagenome-assembled genomes: concepts, analogies, and challenges. **Biophysical reviews**, v. 13, n. 6, p. 905–909, Dec. 2021.

SIEGL, A. et al. Single-cell genomics reveals the lifestyle of Poribacteria, a candidate phylum symbiotically associated with marine sponges. **The ISME journal**, v. 5, n. 1, p. 61–70, Jan. 2011.

SILVEIRA, C. B. et al. Microbial and sponge loops modify fish production in phase-shifting coral reefs. **Environmental microbiology**, v. 17, n. 10, p. 3832–3846, Oct. 2015.

SIMISTER, R. L. et al. **Sponge-specific clusters revisited: a comprehensive phylogeny of sponge-associated microorganisms**. **Environmental Microbiology**, 2012. Disponível em: <http://dx.doi.org/10.1111/j.1462-2920.2011.02664.x>

SLABY, B. M. et al. Metagenomic binning of a marine sponge microbiome reveals unity in defense but metabolic specialization. **The ISME journal**, v. 11, n. 11, p. 2465–2478, Nov. 2017.

SU, J. et al. Phylogenetically diverse ureC genes and their expression suggest the urea utilization by bacterial symbionts in marine sponge Xestospongia testudinaria. **PloS one**, v. 8, n. 5, p. e64848, 31 May 2013.

SUBRAMANIAM, A. et al. Amazon River enhances diazotrophy and carbon sequestration in the tropical North Atlantic Ocean. **Proceedings of the National Academy of Sciences of the United States of America**, v. 105, n. 30, p. 10460–10465, 29 Jul. 2008.

SUTTLE, C. A. **Marine viruses — major players in the global ecosystem**. **Nature Reviews Microbiology**, 2007. Disponível em: <http://dx.doi.org/10.1038/nrmicro1750>

TATUSOV, R. L. et al. The COG database: an updated version includes eukaryotes. **BMC bioinformatics**, v. 4, p. 41, 11 Sep. 2003.

TAYLOR, M. W. et al. Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. **Microbiology and molecular biology reviews: MMBR**, v. 71, n. 2, p. 295–347, Jun. 2007.

TAYLOR, M. W. et al. “Sponge-specific” bacteria are widespread (but rare) in diverse marine environments. **The ISME journal**, v. 7, n. 2, p. 438–443, Feb. 2013.

THOMAS, T. et al. Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. **The ISME journal**, v. 4, n. 12, p. 1557–1567, Dec. 2010.

THOMAS, T. et al. Diversity, structure and convergent evolution of the global sponge microbiome. **Nature communications**, v. 7, p. 11870, 16 Jun. 2016.

THOMAS, T. R. A.; KAVLEKAR, D. P.; LOKABHARATHI, P. A. **Marine Drugs from Sponge-Microbe Association—A Review**. **Marine Drugs**, 2010. Disponível em: <http://dx.doi.org/10.3390/md8041417>

TIAN, R.-M. et al. Effect of copper treatment on the composition and function of the bacterial community in the sponge Haliclona cymaeformis. **mBio**, v. 5, n. 6, p. e01980, 4 Nov. 2014.

TRINDADE-SILVA, A. E. et al. Taxonomic and functional microbial signatures of the endemic marine sponge Arenosclera brasiliensis. **PloS one**, v. 7, n. 7, p. e39905, 2 Jul. 2012.

URITSKIY, G. V.; DIRUGGIERO, J.; TAYLOR, J. MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis. **Microbiome**, v. 6, n. 1, p. 158, 15 Sep. 2018.

VOGEL, S. Current-induced flow through living sponges in nature. **Proceedings of the National Academy of Sciences of the United States of America**, v. 74, n. 5, p. 2069–2071, May 1977.

WASKOM, M. et al. **mwaskom/seaborn: v0.8.1 (September 2017)**. [s.l: s.n.].

WEBSTER, N. S.; HILL, R. T. **The culturable microbial community of the Great Barrier Reef sponge Rhopaloeides odorabile is dominated by an α-Proteobacterium**. **Marine Biology**, 2001. Disponível em: <http://dx.doi.org/10.1007/s002270000503>

WEBSTER, N. S.; TAYLOR, M. W. Marine sponges and their microbial symbionts: love and other relationships. **Environmental microbiology**, v. 14, n. 2, p. 335–346, Feb. 2012.

