Metagenomic Investigation of Deep-Sea Microbial Diversity, Nitrogen Cycling Potential, and Response to Metal Exposure

Mariana da Fonseca Valquaresma
Master's in Bioinformatics and Computational Biology
Biology Department
Faculty of Sciences of the University of Porto
2024

Metagenomic Investigation of Deep-Sea Microbial Diversity, Nitrogen Cycling Potential, and Response to Metal Exposure

Mariana da Fonseca Valquaresma

Dissertation Project carried out as part of the Master's in Bioinformatics and Computational Biology

Biology Department

2024

Supervisor
Miguel Semedo, Junior Researcher, CIIMAR

Co-supervisor
Catarina Magalhães, Senior Researcher, CIIMAR and Auxiliar Professor at FCUP

Acknowledgements

First and foremost, I would like to express my gratitude to my supervisor Dr. Miguel Semedo and to my co-supervisor Dr. Catarina Magalhães for the opportunity to work on this project. I am deeply grateful for the valuable guidance and feedback provided by Dr. Miguel Semedo, which played a crucial role in the progress of my dissertation. Your enthusiasm and your belief in my capabilities has kept me motivated through this period.

Many thanks to Professor Pedro Ferreira and Professor Luís Castro. Your criticism and suggestions helped me refine my dissertation.

I would also like to thank the Microbiome Ecology and Biogeochemistry team of CIIMAR for their cordiality and insightful presentations. A special thanks to Leonor Pizarro for guiding me in the first days at the laboratory.

I am also thankful to all my professors at FCUP for teaching me important bioinformatic and computer science skills that supported the development of my dissertation and will be forever helpful in my future career.

Lastly, I would like to mention my colleagues Clara, Mariana, Margarida and Rafael. Thank you so much for your encouragement, support and friendship through all the hard-working times we spent together.

Fundação para a Ciência e a Tecnologia

This research was supported by national funds provided by FCT (Fundação para a Ciência e Tecnologia), through the projects DeepResist (2022.06475.PTDC) and N-Microarctic (2022.02983.PTDC), and by European funds through the project TwinDEEPS (101160298) and BioPROTEC (HORIZON-MISS-2023-OCEAN-01; Ref: 101157341). Ship time for deep-sea samples collection was supported by the Schmidt Ocean Institute (SOI) and the Crustal Ocean Biosphere Research Accelerator (COBRA) network funded by the US National Science Foundation (OISE-2114593).

Resumo

O crescente interesse na mineração no mar profundo, impulsionado pela necessidade de tecnologias digitais e energia renovável, levanta preocupações ambientais sobre a libertação de metais tóxicos que possam afetar comunidades microbianas. Os microrganismos do mar profundo são essenciais para o ciclo do azoto (N), sustentando a produtividade do ecossistema. Porém, o conhecimento sobre o seu papel e a sobre a sua suscetibilidade quando expostos a metais ainda é limitado. Neste sentido, o objetivo deste estudo é investigar a diversidade e o potencial de participação no ciclo do N em comunidades microbianas presentes em amostras de sedimentos recolhidas de montanhas marinhas em Dorado Outcrop (Oceano Pacífico), assim como avaliar os efeitos causados pela exposição destas amostras a diferentes concentrações de cádmio (Cd), um metal altamente tóxico que pode ser libertado durante operações de mineração no mar profundo. Além disso, dados metagenómicos brutos recuperados de bases de dados públicas (ENA em EMBL-EBI e MG-RAST) complementaram o presente estudo. De modo geral, os sedimentos de Dorado apresentaram evidências genómicas de um ciclo de N quase completo. Embora a abundância bacteriana tenha sido maior, a arqueia *Candidatus Nitrosopumilus* foi o género mais abundante em todas as amostras, sendo o principal responsável pela abundância predominante dos genes *nirK* e *pmoA/amoA*, responsáveis pela redução de nitrito e oxidação de amónia, respetivamente. Verificou-se também que o gene *nosZ*, o qual tem potencial para reduzir óxido nitroso (um potente gás com efeito de estufa), está presente em vários habitats do mar profundo noutras localidades, destacando-se a sua importância em ecossistemas bentónicos do mar profundo. Amostras de sedimentos submetidas a 96h de tratamento com Cd não exibiram uma resposta genómica significativa em termos de diversidade e potencial de ciclagem de N. Ainda assim, recomenda-se avaliar períodos de exposição mais longos de modo a confirmar se o Cd realmente não influência as comunidades microbianas bentónicas.

Palavras-chave: mar profundo, metagenómica, metabarcodificação, ciclo do azoto, cádmio

Abstract

The growing interest in deep-sea mining, driven by digital technology and renewable energy needs, raises environmental concerns about toxic metal release impacting microbial communities. Deep-sea microorganisms are essential for nitrogen (N) cycling, supporting ecosystem productivity, but knowledge about their roles and susceptibility to metal exposure is still limited. In this regard, the aim of this study is to investigate the diversity and N-cycling potential of microbial communities present in sediment samples collected from seamounts at Dorado Outcrop (Pacific Ocean), and evaluate the effects caused by exposure to different concentrations of cadmium (Cd), a highly toxic metal that may be released during deep-sea mining operations. Additionally, raw metagenomic data retrieved from public databases (ENA at EMBL-EBI and MG-RAST) supplemented the current study. Overall, genomic evidence of an almost complete N cycle is present in Dorado's sediments. Even though bacterial abundance was greater, the archaea *Candidatus Nitrosopumilus* was the most abundant genus across all samples, being the main contributor to the predominant genes *nirK* and *pmoA/amoA,* responsible for nitrite reduction and ammonia oxidation, respectively. It was also found that the *nosZ* gene, with the potential to reduce nitrous oxide (a potent greenhouse gas), is widespread across various deep-sea habitats at other locations, highlighting its importance in benthic deep-sea ecosystems. Sediment samples subjected to 96 hours of Cd treatment did not exhibit a significant genomic response in terms of diversity and N-cycling potential. Even so, it is recommended to assess longer exposure periods to confirm if Cd truly does not impact benthic microbial communities.

Keywords: deep-sea, metagenomics, metabarcoding, nitrogen cycle, cadmium

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AMO	Ammonia Monooxygenase
ANOVA	Analysis of Variance
ANR	Assimilatory Nitrate Reduction
AOA	Ammonia-Oxidizing Archaea
AOB	Ammonia-Oxidizing Bacteria
API	Application Programming Interface
ASV	Amplicon Sequence Variant
CCNIR	Cytochrome C Nitrite Reductase
CD	Cadmium
CIIMAR	Interdisciplinary Centre of Marine and Environmental Investigation
CTR	Control Group
CMPL	Completeness
CONT	Contamination
DNA	Deoxyribonucleic Acid
DNR	Dissimilatory Nitrate Reduction
DNRA	Dissimilatory Nitrate Reduction to Ammonium
EMBL-EBI	European Molecular Biology Laboratory – European Bioinformatics Institute
ENA	European Nucleotide Archive
FCUP	Faculty of Sciences of the University of Porto
GTDB	Genome Taxonomy Database
HAO	Hydroxylamine Oxidoreductase
HDH	Hydrazine Dehydrogenase
HSD	Honestly Significant Difference
HZS	Hydrazine Synthase
IQR	Interquartile Range
ISA	International Seabed Authority
KEGG	Kyoto Encyclopedia of Genes and Genomes
KO	KEGG Orthologs
MAG	Metagenome Assembled Genomes
MBSF	Meters Below Seafloor
MBSL	Meters Selow Sea Level
MMO	Methane Monooxygenase

N	Nitrogen
NAP	Periplasmic Nitrate Reductase
NAR	Transmembrane Nitrate Reductase
NAS	Assimilatory Nitrate Reductase
NCD	Non-Cyanobacterial Diazotrophs
NOR	Nitric Oxide Reductase
NOS	Nitrous Oxide Reductase
OMZ	Oxygen-Minimum Zones
PCR	Polymerase Chain Reaction
PCOA	Principal Coordinate Analysis
PERMANOVA	Permutational Analysis of Variance
RDP	Ribosomal Database Project
RNA	Ribonucleic Acid
RRNA	Ribosomal Ribonucleic Acid
ROV	Remotely Operated Vehicle
SOI	Schmidt Ocean Institute

1. Introduction

1.1. Characteristics of the Deep-Sea Environment

The oceans make up 71% of Earth's surface, with half of this vast area submerged at depths that exceed 3000 meters (Ramirez-Llodra et al., 2010). The deep-sea comprises about 90% of the oceans, including ecosystems distributed across an estimated volume of 1 billion km^3 of deep water and 326 million km^2 of deep seafloor (Ramirez-Llodra, 2020). The deep-sea is considered to start at a depth of 200 meters, where sunlight becomes insufficient to sustain primary productivity through photosynthesis, marking the beginning of the mesopelagic or "twilight" zone. Oxygen-minimum zones (OMZs) may occur in the mesopelagic zone, due to oxygen consumption in microbial degradation of organic matter that descends from the upper ocean layer. The largest proportion of the ocean is represented by the bathypelagic zone which starts at 1000 meters deep, comprising 75% of the ocean volume (Ramirez-Llodra et al., 2010). Further down the water column, the abyssopelagic zone extends from depths of 4000 to 6000 meters, followed by the hadalpelagic occurring in trenches and subduction zones (Rogers, 2015).

The benthic deep-sea holds unique habitats representing hot spots of marine diversity, such as hydrothermal vents, cold seeps, and seamounts (Rogers, 2015). The seabed provides both biological and geological resources of great interest (Ramirez-Llodra, 2020). However, despite technological advancements that have led to new discoveries concerning deep-sea biodiversity and ecosystems, scientific knowledge remains very limited (Amon et al., 2022; Miller et al., 2018; Orcutt et al., 2020; Smith et al., 2020) since the deep-sea is still one of the least explored areas on Earth (Ramirez-Llodra, 2020).

1.2. Environmental Impacts of Deep-Sea Mining

Given the global population's dependence on digital processes and the ongoing transition towards renewable energy, the demand for metals and rare earth minerals is expected to rise, resulting in increased interest in deep-sea mining (Braathen & Brekke, 2020; Orcutt et al., 2020). Deep-sea polymetallic minerals include ferro-manganese minerals, which are found in nodules and crusts, and seafloor massive sulfides formed by hydrothermal processes. Mineral deposits may contain metals like nickel, copper, cobalt, zinc, manganese, tellurium, lithium, cadmium, titanium, among other metals and rare earth elements (Braathen & Brekke, 2020; Rogers, 2015). These resources are being explored in the continental shelf of coastal states and in international seabed. The International Seabed Authority (ISA) has granted mining exploration contracts for polymetallic nodules in the Clarion-Clipperton Zone, Central Indian Ocean Basin and

West Pacific Ocean. Mining exploration contracts for sulfides in Atlantic and Indian Oceans, and cobalt-rich ferromanganese crusts in South Atlantic and West Pacific Oceans have also been granted (Amon et al., 2022).

Deep-seabed mining raises serious environmental concerns given the limited knowledge and data available on deep-sea habitats. Such activity is expected to cause benthic fauna removal, sediment plumes, chemical and toxic metal release, and increased noise, vibration, and light (Amon et al., 2022). Cumulative impacts may greatly affect the ecosystem, possibly leading to species extinctions and loss of ecosystem functions and services. Therefore, it is of the utmost importance to better understand the structure and biodiversity of deep-sea habitats in order to make informed decisions and establish appropriate regulations for deep-sea mining.

1.3. The Roles of Deep-Sea Microbial Communities

Most life in the deep-sea relies on heterotrophic processes. In habitats such as whale falls and OMZs, energy is provided by respiration of sinking organic matter (Orcutt et al., 2020). Alternatively, at hydrothermal vents and cold seeps, microorganisms use compounds such as hydrogen sulfide and methane as a source of energy to generate new organic matter from inorganic carbon in a process known as chemosynthesis (Orcutt et al., 2020; Ramirez-Llodra et al., 2010). Microbial communities influence the chemical cycling of carbon, nitrogen, iron and sulfur in the deep-sea (Orcutt et al., 2020), being essential for the sustenance of life. Any significant disruption to their chemosynthetic processes could profoundly impact all organisms inhabiting the ecosystem.

Beyond ecosystem production, deep-sea microbial communities provide various other ecosystem services of great importance. These services may contribute to the environment via nutrient cycling, absorbance of green-house gases, bioremediation and detoxification. Additionally, compounds found in deep-sea microorganisms may be beneficial to human health, due to potential antibiotic, antiviral and anti-cancer properties (Orcutt et al., 2020; Thurber et al., 2014).

1.4. Nitrogen Cycling in the Deep-Sea

Nitrogen (N) is necessary in vital cellular components like proteins and nucleic acids, playing a crucial role in the development of living organisms. In the deep-sea, bioavailable N is scarce (Kuypers et al., 2018), restricting the growth and biomass of organisms. N loss or retention within the ecosystem is mainly regulated by microbial reactions that modify its oxidation state from -3 in ammonium (NH_4^+) to +5 in nitrate (NO_3^-) (Kirchman, 2018). These redox transformations are a part of the N cycle, which can be summarized in 6 pathways (Figure 1): nitrogen fixation, assimilation, ammonification, nitrification, denitrification, anaerobic ammonium oxidation (anammox).

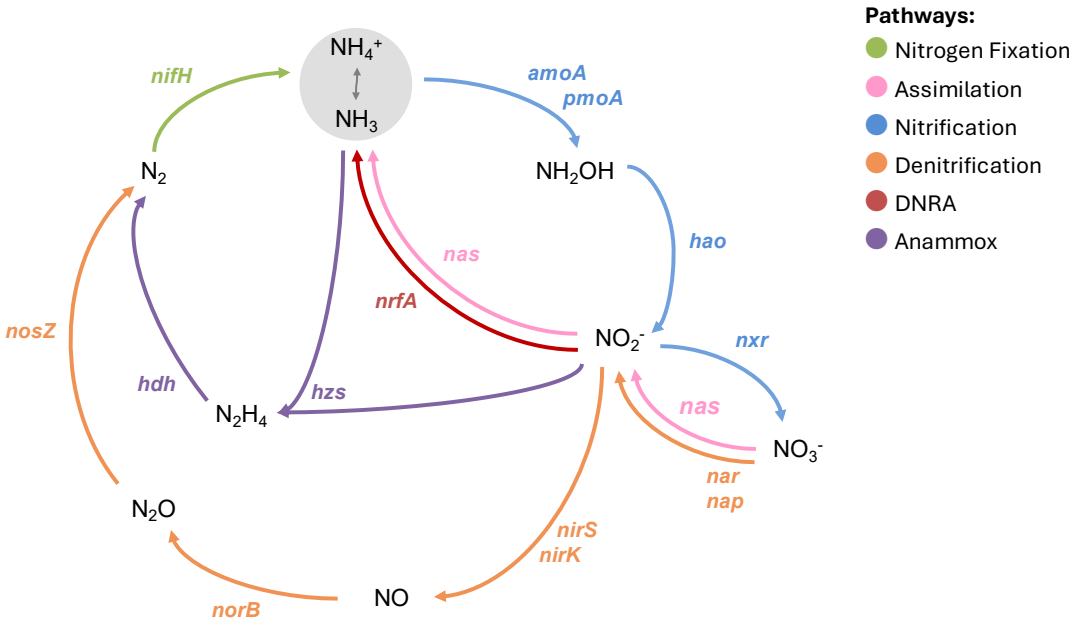

Figure 1 – Overview of the nitrogen cycle pathways and respective genes of interest associated with each redox reaction. This diagram is based on figures present in *Hallin et al.* (2018) and *Kuypers et al.* (2018). Ammonium (NH_4^+) and ammonia (NH_3) are often used interchangeably along explanations, since the difference of one proton between both compounds is solely dependent on ph.

Nitrogen fixation involves the reduction of dinitrogen gas (N_2) to ammonia (NH_3), catalyzed by nitrogenase, which is present in many Bacteria and Archaea (diazotrophs). Dinitrogenase reductase, encoded by *nifH*, is a component of this enzyme complex (Kuypers et al., 2018). It is found in microorganisms of the mesopelagic deep ocean, such as *Trichodesmium*, *Ca. Atelocyanobacterium thalassa* (UCYN-A), and *Crocosphaera watsonii* (UCYN-B) (Bonnet et al., 2023). Reduced N compounds may then be assimilated into organic N. Assimilatory nitrate (NO_3^-) and nitrite (NO_2^-) reduction (ANR) is carried out by assimilatory nitrate reductase (NAS) and assimilatory nitrite reductase, often encoded together on the *nas* operon (Kuypers et al., 2018). Ammonification, contrariwise, consists in the conversion of organic N into ammonium

(NH_4^+) via the mineralization of detrital proteins by deamination of amino acids (Kirchman, 2018; Kuypers et al., 2018).

In the deep-sea, nitrate is one of the predominant inorganic forms of fixed N (Kirchman, 2018). Its formation occurs through nitrification, an aerobic process in which ammonia is oxidized. Complete ammonia oxidation (comammox) is carried out by *Nitrospira* (Kirchman, 2018). Alternatively, nitrification may be divided in a two-step process. During the first step, ammonia oxidation to hydroxylamine (NH_2OH) is carried out mainly by ammonia monooxygenase (AMO) and less efficiently by methane monooxygenase (MMO). The genes *amoA* and *pmoA* encode AMO and MMO enzymes, respectively. Hydroxylamine oxidoreductase (HAO) then oxidizes hydroxylamine to nitric oxide (NO), which is finally oxidized to nitrite. Ammonia-oxidizing archaea (AOA) belong to the Thaumarchaeota phylum, and ammonia-oxidizing Bacteria (AOB) are mainly from Betaproteobacteria and Gammaproteobacteria classes (Kirchman, 2018; Kuypers et al., 2018). In the second step of typical nitrification, nitrite is oxidized to nitrate by nitrite oxidoreductase (NXR), which also catalyzes the reverse reaction (nitrate reduction to nitrite). This enzyme is found in Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria, Chloroflexi, Nitrospinae and Nitrospirae (Kuypers et al., 2018).

Denitrification involves the step-by-step anaerobic reduction of nitrate to dinitrogen. In anoxic environments like OMZs and marine sediments, microorganisms perform dissimilatory nitrate reduction (DNR) to nitrite. This reaction is catalyzed by a transmembrane nitrate reductase (NAR) or a periplasmic nitrate reductase (NAP). Nitrite is then reduced to nitric oxide by microorganisms such as Proteobacteria and Bacteroidetes (Hallin et al., 2018; Kuypers et al., 2018). This reaction is catalyzed by haem-containing or Cu-containing nitrite reductases, encoded by *nirS* and *nirK* genes, respectively. Subsequently, nitric oxide is reduced to nitrous oxide by nitric oxide reductases (NOR), encoded by the *norB* gene, and nitrous oxide (N_2O) is further reduced to dinitrogen by nitrous oxide reductases (NOS), encoded by the *nosZ gene*. NOS is used by Bacteria from Proteobacteria, Bacteroidetes, Chlorobi phylum, and Archaea from Crenarchaeaota and Halobacteria phylum (Hallin et al., 2018; Kuypers et al., 2018).

