



Ecogenomics and metabolic potential of the South Atlantic Ocean microbiome



F.H. Coutinho^{a,b,c,d,1}, F.A.B. von Meijenfildt^c, J.M. Walter^{a,e}, J.M. Haro-Moreno^d, M. Lopéz-Pérez^d, M.C. van Verk^c, C.C. Thompson^a, C.A.N. Cosenza^f, L. Appolinario^a, R. Paranhos^a, A. Cabral^{a,g}, B.E. Dutilh^{b,c}, F.L. Thompson^{a,f,*}

^a Instituto de Biologia (IB), Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil

^b Centre for Molecular and Biomolecular Informatics (CMBI), Radboud University Medical Centre/Radboud Institute for Molecular Life Sciences, Nijmegen, the Netherlands

^c Theoretical Biology and Bioinformatics, Science for Life, Utrecht University (UU), Utrecht, the Netherlands

^d Evolutionary Genomics Group, Departamento de Producción Vegetal y Microbiología, Universidad Miguel Hernández, Alicante, Spain

^e Faculty of Biosciences and Aquaculture, Nord University, Bodo, Norway

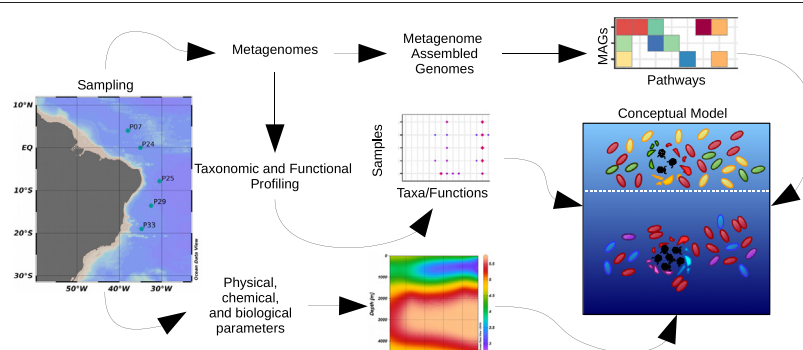
^f COPPE/SAGE, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil

^g Instituto de Microbiologia Paulo de Góes, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil

HIGHLIGHTS

- Metagenomics was applied to unravel the prokaryote and viral diversity in the South Atlantic Ocean.
- Metagenome assembled genomes obtained belonged to novel taxa of Bacteria and Archaea from the mesopelagic and bathypelagic
- The metabolic apparatus bacteria of the aphotic zone revealed a capacity for the use of recalcitrant organic matter
- The metabolic apparatus of mixotrophic bacteria highlighted the relevance of the Calvin-Benson-Bassham cycle
- A model was proposed for the roles of taxa of prokaryotes and viruses to biological processes in the South Atlantic Ocean.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 21 July 2020

Received in revised form 28 September 2020

Accepted 28 September 2020

Available online 17 October 2020

Editor: Frederic Coulon

Keywords:

South Atlantic Ocean

ABSTRACT

The unique combination of depth, salinity, and water masses make the South Atlantic Ocean an ecosystem of special relevance within the global ocean. Yet, the microbiome of this ecosystem has received less attention than other regions of the global Ocean. This has hampered our understanding of the diversity and metabolic potential of the microorganisms that dwell in this habitat. To fill this knowledge gap, we analyzed a collection of 31 metagenomes from the Atlantic Ocean that spanned the epipelagic, mesopelagic and bathypelagic zones (surface to 4000 m). Read-centric and gene-centric analysis revealed the unique taxonomic and functional composition of metagenomes from each depth zone, which was driven by differences in physical and chemical parameters. In parallel, a total of 40 metagenome-assembled genomes were obtained, which recovered one third of the total community. Phylogenomic reconstruction revealed that many of these genomes are derived from poorly

* Corresponding author at: Av. Carlos Chagas Filho, S/N - CCS - IB - BIOMAR - Laboratório de Microbiologia - BLOCO A (Anexo) A3 - sl 102, Cidade Universitária, Rio de Janeiro, RJ 21941-599, Brazil.

E-mail address: fabianothompson1@gmail.com (F.L. Thompson).

¹ Lead contact.

Bathypelagic
Metagenomics
Metagenome assembled genomes
Viruses

characterized taxa of Bacteria and Archaea. Genomes derived from heterotrophic bacteria of the aphotic zone displayed a large apparatus of genes suited for the utilization of recalcitrant organic compounds such as cellulose, chitin and alkanes. In addition, we found genomic evidence suggesting that mixotrophic bacteria from the bathypelagic zone could perform carbon fixation through the Calvin-Benson-Bassham cycle, fueled by sulfur oxidation. Finally, we found that the viral communities shifted throughout the water column regarding their targeted hosts and virus-to-microbe ratio, in response to shifts in the composition and functioning their microbial counterparts. Our findings shed light on the microbial and viral drivers of important biogeochemical processes that take place in the South Atlantic Ocean.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

The South Atlantic Ocean houses a remarkable, yet mostly untapped, microbial biodiversity (Alves Junior et al., 2014). Therefore, this region has great potential for the discovery of novel microorganisms and to describe the contribution of this diversity to ecological processes. Significant efforts have been undertaken to describe the genetic diversity of the global ocean (Williamson et al., 2008; Sunagawa et al., 2015), yet most of these focused on surface waters. In contrast, deep waters of the South Atlantic Ocean (SAO) have received far less attention regarding efforts to link ecological processes to the genomic diversity therein. Previous studies have only begun to scrape the surface of the microbial diversity of this habitat (Martin-Cuadrado et al., 2008; Ghai et al., 2011; Alves Junior et al., 2014) and have mostly focused on samples from the epipelagic zone (0–200 m depth) while the mesopelagic (200–1000 m depth) and bathypelagic (1000–4000 m depth) zones have been largely ignored, mainly due to the intrinsic difficulty of obtaining samples at these depths. Nevertheless, the deep sea is known to be a hot-spot of microbial biodiversity (Sogin et al., 2006). Consequently, more studies are needed to reveal the microbial biodiversity and ecology of the microbiome of the deep South Atlantic Ocean.

Several physical, chemical and biological factors make the SAO a unique ecosystem. The Atlantic is the saltiest of the major oceans and it is also shallower than the Pacific and the Indian oceans at an average depth of 3646 m (Tomczak and Godfrey, 2013). The Atlantic Ocean has higher levels of calcite and aragonite saturation than the Pacific (Zeebe and Wolf-Gladrow, 2001). This phenomenon takes place because the deep waters in the Pacific Ocean are older than the deep waters of Atlantic. The older deep Pacific Ocean has taken up more CO₂ from remineralization of organic matter which lowers its carbonate ion content. In addition the SAO houses multiple water masses that are influenced by the South Atlantic gyre, each with unique salinity and temperature (Emery and Meincke, 1986), both factors that are major drivers of microbial community composition in aquatic ecosystems (Lozupone and Knight, 2007; Sunagawa et al., 2015; Thompson et al., 2017). Meanwhile, circulation of deep water masses has been shown to also be a major driver of microbial community composition (Salazar et al., 2016). Together, this combination of factors might be responsible for extensive diversity of novel genes reported for the SAO specially in the deeper layers (Sunagawa et al., 2015).

Deep oceans are regarded as zones of low primary productivity, which are sustained by the sinking organic matter that is fixed at the epipelagic zone by photoautotrophic organisms using the Calvin-Benson-Bassham (CBB) cycle. In addition, it has been demonstrated that deep oceans have their own primary productivity fueled by alternative carbon fixation pathways that do not rely on light (Sogin et al., 2006; Martín-Cuadrado et al., 2007; Orcutt et al., 2011; Xie et al., 2011). This dark carbon fixation fueled by the oxidation of inorganic compounds (e.g. ammonium and sulfur) was shown to be an important source of primary productivity in the oceans, at both photic and aphotic zones (Baltar and Herndl, 2019). Additional studies provided evidence that chemoautotrophy is widespread among microbes of the dark ocean (Guerrero-Feijóo et al., 2018) such as nitrite-oxidizing bacteria (e.g. Nitrospinae)

(Pachiadaki et al., 2017), SAR324, SAR202 (Chloroflexi) (Guerrero-Feijóo et al., 2018), as well as Crenarchaeota and Euryarchaeota (Herndl et al., 2005; Hügler and Sievert, 2011). Some of these organisms make use of alternative carbon fixation pathways such as the reductive tricarboxylic acid (rTCA) cycle and the 3-hydroxypropanoate cycle to fix carbon in the dark (Hügler and Sievert, 2011).

Environmental DNA sequencing techniques have brought major advancements to the understanding of marine microbial communities of the global ocean (Coutinho et al., 2015; Sunagawa et al., 2015; Roux et al., 2016). These approaches helped elucidate how environmental conditions affect the taxonomic and functional composition of the marine microbiome. Amplicon based sequencing, although extremely useful, is limited in its capacity to provide functional insights about microbial communities. Metagenomics circumvents this limitation as it allows for recovering complete and partial genomes of microorganisms and viruses. Paired with data on physical, chemical and biological parameters this approach has contributed to expanding our understanding of the biological diversity in the oceans and of the ecological processes that take place therein. Specifically, assembly-based approaches can be used to explore the genomic context of microorganisms and their viruses describing their genomic context, simultaneously providing insights into their taxonomic and functional diversity (Brown et al., 2015; Anantharaman et al., 2016; Paez-Espino et al., 2016; Roux et al., 2016; Haro-Moreno et al., 2017; Parks et al., 2017; Zaremba-Niedzwiedzka et al., 2017). Importantly, new tools have emerged to classify this yet unknown diversity and place it within a consistent and robust taxonomic framework established through phylogenomics (Parks et al., 2018; von Meijenfeldt, 2019), allowing for major advancements in our understanding of the diversity, evolution and ecology of uncultured organisms (Soo et al., 2017; Castelle and Banfield, 2018; Rinke et al., 2019).

Our goal was to use metagenomics to gain insights into ecological processes of the deep Atlantic Ocean. We analyzed the taxonomic and functional composition of 31 metagenomes collected across a depth gradient ranging from 0 to 4000 m spanning the epi- meso- and bathypelagic zones at five distinct locations (Fig. 1A). These samples were retrieved during the PIRATAS sampling expedition that took place between August and October of 2011 (Alves Junior et al., 2014). We also performed assembly and binning of these metagenomes which yielded draft genomes of Archaea, Bacteria and viruses. Many of these were classified as members of yet poorly characterized evolutionary lineages, and their genomic content suggested that they contribute to biogeochemical cycles that are fundamental for sustaining life at the deep Atlantic Ocean.

2. Results and discussion

2.1. Shifts in physico-chemical conditions, cell, and viral abundances throughout the water column

Our sampling sites were characterized by measuring physical and chemical water parameters as well as abundances of bacterial cells and viral particles (Table S1). Principal component analysis (Fig. 1B) grouped samples into three clearly distinct groups: epipelagic, mesopelagic and bathypelagic. PC1 explained 70.1% of the variance in the data and separated samples of the photic (epipelagic) from aphotic

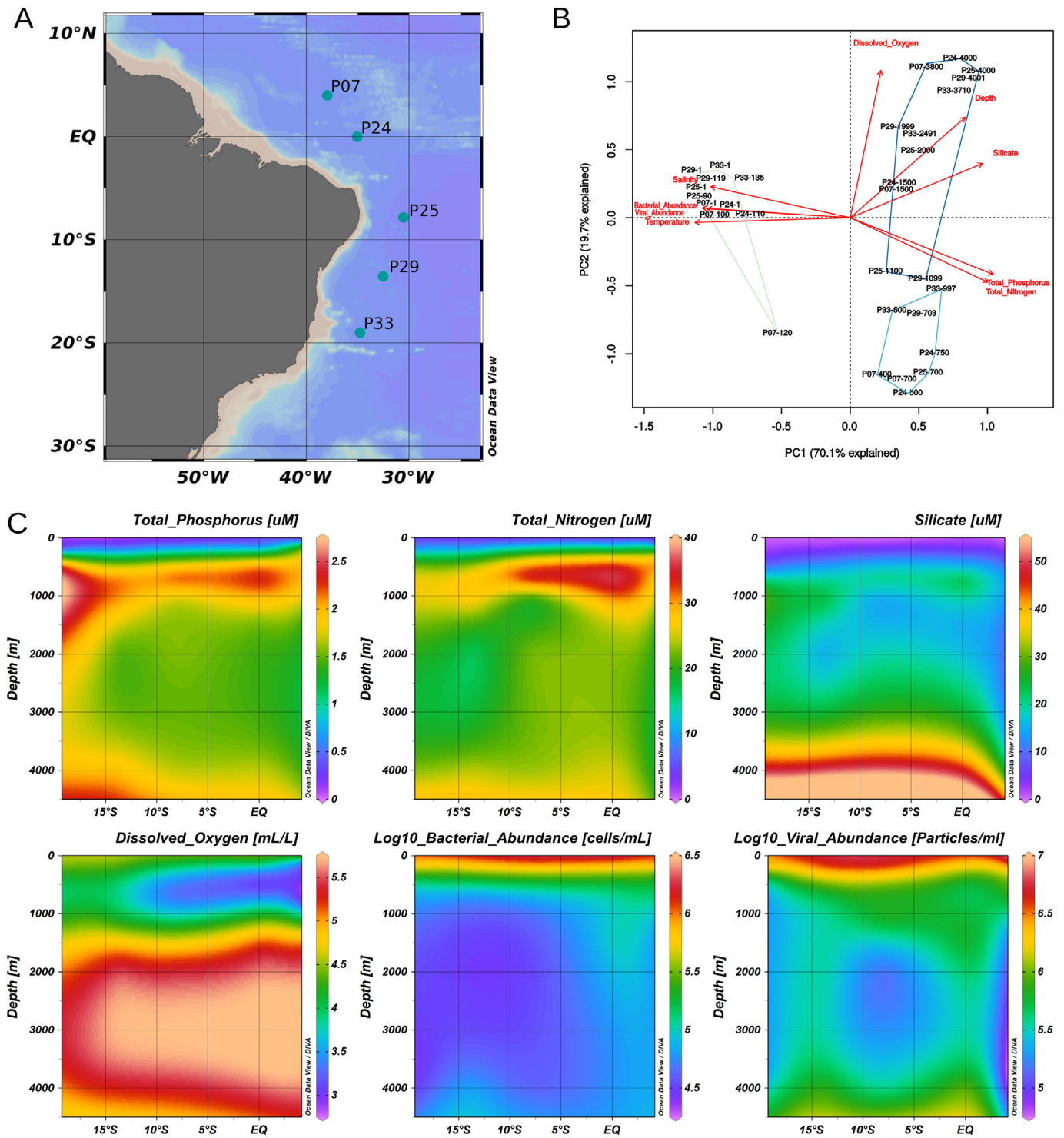


Fig. 1. Sampling sites and physical, chemical and biological parameters of the water column A) Map of sampling sites across the Atlantic Ocean from which metagenomes were obtained. B) Principal component analysis reveals three distinct groups of samples: epipelagic (green), mesopelagic (light blue) and bathypelagic (dark blue). C) Profiles displaying the concentrations of total phosphorus, total nitrogen, silicate, abundance of bacterial cells and abundance of viral particles throughout the latitude and depth gradients spanned by the samples.

(meso- and bathypelagic) zone. This separation was mostly driven by the higher values of temperature, salinity, bacterial and viral abundances among samples of the epipelagic zone. Meanwhile PC2 explained 19.7% of the variance in the data and separated the aphotic zone samples into mesopelagic and bathypelagic groups. This separation was mostly driven by the higher concentrations of total phosphorus and total nitrogen among the mesopelagic samples, and by depth and

higher silicate and dissolved oxygen concentrations of the bathypelagic samples.