WEBSTER, N. S.; THOMAS, T. **The Sponge Hologenome**. **mBio**, 2016. Disponível em: <http://dx.doi.org/10.1128/mbio.00135-16>

WEISZ, J. B. et al. **Linking abundance and diversity of sponge-associated microbial communities to metabolic differences in host sponges**. **Marine Biology**, 2007. Disponível em: <http://dx.doi.org/10.1007/s00227-007-0708-y>

WEISZ, J. B.; LINDQUIST, N.; MARTENS, C. S. Do associated microbial abundances impact marine demosponge pumping rates and tissue densities? **Oecologia**, v. 155, n. 2, p. 367–376, Mar. 2008.

WICKHAM, H. **Ggplot2**. 2. ed. Basel, Switzerland: Springer International Publishing, 2016.

WOOD, D. E.; LU, J.; LANGMEAD, B. Improved metagenomic analysis with Kraken 2. **Genome biology**, v. 20, n. 1, p. 257, 28 Nov. 2019.

WU, Y.-W.; SIMMONS, B. A.; SINGER, S. W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. **Bioinformatics** , v. 32, n. 4, p. 605–607, 15 Feb. 2016.

YANG, Y. et al. Roles of organohalide-respiring Dehalococcoidia in carbon cycling. **mSystems**, v. 5, n. 3, 9 Jun. 2020.

YIN, Y. et al. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. **Nucleic acids research**, v. 40, n. Web Server issue, p. W445-51, Jul. 2012.

YIN, Z. et al. Sponge grade body fossil with cellular resolution dating 60 Myr before the Cambrian. **Proceedings of the National Academy of Sciences of the United States of America**, v. 112, n. 12, p. E1453-60, 24 Mar. 2015.

ZHANG, F. et al. Phosphorus sequestration in the form of polyphosphate by microbial symbionts in marine sponges. **Proceedings of the National Academy of Sciences of the United States of America**, v. 112, n. 14, p. 4381–4386, 7 Apr. 2015.

ZHANG, S. et al. **Comparative Genomics Reveals Ecological and Evolutionary Insights into Sponge-Associated Thaumarchaeota**. **mSystems**, 2019. Disponível em: <http://dx.doi.org/10.1128/msystems.00288-19>

Appendix 1

Appendix 1 List of HMMs used to specific metabolic pathways related to energy production. Custom HMMs are from (ANANTHARAMAN et al., 2016) and are available in https://github.com/banfieldlab/metabolic-hmms.