On the other hand, nitrate reduction can yield ammonium via dissimilatory nitrate reduction to ammonium (DNRA), carried out by the periplasmic cytochrome c nitrite reductase (ccNIR) encoded by the *nrf* gene. During this process, hydroxylamine remains bound to the enzyme as an intermediate until it is completely reduced to ammonium. (Kuypers et al., 2018)

The primary N sink in the ocean is anammox, which is carried out by Planctomycetes. Hydrazine synthase (HZS, encoded by *hzs*) is found in these bacteria, being the only enzyme known to activate ammonium anaerobically. Hydrazine (N_2H_4) is then oxidized to dinitrogen gas by hydrazine dehydrogenase (HDH), encoded by the *hdh* gene. HDH can only oxidize hydrazine and it might be inhibited by hydroxylamine. (Kuypers et al., 2018)

In summary, nitrogen fixation, retention, and loss are mediated by specialized microbial communities in different deep-sea habitats. The activity of enzymes involved in this cycle may be hindered by the availability of trace metals, given that they require metal cofactors, for instance iron in the case of diazotrophs (Casciotti et al., 2024; Pajares & Ramos, 2019). Other environmental factors, such as light, pH, oxygen, temperature, and salinity are known to affect differently each marine N-cycling pathway (Pajares & Ramos, 2019). There still are significant gaps in the understanding of the main N-cycling pathways in various deep-sea habitats, as well as the specific microorganisms that play a key role in these processes (Orcutt et al., 2020). Knowledge about potential impacts of deep-sea mining activities on these crucial N-cycling processes is even more limited, requiring urgent further investigation.

1.5. Research Aims and Objectives

The current study performs a metabarcoding analysis of deep-sea sediment samples from Dorado outcrop (Pacific Ocean), focusing on two main objectives: (1) estimation of alpha and beta diversity and inference of N-cycling potential of microbial communities present in Dorado's sediments; (2) evaluation of genomic responses to toxic metal exposure, more specifically, determine how different concentrations of cadmium (Cd) affect microbial diversity and N-cycling potential. Additionally, metagenomic data of global deep-sea sediments available in public databases was processed and used as support to the analysis. This allowed to explore global patterns of N-cycling potential within deep-sea microbial communities. These assessments are crucial and pertinent as seamounts are one of the least explored habitats targeted for mining (Amon et al., 2022), with limited information available on biodiversity and biogeochemical processes of deep-sea microorganisms. This research carries the potential to enrich the comprehension of deep-sea ecosystems, offering valuable insights into the consequences of mining activities on microbial communities.

2. Materials and Methods

2.1. Sample Collection and Cadmium Exposure

This study analyzes two sediment cores collected from Dorado outcrop in the Costa Rican Pacific Ocean, on June 3rd and 4th, 2023, with ROV SuBastian during the Octopus Odyssey expedition aboard R/V *Falkor (too)* from the Schmidt Ocean Institute (SOI). A researcher of CIIMAR's Microbiome Ecology and Biogeochemistry team participated in the expedition and processed the sediment samples. The samples were obtained at a depth of 3150 meters below sea level, gathering the top 10 cm of sea floor sediments. Sampling occurred at coordinates 9.07972ºS, 87.09722ºW and 9.0784ºS, 87.00983ºW. Upon collection, sediments were transferred to a sterile bag and homogenized. Both samples were divided into 15 × 25 cm^2 culture flasks for shipboard metal exposure experiments and triplicate 2 mL eppendorfs for background microbial genomic analysis. Eppendorfs for microbial genomics (environmental samples) were immediately stored at -80ºC. The 15 flasks correspond to 5 treatments groups with 3 replicates each, including a control group (CTR), without metal addition, and four different dissolved Cd concentrations: 0.015 µM, 0.15 µM, 1.5 µM, and 15 µM. Each flask received, approximately, 1.5 cm of sediment and 40 mL of site water collected with Niskin bottles and filtered with Sterivex filters (0.2 µm). After incubation overnight, 200 µL of the respective treatment were added to each flask. Following 96 hours of Cd exposure at 4°C, to recreate site water temperature, 2/4 of sediment was taken from each flask for DNA extraction and stored in sterile plastic bags at -80ºC.

2.2. DNA Extraction and Amplicon Sequencing

RNA and DNA extraction was carried out at CIIMAR, by members of the Microbiome Ecology and Biogeochemistry team, with the RNeasy PowerSoil Total RNA Kit and the Rneasy PowerSoil DNA Elution Kit (Qiagen), following the manufacturer's instructions. DNA extracts were sent for high-throughput amplicon sequencing at an external service provider. Illumina MiSeq paired-end 2x250bp sequencing was performed using the primers 515F-Y and 926R (Parada et al., 2016), which target the V4-V5 regions of the 16S rRNA gene.

2.3. Bioinformatic Procedure

2.3.1. Amplicon Sequencing Data Analysis Pipeline

16S rRNA sequencing data was processed and analyzed in R environment, utilizing the packages DADA2 v1.30.0 (Callahan et al., 2016), DECIPHER v2.30.0 (E. Wright S., 2016), Phyloseq v1.46.0 (McMurdie & Holmes, 2013), and Vegan v2.6-4 (Oksanen et al., 2022). The package ggplot2 (Wickham, 2016) was used for graphical visualization.

As primers had already been trimmed, the pipeline began with quality analysis, revealing that all samples had an average Q score higher than 30 for all positions. Consequently, the filtering step barely modified raw reads. PCR amplification and sequencing errors were estimated with the function *learnErrors* and amplicon sequence variants (ASVs) were identified in each sample using the denoising function *dada*. Reverse and forward reads were merged, and chimeras were removed. ASV classification was attempted with a combination of two methods and 4 training sets. More specifically, the functions *assignTaxonomy* from DADA2, and *IdTaxa* from DECIPHER were tested with the datasets: SILVA v128 and v138.1 (Quast et al., 2012), GTDB r202 and r214 (Parks et al., 2018), as well as RDP r18 (Cole et al., 2014). Chloroplast, Mitochondrial and Eukaryote sequences were removed from the classification results.

Alpha diversity was calculated with Shannon index, species richness with Chao 1 index, and evenness with Pielou index. Beta diversity was assessed through Principal Coordinate Analysis (PCoA) with Bray-Curtis distances. Homogeneity within samples treated with the same Cd concentration was determined with the function *betadisper* and statistically tested with functions *permutest* and *anova*. Differences between these sample groups were estimated with permutational ANOVA, using the function *adonis2*. Means of different pairs of sample groups were compared with Tukey's Honest Significant Difference (HSD) test.

Functional inference of ASVs was carried out with PICRUSt2 full pipeline (Barbera et al., 2018; Czech et al., 2020; Douglas et al., 2020; Louca & Doebeli, 2018; Mirarab et al., 2011; Ye & Doak, 2009). The resulting KEGG orthologs (KO) were filtered to obtain only those involved in the N cycle, according to the genes of interest described in Table 1. Abundance values were normalized according to *recA* (K03553) abundance, a single-copy gene present in all bacteria. Differences between sample groups were assessed with one-way ANOVA and Tukey's HSD test, utilizing the Vegan package in R.

Table 1 – Overview of genes and proteins of interest involved in the nitrogen cycle, linked to corresponding KEGG Orthologs (KO) and metabolic pathways in which they take part.

Gene	Protein	KO	Pathway
narG, narZ, nxrA	Nitrate Reductase, Nitrite Oxidoreductase	K00370	DNR, Denitrification, Nitrification
napA	Cytochrome Nitrate Reductase	K02567	DNR, Denitrification
narB	Ferredoxin-Nitrate Reductase	K00367	Assimilation
NR	Assimilatory nitrate reductase	K10534	Assimilation
nasA	Catalytic Subunit of Assimilatory Nitrate Reductase	K00372	Assimilation
nrfA	Cytochrome c-552 Nitrite Reductase	K03385	DNRA
nirK	Nitrite Reductase	K00368	Denitrification
nirS	Nitrite Reductase	K15864	Denitrification
norB	Nitric Oxide Reductase	K04561	Denitrification
nosZ	Nitrous Oxide Reductase	K00376	Denitrification
nifH	Nitrogenase	K02588	Nitrogen fixation
pmoA-amoA	Methane/Ammonia Monooxygenase	K10944	Nitrification
hao	Hydroxylamine dehydrogenase	K10535	Nitrification
hzs	Hydrazine synthase	K20932	Anammox
hdh	Hydrazine dehydrogenase	K20935	Anammox

2.3.2. Database Search and Download of Public Metagenomes

The public databases MG-RAST (Glass et al., 2010), European Nucleotide Archive (ENA) at EMBL-EBI, and MGnify (Richardson et al., 2023) were accessed through the provided API service. A python script was developed for the retrieval, filtering, and organization of available metagenome metadata and raw shotgun sequence download. This was accomplished utilizing the packages *Requests* (Reitz, 2023), *Pandas* (McKinney, 2010; The pandas development team, 2020), *NumPy* (Harris et al., 2020), and *Beautiful Soup* (Richardson, 2023), as well as *GeoPandas* (Jordahl et al., 2020), *Cartopy* (Met Office, 2023) and *Matplotlib* (Hunter, 2007) to visualize sampling location.

Due to incoherences and missing metadata, samples and reads were filtered in two steps. Firstly, through API calls, search results were filtered to include public metagenome samples that include “marine” or “sea” as biome and “sediment” as material, containing only reads obtained with whole genome sequencing (shotgun sequencing) by an Illumina platform. Assembled metagenomes were excluded from the results with the intention of only obtaining minimally-processed data, allowing for the comparison between sample sets. After organizing the resulting metagenome metadata in a data frame, samples were further filtered according to depth, aiming to keep only samples of sediments collected at 200 meters or more below sea surface. Note that, in the explored databases, depth of the sediment samples is often defined as the distance below the seafloor and the main target of this filtering step would be elevation, i.e. the distance of the sampling site from mean sea level. However, these values were often not correctly assigned in metadata and sometimes were missing or misplaced. In some rare cases, despite the sample being classified as marine sediment, depth was annotated in the altitude field, which should be a value measured above the sea level. When available, the corresponding published study was consulted in an attempt to fill missing depth values. Samples were manually sorted according to the environmental feature term, excluding those that do not associate with the deep-sea, such as “coastal” or “estuary”, among others. Therefore, the depth filter was based on the depth, elevation, altitude, and environmental feature fields. Finally, raw sequences were retrieved, resulting in the exclusion of samples without available download.

2.3.2. Shotgun Sequencing Data Analysis Pipeline

Firstly, the quality of the samples was assessed with FastQC v0.12.1 (Andrews, 2010) and MultiQC v1.14 (Ewels et al., 2016). Trimmomatic v0.39 (Bolger et al., 2014) was used to trim the raw metagenome sequences, removing adapter sequences, short reads (<36bp), and reads with an average quality score bellow 15 within 4-base windows. Sequences were assembled with MEGAHIT v1.2.9 (D. Li et al., 2016), using the meta-large preset and minimum contig length of 500 bp. Raw sequences related to the same environmental sample were co-assembled during this step. The resulting contigs were binned with MetaBAT2 (Kang et al., 2019) to obtain metagenome assembled genomes (MAGs). The completeness and contamination of each bin was verified utilizing CheckM’s lineage-specific workflow v1.2.2 (Parks et al., 2015). Only bins with medium quality (≥50% completeness and ≤10% contamination) or high quality (≥90% completeness and ≤5% contamination) were kept, constituting the final MAGs. These were taxonomically classified with GTDB-Tk v2.1.1 (Chaumeil et al., 2020; Eddy, 2011;

Hyatt et al., 2010; Jain et al., 2018; Matsen et al., 2010; Ondov et al., 2016; Price et al., 2010; Shaw & Yu, 2023) and assigned to gene calls and protein-coding genes with PROKKA v1.14.6 (Seemann, 2014).

3. Results

There are three main data sets processed and analyzed during this study: (1) environmental 16S rRNA gene sequencing data from Dorado background sediments; (2) 16S rRNA gene sequencing data of samples from Dorado sediments exposed to different Cd concentrations; and (3) raw shotgun metagenome sequences downloaded from public databases.

3.1. Amplicon Sequencing Data Processing

During amplicon data processing, the initial number of amplicon sequence variants (ASVs) inferred with *dada* method was 25994. After removing chimeras, 18626 ASVs remained. Removed sequences only represented approximately 5% of total reads. More detailed results for each sample can be consulted in Attachment 1.

Since the taxonomy assignment was performed with different methods and training sets, the resulting ASVs number differs slightly for each approach. It was observed that the function *assignTaxonomy* from DADA2 (default arguments) with GTDB training set allowed to maintain the 18626 ASVs of the previous step, since no chloroplasts, mitochondria nor eukaryotes were identified (Table 2). Furthermore, this approach was the one with least unknowns in genus classification. In comparison, the results obtained with the function *IdTaxa* from DECIPHER presented the highest percentage of unclassified genera and did not include any mitochondria nor eukaryotes. The combination of the function *assignTaxonomy* with the training set SILVA v138.1 was chosen to proceed with the analysis of amplicon sequencing data, since its result holds the second lowest percentage of unclassified genera (81.37%). This choice is further justified given that this approach results in a lower number of unnamed genera denoted by a code, a common occurrence when utilizing GTDB data.

Table 2 - Final number of ASVs per classification approach and respective percentage of unclassified genera, i.e. ASVs the model was unable to classify. These results include all 6 environmental samples. The approach highlighted in blue was chosen to continue the analysis of amplicon sequencing data.

Function	Training Set	ASVs	Unclassified ASVs at Genus Level (%)
assignTaxonomy	SILVA v128	18598	84.75
assignTaxonomy	SILVA v138.1	18593	81.37
assignTaxonomy	RDP r18	18620	81.90
assignTaxonomy	GTDB r202	18626	59.00
IdTaxa	SILVA v138	18619	89.99
IdTaxa	RDP r18	18621	96.25
IdTaxa	GTDB r214	18624	89.41

3.2. Background Microbial Community Characterization

3.2.1. Dorado's Sediments Taxonomic Composition and Abundance

From the 18593 ASVs of all 6 environmental sample replicates, 16944 belong to Bacteria and 1635 belong to Archaea. The model was unable to classify the remaining 14 ASVs. Bacteria outnumber archaea in each sample, with bacteria having a mean relative abundance of 85.04% compared to 14.96% for archaea.

The most abundant phyla, in descending order of mean relative abundance, are Proteobacteria (29.14%), Crenarchaeota (14.32%), Planctomycetota (9.32%), Chloroflexi (9.05%), NB1-j (7.72%), Actinobacteriota (7.39%), and Acidobacteriota (6.32%). Less abundant phyla (<3%) include Gemmatimonadota, Bacteroidota, Nitrospinota, Myxococcota, Dadabacteria, and Desulfobacterota. Others, like Verrucomicrobiota, Methylomirabilota, Latescibacterota, Schekmanbacteria, Nanoarchaeota, and Nitrospirota have mean relative abundance between 0.4% and 1%. At phylum level, the model was unable to classify 1029 ASVs, which constitute less than 0.5% of each replicate.

Relative abundance was graphically visualized at both class (Figure 2) and genus (Figure 3) levels per sample replicate. The most abundant classes are Gammaproteobacteria (16.22%), Nitrososphaeria (14.29%), Alphaproteobacteria (12.91%), Acidimicrobiia (7.12%), making up roughly half of each sample. Less abundant classes include Planctomycetes (4.12%), Dehalococcoidia (4.02%), Anaerolineae (3.08%), Phycisphaerae (3.07%), and Vicinamibacteria (2.06%). 3233 ASVs were not classified at class level, corresponding once more to less than 0.5% of each replicate.

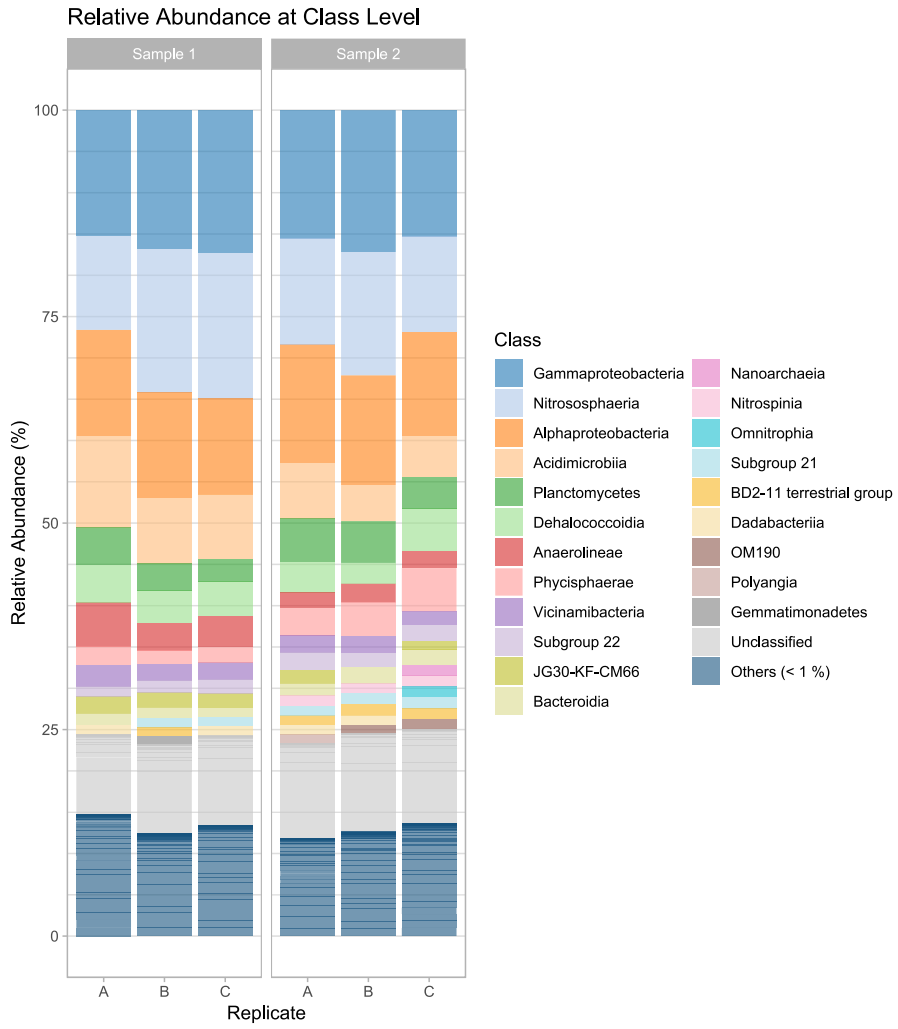

Figure 2 - Relative abundances and taxonomic composition at class level of Bacteria and Archaea present in Dorado's sediments. Includes all classified and unclassified ASVs of environmental samples.