The epipelagic samples were characterized by the highest values of temperature (21–28 °C) and salinity (35.6–37.3 PSU, Fig. 1B), intermediate dissolved oxygen concentrations (3.1–4.5 mL/L, Fig. 1C), and the lowest concentrations of total nitrogen (4.1–14.0 μM), total phosphorus (0.2–1.6 μM) and silicate (0.9–2.4 μM) (Table S1). The mesopelagic zone

samples displayed lower temperature (3.6–10.0 °C), salinity (34.4–34.9 PSU), and dissolved oxygen concentrations (2.9–4.1 mL/L), but the highest concentrations of total phosphorus (1.8–2.7 µM) and total nitrogen (22.7–36.6 µM), and intermediate concentrations of silicate (8.7–28.0 µM). Finally the bathypelagic samples displayed the lowest temperatures (1.7–4.4 °C) but salinity comparable to the mesopelagic zone (34–34.9 PSU). These samples displayed intermediate concentrations of total phosphorus (1.4–2.1 µM) and total nitrogen (15.7–27.8 µM) but the highest concentrations of silicate (11.3–50.6 µM) and dissolved oxygen (3.8–5.5 mL/L).

Epipelagic samples displayed the highest values of bacterial cells ($7.3\text{--}18.0 \times 10^5$) and viral particles ($2.5\text{--}6.9 \times 10^6$). These values decreased steeply among mesopelagic samples (bacterial abundance $0.5\text{--}2.2 \times 10^5$ and viral abundance $0.2\text{--}1 \times 10^6$) and were even lower among bathypelagic samples (bacterial abundance $0.3\text{--}1.0 \times 10^5$ and viral abundance $0.2\text{--}1.0 \times 10^6$). Interestingly, the virus-to-microbe ratio (VMR) tended to increase with depth. The average VMR value was 3.8 among epipelagic samples, 5.1 among mesopelagic samples, and 8.7 among bathypelagic samples (Fig. S1). According to the results obtained through the Mann-Whitney test these differences were significant when comparing the epipelagic against the bathypelagic zone ($p = 0.0086$), and borderline significant when comparing the mesopelagic and bathypelagic zones ($p = 0.098$).

2.2. Taxonomic and functional composition of the SAO microbial community

The taxonomic composition profiles of metagenomes from the epimeso- and bathypelagic zones were markedly different at both the levels of phyla (Fig. 2) and genera (Fig. S2). This pattern is likely a consequence of differences in light availability, temperature and nutrient abundances throughout water column, which affect microbial community composition (Alves Junior et al., 2014; Nunoura et al., 2015; Haro-Moreno et al., 2018; Acinas et al., 2019). Samples from the epipelagic zone (0–120 m) were dominated by *Alphaproteobacteria* followed by *Cyanobacteria* and *Gammaproteobacteria* (Fig. 2). *Actinobacteriota*, *Bacteroidota*, *Thermoplasmata*, and *Crenarchaeota* were present at this zone as well, albeit at lower relative abundances. Among epipelagic samples, the single 120 m metagenome displayed a taxonomic profile slightly different from the samples retrieved at the surface. This sample had a much higher abundance of *Crenarchaeota*. In addition, SAR324, *Marinisomatota* and *Nitrospinota* were also detected at 120 m but at low relative abundances.

Proteobacteria were also the dominant group in samples from the mesopelagic zone (500–750 m), yet *Gammaproteobacteria* often outnumbered *Alphaproteobacteria* at the samples from this zone. Most taxa that were abundant in the epipelagic zone tended to also be abundant at mesopelagic samples, except for *Cyanobacteria* which were not detected in samples from depths below 120 m. In addition, *Chloroflexota* and *Crenarchaeota* were much more abundant in this zone compared to the epipelagic samples. Other taxa with low abundance or undetected at the epipelagic samples that were abundant in the mesopelagic samples included: *Acidobacteriota*, SAR324, *Marinisomatota*, *Planctomycetota* and *Nitrospinota*.

The bathypelagic zone samples (2000–4000 m) were most often dominated by *Gammaproteobacteria* followed by *Alphaproteobacteria* as the second most abundant group. A notable exception to this pattern were two samples retrieved from 3500 m at site P07, which displayed much lower abundances of these classes and higher abundances of SAR324, *Chloroflexota*, and *Crenarchaeota*. These taxa displayed notable abundances in samples retrieved between 2000 and 3500 m, but were most often below 1000 RPKM among the samples retrieved from 4000 m. The unique taxonomic profile of the bathypelagic metagenome from site P07 might be related to concentrations of total phosphorus, total nitrogen, and silicate which were the lowest among the samples retrieved from 3700 to 4000 m deep.

Overall, the changes in community composition observed throughout the depth gradient at the level of phylum could be linked to changes in environmental parameters (i.e. water physical and chemical conditions). According to the Mantel test, these associations were statistically significant (Mantel statistic $r: 0.56, p = 0.001$). Furthermore, we used Canonical Correspondence Analysis (CCA) to explore in detail the associations between samples, taxa and environmental parameters. This revealed an association between *Cyanobacteria* and *Actinobacteria* with the highest values of temperature and salinity, which were observed among epipelagic samples. Meanwhile, *Alphaproteobacteria*, *Gammaproteobacteria*, *Bacteroidota* and *Bdellovibrionota* displayed an association with the highest values of depth, silicate and dissolved oxygen among the bathypelagic samples. Finally, *Chloroflexota*, *Poribacteria*, *Desulfobacterota*, *Acidobacteriota*, *Gemmatimonadota*, SAR324, *Crenarchaeota* and *Thermoplasmata* displayed an association with intermediate values of the aforementioned parameters and slightly positive association with total phosphorus and total nitrogen among samples from the mesopelagic zone (Fig. 3). These findings are in agreement with previous studies focused on the SAO (Alves Junior et al., 2014) and the global oceans (Sunagawa et al., 2015) that reported water masses and temperature as major driving factors of marine microbial community composition.

Functional profiles also displayed marked variation throughout the depth gradient (Fig. 4). The marked differences between depth zones observed through the read-centric approach were also observed through a gene-centric approach (Fig. S3). Several pathways displayed highest relative abundances among the epipelagic zone samples. These included pathways involved in photosynthesis and hydrogen production, processes carried out by *Cyanobacteria* (Tamagnini et al., 2003; Dutta et al., 2005), which were among the dominant taxa at this zone. Alternative pathways of carbon fixation, i.e. the rTCA and the 3-hydroxypropanoate cycle, displayed the highest relative abundances in samples from the mesopelagic and bathypelagic zones. Among the pathways with higher relative abundances in samples from the bathypelagic zone were: degradation of aromatic compounds, octane oxidation, and nitrate reduction. Finally, multiple pathways displayed a more even distribution throughout the depth gradient, with comparable abundances at all zones. These included several constitutive cellular functions, such as the TCA cycle, pentose phosphate pathway, glycolysis, and aerobic respiration. Interestingly, the Calvin-Benson-Bassham cycle was also included among those, suggesting a potential for carbon fixation in both the photic (epipelagic) and aphotic (mesopelagic and bathypelagic) zones. While in the photic zone this process is likely driven by photosynthetic *Cyanobacteria* and photosynthetic Eukaryotes, in the aphotic zone dark carbon fixation can take place (Santoro et al., 2013; Dykstra et al., 2016; Zhou et al., 2017), where this process is mediated by chemoautotrophic organisms such as some of the members of *Nitrospinota* and *Proteobacteria* (Hügler and Sievert, 2011). These findings are in agreement with previous reports of the importance of dark carbon fixation and myxotrophy in the bathypelagic zone (Acinas et al., 2019).

2.3. The deep Atlantic Ocean harbors new species of Bacteria and Archaea

Assembly and binning of metagenomes yielded a total of 40 metagenome assembled genomes (MAGs) with completeness > 25%; contamination < 15%, and completeness > 3× contamination (Table 1). According to the Minimum Information about Metagenome-Assembled Genome (MIMAG) standards (Bowers et al., 2017), these MAGs belonged to the categories of high-quality draft (3), medium-quality draft (15), and low-quality draft (22) genomes. The MAGs ranged in size from 0.2 to 4.1 Mbp and encoded 296 to 3980 proteins (Table 1). These metrics are of course influenced by genome completeness, meaning that they are not representative of the true genome size and number of encoded proteins, especially for the medium-quality and low-quality MAGs. The taxonomic affiliation of the MAGs was determined through phylogenomic reconstruction (Fig. 5) and classification

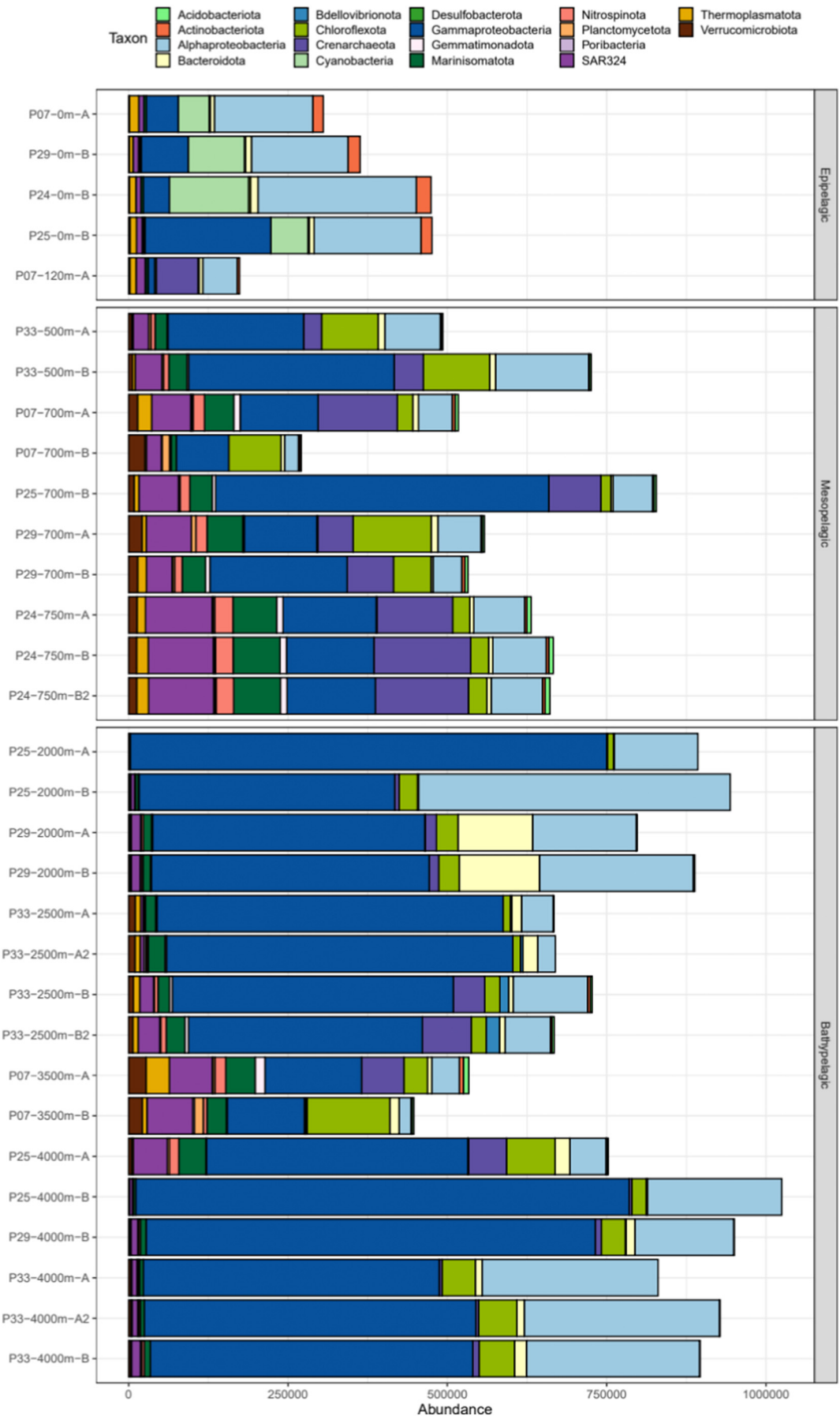


Fig. 2. Atlantic Ocean microbiome taxonomic profiles. Bar plots depict the relative abundances of phyla (or class for Proteobacteria) in the metagenomes expressed as RPKM derived from the gene-centric analysis. Only taxa that displayed relative abundances equal or above 1000 RPKM are shown.

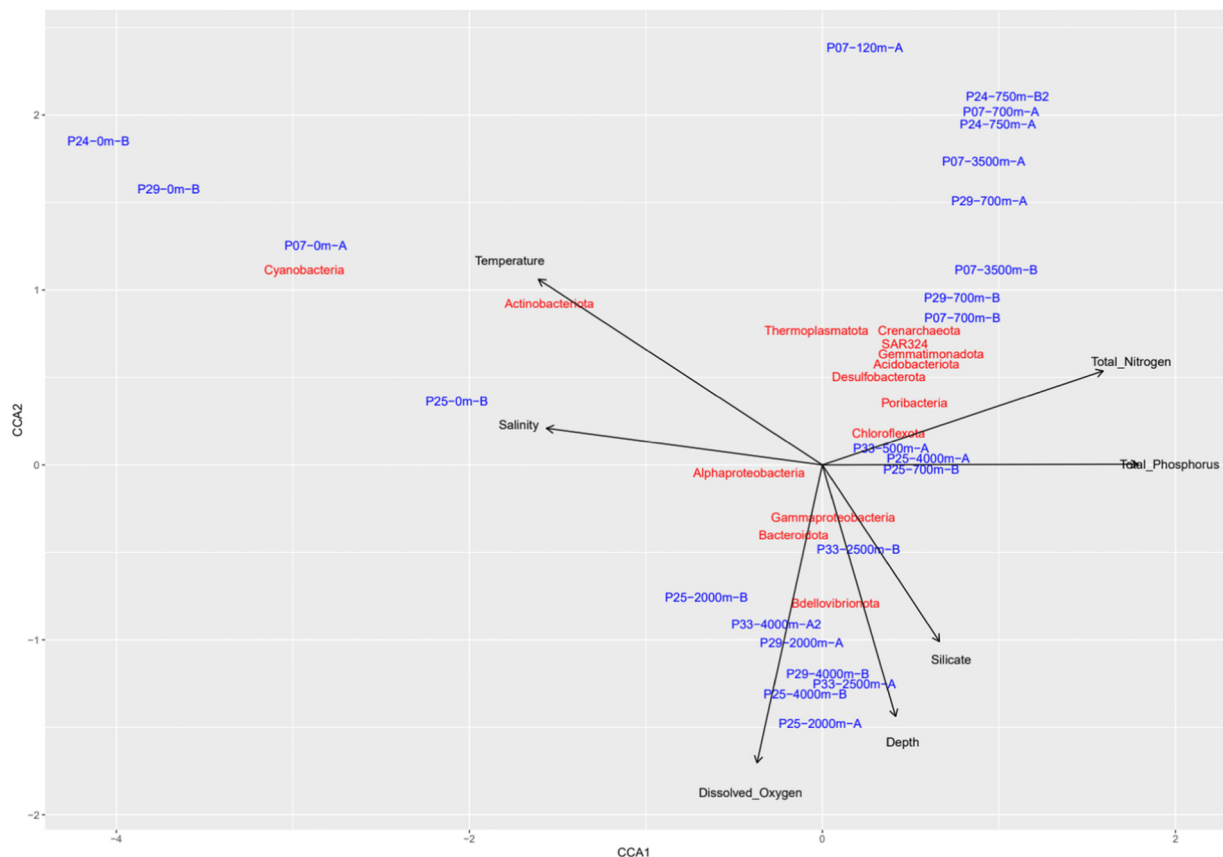


Fig. 3. Canonical correspondence analysis depicting associations among taxa abundances, samples and environmental parameters. Taxon abundances derived from the gene-centric analysis are shown at the level of phylum (or class for Proteobacteria). Only taxa that displayed minimum abundances equal or above to 1000 RPKM were included in this analysis. Taxon and sample labels were slightly shifted from their original positions to avoid overlap.

within the newly proposed genomic taxonomy of Bacteria and Archaea (Parks et al., 2018). The single archaeal MAG obtained was affiliated with marine group II (MG-II). The remaining 39 MAGs were members of the bacterial domain. The taxa most often represented among the bacterial MAGs were *Gammaproteobacteria* (15), *Alphaproteobacteria* (8), *Bacteroidota* (4), *Chloroflexota* (4), and SAR324 (3). Taxa that were less often represented by the MAGs included *Marinisomatota* (2), *Cyanobacteria* (1), *Gemmatimonadota* (1), and *Nitrospina* (1). Notably many MAGs were classified as members of taxa that have only recently been proposed by the genomic taxonomy of bacteria (Parks et al., 2018). Specifically these MAGs were classified as members of newly proposed, yet unnamed, taxa (e.g. UBA3495, UBA10353 and UBA11654). A total of 20 MAGs were classified as new species (Table 1). In addition, some MAGs were affiliated to genera with few genome representatives (e.g. UBA1123 with only four genomes described, and UBA11650 with only five genomes), hence their discovery contributes to expanding the known diversity of organisms within these taxa, which is of special relevance regarding the need to describe the biodiversity of microorganisms that dwell at the deep ocean.