Metabolism	Process/Gene type	Gene	ID	Hmm database	Cutoff Score
Sulfur	Sulfide oxidation	fcc	PF09242	Pfam	266
Sulfur	Sulfide oxidation	Sqr	Custom	Custom	300
Sulfur	Sulfite reduction	dsrD	PF08679	Pfam	50
Sulfur	Sulfur oxidation	dsrA	TIGR02064	TIGRfam	223
Sulfur	Sulfur oxidation	dsrB	TIGR02066	TIGRfam	205
Sulfur	Sulfur oxidation	sor	PF07682	Pfam	300
Sulfur	Sulfur oxidation	sulfur dioxygenase	Custom	Custom	120
Sulfur	thiosulfate oxidation	soxB	TIGR04486	TIGRfam	375
Sulfur	thiosulfate oxidation	soxY	TIGR04488	TIGRfam	125
Sulfur	thiosulfate oxidation	soxC	TIGR04555	TIGRfam	330
Sulfur	Sulfate reduction	asrA	TIGR02910	TIGRfam	223
Sulfur	Sulfate reduction	asrB	TIGR02911	TIGRfam	226
Sulfur	Sulfate reduction	asrC	TIGR02912	TIGRfam	194
Sulfur	Sulfate reduction	aprA	TIGR02061	TIGRfam	326
Sulfur	Sulfate reduction	sat	TIGR00339	TIGRfam	181
Sulfur	Sulfate reduction	cysC	TIGR00455	TIGRfam	133
Sulfur	Sulfate reduction	cysN	TIGR02034	TIGRfam	327
Sulfur	Thiosulfate disproportionation	phsA	Custom	Custom	323
Nitrogen	N2 fixation	Fe only alpha	TIGR01861	TIGRfam	873
Nitrogen	N2 fixation	Fe only beta	TIGR02931	TIGRfam	853
Nitrogen	N2 fixation	Fe only delta	TIGR02929	TIGRfam	124
Nitrogen	N2 fixation	Mo-Fe type Alpha	TIGR01282	TIGRfam	503
Nitrogen	N2 fixation	Mo-Fe type beta	TIGR01286	TIGRfam	414
Nitrogen	N2 fixation	Vanadium-type alpha	TIGR01860	TIGRfam	822
Nitrogen	N2 fixation	Vanadium-type beta	TIGR02932	TIGRfam	821
Nitrogen	N2 fixation	Vanadium-type delta	TIGR02930	TIGRfam	122
Nitrogen	N2 fixation	nifH	TIGR01287	TIGRfam	261
Nitrogen	Nitrite oxidation	nxrA	Custom	Custom	370
Nitrogen	Nitrite oxidation	nxrB	Custom	Custom	252
Nitrogen	Nitrate reduction	napA	TIGR01706	TIGRfam	472
Nitrogen	Nitrate reduction	napB	PF03892	Pfam	23.9
Nitrogen	Nitrate reduction	narG	TIGR01580	TIGRfam	600
Nitrogen	Nitrate reduction	narH	TIGR01660	TIGRfam	348
Nitrogen	Nitrite reduction	cyt c552 small NrfH	TIGR03153	TIGRfam	75
Nitrogen	Nitrite reduction	cyt c552 large NrFA	PF02335	Pfam	57
Nitrogen	Nitrite reduction	cyt c552 large NrFA	PF02335/TIGR03152	Pfam/TIGRfam	576