At order level, Nitrosopumilales exhibits higher mean relative abundance (14.06%), followed by Steroidobacterales (7.06%), Actinomarinales (6.98%), and Kiloniellales (5.13%). Subsequently, the most abundant families are Nitrosopumilaceae (14.06%), Woeseiaceae (7.06%), and Kiloniellaceae (5.13%).

Most ASVs were not classified at genus level, constituting an average of around 69% of each replicate. Most unclassified genera belong to the classes Alphaproteobacteria, Acidimicrobiia and Gammaproteobacteria. The archaea *Candidatus Nitrosopumilus* is the most abundant among classified genera in all samples (9.51%), followed by *Woeseia* (7.02%), Pir4 lineage (2.64%), and AqS1 (1.58%). Urania-1B-19 marine sediment group has much lower relative abundance in replicates of sample 1 (0.47%), when compared to sample 2 (1.99%). Less abundant genera include *Nitrospina*, *Limibaculum*, *Ca. Scalindua*, *Ca. Omnitrophus*, *Nitrosomonas*, *Nitrospira*, and *Pelagibius*.

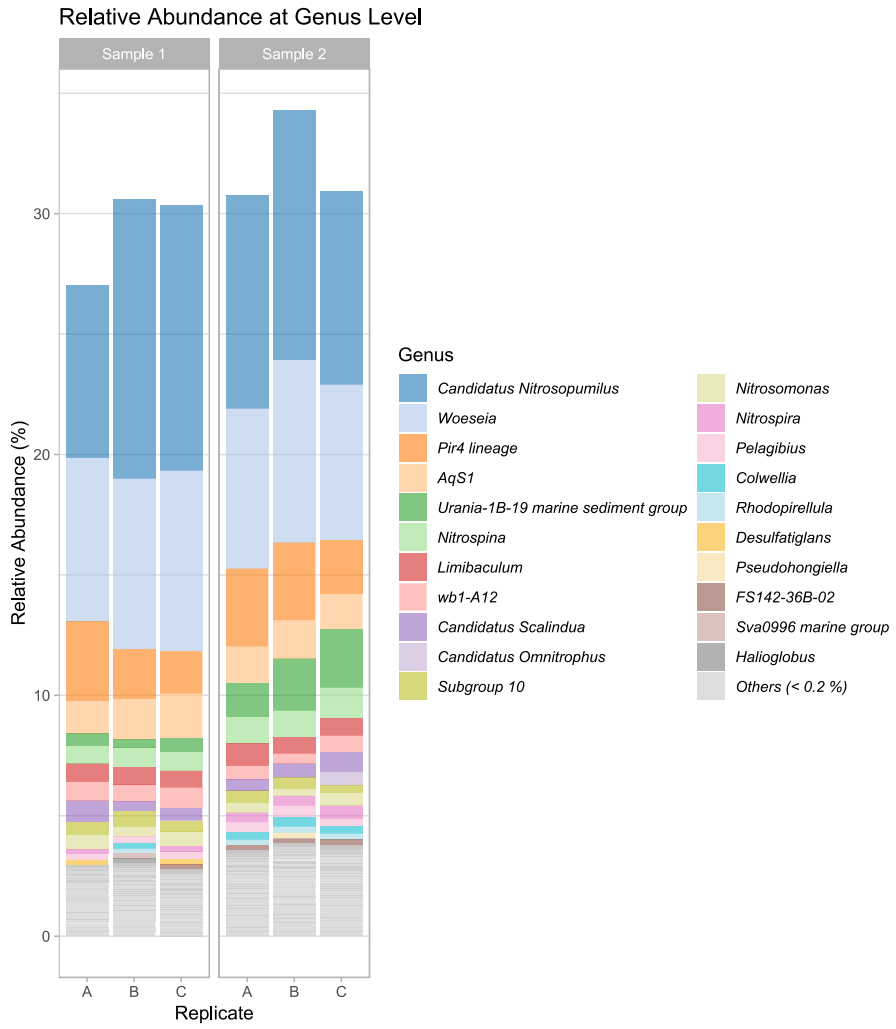

Figure 3 - Relative abundances and taxonomic composition at genus level of Bacteria and Archaea present in Dorado's background sediments. ASVs with unclassified genera were omitted from the graphic to improve clarity, but still considered when calculating relative abundance.

3.2.2. Alpha and Beta Diversity Patterns in Dorado's Sediments

Diversity, species richness and evenness, seem to be higher within sample 2 when compared to sample 1 as showcased by Shannon, Chao 1 and Pielou indexes, respectively (Figure 5). Still, these differences were not statistically significant, as described by the non-parametric Wilcoxon rank-sum test, which resulted in a p-value greater than 0.05. PCoA also seemed to indicate a dissimilarity between samples (Figure 4). The analysis of multivariate homogeneity of group dispersions (*betadisper*) demonstrated that within sample 1 the average distance to the median is 0.1848, while in sample 2 this value equals to 0.22. After testing these results with a permutation test and ANOVA, p-values of 0.1014 and 0.2055 were obtained, respectively. These values possibly indicate high variability between replicates within each sample group. PERMANOVA also showed a statistically non-significant outcome, as indicated by a p-

value of 0.1, failing to support the apparent difference observed during PCoA. However, these results are not very sensitive, as there are only 3 replicates from each environmental sample. Therefore, this analysis is repeated in the following section, with Cd treated samples, with a higher sampling effort.

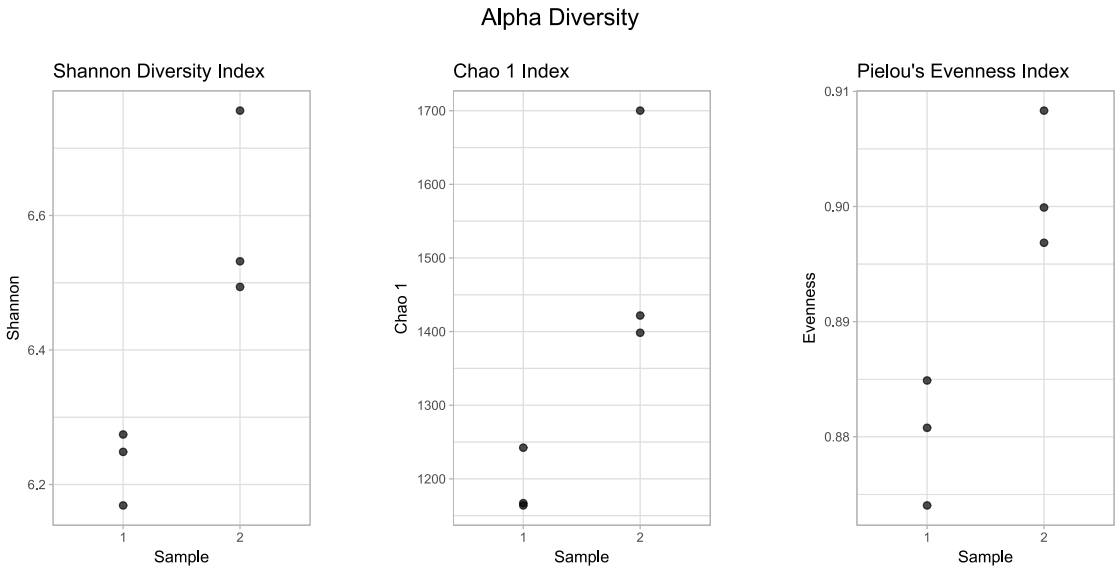

Figure 5 - Alpha diversity of each replicate of samples collected at Dorado outcrop, represented by the Shannon Diversity, Chao 1 and Pielou's Evenness indexes.

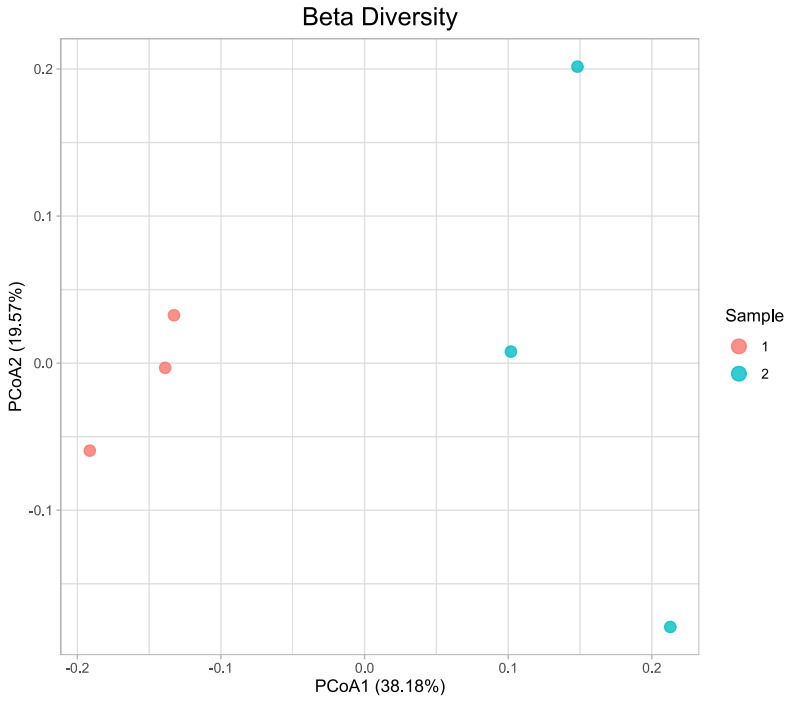

Figure 4 – Beta diversity represented by PCoA with Bray-Curtis distances between each replicate of samples collected at Dorado outcrop, including the percentage of variation explained by each axis.

3.2.3. Nitrogen Cycling Potential of Dorado's Microbial Communities

Based on data obtained from PICRUSt2 analysis, it was found that *nirK* has the highest relative abundance among the genes associated with the N cycle in each sample (Figure 6). This was observed to be even higher than the expectedly more common genes *napA*, *narG*, *narZ*, *nxrA*, and *nasA*, related with the redox reactions of transforming nitrate into nitrite and vice versa. Contrastingly, the *nirS* gene, which also encodes a dissimilatory nitrite reductase, has the lowest mean relative abundance. The genes *pmoA* and *amoA*, are identified as the second most abundant, with *hao* also present completing the nitrification pathway. *NrfA*, n*orB* and *nosZ* have much lower relative abundance, as well as *nifH*. The anammox genes *hdh* and *hzs* were not found by PICRUSt2.

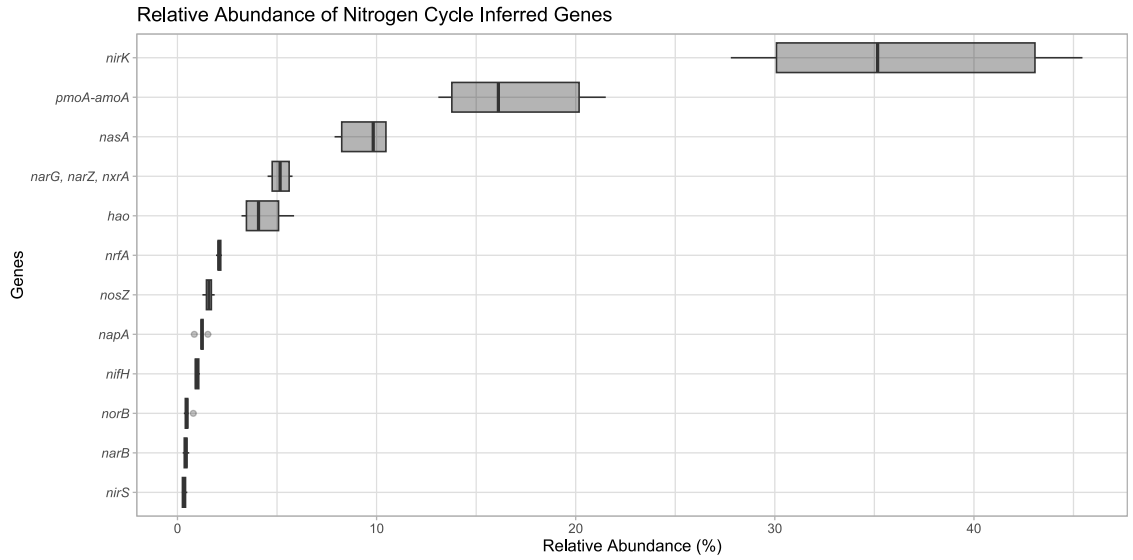

Figure 6 - Relative abundance distribution of genes inferred by PICRUSt2, present in Dorado's sediment samples. Relative abundance was calculated dividing abundance values of each gene by the abundance of *recA* gene in each sample. The relative abundance values of the 6 replicates of each gene are graphically represented by boxplots. Box length depicts the interquartile range (IQR), with whiskers extending to the minimum and maximum values within 1.5 times the IQR, and potential outliers marked as individual points. It is possible to observe that *nirK* displays higher relative abundance values, which also have a larger distribution, when compared to other genes.

In Figure 7, the contribution of each taxon to the different N-cycling genes is shown. Unclassified taxa are prevalent in most N cycle pathways. Unclassified genera belonging to the class Subgroup 22 (phylum Acidobacteriota) and to the order Syntrophobacterales predominantly contribute *nifH* abundance in each sample.

As regards nitrifying microorganisms, *Ca*. *Nitrosopumilus* and *Nitrosomonas* are the main contributors to *pmoA/amoA* abundance, followed by *Ca. Nitrosopelagicus* and unclassified genera belonging to family Nitrosopumilaceae and order Burkholderiales. Additionally, *Ca. Scalindua* and *Nitrosomonas* contribute the most to *hao* abundance.

Hao was also found in other unclassified genera of Gammaproteobacteria, including the orders Burkholderiales and BD7-8 as well as the family Kangiellaceae.

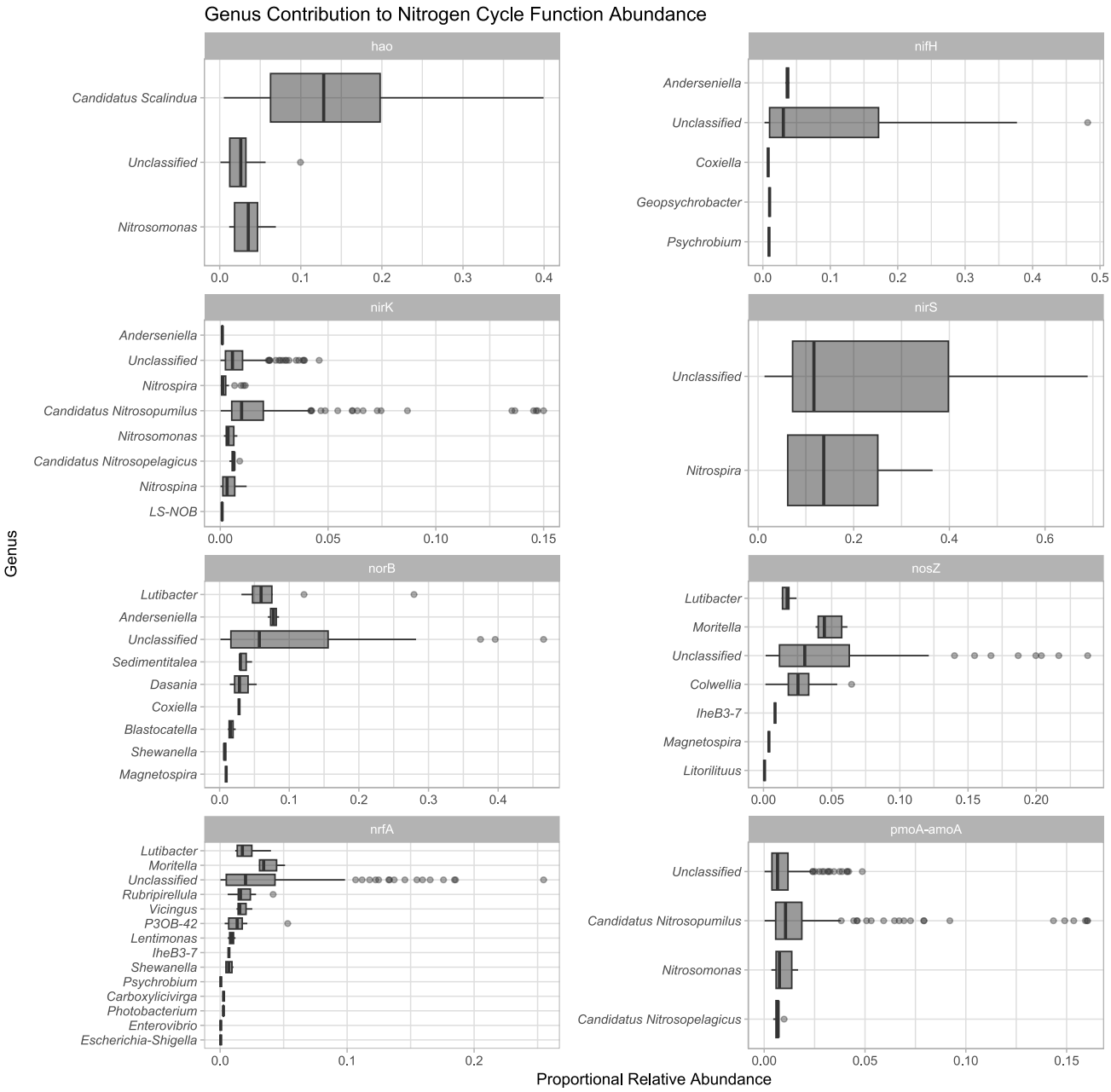

Figure 7 - Proportional relative abundance distribution, representing the contribution of each genus to the abundance of nitrogen cycle related genes in each sample of Dorado's sediment. Each data point corresponds to an ASV and is graphically represented by boxplots. Box length depicts the interquartile range (IQR), with whiskers extending to the minimum and maximum values within 1.5 times the IQR, and potential outliers marked as individual points.

In the context of denitrifying microorganisms, *nirK*'s abundance in each sample is mostly attributable to *Ca. Nitrosopumilus*. Other unclassified archaea from family Nitrosopumilaceae also contain *nirK*. Therefore, at family level, Nitrosopumilaceae

predominantly contribute to *nirK*, alongside *Nitrosomonadaceae*, *Nitrospiraceae* and *Nitrospinaceae*.

