UNIVERSITÀ POLITECNICA DELLE MARCHE

DIPARTIMENTO DI SCIENZE DELLA VITA E DELL'AMBIENTE

Corso di Laurea Magistrale in Biologia Marina

DISTRIBUZIONE GLOBALE DELLA BIODIVERSITÀ DI BACTERIA E ARCHAEA IN AMBIENTI MARINI ESTREMI

GLOBAL BIODIVERSITY OF BACTERIA AND ARCHAEA IN EXTREME MARINE ENVIRONMENTS

Tesi di Laurea Magistrale di:
Elisa Laiolo

Relatore: Chiar.mo
Prof. Roberto Danovaro

Sessione Straordinaria
Anno Accademico 2018/2019

Table of contents

Abstract

I sistemi idrotermali sottomarini sono classificati come ambienti estremi poiché caratterizzati da condizioni fisico-chimiche proibitive per la maggioranza degli organismi viventi.

I fluidi emessi da questi sistemi mostrano elevate temperature e contengono composti ridotti che vengono utilizzati da procarioti chemioautotrofi per la produzione di carbonio organico. Abbondanti comunità di Archaea e Bacteria caratterizzano questi ecosistemi, rendendoli degli hotspot di biodiversità.

I sistemi idrotermali sono distribuiti globalmente, lungo le dorsali oceaniche e in altri regioni tettonicamente attive, come i bacini di retroarco. Questo studio prende in considerazione librerie provenienti da sistemi idrotermali localizzati nei tre oceani e nel Mar Mediterraneo, con caratteristiche fisico-chimiche e geologiche distinte.

L'obiettivo di questo studio è analizzare la biodiversità procariotica in sistemi idrotermali e le potenzialità funzionali delle comunità che li abitano, sia a livello di produzione di metaboliti secondari che di adattamenti metabolici.

I risultati di questo studio hanno mostrato una similarità tra sistemi idrotermali, sia a livello di area geografica di provenienza che di matrice del campione. La presenza pressoché ubiquitaria di geni marker specifici ha confermato la potenziale abilità dei procarioti di adattare le proprie strategie metaboliche in base alle circostanze. La presenza ubiquitaria di geni codificanti per metaboliti secondari implicati in meccanismi di competizione interspecifica come le batteriocine apre nuove prospettive per possibili applicazioni biotecnologiche.

1. Introduction

1.1. Hydrothermal-vent systems

Hydrothermal vents were discovered in 1977 along the spreading axis of the Galapagos Rift in the eastern Pacific Ocean, at a depth of 2500 m. (Corliss and Ballard, 1977).

Deep- sea hydrothermal vents are regions of the sea floor at which hot, anoxic, chemical- rich water is released into the otherwise cold, oxic deep ocean. They form when seawater percolates through cracks in the ocean crust into the subsurface, heating up and reacting with hot rocks. This processes enrich it with a variety of chemicals and volatile gases (Dick, 2019).

Then the now called hydrothermal fluid rises and emerges from orifices in the sea floor, rapidly mixing with seawater and providing a redox interface at which chemical sources of energy support vent ecosystems.

Large and high chimneys (**Figure 1.1)** are often formed due to the precipitation of metals like sulfur, copper, zinc and iron when they meet cold and oxygenated seawater (Von Damm, 1990).

In contrast to most ecosystems, which are fuelled by photosynthesis, vent communities depend on chemosynthesis..

Similar deep-sea ecosystem are the cold seeps (or cold vents), where mineral- or methane-rich water seeps from the seafloor.

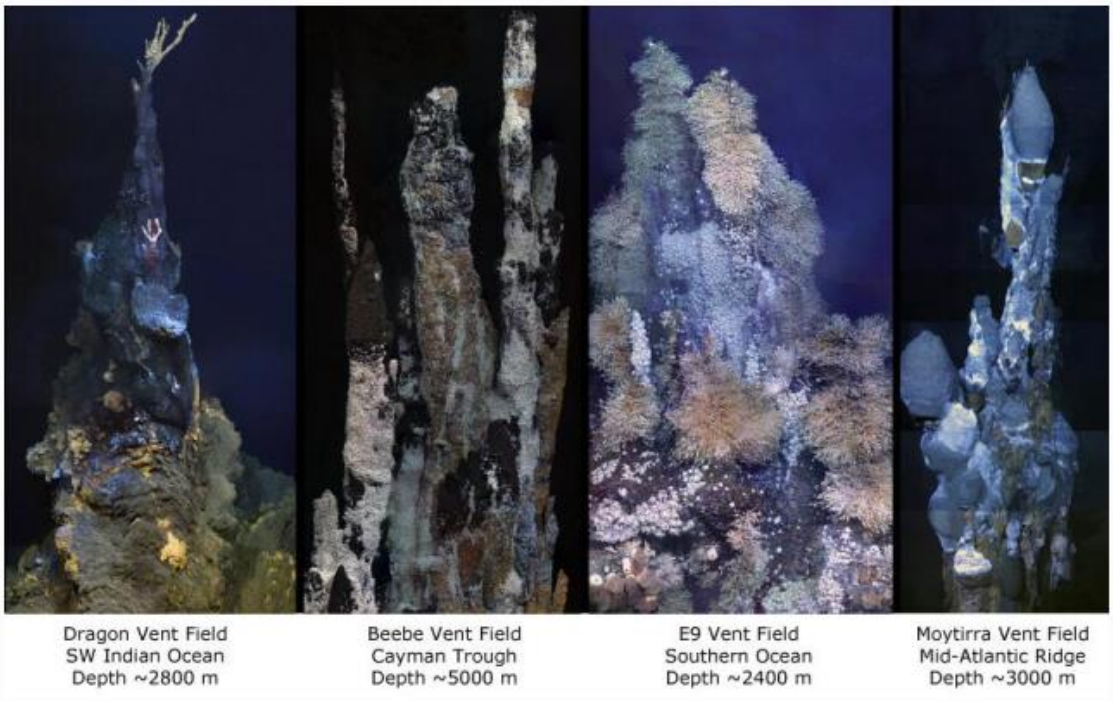

Figure 1.1 Hydrothermal vents chimneys. (modified from University of Southampton archives)

Cold seeps do not require high tectonic activity and may be located more sporadically across the deep sea than hydrothermal vents. Here, specialized bacteria utilize the high density of sulfur or methane compounds as energetic source, forming the bottom of a complex food web.

1.2 Global distribution and characteristics of hydrothermal vents

Hydrothermal vents occur globally along the ridge systems, where the volcanic activity involves the rising of the basaltic magma. Discrete clusters of vent fields exist globally, not only along the mid- ocean ridges but also, in several

other geological settings, such as back-arc spreading centers, hotspot volcanoes and seamounts **(Figure 1.2).**

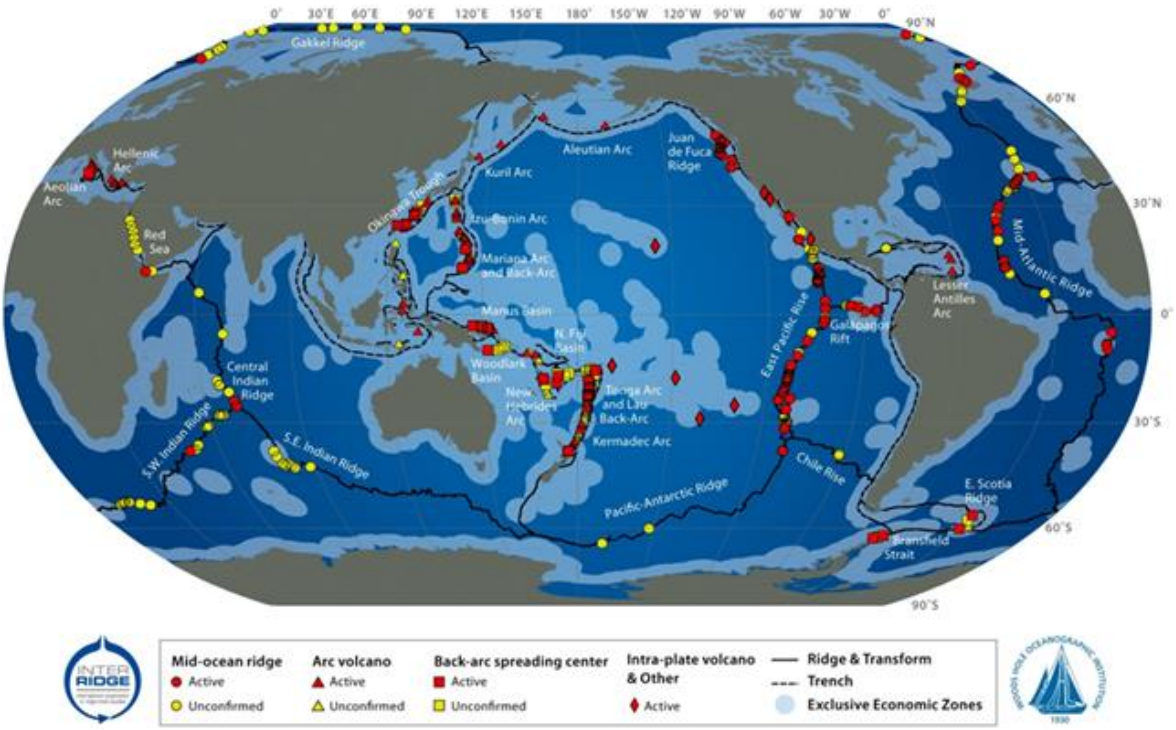

Figure 1.2 Global Distribution of Hydrothermal vent fields (Beaulieu, Joyce, and Soule, 2010. https://vents-data.interridge.org/maps)

Vents are diverse not only in terms of their large- scale geological setting, as mentioned above , but also accordingly to the host rock (for example, basalt versus ultramafic rocks, such as peridotite) and to the seafloor spreading rate **(Figure 1.3).**

The geological setting can be further modified by sediment cover (Schrenk, Huber and Edwards, 2010)

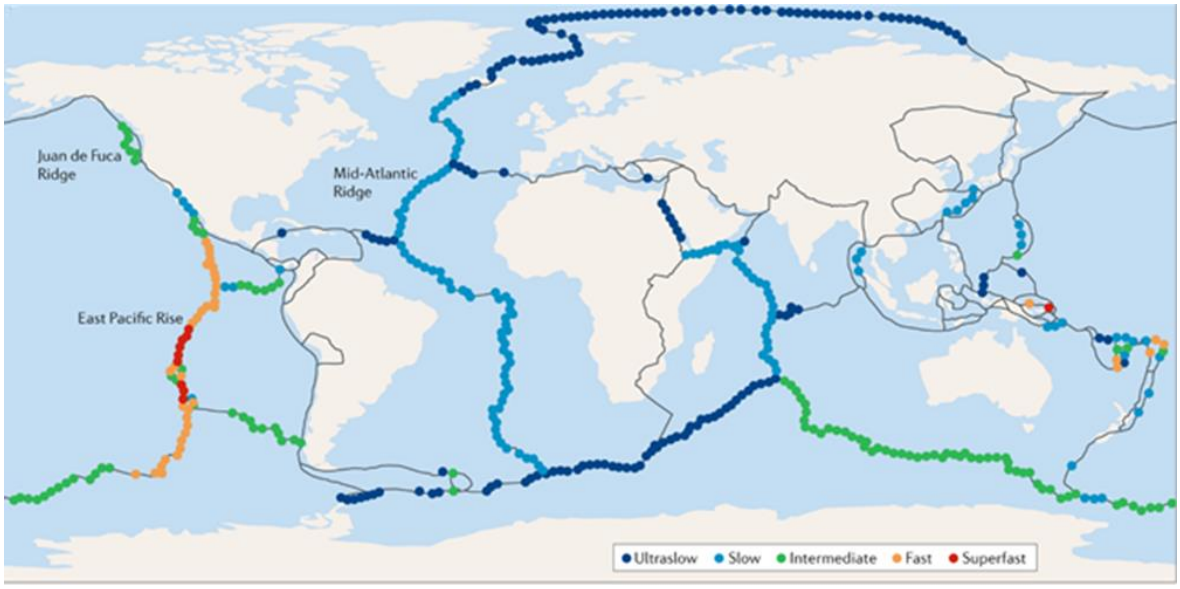

Figure 1.3 Confirmed and inferred active deep-sea hydrothermal-vent fields (Dick et al., 2019) Ridge axes are coloured according to spreading rate: ultraslow (dark blue; <20 mm per year), slow (light blue; 20–50 mm per year), intermediate (green; 50–80 mm per year), fast (orange; 80–140 mm per year) and superfast (red; >140 mm per year).