Nitrogen	Nitrite reduction	nrfD	TIGR03148	TIGRfam	300
Nitrogen	Nitrite reduction	NirB	TIGR02374	TIGRfam	442
Nitrogen	Nitrite reduction	NirD	TIGR02378	TIGRfam	69
Nitrogen	Nitrite reduction	nirK	TIGR02376	TIGRfam	169
Nitrogen	Nitrite reduction	nirS	Custom	Custom	200
Nitrogen	Nitric oxide reduction	norB	Custom	Custom	79
Nitrogen	Nitric oxide reduction	norC	Custom	Custom	50
Nitrogen	Nitrous oxide reduction	nosD	PF05048	Pfam	290
Nitrogen	Nitrous oxide reduction	nosD	TIGR04247	TIGRfam	290
Nitrogen	Nitrous oxide reduction	nosZ	TIGR04246	TIGRfam	550
Nitrogen	Nitrous oxide reduction	nosZ_tat	TIGR04244	TIGRfam	-
Nitrogen	Nitrous oxide reduction	nosZ_sec	TIGR04246	TIGRfam	-
Nitrogen	Anammox	hzoA	Custom	Custom	325
Nitrogen	Anammox	hzsA	Custom	Custom	466

Nitrogen	Octaheme c-type cytochrome Shewanella-type	octR	TIGR04315	TIGRfam	325
Carbon fixation	CBB cycle - Rubisco	Form I	Custom	Custom	500
Carbon fixation	CBB cycle - Rubisco	Form II	Custom	Custom	500
Carbon fixation	CBB cycle - Rubisco	Form II/III	Custom	Custom	500
Carbon fixation	CBB cycle - Rubisco	Form III	Custom	Custom	450
Carbon fixation	3 Hydroxypropionate cycle	propionyl-CoA synthase	Custom	Custom	-
Carbon fixation	4 Hydroxypropionate cycle	malonyl-CoA reductase	Custom	Custom	-
Carbon fixation	3HP/4HB	4-hydroxybutyryl-CoA synthetase	Custom	Custom	-
Carbon fixation	3HP/4HB	4-hydroxybutyryl-CoA dehydratase	Custom	Custom	-
Carbon fixation	Wood Ljungdahl pathway	codhD (delta)	TIGR00381	TIGRfam	197
Carbon fixation	Wood Ljungdahl pathway	codhC (beta)	TIGR00316	TIGRfam	355
Carbon fixation	Wood Ljungdahl pathway	codh (catalytic)	TIGR01702	TIGRfam	210
Carbon fixation	Reverse TCA cycle	aclA	Custom	Custom	215
Carbon fixation	Reverse TCA cycle	aclB	Custom	Custom	177
Urea	Urease	ureC	TIGR01792	TIGRfam	212
Urea	Urease	ureB	TIGR00192	TIGRfam	40
Urea	Urease	ureA	TIGR00193	TIGRfam	31

Appendix 2

Appendix 2 Genomic features of MAGs: Completeness(%), Contamination (%), GC (%), Genome size (Mbp), Contigs (N), N50, Genes (N), rRNA, tRNA, 16S rRNA, tmRNA, CDSs (N) and CDSs assigned to COGs (N)(%.

