UAB
Universitat Autònoma de Barcelona

Ecology of marine aerobic anoxygenic phototrophic bacteria

Carlota Ruiz Gazulla

Ecology of marine aerobic anoxygenic phototrophic bacteria

Carlota Ruiz Gazulla

PhD Thesis

Tesi Doctoral

Programa de doctorat en Microbiologia

UAB

Universitat Autònoma de Barcelona
Departament de Genètica i de Microbiologia
Desembre 2023

Ecology of marine aerobic anoxygenic phototrophic bacteria

Ecología de las bacterias aerobias anoxigénicas fototróficas marinas
Ecologia dels bacteris aeròbics anoxigènics fototròfics marins

Carlota Ruiz Gazulla
Departament de Genètica i de Microbiologia
Universitat Autònoma de Barcelona (UAB)

Directors:

Dra. Isabel Ferrera
Centro Oceanográfico de Málaga,
Instituto Español de Oceanografía (IEO-CSIC)

Dr. Josep M. Gasol
Departament de Biologia Marina i Oceanografia,
Institut de Ciències del Mar (ICM-CSIC)

Dra. Olga Sánchez
Departament de Genètica i de Microbiologia
Universitat Autònoma de Barcelona (UAB)

Tutora acadèmica:

Dra. Olga Sánchez
Departament de Genètica i de Microbiologia
Universitat Autònoma de Barcelona (UAB)

Institut de Ciències del Mar

EXCEL·LÈNCIA SEVERO OCHOA

Málaga, 13th December 2023

During the realization of this thesis, Carlota Ruiz Gazulla held a PhD fellowship (PIF 2018/D/LE/CC/2), funded by the 'Universitat Autònoma de Barcelona' and an internship grant (ESTPIF2022-22) funded by Banco Santander. This thesis was carried out at the 'Institut de Ciències del Mar' (ICM-CSIC), at the 'Universitat Autònoma de Barcelona' and at 'Centro Oceanográfico de Málaga' (IEO-CSIC) and supported by the following projects: ECLIPSE (PID2019-110128RB-I00/AEI/10.13039/501100011033), MALASPINA (CSD2008-00077), POSEIDON (CTM2017-84735-R), MIAU (TI2018-101025-B-I00), and MICOLOR (PID2021-125469NB), funded by the Spanish Government.

Cover design by SHOOK Studio (@shookstudio)

‘Life did not take over the world by combat,
but by networking.’

— Lynn Margulis

Contents

SUMMARY

Aerobic anoxygenic phototrophic (AAP) bacteria are heterotrophic organisms that can harvest light to supply their metabolism. They inhabit the euphotic zone of the oceans, where they play an important role in the recycling of the organic matter. The discovery of their presence in the surface oceans more than two decades ago, marked a change of paradigm in the field of marine microbial ecology. Since then, the exploration of AAP bacteria has progressed in parallel with technological advancements, particularly the emergence of high-throughput sequencing (HTS) techniques.

In this thesis, I explore the biogeographical patterns of AAP bacteria communities across the surface global ocean and throughout the water column, I evaluate and improve the molecular tools available for their study, and analyze their cells abundances, pigment distribution, and bioenergetics patterns across the Atlantic Ocean. Using samples collected during the Malaspina Expedition, I explore the diversity and composition of AAP communities across the Atlantic, Indian, and Pacific Oceans, defining their biogeographical patterns and identifying the environmental selection as the main ecological process governing AAP bacteria communities. The assessment of AAP bacteria diversity traditionally relies on amplicon sequencing of the *pufM* gene, with potential challenges arising from primer-related biases. In this thesis, I comprehensively address this issue by evaluating existing primers, designing new ones, and testing them in a set of environmental samples. This approach has enabled the identification of the optimal set of primers for the *pufM* gene amplification, that then I have applied to study the diversity of AAP bacteria in the water column across the South and Mid Atlantic Ocean. The refinement of primers showed that an important fraction of AAP communities are composed by taxonomic groups that have barely been studied, with limited available ecological information. This thesis explores their distribution and habitat preferences, and shows that the composition of AAP assemblages is strongly influenced by their position along the water column relative to the deep chlorophyll maximum. Furthermore, through an analysis of AAP cell abundances and bacteriochlorophyll a concentrations across different depths, I have been able to estimate the solar energy captured by AAP bacteria in the open ocean.

Overall, this thesis provides new insights into the ecology of AAP bacteria in the marine environment, and pave the way for future research in order to advance our understanding and knowledge on their ecology.

RESUMEN

Las bacterias aerobias anoxigénicas fototróficas (AAPs) son organismos heterótrofos que pueden usar la energía solar para suplementar su metabolismo. Habitan la zona eufótica de los océanos, jugando papel importante en el reciclado de la materia orgánica. El descubrimiento de su presencia en la superficie del océano significó un cambio de paradigma en el campo de la ecología microbiana marina. Desde entonces, la exploración de las AAPs ha progresado en paralelo a los avances tecnológicos, y en particular alentado por el desarrollo de las nuevas técnicas de secuenciación masivas.

En esta tesis, se exploran los patrones biogeográficos de las comunidades de AAPs a lo largo de la superficie del océano global y a través de la columna de agua. Se evalúan y mejoran las herramientas moleculares disponibles para su estudio, y se analizan las abundancias celulares y distribución de pigmento a lo largo de océano Atlántico. Usando muestras recogidas durante la Expedición Malaspina, he estudiado la diversidad y la composición de las comunidades de AAPs en los océanos Atlántico, Índico y Pacífico, definiendo sus principales patrones biogeográficos e identificando la selección ambiental como el proceso ecológico más importante influenciando su composición. El análisis de la diversidad de las AAPs se ha basado tradicionalmente en la amplificación del gen *pufM*, con posibles desafíos asociados a sesgos en los cebadores. En esta tesis, abordo de manera integral este tema al evaluar los cebadores existentes, diseñar nuevos y probarlos en un conjunto de muestras ambientales marinas. Este trabajo ha permitido la identificación de los cebadores más adecuados para el gen *pufM*, que posteriormente se han aplicado al estudio de la diversidad de AAPs en la columna de agua a lo largo del océano Atlántico. La mejora de los cebadores muestra que una fracción importante de las comunidades de AAPs está compuesta por grupos taxonómicos que apenas han sido estudiados, y de los que se tiene poca información ecológica. Esta tesis explora su distribución y su hábitat, y muestra que la composición de las comunidades de AAP está fuertemente influenciada por su posición a lo largo de la columna de agua en relación con el máximo de clorofila profunda. Además, el análisis de las abundancias celulares de AAPs y la concentración de bacterioclorofila a en diferentes profundidades, ha permitido estimar la energía solar capturada por las bacterias AAP en el océano abierto.

En conjunto, esta tesis aporta información nueva relativa a las AAPs en el entorno marino y allana el camino para investigaciones futuras con el fin de avanzar en nuestra comprensión y conocimiento de su ecología.

RESUM

Els bacteris aeròbics anoxigènics fototròfics (AAP) són organismes heterotròfics que poden aprofitar la llum per a proveir el seu metabolisme. Habiten a la zona eufòtica dels oceans, on juguen un paper important en el reciclatge de la matèria orgànica. La descoberta de la seva presència en la superfície dels oceans fa més de dues dècades, va marcar un canvi de paradigma en el camp de l'ecologia microbiana marina. Des de llavors, l'exploració de les AAP ha progressat en paral·lel amb els avenços tecnològics, particularment l'aparició de tècniques de seqüenciació d'alt rendiment (HTS).

En aquesta tesi, s'exploren els patrons biogeogràfics de les comunitats d'AAPs a través de la superfície de l'oceà global i a tota la columna d'aigua, avaluant i millorant les eines moleculars disponibles per al seu estudi, i analitzant les seves abundàncies cel·lulars, distribució de pigments i patrons de bioenergètica a través de l'oceà Atlàntic. Utilitzant mostres recollides durant l'Expedició Malaspina, s'explora la diversitat i composició de les comunitats AAP a través dels oceans Atlàntic, Indi i Pacífic, definint els seus patrons biogeogràfics i identificant la selecció ambiental com el principal procés ecològic que governa les comunitats bacterianes AAP. L'avaluació de la diversitat de bacteris AAP es basa tradicionalment en la seqüenciació d'amplicons del gen pufM, amb potencials reptes derivats de biaixos relacionats amb els encebadors. En aquesta tesi, s'aborda de manera integral aquest tema avaluant els encebadors existents, dissenyant-ne de nous i provant-los en un conjunt de mostres ambientals. Aquest enfoc permet identificar els millors encebadors per a l'amplificació del gen pufM, que després s'ha aplicat per estudiar la diversitat d'AAPs a la columna d'aigua a l'oceà Atlàntic. La millora dels encebadors va demostrar que una fracció important de les comunitats AAPs estan compostes per grups taxonòmics que gairebé no han estat estudiats, i dels quals amb prou feines se'n coneix l'ecològica. Aquesta tesi explora la seva distribució i preferències d'hàbitat, i mostra que la composició dels conjunts d'AAPs està fortament influenciada per la seva posició al llarg de la columna d'aigua en relació amb el màxim de clorofil·la profund. A més, a través del anàlisi de les abundàncies de les cèl·lules AAP i les concentracions de bacterioclorofil·la a a través de diferents profunditats, s'ha pogut estimar l'energia solar capturada pels AAP a l'oceà.

En resum, aquesta tesi proporciona nous coneixements sobre les AAPs en el medi marí, i aplana el camí per a futures investigacions amb la finalitat d'avançar en la nostra comprensió i coneixement sobre la seva ecologia.

INTRODUCTION

INTRODUCTION

The marine environment

The marine environment stands as the largest ecosystem on Earth, covering approximately 71% of the planet's surface with an average depth of 3.7 km (Charette & Smith, 2010). This vast ecosystem is comprised of five main ocean basins —the Atlantic Ocean, the Indian Ocean, the Pacific Ocean, the Arctic Ocean, and the Southern (Antarctic) Ocean— interconnected through the conveyor belt (Broecker, 1991), a global water transport system driven by different density gradients due to changes in temperature and salinity, resulting in a global ocean circulation known as the 'thermohaline circulation'.

The action of winds at the ocean surface stir the uppermost layer, homogenizing temperature and salinity, and creating what is termed the oceanic mixed layer, which extends down to 200 m in depth. Beneath this surface realm, the water temperature decreases and the available light is insufficient to sustain photosynthesis (Figure 1A). The differences between surface and deep waters result in vertical stratification acting as a barrier that prevents the mixing of the surface ocean with nutrient-rich deeper waters. However, occasional events of vertical mixing or upwelling supply the upper ocean with new resources (Anderson and Lucas, 2008). The inputs of nutrients foster primary production, particularly in regions characterized by upwelling events, such as equatorial waters, the Southern Ocean, and coastal areas (Anderson et al., 2009). In contrast, the subtropical oceanic gyres exhibit strong stratification that leads to the depletion of nutrients in the surface ocean, converting these areas in oligotrophic deserts (Irwin and Oliver, 2009). The oceans form a continuous ecosystem but the high degree of environmental heterogeneity, such as nutrient availability, temperature variations, and patterns of upwelling, result in diverse habitats that determine the life that resides there.

Marine microbial ecology

Life in the ocean is predominantly microbial, with microorganisms representing two-thirds of the total biomass of marine organisms. Among these, bacteria and archaea –prokaryotes– represent nearly 30% of the overall marine biomass (Bar-On and Milo, 2019). In total, the ocean harbors around 1029 prokaryotic cells (Whitman et al., 1998). Despite this immense abundance, the significance of microbes and the development of marine microbial ecology did not properly start until the 1970s, with the recognition of

the role of prokaryotes in the carbon (C) and nutrient cycling in the ocean. Historically, investigations on marine bacteria relied on either plate counts or direct microscopy counts (Williams and Ducklow, 2019), leading to discrepancies in the estimation of the total number of bacteria present in the ocean. The application of epifluorescence microscopy revealed that values based on cultivated organisms clearly underestimated the actual values by several orders of magnitude (Jannasch and Jones, 1959), a phenomenon later coined as the 'Great Plate Count Anomaly' (Staley and Konopka, 1985). Hence, after the realization of the high abundance of microorganisms in the marine realm, it became evident that they should be integrated somehow into the marine food web (Pomeroy, 1974). Dissolved organic matter (DOM) becomes available through processes such as phytoplankton exudation, sloppy feeding, and cell lysis caused by viruses (Lønborg et al., 2020). Heterotrophic bacteria actively assimilate this organic carbon and convert it into particulate carbon. In turn, bacteria are grazed by heterotrophic flagellates and pigmented protists (Zubkov and Tarran, 2008), channeling organic carbon to higher trophic systems. This flow of carbon from prokaryotes to upper levels of the food web is known as the 'microbial loop' (Azam et al., 1983) (Figure 1B). Most of the processes integral to the microbial loop develop in the illuminated zone of the ocean, the epipelagic, where the sunlight-mediated photosynthesis takes place. Indeed, about half of the Earth's primary production occurs in the ocean (Field, 1998), mainly mediated by cyanobacteria and eukaryotic microalgae (Duarte and Cebrian, 1996). All in all, bacteria account for a large percentage of the oceanic respiration (del Giorgo and Williams, 2005) and are crucial in key biogeochemical transformations, including primary production, mineralization, respiration, and nutrient cycling (Gasol and Kirchman, 2018). Through these processes, they also contribute to export carbon to the deep ocean, thereby playing a vital role in the oceanic ecosystem functioning.

In recent decades, advancements in technology have significantly enhanced our understanding of bacterioplankton ecology. For example, the use of flow cytometry in the marine environment led to the discovery of the cyanobacterium Prochlorococcus (Chisholm et al., 1988) and provided insights into the cellular activity and growth of marine bacteria (del Giorgo and Gasol, 2008). At the same time, the application of molecular tools, like the 16S rRNA gene PCR amplification in environmental DNA, allowed to dig into the diversity of the 'uncultured majority' (Amman et al., 1995, Rappé and Giovannoni, 2003). Later, the advent of high-throughput sequencing (HTS) techniques increased the sequencing depth while lowering costs (Goodwin et al., 2016), revolutionizing the field of microbial ecology and enabling the study of marine diversity at an unprecedented scale.

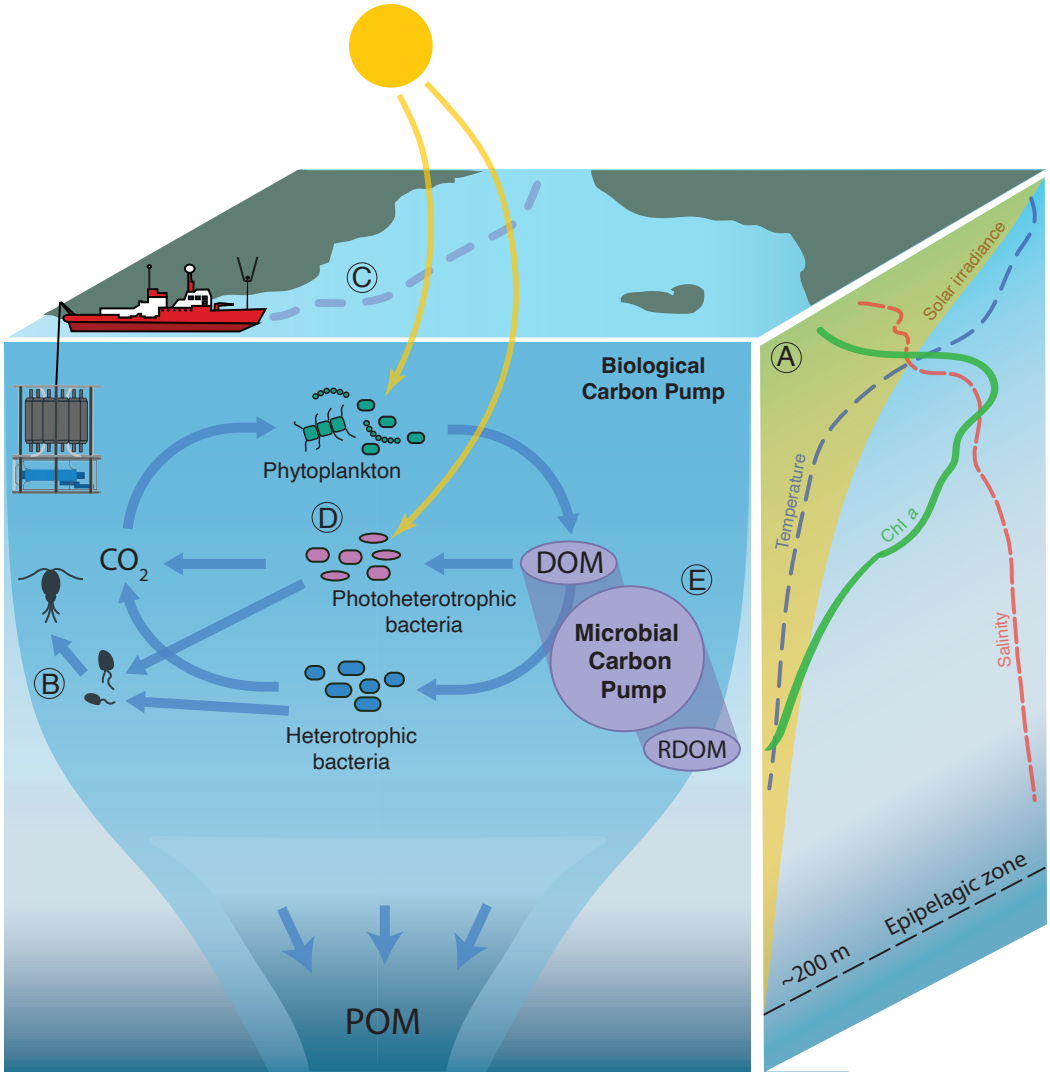

Figure 1. The ocean is a vast and complex three-dimensional space. A) Strong gradients of temperature, salinity, and light, among other factors, shape the structure of the microbial organisms, influencing, for example, the distribution of photosynthetic pigments (chlorophyll *a*, chl *a*) across depth. B) The flow of carbon from prokaryotes to upper levels of the food web is known as the 'microbial loop'. C) Global ocean expeditions are key to unveil the ecological patterns of the ocean microbiota. D) Photoheterotrophic bacteria can harvest light but rely on dissolved organic matter. E) The microbial carbon pump transform labile into recalcitrant organic matter. DOM: dissolved organic matter, POM: particulate organic matter. RDOM: Recalcitrant dissolved organic matter.

In parallel, efforts were invested towards assessing microbial diversity at global scales through initiatives (Figure 1C) such as the pioneering Global Ocean Sampling (GOS) Expedition (Rusch et al. 2007, Venter et al. 2004), or the International Census of Marine Microbes (ICoMM) (Amaral-Zetler et al., 2010), followed by the Tara Oceans (Karsenti et al., 2011) or the Malaspina expeditions (Duarte, 2015). These large-scale surveys, utilizing standardized procedures, facilitated the exploration of the worldwide distribution of marine microorganisms and the identification of the key ecological mechanisms governing planktonic communities (Zinger et al., 2011; de Vargas et al., 2015; Lima-Mendez et al. 2015, Salazar et al., 2015; Sunagawa et al., 2015; Ruiz-González et al., 2019; Ibarbalz et al., 2019, Logares et al., 2020; Obiol et al., 2020).

Photoheterotrophy in the ocean

In the surface ocean, replete with oxygen and lacking reduced nutrients, heterotrophy and autotrophy were considered the predominant prokaryotic metabolisms, and many ecological models designated photoautotrophic microorganisms as primary producers and chemoheterotrophic microorganisms as consumers. This perspective shifted with the groundbreaking discovery that certain heterotrophic bacteria possess the ability to use light energy, documented initially in two independent studies (Béja et al., 2000; Kolber et al., 2000, Box 1). The recognition of the widespread occurrence of photoheterotrophy in the ocean (Box 1) challenged the simplistic trophic level assignments of microbial metabolisms and marked a change of paradigm in our understanding of ocean carbon cycling. These organisms, reliant on dissolved organic matter for growth, have the capacity to harvest light energy (Figure 1D) to supplement different physiological cell functions (Fuhrman et al., 2008). This finding forced to consider direct effects of light on heterotrophic processes, imposing the re-examination of the established models of organic C fluxes in the ocean.

The discovery of the two main groups of marine photoheterotrophic bacteria, proteorhodopsin-containing (PR) bacteria and aerobic anoxygenic phototrophic (AAP) bacteria, occurred nearly simultaneously (Box 1). These groups differ in their light-harvesting mechanisms, ecological roles, phylogeny, and the associated bioenergetic costs and benefits (Kirchman and Hanson, 2013). While the focus of this thesis centers on the ecology of AAP bacteria, it is opportune to expose the contrasting features between these two groups. The simplest mechanism of photoheterotrophy is employed by proteorhodopsin-containing (PR) bacteria. This system involves a protein covalently

bound to a pigment, a retinal (Figure 2A). Upon light absorption, these bacteria generate a proton gradient that helps various cellular functions, such as flagella-based motility (Walter et al., 2007), active transport (Konings 2006), nutrient uptake (Bar-Shalom et al., 2023), or ATP synthesis (Steindler et al., 2011) (Figure 2B). With a single gene coding for proteorhodopsin, the energetic costs of PR bacteria are very low, and horizontal gene transfer processes have been favored. Indeed, PR bacteria are found across diverse lineages of Bacteria and Archaea (Frigaard et al. 2006; de Long and Béjà 2010; Pinhassi et al. 2016), with notable representatives like *Pelagibacter ubique* (Giovannoni et al., 2005), which stands as the most abundant bacterial microorganism in the upper ocean. Metagenomic data from various environments have consistently revealed that PR-based phototrophy is a widespread process across oceanic ecosystems (Finkel et al., 2013; Dubinsky, et al., 2017), potentially absorbing as much light as chlorophyll *a* being therefore major contributors to the solar energy captured in marine environments (Gómez-Consarnau et al., 2019).

On the other side, anoxygenic (non-evolving oxygen) phototrophy based on bacteriochlorophylls is one of the oldest metabolisms on Earth, and comprehends a diverse group of bacteria with different photosynthetic apparatus, pigment composition, and reaction center structures and functioning. This metabolism was associated fundamentally with anaerobic bacteria (Imhoff, 1988), however, the isolation of aerobic species of anoxygenic phototrophs from marine (Harashima, Shiba and Murata 1989; Shiba, Simidu and Taga 1979; Shiba et al. 1991; Nishimura et al. 1994) and freshwater (Fuerst et al. 1993; Suyama et al., 1999; Gich and Overmann 2006) environments, challenged this assumption, revealing their presence in all aquatic systems. AAP bacteria are closely related to purple non-sulfur bacteria (Yurkov and Beatty, 1998a), from which they likely evolved during the Earth's atmospheric oxygenation, adapting to the newly aerobic life. As such, they possess quinone electron acceptors and bacteriochlorophyll *a* (Bchl *a*) based reaction centers, that are only functional under aerobic conditions. Their light-harvesting machinery consists on a diverse set of pigments, such as Bchl *a* and several carotenoids, that vary among species (Yurkov and Csotonyi, 2009), along with proteins conforming the reaction center (Figure 2C). All the genes involved in this function are encoded in the so-called 'photosynthetic gene cluster' (PGC) of around ~45 kb. Within the PGC, there is the photosynthetic unit-fixed (*puf*) operon, with *pufL* and *pufM* genes (Liebetanz et al., 1991) that encode the polypeptides conforming the reaction center (of apparent light (L) and medium (M) molecular weight; Deisenhofer and Michel, 1989). The analyses of this operon in several isolates, compared with the 16S rRNA

gene phylogeny, revealed inconsistencies that could only be explained by horizontal gene transfer processes (Nagashima et al., 1997). Although there have been several hypotheses proposed for the origin and evolution of AAP bacteria (see Discussion), AAP might have experienced a combination of vertical evolution, losses of the photosynthetic gene cluster, and horizontal operon transfer events (Koblížek et al., 2013; Brinkmann et al., 2018). Consequently, marine AAP bacteria form a polyphyletic group of bacteria represented in several families of the Alpha- and the Gammaproteobacteria classes, the last one now including the formerly known Betaproteobacteria.

Box 1. The discovery of photoheterotrophy in the ocean

The isolation of a bacteriochlorophyll *a*-containing bacterium, Citromicrobium bathyomarinum, from the plume waters of a black smoker at the Juan de Fuca Ridge (Yurkov and Beatty 1998b; Yurkov et al. 1999), led to the idea that hydrothermal vents could represent the environment where AAP bacteria originally evolved. Cindy L. Van Dover, a hydrothermal vent expert, shared this idea with Zbigniew Kolber and Paul Falkowski, two oceanographers who had developed a fast repetition rate fluorometer that could detect changes in chlorophyll fluorescence in marine phytoplankton. Enthusiastic about this idea, they modified the equipment to extend its sensitivity to the infrared region and headed to sea to test it with the R/V *Atlantis* and the deep-sea submersible *Alvin*. Their initial attempts to record bacteriochlorophyll signals in hydrothermal vents proved unsuccessful. Unexpectedly, when they tried the device in surface samples, they detected an infrared signal corresponding to Bchl *a*-containing organisms and recorded it along a 1000 km transect. In this pioneering study, they demonstrated that these organisms constituted approximately 10% of total bacteria. Their groundbreaking results were published on 14th September 2000 in the journal *Nature* (Kolber et al., 2000). By coincidence, just a day later, on the 15th of September 2000, Oded Béjà and collaborators published the first evidence of proteorhodopsin-containing bacteria in the surface ocean in the journal *Science* (Béjà et al., 2000). These two landmark reports represented the beginning of a paradigm shift in the field of marine microbial ecology and are regarded as two of the most important findings in microbial oceanography in recent decades.

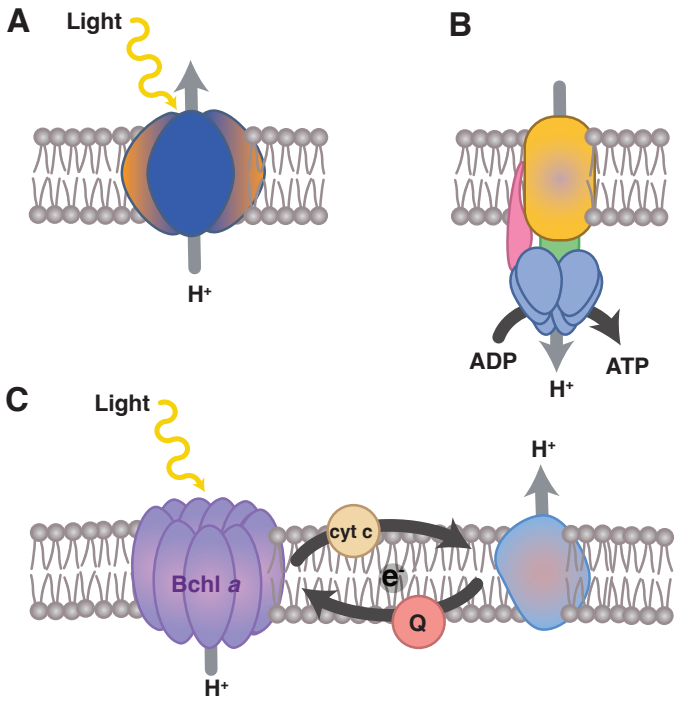

Figure 2. Light harvesting mechanisms in photoheterotrophic bacteria and ATP synthesis generated by the proton motive force. A) Proteorhodopsin-containing (PR) bacteria, B) ATP synthase, C) aerobic anoxygenic phototrophic (AAP) bacteria machinery. Bchl *a*: bacteriochlorophyll *a*; Q: quinone; cyt c: cytochrome c.

Distribution of AAP bacteria in the marine environment

Initially, there was speculation that the advantageous combination of light harvesting ability and heterotrophic metabolism in AAP bacteria would be particularly beneficial in oligotrophic environments where organic carbon is scarce. This led to the hypothesis that AAP bacteria would be more abundant in nutrient-poor, oligotrophic regions (Kolber et al., 2000). However, several independent studies employing various techniques refuted this hypothesis and reported a different trend. Contrary to the initial assumption, these studies indicated that AAP bacteria were more abundant in more productive regions, associated with nutrient-rich waters (e.g. Schwalbach and Fuhrman 2005; Sieracki et al. 2006; Jiao et al. 2007). The debate was fueled by conflicting findings, with different works reporting a large variability in the abundance of AAPs. Based on infrared fast repetition rate fluorescence techniques, Kolber et al., (2001) reported that AAP bacteria constituted ~11% of total bacteria in the North Pacific. Using epifluorescence microscopy and quantitative PCR (qPCR), Schwalbach and Fuhrman (2005) showed that AAP bacteria

only made up around 1-2% of total prokaryotes in the San Pedro Channel in California. Other studies employing qPCR (Du et al., 2006) and metagenomics (Yutin et al., 2007) in various regions, such as the China Sea and in the Atlantic and Pacific Oceans, reported values between 1 and 20% of total bacteria. Since then, epifluorescence microscopy has become the standard methodology to assess the abundance of AAP bacteria, and nowadays there is a common agreement that the abundance of this group generally vary between 0.1 and 10%, as it has been shown in open ocean areas (Sieracki et al., 2006, Cottrell et al., 2006, Jiao et al., 2007, Michelou et al., 2007, Lami et al., 2007), the Baltic Sea (Mašín et al., 2006, Salka et al., 2008), the Mediterranean Sea (Lami et al., 2009, Hojerová et al., 2011, Lamy et al., 2011, Ferrera et al., 2014) or the Arctic Ocean (Boeuf et al., 2013) (Figure 3).

A consistent pattern reported in the marine environment is the positive correlation of AAP bacterial abundance and chlorophyll concentrations (reviewed in Koblížek, 2015). This observation seems to be applicable to contrasting areas such as the Mediterranean Sea (Hojerová et al., 2011), coastal lagoons (Lamy et al., 2011), estuaries (Schwalbach and Fuhrman 2005; Cottrell et al. 2010), and other productive areas (Mašín et al. 2006; Sieracki et al. 2006; Ritchie and Johnson 2012). In these regions, the abundance of AAPs is relatively higher than in more oligotrophic areas, where their relative contribution is often below 2% (Cottrell et al. 2006; Sieracki et al. 2006; Jiao et al. 2007). Light is another variable influencing the distribution of this group of microorganisms. AAP bacteria are known to be typically in the epipelagic zone of the ocean, where the exposure to light may be metabolically beneficial (see next section). Seasonal studies have also reported a positive correlation between AAP cell abundance and day length (Lamy et al., 2011; Ferrera et al., 2014; Auladell et al., 2019). These studies, conducted in the surface Mediterranean Sea, reported higher abundance of AAP bacteria during spring and summer, when days are longer and also when typical phytoplankton spring blooms occur (Nunes et al., 2017). A similar finding of high numbers of AAP cells associated to diatom and dinoflagellate blooms was reported in the East China Sea (Chen et al., 2011). In fact, phytoplankton has been proposed as a potential driver for controlling AAP populations, on one side due to the already explained preference of AAP bacteria for more eutrophic waters, but also due to the similar distribution of chlorophyll and Bchl *a* (Gómez-Consarnau et al., 2019), or the fact that several AAP species have been isolated in association with dinoflagellates (e.g. Biebl et al. 2006). To a lesser extent, the distribution of AAP bacteria has also been observed to be influenced by factors such as temperature or nutrient availability (Mašín et al. 2006; Hojerová et al., 2011; Ferrera et al., 2014).

Overall, although we possess a reasonably good understanding of the distribution of AAP bacteria across the ocean surface, there is a notable gap analyzing their distribution throughout the water column in the epipelagic zone. A thorough investigation into the variations of AAP bacteria abundance across pronounced gradients of light irradiance and chlorophyll concentration within the epipelagic would improve our understanding of the ecology of this functional group.

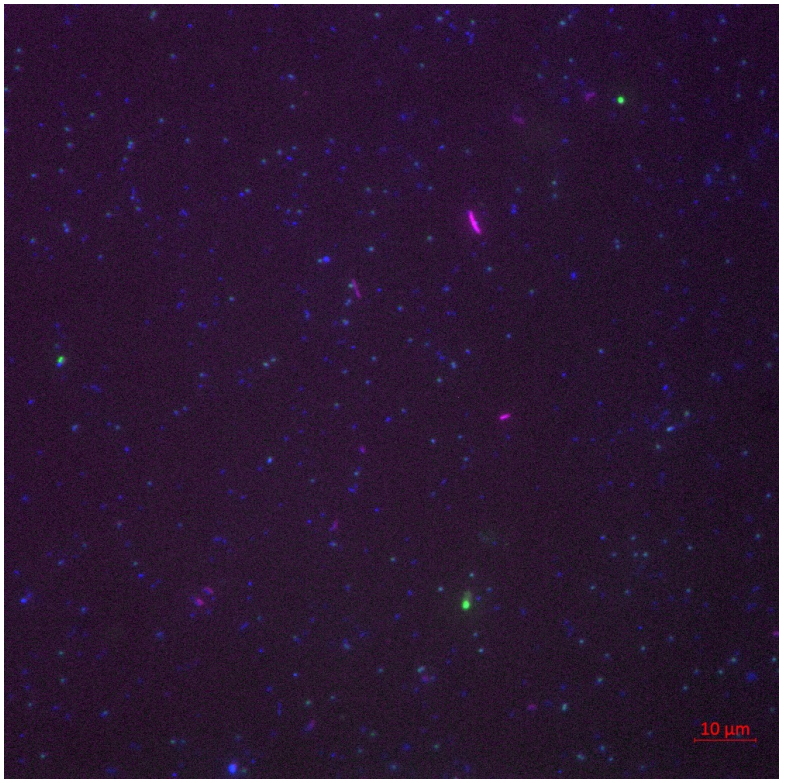

Figure 3. Digital image of prokaryotic cells collected in the equatorial Atlantic Ocean, at 30 m depth, during the Poseidon Expedition. The image is a composition of three images that show the total prokaryotic population stained with DAPI (blue), AAPs bacteria under infrared fluorescence (fuchsia), and cyanobacteria under blue excitation (green).

The ecological role of AAP bacteria

The incorporation of AAP bacteria into the marine microbiota context raises a number of questions regarding their ecological role and their contribution to biogeochemical cycles. Compared to the bulk heterotrophic bacterioplankton, AAP bacteria have large sizes (Mašín et al., 2006; Sieracki et al., 2006; Kirchman et al. 2014; Stegman et al., 2014) (Figure 3), which make them more susceptible to protist grazing. Indeed, manipulation

experiments have proven that grazing is an important factor controlling AAP populations (Koblížek et al., 2007; Ferrera et al., 2011; Ferrera et al., 2017). To cope with this high predation pressure, AAP bacteria display high growth rates (Koblížek et al., 2005; Koblížek et al., 2007; Hojerová et al., 2011; Cepáková et al., 2016; Ferrera et al., 2017; Fecskeová et al., 2021), surpassing those estimated for other bacterial groups by two or threefold (Liu et al., 2010, Ferrera et al., 2011). Their active metabolism is also evidenced by experiments in which AAP incorporate two times more ^{3}H-leucine than average bacteria (Stegman et al., 2014), or by the high expression levels of *puf* and *bchl* transcripts (for Bchl *a* synthesis) at the ALOHA station in the Pacific Ocean (Frias-Lopez et al. 2008). These findings show that AAP bacteria represent a dynamic part of marine microbial communities.