Vents located on fast to superfast spreading ridges are usually mafic hosted (high MgO and FeO content and SiO_2> 45%), while slow to ultraslow spreading ridges tend to present a wider variety of host rocks, and ultramafic (SiO_2 < 45%) settings are not unusual.

Vents not located on ocean ridges, such as hotspot volcanoes and back-arc basins, often present a combination of host-rocks, displaying a combination of basalts and other rocks, ranging from intermediate (52–63% SiO_2), like andesite, to intermediate-felsic (63–69% SiO_2), like dacite, to felsic (>69% SiO_2), such as rhyolite (Wilkinson et al., 2008)

All these factors shape the chemical and physical aspects of microbial habitats (Orcutt et al., 2011; Sievert and Vetriani, 2012; Reysenbach and Shock, 2002; Holden et al., 2012).

Temperature and chemistry of the emitted vent fluids and nature of the formed mineral deposits depend mainly on three factors: the temperature and geology of the local rock, the duration of water-rock reactions and the degree of subsequent mixing with cooler fluids.

This results in a variety of vent types and fluids (black smokers, white smokers and alkaline and carbonate vents), mineral deposits (sulfides and carbonates) and vent structures, such as chimneys, flanges and mounds (Orcutt et al., 2011).

1.3 Microbial communities in hydrothermal vents

1.3.1 Microbial biodiversity and ecological importance

Hydrothermal systems are reducing environments like cold seeps, whale carcasses, oxygen minimum zones, sunken woods, sea grasses and other organic remains (Dando et al., 1992; Van Dover, 2000; Tunnicliffe et al., 2003 a, b).

Prokaryotes play a critical role in channeling the flux of energy in aquatic environments (Azam et al., 1983; Zimmermann et al., 1978).

The microorganisms which feed on sulfur and sulphides provide the compounds of the primary production through chemosynthesis which is at the base of the vent food web (Van Dover, 2001).

The steep thermal and chemical gradients at active hydrothermal chimneys create a diversity of niches. (Reysenbach et al., 2000) (Takai and Nakamura, 2010). **Figure 1.4** offers a generalized, not comprehensive, overview of microbial habitats and niches and most abundant taxa.

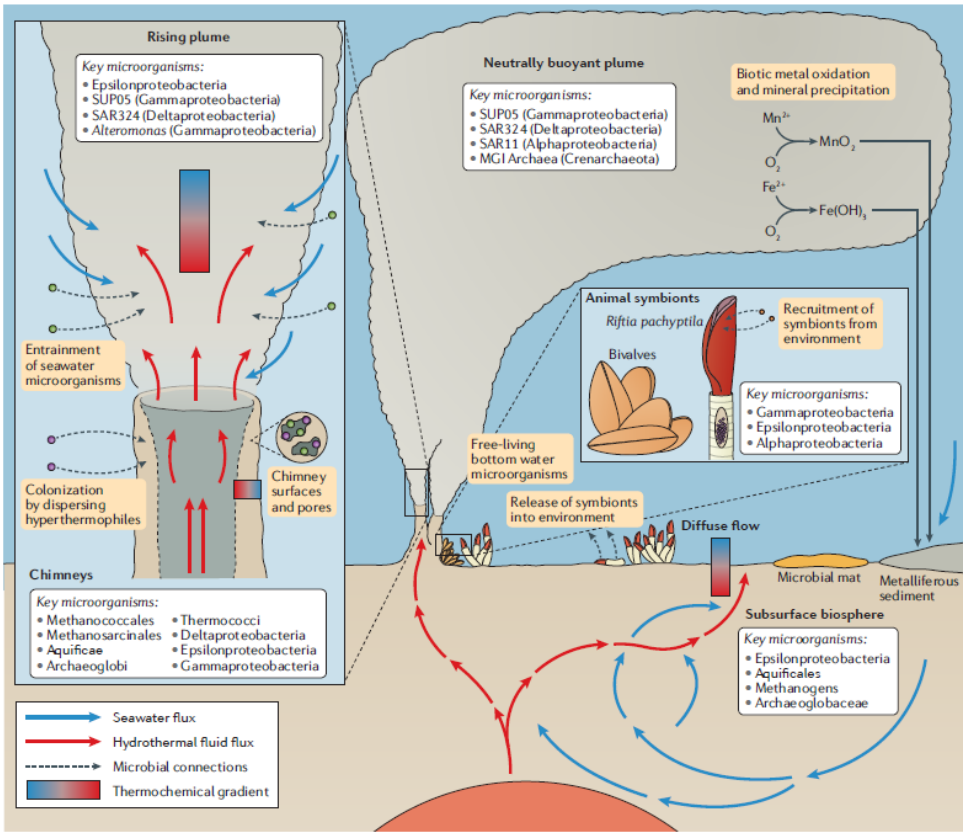

Figure 1.4. Microbial habitats and biodiversity at deep- sea hydrothermal vents. Key microorganisms for each habitat represent abundant taxa observed across multiple vent fields by various cultivation-independent approaches (Dick et al., 2019)

Dominant organisms in chimneys and surrounding sediments include thermophiles and hyperthermophiles from Deltaproteobacteria,

Epsilonproteobacteria , Gammaproteobacteria, Methanococcales, Methanosarcinales, Aquificae, Archaeoglobi and Thermococci, (Orcutt et al., 2011; Reveillaud et al., 2016; Takai and Nakamura, 2010; Flores et al., 2012; Takai et al., 2004)

Hydrothermal plumes undergo rapid cooling and geochemical changes during the turbulent mixing between hydrothermal fluids and seawater, making for a dynamic habitat (Dick et al., 2013; Reed et al., 2015)

The distribution and source of microbial communities within plumes, especially the rising portion, were long elusive owing to sampling challenges that were resolved by recent technological advances (Breier et al.2014). Plumes are composed primarily of organisms derived from the water column, such as Gammaproteobacteria, Deltaproteobacteria), Alphaproteobacteria and Thaumarchaeota (Dick and Tebo, 2010; German et al.,2010; Sheik et al., 2015). Sea floor and/or subsurface organisms such as Epsilonproteobacteria can be present in plumes (German et al.,2010; Sheik et al.,2015).

1.3.2 Metabolic strategies and adaptations

Fluid chemistry and physical characteristics present both opportunities and lethal challenges to hydrothermal- vent organisms. High temperature sets a hard limit on life due to the biochemical constraints of membranes, proteins and nucleic acids (Stetter, 1999). The highest temperature at which microbial

growth has been observed is 122 °C, a record set by a deep- sea hydrothermal-vent archaea. Some hyperthermophiles use flagella to move towards heat sources, to avoid lethally high temperature and to adhere to chimney surfaces (Mora et al., 2014). Temperature has a central role in determining 'who is where'. Some organisms maintain their position in the thermochemical gradient by attaching to surfaces as biofilms (Sievert and Vetriani,2012; Reysenbach and Shock,2002) or finding refuge within animal hosts, although even here temperature and chemistry fluctuate substantially (Johnson et al.,1988).

The energy metabolisms of organisms in the near-vent environment are typically anaerobic and lithoautotrophic. However, fluid mixing and thus the transition to microaerobic and aerobic processes occur over short spatial scales (Tivey, 2004).

Owing to rapid cooling and dynamic chemistry, the temporal and spatial window in which microorganisms can exploit their preferred physical and chemical niches may be narrow. This situation is particularly challenging for free-living organisms that are carried along in fluids in the subsurface, diffuse flow and in rising plumes (Anantharaman et al., 2016; Fortunato et al., 2018; Reed et al., 2015)

Adaptations to such dynamic conditions include motility and metabolic versatility, which provide the ability to find preferred environments and to use

multiple electron acceptors, such as oxygen, nitrate and sulfur (Reysenbach and Shock,2002; Campbell et al.,2006)

Vent fluids range from acidic to highly alkaline, not only demanding capacity for pH homeostasis but also shaping the availability of dissolved inorganic carbon (higher concentrations at lower pH), which is required in excess for the rapid chemoautotrophic growth exhibited by bacteria–animal symbioses as happens in the giant tubeworm *Riftia pachyptila* (Goffredi et al., 1997). Hot, acidic fluids often hold sufficient carbon dioxide, whereas it may be limiting in cold, neutral or basic fluids (Schrenk et al.,2004), thus requiring mechanisms for carbon uptake and concentration (Mangiapia et al., 2017) Examples of distribution according to geochemistry and temperature are the possible pathways for carbon fixation (Hugler and Sievert, 2011) Different carbon-fixation pathways prevail in different temperature ranges: the reductive tricarboxylic acid cycle (rTCA), energy- efficient yet oxygen-sensitive, usually is dominant between 20 and 90°, the Calvin–Benson Bassham (CBB) cycle, energetically costly yet oxygen- tolerant, predominates at temperatures <20 °C (Dick et al. 2019), while above 90 °C methanogenesis or sulfate reduction are preformed using the Wood–Ljungdahl pathway and the dicarboxylate–4- hydroxybutyrate pathway for carbon fixation, respectively (Hugler and Sievert,2011).

Hydrothermal fluids also contain several toxic compounds, such as heavy metals (copper, cadmium and lead) and hydrogen sulfide, that require specific adaptations (Kelley et al.,2002; Sievert and Vetriani, 2012)

1.3.3 Virus-prokaryote interactions

Virus are the most abundant entities on the planet and exceed prokaryotic abundance by, at least, one order of magnitude

They are one of the major cause of mortality for a wide range of marine organisms and the majority of them, in the ocean, appear to infect prokaryotes. Their role is critical in shaping aquatic communities and determining ecosystem dynamics (Fuhrman, 1999; Suttle, 2005).Prophages (viral genomes integrated within a host bacterial genome) can confer phenotypic traits to their hosts, such as enhanced pathogenicity (Rezaei Javan et al., 2019).

Phages affect microbial diversity by mechanisms of horizontal gene transfer (Breitbart, 2012).

During assembly in donor host cells, phages package some part of host genes, which are then transferred and inserted into the genomes of recipient host cells in the next infection (Long et al., 2007; Paul, 2008). Phage-packaged genes can dramatically change the genotype and phenotype of the hosts (Thompson et al.,

2005).

The antagonism between the phages and hosts can be a mechanism to increase the diversity of viral and microbial community structures (Koskella et al., 2014).

Recent studies reveal that phages may be essential to or at least favour microbial survival (Clokie et al., 2011; Anantharaman et al., 2014).

A possible benefit derived from this interactions is an enhanced ability of surviving in unfavorable environments (Steward et al., 2011), thanks to newly encoded genes.

2. Objectives of the study

The aim of this study is to investigate the biodiversity of bacteria and archaea in extreme marine environments, their relative taxonomic abundance and distribution. We investigate if there are differences among hydrothermal vents biogeographic region of provenance and/or among different sample matrices (sediments vs. plumes for example).

