ABSTRACT

Title of Dissertation: NITROGEN, MICROBES, PARTICLES AND OXYGEN DEFICIENT ZONES

Paulina Huanca Valenzuela (Doctor of Philosophy), **2024**

Dissertation directed by: Assistant professor Clara A. Fuchsman, Horn Point Laboratory, UMCES
Assistant professor Jacob A. Cram, Horn Point Laboratory, UMCES

Single-celled microbes mediate most biogeochemical cycling in the ocean. Ammonium is generally the preferred reduced nitrogen form microbes use for assimilation and growth. However, ammonium is often absent from offshore waters. Microorganisms can metabolize alternative organic reduced nitrogen forms in the absence of ammonium, if they possess genes encoding for the enzymes cyanase (*cynS*), and urease (*ureC*), which catalyze the decomposition of cyanate and urea respectively. It is unknown which microbes contain these genes in the environment.

In my first chapter, I quantified the microbes that can use cyanate and/or urea in oxic and anoxic (ODZ) environments by using a phylogenetic read placement technique. First, I explored depth profiles of metagenomes from two Pacific Ocean regions: an oxic region represented by the nutrient limited Hawaii Ocean Time series, and two ODZ environments represented by the Eastern Tropical South and the North Pacific. A larger proportion of N_2 producing anammox bacteria in ODZs have the ability to utilize cyanate than urea, while a larger proportion of nitrite oxidizing *Nitrospina* have the ability to utilize urea than cyanate. Ammonia-oxidizing Thaumarchaeota have the ability to use urea in deep oxic waters. Contrastingly, the majority of heterotrophic SAR11 bacteria have the ability to use urea in surface waters, but none do in deep waters. This structuring of who can utilize which reduced N form could reflect competition between microbes and N availability.

For my second chapter, I examined microbial ability to use urea and cyanate across time and space using metagenomes from two oceanic Geotraces transects in the North Atlantic; GA02 a North-South spring transect, and GA03, a Fall West to East transect. The two transects differed in nutrient concentrations, affecting the composition of phytoplankton

communities. Though eukaryotic phytoplankton were abundant on the spring GA02 transect, they did not have the ability to use urea or cyanate, probably because ammonia was present. However, the ability to use urea was still common in SAR11. Cyanobacteria *Synechococcus* was abundant on this transect and had the ability to use cyanate. In the nutrient limited fall GA03 transect, the results were similar to oxic waters in chapter 1 except that towards the east, cyanobacteria *Prochlorococcus* gained the ability to utilize cyanate. Both seasonal and spatial changes were observed in the distribution of *ureC* and *cynS* genes in microbial groups in the North Atlantic.

My third chapter focuses on organisms living on suspended particles. Marine particles constitute a niche that provides ample nutrient and carbon sources. Large particles have been postulated to support anaerobic metabolism that cannot occur in the surrounding water. We examined how microbial diversity changes among a range of 7 different particle sizes in a depth profile at the East Pacific Rise, an area of the ocean with a distinct oxygen minimum. By combining a quantitative 16S rRNA amplicon sequencing dataset with size-fractionated organic matter concentrations, we estimated numbers of each microbial taxa per gram of carbon. Results show that microbial composition varies with different particle sizes and depths.

NITROGEN, MICROBES, PARTICLES AND OXYGEN DEFICIENT ZONES

by

Paulina Alejandra Huanca Valenzuela

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of

[Doctor of Philosophy]

[2024]

Advisory Committee:
Clara A. Fuchsman, Chair
Jacob A. Cram, Co-Chair
Sairah Malkin
Feng Chen
Wiebke Mohr
Joshua S. Weitz, Dean's Representative

Dedication

I want to dedicate my thesis to my Parents Estrella and Jaime, and to my sister Andrea. Thank you so much for supporting me and motivating me to achieve all my dreams.
Quiero dedicarle mi tesis a mis padres Estrella y Jaime, y a mi hermana Andrea, gracias por quererme apoyarme y motivarme a cumplir todos mis sueños.

Acknowledgments

I thank the Mid-Shore Chapter of the Izaak Walton League of America for providing funding for 5 ODZ metagenomes, Horn Point Laboratory for providing salary funding, and Matthew Hays for the creation of Geotraces metagenomic databases and for sharing his analysis of *psbD* gene phylotypes and SAR11 ecotypes. Ashley Hollins for technical support with the processing of the EPR samples. Additionally, I thank Ben Tully for collecting the EPR samples, and Jason Sylvan, Chief Scientist of Hot2ColdVents expedition.

I would also like to thank my graduate student colleagues Chelsea Fowler, Sophia Ahn, Catherine Fitzgerlad, Carol Kim and Alan Williams for their support.

The GEOTRACES 2021 Intermediate Data Product (IDP2021) represents an international collaboration and is endorsed by the Scientific Committee on Oceanic Research (SCOR). The many researchers and funding agencies responsible for the collection of data and quality control are thanked for their contributions to the IDP2021. Similarly, I thank the researchers at the Hawaii Ocean Time Series for their data collection and quality control. This publication is partially based upon Hawaii Ocean Time-series observations supported by the U.S. National Science Foundation under Award #1756517.

Table of contents

List of Tables

List of Figures

List of Abbreviations

b.d.l. Below Detection Limit
DCM Deep Chlorophyll Maxima
EPR East Pacific Rise
ETNP Eastern Tropical North Pacific
ETSP Eastern Tropical South Pacific
HGT Horizontal Gene Transfer
HOT Hawaii Ocean Time Series
NBP Non-buoyant plume
ODZ Oxygen Deficient Zone
SCM Secondary Chlorophyll maxima
SNM Secondary Nitrite maxima
ST Station

Abstract

Single-celled microbes mediate most biogeochemical cycling in the ocean. Ammonium is generally the preferred reduced nitrogen form microbes use for assimilation and growth. However, ammonium is often removed to undetectable levels from offshore waters. Microorganisms can metabolize alternative organic reduced nitrogen forms in the absence of ammonium, if they possess genes encoding for the enzymes cyanase (*cynS*), and urease (*ureC*), which catalyze the decomposition of cyanate and urea respectively. It is unknown which microbes contain these genes in the environment.

In my first chapter, I quantified the microbes that can use cyanate and/or urea in oxic and anoxic (ODZ) environments by using a phylogenetic read placement technique. First, I explored depth profiles of metagenomes from two Pacific Ocean regions: an oxic region represented by the nutrient limited Hawaii Ocean Time series, and two ODZ environments represented by the Eastern Tropical South and the North Pacific. A larger proportion of N_2 producing anammox bacteria in ODZs have the ability to utilize cyanate than urea, while a larger proportion of nitrite oxidizing *Nitrospina* have the ability to utilize urea than cyanate. Ammonia-oxidizing Thaumarchaeota had the ability to use urea in deep oxic waters. Contrastingly, the majority of heterotrophic SAR11 bacteria had the ability to use urea in surface waters, but none did in deep waters. This structuring of who can utilize which reduced N form could reflect competition between microbes and N availability.

For my second chapter, I examined microbial ability to use urea and cyanate across time and space using metagenomes from two oceanic Geotraces transects in the North Atlantic; GA02 a North-South spring transect, and GA03, a Fall West to East transect. The two transects differed in nutrient concentrations, affecting the composition of phytoplankton communities. Though eukaryotic phytoplankton were abundant on the spring GA02 transect, they did not have the ability to use urea or cyanate, probably because ammonia was present. However, the ability to use urea was still common in SAR11. Cyanobacteria *Synechococcus* was abundant on this transect and had the ability to use cyanate. In the nutrient limited fall GA03 transect, the results were similar to oxic waters in chapter 1 except that towards the east, cyanobacteria *Prochlorococcus* gained the ability to utilize cyanate. Both seasonal and spatial changes were observed in the distribution of *ureC* and *cynS* genes in microbial groups in the North Atlantic.

My third chapter focuses on organisms living on suspended particles. Marine particles constitute a niche that provides ample nutrient and carbon sources. Large particles have been postulated to support anaerobic metabolism that cannot occur in the surrounding water. We examined how microbial diversity changes among a range of 7 different particle sizes in a depth profile at the East Pacific Rise, an area of the ocean with a distinct oxygen minimum. By combining a quantitative 16S rRNA amplicon sequencing dataset with size fractionated organic matter concentrations, we estimated numbers of each microbial taxa per gram of carbon. Results show differences in microbial composition at different particle sizes and depths.

Chapter 0: Introduction

Oxygen is dissolved into water as a result of gas exchange with the atmosphere, and as a result of primary production. The majority of oceans worldwide correspond to water masses with a high concentration of oxygen, however, for ~8% of the ocean, oxygen concentrations are <20 μM whereas, ~1% of ocean waters are anoxic (Paulmier & Ruiz-Pino, 2009). The oceanic regions where oxygen is depleted from the water column, have been often described as Oxygen Minimum Zones (OMZs), which are areas defined by having <20 μM oxygen concentration (Paulmier & Ruiz-Pino, 2009). However, this classification is broad and it does not differentiate low oxygen from anoxia. For example, in three marine regions (Eastern Tropical North Pacific, Eastern Tropical South Pacific, and Arabian Sea), oxygen is completely depleted from the water column (Oxygen <10 nM (undetected by a STOX sensor); (Revsbech et al., 2009; Tiano et al., 2014) allowing for the accumulation of nitrite which can be used as an indicator of anoxia (Banse et al., 2017). These areas are called oxygen deficient zones (ODZs). ODZs and OMZs are areas in the ocean where a confluence of reduced ocean ventilation and aerobic respiration deplete oxygen from the water column causing anoxic conditions (Kessler, 2006; Paulmier & Ruiz-Pino, 2009). In ODZs, bacteria and archaea carry out processes that do not occur in oxygenated waters, in particular nitrate is used instead of oxygen in the water column (Lam et al., 2009), leading to N_2 production by denitrification and anammox (Thamdrup, 2012), and sulfate reduction in particles (Raven et al., 2021; Saunders et al., 2019). ODZs regions are growing and expanding vertically (Horak et al., 2016; Stramma et al., 2008), making it important to study these environments to understand how the ocean will likely change in the future.

Most microorganisms prefer reduced nitrogen forms for assimilation because they do not have to spend energy reducing the N before using it in proteins of DNA (Glibert et al., 2016). Ammonium concentrations are extremely low in ODZs, possibly because of efficient scavenging by anammox bacteria (Widner et al., 2018). Ammonium concentrations are also often quite low in oxic waters due to efficient scavenging of ammonia-oxidizing Thaumarchaeota (Martens-Habbena et al., 2009). One way to adapt to this shortage of ammonium is for microbes to have the ability to use other small organic reduced N sources, such as urea and cyanate. Therefore, in this thesis project, I explore the spatial distribution of microbes that utilize or degrade urea and cyanate either for N assimilation or for dissimilatory processes such as anammox (Babbin et al., 2017; Ganesh et al., 2018) and aerobic ammonia oxidation (Tolar et al., 2017), both at an oxic station in the North Pacific (HOT) and 3 ODZ sites with one station in the Eastern Tropical North Pacific (ETNP) and two stations in the Eastern Tropical South Pacific (ETSP) (Chapter 1) and in a series of spatial transects in the North Atlantic (Chapter 2).

Microbial communities are responsible for breaking down sinking particles, by consuming and remineralizing carbon, thus controlling carbon export to the ocean floor. As the larger particles sink in the water column, they connect the ocean vertically, transporting microorganisms (Mestre et al., 2018) and carbon to the deep ocean (Le Moigne, 2019). Oxygen Deficient Zones have reduced flux attenuation compared to the oxic ocean (Keil et al., 2016; Cram et al., 2022; Roullier et al., 2014; Weber and Bianchi, 2020). Weber and Bianchi (2020) list three hypotheses related to why this reduced flux attenuation would occur: 1) a significant slow-down in respiration when the electron donor switched from oxygen to nitrate, 2) the exclusion of zooplankton from ODZ could prevent the disaggregation of large particles, and 3) diffusion limitation of nitrate in large particles causing the slower sulfate reduction process to occur. Cram et al (2022) used Underwater Vision Profiler (UVP) data to enumerate the size and number of >100 μm particles in the offshore ETNP ODZ. Combined with EK60 backscatter data to examine zooplankton, these data indicated that zooplankton migrated into the ODZ and disaggregated particles at depths shallower than 500 m (Cram et al., 2022), implying that slow remineralization, rather than suppressed disaggregation, drove reduced flux attenuation (Cram et al., 2022). This result is surprising since utilization of nitrate is similar energetically to utilization of oxygen. However, it takes time for microbes to colonize particles (Kiorboe et al 2003). It is possible that the microbes on sinking particles in the ODZ are adapted to oxic conditions rather than the anoxic conditions where they find themselves.

Particles in ODZs have been examined in oxygen deficient regions as two groups: free-living and particle-attached (Fuchsman et al., 2017; Ganesh et al., 2014, 2015; Saunders et al., 2019). In these papers, it was found that denitrifiers and sulfate reducers were enriched in >30 μm particles in ODZ regions. Thus, these particles attached microbes were adapted to the ODZ. However, 30 μm particles are likely suspended rather than sinking. Particle size can affect sinking speed and therefore the source of the particle, with larger particles sinking faster. Data from high latitudes indicate that sinking particles are more reflective of surface communities than suspended particles (Tamelander, 2013).

Similarly, the community particles in the 20-200 μm size range were more reflective of surface waters than those < 20 μm (Mestre et al., 2018). Additionally, particle size can directly affect the diffusion of chemical species between the water column and the inside of the particle (Ploug et al., 1997). Therefore, size could be directly related to internal redox state, changing the types of biogeochemical cycling present. For example, allowing sulfate reduction to occur in particles in nitrate-containing waters. Marine particles exist in a continuum of sizes with the composition and diversity of microbes changing with particle size (Mestre et al., 2017).

Here I quantitatively explore the microbial distribution across particles of 7 different sizes across depth within a hypoxic/ODZ site at the East Pacific Rise (EPR), which is at the Southern edge of the ETNP ODZ region (Chapter

3; Fig. 0.1). At the same time, I examined the distribution of organic C and N across the same size fractions. By combining a quantitative 16S rRNA gene amplicon sequencing dataset created using standards with size fractionated organic matter concentrations, we were able to estimate numbers of each microbial taxa per gram of carbon. Similar to data from the Chesapeake Bay (Cram et al., 2024), there are more small particles than large particles in the EPR, meaning that the majority of particle attached microbes are on small particles.

Together, the analyses in these three chapters give us a better understanding of microbial diversity, ecology and biogeochemical cycling in the anoxic ocean, and how this biology differs from that seen in the open oxic ocean.

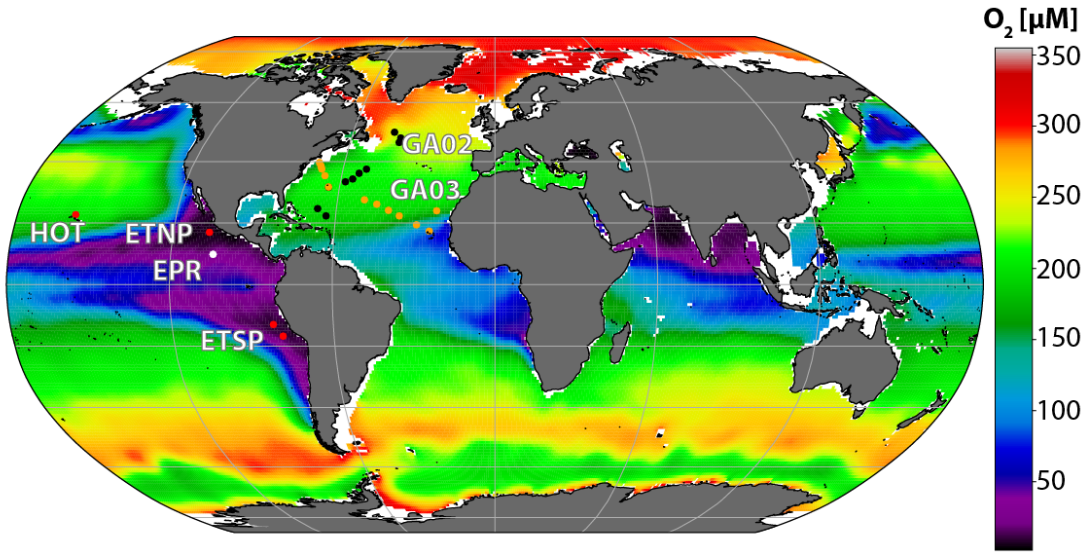


Figure 0.1. Worldwide oxygen concentrations in the ocean at 250 m; data collected from the World Ocean Atlas 2018 (Boyer et al., 2018). Stations: ETNP (Eastern Tropical North Pacific), ETSP (Eastern Tropical South Pacific), and HOT (Hawaii Time Series), are marked in red, and correspond to Chapter 1. Geotraces transect GA03 (orange circles) and transect GA02 (black circles) are in the oxygenated North Atlantic, and correspond to Chapter 2. EPR (East Pacific Rise) station (white circles) is in the ETNP ODZ region, but is not in the anoxic ODZ core, and corresponds to Chapter 3.

Chapter 1: Comparing the microorganisms capable of using alternative reduced nitrogen sources between oxic and anoxic ocean regions

Abstract

Although nitrogen exists in many forms, microorganisms prefer to assimilate reduced nitrogen forms like ammonium (NH_4^+). Therefore, ammonium is generally absent from the water column, including in Oxygen Deficient Zones. In absence of ammonium, some microorganisms may be able to use alternative organic reduced nitrogen forms like urea and cyanate. This capability is based on the presence of genes encoding for the enzymes cyanase (*cynS*), and urease (*ureC*), which catalyze the decomposition of cyanate and urea respectively. This research aims to define which microorganisms, and which proportion of those microorganisms, are capable of using cyanate and/or urea across finely resolved depth profiles, in oxic and anoxic environments. Phylogenetic read placement of metagenomic reads was used to calculate abundance and the proportion of each microbial taxon that contains the urease and cyanase genes. This approach was applied to previously sequenced metagenomes of depth profiles from two Pacific Ocean regions: an oxic region represented by the three timepoints from the Hawaii Ocean Time series-Station; (HOT), and the ODZ environment represented by two stations in the Eastern Tropical South Pacific and one station in the Eastern Tropical North Pacific (ETSP and ETNP respectively). These analyses showed niche partitioning by depth and oxygen and between use of either urea or cyanate. In the oxic euphotic zone, Picocyanobacteria were the main group able to use cyanate. In ODZs, anammox bacteria (Scalindua) showed the highest abundance of *cynS* compared to the other microbial groups, and a higher percentage of anammox bacteria (71-171%) contained *cynS* compared to *ureC* (40-92%). In contrast, all Nitrospina (nitrite oxidizers) contained the urease gene but only ~35% had *cynS* genes in the lower euphotic zone and fewer contained *cynS* in other areas. In deep oxic waters, Thaumarchaeota (ammonia oxidizers) were abundant and had the ability to use urea. Contrastingly, ~40% of SAR11 (metabolically streamlined heterotrophs) genomes contained the *ureC* gene in oxic surface waters, but none at depth. In hypoxic waters, 100% of Verrucomicrobia contained *ureC* compared to 10-20% in anoxic waters while in fully oxic waters 50-100% Verrucomicrobia contained *ureC*. On the other hand, in the ODZ, all S-oxidizing Thioglobaceae contained the *ureC* gene, but few did in oxic waters. Altogether, these profiles likely reflect a combination of N availability and competition between microbes.

Introduction

Oxygen Deficient Zones (ODZs) are regions of the ocean that have <10 nM oxygen (Revsbech et al., 2009; Tiano et al., 2014). They are natural phenomena, but because the oxygen content of the Pacific Ocean has been decreasing since the 1980s (Ito et al., 2017; Stramma et al., 2008). Ocean ODZs host 30-50% of marine fixed-N loss (DeVries et al., 2013). In ODZs, bacteria and archaea utilize nitrate and nitrite as an electron donor instead of oxygen (Lam et al., 2009), leading to N_2 production by heterotrophic denitrification (Org C + $2NO_3^-$ -> N_2 + CO_2) and anammox (NO_2^- + NH_4^+ ->N_2) (Babbin, 2014; Thamdrup, 2012). Most microorganisms prefer reduced nitrogen forms for assimilation because they do not have to spend energy reducing the N before using it in proteins or DNA (Glibert et al., 2016). Ammonium concentrations are extremely low in both ODZs and the oxic oligotrophic ocean, possibly because of efficient scavenging by anammox bacteria and ammonia-oxidizing Thaumarchaeota (Martens-Habbena et al., 2009; Widner et al., 2018, 2018). One way to adapt to this shortage of ammonium is for microbes to have the ability to use small organic reduced N sources, such as urea and cyanate.

Urea

Urea (CH_4N_2O) is a reduced nitrogen compound, which is produced both as an excretion product by crustacean and gastropod zooplankton, but not jellyfish (Pitt et al., 2009; Thibodeau et al., 2020) and as a part of organic matter degradation (Berman et al., 1999; Cho et al., 1996; Price et al., 1985). Microorganisms can metabolize urea by using the enzyme urease, encoded by the gene *ureC.* Urease catalyzes urea hydrolysis, producing one molecule each of ammonia and carbamate (Hausinger, 2004). Urea N uptake experiments generally have uptake maxima in the euphotic zone, where urea averaged 25-30% of total N uptake, but uptake then declined quickly with depth (Painter et al., 2008). In the oxic waters above the Eastern Tropical North Pacific (ETNP) ODZ, uptake experiments using double labeled urea, demonstrated that the N in urea was preferentially assimilated over the carbon, indicating that organisms were using urea as a N source (Widner, Fuchsman, et al., 2018).

Urea concentrations are typically in the nanomolar range. In transects in the North and South Atlantic, euphotic zone urea concentrations were typically between 150 and 350 nM (Painter et al., 2008). Urea in the subtropical North Pacific ranged from 50 to 150 nM in the top 200 m with a maximum at the chlorophyll maximum (Takeda et al., 2020). In the ETNP, urea concentrations as high as 1.5 µM were found at coastal stations, but concentrations were much lower offshore (generally 0-150 nM) (Widner, Fuchsman, et al., 2018). At the station examined further here, measurements showed that urea was generally absent from the water column (Fig. 2), with concentrations below the detection limit (b.d.l.; 70 nM) until below the ODZ, at 958 m where urea concentration reached 0.71 µM, possibly due to migrating

zooplankton (Widner, Fuchsman, et al., 2018). In other marine environments like the Gulf of Mexico, measurement of urea concentration in hypoxic waters reached an average of 69 nM (Widner, Fuchsman, et al., 2018).

Cyanate

Cyanate is a small organic form of reduced nitrogen form (OCN^-) (Dirnhuber & Schütz, 1948) that can be produced abiotically (photoproduction) and via biotic organic matter degradation. Senescent algal cultures have been shown to produce cyanate in culture (Widner et al., 2016). In the marine environment, cyanate can be measured at a nanomolar concentration range using a chromatographic method recently developed by (Widner et al., 2013). Therefore, cyanate concentrations have only been measured in a few places in the ocean. In coastal North Atlantic waters, the concentration of cyanate reaches between 0.4 nM to 11 nM (Widner & Mulholland, 2017). In the Gulf of Mexico, in shelf regions the median cyanate concentration in hypoxic waters was 11.5 nM (Widner, Fuchsman, et al., 2018). In the oxic ocean, cyanate vertical distribution resembles the vertical distribution of ammonium and nitrite with a primary maximum right below the chlorophyll maximum (Widner et al., 2016). These cyanate concentration profiles could be explained by high cyanate uptake and consumption by microbial communities in the upper euphotic zone and lower uptake rates in the deeper euphotic zone (Widner et al., 2016; Widner & Mulholland, 2017).

Cyanate has been measured in and above both the ETNP and ETSP ODZs. In the ETSP cyanate distribution along a transect at 17ºS has been measured where the concentrations ranged below detection limit (b.d.l.; 0.4 nM) to 45 nM above the oxycline (Widner, Mordy, et al., 2018). The cyanate concentrations in the ETSP ODZ were lower, reflected in small peaks of 4 nM or 7 nM. In the ETSP, cyanate uptake was also observed in the euphotic zone, although it accounted for less than 2% of total Nitrogen (N) uptake (Widner, Mordy, et al., 2018) In the ETNP ODZ, cyanate concentrations ranged from b.d.l. to 50 nM in the upper ODZ (Widner, Fuchsman, et al., 2018) (Fig. 1.2). Cyanate uptake was also detected for this region but was estimated to be lower than 0.4 nM/hr (Widner, Fuchsman, et al., 2018).

Cyanase (EC 4.2.1.104) is an enzyme that catalyzes the conversion of cyanate to CO_2 and ammonia $NCO^- + HCO_3^- + 2H^+ \rightarrow 2CO_2 + NH_3$ (W. V. Johnson & Anderson, 1987). Cyanase is encoded by genes in the "cyn" operon, which includes three genes *cynT*, a cyanate permease, *cynS* cyanase, and *cynX* (an unknown protein) all involved in cyanate transport

(P. M. Anderson et al., 1990). The *cynS* gene has spread among microbial lineages via Horizontal Gene transfer (Kamennaya & Post, 2011). In oxic surface waters of the North Atlantic, transcripts of the *cynS* genes have been found in eukaryotic populations, including in Pelagophytes, Dinophytes, Bacillariophyta, and fungi (Mao et al., 2022). In bacteria, cyanase has been found in *Synechococcus*, and *Prochlorococcus* (Kamennaya & Post, 2011; Rocap et al., 2003). In deeper waters cyanase has also been found in single cell genomes and transcripts of nitrite oxidizing bacteria (*Nitrospina*) (Mao et

al., 2022) , and in the anammox bacteria *Scalindua* (Pachiadaki et al., 2017; Widner, Fuchsman, et al., 2018), implying this enzyme could have a significant role in the nitrogen cycle.
Analysis of marine planktonic metatranscriptomes from TARA oceans (global ocean sampling at three depths) has shown that *cynS* transcripts were as prevalent as those for *ureC* (Mao et al., 2022). The authors found *cynS* transcripts from different microbial groups in different ocean regions and in different size fractions. In the large size fraction (>0.8 μm), various eukaryotic algae produced *cynS* transcripts: in the smaller size fraction (<0.8 μm), *Synechococcus cynS* transcript dominated in surface, whereas *Prochlorococcus* and unclassified microbes contribute to the *cynS* transcripts at the DCM. The abundance of *cynS* transcripts negatively correlated with dissolved inorganic N concentrations. Though total *cynS* transcripts were reduced in the mesopelagic region compared to the euphotic zone, *Nitrospina cynS* transcripts were found in the mesopelagic (Mao et al., 2022)

Microorganisms capable of utilizing reduced nitrogen forms urea and cyanate

In the ocean many microbial groups that contain the urease or cyanase gene have been shown either to have key roles in the nitrogen cycle or to be primary producers that need N for assimilation.

Ammonia oxidizing Archaea

Archaeal ammonia oxidizers can grow on cyanate as the only energy source, and nitrogen source, but do not have the cyanase gene (Palatinszky et al., 2015), so we presently do not have a way to trace this ability in the population. Some cultured isolates of Thaumarchaeota can use urea as a sole N source and have urease (Qin et al., 2014). Experimentally, it was shown Thaumarchaeota can utilize nitrogen from urea in the environment, which can be an advantage in environments with low ammonia concentrations (Price et al., 1985) Thaumarchaeota from Arctic waters have the urease (*ureC)* gene with a larger proportion of Thaumarchaeota encoding *ureC* in deep waters compared to the surface (Alonso-Sáez et al., 2012). These Arctic Thaumarchaeota could use urea degradation to fuel dissimilatory process of ammonia oxidation to obtain energy along with carbon assimilation (Shiozaki et al., 2021) The presence of *ureC* is widespread in Thaumarchaeota genomes. It was present in 60-100% of Thaumarchaeota at coastal time series SPOT (Ahlgren et al., 2017). In the ETNP, Thaumarchaeota had one copy of *ureC* per genome (Widner, Fuchsman, et al., 2018). However, in the Gulf of Mexico, only between 10-15% of Thaumarchaeota cells contain an *ureC* gene; therefore, it is possible that only a subpopulation of Thaumarchaeota has *ureC* (Widner, Fuchsman, et al., 2018).

Nitrospina

The marine nitrite-oxidizing bacteria *Nitrospina* are aerobic and obligate chemolithotrophs that oxidize nitrite to nitrate and use CO_2 as the sole C-Source (Spieck et al., 2014). There is one sequenced *Nitrospina* isolate, but single-cell genomes from the deep ocean indicate that Nitrospina is consistently an autotrophic bacteria that uses nitrite oxidation as a main energy source (Pachiadaki et al., 2017). Uncultured Nitrospina-like species have distribution patterns consistent with the nitrite oxidation rate patterns in OMZ regions, both in the ETSP and the ETNP (Fuchsman et al., 2017; Sun et al., 2019). Reads mapping to Nitrospina-like species Metagenome Assembled Genomes (MAG) show that the MAGs from the ODZs are distinct from Nitrospina found in oxygenated waters (Sun et al., 2019). Some *Nitrospina* single-cell genomes contained genes for urea and cyanate degradation (Pachiadaki et al., 2017). In the ETNP ODZ, *Nitrospina* had one copy per genome of urease at the top of the ODZ but only ~50% of Nitrospina had cyanase (*cynS*) (Widner, Fuchsman, et al., 2018).

SAR11

SAR11 are an abundant slow growing group of free-living marine heterotrophic Alphaproteobacteria with small cells and streamlined genomes (Giovannoni et al., 2005; Morris, Rappe, et al., 2002; Morris, Rappé, et al., 2002). Members of the SAR11 clade comprise up to 60% of the microbial population in the ETNP ODZ water column (Fuchsman et al., 2017). Though SAR11 in oxic waters utilize oxygen, SAR11 bacteria in ODZs encode genes for nitrate reductase of the *narG* variety (Tsementzi et al., 2016), responsible for nitrate reduction to nitrite, and copies of the gene nirK which encodes for nitrite reductase, reducing nitrite to NO (Fuchsman et al., 2017). Different SAR11 ecotypes live in ODZs compared to oxic waters (Tsementzi et al., 2016). Stable isotope probing has indicated that some SAR11 can utilize urea in Arctic surface waters (Connelly et al., 2014). In the ETNP ODZ, ~10% of SAR11 contained the urease gene (*ureC*) (Widner, Fuchsman, et al., 2018). Isolates of estuarine ecotypes of SAR11 have also been found to use urea (Lanclos et al., 2023). The fact that SAR11 can utilize urea is particularly interesting as most of the microbes that use urea and cyanate as N sources are autotrophic (Widner, Fuchsman, et al., 2018).

Cand. Scalindua (Anammox)

Cand. Scalindua are free-living bacteria (Fuchsman et al., 2012, 2017) that undergo the anammox process ($NO_2 + NH_4^+ \rightarrow N_2$) in the marine environment. Single cell genomes of *Cand*. Scalindua from the ETNP water column have genes for urea transport, *ureC*, and *cynS* which means these anammox bacteria may have the ability to use cyanate and urea in place of ammonium (Ganesh et al., 2018). Scalindua possessed one copy of the cyanase gene per genome in the ETNP, and were the only bacteria able to use cyanate below the top of the ODZ (Widner, Fuchsman, et al., 2018). When doubly labeled cyanate was used by anammox bacteria in the ODZ,

the C in cyanate was assimilated and the N was not, indicating dissimilatory use of the N (Widner, Fuchsman, et al., 2018). Experimental data showed that anammox N_2 production rates could be supported by the N in cyanate in the ODZ (Babbin et al., 2017). Urea could only stimulate anammox N_2 production rates after a 1.5-day lag time, potentially indicating that urea degradation products were used rather than urea itself (Babbin et al., 2017).

Picocyanobacteria

Cyanobacteria are an important part of the phytoplankton community in aquatic environments. In oligotrophic regions, the Picocyanobacteria *Prochlorococcus* and *Synechococcus* contribute greatly to primary production (Rii et al., 2016). The euphotic zone has strong gradients of light and nutrients. *Prochlorococcus* ecotypes and abundances change dramatically in depth profiles with highlight ecotypes in the surface with Low Light I in the middle of the euphotic zone and Low Light II and Low Light IV (LLIV) at the bottom (Ahlgren et al., 2006; Z. I. Johnson, Zinser, Coe, McNulty, et al., 2006; Zinser et al., 2007). Specific ecotypes of *Prochlorococcus* live at the top of ODZs when >1% blue light overlaps with anoxic water Cepeda-Morales et al., 2009; Ulloa et al., 2021). As primary producers, *Prochlorococcus* and *Synechococcus* need to obtain N from inorganic sources, but can also use small organic compounds cyanate and urea. (Kamennaya & Post, 2011; Rocap et al., 2003). The addition of urea to environmental samples containing *Prochlorococcus* led to an increase in the abundance of *Prochlorococcus* (Shilova et al., 2017). In cyanobacteria, the regulation of *cynS* could help with the utilization of externally acquired cyanate (Kamennaya & Post, 2011).

In this paper, we aim to explore known and new microbial taxonomic groups that can utilize urea and cyanate using the phylogenetic read placement technique. We will look into the spatial distribution of microbes that utilize urea and cyanate, across gradients in depth and oxygen, either for nitrogen assimilation or for dissimilatory processes in 3 ODZ sites (ETNP St 136 and ETSP St 9 and 17) and 3 oxic time-points in the North Pacific from the Hawaii Ocean Time Series (HOT). For each taxon, we calculate the proportion of that taxa which contain *ureC* or *cynS* by dividing by the bacteria/archaea single copy core gene RNA polymerase (*rpoB*) for that taxa. The study of the proportions and abundance of cyanase *cynS* and urease *ureC* has been previously reported by (Widner, Fuchsman, et al., 2018)in and above the ETNP ODZ. The authors found that anammox bacteria were the main bacteria with cyanase genes in the ETNP ODZ, and identified SAR11 ureases for the first time. However, since 2017, more sequences of *ureC* previously classified as “unknown” have been assigned to new taxonomic groups such as Anammox and S-oxidizing Thioglobus (Ganesh et al., 2018; Marshall & Morris, 2013; Shah et al., 2017). The discovery of urease in these new taxonomic groups, require a revision of the previous *ureC* phylogenetic tree that reflects the state of known groups possessing *ureC.*

We re-analyze the (Widner, Fuchsman, et al., 2018) data here because the phylogenetic tree for urease was significantly improved in this

current work to include urease genes for Scalindua (anammox) single cell genomes (Ganesh et al., 2018), S-oxidizing Thioglobus isolates (Marshall & Morris, 2015; Shah et al., 2017), and urease genes for Verrucomicrobia and various Alphaproteobacteria from published MAGs (Sun et al., 2019; Zhang et al., 2023). This reanalysis allows results from the ETNP and ETSP ODZs to be compared to each other and to 3 timepoints at HOT, increasing the number of microbes with known important biogeochemical functions able to be examined. Additionally, this is the first-time urease and cyanase have been examined in this fashion in the oxic ocean. This comparison allows us to understand which microbes are specifically using urea and cyanate in ODZs and which are using them in the oxic and anoxic ocean.

Methods

The metagenomic reads from the genes *ureC* and *cynS* were extracted from the metagenomic databases of interest (ETSP, ETNP and the oxic-time points from HOT), and subsequently placed and aligned by using the Phylogenetic placement method into phylogenetic tree containing aligned amino acid sequences of urease or cyanase enzymes of known taxonomic groups. Once the reads were placed in the tree, we classified the reads according to the taxonomy of the sequence of known taxonomic groups. We then estimated the ratio of *ureC* and/or *cynS* to the housekeeping *rpoB* gene. The ratio *ureC/rpoB* and *cynS/rpoB* were expressed as % of the total prokaryotic community (Archaea + Bacteria) or as the total of specific microbial taxa across depth profiles at each station.

Compilation of Metagenomic data

Metagenomes from microbial communities were obtained from publicly available datasets. Detailed depth profiles of cellular metagenomes from St ALOHA were obtained as part of the Hawaii Ocean Time-series (HOT) (22º45'N and 158ºW) for May (HOT 272), August (HOT 275), and November 2015 (HOT 278) and were downloaded from Bioproject PRJNA352737 (Luo et al., 2020). Nutrient and CTD measurements for these cruises can be downloaded with the Hawaii Ocean Time Series Data Organization and Graphical System (HOT-DOGS) application; University of Hawai'i at Mānoa (https://hahana.soest.hawaii.edu/hot/hot-dogs/), and in the supplement to the original paper (Luo et al., 2020). For Eastern Tropical South Pacific (ETSP) Station 9 (13ºS and 82.2ºW) and Station 17 (16.7ºS and 79ºW) sampled in July 2013, metagenomes were obtained from Bioproject PRJNA704804. There is no metagenomic data available for the upper surface (Top 60 m) (Fuchsman et al., 2022). Hydrographic and nutrient data from this ETSP cruise are previously published (Fuchsman et al., 2022; Peters et al., 2018). Cyanate data from this ETSP cruise can be found in (Widner, Mordy, et al., 2018), but urea was not measured. Metagenomes from ETNP Station 136 (17.04 ºN, 106.54 ºW) in April 2012, can be found in Bioproject PRJNA350692 (Fuchsman et al., 2017) and nutrient data for ETNP St 136 can be seen in (Fuchsman et al., 2017). Cyanate and urea concentration data from this ETNP cruise can be seen in (Widner, Fuchsman, et al., 2018). Station location can be found in Fig. 1.1.

Phylogenetic Read Placement Technique (Metagenomic reads)

Phylogenetic trees for urease (*ureC*) and cyanase *(cynS*) from (Widner, Fuchsman, et al., 2018) and nitrite oxidoreductase (*nxrB*) from (Fuchsman et al., 2017) were updated using assembled sequences from the ETSP and single cell genomes (Berube et al., 2018). Hydrazine oxidoreductase (*hzo*) data was already published (Fuchsman et al., 2022). The RNA polymerase (*rpoB*) phylogenetic tree can be seen in Fuchsman and Hays (2023). A set of reference amino acid sequences (encoding for the

genes *nxrB*, *ureC*, and *cynS* separately) were used as a query to be blasted against the assembled protein ETSP 2013 dataset using blastp (Altschul et al., 1997). References and assembled sequences were aligned using MUSCLE v3.8.1551 (Edgar, 2004). The alignment was examined using SeaView version 4 (Gouy et al., 2010). The alignment was used to build a phylogenetic tree with bootstrapping using RAxML (Berger & Stamatakis, 2012) and later visualized using FigTree v1.4.4 (https://github.com/rambaut/figtree/releases). The tree that has been constructed using RAxML was used as the reference frame to place the metagenomic reads from the ETSP and ETNP datasets.

Read placement was done by recruiting short metagenomic reads via tblastn search of the metagenomes (using an e-value of < -5) (Fuchsman et al., 2017). The reads were later trimmed to remove and converted to amino acid sequences. After the quality trimming, sequences longer than 100 bp (33 amino acids) were kept. The amino acid translated reads were then aligned against the reference sequences using PaPaRa 2.0 (Berger & Stamatakis, 2012). The non-overlapping paired read ends were then combined into one sequence in the same alignment using a python script and they were later placed in the tree using EPA-ng v0.3.6 with filter-max as 1 (Barbera et al., 2019). Placed reads have a pendant length indicating the similarity between a query read and the location it places on the tree. Reads that were placed with a pendant length greater than 2 were removed. 1% or less of reads were removed at this step. To sort the reads into taxonomic groups, the reads in each group were enumerated using the assigned subcommand of Gappa and a taxonomy file listing the taxonomy of the tree reference sequences Gappa v0.6.1 (Czech et al., 2020). Taxonomic read counts were normalized using the method previously described by (Fuchsman, Palevsky, et al., 2019) where normalization factors for each sample were determined by dividing 48556135 (the number of reads in the 100 m ETNP sample) by the number of good quality reads in a sample. The read counts were multiplied by the sample normalization factor, divided by the gene length (1704 nucleotides for *ureC* (Koper et al., 2004), 468 nucleotides for *cynS* (Chin et al., 1983), and 1275 nucleotides for *nxrB*) and then multiplied by 100 to make visualization easier. We refer to these numbers as normalized reads.

We also calculated the percentage of microbial groups containing *cynS* and *ureC* out of the total microbial community (Bacteria and Archaea *rpoB* reads) or proportions of individual taxa using *rpoB* normalized reads for that taxa.

Phylogenetic Read Placement Technique (Transcripts)

Transcripts of sequences corresponding to the transcript of station P2 (16°55 N, 107°09 W) in the ETNP in May 2018 were obtained from the oxycline and top of the ODZ (76-150 m) (Mattes et al., 2022), downloaded from BioProject PRJNA727903 and processed following the same methodology described above. ST P2 is geographically close to ST 136, examined for metagenomes.

Results

Total *ureC* and *cynS* to *rpoB* gene ratios

In the oxic time-points from HOT, >100% of the microbial prokaryotic community (Bacteria and Archaea) had the *ureC* gene in surface waters (top 100 m depth). In contrast, only 30% of the microbial community contained *ureC* in the ODZ (Fig. 1.6). In the ODZ stations, we observed more diverse groups containing *ureC* than in the oxic time-points (Fig. 1.6). In the ODZ stations, we observed the presence of *ureC* of microbial groups that were barely present in the oxic time-points, including anammox bacteria (Scalindua), Verrucomicrobia and members of the Gammaproteobacteria including Thioglobus. In comparison, other microbial groups containing *ureC,* like Nitrospina, Thaumarchaeota and Picocyanobacteria were present in both the ODZs and oxic time-points.

At the ODZ stations, less than 12% of the total bacterial and archaeal communities had *cynS* genes (Fig. 1.5). Among those microbial communities, the highest abundance of *cynS* corresponded to Cand. Scalindua (anammox), with 11.1% in the ETNP at 160 m. In the case of *Nitrospina*, the highest abundance was observed in the ETSP station 17 with 4% of the total microbial community at a depth of 70 m. In contrast, in the oxic time-points up to 25% of the total bacterial and archaeal communities had *cynS* genes, and the majority of *cynS* reads correspond to *Prochlorococcus*; for example, in time point 275 (Aug.), *Prochlorococcus* reached 24.1% of the total community of *cynS* at 100 m, and in the same time, Nitrospina reached 1.4% at 150 m. We also checked the presence of *cynS* in eukaryotic groups like Algae, but we only found 11 normalized reads at 70 m depth at station 136 in the ETNP and 4 normalized reads in the ETSP station 9 at 90 m (Supplementary Figure. 1.1)

Proportion of taxa harboring functional genes

In this section, we will describe the ratios of taxon specific functional genes *ureC* and *cynS* to taxon specific *rpoB* reads for a variety of groups important to biogeochemical cycling. Some taxa have only one of these two genes. We also explore the proportions of taxon specific functional genes nitrite oxidoreductase (*nxrB*; *Nitrospina*) and hydrazine oxidoreductase (*hzo;* Anammox) to support our taxon specific *rpoB* data.

Anammox cynS, *ureC* and *hzo*

Anammox bacteria were found only in the ODZ core. The abundance of Anammox measured *rpoB* gene copy number, was highest in the ETNP station, comprising up to ~10% of the prokaryotic community. In contrast its total abundance peaked at 6% for both ETSP stations (Fig. 1.6 a). Anammox bacteria have both *ureC* and *cynS* genes (Fig. 1.6 b). Anammox *cynS* games are usually almost twice as abundant as Anammox *ureC* genes within each depth. In the ETNP ODZ, the highest Anammox *cynS* to *rpoB* ratio was 116.4% at 180 m: meanwhile the *ureC* to *rpoB* ratio was 56.1%. At the ETSP

ODZ station 17 (ST17), the *cynS* abundance was 157% of Anammox bacteria, versus 80.5% of *ureC* reads at 210 m. The highest abundance of *cynS* reads was found in the ETSP ST9 with 177.6% *cynS* reads and 80.5% *ureC* reads at 275 m. The gene for hydrazine oxidoreductase (*hzo*; a key gene in the anammox process) showed similar patterns and abundance to the *cynS* gene in the ETSP Station (Fig. 1.7 b), however, in the ETNP and the ETSP ST17, the abundance of *hzo* was lower than *cynS* at some depths in the ODZ core.

Nitrospina *cynS, ureC*, and *nxrB*

The abundance of *Nitrospina* (Fig. 1.7 a) estimated using the *rpoB* gene was highest in the ODZ ETSP station 17, reaching 11.4% of the prokaryotic community at 140 m, and it was lower, below 6% of the prokaryotic community, in the other two ODZ stations. The abundance of *Nitrospina* was below 3% at the oxic time-points (HOT). *Nitrospina* has both genes *cynS* and *ureC*. However, a higher proportion of *Nitrospina* had *ureC* than *cynS* in the ODZ and the oxic time-points (Fig. 1.7 b). In the oxic time-points, the abundance of *ureC* reads was higher than *cynS* except in time point 272 at 225 m depth, where the abundance of *cynS* was 118.8% compared to 101.9% *ureC* at the same depth. Within the oxic time-points, the proportion of *Nitrospina* with *ureC* varied between 77.1% to 161.3%, but in the case of *cynS*, the proportions were 37.4% to 118.8%. In the ETNP station, the abundance of *ureC* was also higher than *cynS*; the *ureC* reads were between 25.9-63.4% and *cynS* 20.3-66.2% of *Nitrospina*. In the ETSP ODZ ST9, *ureC* reads were between 46-107.8% and *cynS* 2.6-45.8%. The proportion of both *cynS* and *ureC* genes was lower than the proportion of the gene nitrite oxidoreductase (*nxrB*), a key gene in the nitrite oxidation pathway in *Nitrospina,* which is often found at 2 copies per genome (Fuchsman et al., 2017; Lücker et al., 2013).

Thioglobaceae *ureC*

The abundance of Thioglobaceae estimated using the *rpoB* was higher in the ODZ stations than at the oxic time-points. (Fig. 1.8 a) At the ETSP ST17, Thioglobaceae reached up to 11.6% of the prokaryotic community. However, in the ETNP, the Thioglobaceae abundance was lower than 5.6% of the community. In contrast, Thioglobaceae in the oxic time-points (HOT) reached only 3.6% of the prokaryotic community.

The Thioglobaceae group was composed of subclades SUP05, Arctic-96BD19 and an ODZ clade (Fig. 1.9), using the *rpoB* gene. The ODZ clade had the highest abundance reaching 10.7% at the ETSP ST9 at 350 m and 10.5% at 350 m in the ETSP ST17. The abundance of the Arctic-96BD19 was below 4.5% in the ODZ core but higher (7.7%) at 450 m in hypoxic areas below the ODZ in the ETSP ST17, whereas the sulfide-oxidizing SUP05 clade had a low abundance (<1.4%) in all the ODZ stations. In the ETNP, the ODZ clade was the most abundant (Fig. 1.9), increasing from 1.6% at 110 m to 5% at 300 m depth. The Arctic-96BD19 clade was 1% at 100 m, but the

abundance decreased inside the ODZ core and the SUP05 clade was almost nonexistent in this station. In the oxic time-point HOT 275, only the Arctic-96BD19 clade was present with abundances that increased with depth from 0.4% at 150 m to 3.6% at 250 m depth.

We observed that Thioglobaceae has only *ureC* and no *cynS* genes (Fig. 1.8 b). In the oxic time-point (HOT), the proportion of Thioglobaceae with *ureC* reads was below 7.5% in the majority of the oxic time-points, with the exception of one point, 18.8% at 150 m at time point 278 (November); there was only one depth with Thioglobaceae *ureC* reads at the oxic time-point 272 (May; 4% at 250 m). On the contrary, in the ODZ stations, the proportion of Thioglobaceae with *ureC* increased with depth inside the ODZ core. In the ETNP, the *ureC* reads abundance varied between 27.5% of Thioglobaceae at 90 m in the oxic region increasing towards the ODZ core where the proportion of Thioglobaceae with *ureC* varied between 60.3% and 109.3%. In the ETSP ODZ ST9, the proportion of Thioglobaceae with *ureC* in the oxic region reached 22.1% at 100 m above the ODZ core, whereas in the ODZ core, the proportion of Thioglobaceae with *ureC* varied between 71.6 to 109.8%. In the ETSP ST17 the proportion of Thioglobaceae with *ureC* varied from 9.7 at 110 m depth in oxic waters increasing to 83.7% at 140 m depth in the top of the ODZ core. Within the ODZ core the proportion of Thioglobaceae with *ureC* varied between 87.3-111.8%, below the ODZ core the abundance of Thioglobaceae with *ureC* decreased to 21.7% at 475 m depth.

SAR11 *ureC*

The abundance of SAR11 in the oxic time-points (HOT) according to *rpoB* ranged between 24.9 to 39.4% (Fig. 1.10 a). SAR11 made up to 60% of the microbial community identified with the *rpoB* gene in the ETNP in the ODZ core (Fig. 1.10 a; (Fuchsman et al., 2017)), contrastingly the abundance of SAR11 in the ETSP stations was below 32.3%.

SAR11 only had *ureC*, and not *cynS.* The proportion of the *ureC* gene was higher in the oxic time-points (Fig. 1.10 b), specifically in the surface (30-42.6%), and decreased with depth reaching <7% of SAR11 containing *ureC* at depth. At the ODZ stations, the proportion of the *ureC* reads was below 10%, however in the ODZ core, the *ureC* reads had a slight increase, reaching 6.2% of SAR11 at 260 m at the ETSP ST9 and 5.5% of SAR11 at 250 m in the ETSP ST17. Below the ODZ core, the proportion of SAR11 with *ureC* decreased to below 1%. In the ETNP, the proportion of SAR11 with *ureC* decreased from 6.4% to 1.6% in the oxic region but increased in the ODZ core, reaching 10% of SAR11 at 300 m depth.

We observed that *ureC* reads could be assigned to different SAR11 subgroups Subgroup 1a.3, Subgroup ODZ, Subgroup IV, Subgroup V and two unidentified subgroups (I, II). The proportion of each subgroup varied according to the presence of oxygen. Inside the ODZ core, 100% of the SAR11 *ureC* reads corresponded to the ODZ subgroup (Fig. 1.10 c). In the ETSP ST17, the Subgroup 1a.3 *ureC* was found only in the oxic waters on top of the ODZ (27.8% of SAR11 *ureC* at 110 m) but was below 1.2% at 350

m, increasing to 25% at hypoxic waters at 475 m. All the oxic time-points had similar profiles, for example, in the oxic time-point 272 (May), two subgroups contributed to most of the SAR11 *ureC* abundance: Subgroup V, and Subgroup 1a.3, whereas the other subgroups had *ureC* abundances below 13.5% of SAR11 *ureC*. The subgroup V *ureC* abundance varied between 24.2- 57.9% of the total SAR11 *ureC*, increasing with depth, peaking at 250 m depth. In contrast, subgroup 1a.3 *ureC* abundance varied between 31.4-57.7% of total SAR11 *ureC*, decreasing with depth. The highest *ureC* abundance of 1a.3 at the oxic time-point HOT 272 was 57.1% at a depth of 75 m.

Verrucomicrobia *ureC*

The abundance of Verrucomicrobia, calculated using the *rpoB* gene (Fig. 1.11 a), was extremely low at the oxic time-points, at <1% of the prokaryotic community. Verrucomicrobia abundance estimated using the *rpoB* gene was consistently <5% of the community in the ODZ core in the ETSP ODZ ST9 and ST17. However, the abundance of Verrucomicrobia increased in hypoxic waters below the ODZ reaching 7.9% of the community in station 17 at 400 m and 6.9% at 400 m in ST9.

Verrucomicrobia only had *ureC* but not *cynS* (Fig. 1.11 b). In the ETNP, the abundance of *ureC* reads in the oxic waters on top of the ODZ varied between 5.8% to 8.4% of Verrucomicrobia but at 110 m the abundance of *ureC* reads decreased to 2.3%, and inside the ODZ the abundance of *ureC* varied between 5.9% of Verrucomicrobia at 120 m and a maximum of *ureC* of 10% at 140 m. In the ETSP ST17, the abundance of *ureC* in Verrucomicrobia decreased in the ODZ core; in the oxic waters, *ureC* reached 63.9% of Verrucomicrobia at 110 m, decreasing in ODZ waters (14.2-27.7%), but increased again below the ODZ to 112.5% of Verrucomicrobia at 400 m depth. In the ETSP ST9, the proportion of Verrucomicrobia with *ureC* was higher in the hypoxic waters above the ODZ (66% at 100 m), and below (117.4% at 400 m depth). In the ODZ core, the proportion of Verrucomicrobia with *ureC* varied between 6.7-25.8%. For both ETSP stations, the proportions of Verrucomicrobia with *ureC* were higher in the hypoxic regions above and below the ODZ than in the ODZ core.

Thaumarchaeota *ureC*

Thaumarchaeota abundance, estimated with the *rpoB* gene, was higher in the oxic time-points and increased with depth, reaching ~30% of the prokaryotic community at 250 m in all time-points (Fig. 1.12 a). In contrast, at the ODZ stations, the abundance of Thaumarchaeota was higher in areas above and below the ODZ core. In the oxic waters above the ETNP ODZ, Thaumarchaeota *rpoB* reads reached a maximum of 11.5% of the community at 100 m, decreased to 2.4% of the community at the top of the ODZ core at 110 m and continued decreasing in the ODZ. At ETSP ST9 and ST17, the *rpoB* abundance decreased with depth reaching below 1% of the community in the ODZ core. However, the abundance of Thaumarchaeota in hypoxic

waters below the ODZ core, reaching 6.2% of the community at 400 m depth in ST9, and 28.1% of the community at ST17 at 450 m.

The proportion of Thaumarchaeota with *ureC* generally decreases with depth in each station (Fig. 1.12 b). In the case of the oxic time-points (HOT), the abundance of *ureC* varied between 59.9-138.5% of Thaumarchaeota. In the ODZ stations, the Thaumarchaeota *ureC* abundance decreased in the ODZ, especially in the ETNP station. In the oxic surface, the *ureC* reads had an abundance of 126.2% at 60 m. However, in the ODZ core the *ureC* reads decreased, reaching 26.7% of Thaumarchaeota at 300 m depth. In the ETSP ODZ station, the Thaumarchaeota *ureC* depth profiles had a similar shape, with higher Thaumarchaeota *ureC* reads in the surface, decreasing towards depth, with two small peaks in the ODZ core with 90.7% of Thaumarchaeota containing *ureC* at 200m at ST9 and 121.1% at 165 m at ST17.

Cyanobacteria *cynS* and *ureC*

Picocyanobacteria abundance identified with the *rpoB* gene was highest in the oxic time-points (HOT) (Fig. 1.13 a) particularly at the surface. The oxic time-points 272 (May), 275 (August) and point 278 (November) had similar *rpoB* profiles, so only one profile for time-point 272 is described here. At time point 272 (May), the abundance of Picocyanobacteria estimated with the *rpoB* gene, increased steadily with depth from 32.4% at 5 m, reaching the highest abundance of 53.9% at 100 m; below 100 m, the abundance decreased, reaching 0.6% of the microbial community at 225 m. The abundance of Picocyanobacteria in the ETNP (Fig. 1.13 a) was highest at 60 m comprising 15.4% of the prokaryotic community but then it decreased with depth. Picocyanobacteria in the ETSP were generally low since profiles skipped surface waters: at ST9, it was below 2%, and ST17 had almost no Picocyanobacteria (<1%).

We also calculated the total of Picocyanobacteria that contained the *ureC* gene (Fig. 1.14 b). We observed that in ODZ, Picocyanobacteria *ureC* abundances in the oxic-time points were higher in the surface and decreased with depth from ~150 to 50% close to 100 m depth. In the ODZ however, the ETNP and ETSP ST9 had an increase in the *ureC* with depth from <10% in oxic waters to 100 to 130% in the ODZ.

Among the Picocyanobacteria that had *ureC*, we found four phylotypes: *Synechococcus,* High Light *Prochlorococcus* (HL), Low Light I *Prochlorococcus* (LLI), Low Light IV *Prochlorococcus* (LLIV), and Uncharacterized *Prochlorococcus* (Fig. 1.13 c). The Picocyanobacteria *ureC* groups had similar profiles at the three time-points examined at the oxic station (HOT). We elaborate for time point 272 (May 2015) as an example.

At oxic-time 272, the HL *Prochlorococcus* had the highest proportion of *ureC* close to ~100% out of the total Picocyanobacteria *ureC*, from 5 to 100 m depth (Fig.1.13 c), below 125 m depth the abundance of HL *Prochlorococcus* decreased to 10.4% at 225 m depth. The other Picocyanobacteria *ureC* were observed below 100 m depth with high abundances; the LLI *Prochlorococcus*,

for example corresponded to ~50% of the total Picocyanobacteria *ureC,* increasing from 28% at 125 to 59.7% at 225 m depth.