On average, 33.45% of the reads from the metagenomes could be mapped back to the MAGs, suggesting that these genomes accounted for about one third of sequenced DNA from the samples. We investigated the abundance patterns of the MAGs across depth zones to elucidate their abundance distributions across different layers within the water column (Fig. 6). According to their relative abundance patterns the MAGs could be grouped into three major groups representing the different depth zones: those that displayed highest relative abundance in the epipelagic zone, those that displayed highest relative abundance in the mesopelagic, and those that displayed highest relative abundance in the bathypelagic zone (Table 1). Those that displayed highest median

relative abundances in the epipelagic zone were MAGs M2 (*Pelagibacter*), M4 (*Prochlorococcus*) and MAG 9 (*Owenweeksia*). Those that displayed highest median relative abundances in the mesopelagic zone include those affiliated to: SAR324 (MAGs 24, 38 and 42), MG-II Archaea (MAG 35), multiple MAGs classified as *Gammaproteobacteria* (e.g. MAGs 13, 21 and 33), as well as those classified as *Nitrospina* (MAG39) and *Marinisomatota* (MAGs 19 and 22). Among the MAGs that displayed highest median relative abundances in the bathypelagic zone were several *Alphaproteobacteria* (e.g. MAGs 10, 15 and 16), and *Gammaproteobacteria* (e.g. MAGs 11, 17 and 18).

Deep ocean waters differ from surface waters in several physical, chemical and biological aspects (Fig. 1). Deep water habitats are characterized by the absence of light and consequently of photosynthesis driven primary productivity, as well as low temperatures and high concentrations of inorganic nutrients. These conditions select for microbiomes with the necessary adaptations to thrive in this harsh environment. Therefore, the relative abundance distribution of the MAGs reflects the preference of each organism to different depth zones according to the adaptations encoded in their genomes. The association between MAG abundances and environmental parameters was significant according to the Mantel test (Mantel statistic $r = 0.57$, $p = 0.001$). Furthermore, canonical correspondence analysis revealed the strength and the direction of the associations between MAGs and the tested environmental parameters (Fig. 7). A similar pattern to the one obtained with the CCA of phyla was observed. Namely a group of MAGs associated with higher values of temperature and salinity among epipelagic samples, a second group of MAGs associated with higher values of depth, silicate and dissolved oxygen among bathypelagic samples, and a third group of MAGs associated with intermediate values of the aforementioned variables and slightly

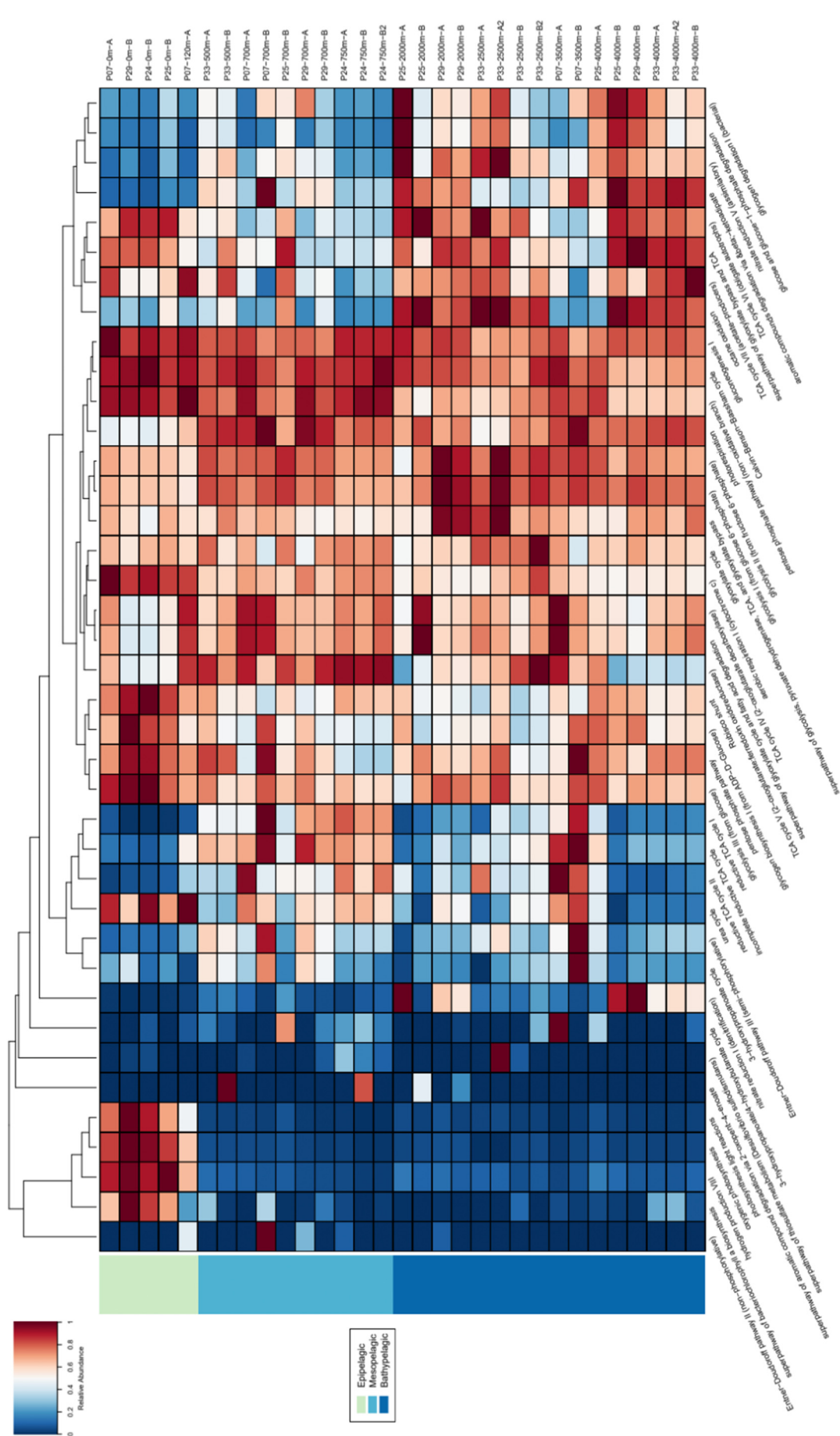


Fig. 4. Atlantic Ocean microbiome functional profiles. Heatmap displays the relative abundances of a selected subset of pathways involved in nutrient utilization and energy metabolism. Pathway abundances were calculated through the annotation of raw reads with Humann2 (read-centric analysis). The cladogram was established through complete hierarchical clustering based on Bray-Curtis distances. Color scale represents abundances of functional pathways relative to the maximum abundance of that pathway across all samples.

Table 1

Detailed description of the obtained MAGs including: Genome size, number of PEGs, completeness, contamination and taxonomic classification. MAGs that required manual curation of the bins are indicated by the letter "M" in their identifier. The preferred depth zones of the MAGs was determined based on the abundance profiles displayed in Fig. 6.

| MAG | Completeness | Contamination | Quality | MIMAG category | GTDB taxonomic classification | Preferred zone | Contigs | Bases | N50 | PEGs |
|--------------------------------|--------------|---------------|---------|----------------------|---|----------------|---------|-----------|--------|------|
| MAG_10 Alphaproteobacteria | 88.09 | 1.46 | 86.63 | Medium-quality draft | d__Bacteria;p__Proteobacteria; c__Alphaproteobacteria; o__Caulobacterales; f__Hyphomonadaceae; g__Hyphomonas; s__GCF_000682775.1 | Bathypelagic | 197 | 3,293,894 | 26,873 | 3358 |
| MAG_11 Gammaproteobacteria | 43.45 | 1.72 | 41.73 | Low-quality draft | d__Bacteria;p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Hahellaceae; g__Marinobacter; s__Marinobacter adhaerens | Bathypelagic | 437 | 3,229,831 | 9052 | 3288 |
| MAG_12 Dehalococcoidia | 87.94 | 0.99 | 86.95 | Medium-quality draft | d__Bacteria;p__Chloroflexota; c__Dehalococcoidia;o__SAR202; f__UBA11138;g__UBA1123;s__ | Mesopelagic | 407 | 3,108,582 | 10,413 | 3285 |
| MAG_13 Gammaproteobacteria | 43.1 | 7.76 | 35.34 | Low-quality draft | d__Bacteria;p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Pseudohongiellaceae; g__UBA9145;s__ | Mesopelagic | 718 | 3,027,884 | 4253 | 3219 |
| MAG_14 Dehalococcoidia | 41.71 | 5.26 | 36.45 | Low-quality draft | d__Bacteria;p__Chloroflexota; c__Dehalococcoidia; o__UBA3495;f__UBA3495; g__UBA9611;s__ | Mesopelagic | 485 | 1,871,753 | 3865 | 2115 |
| MAG_15 Alphaproteobacteria | 82.73 | 1.09 | 81.64 | Medium-quality draft | d__Bacteria;p__Proteobacteria; c__Alphaproteobacteria; o__Caulobacterales; f__Hyphomonadaceae; g__Henriciella; s__GCA_002733705.1 | Bathypelagic | 285 | 2,889,050 | 15,337 | 2980 |
| MAG_16 Alphaproteobacteria | 80.87 | 2.55 | 78.32 | Medium-quality draft | d__Bacteria;p__Proteobacteria; c__Alphaproteobacteria; o__Sphingomonadales; f__Sphingomonadaceae; g__Erythrobacter_A; s__GCA_002377055.1 | Bathypelagic | 300 | 2,843,423 | 13,734 | 2919 |
| MAG_17 Gammaproteobacteria | 84.91 | 0.92 | 83.99 | Medium-quality draft | d__Bacteria;p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Halomonadaceae; g__Halomonas;s__ | Bathypelagic | 335 | 2,755,659 | 10,463 | 2724 |
| MAG_18 Gammaproteobacteria | 76.72 | 1.31 | 75.41 | Medium-quality draft | d__Bacteria;p__Proteobacteria; c__Gammaproteobacteria; o__Pseudomonadales; f__Saccharospirillaceae; g__Oleibacter;s__Oleibacter marinus | Bathypelagic | 339 | 2,819,673 | 11,360 | 2801 |
| MAG_19 Marinisomatia | 42.01 | 12.07 | 29.94 | Low-quality draft | d__Bacteria;p__Marinisomatota; c__Marinisomatia; o__Marinisomatales;f__TCS55; g__TCS55;s__ | Mesopelagic | 509 | 2,086,491 | 3985 | 2235 |
| MAG_2 Alphaproteobacteria | 60.8 | 9.46 | 51.34 | Medium-quality draft | d__Bacteria;p__Proteobacteria; c__Alphaproteobacteria; o__Rhodobacteriales; f__Rhodobacteraceae; g__Roseivivax;s__ | Bathypelagic | 405 | 2,925,088 | 8463 | 3107 |
| MAG_20 Dehalococcoidia | 46.07 | 1.49 | 44.58 | Low-quality draft | d__Bacteria;p__Chloroflexota; c__Dehalococcoidia; o__UBA3495;f__UBA3495; g__UBA9611; s__GCA_002746355.1 | Mesopelagic | 365 | 2,433,737 | 7652 | 2555 |
| MAG_21 Gammaproteobacteria | 43.28 | 3.45 | 39.83 | Low-quality draft | d__Bacteria;p__Proteobacteria; c__Gammaproteobacteria; o__UBA10353;f__LS-SOB; g__UBA11791;s__UBA11791 sp1 | Mesopelagic | 492 | 1,987,824 | 4077 | 2212 |
| MAG_22 Marinisomatia | 30.72 | 3.45 | 27.27 | Low-quality draft | d__Bacteria;p__Marinisomatota; c__Marinisomatia; o__Marinisomatales;f__TCS55; g__TCS55;s__GCA_002716525.1 | Mesopelagic | 368 | 1,465,895 | 3981 | 1498 |
| MAG_23 Alphaproteobacteria | 90.09 | 3.19 | 86.9 | High-quality draft | d__Bacteria;p__Proteobacteria; c__Alphaproteobacteria; o__Micavibrionales; f__Micavibrionaceae;g__;s__ | Bathypelagic | 134 | 2,121,744 | 22,896 | 2019 |