In addition, we focus on their functional potentiality, both in terms of production of secondary metabolites and potential metabolic adaptations, in order to have a better understanding of how prokaryotes are distributed among the various sites; we also investigate possible prophage-host interactions, still poorly studied in these environments.

Mineral-rich sediments such as those of hydrothermal systems are areas of interest for mining activities and need to be investigated for a better understanding of the potential responses of these ecosystems to anthropogenic impacts. Furthermore, prokaryotes in these habitats attract attention for their ability to synthesize diverse classes of bioactive secondary metabolites; the identification of their production methods and effects has become one of the noteworthy trends in modern biotechnology, leading to the potential development of new antibiotics and control methods of pathogenic bacteria.

3. Materials and methods

3.1 Selection of libraries

We selected a total of 82 metagenomic libraries **(Table 3.1)** of marine environments (from both seawater and extreme ecosystems) from the SRA of NCBI (https://www.ncbi.nlm.nih.gov/sra) and we proceeded to download them using the fastq-dump tool within the SRA toolkit.

The SRA is NIH's primary archive of high-throughput sequencing data and is part of the International Nucleotide Sequence Database Collaboration (INSDC) that includes the NCBI Sequence Read Archive (SRA), the European Bioinformatics Institute (EBI), and the DNA Database of Japan (DDBJ).

Thirty-two libraries were from samples collected in the Atlantic Ocean, ten from the Indian Ocean, five from the Mediterranean Sea and thirty-five from the Pacific Ocean.

The geographic provenance of the libraries is shown in **Figure 3.1.**

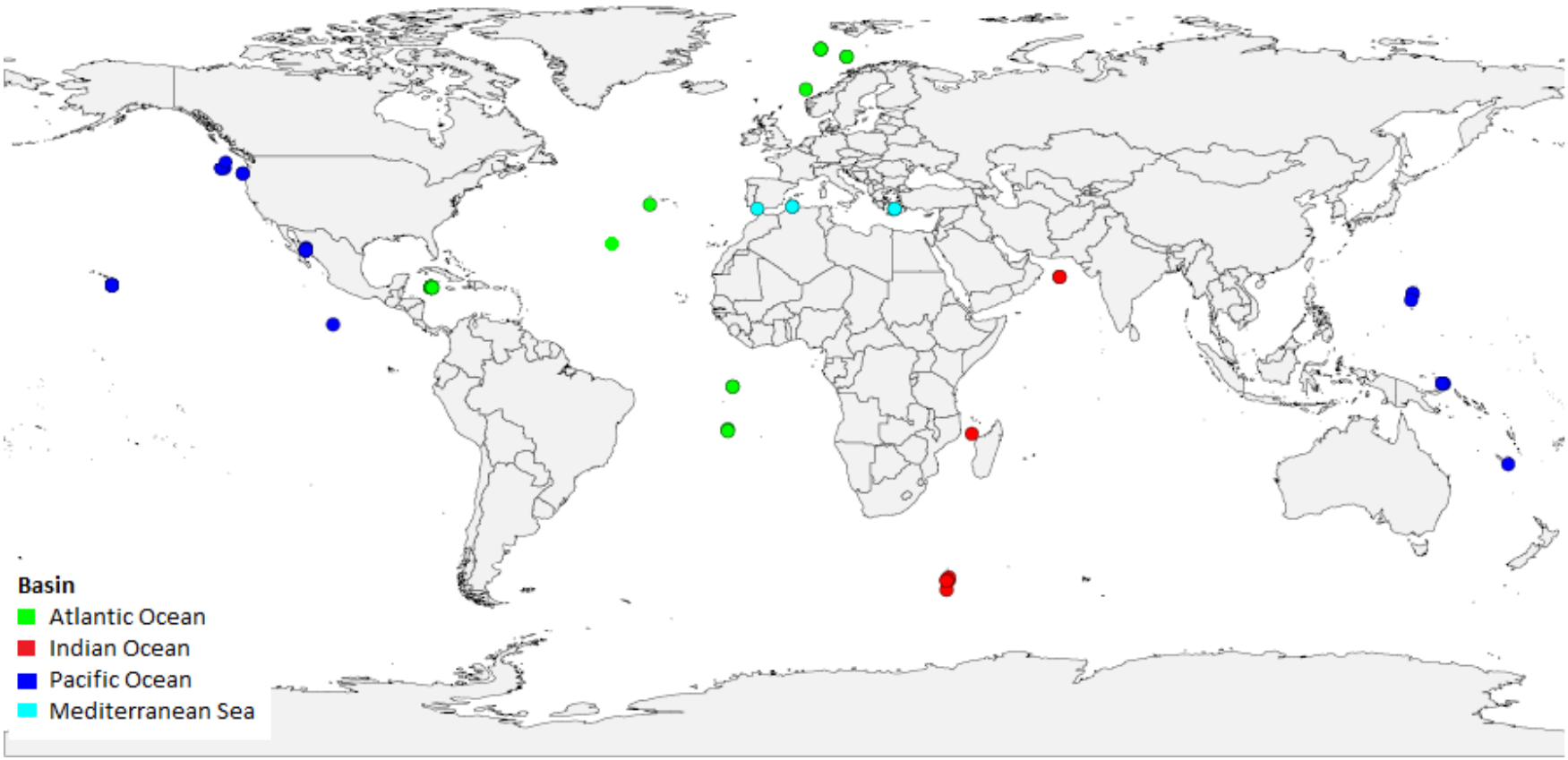

Figure 3.1 Global distribution of the libraries. Different colors highlight the basin of provenance: green for the Atlantic Ocean, red for the Indian Ocean, blue for The Pacific Ocean and cyan for the Mediterranean Sea

Table 3.1 Metagenomic libraries. Libraries are identified by their SRRcode and are classified according to their biogeographic region and their sample matrix.

IDcode	Region	Sample matrix	Latitude	Longitude	Depth
SRR2046237	Atlantic Ocean	H. vent fluid	18.3279	-81.4308	4956
ERR1078300	Atlantic Ocean	H. vent fluid	37.5040	-31.3194	828
ERR1662199	Atlantic Ocean	H. vent fluid	18.3767	-81.7981	2297
ERR1662208	Atlantic Ocean	H. vent fluid	18.5472	-81.7181	4988
ERR1662233	Atlantic Ocean	H. vent fluid	18.5471	-81.7182	4986
ERR1662245	Atlantic Ocean	H. vent fluid	18.3750	-81.7976	2293
ERR1662255	Atlantic Ocean	H. vent fluid	18.3747	-81.7973	2372
ERR1662269	Atlantic Ocean	H. vent fluid	18.3747	-81.7973	2369
ERR1662290	Atlantic Ocean	H. vent fluid	18.3747	-81.7973	2369
ERR1662333	Atlantic Ocean	H. vent fluid	18.3747	-81.7973	2369
ERR1662343	Atlantic Ocean	H. vent fluid	18.3750	-81.7976	2370
ERR1662393	Atlantic Ocean	H. vent fluid	18.3752	-81.7971	2388
ERR1662414	Atlantic Ocean	H. vent fluid	18.3747	-81.7973	2376
ERR1662443	Atlantic Ocean	H. vent fluid	18.3769	-81.7979	2308
ERR1662466	Atlantic Ocean	H. vent fluid	18.5466	-81.7182	4980
ERR1662504	Atlantic Ocean	H. vent fluid	18.5466	-81.7177	4980
SRR2046221	Atlantic Ocean	H. vent fluid	18.3279	-81.4308	4950
SRR1555743	Atlantic Ocean	H. vent sed.	73.7632	8.4640	3283
SRR1555744	Atlantic Ocean	H. vent sed.	73.7632	8.4640	3283
SRR1555748	Atlantic Ocean	H. vent sed.	73.7632	8.4640	3283
SRR6439364	Atlantic Ocean	H. vent sed.	-14.3100	-13.3500	2194
SRR6439366	Atlantic Ocean	H. vent sed.	-14.5100	-13.3500	1820
SRR6439368	Atlantic Ocean	H. vent sed.	-14.7100	-13.3500	2000
EH.MATs	Atlantic Ocean	Volcano vent	27.3700	18.0000	89
SRR043581	Atlantic Ocean	H. vent sed.	72.0000	14.4400	1250
SRR043582	Atlantic Ocean	H. vent sed.	72.0000	14.4400	1250
SRS139630	Atlantic Ocean	H. vent sed.	64.4527	5.0410	746
ERR133679	Atlantic Ocean	H. vent sed.	36.2167	-33.9000	2320
ERR133680	Atlantic Ocean	H. vent sed.	36.2167	-33.9000	2320
ERR133681	Atlantic Ocean	H. vent sed.	36.2167	-33.9000	2320
SRR363374	Atlantic Ocean	H. vent sed.	-4.4800	-12.2200	2996
SRR363375	Atlantic Ocean	H. vent sed.	-4.4800	-12.2200	2996
SRR6439354	Indian Ocean	H. vent sed.	-49.4800	37.9900	2328
SRR6439355	Indian Ocean	H. vent sed.	-49.3800	37.5200	2822
SRR6439356	Indian Ocean	H. vent sed.	-49.3900	37.7800	2742
SRR6439358	Indian Ocean	H. vent sed.	-48.8500	37.9900	1711
SRR6439359	Indian Ocean	H. vent sed.	-51.7300	37.4700	1592
SRR6439360	Indian Ocean	H. vent sed.	-49.4800	37.5200	1973
SRR6439361	Indian Ocean	H. vent sed.	-49.6600	37.4700	1753
ERR598966	Indian Ocean	Seawater	20.8183	63.5047	5

IDcode	Region	Sample matrix	Latitude	Longitude	Depth
ERR598974	Indian Ocean	Seawater	20.8222	63.5133	17
ERR599112	Indian Ocean	Seawater	-15.3379	43.2948	1000
ERR315856	Med.Sea	Seawater	37.0541	1.9478	42
ERR315857	Med.Sea	Seawater	37.0510	1.9378	5
ERR598950	Med.Sea	Seawater	36.5533	-6.5669	40
SRS1611383	Med.Sea	Volcano vent	36.4499	25.4021	336
SRS1611383	Med.Sea	Volcano vent	36.5273	25.4872	444
ERR2021503	Pacific Ocean	H. vent fluid	45.9332	-129.9822	1518
ERR2021505	Pacific Ocean	H. vent fluid	45.9332	-130.0139	1541
ERR2021507	Pacific Ocean	H. vent fluid	45.9332	-129.2822	1518
ERR2021509	Pacific Ocean	H. vent fluid	46.0747	-129.9950	1716
ERR2021511	Pacific Ocean	H. vent fluid	45.9336	-130.0137	1500
ERR1679395	Pacific Ocean	H. vent fluid	-3.7993	152.1013	1199
SRR7168051	Pacific Ocean	H. vent fluid	16.9613	144.8678	3277
SRR7168052	Pacific Ocean	H. vent fluid	16.9618	144.8692	3278
SRR7168054	Pacific Ocean	H. vent fluid	15.4799	144.5075	3913
SRR7968103	Pacific Ocean	H. vent sed.	9.8481	-104.2967	2500
SRR7041831	Pacific Ocean	H. vent sed.	27.0010	-110.5989	2000
SRR7042026	Pacific Ocean	H. vent sed.	27.0010	-110.5989	2000
SRR7042028	Pacific Ocean	H. vent sed.	27.4667	-110.5332	2000
SRR7042271	Pacific Ocean	H. vent sed.	27.0010	-110.5999	2000
SRR7042332	Pacific Ocean	H. vent sed.	27.0010	-110.5998	2000
SRR7042394	Pacific Ocean	H. vent sed.	27.0010	-110.5998	2000
SRR7051036	Pacific Ocean	H. vent sed.	27.0074	-110.5910	2000
SRR7051057	Pacific Ocean	H. vent sed.	27.0114	-110.5931	2000
SRR7051180	Pacific Ocean	H. vent sed.	27.0118	-110.5962	2000
SRR7051207	Pacific Ocean	H. vent sed.	27.0114	-110.5956	2000
SRR7051260	Pacific Ocean	H. vent sed.	27.0114	-110.5956	2000
SRR6439370	Pacific Ocean	H. vent sed.	47.5600	-129.0600	2182
SRR7968105	Pacific Ocean	H. vent sed.	-22.1802	-176.6008	1800
SRR5189948	Pacific Ocean	H. vent sed.	18.9020	-155.2570	1300
SRR5189949	Pacific Ocean	H. vent sed.	18.9020	-155.2570	1300
SRR5189950	Pacific Ocean	H. vent sed.	18.9020	-155.2570	1300
SRR5189951	Pacific Ocean	H. vent sed.	18.9020	-155.2570	1250
SRR5189952	Pacific Ocean	H. vent sed.	18.9020	-155.2570	1200
SRR5189953	Pacific Ocean	H. vent sed.	18.9020	-155.2570	1250
ERR2639421	Pacific Ocean	H. vent sed.	-3.7268	151.6721	1689
ERR2639422	Pacific Ocean	H. vent sed.	-3.7285	151.6725	1719
ERR2639423	Pacific Ocean	H. vent sed.	-3.7269	151.6720	1691
SRR1636517	Pacific Ocean	H. vent sed.	-22.3609	166.8800	43
SRR2133563	Pacific Ocean	H. vent sed.	44.6695	-125.0982	600
SRR2133563	Pacific Ocean	H. vent sed.	44.6695	-125.0982	600