In the ETSP ST9, the High Light *Prochlorococcus ureC* group was basically absent from the water column probably because the upper euphotic zone was not sampled. The LLIV *Prochlorococcus ureC* group had the highest *ureC* abundance out of the total Picocyanobacteria *ureC* (Fig.1.13 c), changing from 49% of Picocyanobacteria at 80 m, to a maximum of 87.1% at 110 m, and 66.7% at 150 m. The LLI *Prochlorococcus ureC* was only present at 2.1% at 80 m depth. The Uncharacterized *Prochlorococcus ureC* was only 21.9% of the total Picocyanobacteria *ureC* at 80 m. In contrast, the *Synechococcus ureC* group had a higher proportion of Picocyanobacteria, ranging from 12 to 33.3% of the total Picocyanobacteria *ureC* in the ODZ (Fig.1.13 c). In the ETNP, the LLIV *Prochlorococcus* group had the highest *ureC* abundance of the total Picocyanobacteria *ureC* reads with 75% at 120 m; within the ODZ the LLIV *Prochlorococcus* abundance varied between 63.1-75% of total Picocyanobacteria *ureC*. The HL *Prochlorococcus ureC* had a low abundance, generally below 3.4% of total Picocyanobacteria *ureC*. The LLI *Prochlorococcus ureC* abundance of total Picocyanobacteria *ureC* decreased with depth from 86.5% at 60 m depth to 0.7% at 120 m depth. The Uncharacterized *Prochlorococcus ureC* increased with depth, reaching a maximum of 25.1% of total Picocyanobacteria *ureC* at 110 m; The *Synechococcus ureC* group abundance increased with depth, reaching 28.1% of Picocyanobacteria at 160 m.

Among the total Picocyanobacteria containing *cynS.* In the ODZ stations the abundance of Picocyanobacteria *cynS* was lower than 40% (Fig.1.13 d), with the highest *cynS* abundance inside the ODZ core. In the oxic time-points, the abundance was lower in the top 50 m (<30%), but increased, reaching 70% at 100 m depth in oxic time-points 275 and 278 (Fig.1.13 d). In the oxic time-point 272, Picocyanobacteria *cynS* had 2 peaks: one of ~50% between 100-150 m depth and a secondary peak of ~70% at around 200 m depth.

Prochlorococcus and *Synechococcus* also contained *cynS* (Fig.1.13 e). In the oxic time-points, *Prochlorococcus cynS* abundance varied between 80-100% of the total Picocyanobacteria *cynS*, whereas *Synechococcus cynS* abundance was below 10% of the total Picocyanobacteria *cynS*. In ETSP ST9 and ST17, the *Synechococcus cynS* was ~90-100% out of the total Picocyanobacteria *cynS* (Fig.1.13 e). In the ETNP, *Synechococcus cynS* abundance peaked in the oxic region (78.3% at 60 m depth, above the ODZ core), decreased with depth and peaked inside the ODZ core (100% at 160 m depth). *Synechococcus cynS* abundance out of the total Picocyanobacteria *ureC* increased from 21.7% at 60% to 100% at 100m in the top of the ODZ and also 100% at 110 (top of the ODZ core), the *Synechococcus cynS* abundance decreased with depth to 9.1% of total Picocyanobacteria *ureC* at 140 m depth.

ureC and *cynS* expression in transcriptional data

ureC transcripts from station P2 in the ETNP were also analyzed using the phylogenetic placement technique. We observed several groups (Fig. 1.14) in the transcripts with Picocyanobacteria having the highest abundance of *ureC* transcripts, with 16.4 normalized *ureC* transcripts in depths 106 and 150 m, and Anammox bacteria (*Cand*. Scalindua) had 2.4 normalized *ureC* transcripts at 112 m depth and 2.7 at 150 m depth. The abundance of *Nitrospina ureC* transcripts was low but present at all the sampled depths, with abundances between 0.2-1.8 normalized *ureC* transcripts. The highest *Nitrospina cynS* abundance was 1.8 normalized *ureC* transcripts at a depth of 106 m.

The *cynS* transcripts in station P2 were dominated by transcripts of Anammox bacteria (*Cand*. Scalindua) (Fig. 1.16), with 104.4 normalized transcripts at 112 m and 83.2 normalized transcripts at 150 m depth. After Anammox bacteria (*Cand.* Scalindua), the second highest number of *cynS* transcripts belonged to Unidentified bacteria. *Nitrospina cynS* transcripts were found in low numbers with a maximum of 2.9 normalized *cynS* transcripts at 112 m depth.

Discussion

Many microorganisms prefer reduced nitrogen forms for assimilation (Glibert et al., 2016). However, ammonium concentrations are extremely low in both ODZs and the oxic oligotrophic ocean (Martens-Habbena et al., 2009; Widner, Fuchsman, et al., 2018). One adaptation to ammonium limitation is for microbes to use small organic reduced N sources, such as urea and cyanate. Urea and cyanate concentrations are also generally in the nanomolar (Kitzinger et al., 2019; Painter et al., 2008; Takeda et al., 2020; Widner, Fuchsman, et al., 2018). However, the ability to use urea or cyanate in addition to ammonium increases microbes' chances to obtain reduced N. The capability of microbes to use small organic reduced nitrogen compounds, which are rarely measured, shifts how we think about N assimilation and dissimilation under limiting inorganic N concentrations, adding important information both about adaptations of individual taxa to N limitation and about the importance of urea and cyanate in the environment.

By analyzing the abundance of normalized reads of the genes for cyanase (*cynS*) and urease (*ureC*) we found that the proportion of each taxonomic group with the ability to utilize urea or cyanate varied among the individual groups with depth. Additionally, there were differences when comparing oxic waters, hypoxic waters and ODZs. This variation implies niche differentiation in cyanate and urea utilization.

Eukaryotic Algae: Mao et al.,(2022) found that when examining transcriptional activity, the transcription of *cynS* in larger size (>0.8 μm) fractionated samples and was dominated by eukaryotic phytoplankton, and the transcription of *cynS* was higher than the transcription of *ureC* in that fraction. We only observed small numbers of algae-derived *cynS* at all stations. This result is not surprising in the ODZs, as eukaryotic algae do not thrive in the ODZs (Fuchsman et al., 2022), but algae *cynS* reads were also only found in low abundances at our oxic time-points (HOT) (Supplementary Figure 1.1). However, HOT is dominated by Picocyanobacteria rather than Eukaryotic algae (Karl et al., 2021). Part of the difference between our results and (Mao et al., 2022) could be because our data was obtained from bulk water samples dominated by bacteria, while Mao et al (2022) examined >0.8 μm fractions enriched in Eukaryotic algae. Our results imply that Eukaryotic algae are not a dominant consumer of cyanate in the systems studied here.

Niche Partitioning of cyanase and urease

Nitrospina has the potential to use cyanate and urea, because it has the genes encoding for the enzymes responsible for cyanate and urea degradation. However, the proportion of Nitrospina with *ureC* was higher than for *cynS* both at oxic time-points at HOT and in ODZ waters. The fact that *Nitrospina* has a higher proportion of *ureC* compared to *cynS* reads could be explained by a metabolic preference of urea as reduced nitrogen form, over cyanate.

On the other hand, **Anammox** bacteria (*Cand*. Scalindua) also have the metabolic capabilities of using urea and cyanate, but a higher proportion of anammox bacteria contained the *cynS* gene compared to *ureC*, which could be explained as Anammox bacteria preferring to use cyanate over urea in the ODZ. The differences in the organic N preferences of these two organisms may prevent them from competing where their depth ranges overlap. In the ETNP station, the only station examined here with measured urea concentration (Fig. 1.1), urea was depleted in the whole water column down to 900 m. In contrast, cyanate was measured at several depths in the surface and upper ODZ (Fig. 1.1; (Widner, Fuchsman, et al., 2018)). However, the detection limit for urea (70 nM) was much higher than the detection limit for cyanate (0.4 nM), so it is difficult to compare (Widner, Fuchsman, et al., 2018).

Although the transcript data and the metagenomic reads data from the ETNP shown here do not correspond to the same sampling collection date (2018 versus 2012), we observed similarities between datasets. In the case of Anammox bacteria (*Cand*. Scalindua), where we observed that *cynS* metagenomic reads dominated over *ureC* reads at the majority of depths inside the ODZ core (Fig. 1.6), similar results were observed with the transcripts (Fig. 1.15), where the highest abundance of *cynS* transcript corresponded to Scalindua in the ODZ core and the abundance of *ureC* in the transcript data was also low in numbers compared to *cynS.* These data imply that Anammox bacteria (*Cand*. Scalindua) is actively transcribing the *cynS* gene to possibly metabolize cyanate from the environment.

In the case of Nitrospina, when we examine the *Nitrospina ureC* transcripts, we observed that *ureC* gene expression was similar in numbers and present at all depths (Fig. 1.14). In contrast to *ureC* transcript data, the *Nitrospina cynS* transcripts showed lower expression <1 *cynS* transcript but increased to 2.2 and 2.9 at 106 and 112 m depth (Fig. 1.15). These observations suggest that *Nitrospina* could be actively transcribing both genes to possibly metabolize both cyanate and urea. The transcript numbers were very low, but contrasted with the distribution of *ureC* and *cynS* reads (Fig. 1.7) when looking into the metagenomic data where the abundance of *ureC* reads was generally higher than *cynS*. A possible explanation for the higher abundance of *cynS* transcripts in *Nitrospina*, could be cells actively transcribing *cynS* as a response to the presence of cyanate in the environment. In the ETNP, cyanate is not evenly distributed in the water columns (Fig. 1.1 (Widner, Mordy, et al., 2018) with peaks of abundance in the surface and top ODZ.

SAR11 was highly abundant in the ocean surface in the oxic time-points (Fig. 1.10), but the abundance of *ureC* from SAR11 decreased constantly with depth, despite the SAR11 abundance (estimated by *rpoB*) remaining fairly constant in the water column. The abundance of the *ureC* reads give an insight of the potential this microbial group has to use urea in oxic waters with low concentrations of other N containing nutrients. At the oxic

time-points for HOT, nitrate was not detected in surface waters and the proportion of SAR11 containing *ureC* negatively correlated with nitrate concentrations (log (nitrate)= -0.053*proportion of SAR11 with urease + 1.1487; R^2 = 0.89, p value = $1E^{-17}$). In contrast in the ETNP and ETSP, the SAR11 *ureC* read abundance was low but tended to increase slightly (5-10%) in the ODZ regions. The phylotypes of SAR11 *ureC* were completely different in oxic waters and in the ODZ with an ODZ specific *ureC* phylotype. SAR11 *ureC* was identified in ETNP *ureC* transcripts, but only at 150 m, the deepest depth sampled for transcripts. Ammonium is often undetectable (<10 nM) in ODZ regions Widner et al., (2018 a; 2018 b), but nitrate concentrations are high (~20 μM) so oxidized sources of N were present. Urease appears to be less advantageous to SAR11 when oxidized sources of inorganic N are present even though energy is needed to convert oxidized N to reduced N.

Thaumarchaeota in oxic time-points had the potential to use urea below surface waters (>100 m depth) because it has the gene *ureC*. Although Thaumarchaeota does not possess the *cynS* gene, it has been reported that it can use cyanate and urea for nitrification and for assimilation (Kitzinger et al., 2019). At the oxic time-points (HOT), it was observed that the Thaumarchaeota abundance increased with depth (Fig. 1.12 a); as Thaumarchaeota increased, the proportion of Thaumarchaeota with *ureC* decreased (Fig. 1.12 b). The decrease in the abundance of *ureC* was also observed in the ODZ stations, where the abundance of Thaumarchaeota (and its *ureC* gene) was high in hypoxic areas surrounding the ODZ core, but extremely low in the ODZ core (Fig. 1.12 a). *ureC* Transcripts from the ETNP mirrored these findings with abundant *ureC* transcripts for Thaumarchaeota in the oxycline, but fewer in the ODZ. In other oxic environments like the Arctic Ocean, it has been found that the abundance of Thaumarchaeota *ureC* increased with depth from the surface down to 100 m, but there was no data for deep waters (Shiozaki et al., 2021). It is possible that the ability to use urea is less useful in the deep ocean (>500 m); perhaps the sources of urea (zooplankton excreta and organic matter degradation) are reduced at these depths.

In **Picocyanobacteria**, we observed different proportions of *cynS* and *ureC* from different Picocyanobacteria groups occurring at different depths. In the case of the *ureC* gene, in the oxic time-points, for example in time-point 275 (August) at HOT, in the surface the majority of reads corresponded to HL *Prochlorococcus* (Fig. 1.14 c) followed by a peak of the LLI *Prochlorococcus* at 100 m depth, the Uncharacterized *Prochlorococcus* peaked at 150 m depth, whereas in the ODZ stations the LLIV *Prochlorococcus* had the highest abundance. We hypothesize that Uncharacterized *Prochlorococcus* corresponds to uncultured NC1/LLVII ecotype of *Prochlorococcus* found in ITS data as this is the only known uncultured *Prochlorococcus* group in the oxic ocean. Our *ureC* data corresponds to the abundances of these ecotypes according to core gene abundances (Fuchsman et al., 2023).

In the case of *cynS*, the distribution of *Picocyanobacteria cynS* varied between stations and also between depths (Fig. 1.13 d). The majority of Picocyanobacteria *cynS* reads correspond to members of the *Prochlorococcus* group but it is not possible to assign *cynS* to specific ecotypes (Fig. 1.13 e). *Prochlorococcus* in ODZs have the ability to use nitrate, nitrite, ammonia and urea for N assimilation, with stable isotopes of ODZ *Prochlorococcus* cells indicating that they primarily utilize nitrite, a partially oxidized form of inorganic N which is abundant in ODZs, reaching 2-3 μM (Aldunate et al., 2020; Widner, Fuchsman, et al., 2018). Similarly, in the oxic ocean some *Prochlorococcus* can utilize nitrate under N limited conditions (Berube et al., 2018; Martiny et al., 2009). However, the presence of urease in *Prochlorococcus* should allow them to utilize reduced N forms, which utilize less energy, when available.

Identification of new urease containing taxa

We identified two new groups of bacteria that use urea in low oxygen conditions: the Gammaproteobacteria **Thioglobus** and an unknown **Verrucomicrobia.**

The Thioglobaceae clade has two distinct published subclades. A cultured strain EF1 of the SUP05 subgroup is capable of respiration with either oxygen and nitrate to oxidize sulfide to fuel carbon fixation (Shah et al., 2017). The EF1 culture can perform nitrate reduction and some steps of the denitrification pathway, but cannot produce nitrogen gas (Shah et al., 2017). A cultured strain of the Arctic96BD-19 subgroup oxidized reduced sulfur, particularly thiosulfate, with oxygen, but was a mixotroph, increasing growth in the presence of glucose (Marshall & Morris, 2013). In the ODZ, we observed three subclades by looking at the abundance of the *rpoB* gene: clades Arctic96-BD19, SUP05 clade, and a new ODZ clade of Thioglobaceae. The third ODZ specific subclade of Thioglobaceae found in our *rpoB* sequences from the ODZs was capable of using urea in the environment, while the other subclades were not.

Thioglobaceae abundance was low in the oxic time-points at HOT (Fig. 1.8 a). Thioglobus was found in very low numbers in all the oxic time-points. It was barely present in oxic time-point 272 and at oxic time-point 275 and 278, it was found in low abundance <3.6% between 150 - 250 m depth. The type of Thioglobaceae found in the oxic waters was Arctic96-BD19, a known oxic clade for this group (Walsh et al., 2009). In contrast, the abundance of Thioglobaceae was higher in the ODZ stations across the water column, and was present at each station with abundances that increased with depth, especially within the ODZ core. The ODZ core was dominated by the new ODZ ecotype.

Thioglobaceae has the potential to use urea, the proportions of Thioglobaceae with *ureC* were higher in ODZ stations (Fig. 1.8 a) The ODZ clade was the clade with the highest abundance in the ODZ core, so it is likely this is the clade contributing to the high number of *ureC* reads we observed in the ODZ core in the ETNP and ETSP stations. Thus, the ODZ Thioglobaceae

may have a greater preference for using urea in the ODZ than the oxic Thioglobaceae (Arctic96-BD19) has at the oxic time-points (HOT). We did not see many *ureC* transcripts for Thioglobaceae in the ETNP ODZ, but the depth range of the transcripts only reached 150 m, and the maximal proportions of Thioglobaceae with *ureC* in the metagenomes were deeper than that in the ODZ.

Verrucomicrobia has the potential to use urea but not cyanate; it has the *ureC* gene, but not *cynS*. (Fig. 1.11 b). We observe a contrasting situation with Verrucomicrobia. In the oxic time-points (HOT Stations), the abundance of Verrucomicrobia is low (<3%) (Fig. 1.11 a), but the proportion of Verrucomicrobia with *ureC* is high (29.8-150%), but in the ODZ, the abundance of Verrucomicrobia examined with the *rpoB* gene, is slightly higher (2-8%) but the proportion of *ureC* is low (<40%) in the ODZ core. In ETSP ST9, the abundance of *ureC* increased from 20.4% of Verrucomicrobia at the bottom of the ODZ core (350 m depth), to 117.4% at 400 m in hypoxic waters below. Similarly, at ETSP ST 17, ~100% of Verrucomicrobia had *ureC* in hypoxic waters below the ODZ. Verrucomicrobia is a phylum among bacteria (Hedlund et al., 1997). The Verrucomicrobia division is characterized by having Chemoheterotrophic growth that contributes to the carbon cycle in the ocean (Freitas et al., 2012). In addition, in Verrucomicrobia MAGs, genes encoding for proteins involved in nitrogen fixation, denitrification, components of the anoxygenic photosystem I and II, and cobalamin biosynthesis have been found (Delmont et al., 2022). Other members of the Verrucomicrobia phylum are methanotrophs, (Schmitz et al., 2021). We do not know the metabolism of the Verrucomicrobia in our system. However, we can see that under low but not zero oxygen conditions some Verrucomicrobia can utilize urea.

In conclusion, microbial access to alternative reduced nitrogen forms such as urea and cyanate, might provide them with a selective advantage, in offshore oceanic regions, contributing to the observed patterns of taxonomic distributions. The differences in cyanase and urease depth profiles between microbes imply niche differentiation. Different microbes have higher proportions of genes for urease in different parts of the water column: SAR11 and *Prochlorococcus* in surface waters, Verrucomicrobia in hypoxic waters, Thioglobus in ODZ waters, and Thaumarchaeota in the lower euphotic zone/mesopelagic. At our oxic station (HOT), *Prochlorococcus* could utilize cyanate in surface waters and Nitrospina could utilize cyanate at depth, while Anammox bacteria (*Cand*. Scalindua) utilized cyanate in the ODZs (Table 1.1). This niche differentiation could be attributed to microbial competition over nitrogen sources, reflecting microbial adaptation to certain parts of the water column. It also, at least in the case of SAR11, is affected by the availability of nitrate, an oxidized inorganic form of N.

An interesting common feature of the microbial taxa we analyzed was the majority correspond to cells that are either photo or chemoautotrophs, including Picocyanobacteria, *Nitrospina*, Thaumarchaeota, Anammox, and

Thioglobaceae (Table 1.1). This is probably because most heterotrophs can assimilate N from organic matter. SAR11 is an exception to this trend. SAR11 is a small free-living heterotroph that has a C1 metabolism and consumes simple DOC (Sun et al., 2011; Tripp et al., 2013). This DOC must not provide enough N to SAR11 for assimilation alone.

Overall, this work adds to our understanding of the N cycle and microbial ecology of the offshore ocean under both oxic and anoxic conditions.

Figures and Table Chapter 1

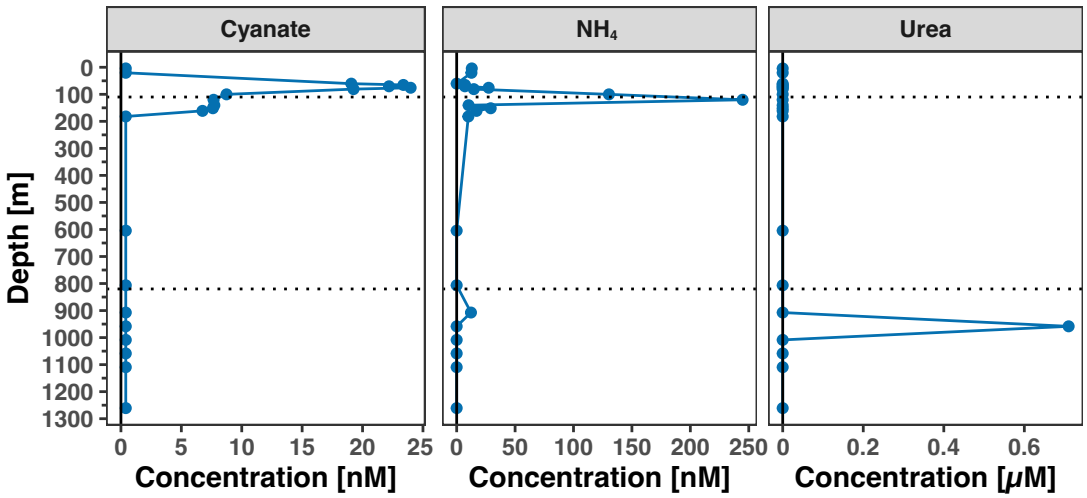


Figure 1.1. Concentration of reduced nitrogen forms in the ETNP. Depth profiles representing the reduced nitrogen in the ETNP at station 136. Cyanate and Ammonium (NH_4) have concentrations in the [nM] range, whereas Urea has concentration in the [μM] range. The area between the dotted lines corresponds to the ODZ core, in which [O_2] < 10 nM (Data from (Widner, Fuchsman, et al., 2018).

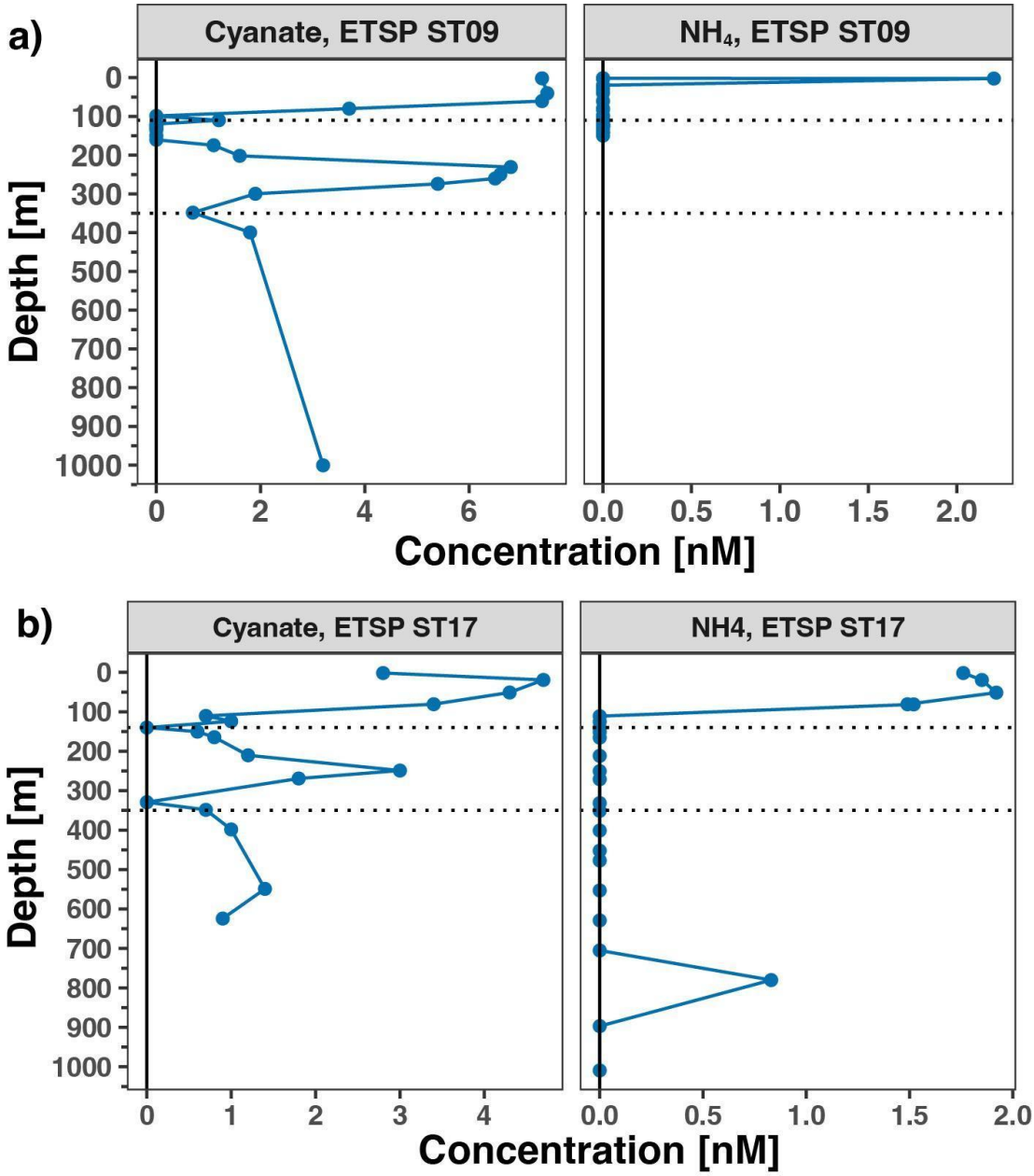


Figure 1.2. Concentration of reduced nitrogen forms in the ETSP. Depth profiles representing the reduced nitrogen in the ETSP a) Station 9, and b) Station 17. Cyanate and Ammonium (NH_4) have concentrations in the [nM] range. There is no urea concentration available for this region. The area between the dotted lines corresponds to the ODZ core (Data from (Widner, Fuchsman, et al., 2018))

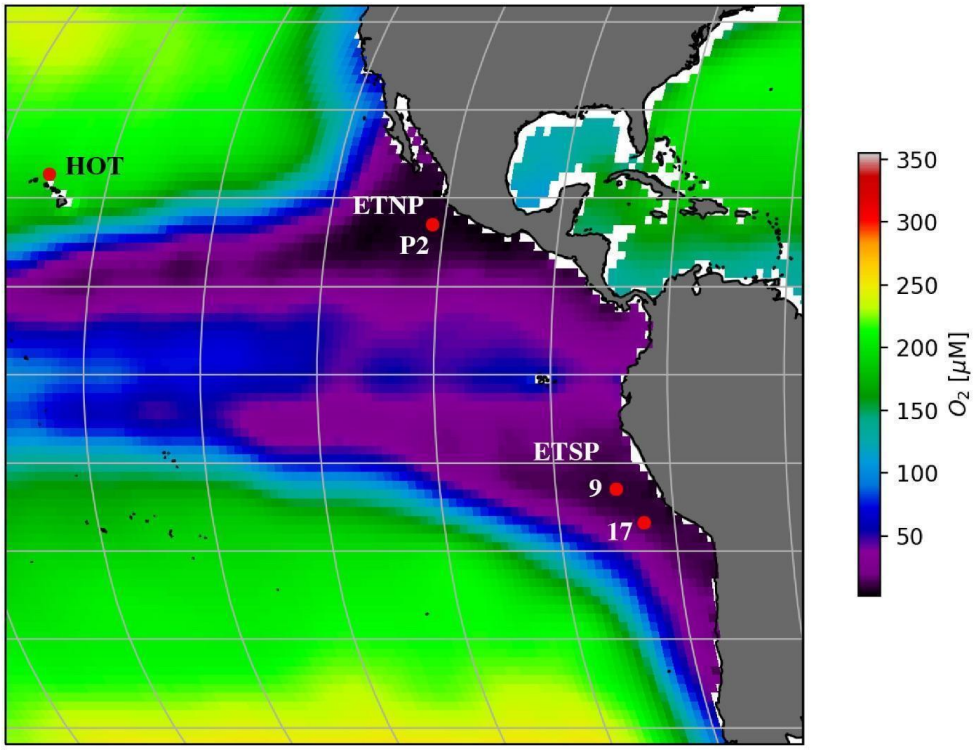


Figure 1.3. Location of the stations studied in Chapter 1. Map of stations, along with World Ocean Atlas oxygen concentration at 250 m. The color scale indicates oxygen concentration in [μM]. The Oxic stations are all different time-points taken at the Hawaii Time Series (HOT), and ODZ stations are from to ETNP (P2) and ETSP (ST9 and ST17). Data collected from the World Ocean Atlas 2018 (Boyer et al., 2018).

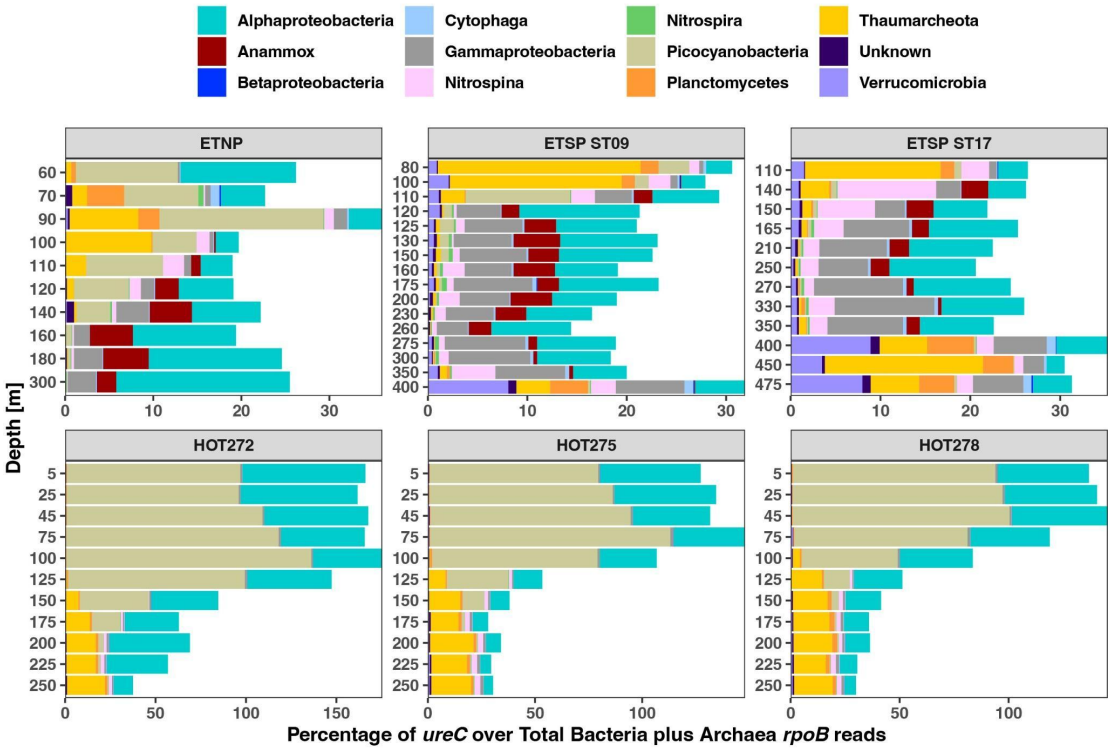


Figure 1.4. *ureC* abundance in the total microbial community. Stacked bar chart showing the percentage of *ureC* copies out of total *rpoB* reads, classified by taxonomic groups according to depth at; ODZ stations (ETNP, ETSP ST09, ETSP ST17) and Oxic time-points at HOT (Stations 272, 275, 278). The depth is not to scale.

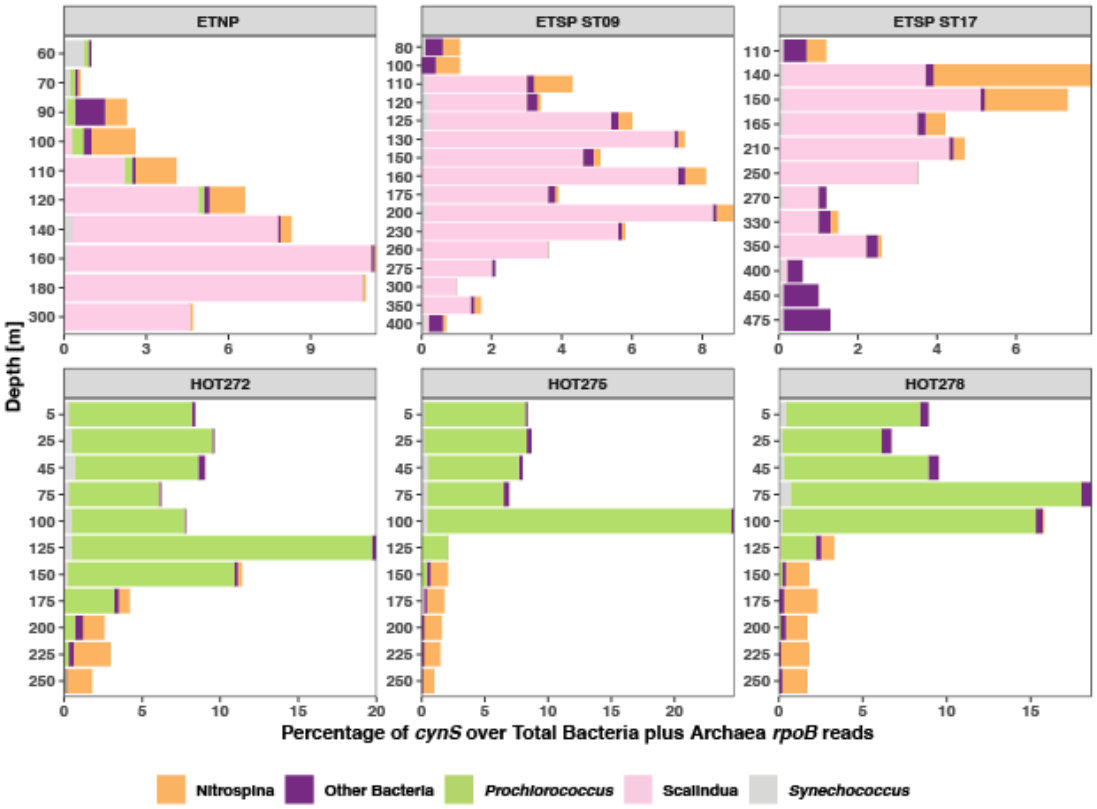


Figure 1.5. *cynS* abundance in the total microbial community. Stacked bar chart showing the normalized read numbers of *cynS* gene copies, classified by taxonomic groups, estimated out of total *rpoB* at the ODZ stations (ETNP, ETSP ST09, ETSP ST17) and Oxic time-points at HOT (Stations 272, 275, 278). Depth is not to scale.

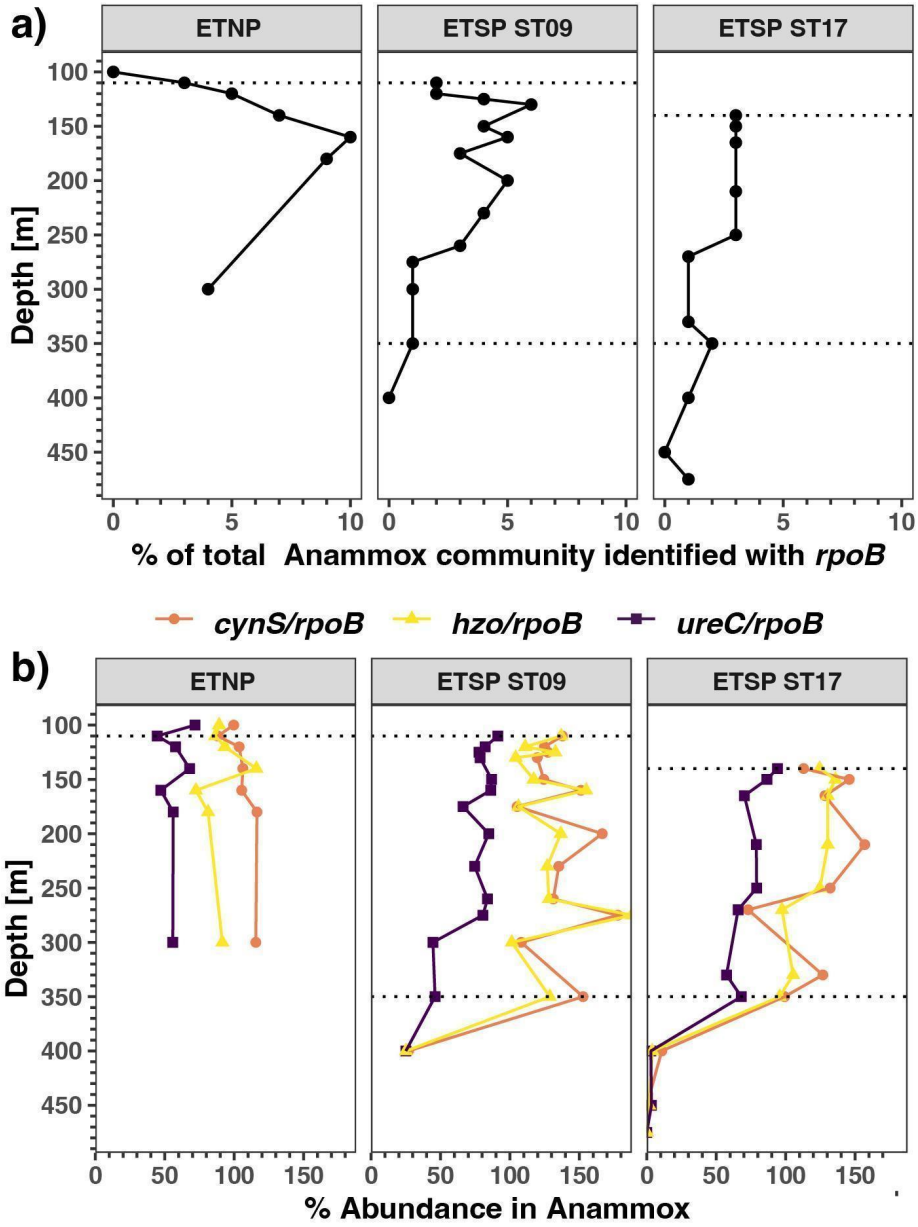


Figure 1.6. Anammox bacteria functional genes distribution in ODZs and oxic time-points. Depth profile of metagenomic reads encoding for enzymes *rpoB* (Beta subunit of RNA polymerase), cyanate (*cynS*), urease subunit alpha (*ureC*), hydrazine oxidoreductase (*hzo*). a) Percentage of the bacteria community identified as Anammox bacteria. b) % functional gene abundance in Anammox bacteria. Anammox specific functional gene to anammox specific *rpoB* gene ratios, expressed as a percentage. Shown are anammox specific *cynS*/*rpoB* ratio (orange), *hzo*/*rpoB* ratio (yellow), and *ureC*/*rpoB* ratio (purple). Each panel corresponds to a different ocean region. The area between the dotted lines corresponds to the ODZ core; in the case of the ETNP, the ODZ extends beyond the region shown in the figure. % Functional gene abundance in Anammox bacteria

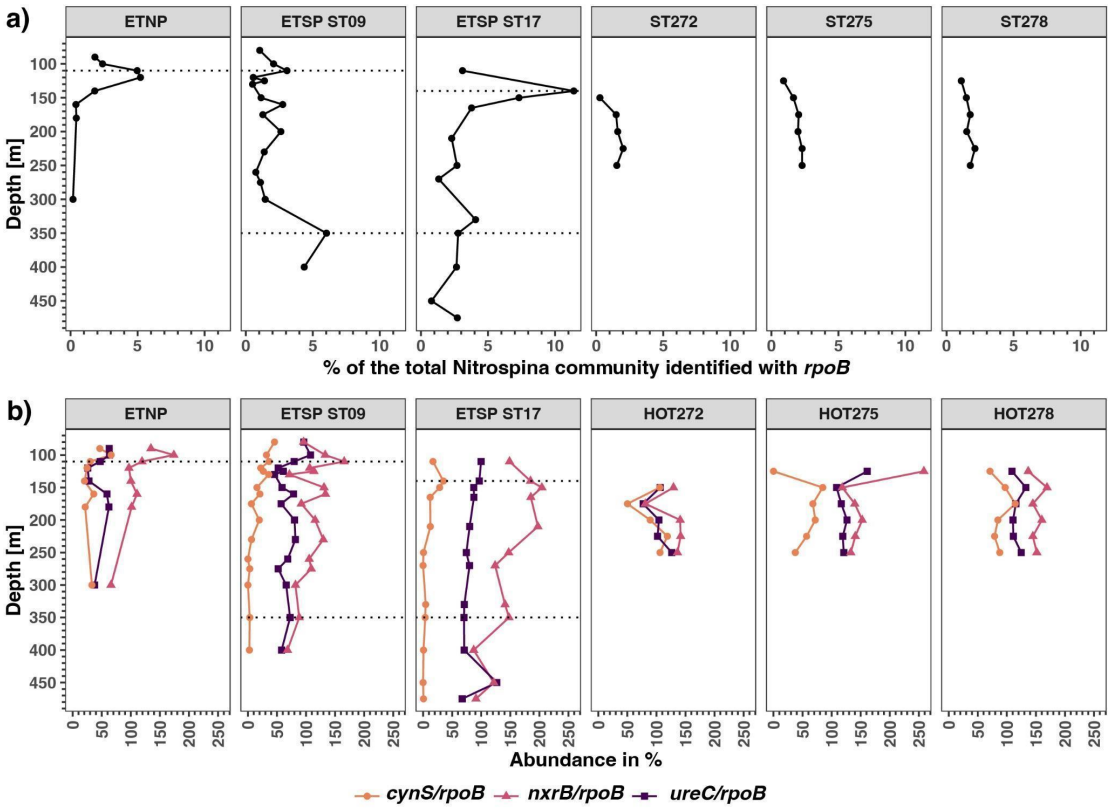


Figure 1.7. *Nitrospina* functional genes distribution in ODZs and oxic time-points. Depth profile of metagenomic reads encoding for enzymes *rpoB* (Beta subunit of RNA polymerase), cyanate (*cynS*), urease subunit alpha (*ureC*), nitrite oxidoreductase (*nxrB*). a) Percentage of the bacteria community identified as Nitrospina Abundance in % of the Nitrospina community identified with *rpoB*. b) Abundance in % of the *cynS/rpoB* ratio (orange), *nxrB/rpoB* ratio (pink), and *ureC/rpoB* ratio (purple). Each panel corresponds to a different ocean region. The area between the dotted lines corresponds to the ODZ core; in the case of the ETNP, the ODZ extends beyond the region shown in the figure.

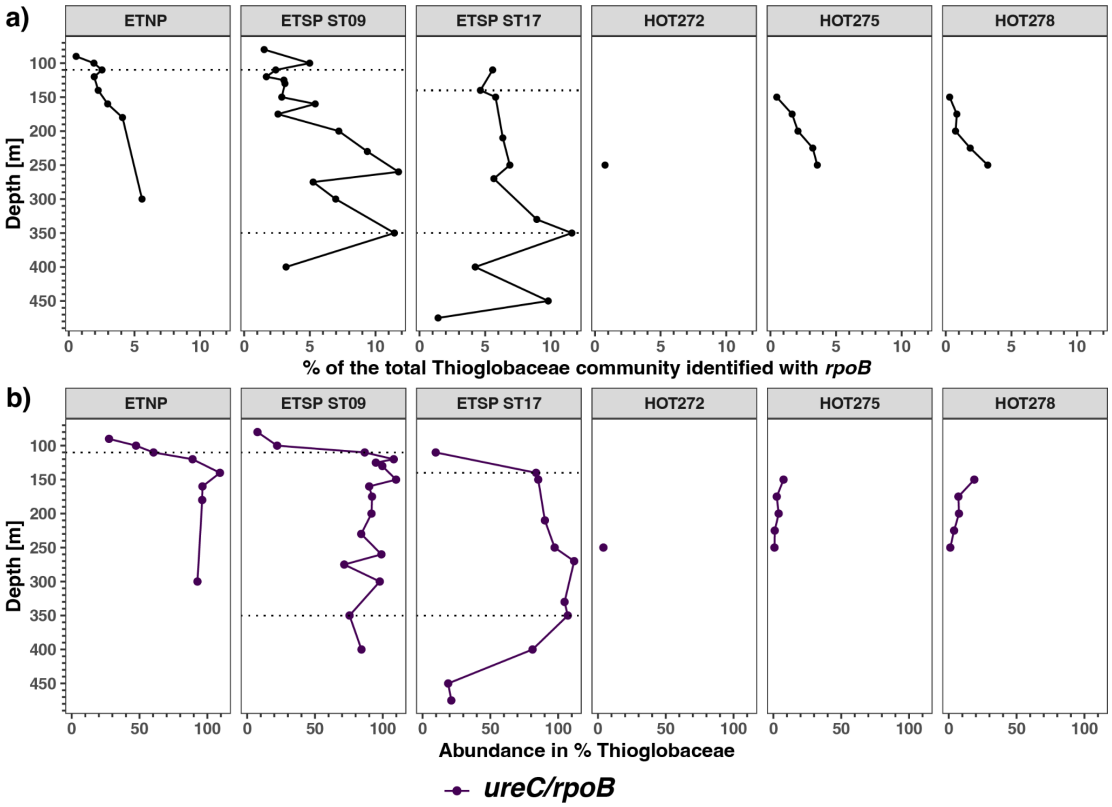


Figure 1.8. Thioglobaceae functional genes distribution in ODZs and oxic time-points. Depth profile of metagenomic reads encoding for enzymes *rpoB* (Beta subunit of RNA polymerase), and urease subunit alpha (*ureC*). a) Abundance in % of the Thioglobaceae community identified with *rpoB*. b) Abundance in % of the Thioglobaceae *ureC/rpoB* ratio. Each panel corresponds to a different ocean region. The area between the dotted lines corresponds to the ODZ core; in the case of the ETNP, the ODZ extends beyond the region shown in the figure.

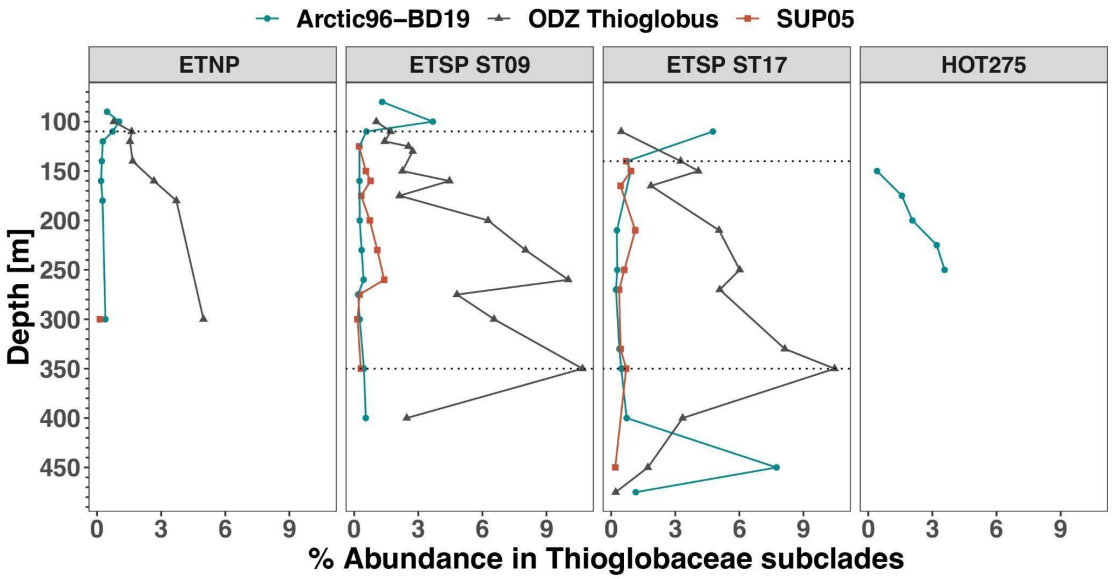


Figure 1.9. Thioglobaceae *rpoB* subclades distribution in ODZs and oxic time-points. Depth profile of metagenomic reads encoding for enzymes *rpoB*, of the Thioglobaceae subclades expressed as % in the ODZ stations (ETNP, ETSP St 09, ETSP ST17) and one Oxic time-point (HOT275). The area between the dotted lines corresponds to the ODZ core; in the case of the ETNP, the ODZ extends beyond the region shown in the figure.

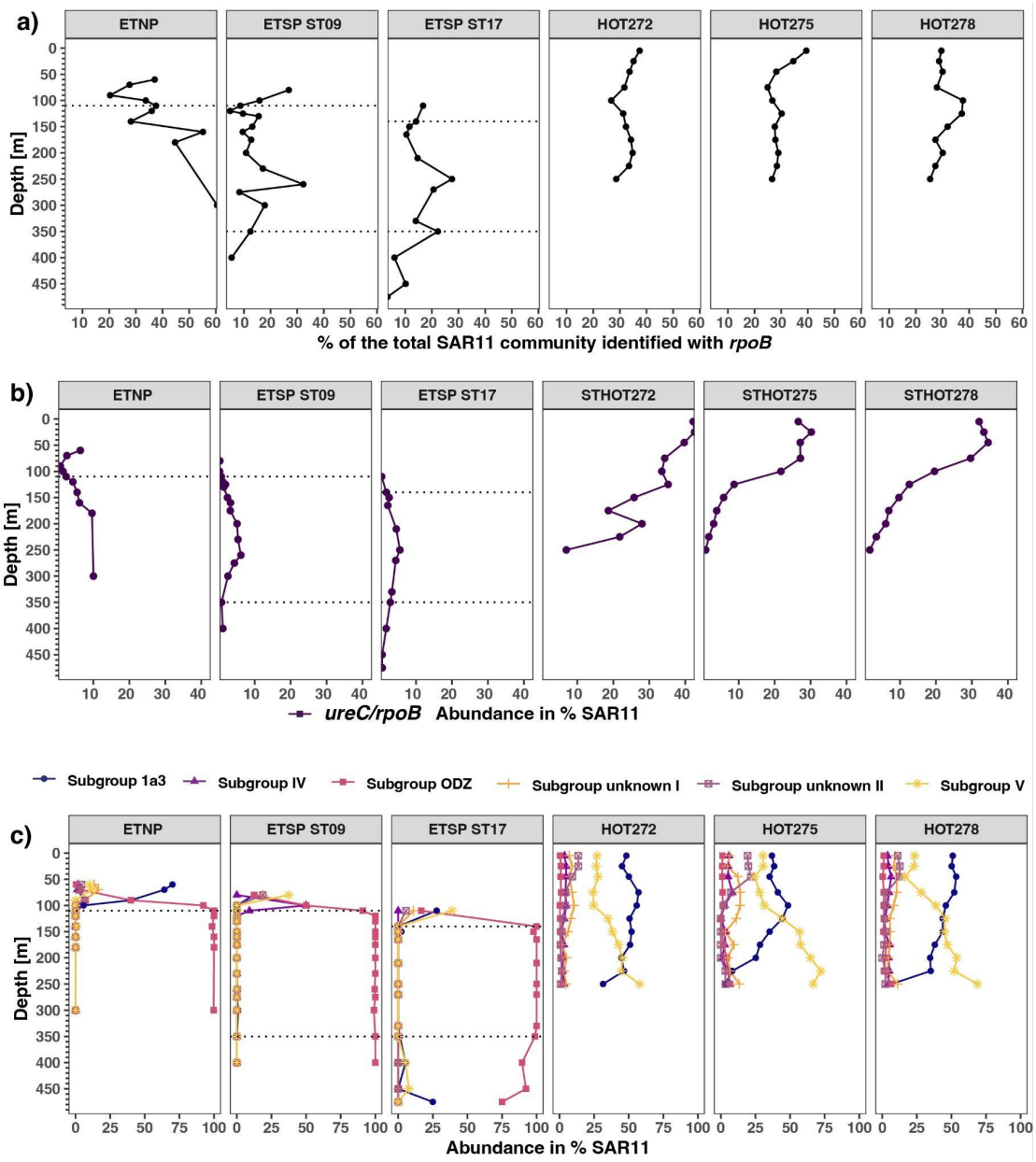


Figure 1.10. SAR11 functional genes distribution in ODZs and oxic time-points. Depth profile of metagenomic reads encoding for enzymes *rpoB* (Beta subunit of RNA polymerase), and urease subunit alpha (*ureC*). a) Abundance the total SAR11 community identified with *rpoB*. b) Abundance in % of the SAR11 *ureC/rpoB* ratio. c) Abundance of each SAR subgroup (*ureC)* calculated of the total SAR11 *ureC.* Each panel corresponds to a different ocean region. The area between the dotted lines corresponds to the ODZ core; in the case of the ETNP, the ODZ extends beyond the region shown in the figure.

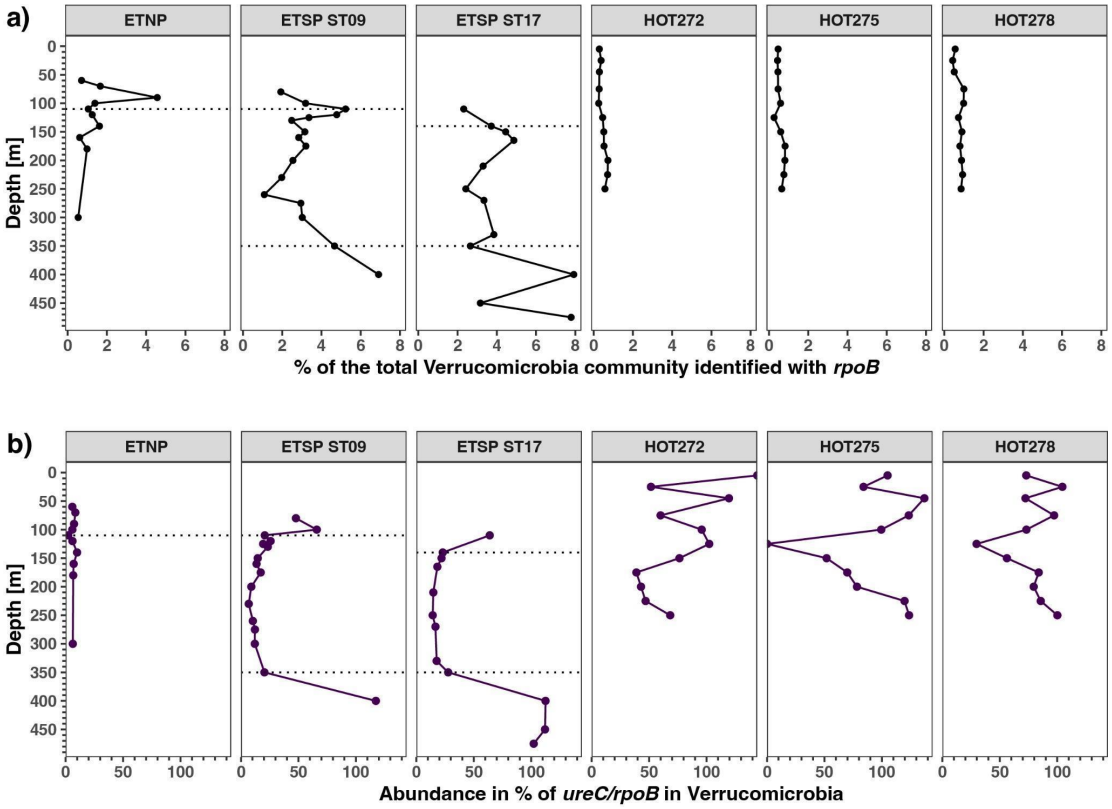


Figure 1.11. Verrucomicrobia functional genes distribution in ODZs and oxic time-points. Depth profile of metagenomic reads encoding for enzymes *rpoB* (Beta subunit of RNA polymerase), and urease subunit alpha (*ureC*). a) Abundance of the total Verrucomicrobia community identified with *rpoB*. b) Abundance in % of the Verrucomicrobia *ureC*/*rpoB* ratio. Each panel corresponds to a different ocean region. The area between the dotted lines corresponds to the ODZ core; in the case of the ETNP, the ODZ extends beyond the region shown in the figure.