Table 1 (continued)

| MAG | Completeness | Contamination | Quality | MIMAG category | GTDB taxonomic classification | Preferred zone | Contigs | Bases | N50 | PEGs |
|----------------------------|--------------|---------------|---------|----------------------|---|----------------|---------|-----------|---------|------|
| MAG_24 SAR324 | 63.27 | 2.2 | 61.07 | Medium-quality draft | d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__Arctic96AD-7;s__ | Mesopelagic | 376 | 1,901,545 | 5754 | 2008 |
| MAG_29 Gemmatimonadetes | 42.84 | 0 | 42.84 | Low-quality draft | d__Bacteria;p__Gemmatimonadota;c__Gemmatimonadetes;o__SG8-23;f__UBA6960;g__UBA887;s__ | Mesopelagic | 368 | 1,680,107 | 5078 | 1687 |
| MAG_3 Gammaproteobacteria | 25.86 | 0 | 25.86 | Low-quality draft | d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Alteromonadaceae;g__Pseudoalteromonas;s__ | Bathypelagic | 666 | 4,138,443 | 7194 | 3954 |
| MAG_33 Gammaproteobacteria | 32.07 | 5.17 | 26.9 | Low-quality draft | d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__HTCC2089;g__UBA11889;s__GCA_002313255.1 | Mesopelagic | 373 | 1,349,224 | 3484 | 1511 |
| MAG_35 MGII | 77.69 | 0.93 | 76.76 | Medium-quality draft | d__Archaea;p__Thermoplasmata;c__MGII;o__MGII;f__MGIIB;g__UBA623;s__GCA_002507125.1 | Mesopelagic | 128 | 1229848 | 13160 | 1112 |
| MAG_38 SAR324 | 26.18 | 0 | 26.18 | Low-quality draft | d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__Arctic96AD-7;s__Arctic96AD-7 sp1 | Mesopelagic | 156 | 1,048,661 | 8865 | 983 |
| MAG_39 Nitrospina | 41.21 | 1.28 | 39.93 | Low-quality draft | d__Bacteria;p__Nitrospina;c__Nitrospina;o__Nitrospinales;f__Nitrospinaceae;g__SCGCAAA288-L16;s__ | Mesopelagic | 288 | 1,021,459 | 3421 | 1181 |
| MAG_4 Gammaproteobacteria | 84.94 | 1.94 | 83 | Medium-quality draft | d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Hahellaceae;g__Marinobacter;s__Marinobacterhydrocarbonoclasticus | Bathypelagic | 345 | 4,042,016 | 17,167 | 3980 |
| MAG_42 SAR324 | 30.07 | 0 | 30.07 | Low-quality draft | d__Bacteria;p__SAR324;c__SAR324;o__SAR324;f__NAC60-12;g__Arctic96AD-7;s__Arctic96AD-7 sp1 | Mesopelagic | 129 | 749,705 | 6636 | 709 |
| MAG_43 Gammaproteobacteria | 41.38 | 3.45 | 37.93 | Low-quality draft | d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__UBA11654;f__UBA11654;g__UBA6731;s__GCA_002453985.1 | Mesopelagic | 145 | 707,796 | 4857 | 765 |
| MAG_49 Gammaproteobacteria | 39.57 | 0.89 | 38.68 | Low-quality draft | d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__SAR86;f__SAR86;g__AEGEAN-183;s__ | Mesopelagic | 152 | 622,162 | 4265 | 733 |
| MAG_5 Dehalococcidia | 67.64 | 13.86 | 53.78 | Low-quality draft | d__Bacteria;p__Chloroflexota;c__Dehalococcidia;o__UBA3495;f__UBA3495;g__UBA11650;s__GCA_002401285.1 | Bathypelagic | 711 | 3,594,278 | 5540 | 3821 |
| MAG_6 Gammaproteobacteria | 82.48 | 1.82 | 80.66 | Medium-quality draft | d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Alteromonadaceae;g__Alteromonas_A;s__ | Bathypelagic | 457 | 4,083,728 | 11,424 | 3884 |
| MAG_63 Gammaproteobacteria | 25.09 | 0 | 25.09 | Low-quality draft | d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Alcanivoracaceae;g__Alcanivorax;s__ | Bathypelagic | 55 | 233,283 | 4490 | 296 |
| MAG_7 Gammaproteobacteria | 88.22 | 1.01 | 87.21 | Medium-quality draft | d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacterales;f__Alteromonadaceae;g__Alteromonas;s__Alteromonasmacleodii | Bathypelagic | 482 | 3,687,390 | 10,053 | 3319 |
| MAG_8 Gammaproteobacteria | 97.4 | 0.37 | 97.03 | High-quality draft | d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria; | Bathypelagic | 31 | 3,467,569 | 188,124 | 3207 |

(continued on next page)

Table 1 (continued)

| MAG | Completeness | Contamination | Quality | MIMAG category | GTDB taxonomic classification | Preferred zone | Contigs | Bases | N50 | PEGs |
|-----------------------------|--------------|---------------|---------|----------------------|--|----------------|---------|-----------|--------|------|
| MAG_9 Bacteroidia | 98.12 | 1.08 | 97.04 | High-quality draft | o_Pseudomonadales; f_Alcanivoraceae; g_Alcanivorax;s_Alcanivorax sp2 d_Bacteria;p_Bacteroidota; c_Bacteroidia; o_Flavobacteriales; f_Cryomorphaceae; g_Owenweeksia;s_ | Epipelagic | 174 | 3,298,885 | 27,380 | 3007 |
| MAG_M1 Alphaproteobacteria | 64.33 | 10.42 | 53.91 | Low-quality draft | d_Bacteria;p_Proteobacteria; c_Alphaproteobacteria; o_Sphingomonadales; f_Sphingomonadaceae; g_Erythrobacter_A;s_ | Bathypelagic | 818 | 2,345,683 | 3495 | 2814 |
| MAG_M2 Alphaproteobacteria | 48.63 | 8.14 | 40.49 | Low-quality draft | d_Bacteria;p_Proteobacteria; c_Alphaproteobacteria; o_Pelagibacterales; f_Pelagibacteraceae; g_Pelagibacter;s_ | Epipelagic | 487 | 842,001 | 1631 | 1278 |
| MAG_M3 Alphaproteobacteria | 50.25 | 1.14 | 49.11 | Medium-quality draft | d_Bacteria;p_Proteobacteria; c_Alphaproteobacteria; o_Rhodobacterales; f_Rhodobacteraceae; g_Donghicola;s_Donghicola eburneus | Bathypelagic | 1065 | 2,245,852 | 2263 | 3020 |
| MAG_M4 Cyanobacteria | 33.73 | 6.6 | 27.13 | Low-quality draft | d_Bacteria;p_Cyanobacteriota; c_Cyanobacteriia; o_Synechococcales_A; f_Cyanobiaceae; g_Prochlorococcus_A;s_ | Epipelagic | 458 | 709,092 | 1467 | 1092 |
| MAG_M5 Bacteroidia | 57.21 | 0 | 57.21 | Medium-quality draft | d_Bacteria;p_Bacteroidota; c_Bacteroidia; o_Flavobacteriales; f_Flavobacteriaceae; g_Leeuwenhoekella;s_ | Bathypelagic | 1335 | 3,003,880 | 2525 | 3722 |
| MAG_M6 Bacteroidia | 30.17 | 0 | 30.17 | Low-quality draft | d_Bacteria;p_Bacteroidota; c_Bacteroidia; o_Flavobacteriales; f_Flavobacteriaceae; g_Muricauda;s_ | Bathypelagic | 889 | 1,633,349 | 1872 | 2200 |
| MAG_M7 Gammaproteobacteria | 79.37 | 3.09 | 76.28 | Medium-quality draft | d_Bacteria;p_Proteobacteria; c_Gammaproteobacteria; o_Enterobacterales; f_Alteromonadaceae; g_Idiomarina;s_Idiomarina baltica | Bathypelagic | 763 | 2,066,970 | 3279 | 2533 |
| MAG_M8 Bacteroidia | 42.24 | 0 | 42.24 | Low-quality draft | d_Bacteria;p_Bacteroidota; c_Bacteroidia; o_Chitinophagales; f_Chitinophagaceae; g_Asinibacterium; s_Asinibacterium sp1 | Mesopelagic | 915 | 1,764,540 | 2017 | 2287 |

higher values of total phosphorus and total nitrogen among mesopelagic samples.

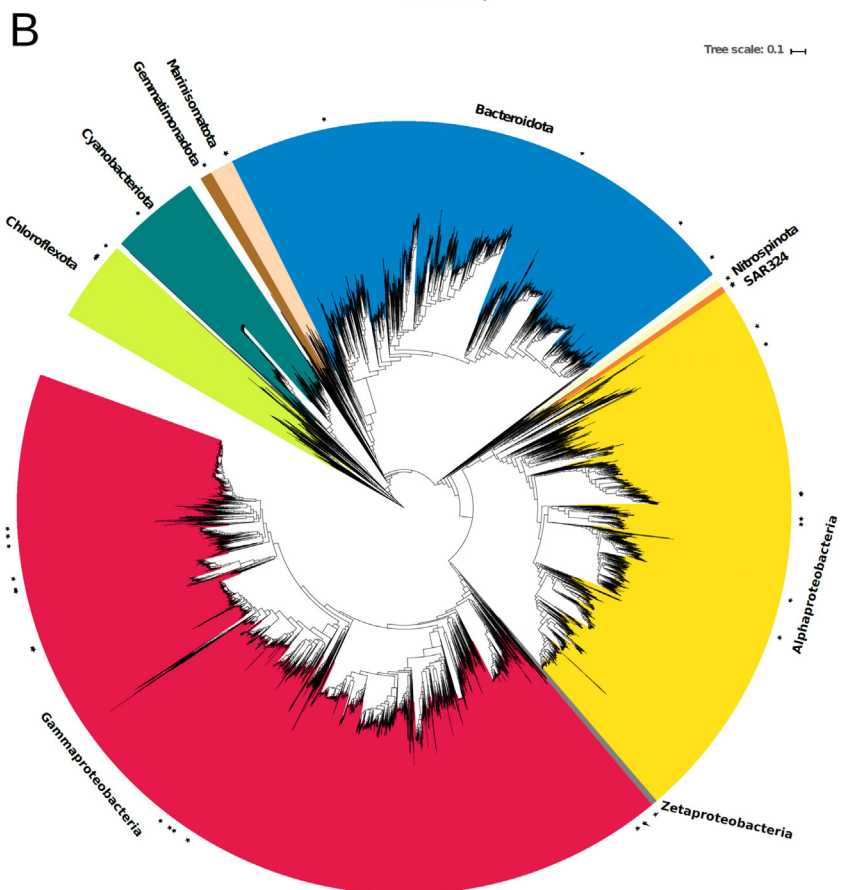
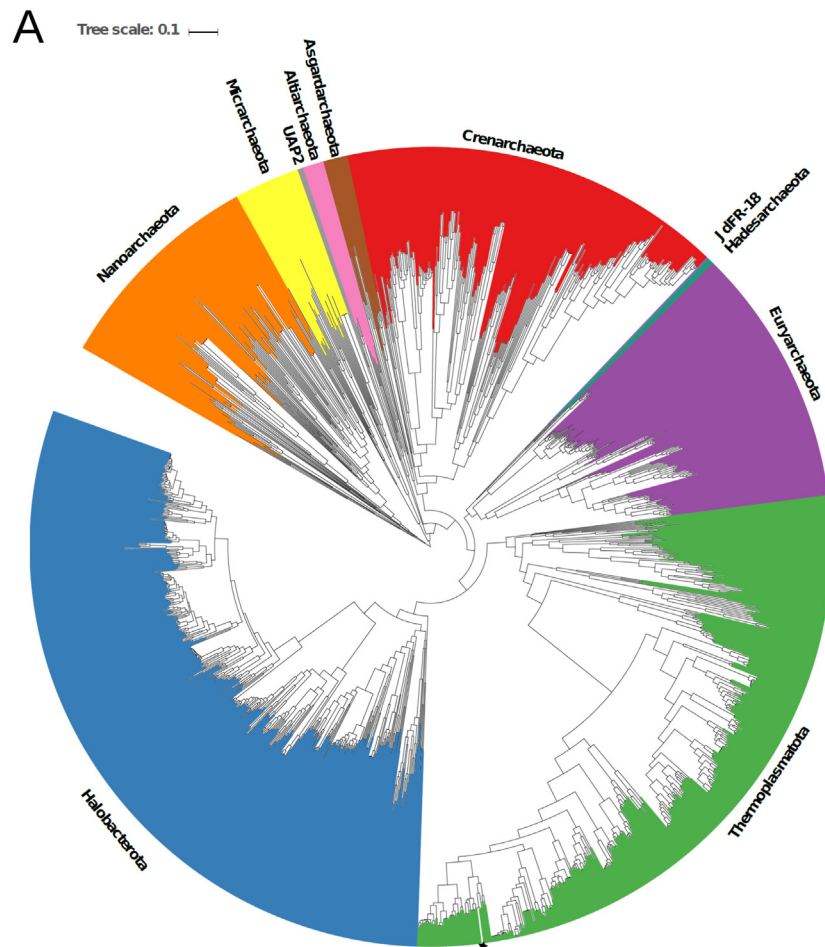
2.4. Potential for carbon fixation and use of alternative energy sources among the SAO microbiome

Following up on the observation that dark carbon fixation pathways were abundant among mesopelagic and bathypelagic samples (Fig. 4), we investigated the genomic content of the MAGs (Fig. 8). We aimed to pinpoint which organisms were responsible for driving the process of carbon fixation throughout the water column, and which energy sources they used to fuel their metabolisms. We identified genes encoding RubisCo and other enzymes of the Calvin-Benson-Bassham

cycle of carbon fixation in two MAGs: M3 (*Donghicola*, which showed highest abundances in the bathypelagic) and M4 (*Prochlorococcus*, which showed highest abundances in the epipelagic). As expected, MAG M4 encoded Photosystem II genes involved in photosynthetic light reactions, highlighting the role of these organisms in light driven carbon fixation at the epipelagic zone. A manual inspection of the RubisCo isoform encoded by MAG M3 revealed it to be a member of the type IV family, which does not have carboxylation/oxygenation activity. Hence, this particular MAG is not likely to be driving dark carbon fixation in the bathypelagic zone of the SAO.

Analysis of single cell genomes of SAR324 bacteria revealed that these organisms perform carbon fixation fueled by sulfur oxidation (Swan et al., 2011). We did find *sox* genes for sulfur oxidation among

Fig. 5. Phylogenomic reconstruction displaying the placement of MAGs in the Archaeal (A) and Bacterial (B) phylogenomic trees. Black stars depict the placement of MAGs within each tree. Branches are color coded according to taxonomic classification of genomes. Bacterial phyla to which no MAGs were assigned were collapsed to better display tree topology.



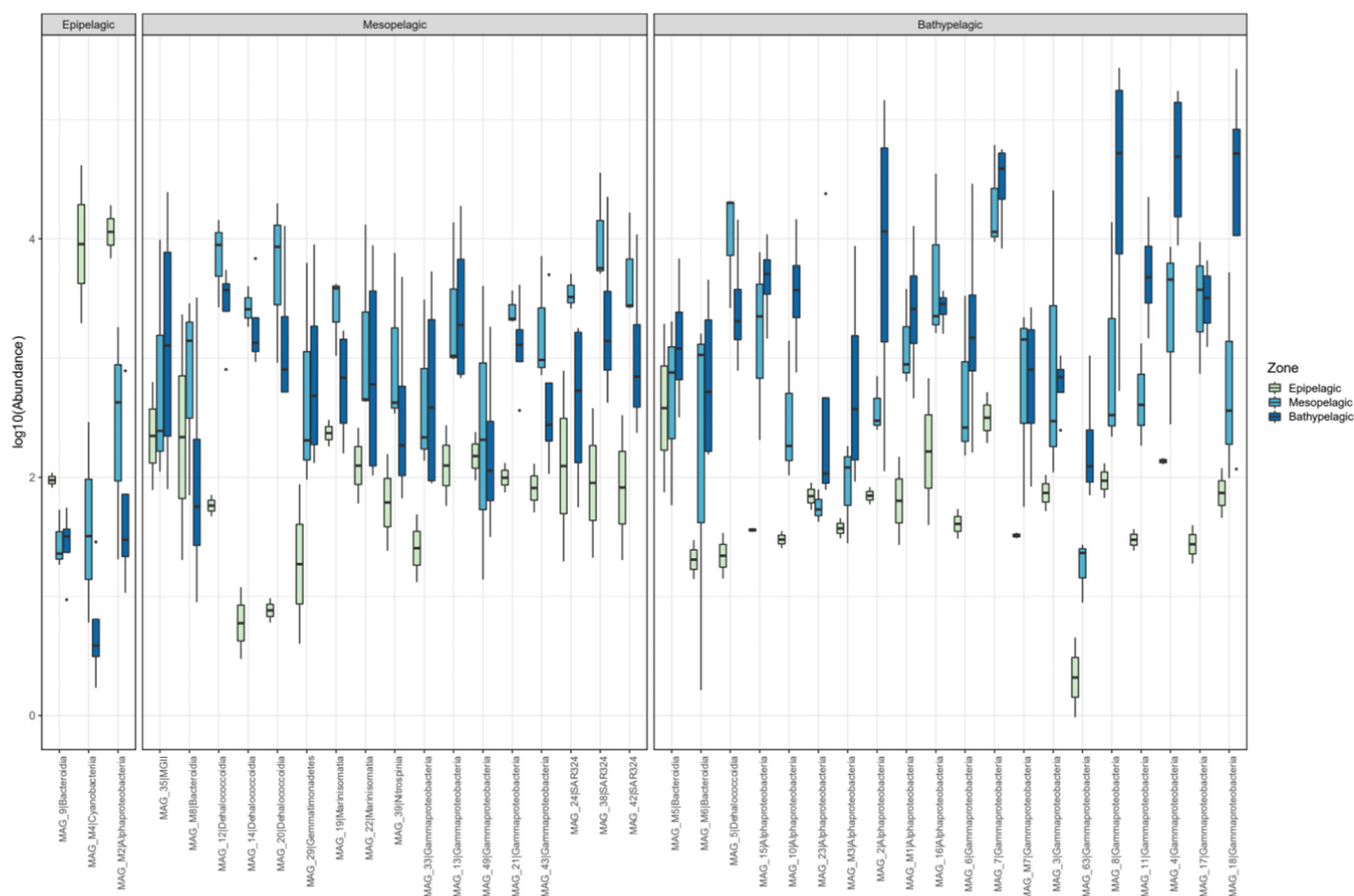


Fig. 6. Relative abundances of MAGs across samples. Box plots depict the relative abundances of each MAG in metagenomes grouped by depth zone. MAGs were grouped according to their preferred depth zone (panels). Preferred depth zone was determined as the one with the highest value of median relative abundance. Boxes depict the median, the first and third quartiles. Whiskers extend to 1.5 of the interquartile ranges. Outliers are represented as dots above or below whiskers.