Nitrospira contributes way more to *nirS* abundance than to *nirK* abundance. Unclassified genera from orders Burkholderiales and BD7-8, as well as families Kangiellaceae and Nitrosomonadaceae contribute almost as much to *nirS* abundance as *Nitrospira*.

Nitric oxide reduction mediated by *norB* is mainly led by *Lucibacter*, *Anderseniella* and unclassified genera that fall on family Magnetospiraceae and orders Rhodospirillales and UBA 10353 marine group. Amongst classified genera, *nosZ* abundance is largely assignable to *Moritella* and *Colwellia*. *Lucibacter* contributes less to *nosZ* than to *norB* abundance. A wide range of genera that substantially partake in *nosZ* abundance is unclassififed. In fact, unclassified genera of families Kangiellaceae, PHOS-HE36 (order Ignavibacteriales), and Magnetospiraceae contribute more to *nosZ* abundance than *Moritellaceae* or *Colwelliaceae*. *Moritella* and *Lucibacter* contribute less to *nrfA* abundance than to *nosZ*.

Most genera that contribute a lot to DNRA via *nrfA* gene are unclassified. These fall on the order Syntrophobacterales, which is the second main taxon involved in *nrfA* abundance, followed by order Ignavibacteriales. Acidobacteriota Subgroup 22 predominantly contribute to *nrfA* abundance.

3.3. Cadmium Exposure

3.3.1. Taxonomic Composition and Abundance after Cadmium Exposure

Taxonomic composition is almost similar, when compared to background samples, with Gammaproteobacteria, Nitrososphaeria and Alphaproteobacteria dominating all sample replicates. However, there is a sudden appearance of the genus *Stenotrophomonas* (class Gammaproteobacteria), especially in sample 2, greatly surpassing *Ca. Nitrosopumilus* (Figure 8). It is possible to observe that sample 2 replicates with *Stenotrophomonas* account for a much smaller prevalence of unclassified genera. Other taxa that were not identified in background samples but are present in Cd experiment samples are *Ralstonia* and *Delftia* (both also from class Gammaproteobacteria), particularly in sample 2. These three Gammaproteobacteria seem to co-occur in replicates and share the same abundance variations.

In terms of mean relative abundance, grouping data by Cd treatment, most genera become gradually less abundant with increase in Cd concentration. Nevertheless, when

looking at samples individually, *Ca. Nitrosopumilus*, *Woeseia* and *Colwellia* relative abundances increase with Cd concentration in replicates B and C of sample 2.

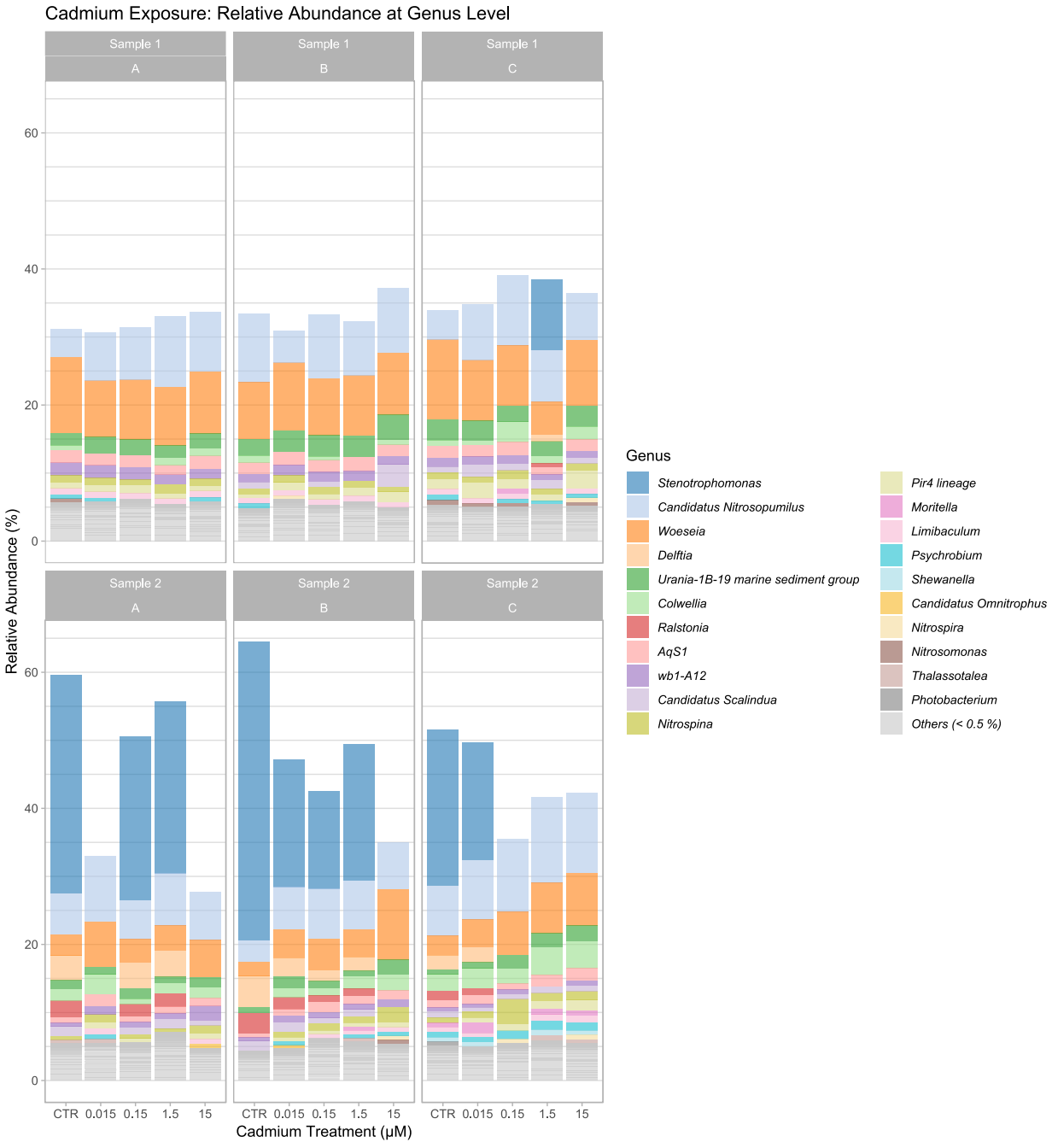

Figure 8 - Relative abundances and taxonomic composition at genus level of Bacteria and Archaea present in Dorado's sediment samples after 96h of Cd exposure at different concentrations. CTR refers to the control group (no cadmium). ASVs with unclassified genera were omitted from the graphic to improve clarity, but still considered when calculating relative abundance.

3.3.2. Alpha and Beta Diversity Responses to Cadmium Exposure

There is not a clear impact in alpha diversity with the increase of Cd concentration (Figure 9). The distribution of Shannon diversity in sample 1 remains similarly even with each different treatment. Despite richness values oscillating, the Shannon index remains high. Sample 2 showed higher variation in its replicate's species richness and diversity, yet still not demonstrating evident treatment patterns. When considering Pielou's index, however, sample 2 appears to have an increase in species evenness with the increase in Cd concentration.

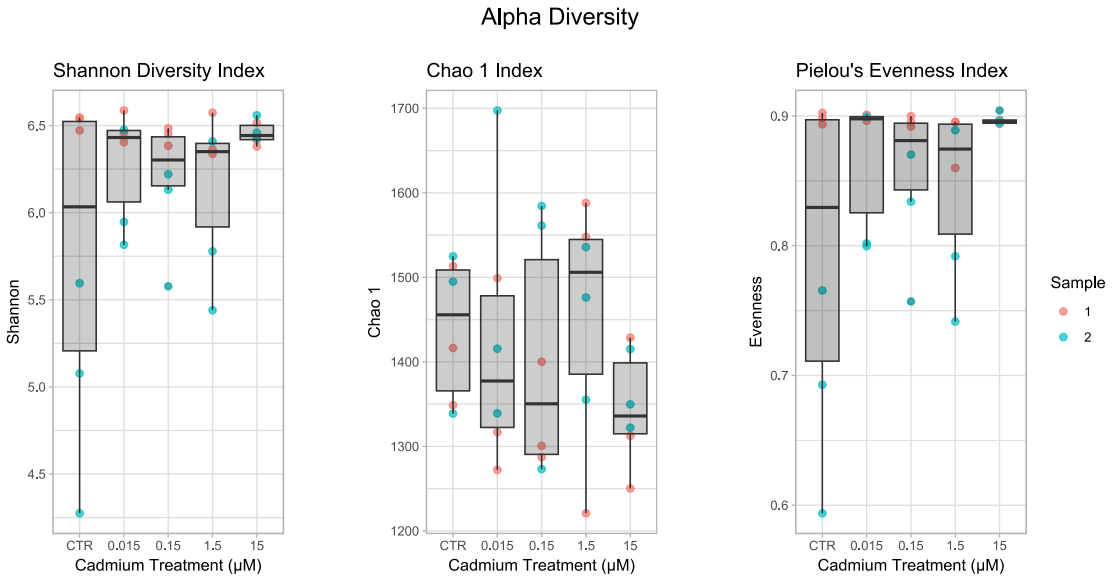

Figure 9 - Alpha diversity of each replicate of samples exposed to different Cd treatments, represented by the Shannon Diversity, Chao 1 and Pielou's Evenness indexes. CTR refers to the control group, i.e. sample not treated with Cd.

The Principal Coordinate Analysis (PCoA) further validates the disparities between samples 1 and 2, and confirms no observable separation according to Cd treatments (Figure 10). It is possible to notice in the results of the analysis of multivariate homogeneity of group dispersions (*betadisper*) (Table 3) that average distance to centroids of each treatment group is similar, even though it appears to be slightly smaller for the highest Cd concentration. The permutation test and ANOVA resulted in p-values much greater than 0.05, specifically 0.3197 and 0.3187, indicating that there is high homogeneity or little variance between samples of the same treatment group. Moreover, PERMANOVA test (*adonis*) also resulted in a p-value greater than 0.05 (0.6116), which suggests once more that Cd concentration may have no significant impact in microbial diversity. Tukey's HSD test was performed, yet also leading to no significant differences between all possible pairs of group means.

Table 3 - Betadisper results: average distance to median of replicates of different Cd treatment groups.

Cadmium Treatment (µL)	CTR	0.015	0.15	1.5	15
Average Distance to Median	0.3771	0.3218	0.3374	0.3306	0.3032

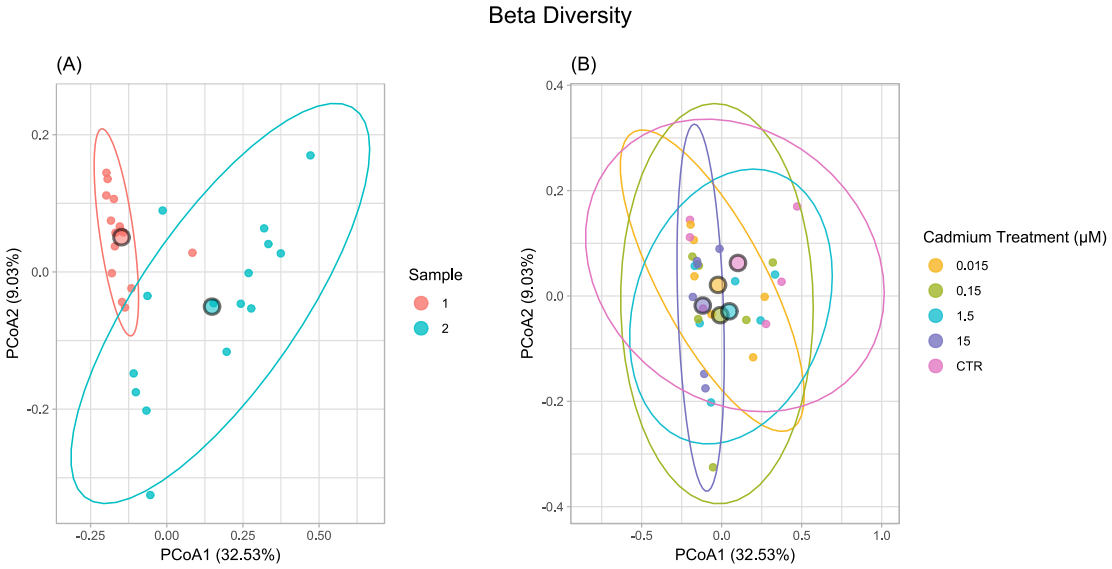

Figure 10 - Beta diversity represented by PCoA with Bray-Curtis distances between each replicate of samples exposed to different Cd treatments, including the percentage of variation explained by each axis. Each group's centroid is represented by a larger point. Ellipses represent a confidence region for the mean of the data points, assuming a multivariate t-distribution, with size determined by a level of confidence of 95%. (A) Beta diversity analysis of each sample group. (B) Beta diversity analysis of each Cd treatment group.

3.3.3. Changes in Nitrogen Cycling Potential with Cadmium Exposure

Increase in Cd concentration does not result in any apparent effect on estimated relative abundance of N cycle genes (Figure 11). ANOVA was performed, returning a p-value of 0.801, possibly confirming the previous statement. As ANOVA considered all N cycle inferred genes, this test was repeated for each individually, although the results remained statistically insignificant as expected. To further compare the differences between specific treatments, Tukey's HSD tests were performed, still not presenting any significant differences between groups.

Despite the lack of significant differences, one can attempt to identify emerging patterns related to the metal exposure. Most detected genes have very similar median values of relative abundance across treatment (Figure 11). However, *nirK* and pmoA/*amoA* appear to have elevated abundances with Cd exposure, while *hao* appears to have lower relative abundances with increasing concentrations of Cd, even though no significant differences were detected. These observations require further testing with more replicates to evaluate if these residual differences are consistent with an increasing sampling effort.

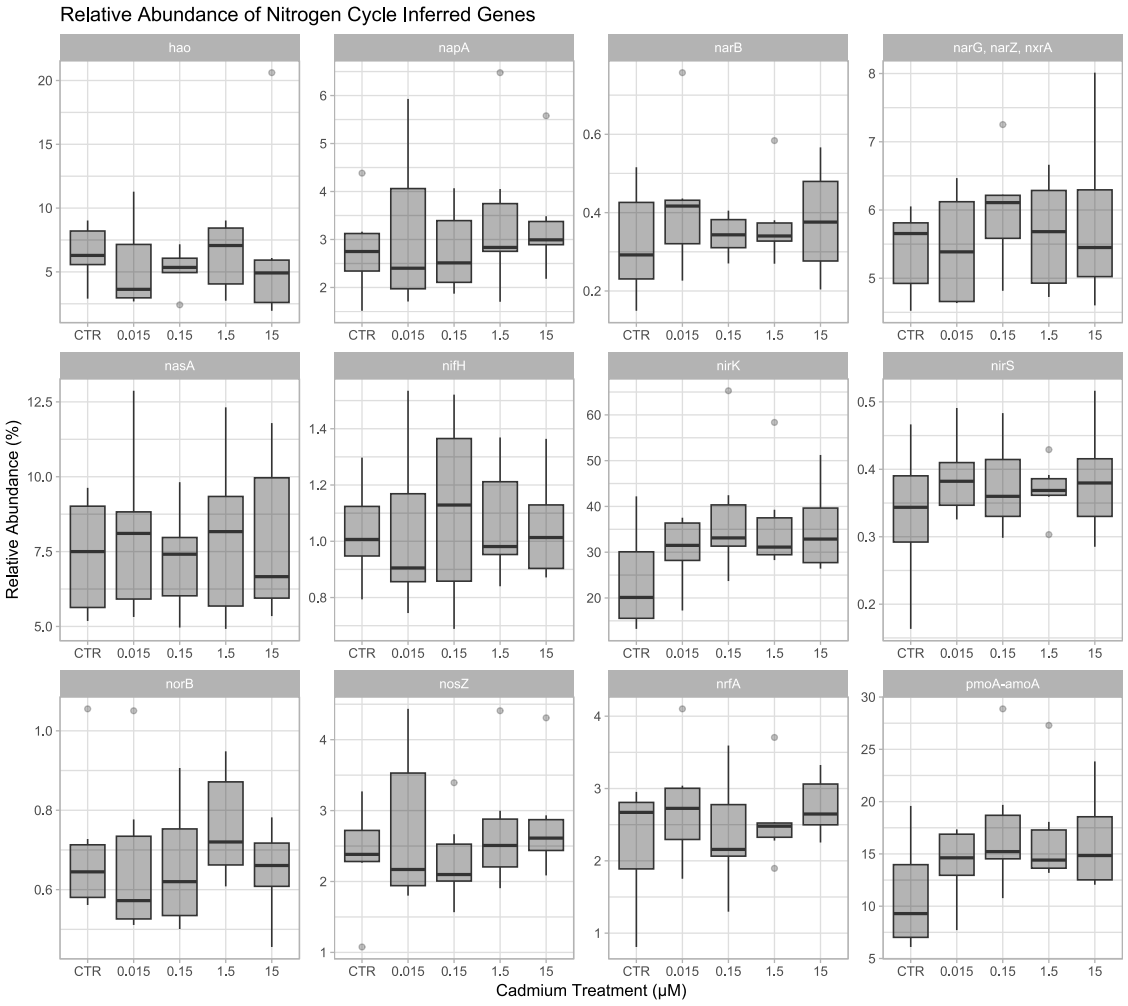

Figure 11 - Relative abundance distribution of nitrogen cycle associated in each Cd treatment group. CTR refers to the control group (no cadmium).

3.4. Public Metagenomic Shotgun Sequencing Data

After querying public metagenome databases for shotgun sequencing data from deep-sea marine sediments and filtering the resulting metadata, 277 metagenome entries were obtained: 79 from MG-RAST, 198 from ENA, and 0 from MGnify. All metagenomes from MGnify ended up being excluded because of unclear metadata. Only 20 MG-RAST entries and 46 ENA entries had available raw reads for download (Figure 12; Attachment 2). These resulting 66 metagenomes were organized and classified according to depth and habitat. As there were only samples collected deeper than 1000 mbsl, they were grouped into the categories: bathypelagic (1000 m – 4000 m), abyssopelagic (4000 m – 6000 m), and hadopelagic (>6000 m). With the available information on metadata and published studies, samples were also divided into different habitat types: abyssal plains and continental slopes, hydrothermal vents, cold seeps (including methane seeps, oil

seeps, and mud volcanos), subduction zones (trenches), and unclassified habitats. There was no confirmed seamount habitat among the retrieved samples.