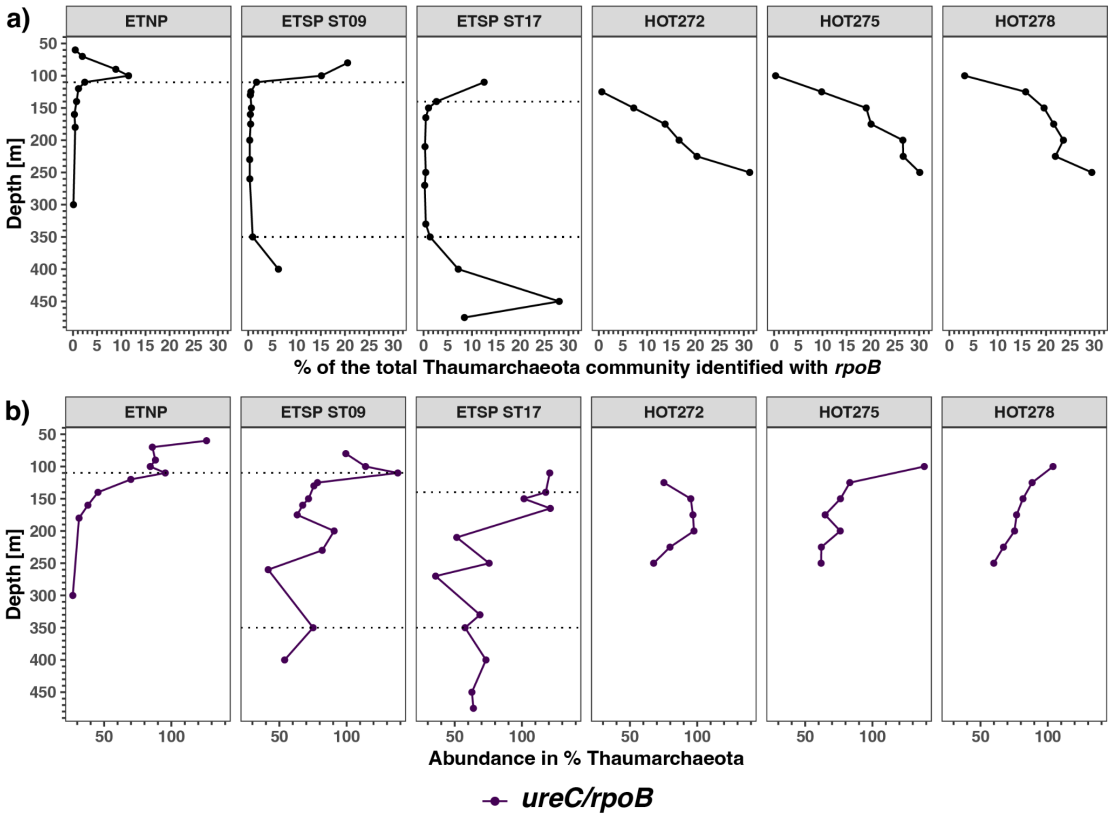


Figure 1.12. Thaumarchaeota functional genes distribution in ODZs and oxic time-points. Depth profile of metagenomic reads encoding for enzymes *rpoB* (Beta subunit of RNA polymerase), and urease subunit alpha (*ureC).* a) Abundance the total Thaumarchaeota community identified with *rpoB*. b) Abundance in % of the Thaumarchaeota *ureC/rpoB* ratio. Each panel corresponds to a different ocean region. The area between the dotted lines corresponds to the ODZ core; in the case of the ETNP, the ODZ extends beyond the region shown in the figure.

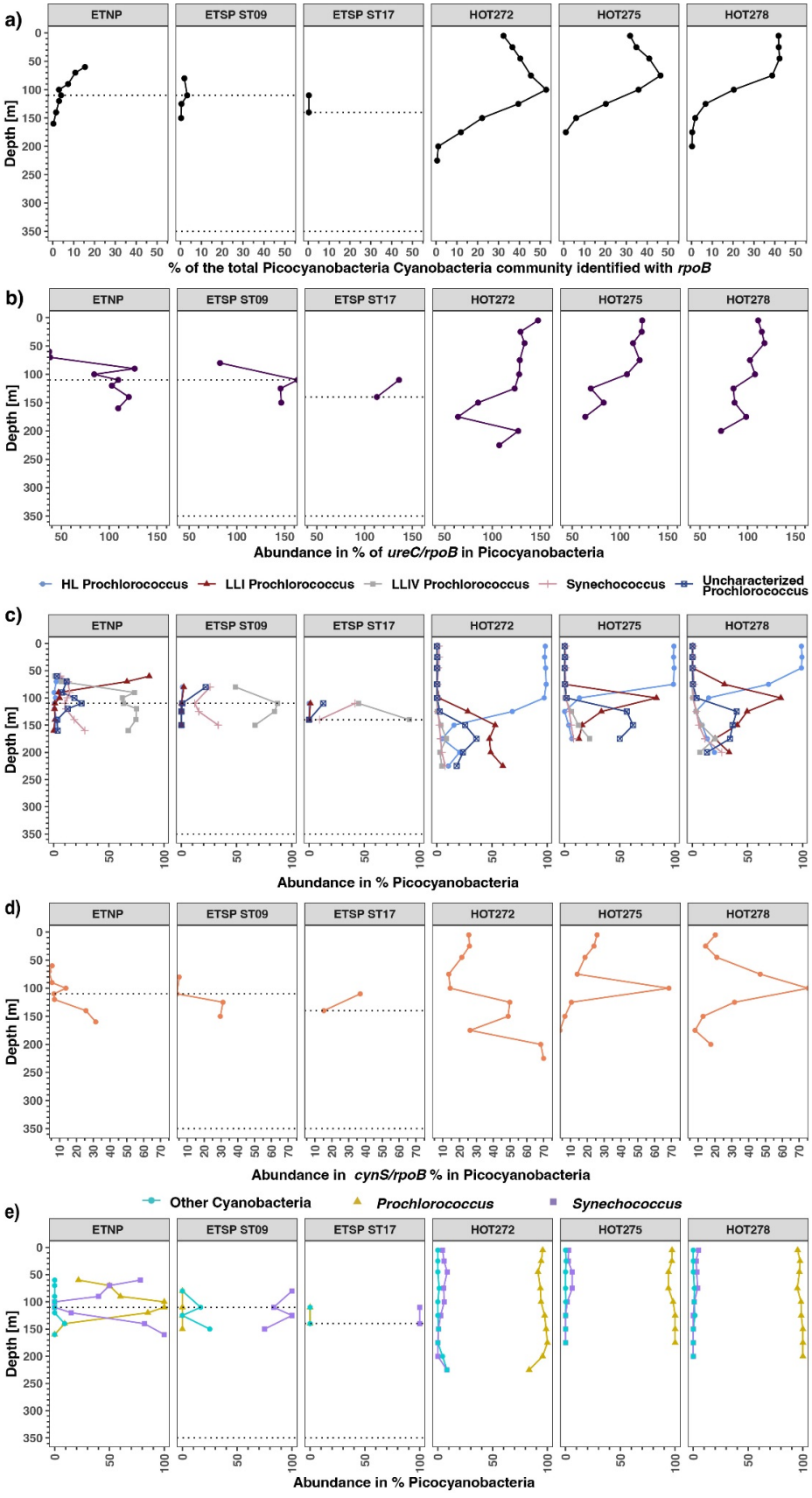

a)
ETNP
ETSP ST09
ETSP ST17
HOT272
HOT275
HOT278
Depth [m]
% of the total Picocyanobacteria Cyanobacteria community identified with rpoB
b)
Abundance in % of ureC/rpoB in Picocyanobacteria
HL Prochlorococcus
LLI Prochlorococcus
LLIV Prochlorococcus
Synechococcus
Uncharacterized Prochlorococcus
c)
Abundance in % Picocyanobacteria
d)
Abundance in cynS/rpoB % in Picocyanobacteria
e)
Other Cyanobacteria
Prochlorococcus
Synechococcus
Abundance in % Picocyanobacteria

Figure 1.13. Picocyanobacteria functional genes distribution in ODZs and oxic time-points. Depth profile of metagenomic reads encoding for enzymes *rpoB* (Beta subunit of RNA polymerase), cyanate (*cynS*), and urease subunit alpha (*ureC*). a) Abundance of the total Picocyanobacteria community identified with *rpoB*. b) Abundance in % of the total Picocyanobacteria *ureC/rpoB* ratio. c) Abundance in % of each the Picocyanobacteria containing *ureC* out of the total Picocyanobacteria *ureC* in *Synechococcus*, and *Prochlorococcus* ecotypes: High Light (HL) *Prochlorococcus,* Low Light (LL) IV *Prochlorococcus*, Low Light (LL) I *Prochlorococcus*, Uncharacterized *Prochlorococcus* and *Synechococcus*, d) Abundance in % of the total Picocyanobacteria *cynS/rpoB* ratio. e) % of the abundance of the *cynS* abundance of *Prochlorococcus*, *Synechococcus*, and "Other Cyanobacteria" calculated over the total Picocyanobacteria *cynS*. Each panel corresponds to a different ocean region. The area between the dotted lines corresponds to the ODZ core; in the case of the ETNP, the ODZ extends beyond the region shown in the figure.

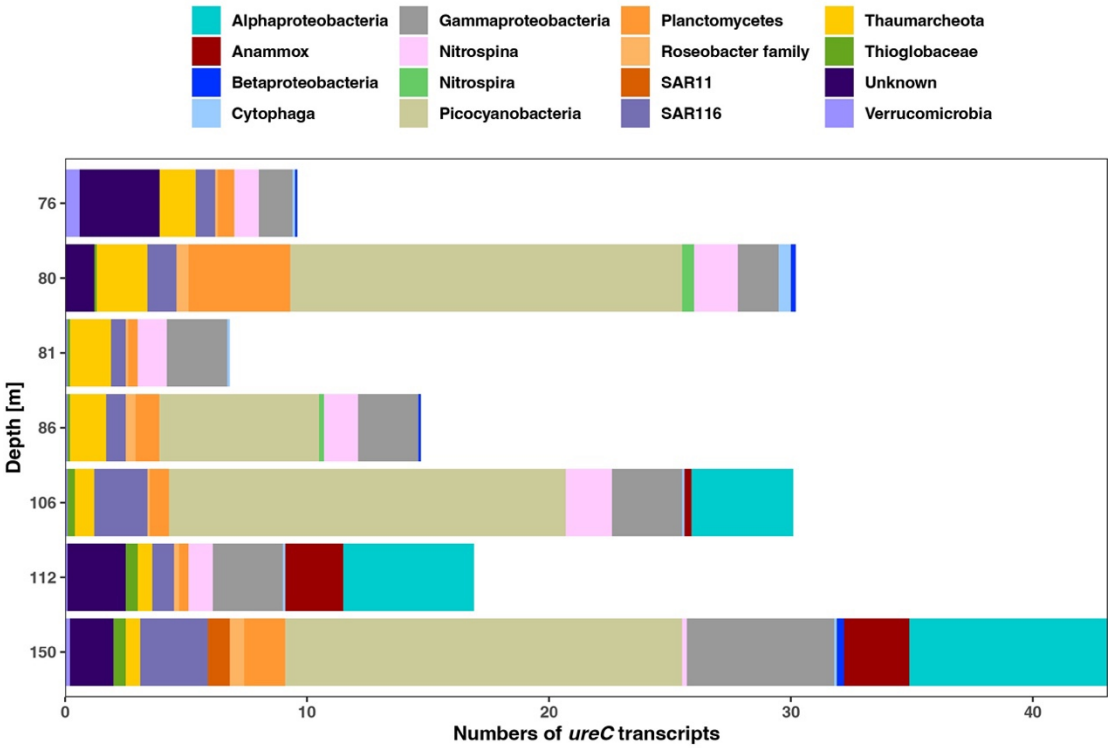


Figure 1.14. *ureC* gene expression in the ETNP. Stacked bar chart showing the normalized *ureC* transcripts of station P2 sampled in May 2018. The *ureC* transcripts are classified by taxonomic groups. Depths include the oxycline and top of the ODZ (76-150 m). The depth is not at a scale, data from (Mattes et al., 2022).

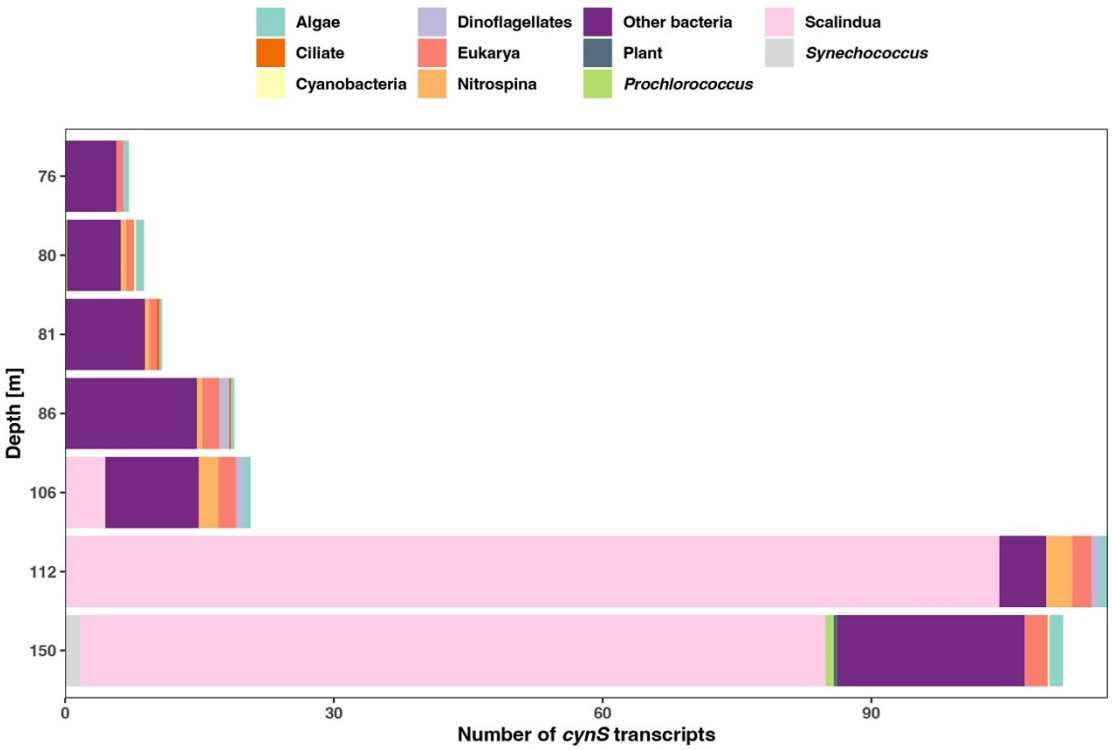
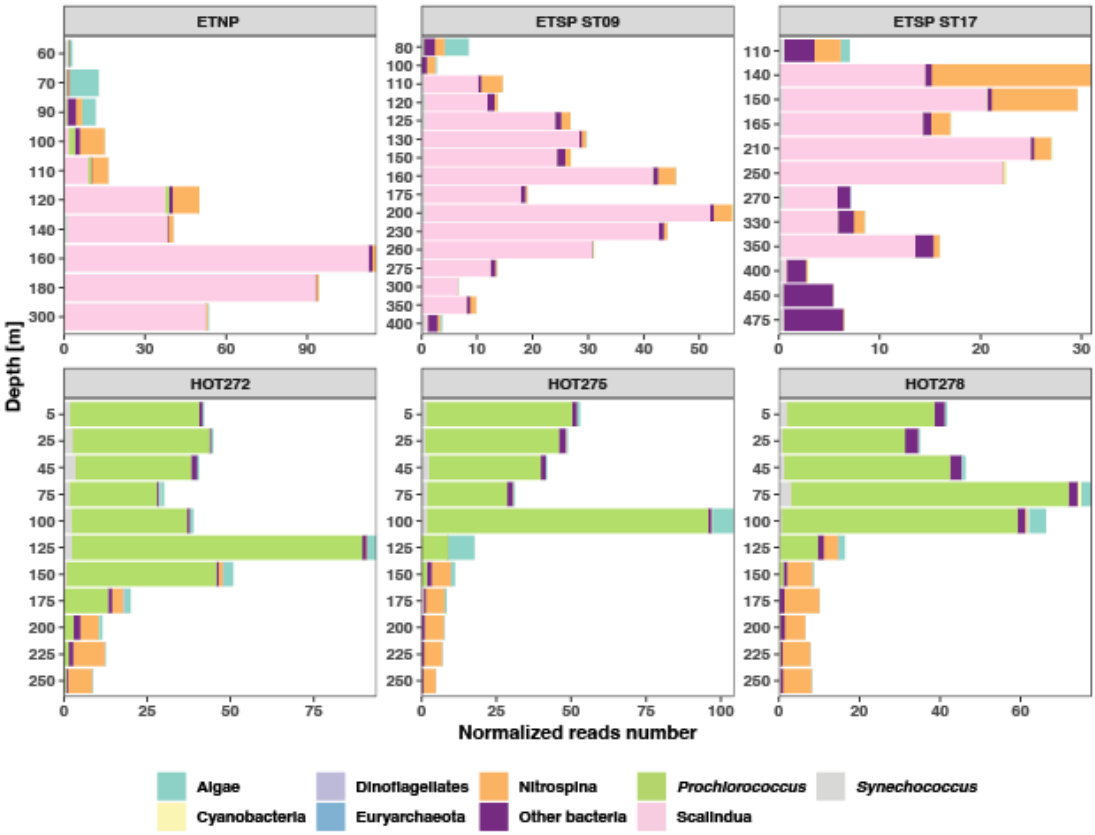


Figure 1.15. *cynS* gene expression in the ETNP. Stacked bar chart showing the normalized *cynS* transcripts of station P2, sampled in May 2018. The *cynS* transcripts are classified by taxonomic groups. The graphs comprehend the oxycline and top of the ODZ (76-150 m). The depth is not at a scale, data from (Mattes et al., 2022).

	Anammox bacteria	Nitrospina	Cyanobacteria	SAR11	Thaumarchaeota	Thioglobus	Verrucomicrobia
Surface oxic	X	X	Cyanate	Urea	Urea	X	Urea
Deep oxic	X	Urea + cyanate?	X	X	Urea	X	Urea
ODZ	Cyanate	Urea	X	A little urea	X	Urea	X
Autotrophic	Yes	Yes	Yes	No	Yes	Yes	?

Table 1.1 Summary of microbial urea/cyanate preference at different ocean regions. ODZ: Oxygen Deficient Zone.



Supplementary Figure 1.1. *cynS* data including eukaryotic algae. Stacked bar chart showing the normalized read numbers of *cynS* gene copies classified by taxonomic groups at the ODZ stations (ETNP, ETSP ST09, ETSP ST17) and Oxic time-points (Stations 272, 275, 278). The depth is not at a scale.

Chapter 2: Microbial ability to utilize cyanate and urea in depth profiles across North-South and West-East transects in the North Atlantic

Abstract

Our understanding of microbial cycles on earth has benefited tremendously by the development of new technologies, and the compilation of massive metagenomic databases, which allow us to look into microbial processes involved in the nitrogen cycle with more detail in vast ocean regions.

Different microbial groups are capable of using different reduced nitrogen forms depending on their metabolic capabilities and the availability of reduced nitrogen forms in the environment. The majority of microorganisms prefer ammonium (NH_4^+), but it is generally absent in the water column offshore. One adaptation to this scarcity is to use other reduced nitrogen forms like urea and cyanate. Two Geotraces transects provide metagenomic data across several depths and stations in the North Atlantic. Eukaryotic algae and the cyanobacteria *Synechococcus* dominated the microbial community along a North-South transect GA02 (Spring, May 2010) and nitrate was present in Northern surface waters. In contrast, the transect orientated West-East GA03 (Fall, December 2011) was nutrient limited and completely dominated by the cyanobacteria *Prochlorococcus* with small amounts of *Synechococcus* in the West (Fuchsman and Hays, 2023). However, the capability of microorganisms to use alternative sources of reduced nitrogen in the North Atlantic over time and space has not been previously examined. These capabilities can be observed by the presence of genes encoding for the enzymes urease (*ureC*) and cyanase (*cynS*), which catalyze the decomposition of urea and cyanate respectively. Phylogenetic read placement of *ureC* and *cynS* metagenomic reads and comparison to single copy core gene RNA polymerase (*rpoB*) was used to calculate abundance and the proportion of each microbial taxon that contains the cyanase and urease genes in depth profiles from in the transects GA02 and GA03 in the North Atlantic. We observed that the abundance of different microbes that had the genes (*cynS* and *ureC*) changed over depth and space. In GA02, SAR11 had the highest proportion of the population having the *ureC* gene. The proportion of *Prochlorococcus* and *Synechococcus* populations that had *cynS* changed across the transect for the North-South stations; *Synechococcus cynS* dominated over *Prochlorococcus cynS* at the Northern stations. In the GA03 transect, Picocyanobacteria, and Thaumarchaeota were the dominant groups possessing *ureC* genes. *Synechococcus cynS* dominated the Western region. Still, as the transect moved East, *Prochlorococcus cynS* appeared but at deeper depths. Despite being present in the GA02 transect, the normalized *cynS* from Eukaryotic Algae were quite low. In the case of *Nitrospina*, in both transects, the proportion of *Nitrospina* with the *ureC* gene was higher than *cynS.* As we analyzed the distribution of microorganisms in the North Atlantic, we observed that the microorganisms capable of using urea and cyanate varied according

to depth, latitude, longitude and seasons. In general, the ability to use cyanate is dominated by bacteria and archaea in the North Atlantic.

Introduction

The study of microbial life is key to understanding microorganisms' role in the global nutrient cycle. In the ocean, Microorganisms play a key role in linking the carbon and nitrogen cycle. However, many aspects of microbial life in the ocean have remained a mystery to scientists. In the last couple of years, the combined multidisciplinary effort of scientists from all over the world has made the collection of data at an unprecedented scale possible. The knowledge and understanding of how microbial communities controls biogeochemical cycles have been continuously expanding thanks to several multidisciplinary, international research campaigns that have compiled metagenomic data, nutrients, and trace elements from vast areas in the ocean globally including TARA (Pesant et al., 2015) and GEOTRACES (R. Anderson et al., 2014), as well as Time-series studies like HOT and BATS (Biller et al., 2018). The GEOTRACES initiative is a multidisciplinary, international effort that aims to study marine biogeochemical cycles of trace elements and how they vary across space and depth (R. Anderson et al., 2014). BioGeotraces is an effort to obtain microbial metagenomic samples on Geotraces cruises (Biller et al., 2018).

Seasonal Changes in Primary producers in the North Atlantic

In the ocean, phytoplankton have a key role in the maintenance of the marine food web as phytoplankton are responsible for producing organic matter via photosynthesis. In the spring in the North Atlantic the increase in light, decrease in mixing, and presence of nutrients lead to an increase in the phytoplankton population, leading to the development of Blooms (Rigby et al., 2020). Phytoplankton Blooms initiate in March/April and progress from South to North (from 40 to 60˚N) and reach a climax in abundance in May (Diaz et al., 2021; Sieracki et al., 1993). During the climax the concentration of nitrate and nitrite reached the highest concentration > 5 μM, indicators indicated increased viral infection, TEP and dissolved organic carbon also accumulated (Diaz et al., 2021). In spring, an increase in the eukaryote-dominated communities and a reduction in cyanobacteria have been observed. Additionally, the phytoplankton community structure shifted along latitude (Bolaños et al., 2020).

In the North Atlantic, there are two Geotraces transects of interest that have metagenomic depth profiles: GEOTRACES GA02 corresponding to a spring (May/June) North-South transect in 2010, and GA03, a fall (November/December) transect from West to East in 2011 (Biller et al., 2018). In the GA02 transect, nitrate and phosphate were detected in the euphotic zone at Northern stations (Fuchsman & Hays, 2023; Middag et al., 2019; Rigby et al., 2020). In the GA03 transect, nutrients like nitrate and phosphate were extremely limited in surface waters, especially in the subtropical gyre (Jenkins et al., 2015). These transects presented very different environmental conditions in the similar geographical area. Additionally, the photosynthesizers were distinct between the two cruises with eukaryotic

phytoplankton dominating in the Northern Stations of GA02, shifting to dominance by *Synechococcus* and then to *Prochlorococcus* as the stations moved South and spring changed to summer (Fuchsman & Hays, 2023). However, in the GA03 transect, *Prochlorococcus* dominated all of the offshore stations (Fuchsman et al., 2023; Fuchsman & Hays, 2023).

Cyanate and urea in the North Atlantic

Most microorganisms prefer reduced nitrogen forms for assimilation because they do not have to spend energy reducing the N before using it in proteins or DNA (Glibert et al., 2016). Ammonium concentrations are extremely low in the oxic oligotrophic ocean because of consumption by phytoplankton in surface waters and efficient scavenging by ammonia-oxidizing Thaumarchaeota in deeper waters (Martens-Habbena et al., 2009:Priddle et al 1997). Ammonium is produced by organic matter degradation and zooplankton excretion, but is quickly consumed (Priddle et al., 1997). While marine autotrophic and heterotrophic bacteria have the ability to transport and use ammonia (Ren et al., 2017), some microbes have adapted to the competition for ammonia by gaining the ability to also use small organic reduced N sources, such as urea and cyanate, and so can adapt to shortages of ammonium. Urea and cyanate were not measured on Geotraces cruises.

Cyanate is a small organic form of reduced nitrogen (OCN^-) (Dirnhuber & Schütz, 1948) that can be produced abiotically (photoproduction) through biotic organic matter degradation, or through algal senescence (Widner et al., 2016). In the North Atlantic, on the coastal North America continental shelf, cyanate has been found at concentrations ranging from below detection limits to 0.4 nM to 11 nM (Widner et al., 2016). However, cyanate has not been quantified offshore. Phytoplankton blooms in the North Atlantic could be a source of cyanate, as a correlation has been shown with the presence of Chlorophyll A and cyanate (Widner et al., 2016).

Urea (CH_4N_2O) is a reduced nitrogen compound, which is produced both as an excretion product by crustacean and gastropod zooplankton (Pitt et al., 2009; Thibodeau et al., 2020) and as a part of organic matter degradation (Berman et al., 1999; Cho et al., 1996; Price et al., 1985). Microorganisms can metabolize urea by using the enzyme urease, encoded by the gene *ureC.* In oxic waters in the Eastern Tropical North Pacific, uptake experiments, which used double labeled urea, demonstrated that the N in urea was preferentially assimilated over the carbon, indicating that organisms were using urea as a N source (Widner, Fuchsman, et al., 2018). Urea concentrations are typically in the nanomolar range. In transects in the North and South Atlantic, upper euphotic zone urea concentrations were typically between 150 and 350 nM, though values could reach as low as 20 nM (Painter et al., 2008). Experimental results in the Atlantic Ocean suggest that the distribution of urea, could be the result of the imbalance between urea consumption and urea production (Painter et al., 2008) . The concentration of urea in the North Atlantic is negatively correlated with spring phytoplankton blooms, where higher uptake and lower urea concentrations were found in

spring (Painter et al., 2008).

Microbial capability of using urea and cyanate in the North Atlantic

In chapter 1, we observed that different taxonomic groups had different capacities for urea and cyanate. The differences were observed not only within microbial groups, but also between surface and deep and oxic, hypoxic and anoxic environments.

Synechococcus and *Prochlorococcus* are Picocyanobacteria prevalent in oligotrophic regions, where they dominate in number, contributing to a significant proportion to primary production (Rii et al., 2016; Scanlan, 2012). *Prochlorococcus* and *Synechococcus* are abundant under different ocean regimes, with *Prochlorococcus* dominating in the oligotrophic ocean nominally from 40°N to 40°S, and *Synechococcus* extending to both cold and high nutrient regimes (Flombaum et al., 2013). The euphotic zone has strong gradients of light and nutrients. *Prochlorococcus* ecotypes and abundances change dramatically in depth profiles with highlight ecotypes in the surface, Low Light I in the middle of the euphotic zone and Low Light II and Low Light IV (LLIV) at the bottom (Ahlgren et al., 2017; Z. I. Johnson, Zinser, Coe, Mcnulty, et al., 2006; Zinser et al., 2007). As primary producers, *Prochlorococcus* and *Synechococcus* need to obtain N from the environment, including both inorganic sources, and the small N containing organic compounds cyanate and urea (Kamennaya & Post, 2011; Rocap et al., 2003)

Several other bacterial and archaeal groups have been implicated in urea and cyanate utilization in the oxic ocean. The heterotrophic Alphaproteobacteria SAR11 is a group capable of using urease (*ureC)* (Connelly et al., 2014; Widner, Fuchsman, et al., 2018). However, the distribution of the *ureC* genes in SAR11 in the ocean is mostly unknown. At HOT in the North Pacific, the proportion of SAR11 with urease correlated negatively with nitrate, with higher proportions of SAR11 with *ureC* in surface waters (Chapter 1). The marine nitrite-oxidizing bacteria *Nitrospina* are aerobic and obligate chemolithotrophs that oxidize nitrite to nitrate (Spieck et al., 2014). Some single-cell *Nitrospina* genomes contained genes for urea and cyanate utilization (Pachiadaki et al., 2017). At HOT and in the ODZs, a higher proportion of Nitrospina had urease than had cyanase (Chapter 1). Additionally, some cultured isolates of ammonia oxidizing Thaumarchaeota can use urea as a sole N source (Qin et al., 2014), and Thaumarchaeota contain urease (*ureC*) in the ocean (Ahlgren et al., 2017; Kitzinger et al., 2019; Shiozaki et al., 2021; Widner, Fuchsman, et al., 2018). Analysis of marine planktonic metatranscriptomes from TARA oceans (global ocean sampling at three depths) showed that, in the >0.8 μm fraction) transcripts for *cynS* were as prevalent as those for *ureC* (Mao et al., 2022). TARA oceans sampled the North Atlantic in February and March of 2012. In the large size fraction (>0.8 μm), various eukaryotic algae, dominated by Pelagophytes, produced *cynS* transcripts in the North Atlantic: in the smaller size fraction (<0.8 μm), *Synechococcus cynS* transcript dominated in surface, whereas *Prochlorococcus* and unclassified microbes contribute to the *cynS* transcripts at the DCM. The abundance of *cynS* transcripts across the global transects is

negatively correlated with dissolved inorganic N concentrations. Though total *cynS* transcripts were reduced in the mesopelagic region compared to the euphotic zone, *Nitrospina cynS* transcripts were found in the mesopelagic (Mao et al., 2022).

We observed niche partitioning for the use of cyanate and/or urea in other Oceanic regions, including HOT and the North and South Pacific ODZs. Although the North Atlantic has been widely studied as part of the GEOTRACES and other projects, there are aspects like the availability of urea and cyanate, in addition to the capability of microbes to use those reduced nitrogen forms that remain uncharacterized. Therefore, in this chapter we aim to examine and compare different microbial groups that are known for being capable of using urea and cyanate, and we look into how microbial groups are using these reduced nitrogen forms across space and time in two transects in the North Atlantic.

Methods

Sample collection

BioGeotraces metagenomic sequences for the GA03 North Atlantic East-West transect, and GA02 North-South transect GA02 (Biller et al., 2018) were downloaded from GenBank bioproject PRJNA385854. GEOTRACES GA02 corresponds to a spring (May/June) North-South transect in 2010, and GA03, a Fall (November/December) transect from West to East in 2011. However, GA03 St 7 was sampled onboard a cruise in October 2010. Nutrient data and CTD information were obtained from the British Oceanographic data Centre (https://www.bodc.ac.uk/geotraces/) as part of the GEOTRACES 2021 Intermediate Data Product (IDP2021).

Phylogenetic Read Placement Technique

Phylogenetic trees for urease (*ureC*) and cyanase *(cynS*) and nitrite oxidoreductase (*nxrB)* were identical to those described in Chapter 1. Read placement was done by recruiting short metagenomic reads via tblastn search of the metagenomes (using an e-value of <-5) (Fuchsman et al., 2017). The reads were later trimmed to remove Ns and converted to amino acid sequences. After the quality trimming, sequences longer than 100 bp (33 amino acids) were kept. The amino acid translated reads were then aligned against the reference sequences using PaPaRa 2.0 (Berger & Stamatakis, 2012) (Berger & Stamatakis, 2012). The non-overlapping paired read ends were then combined into one sequence in the same alignment using a python script and they were later placed in the tree using EPA-ng v0.3.6 with filter-max as 1 (Berger & Stamatakis, 2012). Placed reads have a pendant length indicating the similarity between a query read and the location it places on the tree. Reads that were placed with a pendant length greater than 2 were removed. 1% or less of reads were removed at this step. To sort the reads into taxonomic groups, the reads in each group were enumerated using the assigned subcommand of Gappa and a taxonomy file listing the taxonomy of the tree reference sequences Gappa v0.6.1 (Czech et al., 2020). Taxonomic read counts were normalized using the method previously described by (Fuchsman, Palevsky, et al., 2019) where normalization factors for each sample were determined by dividing 48556135 (the number of reads in the 100 m ETNP sample) by the number of good quality reads in a sample, keeping the data comparable across all publications from the Fuchsman lab. The read counts were multiplied by the sample normalization factor, divided by the gene length, and then multiplied by 100 to make visualization easier. We refer to these numbers as normalized reads.

Additionally, to give context, we include phylogenetic read placement data for photosystem II D2 protein (*psbD*) to examine both eukaryotic and bacterial photosynthesizers, from the Picocyanobacteria Internal Transcribed Spacer region (ITS) to examine Picocyanobacteria ecotypes, and SAR11 ecotypes from RNA polymerase (*rpoB*). Most but not all of these data have been published elsewhere: *psbD* for GA02 Stations 15-27 and GA03 Stations

14-20 are published in Fuchsman and Hays (2023), ITS data for GA03 is all in Fuchsman et al (2023) and for GA02 Stations 15-27 is in Fuchsman and Hays (2023). SAR11 ecotypes from GA02 Stations 10-21 are from Hays and Fuchsman (submitted). New data were calculated exactly as in the original papers. *psbD* is not a single copy core gene. *Synechococcus* has two copies of *psbD* (Golden et al., 1989) and so abundances were divided by 2 to match *Prochlorococcus*, which has one copy of *psbD*. It is unknown how many copies are in the eukaryotic algae so no corrections were made. The uncultured *Prochlorococcus* ITS group NC1, also known as LLVII, was not present on the *psbD* tree (Fuchsman and Hays, 2023).

Results

We were able to examine the presence and distribution of the genes *cynS* and *ureC* across two transects in the North Atlantic. The first analyzed transect was the spring GA02 transect (Fig. 2.1), which was oriented North-South. This transect was composed of 11 stations that only encompassed the top 200 m of the water column. The second transect we studied was the fall GA03 transect (Fig. 2.2), this transect was composed of 13 stations oriented West-East.

Nutrients in the GA02 transect

In the GA02 transect, nitrate was generally absent in ~ top 40 m for ST 17-ST27 (Fig. 2.3 a). However, for ST10-15, nitrate concentrations were low but measurable (0.2 - 0.9 µmol/kg) in surface waters. Along the transect, nitrate was more prevalent in ST10-ST12 (the northernmost stations), but only below ~50 m depth; below 50 m in ST10-ST12 nitrate concentration was ≤4.6 µmol/kg; the highest was found at 200 m depth at ST10 with 16.90 µmol/kg. Along ST15-ST21, the nitrate contraction increased from low numbers in the surface to ~3 µmol/kg 100 m depth. In the Southernmost stations, ST23- ST27, nitrate was found only at ~200 m depth with concentrations < 2.8 µmol/kg. Nitrite concentrations along the transect were low <0.4 µmol/kg (Fig. 2.3 b), and it was absent in the top 40 m but increased to a wide maximum from 50-100 m between Station ST10 to ST18; the highest concentration was 0.38 µmol/kg in ST11 at 75 m depth. While nitrite concentrations were much lower in the Southern stations, we observed a peak in ST21 of 0.22 µmol/kg at 152 m depth.

The temperature in the GA02 transect showed a gradient from cold to warmer waters. In ST10, the temperature was <11ºC throughout the euphotic zone (Fig. 2.3 c); in ST11-ST12, the temperature increased to ~16ºC. Across stations ST15-ST18, the temperature increased to ~19ºC. However, between ST21-ST27, stratification occurred, with larger temperature in the surface than at depth. In ST21, the temperature increased to 23.7ºC at 10 m depth but remained ~18ºC from 75 m depth. In ST23-ST27, there was a region of ~23ºC at 100 m depth, but it decreased at ~200 m.

Phytoplankton in the GA02 transect

Photosystem II gene psbD is found in both eukaryotic algae and Picocyanobacteria. From *psbD* data, eukaryotic algae, specifically from the Stramenopile group and green algae Chlorophyta, were abundant from ST 11-ST15, but at ST10 there was a mix of eukaryotic algae and *Synechococcus* at low abundances with no one group dominant (Supplemental Figure 8). When ITS and photosynthesis gene *psbD* data were examined, Picocyanobacteria at ST1-ST17 were dominated by *Synechococcus*, but at ST18 *Prochlorococcus* became more abundant (Supplemental Figure 8 and 10).

Eukaryotic algae

In the GA02 transect, we observed the presence or *cynS* normalized read that corresponded to Algae (Supplementary Fig. 2.1 a). Algae *cynS* were found in all stations, but in small abundance (< 5 normalized reads). These reads could not be normalized to RNA polymerase (*rpoB*) since our *rpoB* numbers only included bacteria and archaea. However, published *psbD* normalized reads indicated an abundance of eukaryotic algae at the Northern Stations (ST11-15) (Fuchsman & Hays, 2023) (Supplemental figure 2.4) (Supplemental figure 2.4). This implies that few of these eukaryotic algae had the ability to use cyanate.

Total *ureC* and *cynS* to *rpoB* gene ratios in the North-South transect (GA02)

We calculated the abundance of microbes with the *ureC* gene divided by the abundance of the total microbial prokaryotic community (Bacteria *rpoB* plus and Archaea *rpoB*) from *rpoB* to obtain the proportion of the total prokaryotic community containing *ureC*. Along the GA02 transect, these proportions changed over depth, within each station, and over space, across the whole transect. Although we identified several microbial groups containing the *ureC* gene, only a couple of them were present in high abundance across the GA02 transect: Picocyanobacteria, SAR11, other Alphaproteobacteria, and Thaumarchaeota. In ST10 to ST16, ~ 50% or lower of the microbial prokaryotic community (Bacteria and Archaea) have the *ureC* gene (Fig. 2.5 a). In ST18-ST25, the abundance of the *ureC* gene in the microbial community increased, ranging between 50-80%. In the last station of the transect ST27 the abundance of the *ureC* gene in the microbial community increased in abundance, varying between 77-177% of the prokaryotic community.

Spatial variation of the *cynS* ratio to the community *rpoB* in the North-South transect (GA02)

The *cynS* of the microbial prokaryotic community (Bacteria and Archaea) in the GA02 transect was only detected in to a small number of microbial taxa; *Synechococcus, Prochlorococcus,* Nitrospina, and unknown (Unclassified) bacteria. Across the transect the abundance of *cynS* in the prokaryotic community was ~15% (Fig. 2.6 a). The microbial taxa that contributed to most of the *cynS* abundance were *Synechococcus,* and *Prochlorococcus,* with a small number of *cynS* from *Nitrospina* and Unknown bacteria (Fig. 2.6 a).

The *Synechococcus cynS* abundance was highest (~14% of the prokaryotic community) in ST10 to ST17 in the top 10 m, decreasing with depth. In contrast in ST15 the abundance of *cynS* at 10 m depth was <5%. However, as the transect moved south (ST18-ST27), *Synechococcus cynS* abundance decreased in the surface (10 m depth) at the same time *Prochlorococcus cynS* started to increase in abundance at deeper depths. At ST18, the highest *Synechococcus cynS* abundance was 3.8% at 51 m depth,

in the following stations (ST21-ST27) the *Synechococcus cynS* abundance was <3%. *Prochlorococcus cynS*, on the contrary, was almost absent on the first half of the transect (ST10-ST17). In ST21 the *Prochlorococcus cynS* abundance reached 9.7% at 75 m. At station 25, the abundance of *Prochlorococcus cynS* reached 12.7% at 119 m depth, but at station 27 the *Prochlorococcus cynS* abundance decreased to 5.7% at 12 m depth. *Nitrospina cynS* was present, but corresponded to a low number in the community <1.6%.

Nitrospina *cynS* and *ureC*

In the GA02 transect we identified *Nitrospina* using the housekeeping gene *rpoB.* The Nitrospina abundance corresponded to ~3% of the total prokaryotic community, across all depths and stations (Fig, B3 a). *Nitrospina* abundance was higher (~3%) in the North at stations ST10-ST17, and within each station, the abundance increased with depth from <1% to 3%. *Nitrospina rpoB* reads were absent in the top 150 m in ST18, and ST23, below 150 m *Nitrospina* abundance decreased to ~2%, and it was almost absent in the top 200 m examined at the last stations (ST25 and ST27).

The gene *nxrB* that encodes for the enzyme nitrite oxidoreductase, was used as a marker gene for Nitrospina, and calculated out of *rpoB.* The *nxrB* varied between 100-200% at all depths (Fig. 2.7 b). Nitrospina has both genes *ureC* and *cynS* (Fig. 2.7 b). In the majority of the stations in the GA02 transect, the Nitrospina *ureC* was higher and varied between 62.5-239% of Nitrospina having the gene, almost double that of *cynS,* whose abundance varied between 29.4-131.8%, with the exception of ST16. At ST16, at 74 m the *cynS* abundance was 81%, whereas *ureC* was 74.7%, and at 199 m the *ureC/rpoB* ratio was 88.6% whereas *cynS* was 85.2%.

SAR11

In the GA02 transect we identified the SAR11 group by using the *rpoB* housekeeping gene. At ST10, the temperature was low (9.4- 11.4ºC) (Supplementary B0.2.1) and the main SAR11 *rpoB* ecotype was the cold adapted 1a.1 clade (Supplementary figure 2.5). Starting at St 11, 1a.3 clade became dominant in the *rpoB* ecotype data. SAR11 clade IIA.B appeared in surface waters starting at St 15. The SAR11 V ecotype switched from V.y to V.x in the surface waters at station 21 (Supplemental figure 2.5). At the same time, SAR11 *rpoB* ecotype 1a.3 became a smaller proportion of the SAR11 community (Supplemental figure 2.5).

The abundance of SAR11 *rpoB* varied across the transect. We observed that the abundance of SAR11 *rpoB* was <40% of the prokaryotic community across stations ST10 to ST18, but in ST21 the SAR11 *rpoB* abundance increased to 50-80% (Fig. 2.8 a). There was some variation between stations: From ST10 to ST12, the abundance of SAR11 *rpoB* seems to increase towards depth, like in ST12, SAR11 *rpoB* abundance increased from 21% at 9 m to 49% at 101 m depth. The ST21 to ST27 displayed a similar profile, where the SAR11 *rpoB* abundance was higher at the surface

and decreased with depth (Fig. 2.8 a), for example in ST21 the SAR11 *rpoB* abundance at 10 m depth was 60.5% at 10 m decreasing to 31.7% at 202 m depth.

SAR11 only has *ureC*, and not *cynS.* In the ST10, the abundance of total SAR11 *ureC* abundance was very low <3% (Fig. 2.8 b). In the ST11, the abundance of SAR11 with *ureC* was 10.8-16.9% of SAR11, and in ST12 13-14.3% of SAR11 had *ureC*. However, from ST12 onwards we observed more variation of the SAR11 *ureC* within depth and across stations (10.1-48.7%).

We observed that *ureC* reads could be assigned to different SAR11 subgroups, some of which could be related to actual SAR11 ecotypes: Subgroup 1a.3, Subgroup ODZ, Subgroup IV, Subgroup V and two unidentified subgroups (I, II), similar to the groups we identified in Chapter 1. Among SAR11 we observed the different groups contributing to the total SAR11 *ureC*. We observed the different groups have a distinct distribution across the transect. In ST10, the abundance of all *ureC* subgroups was very low <1.1% (Fig. 2.8 b). Across ST11 to ST18, subgroup 1a.3 *ureC* has the highest abundance. For example, in the ST11, the subgroup SAR11 1a.3 abundances varied between 18.4-26.6% of the total SAR11 *ureC*. However, in ST15 the subgroup 1a.3 *ureC* varied between 10.4-15.7% of the total SAR11 *ureC*. But in ST21, there was a decrease of SAR11 1.a3 abundance with it composing only 1.9% of the total SAR11 *ureC* at 10 m depth, but that number increased to 20.2% at 75 m depth.

The SAR11 group V *ureC* abundance was low <7.2% across ST10 to ST18, and found mostly between ~50-200m depth from ST10 to ST18 (Fig. 2.8 b). The SAR11 subgroup V abundances increased from ST21 to ST23, from the surface towards depth, where for example at ST21 the *ureC* abundance at 10 m reached 7.5% of the total SAR11 *ureC* and continued increasing as the transect moved south. The abundance of SAR11 subgroup V reached 30.5% of the total SAR11 *ureC* at 12 m depth. The SAR11 group IV was only present towards the end of the transect at ST23 to ST27, with abundances with 5.6% at 77 m depth and 10.3 and 10.4% at 75 and 100 m depth.

Thaumarchaeota *ureC*

In the transect GA02, the Thaumarchaeota abundance detected with the *rpoB* gene was observed only below ~50 m depth. The Thaumarchaeota *rpoB* abundance increased with depth, reaching a maximum abundance of 21% of the total microbial community in ST10 (Fig. 2.9 a). The depth profile and Thaumarchaeota *rpoB* abundances were generally similar in all the stations. Thaumarchaeota only has *ureC*; in the transect GA02, Thaumarchaeota *ureC* was observed at all the depths where Thaumarchaeota was found (Fig. 2.9 b). The abundance of Thaumarchaeota *ureC* across all stations varied between 40.7% to 237.5%.

Picocyanobacteria *cynS* and *ureC*

In the GA02 transect, Picocyanobacteria were found at the surface in all the stations present in the transect (Fig. 2.10 a). ST10 to ST17 shared a similar Picocyanobacteria *rpoB* depth profile, with higher abundance in the surface that decreased with depth, moreover, the proportion of Picocyanobacteria *rpoB* varied between stations. In ST10 for example, the highest Picocyanobacteria *rpoB* abundance was 12.4% of the prokaryotic community at 10 m depth, reaching 0.1% at 152 m depth. ST18 to ST23 had a similar Picocyanobacteria *rpoB* depth profile (Fig. 2.10 a) with lower Picocyanobacteria *rpoB* on the surface, which increased with depth; for example, at ST23, the abundance of Picocyanobacteria *rpoB* increased from 2.6% of the prokaryotic community at 8 m depth, reaching a peak of 15.6% at 87 m depth, decreasing to 11.9% at 150 m depth.

Picocyanobacteria *Prochlorococcus* and *Synechococcus,* have both genes *ureC* and *cynS*. Picocyanobacteria *ureC* read abundance varied between 100-150% of Picocyanobacteria in the top 150 m across the whole GA02 transect, except at ST27, where the Picocyanobacteria *ureC* abundance reached 287.3% (Fig. 2.10 b) at 75 m depth, which is equivalent to 3 copies of the *ureC* gene. Several *Prochlorococcus* ecotypes had *ureC* including High Light *Prochlorococcus* (HL), Low Light *Prochlorococcus* group I (LLI), Low Light *Prochlorococcus* group IV (LLIV), and an Uncharacterized *Prochlorococcus*.

The HL *Prochlorococcus ureC* depth profiles*,* changed across the GA02 transect. Across ST10-ST16, the HL *Prochlorococcus ureC* abundance was generally low (<16% of the total Picocyanobacteria *ureC*) from ~10 to 90 m depth (Fig. 2.10 c). HL *Prochlorococcus ureC* increased with depth towards ~100m. For example, in ST11, the abundance of HL *Prochlorococcus ureC* at 9 m was <1% of the total Picocyanobacteria *ureC,* however, HL *Prochlorococcus ureC* increased with depth reaching 60% of the total Picocyanobacteria *ureC* at 101 m depth. In ST18 we observed a change in the HL *Prochlorococcus ureC* depth profile, in previous stations the HL *Prochlorococcus ureC* abundance was low at the surface and increased with depth; In contrast, at ST18 the HL *Prochlorococcus ureC* abundance was higher in the surface and decreased with depth. ST18- ST27 shared a similar *ureC* depth profile. For example, at the ST25, at 11 m, the HL *Prochlorococcus ureC* abundance was 99.5% of the total Picocyanobacteria, decreased to 90.3% of the total Picocyanobacteria at 77 m, and then decreased to 11.8% at 119 m depth; there was a small increase to 20.7% of HL *Prochlorococcus ureC* abundance of the total Picocyanobacteria, at 152 m depth.

LLI *Prochlorococcus ureC* was generally low (<11.4% of the total Picocyanobacteria *ureC*) across stations ST10-ST17 (Fig. 2.10 c), with the exception of a peak of 26.7% of LLI *Prochlorococcus ureC* at 26.7% at 101 m at ST11. In ST17-ST21 the LLI *Prochlorococcus ureC* abundance was increased with depth. For example, in ST21 the LLI *Prochlorococcus ureC* abundance increased from 1.1% of the total Picocyanobacteria *ureC* at 10 m

depth to 89.1% at 100 m depth. LLIV *Prochlorococcus* was almost absent across the GA02 transect. It was present in some stations like ST16, ST23, and ST27, but was below 14% of the total Picocyanobacteria *ureC*. Across ST10-ST21, the highest abundance of Uncharacterized *Prochlorococcus ureC* was 12.8%. We observed an increase in the Uncharacterized *Prochlorococcus ureC* in ST23 and ST27 to 53% of the total Picocyanobacteria *ureC* at 150 m depth.

Synechococcus had a higher *ureC* abundance across ST10-ST17 (Fig. 2.10 c). Generally, *Synechococcus ureC* was higher in the surface and decreased towards depth. At these stations, *Synechococcus ureC* composed ~80-96% of the total Picocyanobacteria *ureC* at 10 m depth, and varied between 75-97% between 10-70 m, but below this depth the *Synechococcus ureC* decreased. However, at ST18, we observed a sharp decrease in the *Synechococcus ureC* abundance where it was only 34.5% out of the total Picocyanobacteria *ureC* at ~10 m depth; there was a small increase of 36.6% at 76 m depth in the ST18. *Synechococcus ureC* was very low at ST21 and ST23, with a peak of 12.4% of the total Picocyanobacteria *ureC* in the ST21 and a peak of 37.5% at ST23 at 8 m. *Synechococcus ureC* was <6% of the total Picocyanobacteria *ureC* in ST25-ST27.

In the GA02 transect, the total Picocyanobacteria *cynS* was generally higher on the surface and decreased with depth. We observed that the abundance of *cynS* varied within depth and across space along the transect. The proportion of Picocyanobacteria with *cynS* between ST10 to ST21 was 100-150%. However, in the ST23-ST27, the abundance of Picocyanobacteria *cynS* decreased to 50-100% (Fig. 2.10 d).

The GA02 transect was dominated by the abundance of *Synechococcus cynS* in the majority of the stations. From ST10 to ST18 the abundance of *Synechococcus cynS* corresponded to 50 to 100% of the total Picocyanobacteria *cynS* (Fig.2.10 b). The *Synechococcus cynS* abundance within each of these stations was fairly constant in number. At ST18, however, the *Synechococcus cynS* abundance decreased in proportion compared to the Northern stations; at ST18 the *Synechococcus cynS* varied between 58.6-87% of the total Picocyanobacteria *cynS.* In the following station (ST21), the abundance of *Synechococcus cynS* decreased and in the last station (St 27) was dominated by *Prochlorococcus cynS.*

Prochlorococcus cynS was almost absent between ST10-ST16 (<5% of the total Picocyanobacteria *cynS*). *Prochlorococcus cynS* increased at ST17 with a peak of 11% of the total Picocyanobacteria *cynS* at 100 m depth (Fig. 2.10 e). *Prochlorococcus cynS* continued increasing within depth at each station and across space as the GA02 transect moved South (ST17-ST27). *Prochlorococcus cynS* abundance was similar for the ST21-ST25. We observed that the *Prochlorococcus cynS* depth profiles had a lower abundance of ~10 m at the surface, increasing with depth. For example, in ST21, the *Prochlorococcus cynS* abundance increased from 2.5% of the total Picocyanobacteria *cynS* at 10 m depth to 82.3% of the total Picocyanobacteria *cynS* at 75 m depth, reaching 87.6% of the total

Picocyanobacteria *cynS* at 100 m depth. We observed the presence of *cynS* in other unclassified cyanobacteria in low abundance <22% in ST10-ST18 (Fig. 2.10 e). Thus, there was a transition between *Synechococcus* and *Prochlorococcus* around ST18 or StT21 in both the *ureC* and *cynS* data.

Nutrients and temperature in the GA03 transect

The GA03 transect traveled from the coast of North America (ST1), on a shelf slope transect (ST3-ST8), and then a transect across the subtropical gyre (ST10-20), which approached the coast of Africa (ST22-24, ST7) See Fig. 2.3. Stations across the entire GA03 transect were stratified by temperature. In the GA03 transect, the surface water was warm (24-25ºC) across the whole transect, but decreased with depth. As noted by Jenkins et al (2015), nutrient concentrations shoaled as the GA03 transect approached the African shelf, bringing higher concentrations of nitrate up into shallower water. Nitrate concentrations were unmeasurable in surface waters throughout the transect, except at coastal ST1 where nitrate was 0.6 µmol/kg. For ST3- ST20, nitrate concentrations were <6 µmol/kg in the top 300 m, but at ST22 and ST24, nitrate concentrations reached 20 µmol/kg and 26 µmol/kg. Nitrite maxima were 0.1 µmol/kg or below.

Phytoplankton in the GA03 transect

Besides coastal ST1, *Prochlorococcus* were the dominant phytoplankton on this transect, with eukaryotic algae barely measurable with *psbD* data (Supplemental Figure 7). However, *Synechococcus* were fairly abundant at St 1-St 8, but decreased in the abundance in the subtropical gyre; conversely, *Prochlorococcus* ecotype LLI was not present at St 1-St 8, but was abundant at the Deep Chlorophyll Maximum in the subtropical gyre in both ITS and *psbD* data (Fuchsman et al 2023) (Supplemental Figure 7).

In the GA03 transect we also observed the presence of *cynS* normalized reads that corresponded to Algae (Supplementary Fig. 1 b). However, we found Algal *cynS* in all stations of the transect, but in small abundance, <3 normalized reads.

Total *ureC* and *cynS* to *rpoB* gene ratios in the West-East transect (GA03)

In the GA03 transect, we observed that the abundance of the *ureC* gene in the microbial community varied across depth within stations, and across space along stations in the GA03 transect (Fig. 2.4 b). Similarly, to the GA02 transect, several microbial taxa containing the *ureC* gene were present, however, only a few of them were present in high abundance across the GA03 transect: Picocyanobacteria, Thaumarchaeota, SAR11, and other Alphaproteobacteria. The *ureC* gene abundance in the microbial community varied within depth per station and across the transect, which ranged between 20-119% with higher percentages in the upper water column (Fig. 2.4 b). The group that contributed the most to the *ureC* abundance was Picocyanobacteria that contributed up to 77.8% to the community *ureC*. When

Picocyanobacteria are removed from the graph, the abundance of the *ureC* gene in the community was 22-55% (Supplementary figure 2.2).

The Picocyanobacteria *ureC* was found mainly in the top 150 m in the water column (Fig. 2.4 b). In the ST1 to ST8, the shelf and slope stations, the distribution of the *ureC* gene was similar within those stations; the *ureC* gene abundance was higher in the surface ~30 m depth, and decreased with depth. At ST1, the most coastal station, only 5.9% of the prokaryotic community had Picocyanobacteria *ureC* at 30 m, but from ST3 to ST8 the Picocyanobacteria *ureC* abundance varied between 25-37.8% of the prokaryotic community across similar depths (Fig. 2.4 b). As the transect moved East, within ST10-ST20, the subtropical gyre stations, we observed an increase in the depth where the highest Picocyanobacteria *ureC* abundance was located, to 74-99 m. For example, in ST14, at 40 m the *ureC* maximum was 15.9% of the prokaryotic community, increasing with depth to a maximum of 68.9% at 99 m depth, and then decreasing with depth below 99 m. Towards the last stations in the transect (ST22, ST24) we observed the highest Picocyanobacteria *ureC* gene abundance at ~50 m depth, where for example, in ST22 the *ureC* gene abundance was 77.8% of the prokaryotic community at 50 m depth. At the coastal ST7, the highest Picocyanobacteria *ureC* gene abundance was at 33 m (53.7%) and decreased with depth.

In contrast, Thaumarchaeota *ureC* gene was found below ~80 m depth, with abundances that increased within depth at each station. In general, the abundance of Thaumarchaeota *ureC* gene within ST1 to ST24 and the coastal ST7 varied between 0.1- 29.3% (Fig. 2.4 b). The SAR11 *ureC* corresponded to 12% of the *ureC* genes in the microbial community in the GA03 transect. In general, the abundance of SAR11 *ureC* was higher between 30- 100 m, and it decreased with depth (Fig. 2.1 b). As we observed in the GA02 transect, in the GA03 transect the microbial prokaryotic community (Bacteria and Archaea), *cynS* in the GA03 transect was limited to a small number of microbial taxa; *Synechococcus, Prochlorococcus, Nitrospina*, and unknown (Unclassified) bacteria. Across all the stations in the GA03 transect, the *cynS* was ~12% of the total microbial community determined from *rpoB* (Fig. 2.5 b). Similar to GA02, the microbial taxa that contributed the most to the *cynS* in the community *Synechococcus,* and *Prochlorococcus,* with small numbers of *cynS* reads from *Nitrospina* and Unknown bacteria.

Across the GA03 transect we observed that in ST1-ST8, the Picocyanobacteria *cynS,* although low (~5% of the total microbial community in surface waters). *Synechococcus cynS* was 2- 4% of the community at St 3-St10 but decreased from ST14 onwards East. From ST18 towards the end of the GA03 transect the *Synechococcus cynS* was <1% of the total microbial community. In contrast, the *Prochlorococcus cynS* increased from ST14 onwards as the transect moved East; for example, at ST14, the *Prochlorococcus cynS* abundance was 8.2% of the total microbial community at 69 m (Fig. 2.5 b).