3.2 Metagenomic read assembly and binning

Raw reads were correctly paired (repair.sh), quality-trimmed to remove sequences with a PHRED score <10 while removing potential leftover adapters (BBDuk.sh), error-corrected (tadpole.sh) and digitally normalized to remove high-abundance kmers (BBNorm.sh) using the BBTools suite (Bushnell, 2014; http://sourceforge. net/projects/bbmap). Normalization allowed us to reduce both the computational load and computational time on the HPC cluster (provided by Stazione Zoologica di Napoli "Anton Dohrn"), while reducing complexity in the assembly of metagenomes (Li et al., 2015).

The resulting scaffolds were subsequently binned using metaWRAP (Uritskiy et al., 2018), a modular pipeline software for shotgun metagenomic data analysis. MetaWRAP deploys state-of-the-art software to handle metagenomic data processing starting from raw sequencing reads and ending in metagenomic bins and their analysis. In particular metaWRAP allows to perform binning utilizing different algorithms and to subsequently reconcile binning results to gather the best results across the all possible combinations of binning systems. We performed binning utilizing MetaBAT2 (Kang et al., 2019), MaxBin2 (Wu et al., 2016) and CONCOCT (Alneberg et al., 2013). Bins obtained by each algorithm and by each possible combination of the three selected algorithms were checked for completeness and redundancy using the CheckM pipeline

(Parks et al., 2015); those with completeness >90% and redundancy <5% ("high quality MAGs") were kept for subsequent analyses. Quality scores for metagenome-assembled genomes were assigned following the MIMAG indications (Bowers et al., 2017).

3.3 Taxonomic classification of MAGs

For the taxonomic classification of our sequences we integrated the output of two different approaches: CAT (Contig Annotation Tool, Cambuy et al., 2016) and SSU taxonomic analysis through the SILVA database (Ludwig et al., 2004; Quast et al. 2013). The Contig Annotation Tool (CAT) is a pipeline for robust and accurate taxonomic classification of long metagenomic sequences. The CAT pipeline is implemented in Python, and it uses third party programs for several steps. SILVA provides comprehensive, quality checked and regularly updated datasets of aligned small (16S/18S, SSU) and large subunit (23S/28S, LSU) ribosomal RNA (rRNA) sequences for all three domains of life. SILVA is the official database of the software package ARB.

Annotations provided by the two tools were manually checked at each taxonomic level to generate a consensus taxonomy: annotations shared between the two tools were kept, but we stopped annotating at the first disagreement

between the tools and marked the subsequent taxonomic levels as “no further classification”

3.4 Metabolic marker gene analysis

For the analysis of the presence of marker genes for specific metabolic pathways (e.g. ribulose-1,5-bisphosphate carboxylase oxygenase as a marker for the Calvin-Benson-Bassham cycle; Hügler and Sievert, 2011), we first performed gene annotation of MAGs using Prokka (Seemann, 2014), with domains “Bacteria” and “Archaea” set for each MAG according to its taxonomy. Predicted gene sequences were subsequently utilized as inputs for the GraftM program (Boyd et al., 2018) using all the packages provided.

3.5 Secondary metabolites prediction

For the identification of secondary metabolites, we used antiSMASH (antibiotics & Secondary Metabolite Analysis Shell, Medema et al., 2011), a comprehensive pipeline capable of identifying biosynthetic loci covering the whole range of known secondary metabolite compound classes. It aligns the identified regions at the gene cluster level to their nearest relatives from a database containing all other known gene clusters, and integrates or cross-links

all previously available secondary-metabolite specific gene analysis methods in one interactive view.

3.6 Detection and taxonomic clustering of prophages

Viruses of microbes impact all ecosystems, however, despite this recognized importance, our understanding of viral diversity and impacts remains limited by few model systems and reference genomes. We used VirSorter (Roux et al., 2015), a tool designed to detect viral signal in different types of microbial sequence data in both a reference-dependent and reference-independent manner, leveraging probabilistic models and extensive virome data to maximize detection of novel viruses, in order to get a more complete overview of our site diversity.

3.7 Statistical Analyses

Data are grouped according to their physical, chemical, geological and geographical features, identifying four categories. According to the NCBI project metadata, we classified our libraries in four categories: seawater, for samples not influenced by hydrothermal activities, hydrothermal-vent

influenced sediments, hydrothermal vent plumes and volcanoes not directly part of ridge systems or spreading centres.

To test for differences in the investigated variables between the sites years and between hydrothermal vent sites, a one-way analysis of variance was conducted using distance-based permutational multivariate analysis of variance (PERMANOVA; Anderson, 2005). The biogeographical region of provenance and the sample matrix were used as fixed factor and *post hoc* comparison with PAIRWISE Test was carried out when significant ($p<0.05$) differences were encountered. In order to analyse our metagenomic profiles, we used STAMP (Statistical Analysis of Metagenomic Profiles) (Parks and Beiko,2010).

Statistical analyses were performed using the PRIMER v.6.1. program and the PERMANOVA+ add-on.

We also used CAP (Canonical Analysis of Principal coordinates, Anderson and Willis, 2003) to futher analyze the influence of our variables.

4. Results

4.1 Assembly and binning process

We obtained a total of 1912 bins with different completeness and redundancy levels. A total of 443 bins resulted to have a completeness >90% and a redundancy <5% and were therefore classified as "high quality MAGs" and kept for subsequent analyses (**Figure 4.1)**

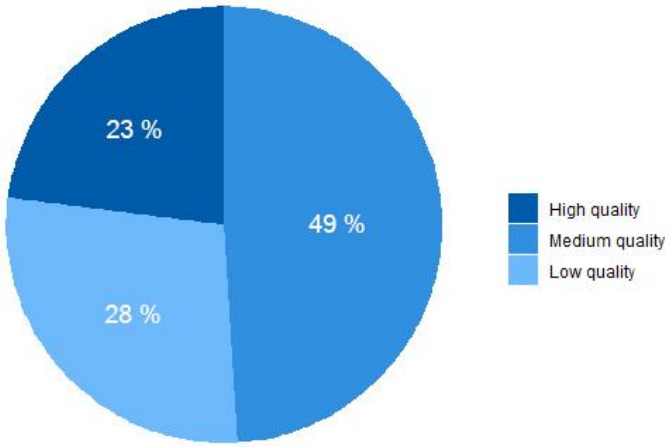

Fig. 4.1 Metagenome-assembled genomes (MAGs) qualitative classification

4.2 Taxonomic classification and relative abundances

4.2.1 Taxonomic classification at domain level

At a domain level, 408 bins (92,1%) were classified as Bacteria and 35 (7,9%) as Archaea **(Figure 4.2)**

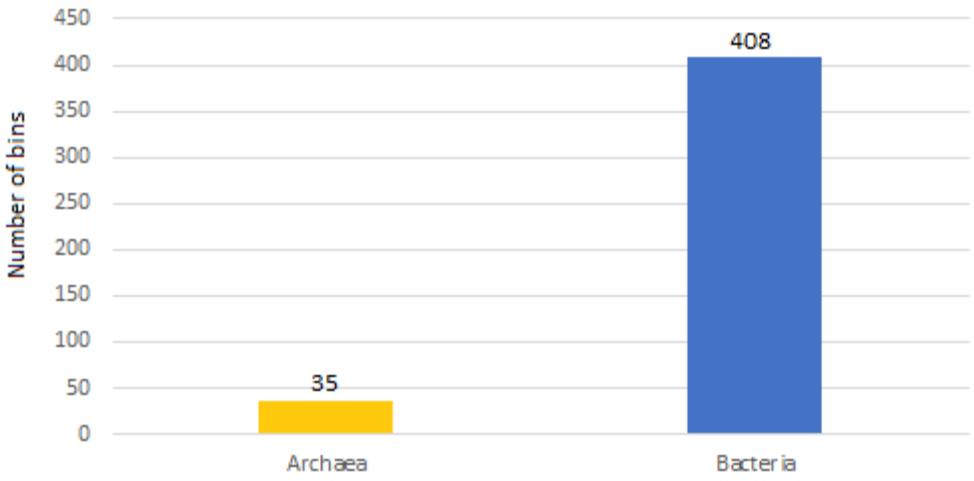

Figure 4.2 Bin taxonomic classification at domain level

4.*2.2. Distribution of phyla*

4.2.2.1 Domain Archaea

In order to compare different biogeographic regions and different sample matrices, we considered the relative abundance of our taxa of interest in each sample, expressed as number of copies of genomes per 10^6 base pairs, for the different sample matrix **(Table 4.1)** and for the different biogeographical regions **(Table 4.2)**

Table 4.1. Archaeal phyla mean relative abundances (copies of genomes /10^6 base pairs) in the different sample matrix.

Phylum	H. vent fluid		H. vent sediments		Seawater		Volcano vent	
	Mean rel. abund.	SD	Mean rel. abund.	SD	Mean rel. abund.	SD	Mean rel. abund.	SD
Crenarchaeota	22.52	19.61	28.66	29.01	8.35	6.55	36.78	22.40
Euryarchaeota	76.26	20.50	70.23	29.08	89.44	7.66	53.76	22.46
Korarchaeota	1.23	1.71	1.11	1.93	2.21	3.69	9.47	9.54

Table 4.2. Archaeal phyla mean relative abundances (copies of genomes /10^6 base pairs) in the different biogeographical regions.

Phylum	Atlantic Ocean		Indian Ocean		Mediterranean Sea		Pacific Ocean	
	Mean rel. abund.	SD	Mean rel. abund.	SD	Mean rel. abund.	SD	Mean rel. abund.	SD
Crenarchaeota	22.85	19.68	12.68	11.57	16.57	11.95	31.50	30.19
Euryarchaeota	75.37	20.59	85.92	11.80	77.35	19.55	67.63	30.30
Korarchaeota	1.78	2.31	1.39	3.02	6.08	8.48	0.87	1.48

We considered only the most abundant phyla in each domain.