It should be noted that there were several raw reads related to the same ENA samples, specifically SAMEA5663119, SAMEA6457547, SAMEA7725637, and SAMN03002196. For each of these samples, raw reads were co-assembled during the assembly step. Therefore, this resulted in a total of 48 contig assemblies. After binning, only 33 samples originated at least 1 bin, summing to a total of 406 bins. Quality assessment revealed that just 81 bins had medium quality (≥50% completeness, and ≤10% contamination). No high-quality bin (≥90% completeness, and ≤5% contamination) was detected. The final 81 MAGs belong to 15 metagenome samples and were used for further N-cycling gene analysis.

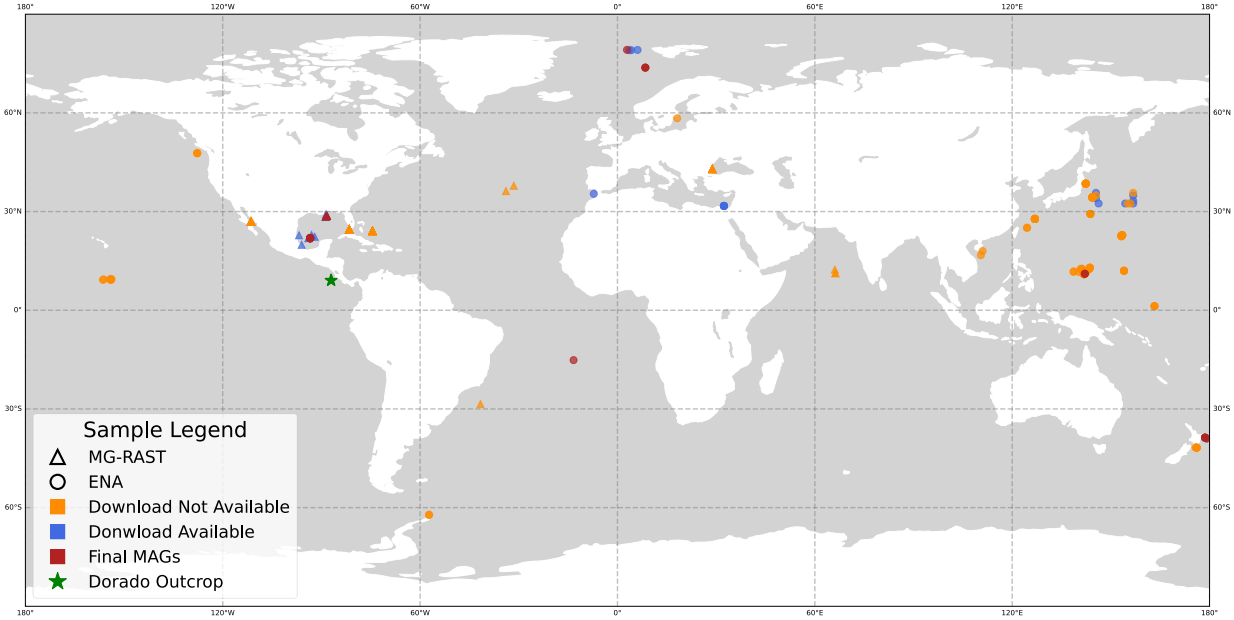

Figure 12 - Location of marine sediment samples available in the databases MG-RAST and ENA, as well as the current study sampling location (Dorado outcrop).

Table 4 and Attachment 3 showcase taxonomic classification, denoting the presence and absence of genes of interest involved in the N cycle. In total, 5 MAGs were classified as Archaea and 76 MAGs were classified as Bacteria. N-cycling genes of interest were only found in 48 of these MAGs. Overall, there were more MAGs of subduction zone and trenches (29 MAGs), though it was only possible to detect N-cycling genes in about half of these MAGs. Hydrothermal vents were the habitat with higher number of MAGs that presented N-cycling genes (25 MAGs). Amongst all 81 final MAGs, the bathypelagic was the most prevalent zone (63 MAGs) and all habitats and pelagic zones had at least one MAG comprising at least one of the N-cycling genes of interest.

Table 4 - Taxonomic classification of MAGs obtained from ENA and MG-RAST data and presence of nitrogen cycle genes, highlighted in blue. This table only lists the genes of interest that were found among PROKKA's results, excluding the undetected genes: *amoA*, *hdh, nxrA*, and NR. The detected genes comprise: N_2 reduction (*nifH*); aerobic and anaerobic NH_3 oxidation (*pmoA*, *hzs*); NO_2^- reduction (*nirK, nirS, nrfA*); assimilatory and dissimilatory NO_3^- reduction (*nasA*, *nar, napA*); NO reduction (*norB*); and N_2O reduction (*nosZ*). Percentage of MAG completeness (CMPL) and contamination (CONT) is colored in orange, with a darker color representing lower CMPL values and higher CONT values. Notice that only 46 genera are represented (out of 48 MAGs), since two *Halopseudomonas* species were found, comprising different bins. Thus, for this genus, CMPL and CONT values indicate an average value between the two MAGs. The only archaeal MAG present in this table belongs to class Methanosarcinia. The pathways associated with each gene are displayed in Table 1 and data of other MAGs can be consulted in Attachment 3.

Zone	Habitat	Sample ID	Class	Genus	CMPL %	CONT %	Nitrogen Cycle Genes												
							narG	*napA*	*narB*	*nasA*	*nrfA*	*nirK*	*nirS*	*norB*	*nosZ*	*nifH*	*pmoA*	*hao*	*hzs*
Bathypelagic (1000m - 4000m)	**Cold Seep**	SAMEA5663119	Methanosarcinia	UBA204	72	0.66		■											
			Spirochaetia	JAHOYX01	88	4.85		■											
			UBA6098	Unclassified	84	2.20		■											
	Continental Slope and Abyssal Plain	mgm4510162.3	Alphaproteobacteria	*Amylibacter*	52	0.86	■	■		■	■								
		mgm4510164.3	Alphaproteobacteria	*Amylibacter*	78	1.68	■	■		■	■								
				Sneathiella	55	5.33		■											
			Gammaproteobacteria	50-400-T64	61	8.10					■								
		mgm4510168.3	Gammaproteobacteria	*Agaribacterium*	93	3.79			■										
		mgm4510171.3	Gammaproteobacteria	*Colwellia*	56	8.62		■							■		■		
				UBA3067	87	9.99	■		■				■	■	■		■		
		mgm4510172.3	UBA6911	Unclassified	69	3.04	■	■											■
	Hydrothermal Vent	SAMN03002195	Alphaproteobacteria	*Aurantimonas*	96	1.61	■	■		■			■		■				■
			Bacteroidia	SM23-62	66	2.15									■				
			Gammaproteobacteria	GCA-002733105	91	0.42				■									■
				Halomonas	98	2.01	■	■		■									
				Halopseudomonas	99	1.06		■		■									
				Marinobacter	98	0.45	■			■									
				Methylophaga	79	8.62													■
				Pseudohongiella	97	0.06				■									
		SAMN03002196	Alphaproteobacteria	*Aurantimonas*	100	7.59	■	■		■			■		■				■
			Aminicenantia	SOIV01	78	4.84		■											
			Aquicultoria	Unclassified	63	1.81		■											
			Gammaproteobacteria	*Halomonas*	100	2.44	■	■		■									
				Halopseudomonas	94	0.73		■		■									
				Pseudoalteromonas	98	6.55				■									
				Pseudohongiella	71	0.06				■									
			Phycisphaerae	HyVt-337	99	5.88		■							■				
	Subduction Zone and Trenches	SAMN19461850	Anaerolineae	Unclassified	92	4.55									■				
		SAMN19461852	Aerophobia	SOJT01	75	1.12		■											
		SAMN19461860	Actinomycetia	*Mycobacterium*	55	0.63	■	■											
			Anaerolineae	E44-bin32	85	5.61		■											
			Desulfobacteria	Unclassified	82	0.97		■											
Abyssopelagic (4000m - 6000m)	**Hydrothermal Vent**	SAMN05571518	Alphaproteobacteria	Unclassified	67	1.03		■											
			Bipolaricaulia	DRJF01	83	0.00	■												■
			Gammaproteobacteria	HyVt-429	86	0.76												■	
			SZUA-79	QIJE01	87	0.86									■	■			
			Thermodesulfovibrionia	BMS3Bbin07	94	2.73										■			
			UBA6919	CAJXGR01	66	2.27		■	■						■				
			UBA7883	DRJW01	89	4.44									■				
	Unclassified	SAMEA5744179	Gammaproteobacteria	GCA-2729495	84	6.65		■											
				JAACFB01	51	8.55		■											
Hadopelagic (>6000m)	**Subduction Zone and Trenches**	SAMN16252685	Gammaproteobacteria	SMWN01	55	1.10		■								■			
			Nitrospinia	SZUA-226	86	2.42						■							
			UBA1144	CR02bin9	53	0.00		■											
		SAMN16252686	Nitrospinia	SZUA-226	64	0.91						■							
			Terriglobia	JACZYA01	58	1.78		■											

As expected, *nar* and *nap* genes are dominant, even though *nxrA* was not found in PROKKA's results. Interestingly, *nosZ* is present in a high number of MAGs (10) and samples (6), followed by *hzs* (6 MAGs and 4 samples). In fact, the *nosZ* gene was the gene of dissimilatory processes most represented across all metagenomes and environments. *Aurantimonas coralicida* and UBA3067 (order Pseudomonadales) revealed the greatest N-cycling potential, presenting a wide variety in terms of genes related to the N cycle, including *narG*, *napA*, *nasA*, *nirS*, *norB*, *nosZ*, and *hzs*.

Among all MAGs, *nosZ* was detected in *Colwellia psychrerythraea* and *Aurantimonas coralicida*. It was also present in phyla Nitrospinota and Spirochaeota, in classes Phycisphaerae and Anaerolineae, in orders Bacteroidales and Pseudomonadales, as well as in family Acidulodesulfobacteraceae. It is also worth mentioning that this gene was detected in MAGs obtained from almost every sampled environment, such as continental slopes, abyssal plains, hydrothermal vents and trenches.

The only *norB* gene was found in the MAG classified as UBA3067, which also contains *nar* genes as well as *nirS*, suggesting a possible reduction of nitrate all the way to dinitrogen gas, performing a complete denitrification. *NirK* was just detected in Nitrospinaceae MAGs. Family Porticoccaceae only exhibits the *nrfA* gene, which was also found in genus *Amylibacter*. It is important to highlight that *nirK* and *nirS* were detected in metagenomes from contrasting depths: *nirK* was only found in hadopelagic sediments and *nirS* was only found in bathypelagic metagenomes.

Genera from family Acidulodesulfobacteraceae and order Thermodesulfovibrionales contained *nifH*¸ participating in N fixation. *Hzs* was the only observed anammox gene, present in *Aurantimonas coralicida* and *Methylophaga aminisulfidivorans*. It was also found in phylum Acidobacteriota, in class Bipolaricaulia, and in family Methylophagaceae.

The gene *pmoA* was only found in one sample, namely in MAGs classified as *Colwellia psychrerythraea* and UBA3067 (order Pseudomonadales). Regarding other nitrification associated genes, *amoA* was not discovered in PROKKA's results, and *hao* was only present in one MAG of order Arenicellales. However, it should be noted that the completeness of each MAG impacts gene identification and other undetected genes may be present in these genomes, including N-cycling genes.

4. Discussion

4.1. Microbial Communities in Deep-Sea Sediments

The sediments collected from Dorado's seamounts revealed a greater proportion of Bacteria over Archaea. The total number of bacterial ASVs (16944) also surpassed archaeal ASVs (1635). Likewise, previous research at Dorado has similarly shown that Bacteria are predominant in basalts (Lee et al., 2015) and in sediments impacted by cold hydrothermal fluids (Zinke et al., 2018). Regarding other locations and habitats, this trend was also seen across MAGs obtained from ENA and MG-RAST data, since the vast majority was classified as Bacteria (76 out of 81 MAGs).

Recent research comprising 2,102 metagenomes has shown that the overall marine diversity is indeed dominated by Bacteria (Laiolo et al., 2024). However, when compared to bacterial abundance, archaeal abundance may be more relevant in benthic microbial communities (Corinaldesi, 2015), as observed in crusts and sediments collected from Takuyo-Daigo seamount (Northwest Pacific Ocean) (Nitahara et al., 2011).

In fact, despite the high number of ASVs with unclassified genera, the most abundant genus in Dorado's samples was the archaea *Candidatus Nitrosopumilus* (Figure 3), as similarly described in prior studies at this location (Lee et al., 2015; Zinke et al., 2018). Accordingly, Crenarchaeota stands out as the second most abundant phylum across all environmental samples, with Nitrososphaeria (formerly Thaumarchaeota) emerging as the predominant class within this phylum (Figure 2). Just like observed at Dorado outcrop, Nitrosopumilales order was also one of the most abundant in sediments from other seamount communities (Bergo et al., 2022; H. Li et al., 2023; Nitahara et al., 2011). In public metagenome analysis, however, no MAGs of this genus were detected. As taxonomic classification of raw reads was not performed, it is not known if this absence is caused by a difficulty in binning these archaeal genomes or by low abundance of AOA in public metagenomes. A metagenomic investigation of Antarctic marine sediments, which also utilized MetaBat2 for binning genomes and GTDB-Tk for taxonomic classification, only found 1 archaeal MAG that was also assigned to *Nitrosopumilus* (Garber et al., 2021).

Still regarding Archaea, a recent study in Northern Arafura Sea (Tapilatu et al., 2024) found that Crenarchaeota were more abundant in sediments at shallow depths, while Nanoarchaeota were more abundant in deep-sea sediments, at 1457 mbsl. However, Dorado's sediments were collected even deeper, at 3150 mbsl, but Nanoarchaeota

comprised less than 0.4% of each sample, while Crenarcheota reached 14.32%, on average.

In terms of bacterial abundance, Proteobacteria is the most prevalent phylum in Dorado's sediments. Within this phylum, the prevalent classes are Gammaproteobacteria and Alphaproteobacteria, identical to what was observed among MAGs obtained with public metagenomic data. Meanwhile, former studies at Dorado outcrop (Lee et al., 2015; Zinke et al., 2018) and research at other seamount locations (Liu et al., 2017; Nitahara et al., 2011) found Deltaproteobacteria to also be one of the most abundant classes in sediments. Inversely, the phyla Desulfobacterota and Dadabacteria (both formerly Deltaproteobacteria) only make up, in average, 2.16% of environmental samples in the present study. The substantial differences observed at such a coarse taxonomic rank between studies and sample locations highlight the under-sampling nature of the deep-sea ecosystem as well as a higher habitat heterogeneity than previously assumed.

Woeseia (class Gammaproteobacteria) was the second most common genus among Dorado's sediments, just after *Ca. Nitrosopumilus*. Bacteria from the order Woeseiales are vastly abundant in seafloor sediments at a global scale (Hoffmann et al., 2020), despite not being reported in past research at Dorado outcrop. Nevertheless, other genera such as Pir4 lineage (family Plantomycetaceae) and Urania.1B.19 marine group were also found in Dorado's basalts, as shown by Zinke et al. (2018), alongside some promising N-cycling bacteria, including *Nitrospina*, *Nitrosomonas*, *Nitrospira*, and *Ca. Scalindua*,

In background samples, most unclassified genera fell under classes Gammaproteobacteria, Alphaproteobacteria and Acidimicrobiia. Lee et al. (2015) also observed a high percentage of unclassified Alphaproteobacteria in Dorado's sediments previously, mainly from Rhodospirillaceae family. Curiously, the abundance of unclassified genera decreased in some control samples of the Cd experiment with the appearance of a dominant taxon that was not identified in background samples, the genus *Stenotrophomonas* (class Gammaproteobacteria). It was accompanied by the presence of other Gammaproteobacteria also not previously observed in background samples, namely *Ralstonia* and *Deftia*. Experiment conditions might have favored these microorganisms, increasing their abundance within the samples, thus allowing the *dada* denoising function to differentiate their sequences from noise.

4.2. Microbial Nitrogen Cycling Network

Even though differences were observed in the relative abundance of different N-cycling genes, genomic evidence of an almost complete N cycle is present in Dorado's sediments, as estimated by metabolic inference from 16S rRNA gene sequencing (Figure 7). On the other hand, in MAGs obtained from ENA and MG-RAST data, N cycle genes are apparently more scattered across samples of different habitats. For instance, even though there is only a single sample of cold seep sediments (oil seep), none of its MAGs presented a diverse N-cycling potential, despite having *napA*. This was the most common gene among all MAGs, being responsible for the reduction of nitrate to nitrite. In comparison, continental slope and abyssal plain samples presented a slightly higher N-cycling potential, with the presence of multiple genes such as *nrfA*, *nirS*, *norB*, *nosZ*, *pmoA*, and *hzs*. This trend follows the one observed in a recent study (Chen et al., 2023), in which samples from active seeps did not showcase N-cycling activity, while extinct seeps and non-seep sites had an increase in N-cycling. However, it is important to mention the only archaeal MAG from cold seep sediments (class Methanosarcinia) was only 72.32% complete, while bacterial MAGs had a completeness score of more than 84%, which might influence the detection of N cycle genes.

It is commonly considered that *nirS* denitrifiers are dominant, whereas *nirK* denitrifiers represent a broader taxonomic diversity (Graf et al., 2014). Interestingly, the relative abundance of *nirK* was found to be the highest in Dorado's samples, when compared to *nirS*. Nitrosopumilaceae emerges as the primary contributor to this phenomenon, considering that *Ca. Nitrosopumilus* is the dominant genus in all samples. In contrast, *nirS* has the lowest value of relative abundance out of all analyzed genes. The accentuated difference between *nirK* and *nirS* abundance might indicate *nirK* as the main protagonist in nitrite reduction at Dorado's sediments, leading to believe that *nirK* is better adapted to environmental conditions at this site.

The *nirK* and *nirS* genes are present in few MAGs obtained from public metagenomic data. *NirK* was solely observed in unclassified MAGs associated with the Nitrospinaceae family, which also made a substantial contribution to *nirK* abundance in Dorado's sediment samples. However, there is a dissimilarity between depths in which *nirK* was found when comparing MAGs (hadalpelagic) with Dorado's samples (bathypelagic). The presence of *nirK* and *nirS* was believed to be mutually exclusive, but there are microorganisms that harbor both genes and each might be transcribed to a functional nitrite reductase under different environmental conditions (Graf et al., 2014). In Dorado's

samples, it was possible to identify the presence of *nirK* and *nirS* in *Nitrospira*, despite its greater contribution to *nirS* abundance.