Nitrospina

In the GA03 transect, *Nitrospina* abundance estimated with the *rpoB* gene ranged between 1- 3% of the total prokaryotic community at all analyzed depth (Fig. 2.11 a). The abundance of the nitrite oxidoreductase marker gene (*nxrB*) varied between 24.3- 212.2%. *Nitrospina* has both genes *ureC* and *cynS*, and in all the stations, the proportion of Nitrospina with *ureC* was higher than for *cynS* (Fig. 2.11 b); *ureC* abundance varied between 75.3 and 178.4% of *Nitrospina*, whereas *cynS* varied between 1.2 to 103.9% of the *Nitrospina* community. In the majority of stations *ureC* abundance was higher than *cynS*, however, an exception was found at ST10 at 414 m depth *cynS* was 103.9%, whereas *ureC* was only 80.4%. In ST18, at 113 m depth and in ST20 at 99 m depth no *Nitrospina cynS* was found.

SAR11

The abundance of SAR11, estimated with the *rpoB* gene, showed that SAR11 was widely distributed along the GA03 transect (Fig. 2.12 a). The SAR11 *rpoB* depth profile was very similar across stations in the GA03 transect. In general, the abundances of SAR11 was higher at the surface and decreased towards depth. SAR11 abundance varied between 13.9 and 48.8% of the prokaryotic community. The highest abundance of SAR11 *rpoB* was found at ST16, where SAR11 *rpoB* was 48.6% of the prokaryotic community at 39 m depth, decreasing to 25.4% at 284 m depth. The lowest SAR11 abundance was at ST22, where the highest SAR11 *rpoB* abundance was 30.5% at 50 m depth, decreasing to 13.9% of SAR11 *rpoB* at 289 m depth.

SAR11 only had *ureC*, and not *cynS.* We examined the abundance of the total SAR11 *ureC* along the GA03 transect. In general, all stations in the GA03 transect shared a similar depth profile, with higher abundance ~30 m depth, and decreasing abundance with depth. In general, the proportion of SAR11 with *ureC* was between 20-30% at ~30 m depth, decreasing with depth to <5% below ~120 m (Fig. 2.12 b).

In the GA03 transect the *ureC* reads could be assigned to different SAR11 subgroups: Subgroup 1a.3, Subgroup ODZ, Subgroup IV, Subgroup V and two unidentified subgroups (I, II), identical to the groups we identified in Chapter 1. In the GA03 transect, the *ureC* subgroup V and subgroup 1a.3 were dominant across the whole transect (Fig. 2.12 c); SAR11 *ureC* subgroup 1a.3 was present in surface waters, with abundance decreasing towards depth, and the SAR11 subgroup V with an opposite depth profile with lower *ureC* abundance in the surface, increasing gradually with depth. The SAR11 subgroup 1a.3, has the highest *ureC* abundance, that generally were higher at ~30 m depth, and decreased with depth. The SAR11 subgroup 1a.3, abundance varied between 3.7% to 68.7% of the total SAR11 *ureC* (Fig. 2.12 c). The SAR11 *ureC* subgroup V had similar depth profiles between ST1-ST10, with low abundance at the surface that increased with depth. For example, in ST4, the abundance of SAR11 subgroup V *ureC* increased with the surface from 19.7% at 39 m depth to 51% of the total SAR11 *ureC* at 285

m depth. Stations ST14 to ST20, shared similar depth profiles where the SAR11 *ureC* subgroup V abundance decreased from 59.1% of the total SAR11 *ureC* at 40 m depth to 37.7% at 99 m depth and then increased with depth until 65.1% of the total SAR11 *ureC* at 285 m depth.

Interestingly, despite being in oxic waters, the SAR11 ODZ *ureC* was found in low numbers in the majority of the transect <5.7% of the total SAR11 *ureC* (Fig. 2.12 c), in ST4 to ST16. However, in ST1 and ST3, the SAR11 ODZ *ureC* was relatively absent in the top 30~100 m depth, increasing in abundance to ~130 m depth, where SAR11 ODZ *ureC* was found at 13.6% of the total SAR11 *ureC* at 185 m depth. Meanwhile, in ST3, the SAR11 ODZ *ureC* abundance increased from 3.3% at a depth of 135 m to 24.2% at a depth of 288 m. In ST18 and 20, the SAR11 *ureC* was present with a peak of 13% of total SAR11 *ureC* at 300 m depth. ST22, ST24 and ST7 also have SAR11 ODZ *ureC*. They have similar depth profiles, where the SAR11 ODZ is almost absent in the 100m and increasing abundance towards depth. In the ST22-ST7 station, the highest SAR11 ODZ *ureC* abundance was 27.3% of the total SAR11 *ureC* at a depth of 285 m.

Thaumarchaeota

Thaumarchaeota presence was examined by observing the variation of Thaumarchaeota *rpoB,* across the GA03 transect. We observed that Thaumarchaeota *rpoB* abundance was generally constant. All stations shared a similar depth profile, where Thaumarchaeota appeared below or close to 50 m depth, with increasing *rpoB* towards depth. Thaumarchaeota corresponded to ~20% of the microbial community (Fig. 2.13 a) examined with the *rpoB* gene. However, on the Easternmost part of the transect (ST7), Thaumarchaeota was found at 30% of the community a few depths below 110 m.

Thaumarchaeota had *ureC* but no *cynS.* Thaumarchaeota *ureC* was observed at every station in the GA03 transect. Generally, the proportion of Thaumarchaeota with *ureC* varied between 29.4% to 114.5%, with an exception of 3% of Thaumarchaeota *ureC* at 99 m depth at ST14 (Fig. 2.13 b). However, in the deep water, the proportions were more constant, varying between 75-114% of Thaumarchaeota.

Picocyanobacteria

The abundance of Picocyanobacteria observed with the variation of the house keeping *rpoB* gene, (Fig. 2.14 a). In general, we observed that Picocyanobacteria *rpoB* abundance decreased with depth. In the ST1 the abundance of Picocyanobacteria *rpoB* was lower than 5% of the total microbial community (Fig. 2.14 a). Across ST3-ST10 the abundance of Picocyanobacteria *rpoB* was ~20% of the total microbial community. Between ST14- ST20 the abundance of Picocyanobacteria *rpoB* at ~50 m depth was <7.3%, and increased to~ 25.6% between 60 - 90 m depth, and then decreased with depth. The Station ST22, ST24 and ST7, shared similar *rpoB*

depth profiles, with higher abundances ~30% of the total microbial community at ~50 m depth, decreasing with depth to <10.7%.

In the GA03 transect, depth profiles of the proportion of Picocyanobacteria containing *ureC* along the whole GA03 transect was generally ~100% (Fig. 2.14 b). Across the GA03 transect the HL *Prochlorococcus ureC* was present in all the stations from ST1 to ST24 and ST7 (Fig. 2.14 c). The HL *Prochlorococcus ureC* abundance was higher between 20-100 m depth and varied between 3.3- 97.5% of the total of Picocyanobacteria *ureC*. The HL *Prochlorococcus ureC* abundance decreased from approximately 60 m towards depth. The highest abundance of the HL *Prochlorococcus ureC* was at ST18 to ST22 and ST7, where HL *Prochlorococcus ureC was >95%* of the total Picocyanobacteria *ureC.*

LLI *Prochlorococcus ureC* was present in all stations, but the *ureC* abundance changed as the transect moved West-East. Generally, the abundance was low between 20~50 m depth, but increased towards depth. We noticed the abundance of LLI *Prochlorococcus ureC* was <6.5% of the total Picocyanobacteria in ST1-ST6, but it had a peak of 22% of Picocyanobacteria *ureC* at 135 m depth at ST4 (Fig.2.14 c). From ST8 towards the end of the transect we observed the peak of LLI *Prochlorococcus ureC* moving up as the transect moved east. For example, in ST16 the highest abundance of LLI *Prochlorococcus ureC* was 67% of the total Picocyanobacteria *ureC* at 110 m depth, but at ST20 the LLI *Prochlorococcus ureC* was 70.6% at 99 m depth. AT ST22 was 59% of the total Picocyanobacteria *ureC* at 84 m depth, in ST24 it was 61.7% of the total Picocyanobacteria *ureC* at 71 m depth. However, the depth profile changed completely in ST7, which is on the coast of Africa, where the highest LLI *Prochlorococcus ureC* abundance was at 40.5% of the total Picocyanobacteria *ureC* at 151 m depth. LLIV *Prochlorococcus ureC* had a low abundance with a small increase towards depth. Generally, the LLIV *Prochlorococcus ureC* was low in the whole transect, reaching ~ 20% of the total Picocyanobacteria *ureC*, and it was higher only between ST1-ST6. Uncharacterized *Prochlorococcus* was highly abundant across the GA03 transect, 0.4- 85.7%, but it was almost absent in ST7.

Synechococcus ureC was found in all stations in the GA03 transect, however, the abundance was higher in the western stations. In ST1-ST6, the abundance of *Synechococcus ureC* was higher at ~20 m depth and decreased with depth. For example, in the ST4 the highest abundance was 24.4% of the total Picocyanobacteria *ureC* at 39 m (Fig.2.14 c) depth and continued decreasing with depth, to 3.4% at 184 m depth.

Prochlorococcus and *Synechococcus* also contained *cynS* in addition to *ureC*. In the GA03 transect, we also estimated the abundance of total *cynS*; in contrast to the *ureC* read abundance, the total Picocyanobacteria *cynS* was generally found in <30% of Picocyanobacteria, only reaching 60% at two depths (Fig. 2.14 d). *Synechococcus cynS* was present across the whole GA03 transect and varied between 25- 92.5% of the total *Picocyanobacteria cynS* (Fig. 2.14 e) in the ST1-ST8. Within *Synechococcus*

cynS abundance was higher in the surface and decreased with depth. However, *Synechococcus cynS* decreased sharply from ST18 towards the station located to the East of the transect. *Prochlorococcus cynS* was found along the entire transect. (Fig. 2.14 e) with high abundances. The *Prochlorococcus cynS* had different depth profiles shape. The lowest *Prochlorococcus cynS* abundance was found in ST1, (<12.5% of the total *Picocyanobacteria cynS*). In other stations in the GA03 transect the *Prochlorococcus cynS* abundance varied between 15.1 to 100% of the total Picocyanobacteria *cynS.*

Discussion

N is a limiting nutrient for much of the oligotrophic ocean (Moore et al., 2013). Many microorganisms prefer reduced nitrogen forms for assimilation (Glibert et al., 2016). However, ammonium concentrations are extremely low in the oxic oligotrophic ocean (Martens-Habbena et al., 2009), and nitrate concentrations are often unmeasurable in surface waters (Moore et al., 2013). One adaptation to ammonium limitation is for microbes to use small organic reduced N sources, such as urea and cyanate. Urea and cyanate concentrations are also generally in the nanomolar range (Kitzinger et al., 2019; Painter et al., 2008; Takeda et al., 2020; Widner, Fuchsman, et al., 2018). However, the ability to use urea or cyanate increases microbes' chances to obtain reduced N or in areas with limited nitrate, any N containing compound. By analyzing the abundance of normalized reads of the genes for cyanase (*cynS*) and urease (*ureC*) for individual taxa and comparing to single copy core gene RNA polymerase (*rpoB*), we find important information both about adaptations of individual taxa to N limitation and about the importance of urea and cyanate in the North Atlantic.

Phytoplankton succession in the spring transect

In the spring transect GA02, with the *rpoB* gene, we observed that Picocyanobacteria, between stations ST1-ST17, had a higher abundance ~10 m depth, and decreased to ~50 m, this profile was similar for all the stations in that transect, there was a change from ST17 to ST18, the depth of maximal abundance increased from ~10 m to ~100 m (Figure 2.9 a). Additionally, when ITS and photosynthesis gene *psbD* data were examined, Picocyanobacteria at ST1-ST17 were dominated by *Synechococcus*, but at ST18 *Prochlorococcus* became more abundant (Supplemental Figure 2.4 and 2.6). When we looked at the proportion of *ureC* and *cynS* genes, we observed that ST1 to ST17 were dominated by *Synechococcus* which had a high abundance of both; 50-90% of the Picocyanobacteria *ureC* corresponded to *Synechococcus* alone. Additionally, ~100% of all the cyanobacteria community *cynS* corresponded to *Synechococcus* (Figure 2.9 c). We identified different *Prochlorococcus* ecotypes based on their different *ureC,* but not all of them were present in the same proportion of total Picocyanobacteria *ureC*. For example, we observed the HL *Prochlorococcus* ecotype present peaking at ~50% at depths, where the *Synechococcus* abundance decreased, but when HL *Prochlorococcus ureC* peaked, the *Prochlorococcus cynS* was almost absent (Figure 2.9 b). The succession of Picocyanobacteria from *Synechococcus* to *Prochlorococcus* in the ITS and psbD profiles occurred at ST17- ST18, however, changes in the abundance and distribution of the different cyanobacteria *ureC* started at ST15, where at ~100 m depth HL *Prochlorococcus ureC* was ~20% at ST15, but ~40% of the total cyanobacteria community *ureC* at St 16. HL *Prochlorococcus* became a larger proportion of Picocyanobacteria *ureC* than *Synechococcus* starting at ST18 (Figure 2.9 b). The changes of *ureC* of the different ecotypes matched the changes in the *cynS* depth profile, for example in ST18 and ST21 the

cynS peak (~50 at 80 m and 80% at 100 m respectively) matched the depth were the LLI *Prochlorococcus ureC* peaked (~0% at 80 m and 90% at 100m) (Figure B6), indicating that LLI *Prochlorococcus* might be the ecotype with *cynS*. Towards the last 3 stations of the transect, HL *Prochlorococcus ureC* was dominant, at the surface (90%), whereas LLI *Prochlorococcus* peaked at ~120 with 80% in ST25 (Fig. 2.10). This *ureC* data was consistent with ecotype data from *psbD* and ITS (Supplemental Figure 2.4 and 2.6). Thus, much of the changes in Picocyanobacteria *ureC* and *cynS* phylotypes are likely related to changes in ecotype profiles in the water column.

In the GA03 transect, *Prochlorococcus cynS* was found along the whole transect, with abundances ranging from 25- 92.5% of the total Picocyanobacteria *cynS*. In contrast, in the GA02 transect, *Prochlorococcus cynS* was almost absent (<10% of the total Picocyanobacteria *cynS* in ST10-16. One of the elements that could be influencing the distribution of *Prochlorococcus* could be related to temperature. In the GA03 transect, the water at the surface was warmer 24-25ºC in the upper 100 m (Supplementary figure 2.4) across the whole transect. In the GA02 transect, we observed a gradient from cold waters in the northern part of the transect (ST10). Towards the end of the transect, the water at the surface of ST23-27 was warmer.

Phytoplankton such as *Synechococcus*, *Prochlorococcus*, and Eukaryotic algae all have physiological responses to the addition of cyanate. Sato et al., (2022) found a high abundance of eukaryotic algae *cynS* and *ureC* transcripts in the North Atlantic; in contrast, in our data the number of Eukaryotic algae *cynS* was extremely low (Supplementary figure 2.1). The >0.8 um size fractionation examined in Mao et al (2022) allowed them to focus on eukaryotes. We examined bulk samples dominated by bacteria/archaea. Additionally, the time of year varied: Mao et al. (2022) sampled in February, GA02 was in May, and GA03 was in November, and the North Atlantic changes substantially between seasons. We barely found Eukaryotic *cynS* in the Geotraces transects we analyzed. This could be because nitrate was measurable in surface waters from ST10-ST15 (0.2 - 0.9 µmol/kg). These are the same stations where eukaryotic algae were found, so it is possible that utilization of cyanate was not necessary. At other stations in the GA02 transect and in the GA03 transect, amounts of eukaryotic algae from *psbD* data were low.

Thaumarchaeota has a ubiquitous presence in the North Atlantic

Thaumarchaeota abundance based on the *rpoB* gene was fairly constant ~20%, only present below ~50 m depth. The Thaumarchaeota abundance was similar for both GA02 and GA03 transects. The Thaumarchaeota *ureC* read abundance was also similar to both transects at ~75-100%. The Coastal ST7 had a slight increase of *rpoB* to 30% at ~350 m depth. Thaumarchaeota has a ubiquitous presence in the North Atlantic, regardless of the season: Spring versus Fall, or the spatial orientation North-South, West-East. The presence of *ureC* is widespread in Thaumarchaeota genomes. It was present in 60-100% of Thaumarchaeota at coastal time

series SPOT (Ahlgren et al., 2017). In the ETNP, Thaumarchaeota had one copy of *ureC* per genome (Widner, Fuchsman, et al., 2018). However, in the Gulf of Mexico, only between 10-15% of Thaumarchaeota cells contain an *ureC* gene (Kitzinger et al., 2019). In Chapter 1, we found that the proportion of Thaumarchaeota with *ureC* was reduced at 1000 m and 4000 m in the North Pacific. Therefore, it appears that only some populations of Thaumarchaeota have *ureC*, but our data indicates that the populations in the lower euphotic zone and mesopelagic open ocean do have this gene.

Nitrospina

Nitrospina abundances detected with the *rpoB* gene showed that in the GA02 transect, the *Nitrospina* comprised 1-3% (Fig, 2.7 a) of the microbial community, but the abundance was not constant across the North Atlantic, because, as the transect moved south, the abundance of Nitrospina decreased, and was absent in the last two stations. The GA02 transect only collected metagenomes in the top 200 m, so the profile of Nitrospina may have deepened below sampled depths. In the GA03 transect, overall, the abundance of Nitrospina increased with depth from ~1 to 3% of the total bacteria community. Since *Nitrospina* is thought to be the main nitrite oxidizer in the open ocean (Pachiadaki et al., 2017), the differences in these distributions across geographical regions could have an impact on the nitrogen cycle. Similar to our data from HOT and the ODZs in Chapter 1, in both Geotraces transects, the proportion of *Nitrospina* with the *ureC* gene proportion was higher than *cynS.*

SAR11 succession across seasons in the North Atlantic

In the GA02 transect, at ST10, the Northern station, the proportion of total SAR11 with *ureC* was very low <3% (Fig. 2.8 b). In the rest of the GA02 transect, we observed variation of the SAR11 with *ureC* between 10.1-48.7% of SAR11. In general, all stations in the GA03 transect shared a similar depth profile, with the highest proportion of SAR11 with *ureC* at ~30 m depth (20-30%), and decreasing abundance with depth. These profiles are quite similar to what we observed at HOT in Chapter 1.

(Bolaños et al., 2022) analyzed the variation of 16S SAR11 ASV to determine changes in SAR11 ecotypes during a span of 27 years at BATS in the North Atlantic subtropical gyre; they found that SAR11 ecotype 1a.3 was the dominant ecotype with a higher dominance in summer, whereas towards winter, other ecotypes were prevalent (Ecotypes Ib, IIa.A). Similarly, we also observed a succession when we analyzed the variation of the *ureC* gene of the SAR11 ecotypes in the spring GA02 transect. At ST10, SAR11 was detected using *rpoB,* however, the abundance of SAR11 *ureC* was almost nonexistent. At that station, the temperature was lower (9.4- 11.4 ºC) (Supplementary Figure 2.3) and the main SAR11 *rpoB* ecotype was the cold adapted 1a.1 clade (Supplementary Figure 2.5). The lower temperature and ecotype change could be an explanation for the lack of SAR11 *ureC* at this station. From ST11 to ST21 (BATS), the ecotype 1a.3 had the highest

number of *ureC* reads out of the SAR11 community ~50-70% (Fig. 2.8), whereas at the same stations the abundance of the other ecotypes was <50 %. The abundance of the subgroup V analyzed by the *ureC* gene, from ST10 to ST21 increased with depth from ~10% of SAR11 *ureC* at ~10 m depth to ~40% at ~200 m depth. In contrast, from S21 to ST27 the abundance of subgroup V was higher (~70% of SAR11 *ureC*) in the surface and decreased with depth to ~40% ~200 m depth. The Ecotype 1a.3 had a low *ureC* abundance in the surface, at the same depth, where the Subgroup V was dominant. The change in the dominant SAR11 *ureC* ecotype towards the southern stations in the GA02 transect could be explained by the change in temperature in the surface waters. Towards the last stations (ST21-ST27) there was a change in the water temperature, where temperature was higher in the surface (Supplementary figure 2.3). The changes in the SAR11 ecotype distribution can be attributed to changes in the environmental conditions such as temperature, or nutrients (Bolaños et al., 2020). In our samples at ST21, the ecotype 1a.3 was almost absent in the surface at 75 m (Fig. 2.7 c), at the same depth the temperature was higher (23.7 ºC). The 1a.3 ecotype reappeared at 75 m depth, which coincided with an area where the temperature decreased to 18.7 ºC (Supplementary figure 2.2). Similarly, the SAR11 subgroup V increased its abundance up to 70% (based on the *ureC* abundance) in ST27 at 12 m the temperature rose to 28.4ºC. Additionally, in the *rpoB* ecotype data, the SAR11 V ecotype switched from V.y to V.x in the surface waters at these same stations (Supplemental figure 2.5). At the same time, SAR11 *rpoB* ecotype 1a.3 became a smaller proportion of the SAR11 community (Supplemental figure 2.5). However, in the GA03 transect, SAR11 *ureC* subgroup profiles were all very similar to each other and also to the southern station of GA02. Thus, temperature, ecotype and the ability to use urea are all intertwined.

Comparison between transects at BATS

The GA02 and GA03 transects corresponded to large areas in the North Atlantic, however they have only one station in common, the BATs station that corresponded to ST21 in GA02 (Fig. 2.1), and ST10 in GA03 (Fig. 2.2). This means that ST21 (from spring transect GA02) and ST10 (from fall transect GA03), can be compared directly. The BATs stations have been described as seasonally stratified (Landolfi et al., 2016). Nutrient concentrations were different at spring and fall timepoints. For example, atST21, the nitrate concentration increased from undetectable at the surface to 0.2 µmol/kg at 75 m and ~3 µmol/kg at 200 m depth, whereas a nitrite peak was observed at ST21 of 0.22 µmol/kg at 152 m depth. In the case of temperature, in the ST21 the temperature on the surface was 23.7ºC at 10 m depth, and below this depth the temperature was ~18ºC. In the GA03 transect the water temperature was warmer (24-25ºC), and nitrate concentrations were still undetectable at 90 m, were ~2 µmol/kg at 200 m and <5 µmol/kg in the top 300 m. The nitrite maximum was 0.1 µmol/kg at 90 m. Thus, more of

the upper water column had undetectable nitrate and nitrite in the fall at BATs.

When we examined the *ureC* calculated over the total community (Bacteria *rpoB* plus Archaea *rpoB*), we noticed that for ST2-GA02 the total proportion of the community with *ureC* was generally below 70%. However, in ST10-GA03, the abundance of the community *ureC* increased to ~100% (Fig. 2.5 b). The increase in the abundance of the *ureC* reads corresponded to Picocyanobacteria and members of the Alphaproteobacteria group. There were also differences between transects when we look at the total *cynS* over the microbial community (Bacteria *rpoB* + Archaea *rpoB)*, in ST21-GA02, the microbial community *cynS* was higher at 75 and 100 m depth in 12 and 7% respectively, however, the increase corresponded to *Prochlorococcus* alone (Fig. 2.6 a); the *cynS* of other microbial groups at other depths was below 2%. In ST10-GA03, the abundance of *cynS* in the total microbial community was lower than 4%, so fewer microbes were capable of using *cynS* in the fall. Thus, a greater proportion of the microbial community could use urea in the more nutrient limited fall, but a smaller proportion of the community could use cyanate. Unfortunately, no seasonal, or any, measurements of cyanate have been made for this station, so it is difficult to explain this data further.

Differences were also observed within individual microbial taxa; when we compared the abundance of SAR11 *ureC* in the BAT station, we observed that in ST21-GA02, the total SAR11 *ureC* reads were between 15-28.2%, whereas in the ST10-GA03, the SAR11 *ureC* reads varied between 8.3-23.7%, at similar depths. The ST10-GA03 had samples below 202 m depth reaching 414 m; however, the total SAR11 *ureC* read abundance was lower than 5.1%. However, even though the total proportion of SAR11 with *ureC* was similar between transects, the proportion and abundance of the SAR11 *ureC* subgroups changed between transects in ST21-GA02, the abundance of SAR1 *ureC* subgroups changed with depth; the subgroups 1a.3 *ureC* has a low abundance (~10%) where water temperature was warmer (24-25ºC at 10 m depth) (Fig. 2.8) and Group V had abundance up to 50% of SAR11. However, in the ST10-GA03 both SAR11 subgroups were present in similar proportions 20-50% of the total SAR11 *ureC*, despite warmer temperatures at this time point.

We expect that the changes in the abundance of the different microbial taxa containing *ureC* and *cynS* changed in response to differences in nutrient concentrations and temperature.

Spatial Variability:

The changes within the distribution of the genes *cynS* and *ureC* are reflecting changes in the community within depth at each station. Microbial groups occupied different niches at different depths in oxic waters. For example, SAR11 was present through the whole water column as we observed with the *rpoB* gene (Fig. 2.8 a), but when we look at the SAR11 *ureC* gene distribution, SAR11 *ureC* was much more abundant in the upper euphotic zone. Both SAR11 and Picocyanobacteria had the ability to use urea

in surface waters. The depth profile of Picocyanobacteria decreased with depth.

In contrast, *Nitrospina* increased abundance with depth. *Nitrospina* abundance examined with the *rpoB* gene was lower in the surface and increased with depth in both transects; GA02 (Fig. 2.7 a) and GA03 (Fig. 2.11 a). *Nitrospina*, like Picocyanobacteria has both genes *ureC* and *cynS*, but despite having both genes Nitrospina had a higher proportion of *ureC* reads over *cynS* reads and we observed this pattern in the GA02 (Fig. 2.7 b), and GA03 (Fig. 2.11 b) transect. However, Nitrospina was the main microorganism that could use *cynS* at depth. Similarly, in the oxic time-points (HOT) Nitrospina showed a similar pattern of distribution, when examining the Nitrospina *rpoB* that varied from low to high abundance with depth (Fig. 1.7 a, Chapter 1) and higher proportion of *ureC* over *cynS* (Fig. 1.7 b, Chapter 1). Thaumarchaeota was present in all stations, with increasing abundances from ~50 m and below. ~100% of Thaumarchaeota had the capability of using urea. Thus, there was a clear switch in the community able to use urea and cyanate with depth.

These key four groups of organisms were also the main organisms found to use urea and cyanate at the Hawaii Ocean Time series in Chapter 1.

We also looked at the distribution of Thioglobaceae that were present in the North Atlantic. Very little of the Thioglobaceae in the North Atlantic and the *ureC* gene. Previously we examined Thioglobaceae in oxic waters and ODZ. Thioglobaceae abundance were extremely low in the oxic time-points at HOT (<5% of the *rpoB*), and the abundance of *ureC* reads were also low <10%, in contrast the highest abundance of Thioglobaceae was in ODZ waters (100%of the *ureC* reads) where the majority of those reads corresponded to an Thioglobaceae subclade specific for ODZ waters (ODZ Thioglobus), in the GA02 and GA03 transect the abundance of this group was also low <10% (Supplementary Figure 2.7 a) and c)), and the abundance of *ureC* was also extremely low <50% with a few exception at two stations (ST21 in the transect GA02 and ST20 in the transect GA03). By looking at Thioglobaceae we can see that microbial groups occupy specific niches in the ocean that are restricted to nutrient availability and the presence of oxygen.

Latitude and longitude changes

Most microbes have *ureC* in the Subtropical gyre in the GA03 transect, but ~ 50% or lower of the microbial prokaryotic community (Bacteria and Archaea) have the *ureC* gene in the GA02 transect, where nutrients were more abundant.

In the case of the GA03 transect, an important aspect to consider is the effect of the subtropical gyre in the region comprehended between ST10-ST22 (Fig. 2.2). The decrease of nutrients in the subtropical gyre and increased stratification could contribute to the increase in the abundance of *ureC* reads between the stations ST10-ST22. The increase in the *ureC* reads was especially high in members of Alphaproteobacteria (SAR11) and Picocyanobacteria, which utilize urea or cyanate in surface waters.

The deeper microbes had less variability across transects in the North Atlantic and between North Atlantic and North Pacific subtropical regions. Thaumarchaeota *rpoB* abundance at the Hawaii Ocean Time Series (HOT) (Fig. 1.12; Chapter 1) reached 30% but the abundance of Thaumarchaeota at GA03 (Fig. 2.13) and GA02 (Fig. 2.9) were generally Thaumarchaeota *rpoB* <20%. In the GA02, the abundance of Thaumarchaeota *ureC* reads was fairly constant in the water column ~100%, (with higher ~200% in the Southern stations). However, in the GA03 transect the depth profile Thaumarchaeota *ureC* reads changed within each station and across the transect. In western shelf/slope stations, the Thaumarchaeota *ureC* reads increased with depth in ST1-ST6 (Fig. 2.13 b) whereas towards the East the Thaumarchaeota *ureC* reads abundance decreased with depth (ST18-ST24) (Fig. 2.13 b). This reflects latitudinal changes in the proportion of Thaumarchaeota with *ureC*.

As discussed earlier, there are clear changes in ecotypes across latitude and longitude. For SAR11 *ureC* subgroups in the GA02 transect, we observed the distribution changed in response to temperature variation as temperature generally increased longitudinally (Fig. 2.8 b) as the transects moved in orientation North (from colder waters) at 40ºN towards South-West until around 20ºN (to waters with higher temperatures in the surface) (Fig. 2.3 c). When we look at latitudinal changes in the GA03 transect, the temperature decreased from >25ºC in the upper 100m, towards lower temperatures ~15ºC at around 300 m depth (Fig. 2.4), within this vertical temperature gradient, we observed changes in patterns of distribution of the SAR11 subgroup with depth. In the SAR11 subgroup V, the abundance of SAR11 *ureC* read increased with depth, whereas for subgroups 1a3 the SAR11 *ureC* abundance decreased with depth (Fig. 2.12 b).

Synechococcus, *Prochlorococcus* and the proportion of them that could use cyanate and urea also changed with latitude and longitude. In the stations dominated by *Synechococcus* in the GA02 transect we observed that *Synechococcus* distribution examined with the *rpoB* gene decreased in abundance from the surface towards depth (Fig. 2.10 a); this pattern was also observed in the distribution of *ureC* reads (calculated out of the total Picocyanobacteria *ureC*) which decreased with depth (Fig. 2.10 c) The total Picocyanobacteria *cynS* community also decreased with depth (Fig. 2.10 d), about ~100% of those *cynS* reads corresponded to *Synechococcus* (Fig. 2.10 e). Different ecotypes of *Prochlorococcus ureC* reads dominated from the middle (ST18, Fig. 2.10 c) section towards the southern stations (ST27) in the GA02 transect. However, the abundance of total Picocyanobacteria *cynS* in that section of the transect (ST18-ST27) was relatively low, <50% compared with the first part (ST10-ST17), was dominated by *Synechococcus cynS* (50-300%). In the GA02 transect in the section ST18-ST27, we observed a change in the depth profile, where *Prochlorococcus cynS* reads abundance increased with depth. In contrast, *Synechococcus cynS* read distribution decreased with depth from the surface (Fig. 2.10 e). Generally, the stations dominated by *Synechococcus ureC* also had a higher proportion of

Synechococcus cynS. In those same stations, we also observed that the total Picocyanobacteria *cynS* abundance was higher 50-300%. In contrast, in stations where the *Prochlorococcus* ecotypes of *ureC* dominated, like the sections between ST18-ST27 in the GA02 transect (Fig. 2.10 c) or the majority of the stations in the GA03 transect (Fig. 2.14 c), the total Picocyanobacteria *cynS* was low <60%. When we looked at the distribution of Picocyanobacteria in the oxic time-points (HOT), the proportion of the Picocyanobacteria examined with *rpoB* was higher ~50% in surface waters (Fig.1.13, Chapter 1), and those stations were dominated by *Prochlorococcus* ecotypes *ureC* reads that sum up to ~100 of the total Picocyanobacteria in surface waters, and, there the abundance of *Synechococcus ureC* reads was rather low ~30% of the total Picocyanobacteria *ureC* ~200 m depth. However, the abundance of the total Picocyanobacteria *cynS* reached ~70% at around 100 m depth. The distribution of the Picocyanobacteria *cynS* and *ureC* genes in subtropical ocean waters showed that *Prochlorococcus* and *Synechococcus* inhabit different niches at different depths and ocean waters. In the environment, in regions where *Synechococcus* lives, it could be using both genes to maximize the use of reduced nitrogen forms in low nutrients waters and colder ocean water like the north part of the GA02 transect. Meanwhile, the different *Prochlorococcus* ecotypes prefer urea over cyanate in warmer waters in the subtropical gyre, but a smaller proportion of the *Prochlorococcus* have *cynS*.

Thus, it appears that changes in depth, season, and latitude and longitude are important for *ureC* and *cynS* abundance and distributions in the water column.

Conclusions

We examined the distribution of microbial groups capable of using cyanate, urea, and, in some cases, both. The microbial capability for using these compounds contributes to elucidating how nitrogen cycles in the ocean, highlighting the importance of the role of microorganisms.

There is clear seasonality and changes with latitude and longitude in use of urea and cyanate in the North Atlantic with significant differences between GA02 and GA03 transects. We find changes in *ureC* and *cynS* across these Geotraces transects, but these changes are mostly related to changes in the microbial community itself, with changes in *ureC* and *cynS* correlating with ecotype changes. Thus, the ability to use urea or cyanate appear to be related to specific ecotypes of SAR11, Picocyanobacteria, and Thaumarchaeota. We expect that the changes in these ecotypes changed in response to differences in nutrient concentrations and temperature.

Most noticeably, *Synechococcus*, and *Prochlorococcus*, switched in dominance across stations during both transects, causing the phylotypes for *ureC* and *cynS* to change at the same time. This switch was likely related to stratification and resulting reduced nutrient concentrations. Interestingly, the abundance of *cynS* in Picocyanobacteria and total microbial community decreased drastically between spring GA02 and nutrient limited fall GA03

cruises, but the abundance of *ureC* in the total community increased in GA03. Thus, the dynamics of urea and cyanate utilization seem more complicated than we would have guessed. It is unfortunate that cyanate and urea were not measured on these Geotraces cruises and that measurements in these areas are sparse to non-existent, so we are missing the information needed to explain our data.

Though there were changes in surface communities using *cynS* and *ureC* between transects, the deeper communities were more consistent. *Nitrospina,* when present, consistently favored *ureC* over *cynS*, and Thaumarchaeota and its *ureC* were evenly distributed across both transects. The findings presented in this chapter contribute to a better understanding of how microorganisms survive oligotrophic conditions and how they are impacting the nitrogen cycle.

Although we only looked at 11 stations in the GA02 transect, which were located in the North Atlantic, the entire GA02 transect covers an extensive region from the North to the Southern Hemisphere with a total of 32 stations (Biller et al., 2018). An expansion of this work should include analyzing the rest of the station of the GA02 transect and incorporating other GEOTRACES transects (GP13 and GA10).

Figures chapter 2

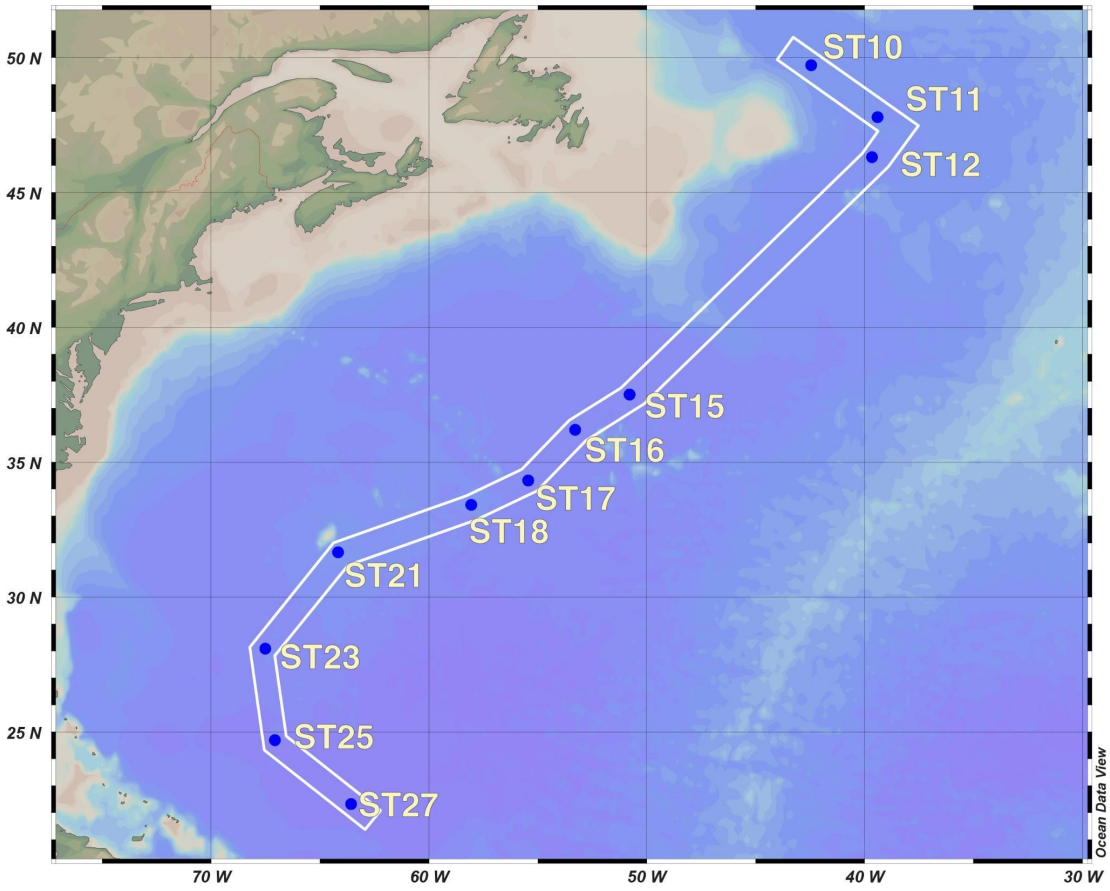


Figure 2.1. Sampling stations in the North Atlantic GA02 transect. The Transect is oriented North-South. ST21 is the same location as BATS. Figure produced using Ocean Data view (Schlitzer, 2023).

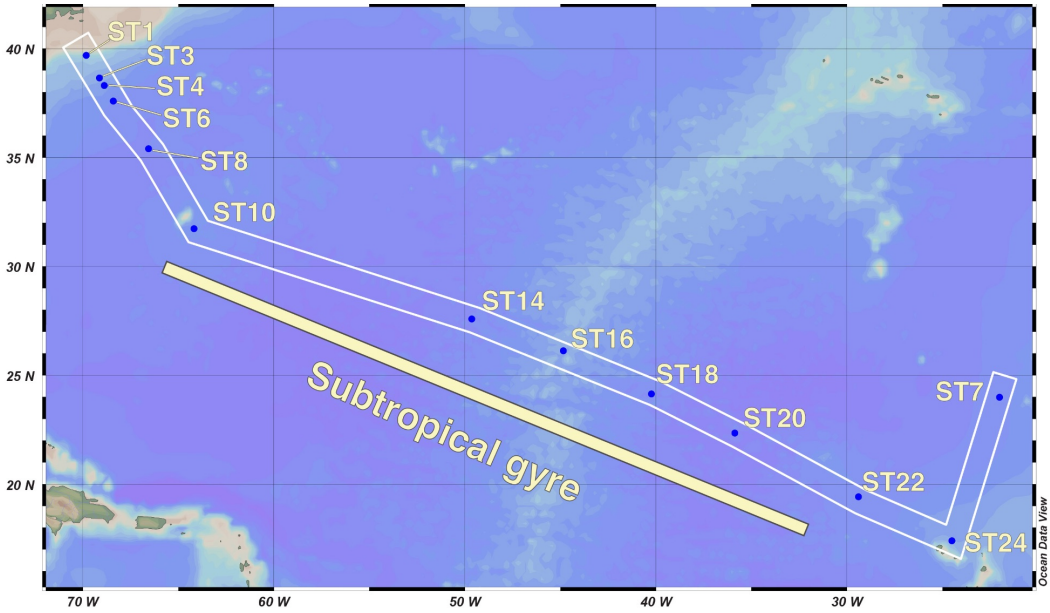


Figure 2.2. Sampling stations in the North Atlantic GA03 transect. The Transect is oriented West-East. ST10 is at the same location as BATS. The yellow bar indicates the location of the subtropical gyre. Figure produced using Ocean Data view (Schlitzer, 2023).

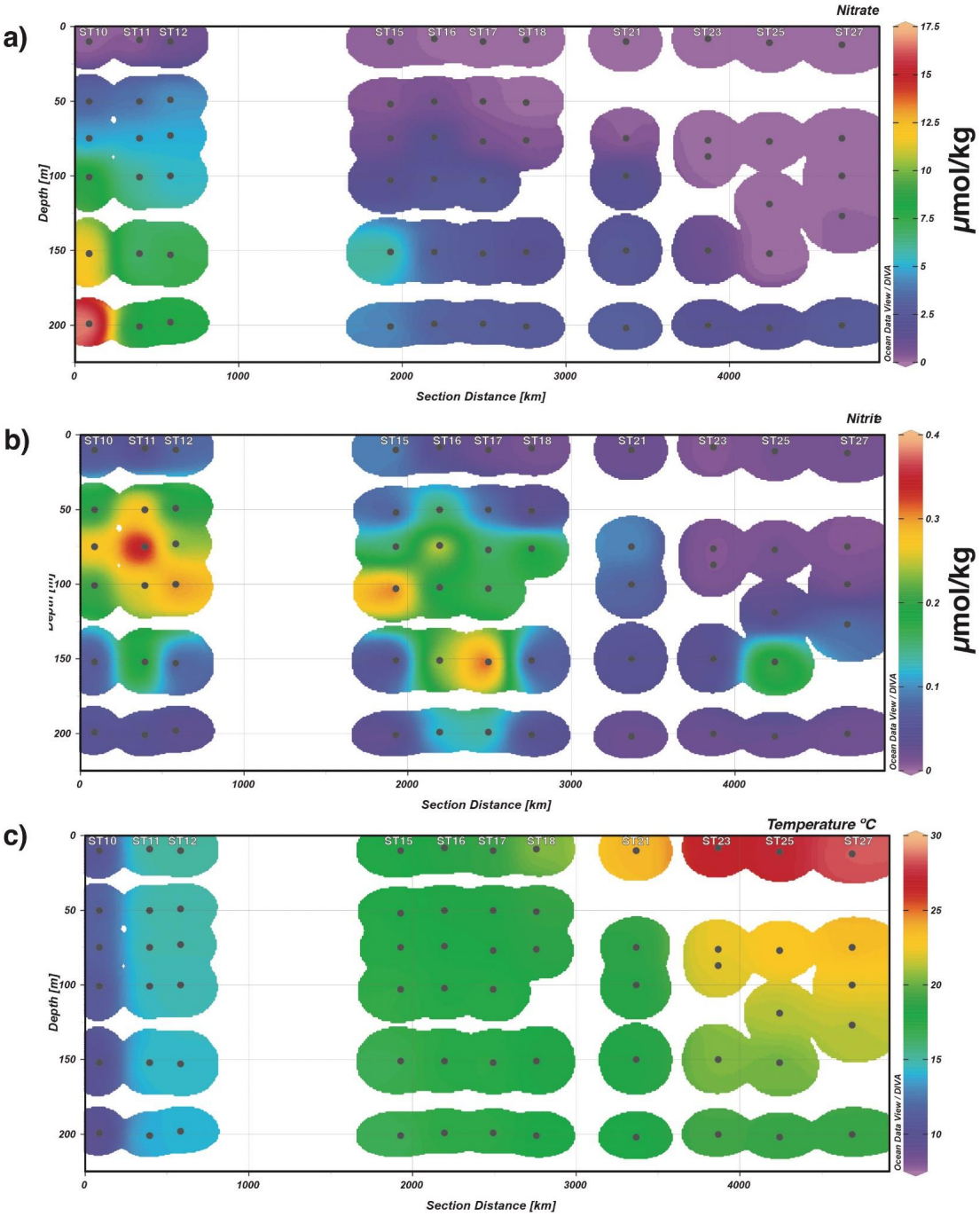


Figure 2.3. Nutrient concentration distribution in the North Atlantic GEOTRACES GA02 transect. The Section distance (y-axis) is oriented North-South (from left to right). a) Nitrate concentration, b) Nitrite concentration. c) Temperature. The station numbers can be found at the top of each graph. Figure produced using Ocean Data view (Schlitzer, 2023).

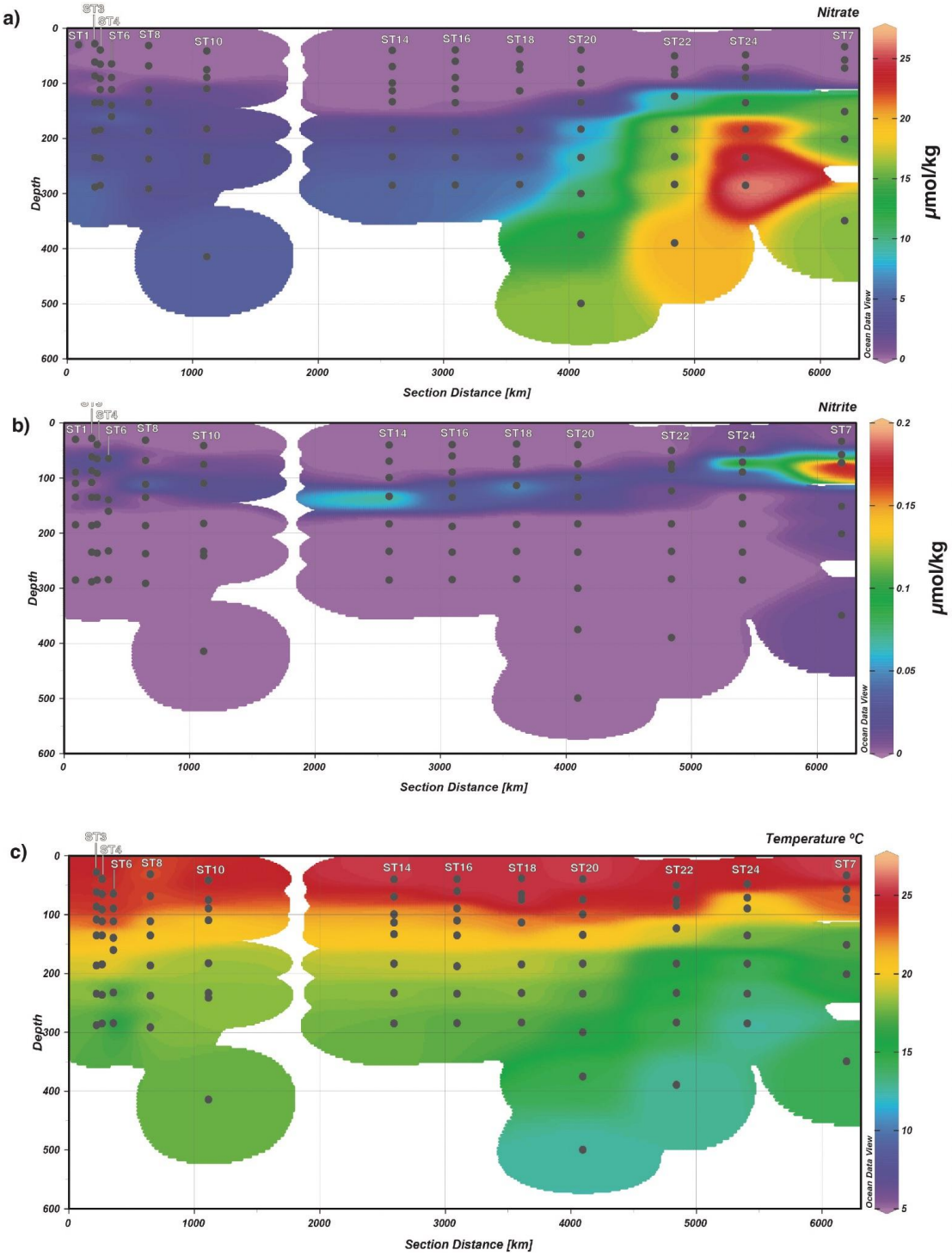


Figure 2.4. Nutrient concentration distribution in the North Atlantic GEOTRACES GA03 transect. The Section distance (y-axis) is oriented West-East (from left to right). a) Nitrate concentration, b) Nitrite concentration, c) Temperature. The station numbers can be found at the top of each graph. Figure produced using Ocean Data view (Schlitzer, 2023).

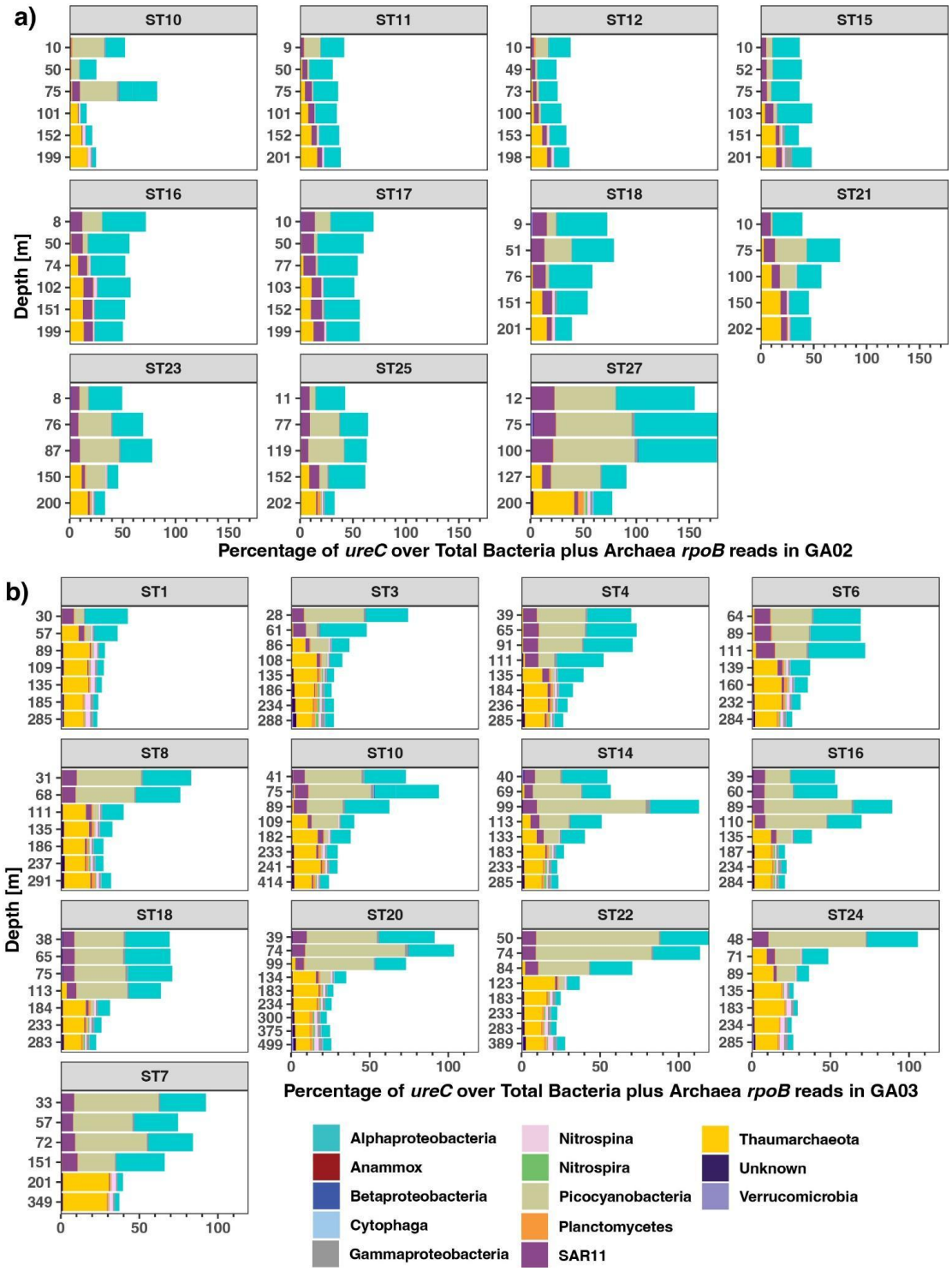


Figure 2.5. *ureC* abundance in the total microbial community in the North Atlantic GA02 transect. Stacked bar chart showing the distribution and proportion of *ureC* reads over *rpoB* (*rpoB*: Archaea + Bacteria). The *ureC*/*rpoB* expressed in % is classified according to the taxonomic group. a) Abundance in % of the *ureC*/*rpoB* ratio in the microbial community, analyzed

in stations located along the GA02 transect (Oriented North-South). b) Abundance in % of the *ureC*/*rpoB* ratio in the microbial community, analyzed in stations located along the GA03 transect (oriented West to East). ST1, ST3, and ST4 are shelf stations. Whereas S24 and S7 are close to the coast. The rest are offshore stations. The depth is not shown at scale.

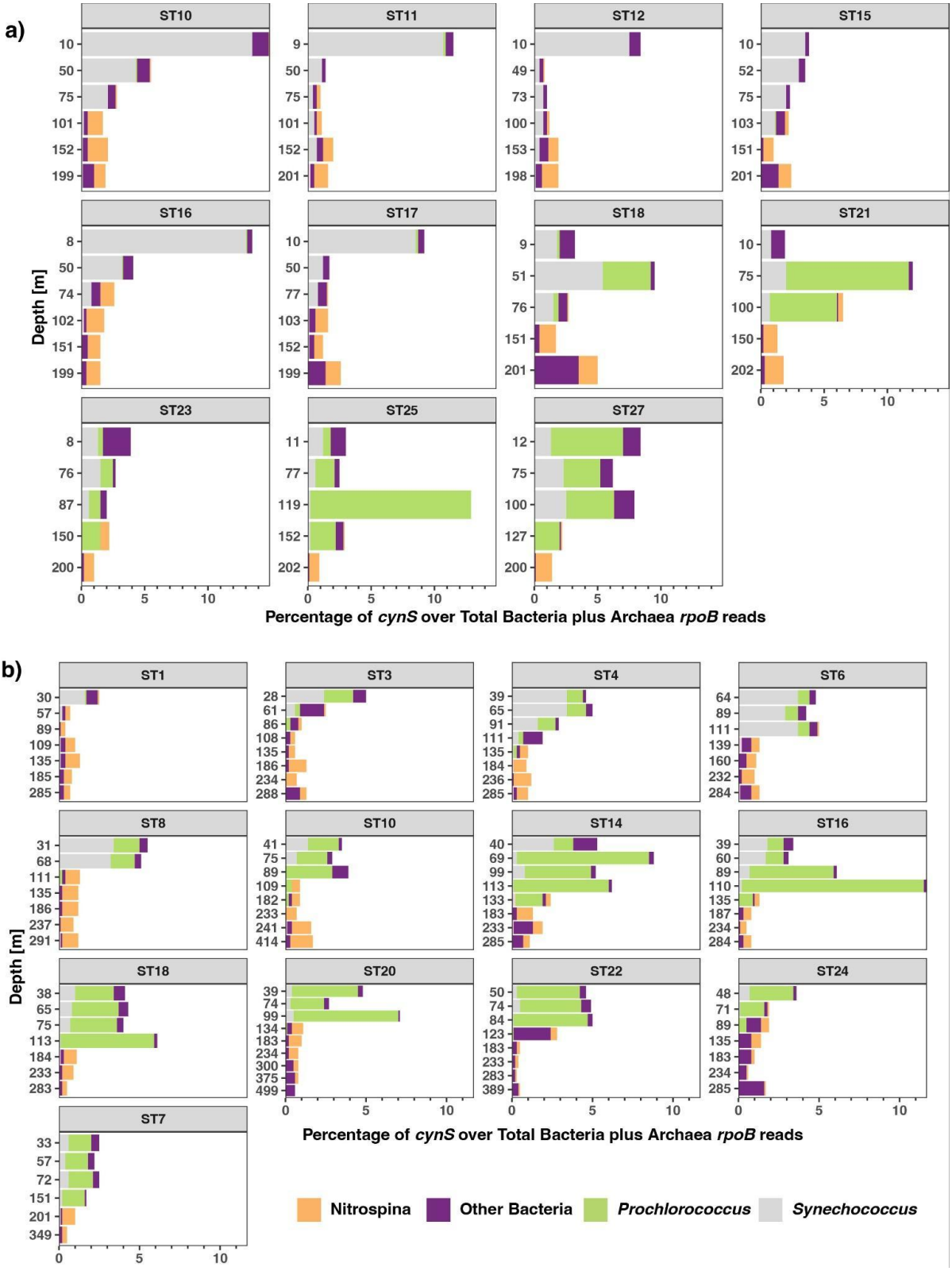


Figure 2.6. *cynS* abundance in the total microbial community in the North Atlantic GA02 transect. Stacked bar chart showing the distribution and proportion of the *cynS/rpoB* ratio (*rpoB*: Archaea + Bacteria). The

cynS/rpoB expressed in % is classified according to the taxonomic group. a) Abundance in % of the *cynS/rpoB* ratio in the microbial community, analyzed in stations located along the GA02 transect (Oriented North-South). b) Abundance in % of the *cynS/rpoB* ratio in the microbial community, analyzed in stations located along the GA03 transect (oriented West to East). ST1, ST3, and ST4 are shelf stations. Whereas S24 and S7 are close to the coast. The rest are offshore stations. The depth is not shown at scale.