MAGs 24 (SAR324) and 42 (SAR324) both from the mesopelagic. Nevertheless, we did not find carbon fixation genes among the SAR324 MAGs, which could be a consequence of the fact that these are not complete genomes. Assuming this is the case, SAR324 could be assigned as mixotrophs driving carbon fixation through sulfur oxidation in the mesopelagic. The *sox* genes were also found in MAGs 2 (*Roseivivax*, bathypelagic), 15 (*Henriciella*, bathypelagic), 21 (Gammaproteobacteria, mesopelagic), and M3 (*Donghicola*, bathypelagic). Since there is no evidence that these organisms are capable of carbon fixation, it is likely that they use the energy derived from sulfur oxidation to fuel heterotrophic pathways in the mesopelagic and bathypelagic zones instead.

Finally, genes for nitrite oxidation (key enzyme: nitrite oxidoreductase) and dissimilatory nitrate reduction (nitrate reductase) were detected in MAGs 4, 13, 17 and 43. Genes for dissimilatory nitrate reduction to ammonium (key enzymes: nitrate reductase and cytochrome *c* nitrite reductase) were detected in MAGs 2, 4, 6, 7, 8, 10, 11, 17. MAG4 (*Marinobacter*) was unique in the sense that besides the aforementioned genes it also encoded genes for nitrite reduction (nitrite reductase), nitric oxide reduction (nitric oxide reductase), and nitrous oxide reduction (nitrous oxide reductase). None of these MAGs encoded the genes for dark carbon fixation and all of them displayed a preference for the mesopelagic or the bathypelagic zone. Hence, we concluded that the reduction of nitrogenated compounds was more relevant for the heterotrophic fraction of the mesopelagic and bathypelagic zone microbiomes.

2.5. A diverse apparatus for degrading recalcitrant organic compounds

In the dark ocean, the concentrations of labile organic compounds are low, forcing these communities to make use of alternative

molecules as energy sources (Orcutt et al., 2011; Herndl and Reinthaler, 2013). An extensive repertoire of enzymes involved in the degradation of recalcitrant organic compounds were observed among the MAGs. Specifically glycoside hydrolases that act as cellulases, xylanases and amylases on polysaccharides were prevalent among MAGs. These enzymes are involved in the degradation of high molecular weight polysaccharides. Interestingly, these genes were prevalent among the genomes derived from organisms that showed a preference for the mesopelagic and bathypelagic zones. Likewise, chitinases were also detected, albeit at lower prevalence among MAGs. These were identified in MAGs 3 (*Pseudoalteromonas*), 8 (*Asinibacterium*), 19 and 22 (both Marinisomatota). This suggests that these MAGs from the mesopelagic and bathypelagic zones survive by utilizing the recalcitrant organic matter from the surface. In addition MAGs 4 (*Marinobacter*), 8 (*Alcanivorax*), 11 (*Marinobacter*) and 18 (*Oleibacter*), encoded alkane 1-monoxygenases and haloalkane dehalogenases, which suggested that these organisms have the capacity to degrade alkanes and possibly utilize them as a source of carbon and energy. All these MAGs were derived from Gammaproteobacteria and showed a preference for the bathypelagic zone, suggesting that alkane metabolism is particularly relevant for this taxon and at these depths. Considering that the majority of the MAGs were medium and low quality drafts it is likely that many of their genes could not be retrieved. This suggests that the apparatus for degrading recalcitrant organic compounds of the SAO microbiome might be even more extensive than reported here. For example, bacteria from marine ecosystems, specially deep-waters, have been shown to be capable of degrading plastics (Urbanek et al., 2018; Ru et al., 2020). Some of the taxa known to display this capacity (e.g. *Vibrio*, *Pseudomonas*, *Alcanivorax*) were identified among our samples.

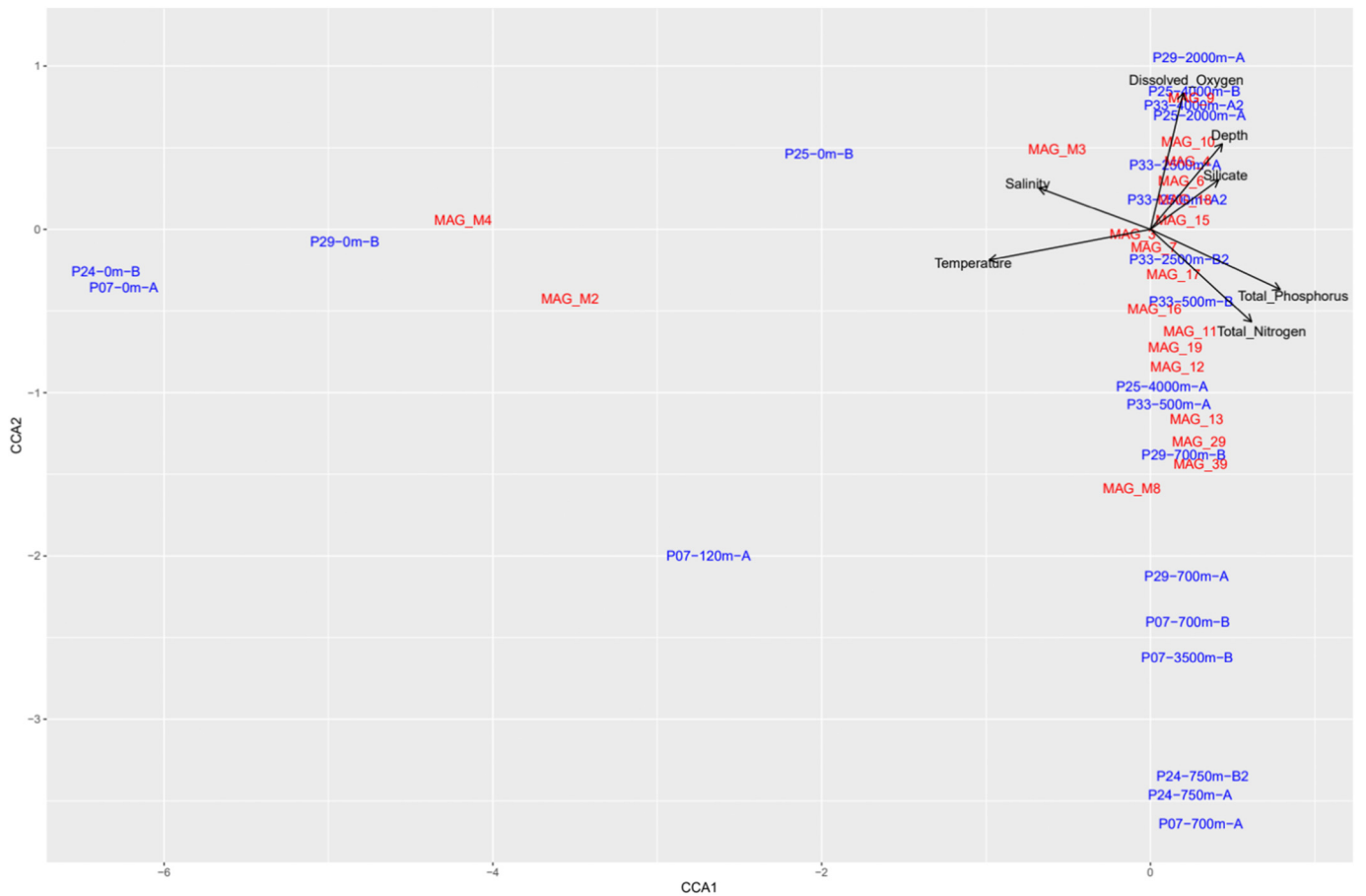


Fig. 7. Canonical correspondence analysis depicting associations among MAG abundances, samples and environmental parameters. Taxon and sample labels were slightly shifted from their original positions to avoid overlap.

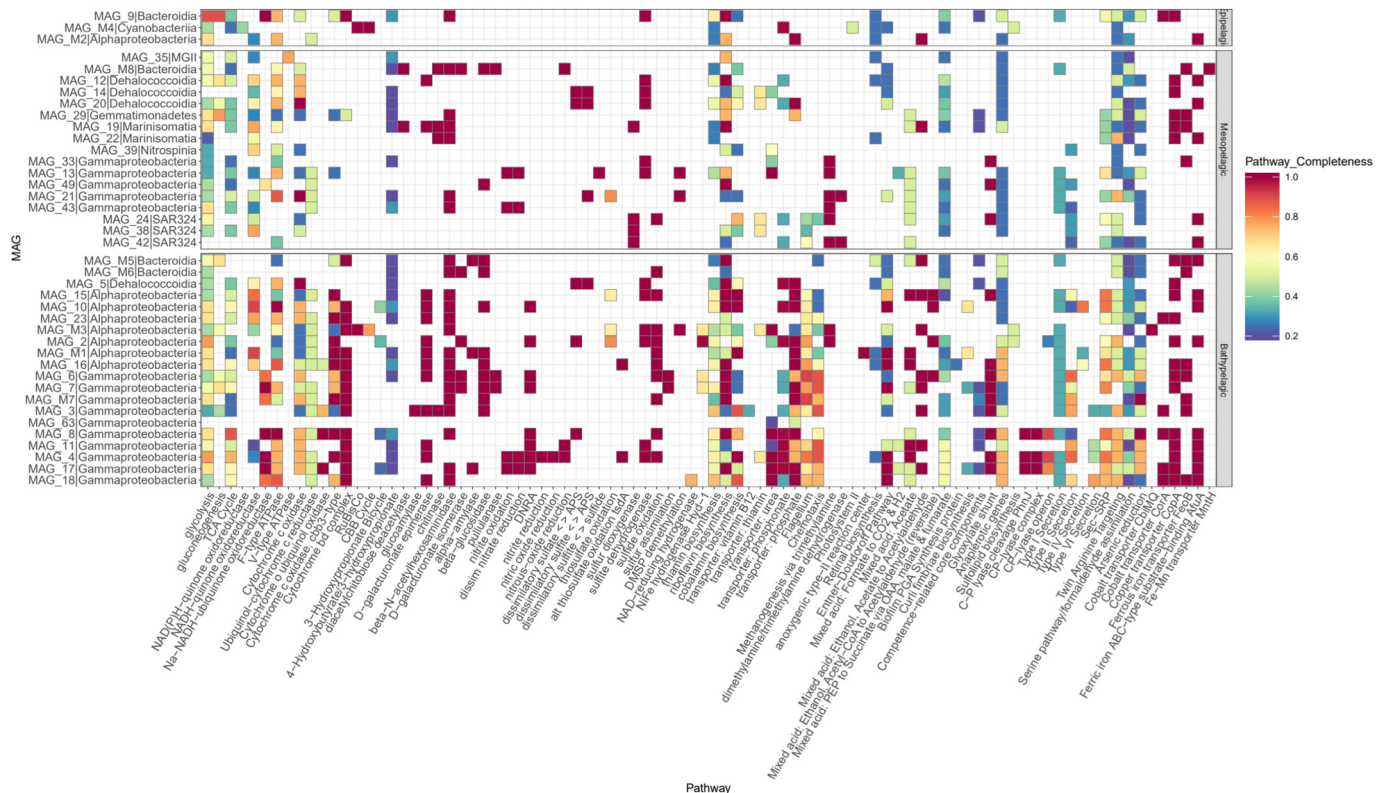


Fig. 8. Heatmap depicting the completeness of key metabolic pathways among SAO MAGs evaluated through KEGG-Decoder. Only values of pathway completeness equal or above 0.2 are shown. A value of 1 indicates a pathway is fully complete. MAGs were grouped according to their preferred depth zone (panels) as defined by the values in Fig. 6.

Hence, plastic degradation might represent an underexplored potential of the SAO microbiome.

2.6. Community composition and ecology of viruses in the deep Atlantic Ocean

Viruses play fundamental roles in the functioning of marine ecosystems. Lytic viral infections lead to the release of nutrients into the water column as well as particulate material, making them readily available to the microbiome through the viral shunt (Breitbart et al., 2018). In addition, viruses also impact the ecosystem by expressing auxiliary metabolic genes that redirect host metabolism, and by acting as agents of horizontal gene transfer by phage mediated genetic transduction (Suttle, 2005; Breitbart, 2012).

We sought to examine viral community composition throughout the depth gradient. We manually inspected the assembled scaffolds to identify sequences of viral origin (by the presence of hallmark viral genes, e.g. terminases, capsid proteins and lysins) (López-Pérez et al., 2017), which identified a total of 188 viral scaffolds. Taxonomic classification indicated that all classified viral sequences were derived from tailed bacteriophages from the order Caudovirales. Among them, the most common viral families were Myoviridae (60 scaffolds), Siphoviridae (35 scaffolds), and Podoviridae (9 scaffolds), which is in agreement with the observed viral community composition across depth gradients (Coutinho et al., 2019).

We also performed automatic host prediction using the generated MAGs as a reference database. This analysis revealed 36 virus-host links covering 19 viral scaffolds (Table S2). The majority of these links associated the viruses with MAGs assigned a *Proteobacteria* (both Alpha and Gamma). Nevertheless, viruses predicted to infect *Chloroflexota*, *Nitrospinota* and SAR324 were also identified. Although restricted to a small set of viruses, this result indicates that the most targeted hosts by the viruses in our dataset are the most abundant bacteria in our samples.

We inspected the gene content of the viral sequences in search of auxiliary metabolic genes (AMGs) that might shed further light into the molecular level interactions between viruses and hosts at the SAO. Among the genes identified as AMGs the most common ones were annotated as DNA (cytosine-5)-methyltransferase 1. This gene encodes a DNA methylase that is likely used by the viruses to methylate their own DNA as a way of escaping the host restriction modification system. Among AMGs we also identified *psbA* (part of photosystem II) which is typically found in the genomes of Cyanophages. We also identified *purA* which is involved in nucleotide metabolism, and commonly observed among phage genomes. Most of the remaining AMGs were linked to either carbohydrate or amino acid metabolism, suggesting that modulating such pathways to enhance viral biosynthesis might be a common strategy applied viruses of the SAO.