Still regarding denitrification, *norB* and *nosZ* showcased much lower relative abundance than *nirK* in Dorado's samples, as *nosZ* is usually more common in *nirS* denitrifiers than in *nirK* denitrifiers (Graf et al., 2014; Pold et al., 2024). Among classified genera, *Lucibacter*, *Colwellia*, and *Magnetospira* contain *norB* and *nosZ*, being able to fully reduce nitric oxide to dinitrogen gas. In comparison, *nosZ* was present in the sole *Colwellia* MAG obtained from ENA and MG-RAST data, despite it only containing *napA* and lacking any of the other genes associated with denitrification. *nosZ* was found in almost 1/4 of MAGs, distributed across various habitats and depths, suggesting the significance of nitrous oxide reduction and highlighting the denitrification pathway in benthic deep ocean ecosystems.

There is also evidence of an alternative pathway to denitrification in Dorado's sediments. DNRA might be predominantly carried out by Acidobacteriota (Subgroup 22), which appears to be the main contributor to *nrfA* abundance. This gene was also found in previously mentioned denitrifying microorganisms such as *Moritella* and *Lucibacter*. On the other hand, *Shewanella* showcased *norB*, *nrfA* and *nap* genes, but lacked *nosZ*. Most *Shewanella* species are ammonifiers (i.e. perform DNRA), while only a few being denitrifiers (Chen & Wang, 2015). Denitrification and DNRA pathways were shown to be often simultaneously present in *nirK* denitrifiers (Helen et al., 2016). In Dorado's sediments, the families Cyclobacteriaceae and Melioribacteraceae were found to possess *nirK*, *nosZ* and *nfrA* genes, alongisde Flavobacteriaceae family, which also had *norB*.

Anammox related genes, like *hdh* and *hzs*, were not found by PICRUSt2 in Dorado's sediments. This was unexpected, as taxonomic composition analysis of these samples revealed the presence of *Ca. Scalindua*. This genus is usually predominant among anammox bacteria found in most marine environments, with species such as *Ca. Scalindua profunda* including the *hzs* gene, as well as *hao*, *nxr*, and *nirS* (Van De Vossenberg et al., 2013). However, in Dorado's sediments, *Ca. Scalindua* only showcased *hao* and *narG/narZ/nxrA,* which may possibly indicate a lack of annotation in PICRUSt2 database. Still regarding anammox, only *hzs* was detected In MAGs obtained from ENA and MG-RAST data. Considering that *hzs* was present in *Aurantimonas coralicida* MAGs, alongside *nosZ* and *nirS*, indicates that this microorganism may reduce nitrate all the way to dinitrogen.

Concerning nitrification, *pmoA* and *amoA* genes rank as the second most prevalent in Dorado's sediments, granted that *Ca. Nitrosopumilus* is the main contributor to its abundance. Lin et al. (2021) also found archaeal *amoA* to be significantly more abundant than bacterial *amoA* in marine sediments. The sole *Colwellia psychrerythraea* MAG demonstrated the presence of *pmoA*, whereas the *Colwellia* species detected in Dorado's did not, although it may be due to MAG completeness of only 56.03%.

As *hao* was also found, having *Ca. Scalindua* and *Nitrosomonas* as main contributors to its abundance, there is an indication of complete nitrification in Dorado's sediments, carried out whether by distinct organisms or by a single organism (complete ammonia oxidation, i.e. comammox). *Nitrospira* is known for including comammox species, typically containing both *amoA* and *hao* genes (Koch et al., 2019). This genus was indeed detected with low relative abundance in Dorado's samples but PICRUSt2 did not assign any of these metabolic functions to the taxon. In fact, until this day, commamox *Nitrospira* were not observed in marine benthic ecosystems (Palomo et al., 2022), apart from coastal or estuarine habitats (D. Sun et al., 2020) and possibly recently discovered in the water column (W. Sun et al., 2022). Instead, Dorado's samples present *Nitrosomonas* with both *pmoA/amoA* and *hao* genes.

As previously mentioned, *Ca. Nitrosopumilus* was predominant in Dorado's sediments, being involved in the initial steps of denitrification (*nirK*) and nitrification (*pmoA/amoA*). It is known that species such as *Ca. Nitrosopumilus maritimus* can produce oxygen via nitric oxide dismutation, with *nirK*-mediated reduction of nitrite being the primary source of nitric oxide (C. L. Wright & Lehtovirta-Morley, 2023). The oxygen is mostly directly consumed by ammonia oxidation carried out by *amoA* (Kraft et al., 2022). Just like observed in *Ca. Nitrosopumilus* in Dorado's samples, *Ca. Nitrosopumilus maritimus* lacks *hao* genes, but it might follow an alternative pathway as described by Walker et al. (2010).

In what regards nitrogen fixation, *nifH* appears with very low relative abundance. In each sample, the main phyla contributing to this function are Acidobacteriota, Desulfobacterota (Deltaproteobacteria) and Thermoplasmatota. Among MAGs obtained with public metagenomic data, the presence of *nifH* was also observed in Deltaproteobacteria, specifically in the family Acidulodesulfobacteraceae, in addition to the order Thermodesulfovibrionales. Research on diazotrophs usually highlights the abundance of *Trichodesmium* and other Cyanobacteria in surface waters and associated to sinking particles (Bonnet et al., 2023; Zehr & Capone, 2024). In Baltic Sea waters, despite cyanobacteria accounting for 99% of *nifH* transcripts, Deltaproteobacteria

dominated non-cyanobacterial diazotrophs (NCD) containing *nifH* genes (Salamon Slater et al., 2023). Moreover, similarly to what was observed at Dorado outcrop, deep-sea sediments from Monterey Canyon (Pacific Ocean) exhibit a higher relative abundance of Deltaproteobacteria, Firmicutes and Acidobacteria among all diazotrophs (Kapili et al., 2020). NCDs have been reported in various benthic environments, including coastal and continental shelf sediments, hydrothermal vents, methane seeps, and others (Turk-Kubo et al., 2023). Remarkably, beyond hydrothermal vents, *nifH* was also detected in a MAG from a sample collected in trenches, at hadalpelagic depths.

4.3. Genomic Responses to Cadmium Treatment

Cd has been proven to be more toxic to benthic microorganisms when compared to other heavy metals (Alvarado-Campo et al., 2023; Gillard et al., 2019). Research about the impact of heavy metals on the N cycle demonstrated that Cd treatment inhibits denitrification in wetland, coastal and estuary sediments (Magalhães et al., 2007; Sakadevan et al., 1999), as well as in simulated pond systems (Ganguly & Jana, 2002). In addition, a study with *Shewanella loihica* revealed that this microorganism's *nirK* and *nosZ* expression was strongly inhibited by Cd exposure (Pizarro et al., 2023). However, this trend was not observed in Dorado's DNA samples, since the increase in Cd concentration did not produce any observable pattern of inhibition of *nirK* or *nosZ* estimated abundance in the overall microbial community. The case is the same for the rest of the inferred N cycle genes.

A Cd concentration of 500 µg/L was shown to influence the microbial N cycle in coastal sediments, particularly increasing N_2 production (Broman et al., 2019). One of the proposed causes for this behavior was the increase of NO_3^- as a product of nitrification, which in turn stimulates denitrification. In the present study, an apparent emerging increase in *amoA/pmoA* and *nirK* abundance was observed with Cd treatment, despite *hao* abundance seemingly decreasing. Yet, it is important to notice that all these observations do not entail significant changes. An actual increase in these genes could further support the previous statement, potentially indicating a stimulation of the initial steps of nitrification and denitrification in the overall community with Cd exposure. *Ca. Nitrosopumilus* was the main contributor to *amoA/pmoA* and *nirK* abundances, but its mean relative abundance gradually decreased with the increase in Cd concentration, as similarly observed for most other genera as well as other taxon at other levels.

In the current study, beta diversity analysis did not establish any apparent differences between different Cd concentrations and the control group (no cadmium). All statistic

tests confirmed these observations with a p-value greater than 0.05, proving that differences between treatment groups were not statistically significant. Still, it was observed an increase in Shannon diversity and evenness with Cd concentration, whereas species richness did not show any apparent pattern. Similarly, bacterial communities from coastal seawater exhibited higher alpha diversity after Cd exposure, despite showing a decrease of Actinobacteria and Alphaproteobacteria abundances and a increase of Bacteroidetes and Planctomycetes abundances (Wang et al., 2015). In contrast, archaeal community richness in coastal sediment samples decreased with the increase in Cd concentration, though these samples similarly displayed Crenarchaeota dominance, alongside other unclassified Archaea (Yu et al., 2024).

Various isolated deep-sea strains were shown to be resistant to different Cd concentrations, including Bacteria and Archaea from sediments and seawater (Alvarado-Campo et al., 2023; Ares et al., 2022; Jasso-Chávez et al., 2019; Kalkan, 2022; Lira-Silva et al., 2012; Ma & Sun, 2021). Additionally, recent genomic analysis of *Rossellomorea,* isolated from deep-sea sediments, showed Cd resistance potential (Bai et al., 2024). Identification of microbial resistance as well as Cd uptake and removal mechanisms are crucial for bioremediation of contaminated environments (Alabssawy & Hashem, 2024; Khan et al., 2022).

It is tempting to conclude that Cd does not influence the diversity and N-cycling potential of Dorado's benthic microbial communities. However, these results reflect only the visible effects after 96h of exposure. It would be interesting to recreate this experiment with a prolonged duration of Cd exposure, alongside other toxic metals and accounting for the variable of time, in order to ascertain whether impacts on microorganisms remain negligible and/or completely absent. Moreover, a metatranscriptomic analysis could shed more light on how N-cycling potential is affected. This would supplement the current study, which displays some limitations in determining N-cycling potential considering that functional prediction relied on 16S metabarcoding data that showcased mostly unclassified genera. Available annotated data in reference databases may not include most of the microorganisms found in the benthic deep sea, as these are often difficult to cultivate in the laboratory.

4.4. Accessibility to Metagenomic Data of Deep-Sea Samples

During the course of this study, it was particularly challenging to gather and collect raw metagenomic sequencing data specifically from deep-sea sediments, even more so when it comes to seamount habitats. One of the reasons points to incomplete or incorrect

metadata. Moreover, the inevitable inconsistency of categorical fields such as biome and environmental feature complicated the filtering steps, making it difficult to create a standardized data retrieval script that could be used in future research.

Nevertheless, it was possible to observe that gaps of knowledge exist in regards to benthic microbial communities present in seamounts. Studies that retrieved public marine metagenomic data (Laiolo et al., 2024; Zhong et al., 2024) include mainly samples of the water column (mostly epipelagic and mesopelagic) rather than sediment samples. In comparison, the current study improved on aggregating deep-sea benthic metagenomic data, although there are still several aspects that can be further perfected.

To facilitate ongoing scientific research of benthic ecosystems in the light of deep-sea mining emergence, there should be a collaborative approach towards adopting a standard and cohesive terminology, along with thorough metadata annotation. Currently, the Planet Microbe (Ponsero et al., 2021) is engaged in aggregating existing marine environmental data in an accessible and interactive web-platform.

Conclusion

Marine seamount habitats host diverse microbial communities, and nutrient cycling within these ecosystems is essential to the sustenance of life. This study focused on investigating the benthic microbial communities of Dorado Outcrop, specifically their taxonomic composition and N-cycling potential, through metabarcoding and metabolic inference analysis of sediment samples and comparison with publicly available metagenomic data. It was observed that the dominant microorganisms at Dorado Outcrop are primarily Gammaproteobacteria, Nitrososphaeria and Alphaproteobacteria. *Ca. Nitrosopumilus* was the most abundant among classified genera, contributing significantly to the abundance of *nirK* and *pmoA/amoA* genes. It was possible to delineate a microbial N-cycling network in Dorado's sediments, although anammox related genes such as *hzs* and *hdh* were absent. Surprisingly, public metagenomic data revealed a vast presence of the *nosZ* gene across various deep-sea habitats, followed by *hzs*.

Given the possible risks of deep-sea mining to the marine environment, this study also examined the genomic response of microbial communities to Cd in order to assess potential effects on habitat diversity and N-cycling. However, the short duration of Cd exposure did not show significant impacts at the DNA level. Studies that include shorter term responses, such as transcriptomic responses, are needed to evaluate more acute effects on ecosystem functioning. Additionally, long-term exposure studies involving Cd and other toxic metals are essential to evaluate chronic and, eventually, longer-lasting changes in microbial community diversity and N-cycling potential.

To conclude, this study contributes to the ongoing efforts of deep-sea exploration as well as to the comprehension of deep-sea microbial communities and their ecological functions.

References

Alabssawy, A. N., & Hashem, A. H. (2024). Bioremediation of hazardous heavy metals by marine microorganisms: A recent review. *Archives of Microbiology*, *206*(3), 103. https://doi.org/10.1007/s00203-023-03793-5

Alvarado-Campo, K. L., Quintero, M., Cuadrado-Cano, B., Montoya-Giraldo, M., Otero-Tejada, E. L., Blandón, L., Sánchez, O., Zuleta-Correa, A., & Gómez-León, J. (2023). Heavy Metal Tolerance of Microorganisms Isolated from Coastal Marine Sediments and Their Lead Removal Potential. *Microorganisms*, *11*(11), 2708. https://doi.org/10.3390/microorganisms11112708

Amon, D. J., Gollner, S., Morato, T., Smith, C. R., Chen, C., Christiansen, S., Currie, B., Drazen, J. C., Fukushima, T., Gianni, M., Gjerde, K. M., Gooday, A. J., Grillo, G. G., Haeckel, M., Joyini, T., Ju, S.-J., Levin, L. A., Metaxas, A., Mianowicz, K., … Pickens, C. (2022). Assessment of scientific gaps related to the effective environmental management of deep-seabed mining. *Marine Policy*, *138*, 105006. https://doi.org/10.1016/j.marpol.2022.105006

Andrews, S. (2010). *FastQC: A Quality Control tool for High Throughput Sequence Data*. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Ares, Á., Sakai, S., Sasaki, T., Shimamura, S., Mitarai, S., & Nunoura, T. (2022). Sequestration and efflux largely account for cadmium and copper resistance in the deep-sea Nitratiruptor sp.SB155-2 (phylum Campylobacterota). *Environmental Microbiology*, *24*(12), 6144–6163. https://doi.org/10.1111/1462-2920.16255

Bai, S., Huang, Z., & Li, X.-G. (2024). Genome analysis of Rossellomorea sp. Y25, a deep sea bacterium isolated from the sediments of South China Sea. *Marine Genomics*, *75*. https://doi.org/10.1016/j.margen.2024.101110

Barbera, P., Kozlov, A. M., Czech, L., Morel, B., Darriba, D., Flouri, T., & Stamatakis, A. (2018). *Data from: EPA-ng: massively parallel evolutionary placement of genetic sequences* (Version 1, p. 16601287 bytes) [dataset]. Dryad. https://doi.org/10.5061/DRYAD.KB505NC

Bergo, N. M., Torres-Ballesteros, A., Signori, C. N., Benites, M., Jovane, L., Murton, B. J., Da Rocha, U. N., & Pellizari, V. H. (2022). Spatial patterns of microbial diversity in Fe-Mn deposits and associated sediments in the Atlantic and Pacific oceans. *Science of The Total Environment*, *837*, 155792. https://doi.org/10.1016/j.scitotenv.2022.155792

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bonnet, S., Benavides, M., Le Moigne, F. A. C., Camps, M., Torremocha, A., Grosso, O., Dimier, C., Spungin, D., Berman-Frank, I., Garczarek, L., & Cornejo-Castillo, F. M. (2023). Diazotrophs are overlooked contributors to carbon and nitrogen export to the deep ocean. *The ISME Journal*, *17*(1), 47–58. https://doi.org/10.1038/s41396-022-01319-3

Braathen, A., & Brekke, H. (2020). Characterizing the Seabed: A Geoscience Perspective. In C. Banet (Ed.), *The Law of the Seabed* (pp. 21–35). Brill | Nijhoff. https://doi.org/10.1163/9789004391567_003

Broman, E., Motwani, N. H., Bonaglia, S., Landberg, T., Nascimento, F. J. A., & Sjöling, S. (2019). Denitrification responses to increasing cadmium exposure in Baltic Sea sediments. *Aquatic Toxicology*, *217*, 105328. https://doi.org/10.1016/j.aquatox.2019.105328

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Casciotti, K., Marshall, T., Fawcett, S., & Knapp, A. (2024). Advances in Understanding the Marine Nitrogen Cycle in the GEOTRACES Era. *Oceanography*, *37*(2). https://doi.org/10.5670/oceanog.2024.406

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2020). GTDB-Tk: A toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*, *36*(6), 1925–1927. https://doi.org/10.1093/bioinformatics/btz848

Chen, Y., Feng, X., He, Y., & Wang, F. (2016). Genome Analysis of a Limnobacter sp. Identified in an Anaerobic Methane-Consuming Cell Consortium. *Frontiers in Marine Science*, *3*. https://doi.org/10.3389/fmars.2016.00257

Chen, Y., & Wang, F. (2015). Insights on nitrate respiration by Shewanella. *Frontiers in Marine Science*, *1*. https://doi.org/10.3389/fmars.2014.00080

Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., Brown, C. T., Porras-Alfaro, A., Kuske, C. R., & Tiedje, J. M. (2014). Ribosomal Database Project:

Data and tools for high throughput rRNA analysis. *Nucleic Acids Research*, *42*(D1), D633–D642. https://doi.org/10.1093/nar/gkt1244

Corinaldesi, C. (2015). New perspectives in benthic deep-sea microbial ecology. *Frontiers in Marine Science*, *2*. https://doi.org/10.3389/fmars.2015.00017

Czech, L., Barbera, P., & Stamatakis, A. (2020). Genesis and Gappa: Processing, analyzing and visualizing phylogenetic (placement) data. *Bioinformatics*, *36*(10), 3263–3265. https://doi.org/10.1093/bioinformatics/btaa070

Douglas, G. M., Maffei, V. J., Zaneveld, J. R., Yurgel, S. N., Brown, J. R., Taylor, C. M., Huttenhower, C., & Langille, M. G. I. (2020). PICRUSt2 for prediction of metagenome functions. *Nature Biotechnology*, *38*(6), 685–688. https://doi.org/10.1038/s41587-020-0548-6

Eddy, S. R. (2011). Accelerated Profile HMM Searches. *PLoS Computational Biology*, *7*(10), e1002195. https://doi.org/10.1371/journal.pcbi.1002195

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Ganguly, S., & Jana, B. B. (2002). Cadmium induced adaptive responses of certain biogeochemical cycling bacteria in an aquatic system. *Water Research*, *36*(7), 1667–1676. https://doi.org/10.1016/S0043-1354(01)00369-4

Garber, A. I., Zehnpfennig, J. R., Sheik, C. S., Henson, M. W., Ramírez, G. A., Mahon, A. R., Halanych, K. M., & Learman, D. R. (2021). Metagenomics of Antarctic Marine Sediment Reveals Potential for Diverse Chemolithoautotrophy. *mSphere*, *6*(6), e00770-21. https://doi.org/10.1128/mSphere.00770-21