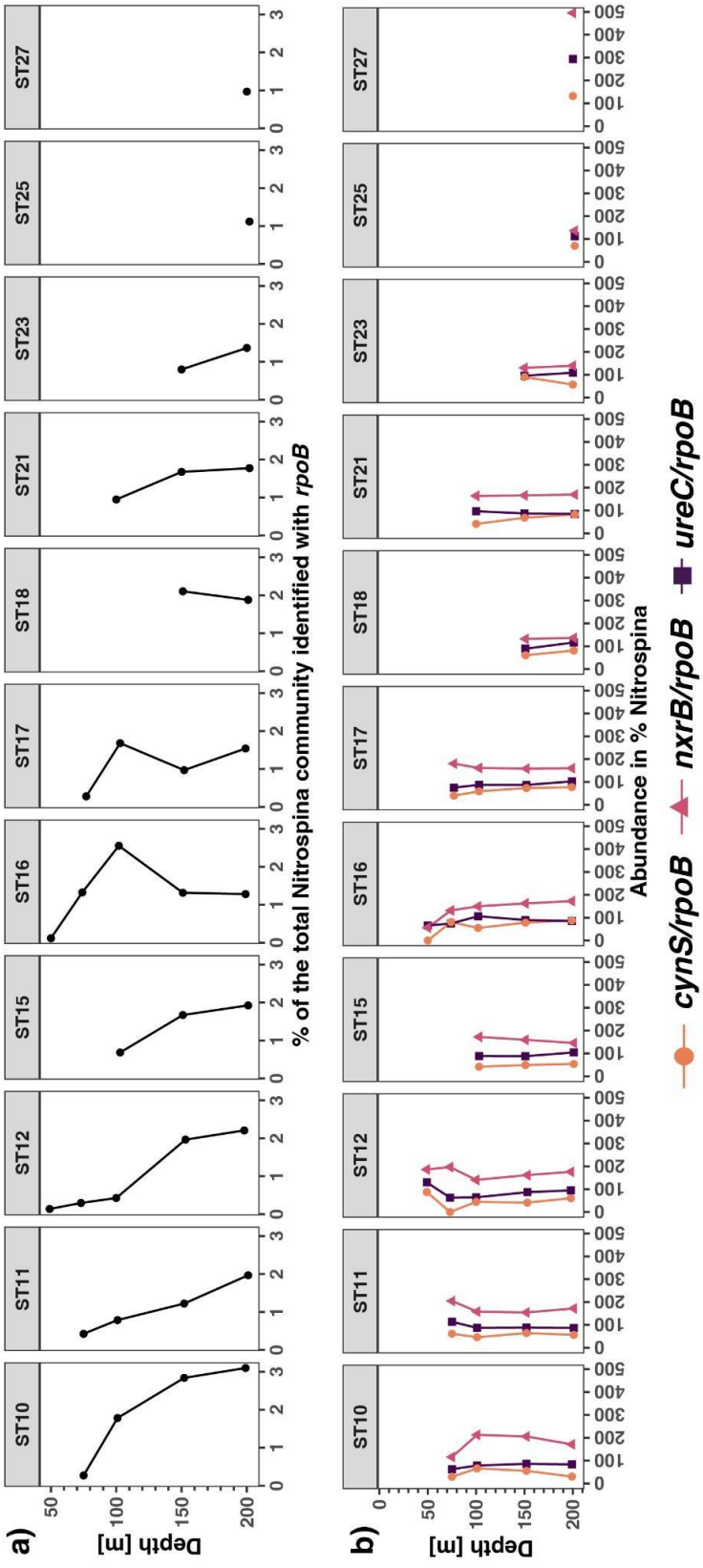


Figure 2.7. *Nitrospina* functional genes distribution in the North Atlantic GA02 transect. Depth profile of metagenomic reads encoding for enzymes *rpoB* (Beta subunit of RNA polymerase), cyanate (*cynS*), urease subunit alpha (*ureC*), nitrite oxidoreductase (*nxrB*) in the GA02 transect. a) Percentage of the bacteria community identified as Nitrospina Abundance in % of the Nitrospina community identified with *rpoB*. b) Abundance in % of the *cynS/rpoB* ratio (orange), *nxrB/rpoB* ratio (pink), and *ureC/rpoB* ratio (purple). Each panel corresponds to a different station. The stations are oriented with increasing numbers from North to South.

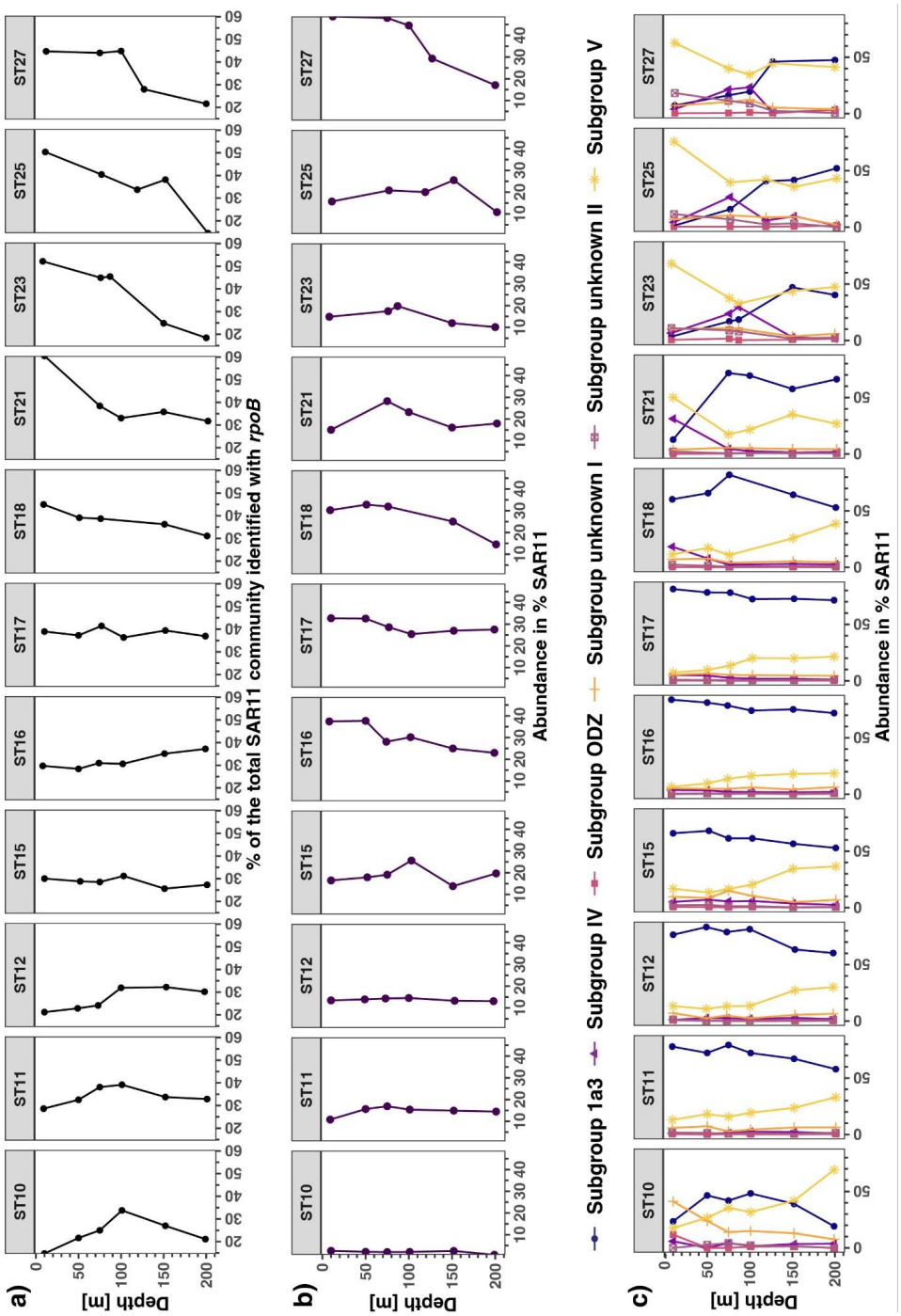


Figure 2.8. SAR11 functional genes distribution in the North Atlantic GA02 transect. Depth profile of metagenomic reads encoding for enzymes *rpoB* (Beta subunit of RNA polymerase), and urease subunit alpha (*ureC*) in the GA02 transect. a) Abundance the total SAR11 community identified with *rpoB*, b) Abundance in % of the SAR11 *ureC/rpoB* ratio. c) Abundance of each SAR subgroup (*ureC)* calculated of the total SAR11 *ureC.* Each panel corresponds to a different station. The stations are oriented with increasing numbers from North to South.

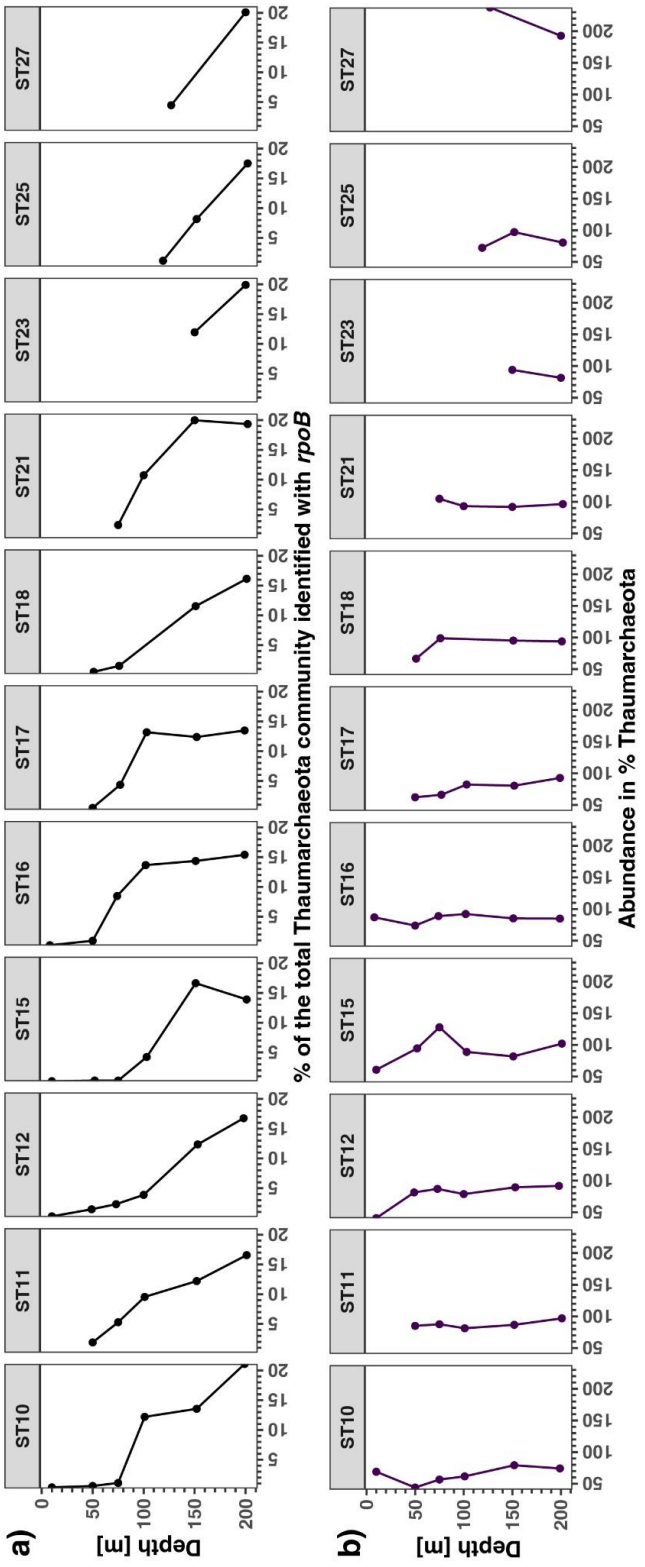


Figure 2.9. Thaumarchaeota functional genes distribution in the North Atlantic GA02 transect. Depth profile of metagenomic reads encoding for enzymes *rpoB* (Beta subunit of RNA polymerase), and urease subunit alpha (*ureC)* in the GA02 transect. a) Abundance the total Thaumarchaeota community identified with *rpoB*. b) Abundance in % of the Thaumarchaeota *ureC/rpoB* ratio. Each panel corresponds to a different station. The stations are oriented with increasing numbers from North to South.

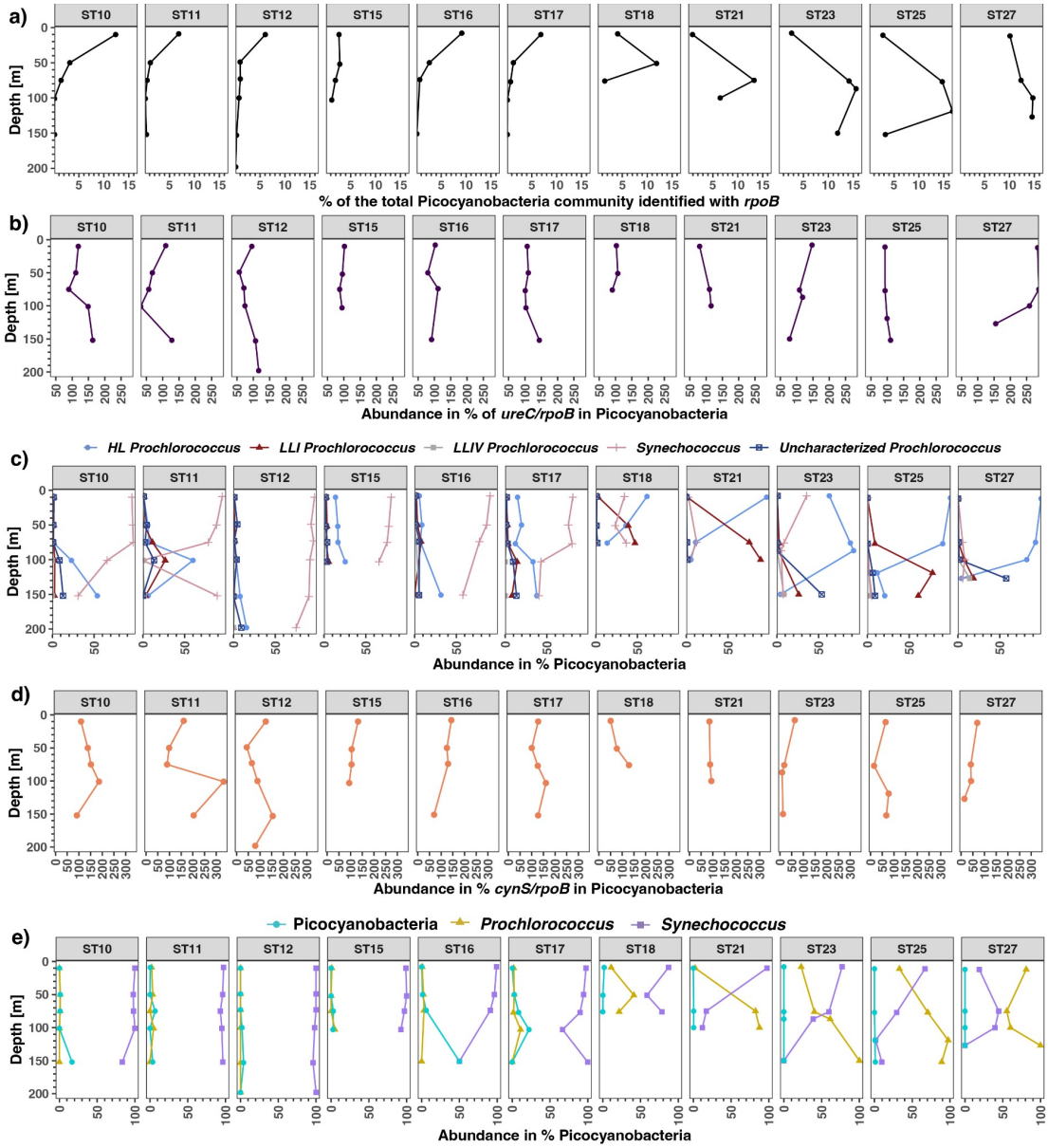


Figure 2.10. Picocyanobacteria functional genes distribution in the North Atlantic GA02 transect. Depth profile of metagenomic reads encoding for enzymes *rpoB* (Beta subunit of RNA polymerase), cyanate (*cynS*), and urease subunit alpha (*ureC*) in the GA02 transect. a) Abundance of Picocyanobacteria in the total microbial community identified with *rpoB*. b) Abundance of *ureC* in % of the total Picocyanobacteria from the *ureC/rpoB* ratio. c) Abundance in % of each the Picocyanobacteria containing *ureC* out of the total Picocyanobacteria *ureC* in *Synechococcus*, and *Prochlorococcus* ecotypes: High Light (HL) *Prochlorococcus,* Low Light (LL) IV *Prochlorococcus*, Low Light (LL) I *Prochlorococcus*, Uncharacterized *Prochlorococcus* and *Synechococcus*, d) Abundance of *cynS* in % of the total Picocyanobacteria from the *cynS/rpoB* ratio. e) % of the abundance of the *cynS* abundance of *Prochlorococcus*, *Synechococcus*, and "Other Cyanobacteria" calculated over the total Picocyanobacteria *cynS*. Each panel

corresponds to a different station. The stations are oriented with increasing numbers from North to South.

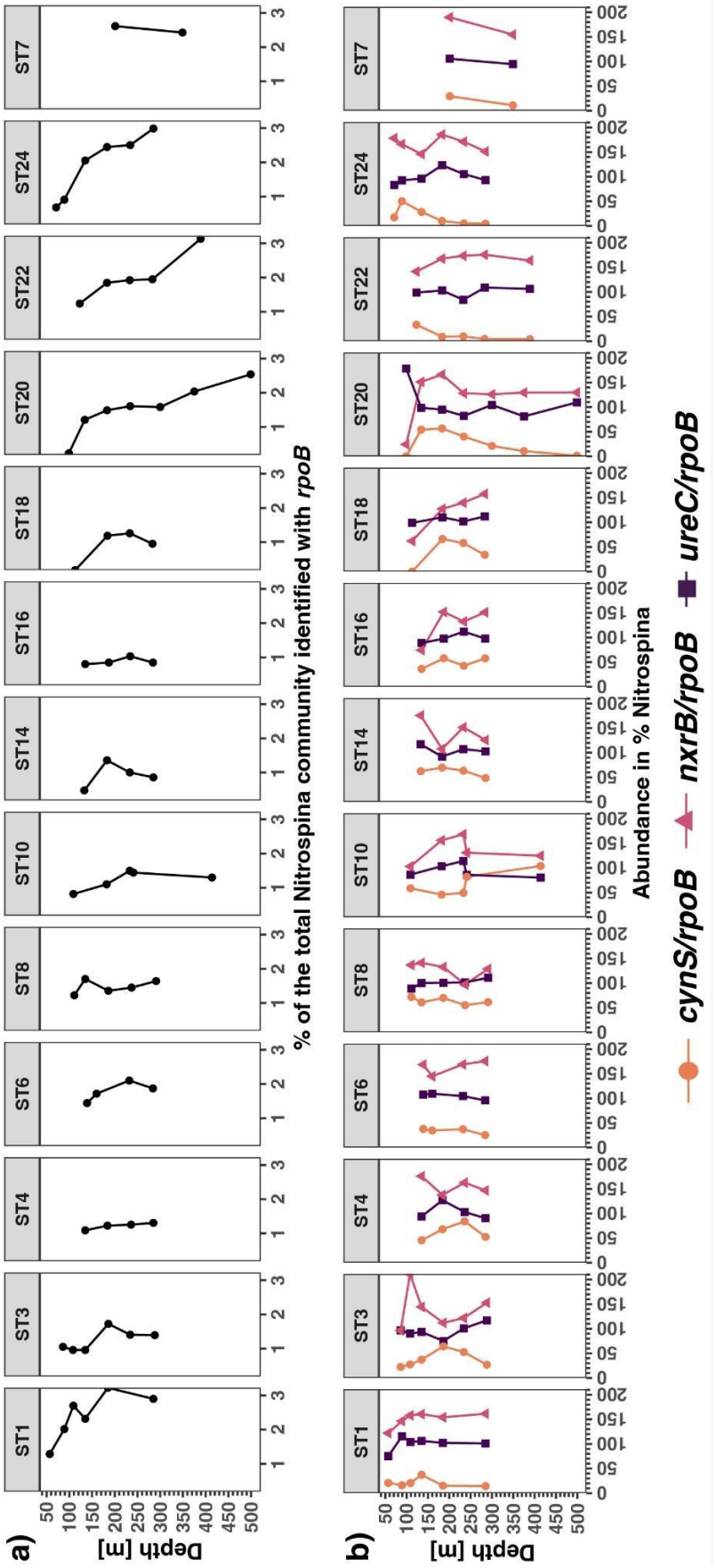


Figure 2.11. Nitrospina functional genes distribution in the North Atlantic GA03 transect. Depth profile of metagenomic reads encoding for enzymes *rpoB* (Beta subunit of RNA polymerase), cyanate (*cynS*), urease subunit alpha (*ureC*), nitrite oxidoreductase (*nxrB*). a) Percentage of the bacteria community identified as Nitrospina Abundance in % of the Nitrospina community identified with *rpoB*. b) Abundance in % of the *cynS/rpoB* ratio (orange), *nxrB/rpoB* ratio (pink), and *ureC/rpoB* ratio (purple). The stations are arranged from West to East (left to right); from station 1 to station 24 Stations 1, 3, and 4 are shelf stations. Stations 24 and 7 are close to the coast. ST10-20 are subtropical gyre stations.

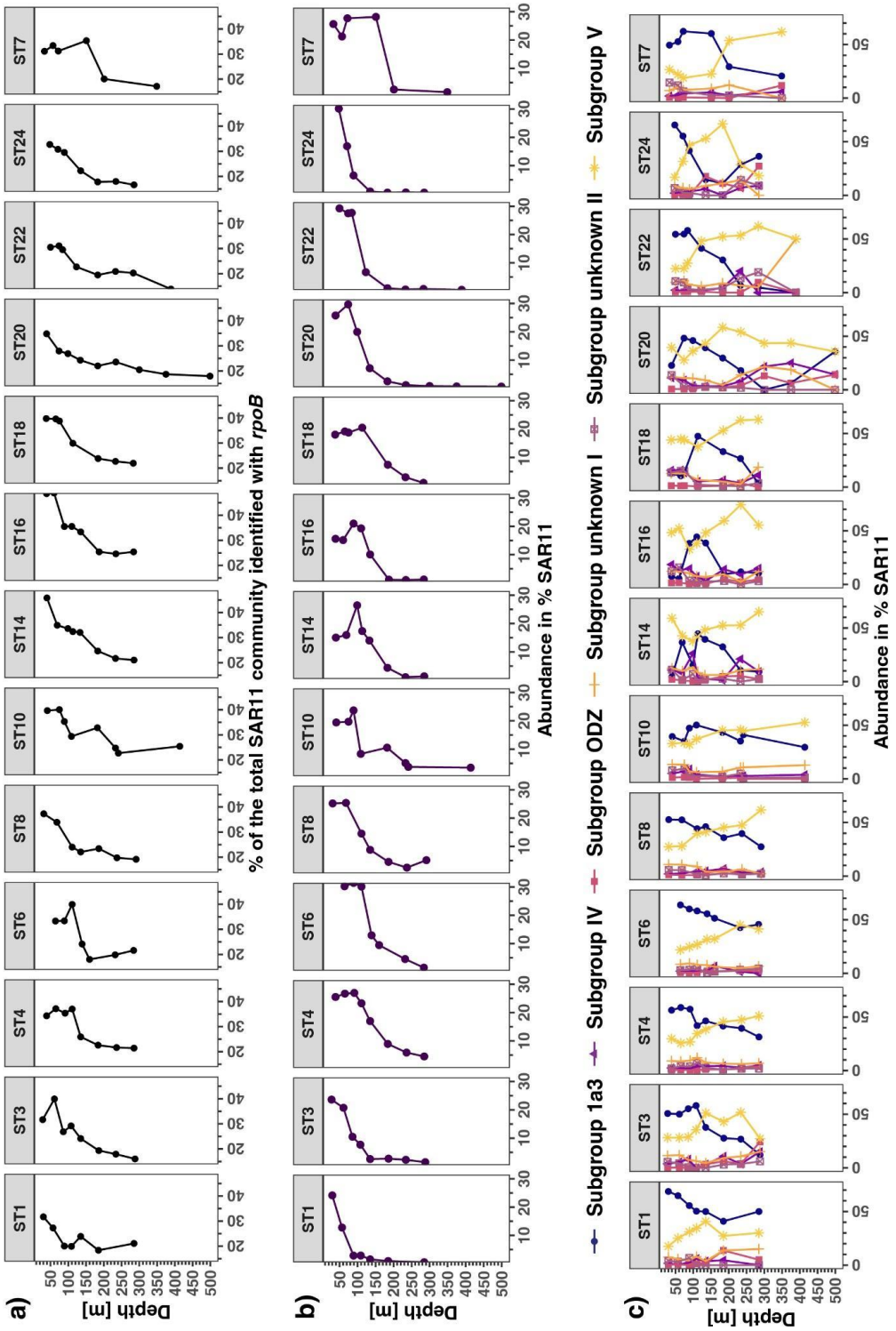


Figure 2.12. SAR11 functional genes distribution in the North Atlantic GA03 transect. Depth profile of metagenomic reads encoding for enzymes *rpoB* (Beta subunit of RNA polymerase), and urease subunit alpha (*ureC*) in the GA03 transect. a) Abundance the total SAR11 community identified with *rpoB*. b) Abundance in % of the SAR11 *ureC/rpoB* ratio. c) Abundance of each SAR subgroup (*ureC)* calculated of the total SAR11 *ureC.* The stations

are arranged from West to East (left to right); from station 1 to station 24. Stations 1, 3, and 4 are shelf stations. ST10-20 are subtropical gyre stations. Stations 24 and 7 are close to the coast.

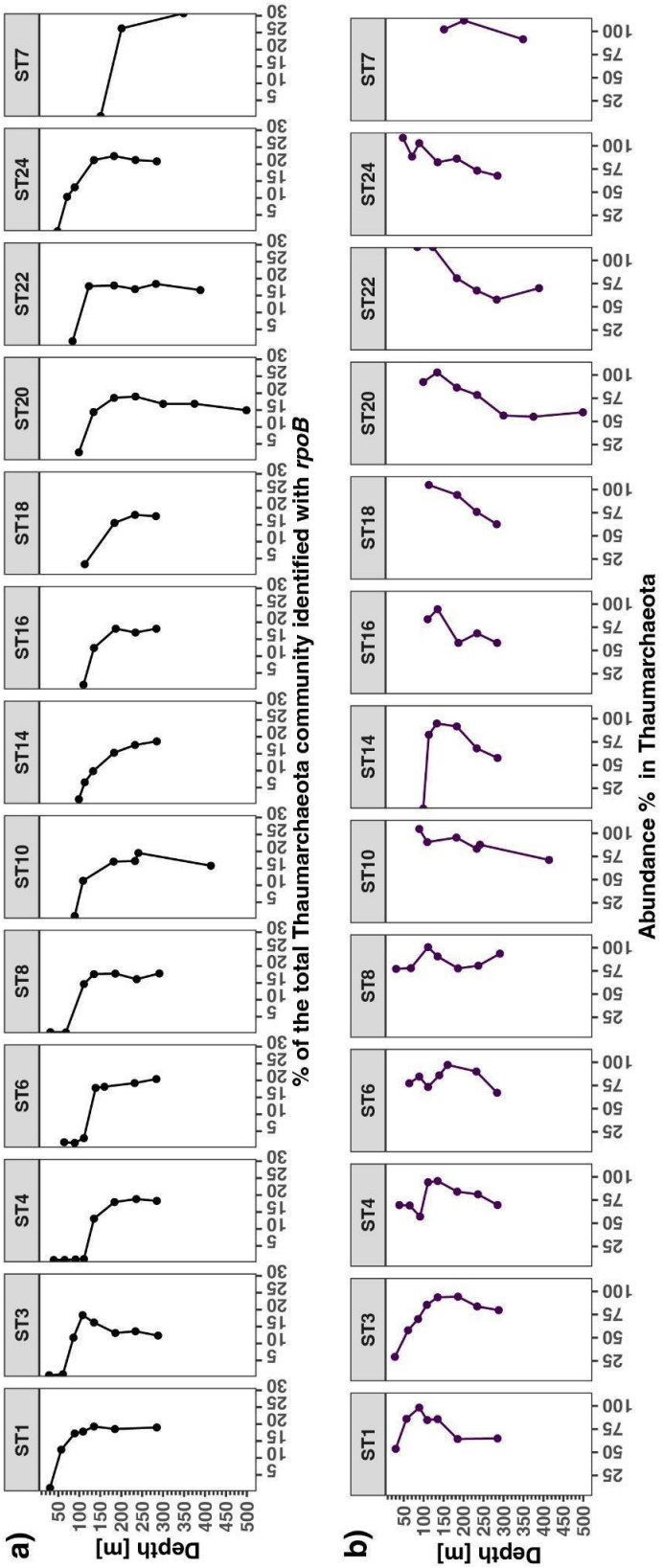


Figure 2.13. Thaumarchaeota functional genes distribution in the North Atlantic GA03 transect. Depth profile of metagenomic reads encoding for enzymes *rpoB* (Beta subunit of RNA polymerase), and urease subunit alpha (*ureC)* in the GA03 transect. a) Abundance the total Thaumarchaeota community identified with *rpoB*. b) Abundance in % of the Thaumarchaeota *ureC/rpoB* ratio. The stations are arranged from West to East (left to right); from station 1 to station 24. Stations 1, 3, and 4 are shelf stations. Stations 24 and 7 are close to the coast. ST10-20 are subtropical gyre stations.

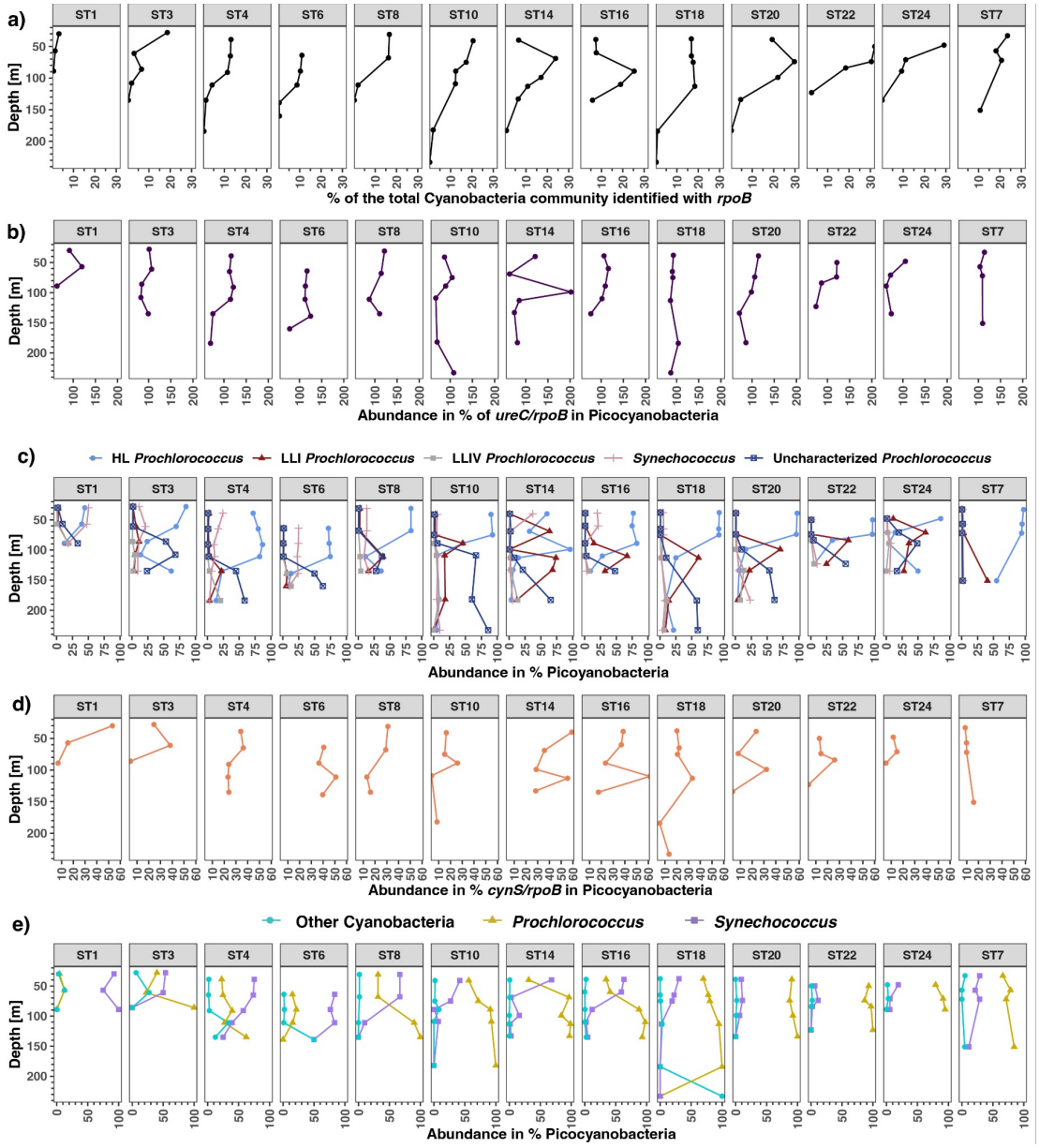
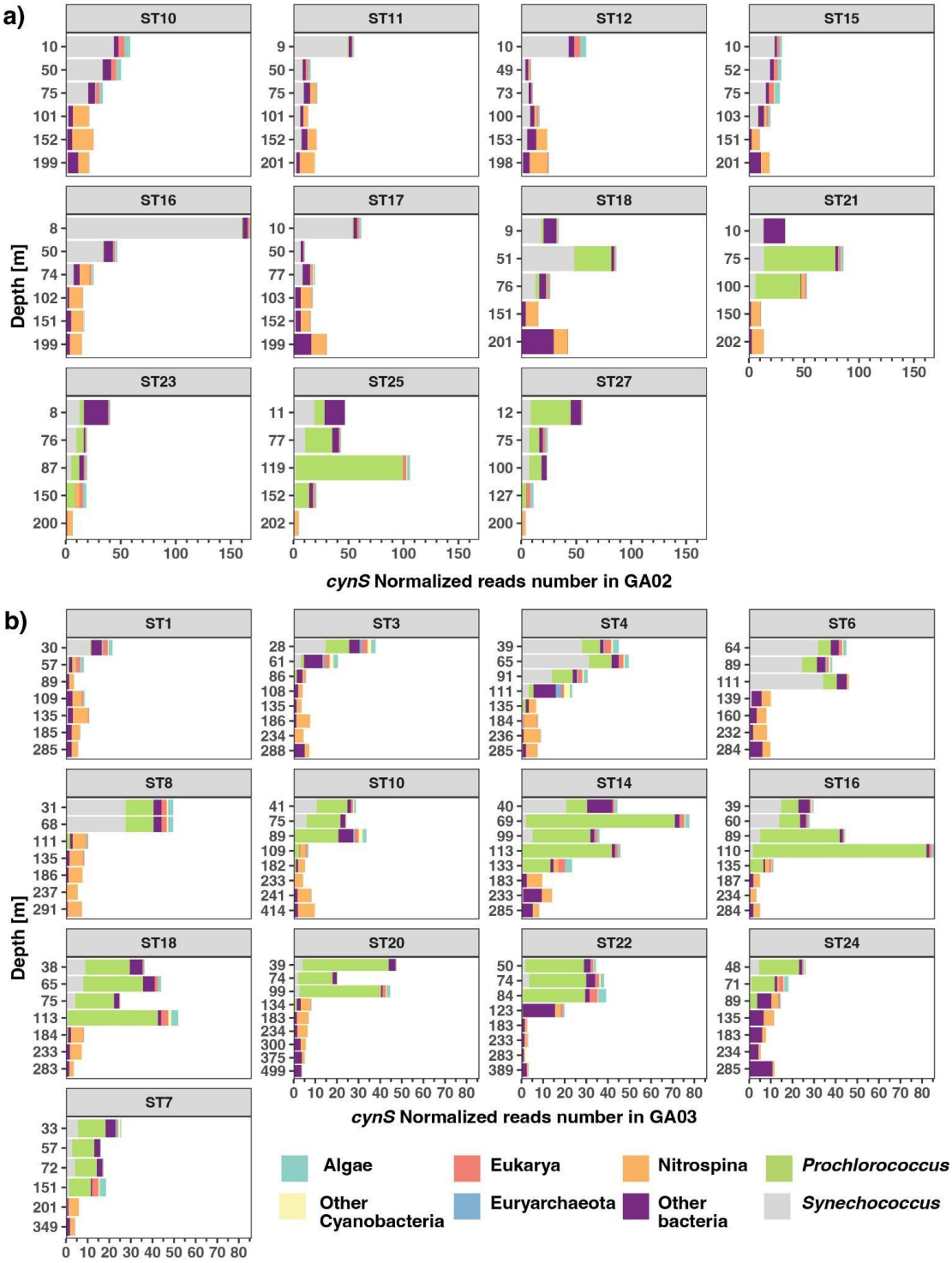


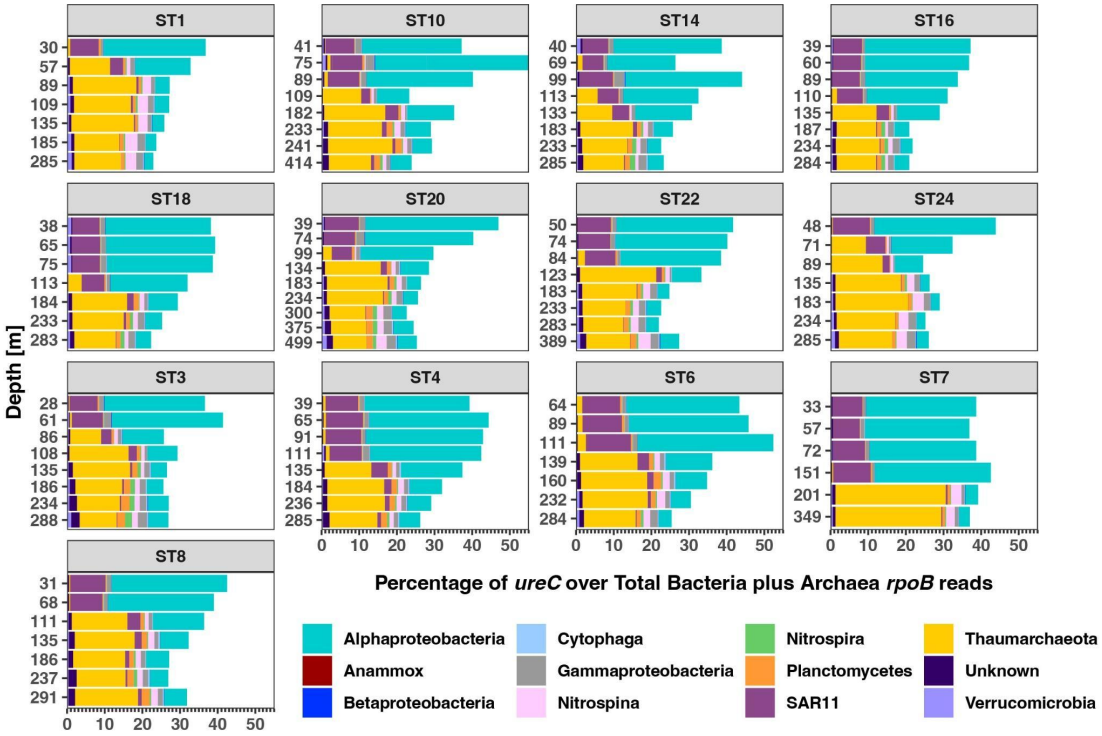
Figure 2.14. Picocyanobacteria functional genes distribution in the North Atlantic GA03 transect. Depth profile of metagenomic reads encoding for enzymes *rpoB* (Beta subunit of RNA polymerase), cyanate (*cynS*), and urease subunit alpha (*ureC*) in the GA03 transect. a) Abundance of the total Picocyanobacteria community identified with *rpoB*. b) Abundance of *ureC* in % of the total Picocyanobacteria from the *ureC/rpoB* ratio. c) Abundance in % of each the Picocyanobacteria containing *ureC* out of the total Picocyanobacteria *ureC* in *Synechococcus*, and *Prochlorococcus* ecotypes: High Light (HL) *Prochlorococcus,* Low Light (LL) IV *Prochlorococcus*, Low Light (LL) I *Prochlorococcus*, Uncharacterized *Prochlorococcus* and *Synechococcus*, d) Abundance of *cynS* in % of the total Picocyanobacteria from the *cynS/rpoB* ratio. e) % of the abundance of the *cynS* abundance of *Prochlorococcus*, *Synechococcus*, and "Other Cyanobacteria" calculated over

the total Picocyanobacteria *cynS.* GA03 transect, the stations are arranged from West to East (left to right); from station 1 to station 24. Stations 1, 3, and 4 are shelf stations. Stations 24 and 7 are close to the coast. ST10-20 are subtropical gyre stations.

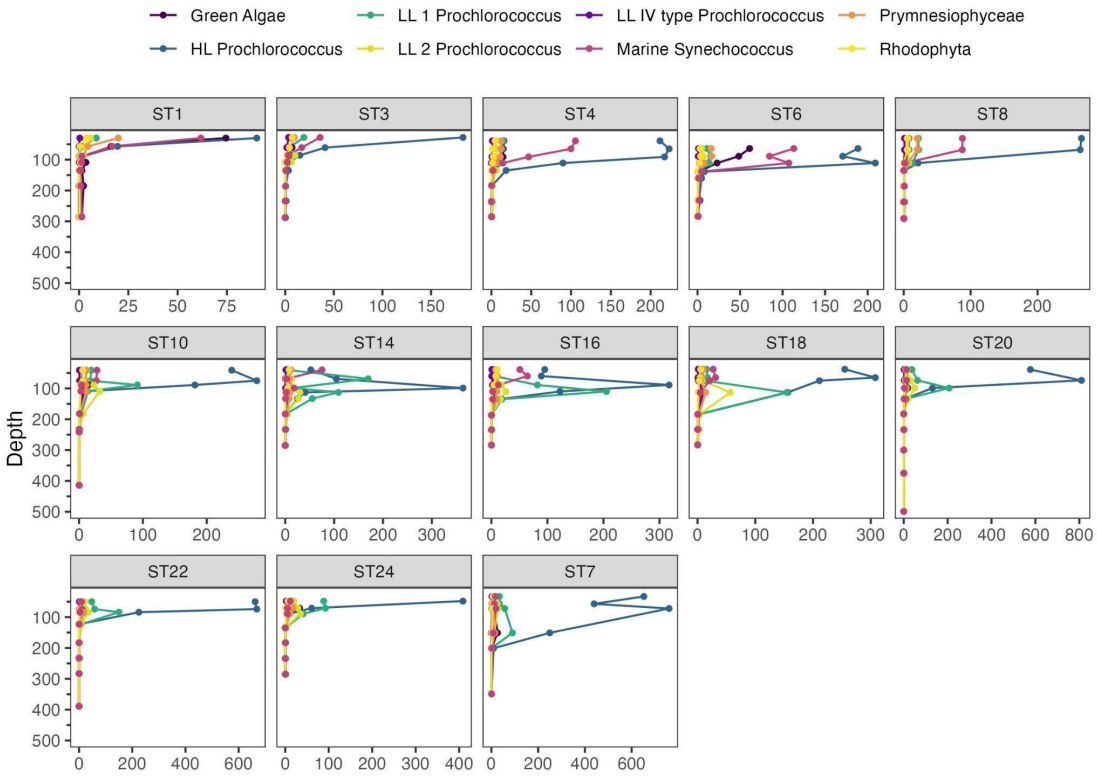


Supplementary figure 2.1. *cynS* normalized read abundance in the North Atlantic GEOTRACES GA02 and GA03. Stacked bar chart showing the percentage of *cynS* normalized read number, classified by taxonomic groups according to depth. a) GA02 transect. The stations are oriented from North to

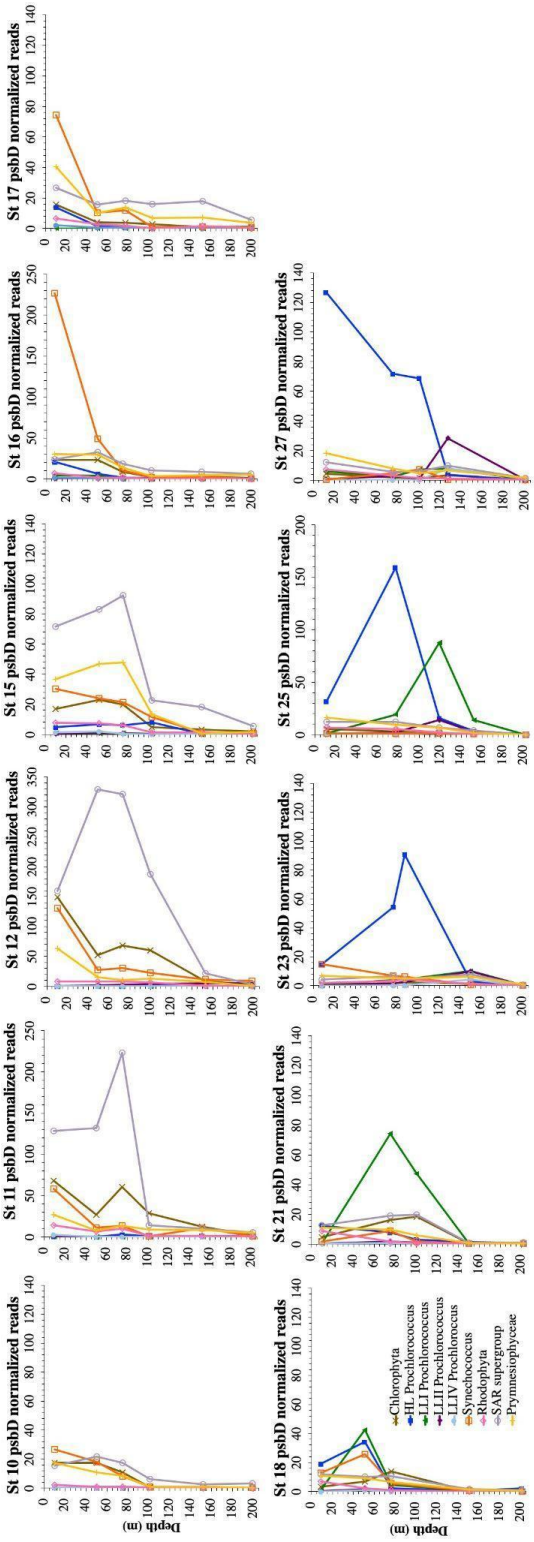
South. The station number increases from North to South. b) GA03 transect, the stations are arranged from West to East (left to right); from station 1 to station 24. Stations 1 ,3, and 4 are shelf stations. Station 24 and 7 are close to the coast. ST10-20 are offshore subtropical gyre stations. The depth is not shown at scale. These graphs include Eukaryotic Algae reads.



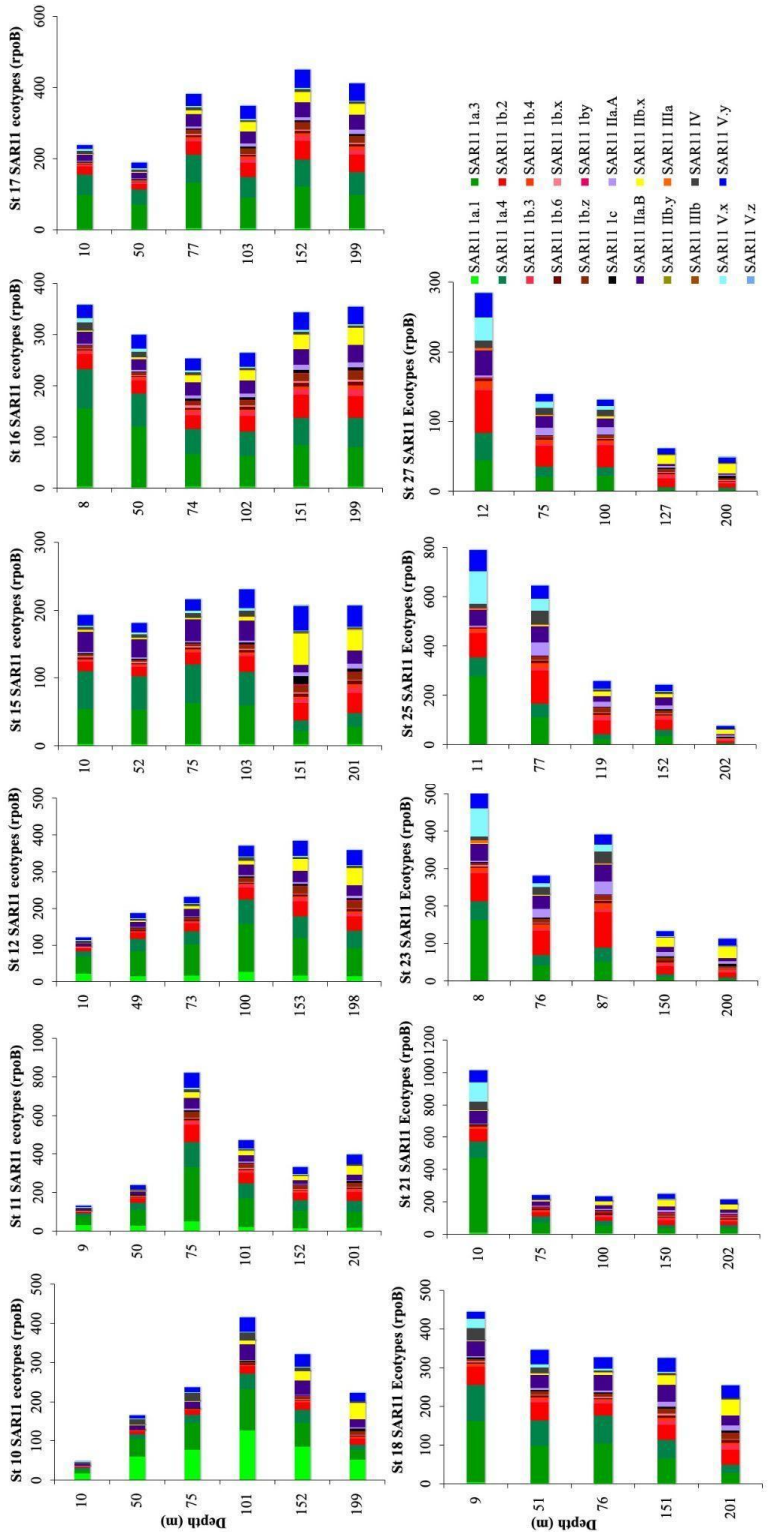
Supplementary figure 2.2. Supplementary figure 2.1. *cynS* normalized read abundance in the North Atlantic GEOTRACES GA02 and GA03. Stacked bar chart showing the percentage of *ureC* copies out of total *rpoB* reads, classified by taxonomic groups according to depth a) for the GA02 transect the stations are oriented from North to South. The station number increases from North to South. b) GA03 transect, the stations are arranged from West to East (left to right); from station 1 to station 24. Stations 1 ,3, and 4 are shelf stations. Station 24 and 7 are close to the coast. ST10-20 are offshore subtropical gyre stations. The depth is not shown at scale.



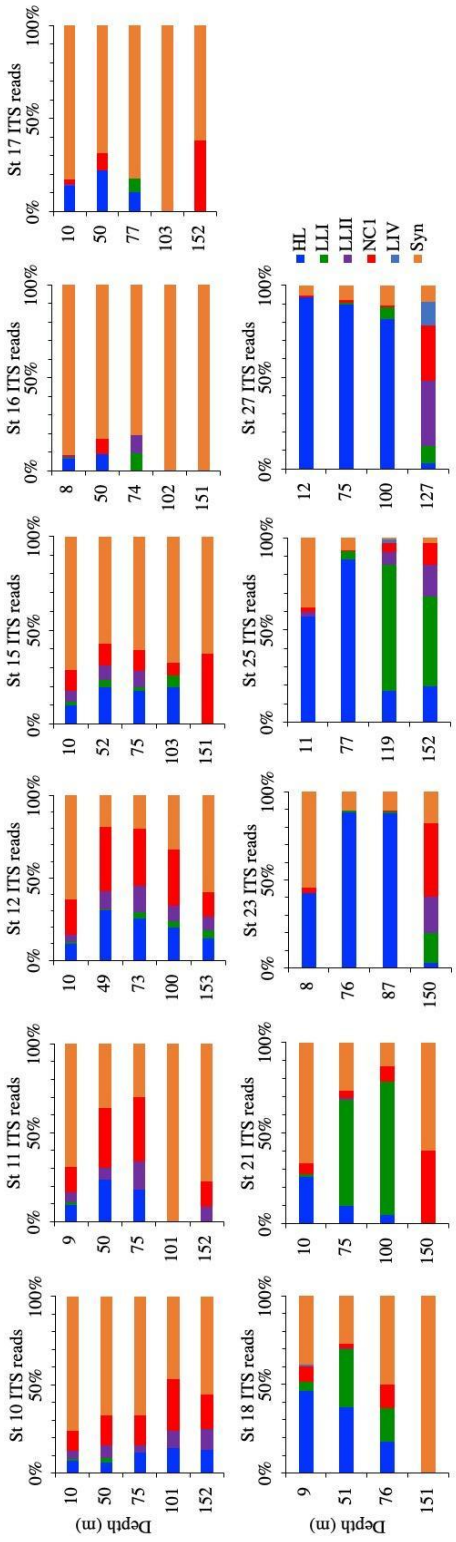
Supplementary figure 2.3. Depth profiles of the abundance of the gene *psbD*, representing both bacterial and eukaryotic photosynthesizers, for GA03. Data from stations 14-20 are from Fuchsman and Hays (2023). ST1, ST3, and ST4 are shelf stations. ST24 and ST7 are close to the coast. ST10-20 are offshore subtropical gyre stations.



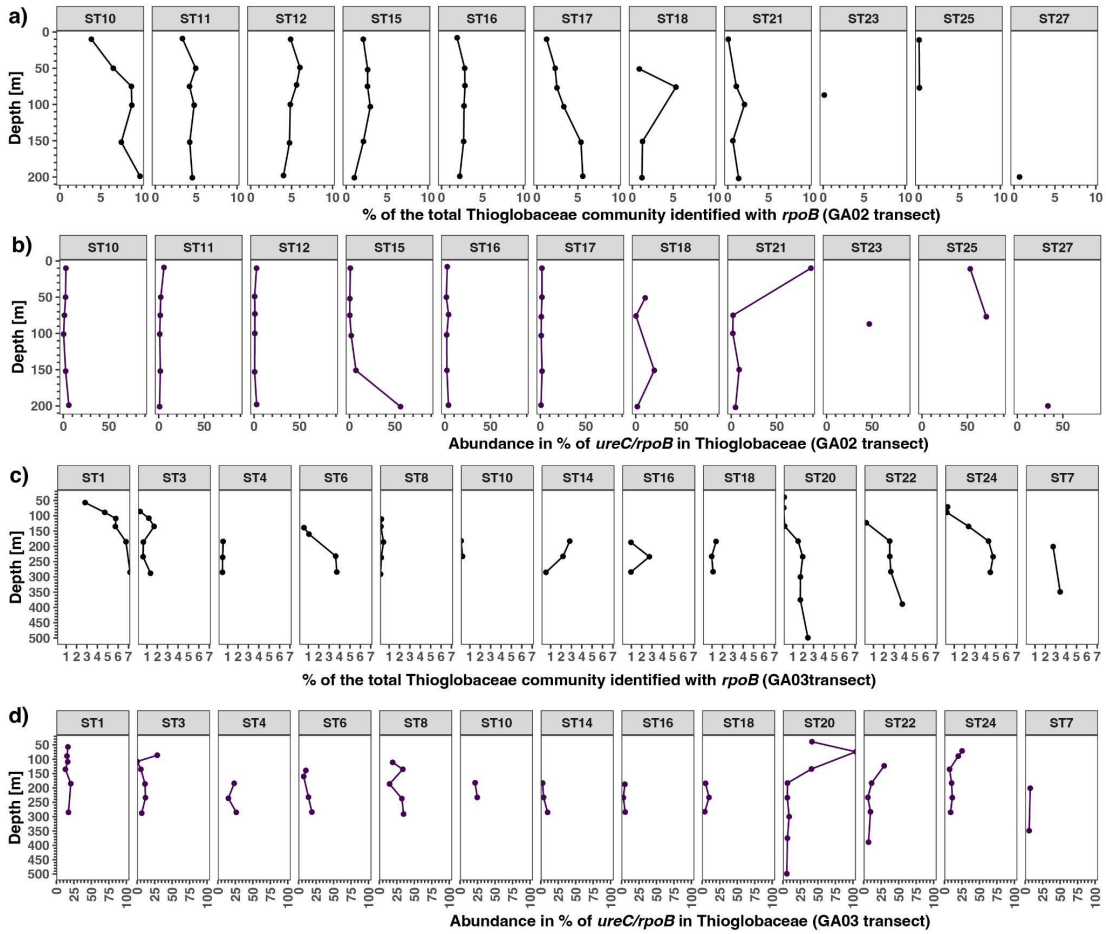
Supplementary figure 2.4. Depth profiles of the abundance of the gene *psbD*, representing both bacterial and eukaryotic photosynthesizers, for GA02. Data are from Fuchsman and Hays (2023) where only stations 15-27 were shown in detail. The station number increases from North to South.