The ecological relevance of AAP bacteria is further underscored by their active metabolism. Although not constituting a large fraction of bacterioplankton, AAP bacteria consume large quantities of dissolved organic carbon (DOC) to sustain their high metabolic activity (Koblížek et al., 2003; Biebl and Wagner-Döbler 2006). Under the high grazing pressure they are subjected to, AAP bacteria contribute to the incorporation of organic carbon into higher trophic levels of the food web (García-Chaves et al., 2015; Cepáková et al., 2016), being this flow of carbon from AAP bacteria to grazers likely more substantial than their abundance alone would suggest (Koblížek, 2015).

Regarding their ecological role, an important finding was the realization that the consumption of carbon by AAP bacteria was intricately linked to their phototrophic activity. Traditionally, microbial respiration was regarded as independent of light (del Giorgio and Duarte, 2002). However, it was soon proved that AAP bacteria could cover some of their energy requirements through light harvesting. Laboratory experiments demonstrated that exposure to light enhanced the growth efficiency of AAP bacteria (Biebl and Wagner-Döbler 2006; Hauruseu and Koblížek 2012), facilitating the accumulation of biomass in terms of carbon, reducing their respiration rate (Koblížek et al., 2010; Soora et al., 2013), and increasing the efficiency of carbon metabolism (Piwosz et al., 2018). All in all, their efficient phototrophic metabolism coupled with their high assimilatory rates of organic carbon, suggest that AAP bacteria could have an impact on the flow of carbon in the ocean (Figure 1E), on the one hand by processing large amounts of DOM, and on the other by substantially reducing respiration and the release of CO_2 in the light. Their ability to process diverse types of organic matter and to form polymers (Xiao and Jiao, 2011) suggest that they might have a significant role in the microbial carbon pump (Jiao and

Azam, 2011), although the link between AAP bacteria and the production of recalcitrant DOM have not been explored. While effects of AAP bacteria on the global carbon fluxes have been demonstrated in freshwater lakes (Piwosz et al., 2022), there is currently limited information regarding their impact in marine ecosystems. Further exploration of AAP bacterial metabolism in marine environments could provide valuable insights into the broader implications of AAP bacteria in carbon dynamics.

Diversity of AAP marine communities

The isolation of the first marine AAP species identified the genera *Erythrobacter* (Shiba and Shimidu, 1982) and *Roseobacter* (Shiba, 1991), in the order Sphingomonadales and Rhodobacterales, respectively (class Alphaproteobacteria) as members of this microbial guild. While early studies reliant on culture-dependent techniques continued to unveil novel AAP species within the Alphaproteobacteria class (e.g. Allgaier et al., 2003), the advent of molecular techniques transformed our ability to investigate the diversity of AAP communities on a broader scale. In this sense, the evolution of our understanding of AAP diversity has been intricately linked with the progress of the molecular methodologies. The possibility of targeting a specific functional gene from the environment to study the ecology of a certain functional group led to the use of the *pufM* gene as the genetic marker for AAP bacteria.

The first primers proposed to study AAP bacterial communities were designed by Nagashima et al. (1997) and Achenbach et al. (2001), using sequences from cultured bacteria. Derived from these, Béjà et al. (2002) designed a new set of primers and using bacterial artificial chromosome (BAC) libraries, unveiled the existence of various uncultured marine AAP bacteria tentatively belonging to the Beta- and Gammaproteobacteria (the Betaproteobacteria class is now considered part of the Gammaproteobacteria). This was later confirmed by Cho et al. (2007) and Fuchs et al. (2007), who isolated several strains belonging to the NOR5/OM6 cluster, a common clade in the marine environment which significantly contributes to bacterioplankton, particularly during summer. Among these strains, the gammaproteobacterium KT71 was identified, characterized, and named *Congregibacter litoralis* (Fuchs et al., 2007). In spite of the widespread distribution of this clade, clone libraries created from different oceanic regions showed a predominant presence of *Roseobacter*-like species in AAP communities. In contrast, Gammaproteobacteria represented only a small fraction of the populations in following studies based on BAC clone libraries (Oz et al., 2005; Yutin et al., 2005).

While efforts to analyze and describe the diversity of AAP assemblages were underway, little information was available on how these different groups were distributed across different oceanic regions and environments. The initial exploration of variations in AAP assemblages across distinct oceanic regions was based on the metagenomic data collected during the Global Ocean Sampling (GOS) Expedition (Rusch et al., 2007). In their study, Yutin and coauthors (2007) analyzed the phylogeny of the *pufM* gene and the structure of the *puf* operon, and defined 12 phylogroups of AAP bacteria (designated A to L). Some of these phylogroups represented novel groups of AAP bacteria that appeared to be abundant in the open ocean according to the presence of the gene, but lacked cultured representatives. Other, in turn, were associated to previously reported genera within the Alpha- and Gammaproteobacteria. Since then, the nomenclature based on these phylogroups has been commonly employed in studies on AAP diversity.

Despite the established use in recent years of metagenomics for studying bacterial communities, nearly all studies targeting AAP bacteria have been based on the PCR-amplification of the *pufM* gene due to the high costs of using metagenomics when targeting a group characterized by low relative abundances, like the AAP bacteria. Employing a combination of primers from two previous studies (Béjà et al., 2002 and Yutin et al., 2005), Lehours et al. (2010) described for the first time the prevalence of gammaproteobacterial AAPs in the Mediterranean Sea. This finding was followed by subsequent works confirming the dominance of this class in the Mediterranean Sea (Jeanthon et al., 2011; Ferrera et al., 2014; Auladell et al., 2019), as well as in different open ocean areas (Jiao et al., 2007; Lehours et al., 2010), or in Australian coastal waters (Bibiloni-Isaksson et al., 2016).

The above-mentioned studies, along with numerous others, have covered a wide range of environmental variability, spanning from the cold waters of the Arctic Ocean (Boeuf et al., 2013, Lehours and Jeanthon 2015) to the temperate Mediterranean Sea (Lehours et al., 2010, Jeanthon et al., 2011, Ferrera et al., 2014, Auladell et al., 2019), or tropical and subtropical waters (Jiao et al., 2007, Yutin et al., 2007, Bibiloni-Isaksson et al., 2016). While these studies have provided valuable insights into the ecological distribution of AAP assemblages across different environments (Figure 4), a comprehensive global view is currently lacking. Global ocean studies have proven very useful in addressing questions related to biogeographical patters of the marine microbiota, for example, explaining the ecological processes governing bacterioplankton community assembly (Logares et al., 2020), or defining ubiquitous and rare species within marine bacteria (Ruiz-González et

al., 2019). In the case of AAP bacteria, our knowledge on their diversity is still fragmented, with some areas extensively studied (e.g. the Mediterranean Sea), while others remaining unexplored (Figure 4). Besides, the improvement of sequencing techniques over the years, coupled to the appearance of threshold-free algorithms for amplicon sequence variants (ASVs) analysis, which surpass the traditional clustering of sequences based on similarity cutoffs into operational taxonomic units (OTUs), hampers a proper comparison of the existing datasets.

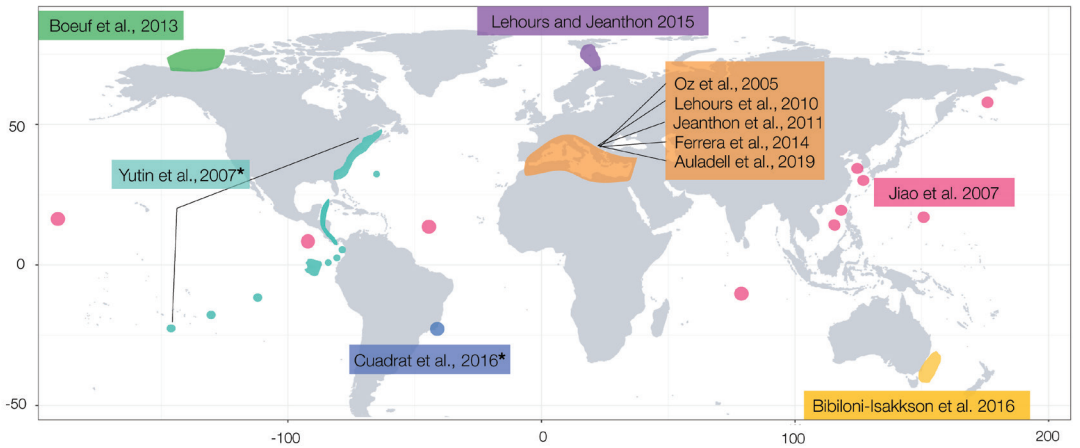

Figure 4. Sampling locations in the key studies exploring the distribution of AAP communities in the marine environment. These studies exhibit a wide geographic range, including the Arctic Ocean, the temperate WMediterranean Sea, the tropical and subtropical oceans, as well as the waters surrounding Australia.

Finally, while the current molecular techniques have significantly advanced our understanding of the diversity of AAP assemblages in the marine ecosystem, a discussion about the suitability of the commonly employed primers is still alive, as the few studies based on metagenomics (Yutin et al., 2007 or Cuadrat et al., 2016) show community compositions that differ from those based on PCR approaches. The main differences refer to the high abundance of Gammaproteobacteria, that is frequently reported in amplicon-based studies, as opposed to the prevalence of certain phylogroups with no cultured representatives reported in metagenomics studies. The primer pair consistently used in all amplicon-based studies since their introduction in Lehours et al. (2010) could suffer from amplification biases. While a recent comparison of the seasonal trends based on both methodologies showed that the patterns were conserved (Auladell et al., 2019), this study also identified some groups not appearing in the PCR approaches that deserved a careful revision of the primers used and, ultimately, of the AAP community composition image that these techniques generate.

References:

Achenbach LA, Carey J, & Madigan MT (2001) Photosynthetic and phylogenetic primers for detection of anoxygenic phototrophs in natural environments. *Appl Environ Microbiol* 67:2922–2926. https://doi.org/10.1128/AEM.67.7.2922-2926.2001

Allgaier, M., H. Uphoff, Felske, A., I. & Wagner-Dobler. (2003). Aerobic anoxygenic photosynthesis in *Roseobacter* clade bacteria from diverse marine habitats. *Appl. Environ. Microbiol.* 69:5051–5059

Amaral-Zettler, L., Artigas, L.F., Baross, J., Bharathi, L., Boetius, A., Chandramohan, D., ... & Sogin, M. (2010). A global census of marine microbes. *Life in the world's Oceans: Diversity, Distribution and Abundance*, 223-245.

Anderson, T. R. & M. I. Lucas (2008) Upwelling Ecosystems. In Encyclopedia of Ecology, S. E. Jørgensen and B. D. Fath, eds., Pp. 3651–3661. Oxford: Academic Press.

Anderson R.F. et al. (2009) Wind-Driven Upwelling in the Southern Ocean and the Deglacial Rise in Atmospheric CO_2. *Science* 323,1443-1448 DOI:10.1126/science.1167441

Auladell, A., Sánchez, P., Sánchez, O. et al. (2019) Long-term seasonal and interannual variability of marine aerobic anoxygenic photoheterotrophic bacteria. *ISME* 13, 1975–1987. https://doi.org/10.1038/s41396-019-0401-4

Azam, F., Fenchel, T., Field, J., Gray, J., Meyer-Reil, L., & Thingstad, F. (1983). The Ecological Role of Water-Column Microbes in the Sea. *Marine Ecology Progress Series*, 10, 257–263. https://doi.org/10.3354/meps010257

Bar-On, Y.M. & Milo, R. (2019) The biomass composition of the oceans: a blueprint of our blueplanet. Cell 179: 1451–1454.

Bar-Shalom, R., Rozenberg, A., Lahyani, M. et al. (2023) Rhodopsin-mediated nutrient uptake by cultivated photoheterotrophic *Verrucomicrobiota*. *ISME* J 17, 1063–1073 https://doi.org/10.1038/s41396-023-01412-1

Béjà, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A., Nguyen, L.P., et al. (2000) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* 289: 1902–1906.

Béjà O, Suzuki MT, Heidelberg JF et al (2002) Unsuspected diversity among marine aerobic anoxygenic phototrophs. *Nature* 415:630–633

Bibiloni-Isaksson, J., Seymour, J.R., Ingleton, T., van de Kamp, J., Bodrossy, L., & Brown, M.V. (2016) Spatial and temporal variability of aerobic anoxygenic photoheterotrophic bacteria along the East coast of Australia. *Environ Microbiol* 18: 4485–4500. https://doi.org/10.1111/1462-2920.13436

Biebl H, Tindall BJ,Wagner-Döbler I. (2006) *Hoeflea phototrophica* sp. nov., a novel marine aerobic alphaproteobacterium that forms bacteriochlorophyll *a. Int J Syst Evol Micr*. 56:821–6.

Biebl H, & Wagner-Döbler I. (2006) Growth and bacteriochlorophyll a formation in taxonomically diverse aerobic anoxygenic phototrophic bacteria in chemostat culture: influence of light regiment and starvation. *Proc Biochem*;41: 2153–9.

Boeuf D, Cottrell MT, Kirchman DL, et al. (2013) Summer community structure of aerobic anoxygenic phototrophic bacteria in the western Arctic Ocean. *FEMS Microbiol Ecol* 85:417–32.

Brinkmann, H., Göker, M., Koblížek, M. et al. (2018) Horizontal operon transfer, plasmids, and the evolution of photosynthesis in *Rhodobacteraceae*. *ISME* J 12, 1994–2010. https://doi.org/10.1038/s41396-018-0150-9

Broecker, W. S. (1991). The Great Ocean Conveyor. *Oceanography*, 4(2), 79–89. http://www.jstor.org/stable/43924572

Cepáková, Z., Hrouzek, P., Žišková, E., Nuyanzina-Boldareva, E., Šorf, M., Kozlíková-Zapomělová, E., Salka, I., Grossart, H.-P. and Koblížek, M. (2016), High turnover rates of aerobic anoxygenic phototrophs in European freshwater lakes. *Environ Microbiol*, 18: 5063-5071. https://doi.org/10.1111/1462-2920.13475

Charette, M. A., & Smith, W. H. F. (2010). The Volume of Earth's Ocean. *Oceanography*, 23(2), 112–114.

Chen Y, Zhang Y, & Jiao N. (2011) Responses of aerobic anoxygenic phototrophic bacteria to algal blooms in the East China Sea. *Hydrobiologia* 661:435–43.

Cho J-C, Stapels MD, Morris RM, et al. (2007) Polyphyletic photosynthetic reaction centre genes in oligotrophic marine Gammaproteobacteria. *Environ Microbiol* 2007 9:1456–63.

Cottrell MT, Mannino A, & Kirchman DL. (2006) Aerobic anoxygenic phototrophic bacteria in the Mid-Atlantic Bight and the North Pacific Gyre. *Appl Environ Microb* 72:557–64.

Cottrell MT, Ras J, & Kirchman DL. (2010) Bacteriochlorophyll and community structure of aerobic anoxygenic phototrophic bacteria in a particle-rich estuary. *ISME* 4:945–54.

Cuadrat, R., Ferrera, I., Grossart, H., and D avila, A. (2016) Picoplankton bloom in global south? Ahigh fraction of aerobic & phototrophic bacteria in metagenomes from a coastal bay (Arraial do Cabo-Brazil). *Omi A J Integr Biol* 20: 76–87. https://doi.org/10.1089/omi.2015.0142

Deisenhofer, J., & Michel, H. (1989). The Photosynthetic Reaction Center from the Purple Bacterium *Rhodopseudomonas viridis*. *Science*, 245(4925), 1463–1473. https://doi.org/10.1126/science.245.4925.1463

del Giorgio, P.A. & Duarte, C.M. (2002) Respiration in the open ocean. *Nature* 420: 379–384.

del Giorgio, P.A. & Williams, P. (2005). Respiration in aquatic ecosystems. OUP Oxford.

del Giorgio, P.A. & Gasol. J.M. (2008). Physiological structure and single-cell activity in marine

bacterioplankton. In D. L. Kirchman (ed.), *Microbial Ecology of the Oceans* (2nd ed., pp. 243–298). New York: Wiley-Liss.

DeLong EF, & Béjà O (2010) The Light-Driven Proton Pump Proteorhodopsin Enhances Bacterial Survival during Tough Times. *PLoS Biol* 8(4): e1000359. https://doi.org/10.1371/journal.pbio.1000359

Du H, Jiao N, Hu Y, et al. (2006) Real-time PCR for quantification of aerobic anoxygenic phototrophic bacteria based on *pufM* gene in marine environment. *J Exp Mar Biol Ecol* 329:113–21.

Duarte, C. M., & J. Cebrián. (1996) The fate of marine autotrophic production. *Limnol. Oceanogr.* 41: 1758–1766.

Dubinsky, V., Haber, M. Burgsdorf, I., Saurav, K. Lehahn, Y., Malik, A., Sher, D., Aharonovich, D., & Steindler, L. (2017) Metagenomic analysis reveals unusually high incidence of proteorhodopsin genes in the ultraoligotrophic Eastern Mediterranean Sea. *Environ. Microbiol.* 19, 1077–1090.

Fecskeová LK, Piwosz K, Šantic D, Šestanovic S, Tomaš AV, Hanusová M, Šolic M, Koblížek M. (2021) Lineage-specific growth curves document large differences in response of individual groups of marine bacteria to the top-down and bottom-up controls. *mSystems* 6:e00934-21. https://doi.org/10.1128/ mSystems.00934-21.

Ferrera I, Gasol JM, Sebastian M, Hojerová E, Koblížek M. (2011) Comparison of growth rates of aerobic anoxygenic phototrophic bacteria and other bacterioplankton groups in coastal Mediterranean waters. *Appl Environ Microbiol* 77:7451–7458. https://doi.org/10.1128/AEM.00208-11.

Ferrera I, Borrego CM, Salazar G, et al. (2014) Marked seasonality of aerobic anoxygenic phototrophic bacteria in the coastal NW Mediterranean Sea as revealed by cell abundance, pigment concentration and pyrosequencing of pufM gene. *Environ Microbiol* 16:2953–65.

Ferrera I, Sánchez O, Kolárová E, Koblížek M, Gasol JM. (2017) Light enhances the growth rates of natural populations of aerobic anoxygenic phototrophic bacteria. *ISME* 11:2391–2393. https://doi.org/10.1038/ismej.2017.79.

Field, C.B., Behrenfeld, M.J., Randerson, J.T., & Falkowski, P. (1998) Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* 281: 237–240.

Frigaard, N.U., Martinez, A., Mincer, T.J., & DeLong, E.F. (2006) Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and Archaea. *Nature* 439: 847–850.

Finkel, O. Béjà, O. & Belkin, S. (2013) Global abundance of microbial rhodopsins. *ISME* 7, 448–45.

Frias-Lopez J, Shi Y, Tyson GW, et al. (2008) Microbial community gene expression in ocean surface waters. *P Natl Acad Sci USA* 105:3805–10.

Fuchs BM, Spring S, Teeling H, et al. (2007) Characterization of a marine gammaproteobacterium

capable of aerobic anoxygenic photosynthesis. *P Natl Acad Sci USA* 104: 2891–6.

Fuerst JA, Hawkins JA, Holmes A, et al. (1993) *Porphyrobacter neustonensis* gen. nov., sp. nov., an aerobic bacteriochlorophyll synthesizing budding bacterium from fresh water. *Int J Syst Bacteriol* 1993;43:125–34.

Fuhrman, J., Schwalbach, M. & Stingl, U. (2008) Proteorhodopsins: an array of physiological roles?. *Nat Rev Microbiol* 6, 488–494. https://doi.org/10.1038/nrmicro1893

Garcia-Chaves, M.C., Cottrell, M.T., Kirchman, D.L., Derry, A.M., Bogard, M.J., & del Giorgio, P.A. (2015) Major contribution of both zooplankton and protists to the top-down regulation of freshwater aerobic anoxygenic phototrophic bacteria. *Aquat Microb Ecol* 76: 71–83.

Gasol, JM., & Kirchman, DL. (2018) *Microbial Ecology of the Oceans*, Third Edition. John Wiley & Sons. 528 pp.ISBN-13: 978-1119107187

Gich F, & Overmann J. (2006) *Sandarakinorhabdus limnophila* gen. nov., sp. nov., a novel bacteriochlorophyll a-containing, obligately aerobic bacterium isolated from freshwater lakes. *Int J Syst Evol Micr* 56:847–54.

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. *Science* 309(5738), 1242–1245. https://doi.org/10.1126/science.1114057

Gómez-Consarnau, L.,Raven, JA., Levine, NM., Cutter, LS., Wang, D., Seegers, B., Arístegui, J., Fuhrman, JA:, Gasol, JM., & Sañudo-Wilhelmy, SA. (2019) Microbial rhodopsins are major contributors to the solar energy captured in the sea. *Science Advances* 10.1126/sciadv. aaw8855, 5, 8.

Harashima K, Shiba T, Murata N. (1989) Aerobic Photosynthetic Bacteria. Tokyo: Japan Scientific Societies Press.

Hauruseu D, & Koblížek M. (2012) The influence of light on carbon utilization in aerobic anoxygenic phototrophs. *Appl Environ Microb* 2012;78:7414–9.

Hojerová E, Mašín M, Brunet C, et al. (2011) Distribution and growth of aerobic anoxygenic phototrophs in the Mediterranean Sea. *Environ Microbiol*;13:2717–25.

Imhoff, J. F. (1988) Anoxygenic phototrophic bacteria, p. 207–240. In B. Austin (ed.), *Methods in aquatic bacteriology*. John Wiley & Sons, Inc., New York, N.Y.

Irwin, A. J., & M. J. Oliver. 2009. Are ocean deserts getting larger? *Geophys. Res. Lett.* 36. doi: 10.1029/2009gl039883

Jannasch HW & Jones A (1959) Bacterial populations in sea water as determined by different methods of enumeration. *Limnol Oceanogr* 4:128–139

Jeanthon, C., Boeuf, D., Dahan, O., le Gall, F., Garczarek, L., Bendif, E.M., & Lehours, A.C. (2011) Diversity of cultivated and metabolically active aerobic anoxygenic phototrophic bacteria along an oligotrophic gradient in the Mediterranean Sea. *Biogeosciences* 8: 1955–1970. https://doi.org/10.5194/bg-8-1955-2011

Jiao, N., Y. Zhang, Y. Zeng, N. Hong, R. Liu, F. Chen, & P. Wang. (2007) Distinct distribution pattern of abundance and diversity of aerobic anoxygenic phototrophic bacteria in the global ocean. *Environ Microbiol* 9: 3091–3099. doi:10.1111/j.1462-601 2920.2007.01419.x

Kirchman, D. L., & Hanson T. E. (2013) Bioenergetics of photoheterotrophic bacteria in the oceans. *Environ Microbiol Rep* 5: 188–199. doi:10.1111/j.1758-2229.2012.00367.x

Kirchman DL, Stegman MR, Nikrad MP, et al. (2014) Abundance, size, and activity of aerobic anoxygenic phototrophic bacteria in coastal waters of the West Antarctic Peninsula. *Aquat Microb Ecol* 73:41–9.

Koblížek M, Béjà O, Bidigare RR, et al. (2003) Isolation and characterization of Erythrobacter sp. strains from the upper ocean. *Arch Microbiol* 180:327–38.

Koblížek, M., Stoń-Egiert, J., Sagan, S. & Kolber, Z.S. (2005) Diel changes in bacteriochlorophyll a concentration suggest rapid bacterioplankton cycling in the Baltic Sea. *FEMS Microbiology Ecology* 51: 353-361. https://doi.org/10.1016/j.femsec.2004.09.016

Koblížek M, Mašín M, Ras J, Poulton AJ, Prášil O. (2007) Rapid growth rates of aerobic anoxygenic phototrophs in the ocean. *Environ Microbiol* 9: 2401–2406. https://doi.org/10.1111/j.1462-2920.2007.01354.x

Koblížek, M, Mlčoušková, J., Kolber Z, et al. (2010) On the photosynthetic properties of marine bacterium COL2P belonging to *Roseobacter* clade. *Arch Microbiol* 192:41–9.

Koblížek M, Zeng Y, Horák A, Oborník M. (2013) Regressive evolution of photosynthesis in the *Roseobacter* clade. *Adv Bot Res.* 66:385–405.

Koblížek, M. (2015) Ecology of aerobic anoxygenic phototrophs in aquatic environments. *FEMS Microbiol Rev* 39: 854–870.

Kolber, Z.S., Van Dover, C.L., Niederman, R. and Falkowski, P.G. (2000) Bacterial photosynthesis in surface waters of the open ocean. *Nature* 407: 177–179.

Kolber, Z. S., F. G. Plumley, A. S. Lang, and others. (2001) Contribution of aerobic photoheterotrophic bacteria to the carbon cycle in the ocean. *Science* (1979) 292:2492–2495. doi:10.1126/science.1059707

Konings, W.N. (2006) Microbial transport: adaptations to natural environments. *Antonie Van Leeuwenhoek Int. J. Gen. Mol. Microbio*l., 90, 325–342.

Lami R, Cottrell MT, Ras J, et al. (2007) High abundances of aerobic anoxygenic photosynthetic bacteria in the South Pacific Ocean. *Appl Environ Microb* 73:4198–205.

Lami R, Cûperová Z, Ras J, et al. (2009) Distribution of free-living versus particle-attached aerobic anoxygenic phototrophic bacteria in marine environments. *Aquat Microbial Ecol* 55:31–8.

Lamy D, Jeanthon C, Cottrell MT, et al. (2011) Ecology of aerobic anoxygenic phototrophic bacteria along an oligotrophic gradient in the Mediterranean Sea. *Biogeosciences* 8:973–85.

Lehours, A., & Jeanthon, C. (2015) The hydrological context determines the beta-diversity of aerobic anoxygenic phototrophic bacteria in European Arctic seas but does not favor endemism. *Front Microbiol* 6: 638. https://doi.org/10.3389/fmicb.2015.00638

Lehours, A.C., Cottrell, M.T., Dahan, O., Kirchman, D.L., & Jeanthon, C. (2010) Summer distribution and diversity of aerobic anoxygenic phototrophic bacteria in the Mediterranean Sea in relation to environmental variables. *FEMS Microbiol Ecol* 74: 397–409. https://doi.org/10.1111/j.1574-6941.2010.00954.x

Lehours, A.C., Enault, F., Boeuf, D., & Jeanthon, C. (2018) Biogeographic patterns of aerobic anoxygenic phototrophic bacteria reveal an ecological consistency of phylogenetic clades in different oceanic biomes. *Sci Rep* 8: 1–10. https://doi.org/10.1038/s41598-018-22413-7

Liebetanz, R., Hornberger, U., & Drews, G. (1991). Organization of the genes coding for the reaction-centre L and M subunits and B870 antenna polypeptides alpha and beta from the aerobic photosynthetic bacterium *Erythrobacter species* OCH114. *Molec Microbiol* 5(6), 1459–1468. https://doi.org/10.1111/j.1365-2958.1991.tb00792.x

Liu R, Zhang Y, Jiao N. (2010) Diel variations in frequency of dividing cells and abundance of aerobic anoxygenic phototrophic bacteria in a coral reef system of the South China Sea. *Aquat Microb Ecol* 2010;58:303–10.

Logares, R., Deutschmann, I.M., Junger, P.C., Giner, C.R., Krabberød, A.K., Schmidt, T.S.B., et al. (2020) Disentangling the mechanisms shaping the surface ocean microbiota. *Microbiome* 8: 1–17.

Lønborg, C., Carreira, C., Jickells, T., & Álvarez-Salgado, X. A. (2020). Impacts of Global Change on Ocean Dissolved Organic Carbon (DOC) Cycling. *Frontiers in Marine Science*, 7. https://www.frontiersin.org/article/10.3389/fmars.2020.00466

Mašín M, Zdun A, Stoń-Egiert J, et al. (2006) Seasonal changes and diversity of aerobic anoxygenic phototrophs in the Baltic Sea. *Aquat Microbial Ecol* 45:247–54.

Michelou VK, Cottrell MT, Kirchman DL. (2007) Light-stimulated organic matter assimilation by cyanobacteria and other microbes in the North Atlantic Ocean. *Appl Environ Microb*73:5539–46.

Nagashima KVP, Hiraishi A, Shimada K, Matsuura K (1997) Horizontal transfer of genes coding for the photosynthetic reaction centers of purple bacteria. *J Mol Evol* 45:131–136. https://doi.org/ 10.1007/ PL000 06212

Nishimura Y, Muroga Y, Saito S, et al. (1994) DNA relatedness and chemotaxonomic feature of aerobic bacteriochlorophyll-containing bacteria isolated from coast of Australia. *J Gen Appl Microbiol* 1994;40:287–96.

Nunes S, Latasa M, Gasol JM, & Estrada M. (2017) Seasonal and interannual variability of phytoplankton community structure in a Mediterranean coastal site. *Mar Ecol Prog Ser.*;592:57–75.

Oz, A., Sabehi, G., Koblízek, M., Massana, R., & Béjà., O. (2005) *Roseobacter*-Like Bacteria in Red and Mediterranean Sea Aerobic Anoxygenic Photosynthetic Populations. *Applied and Environmental Microbiology.* 71, 1. 344-353. doi:10.1128/AEM.71.1.344–353.2005

Pinhassi J, Delong EF, Béjà O, González JM, & Pedrós-Alió C. (2016) Marine bacterial and archaeal ion-pumping rhodopsins: genetic diversity, physiology, and ecology. *Microbiol Mol Biol Rev* 80:929 –954. doi:10.1128/MMBR.00003-16.

Piwosz, K., D. Kaftan, J. Dean, J. Šetlík, & M. Koblížek. (2018) Nonlinear effect of irradiance on photoheterotrophic activity and growth of the aerobic anoxygenic phototrophic bacterium Dinoroseobacter shibae. *Environ Microbiol* 20: 724–733. doi:10.1111/1462-2920.14003

Piwosz, K., Villena-Alemany, C. & Mujakić, I. (2022) Photoheterotrophy by aerobic anoxygenic bacteria modulates carbon fluxes in a freshwater lake. *ISME* 16, 1046–1054. https://doi.org/10.1038/s41396-021-01142-2

Pomeroy, L. R. (1974). The Ocean's Food Web, A Changing Paradigm. BioScience, 24(9), 499–504. https://doi.org/10.2307/1296885

Ruiz-González, C., Logares, R., Sebastián, M., Mestre, M., Rodríguez-Martínez, R., Galí, M., et al. (2019) Higher contribution of globally rare bacterial taxa reflects environmental transitions across the surface ocean. *Mol Ecol* 28: 1930–1945. https://doi.org/10.1111/mec.15026

Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, et al. (2007) The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biol* 5(3): e77. https://doi.org/10.1371/journal.pbio.0050077

Salka I, Moulisová V, Koblízek M, et al. (2008) Abundance, depth distribution, and composition of aerobic bacteriochlorophyll aproducing bacteria in four deeps of the central Baltic Sea. *Appl Environ Microb* 74:4398–404.

Shiba T. (1991) *Roseobacter litoralis* gen. nov., sp. nov. and *Roseobacter denitrificans* sp. nov., aerobic pink-pigmented bacteria which contain bacteriochlorophyll *a. Syst Appl Microbiol* 14:140–5

Shiba T, Simidu U, Taga N. (1979) Another aerobic bacterium which contains bacteriochlorophyll a. *B Jpn Soc Sci Fish*, 45:801.

Shiba T, Simidu U. (1982) *Erythrobacter longus* gen. nov., sp. nov., an aerobic bacterium which contains bacteriochlorophyll a. *Int J Syst Bacteriol* 32:211–7.

Schwalbach MS, Fuhrman JA. (2005) Wide-ranging abundances of aerobic anoxygenic phototrophic bacteria in the world ocean revealed by epifluorescence microscopy and quantitative PCR. *Limnol Oceanogr* 50:620–8.

Sieracki, M. E., I. C. Gilg, E. C. Thier, N. J. Poulton, and R. Goericke. (2006) Distribution of planktonic aerobic anoxygenic photoheterotrophic bacteria in the northwest Atlantic. *Limnol Oceanogr* 51: 38–46. doi:10.4319/lo.2006.51.1.0038

Soora M, & Cypionka H. (2013) Light enhances survival of *Dinoroseobacter shibae* during long-term starvation. *PLOS One* 2013;8:e83960.

Staley JT, & Konopka A (1985) Measurements of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Ann Rev Microbiol* 39:321–346

Stegman MR, Cottrell MT, & Kirchman DL. (2014) Leucine incorporation by aerobic anoxygenic phototrophic bacteria in the Delaware estuary. *ISME* 2014;8:2339–48.

Steindler, L., Schwalbach, M.S., Smith, D.P., Chan, F., & Giovannoni, S.J. (2011) Energy starved *Candidatus* Pelagibacter ubique substitutes light-mediated ATP production for endogenous carbon respiration. *PLoS ONE* 6: e19725.

Suyama T, Shigematsu T, Takaichi S, et al. (1999) *Roseateles depolymerans* gen. nov., sp. nov., a new bacteriochlorophyll a-containing obligate aerobe belonging to the β-subclass of the Proteobacteria. *Int J Syst Bacteriol* 49:449–57.

Walter, J.M., Greenfield, D., Bustamante, C., and Liphardt, J. (2007) Light-powering Escherichia coli with proteorhodopsin. *Proc Natl Acad Sci USA* 104: 2408–2412.

Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998) Prokaryotes: The unseen majority. *PNAS* 95:6578–6583.

Williams, P. J. le B., & Ducklow, H. W. (2019). The microbial loop concept: A history, 1930–1974. *Journal of Marine Research*, 77(2), 23–81. https://doi.org/10.1357/002224019828474359

Xiao N., Jiao N. (2011) Formation of polyhydroxyalkanoate in aerobic anoxygenic phototrophic bacteria and its relationship to carbon source and light availability. *Appl. Environ. Microbiol.* 77:7445–7450

Yurkov, V.V. & Beatty, J.T. (1998a) Aerobic anoxygenic phototrophic bacteria. *Microbiol Mol Biol R* 62:695–724.

Yurkov, V.V. & Beatty, J.T. (1998b) Isolation of aerobic anoxygenic photosynthetic bacteria from Black Smoker plume waters of the Juan de Fuca Ridge in the Pacific Ocean. *Appl Environ Microb* 64:337–41.

Yurkov VV, Krieger S, Stackebrandt E, et al. (1999) *Citromicrobium bathyomarinum*, a novel aerobic bacterium isolated from deep-sea hydrothermal vent plume waters that contains

photosynthetic pigment-protein complexes. *J Bacteriol* 181:4517–25.