Gillard, B., Chatzievangelou, D., Thomsen, L., & Ullrich, M. S. (2019). Heavy-Metal-Resistant Microorganisms in Deep-Sea Sediments Disturbed by Mining Activity: An Application Toward the Development of Experimental in vitro Systems. *Frontiers in Marine Science*, *6*, 462. https://doi.org/10.3389/fmars.2019.00462

Glass, E. M., Wilkening, J., Wilke, A., Antonopoulos, D., & Meyer, F. (2010). Using the Metagenomics RAST Server (MG-RAST) for Analyzing Shotgun Metagenomes. *Cold Spring Harbor Protocols*, *2010*(1), pdb.prot5368. https://doi.org/10.1101/pdb.prot5368

Graf, D. R. H., Jones, C. M., & Hallin, S. (2014). Intergenomic Comparisons Highlight Modularity of the Denitrification Pathway and Underpin the Importance of Community

Structure for N2O Emissions. *PLoS ONE*, *9*(12), e114118. https://doi.org/10.1371/journal.pone.0114118

Hallin, S., Philippot, L., Löffler, F. E., Sanford, R. A., & Jones, C. M. (2018). Genomics and Ecology of Novel N2O-Reducing Microorganisms. *Trends in Microbiology*, *26*(1), 43–55. https://doi.org/10.1016/j.tim.2017.07.003

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Rio, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Helen, D., Kim, H., Tytgat, B., & Anne, W. (2016). Highly diverse nirK genes comprise two major clades that harbour ammonium-producing denitrifiers. *BMC Genomics*, *17*(1), 155. https://doi.org/10.1186/s12864-016-2465-0

Hoffmann, K., Bienhold, C., Buttigieg, P. L., Knittel, K., Laso-Pérez, R., Rapp, J. Z., Boetius, A., & Offre, P. (2020). Diversity and metabolism of *Woeseiales* bacteria, global members of marine sediment communities. *The ISME Journal*, *14*(4), 1042–1056. https://doi.org/10.1038/s41396-020-0588-4

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, *11*(1), 119. https://doi.org/10.1186/1471-2105-11-119

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications*, *9*(1), 5114. https://doi.org/10.1038/s41467-018-07641-9

Jasso-Chávez, R., Lira-Silva, E., González-Sánchez, K., Larios-Serrato, V., Mendoza-Monzoy, D. L., Pérez-Villatoro, F., Morett, E., Vega-Segura, A., Torres-Márquez, M. E., Zepeda-Rodríguez, A., & Moreno-Sánchez, R. (2019). Marine Archaeon Methanosarcina acetivorans Enhances Polyphosphate Metabolism Under Persistent Cadmium Stress. *Frontiers in Microbiology*, *10*, 2432. https://doi.org/10.3389/fmicb.2019.02432

Jordahl, K., Van den Bossche, J., Fleischmann, M., Wasserman, J., McBride, J., Gerard, J., Tratner, J., Perry, M., Badaracco, A. G., Farmer, C., Hjelle, G. A., Snow, A. D., Cochran, M., Gillies, S., Culbertson, L., Bartos, M., Eubank, N., maxalbert, Bilogur, A., … Leblanc, F. (2020). *geopandas/geopandas: V0.8.1* (v0.8.1) [Computer software]. https://doi.org/10.5281/zenodo.3946761

Kalkan, S. (2022). Heavy metal resistance of marine bacteria on the sediments of the Black Sea. *Marine Pollution Bulletin*, *179*. https://doi.org/10.1016/j.marpolbul.2022.113652

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. https://doi.org/10.7717/peerj.7359

Kapili, B. J., Barnett, S. E., Buckley, D. H., & Dekas, A. E. (2020). Evidence for phylogenetically and catabolically diverse active diazotrophs in deep-sea sediment. *The ISME Journal*, *14*(4), 971–983. https://doi.org/10.1038/s41396-019-0584-8

Khan, Z., Elahi, A., Bukhari, D. A., & Rehman, A. (2022). Cadmium sources, toxicity, resistance and removal by microorganisms-A potential strategy for cadmium eradication. *Journal of Saudi Chemical Society*, *26*(6), 101569. https://doi.org/10.1016/j.jscs.2022.101569

Kirchman, D. L. (2018). *Processes in Microbial Ecology* (2nd ed.). Oxford University PressOxford. https://doi.org/10.1093/oso/9780198789406.001.0001

Koch, H., Van Kessel, M. A. H. J., & Lücker, S. (2019). Complete nitrification: Insights into the ecophysiology of comammox Nitrospira. *Applied Microbiology and Biotechnology*, *103*(1), 177–189. https://doi.org/10.1007/s00253-018-9486-3

Kraft, B., Jehmlich, N., Larsen, M., Bristow, L. A., Könneke, M., Thamdrup, B., & Canfield, D. E. (2022). Oxygen and nitrogen production by an ammonia-oxidizing archaeon. *Science*, *375*(6576), 97–100. https://doi.org/10.1126/science.abe6733

Kuypers, M. M. M., Marchant, H. K., & Kartal, B. (2018). The microbial nitrogen-cycling network. *Nature Reviews Microbiology*, *16*(5), 263–276. https://doi.org/10.1038/nrmicro.2018.9

Laiolo, E., Alam, I., Uludag, M., Jamil, T., Agusti, S., Gojobori, T., Acinas, S. G., Gasol, J. M., & Duarte, C. M. (2024). Metagenomic probing toward an atlas of the taxonomic

and metabolic foundations of the global ocean genome. *Frontiers in Science*, *1*, 1038696. https://doi.org/10.3389/fsci.2023.1038696

Lee, M. D., Walworth, N. G., Sylvan, J. B., Edwards, K. J., & Orcutt, B. N. (2015). Microbial Communities on Seafloor Basalts at Dorado Outcrop Reflect Level of Alteration and Highlight Global Lithic Clades. *Frontiers in Microbiology*, *6*. https://doi.org/10.3389/fmicb.2015.01470

Li, D., Luo, R., Liu, C.-M., Leung, C.-M., Ting, H.-F., Sadakane, K., Yamashita, H., & Lam, T.-W. (2016). MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods*, *102*, 3–11. https://doi.org/10.1016/j.ymeth.2016.02.020

Li, H., Zhou, H., Yang, S., & Dai, X. (2023). Stochastic and Deterministic Assembly Processes in Seamount Microbial Communities. *Applied and Environmental Microbiology*, *89*(7), e00701-23. https://doi.org/10.1128/aem.00701-23

Li, Q., Lei, Y., Morard, R., Li, T., & Wang, B. (2020). Diversity hotspot and unique community structure of foraminifera in the world's deepest marine blue hole – Sansha Yongle Blue Hole. *Scientific Reports*, *10*(1), 10257. https://doi.org/10.1038/s41598-020-67221-0

Lin, G., Huang, J., Lu, J., Su, M., Hu, B., & Lin, X. (2021). Geochemical and microbial insights into vertical distributions of genetic potential of N-cycling processes in deep-sea sediments. *Ecological Indicators*, *125*, 107461. https://doi.org/10.1016/j.ecolind.2021.107461

Lira-Silva, E., Santiago-Martínez, M. G., Hernández-Juárez, V., García-Contreras, R., Moreno-Sánchez, R., & Jasso-Chávez, R. (2012). Activation of Methanogenesis by Cadmium in the Marine Archaeon Methanosarcina acetivorans. *PLoS ONE*, *7*(11), e48779. https://doi.org/10.1371/journal.pone.0048779

Liu, J., Zhang, W., Li, X., Li, X., Chen, X., Li, J.-H., Teng, Z., Xu, C., Santini, C.-L., Zhao, L., Zhao, Y., Zhang, H., Zhang, W.-J., Xu, K., Li, C., Pan, Y., Xiao, T., Pan, H., & Wu, L.-F. (2017). Bacterial community structure and novel species of magnetotactic bacteria in sediments from a seamount in the Mariana volcanic arc. *Scientific Reports*, *7*(1), 17964. https://doi.org/10.1038/s41598-017-17445-4

Louca, S., & Doebeli, M. (2018). Efficient comparative phylogenetics on large trees. *Bioinformatics*, *34*(6), 1053–1055. https://doi.org/10.1093/bioinformatics/btx701

Ma, N., & Sun, C. (2021). Cadmium sulfide nanoparticle biomineralization and biofilm formation mediate cadmium resistance of the deep-sea bacterium *Pseudoalteromonas* sp. MT33b. *Environmental Microbiology Reports*, *13*(3), 325–336. https://doi.org/10.1111/1758-2229.12933

Magalhães, C., Costa, J., Teixeira, C., & Bordalo, A. A. (2007). Impact of trace metals on denitrification in estuarine sediments of the Douro River estuary, Portugal. *Marine Chemistry*, *107*(3). https://doi.org/10.1016/j.marchem.2007.02.005

Mason, O. U., Scott, N. M., Gonzalez, A., Robbins-Pianka, A., Bælum, J., Kimbrel, J., Bouskill, N. J., Prestat, E., Borglin, S., Joyner, D. C., Fortney, J. L., Jurelevicius, D., Stringfellow, W. T., Alvarez-Cohen, L., Hazen, T. C., Knight, R., Gilbert, J. A., & Jansson, J. K. (2014). Metagenomics reveals sediment microbial community response to Deepwater Horizon oil spill. *The ISME Journal*, *8*(7), 1464–1475. https://doi.org/10.1038/ismej.2013.254

Matsen, F. A., Kodner, R. B., & Armbrust, E. V. (2010). pplacer: Linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics*, *11*(1), 538. https://doi.org/10.1186/1471-2105-11-538

McKinney, W. (2010). Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*, 56–61. https://doi.org/10.25080/Majora-92bf1922-00a

McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*, *8*(4), e61217. https://doi.org/10.1371/journal.pone.0061217

Met Office. (2023). *Cartopy: A cartographic python library with a Matplotlib interface* (v0.23.0) [Computer software]. https://scitools.org.uk/cartopy/docs/latest/

Miller, K. A., Thompson, K. F., Johnston, P., & Santillo, D. (2018). An Overview of Seabed Mining Including the Current State of Development, Environmental Impacts, and Knowledge Gaps. *Frontiers in Marine Science*, *4*, 418. https://doi.org/10.3389/fmars.2017.00418

Mirarab, S., Nguyen, N., & Warnow, T. (2011). SEPP: SATé-Enabled Phylogenetic Placement. *Biocomputing 2012*, 247–258. https://doi.org/10.1142/9789814366496_0024

Nitahara, S., Kato, S., Urabe, T., Usui, A., & Yamagishi, A. (2011). Molecular characterization of the microbial community in hydrogenetic ferromanganese crusts of the Takuyo-Daigo Seamount, northwest Pacific: Microorganisms in Mn crusts. *FEMS Microbiology Letters*, *321*(2), 121–129. https://doi.org/10.1111/j.1574-6968.2011.02323.x

Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., Caceres, M. D., Durand, S., … Stier, A. (2022). *Vegan: Community Ecology Package*.

Ondov, B. D., Treangen, T. J., Melsted, P., Mallonee, A. B., Bergman, N. H., Koren, S., & Phillippy, A. M. (2016). Mash: Fast genome and metagenome distance estimation using MinHash. *Genome Biology*, *17*(1), 132. https://doi.org/10.1186/s13059-016-0997-x

Orcutt, B. N., Bradley, J. A., Brazelton, W. J., Estes, E. R., Goordial, J. M., Huber, J. A., Jones, R. M., Mahmoudi, N., Marlow, J. J., Murdock, S., & Pachiadaki, M. (2020). Impacts of deep-sea mining on microbial ecosystem services. *Limnology and Oceanography*, *65*(7), 1489–1510. https://doi.org/10.1002/lno.11403

Pajares, S., & Ramos, R. (2019). Processes and Microorganisms Involved in the Marine Nitrogen Cycle: Knowledge and Gaps. *Frontiers in Marine Science*, *6*, 739. https://doi.org/10.3389/fmars.2019.00739

Palomo, A., Dechesne, A., Pedersen, A. G., & Smets, B. F. (2022). Genomic profiling of Nitrospira species reveals ecological success of comammox Nitrospira. *Microbiome*, *10*(1), 204. https://doi.org/10.1186/s40168-022-01411-y

Parada, A. E., Needham, D. M., & Fuhrman, J. A. (2016). Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology*, *18*(5), 1403–1414. https://doi.org/10.1111/1462-2920.13023

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., & Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology*, *36*(10), 996–1004. https://doi.org/10.1038/nbt.4229

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. https://doi.org/10.1101/gr.186072.114

Pizarro, L., Magalhães, C., Almeida, C. M. R., Carvalho, M. D. F., & Semedo, M. (2023). Cadmium effects on net N2O production by the deep-sea isolate *Shewanella loihica* PV-4. *FEMS Microbiology Letters*, *370*, fnad047. https://doi.org/10.1093/femsle/fnad047

Pold, G., Bonilla-Rosso, G., Saghaï, A., Strous, M., Jones, C. M., & Hallin, S. (2024). Phylogenetics and environmental distribution of nitric oxide-forming nitrite reductases reveal their distinct functional and ecological roles. *ISME Communications*, *4*(1), ycae020. https://doi.org/10.1093/ismeco/ycae020

Ponsero, A. J., Bomhoff, M., Blumberg, K., Youens-Clark, K., Herz, N. M., Wood-Charlson, E. M., Delong, E. F., & Hurwitz, B. L. (2021). Planet Microbe: A platform for marine microbiology to discover and analyze interconnected 'omics and environmental data. *Nucleic Acids Research*, *49*(D1), D792–D802. https://doi.org/10.1093/nar/gkaa637

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE*, *5*(3), e9490. https://doi.org/10.1371/journal.pone.0009490

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, *41*(D1), D590–D596. https://doi.org/10.1093/nar/gks1219

Ramirez-Llodra, E. (2020). Deep-Sea Ecosystems: Biodiversity and Anthropogenic Impacts. In C. Banet (Ed.), *The Law of the Seabed* (pp. 36–60). Brill | Nijhoff. https://doi.org/10.1163/9789004391567_004

Ramirez-Llodra, E., Brandt, A., Danovaro, R., De Mol, B., Escobar, E., German, C. R., Levin, L. A., Martinez Arbizu, P., Menot, L., Buhl-Mortensen, P., Narayanaswamy, B. E., Smith, C. R., Tittensor, D. P., Tyler, P. A., Vanreusel, A., & Vecchione, M. (2010). Deep, diverse and definitely different: Unique attributes of the world's largest ecosystem. *Biogeosciences*, *7*(9), 2851–2899. https://doi.org/10.5194/bg-7-2851-2010

Reitz, K., & contributors. (2023). *Requests: HTTP for Humans* (v 2.31.0) [Computer software]. https://pypi.org/project/requests/

Richardson, L. (2023). *Beautiful Soup: Screen-scraping library* (v4.12.2) [Computer software]. https://pypi.org/project/beautifulsoup4/

Richardson, L., Allen, B., Baldi, G., Beracochea, M., Bileschi, M. L., Burdett, T., Burgin, J., Caballero-Pérez, J., Cochrane, G., Colwell, L. J., Curtis, T., Escobar-Zepeda, A., Gurbich, T. A., Kale, V., Korobeynikov, A., Raj, S., Rogers, A. B., Sakharova, E., Sanchez, S., … Finn, R. D. (2023). MGnify: The microbiome sequence data analysis resource in 2023. *Nucleic Acids Research*, *51*(D1), D753–D759. https://doi.org/10.1093/nar/gkac1080

Rogers, A. D. (2015). Environmental Change in the Deep Ocean. *Annual Review of Environment and Resources*, *40*(1), 1–38. https://doi.org/10.1146/annurev-environ-102014-021415

Sakadevan, K., Zheng, H., & Bavor, H. J. (1999). Impact of heavy metals on denitrification in surface wetland sediments receiving wastewater. *Water Science and Technology*, *40*(3), 349–355. https://doi.org/10.1016/S0273-1223(99)00471-0

Salamon Slater, E. R., Turk-Kubo, K. A., Hallstrøm, S., Kesy, K., Laas, P., Magasin, J., Zehr, J. P., Labrenz, M., & Riemann, L. (2023). Composition and distribution of diazotrophs in the Baltic Sea. *Estuarine, Coastal and Shelf Science*, *294*, 108527. https://doi.org/10.1016/j.ecss.2023.108527

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Shaw, J., & Yu, Y. W. (2023). Fast and robust metagenomic sequence comparison through sparse chaining with skani. *Nature Methods*, *20*(11), 1661–1665. https://doi.org/10.1038/s41592-023-02018-3

Smith, C. R., Tunnicliffe, V., Colaço, A., Drazen, J. C., Gollner, S., Levin, L. A., Mestre, N. C., Metaxas, A., Molodtsova, T. N., Morato, T., Sweetman, A. K., Washburn, T., & Amon, D. J. (2020). Deep-Sea Misconceptions Cause Underestimation of Seabed-Mining Impacts. *Trends in Ecology & Evolution*, *35*(10), 853–857. https://doi.org/10.1016/j.tree.2020.07.002

Sun, D., Tang, X., Zhao, M., Zhang, Z., Hou, L., Liu, M., Wang, B., Klümper, U., & Han, P. (2020). Distribution and Diversity of Comammox Nitrospira in Coastal Wetlands of China. *Frontiers in Microbiology*, *11*, 589268. https://doi.org/10.3389/fmicb.2020.589268

Sun, W., Jiao, L., Wu, J., Ye, J., Wei, M., & Hong, Y. (2022). Existence and distribution of novel phylotypes of Nitrospira in water columnsof the South China Sea. *iScience*, *25*(9), 104895. https://doi.org/10.1016/j.isci.2022.104895

Tapilatu, Y., Fauzan, I., Pradipta, A., & Kusuma, A. B. (2024). A first report on prokaryotic diversity in northwestern Arafura deep-sea sediments, Indonesia. *Scientific Reports*, *14*(1), 895. https://doi.org/10.1038/s41598-024-51614-6

The pandas development team. (2020). *pandas-dev/pandas: Pandas* (v2.1.3) [Computer software]. https://doi.org/10.5281/zenodo.3509134

Thurber, A. R., Sweetman, A. K., Narayanaswamy, B. E., Jones, D. O. B., Ingels, J., & Hansman, R. L. (2014). Ecosystem function and services provided by the deep sea. *Biogeosciences*, *11*(14), 3941–3963. https://doi.org/10.5194/bg-11-3941-2014

Torres-Beltrán, M., Vargas-Gastélum, L., Magdaleno-Moncayo, D., Riquelme, M., Herguera-García, J. C., Prieto-Davó, A., & Lago-Lestón, A. (2021). The metabolic core of the prokaryotic community from deep-sea sediments of the southern Gulf of Mexico shows different functional signatures between the continental slope and abyssal plain. *PeerJ*, *9*, e12474. https://doi.org/10.7717/peerj.12474

Turk-Kubo, K. A., Gradoville, M. R., Cheung, S., Cornejo-Castillo, F. M., Harding, K. J., Morando, M., Mills, M., & Zehr, J. P. (2023). Non-cyanobacterial diazotrophs: Global diversity, distribution, ecophysiology, and activity in marine waters. *FEMS Microbiology Reviews*, *47*(6), fuac046. https://doi.org/10.1093/femsre/fuac046

Van De Vossenberg, J., Woebken, D., Maalcke, W. J., Wessels, H. J. C. T., Dutilh, B. E., Kartal, B., Janssen-Megens, E. M., Roeselers, G., Yan, J., Speth, D., Gloerich, J., Geerts, W., Van Der Biezen, E., Pluk, W., Francoijs, K., Russ, L., Lam, P., Malfatti, S. A., Tringe, S. G., … Jetten, M. S. M. (2013). The metagenome of the marine anammox bacterium ‘ *Candidatus* Scalindua profunda’ illustrates the versatility of this globally important nitrogen cycle bacterium. *Environmental Microbiology*, *15*(5), 1275–1289. https://doi.org/10.1111/j.1462-2920.2012.02774.x

Walker, C. B., De La Torre, J. R., Klotz, M. G., Urakawa, H., Pinel, N., Arp, D. J., Brochier-Armanet, C., Chain, P. S. G., Chan, P. P., Gollabgir, A., Hemp, J., Hügler, M., Karr, E. A., Könneke, M., Shin, M., Lawton, T. J., Lowe, T., Martens-Habbena, W., Sayavedra-Soto, L. A., … Stahl, D. A. (2010). *Nitrosopumilus maritimus* genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea.