For the domain Archaea the phyla considered are Euryarchaeota, Crenarchaeota and Korarchaeota. The most abundant phylum in the 4 sample sources (hydrothermal vent fluids, hydrothermal vent sediments, single volcanoes and seawater) resulted to be Euryarchaeota, followed by members of Crenarchaeota and Korarchaeota. The three phylum showed the same order of abundance considering the four biogeographical regions (Atlantic Ocean, Indian Ocean, Pacific Ocean and Mediterranean Sea).

Korarchaeota is the only archaeal phylum which showed significant differences among biogeographical regions and sample matrices **(Figure 4.3, 4.4)**

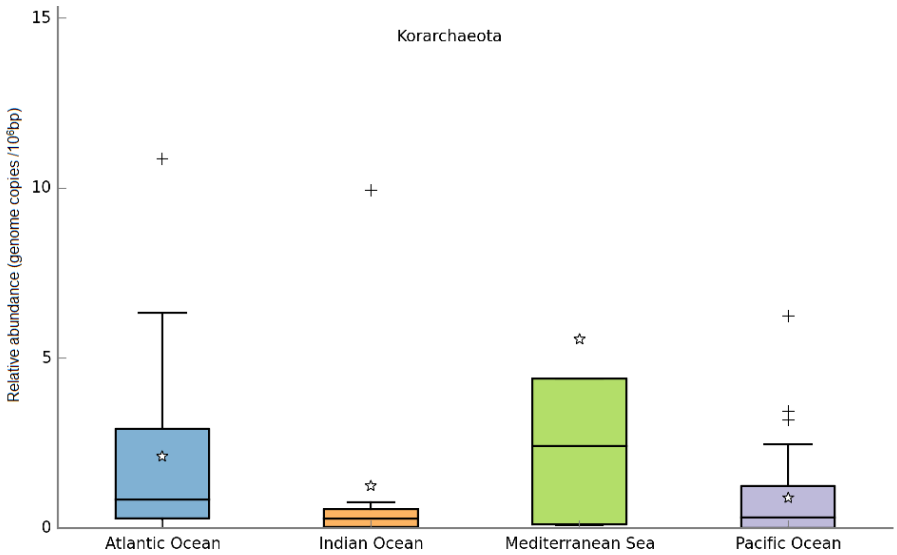

Figure 4.3 Phylum Korarchaeota relative abundance across biogeographical regions (Statistical test: ANOVA, Post-hoc test: Tukey-Kramer, p<0.005)

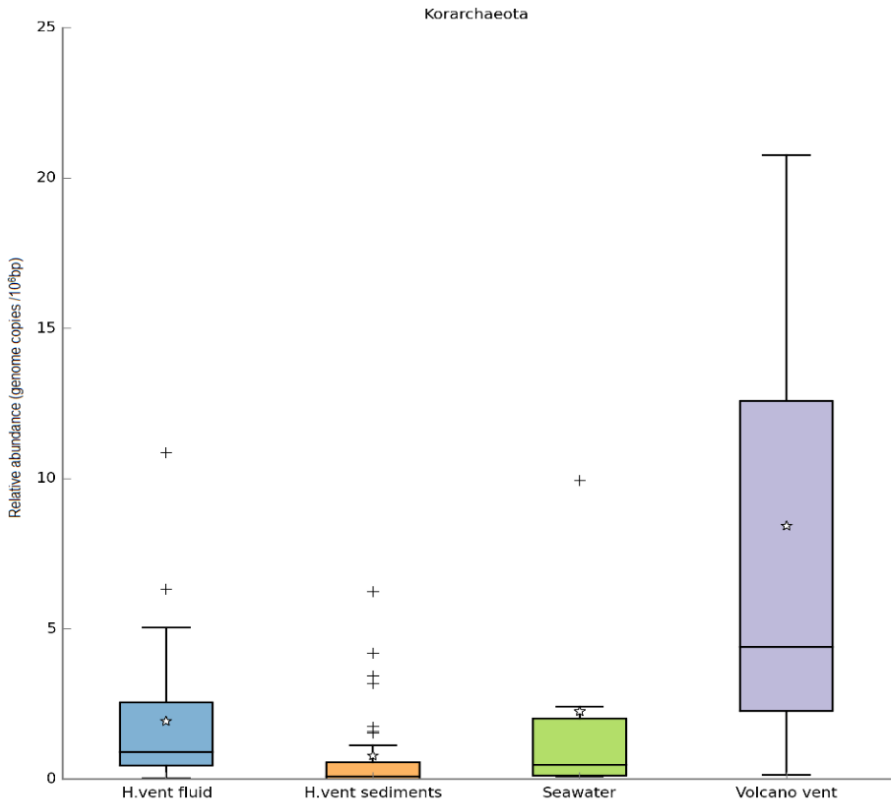

Figure 4.4 Phylum Korarchaeota relative abundance across sample matrices (Statistical test: ANOVA, Post-hoc test: Tukey-Kramer, p<0.005)

4.2.2.2 Domain Bacteria

For the domain Bacteria we considered the phyla Aquificae and Proteobacteria, which were ubiquitously the most abundant.

The phylum Proteobacteria was analyzed at class level, considering classes Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria, Deltaproteobacteria and Epsilonproteobacteria.

In order to compare different biogeographical regions and different sample sources, we considered the relative abundance of our taxa of interest in each sample, expressed as number of copies of genomes per 10^6 base pairs, for the different sample matrices **(Table 4.3)** and for the different biogeographical regions **(Table 4.4).**

The most abundant group in the 4 sample matrices (hydrothermal vent fluids, hydrothermal vent sediments, single volcanoes and seawater) resulted to be Gammaproteobacteria. Epsilonbacteria resulted to be the second group for abundance, with the only exception of seawater samples where Alphaproteobacteria outnumbered them.

Considering the four biogeographical regions (Atlantic Ocean, Indian Ocean, Pacific Ocean and Mediterranean Sea), Gammaproteobacteria dominated across all of them, followed in abundance by Epsilonproteobacteria. The only exception was represented by the samples from the Indian Ocean, where the second most abundant group were Alphaproteobacteria.

Table 4.3. Bacterial phyla mean relative abundances (copies of genomes /10^6 base pairs) in the different sample matrices.

Phylum/Class	H. vent fluid		H. vent sediments		Seawater		Volcano vent	
	Mean rel. abund.	SD	Mean rel. abund.	SD	Mean rel. abund.	SD	Mean rel. abund.	SD
Alphaproteobacteria	6.84	12.33	13.49	11.81	26.06	10.16	20.97	17.32
Betaproteobacteria	4.79	13.00	3.59	10.78	1.14	0.94	0.38	0.27
Deltaproteobacteria	7.34	11.72	19.93	24.17	5.98	1.46	8.28	5.38
Epsilonproteobacteria	42.85	33.11	10.41	18.93	18.30	8.06	35.82	35.43
Gammaproteobacteria	24.64	26.76	48.07	28.10	35.50	7.35	28.14	18.89
Zetaproteobacteria	1.56	5.98	3.28	7.30	1.06	0.37	1.54	1.56
Aquificae	11.98	20.31	1.23	1.57	11.96	4.35	4.87	3.09

Table 4.4. Bacterial phyla mean relative abundances (copies of genomes /10^6 base pairs) in the different biogeographic regions.

Phylum/Class	Atlantic Ocean		Indian Ocean		Mediterranean Sea		Pacific Ocean	
	Mean rel. abund.	SD	Mean rel. abund.	SD	Mean rel. abund.	SD	Mean rel. abund.	SD
Alphaproteobacteria	11.97	15.25	12.01	9.39	25.65	16.57	10.74	11.27
Betaproteobacteria	2.83	9.04	0.41	0.14	1.20	1.05	5.70	14.11
Deltaproteobacteria	9.12	20.52	7.45	5.15	7.25	4.49	19.30	21.40
Epsilonproteobacteria	28.59	34.44	7.41	8.19	28.63	29.60	26.22	28.69
Gammaproteobacteria	35.05	30.59	67.43	21.86	27.19	12.07	30.98	23.83
Zetaproteobacteria	0.34	0.38	0.91	0.21	1.46	1.20	4.35	8.94
Aquificae	12.09	20.89	4.38	5.63	8.62	5.16	2.70	7.07

The only significant differences across sample matrices regarded Alphaproteobacteria **(Figure 4.5),** Epsilonproteobacteria **(Figure 4.6)** and Gammaproteobacteria **(Figure 4.7)** classes.

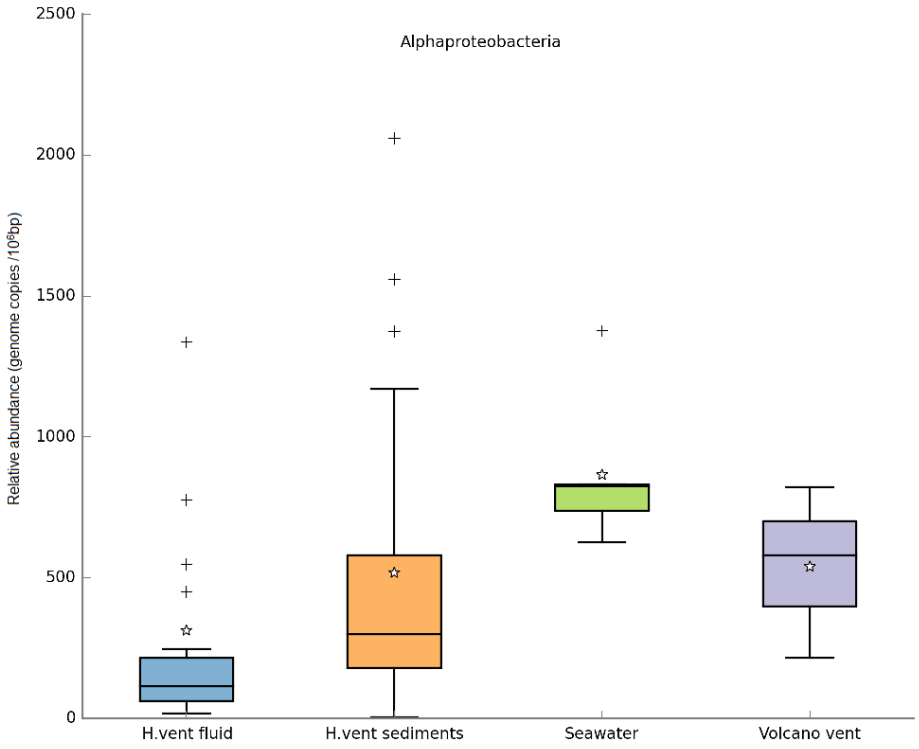

Figure 4.5 Class Alphaproteobacteria relative abundance across sample matrices (Statistical test: ANOVA, Post-hoc test: Tukey-Kramer, $p<0.005$)