Yurkov, V.V. & Csotonyi, J.T. (2009) New light on aerobic anoxygenic phototrophs. In: Hunter CN, Daldal F, Thurnauer MC, et al. (eds). The Purple Phototrophic Bacteria. *Advances in Photosynthesis and Respiration*, Vol. 28. Dordrecht, the Netherlands: Springer, 31–55.

Yutin N, Suzuki MT, & Béjà O (2005) Novel primers reveal wider diversity among marine aerobic anoxygenic phototrophs. *Appl Environ Microbiol* 71:8958–8962. https://doi.org/10.1128/AEM.71.12.8958-8962.2005

Yutin N, Suzuki M, Teeling H, Weber M, Venter JC, Rusch DB, & Béjà O (2007) Assessing diversity and biogeography of aerobic anoxygenic phototrophic bacteria in surface waters of the Atlantic and Pacific Oceans using the Global Ocean Sampling expedition metagenomes. *Environ Microbiol* 9:1464–1475. https://doi.org/10.1111/j.1462-2920.2007.01265.x

Zubkov, M. V., & Tarran, G. A. (2008). High bacterivory by the smallest phytoplankton in the North Atlantic Ocean. *Nature*, 455(7210), 224–226. https://doi.org/10.1038/nature07236

OBJECTIVES

AIMS AND OBJECTIVES OF THE THESIS

The main aim of this thesis is to gain insights into the ecology of marine aerobic anoxygenic phototrophic (AAP) bacteria on a large scale. While the mechanisms operating globally in major planktonic groups are being elucidated, our knowledge on AAP bacteria is still fragmented and biased towards some specific areas of the ocean. To achieve this goal, this thesis has been structured into four chapters.

Chapter I (*Global diversity and distribution of aerobic anoxygenic phototrophs in the tropical and subtropical oceans*) provides a global assessment of the diversity of AAP bacteria and of the biogeographic patterns governing their communities in most of the world oceans. Taking advantage of the Malaspina Expedition that covered the tropical and subtropical global ocean, I analyze AAP bacterial communities in the surface ocean. While in this first **chapter I** introduce potential primer-related challenges, I thoroughly address this issue in **Chapter II** (*A metagenomic and amplicon sequencing combined approach reveals the best primers to study marine aerobic anoxygenic phototrophs*). In this second chapter, I conduct an exhaustive evaluation of the commonly used primers in PCR-based approaches, design new ones, and compare their performance with metagenomics results. The assessment considers temporal and spatial variability of AAP communities, by employing samples from the Blanes Bay Microbial Observatory (BBMO), collected at different times of the year, along with the global samples from the Malaspina Expedition examined in **Chapter I**. This investigation has allowed to identify the optimal set of primers that captures the broadest AAP diversity. Once identified, I use these primers in **Chapter III** (*Diversity and community structure of AAP communities across the Atlantic Ocean epipelagic waters*) to analyze AAP communities of the Atlantic Ocean at both, the longitudinal and the vertical scale. I analyze the composition of AAP assemblages along the deep chlorophyll maximum (DCM) variation, and across areas of contrasting productivity in the Atlantic Ocean to fill the knowledge gap existing on their distribution throughout the water column. Finally, in **Chapter IV** (*Vertical distribution and light energy capture by AAP bacteria along contrasted areas of productivity in the Atlantic Ocean*), I move beyond diversity analyses, and building on the same transect introduced in **Chapter III**, I take a significant step forward on the study of these bacteria. Focusing on the cell abundances and pigment concentrations of AAP bacteria, I aim at estimating the solar energy captured by AAP communities across the Atlantic Ocean.

In the abovementioned context, the specific objectives of this thesis are:

Objective 1. To provide a global assessment of the diversity and biogeography of AAP communities in the surface global ocean.

- To obtain a detailed description of the diversity and biogeographical patterns of surface AAP assemblages using Amplicon Sequence Variants (ASVs) at a fine scale.
- To unravel the underlying factors influencing global AAP patterns.
- To investigate whether AAP are governed by the same ecological processes that are observed to determine surface ocean microbiota.

Objective 2. To assess the efficiency of current molecular tools for studying AAP bacteria in the marine environment.

- To design and optimize novel primers for the efficient amplification of the *pufM* gene.
- To evaluate the performance of both existing and newly designed primers.
- To conduct a comparative analyses of the effectiveness of amplicon sequencing vs. metagenomics.

Objective 3. To investigate the composition of AAP bacteria across both horizontal and vertical scales within an area characterized by contrasting productivity.

- To assess the composition of AAP communities across a latitudinal transect spanning the South and Mid Atlantic Ocean, utilizing primers with a broad phylogenetic coverage.
- To examine the composition of AAP communities at various depths along the Depth Chlorophyll Maximum (DCM).
- To investigate the biotic and abiotic factors influencing the vertical distribution of AAP bacteria.

Objective 4. To examine the cell abundances of AAP bacteria and bacteriochlorophyll *a* concentrations to estimate their contribution to total ocean light energy capture.

- To quantify AAP bacteria and determine bacteriochlorophyll *a* concentrations along the photic zone of a latitudinal transect in the South and Mid Atlantic Ocean.
- To estimate the light energy captured by AAP bacteria throughout the sunlit water column, and compare to that of phytoplankton.

CHAPTER I

CHAPTER I

Global diversity and distribution of aerobic anoxygenic phototrophs in the tropical and subtropical oceans

Carlota R. Gazulla, Adrià Auladell, Clara Ruiz-González, Pedro C. Junger, Marta Royo-Llonch, Carlos M. Duarte, Josep M. Gasol, Olga Sánchez, Isabel Ferrera

Environmental Microbiology (2022) DOI: https://doi.org/10.1111/1462-2920.15835

Abstract

We studied the long-term temporal dynamics of the aerobic anoxygenic phototrophic (AAP) bacteria, a relevant functional group in the coastal marine microbial food web, using high-throughput sequencing of the *pufM* gene coupled with multivariate, time series and co-occurrence analyses at the Blanes Bay Microbial Observatory (NW Mediterranean). Additionally, using metagenomics, we tested whether the used primers captured accurately the seasonality of the most relevant AAP groups. Phylogroup K (Gammaproteobacteria) was the greatest contributor to community structure over all seasons, with phylogroups E and G (Alphaproteobacteria) being prevalent in spring. Diversity indices showed a clear seasonal trend, with maximum values in winter, which was inverse to that of AAP abundance. Multivariate analyses revealed sample clustering by season, with a relevant proportion of the variance explained by day length, temperature, salinity, phototrophic nanoflagellate abundance, chlorophyll *a*, and silicate concentration. Time series analysis showed robust rhythmic patterns of co-occurrence, but distinct seasonal behaviors within the same phylogroup, and even within different amplicon sequence variants (ASVs) conforming the same operational taxonomic unit (OTU). Altogether, our results picture the AAP assemblage as highly seasonal and recurrent but containing ecotypes showing distinctive temporal niche partitioning, rather than being a cohesive functional group.

1.1 Introduction

The discovery of marine aerobic photoheterotrophs (i.e., aerobic anoxygenic phototrophic (AAP) bacteria and proteorhodopsin-containing bacteria) (Béjà et al., 2000; Kolber et al., 2000) challenged the classic view of bacterioplankton being composed of photoautotrophic microorganisms as primary producers and of chemoheterotrophs as consumers. Since then, many studies have investigated their abundance, diversity and distribution in the ocean, and ultimately tried to understand their role in the marine ecosystem (DeLong and Béjà, 2010; Kirchman and Hanson, 2013; Koblížek, 2015; Pinhassi et al., 2016). AAP bacteria are photoheterotrophs that use dissolved organic matter but harvest solar energy using bacteriochlorophyll *a* (Bchl *a*) to supplement their metabolism. In the marine environment, these organisms can typically constitute up to 10% of total prokaryotes (Schwalbach and Fuhrman, 2005; Sieracki et al., 2006; Jiao et al., 2007; Hojerová et al., 2011), and are an active part of the community because they consist of large cells that display higher growth rates and receive higher grazing pressure than most bacteria (Sieracki et al., 2006; Koblížek et al., 2007; Ferrera et al., 2011, 2017). It has thus been hypothesized that this functional group plays a remarkably important role in the processing of organic matter and, as a consequence, in the global carbon cycle (see review by Koblížek, 2015).

Phylogenetically, marine AAP bacteria belong mainly to the Alpha- and Gammaproteobacteria classes. The *pufM* gene, which encodes the M subunit of the photosynthetic reaction centre, is commonly used to screen the diversity of AAPs in environmental samples and to describe their distribution patterns. The first studies showed AAP communities as being mainly affiliated to the alphaproteobacterial *Roseobacter*-like clade (Béjà et al., 2002; Oz et al., 2005) but the global ocean sampling (GOS), based on metagenomic data, unveiled that an important fraction of marine AAP bacteria was associated to phylogroups without cultured representatives (Yutin et al., 2007). The later study also showed that, while the *Roseobacter*-like AAPs were the most ubiquitous clade, unidentified uncultured groups dominated in open ocean areas, while Gammaproteobacteria dominated in coastal sites (Yutin et al., 2007). Later investigations showed that Gammaproteobacteria have in fact a widespread distribution and can constitute an important fraction of AAP communities in diverse sites of contrasting trophic status (Mašín et al., 2006; Lehours et al., 2010; Ferrera et al., 2014; Lehours and Jeanthon, 2015; Auladell et al., 2019). In contrast, AAPs from the Betaproteobacteria clade –that currently belongs to the Gammaproteobacteria class, according to the

Genome Taxonomy Database, GTDB (Parks et al., 2018)– are rarely prevalent in marine environments and they seem to prefer low-salinity waters (Waidner and Kirchman, 2008; Cottrell and Kirchman, 2009; Boeuf et al., 2013).

Although most AAP diversity studies have been restricted to particular areas of the world's ocean, a few studies have already compared communities across different oceanic regions. The pioneering metagenomic study by Yutin et al. (2007), which covered a transect between 45ºN in the Atlantic Ocean and 15ºS in the Pacific Ocean, showed that the composition of AAP communities varied between different biogeographical regions. By constructing clone libraries in a limited number of samples (N=10) from the Pacific, Atlantic and Indian oceans, Jiao et al. (2007) reported diversity patterns linked to the trophic regime of the oceanic region. Later, another study compared clone libraries of different seas encompassing a very large environmental variability (Mediterranean Sea, North Pacific Ocean, Western Beaufort Sea, Barents Sea and Norwegian Sea), and found that deterministic processes largely influenced the structuring of AAP assemblages (Lehours et al., 2018). This study further concluded that diverse AAP lineages showed some habitat preference, suggesting the existence of a certain degree of ecological cohesiveness for AAP clades, at least when comparing contrasting biomes. Besides, a study applying high-throughput sequencing to coastal Australian waters concluded that AAP communities exhibited niche partitioning whereas others shared their preferred niches (Bibiloni-Isaksson et al., 2016). Altogether, these results indicate that AAP assemblages –and the taxa within them– display complex spatial patterns (Jiao et al., 2007; Yutin et al., 2007; Lehours et al., 2010; Jeanthon et al., 2011; Boeuf et al., 2013; Lehours and Jeanthon, 2015; Bibiloni-Isaksson et al., 2016), probably driven by environmental selection (Lehours et al., 2018). Nevertheless, these conclusions are drawn from studies performed at different scales, using various methodologies and biased towards particular –and often coastal– ocean regions, so a coherent global assessment is still lacking. The exploration of the worldwide distribution of marine microorganisms, and thus, the definition of global biogeographical patterns, has become feasible in the last decade thanks to contemporary global oceanographic circumnavigations like the Malaspina Circumnavigation Expedition (Duarte, 2015) or the Tara Oceans Expedition (Karsenti et al., 2011), that used standardized procedures in a large collection of samples, coupled with recent advances in sequencing methodologies. Large scale surveys have also been key in the definition of the underlying ecological mechanisms in bulk prokaryotic and small eukaryotic communities (de Vargas et al., 2015; Salazar et al., 2015; Sunagawa et al., 2015; Ruiz-Gonzalez et al., 2019; Logares et al., 2020; Obiol et al., 2020). Data

generated from large sequencing initiatives have also been used to retrieve new diversity (Tully et al., 2018; Nayfach et al., 2020), including that within the AAPs (from the Tara Oceans expedition, Graham et al., 2018). Hence, a comprehensive study defining the global ocean biogeography of AAP assemblages and the mechanisms underlying their patterns is now feasible, but yet to be performed.

Here, we present a global assessment of AAP bacteria communities across the global tropical and subtropical ocean based on the Malaspina Circumnavigation Expedition. In particular, we studied the diversity and biogeography of AAP communities at a fine scale in the surface ocean using Amplicon Sequence Variants (ASVs) of the *pufM* gene. Our objectives were three-fold: (i) to describe the diversity and biogeography of the surface AAP assemblages along the global tropical and subtropical ocean, (ii) to disentangle the factors driving global patterns of AAP communities, and (iii) to compare the trends observed in the AAP communities with those of the broader surface ocean microbiota (i.e., whole prokaryotic and picoeukaryotic communities). For this purpose, we analyzed the composition of AAP communities based on the dominance or rarity of each individual taxa in an approach based on the spatial abundance distribution of each ASV. Furthermore, we estimated the role of different ecological processes shaping the structure of AAP communities. Since AAP bacteria, as a whole, display ecological traits that differentiate them from the rest of the bacterioplankton (i.e., photoheterotrophy, higher growth rates and higher susceptibility to predation than other prokaryotes), we hypothesize that their ecological patterns may deviate from those of the bulk communities.

1.2 Results and Discussion

Oceanographic context

The 113 studied stations were representative of the tropical and subtropical regions of the three major oceans, the Pacific, the Atlantic and the Indian Ocean (Table 1, Figure S1). The cruise track spanned across all five subtropical oceanic gyres, characterized by their oligotrophic conditions, as well as over relatively more productive areas such as the Equatorial Pacific, the Caribbean Sea, the Benguela Coastal province or the South Subtropical Convergence Zone, in the South Australian Bight (Estrada et al., 2016). The schedule of the cruise was planned so that most of the stations were sampled during spring and summer in order to avoid adverse weather conditions and allowing seasonal comparability. Across this route, temperatures ranged between 15.8 and 29.3 ^{o}C (mean 24.5 ^{o}C), with the coldest waters found in the South Australian Bight and the warmest

temperatures in stations located along the Equatorial Pacific and Atlantic Oceans (Figure S2). Salinity ranged from 33.15 to 37.65, being the highest in stations from the Atlantic Ocean and lowest in certain stations from the Indian and Pacific Oceans. Chlorophyll *a* (Chl *a*) ranged between 0.034 $mg \cdot m^3$ (station 38 in the South Atlantic) and 0.647 $mg \cdot m^3$ (station 45 in the Benguela Current Coastal province) with a mean value of 0.155 $mg \cdot m^3$. Phosphate, nitrate and silicate had higher concentrations in the Equatorial Pacific, in the South African stations and in the South of Australia (Figure S2). Water mass properties and productivity regimes for the stations sampled in the Malaspina Circumnavigation Expedition have been previously described in detail (e.g., Estrada et al., 2016; Regaudiede-Gioux et al., 2019; Teira et al., 2019 and Villamaña et al., 2019).

Contrasting patterns of alpha diversity in distinct biogeographical provinces

Our survey of the *pufM* gene allowed us to generate the largest dataset of ASVs of the *pufM* gene so far available. Partial sequencing of this marker resulted in 1119 distinct ASVs that clustered into 229 OTUs (Operational Taxonomic Units of 94% similarity). Rarefaction curves reached a plateau for all samples (Figure S3A), indicating that we obtained a fair representation of the AAPs' surface ocean diversity for each individual sample, while the global sample-based rarefaction curve (Figure S3B) suggested that the number of ASVs would rise had more stations been sampled. We observed a large variability in the richness estimates (Chao1 index) per community (Figure 1), which varied between 14 and 132 ASVs (mean 61), while the Shannon diversity index ranged between 0.9 and 3.9 (mean 2.9). Overall, richness values were within the same range than those previously reported from the Mediterranean Sea or Australian coastal waters using similar methodologies (Bibiloni-Isaksson et al., 2016; Auladell et al., 2019).

Richness and diversity of AAP communities were highest in the North Atlantic (mean richness 81, mean Shannon diversity 3.1) compared to other regions (Tukey test, $p<0.001$, Figure 1). Taxonomic richness and diversity varied between and within some Longhurst provinces. In general, AAP bacteria diversity was lower in eutrophic areas (correlation between Shannon and Chl *a* concentration, N=107, R=0.33, $p<0.001$ and primary production, N=96, R=0.38, $p<0.001$), consistent with previous observations (Jiao et al., 2007; Jeanthon et al., 2011). In contrast, AAP communities having higher richness values were associated with low concentrations of phosphate (N=89, R=0.48, $p<0.0001$) and nitrate (N=89, R=0.34, $p=0.001$), and correlated positively with temperature and salinity (N=113, R =0.24, $p=0.011$; R=0.29, $p=0.002$, respectively, see Table S1). Temperature and salinity had been shown to influence AAP bacterial richness at local scales (Lehours and

Jeanthon, 2015; Bibiloni-Isaksson et al., 2016). Our results demonstrate that temperature, salinity and trophic status govern patterns of AAP bacterial alpha diversity at the global scale.

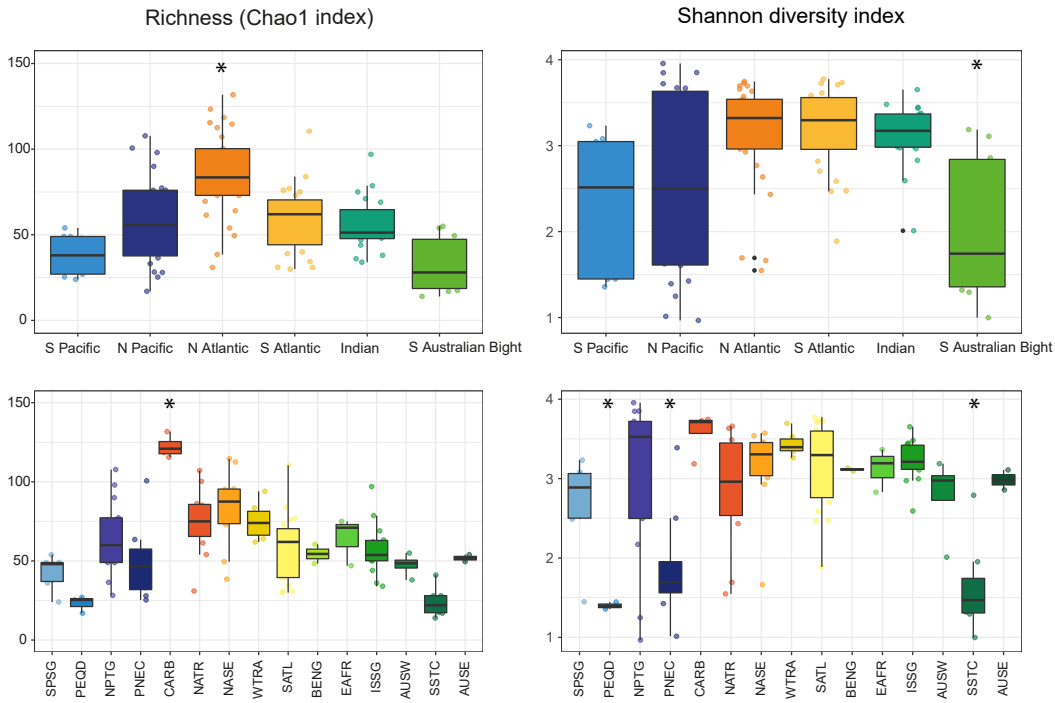

Figure 1. AAP alpha diversity measured as richness (Chao1 index) and Shannon diversity index within each oceanic region (top panels) and each Longhurst province (bottom panels) sampled during the Malaspina Expedition. The complete names of the Longhurst provinces are listed in Table 1 and Figure S1. *Asterisks indicate regions or provinces that are statistically different from the others, after a post-hoc Tukey test (p<0.001).

We also explored whether the patterns of AAP diversity were similar to the trends observed for other picoplanktonic groups. To this end, we compared the Shannon diversity of prokaryotes and picoeukaryotes (previously determined in the same sample set, Ruiz-González et al., 2019; Logares et al., 2020) with the values obtained for AAP bacteria. We observed a significant negative correlation between the Shannon diversity index of AAP communities and that of total prokaryotes (N=104, R=0.32, p=0.001, Figure S4), while no significant correlation was found with the picoeukaryotic community values. However, the low diversity values observed for AAP bacteria in some eutrophic regions (PEQD, PNEC and SSTC provinces, cf. Table 1 for complete names) were not observed in the whole prokaryotic dataset, suggesting that trophic status may exert a strongest role in shaping the diversity of the AAP subcommunity than of the bulk prokaryotic assemblage.

Table 1. Provinces covered during the 2010 Malaspina Expedition and values (average ± standard deviation, minimum to maximum) of temperature, salinity and chlorophyll *a* concentration measured in each province. Names and abbreviations according to Longhurst (1998). N = number of stations visited in each Longhurst province.

Provinces	Province abbreviations	N	Temperature (ºC)	Salinity	Chlorophyll *a* (mg·m^{-3})
East Australian Coastal	AUSE	2	21.39 ± 0.35 (21.14 to 21.64)	35.52 ± 0.08 (35.47 to 35.58)	0.34 ± 0.02 (0.32 to 0.36)
Australia-Indonesia Coastal	AUSW	4	23.05 ± 1.41 (21.36 to 24.8)	35.48 ± 0.09 (35.34 to 35.54)	0.13 ± 0.03 (0.1 to 0.16)
Benguela Current Coastal	BENG	2	20.55 ± 0.16 (20.44 to 20.66)	35.52 ± 0.06 (35.48 to 35.56)	0.14 ± 0.11 (0.06 to 0.22)
Caribbean	CARB	4	28.73 ± 0.29 (28.38 to 29.09)	35.6 ± 0.08 (35.54 to 35.71)	0.14 ± 0.04 (0.09 to 0.19)
East Africa Coastal	EAFR	3	23.94 ± 1.82 (22.56 to 26)	35.41 ± 0.12 (35.31 to 35.54)	0.3 ± 0.3 (0.09 to 0.65)
Indian South Subtropical Gyre	ISSG	14	23.57 ± 1.36 (21.74 to 25.92)	35.65 ± 0.25 (35.23 to 36.14)	0.07 ± 0.03 (0.04 to 0.14)
Northeast Atlantic Subtropical Gyre	NASE	10	21.35 ± 1.88 (18.45 to 24.31)	37.03 ± 0.39 (36.43 to 37.65)	0.1 ± 0.07 (0.04 to 0.25)
North Atlantic Tropical Gyre	NATR	11	26.82 ± 1.32 (24.83 to 28.85)	36.68 ± 0.67 (35.53 to 37.57)	0.14 ± 0.1 (0.05 to 0.31)
North Pacific Tropical Gyre	NPTG	13	23.86 ± 1.49 (21.65 to 26.35)	34.66 ± 0.2 (34.2 to 34.94)	0.17 ± 0.09 (0.08 to 0.44)
Pacific Equatorial Divergence	PEQD	3	27.5 ± 0.62 (26.89 to 28.13)	35.21 ± 0.31 (34.85 to 35.39)	0.24 ± 0.05 (0.19 to 0.29)
North Pacific Equatorial Countercurrent	PNEC	8	28.37 ± 0.58 (27.61 to 29.28)	33.84 ± 0.4 (33.15 to 34.28)	0.34 ± 0.11 (0.18 to 0.52)
South Atlantic Gyre province	SATL	19	24.7 ± 2.26 (20.9 to 27.33)	36.49 ± 0.45 (35.79 to 37.25)	0.07 ± 0.03 (0.03 to 0.12)
South Pacific Subtropical Gyre	SPSG	7	28 ± 2 (23.99 to 29.31)	35.04 ± 0.41 (34.43 to 35.59)	0.11 ± 0.05 (0.06 to 0.18)
South Subtropical Convergence Province	SSTC	7	17.29 ± 1.35 (15.75 to 19.55)	35.3 ± 0.2 (34.99 to 35.61)	0.25 ± 0.14 (0.1 to 0.52)
Western Tropical Atlantic Province	WTRA	6	27.6 ± 0.29 (27.27 to 28.05)	35.77 ± 0.38 (35.42 to 36.33)	0.23 ± 0.11 (0.11 to 0.44)

Spatially structured communities dominated by distinct taxonomic groups

We classified all the ASVs into seven broad taxonomic groups based on their placement in a reference phylogenetic tree (Figure S5). One group contained sequences assigned to the family Halieaceae of the Gammaproteobacteria (here-after 'Gamma-Halieaceae' group), while the 'Gamma-Burkholderiales' group included sequences from Burkholderiales order (formerly affiliated to the Betaproteobacteria class, but now reassigned within the Gammaproteobacteria class, according to the GTDB; Parks et al., 2018). Members of the Alphaproteobacteria were distinguished into four subgroups: 'Methylobacteriaceae' (sequences from order Rhizobiales, family Methylobacteriaceae), 'Rhodobacterales' (order Rhodobacterales), 'Sphingomondales' (order Sphingomonadales), and 'Alpha-Others', which grouped other members of the Alphaproteobacteria that could not be further assigned. Finally, sequences that did not belong to any of these groups were classified as 'Others'.

Most of the studied communities (75 out of 113 sampled stations, Figure 2 and Figure S6) were dominated by Gamma-Halieaceae, followed by 24 stations dominated by Alpha-Rhodobacterales. The overall dominance of these groups is in agreement with previous studies from the Mediterranean Sea (Lehours et al., 2010; Jeanthon et al., 2011; Ferrera

et al., 2014; Auladell et al., 2019), the Baltic Sea (Mašín et al., 2006), the Arctic Ocean (Lehours and Jeanthon, 2015), and Australian waters (Bibiloni-Isaksson et al., 2016).

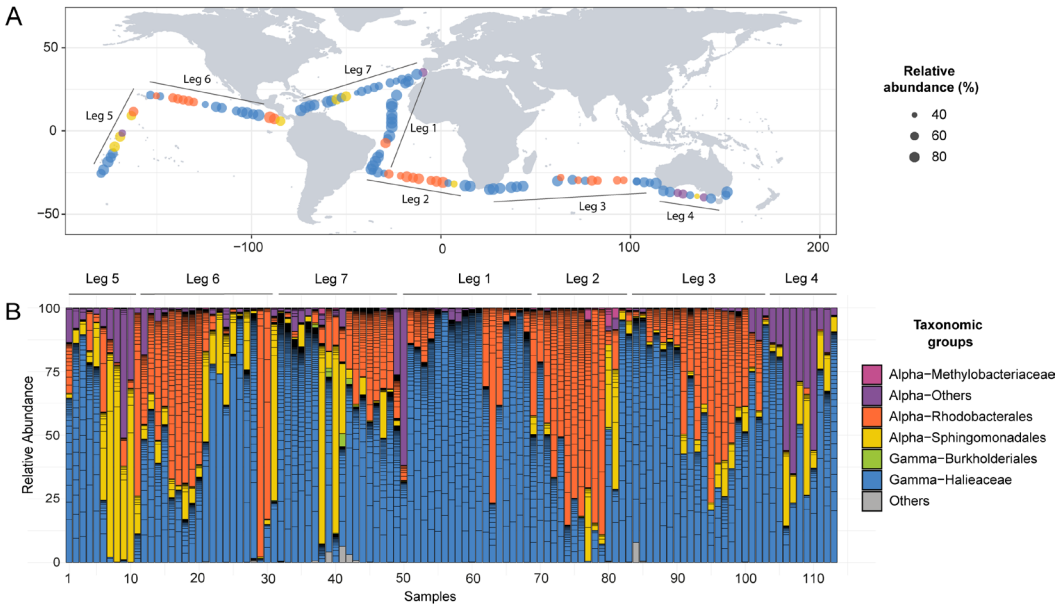

Figure 2. A) Dominant AAP taxonomic groups across the global tropical and subtropical surface ocean. Each station is colour-coded by the most abundant taxonomic group in the sample (see taxonomy legend in panel B), and the size of the dot is proportional to the relative abundance of the dominant taxon. The Malaspina Expedition legs are indicated to help visualize the cruise track. B) Community composition at each station, expressed as the relative contribution of each *pufM* sequence colour-coded by its taxonomic affiliation. Samples are ordered following the cruise path as in panel A.

The large dominance of gammaproteobacterial clades in marine AAP communities has been a matter of debate; it has been argued that it could be due to possible primer biases in amplicon-based studies (Lehours et al., 2010; Ferrera et al., 2014). In fact, PCR-based approaches can suffer from amplification biases that could result in misrepresentation of the relative abundances of various taxa as well as in low phylogenetic coverage. Nevertheless, a recent comparison of AAP assemblages in the Mediterranean Sea using metagenomics and *pufM* amplicon sequencing showed that, despite there were some discrepancies in the relative abundance of certain taxa, Gammaproteobacteria were abundant in both the amplicon and metagenomic datasets, which showed comparable patterns of diversity and community structure (Auladell et al., 2019). This study also showed that despite that the amplicon approach –identical to the one used here– missed some phylogenetic groups, it allowed the identification of other groups that were

overlooked by metagenomics because they were present in low abundances, as well as the retrieval of more variants, enabling the definition of distinct ecotypes among very similar sequences (Auladell et al., 2019). Metagenomics is often considered the least biased approach for functional gene analysis, but it is limited in its capacity to retrieve the least abundant members of the communities, and AAP taxa are generally present at relatively low abundances in natural samples (often below 10%). Although technically possible, the cost of conducting a high-resolution global ecological study based on a specific functional gene using metagenomics would be prohibitive and unfeasible for most researchers because, among other reasons, metagenomes retrieve less copies of specific marker genes for a given sequencing investment. Given that the goal of this work was to establish the global ecological patterns of AAP communities, and to understand how these are assembled at the fine scale, we consider that using *pufM* amplicon sequencing, despite not free of biases, was the most suitable approach to address our questions.

Interestingly, Gamma-Halieaceae and Alpha-Rhodobacterales-dominated communities were not randomly distributed but appeared to be spatially structured, with a marked succession of samples dominated by either one or the other group (Figure 2A). Gamma-Halieaceae contributed between 0.1% and 99.7% of total community *pufM* sequences (mean 58.42%). In locations where they were not abundant, the contribution of Alpha-Rhodobacterales was high, suggesting a replacement of the dominant taxonomic groups across space (Figure 2). Both groups also showed high intragroup diversity across samples (Figure 2), yet we observed that in stations 114, 115, 118 and 119, in the North Pacific region, this intragroup diversity decreased, and one single sequence assigned to the Gamma-Halieaceae (ASV217) represented abundances over 50% coinciding with a decrease in salinity. The relative contribution of Alpha-Rhodobacterales increased towards ultraoligotrophic gyre waters, characterized by low Chl *a* concentrations (N = 107, R=0.42, $p<0.001$) and deeper chlorophyll maxima (N=113, R=0.42, $p<0.001$). While the negative correlation between the contribution of Gamma-Halieaceae and Alpha-Rhodobacterales was observed before (Ferrera et al., 2014; Bibiloni-Isaksson et al., 2016; Auladell et al., 2019), in those cases, higher relative abundances of Alpha-Rhodobacterales were linked to higher concentrations of Chl *a* and, in general, to higher nutrient levels. While those studies were conducted in coastal stations, the Malaspina Circumnavigation Expedition occupied open-ocean stations, yet covered contrasting regions, from some relatively eutrophic areas (such as the Equator, South African provinces or the South Australian Bight) to the oligotrophic open ocean gyres. Just like seasonal ecotypes have been defined within the Alpha-Rhodobacterales based on 16S rRNA gene sequencing (Mena

et al., 2020), one possible explanation for the observed contrasting results is that closely related, but ecologically different Alpha-Rhodobacterales could be divided into an ecotype with a preference for productive regions such as coastal areas and an ecotype dominant in oligotrophic environments like the oceanic gyres.

Representatives from the Alpha-Sphingomonadales and 'Other Alpha' were scarce across the surface ocean with some localized exceptions (Figure 2B). The relative abundances of Alpha-Sphingomonadales members correlated positively with prokaryotic heterotrophic production (N=113, R=0.44, $p<0.005$), prokaryotic cell volume and total biomass (N=113, R=0.52, $p<0.001$ and R=0.46; $p<0.005$ respectively). Interestingly, in stations where Sphingomonadales dominated, this dominance was due to a single ASV (ASV512), which contributed up to 50% of the total AAP community reads. This ASV is related to an uncultured bacterial sequence (96% of identity in the *pufM* nucleotide sequence) detected in the Beaufort Sea (Boeuf et al., 2013), but does not resemble any cultured AAP bacteria. Thus, information on the physiology of the organism behind this sequence is missing. In any case, its widespread distribution from the Arctic to the tropical oceans and its ability to dominate communities under different conditions are remarkable. Other ASVs that could only be assigned to the Alpha-Proteobacteria level (and were thus grouped as 'Alpha-Others') dominated communities in some stations across the whole transect (Figure 2), such as station 93 in the Pacific Ocean, station 1 in the Atlantic Ocean, adjacent to the Strait of Gibraltar, and stations 71, 72 and 75 in the South Australian Bight, coinciding with the South Subtropical Convergence Zone (SSTC Longhurst province). In these stations, two ASVs were dominant, ASV860 in the Atlantic Ocean and ASV1102 in the Pacific and in the South of Australia. Although we could not classify them further and they do not have close cultured representatives, they are very similar to sequences from previous studies. In particular, ASV860 is very similar (99.5% identity) to a sequence retrieved from the Atlantic Ocean (OTU SPIT34 in Lehours and Jeanthon (2015), GenBank accession number KM654597) and ASV1102 is identical to an uncultured bacterium found in the East coast of Tasmania (Bibiloni-Isaksson et al., 2016). This ASV appears to be associated to low water temperature (correlation with temperature, N=113, R=0.40, $p<0.001$) and higher concentrations of nitrate (N=89, NO_3, R=0.47, $p<0.001$). Finally, Gamma-Burkholderiales representatives were scarce along the dataset (only 11 sequences with very low abundances) as expected, since this group is mostly absent in the marine environment (Ferrera et al., 2014; Bibiloni-Isaksson et al., 2016; Lehours et al., 2018; Auladell et al., 2019).

The taxonomic composition hitherto described here is based on the relative abundances of ASVs. In order to obtain data on the absolute abundance of AAP bacteria, we quantified them by epifluorescence microscopy. Unfortunately, we were not able to quantify AAP abundance along the entire transect, but only in a subset of 21 stations (samples for other stations were either not available or of insufficient quality). Yet, the stations for which the abundance was quantified were uniformly distributed along the transect (except for the Indian Ocean for which samples were not available) and should provide a good representation of the abundance variation along the tropical and subtropical oceans. Abundances ranged between $5.52 \cdot 10^2$ and $6.2 \cdot 10^4$ cells·ml^{-1} and the percentage of AAP bacteria within the prokaryotic community varied between 0.1% and 10% (Figure S7). Although we estimated AAP abundances for a subgroup of samples, their absolute and relative cell abundances are in line with the abundances reported in previous studies using the same methodology (see data reviewed in Koblížek, 2015). We observed higher AAP bacteria concentrations at lower latitudes (correlation between latitude and %AAP, N=21, R=0.50, p=0.024, Figure S7), and interestingly, we did not find any relationship between the abundance of AAP bacteria and the taxonomic composition of the AAP communities (Figure S7). This observation indicates that despite several communities were dominated by different ASVs, there was not a single dominant taxonomic group associated to the increase in absolute AAP bacterial abundances.