MAGs	Completeness (%)	Contamination (%)	GC (%)	Genome size (Mbp)	Contigs (N)	N50	Genes (N)	rRNA (N)	tRNA (N)	5S rRNA (N)	16S rRNA (N)	5.8S rRNA (N)	5S rRNA (partial) (N)	tmRNA (N)	16S rRNA (partial) (N)	CDSs (N)	CDSs assigned to COGs (N)(%)
GARS001	94.17	1.94	59.9	1.77	457	5530	2153	1	39	1	0	0	0	0	0	2149	1188 (55.28)
GARS002	93.44	0.97	67.8	1.43	222	8470	1644	2	37	1	1	0	0	0	0	1648	902 (54.73)
GARS003	88.83	4.37	47.4	1.49	288	6464	1708	0	32	0	0	0	0	0	0	1852	1012 (54.64)
GARS004	76.78	0.48	59.5	1.57	730	2373	1962	2	26	1	0	1	0	0	0	1935	987 (51.01)
GARS005	73.2	1.29	48.6	1.32	353	4074	1468	0	26	0	0	0	0	0	0	1616	823 (50.93)
GARS006	69.74	1.85	62.9	1.11	414	2791	1466	1	35	1	0	0	0	0	0	1409	836 (59.33)
GARS007	62.13	1.46	65.3	1.31	645	2144	1857	0	11	0	0	0	0	0	0	1462	763 (52.19)
GARS008	97.08	0.0	46.0	2.1	165	16063	2425	1	42	1	0	0	0	0	0	2511	1173 (46.71)
GARS009	93.65	0.4	49.8	02.02	866	2688	2398	3	30	2	0	0	1	0	0	2718	1193 (43.89)
GARS010	89.32	0.0	45.9	1.71	643	3252	1835	0	24	0	0	0	0	0	0	2094	1034 (49.38)
GARS011	56.15	0.91	66.4	2.28	608	3887	1972	0	15	0	0	0	0	0	0	2169	1353 (62.38)
GARS012	51.62	2.56	62.5	3.69	1185	3331	3586	0	19	0	0	0	0	1	0	3863	1989 (51.49)
GARS013	81.53	4.27	62.4	3.27	936	3861	2996	1	26	1	0	0	0	0	0	3380	2243 (66.36)
GARS014	67.95	0.5	63.1	2.39	878	2873	2196	1	16	1	0	0	0	0	0	2501	1537 (61.46)
GARS015	50.25	1.71	62.6	1.87	959	2027	1796	0	14	0	0	0	0	0	0	2262	1414 (62.51)
GARS016	73.78	0.5	63.3	2.78	967	3097	2633	1	22	1	0	0	0	1	0	3070	1946 (63.39)
GARS017	71.86	3.58	69.0	5.66	1973	3054	5971	0	42	0	0	0	0	0	0	6784	3442 (50.74)
GARS018	87.8	2.68	67.1	2.53	578	5427	2256	1	40	1	0	0	0	0	0	2428	1720 (70.84)
GARS019	77.4	1.76	68.6	2.56	420	6906	2413	1	37	1	0	0	0	0	0	2568	1750 (68.15)
GARS020	67.32	0.92	69.5	2.69	786	3739	2568	1	22	1	0	0	0	0	0	2875	1910 (66.43)
GARS021	58.59	1.88	70.1	1.61	453	3717	1463	0	19	0	0	0	0	0	0	1666	1150 (69.03)
GARS022	55.24	0.91	68.5	1.8	696	2631	1813	0	16	0	0	0	0	0	0	2079	1319 (63.44)
GARS023	62.44	1.92	65.8	1.93	628	3261	2065	1	25	1	0	0	0	1	0	2334	1722 (73.78)
GARS024	79.81	1.8	69.6	2.81	620	5365	2823	1	44	1	0	0	0	0	0	3057	2104 (68.83)
GARS025	68.82	2.99	56.2	1.74	781	2486	1695	1	30	1	0	0	0	1	0	2232	1604 (71.86)
GARS026	55.21	3.28	67.3	2.15	819	2676	2362	0	18	0	0	0	0	0	0	2651	1726 (65.11)
GARS027	79.43	3.85	69.6	3.83	595	8766	3835	1	39	1	0	0	0	0	0	4029	2867 (71.16)
GARS028	77.54	3.83	64.9	02.04	649	3408	2172	1	34	1	0	0	0	0	0	2427	1794 (73.92)
GARS029	89.28	0.62	42.9	02.07	461	6140	1759	1	25	1	0	0	0	1	0	1972	1136 (57.61)
GARS030	68.37	1.36	58.9	3.31	1004	3518	3129	1	24	1	0	0	0	0	0	3455	2473 (71.58)
GARS031	84.77	3.74	55.8	3.77	1501	2980	4086	1	32	1	0	0	0	0	0	4394	2827 (64.34)
GARS032	83.08	01.01	61.3	3.77	833	6462	3628	1	37	1	0	0	0	1	0	3924	2499 (63.69)
GARS033	67.86	0.45	56.9	2.69	768	4015	2898	1	33	1	0	0	0	1	0	2955	1991 (67.38)
GARS034	59.34	0.91	59.5	2.8	1305	2406	2976	0	26	0	0	0	0	0	0	3533	2335 (66.09)
GARS035	76.45	4.29	57.9	2.39	488	5333	2464	1	29	1	0	0	0	0	0	2738	1875 (68.48)
GARS036	54.25	3.96	53.3	1.54	676	2649	1601	1	18	1	0	0	0	1	0	2031	1393 (68.59)