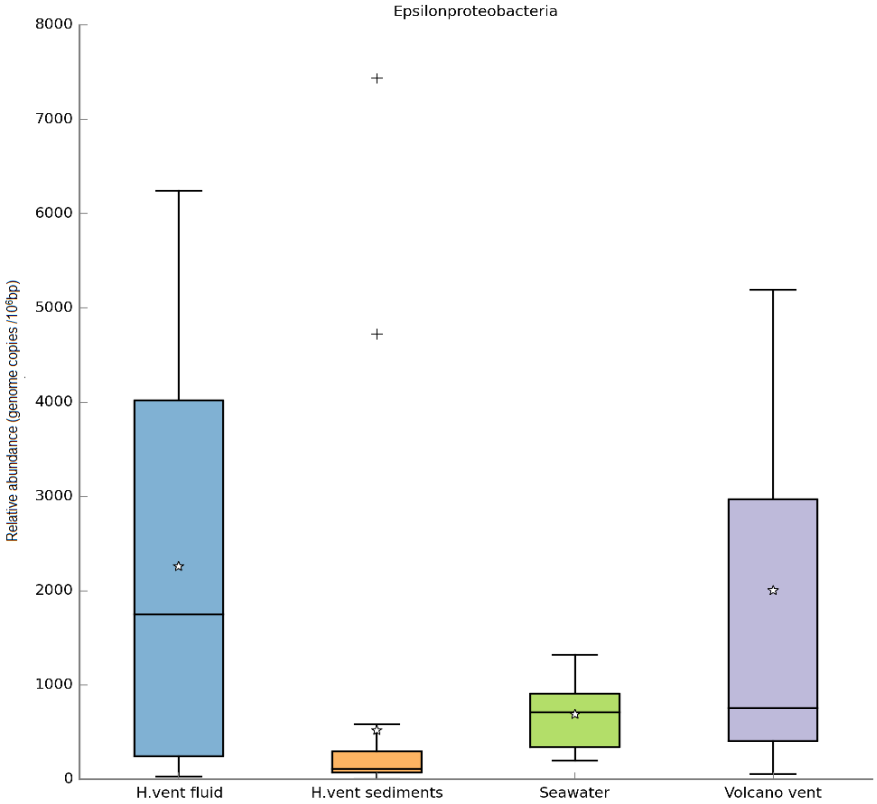

Figure 4.6 Class Epsilonproteobacteria relative abundance across sample matrices (Statistical test: ANOVA, Post-hoc test: Tukey-Kramer, p<0.005)

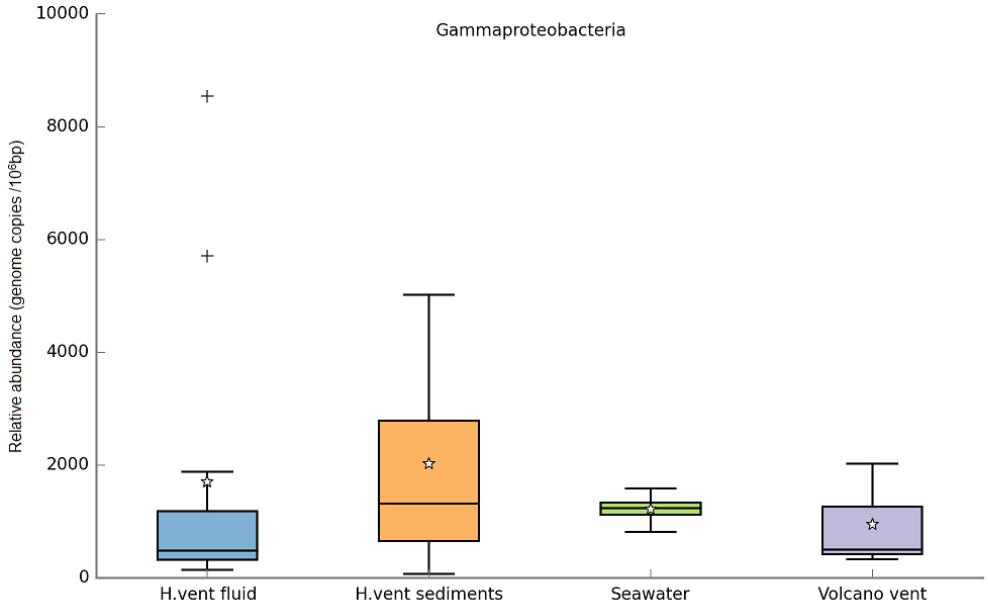

Figure 4.7 Class Gammaproteobacteria relative abundance across sample matrices (Statistical test: ANOVA, Post-hoc test: Tukey-Kramer, p<0.005)

The only significant differences across biogeographical regions regarded only the Gammaproteobacteria class **(Figure 4.8)**

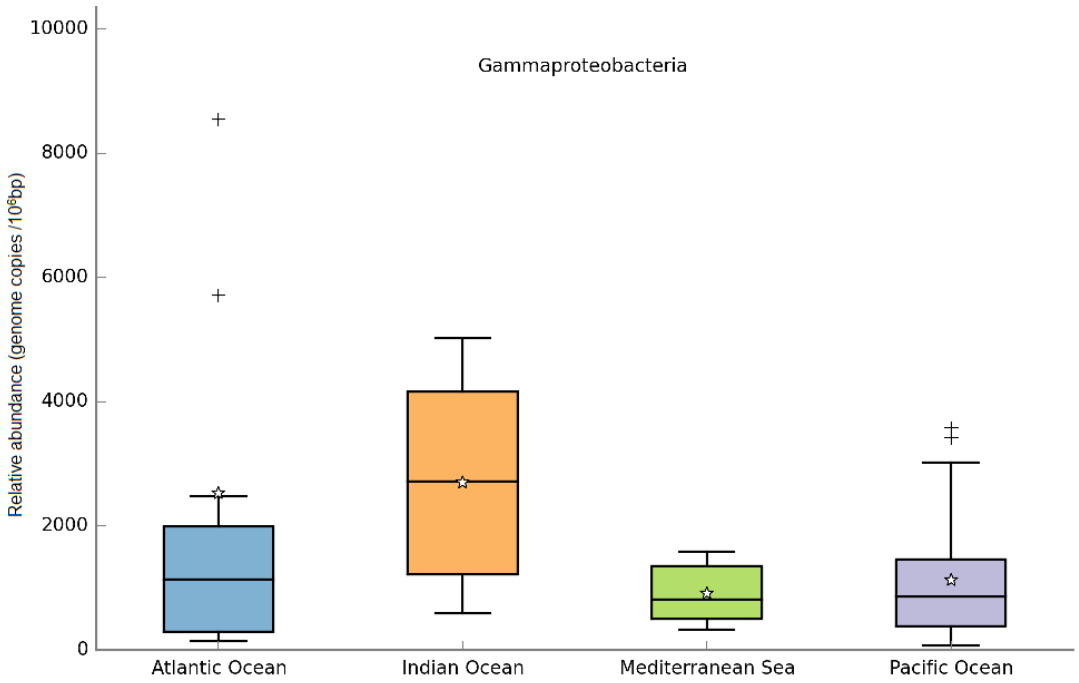

Figure 4.8 Class Gammaproteobacteria relative abundance across biogeographical regions (Statistical test: ANOVA, Post-hoc test: Tukey-Kramer, $p<0.005$)

4.2.2.3 General distribution

In order to correlate sample matrices, biogeographical region of provenance and biodiversity in our samples, we performed a Canonical Analysis of Principal Coordinates, both for biogeographical regions **(Figure 4.9)** and sample matrix **(Figure 4.10)**.

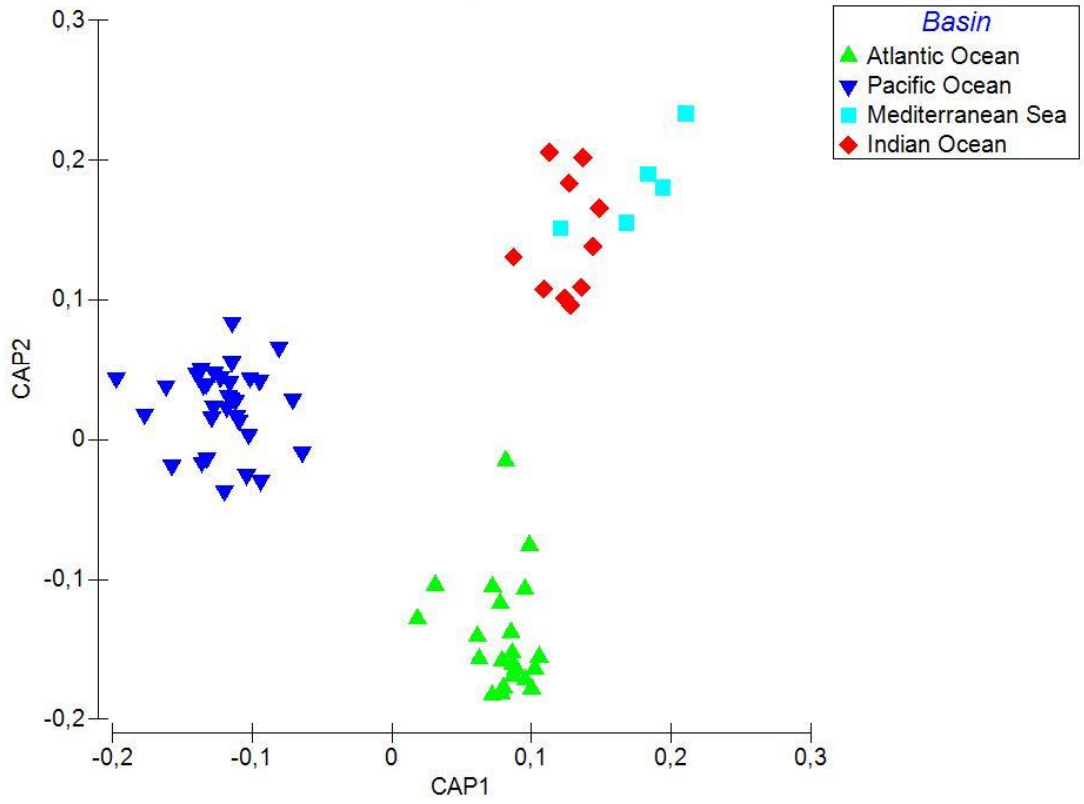

Figure 4.9. Canonical Analysis of Principal Coordinates, Biogeographical regions. Data transformation: Square root, resemblance: Bray Curtis similarity

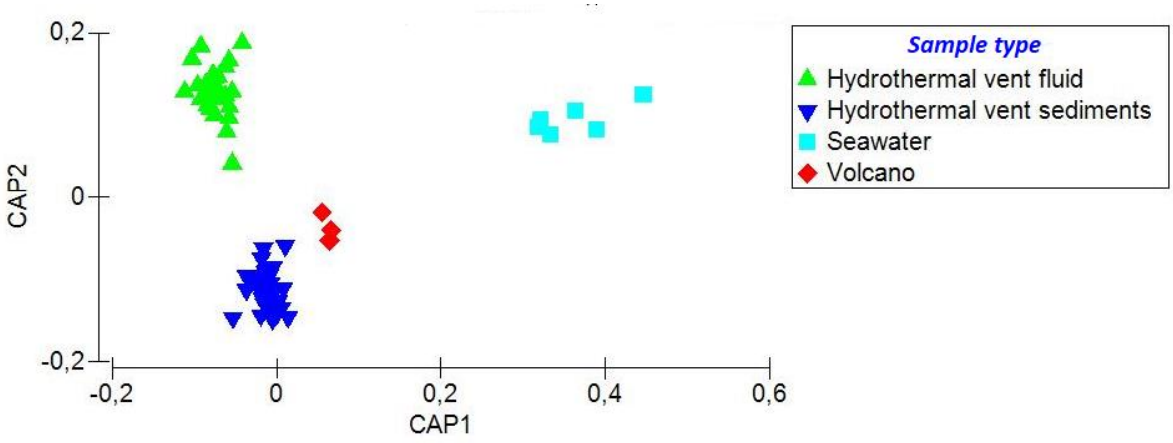

Figure 4.10. Canonical Analysis of Principal Coordinates, Sample matrix. Data transformation: Square root, resemblance: Bray Curtis similarity

4.3 Metabolic marker genes

We considered a total of 34 marker genes, involved in cellular processes and several metabolic pathways.