ASVs displaying bimodal and lognormal spatial abundance distributions (SpADs) dominate AAP assemblages.

We explored the spatial patterns of AAPs and found that most of the individual taxa (64%) were only found in one oceanic region, and these sequences represented only around 10% of the total number of reads. On the contrary, very few sequences (30 ASVs) appeared in all sampled areas, and they represented almost 50% of the total number of reads. Within this group of prevalent sequences, we found representatives of all the taxonomic groups defined above (data not shown), and thus, dominance or rarity of individual sequences does not seem to be linked to taxonomy. For this reason, to better understand the ecological behaviour of AAP taxa, we went beyond their taxonomic affiliations by analyzing the Spatial Abundance Distribution (SpAD) of the individual taxa, an approach that has proven as a useful tool for identifying groups of bacteria sharing similar spatial patterns regardless of their identity (Niño-García et al., 2016; Ruiz-González et al., 2019). In particular, the SpADs analysis classifies individual taxa into different categories according to the shape of their abundance distribution (see

Experimental Procedures). The different shapes can be interpreted as ecological traits because the abundance distribution of a given taxon will be the result of the combination of its physiological capacities, environmental tolerances or ability to persist under unfavourable conditions, but also of the external factors controlling its abundance. This approach has been previously used to explore the mechanisms behind the ubiquity or rarity of taxa within aquatic prokaryotic or picoeukaryotic communities (Niño-García et al., 2016; Mangot et al., 2018; Ruiz-González et al., 2019; LaBrie et al., 2021), but to our knowledge, this is the first time that it is restrictively applied to a functional group.

We only found two ASVs displaying normal-like distributions presenting high abundances and broad environmental tolerances (Figure S8A,C); the bimodal category (N=15 ASVs) included ASVs with lower average abundances and occurrence, likely representing less generalist taxa whose presence is restricted to specific areas, while lognormal (N=28) and logistic (N=872) distributions, which represented the majority of cases, were characteristic of globally rare and endemic AAPs (Figure S8). AAP assemblages in the surface ocean were dominated by bimodal and lognormal ASVs (Figure 3C), mostly associated to the Gamma-Halieaceae, Alpha-Rhodobacterales and Alpha-Sphingomonadales groups (Figure S8), and only few communities were dominated by either normal-like or logistic taxa. The two normal-like ASVs were Sphingomonadales-like (Figure S8), suggesting large environmental tolerances for this category, regardless of its relatively low contribution in most stations (Figure 2B).

Communities dominated by logistic ASVs in our study appeared spatially clustered and coincided with productive Atlantic, the Caribbean Sea, the Equatorial Pacific and the station nearest to the Strait of Gibraltar (Figure 3C). In fact, the relative abundances of logistic ASVs showed a significant positive correlation with the mean Chl *a* concentration across stations (N=107, R=0.43, $p<0.0001$), pointing to local selection of globally-rare opportunistic AAP bacteria in nutrient-rich areas, as shown for prokaryotic and picoeukaryotic bloomers (Ruiz-González et al., 2019; Logares et al., 2020).

Yet, the overall distribution of SpADs in our study differs from that reported by Ruiz-González et al. (2019) for the whole prokaryotic communities from the same Malaspina Circumnavigation Expedition surface samples. Whereas bimodal and lognormal ASVs were prevalent in AAP communities, bulk prokaryotic assemblages were dominated by a few cosmopolitan normal-like OTUs. For the bulk community, bimodal and logistic OTUs increased in stations with anomalies in temperature and productivity with respect

to the average values. This different distribution suggests that AAP bacteria are less homogeneously distributed than the bulk bacterioplankton (or at least than their dominant members) and that changes in the environment have a large effect on the AAP communities, promoting larger compositional shifts across environmental gradients and the increase of habitat specialists within this functional group.

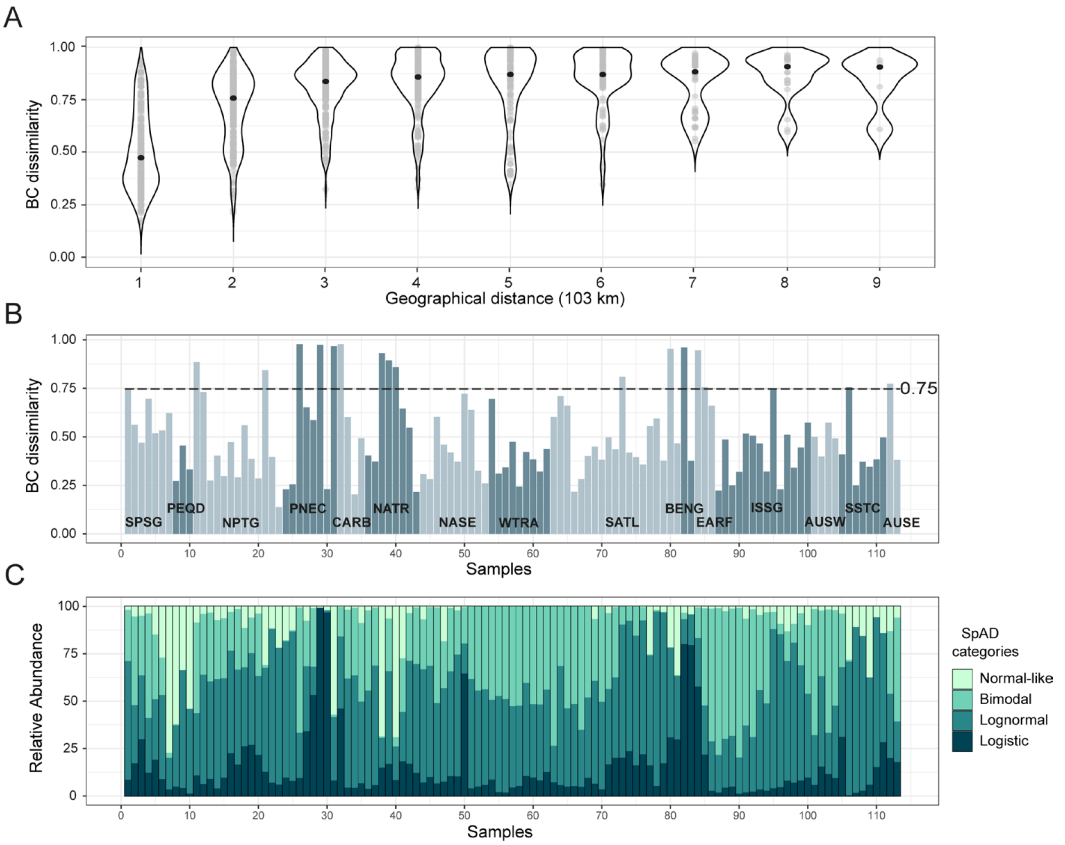

Figure 3. A) Changes in community dissimilarity between AAP assemblages, measured as Bray-Curtis (BC) dissimilarity with regard to the geographical distance among samples. All comparisons are represented by grey dots while black dots indicate the median value of dissimilarity at each distance. We only considered pairwise comparisons between samples located in the same ocean. B) Sequential change in community composition across space (sequential beta diversity). Bars represent BC dissimilarity between each community and the one sampled immediately before (e.g., first bar represents BC dissimilarity between stations 113 and 1, second bar represents BC dissimilarity between stations 1 and 2, and so on, up to stations 112 and 113). Samples are ordered following the cruise path as in Figure 2 for comparison. Alternating light and dark colour represent a change in Longhurst provinces along the transect and the provinces are indicated according to Longhurst (1998) abbreviations. C) Relative contribution in each community of the four Spatial Abundance Distribution (SpAD) categories of ASVs throughout the surface ocean, displayed using the same sample order as in panel B.

Environmental setting drives marked differences in community structure among oceanic regions

We further analyzed the AAP community structure along the Malaspina transect using Bray–Curtis (BC) dissimilarity metrics. The overall BC dissimilarity (mean=0.85 ± 0.15) was significantly higher than that described for prokaryotic and picoeukaryotic assemblages in the same transect (prokaryotes mean=0.61 ± 0.19; picoeukaryotes mean=0.74 ± 0.08, Logares et al., 2020), meaning that changes in the species composition and abundance distributions across AAP communities are larger than across bulk microbial groups. The higher beta diversity observed is consistent with these results, showing that AAP communities are mainly composed of habitat specialists (ASVs with a bimodal distribution) and rare taxa (lognormal distribution), while bulk prokaryotic communities were dominated by few abundant and ubiquitous species (Ruiz-González et al., 2019).

Moreover, we explored which abiotic and biotic variables influenced AAP community structure across the global ocean through PERMANOVA (p<0.001), and temperature, salinity and Chl *a* emerged as the most important variables (Table S2). When we pulled all samples together in a distance-based redundancy analysis (dbRDA, Figure S9), the first two axis explained only 16% of the variation, and there was no obvious clustering of communities based on region or province, even though simple analyses of variance showed statistical differences (p=0.001, Table S3). Thus, we further analyzed the samples for each ocean separately (Figure 4). Higher percentages of variation were explained by the two first axis (28.9%, Pacific Ocean, 29.1%, Atlantic Ocean and 40.2%, Indian Ocean), and the main variables associated were temperature (with the first axis) and salinity (with the second axis) (Figures 4 and S10). Stations from the same Longhurst province clustered together in most cases and, in general, we observed that communities from adjacent locations where more similar to each other than communities from distant stations, pointing to gradual changes in community structure along areas of the surface ocean (Figure 4). Previous studies restricted to specific areas of different ocean basins observed that the composition of AAP bacteria varied with the trophic conditions (Jiao et al., 2007; Yutin et al., 2007), while studies from the Arctic Sea showed that the hydrological context of the water masses was also relevant (Boeuf et al., 2013; Lehours and Jeanthon, 2015). Our results indicate that temperature, salinity and the general environmental context (as defined by the Longhurst provinces) largely structure AAP surface communities in the global tropical and subtropical ocean.

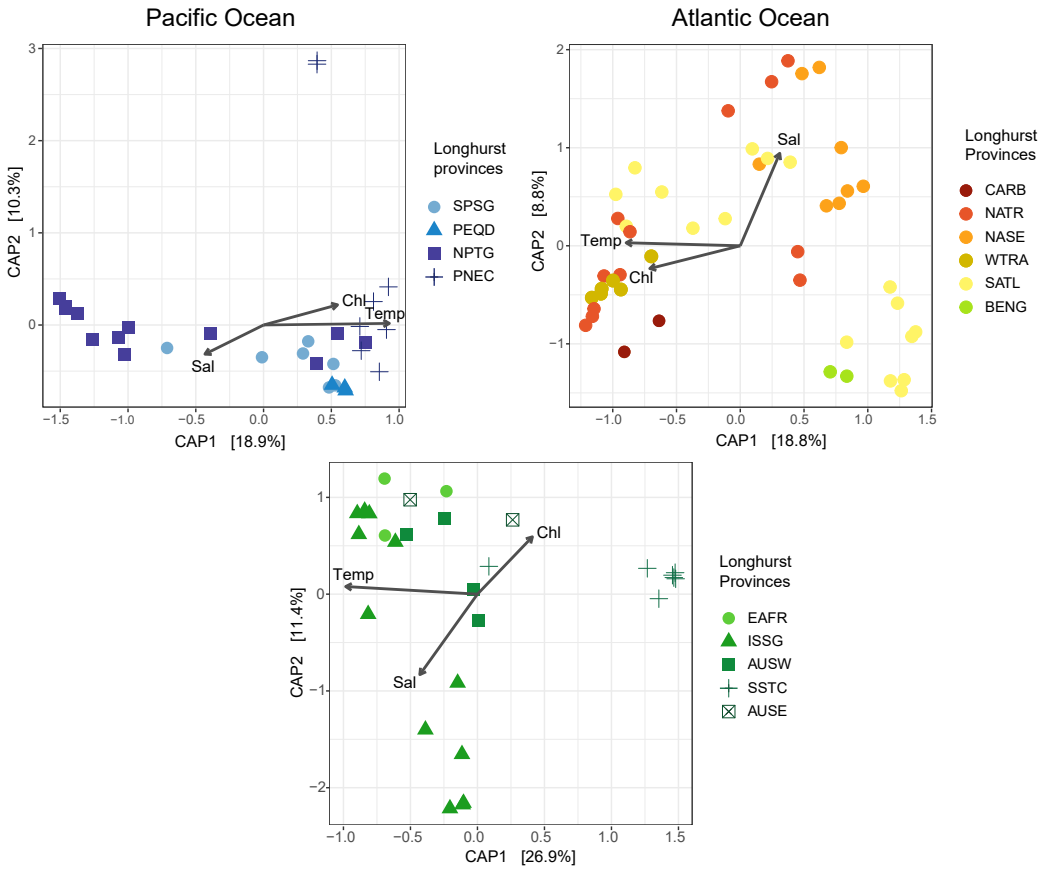

Figure 4. Distance-based redundancy analysis (dbRDA) performed separately for the Pacific, Atlantic and Indian Ocean stations. Samples are colour-coded according to the Longhurst provinces to which they belong. Temperature, salinity, and chlorophyll *a* were the three variables that explained the largest fraction of community variance, and they are represented by arrows, where 'Temp' is temperature, 'Sal' is salinity, and 'Chl' is chlorophyll *a* concentration.

Community dissimilarity increases with increasing geographic distance

To visualize the turnover of AAP communities along the Malaspina track, we plotted taxonomic community dissimilarities versus geographic distance (Figure 3A) –considering only pairwise comparisons within the same ocean– which unveiled a strong pattern of biogeography, that is, a remarkable increase of community dissimilarity with increasing distance within each ocean. To further explore community turnover at a fine scale, we explored the sequential changes of beta diversity across the whole sampling transect and found 17 stations displaying BC dissimilarity values > 0.75 which can be interpreted as sites of abrupt changes in community structure (Figure 3B). In general, the pattern of sequential beta diversity followed the changes shown through the SpADs analysis (Figure 3B and C). Stations showing the highest dissimilarity values (BC > 0.9) were located in

the South African Atlantic Coast (BENG and EARF) and the Costa Rica Dome (PNEC), where some sequences –belonging to logistic ASVs– presented remarkably high relative abundances, associated to an increase in Chl *a* concentration (see above). Other sites (BC values 0.75–0.9) were in the borders of several Longhurst provinces, such as the South Subtropical Convergence (SSTC), the Pacific Equatorial Divergence (PEQD), the North Atlantic tropical gyre (NATR) or provinces in the South Atlantic (SATL, BENG and EARF) (Figure 3B). The partition of the surface ocean into biogeographical provinces was proposed by Longhurst (2007) based on changes in environmental oceanic variables and their annual dynamics. This subdivision has been extensively used in several studies analyzing the surface ocean microbiota and has been proven to explain their biogeographic structure (see for example Friedline et al., 2012; Frank et al., 2016; Milici et al., 2016; Logares et al., 2020; Ruiz-González et al., 2020). We indeed observed that different Longhurst provinces harboured distinct AAP communities but it should also be considered that the borders between these provinces are dynamic and change seasonally (Reygondeau et al., 2013). For example, during the boreal summer, the Northwest Atlantic subtropical gyre (NASW, not included in this sampling) and North Atlantic tropical gyre (NATR) provinces tend to become mixed and an infiltration from the NASW province into the NATR province has clearly been observed (Figure 4 in Reygondeau et al., 2013). In this same area (NATR province) and timing (during June and July), we observed two stations (133 and 135) that differed largely from the rest, as seen by their different taxonomic composition (Leg 7 in Figure 2) and high BC sequential dissimilarities (Figure 3B). This difference could not be attributed to any measured environmental variable. Although this is speculative, the infiltration of water from a different province or some other physical oceanographic feature (Baltar et al., 2010, 2016; Bagnaro et al., 2020) could explain the abrupt changes seen in the North Atlantic in our study. Overall, we observed that AAP communities displayed strong biogeographic patterns, with large dissimilarities across the surface ocean which surpassed in magnitude those described for the bulk surface ocean microbiota.

Selection has a prominent role in structuring AAP communities

Our analyses of AAP community turnover clearly showed a biogeographical pattern across the surface ocean. The different patterns of diversity and species composition across spatial scales result from the combination of different ecological processes, such as selection, dispersal, or drift (Vellend, 2016). Changes in microbial species composition across space could be related to selection processes driven by changes in environmental variables (Figure 4). Nevertheless, we observed that environmental

conditions at adjacent stations were generally comparable, so these changes could also arise from dispersal limitation imposed by physical oceanographic features (Baltar et al., 2010, 2016; Bagnaro et al., 2020). In fact, previous studies have shown that oceanic features such as boundaries between different ocean regions can act as strong barriers and delimit the distribution of microbes in the ocean (Baltar et al., 2016; Raes et al., 2018).

Whether the pattern observed is the result of environmental selection and/or dispersal limitation cannot be determined based on our previous analysis (see also Hanson et al., 2012). Thus, to further investigate the ecological processes shaping AAP communities across the global surface ocean, we applied the approach proposed by Stegen et al. (2013), which quantitatively estimates the influence of selection, dispersal and drift based on the phylogenetic turnover of communities. Since this method relies solely on the phylogeny of the *pufM* gene and on null models (randomization), it avoids the problem of unmeasured environmental variables that can potentially be associated with selection or dispersal (Stegen et al., 2013). The influence of selection was estimated using the β-nearest taxon index (βNTI), which is the difference between the observed phylogenetic turnover for a given pair of communities and the null model after 999 randomizations (see Experimental Procedures). The values of βNTI were calculated for all the pairwise community comparisons possible in the dataset. We found that approximately 23% of the pairwise comparisons had values of $\beta NTI < 2$, which implies that there is a shorter phylogenetic distance within these pairs of communities, than expected by chance (Stegen et al., 2012). Lower turnover between communities is expected when environmental conditions are very similar and there is a homogeneous selection of closely related taxa in these communities. Likewise, approximately 27% of the pairwise comparisons had $\beta NTI > 2$, which is associated with a greater phylogenetic distance than the expected under a null model and can be interpreted as different environmental conditions heterogeneously selecting distantly related taxa (Stegen et al., 2012). Overall, approximately 50% of the observed turnover could be explained by selection, with homogeneous and heterogeneous selection being almost equally important at a global scale (Figure 5). Within samples located in the same Longhurst province, homogeneous selection had an important role, as the main ecological process in most provinces (Figure S11). In turn, heterogeneous selection had a modest role within Longhurst provinces, and only operated in some provinces. Based on βNTI values of comparisons between adjacent stations, heterogeneous selection was high in areas where logistic ASVs dominated (data not shown), pointing towards the selection of rare taxa in productive areas. These results are in line with previous studies that already pointed to selection as a major ecological

process structuring AAP communities in both spatial (Lehours et al., 2018) and temporal studies (Auladell et al., 2019).

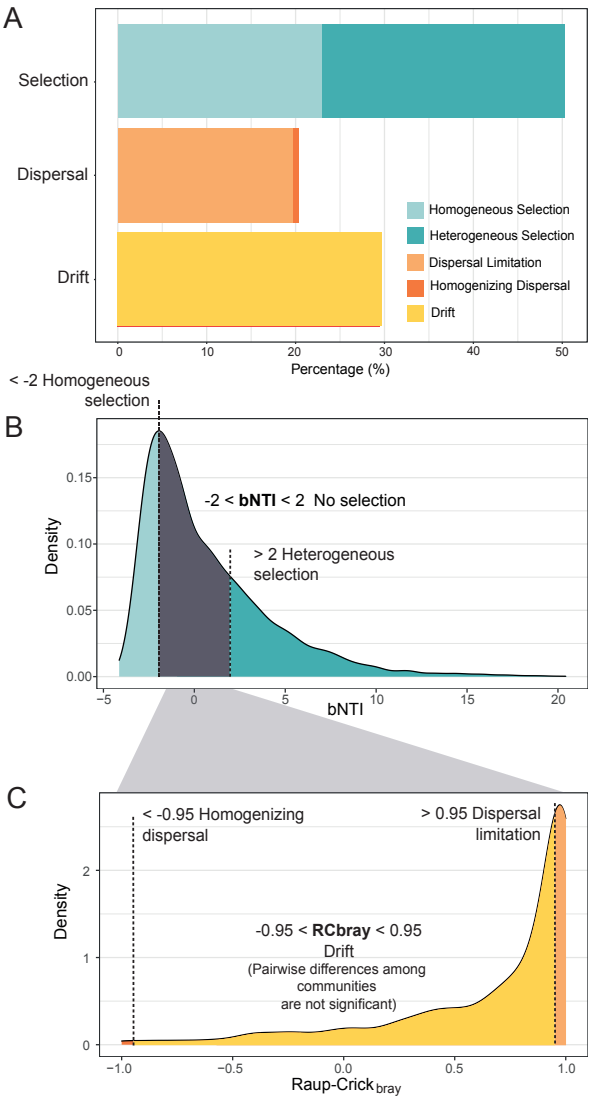

Figure 5. A) Percentage of the AAP bacterial community turnover associated to each ecological process in the tropical and subtropical surface ocean. B) Distribution of βNTI estimates for the total number of comparisons between all samples in the dataset. Absolute values of βNTI above 2 are considered as significant departures from random phylogenetic turnover and are associated to homogeneous and heterogeneous selection (blue areas). The grey area represents the fraction of nonsignificant βNTI values. To disentangle whether drift or dispersal are the main ecological processes shaping the turnover between these communities, Bray–Curtis-based Raup–Crick (RCbray) was calculated. C) Distribution of RCbray for the pairwise community comparisons that are not structured by selection. RCbray values between –0.95 and +0.95 point to a community assembly governed by drift (yellow area). On the contrary, RCbray values > +0.95 or <–0.95 indicate that community turnover is driven by dispersal limitation or homogenizing dispersal respectively (orange areas).

For the remaining pairwise comparisons, the value of Bray–Curtis-based Raup–Crick (RCbray) characterized the magnitude of deviation between the observed BC and the null BC. RCbray distribution varied between -1 and 1, and only values |RCbray| > 0.95 were considered as significant departures from drift (see Experimental Procedures and Figure 5). Dispersal limitation explained approximately 20% of the community turnover (RCbray > 0.95) while homogenizing dispersal was observed only for 18 pairwise comparisons (0.7%). The limited role of dispersal limitation structuring AAP communities could be expected, since distant microbial communities are known to be connected on a global scale, under what has been described as the 'Microbial Conveyor Belt' (Mestre and Höfer, 2021).

Finally, almost 30% of the community turnover was explained by drift (stochastic processes), as the differences between the null and the observed beta diversity were not significant. Stochastic processes are difficult to predict and to distinguish from other ecological processes (Zhou and Ning, 2017); however, they play an important role in microbial community assembly (Evans et al., 2017, Graham and Stegen, 2017) and their importance increases under high selection and low dispersal (Fodelianakis et al., 2021), as it happens in AAP communities across the surface ocean. Remarkably, the observed pattern is different from that reported for whole prokaryotic communities along the same transect (Logares et al., 2020), which appeared to be structured to a similar extent by both selection and dispersal (representing each process approximately 25% of the community turnover). In contrast, dispersal limitation had a much more important role in structuring picoeukaryotic communities (approximately 65%), likely due to their larger cell sizes and lower abundances (Logares et al., 2020). The relatively higher importance of selection mechanisms in AAP communities suggests that AAP bacteria are more affected by small changes in environmental conditions than the prokaryotic community as a whole. As we have shown above, the community turnover measured as BC dissimilarity is higher in this functional group than in the bulk picoplankton, pointing to higher changes in the composition and structure of AAP communities over short distances. Besides, while prokaryotic assemblages are dominated by few cosmopolitan and very abundant taxa, AAP assemblages are mainly composed by taxa classified as rare or habitat specialists, with more restricted environmental tolerances.

1.3 Conclusions

In this study, we describe the global diversity and community structure patterns of marine AAP bacteria in the tropical and subtropical oceans. Alpha diversity varied

across biogeographical provinces mainly related to temperature, salinity and trophic status and showed remarkably low values in the more productive Longhurst provinces. AAP communities along the surface ocean were mainly composed of members of the Halieaceae (Gammaproteobacteria), which were adapted to a large range of environmental conditions, and by different clades of the Alphaproteobacteria, that seemed to dominate under particular circumstances, such as in the oligotrophic gyres. These taxa were not randomly distributed but appeared to be spatially structured, with a marked succession of samples dominated by either one or the other class. Communities from adjacent stations shared more taxonomic similarities, that is, community dissimilarity increased with increasing distance, which resulted in a remarkable biogeographical pattern. However, this pattern was to a large extent the result of homogeneous and heterogeneous selection of individual taxa, while dispersal and drift had less of a role in shaping the structure of AAP bacterial communities. While the seasonal patterns of AAPs have been shown to be notably comparable to those of the bulk bacterioplankton, at a large-scale, AAP communities seem to have their own spatial patterns that do not mimic those of the bulk picoplankton. Of the measured environmental variables, temperature, salinity and Chla were found to influence AAP community structure. Small changes in environmental conditions translated into significant changes in AAP communities, and therefore, several habitat specialists and many rare species dominated their communities. The photoheterotrophic metabolism, high growth rates and high predation pressure on AAP bacteria, among other attributable traits to this functional group, could explain the stronger role of selection for this group as compared to the bulk surface ocean microbiota. Overall, our results represent the most comprehensive study investigating the global biogeography of AAP communities and shows how different ecological processes explain these patterns.

1.4 Methods

Sample collection

The Malaspina Circumnavigation Expedition took place between December 2010 and July 2011 (Duarte, 2015). Samples were collected in 113 stations across the tropical and subtropical waters of the Pacific, Atlantic and Indian Oceans. At each station, about 12 L of surface seawater (3 m depth) were collected with large (30 L) oceanographic bottles. Simultaneously, a CTD profiler was used to profile temperature, salinity, conductivity, fluorescence and dissolved oxygen. Seawater was prefiltered through a 20 µm nylon mesh and a 3 µm filter onto a 0.2 µm Millipore polycarbonate filter. Samples were

conserved at -80 ºC until further processing. Samples for enumerating AAP cells were pre-filtered through a 200 µm mesh and 10 ml of each sample were filtered onto 0.2 µm polycarbonate filters. Cells were enumerated by infrarred epifluorescence microscopy in 21 stations as described in Ferrera et al. (2014). The environmental biotic and abiotic parameters used in this study were determined as reported in Estrada et al. (2016) and Ruiz-González et al. (2019).

DNA extraction, *pufM* amplification, sequencing and ASV generation

DNA was extracted from the 0.2 µm filter using the phenol–chloroform protocol as described in Massana et al., (1997). Partial amplification of the *pufM* gene (~245 bp fragment) was performed in 50 µl reactions using primers pufM forward (5'-TACGGSAACCTGTWCTAC-3'; Béjà et al., 2002) and pufM_WAW reverse (5'-AYNGCRAACCACCANGCCCA-3'; Yutin et al., 2005) as described in Auladell et al. (2019). DNA was sequenced in an Illumina MiSeq sequencer (2 x 250 bp, Research and Testing Laboratory; http://rtlgenomics.com/). After sequencing, we used cutadapt v1.16 (Martin, 2013) to remove primers and DADA2 v1.10 (Callahan et al., 2016) to infer amplicon sequence variants (ASV) with the following parameters: maxEE = c(2,6) and trunclen = c(210,150). After filtering chimeras and spurious sequences, we kept 82% of the initial number of reads (mean 24,173, min 4503, max 79,968). To be able to compare our data with previous studies that used OTUs, we clustered the ASVs with UCLUST v10.0 (Edgar, 2010) at 94% similarity, the threshold usually employed for the *pufM* gene (Zeng et al., 2007).

Phylogenetic classification

We used phylogenetic placement for predicting the taxonomic assignment of the *pufM* gene short sequences. Due to the lack of comprehensive public databases for AAP bacteria, we built a custom made *pufM* database retrieving more than 750 sequences longer than 600 bp from the GTDB and GenBank, as well as from metagenomic datasets from the Tara Oceans Expedition (Sunagawa et al., 2015), the Malaspina Expedition (unpublished), the Global Ocean Survey (GOS, Yutin et al., 2007; Cuadrat et al., 2016), and the Blanes Bay Microbial Observatory (BBMO, Auladell et al., 2019). The alignment was done using the Decipher R package (Wright, 2016) and MAFFT v7 (Katoh and Standley, 2013). After a manual curation using AliView v1.26 (Larsson, 2014), we kept 673 sequences. A phylogenetic tree was constructed using RAxML v8.2 (Stamakis, 2014) (GTRGAMMA model, 100 bootstraps), and visualized using iTOL (Letunic and Bork, 2019), (Figure S5). Finally, to infer the phylogeny of the amplicon sequence variants, we applied the Evolutionary Placement Algorithm v0.3.5 (Barbera et al., 2019).

Data analyses

All statistical analyses were performed using R v3.6.3 (R Core Team 2021). The ASV table was rarefied down to 4500 reads per sample using the *vegan* package. Alpha diversity was estimated using Chao1 and Shannon diversity indices (Chao and Lee, 1992), with the *phyloseq* package. Post-Hoc Tukey tests were employed to see if there were statistically significant differences between the diversity of different regions. To test whether diversity was influenced by environmental conditions, we performed Pearson correlations between a selection of environmental variables and the diversity indices and applied false discovery rate (FDR)-correction to all *p*-values. We also compared the diversity of AAP bacteria with those of the bulk prokaryotic communities using the 16S rRNA gene data presented in Ruiz-González et al. (2019) and the picoeukaryotic community data (18S rRNA gene) presented in Logares et al. (2020), both from the same samples taken during the Malaspina Circumnavigation Expedition. Community composition was analyzed and described using the *phyloseq* package in R.

In order to explore the spatial patterns of individual AAP bacteria across space, we analyzed the abundance distribution of each *pufM* sequence across all samples. We used the rarefied table of reads ($\log_{10}(x + 1)$ transformed) to select the statistical distribution that best fitted the spatial abundance distribution (SpAD) of each ASV, as described in Niño-García et al. (2016) and Ruiz-González et al. (2019). We could classify all ASVs into four SpAD categories: 'normal-like' ASVs showed a normal statistical distribution, which has previously been associated with globally abundant and widespread taxa, and which might represent habitat generalists (Niño-García et al., 2016; Ruiz-González et al., 2019). The distribution of 'bimodal' ASVs is characterized by two density peaks, with the first one commonly corresponding to zero cases. This group could be considered less generalists because they are only detected in certain regions and their average abundances are also lower. Finally, ASVs classified as 'logistic' and 'lognormal' present a distribution with a zero-abundance mode, and they have been shown to comprise mostly rare sequences (for more details on the analysis see Niño-García et al., 2016 and Ruiz-González et al., 2019). For each SpAD category, we calculated the mean relative abundance of the constituent ASVs across all stations, as well as their occurrence. We also estimated the individual environmental breath as the range of temperature, salinity, Chl *a*, and dissolved oxygen concentration in which each of the ASVs within the different categories were detected.

The exploration of the main environmental drivers explaining the structure of AAP communities was done using a BC dissimilarity matrix, built with the *vegdist* function from the *vegan* package and visualized in a distance-base redundancy analysis (dbRDA), with a previous selection of significant environmental variables (PERMANOVA $p< 0.01$). Permutation tests (adonis function from *vegan* package) were employed to examine community differences among the six oceanic regions (South Pacific, North Pacific, North Atlantic, South Atlantic, Indian and South Australian Bight) and Longhurst oceanographic provinces (Longhurst, 2007). We used Mantel tests (999 permutations) to compare the changes in the structure of AAP communities between stations with differences in temperature, salinity and Chl *a*. In addition, we performed partial Mantel tests to compare the community structures of AAP, prokaryotes and picoeukaryotes, removing the effect of temperature, salinity and Chl *a*. The BC dissimilarity matrix was also used to analyze the spatial community structure turnover, and to explore sequential changes along the Malaspina transect, by comparing each sample with the one sampled immediately before.

Finally, to quantify the relative importance of selection, dispersal and drift as processes structuring the communities of AAP bacteria, we followed the framework developed by Stegen et al. (2013). This approach assumes that there is a phylogenetic signal (Cavender-Bares et al. 2009) in the ASVs optimal habitat conditions (i.e., the habitat preferences of closely related taxa are more similar than the preferences of distantly related taxa). To confirm this assumption, we firstly compared ASVs niche distances (using temperature, salinity and Chl *a*) and ASVs phylogenetic distances using a Mantel correlogram test. We detected phylogenetic signal in the pufM gene marker over relatively short phylogenetic distances (Fig. S12), as previously shown with other marker genes (e.g., Stegen et al., 2013; Dini-Andreote et al., 2015; Huber et al., 2020; Logares et al., 2020).

Then, to analyze the influence of selection, we calculated the β-mean nearest taxon distance (βMNT) metric, which quantifies the mean phylogenetic distances between two communities, and compared them to a random expectation (999 randomizations). The difference between the observed phylogenetic turnover (or βMNT) and the values obtained with the null model are denoted as β-nearest taxon index (βNTI). Absolute βNTI values above 2 ($|\beta NTI| > 2$) indicate that coexisting taxa are more closely related than expected by chance; thus, pointing to the action of selection. Afterwards, to differentiate whether drift or dispersal were the main structuring processes, we calculated the Raup–Crick metric (Chase et al., 2011) using BC dissimilarities (RCbray) (Chase et al., 2011; Stegen et al., 2013). RCbray compares the measured

beta diversity to the beta diversity obtained by the null model (999 randomizations) that would be obtained under random community assembly (drift). RCbray values between 0.95 and + 0.95 point to a community assembly governed by drift. On the contrary, RCbray values ≥ 0.95 or ≤ 0.95 indicate that community turnover is driven by dispersal limitation or homogenizing dispersal, respectively (Stegen et al., 2013). For this analysis, raw ASV sequences were aligned with AliView v1.26 (Larsson, 2014), aligned sequences were visually curated with Seaview (Gouy et al., 2010) and the phylogenic tree was constructed using FastTree v2.1.9 (Price et al., 2010). The βMNTD and βNTI metrics were calculated using the R package *picante* (Kembel et al., 2010) and the RCbray was calculated with the *raup_crick_abundance* function following Stegen et al. (2013). These analyses were performed in R v3.6.3 (R Core Team 2021) and codes are available in Github (https://gitlab.com/crgazulla/malaspina_aaps). Sequence data have been deposited in the NCBI Sequence Read Archive (SRA) under BioProject ID PRJNA736051.