The scaffolds identified as viral along with previously published viral sequences from epi-, meso- and bathypelagic zones (that were also described using metagenomics) (Mizuno et al., 2016; Luo et al., 2017) were used as a reference database to annotate reads and estimate relative abundances of viral genomes. The viral communities also displayed a shift in composition throughout the water column (Fig. 9), similar to the trends observed for their microbial counterparts. Among the epipelagic metagenomes, viruses of *Cyanobacteria* and *Alphaproteobacteria* specifically those infecting *Pelagibacter* were consistently the most abundant ones. Yet among mesopelagic samples viruses of *Cyanobacteria* displayed negligible abundances. Instead these communities showed high abundances of viruses of *Alphaproteobacteria* and *Planctomycetota*, with some samples displaying high abundances of viruses of *Gammaproteobacteria*. Finally the bathypelagic samples often displayed highest abundances of viruses of *Gammaproteobacteria*, specially among the 4000 m samples in which viruses of *Bacteroidota* displayed highest relative abundances as well. At the bathypelagic zone the viruses of *Alphaproteobacteria* and *Planctomycetota*, were also

detected, with abundances similar to those observed among the mesopelagic zone samples. This shift in viral community composition reflects differences among the organisms targeted by viral infections at different zones (Coutinho et al., 2017). Consequently, the organisms targeted by viruses that represent the sources of organic matter released through the viral shunt also shift throughout the depth gradient, namely from *Cyanobacteria* in the photic zone to *Alphaproteobacteria*, *Gammaproteobacteria* and *Bacteroidota* in the aphotic zone. The mantel test was used to determine if the association between viral and prokaryotic community composition was significant. This association was found to be significant when comparing taxon (phylum level) and viral abundances (Mantel statistic $r = 0.76$, $p = 0.001$) as well as MAG and viral abundances ($r = 0.75$, $p = 0.001$).

Both the concentrations of bacterial cells and viral particles decreased with depth (Fig. 1C and Table S1). The increase in depth was also associated with an increase in the virus-to-microbe ratio (Table S1 and Fig. S1B), which was lowest among epipelagic samples (mean 3.81), intermediate among mesopelagic samples (5.15), and highest among bathypelagic samples (8.7). We hypothesize that these shifts in VMR were linked the observed differences in viral community composition across each depth zone. Those viruses that dominated the community in the bathypelagic zone might produce more viral particles per infection than those at the mesopelagic and the epipelagic zones, which would explain the observed differences in virus-to-microbe ratios among zones. In addition, the differences in virus-to-microbe ratios might be related to differences in the prevalence of lytic or lysogenic infections at each depth zone, and the switch between these two strategies, which is dependent on microbial cell densities (Knowles et al., 2016; Erez et al., 2017).

2.7. Ecological processes throughout the depth gradient: a conceptual model

We coupled the information derived from the taxonomic and functional composition of the metagenomes and the genomic content of the MAGs. This allowed us to determine the taxon composition across the depth profile and to identify some of the taxa that drove processes of ecological relevance. Based on these findings we propose a conceptual model of the ecology of the South Atlantic Ocean (Fig. 10). In the photic zone, photoautotrophs (i.e. *Cyanobacteria*) and photoheterotrophs (e.g. *Pelagibacter* and *Puniceispirillum*) are dominant members of the community (Fig. S2). Both these groups rely on sunlight as a source of energy. In this habitat, *Cyanobacteria* perform carbon fixation through the Calvin-Benson-Bassham cycle, fueled by light-derived energy. This process is the main source of organic matter that feeds into the dissolved organic carbon (DOC) pool, which is eventually transferred to the aphotic zone. At the epipelagic zone, viruses target the photoautotrophs (e.g. *Cyanobacteria*), chemo- (e.g. *Gammaproteobacteria*) and photoheterotrophs (e.g. *Pelagibacter*) that dwell at this zone. These infections culminate with the lysis of microbial cells and release of nutrients from their cytoplasm into the water column which also feeds into the DOC pool, eventually reaching the deep ocean (Hurwitz et al., 2015; Breitbart et al., 2018). At this zone the number of lytic viral infections taking place, and the density of microbial cells lead to the lowest virus-to-microbe ratio observed throughout the water column.

In the aphotic zone, dark carbon fixation can take place, albeit driven by organisms that do not rely on light as their primary energy source (Swan et al., 2011). This takes place through the CBB pathway, and also possibly with contributions of the 3-hydroxypropanoate cycle and the reverse TCA cycle. Yet in the absence of light these processes are driven by mixotrophic and chemoautotrophic bacteria (e.g. SAR324) that rely on oxidation of sulfur and nitrogenated compounds as energy sources. These alternative carbon fixation pathways contribute to the DOC pool as well. In this habitat viral lysis likely plays an important role by releasing labile nutrients into the water column following infections of the chemoautotrophs, mixotrophs and

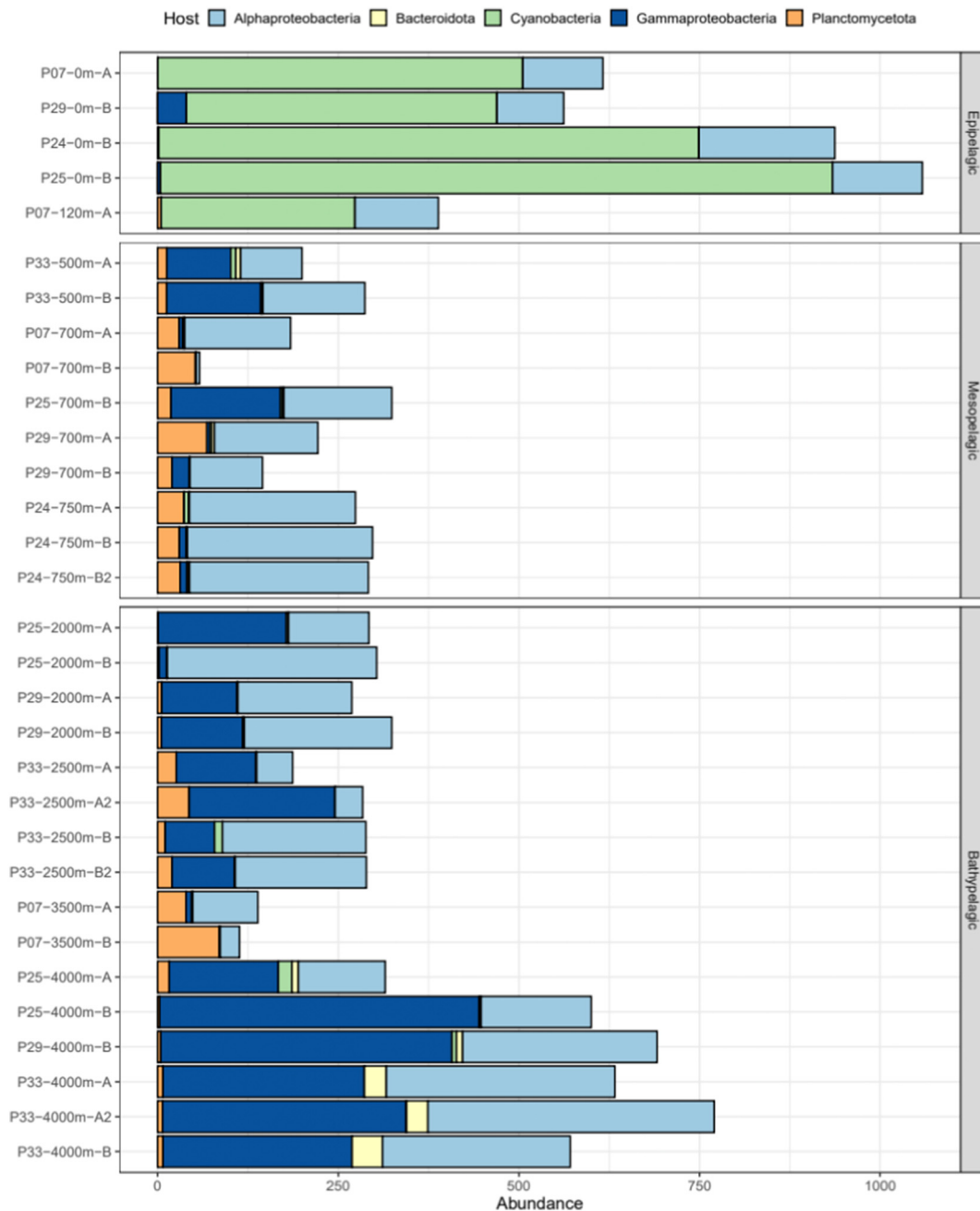


Fig. 9. South Atlantic Ocean viral community composition profiles. Bar plots depict the relative abundances of viruses in the metagenomes expressed as RPKG and summed up according to their putative hosts.

chemoheterotrophs (e.g. *Bacteroidota* and *Gammaproteobacteria*) that dwell in the aphotic zone, which also feeds into the DOC pool. At this zone the number of lytic viral infections taking place, and the density of microbial cells lead to the highest virus-to-microbe ratio observed throughout the water column.

Another survival strategy of microbes from the aphotic zone is the degradation of recalcitrant organic matter, such as cellulose (mediated by *Alteromonas*, *Hyphomonas* and *Muricauda*, among others) and chitin (mediated by *Marinisomatia* and *Pseudoalteromonas*). Likewise, the oxidation of alkanes is mediated by *Alcanivorax*, *Marinobacter* and *Oleibacter*. Thus, several members of the aphotic zone microbiome survive by degrading polysaccharides and alkanes. These processes are of ecological relevance because they convert recalcitrant organic matter into labile forms that can readily be utilized by members of the

community. Thus, these organisms play a fundamental role in the transfer of organic matter and energy to higher levels of the trophic web (Appolinario et al., 2019).

Due to the lower primary productivity in the aphotic zone the chemoheterotrophs that dwell in this habitat rely on alternative strategies for acquiring nutrients and energy. The parasitism practiced by *Micavibrio* (MAG 23) is an example of that. These organisms are obligate epibiotic parasites that survive by attaching to a host bacteria and leeching nutrients from its cytoplasm (Kadouri et al., 2007; Wang et al., 2011). Our genomic analysis does not allow us to determine which hosts are preyed upon by the identified *Micavibrio* MAG. Nevertheless, previous findings have demonstrated that *Micavibrio* can prey on a diverse array of hosts, which includes both chemoheterotrophic and chemoautotrophic bacteria (Dashiff et al., 2011; Dolinšek et al.,

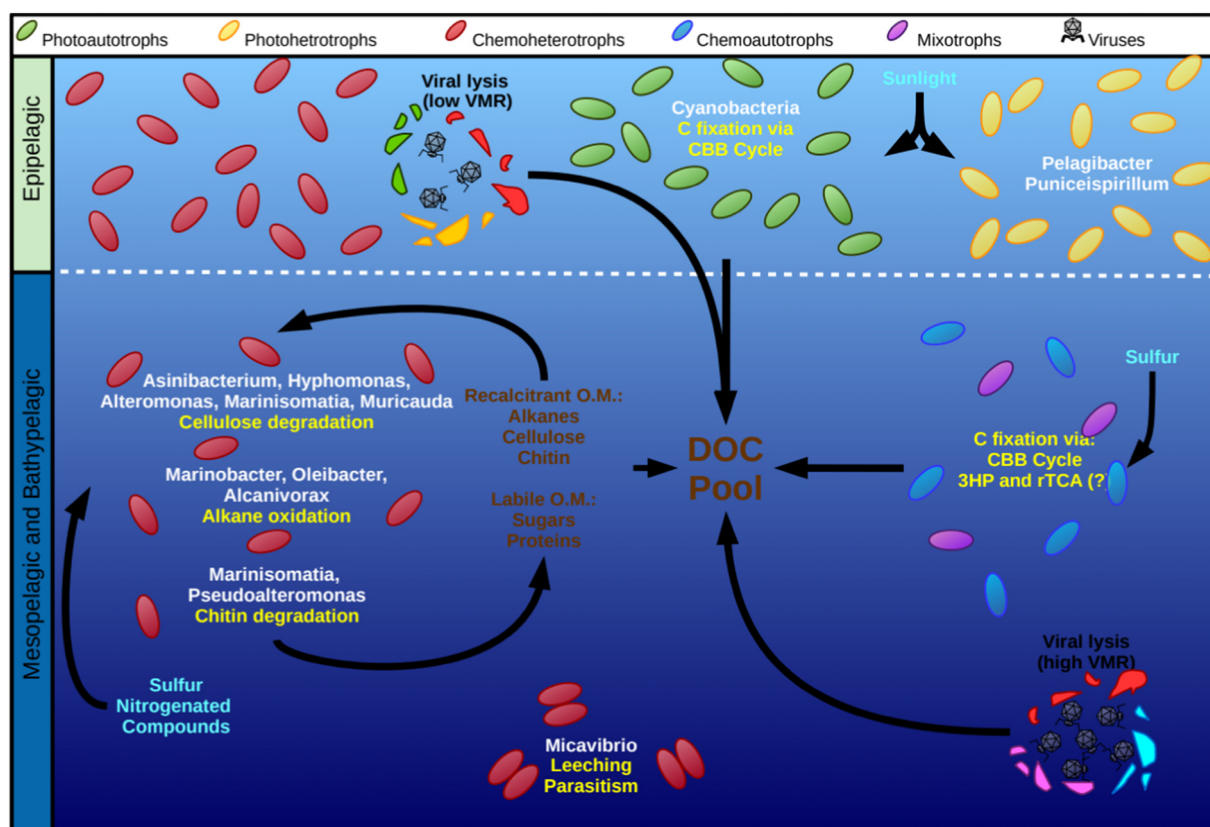


Fig. 10. A conceptual model depicting metabolic and ecological processes taking place among the South Atlantic Ocean microbiome. Taxon names are written in white. Metabolic processes are written in yellow. Energy sources are written in cyan.

2013). Therefore, we suggest that members of the genus *Micavibrio* might prey on autotrophic, heterotrophic and mixotrophic bacteria that dwell in the deep Atlantic Ocean. This would allow these predatory bacteria to indirectly benefit from the capacity of their hosts to utilize alternative energy sources and recalcitrant organic matter. In addition, this would place *Micavibrio* as an important transfer agent of energy and organic matter in the aphotic zone.

In conclusion, the obtained MAGs were derived from members of recently proposed taxa with few genome representatives. Thus, understanding their genomic content helps to elucidate the metabolic potential and the ecological roles of poorly characterized branches of the tree of life. In addition, we obtained insights about important ecological processes taking place at different depths in the South Atlantic Ocean, and revealed some of the microbial players that drive them. Namely, we have provided further evidence that dark carbon fixation is an important process that takes place at this ecosystem, challenging the idea that this is a low activity zone in which low cell densities leads to no primary productivity. Furthermore, our results suggested that processes such as leeching parasitism, alkane oxidation and degradation of recalcitrant organic matter may be among the survival strategies utilized by the microbiome of the deep Atlantic Ocean. Cultivation and isolation of these organisms will be fundamental to further characterize their metabolism to confirm our predictions regarding their metabolic potential.

3. Limitations of the study

Due to the low amounts of DNA obtained from the deep-water samples, our sequences were amplified by MDA to yield sufficient amounts of nucleic acids for sequencing, which is known to reduce diversity within samples (Roux et al., 2019) and likely affected the number and diversity of MAGs obtained. Approaches that do not rely on MDA can

provide a more detailed picture of MAG composition in deep ocean samples (Acinas et al., 2019). Like any other metagenomic study, our metagenomes do not encompass the entirety of the diversity of the analyzed samples. In the future, more advanced sequencing technologies will allow for an even more comprehensive understanding of marine microbial diversity. In addition, some of the diversity is also lost during the assembly and binning, as seen by the fact that some taxa that were abundant in the samples were not represented in the MAGs. Again, this is a limitation of assembly and binning algorithms and their advancement will allow for a better representation of the diversity in marine microbiomes. Also, because not all of the MAGs are nearly complete genomes we cannot make inferences about the metabolisms that they do not have, hence we limited our discussion to the metabolisms that we did find even among these incomplete genomes. Likewise, our findings regarding the novelty of the MAGs and their taxonomic affiliation could also be affected by their completeness. Hence, these results should be reassessed once complete or nearly-complete genomes of these organisms are obtained. Another methodological limitation pertains to the read-based analysis of functional profiles. Reference databases are underrepresented when it comes to genomes derived from organisms of the mesopelagic and bathypelagic. This limitation might have impacted the functional profiles observed for these samples. As more genomes from these zones are discovered and integrated into databases, they will be able to produce more accurate functional profile representations. Finally, our data is metagenomic only, meaning that other types of omics data such as transcriptomics and metabolomics were not analyzed because retrieving and preserving samples for those types of analyses from 4000 m is problematic at best. Nevertheless, including this type of data would have allowed us to go a step further, from revealing what these organisms have the potential to do, to revealing what they are actually doing at each sampling site.