Only five genes were found in every bin: *recA, hyd, aprA, aprB* and *nifH.*

- *recA*, involved in the structural maintenance of DNA.
- *hyd*, a gene encoding for [FeFe]- hydrogenases, typical of anaerobic bacteria.
- *aprA* and *aprB*, encoding respectively for subunit A and B of adenylylsulfate reductase, a key enzyme of sulfur metabolism pathways.
- *nifH*, encoding for a structural polypeptides of nitrogenase, the (Fe)-protein or dinitrogenase reductase, which couples hydrolysis of ATP to electron transfer.

aclA, encoding for ATP-citrate lyase subunit A, key enzyme of the rTCA cycle, was found in 96.7% of the bins.

4.4 Secondary metabolites

We decided to concentrate on secondary metabolites that can have interesting biotechnological applications. For this reason, we focused on ectoine and the family of bacteriocins.

Ectoine (1,4,5,6-tetrahydro-2-methyl-4-pyrimidinecarboxylic acid) is a natural compound found in several species of bacteria (Peters et al., 1990). It serves as a protective substance by acting as an osmolyte and thus helps organisms survive extreme osmotic stress. It confers resistance towards high salt concentrations and high temperatures (Bernard et al. 1993).

Ectoine is used as an active ingredient in skin care and sun protection products. It stabilizes proteins and other cellular structures and protects the skin from stresses like UV irradiation and dryness (Stöveken et al.,2011)

Bacteriocins are ribosomally synthesized proteinaceous or peptidic toxins produced by bacteria to inhibit the growth of similar or closely related bacterial strains.

They are structurally, functionally, and ecologically diverse.

Applications of bacteriocins are being tested to assess their application as narrow-spectrum antibiotics (Cotter et al., 2012) and in food industry (Silva et al., 2018) As expected, bacteriocins were ubiquitarious, while ectoine was found in samples from eight different locations.

4.5 Prophages detection and clustering

Using VIRSorter we were able to detect 26 unclassified prophages in our bins. We concentrated our efforts on 12 of them which created novel clusters or were grouped with already identified ones, trying to understand more about their origin and their preferential hosts.

We found 5 prophages forming two novel exclusive clusters, VC_292(3 prophages) and VC_316 (2 prophages). We were not able to define preferential hosts for these two clusters.

The remaining 7 were part of existing ones. Preferential hosts of the prophages in this clusters are Bacteria belonging to Pseudomonas, Burkholderia and Salmonella genera.

5. Discussion

Deep-sea hydrothermal vents have been intensively investigated since their discovery (Corre et al., 2001; Juniper et al., 1998; Moyer et al., 1998; Nakagawa et al., 2004; Nercessian et al., 2005)

It is known that there is a wide range of microorganisms which can tolerate hot temperatures, low concentration of oxygen and high concentrations of reduced compounds. The study of the microbial community in extreme environments is particularly critical for a better comprehension of the adaptations of these organisms to non conventional life conditions.

Microbial colonization of freshly formed hydrothermal vents remains poorly understood but presumably involves hyperthermophiles that can disperse in cold seawater (Wirth, 2017). This scenario is supported by observations that hyperthermophiles are present in cold plumes (Huber et al., 1990) and that endemic vent taxa can be found in ambient seawater (Gonnella et al., 2016)

In this study, we investigated prokaryotic abundance and diversity in different hydrothermal vent areas characterized by different geographical locations, settings and type of fluid flow.

We found that the community structure composition was always dominated by the Bacteria , in every biogeographical region or sample matrix.

The differences in the presence of different phyla could be related to the nature and temperature of the fluid flow which could support different microbial taxa (Price et al., 2015).

The surprisingly high abundance of Epsilonproteobacteria in single volcanic samples and in hydrothermal vent plumes can be due to a transport caused by the rising plume from the chimney structures (Dick et al. 2019)

The significant difference in the abundance of Korarchaeota in single volcano samples and in the Mediterranean Sea is probably due to the reduced number of sampled sites and to peculiar geochemical characteristics which allow them to prosper. Notice that their mean relative abundance is still one order of magnitude below the one of Euryarchaeota. Further studies are needed in order to highlight this particular distribution of the phylum.

Considering the biogeographical region of provenance, the sample matrix and the taxonomical groups relative abundances, we saw that our samples cluster according to the abiotic variables.

The proximity and partial sovrapposition of the clusters from the Mediterranean Sea and Indian Ocean can be due to similar geochemical features between the sampling sites, but the low number of samples in the two biogeographical regions and a lack of data about the emissions and sediment nature in the areas doesn’t allow us to formulate further hypotheses.

Regarding the presence of metabolic marker genes, the ubiquitarious genes regarded sulfur and nitrogen metabolism, elements known for driving habitats based on chemosynthesis, showing the potential ability of different prokaryotes of using different energy sources if needed.

The high presence of the *aclA* gene, involved in the rTCA cycle, showed how, even if alternative carbon-fixation pathways exist, almost all prokaryotes mantain the potential capacity of performing it. In reality, other pathways are usually preferred in different conditions, because of their energetic efficiency and for kinetic reasons in the enzymatic processes.

The reductive tricarboxylic acid cycle (rTCA) in fact is a metabolic pathway for carbon fixation in which two molecules of carbon dioxide are converted into acetyl coenzyme A; it uses most of the same enzymes as the oxidative tricarboxylic acid cycle but runs it in reverse by using three alternative enzymes: fumarate reductase, 2-oxoglutarate synthase and ATP citrate lyase. Being energy- efficient yet oxygen- sensitive, rTCA often predominates at 20–90 °C, where oxygen levels are sufficiently low (yet potentially high enough to support microaerobic growth with oxygen as an electron acceptor). Alternative carbon-fixation pathways are the Calvin-Benson Bassham cycle, the Wood-Ljungdal pathway and the dicarboxylate–4-hydroxybutyrate pathway.

The Calvin–Benson Bassham cycle is a carbon fixation pathway in which carbon dioxide is converted into glyceraldehyde- 3-phosphate using a Rubisco isoform as a key enzyme. The CBB cycle is energetically costly yet oxygen-tolerant and predominates at temperatures <20 °C. Above 90 °C, most growing microorganisms are hyperthermophilic archaea that perform methanogenesis or sulfate reduction using the Wood–Ljungdahl pathway and the dicarboxylate–4- hydroxybutyrate pathway for carbon fixation, respectively. (Hugler and Sievert,2011). The Wood–Ljungdahl pathway is a metabolic pathway for carbon fixation in which two molecules of carbon dioxide are converted into acetyl coenzyme A by the key enzyme carbon monoxide dehydrogenase–acetyl coenzyme A synthase. Dicarboxylate–4-hydroxybutyrate pathway is a recently described carbon fixation pathway found in Archaea in which a molecule of bicarbonate is fixed onto acetyl coenzyme A via a combination of enzymes from the reductive tricarboxylic acid cycle and the 4-hydroxybutyrate part of the 3 hydroxypropionate– 4-hydroxybutyrate cycle.

Further studies on this metabolic pathways could provide a more comprehensive view of the different adaptation startegies used by Archaea and Bacteria in extreme environments.

Regarding the detection of secondary metabolites, we decided to include them in our study because of their potential biotechnological applications. The identification of their production methods and effects could lead to the development of new antibiotics and control methods of pathogenic bacteria, finding applications in the medical field.

Prophages were detected to try to understand if the antagonism between the phages and hosts can be a mechanism to increase the diversity of viral and microbial community structures. Further studies are needed in order to have a more comprehensive view of their abundance and of their host preferences.

6. Conclusions

- There is a similarity between vents from the same biogeographic region and among samples from the same matrix, even if it's not completely reflected considering the single taxonomical levels
- The almost ubiquitarious presence of specific marker genes confirms the potential ability of prokaryotes of adapting their metabolic strategies
- The ubiquitarious presence of specific secondary metabolites such as bacteriocins provides a novel perspective in the competition dinamics in this habitats

Further studies are needed to investigate more deeply the microbial diversity in such extreme environments, characterized by high spatial and temporal variability. These results add new insights for the comprehension of the microbial diversity in extreme marine environments.

7. References

Alneberg, J., Bjarnason, B. S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., ... & Quince, C. (2013). CONCOCT: clustering contigs on coverage and composition. arXiv preprint arXiv:1312.4038.

Anantharaman, K., Breier, J. A., & Dick, G. J. (2016). Metagenomic resolution of microbial functions in deep-sea hydrothermal plumes across the Eastern Lau Spreading Center. *The ISME journal*, 10(1): 225-239.

Anantharaman, K., Duhaime, M.B., Breier, J.A., Wendt, K.A., Toner, B.M., Dick, G.J. (2014). Sulfur oxidation genes in diverse deep-sea viruses. *Science* 344: 757–760.

Anderson, M. & Willis, Trevor. (2003). Canonical analysis of principal coordinates: A useful method of constrained ordination for ecology. *Ecology* 84: 511-525.

Anderson, M.J. (2005). Permutational Multivariate Analysis of Variance, thesis, Department of Statistics, University of Auckland, Auckland.

Azam, F., Fenchel, T., Field, J.G., Gray, J.S., Meyer-Reil, L.A., Thingstad, F., (1983). The ecological role of water-column microbes in the sea. *Marine Ecology Progress Series* 10: 257-263.

Bernard, T.; Jebbar, M.; Rassouli, Y.; Himdi-Kabbab, S.; Hamelin, J.; Blanco, C. (1993). "Ectoine accumulation and osmotic regulation in Brevibacterium linens" *(PDF). Journal of General Microbiology. **139**: 129–136.*

Bowers, R. M., Kyrpides, N. C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T. B. K., ... & Tringe, S. G. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature biotechnology*, 35(8): 725-731.

Boyd, J. A., Woodcroft, B. J., & Tyson, G. W. (2018). GraftM: a tool for scalable, phylogenetically informed classification of genes within metagenomes. *Nucleic Acids Research* 46(10): e59-e59.

Breier, J. A. et al. (2014). A large volume particulate and water multi- sampler with in situ preservation for microbial and biogeochemical studies. *Deep Sea Res.Part I Oceanogr. Res*. Pap. 94: 195–206

Breitbart M. (2012). Marine viruses: truth or dare. *Annu. Rev. Mar. Sci.* 4: 425– 448.

Cambuy, D.D., Coutinho, F.H., and Dutilh, B.E. (2016). Contig annotation tool CAT robustly classifies assembled metagenomic contigs and long sequences. bioRxiv

Campbell, B. J., Summers Engel, A., Porter, M. L. & Takai, K. (2006). The versatile ε- proteobacteria: key players in sulphidic habitats. *Nat. Rev. Microbiol*. 4: 458–468.

Clokie, M.R., Millard, A.D., Letarov, A.V., Heaphy S. (2011). Phages in nature. *Bacteriophage* 1:31– 45.

Corliss, J.B., and R.D. Ballard (1977). Oases of Life in the cold abyss, *Natl. Geogr. Mag*., 152 (4): 441-453.

Corre, E, Reysenbach AL, Prieur D (2001) ε-Proteobacterial diversity from a deep sea hydrothermal vent on the Mid-Atlantic Ridge. *FEMS Microbiol Lett* 205: 329–335

Cotter, Paul D.; Ross, R. Paul; Hill, Colin (2012). "Bacteriocins — a viable alternative to antibiotics?". *Nature Reviews Microbiology*. **11** (2): 95–105.

Dando, P.R., Dixon, D.R., Southward, A.J., Southward, E.C., Crawford, A., Crawford, M., (1992). Shipwrecked tube worms. *Nature* 356,667.

Dick, G. J. & Tebo, B. M. (2010). Microbial diversity and biogeochemistry of the Guaymas Basin deep- sea hydrothermal plume. *Environ. Microbiol*., 12: 1334–1347

Dick, G. J. et al. (2013). Hydrothermal vent plume microbiology: ecological and biogeographic linkages to seafloor and water column habitats. *Front. Microbiol* 4: 124

Dick, G. J. (2019). The microbiomes of deep-sea hydrothermal vents: distributed globally, shaped locally. *Nature Reviews Microbiology*, *17*(5), 271-283.