1.5 Acknowledgements

We thank all scientists and crew involved in the Malaspina Expedition, particularly those participating in DNA sample collection and extraction, those collecting samples for AAP abundance, and those involved in generating the accompanying environmental data used here. This work was funded by the Spanish Ministry of Economy and Competitiveness (MINECO) through the Consolider-Ingenio program (Malaspina 2010 Expedition, CSD2008-00077), with contributions from grants ECLIPSE (PID2019-110128RB-I00) funded to IF, MIAU (RTI2018-101025-B-I00) to JMG and OS, and GRAMMI (RTI2018-099740-J-I00) to Clara RG, all from the Spanish Ministry of Science and Innovation (MICIN). Carlota RG was supported with a contract for research staff training from the Universitat Autònoma de Barcelona. Authors affiliated to the Institut de Ciències del Mar had the institutional support of the 'Severo Ochoa Centre of Excellence' accreditation (CEX2019-000928-S), and IF received the support of the BBVA Foundation through the 'Becas Leonardo a Investigadores y Creadores Culturales' 2019 Program. The Foundation takes no responsibility for the contents of this publication, which are entirely the responsibility of its authors. PCJ was supported by São Paulo Research Foundation–FAPESP (PhD grants #2017/26786-1 and #2020/02517-4).

1.6 References

Auladell, A., Sánchez, P., Sánchez, O., Gasol, J.M., & Ferrera, I. (2019) Long-term seasonal and inter annual variability of marine aerobic anoxygenic photoheterotrophic bacteria. *ISME* 13: 1975–1987.

Bagnaro, A., Baltar, F., Brownstein, G. Lee, W.G., Morales, S.E., Pritchard, D.W., & Hepburn, C.D. (2020) Reducing the arbitrary: fuzzy detection of microbial ecotones and ecosystems – focus on the pelagic environment. *Environ Microbiome* 15, 16 https://doi.org/10.1186/s40793-020-00363-w

Baltar, F., Arístegui, J., Gasol, J. Lekunberri, I., & Herndl, G.J. (2010) Mesoscale eddies: hotspots of prokaryotic activity and differential community structure in the ocean. *ISME* 4, 975–988. https://doi.org/10.1038/ismej.2010.33

Baltar, F., Currie, K., Stuck, E., Roosa, S., & Morales, S.E. (2016) Oceanic fronts: Transition zones for bacterioplankton community composition. *Environ Microbiol Rep* 8: 132–138.

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: 365–369.

Béjà, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A., Nguyen, L.P., et al. (2000) Bacterial Rhodopsin : Evidence for a New Type of Phototrophy in the Sea. *Science* 289: 1902–1906.

Béjà, O., Suzuki, M.T., Heidelberg, J.F., Nelson, W.C., Preston, C.M., Hamada, T., et al. (2002) Unsuspected diversity among marine aerobic anoxygenic phototrophs. *Nature* 415: 630–633.

Bibiloni-Isaksson, J., Seymour, J.R., Ingleton, T., van de Kamp, J., Bodrossy, L., & Brown, M.V. (2016) Spatial and temporal variability of aerobic anoxygenic photoheterotrophic bacteria along the east coast of Australia. *Environ Microbiol* 18: 4485–4500.

Boeuf D, Cottrell MT, Kirchman DL, et al. (2013) Summer community structure of aerobic anoxygenic phototrophic bacteria in the western Arctic Ocean. *FEMS Microbiol Ecol* 85:417–32.

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. *Nat Meth* 13, 581-583. doi: 10.1038/nmeth.3869.

Cavender-Bares J, Kozak K.H, Fine P.V.A, & Kembel S.W. (2009). The merging of community ecology and phylogenetic biology. *Ecol Lett* 12: 693–715.

Chao, A. & Lee, S.M. (1992) Estimating the number of classes via sample coverage. *Journal of the American Statistical Association* 87: 210–217.

Chase, J.M., Kraft, N.J.B., Smith, K.G., Vellend, M., & Inouye, B.D. (2011) Using null models to disentangle variation in community dissimilarity from variation in α-diversity. *Ecosphere* 2(2), 1-11 doi: https://doi.org/10.1890/ES10-00117.1

Cottrell, M.T. & Kirchman, D.L. (2009) Photoheterotrophic microbes in the arctic ocean in summer and winter. *App and Environ Microbiol* 75: 4958–4966.

Cuadrat, R., Ferrera, I., Grossart, H., & Dávila, A. (2016) Picoplankton Bloom in Global South? A High Fraction of Aerobic Anoxygenic Phototrophic Bacteria in Metagenomes from a Coastal Bay. *Omi A J Integr Biol.* 20: 76–87.

DeLong, E.F. & Béjà, O. (2010) The light-driven proton pump proteorhodopsin enhances bacterial survival during tough times. *PLoS Biology* 8: 1–5.

Dini-Andreote, F., Stegen, J.C., van Elsas J.D., S& alles, J.F. (2015) Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *PNAS* 112: E1326-E1332, doi: https://doi.org/10.1073/pnas.1414261112

Duarte, C.M. (2015) Seafaring in the 21st century: The Malaspina 2010 circumnavigation expedition. *Limnology and Oceanography Bulletin* 24: 11–14.

Edgar R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 26:2460–1.

Estrada, M., Delgado, M., Blasco, D., Latasa, M., Cabello, A.M., Benítez-Barrios, V., et al. (2016) Phytoplankton across tropical and subtropical regions of the Atlantic, Indian and Pacific oceans. *PLoS ONE* 11: 1–29.

Evans, S., Martiny, J. & Allison, S. (2017) Effects of dispersal and selection on stochastic assembly in microbial communities. *ISME* 11: 176–185. https://doi.org/10.1038/ismej.2016.96

Ferrera, I., Borrego, C.M., Salazar, G., & Gasol, J.M. (2014) Marked seasonality of aerobic anoxygenic phototrophic bacteria in the coastal NW Mediterranean Sea as revealed by cell abundance, pigment concentration and pyrosequencing of pufM gene. *EnvironMicrobiol* 16: 2953–2965.

Ferrera, I., Gasol, J.M., Sebastián, M., Hojerova, E., & Koblížek, M. (2011) Comparison of Growth Rates of Aerobic Anoxygenic Phototrophic Bacteria and Other Bacterioplankton Groups in Coastal Mediterranean Waters. *Appl Environ Microbiol*. 77: 7451–7458.

Ferrera, I., Sánchez, O., Kolářová, E., Koblížek, M., & Gasol, J.M. (2017) Light enhances the growth rates of natural populations of aerobic anoxygenic phototrophic bacteria. *ISME* 11: 2391–2393.

Fodelianakis, S., Valenzuela-Cuevas, A., Barozzi, A. et al. (2021) Direct quantification of ecological drift at the population level in synthetic bacterial communities. *ISME* 15, 55–66. https://doi.org/10.1038/s41396-020-00754-4

Frank, A.H., Garcia, J.A.L., Herndl, G.J., & Reinthaler, T. (2016) Connectivity between surface and deep waters determines prokaryotic diversity in the North Atlantic Deep Water. *Environ Microbiol* 18: 2052–2063.

Friedline, C.J., Franklin, R.B., McCallister, S.L., & Rivera, M.C. (2012) Bacterial assemblages of the eastern Atlantic Ocean reveal both vertical and latitudinal biogeographic signatures. *Biogeosciences* 9: 2177–2193.

Gouy M, Guindon S, & Gascuel O. (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol*. 27(2):221-4. doi: 10.1093/molbev/msp259.

Graham, E.D., Heidelberg, J.F., & Tully, B.J. (2018) Potential for primary productivity in a globally-distributed bacterial phototroph. *ISME* 12: 1861–1866.

Graham, Emily B.; Stegen, & James C. (2017). Dispersal-Based Microbial Community Assembly Decreases Biogeochemical Function *Processes* 5: 65. https://doi.org/10.3390/pr5040065

Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C., & Martiny, J.B.H. (2012) Beyond biogeographic

patterns: Processes shaping the microbial landscape. *Nature Reviews Microbiology* 10: 497–506.

Hojerová, E., Mašín, M., Brunet, C., Ferrera, I., Gasol, J.M., & Koblížek, M. (2011) Distribution and growth of aerobic anoxygenic phototrophs in the Mediterranean Sea. *Environ Microbiol* 13: 2717–2725.

Huber, P., Metz, S., Unrein, F. et al. (2020) Environmental heterogeneity determines the ecological processes that govern bacterial metacommunity assembly in a floodplain river system. *ISME* 14, 2951–2966. https://doi.org/10.1038/s41396-020-0723-2

Jeanthon, C., Boeuf, D., Dahan, O., le Gall, F., Garczarek, L., Bendif, E.M., & Lehours, A.C. (2011) Diversity of cultivated and metabolically active aerobic anoxygenic phototrophic bacteria along an oligotrophic gradient in the Mediterranean Sea. *Biogeosciences* 8:1955–1970.

Jiao, N., Zhang, Y., Zeng, Y., Hong, N., Liu, R., Chen, F., & Wang, P. (2007) Distinct distribution pattern of abundance and diversity of aerobic anoxygenic phototrophic bacteria in the global ocean. *Environ Microbiol* 9: 3091–3099.

Karsenti, E., Acinas, S.G., Bork, P., Bowler, C., de Vargas, C., Raes, J., et al. (2011) A holistic approach to marine Eco-systems biology. *PLoS Biology* 9: 7–11.

Katoh, K. & Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780.

Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., et al. (2010) Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26: 1463–1464.

Kirchman, D.L. & Hanson, T.E. (2013) Bioenergetics of photoheterotrophic bacteria in the oceans. *Environ Microbiol Rep* 5: 188–199.

Koblížek, M. (2015) Ecology of aerobic anoxygenic phototrophs in aquatic environments. *FEMS Microbiology Reviews* 39: 854–870.

Koblížek, M., Mašín, M., Ras, J., & Poulton, A.J. (2007) Rapid growth rates of aerobic anoxygenic phototrophs in the ocean. *Environ Microbiol* 9: 2401–2406.

Kolber, Z.S., van Dover, C.L., Niederman, R.A., & Falkowski, P.G. (2000) Bacterial photosynthesis in surface waters of the open ocean. *Nature* 407: 177–179.

LaBrie, R., Bélanger, S., Benner, R., & Maranger, R. (2021) Spatial abundance distribution of prokaryotes is associated with dissolved organic matter composition and ecosystem function. *Limnology and Oceanography* 66: 575–587.

Larsson, A. (2014) AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30: 3276–3278.

Lehours, A. & Jeanthon, C. (2015) The hydrological context determines the beta-diversity of aerobic anoxygenic phototrophic bacteria in European Arctic seas but does not favor endemism. *Front Microbiol* 6: 638.

Lehours, A.C., Cottrell, M.T., Dahan, O., Kirchman, D.L., & Jeanthon, C. (2010) Summer distribution and diversity of aerobic anoxygenic phototrophic bacteria in the Mediterranean Sea in relation to environmental variables. *FEMS Microbiol Ecol* 74: 397–409.

Lehours, A.C., Enault, F., Boeuf, D., & Jeanthon, C. (2018) Biogeographic patterns of aerobic

anoxygenic phototrophic bacteria reveal an ecological consistency of phylogenetic clades in different oceanic biomes. *Sc Rep* 8: 1–10.

Letunic, I. & Bork, P. (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* doi: 10.1093/nar/gkz239

Logares, R., Deutschmann, I.M., Junger, P.C., Giner, C.R., Krabberød, A.K., Schmidt, T.S.B., et al. (2020) Disentangling the mechanisms shaping the surface ocean microbiota. *Microbiome* 8: 1–17.

Longhurst, A. (1998). Ecological geography of the sea. San Diego, CA: Academic Press.

Mangot, J.F., Forn, I., Obiol, A., & Massana, R. (2018) Constant abundances of ubiquitous uncultured protists in the open sea assessed by automated microscopy. *Environ Microbiol* 20: 3876–3889.

Martin, M. (2013) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17:10.

Mašín, M., Zdun, A., Stoń-Egiert, J., Nausch, M., Labrenz, M., Moulisová, V., & Koblížek, M. (2006) Seasonal changes and diversity of aerobic anoxygenic phototrophs in the Baltic Sea. *Aquatic Microbial Ecology* 45: 247–254.

Massana, R., Murray, A.E., & Preston, C.M. (1997) Vertical Distribution and Phylogenetic Characterization of Marine Planktonic. *Microbiology* 63: 50–56.

Mena, C., Reglero, P., Balbín, R., Martín, M., Santiago, R., & Sintes, E. (2020) Seasonal Niche Partitioning of Surface Temperate Open Ocean Prokaryotic Communities. *Frontiers in Microbiology* 11 28;11:1749. doi: 10.3389/fmicb.2020.01749.

Mestre, M., & Höfer, J. (2021) The Microbial Conveyor Belt: Connecting the Globe through Dispersion and Dormancy. *Trends in Microbiology* 29: 482-492. https://doi.org/10.1016/j.tim.2020.10.007.

Milici, M., Tomasch, J., Wos-Oxley, M.L., Decelle, J., Jáuregui, R., Wang, H., et al. (2016) Bacterioplankton biogeography of the Atlantic ocean: A case study of the distance-decay relationship. *Frontiers in Microbiology* 7: 1–15.

Nayfach, S., Roux, S., Seshadri, R., Udwary, D., Varghese, N., Schulz, F., et al. (2020) A genomic catalog of Earth's microbiomes. *Nature Biotechnology* 39: 499–509. doi: 10.1038/s41587-020-0718-6.

Niño-García, J.P., Ruiz-González, C., & del Giorgio, P.A. (2016) Landscape-scale spatial abundance distributions discriminate core from random components of boreal lake bacterioplankton. *Ecology Letters* 19:1506–1515.

Obiol, A., Giner, C.R., Sánchez, P., Duarte, C.M., Acinas, S.G., & Massana, R. (2020) A metagenomic assessment of microbial eukaryotic diversity in & global ocean. *Molecular Ecology Resources* 20: 718–731.

Oz, A., Sabehi, G., Koblízek, M., Massana, R., & Béjà, O. (2005) Roseobacter-like bacteria in Red and Mediterranean Sea aerobic anoxygenic photosynthetic populations. *Applied and Environmental Microbiology* 71: 344–353.

Parks D.H., Chuvochina M., Waite D.W., Rinke C., Skarshewski A., Chaumeil P.-A., et al. (2018) A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree

of life. *Nat Biotechnol*. 36: 996–1004

Pinhassi, J., DeLong, E.F., Béjà, O., González, J.M., & Pedrós-Alió, C. (2016) Marine Bacterial and Archaeal Ion-Pumping Rhodopsins: Genetic Diversity, Physiology, and Ecology. *Microbiology and Molecular Biology Reviews* 80: 929–954.

Price MN, Dehal PS, & Arkin AP (2010) FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE* 5: e9490. https://doi.org/10.1371/journal.pone.0009490

R Core Team (2014) R: a language and environment for statistical computing.

Raes, E.; Bodrossy, L., van de Kamp, J.; Bissett, A.; Ostrowski, M.; Brown, M.V. et al. (2018) Oceanographic boundaries constrain microbial diversity gradients in the South Pacific Ocean. *PNAS* 115: E8266-E8275; doi: 10.1073/pnas.1719335115

Regaudie-de-Gioux A., Huete-Ortega M., Sobrino C., López-Sandoval D. C., González N., Fernández-Carrera A., Vidal M., Marañón E., Cermeño P., Latasa, M., Agustí S., and Duarte, C.M. (2019). Multi-model remote sensing assessment of primary production in the subtropical gyres. *J. Marine Systems* 196:97–106 https://doi.org/10.1016/j.jmarsys.2019.03.007

Reygondeau, G., Longhurst, A., Martinez, E., Beaugrand, G., Antoine, D., & Maury, O. (2013) Dynamic biogeochemical provinces in the global ocean. *Global Biogeochemical Cycles* 27: 1046–1058.

Ruiz-González, C., Logares, R., Sebastián, M., Mestre, M., Rodríguez-Martínez, R., Galí, M., et al. (2019) Higher contribution of globally rare bacterial taxa reflects environmental transitions across the surface ocean. *Molecular Ecology* 28: 1930–1945.

Ruiz-González, C., Mestre, M., Estrada, M., Sebastián, M., Salazar, G., Agustí, S., et al. (2020) Major imprint of surface plankton on deep ocean prokaryotic structure and activity. *Molecular Ecology* 29: 1820–1838.

Salazar, G., Cornejo-Castillo, F.M., Borrull, E., Díez-Vives, C., Lara, E., Vaqué, D., et al. (2015) Particle-association lifestyle is a phylogenetically conserved trait in bathypelagic prokaryotes. *Molecular Ecology* 24: 5692–5706.

Schwalbach, M.S. & Fuhrman, J.A. (2005) Wide-ranging abundances of aerobic anoxygenic phototrophic bacteria in the world ocean revealed by epifluorescence microscopy and quantitative PCR. *Limnology and Oceanography* 50: 620–628.

Sieracki, M.E., Gilg, I.C., Thier, E.C., Poulton, N.J., & Goericke, R. (2006) Distribution of planktonic aerobic anoxygenic photoheterotrophic bacteria in the northwest Atlantic. *Limnology and Oceanography* 51: 38–46.

Stamatakis A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30:1312–3.

Stegen, J.C., Lin, X., Fredrickson, J.K., Chen, X., Kennedy, D.W., Murray, C.J., et al. (2013) Quantifying community assembly processes and identifying features that impose them. *ISME* 7: 2069–2079.

Stegen, J.C., Lin, X., Konopka, A.E., and Fredrickson, J.K. (2012) Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME* 6: 1653–1664.

Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G., et al. (2015) Structure and function of the global ocean microbiome. *Science* 348: 1–10.

Teira, E., Logares, R., Gutiérrez-Barral, A., Ferrera, I., Varela, M.M., Morán, X.A.G., & Gasol, J.M. (2019). Impact of grazing, resource availability and light on prokaryotic growth and diversity in the oligotrophic surface global ocean. *Environmental microbiology* 21: 1482-1496.

Tully, B.J., Graham, E.D., & Heidelberg, J.F. (2018) The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. *Scientific Data* 5: 1–8.

de Vargas, C, Audie, S., Henry, N., Decelle, J., Mahé, F., Logares, R., et al. (2015) Eukaryotic plankton diversity in the sunlit ocean. *Science* 348: 1261605–1/11.

Vellend M. (2016) *The theory of ecological communities*. 1st ed. Woodstock, UK: Princeton University Press.

Villamaña, M., Marañón, E., Cermeño, P., Estrada, M., Fernández-Castro, B., Figueiras, F.G., Latasa, M., Otero-Ferrer, J.L., Reguera, B. & Mouriño-Carballido, B. (2019). The role of mixing in controlling resource availability and phytoplankton community composition. *Progress in Oceanography*, 178, p.102181.

Waidner, L.A. & Kirchman, D.L. (2008) Diversity and distribution of ecotypes of the aerobic anoxygenic phototrophy gene pufM in the Delaware estuary. *Applied and Environmental Microbiology* 74: 4012–4021.

Wright, E.S. (2016) Using DECIPHER v2.0 to analyze big biological sequence data in R. *R Journal* 8: 352–359.

Yutin, N., Suzuki, M.T., & Béjà, O. (2005) Novel primers reveal wider diversity among marine aerobic anoxygenic phototrophs. *Applied and Environmental Microbiology* 71: 8958–8962.

Yutin, N., Suzuki, M.T., Teeling, H., Weber, M., Venter, J.C., Rusch, D.B., and Béjà, O. (2007) Assessing diversity and biogeography of aerobic anoxygenic phototrophic bacteria in surface waters of the Atlantic and Pacific Oceans using the Global Ocean Sampling expedition metagenomes. *Environmental Microbiology* 9: 1464–1475.

Zeng, Y.H., Chen, X.H., and Jiao, N.Z. (2007) Genetic diversity assessment of anoxygenic photosynthetic bacteria by distance-based grouping analysis of pufM sequences. *Letters in Applied Microbiology* 45: 639–645.

Zhou J, Ning D. (2017) Stochastic community assembly: does it matter in microbial ecology? Microbiol *Mol Biol Rev* 81: e00002-17. https://doi.org/10.1128/MMBR.00002-17.

1.7 Supplementary figures

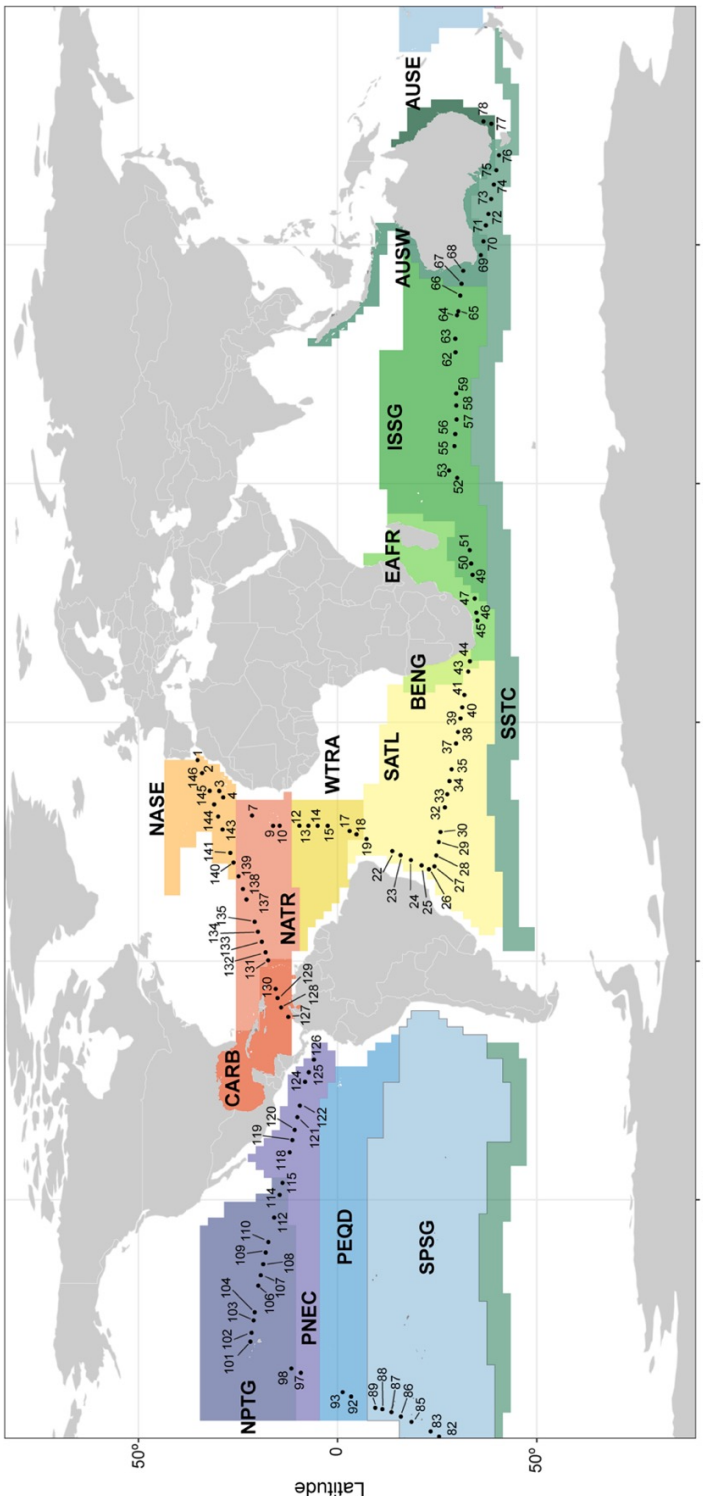

Figure S1. Stations sampled during the Malaspina Expedition included in this study. Station numbers follow the cruise numbering scheme for comparison with other studies using the same dataset. Longhurst provinces are colored-coded and indicated with their four-letters geocode. Longhurst provinces sampled from Station 1 to Station 146 are described in Table S1.

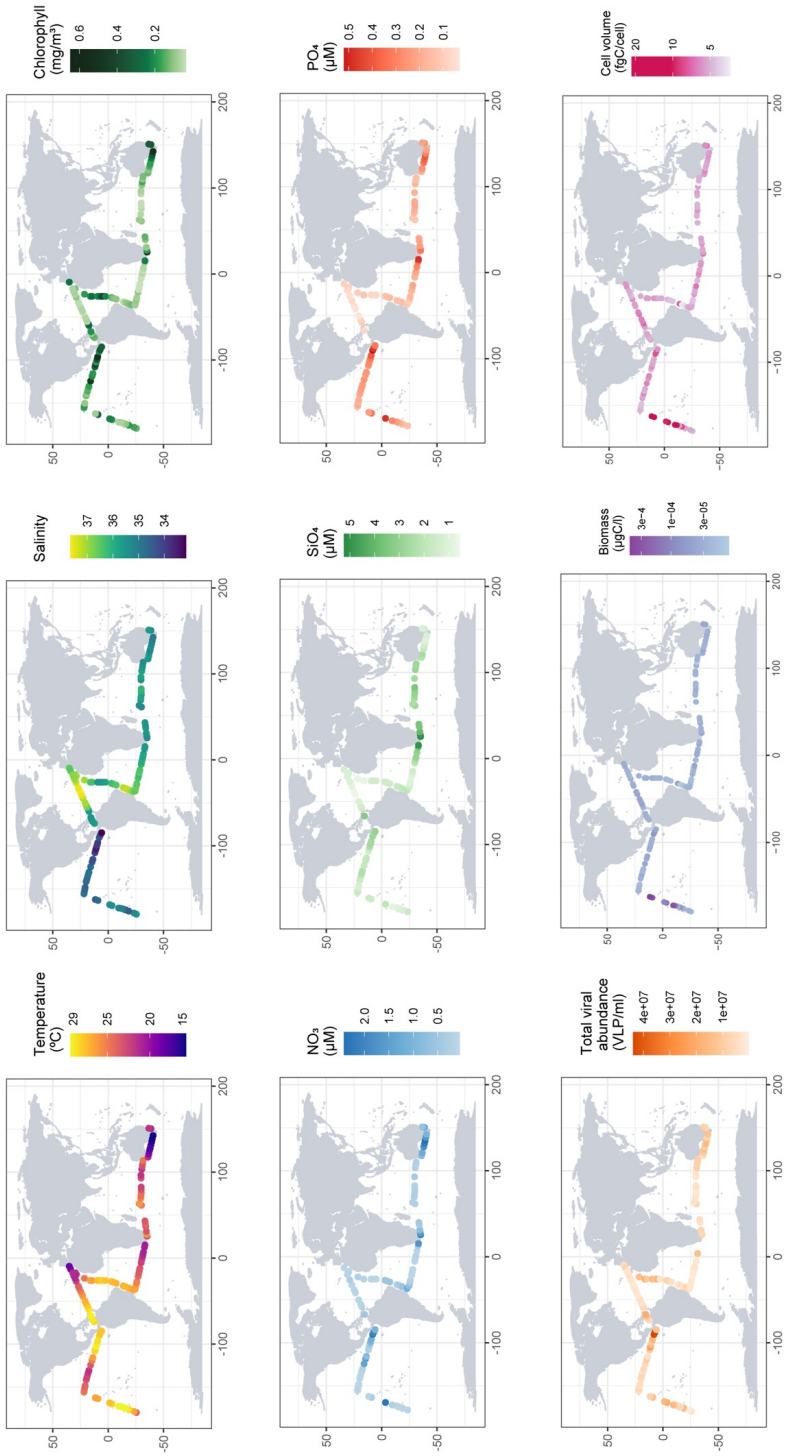

Figure S2. Maps showing the spatial variation of the main abiotic (temperature, salinity, nitrate, phosphate, and silicate concentrations) and biotic (chlorophyll *a* concentration, prokaryotic biomass and cellular carbon content, and viral abundance) parameters measured at the surface along the Malaspina Circumnavigation Expedition cruise track.

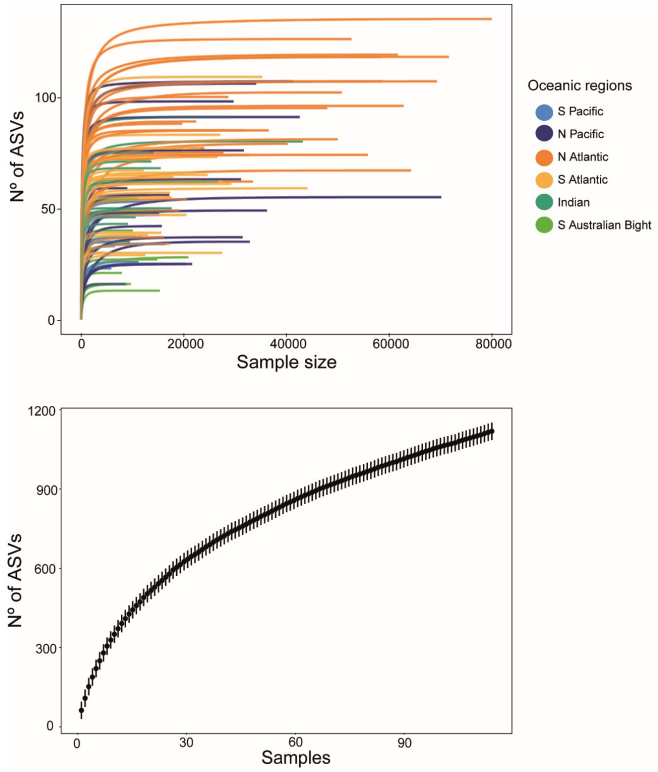

Figure S3. A) Rarefaction curve of each sample, coloured by oceanic region. B) Sample-based rarefaction curve, considering all sampled communities together.

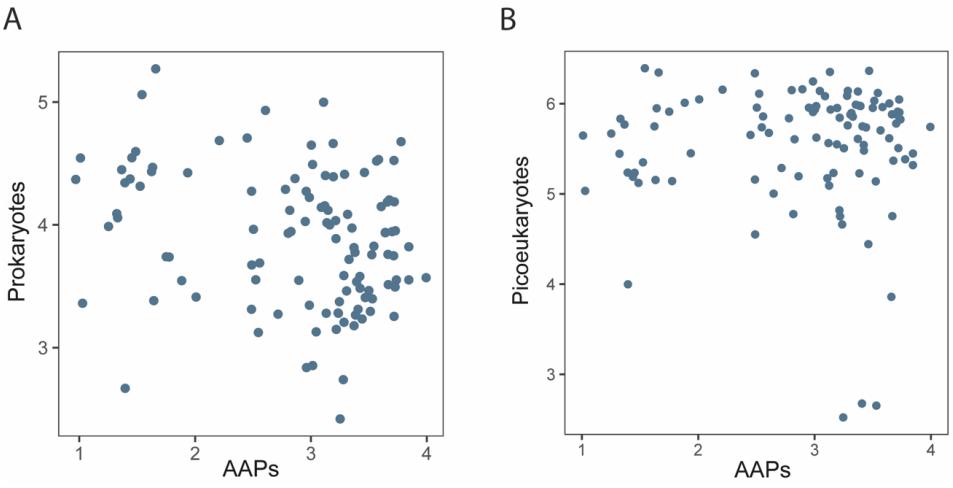

Figure S4. Correlation between the Shannon diversity index of AAP communities and that of the total prokaryotic (A) and picoeukaryotic (B) assemblages.

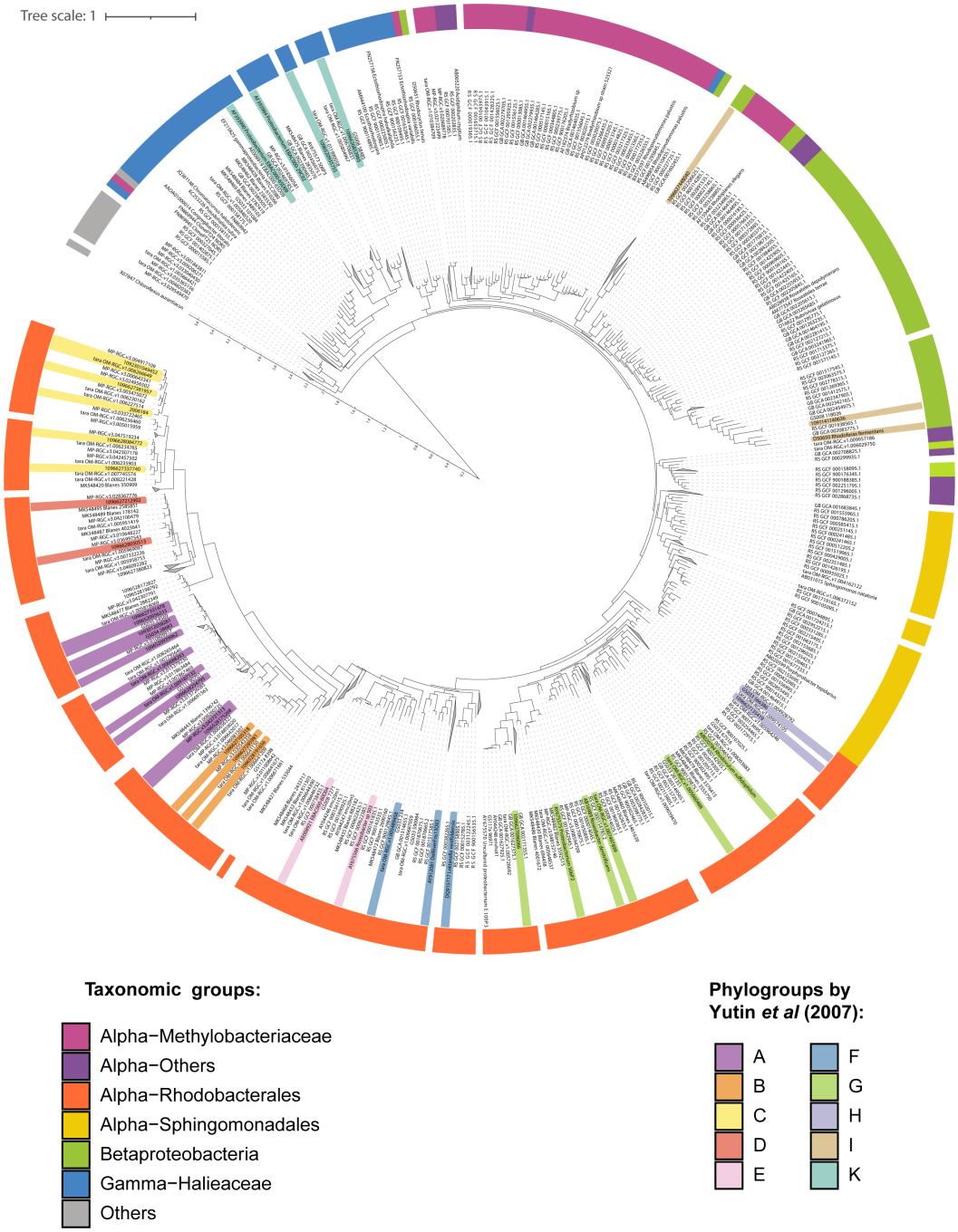

Figure S5. Phylogenetic tree of *pufM* sequences. Tree was constructed with RAxML v8.2 (Stamakis 2014), and visualized using iTOL (Letunic and Bork, 2011). Legend in the left shows the taxonomic groups defined in this manuscript. Note that the group 'Alpha-Others' is missing, since we used it for ASVs that could only be assigned to the Alpha-Proteobacteria level. For comparison, the phylogroups defined by Yutin et al., (2007) are shown in the right legend.