GARS037	50.38	0.0	57.3	1.17	410	2947	1278	0	26	0	0	0	0	0	0	1458	1054 (72.29)
GARS038	83.96	0.77	58.8	2.43	907	3281	2514	1	23	1	0	0	0	0	0	2950	2051 (69.53)
GARS039	87.95	0.09	56.5	1.32	153	20217	1215	2	48	1	1	0	0	1	0	1173	862 (73.49)
GARS040	52.85	4.8	54.9	1.7	962	1846	1949	1	17	1	0	0	0	1	0	2343	1362 (58.13)
GARS041	67.21	0.99	58.1	1.61	483	3609	1680	0	32	0	0	0	0	1	0	1850	1346 (72.76)
GARS042	63.96	1.98	55.1	1.98	535	4143	2103	1	40	1	0	0	0	1	0	2301	1494 (64.93)
GARS043	50.55	2.33	54.5	02.08	1299	1569	2562	0	18	0	0	0	0	0	0	3039	1628 (53.57)
GARS044	53.21	0.5	57.9	0.95	232	4368	1014	0	26	0	0	0	0	1	0	1096	824 (75.18)
GARS045	50.51	4.95	70.7	01.05	634	1684	1186	1	7	0	0	0	1	0	0	1536	1059 (68.95)
GARS046	53.51	0.0	60.1	1.78	615	3052	1855	1	15	1	0	0	0	0	0	2224	1463 (65.78)
GARS047	80.76	2.42	60.5	2.43	713	3783	2497	1	35	1	0	0	0	0	0	2797	1941 (69.4)
GARS048	63.41	2.67	62.2	1.95	650	3194	2093	0	20	0	0	0	0	0	0	2394	1603 (66.96)
GARS049	51.54	3.28	61.4	2.29	882	2677	2621	1	29	1	0	0	0	0	0	3002	1886 (62.82)
GARS050	93.72	02.09	62.1	3.19	289	16399	3116	1	42	1	0	0	0	1	0	3241	2313 (71.37)
GARS051	83.38	0.11	60.9	2.74	581	5879	2863	0	35	0	0	0	0	0	0	3104	2147 (69.17)
GARS052	56.21	4.12	61.7	2.24	855	2730	2337	0	22	0	0	0	0	0	0	2757	1869 (67.79)
GARS053	53.74	2.31	66.1	01.01	294	3872	1068	0	8	0	0	0	0	0	0	1177	911 (77.4)
GARS054	82.88	03.08	65.1	2.1	568	4091	2167	0	29	0	0	0	0	2	0	2435	1806 (74.17)
GARS055	75.59	3.46	63.9	2.8	778	3921	2977	0	38	0	0	0	0	1	0	3302	2160 (65.41)
GARS056	50.24	4.95	62.4	1.58	887	1882	1844	0	13	0	0	0	0	1	0	2256	1484 (65.78)
GARS057	91.31	0.84	43.6	1.44	162	11193	1346	1	30	1	0	0	0	0	0	1384	1064 (76.88)
GARS058	86.14	3.39	51.7	1.4	215	8835	1393	1	30	1	0	0	0	0	0	1465	1146 (78.23)
GARS059	83.67	1.83	51.0	1.78	406	5135	1795	1	28	1	0	0	0	1	0	1990	1293 (64.97)
GARS060	68.46	4.25	51.0	1.28	348	4640	1322	1	26	1	0	0	0	0	0	1482	1086 (73.28)
GARS061	60.41	4.31	42.8	0.99	486	2155	926	1	13	1	0	0	0	0	0	1237	886 (71.62)
GARS062	53.48	0.84	43.7	01.02	525	2118	1041	0	18	0	0	0	0	0	0	1377	822 (59.69)
GARS063	83.33	1.68	51.4	1.52	222	8847	1591	1	30	1	0	0	0	0	0	1679	1222 (72.78)
GARS064	64.36	3.25	62.9	1.73	389	5014	1683	1	26	1	0	0	0	1	0	1837	1388 (75.56)
GARS065	68.1	0.0	63.0	3.19	780	4693	3263	1	32	1	0	0	0	0	0	3656	2466 (67.45)
GARS066	53.06	2.99	66.8	1.88	758	2537	1896	0	12	0	0	0	0	1	0	2332	1686 (72.3)
GARS067	50.34	2.36	64.5	1.71	616	2933	1662	1	12	1	0	0	0	0	0	1891	1262 (66.74)
GARS068	56.85	3.85	67.1	2.19	625	3752	2175	0	16	0	0	0	0	1	0	2392	1477 (61.75)
GARS069	50.9	1.2	65.1	1.45	546	2788	1454	0	10	0	0	0	0	1	0	1608	1072 (66.67)
GARS070	80.84	0.1	60.3	3.18	748	4978	2792	2	28	0	0	0	1	0	1	3083	2104 (68.25)
GARS071	77.4	4.61	61.1	3.42	1179	3234	3131	1	20	1	0	0	0	0	0	3683	2486 (67.5)
GARS072	62.18	04.04	51.0	2.96	1608	1897	3000	1	11	1	0	0	0	0	0	3982	2455 (61.65)
GARS073	64.18	0.2	51.1	3.32	1726	1981	3304	1	21	1	0	0	0	0	0	4335	2559 (59.03)
GARS074	59.88	0.2	53.7	2.51	730	3660	2739	0	19	0	0	0	0	1	0	2982	1856 (62.24)
GARS075	54.93	0.98	50.8	3.11	1475	2289	3008	0	15	0	0	0	0	0	0	3920	2383 (60.79)
GARS076	73.49	4.5	59.7	3.76	715	6438	3340	0	31	0	0	0	0	0	0	3659	2236 (61.11)
GARS077	85.25	0.2	58.7	2.74	477	7524	2640	2	33	1	0	0	0	0	1	2756	2131 (77.32)
GARS078	80.5	1.36	63.4	2.75	387	9481	2505	0	32	0	0	0	0	0	0	2660	2061 (77.48)