Supplementary figure 2.5. SAR11 ecotypes for the GA02 transect calculated using phylogenetic read placement of the *rpoB* gene. Stations are ordered from North to South. Data from St 10-21 is from Hays and Fuchsman (submitted). The station number increases from North to South.



Supplementary Figure 2.6.
Prochlorococcus ecotypes and total *Synechococcus* as percent of Picocyanobacteria reads for the GA02 transect calculated using phylogenetic read placement of the Internal Transcribed Spacer (ITS) region of DNA. Stations are ordered from North to South. Data from Stations 15-27 are from Fuchsman and Hays (2023). The station number increases from North to South.



Supplementary Figure 2.7 Thioglobaceae functional genes distribution in the North Atlantic. Depth profile of metagenomic reads encoding for enzymes *rpoB* (Beta subunit of RNA polymerase), and urease subunit alpha (*ureC*). Abundance in % of the Thioglobaceae community identified with *rpoB* in a) GA02 transect, b) GA03 transect. Abundance in % of the Thioglobaceae *ureC/rpoB* ratio in c) GA02 transect, and d) GA03 transect. In the GA02 transect, the stations are oriented with increasing numbers from North to South. Whereas in the GA03 The stations are arranged from West to East (left to right); from station 1 to station 24. Stations 1, 3, and 4 are shelf stations. Stations 24 and 7 are close to the coast. ST10-20 are subtropical gyre stations.

Chapter 3: Quantitative microbial taxonomy across particle size, depth, and oxic, hypoxic, and anoxic conditions

Abstract

Particles that form in the ocean surface sink through the water column into the deep ocean, thereby sequestering carbon in the ocean. Microorganisms inhabit and consume these particles, thereby regulating carbon sequestration. Marine particles vary in size and origin and are found in different marine habitats, with each of these differences affecting community structure. One important habitat distinction is the oxygen content of the water, while two distinct particle sources are the photic zone and hydrothermal vents. The East Pacific Rise (EPR), harbors both an Oxygen Minimum Zone (OMZ) and a non-buoyant plume formed from a hydrothermal vent located in the ocean floor and is therefore an ideal environment to explore effects of oxygen, differences between surface derived and hydrothermal plume particles, and the relationships between these differences and particle size.

Previous studies of microbial diversity classify microbes associated with particles into two or three size classes. Furthermore, the classification of microbes in particles has been semi-quantitative, reporting microbial diversity as relative abundance. In this study, we examined microbial diversity using a new approach. First, we used a recently developed fractionation method that separates particles at a finer scale fraction, then we used a spike-in standard to quantify microbial composition using amplicon sequencing, allowing the estimation of numbers of bacteria per unit of volume and weight. Fine scale size fractionation into 7 fractions allows us to compare free-living microbes to those on small, intermediate and large particles. Finally, we size fractionated organic matter into the same fractions as the DNA, so the amount of particles in each size fraction was known.

We observed that most bacteria were free living or on the smallest particle fractions. This appeared to be because the smallest particle size fractions harbored the most mass, carbon and nitrogen. Community structure varied between our samples of different sizes and collected at different depths. Quantitative abundance estimates found that ostensibly obligate free-living bacteria, such as SAR11, were more abundant in the free-living fraction but also present in abundance in the particulate size fractions and ostensibly obligate particle attached bacteria such as Flavobacteria and Planctomycetes were actually equally abundant in the free-living fraction and suspended particles. Such patterns would not have been evident if we had looked at relative abundance data only. Additionally, it was clear that the total and group specific abundances increased in the OMZ region, and that many species were unusually abundant in the non-buoyant plume. Thus, quantitative examination of microbial communities suggest that most microbes exist both in particles and in the water column.

Introduction

In the upper ocean, phytoplankton such as cyanobacteria and microalgae use sunlight to convert CO_2 into organic molecules via photosynthesis. Photosynthesis is a key process in the ocean which is responsible for about 50% of the primary production on earth (Guidi et al., 2009). A fraction of the CO_2 incorporated into these organic molecules sinks to the deep ocean, removing it from the atmosphere for several centuries, and a small fraction of that sinking carbon is entrained in the sea floor removing that carbon from the atmosphere for millions of years (Le Moigne, 2019). The sinking of organic matter into deep waters constitutes a process known as the gravitational biological carbon pump (BCP)(Le Moigne, 2019). Ocean regions are connected vertically by marine particles that transport microorganisms from the surface towards depth (Mestre et al., 2017).

However, not all particles sink, especially those particles smaller than <53 μm (Clegg & Whitfield, 1990). Chemical data indicates that suspended and sinking sediment trap material have different isotopic, lipid, pigment and amino acid compositions (Abramson et al., 2010; Altabet, 1988; Wakeham & Canuel, 1988). Older microbial community fingerprinting techniques indicated that communities on sinking particles from sediment traps were significantly different from communities on suspended particles at depth, with trap material more reflective of surface waters (Tamelander, 2013). However, there is connectivity between sinking and suspended particles with disaggregation of sinking particles forming suspended particles (Alldredge, 1998) and aggregation of suspended particles to form sinking particles (Cavan et al., 2018).

Particulate Organic Carbon (POC) forms through several mechanical and biological processes including zooplankton ingestion/excretion and molting, aggregation of phytoplankton or of smaller particles into marine snow (Simon et al., 2002). Different members of the phytoplankton group contribute differently to the POC export to deep ocean; for example, modeling of carbon export has shown that diatoms could be contributing to 40% of carbon export whereas coccolithophorid contribution reaches 10% (Jin et al., 2006). Picocyanobacteria can also contribute to export, but by forming small particles that are aggregated into larger particles below the mixed layer

(Cavan et al., 2018; Fuchsman et al., 2019). The size, and chemical composition of marine particles will depend on the combination of factors that contributed to their formation.

Marine particles can be colonized by microbial communities that use the organic matter for microbial respiration. From metagenomic data, sinking particles are enriched in bacteria that have extracellular peptidases, which degrade proteins, and carbohydrate degrading enzymes as well as large numbers of ABC transporters (Leu et al 2022). In the presence of oxygen, the organic matter in marine particles can be consumed via aerobic respiration. In contrast, in the absence of oxygen, the organic matter can be respired via anaerobic respiration such as denitrification, nitrate reduction or sulfate reduction (Lam et al 2009; Fuchsman et al., 2017; Saunders et al., 2019). The

presence or absence of high oxygen concentrations in the water appears to impact the microbial diversity found in particles (Fuchsman et al., 2011). However, particles themselves have a diffusive boundary layer, and thus can have zones of anoxia despite being in oxic waters (Bianchi et al., 2018; Ploug et al., 1997). Oxygen consumption due to the degradation of organic matter in particles leads to the development of microenvironments within the particles that can support metabolisms that would not be able to take place in the water column, such as denitrification in hypoxic waters (Bianchi et al., 2018; Fuchsman, Paul, et al., 2019) and sulfate, and metal reduction (Raven et al., 2021; Saunders et al., 2019) in nitrogenous anoxic zones.

Most studies tend to assign microbial diversity to two main categories; free-living (FL) or particle-attached (PA). Groups consistently enriched on particles include Planctomycetes, Rhodobacterales (Alphaproteobacteria), Cytophaga-Flavobacteria-Bacteroidetes and specific Gammaproteobacteria (Eloe et al., 2011; Fuchsman et al., 2011, 2012; Ganesh et al., 2014; Lecleir et al., 2014; Leu et al., 2022). It has been found that while some taxa can be classified as strictly FL or PA, other taxa are present in both (Liu et al., 2018). However, in the ocean, marine particles exist in a continuum of sizes through the water column (Nguyen and Harvey, 1997). A simple free-living or particle-attached characterization misses the fact that diffusion into particles is directly related to particle size and surface area to volume ratio (Ploug et al., 1997), and therefore the interior of large particles may have different redox and nutrient concentrations than found in smaller particles (Ploug et al 1997). Additionally, particle size directly affects sinking rates and by extension is thereby related to the origin and age of the particle.

A recent approach to fractionating particles classified them according to 5 size classes and showed that different sized particles host different bacterial communities. The diversity was highest in the larger particles (20-200 μm), presumably because bigger particles can provide more niches for microorganisms (Mestre et al., 2017). Mestre et al., (2018) found that communities on 3-20 μm and 20-200 μm size fractions at depth were distinct with the 20-200 μm size fraction more reflecting surface communities. Observations in the lab (Bižić-Ionescu et al., 2018; Datta et al., 2016) and models (Nguyen et al., 2022) suggest that microbial communities likely vary from particle to particle even within size classes.

Study Site

The East Pacific Rise (EPR) study site is of particular interest due to the presence of both an Oxygen Minimum Zone in the mesopelagic and a hydrothermal plume in the bathypelagic in the same water column.

Oxygen Deficient Zones (ODZs) are regions of the ocean that have undetectable oxygen (Revsbech et al., 2009; Tiano et al., 2014). Ocean ODZs host 30-50% of marine fixed-N loss (DeVries et al., 2013). In ODZs, bacteria and archaea use nitrate and nitrite as an electron acceptor instead of oxygen (Lam et al., 2009), leading to N_2 production by heterotrophic denitrification and anammox (Babbin, 2014; Thamdrup, 2012). Because the

oxygen content of the Pacific Ocean has been decreasing since the 1980s (Ito et al., 2017), so in consequence ODZs are expanding (Stramma et al., 2008).

ODZs have a large anoxic core (O_2 <10 nM) surrounded by shells of suboxic (O_2 <5 μM), severely hypoxic (O_2 <20 μM) and then hypoxic waters (O_2 <60 μM). In the geographic core of the ODZ, the vertical width of anoxic waters can be >700 m. However, the vertical width of the ODZ is much smaller at the edges of the ODZ region. The East Pacific Rise is a hydrothermal vent field that underlies the Southern edge of the Eastern Tropical North Pacific (ETNP) ODZ (Fig, 1, Chapter 0). It is unclear if the oxygen minimum zone in this edge region is truly anoxic or if oxygen is present but not measurable with Seabird sensors typically found on CTD rosettes (~1 μM oxygen). In regions and datasets where determination of oxygen concentration with a STOX sensor (Revsbech et al., 2009) or equivalent is not available, Banse et al., (2017) has determined that the presence of 0.5 μM nitrite or greater indicates functional anoxia. Thus, we made sure to measure nitrite concentrations in this work.

Deep sea hydrothermal vents, are regions in the ocean floor where ocean water that has percolated to the subsurface of the ocean floor gets heated and enriched with chemicals and gas, and then forms a fluid that is released through orifices in the seafloor to the cold deep ocean waters (Dick, 2019). The hydrothermal vent plume at the bottom of the ocean floor at the EPR affects the chemistry and particle composition of the area surrounding the vent and the water column at depth (Dick, 2019). The analysis of hydrothermal plumes by using sediment traps in the EPR revealed the presence of several microbial groups: Firmicutes, Actinobacteria, Planctomyces, and Epsilon proteobacteria (Sylvan et al., 2012). The Epsilonproteobacteria had the most diversity with members of Sulfurovumaceae or Sulfuricurvaceae, Arcobacteraeceae, Hydrogenimonaceae, and Nautillaceae were found (Sylvan et al., 2012). However, the microbial communities associated with particles that sink from the surface to the water column, reaching the hydrothermal vent are poorly understood.

Marine Particle and microbial abundance

The analysis of microbial amplicon sequences has been limited by the use of relative abundance as an indicator of microbial diversity; it makes it difficult to directly compare the changes in abundance of microbial groups across different samples. Indeed, relative abundance community data have been shown to provide misleading patterns, when conventional statistical approaches are applied with changes in the actual abundances of one organism creating apparent changes in the opposite direction of all other organisms, and leading to both false positive and negative associations (Gloor et al., 2016). Recently, a range of studies in microbial ecology have been adding quantitative elements, such as collecting microscopy data in concert with metagenomic or amplicon based measurements of relative

abundance (Fu et al., 2023; Yeh & Fuhrman, 2022), combining qPCR with amplicon sequencing (Shelton et al., 2023; Tettamanti Boshier et al., 2020), and more recently using internal spikes (Cram et al., 2024; Tourlousse et al., 2016; Wu et al., 2020)). This last newly developed approach has aimed to address this limitation by incorporating an artificially designed 16S rRNA gene sequence that amplifies like other 16S rRNA genes, but that is distinct from all known microorganisms (Tourlousse et al., 2016). The use of spike-in allows assaying microbial diversity quantitatively, allowing the estimation of numbers of bacteria per unit of volume and mass (Cram et al., 2024) . Additionally, fine scale size fractionation into 7 fractions allows us to compare free-living microbes to those on small (0.2, 1, 2, 5 μm), intermediate (20, 53 μm) and large particles (180, 500 μm). Compared to the previous work by Mestre et al., (2017, 2018), we have added two larger size fractions (180-500 μm and >500 μm), increasing our chances of obtaining sinking particles. Thirdly, we size fractionated organic matter into the same fractions as the DNA, so the mass, carbon and nitrogen content of particles in each size fraction were known. The combination of these three techniques has been used in the Chesapeake Bay Estuary where it was found that the majority of particle-attached bacteria were found on small particles due to the fact that small particles greatly outnumbered large particles (Cram et al., 2024). However, the Chesapeake Bay is a shallow coastal system, and is therefore fundamentally different from the deep and oligotrophic open ocean. This chapter is the first use of these techniques in a depth profile in the open ocean.

This chapter addresses how microbial diversity changes among particle sizes in a depth profile at the EPR (in the water column over the hydrothermal vents). First, the changes in the microbial composition are evaluated quantitatively using particles separated according to a range of size fractions, followed by 16S rRNA gene amplicons sequencing using: 16S rRNA gene synthetic standards (spike-in), nutrient analysis, particulate carbon and nitrogen content, and mass analysis to estimate changes among seven different particle sizes.

Methods

Hydrographic data

Samples were collected as part of the Hot2ColdVents expedition https://sylvangeomicrobiology.wordpress.com/blog/page/2/ aboard the R/V Atlantis in April 2019. The expedition aimed to study a station in the EPR located at 9°50'N, 104°18.144'W (Table 3.1).

A CTD-Rosette was used to collect water samples, and attached sensors were used to record oxygen (Seabird 43 Dissolved oxygen sensor), chlorophyll, fluorescence, turbidity, beam transmission, temperature, depth, conductivity and salinity. The hydrographic data from the cruise can be found at https://www.rvdata.us/search/fileset/144040.

The EPR station was sampled in several CTD-casts that targeted 6 different depths. Depths were categorized into the following zones: that were Surface at 4 m, Deep Chlorophyll Maximum (DCM) at 92 m, the mesopelagic (MESO) with two depths —316 m and 490 m, the OMZ with two depths —490 m and 527 m, bathypelagic at 1001 m and Non-buoyant Plume (NBP) at 2369 m (Table 3.1).

Samples from all different casts were combined as they were part of the same water column.

Table 3.1. Date, coordinates, CTD cast number, Sample category, Depth and Volume of water collected (L) for each water sample.

Date	Latitude	Longitude	CTD Cast #	Sample	Depth (m)	Volume water collected (L)
6-Apr-19	9°50.869'N	104°18.144'W	AT4209005	Deep Chlorophyll Maxima (DCM)	92	80
				Mesopelagic (MESO)	275	80
				Surface	4	80
9-Apr-19	9°50.850'N	104°18.157'W	AT4209008	DCM	92	80
				Mesopelagic (MESO)	316	80
11-Apr-19	9°50.874'N	104°18.158'W	AT4209010	OMZ	490	80
				Bathypelagic (BATHY)	1001	160
11-Apr-19	9°50.874'N	104°18.158'W	AT4209012	OMZ	527	120
				Non-buoyant plume (NBP)	2369	120

Sample preparation

Between 80-160 L of seawater was collected at each depth by the shipboard rosette (Table 1). The water was gravity filtered through a stacked series of nylon filters fixed between PVC connectors. The nylon filters were organized and placed one on top of each other in a gradient from the largest pore size at the top towards the smallest size pore at the bottom (500, 180, 53, 20, 5 µm). The diameter of all nylon filters was 142 mm.

By pouring the water through the stack of filters, the particles were separated by size on the nylon filters. After filtration, each nylon filter was back rinsed thoroughly using an air powered power water pistol filled with prefiltered (0.2 µm Pall AcroPak 1500 Capsule with a Supor Polyethersulfone membrane) seawater. The marine particles were removed from the filters and collected into a funnel and fractionated according to particle size (Table 3.2). All the fractions were divided into two parts: one fraction was used for DNA analysis and the second fraction was used to analyze particulate organic matter mass (by using weighted pre-combusted filters). The 5 µm filtrate that passes through all the filters was collected and filtered again through a 1.2 µm (Polyethersulfone filter Sterlitech PES1247100) and 0.2 µm filter using a vacuum manifold. This final 0.2 µm filter collected the free-living microbial fraction.

The resuspended particulate organic matter from each sample and different size classes were split, with 40 ml reserved for microscopy (data not included herein), and the remainder split in half. Each half was filtered using vacuum via a 1.2 µm nominal pore size, 25 mm diameter, GF/C glass fiber filter (Whatman WHA1822025), that had been pre-combusted at 400˚C. One filter was saved for DNA extraction and analysis, while the other was saved for particulate organic matter measurements. The filters used for Particulate Organic Matter (POM) analysis had been pre-weighed using a Sartorius microbalance (precision of 1 µg) these fractions corresponded to the microbial community attached to particles. Filters for DNA analysis were placed inside a 2 mL sterile screw cap cryogenic tube and frozen in liquid nitrogen and stored at -80 ºC until DNA extraction. Samples for POM analysis were stored in foil packs and also stored at -80 ºC.

Table 3.2. Filter sizes, particles range, filter type and treatment used to recover particles from filters.

Filter size (µm)	Particle range (µm)	Filter type	Treatment used to recover particles from filters
500	> 500	Nylon filter (Gilson Inc.)	Back rinse (nylon filter)
180	180-500	Nylon filter (Gilson Inc.)	Back rinse (nylon filter)
53	53-180	Nylon filter (Gilson Inc.)	Back rinse (nylon filter)
20	20-53	Nylon filter (Gilson Inc.)	Back rinse (nylon filter)
5	5-20	Nylon filter (Spectrum Laboratories)	Back rinse (nylon filter)
1.2	1.2-5	Polyethersulfone filter Sterlitech PES1247100	5 µm filtered that passed through the nylon filters using vacuum manifold
0.2	0.2-1.2	Polyethersulfone filter Sterlitech PES1247100	5 µm filtered that passed through the nylon filters using vacuum manifold

Nutrient Analysis

Dissolved nitrate, nitrite, and phosphate from the water samples were analyzed at the Nutrient Analytical Services Laboratory at the Chesapeake Biological Laboratory (CBL) where Nitrite plus nitrate (Nitrogen) were measured by using the Method EPA 353.2. Phosphate by using method EPA 365.1, and Nitrate with the method ASTM D-7781-14. These values were used to calculate N deficit as described in (Chang et al., 2012). N deficit, a variation on N* that uses N:P ratios from nearby oxic waters, which are ~14, rather than assuming a ratio of 16, was calculated from these values.

Particulate Organic Matter analysis

GF/C filters (pore size 1.2 µm) were combusted and pre-weighed. A fraction of the 500,180, 53, 20, 5, 1.2 µm samples were filtered onto these pre-combusted filters. After the cruise, POM on filters was dried at 40ºC for >24 hours and weighed. Then samples were wafted with HCl for 24 hours to remove carbonates and sent to the UC Davis Stable Isotope Facility to obtain C and N concentrations and the natural stable isotopes of $d^{15}N$ and $d^{13}C$. Isotope data are not shown in this paper. Samples were processed for mass spectrometry following their Difficult Combustion of Solid Samples Protocol. Both carbon and nitrogen were quantified in an Elementar Vario EL Cube coupled with an Isoprim VisION isotope ratio mass spectrometer.

DNA Analysis

As described in detail in Cram et al. (2024), frozen samples were extracted from DNA filters using a phenol-chloroform method. A difference from that protocol was that these samples were on glass fiber filters, while the Cram et al., (2024) samples were on Supore filters. In this case, lysis buffer was added to the samples, they were incubated, and the lysis buffer was removed from the filters by squeezing the filters with a pipette tip against the side of the tube. The DNA concentration of the samples was measured using a Qubit 4.0.(ThermoFisher). A sample aliquot was then diluted to a final concentration of 2 ng/µl using PCR grade water. Samples were then preserved at -80 °C to further analysis.

Spike-In preparation

A synthetic 16S rRNA gene sequence (LC140931.1 Synthetic 16S Spike-In (Tourlousse et al., 2016) was synthesized as gBlocks Gene Fragments (Integrated DNA Technologies, Inc.). The gBlock was resuspended in a solution of 0.1mg/ml of PolyA (Roche Diagnostics Deutschland GmbH) in PCR grade water (DEPC-Treated Water Ambion™) and then serially diluted to reach a final concentration of 10^6 copies/µl following the manufacturer's instruction.

DNA concentration per sample was diluted using PCR-grade water to reach a concentration of 2ng/µl DNA (in the case of samples with higher DNA concentrations); then the spike resuspended in PolyA was added to each sample to reach a final concentration of 10^4 copies/µl or $5*10^3$ copies/µl depending on the sample.

DNA sequencing

The DNA extracted from the free-living fraction and the particle associated bacteria, was sequenced with Illumina MiSeq sequencing by a third-party sequencing company, Molecular Research (MRDNA) Shallowater, Texas, USA. In brief, at MRDNA, 1 ng of sample DNA plus the 10^4 or $5 * 10^3$ copies of spike were amplified using universal 515FY-926R primers (Parada et al., 2016), which have been shown to have low amplification bias and broad coverage. It is important to note these sets of primers do not target anammox bacteria (McNichol et al., 2021). The samples were processed following MRDNA's in house pipeline for amplicon sequencing on an Illumina MiSeq instrument. Samples were returned to us as FASTA files.

Amplicon Bioinformatics

The EPR sequenced amplicons were processed following the same pipeline described in Cram et al. (2024), which will be briefly summarized here. The sequenced amplicons were processed using the DADA2 pipeline (Callahan et al., 2016). The primers were removed from the sequences using Cutadapt (Martin, 2011), with the following modification, the filterandtrim() algorithm was used to retain sequence with < three errors in the forward read and fewer than five errors on the reverse read (Cram et al., 2024). Then DADA2 was used to dereplicate, error rates, and merge forward and reverse reads. The taxonomy of the ASV was assigned using the Silva database version 132 released Dec 2017 (Quast et al., 2013; Yilmaz et al., 2006) This database was modified by adding the spike-in sequences (Cram et al., 2024). Adding the spike-ins allowed us to classify sequences into Archaea, Bacteria, Eukaryotic and Spike-in Sequences. The data was analyzed using the R software for statistical analysis (V4.0.2) using the packages Vegan, Phyloseq for data analysis, Dplyr for data curation and ggplot2 to visualize data.

Calculating taxon abundance

The calculation of the taxon abundance (16S + 18S rRNA gene copies/L) of each microbial group in each sample was estimated by normalizing with the spike in read numbers, DNA extracted and volumes filtered, following Equation 1.

(Equation 1)

$$\text{Taxon Abundance (16S + 18S Copies/L)} = \frac{\text{ASV Reads}}{\text{Spike Reads}} * \frac{10^5 \text{ Spikes Copies}}{\text{1ng DNA}} * \frac{\text{DNA Extracted (ng)}}{\text{Volume Filtered (L)}} * \frac{\text{Total Volume of Rinse water (L)}}{\text{Volume Rinse water used for DNA filter (L)}}$$

Calculating normalized taxon abundance

The abundance of taxon was normalized to the width of the particle size fraction bins in order to better characterize the shape and slope of the size to abundance relationships as described by Cram et al. in 2024, using the equation 2.

(Equation 2)

$$\text{Normalized Taxon Abundance (copies/L/μm)} = \frac{\text{Taxon Abundance (Copies/L)}}{[\text{Size Class Upper Bound (μm)}] - [\text{Size Class Lower Bound (μm)}]}$$

In equation 2 we divided the Taxon abundance (Copies/L), by the difference between the upper and lower size classes which is the normalized bin size. That difference accounts for the large variation in particles sized between each filter size. A scale depicting the difference in range can be seen in Supplementary Figure 3.6.

The quantitative examination of microbial amplicons in particles using spike-in allowed for calculating the number of bacterial gene copy numbers (16S rRNA gene copies/L/μm). To interpret our findings, we made the assumption that a microbial cell contains only 1 copy of the 16S rRNA gene; this allows us to interpret our result as direct microbial abundance, as 1 copy of 16S rRNA gene corresponds to one cell. It has been noted that microbial abundance is directly related to bacterial cell copy numbers. as it has been shown in Cram et al. (2024).

Calculating particle mass

The particle mass per each sample (from the comparison of the f pre- and post-weights (Cram et al., 2024) was normalized to the volume of the water that went through the stacked set of nylon filters and the fraction of water that was rinsed through the GF/F filter. The total particle mass was calculated following equation 3.

(Equation 3)

$$\text{Particle mass (mg/L)} = \frac{\text{Filter Post Weight (mg)} - \text{Filter Pre Weight (mg)}}{\text{Volume Filtered (L)}} * \frac{\text{Total Volume of Rinse water (L)}}{\text{Volume Rinse Water Used for POM Measurement (L)}}$$

Calculating taxon abundance per milligram particle mass

Microbial abundance was further normalized to particulate material by dividing microbial concentration by POC concentration. In order to estimate the number of cells per mg of particle mass, the taxon abundance was normalized by using the change in particle mass (Equation 4).

(Equation 4)

$$\text{Taxon Abundance (copies/mg)} = \frac{\text{Abundance (Copies/L)}}{\text{Particle Mass (mg/L)}}$$

Redundancy analysis

We used ANOVA like Permutation Test for Redundancy Analysis (rda) and a Constrained Analysis of Principal Coordinates. We used R-Studio, with the Package Vegan, to analyze the quantitative relationship between the 16S

Amplicon sequencing fractionated according to: particle size, (0.2, 1.2, 5, 20, 54, 189, 500 μm), and according to depth (m).

Results

Oxygen and nitrite concentration in the EPR

The oxygen concentration in the top 100 m of the water column, as measured by the Seabird 43 oxygen sensor, was ~200 μM (Fig. 3.1 a). Oxygen decreased abruptly to <20 μM in the region between 100-120 m depth. Between 100-400 m depth, the oxygen concentration remained below 15 μM in the area between 400-700 m, the oxygen concentration reached the lower detection limit for the oxygen sensor (detection limit ~1 μM).

The concentration of Chlorophyll A, measured as relative fluorescence units (rfu), was low in the top 10 m (~0.1 rfu). There were two fluorescence peaks, one of close to 1.3 rfu between 60- 80 m depth (Fig. 3.1 a). Below 80 m depth, the concentration of Chlorophyll A decreased, coinciding with the decrease in oxygen concentration around the same depth. At 120 m depth a secondary peak of Chlorophyll A was observed, although the concentration measured was low (~0.2 rfu). Chlorophyll A was not detected below at 200 m.

The presence of nitrite was observed in the water column in the EPR (Fig. 3.1 b). Nitrite accumulation was observed at two peaks. The primary nitrite maximum was observed at 103 m, where nitrite reached 0.71 μM. A Secondary Nitrite Maximum (SNM) was measured at 0.32 μM at 530 m. Banse et al., (2017) uses 0.5 μM nitrite as a cut off for using nitrite to indicate true anoxia. The Secondary Nitrite Maximum found here is close but less than this cut off. In addition to nitrite accumulation, an area of nitrogen deficit was observed at 530 m depth in the ODZ region, with 10.9 μM of deficit (Figure 1) (Banse et al., 2017) This N deficit value is typical for those measured in the ETNP ODZ (Fuchsman et al., 2018).

Nitrate and phosphate were also measured in the EPR water column; in the case of nitrate, the abundance was low (<10 μM) in the top 100 m depth (Supplementary Figure 3.7); however, it increased to 40 μM at ~100 m depth, and continued increasing with depth to a peak ~55% at 100 m depth and then decreased with depth. Phosphate was measured and generally has a depth profile pattern similar to nitrate. However, the abundance of phosphate was generally <4 μM (Supplementary Figure 3.7).

Particle total and carbon and nitrogen mass

When we examined total particle mass (Fig. 3.2), we observed that the smaller particle fraction contained more mg particles/L/μm than larger particle sizes. For example, in the surface at 4 m depth, the mass of particles in the small fraction (5 μm) was higher 10^{-2} (mg Particles)/L/μm, and decreased almost linearly as a particle increased in size, reaching ~10^{-4} (mg Particles)/L/μm in the fraction 500 μm. When we observed changes in particle mass, we observed a decrease in mass in the bathypelagic (1000 m depth) where particle abundance for the small size fraction was 10^{-3} (mg Particles)/L/μm decreasing as particle size increased. However, at the deepest sample at the 2369 m depth (non-buoyant plume), we observed an

increase in the particle mass, to a mass similar to the particle mass in the OMZ region.

Organic Carbon concentrations were highest at 275 m, but were also higher in the surface and DCM than at 400 m and below (Supplemental Figure 3.8). Most of the carbon was in the 5 µm and 20 µm fractions (Supplemental Figure 3.8). When normalized by bin width, the carbon (Fig. 3.3 a) and nitrogen content was higher in small particle sizes and decreased as particles increased in size. In the case of carbon content, for example, at 4 m depth it was $<10^0$ µg/L/µm and decreased almost linearly as particles increased in size. At 4 m the total N µg/L/µm was ~$>10^{-2}$ µg/L/µm (Fig. 3.3 b), however in the largest particle fraction (≥ 500 µm), the N abundance decreased to 10^{-4} µg/L/µm. This pattern of high abundance in smaller particles rather than larger particles, was observed in all depths sampled, from the surface to the bathypelagic and even in the Non-buoyant plume.

Quantitative examination of microbial amplicons in particles

Bacterial 16S rRNA gene copies/mg abundance was higher in the surface top 100 m depth with $>10^9$ copies/mg, decreasing with depth to ~10^6 copies/mg for the small particle sized ≤5µm in the mesopelagic on top of the OMZ region 400-700 m depth (Fig. 3.4). In contrast, for particles sized ≥53 the abundance of 16S rRNA gene copies was higher ~10^8 copies/mg in the OMZ core than in the mesopelagic (~10^7 copies/mg) on top of the OMZ. In the OMZ for particles ≥180 µm, we found that Archaea abundance dominated the OMZ region $>10^8$ copies/mg, whereas Bacteria abundance was $<10^7$ copies/mg. In the NBP (at 2,369 m depth), the bacterial abundance did not present a clear pattern. Archaea abundance was lower than bacteria abundance by one or two orders of magnitude in all sizes, and the depth profiles had similar shape to the bacteria one (Fig. 3.4).

Abundance of bacterial taxa on particles

The most abundant bacterial phyla on particles were Bacteroidetes, Cyanobacteria, Firmucutes, Planctomycetes and Proteobacteria (Fig. 3.6). Firmicutes and Cyanobacteria were the phyla whose abundances changed the most with depth. Cyanobacteria abundance decreased consistently with depth, going from 10^9 16S rRNA gene copies/L/µm in the surface to 10^5 16S rRNA gene copies/L/µm in the mesopelagic region, for particles of sizes 0.2 and 1.2 µm. The decrease was almost linear for particles of size 5 µm changing from 10^7 16S rRNA gene copies/L/µm in the surface to $<10^4$ in the mesopelagic. Among Archaea, the phylum with the highest 16S rRNA gene copies/L/µm corresponded to Thaumarchaeota (Fig. 3.7), where the abundance also changes with depth and across particle sizes. For example, in the free-living size fraction (0.2 µm) the 16S rRNA gene copies/L/µm of Thaumarchaeota increased from ~10^6 16S rRNA gene copies/L/µm on the surface to ~10^8 16S rRNA gene copies/L/µm in the OMZ region, and a similar pattern was observed in particle size 1.2 µm. However, in particle 5 µm, the abundance of Thaumarchaeota decreased with depth from ~10^6 16S rRNA

gene copies/L/μm close to the surface to ~10^4 16S copies/L/μm in the OMZ region to <10^2 16S rRNA gene copies/L/μm in the bathypelagic region (1000 m depth). This was the only particle size that showed such a steep decrease. In the OMZ region we observed an increase only in Thaumarchaeota abundance in the free-living fraction ~10^8 16S rRNA gene copies/L/μm. In particle sizes ≥20 μm the Thaumarchaeota abundance remained fairly constant (~10^4–10^6 16S rRNA gene copies/L/μm) in the water column including the bathypelagic region (1000 m). The abundance of Archaea in the non-buoyant plume depth was different at each particle size: It was high ~10^8 copies/L/μm in the small size fractions (from 0.2 to 20 μm). In contrast, in the 53 μm and 180 μm size fractions, there were almost no Thaumarchaeota. Finally, on the >500 μm size fraction, there were ~10^5 16S rRNA gene copies/L/μm.

Proteobacteria abundance

Among the Proteobacteria, the most abundant class are the Gammaproteobacteria with ~10^8 copies/L/μm followed by Alphaproteobacteria in all fractions except free living (0.2 μm), in the free-living fraction Alphaproteobacteria abundance was > 10^8 copies/L/μm (Fig. 3.8). Gammaproteobacteria and Alphaproteobacteria appeared to co-vary in abundance (Fig. 3.8). Gammaproteobacteria were higher, by up to one order of magnitude than Alphaproteobacteria on all attached size fractions (≥1.2 μm). The abundance of Alphaproteobacteria did not show a consistent trend with depth, but rather varied substantially from 10^6~10^8 copies/L/μm in the water column from the surface towards the OMZ and bathypelagic region in small particles <1.2 μm, and in the larger particles >180 μm. In the 5 μm size class Alphaproteobacterial abundance decreased almost linearly from the surface ~10^8 to ~10^4 copies/L/μm in the Bathypelagic region. Although particles in sizes 1.2 and 53 μm varied in the same range (~10^8 to ~10^4 copies/L/μm) as 5 μm particles, there was not a clear pattern. Deltaproteobacteria were present throughout the water column, and in all size fractions at lower abundances than the Gammaproteobacteria and Alphaproteobacteria. The abundance of Deltaproteobacteria decreased as particles increased in size, for example in small particles (≥1.2 μm) the abundance varied ~10^6 to ~10^8 copies/L/μm but, at fraction 180 μm the abundance was between ~10^2 to ~10^4 copies/L/μm. Members of Zetaproteobacteria were only found in low numbers <10^4 as free living (fraction 0.2 μm) and associated with particles at 20 and 53 μm in the OMZ region (Fig. 3.8).

Planctomycetes abundance

The majority of taxa from the Planctomycetes phylum are known particle attached bacteria except for some anammox bacteria (Fuchsman et al., 2012). Examination of the Planctomycetes phylum as a whole, indicates that the samples with the most Planctomycetes were the free-living fractions in the OMZ region (~10^8 copies/L/μm) (Fig. 3.6). In the surface waters, when

examined in comparison to Proteobacteria (>10^9), there were few Planctomycetes (~10^6). At non-OMZ depths, Planctomycetes remained at (~10^6) in particles <53 μm while Proteobacteria decreased to (~10^7). In larger particles, abundances shifted to ~10^5 and ~10^4 (Fig. 3.6). Thus, in terms of relative abundance, Planctomycetes were enriched in particles, but in actual abundances, they were similar between free-living and suspended particles, but lower in abundance in large particles.

We looked into the abundance of classes inside Planctomycetes. We note that the primers used do not amplify anammox bacteria (McNichol et al., 2021). The class Planctomycetacia, which includes Pirellulaceae and Planctomycetales orders among others, has the highest abundance in all particle sizes, with highest abundance in the free living (0.2 μm) and smallest particle sizes 1.2 μm with 10^4~10^8 16S rRNA gene copies/L/μm (Fig. 3.11). The Planctomycetacia abundance did not present a defined profile. Like the Planctomycetes as a whole, we observed an increase in the 16S rRNA gene copies/L/μm number in the OMZ region for the free-living (<0.2 μm fraction) (Fig. 3.11). However, in the ≥500 μm fraction size the abundance of Planctomycetacia decreased to 10^2 - 10^4 16S rRNA gene copies/L/μm. This taxon, along with most others, was evident in the free-living fraction as well. Other taxa generally tracked the abundance of Planctomycetacia, but at lower abundances on the attached fractions. An exception is the class BD7-11 which is abundant throughout the water column on particles 53 μm and larger, but absent or rare in the free-living fraction and smaller particles. Contrastingly the class Phycisphaerae is rare in sizes fraction 53 μm and larger, but abundant in smaller size fractions. In the non-buoyant plume, different classes of Planctomycetes appear to dominate different size fractions, in contrast to the main water column, where most phyla are abundant on multiple size classes.

SAR11 abundance

SAR11 bacteria are an order level clade of Alphaproteobacteria that are thought to be free-living (Giovannoni et al., 2005). In general, the total SAR11 abundance was higher in the smallest size classes (0.2 μm; free-living) where it varied from 10^7 to 10^9 16S rRNA gene copies/L/μm across the water column (Fig. 3.9). Total free living SAR11 had highest abundances in the surface and the OMZ, as seen previously in the ETNP (Fuchsman et al., 2017). SAR11 and Rhodobacteriales were the most abundant Alphaproteobacteria present on all sizes of particles (Fig. 3.9). SAR11 had highest abundance on 1.2 μm particles (10^5 - 10^6 copies/L/μm) with lower abundance on larger particles (10^4 copies/L/μm on >500 μm particles) (Fig. 3.9).

We looked into the abundance of members of the SAR11 Clade on particles (Fig. 3.10). Within SAR11, we detected four clades (Clades I, II, III, and IV) as well as Unclassified SAR11 that did not fall into any clade (Fig. 3.10). The clade that had the highest abundances both in the free-living and particle fractions was Clade I, closely followed by Clade II. Both clades were found at

almost every analyzed depth, whereas Clades III, IV, and the unclassified SAR11 were consistently lower in abundance. Clade I reached a maximum of >10^8 16S rRNA gene copies/L/μm in the free-living fraction (0.2μm) at 4 and 92 m depth (Fig. 3.10), decreased in the mesopelagic to ~10^6 copies/L/μm but then in the OMZ region increased to >10^8 copies/L/μm. The abundance of all SAR11 clades decreased as particle size increased, for example in particles >500 μm the abundance of all the SAR11 Clades was lower than 10^4 copies/L/μm.

Nitrospina abundance

We also examined the abundance of nitrite-oxidizing *Nitrospina* in our samples due to its ecological relevance in the nitrogen cycle. Nitrospina abundance was higher in the free-living fraction (0.2 μm) specially inside the OMZ region, where the Nitrospina abundance increased to 10^7 copies/L/μm (Supplementary Fig. 3.2), whereas in the surface and in the non-buoyant plume the abundance was one order of magnitude lower ~10^6 copies/L/μm. The abundance of Nitrospina decreased one order of magnitude in the larger particle sizes, at 1.2 μm the highest abundance was in the OMZ with ~10^6 copies/L/μm, and in the larger particle size (≥20 μm) the Nitrospina abundance decreased to <10^4 16S rRNA gene copies/L/μm.

Relationship between depth, size and community structure

Principal Component Analysis (PCA) from the Taxonomic diversity compared against depth and size fraction generally suggested that large particles (≥180 μm) appear to be more similar to each other, regardless of location, and that smaller particles appeared to be more dissimilar between depth regions (Fig. 3.12). However, permutational analysis of variance indicated no consistent effect of depth ($p = 0.593$) nor size fraction ($p = 0.542$) on community structure, indicating variability between samples that was more complex than could be predicted by a linear function of size and depth. Moreover, more complex interaction effects (interaction term between size and depth; $p = 0.499$) were also not statistically significant, likely reflecting both inter-sample variability and a relatively small sample size. Because the non-buoyant plume is a profoundly different environment from the other samples, we performed a similar analysis in which it was removed from the analysis (Supplementary Figure 3.5). Again, ANOVA results indicated a non-significant relationship between size plus depth and community structure (size marginal $p = 0.295$, depth marginal $p = 0.055$). Additionally removing the bathypelagic samples resulted in a similar non-relationship.

Discussion

Marine particles constitute habitat that can be colonized and inhabited by microorganisms (Azam & Malfatti, 2007). The EPR region examined here had a defined low oxygen zone OMZ (400-700 m depth), and in the deepest region sampled reached a non-buoyant plume (2369 m depth) above a hydrothermal vent. We use a size fractionated quantitative amplicon dataset to examine changes in microbial community and quantitative abundance across particle size and depth.

The low oxygen zone had oxygen concentrations below the detection limit of the Seabird oxygen sensor on the CTD (~1 μM), harbored a N deficit and a secondary nitrite maximum, suggesting the presence of denitrification (Fig. 3.1). The secondary nitrite maximum was only 0.3 μM, which is close to but below the threshold of 0.5 μM suggested to confirm anoxia by (Banse et al., 2017). However, our nutrient sampling was sparse with 1 point in the OMZ, so depths with higher concentrations could easily have been missed. The N deficit was in the same concentration range as previously seen in the core of the ETNP ODZ (Fuchsman et al 2018). With this dataset, we cannot prove that the N deficit signal was produced in situ in the EPR region rather than advected from the ETNP core. However, we strongly suspect that the OMZ at this station was anoxic with anaerobic metabolisms dominating over aerobic metabolisms.

Marine Particles

Organic carbon and nitrogen concentrations decreased with depth. Organic matter is produced in the euphotic zone from photosynthesis. The decrease in the carbon and nitrogen content with depth can be explained by microbial respiration in the particles that consume carbon and nitrogen in its labile form, leaving more recalcitrant forms of organic matter to sink to the deep ocean.

The particle mass was higher in the smaller particle size classes and decreased as particles increased in size (Fig. 3.2). For example, for particles in the surface (4 m depth) the carbon content at 5 μm particle size was higher ~10^{-2} mg particles/L/μm compared to the content at 500 μm particle size with ~10^{-4} mg particles/L/μm. This pattern was similar in different regions of the water column including the surface and at deeper regions. While particle number counts were not obtained on this cruise, this phenomenon is probably due to there being more numbers of small particles than large particles, as seen nearby in the core of the ETNP ODZ (Cram et al., 2022).

Microbial abundance was also several orders of magnitude higher in particles of 1.2 μm than in larger sizes. We observed that small particles contained a higher microbial abundance >10^8 16S rRNA gene copies/L/μm that generally decreased towards depth. In larger particle sizes ≥53 μm we observed a decrease in the abundance from the surface (10^6 to 10^4 16S rRNA gene copies/L/μm) towards the OMZ (Fig. 3.4), but in the OMZ the abundance increased, and it was similar in abundance to the bathypelagic sample (10^8 16S rRNA gene copies/L/μm).

We observed a general non-size dependence of bacterial abundance per mg of particles (Fig. 3.4), indicating there is a consistent relationship between numbers of bacteria and amount of particles. These data are consistent with observations in the Chesapeake Bay (Cram et al., 2024). Together these observations reflect the observed power law relationship in which there were orders of magnitude more mass within smaller particles than larger ones. Such a relationship was also seen in the substantially more eutrophic, but also anoxic Chesapeake Bay (Cram et al., 2024). As marine snow particles are believed to follow power law size to abundance relationships throughout the oceans (Cram et al., 2018; Guidi et al., 2008), it is likely that this pattern where most particle attached microbes are on the smallest particles is a global phenomenon.

Ecology of sinking particles in the EPR OMZ

In our dataset, the most abundant bacterial phyla on particles were Bacteroidetes, Cyanobacteria, Planctomycetes, Gammaproteobacteria and Alphaproteobacteria (Fig. 3.6). These results are generally comparable to those found previously, with some key exceptions. Boeuf et al., (2019) found Epsilonproteobacteria, and Gammaproteobacteria dominating particles at 4000 m depth at HOT in the North Pacific subtropical gyre. The Analysis of particle attached microorganism of metagenomic samples in the ALOHA station, which examined sediment trap sampled from 150 m depth, found that the most abundant phyla were Alphaproteobacteria, Gammaproteobacteria, Planctomycetota, Bacteroidota, and Verrucomicrobiota (Leu et al., 2022). Thus, the shallower sediment traps showed more phylum level diversity. In the ETNP ODZ, analysis of metagenomes of fractionated particles showed that in fractionated particles >30 μm, the microbial community was dominated by Deltaproteobacteria, Planctomycetes, Verrucomicrobia, Flavobacteria (Bacteroidetes), MGII Archaea (Fuchsman et al., 2017).

We did not find Epsilonproteobacteria in abundance on shallow or deep particles in this dataset Boeuf et al., (2019) found Epsilonproteobacteria (order Campylobacterales), and Gammaproteobacteria (order Alteromonadales) dominating particles at 4000 m depth at HOT in the North Pacific subtropical gyre. Additionally, Epsilonproteobacteria have been particularly enriched in hydrothermal vent plumes (Sylvan et al., 2012). In our 2369 m depth particles, Epsilonproteobacteria, were barely present at $<10^4$ copies/L/μm, but Alteromonadales were present in high numbers 10^8 copies/L/μm. The lack of Epsilonproteobacteria was surprising since they have been described as part of deep sea ecosystems and hydrothermal vent plumes (Boeuf et al., 2019; Sylvan et al., 2012), and should have amplified with the primers used (McNichols et al 2021).

Additionally, Verrucomicrobia were not abundant in our EPR particle samples, but were abundant in particle metagenomes from the ETNP ODZ and from shallow HOT sediment traps (Fuchsman et al 2017; Leu et al 2022). Verrucomicrobia may not be abundant in the EPR system. However, members of the Planctomycetes-Verrucomicrobia-Chlamydia supergroup

often have primer mismatches with universal primers (Fuchsman et al 2012), so it is possible that key Verrucomicrobia were missed with this amplification approach.

Quantifying microbial abundance in marine particles quantitatively

We used the unit 16S rRNA gene copies/L/μm, which correspond to 16S rRNA gene copies normalized to volume filtered (L), normalized to bin width (μm). The range of μm between filters varied significantly, from 1 μm to 320 μm (Supplementary Fig. 3.6). Normalizing to bin width allows us to examine and compare direct distribution patterns without this bias. Without the normalization to bin size, we noticed that the depth profile changes and the abundance of 16S rRNA gene copy numbers/L was 10^{10} 16S rRNA gene copy numbers in the larger (>500 μm) particle size, which was similar to numbers calculated for the free-living fractions (Supplementary Fig. 3.9). While when the same data was examined in copies/L/μm, the free-living fraction was more abundant (10^8) than the >500 μm fraction (10^6) (Fig. 3.8). This difference is because the free-living fraction had a bin width of 1 μm and the >500 μm fraction had a bin width of 500 μm.

We looked at microbial diversity using 7 particle size ranges and spike-in standards to estimate quantitative microbial abundance as 16S rRNA gene copies/L/μm. Our combined approach allowed us to look into microbial distribution in particles quantitatively, allowing us to directly compare the abundance of microbial across depth, and across different particle size ranges. When we look at the variation of individual groups, we can see changes taking place across depth and particle sizes. For example, in the free-living fraction, there was an increase in the abundance in the OMZ region of Proteobacteria >10^8 copies/L/μm, but we can also see the changes in the lower abundance Planctomycetes across depth in the free-living fraction, including a maximum in the OMZ (Fig. 3.6).

If we compare our combined approach to traditional approaches which use relative abundance to study microbial communities, we can observe key differences. We calculated abundance as relative abundance and constructed a phylum level plot similar to the plot we used to present our quantitative data and an even more traditional stacked bar plot (Supplementary Figure 3.3 and 3.4). Proteobacteria were the dominant group at each particle size and across all depths, and we can see changes in Proteobacteria abundance (reaching almost 1 in the particles ≥500 μm). However, the less abundant groups are clustered together and are hard to differentiate in the stacked bar plot because their abundances are so close to zero. The variation of these smaller groups appears to be determined by variations in the large Proteobacteria group. With relative abundance, we lost the information about the presence and variation of the low abundance group, like Planctomycetes, compared to the most abundant group (Proteobacteria).

Considering traditional free-living and particle attached bacteria quantitatively

Marine particles were inhabited by both, sessile Planctomycetes, as well as by free-living SAR11 and Thaumarchaea. We observed that microbial groups like SAR11, which were highly abundant in the free-living fraction (0.2 μm), were also present in larger particle sizes, but in lower abundances (10^4 copies/L/μm on >500 μm particles) (Fig. 3.9). SAR11 has been previously described as a free-living organism (Giovannoni, 2017) In other studies, like in the Meridional Atlantic, SAR11 was also found in the free living fraction (Defined between 0.2 - 3 μm in Reintjes et al.,(2023)). This is in concordance with the high abundances in the free-living fraction we observed in our samples (Fig. 3.9). However, in the EPR SAR11 was often the most abundant Alphaproteobacteria on particles (Fig. 3.9). Several other papers with size fractionated amplicon data have found SAR11 on particles, casting doubt that SAR11 are only free-living bacteria (Mester et al 2020; Yeh et al. 2022; Cram et al 2024). The presence of SAR11 on particles could suggest they might play a role in the nutrient cycle on these particles. However, it is also possible that abundant free-living microbes are entrained in particles as a result of viral infection (Fuchsman, Palevsky, et al., 2019) or physical aggregation processes.

Planctomycetes are considered a particle attached group (Eloe et al., 2011; Fuchsman et al., 2011, 2012; Ganesh et al., 2014; Lecleir et al., 2014; Leu et al., 2022; Luo et al., 2020). The abundance of Planctomycetes in particles, could be explained by their ecological role as they respire complex organic matter (Fuchsman et al., 2012). It has been described that Planctomycetes have a role in the degradation of polysaccharides (Reintjes et al., 2023). In the surface waters of the EPR, there were (~10^6) copies/L/μm) Planctomycetes in the free-living fractions. At non-OMZ depths, Planctomycetes remained at (~10^6 copies/L/μm) in particles <53 μm, but for particles ≥53 μm, abundances shifted to ~10^5 and ~10^4 copies/L/μm (Fig. 3.6). The highest abundance of Planctomycetes was actually found in the free-living fractions in the OMZ (~10^8 copies/L/μm). Thus, in quantified abundance Planctomycetes were not more abundant on particles than in the water column, but rather their abundances were fairly similar in free-living and suspended particle fractions but lower in large particle fractions.

We can examine a second group that is generally considered particle attached, Flavobacteria (Eloe et al., 2011; Fuchsman et al., 2011, 2012; Ganesh et al., 2014; Lecleir et al., 2014; Luo et al., 2022). Our data indicate that the Flavobacteriales group (Bacteroidetes) was more prevalent in the particles in surface waters than at depth. We observed that Flavobacteria had a higher number in small particle fraction with 10^8 copies/L/μm in both free-living and < 20 μm particles but ~10^6 copies/L/μm in 20-53 μm and ~10^5 copies/L/μm in ≥500 μm particles. Thus, the pattern was fairly similar between Planctomycetes and Flavobacteria, both of which are generally considered particle attached. These data, in which particle associated organisms are actually most abundant in the free-living phase, cause us to rethink the conventional understanding of the habitats of “particle attached” organisms.

Marine particles crossing the OMZ

When we examined the depth profile, we noticed that in general the microbial abundance decreased with depth, however the OMZ samples showed a different pattern (Fig. 3.5). In the OMZ region, the abundance of Bacteria was >10^8 copies/L/µm, in the free-living fraction, and in fractions >20 µm. The increase in abundance in the OMZ region was also observed in Archaea. In the free living and small size fraction, the numbers for bacteria and Archaea were similar, but in particle sizes >5 µm the Archaea copies/L/µm were 2 orders of magnitude smaller than in Bacteria (Fig. 3.5). When we considered microbial abundance normalized to particle mass, in terms of 16S rRNA gene copies/mg in the small particle size range ≤5µm, we observed a linear decrease from surface 10^{10} copies/mg towards 10^8 copies/mg for bacteria at 1001 m depth In the Bathypelagic) (Fig. 3.4). However, on larger particle sizes ≥20 µm we observed a decrease in abundance from the surface (in the oxic region) towards the oxic mesopelagic water above the OMZ. In the OMZ, (400~700 m depth), where particles were exposed to low oxygen concentration, and nitrite was available (Fig. 3.1), the abundance increased in all fractions ≥20 µm to ~10^{10} copies/mg. In the larger particles the abundance in the OMZ was > 10^{10} copies/mg.

In the OMZ region, the microbial abundance numbers changed depending on particle size and according to taxonomy. One of the groups that only appears in the OMZ were Zetaproteobacteria (Fig. 3.8). Although it is present in low numbers in the OMZ, it was almost absent in the oxic part of the water column. Zetaproteobacteria have been found distributed in deep-sea environments (Emerson, 2014). Additionally, some other groups like Planctomycetes and Nitrospina had unusually high abundances in the OMZ water column. Planctomycetes, abundance in the free-living fraction, was low in the area above the ODZ (Fig. 3.11). However, its abundance increased in the OMZ to >10^8 16S rRNA gene copies/L/µm. This increase in abundance could be explained by the increase in organic matter that Planctomycetes could be respiring in that area (Fig. 3.3). It may be surprising that Planctomycetes increased in the OMZ region despite primers not amplifying anammox bacteria, but actually many Planctomycetes have been identified in anoxic environments (Fuchsman et al 2012). Inside the OMZ region, the nitrite oxidizer Nitrospina had higher abundance in the free-living fraction with ~10^7 copies/L/µm, (Supplementary Fig. 3.2). This increase in Nitrospina abundance can be related to the presence of nitrite concentrations at similar depth ~500 m depth (Fig. 3.1). Although we observed Nitrospina in larger fractions, the abundance was rather low. The presence of Nitrospina in OMZs, were nitrite is found, is similar to what has been observed in the ETNP and ETSP ODZs (Fuchsman et al., 2017; Sun et al., 2019) This could indicate that, although Nitrospina is not highly abundant in particles, it could be contributing to nitrification in the EPR OMZ water column (Beman & Carolan, 2013).

Previous work has also found increases in particle number (Cram et al 2022), in particle flux (Fuchsman et al 2019) and cell counts (Ulloa et al.,

2012) in OMZ regions. Potential reasons hypothesized for these increases include higher rates of chemosynthesis, photosynthesis when OMZs are at shallower depths than seen here and active transport by zooplankton who migrate into OMZs to avoid predation (Cram et al 2022; Fuchsman et al 2019; Ulloa et al 2012).

The Non-Buoyant plume (NBP) had a unique microbial community structure

Although our sampling did not reach the ocean floor, where the hydrothermal vents are located in the EPR, we analyzed a sample at 2369 m depth which was characterized using beam transmission as containing particles from the hydrothermal vent's non-buoyant plume (Tully pers. comm.). This sample had a unique structure, where the abundance of cells did not follow the trends or patterns, we observed through the water column at shallower depths.

When we examined the depth profiles of the Domain bacteria we observed a decrease of microbial abundance towards depth, followed by an increase in microbial abundance in the OMZ region, and in the bathypelagic (1001 m depth) change depending on fraction size. However, the microbial abundance in the NBP contrasted to the trend observed in the surface and OMZ, for example for the free-living fraction the abundance of bacteria in the NBP region was similar to the numbers in the OMZ ~10^8 copies/L/μm (Fig. 3.5), but at the fraction 180 μm for example, the abundance was ~10^8 copies/L/μm in the OMZ, but at the NPB the cell abundance was <10^4 copies/L/μm.