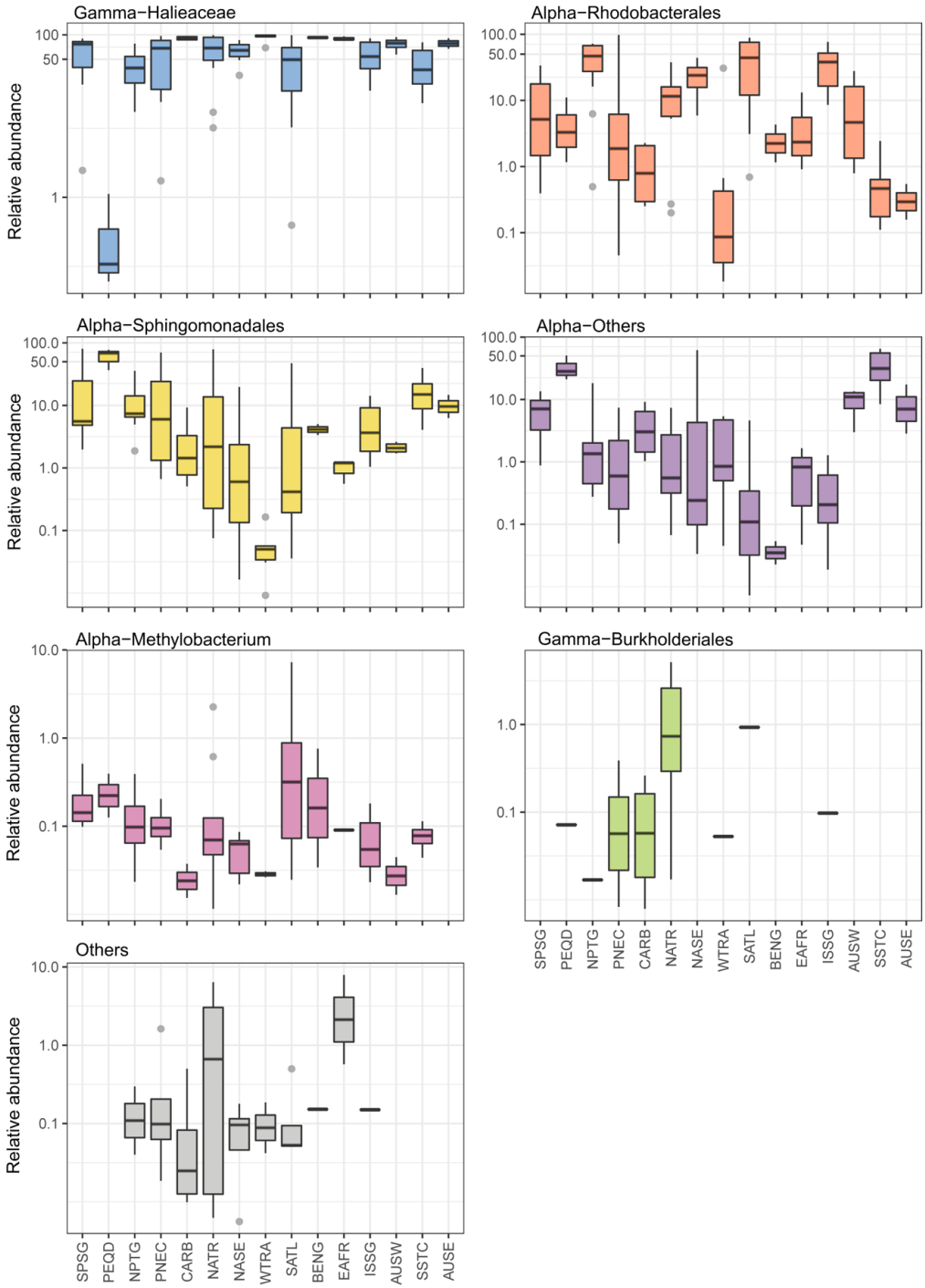

Figure S6. Relative contributions of the different AAP taxonomic groups in each Longhurst province. Full names of Longhurst provinces can be found in Figure S1. Each boxplot presents the median and the 25 and 75% limits of the distribution. Outliers are indicated by grey dots.

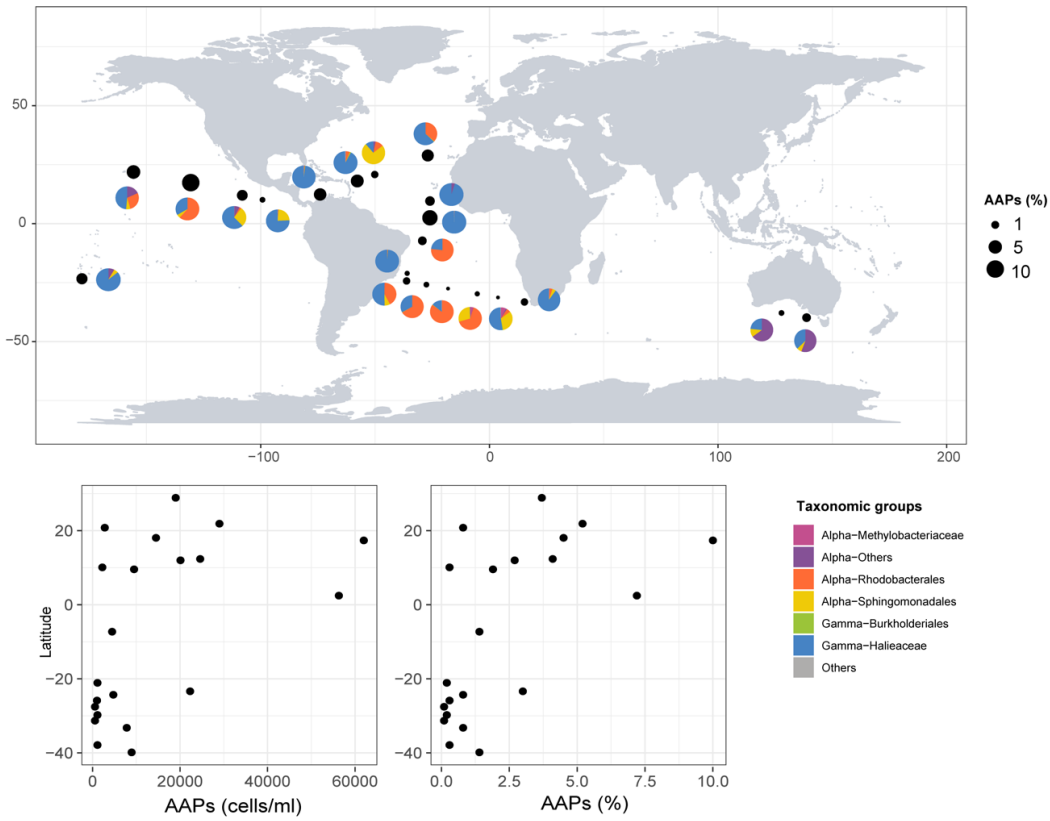

Figure S7. Percentage of AAP bacteria in the bacterioplankton, determined by epifluorescence microscopy, in 21 stations along the Malaspina transect. Pie charts next to each sample (black dots) show the taxonomic composition of each station while bottom dot plots show the absolute numbers of AAPs and percentage of AAPs within the total community in the latitudinal gradient.

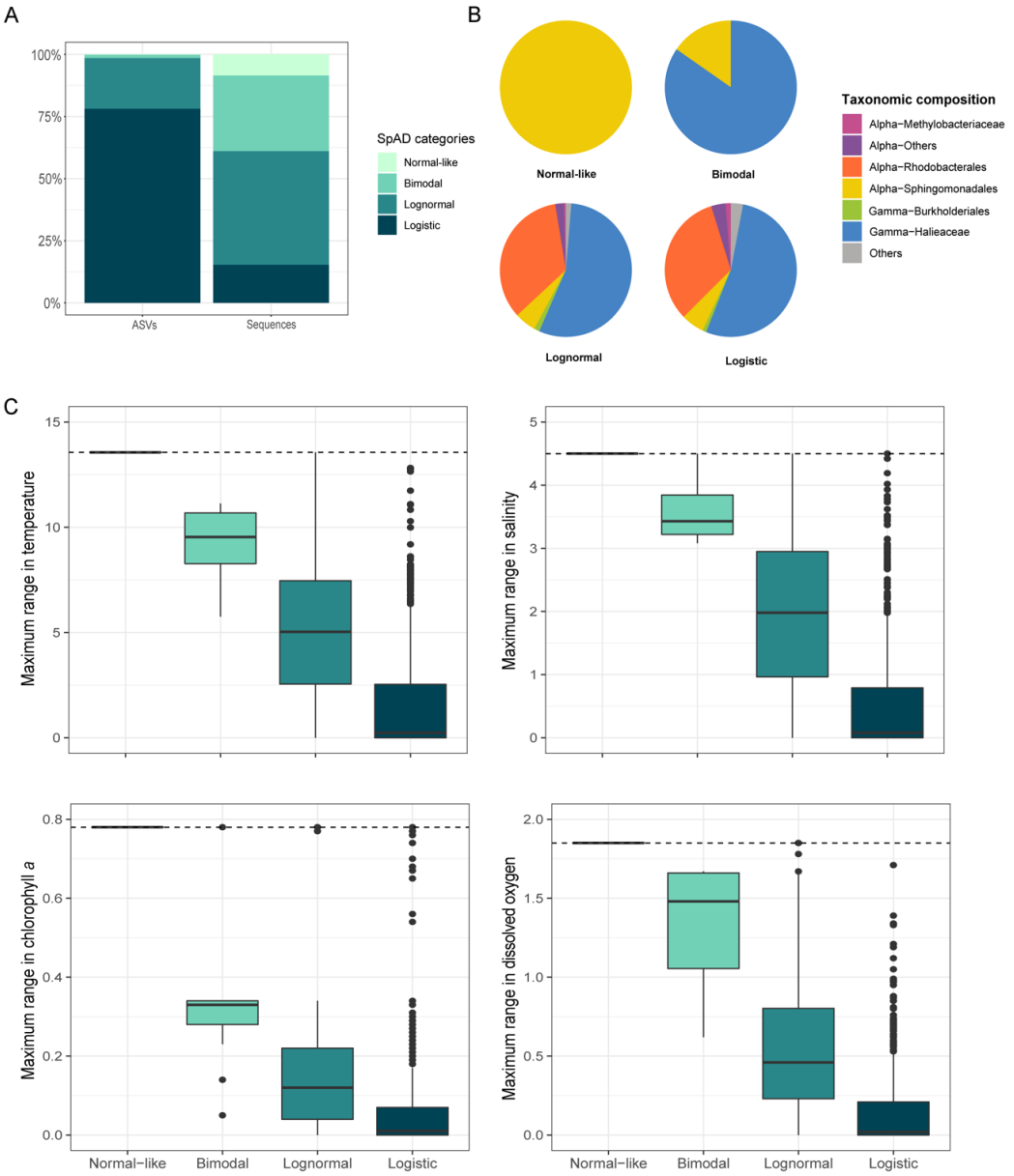

Figure S8. Spatial Abundance Distributions (SpADs) analysis. A) Percentage of ASVs and sequences within each SpAD category: normal-like (N=2 ASVs), bimodal (N=15), lognormal (N=228) and logistic (N=872). B) Pie charts showing the taxonomic composition of ASVs within each SpAD category. C) Environmental breadth of ASVs within each category regarding temperature, salinity, chlorophyll a, and dissolved oxygen concentrations. For each environmental variable, the maximum range is calculated considering the values of all samples in which an ASV appears, and calculating the range of those values. The dashed line represents the maximum range considering all samples.

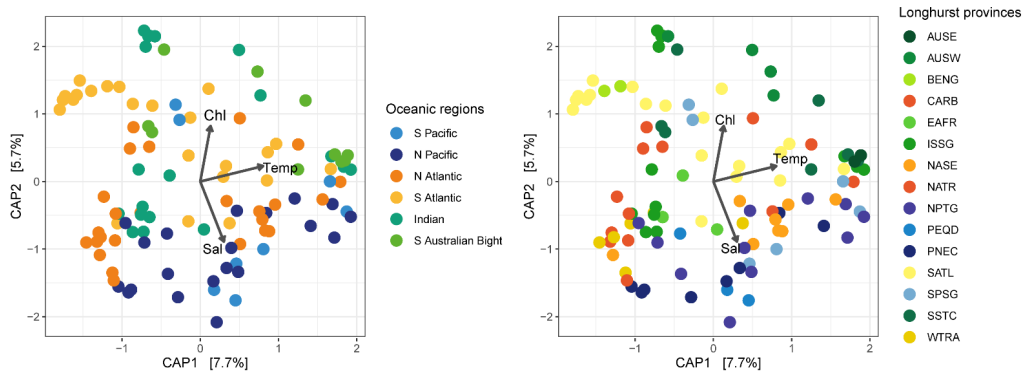

Figure S9. Distance-b2ased redundancy analysis (dbRDA) of all studied AAP communities. Samples are color-coded according to the oceanic region (A) or the Longhurst province (B) to which they belong. The environmental variables explaining community distribution are represented by arrows, where 'Temp' is temperature, 'Sal' is salinity, and 'Chl' is chlorophyll *a* concentration.

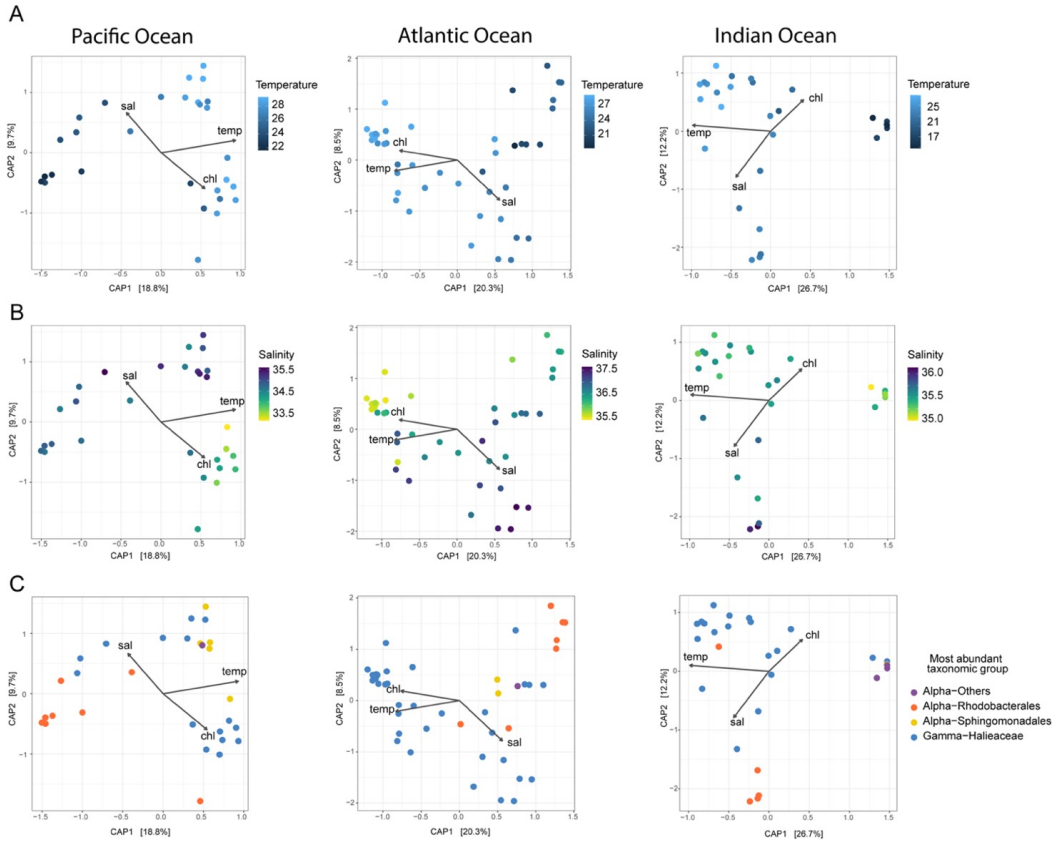

Figure S10. Distance-based redundancy analysis (dbRDA) for all samples separated by oceans. Stations are coloured according to their temperature (A) and salinity (B). In C) the colour represents the most abundant taxonomic group for that station.

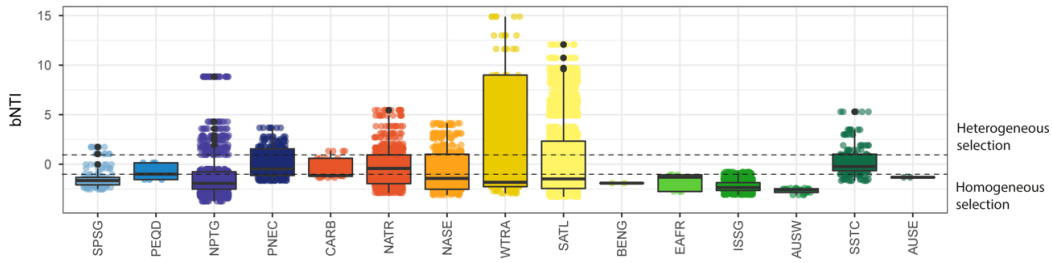

Figure S11. βNTI values for pairwise comparisons of samples located within the same Longhurst provinces. Dotted lines highlight βNTI values below −2 (which are associated to homogenous selection) and above +2 (associated to heterogeneous selection). Longhurst provinces are ordered following the cruise path.

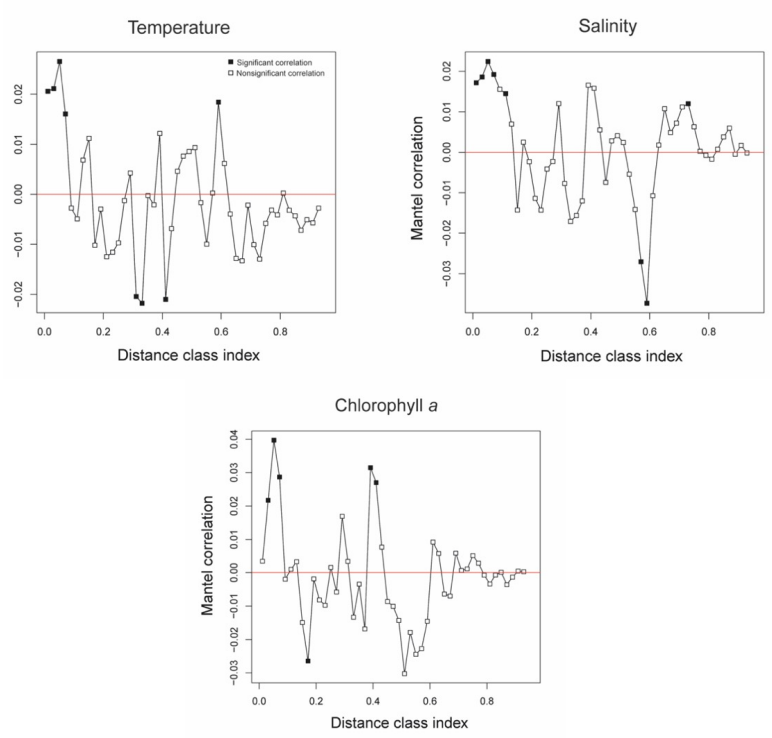

Figure S12. Mantel correlograms relating between-ASVs niche differences and between-ASVs phylogenetic distance across a given phylogenetic distance range and depicting a significant phylogenetic signal for the pufM gene marker across relatively short phylogenetic distances. The phylogenetic signal was tested using temperature, salinity, and chlorophyll *a*, the three variables that explained the highest fraction of the AAP community variance. Solid and open squares denote significant and nonsignificant (p=0.05) correlations, respectively. Correlations that are significantly positive indicate that the between-ASVs phylogenetic distance increases as their environmental niche distance increases for the specific phylogenetic range being evaluated.

1.8 Supplementary tables

Table S1. Pearson correlation values among environmental variables and AAP alpha diversity indices (i.e., Chao1 and Shannon). Only significant correlations are shown. False discovery rate (FDR)-correction was applied to all *p*-values.

Primer candidate name	Sequence	Length (bp)	Hybridization(%)		Tm (ºC)	GC %
			0 mismt	1-2 mismt		
pufMF_V	5'-GGS AAY CTS TWY TAY AAY CCV T-3'	22	54.3	93.3	53.6	44
pufMF_Y	5'-GGS AAY CTS TWY TAY AAY C-3'	19	55.5	94.62	47.5	42.1
pufM_F1A	5'- GAR GCN GTK CCN TWY GGN RT-3'	20	61	94	58.9	57.5
pufM_F1B	5'- GAR GCN GTK CCN TWY GGN AT-3'	20	57	93	58.9	57.5
pufM_F2	5'- TGG GCN TTY GCN DSN GCV AT-3'	20	60	75.5	62.9	60

Table S2. Significant environmental variables determining AAP community structure, after a PERMANOVA test with the adonis function and 999 permutations. Df, degrees of freedom; Sum of Sqs, sequential sum of squares; Mean Sqs, Mean Squares; F.Model, F value by permutation; R2, partial R-squared. Pr(>F), p-values based on 999 permutations (the lowest *p*-value possible is obtained).

	Df	SumsOfSq	MeanSqs	F.Model	R^2	Pr(>F)
Temperature	1	2.260	2.260	8.22	0.063	0.001
Chlorophyll *a*	1	1.739	1.739	6.32	0.049	0.001
Salinity	1	1.804	1.804	6.56	0.051	0.001
Virus	1	0.933	0.932	3.39	0.026	0.001
Cell volume	1	1.383	1.383	5.03	0.039	0.001
Synechococcus ab.*	1	1.099	1.099	3.99	0.031	0.001
Picoeukaryote ab.	1	0.189	0.189	0.69	0.005	0.83
Residuals	86	23.652	0.275		0.664	
Total	97	35.613			1	

*ab. = abundance.

Table S3. Analysis of variance based on Bray-Curtis dissimilarities, comparing the community structure within oceanic regions and Longhurst provinces. Performed with a PERMANOVA test with the adonis function and 999 permutations. Abbreviations in Table S3.

	Df	SumsOfSqs	MeanSqs	F.Model	R^2	Pr(>F)
Oceanic regions	5	8,341	1,67	6.0358	0.22	0.001
Residuals	107	29,572	0,27637	0.78		
Total	112	37,912	1			
Longhurst provinces	14	17.721	1,26579	6.1436	0.46742	0.001
Residuals	98	20.191	0.20603	0.53258		
Total	112	37.912	1			

CHAPTER II

CHAPTER II

A metagenomic and amplicon sequencing combined approach reveals the best primers to study marine aerobic anoxygenic phototrophs

Carlota R. Gazulla, Ana María Cabello, Pablo Sánchez, Josep M. Gasol, Olga Sánchez, Isabel Ferrera

Microbial Ecology (2023) DOI: https://doi.org/10.1007/s00248-023-02220-y

Abstract:

Studies based on protein-coding genes are essential to describe the diversity within bacterial functional groups. In the case of aerobic anoxygenic phototrophic (AAP) bacteria, the *pufM* gene has been established as the genetic marker for this particular functional group, although available primers are known to have amplification biases. We review here the existing primers for *pufM* gene amplification, design new ones, and evaluate their phylogenetic coverage. We then use samples from contrasting marine environments to evaluate their performance. By comparing the taxonomic composition of communities retrieved with metagenomics and with different amplicon approaches, we show that the commonly used PCR primers are biased towards the Gammaproteobacteria phylum and some Alphaproteobacteria clades. The metagenomics approach, as well as the use of other combinations of the existing and newly designed primers, show that these groups are in fact less abundant than previously observed, and a great proportion of *pufM* sequences are affiliated to uncultured representatives, particularly in the open ocean. Altogether, the framework developed here becomes a better alternative for future studies based on the *pufM* gene and, additionally, serves as a reference for primer evaluation of other functional genes.

2.1 Introduction

The open ocean microbiota consists of approximately 10^{29} organisms that perform key biogeochemical processes essential for ecosystem functioning (Whitman et al., 1998). However, only a small portion can be isolated and culture-independent techniques based on their genetic content are fundamental to study them. Indeed, the sequencing of the ribosomal 16S RNA gene allowed the first studies on the biogeography of marine bacterial communities (Hagström et al., 2002; Fuhrman et al., 2006; Pommier et al., 2007). In the last decade, the development of high-throughput sequencing (HTS) methods together with worldwide oceanic surveys (Karsenti et al., 2011; Duarte, 2015) have generated a massive amount of sequencing data obtained with standardized methodologies, which has facilitated studying the marine microbiome at an unprecedent scale and has elucidated patterns of prokaryotic diversity, interactions, and connectivity around the globe (e.g. Lima-Mendez et al., 2015; Sunagawa et al., 2015; Louca et al., 2016; Mestre et al., 2018; Logares et al., 2020; Ruiz-González et al., 2020; Herndl et al., 2022). While most of the efforts have been performed at the whole bacterioplankton community level, the focus on specific functional groups allows the identification of microorganisms involved in a wide range of functions, such as carbon and nitrogen fixation, ammonia oxidation or light harvesting, which are key to understand global biogeochemical cycles (Ferrera et al., 2015; Louca et al., 2016). Studies based on protein-coding genes are essential for this endeavor, since they may have experienced horizontal gene transfer (HGT) processes (Wiedenbeck & Cohan, 2011), and their phylogeny differs from the observed with the canonical ribosomal 16S RNA gene.

A polyphyletic group that has been extensively studied in the last two decades is that of the aerobic anoxygenic phototrophic (AAP) bacteria. Their discovery in the ocean surface (Kolber et al., 2000) implied a change of paradigm in our understanding of carbon cycling since they are heterotrophic organisms that can also obtain energy from light. Although they derive a fraction of their energy needs harvesting light using bacteriochlorophyll *a*, AAP bacteria are thought to be unable to fix inorganic carbon, relying thus on dissolved organic matter. Studies of the diversity of AAP communities are based on the sequencing of the *pufM* gene that encodes the M subunit of the AAP reaction center. The first versions of *pufM* primers were designed based on sequences from cultured bacteria (Nagashima et al., 1997; Achenbach et al., 2001; Béjà et al., 2002; Tank et al., 2009). A comparison of around 200 sequences from cultivated bacteria and environmental samples carried out by Yutin et al. (2005) indicated that the environmental sequences of the *pufM* gene had

a greater variability than the ones from cultured bacteria and these authors proposed new universal primers: pufM_uniF (forward) and pufM_uniR (reverse), hereafter called UniF and UniR, and an additional reverse primer called pufM_WAW. Although they were originally designed for marine environments, primers UniF and UniR were discarded in subsequent studies due to PCR amplification problems (e.g. Koh et al., 2011) and have mostly been used in freshwater ecosystems (Piwosz et al., 2020, 2022; Villena-Alemany et al., 2022). The combination of primers pufMF (forward), designed by Béjà et al. (2002), and pufM_WAW (reverse), designed by Yutin et al. (2005), was first proposed by Lehours et al. (2010) on the basis of their specificity and efficiency after testing multiple primer combinations. Since then, this combination has been used in posterior studies analyzing AAP communities in the marine environment. Most of these studies pictured AAP communities as mainly composed by Gammaproteobacteria and Alphaproteobacteria clades (Jiao et al., 2007; Lehours et al., 2010, 2018; Jeanthon et al., 2011; Ferrera et al., 2014; Bibiloni-Isaksson et al., 2016; Auladell et al., 2019; Gazulla et al., 2022). The few studies based on metagenomics showed however far more diversity and a large fraction of AAP assemblages composed of members with no cultured representatives. For example, Yutin et al. (2007), using the Global Ocean Sampling (GOS) metagenomic shotgun data, described several groups of AAP bacteria that were abundant in some specific areas of the ocean and that had hardly been recovered with amplicon-based methodologies. Some of these groups were also described to be abundant in samples from a Brazilian coastal bay using metagenomics (Cuadrat et al., 2016). A new group of AAP bacteria named "*Candidatus* Luxescamonaceae" (class Alphaproteobacteria), with a putative potential for carbon fixation, was described from the Tara Oceans metagenomic dataset (Graham et al., 2018). These differences among diversity surveys based on metagenomics and amplicon-sequencing approaches are likely due to primer biases (Polz & Cavanaugh, 1998). In fact, previous discussions regarding possible biases in *pufM* amplification argued that primers pufMF and pufM_WAW may overestimate some groups to the detriment of others (Yutin et al., 2005; Lehours et al., 2010; Ferrera et al., 2014; Auladell et al., 2019; Gazulla et al., 2022). While metagenomics overcomes some of the PCR limitations, it generally only retrieves the most abundant members of the bacterial community. In the case of a functional group like the AAP bacteria, whose relative abundance ranges between 0.1–10% of the total bacterioplankton (Koblížek, 2015), metagenomics can limit a comprehensive knowledge of AAPs diversity.

In this context, the aim of this study is to evaluate the performance of existing and newly designed primers of the *pufM* gene to characterize the diversity of AAPs in marine

samples. We employed several combinations of existing and novel *pufM* primers and determined their phylogenetic coverage. Then, using a selection of these primers, we compared the taxonomic composition of marine AAP communities based on amplicon sequencing vs. metagenomics. The combination of in silico tests and phylogenetic coverage analyses, together with its application to natural samples, allows us to propose the optimal combination of primers for future studies targeting the *pufM* gene. In addition, the approach developed here can serve as reference for future studies involving primer evaluation of functional genes.

2.2 Methods

Building a *pufM* database

We built a *pufM* gene database containing around 1300 sequences from isolates and metagenomes from marine environments. For that purpose, we downloaded 697 *pufM* sequences from the Genome Taxonomy Database (GTDB, https://gtdb.ecogenomic.org/, release 202) using AnnoTree (Mendler et al., 2019) and then added to the database sequences from metagenomic datasets such as those from the Tara Oceans Expedition (Karsenti et al., 2011), the Malaspina Expedition (Duarte, 2015), the Global Ocean Survey (GOS) (Yutin et al., 2007; Cuadrat et al., 2016), and the Blanes Bay Microbial Observatory (Auladell et al., 2019). The taxonomic assignation of the metagenomic sequences was based on a phylogenetic tree generated in Gazulla et al. (2022). We classified all sequences into different groups based on their class or order ranks and the phylogroups A to L, previously established by Yutin et al. (2007). These phylogroups were defined based on the *puf* operon organization and on the *pufM* gene phylogeny. Around 100 sequences could not be assigned to any phylogroup and were clustered together in the 'Others' group.

Phylogenetic coverage, primer design, and in vitro performance

To evaluate the phylogenetic coverage of the primers pufMF (Béjà et al., 2002), UniF, UniR, and pufM_WAW (Yutin et al., 2005), we aligned them against our *pufM* database using the *AlignTranslation* function in the Decipher R package (Wright, 2016). We visualized the alignment with Geneious Prime® 2021.1.1. and calculated the percentage of sequences showing between 0 and 7 mismatches for each primer region within the whole database. To further evaluate these primers, we calculated in silico parameters such as the mean melting temperature, GC content, and ΔG values for hairpin, self-dimer and hetero-dimer formation, with the OligoAnalyzer tool from IDT (Integrated DNA Technologies,

https://eu.idtdna.com/calc/analyzer). Based on the results of these analyses (Table 1), we decided to design forward primers that would combine with the reverse primer pufM_WAW, which showed good performance. The designing was carried out using the *DesignPrimers* function of the *Decipher* R package (Wright, 2016) and Geneious Prime® 2021.1.1. All primer proposals went through the same in-silico tests as the existing primers and their phylogenetic coverage was calculated as explained above. We came up with five candidates (Table S1) that were synthesized by ©Metabion International AG (https://www.metabion.com/). We performed PCR amplifications of the *pufM* gene using the different combinations of the newly designed primers as well as the primers from Béjà et al. (2002) and Yutin et al. (2005). DNA from the AAP strains *Congregibacter litoralis* (Gammaproteobacteria), *Sandarakinorhabdus limnophila* (Alphaproteobacteria), and from the purple sulfur bacteria *Allochromatium vinosum* (Gammaproteobacteria), all three containing the *pufM* gene, was used as positive controls, as well as several marine environmental samples from the Blanes Bay Microbial Observatory (BBMO) and the Alboran Sea, which were known to contain AAP bacteria. After several attempts of amplification using various PCR conditions (see Table S2), only the following set of primers showed positive results (forward/reverse): pufMF/pufM_WAW, UniF/UniR, and pufMF_Y/pufM_WAW. The forward primer pufMF_Y was the only candidate from our designed primer proposals that successfully amplified a gene fragment. We confirmed the amplification of *pufM* fragments using Sanger sequencing, performed at the Genomics Unit of the University of Málaga, Spain (www.scai.uma.es). The following material and methods refer only to these three combinations of primers.

DNA extraction, *pufM* amplification, sequencing, and sequence processing

To analyze the performance of these primers, we selected 17 environmental samples belonging to datasets whose AAP communities had previously been analyzed: 9 samples from a seasonal study of the BBMO (Auladell et al., 2019) and 8 samples from the surface global ocean Malaspina Expedition (Gazulla et al., 2022) (Table S3). For this subset of samples, both amplicon (with primers pufMF/pufM_WAW) and metagenomic data were available and they represented a comprehensive picture of the diversity at a seasonal and at a spatial scale. DNA extraction and amplification of the *pufM* gene with primers pufMF/pufM_WAW was done as explained in Auladell et al. (2019) and in Gazulla et al. (2022) in samples from the BBMO and the Malaspina Expedition, respectively. Primers pufMF_Y/pufM_WAW and UniF/UniR were used to amplify different size fragments of the *pufM* gene following the conditions described in Supplementary Information 1. Sequencing was performed in an Illumina MiSeq sequencer (2 × 250 bp) at the Research and Testing

Laboratory (http://rtlgenomics.com/). Noteworthy, amplification with primers UniF/UniR was only possible after a cleaning step performed in the sequencing house using the TaKaRa ExTaq DNA polymerase (TaKaRa Bio Inc., Shiga, Japan). Sequences of the *pufM* gene from the BBMO metagenomes were generated as described in Auladell et al. (2019). Those from the Malaspina Expedition were retrieved from the Malaspina gene catalog when annotated as any of the following: *pufM* (prokka 1.14.6; Seemann, 2014), K08928 for the Kyoto Encyclopedia of Genes and Genomes orthologs (KEGG; Kanehisa et al., 2014) and PF00124 (Protein Families; El-Gebali et al., 2019). Annotations were manually curated to filter out possible false positives. Sequences of the *pufM* gene from the metagenomic assemblies of each of the 8 Malaspina samples were extracted likewise. The generation of the Malaspina Gene Database and their annotation is described in Supplementary Information 2 and in Sánchez et al. (2023).