Proceedings of the National Academy of Sciences, *107*(19), 8818–8823. https://doi.org/10.1073/pnas.0913533107

Wang, K., Zhang, D., Xiong, J., Chen, X., Zheng, J., Hu, C., Yang, Y., & Zhu, J. (2015). Response of Bacterioplankton Communities to Cadmium Exposure in Coastal Water Microcosms with High Temporal Variability. *Applied and Environmental Microbiology*, *81*(1), 231–240. https://doi.org/10.1128/AEM.02562-14

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wright, C. L., & Lehtovirta-Morley, L. E. (2023). Nitrification and beyond: Metabolic versatility of ammonia oxidising archaea. *The ISME Journal*, *17*(9), 1358–1368. https://doi.org/10.1038/s41396-023-01467-0

Wright, E., S. (2016). Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R. *The R Journal*, *8*(1), 352. https://doi.org/10.32614/RJ-2016-025

Ye, Y., & Doak, T. G. (2009). A Parsimony Approach to Biological Pathway Reconstruction/Inference for Genomes and Metagenomes. *PLoS Computational Biology*, *5*(8), e1000465. https://doi.org/10.1371/journal.pcbi.1000465

Yu, C., Meng, K., Zhu, Z., Liu, S., Zhou, Z., Zhang, H., & Xu, M. (2024). Impacts of cadmium accumulation on the diversity, assembly processes, and co-occurrence patterns of archaeal communities in marine sediments. *Science of The Total Environment*, *926*. https://doi.org/10.1016/j.scitotenv.2024.171936

Zehr, J. P., & Capone, D. G. (2024). Unsolved mysteries in marine nitrogen fixation. *Trends in Microbiology*, *32*(6), 532–545. https://doi.org/10.1016/j.tim.2023.08.004

Zhong, J., Osborn, T., Del Rosario Hernández, T., Kyrysyuk, O., Tully, B. J., & Anderson, R. E. (2024). Increasing transposase abundance with ocean depth correlates with a particle-associated lifestyle. *mSystems*, *9*(3), e00067-24. https://doi.org/10.1128/msystems.00067-24

Zhou, Y.-L., Mara, P., Cui, G.-J., Edgcomb, V. P., & Wang, Y. (2022). Microbiomes in the Challenger Deep slope and bottom-axis sediments. *Nature Communications*, *13*(1), 1515. https://doi.org/10.1038/s41467-022-29144-4

Zinke, L. A., Reese, B. K., McManus, J., Wheat, C. G., Orcutt, B. N., & Amend, J. P. (2018). Sediment Microbial Communities Influenced by Cool Hydrothermal Fluid Migration. *Frontiers in Microbiology*, *9*, 1249. https://doi.org/10.3389/fmicb.2018.01249

Attachments

Attachment 1

A.1 - Total number of ASVs per sample along every step of amplicon sequencing data processing. Samples with treatment denoted as "*none"* are background sediment samples collected at Dorado outcrop. CTR is the control group of Cd experiments (no metal), and Cd concentrations are measured in µM.

Sample	Treatment	Replicate	Dada2 Input ASVs	Filtered ASVs	Dada Forward ASVs	Dada Reverse ASVs	Merged ASVs	ASVs Without Chimeras	Final Reads Retained (%)
1	CTR	A	65748	65727	58179	59837	34643	31424	47.8
1	Cd 0.015	A	65914	65904	58576	59887	35337	32293	49.0
1	Cd 0.15	A	70053	70034	63041	64228	40129	36939	52.7
1	Cd 1.5	A	65467	65454	59106	60376	39493	36809	56.2
1	Cd 15	A	62373	62365	55335	56633	31261	27943	44.8
1	CTR	B	76669	76655	69808	71233	48517	45666	59.6
1	Cd 0.015	B	66498	66480	59985	61237	40312	37928	57.0
1	Cd 0.15	B	62680	62666	56221	57460	37069	34816	55.5
1	Cd 1.5	B	78897	78881	70951	72063	44494	40598	51.5
1	Cd 15	B	65373	65363	58840	59863	39397	36646	56.1
1	CTR	C	63720	63705	57405	58501	39213	36969	58.0
1	Cd 0.015	C	66268	66249	59746	61007	40078	37389	56.4
1	Cd 0.15	C	66772	66756	59950	61075	37711	34141	51.1
1	Cd 1.5	C	64991	64989	60552	61676	51024	49760	76.6
1	Cd 15	C	69113	69105	62629	63556	40787	37454	54.2
1	none	A	63855	63843	57913	58983	39408	37315	58.4
1	none	B	64145	64139	58152	59148	38460	35825	55.9
1	none	C	62595	62591	56810	57749	38513	36050	57.6
2	CTR	A	70158	70156	66875	67561	59532	58589	83.5
2	Cd 0.015	A	69565	69553	61525	62606	33725	29612	42.6
2	Cd 0.15	A	69987	69986	66319	66921	57569	56406	80.6
2	Cd 1.5	A	63873	63870	60536	61174	53068	52033	81.5
2	Cd 15	A	63328	63321	57047	57714	34215	31697	50.1
2	CTR	B	66215	66213	63725	63956	57538	56893	85.9
2	Cd 0.015	B	66406	66402	62837	63584	54726	53821	81.0
2	Cd 0.15	B	62692	62689	59013	59607	50887	49664	79.2
2	Cd 1.5	B	64818	64817	60819	61780	52634	51288	79.1
2	Cd 15	B	64073	64061	58891	60194	45667	43551	68.0
2	CTR	C	64284	64283	60655	61596	53013	51820	80.6
2	Cd 0.015	C	64832	64830	60779	61763	52276	50914	78.5
2	Cd 0.15	C	65620	65613	61516	62132	49455	47848	72.9
2	Cd 1.5	C	65674	65663	60625	61174	45608	43201	65.8
2	Cd 15	C	66306	66291	61243	62131	45988	43236	65.2
2	none	A	65497	65486	59398	60389	40312	38162	58.3
2	none	B	65214	65200	58847	60281	40885	38644	59.3
2	none	C	72075	72062	65627	66996	48020	45911	63.7
Average Values			**66437**	**66428**	**60819**	**61835**	**44193**	**41923**	**63.2**

Attachment 2

A.2 – Deep-sea sediment samples retrieved from public databases. This table only includes samples with available download of raw shotgun sequencing data.

Database	Sample ID	Run ID	mbsf	mbsl	Habitat	Location	Study ID	Reference
MG-RAST	mgm4636854.3		0.05	1937	continental slope and abyssal plain	Gulf of Mexico	mgp13888	(Torres-Beltrán et al., 2021)
	mgm4636855.3		0.05	3715				
	mgm4636856.3		0.05	1598				
	mgm4636857.3		0.05	3559				
	mgm4636858.3		0.05	3569				
	mgm4508047.3			1600.79		Gulf of Mexico	mgp2734	
	mgm4510162.3		0.01	1500	continental slope and abyssal plain	Gulf of Mexico	mgp3023	(Mason et al., 2014)
	mgm4510163.3		0.01	1500				
	mgm4510164.3		0.01	1500				
	mgm4510165.3		0.01	1500				
	mgm4510166.3		0.01	1500				
	mgm4510167.3		0.01	1500				
	mgm4510168.3		0.01	1500				
	mgm4510169.3		0.01	1500				
	mgm4510170.3		0.01	1500				
	mgm4510171.3		0.01	1500				
	mgm4510172.3		0.01	1500				
	mgm4510173.3		0.01	1500				
	mgm4510174.3		0.01	1500				
	mgm4510175.3		0.01	1500				
ENA	SAMEA5663117	ERR3341263	0.15	3092	cold seep	Atlantic Ocean	PRJEB32776	
	SAMEA5663119	ERR3341264	0.03	2925				
		ERR3341265	0.03	2925				
		ERR3341266	0.03	2925				
		ERR3341267	0.03	2925				
		ERR3341270	0.03	2925				
		ERR3365594	0.03	2925				
		ERR3365595	0.03	2925				
		ERR3365596	0.03	2925				
	SAMEA5744179	ERR3403595	0.01	5525		Arctic Ocean	PRJEB33248	
	SAMEA5744180	ERR3403596	0.01	3531				
	SAMEA5744182	ERR3403597	0.01	2403				
	SAMEA5744185	ERR3403598	0.01	1244				
	SAMEA6426302	ERR3771508	0.01	5704	blue hole	Pacific Ocean	PRJEB35877	(Q. Li et al., 2020)
	SAMEA6426305	ERR3771511	0.01	5121				
	SAMEA6426306	ERR3771512	0.01	5720				
	SAMEA6426307	ERR3771513	0.01	5040				

ENA	SAMEA6426308	ERR3771514	0.01	5830				
	SAMEA6426309	ERR3771515	0.01	4858				
	SAMEA6426312	ERR3771518	0.01	4080				
	SAMEA6457547	ERR3799536	0.1	2925	cold seep	Atlantic Ocean	PRJEB36096	
		ERR3799537	0.1	2925				
		ERR3799538	0.1	2925				
		ERR3799539	0.1	2925				
		ERR3799540	0.1	2925				
		ERR3799541	0.1	2925				
		ERR3799542	0.1	2925				
	SAMEA7725637	ERR5000361	0.05	1122	cold seep	Mediterranean Sea		
		ERR5000362	0.05	1122				
		ERR5000363	0.05	1122				
		ERR5000364	0.05	1122				
		ERR5000365	0.05	1122				
		ERR5000366	0.05	1122				
	SAMN03002195	SRR1555744	0.75	3283	hydrothermal vent	Arctic Ocean: Gakkel Ridge	PRJNA259156	
	SAMN03002196	SRR1555743	0.74	3283				
		SRR1555748	0.74	3283				
	SAMN04004108	SRR2187304	0.5	1200	cold seep	Spain: Gulf of Cadiz, Capt Aryutinov Mud volcano	PRJNA293476	(Chen et al., 2016)
	SAMN05571518	SRR4028170	2500	4200	hydrothermal vent	Atlantic Ocean	PRJNA338830	
	SAMN16252685	SRR13077602	0.03	7850	subduction zone and trenches	Pacific Ocean: Mariana Trench	PRJNA635214	(Zhou et al., 2022)
	SAMN16252686	SRR13077601	0.03	7850				
	SAMN19461849	SRR14803106	80.64	1000.3	subduction zone and trenches	New Zealand: Hikurangi Subduction Margin	PRJNA678552	
	SAMN19461850	SRR14803105	280.17	1000.3				
	SAMN19461851	SRR14803101	10.35	1000.3				
	SAMN19461852	SRR14803100	110.34	1000.3				
	SAMN19461859	SRR14803104	265.73	3520.3				
	SAMN19461860	SRR14803103	18.52	3520.3				
	SAMN19461861	SRR14803102	51.39	3520.3				

Attachment 3

A.3 - Depth, habitat, taxonomy, completeness (CMPL), and contamination (CONT) of all final MAGs and corresponding samples. In the last column, it is indicated whether nitrogen cycle genes of interest were detected in the MAG. Archaea are highlighted in blue, whereas the remaining taxa are Bacteria.

Zone	Habitat	Sample	Class	Genus	CMPL %	CONT %	N-Cycling?
Bathypelagic (1000m – 4000m)	Cold Seep	SAMEA5663117	JS1	34-128	51.04	3.54	
			Planctomycetia	Unclassified	94.83	2.37	
		SAMEA5663119	Anaerolineae	E44-bin32	82.73	3.94	
			Dehalococcoidia	Unclassified	90.76	5.94	
			Desulfobacteria	*Desulfosarcina*	62.50	0.00	
			Methanosarcinia	UBA204	72.32	0.66	✓
			MSB-5A5	Unclassified	76.73	7.79	
			Spirochaetia	JAHOYX01	88.40	4.85	✓
			Syntropharchaeia	Unclassified	72.85	4.25	
			Thorarchaeia	WTCK01	83.64	4.21	
			UBA6098	Unclassified	84.36	2.20	✓
	Continental Slope and Abyssal Plain	mgm4510162.3	Alphaproteobacteria	*Amylibacter*	52.02	0.86	✓
		mgm4510164.3	Alphaproteobacteria	*Amylibacter*	77.91	1.68	✓
				Sneathiella	54.69	5.33	✓
				Unclassified	82.93	0.28	
			Gammaproteobacteria	50-400-T64	61.02	8.10	✓
		mgm4510168.3	Gammaproteobacteria	*Agaribacterium*	92.86	3.79	✓
		mgm4510170.3	Gammaproteobacteria	*Agaribacterium*	85.87	3.05	
		mgm4510171.3	Gammaproteobacteria	*Agaribacterium*	90.38	3.41	
				Colwellia	56.03	8.62	✓
				UBA3067	87.35	9.99	✓
		mgm4510172.3	UBA6911	Unclassified	68.92	3.04	✓
	Hydrothermal Vent	SAMN03002195	Alphaproteobacteria	*Aurantimonas*	96.30	1.61	✓
			Bacteroidia	*Leeuwenhoekiella*	65.28	1.72	
			Bacteroidia	SM23-62	65.76	2.15	✓
			Gammaproteobacteria	GCA-002733105	90.86	0.42	✓
				Halomonas	97.84	2.01	✓
				Halopseudomonas	99.32	1.12	✓
					99.14	1.00	✓
				Marinobacter	98.10	0.45	✓
				Methylophaga	78.78	8.62	✓
				Pseudohongiella	96.67	0.06	✓
		SAMN03002196	Alphaproteobacteria	*Aurantimonas*	99.60	7.59	✓
			Aminicenantia	SOIV01	78.00	4.84	✓
			Aquicultoria	*Unclassified*	63.28	1.81	✓
			Bacteroidia	*Pricia*	61.68	0.91	
			Gammaproteobacteria	*Halomonas*	100.00	2.44	✓
				Halopseudomonas	93.10	0.86	✓
					95.69	0.60	✓

				Pseudoalteromonas	97.94	6.55	✓
				Pseudohongiella	71.30	0.06	✓
			Phycisphaerae	HyVt-337	98.80	5.88	✓
	Subduction Zone and Trenches	SAMN19461849	Phycisphaerae	HyVt-337	89.20	1.77	
		SAMN19461850	Anaerolineae	Unclassified	92.27	4.55	✓
			Dehalococcoidia	G020351185	88.12	0.00	
				SKVJ01	78.38	1.98	
		SAMN19461851	Dehalococcoidia	E44-bin88	66.67	2.39	
			Hadarchaeia	DG-33	89.60	1.60	
		SAMN19461852	Aerophobia	SOJT01	74.91	1.12	✓
		SAMN19461859	Dehalococcoidia	E44-bin89	61.77	5.61	
		SAMN19461860	Actinomycetia	*Mycobacterium*	55.02	0.63	✓
			Anaerolineae	B3-Chlor	86.24	1.10	
				E44-bin32	85.27	5.61	✓
			Bipolaricaulia	Unclassified	91.53	0.00	
			Dehalococcoidia	E44-bin89	56.11	1.98	
				JAFGPF01	77.78	1.43	
				Unclassified	82.42	0.97	✓
			MSB-5A5	DG-27	95.54	3.30	
				Unclassified	72.13	5.41	
		SAMN19461861	Bathyarchaeia	BS750m-G28	74.47	1.94	
			Bipolaricaulia	Unclassified	87.29	0.00	
			Dehalococcoidia	E29-bin54	88.28	1.98	
				E44-bin89	63.20	0.50	
Abyssopelagic (4000m - 6000m)	Hydrothermal Vent	SAMN05571518	Alphaproteobacteria	Unclassified	66.53	1.03	✓
			Bipolaricaulia	DRJF01	83.05	0.00	✓
			Gammaproteobacteria	HyVt-429	85.98	0.76	✓
			Paceibacteria	C7867-001	69.10	3.22	
			SZUA-79	QIJE01	87.07	0.86	✓
			Thermodesulfovibrionia	BMS3Bbin07	94.09	2.73	✓
			UBA6919	CAJXGR01	65.60	2.27	✓
			UBA7883	DRJW01	88.77	4.44	✓
	Unclassified	SAMEA5744179	Gammaproteobacteria	GCA-2729495	84.00	6.65	✓
				JAACFB01	50.84	8.55	✓
Hadopelagic (>6000m)	Subduction Zone and Trenches	SAMN16252685	Acidimicrobiia	SZUA-217	51.09	0.09	
			Gammaproteobacteria	SMWN01	55.35	1.10	✓
			Nitrospinia	SZUA-226	86.25	3.42	✓
			Terriglobia	JACZYA01	63.88	1.07	
			UBA1144	CR02bin9	52.75	0.00	✓
		SAMN16252686	Acidimicrobiia	SZUA-217	69.61	0.95	
			Nitrospinia	SZUA-226	64.10	0.91	✓
			Terriglobia	JACZYA01	57.55	1.78	✓