It has been previously described that Alteromonas, Gammaproteobacteria and Epsilonproteobacteria are abundant taxa on sinking particles in the deep sea environment (Boeuf et al., 2019; Sylvan et al., 2012) In our samples, in the Non-Buoyant plume, we found Gammaproteobacteria (Fig. 3.8), but all Epsilonproteobacteria were almost absent in all of our samples. We did observe SAR11 in our free-living non-buoyant plume sample (2369 m depth) in high abundances (~10^8 copies/L/μm) which was similar to abundances in the surface (Fig. 3.9). SAR11 has been reported as free living in deep sea samples (~4000 m depth) (J. Li et al., 2021), and was also much less abundant in plume particle samples (Fig. 3.9). Non-Buoyant plumes evolve as they move away from the hydrothermal vent, with fewer vent related Epsilonproteobacteria and more SAR11, S-oxidizing Thioglobus Gammaproteobacteria and S-oxidizing Sulfurimonas Epsilonproteobacteria farther from the vent (G. Li et al., 2020). Perhaps the plume sampled here was even farther in its evolution from vent plume to seawater.

Methodological advances and limitations

Advances

1) The use of 16S rRNA gene sequence spike-in allowed us to look quantitatively into microbial abundance across different particle sizes at

different ocean depth ranges. This approach allowed us to compare microbial abundance of less abundant taxa directly across particle samples, rather than using relative abundance which is greatly affected by the abundance of the most abundant microbes.

2) We describe the microbial diversity and abundance in marine particles separated according to seven size-fractions from 0.2 to > 500 μm. This is many more particle sizes compared to other studies in ODZs that have only focused into 2 size fractions: >1.6 μm and 0.2-1.6 μm (Ganesh et al., 2014) or 0.2-30 μm and >30 μm (Fuchsman et al., 2017). This allowed us to examine how microbial communities change across the full particle size spectrum, and consider small, intermediate and large particles separately.

Limitations:

1) Many bacteria contain multiple copies of the 16S rRNA gene (1-15) (Venter et al 2004; Espejo and Plaza, 2018): Multiple gene copies per cell which would cause our gene copy counts to be higher than true microbial cellular abundances. Indeed in (Cram et al., 2024), the amplicon-based estimates of microbial abundance were generally higher than microscopy counts from the particles. Therefore, we have been careful herein to report gene copy abundances, rather than cellular abundances. We note that bacterial species are not known to have 16S rRNA gene copy numbers that vary between environments, so we contend that the *differences* in gene copy numbers are directly related to *differences* in cellular abundances between samples.

2) Amplification causes biases: PCR based approaches are known to under-represent some species and over-represent others (Elbrecht & Leese, 2015). For example, the primers used here did not amplify anammox bacteria (McNichol et al 2021), which are important to OMZ systems.

3) 16S rRNA gene corresponds to identity not function: In previous particle metagenomes, there were found several genes involved in denitrification (*nosZ, nirK, qnorB*) enriched in particles (Fuchsman et al., 2017; Ganesh et al., 2014). However, denitrifiers are phylogenetically diverse and it is very difficult to identify them based on 16S rRNA gene sequences (Zhang et al., 2023; Zumft, 1997).

4) Filtering and fractioning marine particles may disrupt particles: Passing through a filter may break large particles into smaller ones or cause small particles to stick together. Due to low organic matter concentrations in the EPR region, to carry out this study ~100 L of water per depth had to be filtered in order to collect enough bacteria and particles to be measured. This means that filter clogging was not an issue in this study, but that the large amount of water passing through each filter could theoretically cause disaggregation.

5) This study did not show clear trends in microbial community structure between size and depth: This non-relationship likely reflects the different depths being quite different environments from each other. In this study we collected replicate samples, from the surface, mesopelagic, oxygen

minimum zone, deep ocean and particle plume over a >2000 m depth range, and these environments are quite different from each other. Thus, we contend that we were statistically underpowered to identify trends, because the different environments contributed so much to community structure variability.

Concluding remarks

In our study, combining 16S rRNA gene standard spike in amplicon sequencing and fractionation of particles at a finer scale allowed us to look into the diversity and abundance of free-living and particle-attached microbes from a new perspective. When we compared the quantitative abundance of different microbial taxa across depths and particle sizes, we observed increases in abundance of many groups in the OMZ region and in the non-buoyant hydrothermal vent plume, both regions that promote chemosynthesis.

We observed that most bacteria were free-living or on the smallest particle fractions likely because the smallest particle size fractions had the most mass, carbon and nitrogen. Quantitative abundance estimates found that ostensibly particle attached bacteria such as Flavobacteria and Planctomycetes were actually equally abundant in the free-living fraction and suspended particles. Such patterns would not have been evident if we had looked at relative abundance data only, where the changes in total numbers of microbes between the water and particles are not taken into account.

Although this approach has its limitations, particularly due to amplification biases, it constitutes a valuable tool that can be used with other techniques to look into the microbial world in unexplored environments, like the EPR which we just described, or other marine environments. As OMZs extend as a consequence of climate change, it is crucial to understand the microbiology of these environments to comprehend how changes in the OMZ environment will impact nutrient cycles globally, particularly the Nitrogen cycle.

Future Directions

The general lack of clear patterns with respect to depth and size in the EPR environment suggests that this data set was “under sampled,” with relatively few samples, spread into many disparate environments. Therefore, we advocate applying this approach either with greater depth resolution or with a similar number of samples collected within environments that have more linear gradients in chemical and physical parameters.

Additional analysis could be applied to this same data. In particular, it would be worth investigating patterns in alpha diversity to see how species richness and evenness vary along the particle size spectrum. Next, while it was shown here that most bacteria are most abundant on small particles, it would be worth systematically investigating whether some bacterial taxa deviate from this pattern. Indeed Cram et al. (2024) showed that some bacterial taxa are known to carry out nitrogen and sulfur cycling processes associated primarily with large particles and so it would also be valuable to

focus on the abundance patterns of taxa known to carry out key biogeochemical processes.

Figures chapter 3

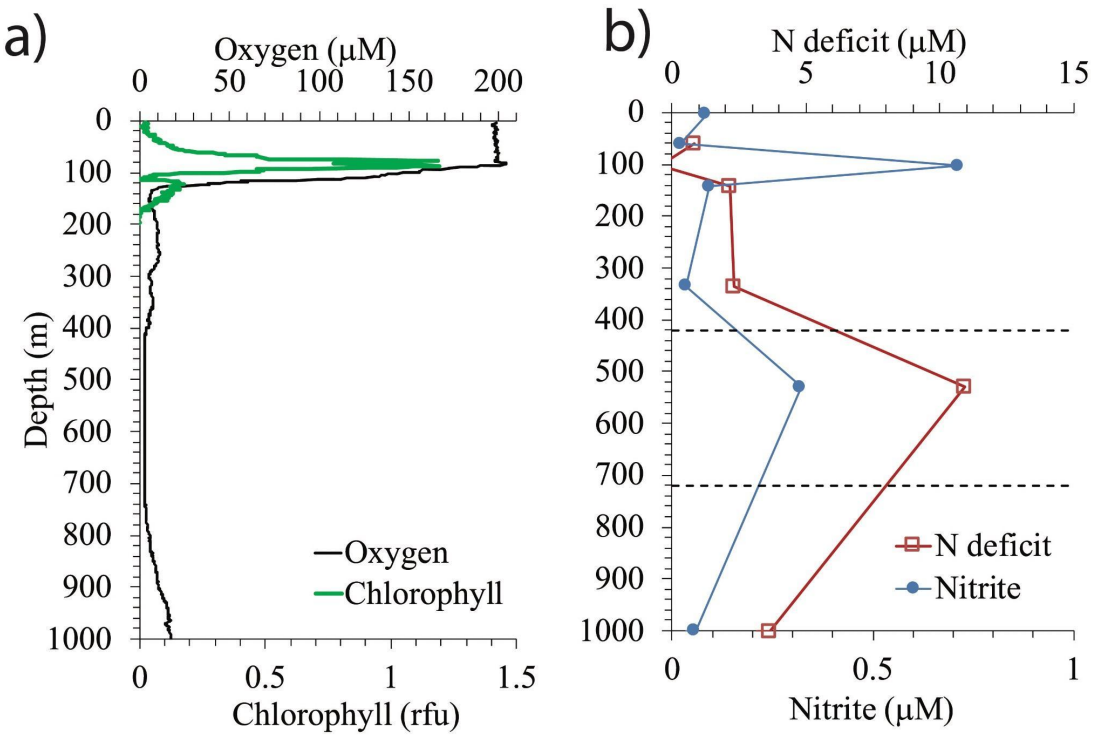


Figure 3.1. Nutrient depth profiles at the EPR station. a) Oxygen concentration (µM), as measured by a Seabird 43 sensor, and Chlorophyll concentration (rfu). b) Nitrite concentration and nitrogen deficit, through the top 1000m of the water column. The area between the dashed lines corresponds to the Oxygen Minimum Zone.

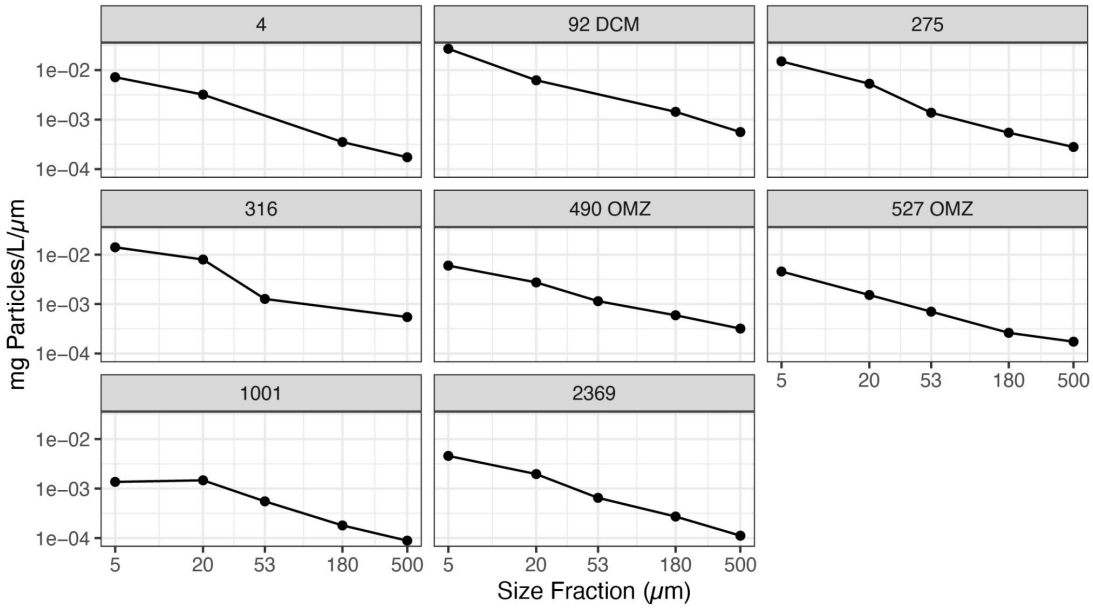


Figure 3.2. Marine particle size to mass. Each panel represents increasing depth (from left to right). Y axis corresponds to volume and filter-size-bin normalized particle mass (mg Particles /L/µm). The X axis corresponds to the particle size fraction (in µm). DCM: Deep Chlorophyll Maxima, OMZ: Oxygen Minimum Zone.

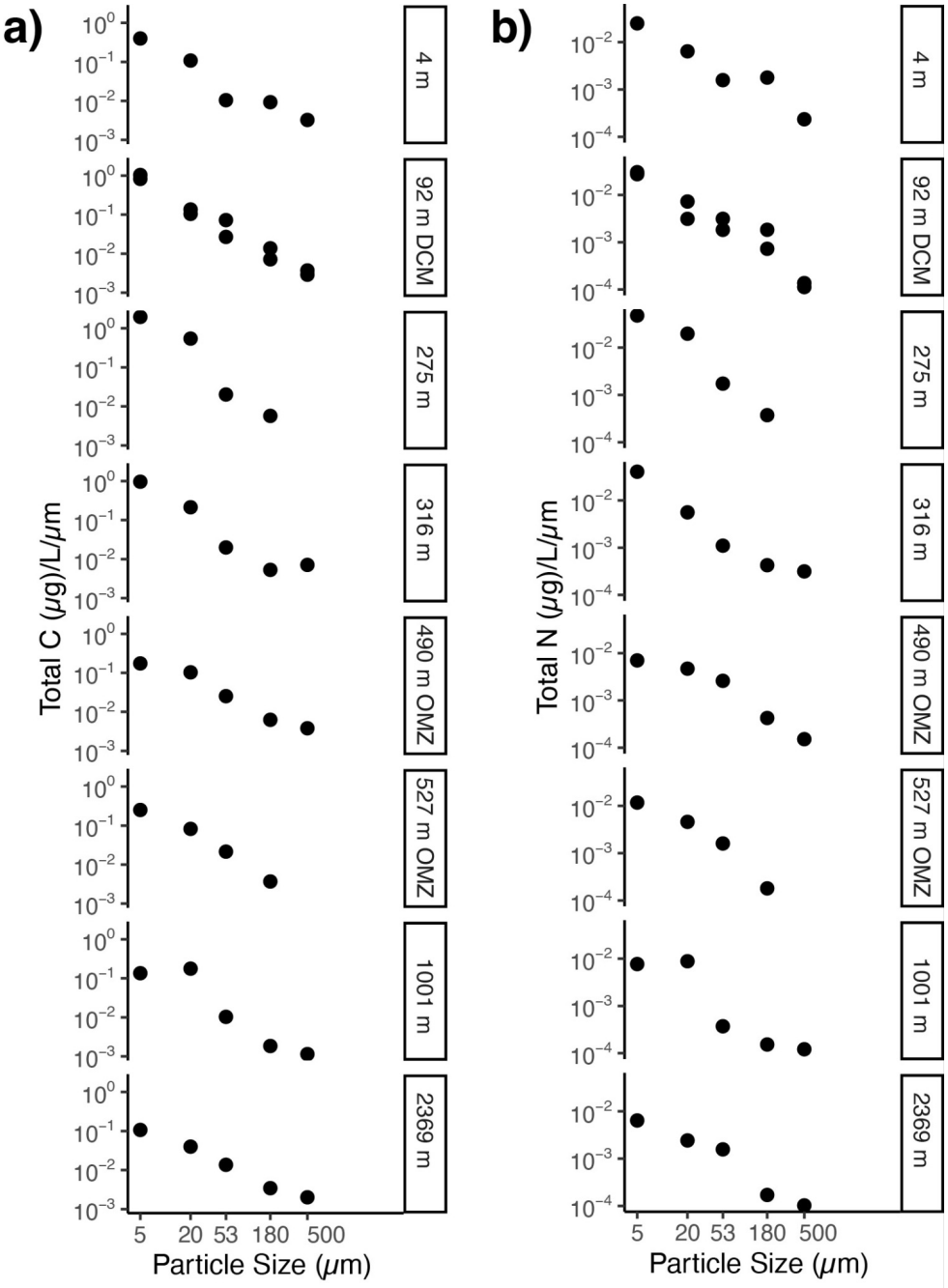


Figure 3.3. Carbon and Nitrogen composition of marine particles in the EPR. Each panel corresponds to different ocean regions (Classified according to depth in m). x-Axis represents particle size range, y-axis indicates size and bin size normalized (a) Particulate organic carbon (μg)/L/μm, b) Particulate organic Nitrogen (μg)/L/μm. The depths containing 2 points per particle size correspond to replicates. DCM: Deep Chlorophyll Maxima, OMZ: Oxygen Minimum Zone.

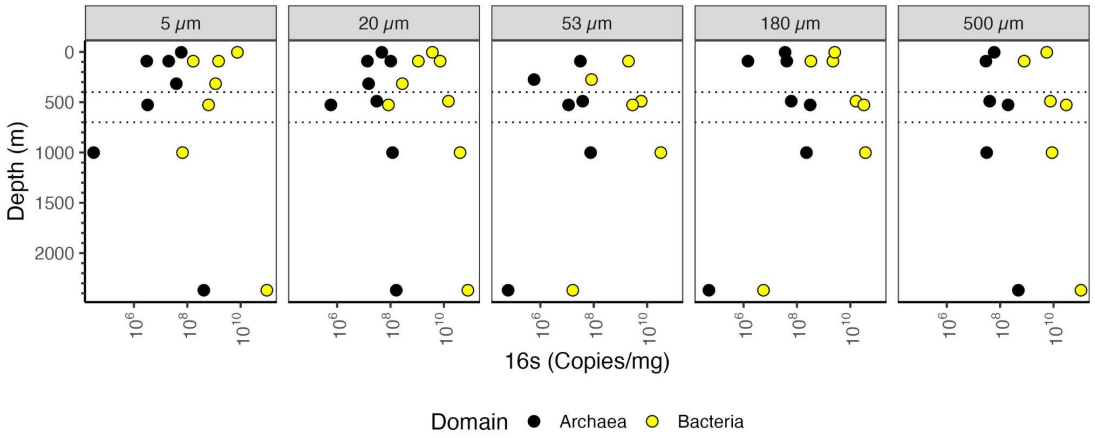


Figure 3.4. Depth profile of 16S rRNA gene copy numbers per mg particle mass, aggregated to domain level. As in figure 3, this time normalized to particle mass. Each panel represents particle size (in μm). The Y axis corresponds to depth (m), and the X axis corresponds to 16S rRNA gene copies/mg particle mass. The area between the dashed lines corresponds to the Oxygen Minimum Zone.

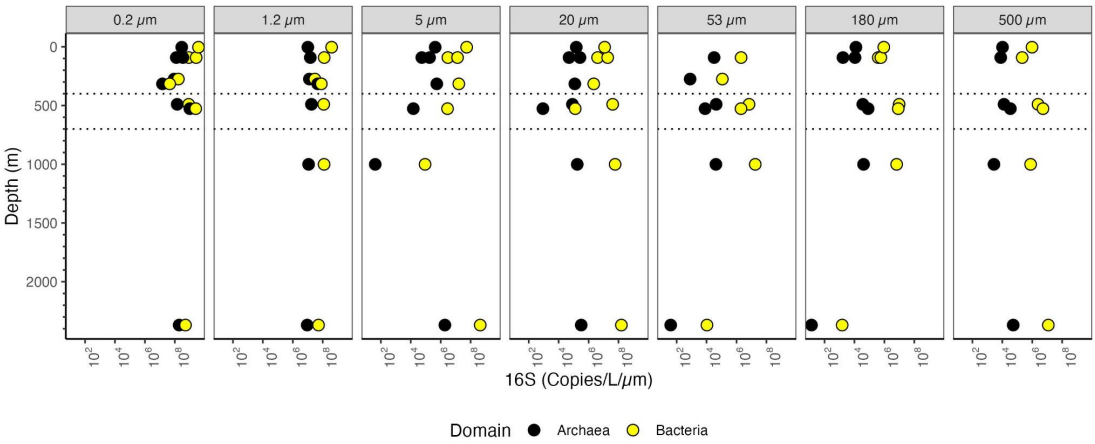


Figure 3.5. Depth profile of volume and bin size normalized 16S rRNA gene copies aggregated to domain level in the EPR. Depth profile of 16S amplicon sequences, aggregated to domain level along a depth profile in the EPR. Each panel represents different particle sizes (in μm). The Y axis represents depth in meters. The x-axis indicates the bin-size normalized abundance of the sum of all ASVs in each domain. The area between the dashed lines corresponds to the Oxygen Minimum Zone.

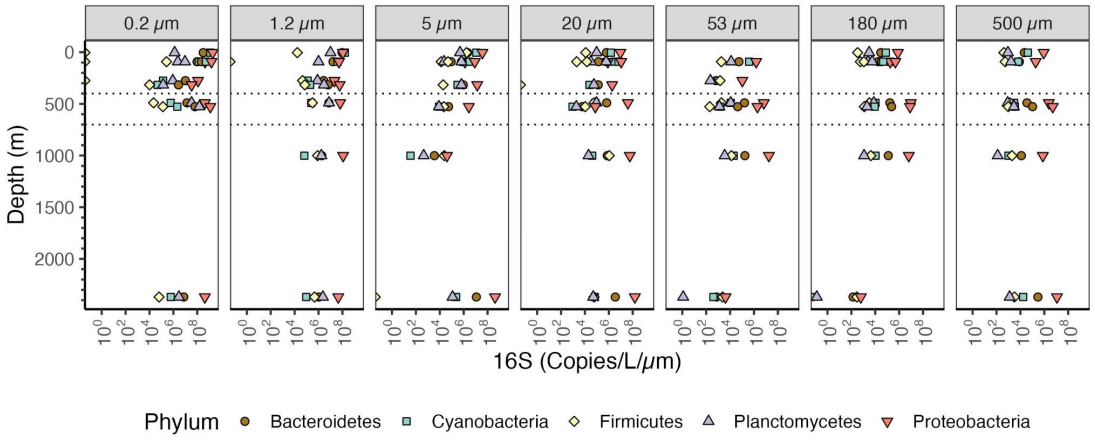


Figure 3.6. Depth profile of volume and bin size normalized 16S rRNA gene copy abundance, aggregated to Phylum level in the EPR. The five most abundant phyla are shown. The area between the dashed lines corresponds to the Oxygen Minimum Zone.

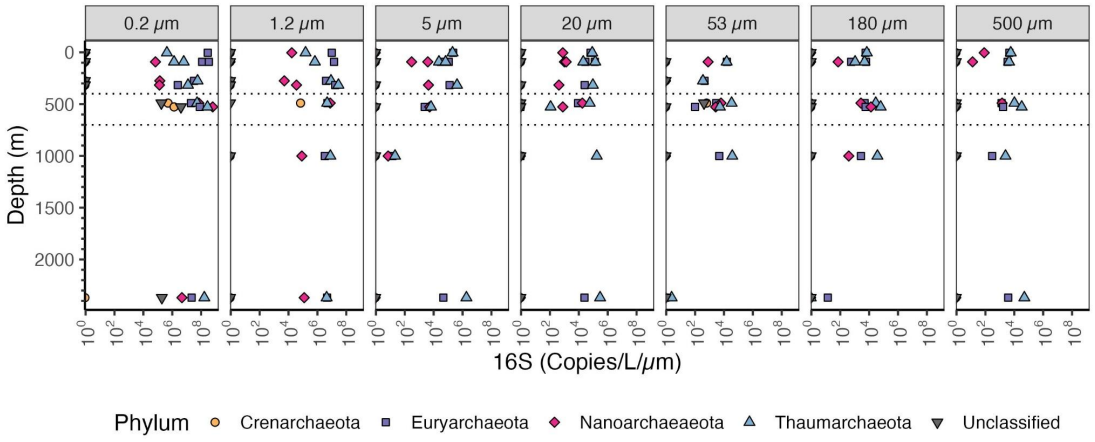


Figure 3.7. Depth profile of volume and bin size normalized 16S rRNA gene copy abundance of Archaea, aggregated to Phylum level in the EPR. The five most abundant phyla are shown. Each panel represents different particle sizes (in μm). The Y axis represents depth in meters, and the x axis represents 16S rRNA gene copies per L per μm. The area between the dash lines corresponds to the Oxygen Minimum Zone.

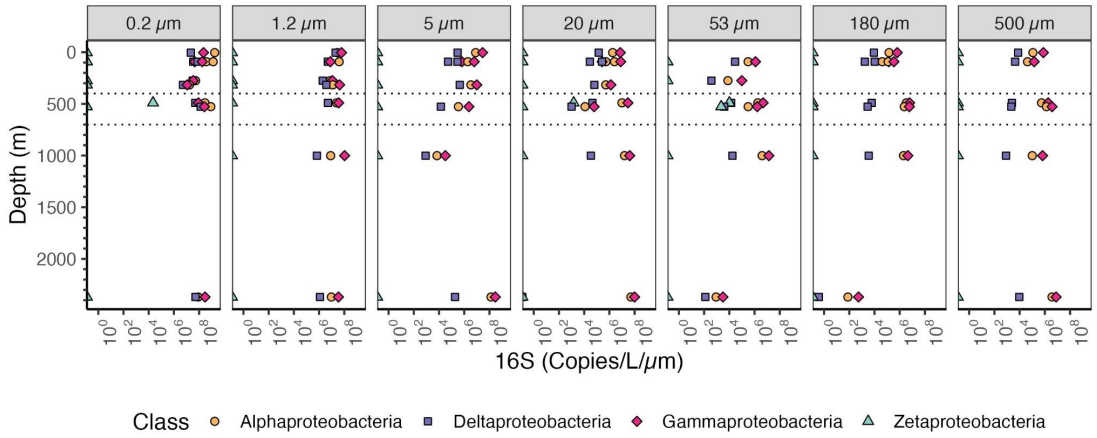


Figure 3.8. Depth profile of volume and bin size normalized 16S rRNA gene copy abundance of Proteobacteria, aggregated to Class level in the EPR. The five most abundant Classes are shown. Each panel represents different particle sizes (in μm). The Y axis represents depth in meters, and the x axis represents 16S rRNA gene copies per L per μm. The area between the dashed lines corresponds to the Oxygen Minimum Zone.

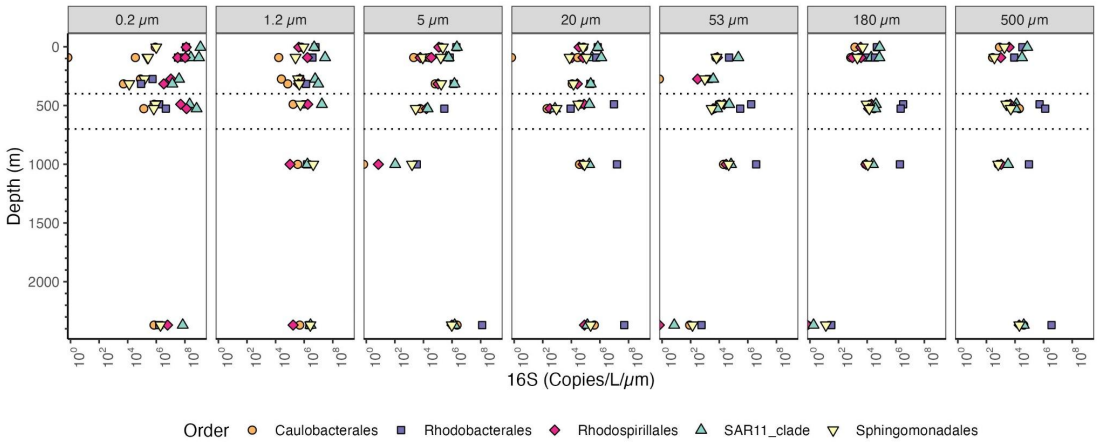


Figure 3.9. Depth profile of volume and bin size normalized 16S rRNA gene copy abundance of Alphaproteobacteria, aggregated to Order level in the EPR. The five most abundant orders are shown. Each panel represents different particle sizes (in μm) The Y axis represents depth in meters, and the x axis represents 16S rRNA gene copies per L per μm. The area between the dashed lines corresponds to the Oxygen Minimum Zone.

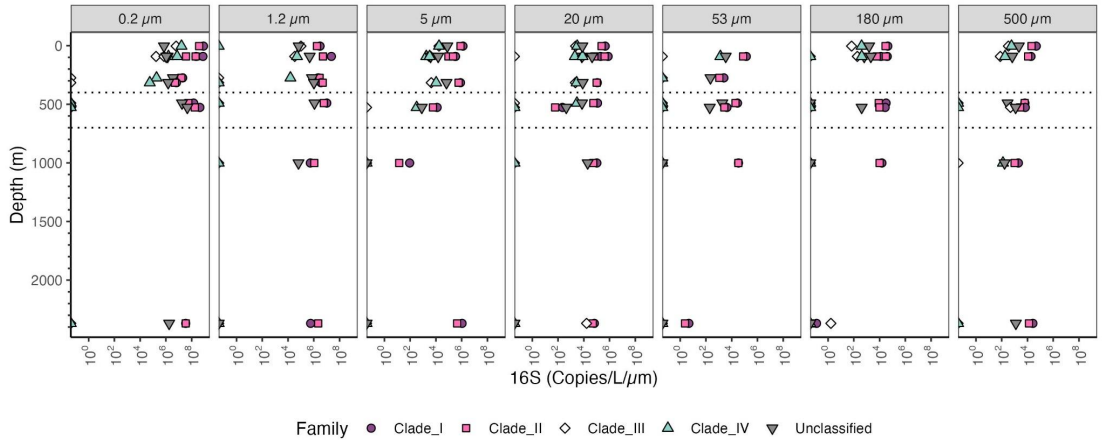


Figure 3.10. Depth profile of volume and bin size normalized 16S rRNA gene copy abundance of SAR11, aggregated to Family level. Each panel represents different particle sizes (in μm). The Y axis represents depth in meters, and the x axis represents 16S rRNA gene copies per L per μm. The area between the dashed lines corresponds to the Oxygen Minimum Zone.

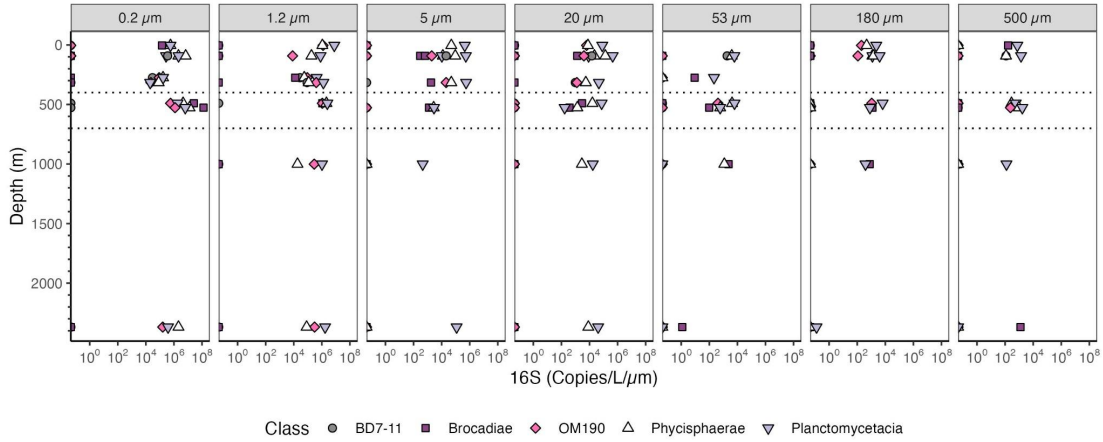


Figure 3.11. Depth profile of volume and bin size normalized 16S rRNA gene copy abundance of Planctomycetes, aggregated to Class level. The five most abundant phyla are shown. Each panel represents different particle sizes (in μm). The Y axis represents depth in meters, and the x axis represents 16S rRNA gene copies per L per μm. The area between the dashed lines corresponds to the Oxygen Minimum Zone.

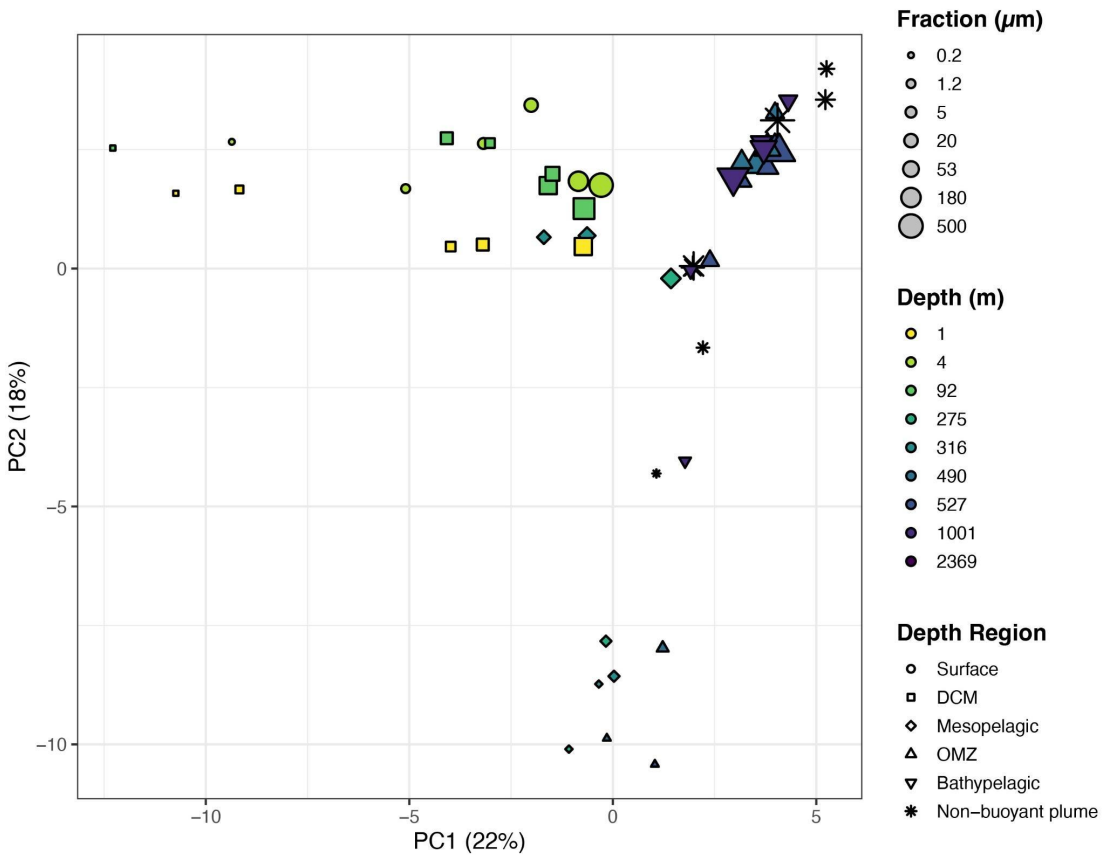
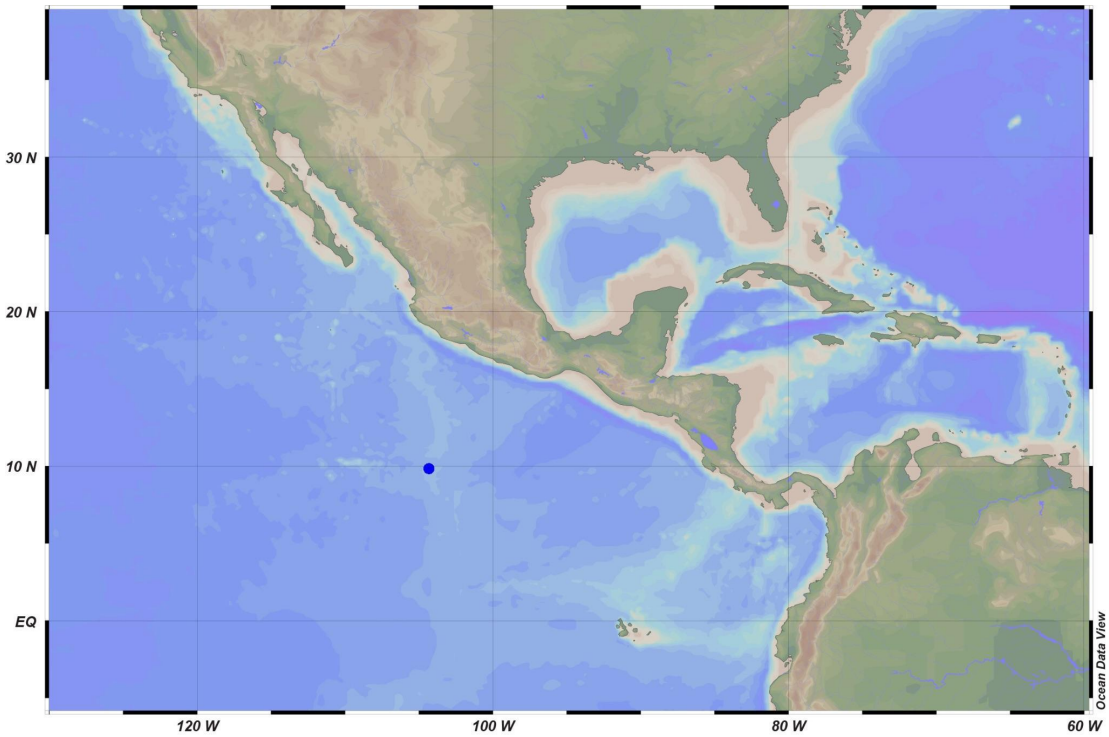
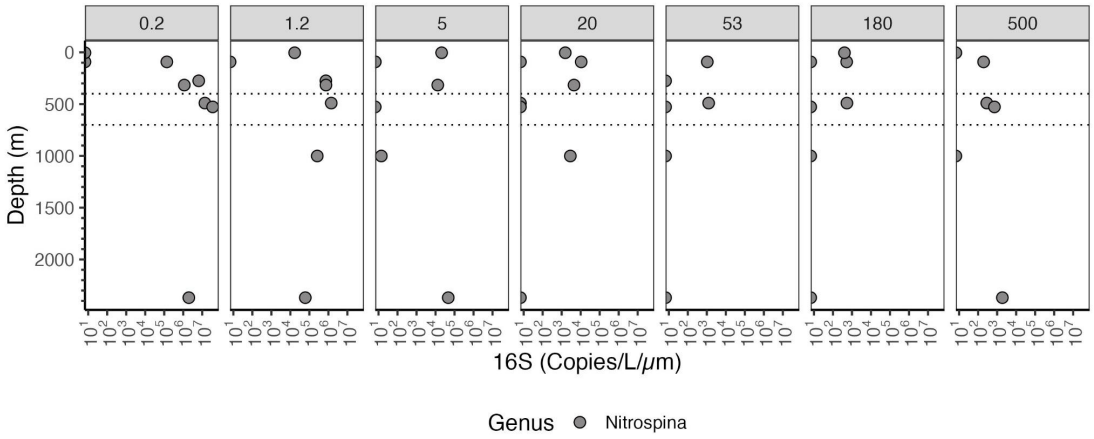


Figure 3.12. Principal component analysis plot (PCA) of the 16S rRNA gene amplicon sequencing fractionated according to size variation and depth. Points that are closer together in space have more similar community structure. The axes indicate principal component scores of each sample. Point size corresponds to the particle size, fraction, color to the depth in meters, and shape to the region of the water column from which the sample was collected. OMZ (Oxygen Minimum Zone, DCM (Deep Chlorophyll Maximum).

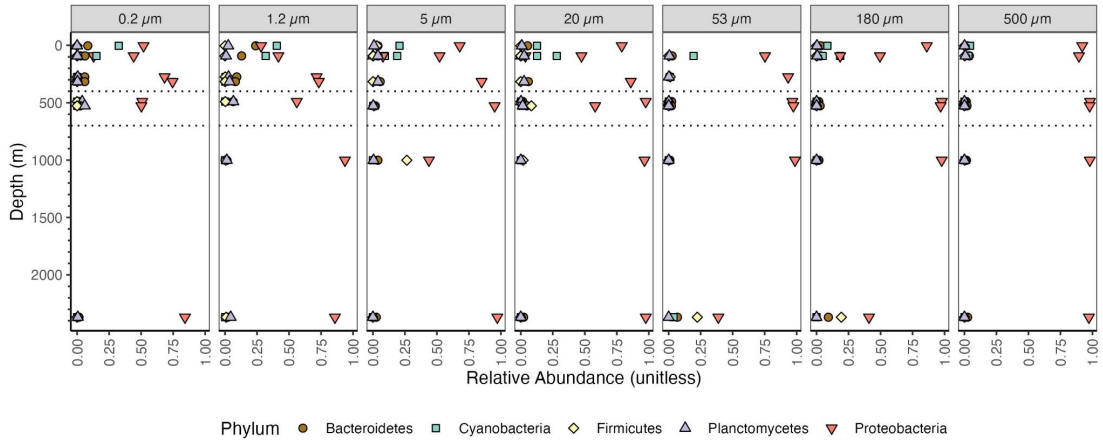
Supplementary Figures chapter 3



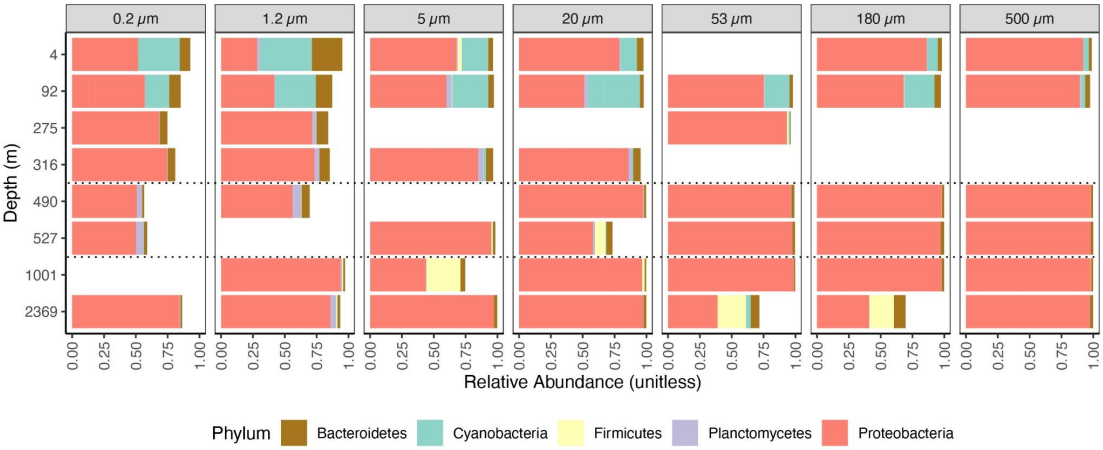
Supplementary Figure 3.1 Location of the EPR station. Figure produced using Ocean Data view (Schlitzer, 2023).



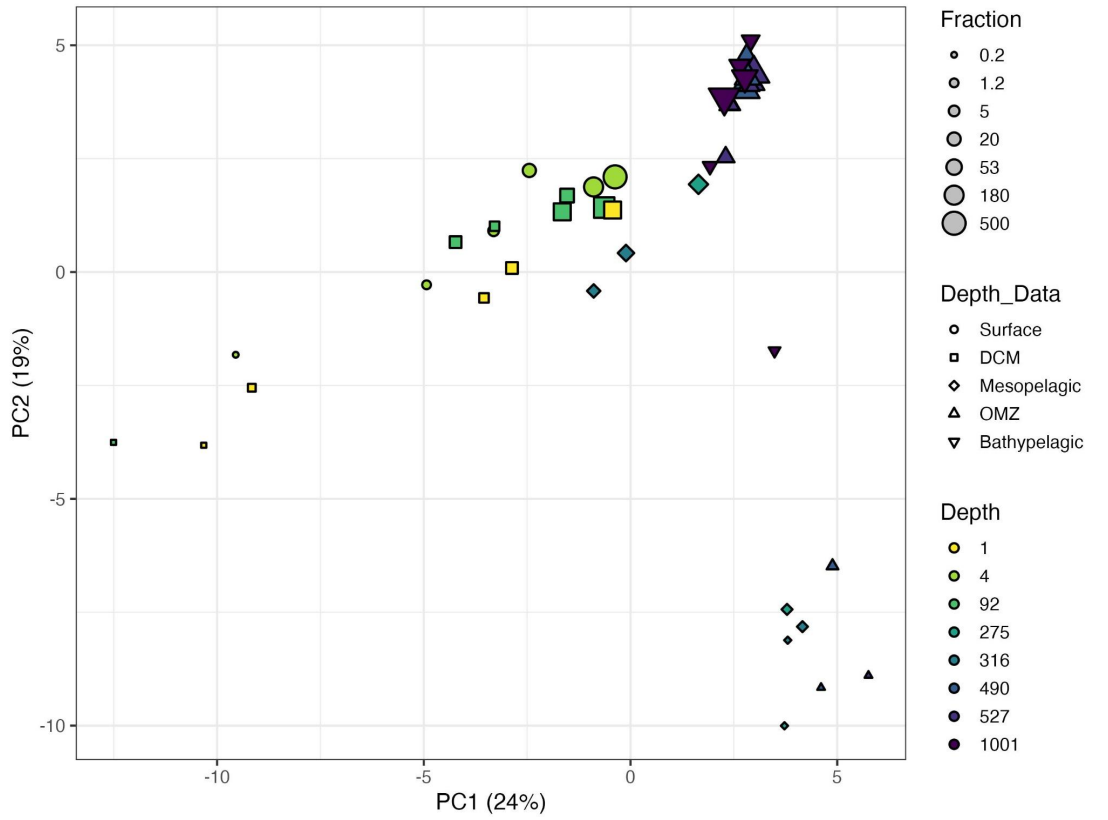
Supplementary Figure 3.2 Abundance of 16S gene copy numbers of Nitrospina. Each panel represents different particle sizes (in µm). The Y axis represents depth in meters, and the x axis represents 16S rRNA gene copies per L per µm. The area between the dash lines corresponds to the Oxygen Minimum Zone.



Supplementary Figure 3.3 Relative abundance of Bacterial 16S rRNA gene copy numbers aggregated to the Phylum level. The five most abundant phyla are shown. The area between the dash lines corresponds to the Oxygen Minimum Zone.

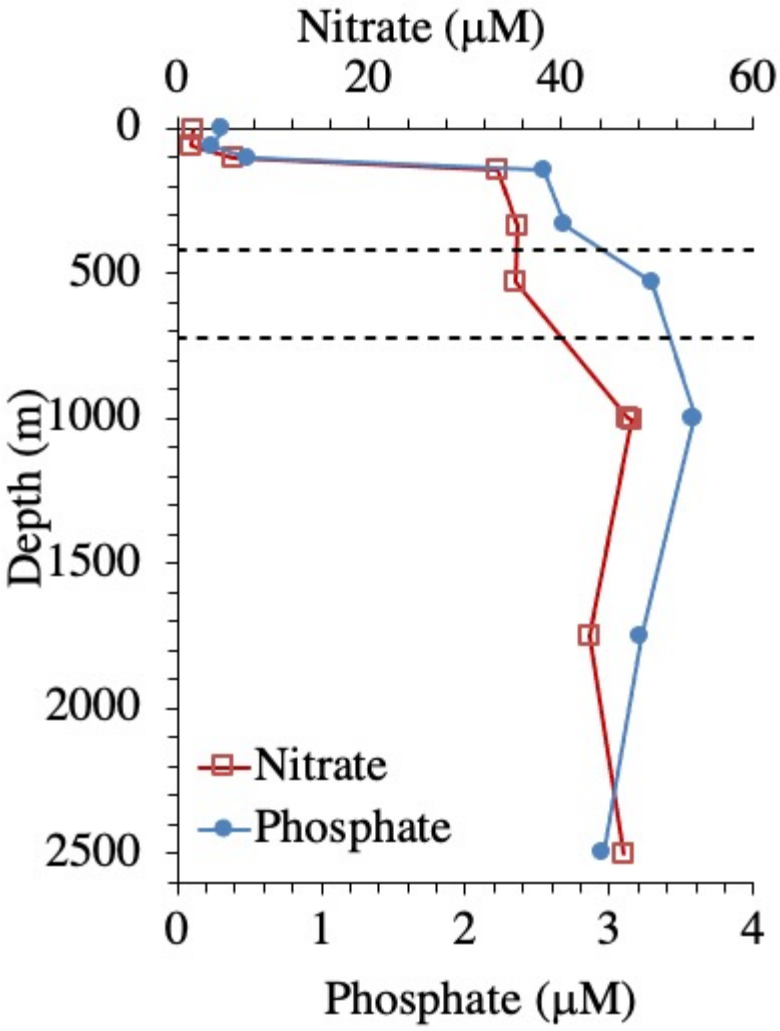


Supplementary Figure 3.4. Stacked bar chart showing relative abundance of 16S rRNA genes aggregated to the Phylum level. The five most abundant phyla are shown. The area between the dash lines corresponds to the Oxygen Minimum Zone. The depth is not to scale.

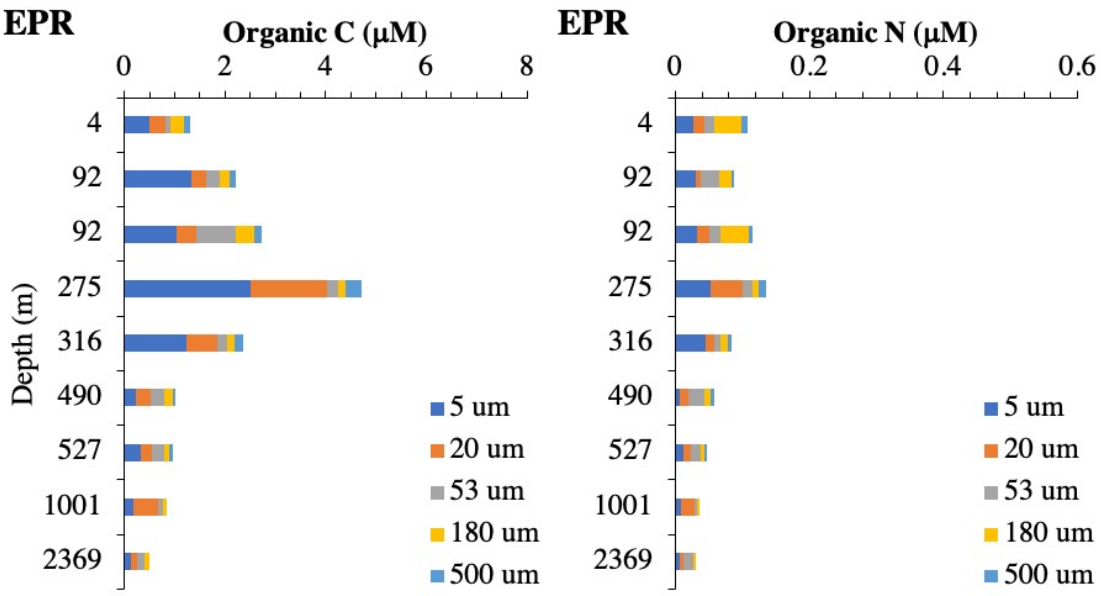


Supplementary Figure 3.5. Principal component analysis plot (PCA) of the 16S amplicon sequencing fractionated according to size variation and depth. For this analysis the Non-buoyant plume was removed Points that are closer together in space have more similar community structure. The axes indicate principal component scores of each sample. Point size corresponds to the particle size, fraction, color to the depth in meters, and shape to the region of the water column from which the sample was collected. OMZ (Oxygen Minimum Zone, DCM (Deep Chlorophyll Maxima).

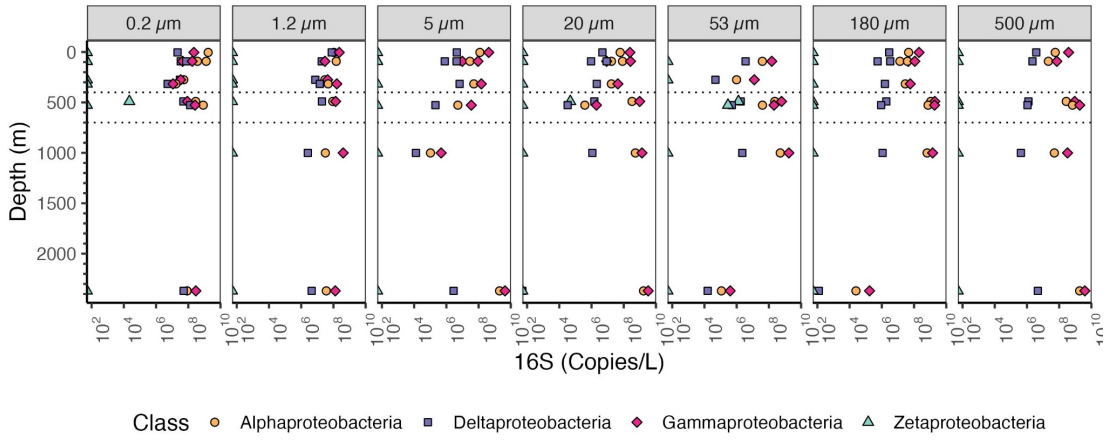
Supplementary Figure 3.6. Scale bar showing spacing of the filter sizes used for normalization to bin width.



Supplementary Figure 3.7. Nitrate and phosphate depth profile at the EPR station. The area between the dash lines corresponds to the Oxygen Minimum Zone. The graph only shows down to 1000m.



Supplementary Figure 3.8. Organic C and N concentrations separated by bin size over the EPR depth profile. These values are not normalized by bin width.



Supplementary Figure 3.9. Depth profile of volume normalized 16S rRNA gene copy abundance of Proteobacteria, aggregated to Class level in the EPR. The five most abundant Classes are shown. Each panel represents different particle sizes (in µm). The Y axis represents depth in meters, and the x axis represents 16S rRNA gene copies per L per µm. The area between the dashed lines corresponds to the Oxygen Minimum Zone.

Conclusion

In the three chapters of this dissertation, we used quantitative approaches to studying microbial ecology in the ocean. In Chapters 1 and 2 we examined ratios of functional genes to single copy core genes to quantify how common given functions were among water column bacteria and archaea. Specifically, we looked at the proportion of microbes with key functional genes involved in using alternative reduced nitrogen sources. We used two functional marker genes, the gene cyanase (*cynS)* and urease (*ureC)*, and we examined the presence and abundance of those marker genes in a ratio with single copy core gene RNA polymerase (*rpoB*) to identify the proportion of each taxa contained urease (*ureC*) and cyanase (*cynS*) genes and which reduced nitrogen form was preferred. In chapter three, we identified the total abundance of microorganisms that were in different particle microenvironments, normalizing to volumes of water and concentrations of particles.

In Chapter 1, we examined how different microbial taxa preferred to use urea or cyanate, and we compared how that preference changed across the ETNP ODZs located in the North and two ETSP stations in the South Pacific. We compared them with an oxic ocean region represented by the oxic-time points from the Hawaii Time series (HOT). Microbial access to alternative reduced nitrogen forms such as urea and cyanate, might provide them with a selective advantage, in these offshore oceanic regions, contributing to the observed patterns of taxonomic distributions. We identified two new taxa containing urease in ODZ regions: Verrucomicrobia and a specific ODZ Clade of Thioglobaceae. The differences in cyanase and urease depth profiles between microbes imply niche differentiation. Different microbes have higher proportions of genes for urease in different parts of the water column: SAR11 and Picocyanobacteria in surface waters, Verrucomicrobia in hypoxic waters, Thioglobus in ODZ waters, and Thaumarchaeota in the lower euphotic zone/mesopelagic. At our oxic station (HOT), *Prochlorococcus* could utilize cyanate in surface waters and Nitrospina could utilize cyanate at depth, while Anammox bacteria (*Cand*. Scalindua) utilized cyanate in the ODZs (Table 1.1). This niche differentiation could be attributed to microbial competition over nitrogen sources, reflecting microbial adaptation to certain parts of the water column. It also, at least in the case of SAR11, is affected by the availability of nitrate, an oxidized inorganic form of N. All of these microbes are autotrophs, except SAR11, which is a heterotroph utilizing simple DOC.

In Chapter 2, we examined the microbial use of alternative reduced nitrogen forms, urea, and cyanate across the oxic ocean. We focused on two oceanic transects. The first transect took place in spring and corresponded to a transect oriented North-South, and we compared that transect with another transect in Fall; however, the orientation was West-East. Our results indicate seasonality in the preference for urea or cyanate and in stratification and nutrient concentrations. In the fall transect there was an influence of the

subtropical gyre related to changes in nutrients and temperature. Here there were distinct changes in composition of organisms able to use urea and cyanate with depth and with latitude and longitude and depth. Particularly *Synechococcus*, and *Prochlorococcus*, switched in dominance across stations during both transects, causing the phylotypes for *ureC* and *cynS* to change at the same time. This switch was likely related to stratification and resulting reduced nutrient concentrations. More interestingly, the abundance of *cynS* in Picocyanobacteria and total microbial community decreased drastically between spring GA02 and nutrient limited fall GA03 cruises, but the abundance of *ureC* in the total community increased in GA03. Thus, the dynamics of urea and cyanate utilization seem more complicated than we would have guessed. Though there were changes in surface communities using *cynS* and *ureC*, the deeper communities were more consistent between transects. *Nitrospina* was present in both transects, but it was absent towards the southern stations in the GA02 transect, but when present, consistently favored *ureC* over *cynS*. Thaumarchaeota and the *ureC* were evenly distributed across both transects.

In summary of Chapter 1 and 2, the use of alternative reduced nitrogen forms, like cyanate and urea, could provide a selective advantage in the analyzed offshore oceanic regions. In addition, the use of urea and cyanate is observed in patterns that reflect niche differentiation, where different microbial taxa, living in different regions of the water column, have preferences for urea or cyanate (for those microbial taxa harboring both genres), or can use urea (for those that have only *ureC* gene). The results presented here contribute to expanding our understanding of the use of alternative reduced nitrogen forms in the ocean. Understanding the fate of nitrogen formation can constitute an important aspect of the nitrogen cycle in oceans that are changing in response to climate change. Climate change is affecting the ocean globally, influencing the increase of ODZs. As ODZs expand, the niche for N_2 producing bacteria such as anammox will increase. Use of urea and cyanate as an energy source by Anammox bacteria expand their options beyond those typically considered by scientists and their models.