Sequence data processing and statistical analyses

Each amplicon dataset was processed separately with cutadapt v1.16 (Martin, 2013) to remove primers and spurious sequences, and with DADA2 v1.10 (Callahan et al., 2016) to differentiate exact sequence variants and remove chimeras (Supplementary Information 1). In total, we obtained three amplicon sequence variant (ASV) tables, each one corresponding to each primer set combination. Sample BL110412 from the BBMO dataset was discarded due to a low number of reads (103 reads in the UniF/UniR assay). In addition, to compare the performance of the different primer pairs, we joined the three amplicon datasets, by cutting all sequences to the same length (145 bp, the size of the smallest amplicon, obtained with primers UniF/UniR) and using the *mergeSequenceTables* function in DADA2, to analyze them as a single dataset. The phylogeny of the sequences was inferred using the phylogenetic tree from Gazulla et al. (2022) and the Evolutionary Placement Algorithm v0.3.5 (Barbera et al., 2019). Community composition, statistical analyses and figures were performed in R v4.2.0 (R Core Team 2022) using packages *phyloseq* (McMurdie & Holmes, 2013), *tidyverse* (Wickham et al., 2019), *vegan* (Oksanen et al., 2022), and *ggplot2* (Wickham, 2016) (see details in Supplementary Information 3).

Data availability

Most of the analyses were performed in R version 4.2.0 (R Core Team 2022) and code is available in the following repository: https://gitlab.com/crgazulla/analysing-pufm-primers-for-marine-aap-studies. The *pufM* database is available in the Supplementary Material. Amplicon sequences have been deposited in the NCBI Sequence Read Archive (SRA) under BioProject ID PRJNA919028. Sequences from the BBMO metagenomic dataset

were published in Auladell et al. (2019), and those from the Malaspina Expedition dataset are deposited under BioProject ID PRJEB52452.

2.3 Results

Phylogenetic coverage and PCR parameters of the existing primers for the amplification of the *pufM* gene

We examined the phylogenetic coverage of existing primers *pufMF* (Béjà et al., 2002), UniF, UniR, and pufM_WAW (Yutin et al., 2005) against an in-house built database of the *pufM* gene consisting of >1300 sequences from isolates and metagenomes, by calculating the number of mismatches for each position. Then, to determine the frequency of mismatches in different AAP assemblages, we classified our database into groups according to their taxonomic rank and into the phylogroups A to L previously established by Yutin et al. (2007) and commonly used in AAP diversity surveys (Figure 1). The forward primer pufMF had the highest number of mismatches, especially for phylogroups A, B, C, D, and G, for which most of the sequences had more than three mismatches in that primer region. In total, the pufMF primer showed perfect matches to only 21.6% of the sequences in our database. This primer has been commonly paired with the reverse primer pufM_WAW, which shows a better performance, and for which most of the sequences present zero mismatches (80.4%). On the other hand, primers UniF and UniR showed a higher coverage for all the taxonomic groups, with a hybridization ratio (zero mismatches) of 94.9% for UniF and 88.5% for UniR in our database. To further evaluate these primers, we calculated in silico parameters such as the mean melting temperature, GC content, and ΔG values for hairpin, self-dimer, and hetero-dimer formation individually for each primer (Table 1). All primers have similar characteristics in terms of length, GC content or degenerate nucleotides. However, the forward primer UniF stands out as the longest oligonucleotide, with a very low percentage of GC and a high number (ten) of degenerate nucleotides.

Design of new primers for the amplification of *pufM* gene

The design of new primers for the *pufM* gene was addressed from two perspectives: on one hand, we attempted to design primers for an upstream region, that combined with the reverse primer pufM_WAW, would generate longer amplicons and improve the taxonomic resolution. The pufM_WAW primer region is located at the end of the *pufM* gene (Figure 2A) and since the phylogenetic coverage is high (Figure 1, Table 1) it was a good candidate for a reverse primer. On the other hand, we intended to improve the

phylogenetic coverage of the primer pufMF and the in vitro performance of primer UniF, both hybridizing in the same conserved region of the gene, by revising their design and including modifications. For this purpose, we used the nucleotide and amino-acid alignment of the sequences, we analyzed the percentage of each nucleotide at each position, and we associated these changes to the different phylogroups when possible (Figure 2B). A total of five forward primer candidates (Table S1) were tested in vitro using cultures and environmental samples and by varying the PCR conditions (annealing temperature, Mg^{+2} concentration, and primer concentration, Table S2). Three primers were designed in upstream regions of the *pufM* gene while two were improved versions of the

Table 1. In silico parameters of primers pufMF (Béjà et al., 2002), pufM_WAW, UniF, UniR (Yutin et al., 2005), and pufMF_Y (this study). Abbreviations; Tm = mean melting temperature, mismt = mismatches. The hybridization percentage refers to the percentage of sequences from our database that hybridize with 0 mismatches (0 mismt) and 0, 1, or 2 mismatches (0–2 mismt). Degenerate nucleotides are underlined.

Primer name	Sequence	Length	%GC	Tm (°C)	Hybridization (%)		Hairpin*	Self-dimer*	Hetero-dimer *
					0 mismt	0–2 mismt			
pufMF	TAC GGS AAC CTG TWC TAC	18 bp	50%	62.9°C	21.59%	69.42%	-0.3	-5.05	-7.94***
pufM_WAW	AYN GCR AAC CAC CAN GCC CA	20 bp	60%	73°C	82.50%	91.01%	1.25	-12.32	-
UniF	GGN AAY YTN TWY TAY AAY CCN TTY CA	26 bp	36.5%	54.2°C	94.86%	99.9%	2.56	-12.67	-7.22***
UniR	YCC ATN GTC CAN CKC CAR AA	20 bp	52.5%	56.9°C	88.50%	95.75%	**	-10.66	-
pufMF_Y	GGS AAY CTS TWY TAY AAY C	19 bp	42.1	47.5°C	55.5%	94.62%	2.3	-10.53	-8.22***

* Maximum value of ΔG (kcal/mole)
** No structure found for this sequence
*** Hetero-dimer values are calculated for the following pairs of primers: pufMF/pufM_WAW, UniF/UniR, and pufMF_Y/pufM_WAW.

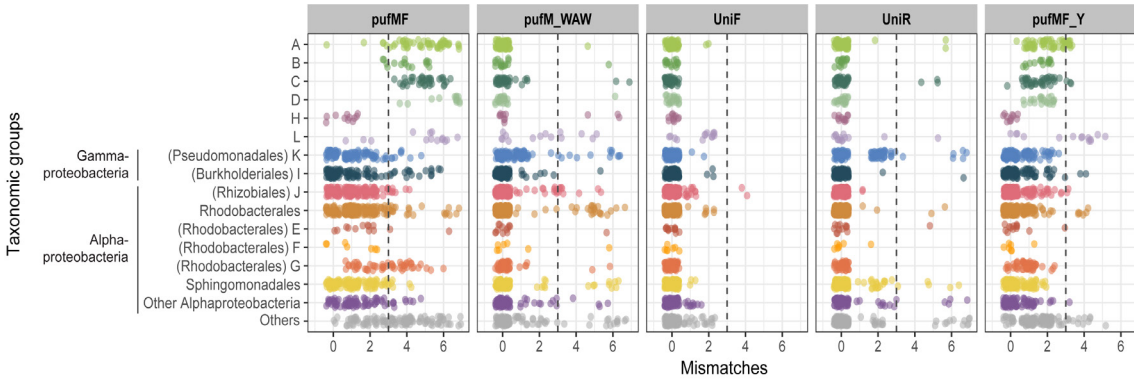

Figure 1. Mismatches of primers pufMF (Béjà et al., 2002), UniF, UniR, and pufM_WAW (Yutin et al., 2005), and pufMF_Y (this study) for the different AAP groups defined in this study. Dashed lines separate sequences with three or more mismatches.

existing pufMF primer. We successfully obtained one forward primer, named 'pufMF_Y' that combined with the existing reverse primer pufM_WAW, amplified a 203 bp-fragment of the *pufM* gene, while the others were discarded due to their poor performance. This primer hybridizes in the same conserved region as primers pufMF and UniF (Figure 2A), it improves the phylogenetic coverage of primer pufMF, and it has a lower number of degenerate nucleotides that primer UniF. Although the hybridization ratio is relatively low (55.45%) compared to primer UniF, when we consider sequences with zero, one or two mismatches (that are likely to amplify), the hybridization ratio increases up to 95%. Additionally, we performed PCR amplification with different combinations of the existing primers. Primers pufMF/pufM_WAW and UniF/UniR amplified fragments of 207 and 145 bp respectively. Although it has been used before in marine (Fecskeová et al., 2021) and freshwater ecosystems (Piwosz et al., 2022) , the combination of primers UniF/pufM_WAW did not result in positive amplification of our marine samples after several attempts under different PCR conditions.

Oceanographic context of the environmental samples used for primer comparison and sequencing results:

To test the performance of the different primers in natural samples, we compared the composition of AAP communities retrieved by metagenomics and that retrieved by amplicon sequencing using different primer combinations. Based on the in vitro results explained above, we selected primers pufMF/pufM_WAW, which have been routinely used in marine environments (e.g. Mašín et al., 2006; Lehours et al., 2010; Jeanthon et al., 2011; Ferrera et al., 2014; Auladell et al., 2019; Gazulla et al., 2022), primers UniF/UniR, mainly used in freshwater ecosystems (Piwosz et al., 2020; Villena-Alemany et al., 2022), and the newly designed forward primer pufMF_Y with the reverse pufM_WAW. Illumina sequencing with the three primer combinations was performed for 17 samples that covered both spatial and temporal variability: eight open ocean samples from the Malaspina Expedition (Duarte, 2015), from the Pacific, the Atlantic, and the Indian Oceans, and nine coastal samples from the Blanes Bay Microbial Observatory (BBMO) in the Mediterranean Sea collected at different seasons during years 2011 and 2012 (Gasol et al., 2016) (Figure 3A, Table S3). These samples were part of previous studies analyzing AAP communities by means of amplicon sequencing with primers pufMF/pufM_WAW: Malaspina samples in Gazulla et al. (2022) and BBMO samples in Auladell et al. (2019). Besides, we obtained 176 and 62 predicted genes from metagenomic assemblies associated to the *pufM* gene from the Malaspina and the BBMO metagenomic datasets, respectively. Predicted genes from Malaspina were between 101 bp and 1040

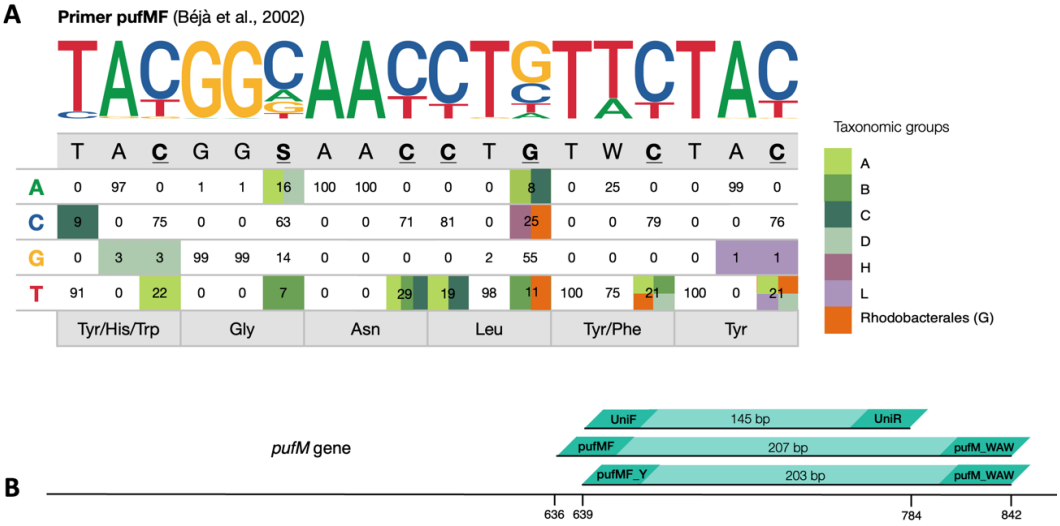

Figure 2. A) Sequence logo and phylogenetic coverage of primer pufMF (Béjà et al., 2002). The table shows the percentage of each nucleotide in each position of the primer, based on the alignment of our *pufM* database. Nucleotides representing positions with a high number of mismatches (≥ 19% of sequences) that could be associated with specific taxonomic groups (see color legend) are in bold and underlined. B) Schematic representation of the *pufM* gene and the primers used in this study.

datasets, respectively. Predicted genes from Malaspina were between 101 bp and 1040 bp (N50 = 804bp) in length, while genes from the BBMO were between 708 and 1044 bp long (N50 = 966bp). Regarding the amplicon analyses, for the primer combination pufMF_Y/pufM_WAW we obtained a total of 1904 ASVs, for the UniF/UniR we retrieved 1294 ASVs, and with primers pufMF/pufM_WAW we obtained a total of 418 ASVs. There were almost no shared ASVs between oceanic and coastal environments (~3%, see Table S4 for details). In terms of primer efficiency, we obtained 1.2 million reads with primers pufMF_Y/pufM_WAW, 0.5 million reads with primers UniF/UniR, and less than 0.3 million reads with primers pufMF/pufM_WAW. The high efficiency of primers pufMF_Y/pufM_WAW and the amplicon size of 203 bp (vs 145 bp with UniF/UniR primers) led to higher values of alpha-diversity for this primer combination (Figure S1). Communities amplified with primers pufMF_Y/pufM_WAW had the highest richness (mean 173.5 ± 0.27) values, followed by primers UniF/UniR, with a significantly lower mean observed diversity (125.1 ± 2.5; Tukey test, $p<0.05$, Figure S1). Instead, the mean Shannon index value (pufMF_Y/pufM_WAW, 3.58; UniF/UniR, 3.36 primer) was comparable for both approaches. Primers pufMF/pufM_WAW failed to amplify many sequences as compared to the previous primers, which resulted in a significantly lower observed diversity (mean 43.9 ± 0.3; Tukey test, $p<0.05$) and Shannon index values (mean 2.47; Tukey test, $p<0.05$, Figure S1).

We classified taxonomically all ASVs into 14 broad taxonomic groups, according to their order within the Alphaproteobacteria ('Rhizobiales', 'Rhodobacterales', 'Sphingomonadales' and 'Other Alphaproteobacteria') and the Gammaproteobacteria classes ('Burkholderiales' and 'Pseudomonadales', with sequences from the family Halieaceae). Besides, we also used the taxonomic groups proposed by Yutin et al. (2007) –A to L– to assign taxonomy to our dataset for comparison with previously published studies. Some of these phylogroups can be associated to known groups: phylogroup K contains sequences affiliated to the NOR5/OM60 clade, from family Halieaceae, order Pseudomonadales, and phylogroup I is related to the Burkholderiales order, both belonging to the Gammaproteobacteria. Phylogroup J has been associated to the Rhizobiales order, while phylogroups E, F, and G are associated to the Rhodobacterales order. Phylogroups A, B, C, D, H, and L have no taxonomically described representative. Finally, some of our ASVs could be associated to the recently described '*Candidatus* Luxescamonaceae' family. Since this family clustered within phylogroups C and D in our phylogenetic tree, we classified them as C or D depending on their position in the tree. Sequences that could not be further classified, were categorized as 'Others'.

Performance of the primers as compared to metagenomics in marine environmental samples.

The composition of AAP communities varied largely depending on the methodology (amplicon sequencing vs. metagenomics) and the primer combination in the amplicon approaches. In the Blanes Bay metagenomes, phylogroup K was always present and abundant in some months, while other groups peaked at specific times of the year (Figure 3B). This was the case of phylogroup A, whose relative abundance increased from <1% up to 30% during winter, or the Rhodobacterales, which increased after the spring bloom. In the case of the Malaspina metagenomes, the uncultured phylogroups A, B, C, and D dominated in all samples while Gammaproteobacteria and Alphaproteobacteria sequences were always present but scarce (Figure 3B).

The different amplicon approaches resulted in communities with very different taxonomic composition (Figure 3CDE; Adonis test, $p<0.05$), especially for primers pufMF/pufM_WAW. The communities described with this primer combination were dominated by phylogroup K (Pseudomonadales) and only samples BL110208, MP0311, MP1176, and MP1672 were dominated by ASVs associated to the Alphaproteobacteria class (Figure 3C). Noteworthy, we only retrieved three sequences with very low abundances of phylogroup C (in sample MP0778) and none from phylogroups A, B or D. Communities

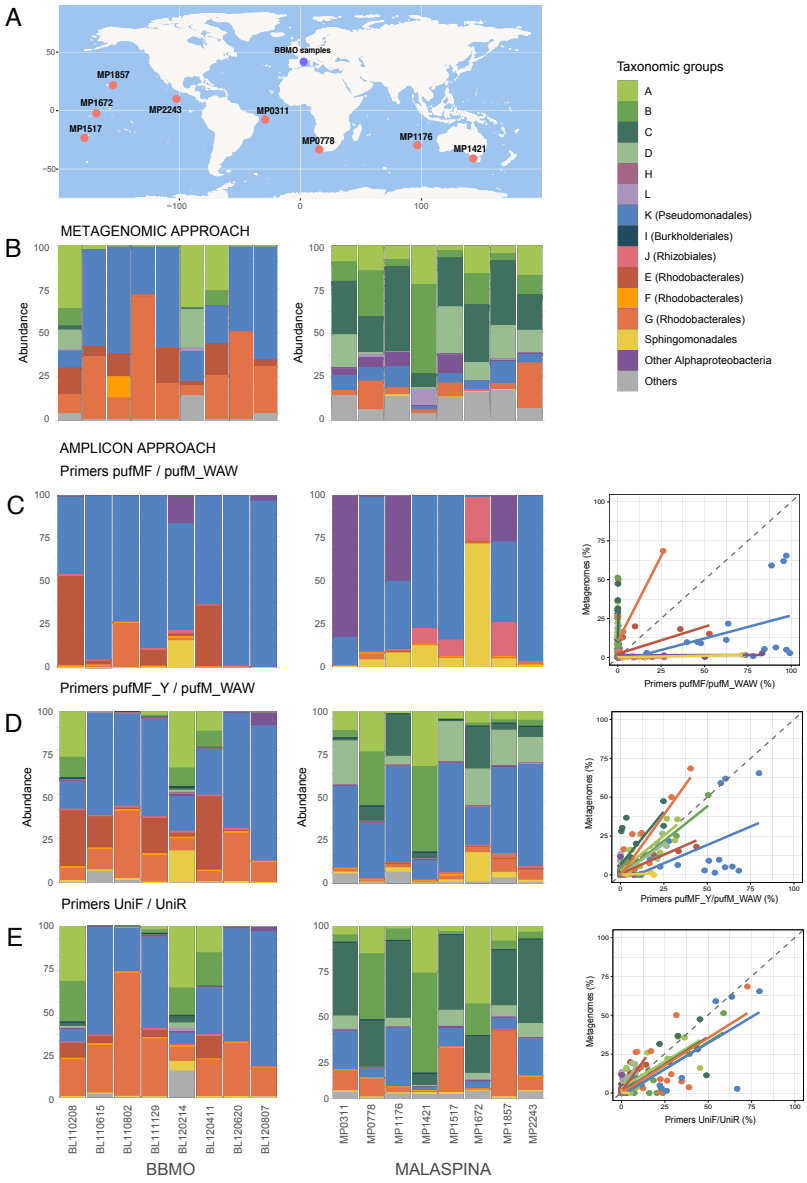

Figure 3. A) Stations from the Malaspina Expedition used in this study ("MP" code). Samples from Blanes Bay are all from the same coastal site, yet collected at different times of the year. The code of Blanes samples indicates Blanes-year-month-day (i.e., BL110208 is from the 8th of February 2011). B) *pufM* taxonomic composition of the Blanes Bay (left) and Malaspina (right) samples with the metagenomic approach. Below, the community composition at each station retrieved with the amplicon approach and with the different primers combinations: primers pufMF/pufM_WAW (C), pufMF_Y/pufM_WAW (D), and UniF/UniR (E). For each primer combination, and each sample, we have represented the relative abundances of the taxonomic groups retrieved with metagenomics vs those retrieved with the amplicon approach. The dashed lines represent the 1:1 lines in which both the metagenomic and the amplicon approach would indicate the same taxonomic community composition.

inferred using primers pufMF_Y/pufM_WAW were more diverse (Tukey test, $p < 0.05$) and contained more groups (Figure 3D). Although Pseudomonadales were still prevalent, especially in coastal samples, other groups appeared as relevant, such as phylogroups A and B and Rhodobacterales from phylogroups E and G. Finally, primers UniF/UniR also succeeded in amplifying phylogroups A, B, C, and D. The communities retrieved with these primers were similar to those described with primers pufMF_Y/pufM_WAW, albeit with some differences. In samples from Blanes Bay, phylogroup G was predominant within the Rhodobacterales. In turn, Malaspina samples were mainly dominated by phylogroups A, B, and C, while the relative abundance of Pseudomonadales was much lower.

To further compare the performance of the three primers pairs, we plotted the relative abundance of the different taxonomic groups retrieved with each amplicon approach vs. the metagenomic approach (Figure 3CDE) and summarize the phylogenetic coverage of the metagenomic and amplicon approaches (Figure S2). All in all, primer pufMF failed at amplifying groups with no cultured representatives (phylogroups A, B, C, and D), while it overestimated the abundance of Gammaproteobacteria and Rhizobiales. In turn, primers pufMF_Y/pufM_WAW and UniF/UniR better reflected the composition of the communities as observed with the metagenomic approach, assumed to be less biased than amplicon approaches. Although primers pufMF_Y/pufM_WAW seemed to overestimate the abundance of Gammaproteobacteria, especially in open ocean samples (Figure 3D), their performance was comparable to that observed with metagenomes (no statistical differences after Tukey test, $p>0.05$; Figure S2). In turn, the performance of primers UniF/UniR was quite good for the most abundant phylogroups.

Given the differences in the taxonomic composition of samples analyzed with different primers combinations (Figure S2, Tukey test, $p<0.05$ in seven phylogroups with primers pufMF/pufM_WAW), we wanted to test whether the community structure was conserved with primers UniF/UniR and pufMF_Y/pufM_WAW. To do so, we computed Bray-Curtis dissimilarity matrices for each amplicon approach to compare the structure of communities and performed Mantel test correlations. Besides, we Hellinger-transformed the data and applied a Procrustes test to assess the statistical significance between each ordination. The matrices obtained with the primers traditionally employed in AAP diversity surveys and the primer combinations proposed in this study were highly correlated (Mantel tests, Table S5). Besides, the Procrustes tests showed no statistical differences between their spatial ordination (Table S5). For further comparison of primer's performance, we combined all the samples into non-metric multidimensional scaling (NMDS) plots, first by

merging the samples from the three amplicon approaches into one matrix (Figure S3A), and then including the metagenomic data (Figure S3B). Samples from coastal and open ocean environments appeared in two clear distinct clusters (dispersion between samples is statistically significant, betadisper test, $p<0.001$, Figure S3D) for both plots. However, the clustering of samples based on the approach (metagenomics and different primer combinations) was less clear and not significant after testing the dispersion between samples (betadisper test, $p=0.0793$, Figure S3D). All in all, this indicates that even though the diversity and taxonomic composition varied depending on the primer, the community structure was conserved in the three approaches, and general ecological patterns could be observed indistinctively of the used primer.

2.4 Discussion

The study of key marker genes together with the development of the "-omics" techniques has increased our understanding of marine diversity and biogeochemical cycles (Ferrera et al., 2015; Grossart et al., 2020). Both metagenomics and amplicon approaches are commonly used to target functional genes and to describe their ecological patterns. Amplicon sequencing is easy to implement, relatively cheap, and effective in capturing large numbers of sequence variants. However, due to the high sequence variability of protein-coding genes, primer biases are common and can result in the misrepresentation of the relative contribution of certain taxa. In contrast, metagenomics is PCR-free and less biased for functional gene analysis, but it generally retrieves fewer copies of marker genes and hinders the comprehensive study of functional groups with low abundances in the environment. For example, the *nifH* gene, a genetic marker of nitrogen-fixing populations, is usually studied using amplicon sequencing since the number of variants retrieved with metagenomic surveys is too low, if not undetectable (Delmont et al., 2018; Cornejo-Castillo & Zehr, 2021). Likewise, most of the studies of AAP bacteria are based on the partial amplification of the *pufM* gene (e.g., Achenbach et al., 2001; Béjà et al., 2002; Karr et al., 2003; Oz et al., 2005; Ferrera et al., 2014; Bibiloni-Isaksson et al., 2016; Lehours et al., 2018; Auladell et al., 2019; Gazulla et al., 2022) while only few have approached their study based solely on metagenomics (Yutin et al., 2007; Cuadrat et al., 2016). Yet, the application of metagenomics allows the description of new metabolisms and new taxa in marine microbiomes, such as the discovery of new nitrogen fixation pathways in surface ocean heterotrophs (Delmont et al., 2018), and the description of new AAP phylogroups previously not reported (Yutin et al., 2007). In this study we combined both approaches to unveil the biases of existing primers for the *pufM* gene, design new ones, and test the

performance of different primer combinations in a variety of marine environments.

Debate regarding primer biases in the case of the *pufM* gene is not new, it arose in studies that used the primer pair pufMF/pufM_WAW and showed a large dominance of phylogroup K (Gammaproteobacteria) in AAP communities (Lehours et al., 2010; Ferrera et al., 2014; Auladell et al., 2019; Gazulla et al., 2022). Lehours et al. (2010) reported >80% of relative abundance of this phylogroup in samples from the Mediterranean Sea and considered possible primer biases favoring the amplification of the Gammaproteobacteria but disregarded that problem because the same primer pair in Arctic waters led to low abundances of phylogroup K. Still, it is likely that arctic bacterial communities have a very different composition, due to temperature differences and the effect of ice melt in salinity. Gammaproteobacteria were also predominant in samples from the coastal Blanes Bay (Ferrera et al., 2014; Auladell et al., 2019). Although the seasonal trends of communities retrieved by both metagenomics and amplicon sequencing had similar trends for the predominant groups, primers pufMF/pufM_WAW overestimated the contribution of phylogroup K and failed to amplify sequences of some groups that appeared in the metagenomic approach (Auladell et al., 2019). In addition to the prevalence of Gammaproteobacteria in marine AAP communities, most studies also reported the presence of members of the Alphaproteobacteria (Mašín et al., 2006; Lehours et al., 2010; Boeuf et al., 2013; Ferrera et al., 2014; Lehours & Jeanthon, 2015; Bibiloni-Isaksson et al., 2016; Auladell et al., 2019; Gazulla et al., 2022). Our results show that for the pufMF primer, the number of mismatches of sequences affiliated to Gammaproteobacteria and some Alphaproteobacteria orders are low in comparison to the mismatches in phylogroups A, B, C, and D, that are almost absent in amplicon-based studies (Figure 1). In fact, these phylogroups were described for the first time following the shotgun metagenomic sequencing approach of Yutin et al. (2007). They have no cultured representatives and even though they were abundant in samples from the Global Ocean Expedition (GOS; Yutin et al., 2007), they have been barely retrieved in other studies, which might be explained by the high number of mismatches within the region of the commonly used pufMF primer (Figure 1, 2B). On the contrary, the universal primers UniF, UniR, and pufM_WAW (Yutin et al., 2005) have a very good phylogenetic coverage (Figure 1). While the reverse primer pufM_WAW has been extensively used in combination with pufMF, primers UniF/UniR have barely been used in marine environments and were discarded in previous studies as they repeatedly failed in amplifying under different conditions (Lehours et al., 2010; Koh et al., 2011; Ferrera et al., 2014; and authors unpublished observations). Both primers, pufMF and UniF, hybridize in the same region of the *pufM*

gene (with 3-nucleotides shift between them), but UniF has 10 degenerate nucleotides, a very low GC content, and low melting temperature (Table 1), which might explain why it is problematic in vitro.

The primers designed in this study aimed at both improving the coverage of the commonly used ones and producing longer amplicons. While we did not succeed in the design of primers in the upstream region, we were able to produce an oligonucleotide with a higher hybridization ratio than pufMF, and a lower number of degenerate nucleotides than UniF, while maintaining the amplicon size of ~200 bp. The detailed analysis of the nucleotide composition of primer pufMF (Figure 2b) indicate that most of the mutations in the primer region can be associated to different phylogroups and generally represent silent mutations. The only exceptions are sequences from phylogroup C and D that have a histidine and a tryptophan respectively, instead of a tyrosine in the first position (of the primer region), and sequences from phylogroup D that encode for a tyrosine instead of a phenylalanine in position five. We identified 7 nucleotides with problematic mismatches, represented in bold and underlined in Figure 2, that were considered when redesigning the new primer pufMF_Y. Nevertheless, we cannot discard that other regions or combinations might work, nor that designs of primers that include both the *pufL* (upstream gene that encodes the L subunit of the reaction center) and the *pufM* gene might produce better tools.

Finding universally conserved regions in a functional gene is challenging and sometimes it is not possible to generate universal primers, as it happens, among others, with genes *nirS* and *nirK* (NO-forming nitrite reductase genes) that rely on the use of clade-specific primers (Bonilla-Rosso et al., 2016). Although the use of universal primers is appropriate to describe microbial communities, even perfectly matched primers can exhibit preferential amplification, thus beyond the in silico testing, analyses with environmental samples is also important for primer evaluation (Parada et al., 2016). To test different primer combinations, we used samples from different marine environments (coastal vs open ocean) and from different seasons, to include the spatial and seasonal variability that exists in AAP assemblages, as reported previously (Yutin et al., 2007; Lehours et al., 2018; Auladell et al., 2019; Gazulla et al., 2022). The metagenomic assay provided a bias-free representation of the most abundant components of AAP communities in samples from Blanes Bay and Malaspina (Figure 3B). Even though the number of copies retrieved was low, comparing these communities to those obtained through amplicon-sequencing was the best approach to analyze the biases of each primer combination. A

previous analysis with samples from Blanes Bay already identified discrepancies in the taxonomic composition of communities with the different methods, such as sequences from phylogroups A, B, and C that were only retrieved by metagenomics and were absent in the amplicon approach (Auladell et al., 2019). In this study, using the same extracted DNA from BBMO samples and different primer combinations, we were able to amplify sequences from these phylogroups, which in fact constitute over 50% of the relative abundance in two samples (Figure 3DE). Likewise, we obtained a great proportion of these groups in samples from Malaspina, that were completely missed with primers pufMF/pufM_WAW, as shown in Figure 3C and in Gazulla et al. (2022). These results are consistent with the reports from Yutin et al. (2007) and Cuadrat et al. (2016), in which they describe a high proportion of sequences affiliated to phylogroups A, B, C, and D in AAP communities from different marine environments (Graham et al., 2018). The alpha diversity estimates obtained with primers pufMF_Y/pufM_WAW and UniF/UniR (Figure S1) surpassed by far the estimates described in samples from the BBMO (Auladell et al., 2019) and the Malaspina Expedition (Gazulla et al., 2022) with primers pufMF/pufM_WAW. Interestingly, these previous studies had provided the most comprehensive datasets for AAP bacteria but were clearly underestimating AAP diversity.

Overall, our results indicate that the taxonomic composition of primers pufMF/pufM_WAW is biased towards phylogroup K (Pseudomonadales) (Figure 3C), that is overestimated in almost all samples. The same happens with Sphingomonadales-like and Rhizobiales representatives, and with a small cluster of other Alphaproteobacteria. Previous studies reported high abundances of Gammaproteobacteria in the Mediterranean Sea (Lehours et al., 2010; Ferrera et al., 2014; Auladell et al., 2019), the Baltic Sea (Mašín et al., 2006), the North Pacific Ocean (Boeuf et al., 2013), the Arctic Sea (Lehours & Jeanthon, 2015), the east coast of Australia (Bibiloni-Isaksson et al., 2016), or the tropical and subtropical global ocean (Gazulla et al., 2022). Since these studies analyzed AAP communities with the pufMF/pufM_WAW primers, it is likely that some of these results misrepresent the real composition of AAP communities, just as we have shown for samples of the Malaspina Expedition or Blanes Bay. Yet, albeit the exposed primer biases, the community structure of the different approaches was conserved in different ordination tests and the matrices strongly correlated, so previously published results (e.g., the adaptation of different phylogenetic clades (Lehours et al., 2018), their seasonal trends (Auladell et al., 2019), or the ecological processes operating in the surface global ocean (Gazulla et al., 2022) should not be discarded.

To conclude, we used an extensive *pufM* database to show the limitations of the forward primer pufMF and propose some alternatives to determine the composition and diversity of AAP communities in marine environments. We revised existing primers for the *pufM* gene in the literature, designed new ones and selected those with the best performance that were tested with environmental samples and benchmarked against metagenomics. We show that the phylogenetic coverage of primer pufMF is very low for some taxonomic groups and, as a result, amplification with this primer is biased towards phylogroup K (Gammaproteobacteria) and some orders of the Alphaproteobacteria class. Although Gammaproteobacteria are relevant components of AAP communities, several species with no cultured representatives, like phylogroups A, B, C, and D, have been entirely underrepresented in the past and are in fact an important fraction of AAP assemblages. For future studies analyzing marine AAP bacteria, we recommend using either primers pufMF_Y/pufM_WAW or UniF/UniR, to guarantee an unbiased representation of their taxonomic composition.

2.5 Acknowledgments

The authors thank the Group on Ecology of Plankton and Environmental Challenges and the Laboratory of Genetics and Molecular Biology for sharing their facilities at the 'Centro Oceanográfico de Málaga' (COMA, IEO-CSIC), where the experimental work of this manuscript was performed. We also thank Prof. Michal Koblížek for providing a culture of *Congregibacter litoralis* and Dr. Frederic Gich for doing so with *Sandarakinorhabdus limnophila*. We would also like to thank Dr. Adrià Auladell for help with preliminary sequence analyses. Computing analyses were performed at the Marine Bioinformatics Core facility (MARBITS, https://marbits.icm.csic.es) of the Institut de Ciències del Mar (ICM-CSIC) in Barcelona, Spain.

2.6 References

Achenbach, L.A., Carey, J., & Madigan, M.T. (2001) Photosynthetic and Phylogenetic Primers for Detection of Anoxygenic Phototrophs in Natural Environments. *Appl Environ Microbiol* 67: 2922–2926.

Auladell, A., Sánchez, P., Sánchez, O., Gasol, J.M., & Ferrera, I. (2019) Long-term seasonal and interannual variability of marine aerobic anoxygenic photoheterotrophic bacteria. *ISME J* 13: 1975–1987.

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. *Syst Biol* 68: 365–369.

Béjà, O., Suzuki, M.T., Heidelberg, J.F., Nelson, W.C., Preston, C.M., Hamada, T., et al. (2002) Unsuspected diversity among marine aerobic anoxygenic phototrophs. *Nature* 415: 630–633.