GARS079	61.21	3.3	64.6	1.96	747	2681	1918	0	16	0	0	0	0	0	0	2215	1637 (73.91)
GARS080	54.86	1.71	61.8	2.44	681	3916	2313	0	12	0	0	0	0	0	0	2648	1907 (72.02)
GARS081	81.43	04.09	57.0	1.64	333	5851	1712	1	20	1	0	0	0	0	0	1839	1340 (72.87)
GARS082	62.29	1.19	52.7	1.27	355	3879	1405	1	24	1	0	0	0	0	0	1555	1023 (65.79)
GARS083	75.81	4.2	49.2	5.16	758	8648	4356	1	21	1	0	0	0	0	0	4847	3140 (64.78)
GARS084	51.55	2.4	50.4	2.1	612	3662	1916	0	15	0	0	0	0	0	0	2204	1313 (59.57)
GARS085	97.58	1.12	47.8	4.41	199	32406	3775	1	41	1	0	0	0	1	0	3856	2431 (63.04)
GARS086	80.22	1.65	42.5	3.24	680	5711	2585	1	26	1	0	0	0	1	0	2931	1905 (64.99)
GARS087	79.74	4.3	38.0	3.84	908	4985	3109	0	29	0	0	0	0	0	0	3557	2256 (63.42)
GARS088	70.94	2.2	43.2	4.61	928	5670	3840	0	17	0	0	0	0	0	0	4323	2487 (57.53)
GARS089	57.32	0.2	50.1	3.2	515	7406	2740	0	27	0	0	0	0	0	0	3030	1878 (61.98)
GARS090	74.85	3.37	47.8	3.33	606	6192	2859	0	32	0	0	0	0	0	0	3329	2054 (61.7)
GARS091	70.34	2.3	47.4	05.04	511	12243	4135	0	27	0	0	0	0	0	0	4511	2595 (57.53)
GARS092	69.45	4.39	48.4	4.55	812	6723	3975	1	20	1	0	0	0	2	0	4533	2597 (57.29)
GARS093	52.61	3.66	47.5	2.75	594	5102	2265	0	26	0	0	0	0	0	0	2746	1742 (63.44)
GARS094	51.29	1.1	47.5	1.9	349	6055	1593	1	25	1	0	0	0	1	0	1847	1183 (64.05)
GARS095	79.74	1.8	59.5	3.79	975	4383	3981	0	35	0	0	0	0	0	0	4327	3108 (71.83)
GARS096	86.59	1.74	63.1	2.91	407	7980	2932	0	27	0	0	0	0	1	0	3086	2525 (81.82)
GARS097	51.86	0.12	56.5	1.42	504	2940	1558	0	13	0	0	0	0	0	0	1782	1218 (68.35)
GARS098	80.32	0.2	53.9	1.54	250	8392	1631	0	35	0	0	0	0	1	0	1630	1333 (81.78)
GARS099	70.72	02.03	65.2	1.47	383	5122	1532	0	35	0	0	0	0	1	0	1679	1329 (79.15)
GARS100	80.49	1.26	55.3	1.7	311	7375	1707	1	39	1	0	0	0	1	0	1867	1412 (75.63)
GARS101	53.63	3.72	61.3	01.03	595	1732	1190	0	26	0	0	0	0	0	0	1441	1126 (78.14)
GARS102	68.19	2.68	55.1	2.69	949	3333	2733	0	29	0	0	0	0	0	0	3134	2140 (68.28)
GARS103	63.44	2.25	51.0	2.32	699	3600	2169	0	25	0	0	0	0	0	0	2566	1837 (71.59)
GARS104	58.64	2.99	69.2	2.38	776	3383	2382	0	15	0	0	0	0	0	0	2796	1925 (68.85)
GARS105	67.69	3.8	52.4	1.3	514	2909	1515	0	12	0	0	0	0	1	0	1686	1336 (79.24)
GARS106	84.87	0.47	60.3	1.95	89	39449	1728	1	33	1	0	0	0	1	0	1660	1318 (79.4)
GARS107	76.03	03.01	42.8	1.61	561	3536	1418	1	28	1	0	0	0	1	0	1805	1244 (68.92)
GARS108	52.71	2.44	59.1	01.01	352	2972	1074	0	14	0	0	0	0	1	0	1219	887 (72.76)
GARS109	96.21	0.0	42.5	2.71	125	35279	2434	1	34	1	0	0	0	1	0	2483	2018 (81.27)
GARS110	74.65	2.44	50.7	2.4	656	4021	2278	1	20	1	0	0	0	1	0	2671	1982 (74.2)
GARS111	66.61	0.05	52.1	2.0	656	3321	2114	1	20	1	0	0	0	0	0	2363	1644 (69.57)
GARS112	52.34	4.94	64.8	02.09	736	2916	2109	0	17	0	0	0	0	0	0	2455	1699 (69.21)
GARS113	78.44	04.03	67.5	2.95	850	3910	2959	0	42	0	0	0	0	0	0	3408	2315 (67.93)
GARS114	51.53	2.59	63.0	03.09	1191	2614	3283	0	23	0	0	0	0	0	0	3912	2663 (68.07)
GARS115	71.29	04.08	57.9	2.77	894	3217	3137	0	30	0	0	0	0	0	0	3394	2494 (73.48)