Finally, in **Chapter 3**, we looked at the quantitative abundance of microbial taxa in marine particles at the EPR. We used a combined approach that included using 16S spike alongside rRNA amplicon sequencing and particle fractionation. This approach allowed us to look at microbial abundance quantitatively. We examined the abundance of different microbial groups in free-living and particle-attached fractions. We quantified microbial abundance in different particle-attach fractions; this allows us to look at microbial distribution at a finer scale than traditional methods. In this work, we observed increases in abundance of many groups in the OMZ region and in the non-buoyant hydrothermal vent plume, both regions that promote chemosynthesis. Additionally, we observed that most bacteria were free-living or on the smallest particle fractions likely because the smallest particle size fractions had the most mass, carbon and nitrogen. Quantitative abundance estimates found that ostensibly particle attached bacteria such as

Flavobacteria and Planctomycetes were actually equally abundant in the free-living fraction and suspended particles. Such patterns would not have been evident if we had looked at relative abundance data only, where the changes in total numbers of microbes between the water and particles are not taken into account.

Why is quantification important to microbial ecology?

Using tools for the quantification of microbial abundances can provide us with a more accurate view of microbial distribution in the environment. A reference on how microorganisms distribute in the environment can help answer questions in marine microbial ecology. The quantitative approaches applied here advance common sequencing-based methods used in microbial ecology. In less quantitative metagenomics approaches, a gene is identified as present, but it is unknown how common a gene is in each taxa. Here, and in previous papers from the Fuchsman lab (Widner et al 2018; Fuchsman et al 2021, 2023), we quantified the proportion of each taxa with specific functional genes, which better indicates the importance of the functional gene to the taxa in question. In relative abundance focused 16S rRNA community approaches, the relative proportions of each taxon are known, but not their total abundance. Indeed, relative abundance community data have been shown to provide misleading patterns, when conventional statistical approaches are applied (Gloor et al., 2016). Chapter 3 used a relatively new quantitative approach to quantify microbial abundances across the particle size gradient, quantifying the particles as well as using an internal 16S spike to quantify the bacteria and archaea (as in Cram et al 2024) and applied it for the first time to particles in the oligotrophic ocean.

Quantification of microbial ecology is necessary to properly compare microbial information to chemical parameters and fluxes and to incorporate microbial ecology data into models. Thus, quantification is essential to allow microbial ecology to interact with other branches of oceanography.

Future directions:

In order to have a more comprehensive understanding of how the use of cyanate and urea changes across different ocean regions, future work could focus on examining the abundance and distribution of the *cynS* and *ureC* genes in more areas of the ocean; in Chapter 1, we looked at ODZs and the Hawaii Time series, whereas in Chapter 2, we looked at stations in the GA03, and the GA02 transect both located in the North Atlantic. Although we only looked at 11 stations in the GA02 transect, which were located in the North Atlantic, the entire GA02 transect covers an extensive region from the North to the Southern Hemisphere with a total of 32 stations. An expansion of this work should include analyzing the rest of the station of the GA02 transect and incorporating other GEOTRACES transects, even expanding to the less detailed sampled stations (3 depths) examined as part of the TARA expedition. An analysis of *cynS* and *ureC* genes at a global scale

(GEOTRACES + TARA) could give us a necessary reference frame for understanding the global use of urea and cyanate.

In addition to the analysis of metagenomic data, it could be a good practice to incorporate the measurement of cyanate and urea as part of the standard nutrient measurements in ocean surveys. Incorporating the routine measurement of cyanate and urea, and their fluxes, would improve our understanding of how nitrogen is recycled in the water column, which might impact the calculations of the nitrogen budget.

A clear opportunity for future progress falls in unifying the metagenomic and particle focused approaches provided here; specifically in using standards with metagenomic data to explore microbial functional gene content along the particle size spectrum, especially in anoxic environments. Models suggest that in the hypoxic ocean, the niche of many microbial processes, including denitrification and sulfate and metal reduction are constrained to ranges of particles larger than a threshold (Bianchi et al 2018). Metagenomic analysis has identified genes for key elemental processes associated with particles under anoxic conditions (Fuchsman et al 2017; Saunders et al 2019) but these studies have not identified the sizes of particles in which these processes take place. Such distinctions are important because particle size can impact which microbes interact with each other and their residence time in the water column. Therefore, we advocate quantitative sequencing the metagenomes and/or metatranscriptomes of a size spectrum of particles across oxic and anoxic environments and identifying the abundance of key functional genes in microbial populations.

Bibliography

Abramson, L., Lee, C., Liu, Z., Wakeham, S. G., & Szlosek, J. (2010). Exchange between suspended and sinking particles in the northwest Mediterranean as inferred from the organic composition of in situ pump and sediment trap samples. *Limnology and Oceanography*, *55*(2), 725–739.

Ahlgren, N. A., Chen, Y., Needham, D. M., Parada, A. E., Sachdeva, R., Trinh, V., Chen, T., & Fuhrman, J. A. (2017). Genome and epigenome of a novel marine Thaumarchaeota strain suggest viral infection, phosphorothioation DNA modification and multiple restriction systems. *Environmental Microbiology*, *19*(6), 2434–2452. https://doi.org/10.1111/1462-2920.13768

Ahlgren, N. A., Rocap, G., & Chisholm, S. W. (2006). Measurement of *Prochlorococcus* ecotypes using real-time polymerase chain reaction reveals different abundances of genotypes with similar light physiologies. *Environmental Microbiology*, *8*(3), 441–454. https://doi.org/10.1111/j.1462-2920.2005.00910.x

Aldunate, M., Henríquez-Castillo, C., Ji, Q., Lueders-Dumont, J., Mulholland, M. R., Ward, B. B., von Dassow, P., & Ulloa, O. (2020). Nitrogen assimilation in picocyanobacteria inhabiting the oxygen-deficient waters of the eastern tropical North and South Pacific. *Limnology and Oceanography*, *65*(2), 437–453. https://doi.org/10.1002/lno.11315

Alldredge, A. (1998). The carbon, nitrogen and mass content of marine snow as a function of aggregate size. *Deep-Sea Research Part I: Oceanographic Research Papers*, *45*(4–5), 529–541. https://doi.org/10.1016/S0967-0637(97)00048-4

Alonso-Sáez, L., Waller, A. S., Mende, D. R., Bakker, K., Farnelid, H., Yager, P. L., Lovejoy, C., Tremblay, J.-É., Potvin, M., Heinrich, F., Estrada, M., Riemann, L., Bork, P., Pedrós-Alió, C., & Bertilsson, S. (2012). Role for urea in nitrification by polar marine Archaea. *Proceedings of the National Academy of Sciences*, *109*(44), 17989–17994. https://doi.org/10.1073/pnas.1201914109

Altabet, M. A. (1988). Variations in nitrogen isotopic composition between sinking and suspended particles: Implications for nitrogen cycling and particle transformation in the open ocean. *Deep Sea Research Part A. Oceanographic Research Papers*, *35*(4), 535–554. https://doi.org/10.1016/0198-0149(88)90130-6

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402. https://doi.org/10.1093/nar/25.17.3389

Anderson, P. M., Sung, Y.-C., & Fuchs, J. A. (1990). The cyanase operon and cyanate metabolism. *FEMS Microbiology Reviews*, *7*(3–4), 247--252.

Anderson, R., Mawji, E., Cutter, G., Measures, C., & Jeandel, C. (2014). GEOTRACES: Changing the Way We Explore Ocean Chemistry.

Oceanography, *27*(1), 50–61. https://doi.org/10.5670/oceanog.2014.07
Azam, F., & Malfatti, F. (2007). Microbial structuring of marine ecosystems. *Nature Reviews Microbiology*, *5*(10), 782–791. https://doi.org/10.1038/nrmicro1747
Babbin, A. R. (2014). *Biogeochemical Controls on Fixed Nitrogen Loss Processes in the Marine Environment*.
Babbin, A. R., Peters, B. D., Mordy, C. W., Widner, B., Casciotti, K. L., & Ward, B. B. (2017). Multiple metabolisms constrain the anaerobic nitrite budget in the Eastern Tropical South Pacific. *Global Biogeochemical Cycles*, *31*(2), 258–271. https://doi.org/10.1002/2016GB005407
Banse, K., Naqvi, S. W. A., & Postel, J. R. (2017). A zona incognita surrounds the secondary nitrite maximum in open-ocean oxygen minimum zones. *Deep Sea Research Part I: Oceanographic Research Papers*, *127*, 111–113. https://doi.org/10.1016/j.dsr.2017.07.004
Barbera, P., Kozlov, A. M., Czech, L., Morel, B., Darriba, D., Flouri, T., & Stamatakis, A. (2019). EPA-ng: Massively Parallel Evolutionary Placement of Genetic Sequences. *Systematic Biology*, *68*(2), 365–369. https://doi.org/10.1093/sysbio/syy054
Beman, J. M., Leilei Shih, J., & Popp, B. N. (2013). Nitrite oxidation in the upper water column and oxygen minimum zone of the eastern tropical North Pacific Ocean. *The ISME Journal*, *7*(11), 2192–2205. https://doi.org/10.1038/ismej.2013.96
Berger, S. A., & Stamatakis, A. (2012). PaPaRa 2.0: A vectorized algorithm for probabilistic phylogeny-aware alignment extension. *Heidelberg Institute for Theoretical Studies*, *12*.
Berman, T., Béchemin, C., & Maestrini, S. Y. (1999). Release of ammonium and urea from dissolved organic nitrogen in aquatic ecosystems. *Aquatic Microbial Ecology*, *16*(3), 295–302.
Berube, P. M., Biller, S. J., Hackl, T., Hogle, S. L., Satinsky, B. M., Becker, J. W., Braakman, R., Collins, S. B., Kelly, L., & Berta-Thompson, J. (2018). Single cell genomes of *Prochlorococcus*, *Synechococcus*, and sympatric microbes from diverse marine environments. *Scientific Data*, *5*(1), 1–11.
Bianchi, D., Weber, T. S., Kiko, R., & Deutsch, C. (2018). Global niche of marine anaerobic metabolisms expanded by particle microenvironments. *Nature Geoscience*, *11*(4), 263–268. https://doi.org/10.1038/s41561-018-0081-0
Biller, S. J., Berube, P. M., Dooley, K., Williams, M., Satinsky, B. M., Hackl, T., Hogle, S. L., Coe, A., Bergauer, K., Bouman, H. A., Browning, T. J., De Corte, D., Hassler, C., Hulston, D., Jacquot, J. E., Maas, E. W., Reinthaler, T., Sintes, E., Yokokawa, T., & Chisholm, S. W. (2018). Marine microbial metagenomes sampled across space and time. *Scientific Data*, *5*(1), 180176. https://doi.org/10.1038/sdata.2018.176
Bižić-Ionescu, M., Ionescu, D., & Grossart, H.-P. (2018). Organic Particles: Heterogeneous Hubs for Microbial Interactions in Aquatic Ecosystems.

Frontiers in Microbiology, *9*. https://www.frontiersin.org/articles/10.3389/fmicb.2018.02569

Boeuf, D., Edwards, B. R., Eppley, J. M., Hu, S. K., Poff, K. E., Romano, A. E., Caron, D. A., Karl, D. M., & DeLong, E. F. (2019). Biological composition and microbial dynamics of sinking particulate organic matter at abyssal depths in the oligotrophic open ocean. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(24), 11824–11832. https://doi.org/10.1073/pnas.1903080116

Bolaños, L. M., Karp-Boss, L., Choi, C. J., Worden, A. Z., Graff, J. R., Haëntjens, N., Chase, A. P., Della Penna, A., Gaube, P., & Morison, F. (2020). Small phytoplankton dominate western North Atlantic biomass. *The ISME Journal*, *14*(7), 1663–1674.

Bolaños, L. M., Tait, K., Somerfield, P. J., Parsons, R. J., Giovannoni, S. J., Smyth, T., & Temperton, B. (2022). Influence of short and long term processes on SAR11 communities in open ocean and coastal systems. *ISME Communications*, *2*(1), 116. Publicly Available Content Database. https://doi.org/10.1038/s43705-022-00198-1

Boyer, T. P., Baranova, O. K., Coleman, C., Garcia, H. E., Grodsky, A., Locarnini, R. A., Mishonov, A. V., Paver, C. R., Reagan, J. R., & Seidov, D. (2018). Noaa atlas nesdis 87. *World Ocean Database*, 1–207.

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

Cavan, E. L., Giering, S. L. C., Wolff, G. A., Trimmer, M., & Sanders, R. (2018). Alternative Particle Formation Pathways in the Eastern Tropical North Pacific ' s Biological Carbon Pump. *Jounral of Geophysical Research: Biogeosciences*, *123*, 2198–2211. https://doi.org/10.1029/2018JG004392

Cepeda-Morales, J., Beier, E., Gaxiola-Castro, G., Lavín, M., & Godínez, V. (2009). Effect of the oxygen minimum zone on the second chlorophyll maximum. *Ciencias Marinas*, *35*(4), 389–403. https://doi.org/10.7773/cm.v35i4.1622

Chang, B. X., Devol, A. H., & Emerson, S. R. (2012). Fixed nitrogen loss from the eastern tropical North Pacific and Arabian Sea oxygen deficient zones determined from measurements of N_2:Ar. *Global Biogeochemical Cycles*, *26*(3), 2011GB004207. https://doi.org/10.1029/2011GB004207

Chin, C. C., Anderson, P. M., & Wold, F. (1983). The amino acid sequence of Escherichia coli cyanase. *Journal of Biological Chemistry*, *258*(1), 276–282. https://doi.org/10.1016/S0021-9258(18)33253-8

Cho, B. C., Park, M. G., Shim, J. H., & Azam, F. (1996). Significance of bacteria in urea dynamics in coastal surface waters. *Marine Ecology Progress Series*, *142*, 19–26.

Clegg, S. L., & Whitfield, M. (1990). A generalized model for the scavenging

of trace metals in the open ocean—I. Particle cycling. *Deep Sea Research Part A. Oceanographic Research Papers*, *37*(5), 809–832. https://doi.org/10.1016/0198-0149(90)90008-J

Connelly, T. L., Baer, S. E., Cooper, J. T., Bronk, D. A., & Wawrik, B. (2014). Urea Uptake and Carbon Fixation by Marine Pelagic Bacteria and Archaea during the Arctic Summer and Winter Seasons. *Applied and Environmental Microbiology*, *80*(19), 6013–6022. https://doi.org/10.1128/AEM.01431-14

Cram, J. A., Fuchsman, C. A., Duffy, M. E., Pretty, J. L., Lekanoff, R. M., Neibauer, J. A., Leung, S. W., Huebert, K. B., Weber, T. S., Bianchi, D., Evans, N., Devol, A. H., Keil, R. G., & McDonnell, A. M. P. (2022). Slow particle remineralization, rather than suppressed disaggregation, drives efficient flux transfer through the Eastern Tropical North Pacific Oxygen Deficient Zone. *Global Biogeochemical Cycles*, *36*, e2021GB007080-e2021GB007080. https://doi.org/10.1002/essoar.10507130.1

Cram, J. A., Hollins, A., McCarty, A. J., Martinez, G., Cui, M., Gomes, M. L., & Fuchsman, C. A. (2024). *Microbial diversity and abundance vary along salinity, oxygen, and particle size gradients in the Chesapeake Bay*. https://ami-journals.onlinelibrary.wiley.com/doi/abs/10.1111/1462-2920.16557

Cram, J. A., Weber, T., Leung, S. W., McDonnell, A. M. P., Liang, J. H., & Deutsch, C. (2018). The Role of Particle Size, Ballast, Temperature, and Oxygen in the Sinking Flux to the Deep Sea. *Global Biogeochemical Cycles*, *32*(5), 858–876. https://doi.org/10.1029/2017GB005710

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

Datta, M. S., Sliwerska, E., Gore, J., Polz, M. F., & Cordero, O. X. (2016). Microbial interactions lead to rapid micro-scale successions on model marine particles. *Nature Communications*, *7*(May), 1–7. https://doi.org/10.1038/ncomms11965

Delmont, T. O., Pierella Karlusich, J. J., Veseli, I., Fuessel, J., Eren, A. M., Foster, R. A., Bowler, C., Wincker, P., & Pelletier, E. (2022). Heterotrophic bacterial diazotrophs are more abundant than their cyanobacterial counterparts in metagenomes covering most of the sunlit ocean. *The ISME Journal*, *16*(4), 927–936. https://doi.org/10.1038/s41396-021-01135-1

DeVries, T., Deutsch, C., Rafter, P. A., & Primeau, F. (2013). Marine denitrification rates determined from a global 3-D inverse model. *Biogeosciences*, *10*(4), 2481–2496. https://doi.org/10.5194/bg-10-2481-2013

Diaz, B. P., Knowles, B., Johns, C. T., Laber, C. P., Bondoc, K. G. V., Haramaty, L., Natale, F., Harvey, E. L., Kramer, S. J., Bolaños, L. M.,

Lowenstein, D. P., Fredricks, H. F., Graff, J., Westberry, T. K., Mojica, K. D. A., Haëntjens, N., Baetge, N., Gaube, P., Boss, E., … Bidle, K. D. (2021). Seasonal mixed layer depth shapes phytoplankton physiology, viral production, and accumulation in the North Atlantic. *Nature Communications*, *12*(1), 6634. https://doi.org/10.1038/s41467-021-26836-1

Dick, G. J. (2019). The microbiomes of deep-sea hydrothermal vents: Distributed globally, shaped locally. *Nature Reviews Microbiology*, *17*(5), 271–283. https://doi.org/10.1038/s41579-019-0160-2

Dirnhuber, P., & Schütz, F. (1948). The isomeric transformation of urea into ammonium cyanate in aqueous solutions. *Biochemical Journal*, *42*(4), 628.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Elbrecht, V., & Leese, F. (2015). Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass-sequence relationships with an innovative metabarcoding protocol. *PLoS ONE*, *10*(7), e0130324–e0130324. https://doi.org/10.1371/journal.pone.0130324

Eloe, E. A., Shulse, C. N., Fadrosh, D. W., Williamson, S. J., Allen, E. E., & Bartlett, D. H. (2011). Compositional differences in particle-associated and free-living microbial assemblages from an extreme deep-ocean environment. *Environmental Microbiology Reports*, *3*(4), 449–458. https://doi.org/10.1111/j.1758-2229.2010.00223.x

Emerson, S. (2014). Annual net community production and biological carbon flux in the ocean. *Global Biogeochemical Cycles*, *28*, 14–28. https://doi.org/10.1002/2013GB004680.Received

Emerson, D., Rentz, J. A., Lilburn, T. G., Davis, R. E., Aldrich, H., Chan, C., & Moyer, C. L. (2007). A Novel Lineage of Proteobacteria Involved in Formation of Marine Fe-Oxidizing Microbial Mat Communities. *PLOS ONE*, *2*(8), e667. https://doi.org/10.1371/journal.pone.0000667

Flombaum, P., Gallegos, J. L., Gordillo, R. A., Rincón, J., Zabala, L. L., Jiao, N., Karl, D. M., Li, W. K., Lomas, M. W., & Veneziano, D. (2013). Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proceedings of the National Academy of Sciences*, *110*(24), 9824–9829.

Freitas, S., Hatosy, S., Fuhrman, J. A., Huse, S. M., Mark Welch, D. B., Sogin, M. L., & Martiny, A. C. (2012). Global distribution and diversity of marine Verrucomicrobia. *The ISME Journal*, *6*(8), 1499–1505. https://doi.org/10.1038/ismej.2012.3

Fu, S., Wang, R., Xu, Z., Zhou, H., Qiu, Z., Shen, L., & Yang, Q. (2023). Metagenomic sequencing combined with flow cytometry facilitated a novel microbial risk assessment framework for bacterial pathogens in municipal wastewater without cultivation. *iMeta*, *2*(1), e77. https://doi.org/10.1002/imt2.77

Fuchsman, C. A., Cherubini, L., & Hays, M. D. (2022). An analysis of protists in Pacific oxygen deficient zones: Implications for *Prochlorococcus* and N_2-producing bacteria. *Environmental Microbiology*, *24*(4), 1790–1804. https://doi.org/10.1111/1462-2920.15893

Fuchsman, C. A., Devol, A. H., Casciotti, K. L., Buchwald, C., Chang, B. X., & Horak, R. E. A. (2018). An N isotopic mass balance of the Eastern Tropical North Pacific oxygen deficient zone. *Deep Sea Research Part II: Topical Studies in Oceanography*, *156*, 137–147. https://doi.org/10.1016/j.dsr2.2017.12.013

Fuchsman, C. A., Devol, A. H., Saunders, J. K., McKay, C., & Rocap, G. (2017). Niche Partitioning of the N Cycling Microbial Community of an Offshore Oxygen Deficient Zone. *Frontiers in Microbiology*, *8*, 2384. https://doi.org/10.3389/fmicb.2017.02384

Fuchsman, C. A., Garcia Prieto, D., Hays, M. D., & Cram, J. A. (2023). Associations between picocyanobacterial ecotypes and cyanophage host genes across ocean basins and depth. *PeerJ*, *11*, e14924. https://doi.org/10.7717/peerj.14924

Fuchsman, C. A., & Hays, M. D. (2023). Increased cyanophage infection at the bottom of the euphotic zone, especially in the fall. *Environmental Microbiology*, 1462-2920.16525. https://doi.org/10.1111/1462-2920.16525

Fuchsman, C. A., Kirkpatrick, J. B., Brazelton, W. J., Murray, J. W., & Staley, J. T. (2011). Metabolic strategies of free-living and aggregate-associated bacterial communities inferred from biologic and chemical profiles in the Black Sea suboxic zone. *FEMS Microbiology Ecology*, *78*(3), 586–603.

Fuchsman, C. A., Palevsky, H. I., Widner, B., Duffy, M., Carlson, M. C. G., Neibauer, J. A., Mulholland, M. R., Keil, R. G., Devol, A. H., & Rocap, G. (2019). Cyanobacteria and cyanophage contributions to carbon and nitrogen cycling in an oligotrophic oxygen-deficient zone. *The ISME Journal*, *13*(11), 2714–2726. https://doi.org/10.1038/s41396-019-0452-6

Fuchsman, C. A., Paul, B., Staley, J. T., Yakushev, E. V., & Murray, J. W. (2019). Detection of transient denitrification during a high organic matter event in the Black Sea. *Global Biogeochemical Cycles*, *33*(2), 143–162.

Fuchsman, C. A., Staley, J. T., Oakley, B. B., Kirkpatrick, J. B., & Murray, J. W. (2012). Free-living and aggregate-associated Planctomycetes in the Black Sea. In *FEMS Microbiology Ecology* (Vol. 80, Issue 2, pp. 402–416).

Ganesh, S., Bertagnolli, A. D., Bristow, L. A., Padilla, C. C., Blackwood, N., Aldunate, M., Bourbonnais, A., Altabet, M. A., Malmstrom, R. R., Woyke, T., Ulloa, O., Konstantinidis, K. T., Thamdrup, B., & Stewart, F. J. (2018). Single cell genomic and transcriptomic evidence for the use of alternative nitrogen substrates by anammox bacteria. *The ISME Journal*, *12*(11), 2706–2722. https://doi.org/10.1038/s41396-018-0223-

9
Ganesh, S., Bristow, L. A., Larsen, M., Sarode, N., Thamdrup, B., & Stewart, F. J. (2015). Size-fraction partitioning of community gene transcription and nitrogen metabolism in a marine oxygen minimum zone. *The ISME Journal*, *9*(12), 2682–2696. https://doi.org/10.1038/ismej.2015.44
Ganesh, S., Parris, D. J., DeLong, E. F., & Stewart, F. J. (2014). Metagenomic analysis of size-fractionated picoplankton in a marine oxygen minimum zone. *The ISME Journal*, *8*(1), 187–211. https://doi.org/10.1038/ismej.2013.144
Giovannoni, S. J. (2017). SAR11 Bacteria: The Most Abundant Plankton in the Oceans. *Annual Reviews of Marine Science*, *9*, 231–255. https://doi.org/10.1146/annurev-marine-010814-015934
Giovannoni, S. J., Tripp, H. J., Givan, S., Podar, M., Vergin, K. L., Baptista, D., Bibbs, L., Eads, J., Richardson, T. H., Noordewier, M., Rappé, M. S., Short, J. M., Carrington, J. C., & Mathur, E. J. (2005). Genome Streamlining in a Cosmopolitan Oceanic Bacterium. In *Science* (Vol. 309, Issue 5738, pp. 1242–1245).
Glibert, P. M., Wilkerson, F. P., Dugdale, R. C., Raven, J. A., Dupont, C. L., Leavitt, P. R., Parker, A. E., Burkholder, J. M., & Kana, T. M. (2016). Pluses and minuses of ammonium and nitrate uptake and assimilation by phytoplankton and implications for productivity and community composition, with emphasis on nitrogen-enriched conditions: Pluses and minuses of NH4+ *and* NO3−. *Limnology and Oceanography*, *61*(1), 165–197. https://doi.org/10.1002/lno.10203
Gloor, G. B., Wu, J. R., Pawlowsky-Glahn, V., & Egozcue, J. J. (2016). It's all relative: Analyzing microbiome data as compositions. *Annals of Epidemiology*, *26*(5), 322–329. https://doi.org/10.1016/j.annepidem.2016.03.003
Golden, S. S., Cho, D. S., & Nalty, M. S. (1989). Two functional psbD genes in the cyanobacterium *Synechococcus* sp. Strain PCC 7942. *Journal of Bacteriology*, *171*(9), 4707–4713. https://doi.org/10.1128/jb.171.9.4707-4713.1989
Gouy, M., Guindon, S., & Gascuel, O. (2010). SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Molecular Biology and Evolution*, *27*(2), 221–224. https://doi.org/10.1093/molbev/msp259
Guidi, L., Jackson, G. A., Stemmann, L., Miquel, J. C., Picheral, M., & Gorsky, G. (2008). Relationship between particle size distribution and flux in the mesopelagic zone. *Deep-Sea Research Part I: Oceanographic Research Papers*, *55*(10), 1364–1374. https://doi.org/10.1016/j.dsr.2008.05.014
Guidi, L., Stemmann, L., Jackson, G. A., Ibanez, F., Claustre, H., Legendre, L., Picheral, M., & Gorsky, G. (2009). Effects of phytoplankton community on production, size and export of large aggregates: A world-ocean analysis. *Limnology and Oceanography*, *54*(6), 1951–1963.

Hausinger, R. P. (2004). Metabolic Versatility of Prokaryotes for Urea Decomposition. *Journal of Bacteriology*, *186*(9), 2520–2522. https://doi.org/10.1128/JB.186.9.2520-2522.2004

Hedlund, B. P., Gosink, J. J., & Staley, J. T. (1997). Verrucomicrobia div. Nov., a new division of the Bacteria containing three new species of Prosthecobacter. *Antonie van Leeuwenhoek*, *72*(1), 29–38. https://doi.org/10.1023/A:1000348616863

Horak, R. E. A., Ruef, W., Ward, B. B., & Devol, A. H. (2016). Expansion of denitrification and anoxia in the eastern tropical North Pacific from 1972 to 2012. *Geophysical Research Letters*, *43*(10), 5252–5260. https://doi.org/10.1002/2016GL068871

Ito, T., Minobe, S., Long, M. C., & Deutsch, C. (2017). Upper ocean O_2 trends: 1958–2015. *Geophysical Research Letters*, *44*(9), 4214–4223. https://doi.org/10.1002/2017GL073613

Jenkins, W. J., Smethie, W. M., Boyle, E. A., & Cutter, G. A. (2015). Water mass analysis for the U.S. GEOTRACES (GA03) North Atlantic sections. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *116*, 6–20. https://doi.org/10.1016/j.dsr2.2014.11.018

Jin, X., Gruber, N., Dunne, J. P., Sarmiento, J. L., & Armstrong, R. A. (2006). Diagnosing the contribution of phytoplankton functional groups to the production and export of particulate organic carbon, $CaCO_3$, and opal from global nutrient and alkalinity distributions. *Global Biogeochemical Cycles*, *20*(2), 2005GB002532. https://doi.org/10.1029/2005GB002532

Johnson, W. V., & Anderson, P. M. (1987). Bicarbonate is a recycling substrate for cyanase. *Journal of Biological Chemistry*, *262*(19), 9021–9025. https://doi.org/10.1016/S0021-9258(18)48040-4

Johnson, Z. I., Zinser, E. R., Coe, A., Mcnulty, N. P., Malcolm, E. S., Chisholm, S. W., Woodward, E. M. S., & Chisholm, S. W. (2006). Partitioning Among *Prochlorococcus* Ecotypes Along Environmental Gradients. *Science*, *311*(March), 1737–1740. https://doi.org/10.1126/science.1118052

Johnson, Z. I., Zinser, E. R., Coe, A., McNulty, N. P., Woodward, E. M. S., & Chisholm, S. W. (2006). Niche Partitioning Among *Prochlorococcus* Ecotypes Along Ocean-Scale Environmental Gradients. In *Science* (Vol. 311, Issue 5768, pp. 1737–1740).

Kamennaya, N. A., & Post, A. F. (2011). Characterization of cyanate metabolism in marine *Synechococcus* and *Prochlorococcus* spp. *Applied and Environmental Microbiology*, *77*(1), 291--301.

Karl, D., Letelier, R., Bidigare, R., Björkman, K., Church, M., Dore, J., & White, A. (2021). Seasonal-to-decadal scale variability in primary production and particulate matter export at Station ALOHA. *PROGRESS IN OCEANOGRAPHY*, *195*. https://doi.org/10.1016/j.pocean.2021.102563

Kessler, W. S. (2006). The circulation of the eastern tropical Pacific: A review. *Progress in Oceanography*, *69*(2–4), 181–217. https://doi.org/10.1016/j.pocean.2006.03.009

Kitzinger, K., Padilla, C. C., Marchant, H. K., Hach, P. F., Herbold, C. W., Kidane, A. T., Könneke, M., Littmann, S., Mooshammer, M., Niggemann, J., Petrov, S., Richter, A., Stewart, F. J., Wagner, M., Kuypers, M. M. M., & Bristow, L. A. (2019). Cyanate and urea are substrates for nitrification by Thaumarchaeota in the marine environment. *Nature Microbiology*, *4*(2), 234–243. https://doi.org/10.1038/s41564-018-0316-2

Koper, T. E., El-Sheikh, A. F., Norton, J. M., & Klotz, M. G. (2004). Urease-Encoding Genes in Ammonia-Oxidizing Bacteria. *Applied and Environmental Microbiology*, *70*(4), 2342–2348. https://doi.org/10.1128/AEM.70.4.2342-2348.2004

Lam, P., Lavik, G., Jensen, M. M., Van De Vossenberg, J., Schmid, M., Woebken, D., Gutiérrez, D., Amann, R., Jetten, M. S. M., & Kuypers, M. M. M. (2009). Revising the nitrogen cycle in the Peruvian oxygen minimum zone. *Proceedings of the National Academy of Sciences*, *106*(12), 4752–4757. https://doi.org/10.1073/pnas.0812444106

Lanclos, V. C., Rasmussen, A. N., Kojima, C. Y., Cheng, C., Henson, M. W., Faircloth, B. C., Francis, C. A., & Thrash, J. C. (2023). Ecophysiology and genomics of the brackish water adapted SAR11 subclade IIIa. *The ISME Journal*, *17*(4), 620–629. https://doi.org/10.1038/s41396-023-01376-2

Landolfi, A., Dietze, H., & Volpe, G. (2016). Longitudinal variability of organic nutrients in the North Atlantic subtropical gyre. *Deep Sea Research Part I: Oceanographic Research Papers*, *111*, 50–60.

Le Moigne, F. A. C. (2019). Pathways of Organic Carbon Downward Transport by the Oceanic Biological Carbon Pump. *Frontiers in Marine Science*, *6*, 634. https://doi.org/10.3389/fmars.2019.00634

Lecleir, G. R., Debruyn, J. M., Maas, E. W., Boyd, P. W., & Wilhelm, S. W. (2014). Temporal changes in particle-associated microbial communities after interception by nonlethal sediment traps. *FEMS Microbiology Ecology*, *87*, 153–163. https://doi.org/10.1111/1574-6941.12213

Leu, A. O., Eppley, J. M., Burger, A., & Delong, E. F. (2022). Diverse Genomic Traits Differentiate Sinking-Particle-Associated versus Free-Living Microbes throughout the Oligotrophic Open Ocean Water Column. *mBio*, *13*, e01569-22. https://doi.org/10.1128/mbio.01569-22

Li, G., Cheng, L., Zhu, J., Trenberth, K. E., Mann, M. E., & Abraham, J. P. (2020). Increasing ocean stratification over the past half-century. *Nature Climate Change*, *10*(12), 1116–1123. https://doi.org/10.1038/s41558-020-00918-2

Li, J., Gu, L., Bai, S., Wang, J., Su, L., Wei, B., Zhang, L., & Fang, J. (2021). Characterization of particle-associated and free-living bacterial and archaeal communities along the water columns of the South China Sea. *Biogeosciences*, *18*(1), 113–133. https://doi.org/10.5194/bg-18-113-2021

Liu, R., Wang, L., Liu, Q., Wang, Z., Li, Z., Fang, J., Zhang, L., & Luo, M.

(2018). Depth-Resolved Distribution of Particle-Attached and Free-Living Bacterial Communities in the Water Column of the New Britain Trench. *Frontiers in Microbiology*, *9*, 625. https://doi.org/10.3389/fmicb.2018.00625

Lücker, S., Nowka, B., Rattei, T., Spieck, E., & Daims, H. (2013). The Genome of Nitrospina gracilis Illuminates the Metabolism and Evolution of the Major Marine Nitrite Oxidizer. *Frontiers in Microbiology*, *4*. https://doi.org/10.3389/fmicb.2013.00027

Luo, E., Eppley, J. M., Romano, A. E., Mende, D. R., & DeLong, E. F. (2020). Double-stranded DNA virioplankton dynamics and reproductive strategies in the oligotrophic open ocean water column. *The ISME Journal*, *14*(5), 1304–1315. https://doi.org/10.1038/s41396-020-0604-8

Luo, E., Leu, A. O., Eppley, J. M., Karl, D. M., & DeLong, E. F. (2022). Diversity and origins of bacterial and archaeal viruses on sinking particles reaching the abyssal ocean. *ISME Journal*, *16*, 1627–1635. https://doi.org/10.1038/s41396-022-01202-1

Mao, X., Chen, J., van Oosterhout, C., Zhang, H., Liu, G., Zhuang, Y., & Mock, T. (2022). Diversity, prevalence, and expression of cyanase genes (*cynS*) in planktonic marine microorganisms. *The ISME Journal*, *16*(2), 602–605. https://doi.org/10.1038/s41396-021-01081-y

Marshall, K. T., & Morris, R. M. (2013). Isolation of an aerobic sulfur oxidizer from the SUP05/Arctic96BD-19 clade. *The ISME Journal*, *7*(2), 452–455. https://doi.org/10.1038/ismej.2012.78

Marshall, K. T., & Morris, R. M. (2015). Genome Sequence of " *Candidatus* Thioglobus singularis" Strain PS1, a Mixotroph from the SUP05 Clade of Marine Gammaproteobacteria. *Genome Announcements*, *3*(5), e01155-15. https://doi.org/10.1128/genomeA.01155-15

Martens-Habbena, W., Berube, P. M., Urakawa, H., de la Torre, J. R., & Stahl, D. A. (2009). Ammonia oxidation kinetics determine niche separation of nitrifying Archaea and Bacteria. *Nature*, *461*(7266), 976–979. https://doi.org/10.1038/nature08465

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet*, *17*(1), 10–12.

Martiny, A. C., Kathuria, S., & Berube, P. M. (2009). Widespread metabolic potential for nitrite and nitrate assimilation among *Prochlorococcus* ecotypes. *Proceedings of the National Academy of Sciences*, *106*(26), 10787–10792. https://doi.org/10.1073/pnas.0902532106

Mattes, T. E., Burke, S., Rocap, G., & Morris, R. M. (2022). Two Metatranscriptomic Profiles through Low-Dissolved-Oxygen Waters (DO, 0 to 33 µM) in the Eastern Tropical North Pacific Ocean. *Microbiology Resource Announcements*, *11*(2), e01201-21. https://doi.org/10.1128/mra.01201-21

McNichol, J., Berube, P. M., Biller, S. J., & Fuhrman, J. A. (2021). Evaluating and improving small subunit rRNA PCR primer coverage for bacteria, archaea, and eukaryotes using metagenomes from global ocean surveys. *Msystems*, *6*(3), 10.1128/msystems. 00565-21.

Mestre, M., Borrull, E., Sala, M. M., & Gasol, J. M. (2017). Patterns of bacterial diversity in the marine planktonic particulate matter continuum. *The ISME Journal*, *11*(4), 999–1010. https://doi.org/10.1038/ismej.2016.166

Mestre, M., Ruiz-González, C., Logares, R., Duarte, C. M., Gasol, J. M., & Sala, M. M. (2018). Sinking particles promote vertical connectivity in the ocean microbiome. *Proceedings of the National Academy of Sciences*, *115*(29). https://doi.org/10.1073/pnas.1802470115

Middag, R., De Baar, H. J. W., & Bruland, K. W. (2019). The Relationships Between Dissolved Zinc and Major Nutrients Phosphate and Silicate Along the GEOTRACES GA02 Transect in the West Atlantic Ocean. *Global Biogeochemical Cycles*, *33*(1), 63–84. https://doi.org/10.1029/2018GB006034

Moore, C. M., Mills, M. M., Arrigo, K. R., Berman-Frank, I., Bopp, L., Boyd, P. W., Galbraith, E. D., Geider, R. J., Guieu, C., Jaccard, S. L., Jickells, T. D., La Roche, J., Lenton, T. M., Mahowald, N. M., Marañón, E., Marinov, I., Moore, J. K., Nakatsuka, T., Oschlies, A., … Ulloa, O. (2013). Processes and patterns of oceanic nutrient limitation. *Nature Geoscience*, *6*(9), 701–710. https://doi.org/10.1038/ngeo1765

Morris, R. M., Rappe, M. S., Connon, S. A., Vergin, K. L., Siebold, W. A., Carlson, C. A., & Giovannoni, S. J. (2002). SAR11 clade dominates ocean surface bacterioplankton communities. *Nature*, *420*(December), 806–810. https://doi.org/10.1038/nature01281.1

Morris, R. M., Rappé, M. S., Connon, S. A., Vergin, K. L., Siebold, W. A., Carlson, C. A., & Giovannoni, S. J. (2002). SAR11 clade dominates ocean surface bacterioplankton communities. *Nature*, *420*(6917), 806–810. https://doi.org/10.1038/nature01240

Nguyen, R. T., & Harvey, H. R. (1997). Protein and amino acid cycling during phytoplankton decomposition in oxic and anoxic waters. *Organic Geochemistry*, *27*(3–4), 115–128. https://doi.org/10.1016/S0146-6380(97)00076-4

Nguyen, T. T. H., Zakem, E. J., Ebrahimi, A., Schwartzman, J., Caglar, T., Amarnath, K., Alcolombri, U., Peaudecerf, F. J., Hwa, T., Stocker, R., Cordero, O. X., & Levine, N. M. (2022). Microbes contribute to setting the ocean carbon flux by altering the fate of sinking particulates. *Nature Communications*, *13*(1), 1657. https://doi.org/10.1038/s41467-022-29297-2

Pachiadaki, M. G., Sintes, E., Bergauer, K., Brown, J. M., Record, N. R., Swan, B. K., Mathyer, M. E., Hallam, S. J., Lopez-Garcia, P., Takaki, Y., Nunoura, T., Woyke, T., Herndl, G. J., & Stepanauskas, R. (2017). Major role of nitrite-oxidizing bacteria in dark ocean carbon fixation. *Science*, *358*(6366), 1046–1051. https://doi.org/10.1126/science.aan8260

Painter, S., Sanders, R., Waldron, H., Lucas, M., & Torres-Valdes, S. (2008). Urea distribution and uptake in the Atlantic Ocean between 50°N and 50°S. *Marine Ecology Progress Series*, *368*, 53–63.

https://doi.org/10.3354/meps07586
Palatinszky, M., Herbold, C., Jehmlich, N., Pogoda, M., Han, P., von Bergen, M., Lagkouvardos, I., Karst, S. M., Galushko, A., Koch, H., Berry, D., Daims, H., & Wagner, M. (2015). Cyanate as an energy source for nitrifiers. *Nature*, *524*(7563), 105–108. https://doi.org/10.1038/nature14856
Parada, A. E., Needham, D. M., & Fuhrman, J. A. (2016). Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology*, *18*(5), 1403–1414. https://doi.org/10.1111/1462-2920.13023
Paulmier, A., & Ruiz-Pino, D. (2009). Oxygen minimum zones (OMZs) in the modern ocean. *Progress in Oceanography*, *80*(3), 113–128. https://doi.org/10.1016/j.pocean.2008.08.001
Pesant, S., Not, F., Picheral, M., Kandels-Lewis, S., Le Bescot, N., Gorsky, G., Iudicone, D., Karsenti, E., Speich, S., Troublé, R., Dimier, C., Searson, S., Acinas, S. G., Bork, P., Boss, E., Bowler, C., De Vargas, C., Follows, M., Gorsky, G., … Tara Oceans Consortium Coordinators. (2015). Open science resources for the discovery and analysis of Tara Oceans data. *Scientific Data*, *2*(1), 150023. https://doi.org/10.1038/sdata.2015.23
Peters, B., Horak, R., Devol, A., Fuchsman, C., Forbes, M., Mordy, C. W., & Casciotti, K. L. (2018). Estimating fixed nitrogen loss and associated isotope effects using concentration and isotope measurements of NO3–, NO2–, and N2 from the Eastern Tropical South Pacific oxygen deficient zone. *Deep Sea Research Part II: Topical Studies in Oceanography*, *156*, 121–136. https://doi.org/10.1016/j.dsr2.2018.02.011
Pitt, K. A., Welsh, D. T., & Condon, R. H. (2009). Influence of jellyfish blooms on carbon, nitrogen and phosphorus cycling and plankton production. *Hydrobiologia*, *616*, 133–149.
Ploug, H., Kühl, M., Buchholz-Cleven, B., & Jørgensen, B. B. (1997). Anoxic aggregates-an ephemeral phenomenon in the pelagic environment? *Aquatic Microbial Ecology*, *13*(3), 285–294.
Price, N. M., Cochlan, W. P., & Harrison, P. J. (1985). Time course of uptake of inorganic and organic nitrogen by phytoplankton in the Strait of Georgia: Comparison of frontal and stratified communities. *Marine Ecology Progress Series*, 39–53.
Priddle, J., Whitehouse, M. J., Atkinson, A., Brierley, A. S., & Murphy, E. J. (1997). Diurnal changes in near-surface ammonium concentration—Interplay between zooplankton and phytoplankton. *Journal of Plankton Research*, *19*(9), 1305–1330. https://doi.org/10.1093/plankt/19.9.1305
Qin, W., Amin, S. A., Martens-Habbena, W., Walker, C. B., Urakawa, H., Devol, A. H., Ingalls, A. E., Moffett, J. W., Armbrust, E. V., & Stahl, D. A. (2014). Marine ammonia-oxidizing archaeal isolates display obligate mixotrophy and wide ecotypic variation. *Proceedings of the National*

Academy of Sciences, *111*(34), 12504–12509. https://doi.org/10.1073/pnas.1324115111

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, *41*(D1), 590–596. https://doi.org/10.1093/nar/gks1219

Raven, M. R., Keil, R. G., & Webb, S. M. (2021). Microbial sulfate reduction and organic sulfur formation in sinking marine particles. *Science*, *371*(6525), 178–181. https://doi.org/10.1126/science.abc6035

Reintjes, G., Heins, A., Wang, C., & Amann, R. (2023). Abundance and composition of particles and their attached microbiomes along an Atlantic Meridional Transect. *Frontiers in Marine Science*, *10*(March), 1051510–1051510. https://doi.org/10.3389/fmars.2023.1051510

Ren, J., Ahlgren, N. A., Lu, Y. Y., Fuhrman, J. A., & Sun, F. (2017). VirFinder: A novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome*, *5*, 69–69. https://doi.org/10.1186/s40168-017-0283-5

Revsbech, N. P., Larsen, L. H., Gundersen, J., Dalsgaard, T., Ulloa, O., & Thamdrup, B. (2009). Determination of ultra-low oxygen concentrations in oxygen minimum zones by the STOX sensor. *Limnology and Oceanography: Methods*, *7*(5), 371–381. https://doi.org/10.4319/lom.2009.7.371

Rigby, S. J., Williams, R. G., Achterberg, E. P., & Tagliabue, A. (2020). Resource Availability and Entrainment Are Driven by Offsets Between Nutriclines and Winter Mixed-Layer Depth. *Global Biogeochemical Cycles*, *34*(7), e2019GB006497. https://doi.org/10.1029/2019GB006497

Rii, Y., Karl, D., & Church, M. (2016). Temporal and vertical variability in picophytoplankton primary productivity in the North Pacific Subtropical Gyre. *Marine Ecology Progress Series*, *562*, 1–18. https://doi.org/10.3354/meps11954

Rocap, G., Larimer, F. W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N. A., Arellano, A., Coleman, M., Hauser, L., Hess, W. R., Johnson, Z. I., Land, M., Lindell, D., Post, A. F., Regala, W., Shah, M., Shaw, S. L., Steglich, C., Sullivan, M. B., … Chisholm, S. W. (2003). Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature*, *424*(6952), 1042–1047. https://doi.org/10.1038/nature01947

Saunders, J. K., Fuchsman, C. A., McKay, C., & Rocap, G. (2019). Complete arsenic-based respiratory cycle in the marine microbial communities of pelagic oxygen-deficient zones. *Proceedings of the National Academy of Sciences*, *116*(20), 9925–9930. https://doi.org/10.1073/pnas.1818349116

Scanlan, D. J. (2012). Marine Picocyanobacteria. In *Ecology of Cyanobacteria II: Their Diversity in Space and Time* (pp. 503–533).

Springer.
Schlitzer, R. (2023). Ocean Data View. *Odv.Awi.De*.
Schmitz, R. A., Peeters, S. H., Versantvoort, W., Picone, N., Pol, A., Jetten, M. S. M., & Op den Camp, H. J. M. (2021). Verrucomicrobial methanotrophs: Ecophysiology of metabolically versatile acidophiles. *FEMS Microbiology Reviews*, *45*(5), fuab007. https://doi.org/10.1093/femsre/fuab007
Shah, V., Chang, B. X., & Morris, R. M. (2017). Cultivation of a chemoautotroph from the SUP05 clade of marine bacteria that produces nitrite and consumes ammonium. *The ISME Journal*, *11*(1), 263–271. https://doi.org/10.1038/ismej.2016.87
Shelton, A. O., Gold, Z. J., Jensen, A. J., D′Agnese, E., Andruszkiewicz Allan, E., Van Cise, A., Gallego, R., Ramón-Laca, A., Garber-Yonts, M., Parsons, K., & Kelly, R. P. (2023). Toward quantitative metabarcoding. *Ecology*, *104*(2), e3906. https://doi.org/10.1002/ecy.3906
Shilova, I. N., Mills, M. M., Robidart, J. C., Turk-Kubo, K. A., Björkman, K. M., Kolber, Z., Rapp, I., van Dijken, G. L., Church, M. J., Arrigo, K. R., Achterberg, E. P., & Zehr, J. P. (2017). Differential effects of nitrate, ammonium, and urea as N sources for microbial communities in the North Pacific Ocean. *Limnology and Oceanography*, *62*(6), 2550–2574. https://doi.org/10.1002/lno.10590
Shiozaki, T., Hashihama, F., Endo, H., Ijichi, M., Takeda, N., Makabe, A., Fujiwara, A., Nishino, S., & Harada, N. (2021). Assimilation and oxidation of urea-derived nitrogen in the summer Arctic Ocean. *Limnology and Oceanography*, *66*(12), 4159–4170. https://doi.org/10.1002/lno.11950
Sieracki, M. E., Verity, P. G., & Stoecker, D. K. (1993). Plankton community response to sequential silicate and nitrate depletion during the 1989 North Atlantic spring bloom. *Deep Sea Research Part II: Topical Studies in Oceanography*, *40*(1–2), 213–225.
Simon, M., Grossart, H.-P., Schweitzer, B., & Ploug, H. (2002). Microbial ecology of organic aggregates in aquatic ecosystems. *Aquatic Microbial Ecology*, *28*(2), 175–211.
Spieck, E., Keuter, S., Wenzel, T., Bock, E., & Ludwig, W. (2014). Characterization of a new marine nitrite oxidizing bacterium, Nitrospina watsonii sp. Nov., a member of the newly proposed phylum “Nitrospinae.” *Systematic and Applied Microbiology*, *37*(3), 170–176. https://doi.org/10.1016/j.syapm.2013.12.005
Stramma, L., Johnson, G. C., Sprintall, J., & Mohrholz, V. (2008). Expanding Oxygen-Minimum Zones in the Tropical Oceans. *Science*, *320*(5876), 655–658. https://doi.org/10.1126/science.1153847
Sun, X., Kop, L. F. M., Lau, M. C. Y., Frank, J., Jayakumar, A., Lücker, S., & Ward, B. B. (2019). Uncultured Nitrospina-like species are major nitrite oxidizing bacteria in oxygen minimum zones. *The ISME Journal*, *13*(10), 2391–2402. https://doi.org/10.1038/s41396-019-0443-7
Sylvan, J. B., Pyenson, B. C., Rouxel, O., German, C. R., & Edwards, K. J.

(2012). Time-series analysis of two hydrothermal plumes at 9°50′N East Pacific Rise reveals distinct, heterogeneous bacterial populations. *Geobiology*, *10*(2), 178–192. https://doi.org/10.1111/j.1472-4669.2011.00315.x

Takeda, N., Hashihama, F., & Kanda, J. (2020). Automated colorimetric determination of nanomolar urea in seawater by gas-segmented continuous flow analysis using a liquid waveguide capillary cell. *Talanta*, *208*, 120371. https://doi.org/10.1016/j.talanta.2019.120371

Tamelander T. (2013). Community composition and extracellular enzyme activity of bacteria associated with suspended and sinking particles in contrasting arctic and sub-arctic marine environments. *Aquatic Microbial Ecology*, *69*(3), 211–221.

Tettamanti Boshier, F. A., Srinivasan, S., Lopez, A., Hoffman, N. G., Proll, S., Fredricks, D. N., & Schiffer, J. T. (2020). Complementing 16S rRNA Gene Amplicon Sequencing with Total Bacterial Load To Infer Absolute Species Concentrations in the Vaginal Microbiome. *mSystems*, *5*(2), e00777-19. https://doi.org/10.1128/mSystems.00777-19

Thamdrup, B. (2012). New Pathways and Processes in the Global Nitrogen Cycle. *Annual Review of Ecology, Evolution, and Systematics*, *43*(1), 407–428. https://doi.org/10.1146/annurev-ecolsys-102710-145048

Thibodeau, P. S., Steinberg, D. K., & Maas, A. E. (2020). Effects of temperature and food concentration on pteropod metabolism along the Western Antarctic Peninsula. *Journal of Experimental Marine Biology and Ecology*, *530*, 151412.

Tiano, L., Garcia-Robledo, E., Dalsgaard, T., Devol, A. H., Ward, B. B., Ulloa, O., Canfield, D. E., & Peter Revsbech, N. (2014). Oxygen distribution and aerobic respiration in the north and south eastern tropical Pacific oxygen minimum zones. *Deep Sea Research Part I: Oceanographic Research Papers*, *94*, 173–183. https://doi.org/10.1016/j.dsr.2014.10.001

Tolar, B. B., Wallsgrove, N. J., Popp, B. N., & Hollibaugh, J. T. (2017). Oxidation of urea-derived nitrogen by thaumarchaeota-dominated marine nitrifying communities. *Environmental Microbiology*, *19*(12), 4838–4850. https://doi.org/10.1111/1462-2920.13457

Tourlousse, D. M., Yoshiike, S., Ohashi, A., Matsukura, S., Noda, N., & Sekiguchi, Y. (2016). Synthetic spike-in standards for high-throughput 16S rRNA gene amplicon sequencing. *Nucleic Acids Research*, gkw984. https://doi.org/10.1093/nar/gkw984

Tsementzi, D., Wu, J., Deutsch, S., Nath, S., Rodriguez-R, L. M., Burns, A. S., Ranjan, P., Sarode, N., Malmstrom, R. R., Padilla, C. C., Stone, B. K., Bristow, L. A., Larsen, M., Glass, J. B., Thamdrup, B., Woyke, T., Konstantinidis, K. T., & Stewart, F. J. (2016). SAR11 bacteria linked to ocean anoxia and nitrogen loss. *Nature*, *536*(7615), 179–183. https://doi.org/10.1038/nature19068

Ulloa, O., Delong, E. F., Letelier, R. M., & Stewart, F. J. (2012). Microbial

oceanography of anoxic oxygen minimum zones. *Proceedings of the National Academy of Sciences*, *109*(40), 15996–16003. https://doi.org/10.1073/pnas.1205009109

Ulloa, O., Henríquez-Castillo, C., Ramírez-Flandes, S., Plominsky, A. M., Murillo, A. A., Morgan-Lang, C., Hallam, S. J., & Stepanauskas, R. (2021). The cyanobacterium *Prochlorococcus* has divergent light-harvesting antennae and may have evolved in a low-oxygen ocean. In *Proceedings of the National Academy of Sciences* (Vol. 118, Issue 11, p. e2025638118).

Wakeham, S. G., & Canuel, E. A. (1988). Organic geochemistry of particulate matter in the eastern tropical North Pacific Ocean: Implications for particle dynamics. *Journal of Marine Research*, *46*(1), 183–213.

Walsh, D. A., Zaikova, E., Howes, C. G., Song, Y. C., Wright, J. J., Tringe, S. G., Tortell, P. D., & Hallam, S. J. (2009). Metagenome of a Versatile Chemolithoautotroph from Expanding Oceanic Dead Zones. *Science*, *326*(5952), 578–582. https://doi.org/10.1126/science.1175309

Widner, B., Fuchsman, C. A., Chang, B. X., Rocap, G., & Mulholland, M. R. (2018). Utilization of urea and cyanate in waters overlying and within the eastern tropical north Pacific oxygen deficient zone. *FEMS Microbiology Ecology*, *94*(10). https://doi.org/10.1093/femsec/fiy138

Widner, B., Mordy, C. W., & Mulholland, M. R. (2018). Cyanate distribution and uptake above and within the Eastern Tropical South Pacific oxygen deficient zone. *Limnology and Oceanography*, *63*(S1). https://doi.org/10.1002/lno.10730

Widner, B., & Mulholland, M. R. (2017). Cyanate distribution and uptake in North Atlantic coastal waters. *Limnology and Oceanography*, *62*(6), 2538–2549. https://doi.org/10.1002/lno.10588

Widner, B., Mulholland, M. R., & Mopper, K. (2013). Chromatographic Determination of Nanomolar Cyanate Concentrations in Estuarine and Sea Waters by Precolumn Fluorescence Derivatization. *Analytical Chemistry*, *85*(14), 6661–6666. https://doi.org/10.1021/ac400351c

Widner, B., Mulholland, M. R., & Mopper, K. (2016). Distribution, Sources, and Sinks of Cyanate in the Coastal North Atlantic Ocean. *Environmental Science & Technology Letters*, *3*(8), 297–302. https://doi.org/10.1021/acs.estlett.6b00165

Wu, Z., Wang, S., Zhang, Q., Hao, J., Lin, Y., Zhang, J., & Li, A. (2020). Assessing the intestinal bacterial community of farmed Nile tilapia (Oreochromis niloticus) by high-throughput absolute abundance quantification. *Aquaculture*, *529*, 735688. https://doi.org/10.1016/j.aquaculture.2020.735688

Yeh, Y. C., & Fuhrman, J. A. (2022). Effects of phytoplankton, viral communities, and warming on free-living and particle-associated marine prokaryotic community structure. *Nature Communications*, *13*(1), 7905–7905. https://doi.org/10.1038/s41467-022-35551-4

Yilmaz, A., Coban-Yildiz, Y., Tellikarakoc, F., & Bologa, A. (2006). Surface and mid-water sources of organic carbon by photoautotrophic and

chemoautotrophic production in the Black Sea. *Deep Sea Research Part II: Topical Studies in Oceanography*, *53*(17–19), 1988–2004. https://doi.org/10.1016/j.dsr2.2006.03.015

Zhang, I. H., Sun, X., Jayakumar, A., Fortin, S. G., Ward, B. B., & Babbin, A. R. (2023). Partitioning of the denitrification pathway and other nitrite metabolisms within global oxygen deficient zones. *ISME Communications*, *3*(1), 76. https://doi.org/10.1038/s43705-023-00284-y

Zinser, E. R., Johnson, Z. I., Coe, A., Karaca, E., Veneziano, D., & Chisholm, S. W. (2007). Influence of light and temperature on *Prochlorococcus* ecotype distributions in the Atlantic Ocean. *Limnology and Oceanography*, *52*(5), 2205–2220. https://doi.org/10.4319/lo.2007.52.5.2205

Zumft, W. G. (1997). Cell biology and molecular basis of denitrification. *Microbiology and Molecular Biology Reviews : MMBR*, *61*(4), 533–616.