Flores, G. E. et al. (2012). Inter- field variability in the microbial communities of hydrothermal vent deposits from a back- arc basin. *Geobiology*, 10: 333–346.

Fortunato, C. S., Larson, B., Butterfield, D. A., & Huber, J. A. (2018). Spatially distinct, temporally stable microbial populations mediate biogeochemical cycling at and below the seafloor in hydrothermal vent fluids. *Environmental microbiology*, 20(2): 769-784.

Fuhrman, J. A. (1999). Marine virus and their biogeochemical and ecological effects. *Nature* 399: 541-548.

German, C. R. et al. (2010). Diverse styles of submarine venting on the ultraslow spreading Mid- Cayman Rise. *Proc. Natl Acad. Sci*. USA 107: 14020–14025

Goffredi, S. et al. (1997). Inorganic carbon acquisition by the hydrothermal vent tubeworm Riftia pachyptila depends upon high external PCO2 and upon proton equivalent ion transport by the worm. *J. Exp. Biol.,* 200: 883–896.

Gonnella, G. et al. Endemic hydrothermal vent species identified in the open ocean seed bank. *Nat. Microbiol.***1**, 16086 (2016).

Holden, J. F., Breier, J. A., Rogers, K. L., Schulte, M. D., & Toner, B. M. (2012). Biogeochemical processes at hydrothermal vents: microbes and minerals, bioenergetics, and carbon fluxes. *Oceanography*, 25(1): 196-208.

Huber, R., Stoffers, P., Cheminee, J. L., Richnow, H. H. & Stetter, K. O. Hyperthermophilic archaebacteria within the crater and open- sea plume of erupting Macdonald Seamount. *Nature* **345**, 179–182 (1990).

Hugler, M. & Sievert, S. M. (2011). Beyond the Calvin cycle:autotrophic carbon fixation in the ocean. *Ann. Rev. Mar. Sci*., 3: 261–289.

Johnson, K. S., Childress, J. J. & Beehler, C. L. (1988). Short- term temperature variability in the rose garden hydrothermal vent field - an unstable deep-sea environment. *Deep Sea Res*. A 35: 1711–1721.

Juniper, SK, Bird DF, Summit M, Vong MP, Baker ET (1998) Bacterial and viral abundances in hydrothermal event plumes over northern Gorda Ridge. *Deep-Sea Res* 45: 2739–2749.

Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., & Wang, Z. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 7: e7359.

Kelley, D. S., Baross, J. A., & Delaney, J. R. (2002). Volcanoes, fluids, and life at mid-ocean ridge spreading centers. *Annual Review of Earth and Planetary Sciences*, 30(1): 385-491.

Koskella, B., Brockhurst, M.A. (2014). Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol Rev* 38:916 –931.

Li, D., Liu, C. M., Luo, R., Sadakane, K., & Lam, T. W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, 31(10): 1674-1676.

Long, A., Patterson, S.S., Paul J.H. (2007). Macroarray analysis of gene expression in a marine pseudotemperate bacteriophage. *Aquat Microb Ecol* 49: 1–14.

Lonsdale, P. (1977), Clustering of suspension feeding macrobenthos near abyssal hydrothermal vents at oceanic spreading centers, *Deep Sea Res,* 24: 857-863.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, ... & Förster, W. (2004). ARB: a software environment for sequence data. *Nucleic acids research* 32(4): 1363-1371.

Mangiapia, M. et al. (2017). Proteomic and mutant analysis of the CO2 concentrating mechanism of hydrothermal vent chemolithoautotroph Thiomicrospira crunogena. *J. Bacteriol*. 199: e00871–16.

Medema, M. H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M. A., ... & Breitling, R. (2011). antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic acids research* 39(suppl_2): W339-W346.

Mora, M., Bellack, A., Ugele, M., Hopf, J. & Wirth, R. (2014). The temperature gradient forming device, an accessory unit for normal light microscopes to study the biology of hyperthermophilic microorganisms. *Appl. Environ. Microbiol*. 80: 4764–4770.

Moyer, CL, Tiedje JM, Dobbs FC, Karl DM (1998) Diversity of deep-sea hydrothermal vent Archaea from Loihi Seamount, Hawaii. *Deep-Sea Res, Part II* 45: 303–317

Nakagawa, T, Nakagawa S, Inagaki F, Takai K, Horikoshi K (2004) Phylogenetic diversity of sulfate-reducing prokaryotes in active deep-sea hydrothermal vent chimney structures. *FEMS Microbiol Lett* 232: 145–152

Nercessian, O, Bienvenu N, Moreira D, Prieur D, Jeanthon C (2005) Diversity of functional genes of methanogens, methanotrophs and sulphate reducers in deep-sea hydrothermal environments. *Environ Microbiol* 7: 118–132

Nunoura, T. et al. (2018). A primordial and reversible TCA cycle in a facultatively chemolithoautotrophic thermophile. *Science* 359: 559–563.

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

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

Paul, J.H. (2008). Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? *ISME J* 2:579 –589.

Peters, P; Miwatani, T; Honda, T (1990). "The biosynthesis of ectoine". *FEMS Microbiology Letters. **71** (2): 157–61.*

Price, R.E., LaRowe D.E., Italiano F, Savov I., Pichler T., Amend J.P., 2015. Subsurface hydrothermal processes and the bioenergetics of chemolithoautotrophy at the shallow-sea vents off Panarea Island (Italy) *Chemical Geology* 407–408: 21– 45.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., ... & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, 41(Database issue): D590.

Reed, D. C. et al. (2015). Predicting the response of the deep ocean microbiome to geochemical perturbations by hydrothermal vents. *ISME J.*, 9: 1857–1869.

Reveillaud, J., Reddington, E., McDermott, J., Algar, C., Meyer, J. L., Sylva, S., … & Huber, J. A. (2016). Subseafloor microbial communities in hydrogen-rich vent fluids from hydrothermal systems along the Mid-Cayman Rise. *Environmental microbiology*, 18(6): 1970-1987.

Reysenbach, A. L., & Shock, E. (2002). Merging genomes with geochemistry in hydrothermal ecosystems. *Science*, 296(5570): 1077-1082.

Reysenbach, A. L., Banta, A. B., Boone, D. R., Cary, S. C., & Luther, G. W. (2000). Microbial essentials at hydrothermal vents. *Nature*, 404(6780): 835-835.

Rezaei Javan, R., Ramos-Sevillano, E., Akter, A. et al. (2019). Prophages and satellite prophages are widespread in Streptococcus and may play a role in pneumococcal pathogenesis. *Nat Commun* 10: 4852

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

Schrenk, M. O., Huber, J. A., & Edwards, K. J. (2010). Microbial provinces in the subseafloor. *Annual review of marine science,* 2: 279-304.

Schrenk, M. O., Kelley, D. S., Bolton, S. A. & Baross, J. A. (2004). Low archaeal diversity linked to subseafloor geochemical processes at the lost city hydrothermal field, Mid- Atlantic Ridge. *Environ. Microbiol*., 6: 1086–1095.

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

Sheik, C. S. et al. (2015). Spatially resolved sampling reveals dynamic microbial communities in rising hydrothermal plumes across a back- arc basin. *ISME J*. 9: 1434–1445

Sievert, S. M., & Vetriani, C. (2012). Chemoautotrophy at deep-sea vents: past, present, and future. *Oceanography*, 25(1): 218-233.

Silva, C. C., Silva, S. P., & Ribeiro, S. C. (2018). Application of bacteriocins and protective cultures in dairy food preservation. *Frontiers in microbiology*, *9*, 594.

Stetter, K. O. (1999). Extremophiles and their adaptation to hot environments. *FEBS Lett*. 452: 22–25.

Steward, G.F., Preston, C.M. (2011). Analysis of a viral metagenomic library from 200 m depth in Monterey Bay, California constructed by direct shotgun cloning. *Virol J.* 8: 287

Stöveken, N; Pittelkow, M; Sinner, T; Jensen, R. A.; Heider, J; Bremer, E (2011). "A specialized aspartokinase enhances the biosynthesis of the osmoprotectants ectoine and hydroxyectoine in *Pseudomonas stutzeri* A1501". *Journal of Bacteriology*. **193** (17): 4456–68.

Suttle, C. A., (2005). Virus in the sea. *Nature* 437: 356-361.

Takai, K. & Nakamura, K. in Geomicrobiology: Molecular and Environmental Perspective (eds Loy, A., Mandl, M. & Barton, L. L.) 251–283 (Springer Netherlands, 2010).

Takai, K. et al. (2004).Geochemical and microbiological evidence for a hydrogen- based, hyperthermophilic subsurface lithoautotrophic microbial

ecosystem (HyperSLiME) beneath an active deep- sea hydrothermal field. *Extremophiles*, 8: 269–282

Thompson, J.R., Pacocha, S., Pharino, C., Klepac-Ceraj, V., Hunt, D.E., Benoit, J., Sarma-Rupavtarm, R., Distel, D.L., Polz, M.F. (2005). Genotypic diversity within He et al. ® July/August 2017 Volume 8 Issue 4 e00893-17 mbio.asm.org 12 on February 11, 2020 by guest http://mbio.asm.org/ Downloaded from a natural coastal bacterioplankton population. *Science* 307: 1311–1313.

Tivey, M. K. in The Subseafloor Biosphere at Mid-Ocean Ridges (eds Wilcock, W. S. D., DeLong, E. F. ,Kelley, D. S., Baross, J. A. & Cary, S. C.) 137–152 (*American Geophysical Union*, 2004).

Tunnicliffe, V., Barross, J.A., Gebruk, A.V., Giere, O., Holland, M.E., Koschinsky, A., Reysenbach, A.-L., Shank, T.M., (2003)a. Group report: what are the interactions between biotic processes at vent and physical, chemical, and geological conditions? In: Halbach, P.E., Tunnicliffe, V., Hein, J.R. (Eds.), Energy and Mass Transfer in Marine Hydrothermal Systems, *Dahlem Workshop Report, vol. 89. Dahlem Univ. Press, Berlin*, 251– 270.

Tunnicliffe, V., Juniper, S.K., Sibuet, M., (2003) b. Reducing environments of the deep-sea floor. In: Tyler, P.A. (Ed.), Ecosystems of the Deep Oceans. Elsevier, Amsterdam, 81– 110.

Uritskiy, G. V., DiRuggiero, J., & Taylor, J. (2018). MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 6(1): 1-13.

Van Dover, C.L., (2000). The Ecology of Deep-Sea Hydrothermal Vents. *Princeton University Press, Princeton.* 424 pages.

Von Damm, K. L. (1990). Seafloor hydrothermal activity: black smoker activity and chimneys. *Annual Review of Earth and Planetary Science*, 18: 173–204

Wilkinson, Bruce H.; McElroy, Brandon J.; Kesler, Stephen E.; Peters, Shanan E.;Rothman, Edward D. (2008). "Global geologic maps are tectonic speedometers—Rates of rock cycling from area-age frequencies". *Geological Society of America Bulletin*, 121(5–6): 760–779.

Wirth, R. Colonization of black smokers by hyperthermophilic microorganisms. *Trends Microbiol.* **25**, 92–99 (2017).

Wu, Y. W., Simmons, B. A., & Singer, S. W. (2016). MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32(4): 605-607.

Zimmermann, R., R. Iturriaga, and J. Becker-Birck. (1978). Simultaneous determination of the total number of aquatic bacteria and the number thereof involved in respiration. *Applied and Environmental Microbiology* 36: 926-935.