4. Methods

4.1. Sampling sites

Samples were collected at five distinct sampling sites along a latitude gradient between August and October 2011 (Fig. 1): P07 (4°00,49' N/037°59,05' W), P24 (00°00,06' N/035°02,88' W), P25 (07°48,68' S/030°34,63' W), P29 (13°33,10' S/032°31,78' W) and P33 (18°58,13' S/034°47,72' W). Seawater samples were collected at depths ranging from the surface layer to 4000 m deep. We selected the depths for metagenome sampling to cover the major pelagic zones: epipelagic, mesopelagic and bathypelagic. Due to the weather conditions and limited availability of the sampling vessel, not all sites could be sampled at the exact same depths. Hence, we directed our efforts to obtain more samples the bathypelagic because this is a particularly underexplored zone of the SAO.

4.2. Water quality and cytometry analysis

Seawater samples were subjected to water quality analysis through standard methods (Grasshoff et al., 2009). One liter of water from each site was frozen and stored for analysis of inorganic nutrients through these methods: total phosphorous by acid digestion to phosphate; orthophosphate by reaction with ascorbic acid; total nitrogen by digestion with potassium persulfate followed by nitrate determination; ammonia by indophenol; nitrite by diazotization; nitrate by reduction in Cd—Cu column and subsequent diazotization; and silicate by reaction with molybdate. Chlorophyll-a was measured through positive pressure filtration of 2 L of seawater. Filters (glass fiber Whatman GF/F) were kept overnight in a solution of 90% acetone at 4 °C for extraction, and analyzed by spectrophotometry or fluorimetry. Flow cytometry counts were performed using a FACSCalibur flow cytometer (BD Biosciences) equipped with an aircooled argon ion laser emitting at 488 nm (power at 20 mW), fixed laser alignment, and fixed optical components as previously described (Cabral et al., 2020).

4.3. DNA sequencing and quality-control

At each depth from each site, samples of approximately 2.5 L of seawater were filtered through 0.22 µm Sterivex filters aboard the ship. Metagenome samples were collected in duplicates (i.e. two separate rounds of filtration in different Sterivex filters). Filters were kept in liquid nitrogen until processing at the lab. DNA was extracted from Sterivex filters through the phenol-chloroform method. Low DNA yields made some samples inadequate for sequencing. Therefore, all samples were subjected to multiple displacement amplification using the GenomiPhi amplification kit following manufacturer's recommendations, which yielded the necessary DNA concentrations to allow all samples to be sequenced. Next, the amplified DNA was used for Nextera library preparation and sequenced with an Illumina MiSeq platform. Metagenome sequencing yielded an average of 1,492,659 (±588,899) reads and 326,150,356 (±125,185,621) base pairs per sample for a total approximately 92.5 million sequences and 20.2 Tbp of data. Replicated metagenomes were indicated by the "A" and "B" letters in their identifiers. Not all replicates of metagenomes could be sequenced, hence for some samples only a single replicate is available. To improve the quality of the obtained MAGs two of the libraries were sequenced twice, these are indicated by the number "2" appended to the end of their identifiers.

4.4. Metagenomic sequence analysis

Metagenome analysis was performed using both read-centric and gene-centric methods. First, a custom database of proteins derived from representative genomes of GTDB species (Parks et al., 2018) was built to be used as reference for sequence annotation. For the read-

centric approach, taxonomic profiles were obtained by querying quality controlled metagenome reads against the GTDB reference database using DIAMOND (Buchfink et al., 2015) with the following parameters: Maximum e-value of 10^{-5} , minimum bitscore of 50 and minimum identity of 30%. Meanwhile, read-centric functional profiles were obtained by annotating reads through the Humann2 pipeline (Franzosa et al., 2018), which was ran using the EC filtered Uniref90 database with the following parameters: bypassing nucleotide searches, evalue ≤ 0.001 , translated subject coverage threshold 0, translated query coverage threshold 20%, and identity threshold 30%. With these parameters an average of 61.2% (±9.7%) of the reads were mapped for each sample.

For the gene-centric approach a co-assembly of all metagenomes was performed with metaSPADES v3.9.0 (Nurk et al., 2017) using default parameter settings. Protein encoding genes were predicted from the assembled scaffolds using the metagenomic mode of Prodigal (Hyatt et al., 2010). Predicted proteins were subjected to functional annotation by querying them against the Pfam (El-Gebali et al., 2019) database using Hmmer (Finn et al., 2015), and against the NCBI-nr and the SEED databases using DIAMOND (Buchfink et al., 2015) (Identity $\geq 30\%$, bitscore ≥ 30 , e-value $\leq 10^{-3}$). In addition, the protein sequences were queried against the aforementioned custom database derived from GTDB species for taxonomic annotation. At each taxonomic level the scaffolds were assigned a closest relative (CR) which was the taxon with the highest number of matched proteins. Potential ties between taxa were resolved based on the average identity of the hits by selecting the taxon with the highest value.

Next, quality controlled reads from metagenomic samples were queried against the DNA sequences of the protein encoding genes derived from the co-assembly using Bowtie2 (Langmead and Salzberg, 2012) in sensitive-local mode. For each sample, the relative abundance of each gene was estimated in Reads Per Kilobase Million (RPKM). A matrix of taxon abundances was generated by adding up the RPKM values of each gene according to the taxonomic affiliation (CR) of the scaffold from which the gene was derived. Likewise, a matrix of function abundances was generated by adding up the RPKM values of each gene according to the SEED subsystem to which the gene was assigned. All comparative analyses of taxonomic and functional profiles and visualization of the data were performed in R (R Core Team, 2016).

4.5. Metagenome assembled genomes

Assembled scaffolds were binned based on tetranucleotide frequencies and coverage profile with MetaBAT v0.32.4 (Kang et al., 2015) set to the -super-specific configuration (aimed at minimizing bin contamination) and -B set to 20. Preliminary bins generated by MetaBAT that displayed contamination above 5% were manually curated to improve their quality by including or removing scaffolds to bins based on coverage and taxonomic affiliation of annotated proteins. Average Nucleotide Identity (ANI) was computed among the obtained MAGs using fastANI (Jain et al., 2018). No pairs of MAGs displayed more than 95% ANI, hence they were considered non-redundant and all were kept for subsequent steps (i.e. no MAG de-replication step was necessary).

Metagenome assembled genomes (post curation bins) were assigned taxonomic classification using the GTDB-Tk pipeline, which classifies genomes by using marker gene sets to place genomes in the GTDB reference tree (Parks et al., 2018). MAGs were also analyzed with CheckM to assess their levels of completeness and contamination (Parks et al., 2015). A matrix of MAG abundances was generated by adding up the RPKM values of each gene according to the MAG affiliation of the scaffold from which the gene was derived.

4.6. Metabolic potential of metagenome assembled genomes

To infer the metabolic of the metagenome assembled genomes, the protein encoding genes of the MAGs were queried against the KEGG database HMM profiles through KofamKOALA (Aramaki et al., 2019) using

default parameters. The results of this analysis were used as input for KEGG-Decoder (Graham et al., 2018) to estimate the completeness of key metabolic pathways among MAGs. Additionally, the presence of the key enzymes of metabolic pathways among MAGs was manually inspected based on the results of the annotation against the NCBI-nr and Pfam databases.

Phylogenetic reconstructions were performed to assess the validity of RubisCO protein annotations. The RubisCO encoded by MAG M3 was compared against protein sequences spanning the four known isoforms (Tabita et al., 2007). This reference database was constructed with the sequences described in (Tabita et al., 2008) and downloaded from the NCBI protein database. Sequences were aligned with muscle (Edgar, 2004) and a maximum likelihood phylogenetic tree was conducted with MEGA X (Kumar et al., 2018) with the following parameters: Jones-Taylor-Thornton model, gamma distribution with five discrete categories, and 100 bootstraps. Positions with less than 80% site coverage were eliminated.

4.7. Viral sequence analysis

Assembled scaffolds were run through VirSorter (Roux et al., 2015) to identify sequences of phage and prophage origin. This analysis was performed using the default parameters set for metagenomic datasets and only those sequences under the categories “Pretty Sure” and “Quite Sure” were considered putative viruses. In addition, we manually inspected the annotation of protein sequences derived from the assembled scaffolds to identify bona fide viral sequences (i.e. those with a strong viral signature defined by the presence of hallmark viral genes, such as capsid and tail proteins and terminases). Protein sequences derived from viral scaffolds were queried against the NCBI-nr viral database. Viruses were assigned taxonomic affiliation based on closest relative affiliation as described above. Only taxonomic affiliations based on at least three protein hits were considered valid. Viral sequences were also analyzed through VIBRANT with default parameters to identify putative auxiliary metabolic genes (AMGs).

We sought to assign hosts to the obtained viral sequences. To this end, we used a previously described methodology (Coutinho et al., 2019) designed to assign putative hosts to viruses based on shared genetic content between virus and host genomes. The obtained MAGs were used as the database of putative hosts. Searches for homology matches, shared tRNAs, and CRISPR spacers were performed. For the homology matches, viral sequences queried against the databases of MAGs through BLASTn (Altschul et al., 1990). The cut-offs defined for these searches were: minimum alignment length of 300 bp, minimum identity of 50% and maximum e-value of 0.001. tRNAscan-SE (Lowe and Chan, 2016) was used to identify tRNAs in viral scaffolds using the bacterial models. Next, viral tRNAs were queried against the database of MAGs using BLASTn. The cut-offs defined for these searches were: minimum alignment length of 60 bp, minimum identity of 90%, minimum query coverage of 95%, maximum of 10 mismatches and maximum e-value of 0.001. CRISPR spacers were identified among MAGs using CRISPRDetect v2.2 (Biswas et al., 2016). These CRISPR spacers were queried against the viral sequences using BLASTn. The cut-offs defined for these searches were: minimum identity of 80%, minimum query coverage of 100%, maximum of 1 mismatch and maximum e-value of 1.

The identified viral sequences were analyzed together with those derived from studies that recovered viral genomes from the epi-, meso- and bathypelagic zones (Mizuno et al., 2016; López-Pérez et al., 2017; Luo et al., 2017). Quality filtered reads from the metagenomes were queried against the aforementioned viral genomes using BLASTN (Altschul et al., 1990) setting minimum identity to 99% and minimum alignment length to 50 bp. Next, the relative abundances of these genomes were calculated as the number of matched reads per kilobase of genome per gigabase of metagenome (RPKG). For this analysis, we

used the host assignments obtained through manual curation, as those derived from the automatic pipeline are more prone to yielding false-positives (Edwards et al., 2016).

4.8. Statistical analyses

All statistical analyses were performed using the R software (R Core Team, 2016). The virus-to-microbe ratios were compared between samples from the epipelagic, mesopelagic and bathypelagic zones. The significance of the differences in the distribution of these ratios among each group was assessed through the Mann-Whitney test. Associations between environmental parameters, microbial community composition, and viral community composition were assessed through the Mantel test. For the environmental parameters, Euclidean distances were calculated based on a scaled matrix of sample depth, temperature, salinity, total phosphorus, total nitrogen, silicate and dissolved oxygen. For the microbial and viral communities Bray-Curtis distances were calculated based on the relative abundances of taxa/viruses. We applied canonical correspondence analysis (CCA), to determine how the microbial community composition was associated with environmental parameters. This analysis was performed through the Vegan package (Oksanen, 2008).

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.scitotenv.2020.142758>.

CRediT authorship contribution statement

FHC, BED and FLT conceived and designed the experiments. RP and AC performed sampling and sample preparation. FHC, RP and AC performed the experiments. FHC, FABvM, JMW, JHM, MLP, MvVerk, CCT analysed the data. All authors contributed to writing and revising the manuscript.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgements

We are grateful to the Brazilian Navy for providing their time and the crew of RV Antares and to CMG Obino for valuable help aboard during sampling. Cruises were organized by the Brazilian PIRATA project. FHC, FLT, JMW, CCT and RP were supported by grants from CNPq, CAPES and FAPERJ. FHC was supported by a grant from “Ciência sem fronteiras” from CNPq and by APOSTD/2018/186 Post-doctoral fellowship from Generalitat Valenciana. BED and FABvM were supported by the Netherlands Organisation for Scientific Research (NWO) Vidi grant 864.14.004. FHC, MLP and JHM were supported by the grants “VIREVO” CGL2016-76273-P [AEI/FEDER, EU], (co-funded with FEDER funds); Acciones de dinamización “REDES DE EXCELENCIA” CONSOLIDER-CGL2015-71523-REDC from the Spanish Ministerio de Economía, Industria y Competitividad. AC was financed by CAPES - Finance Code 001.

Data availability

All the metagenomes generated and analyzed in this work are publicly available through the European Nucleotide Archive (ENA) under project number PRJEB32934.

References

- Acinas, S.G., et al., 2019. 'Metabolic Architecture of the Deep Ocean Microbiome', bioRxiv. Cold Spring Harbor Laboratory, p. 635680 <https://doi.org/10.1101/635680>.
- Altschul, S.F., et al., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215 (3), 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).

- Alves Junior, N., et al., 2014. Microbial community diversity and physical-chemical features of the Southwestern Atlantic Ocean. *Arch. Microbiol.* <https://doi.org/10.1007/s00203-014-1035-6>.
- Anantharaman, K., et al., 2016. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nature Communications*. vol. 7. Nature Publishing Group, p. 13219. <https://doi.org/10.1038/ncomms13219>.
- Appolinario, L.R., et al., 2019. Metagenomics sheds light on the metabolic repertoire of oil-biodegrading microbes of the South Atlantic Ocean. *Environ. Pollut.* 249, 295–304. <https://doi.org/10.1016/j.envpol.2019.03.007>.
- Aramaki, T., et al., 2019. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *bioRxiv*, 602110 <https://doi.org/10.1101/602110>.
- Baltar, F., Herndl, G.J., 2019. Is dark carbon fixation relevant for oceanic primary production estimates? *Biogeosci. Discuss.*, 1–12 <https://doi.org/10.5194/bg-2019-223>.
- Biswas, A., et al., 2016. CRISPRDetect: a flexible algorithm to define CRISPR arrays. *BMC Genomics* 17 (1), 356. <https://doi.org/10.1186/s12864-016-2627-0>.
- Bowers, R.M., et al., 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* 35 (8), 725–731. <https://doi.org/10.1038/nbt.3893>.
- Breitbart, M., 2012. Marine viruses: truth or dare. *Annu. Rev. Mar. Sci.* 4 (1), 425–448. <https://doi.org/10.1146/annurev-marine-120709-142805>.
- Breitbart, M., et al., 2018. Phage puppet masters of the marine microbial realm. *Nature Microbiology*. vol. 3(7). Springer US, pp. 754–766. <https://doi.org/10.1038/s41564-018-0166-y>.
- Brown, C.T., et al., 2015. Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature*. vol. 523(7559). Nature Publishing Group, pp. 208–211. <https://doi.org/10.1038/nature14486>.
- Buchfink, B., Xie, C., Huson, D.H., 2015. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12 (1), 59–60. <https://doi.org/10.1038/nmeth.3176>.
- Cabral, A.S., et al., 2020. Occurrence and role of viroplankton in a tropical estuarine system. *Hydrobiologia*. Springer International Publishing, 0123456789 <https://doi.org/10.1007/s10750-020-04404-w>.
- Castelle, C.J., Banfield, J.F., 2018. Major new microbial groups expand diversity and alter our understanding of the tree of life. *Cell*. vol. 172(6). Elsevier Inc., pp. 1181–1197. <https://doi.org/10.1016/j.cell.2018.02.016>.
- Coutinho, F.H., et al., 2015. Niche distribution and influence of environmental parameters in marine microbial communities: a systematic review. *PeerJ* 3 (6), e1008. <https://doi.org/10.7717/peerj.1008>.
- Coutinho, F.H., et al., 2017. Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. *Nature Communications*. vol. 8(1). Nature Publishing Group, p. 15955. <https://doi.org/10.1038/ncomms15955>.
- Coutinho, F.H., Rosselli, R., Rodríguez-Valera, F., 2019. Trends of microdiversity reveal depth-dependent evolutionary strategies of viruses in the Mediterranean. In: Zhaxybayeva, O. (Ed.), *mSystems*. vol. 4(6), pp. 1–17. <https://doi.org/10.1128/mSystems.00554-19>.
- Dashiff, A., et al., 2011. Predation of human pathogens by the predatory bacteria *Micavibrio aeruginosavorus* and *Bdellovibrio bacteriovorus*. *J. Appl. Microbiol.* 110 (2), 431–444. <https://doi.org/10.1111/j.1365-2672.2010.04900.x>.
- Dolinsek, J., et al., 2013. Interactions of nitrifying bacteria and heterotrophs: identification of a *Micavibrio*-like putative predator of *Nitrospira* spp. *Appl. Environ. Microbiol.* 79 (6), 2027–2037. <https://doi.org/10.1128/aem.03408-12>.
- Dutta, D., et al., 2005. Microbial Cell Factories Hydrogen Production by Cyanobacteria. vol. 11, pp. 1–11. <https://doi.org/10.1186/1475-2859-4-36>.
- Dyksma, S., et al., 2016. Ubiquitous Gammaproteobacteria dominate dark carbon fixation in coastal sediments. *The ISME Journal*. vol. 10(8). Nature Publishing Group, pp. 1–15. <https://doi.org/10.1038/ISMEJ.2015.257>.
- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32 (5), 1792–1797. <https://doi.org/10.1093/nar/gkh340>.
- Edwards, R.A., et al., 2016. Computational approaches to predict bacteriophage–host relationships. In: Smith, M. (Ed.), *FEMS Microbiology Reviews*. vol. 40(2), pp. 258–272. <https://doi.org/10.1093/femsrev/fuv048>.
- El-Gebali, S., et al., 2019. The Pfam protein families database in 2019. *Nucleic Acids Res.* 47 (D1), D427–D432. <https://doi.org/10.1093/nar/gky995>.
- Emery, W.J., Meincke, J., 1986. Global water masses—summary and review. *Oceanologica Acta*. vol. 9(4). Gauthier-Villars, pp. 383–391.
- Erez, Z., et al., 2017. Communication between viruses guides lysis–lysogeny decisions. *Nature*. vol. 541(7638). Nature Publishing Group, pp. 488–493. <https://doi.org/10.1038/nature21049>.
- Finn, R.D., et al., 2015. HMMER web server: 2015 update. *Nucleic Acids Research*. vol. 43 (W1). Oxford University Press, pp. W30–W38.
- Franzosa, E.A., et al., 2018. Species-level functional profiling of metagenomes and metatranscriptomes. *Nature Methods*. vol. 15(11). Springer US, pp. 962–968. <https://doi.org/10.1038/s41592-018-0176-y>.
- Ghai, R., et al., 2011. Metagenomics of the water column in the pristine upper course of the Amazon River. *PLoS One* 6 (8). <https://doi.org/10.1371/journal.pone.0023785>.
- Graham, E.D., Heideberg, J.F., Tully, B.J., 2018. Potential for primary productivity in a globally-distributed bacterial phototroph. *The ISME Journal*. vol. 12(7). Springer US, pp. 1861–1866. <https://doi.org/10.1038/s41396-018-0091-3>.
- Grasshoff, K., Kremling, K., Ehrhardt, M., 2009. *Methods of Seawater Analysis*. John Wiley & Sons.
- Guerrero-Feijóo, E., et al., 2018. High dark inorganic carbon fixation rates by specific microbial groups in the Atlantic off the Galician coast (NW Iberian margin). *Environ. Microbiol.* 20 (2), 602–611. <https://doi.org/10.1111/1462-2920.13984>.
- Haro-Moreno, J.M., et al., 2017. New insights into marine group III Euryarchaeota, from dark to light. *The ISME Journal*. Nature Publishing Group, pp. 1–16 <https://doi.org/10.1038/ismej.2016.188>.
- Haro-Moreno, J.M., et al., 2018. Fine metagenomic profile of the Mediterranean stratified and mixed water columns revealed by assembly and recruitment. *Microbiome* 6 (1), 128. <https://doi.org/10.1186/s40168-018-0513-5>.
- Herndl, G.J., Reinthaler, T., 2013. Microbial control of the dark end of the biological pump. *Nature Geoscience*. vol. 6(9). Nature Publishing Group, pp. 718–724. <https://doi.org/10.1038/ngeo1921>.
- Herndl, G.J., et al., 2005. Contribution of Archaea to total prokaryotic production in the deep Atlantic Ocean. *Appl. Environ. Microbiol.* 71 (5), 2303–2309. <https://doi.org/10.1128/AEM.71.5.2303-2309.2005>.
- Hügler, M., Sievert, S.M., 2011. Beyond the Calvin cycle: autotrophic carbon fixation in the ocean. *Annu. Rev. Mar. Sci.* 3 (1), 261–289. <https://doi.org/10.1146/annurev-marine-120709-142712>.
- Hurwitz, B.L., Brum, J.R., Sullivan, M.B., 2015. Depth-stratified functional and taxonomic niche specialization in the “core” and “flexible” Pacific Ocean Virome. *ISME J.* 9, 472–484. <https://doi.org/10.1038/ismej.2014.143>.
- Hyatt, D., et al., 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinforma.* 11, 119. <https://doi.org/10.1186/1471-2105-11-119>.
- Jain, C., et al., 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9 (1), 1–8. <https://doi.org/10.1038/s41467-018-07641-9>.
- Kadouri, D., Venzon, N.C., O’Toole, G.A., 2007. Vulnerability of pathogenic biofilms to *Micavibrio aeruginosavorus*. *Appl. Environ. Microbiol.* 73 (2), 605–614. <https://doi.org/10.1128/AEM.01893-06>.
- Kang, D.D., et al., 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 3, e1165. <https://doi.org/10.7717/peerj.1165>.
- Knowles, B., et al., 2016. Lytic to temperate switching of viral communities. *Nature*. vol. 531(7595). Nature Publishing Group, pp. 466–470. <https://doi.org/10.1038/nature17193>.
- Kumar, S., et al., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35 (6), 1547–1549. <https://doi.org/10.1093/molbev/msy096>.
- Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. vol. 9(4). Nature Publishing Group, a division of Macmillan Publishers Limited, pp. 357–359. <https://doi.org/10.1038/nmeth.1923> All Rights Reserved.
- López-Pérez, M., et al., 2017. Genome diversity of marine phages recovered from Mediterranean metagenomes: size matters. *PLoS Genetics*. vol. 13(9). Public Library of Science, e1007018.
- Lowe, T.M., Chan, P.P., 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* 44 (W1), W54–W57. <https://doi.org/10.1093/nar/gkw413>.
- Lozupone, C.A., Knight, R., 2007. Global patterns in bacterial diversity. *Proc. Natl. Acad. Sci. U. S. A.* 104 (27), 11436–11440.
- Luo, E., et al., 2017. Bacteriophage distributions and temporal variability in the ocean’s interior. In: Moran, M.A. (Ed.), *mBio*. vol. 8(6), pp. 1–13. <https://doi.org/10.1128/mBio.01903-17>.
- Martín-Cuadrado, A.-B., et al., 2007. Metagenomics of the deep Mediterranean, a warm bathypelagic habitat. *PLoS One* 2 (9), e914. <https://doi.org/10.1371/journal.pone.0000914>.
- Martín-Cuadrado, A.-B., et al., 2008. Hindsight in the relative abundance, metabolic potential and genome dynamics of uncultivated marine archaea from comparative metagenomic analyses of bathypelagic plankton of different oceanic regions. *ISME J.* 2 (8), 865–886. <https://doi.org/10.1038/ismej.2008.40>.
- Mizuno, C.M., et al., 2016. Genomes of abundant and widespread viruses from the deep ocean. *mBio* 7 (4), e00805–e00816. <https://doi.org/10.1128/mBio.00805-16>.
- Nunoura, T., et al., 2015. Hadal biosphere: insight into the microbial ecosystem in the deepest ocean on Earth. *Proc. Natl. Acad. Sci. U. S. A.* 112 (11), E1230–E1236. <https://doi.org/10.1073/pnas.1421816112>.
- Nurk, S., et al., 2017. MetaSPAdes: a new versatile metagenomic assembler. *Genome Res.* 27 (5), 824–834. <https://doi.org/10.1101/gr.213959.116>.
- Oksanen, J., 2008. *Vegan: an introduction to ordination*. *Management* 1, 1–10 doi: <https://doi.org/10.18607/1260-08-17-12-11-04Z-jarioksa-processed-with-vegan-1.17-6-in-R-version-2.12.1-2010-12-16> on January 10, 2011.
- Orcutt, B.N., et al., 2011. Microbial ecology of the dark ocean above, at, and below the seafloor. *Microbiol Mol Biol Rev* 75 (2), 361–422. <https://doi.org/10.1128/MMBR.00039-10>.
- Pachiadaki, M.G., et al., 2017. Major role of nitrite-oxidizing bacteria in dark ocean carbon fixation. *Science* 358 (6366), 1046–1051. <https://doi.org/10.1126/science.aan8260>.
- Paez-Espino, D., et al., 2016. Uncovering Earth’s virome. *Nature*. vol. 536(7617). Nature Publishing Group, pp. 425–430. <https://doi.org/10.1038/nature19094>.
- Parks, D.H., et al., 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25 (7), 1043–1055. <https://doi.org/10.1101/gr.186072.114>.
- Parks, D.H., et al., 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature Microbiology*. vol. 903. Springer US, pp. 1–10. <https://doi.org/10.1038/s41564-017-0012-7>.
- Parks, D.H., et al., 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology*. vol. 36(10). Nature Publishing Group, pp. 996–1004. <https://doi.org/10.1038/nbt.4229>.
- R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria Available at: <https://www.r-project.org/>.
- Rinke, C., et al., 2019. A phylogenomic and ecological analysis of the globally abundant Marine Group II archaea (Ca. Poseidoniales ord. nov.). *ISME Journal*. vol. 13(3). Springer US, pp. 663–675. <https://doi.org/10.1038/s41396-018-0282-y>.

- Roux, S., et al., 2015. Viral dark matter and virus-host interactions resolved from publicly available microbial genomes. *eLife*. vol. 4. eLife Sciences Publications Limited, p. e08490. <https://doi.org/10.7554/eLife.08490>.
- Roux, S., et al., 2016. Ecogenomics and biogeochemical impacts of uncultivated globally abundant ocean viruses. *Nature*. vol. 537(7622). Nature Publishing Group, pp. 589–693. <https://doi.org/10.1101/053090>.
- Roux, S., et al., 2019. Optimizing De Novo Genome Assembly From PCR-amplified Metagenomes. , pp. 1–18 <https://doi.org/10.7717/peerj.6902>.
- Ru, J., Huo, Y., Yang, Y., 2020. Microbial degradation and valorization of plastic wastes. *Front. Microbiol.* 11 (April), 1–20. <https://doi.org/10.3389/fmicb.2020.00442>.
- Salazar, G., et al., 2016. Global diversity and biogeography of deep-sea pelagic prokaryotes. *ISME J.* 10 (3), 596–608. <https://doi.org/10.1038/ismej.2015.137>.
- Santoro, A.L., et al., 2013. Dark carbon fixation: an important process in lake sediments. *PLoS One* 8 (6), 1–7. <https://doi.org/10.1371/journal.pone.0065813>.
- Sogin, M.L., et al., 2006. Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proc. Natl. Acad. Sci.* 103 (32), 12115–12120. <https://doi.org/10.1073/pnas.0605127103>.
- Soo, R.M., et al., 2017. On the origins of oxygenic photosynthesis and aerobic respiration in Cyanobacteria. *Science* 355 (6332), 1436–1440. <https://doi.org/10.1126/science.aal3794>.
- Sunagawa, S., et al., 2015. Structure and function of the global ocean microbiome. *Science* 348 (6237), 1261359. <https://doi.org/10.1126/science.1261359>.
- Suttle, C.A., 2005. Viruses in the sea. *Nature* 437 (7057), 356–361. <https://doi.org/10.1038/nature04160>.
- Swan, B.K., et al., 2011. Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science* 333 (6047), 1296–1300. <https://doi.org/10.1126/science.1203690>.
- Tabita, F.R., et al., 2007. Function, structure, and evolution of the RubisCO-like proteins and their RubisCO homologs. *Microbiol. Mol. Biol. Rev.* 71 (4), 576–599. <https://doi.org/10.1128/mmb.00015-07>.
- Tabita, F.R., et al., 2008. Phylogenetic and evolutionary relationships of RubisCO and the RubisCO-like proteins and the functional lessons provided by diverse molecular forms. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363 (1504), 2629–2640. <https://doi.org/10.1098/rstb.2008.0023>.
- Tamagnini, P., et al., 2003. Hydrogenases and hydrogen metabolism of cyanobacteria. *Microbiol. Mol. Biol. Rev.* 66 (1), 1–20. <https://doi.org/10.1128/mmb.66.1.1-20.2002>.
- Thompson, L.R., et al., 2017. A communal catalogue reveals Earth’s multiscale microbial diversity. *Nature* 551 (7681), 457–463. <https://doi.org/10.1038/nature24621>.
- Tomczak, M., Godfrey, J.S., 2013. *Regional Oceanography: An Introduction*. Elsevier.
- Urbanek, A.K., Rymowicz, W., Mironczuk, A.M., 2018. Degradation of plastics and plastic-degrading bacteria in cold marine habitats. *Appl. Microbiol. Biotechnol.* 102 (18), 7669–7678. <https://doi.org/10.1007/s00253-018-9195-y>.
- von Meijenfeldt, F.A.B., et al., 2019. Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. *Genome Biol.* 20, 217. <https://doi.org/10.1186/s13059-019-1817-x>.
- Wang, Z., Kadouri, D.E., Wu, M., 2011. Genomic insights into an obligate epibiotic bacterial predator: *Micavibrio aeruginosavorus* ARL-13. *BMC Genomics* 12 (1), 453. <https://doi.org/10.1186/1471-2164-12-453>.
- Williamson, S.J., et al., 2008. The Sorcerer II Global Ocean Sampling Expedition: metagenomic characterization of viruses within aquatic microbial samples. *PLoS One* 3 (1), e1456. <https://doi.org/10.1371/journal.pone.0001456>.
- Xie, W., et al., 2011. Comparative metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with contrasting chemistries. *ISME J.* 5 (3), 414–426. <https://doi.org/10.1038/ismej.2010.144>.
- Zaremba-Niedzwiedzka, K., et al., 2017. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature*. vol. 541(7637). Nature Publishing Group, pp. 353–358. <https://doi.org/10.1038/nature21031>.
- Zeebe, R.E., Wolf-Gladrow, D., 2001. *CO₂ in Seawater: Equilibrium, Kinetics, Isotopes*. Gulf Professional Publishing.
- Zhou, W., et al., 2017. High dark carbon fixation in the tropical South China Sea. *Continental Shelf Research*. vol. 146(August). Elsevier Ltd, pp. 82–88. <https://doi.org/10.1016/j.csr.2017.08.005>.