Bibiloni-Isaksson, J., Seymour, J.R., Ingleton, T., van de Kamp, J., Bodrossy, L., & Brown, M. V. (2016) Spatial and temporal variability of aerobic anoxygenic photoheterotrophic bacteria along the east coast of Australia. *Environ Microbiol* 18: 4485–4500.

Boeuf, D., Cottrell, M.T., Kirchman, D.L., Lebaron, P., & Jeanthon, C. (2013) Summer community structure of aerobic anoxygenic phototrophic bacteria in the western Arctic Ocean. *FEMS Microbiol Ecol* 85:417–432. https:// doi. org/ 10. 1111/ 1574- 6941. 12130

Bonilla-Rosso, G., Wittorf, L., Jones, C.M., and Hallin, S. (2016) Design and evaluation of primers targeting genes encoding NO-forming nitrite reductases: Implications for ecological inference of denitrifying communities. *Sci Rep* 6: 1–8. https:// doi. org/ 10. 1038/ srep39208

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. *Nat Methods* 13:7. https://doi.org/10.1038/nmeth.3869

Cornejo-Castillo, F.M. & Zehr, J.P. (2021) Intriguing size distribution of the uncultured and globally widespread marine non-cyanobacterial diazotroph Gamma-A. *ISME J* 15: 124–128. https:// doi.org/10.1038/s41396- 020- 00765-1

Cuadrat, R., Ferrera, I., Grossart, H., & Dávila, A. (2016) Picoplankton Bloom in Global South ? A High Fraction of Aerobic Anoxygenic Phototrophic Bacteria in Metagenomes from a Coastal Bay (Arraial do Cabo—Brazil). *OMICS* 20:76–87. https://doi.org/10.1089/omi.2015.0142

Delmont, T.O., Quince, C., Shaiber, A., Esen, Ö.C., Lee, S.T., Rappé, M.S., et al. (2018) Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nat Microbiol* 3: 804–813. https://doi.org/10.1038/s41564-018-0176-9

Duarte, C.M. (2015) Seafaring in the 21st century: The Malaspina 2010 circumnavigation expedition. *Limnol Oceanogr Bull* 24: 11–14. https://doi.org/10.1002/lob.10008

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., et al. (2019) The Pfam protein families database in 2019. *Nucleic Acids Res* 47: D427–D432.

Fecskeová, L.K., Piwosz, K., Šantic, D., Šestanovic, S., Vrdoljak, A.T., Hanusová, M., et al. (2021) Lineage-Specific growth curves document large differences in response of individual groups of marine bacteria to the top-down and bottom-up controls. *mSystems* 6:e00934-21. https://doi.org/10.1128/mSystems.00934-21

Ferrera, I., Borrego, C.M., Salazar, G., & Gasol, J.M. (2014) Marked seasonality of aerobic anoxygenic phototrophic bacteria in the coastal NW Mediterranean Sea as revealed by cell abundance, pigment concentration and pyrosequencing of *pufM* gene. *Environ Microbiol* 16: 2953–2965. https://doi.org/10.1111/1462-2920.12278

Ferrera, I., Sebastian, M., Acinas, S.G., & Gasol, J.M. (2015) Prokaryotic functional gene diversity in the sunlit ocean: Stumbling in the dark. *Curr Opin Microbiol* 25: 33–39. https://doi.org/ 10.1016/j.mib.2015.03.007

Fuhrman, J.A., Hewson, I., Schwalbach, M.S., Steele, J.A., Brown, M. V., & Naeem, S. (2006) Anually reocurring bacterial communities are predictable from ocean conditions. *Proc Natl Acad Sci U S A* 103: 13104–13109. https://doi.org/10.1073/pnas.0602399103

Gasol, J.M., Cardelús, C., Morán, X.A.G., Balagué, V., Forn, I., Marrasé, C., et al. (2016) Seasonal patterns in phytoplankton photosynthetic parameters and primary production at a coastal NW Mediterranean site. *Sci Mar* 80:63–77 https://doi.org/10.3989/scimar.04480.06E

Gazulla, C.R., Auladell, A., Ruiz-González, C., Junger, P.C., Royo-Llonch, M., Duarte, C.M., et al. (2022) Global diversity and distribution of aerobic anoxygenic phototrophs in the tropical and subtropical oceans. *Environ Microbiol* 24: 2222–2238.

Graham, E.D., Heidelberg, J.F., & Tully, B.J. (2018) Potential for primary productivity in a globally-distributed bacterial phototroph. *ISME J* 12: 1861–1866. https://doi.org/10.1038/s41396-018-0091-3

Grossart, H.P., Massana, R., McMahon, K.D., & Walsh, D.A. (2020) Linking metagenomics to aquatic microbial ecology and biogeochemical cycles. *Limnol Oceanogr* 65: S2–S20. https://doi.org/10.1002/lno.11382

Hagström, Å., Pommier, T., Rohwer, F., Simu, K., Stolte, W., Svensson, D., & Zweifel, U.L. (2002) Use of 16S ribosomal DNA for delineation of marine bacterioplankton species. *Appl Environ Microbiol* 68: 3628–3633. https://doi.org/10.1128/AEM.68.7.3628-3633.2002

Herndl, G.J., Bayer, B., Baltar, F., & Reinthaler, T. (2022) Prokaryotic Life in the Deep Ocean's Water Column. *Ann Rev Mar Sci* 15:1. https://doi.org/10.1146/annurev-marine-032122-115655

Jeanthon, C., Boeuf, D., Dahan, O., Le Gall, F., Garczarek, L., Bendif, E.M., & Lehours, A.C. (2011) Diversity of cultivated and metabolically active aerobic anoxygenic phototrophic bacteria along an oligotrophic gradient in the Mediterranean Sea. *Biogeosciences* 1955–1970. https://doi.org/10.5194/bg-8-1955-2011

Jiao, N., Zhang, Y., Zeng, Y., Hong, N., Liu, R., Chen, F., & Wang, P. (2007) Distinct distribution pattern of abundance and diversity of aerobic anoxygenic phototrophic bacteria in the global ocean. *Environ Microbiol* 9:3091–3099. https://doi.org/10.1111/j.1462-2920.2007. 01419.x

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2014) Data, information, knowledge and principle: Back to metabolism in KEGG. *Nucleic Acids Res* 42: 199–205. https://doi.org/10.1093/nar/gkt1076

Karr, E.A., Sattley, W.M., Jung, D.O., Madigan, M.T., & Achenbach, L.A. (2003) Remarkable diversity of phototrophic purple bacteria in a permanently frozen Antarctic lake. *Appl Environ Microbiol* 69: 4910–4914.

Karsenti, E., Acinas, S.G., Bork, P., Bowler, C., de Vargas, C., Raes, J., et al. (2011) A holistic approach to marine Eco-systems biology. *PLoS Biol* 9: 7–11. https://doi.org/10.1371/journal.pbio.1001177

Koblížek, M. (2015) Ecology of aerobic anoxygenic phototrophs in aquatic environments. *FEMS Microbiol Rev* 39:854–870. https://doi.org/10.1093/femsre/fuv032

Koh, E.Y., Phua, W., & Ryan, K.G. (2011) Aerobic anoxygenic phototrophic bacteria in Antarctic sea ice and seawater. Environ Microbiol Rep 3: 710–716.

Kolber, Z.S., Van Dover, C.L., Niederman, R.A., & Falkowski, P.G. (2000) Bacterial photosynthesis in surface waters of the open ocean. *Nature* 407: 177–179.

Lehours, A. & Jeanthon, C. (2015) The hydrological context determines the beta-diversity of aerobic anoxygenic phototrophic bacteria in European Arctic seas but does not favor endemism.

Front Microbiol 6:1–9. https://doi.org/10.3389/fmicb.2015.00638

Lehours, A.C., Cottrell, M.T., Dahan, O., Kirchman, D.L., & Jeanthon, C. (2010) Summer distribution and diversity of aerobic anoxygenic phototrophic bacteria in the Mediterranean Sea in relation to environmental variables. *FEMS Microbiol Ecol* 74: 397–409.

Lehours, A.C., Enault, F., Boeuf, D., & Jeanthon, C. (2018) Biogeographic patterns of aerobic anoxygenic phototrophic bacteria reveal an ecological consistency of phylogenetic clades in different oceanic biomes. *Sci Rep* 8:1–10. https://doi.org/10.1038/s41598-018-22413-7

Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., et al. (2015) Determinants of community structure in the global plankton interactome. *Science* 348(6237):1262073. https://doi.org/10.1126/science.1262073

Logares, R., Deutschmann, I.M., Junger, P.C., Giner, C.R., Krabberød, A.K., Schmidt, T.S.B., et al. (2020) Disentangling the mechanisms shaping the surface ocean microbiota. *Microbiome* 8: 1–17. https://doi.org/10.1186/s40168-020-00827-8

Louca, S., Wegener Parfrey, L., & Doebeli, M. (2016) Decoupling function and taxonomy in the global ocean microbiome. *Science* (1979) 353: 1272–1277.

Martin, M. (2013) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17.1: 10.

Mašín, M., Zdun, A., Stoń-Egiert, J., Nausch, M., Labrenz, M., Moulisová, V., & Koblížek, M. (2006) Seasonal changes and diversity of aerobic anoxygenic phototrophs in the Baltic Sea. Aquatic *Microbial Ecology* 45: 247–254.

McMurdie, P.J. & Holmes, S. (2013) Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:4. https://doi.org/10.1371/journal.pone.0061217

Mendler, K., Chen, H., Parks, D.H., Lobb, B., Hug, L.A., & Doxey, A.C. (2019) Annotree: Visualization and exploration of a functionally annotated microbial tree of life. *Nucleic Acids Res* 47: 4442–4448.

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. *Proc Natl Acad Sci U S A* 115: E6799–E6807.

Nagashima, K.V.P., Hiraishi, A., Shimada, K., & Matsuura, K. (1997) Horizontal transfer of genes coding for the photosynthetic reaction centers of purple bacteria. *J Mol Evol* 45: 131–136.

Oksanen, J., Simpson, G.L., Blanchet, F., Kindt, R., Legendre, P., Minchin, P., et al. (2022) vegan: Community Ecology Package_. R package version 2.6–2. https://cran.r-project.org/package=vegan.

Oz, A., Sabehi, G., Koblízek, M., Massana, R., & Béjà, O. (2005) *Roseobacter*-like bacteria in Red and Mediterranean Sea aerobic anoxygenic photosynthetic populations. *Appl Environ Microbiol* 71: 344–353.

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. *Environ Microbiol* 18: 1403–1414.

Piwosz, K., Villena-Alemany, C., & Mujakić, I. (2022) Photoheterotrophy by aerobic anoxygenic

bacteria modulates carbon fluxes in a freshwater lake. *ISME* 16:1046–1054. https://doi.org/10.1038/s41396-021-01142-2

Piwosz, K., Vrdoljak, A., Frenken, T., & Manuel, J. (2020) Light and primary production shape bacterial activity and community composition of aerobic anoxygenic phototrophic bacteria in a microcosm experiment. *mSphere* 5:4. https://doi.org/10.1128/mSphere.00354-20

Polz, M.F. & Cavanaugh, C.M. (1998) Bias in template-to-product ratios in multitemplate PCR. *Appl Environ Microbiol* 64: 3724–3730.

Pommier, T., Canbäck, B., Riemann, L., Boström, K.H., Simu, K., Lundberg, P., et al. (2007) Global patterns of diversity and community structure in marine bacterioplankton. *Mol Ecol* 16: 867–880.

Ruiz-González, C., Mestre, M., Estrada, M., Sebastián, M., Salazar, G., Agustí, S., et al. (2020) Major imprint of surface plankton on deep ocean prokaryotic structure and activity. *Mol Ecol* 29: 1820–1838. https://doi.org/10.1111/mec.15454

Sánchez, P., Sebastián, M., Pernice, M., Rodríguez-Martínez, R., Pesant, S., Agustí, S., et al. (2023) Marine Picoplankton Metagenomes From Eleven Vertical Profiles Obtained By the Malaspina Expedition in the Tropical and Subtropical Oceans. bioRxiv. https://doi.org/10.1101/2023.02.06.526790

Seemann, T. (2014) Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 30: 2068–2069.

Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G., et al. (2015) Structure and function of the global ocean microbiome. *Science* 348:6237 https://doi.org/10.1126/science.1261359

Tank, M., Thiel, V., & Imhoff, J.F. (2009) Phylogenetic relationship of phototrophic purple sulfur bacteria according to *pufL* and *pufM* genes. *Intern Microbiol* 12:175–185. https://doi.org/10.2436/20.1501.01.96

Villena-Alemany, C., Mujakić, I., Porcal, P., Koblížek, M., & Piwosz, K. (2022) Diversity dynamics of aerobic anoxygenic phototrophic bacteria in a freshwater lake. *Environ Microbiol Rep* 15:60–71.https://doi.org/10.1111/1758-2229.3131

Whitman, W.B., Coleman, D.C., & Wiebe, W.J. (1998) Prokaryotes: The unseen majority. *Proc Natl Acad Sci USA* 95:6578–6583. https://doi.org/10.1038/s41437-017-0034-1

Wickham, H. (2016) ggplot2: Elegant graphics for data analysis. Springer-Verlag, New York. https://ggplot2.tidyverse.org.

Wickham, H., Averick, M., & Bryan, J. (2019) Welcome to the Tidyverse. *J Open Source Softw* 4:1686. https://doi.org/10.21105/joss.01686

Wiedenbeck, J. & Cohan, F.M. (2011) Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiol Rev* 35: 957–976.

Wright, E.S. (2016) Using DECIPHER v2.0 to analyze big biological sequence data in R. R Journal 8: 352–359. https://doi.org/10.32614/rj-2016-025

Yutin, N., Suzuki, M.T., & Béjà, O. (2005) Novel primers reveal wider diversity among marine aerobic anoxygenic phototrophs. *Appl Environ Microbiol* 71: 8958–8962.

Yutin, N., Suzuki, M.T., Teeling, H., Weber, M., Venter, J.C., Rusch, D.B., & Béjà, O. (2007) As-

sessing diversity and biogeography of aerobic anoxygenic phototrophic bacteria in surface waters of the Atlantic and Pacific Oceans using the Global Ocean Sampling expedition metagenomes. *Environ Microbiol* 9: 1464–1475.

2.7 Supplementary figures

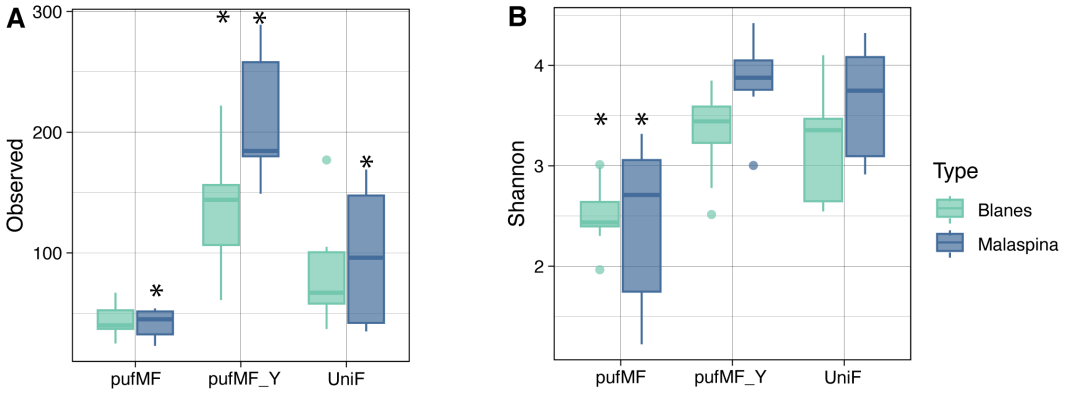

Figure S1. Alpha diversity as observed richness (A) and Shannon index (B) for each primer combination. * Asterisks indicate groups that are statistically different from the others, after an analysis of variance and a post-hoc Tukey test (* $p<0.05$).

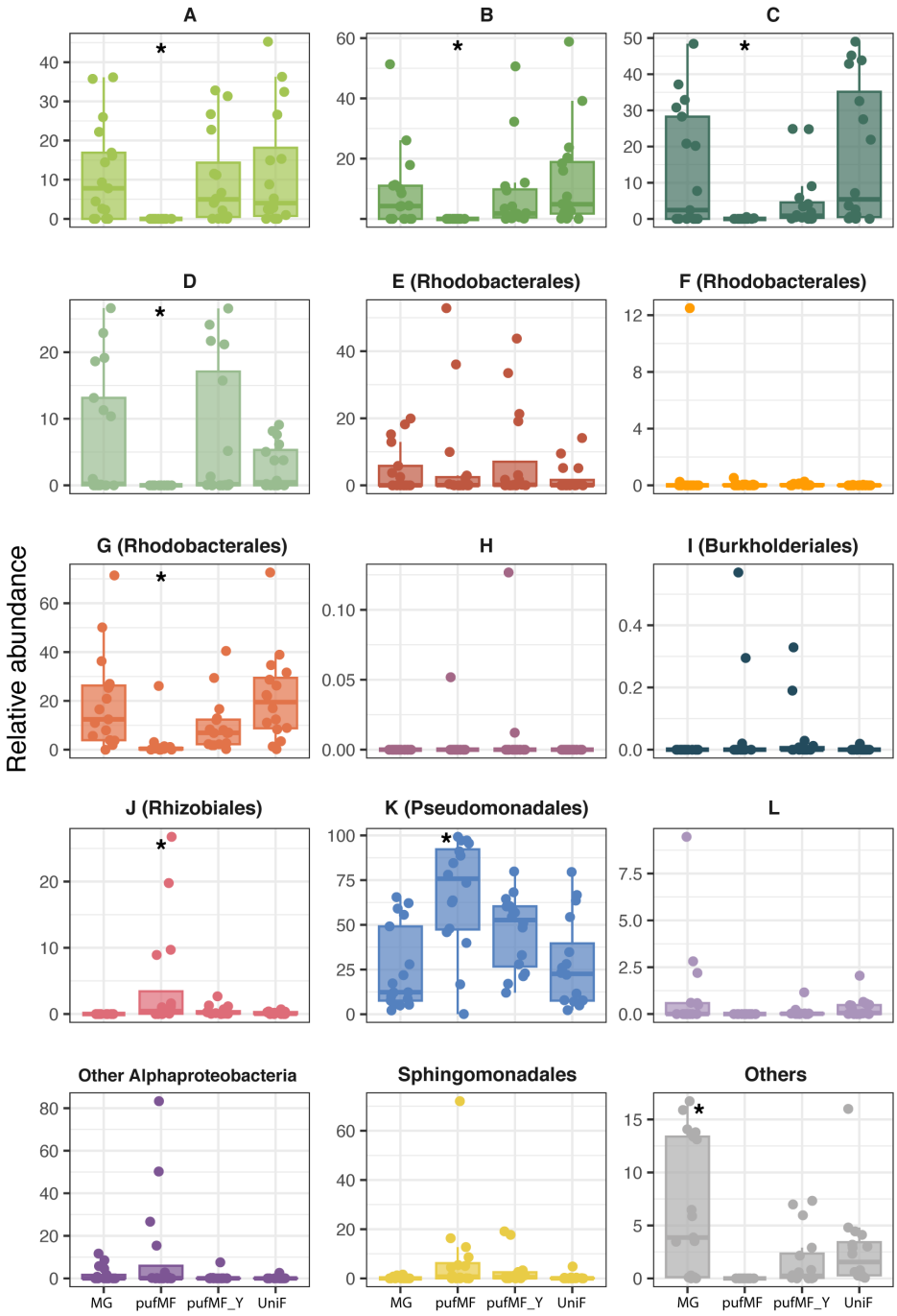

Figure S2. Relative abundance of the different taxonomic groups, retrieved in each amplicon approach and from the metagenomes. Samples of coastal and open ocean studies are pooled together in these plots. *Asterisks indicate groups that are statistically different from the others, after an analysis of variance and a post-hoc Tukey test (* $p<0.05$). MG: metagenomes; The amplicon approaches are identified with the forward primers.

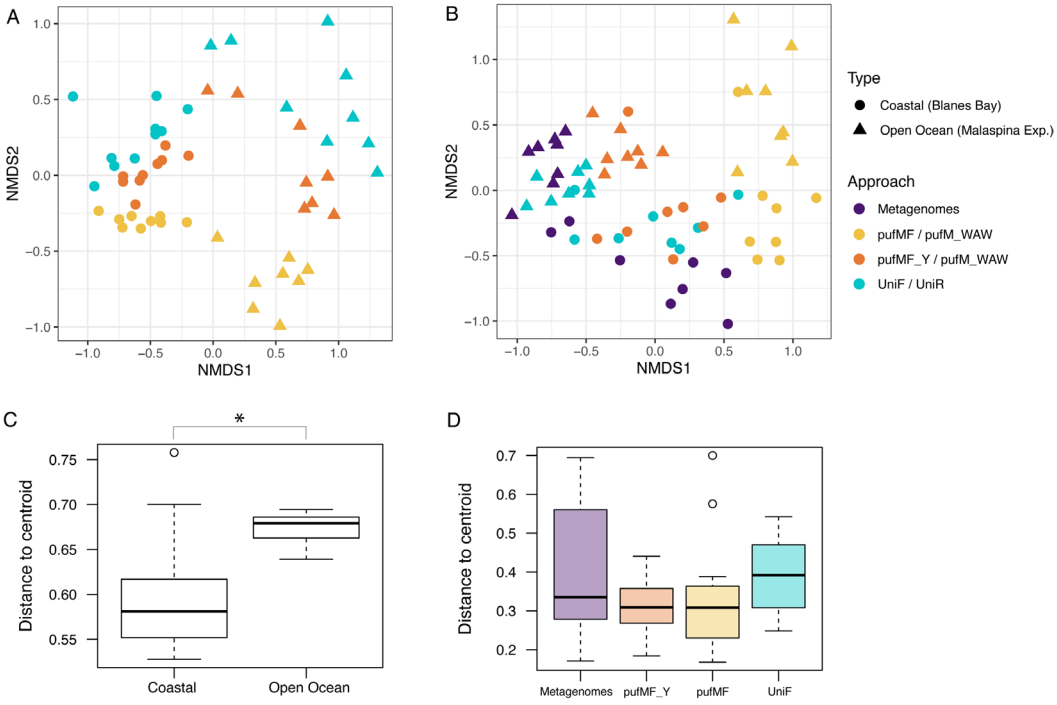

Figure S3. A) Non-metric multidimensional scaling (NMDS) of samples from the three amplicon approaches based on Bray-Curtis distances. B) NMDS of samples from the amplicon and the metagenomic approaches. This ordination is based on the abundance of the taxonomic groups instead of the abundance of ASVs (for the amplicon approaches) or contigs (for the metagenomes). Samples are color-coded based on the approach (see legend), and the type of environment (coastal and open ocean) is represented by circles and triangles respectively. C) Distance to centroids of coastal and open ocean environments samples (p<0.0001), and D) Distance to centroids of the different amplicon and metagenomic approaches (p=0.059) after analysis with the betadisper and permutest fucntions.

2.8 Supplementary figures

Table S1. Primer candidates designed for *pufM* amplification. Abbreviations; Tm = mean melting temperature, mismt = mismatches.

Primer candidate name	Sequence	Length (bp)	Hybridization(%)		Tm (ºC)	GC %
			0 mismt	1-2 mismt		
pufMF_V	5'-GGS AAY CTS TWY TAY AAY CCV T-3'	22	54.3	93.3	53.6	44
pufMF_Y	5'-GGS AAY CTS TWY TAY AAY C-3'	19	55.5	94.62	47.5	42.1
pufM_F1A	5'- GAR GCN GTK CCN TWY GGN RT-3'	20	61	94	58.9	57.5
pufM_F1B	5'- GAR GCN GTK CCN TWY GGN AT-3'	20	57	93	58.9	57.5
pufM_F2	5'- TGG GCN TTY GCN DSN GCV AT-3'	20	60	75.5	62.9	60

Table S2. PCR conditions for the primer candidates pufMF_V, pufMF_Y, pufM_F1A, pufM_F1B, pufM_F2 and pufM_WAW from Bejá et al. (2002), and UniF and UniR from Yutin et al. (2005). Only two combinations were successful in the amplification of the *pufM* gene fragment, the optimal PCR conditions of both assays appear in bold and underlined.

Forward primer	Reverse primer	Temperature (ºC)	Mg^{+2} (mM)	Primer concentration (µM)	Amplicon length (bp)	Amplification results
pufMF_V	pufM_WAW	51ºC – 63ºC	2 – 2.5	0.2 – 0.8	203 bp	No product
pufMF_Y	pufM_WAW	51ºC – 63ºC	2 – 2.5	0.2 – 0.8	203 bp	Amplification
pufM_F1A	pufM_WAW	50ºC – 59ºC	2 – 2.5	0.2 – 0.8	265 bp	No product
pufM_F1B	pufM_WAW	50ºC – 59ºC	2 – 2.5	0.2 – 0.8	265 bp	No product
pufM_F2	pufM_WAW	50ºC – 59ºC	2 – 2.5	0.2 – 0.8	340 bp	No product
UniF	UniR	48ºC – 58ºC	2 – 2.5	0.5 – 0.8	145 bp	Amplification
UniF	pufM_WAW	48ºC – 58ºC	2 – 2.5	0.5 – 0.8	198 bp	No product

Table S3. Oceanographic context of the samples used in this study.

A. Open Ocean samples from the Malaspina Expedition (all collected at 3 m deep)

Sample	Date	Oceanic region	Latitude	Longitude
MP0311	4th January 2011	South Atlantic	-7.28	-29.32
MP0778	5th February 2011	South Atlantic	-33.23	15.34
MP1176	7th March 2011	Indian Ocean	-29.57	96.41
MP1421	25th March 2011	South Australian Bight	-40.55	142.49
MP1517	20th April 2011	South Pacific	-23.38	-178.21
MP1672	29th April 2011	South Pacific	-3.41	-169.46
MP1857	14th May 2011	North Pacific	21.89	-155.66
MP2243	2nd June 2011	North Pacific	10.76	-102.44

B. Surface coastal samples from the Blanes Bay Microbial Observatory (41°40′N, 2°48′E), in the NW Mediterranean Sea.

Sample	Date
BL110208	8th February 2011
BL110412	12th April 2011
BL110615	15th June 2011
BL110802	2nd August 2011
BL111129	29th November 2011
BL120214	14th February 2012
BL120411	11th April 2012
BL120620	20th June 2012
BL120807	7th August 2012

Table S4. Number of amplicon sequence variants (ASVs), or predicted number of genes in the metagenomic assay, retrieved by each method and environment.

Assays	ASVs			
	Total	Open Ocean (Malaspina Exp)	Coastal (BBMO)	Common to both environments
pufMF / pufM_WAW	418	252	180	14
pufMF_Y / pufM_WAW	1904	1397	563	56
UniF / UniR	1294	819	518	43
Metagenomic Assay	238	176	62	-

Table S5. Mantel test and Procrustes test statistics. Correlation in the Procrustes test refers to the correlation in a symmetric Procrustes rotation. The Mantel statistic is based on Pearson's product-moment correlation. Both tests were performed with 999 permutations.

			Mantel Test		Procrustes Test		
			Mantel R	Significance	m_{12} squares	Correlation	Significance
pufMF/pufM_WAW	vs	pufMF_Y/pufM_WAW	0.8275	$p < 0.001$	0.0700	0.9643	$p = 0.001$
pufMF/pufM_WAW	vs	UniF/UniR	0.7204	$p < 0.001$	0.2043	0.8920	$p = 0.001$
pufMF_Y/pufM_WAW	vs	UniF/UniR	0.7938	$p < 0.001$	0.0728	0.9629	$p = 0.001$

2.9 Supplementary Information

Supplementary Information 1.

PCR conditions for primers pufMF_Y/pufM_WAW and UniF/UniR

Partial amplification of the *pufM* gene was performed with forward primer pufMF_Y and reverse primer pufM_WAW, each at 0.8 µM, and a final concentration of $MgCl_2$ of 2.5 mM. PCR conditions started with an initial denaturation step at 95 °C for 2 min, followed by 35 cycles at 95 °C (30s), 51°C (30s), 72 °C (40s) and a final elongation step at 72 °C for 10 min. PCR conditions for primers UniF and UniR were the following: an initial denaturation step at 95 °C for 2 min and 35 cycles at 95 °C (30s), 48°C (30s), 72 °C (40s) and a final elongation step at 72 °C for 10 min. The concentration of primers was 0.8 µM and the concentration of $MgCl_2$ was 2.5 mM.

Inferring amplicon sequence variants (ASVs) with the DADA2 pipeline:

The dataset from each combination of primers was processed separately in DADA2 v1.10 (Callahan et al., 2016). Sequences from primers pufMF and pufM_WAW were already analyzed in Auladell et al. (2019) and Gazulla et al. (2022), for the BBMO and the Malaspina Expedition samples respectively. In this study, we repeated the analyses with DADA2 for the 17 samples with the following parameters maxEE = c(2,2), trunclen = c(200,160). We kept 85% of total reads (mean 21311, min 8724, max 48334) after removing chimeras.

DADA2 parameters for the sequences obtained with primers pufMF_Y/pufM_WAW were maxEE = c(2,2), trunclen = c(190,120). After filtering we kept 83% of the total reads (mean 78047, min 26061, max 124270). Finally, for primers UniF/UniR DADA2 parameters were maxEE = c(2,2), trunclen = c(125,100) (the amplicon generated with this primer set is shorter). We kept 97% of total reads (mean 12319, min 1131, max 57444). Sample BL110412 had very low read counts (103) and was discarded from further analyses.

Supplementary Information 2.

Malaspina metagenomic analysis

A totalof 76 DNA samples from 11 stations of the global tropical and subtropical oceans (water filtered through 3.0 and 0.22 µm polycarbonate filters) were sequenced on an Illumina HiSeq 2000 platform (2 x 101 bp) at the Centre Nacional d'Anàlisi Genòmica (CNAG, Barcelona, Spain). Raw reads were trimmed in trimmomatic 0.38 (Bolger et al., 2014) with options ILLUMINACLIP:2:30:10 LEADING:3 SLIDINGWINDOW:4:20 MINLEN:45. Clean reads were assembled in megahit 1.1.3 (Li et al., 2016) with options --presets meta-large and --min-contig-len 500. Assembled contigs were annotated in prokka 1.14.6 for gene prediction (Seeman, 2014), clusters of orthologous groups (COGs), Enzyme Commission numbers (EC) and gene product name. Additionally, predicted genes were annotated for protein families' domains (PFAM v34, El-Gebali et al., 2019) using HMMER 3.33 (Eddy, 2011), the Kyoto Encyclopedia of Genes and Genomes Orthologs (KEGG KO) release v98.0 using kofamscan 1.3.0 (Aramaki et al., 2019), and carbohydrate active enzymes (CAZyme) using HMMER 3.3 against dbCAN v10 (Yin et al., 2012).

Malaspina gene catalog

All predicted genes longer than 100 bp were pooled and clustered in cd-hit-est 4.6.1 (Li and Godzik 2006) with options -c 0.95 -G 0 -aS 0.9 -g 1 -r 1 -d 0 -s 0.8. Clean reads were back-mapped to the catalog with bowtie2 2.2.9 (Langmead and Salzberg 2012) and alignments were filtered

with samtools 1.3.1 (Li et al., 2009) with options -q 10 -F 260. Reads mapping to catalog genes were counted in htseq-count 0.10.0 (Anders et al., 2015) with options --nonunique all to build gene profiles per sample. Counts were normalized by gene length in kbp and then normalized by the median abundance of 10 single copy COGs to obtain a proxy to per-cell number of gene copies as in Salazar et al. (2019).

Supplementary Information 3.

Statistical analyses

Data processing, statistical analyses and figures were performed in R v4.2.0 (R Core Team 2022). Alpha diversity values were calculated with *estimate_richness* function from phyloseq (McMuride and Holmes, 2013). The taxonomic composition of samples was represented with packages *phyloseq* (McMuride and Holmes, 2013) and *ggplot2* (Wickham, 2016), and we performed several statistical tests: we tested whether there were statistical differences between the taxonomic composition of samples retrieved with each primer combination using *adonis2* function (permutational multivariate analysis of variance) with Bray-Curtis distance matrices and 999 permutations. Also, for each taxonomic group, we performed analyses of variance and post-hoc Tukey tests (*aov* and *TukeyHDS* functions from *stats* package, version 3.6.2) to examine the relative abundance variation observed with each primer combination and with metagenomics.

For the community structure analyses we calculated Bray-Curtis distances with *vegdist* function, and Mantel correlations were performed with mantel function (999 permutations). We performed Procrustes tests to the three possible combinations between the three amplicon approaches (pufMF/pufM_WAW vs pufMF_Y/pufM_WAW, pufMF/pufM_WAW vs UniF/UniR and pufMF_Y/pufM_WAW vs UniF/UniR). The data was Hellinger-transformed using *decostand* function and we calculated the m_{12} statistic using *procrustes* function. The significance of the m_{12} statistic was then tested with the *protest* function and 999 permutations. All the functions mentioned in this paragraph are from *vegan* package (Oksanen, 2022).

Finally, the ordination plots were based on Bray-Curtis distances and calculated and represented with *ordinate* and *plot_ordination* from *phyloseq* (McMuride and Holmes, 2013). We used *betadisper* and *permutest* functions, both from *vegan* (Oksanen, 2022), to test the dispersion between samples from different environments (coastal vs open ocean) and assays (the three amplicon approaches and the metagenomic approach).

References from the Supplementary Information:

Anders, S., Pyl, P.T., Huber, W., (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. https://doi.org/10.1093/bioinformatics/btu638

Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., Ogata, H., (2019). KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* 1–2. https://doi.org/10.1093/bioinformatics/btz859

Béjà, O., Suzuki, M.T., Heidelberg, J.F., Nelson, W.C. Preston, C.M., Hamada, T., Elsen, J.A., Fraser, C.M., DeLong, E.F. (2002) Unsuspected diversity among marine aerobic anoxygenic phototrophs, *Nature*, 415, no. February, pp. 630–633, 2002.

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

Eddy, S.R., (2011). Accelerated profile HMM searches. *PLoS Computational Biology* 7, e1002195, 1–16. https://doi.org/10.1371/journal.pcbi.1002195

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., Sonnhammer, E.L.L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S.C.E., Finn, R.D., (2019). The Pfam protein families database in 2019. *Nucleic Acids Research* 47, D427–D432. https://doi.org/10.1093/nar/gky995

Langmead, B., Salzberg, S.L., (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9, 357–359. https://doi.org/10.1038/nmeth.1923

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, W., Godzik, A., (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. https://doi.org/10.1093/bioinformatics/btl158

McMurdie PJ, Holmes S (2013) Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* 8:4 DOI: 10.1371/journal.pone.0061217.

Oksanen J, Simpson G, Blanchet F et al (2022) vegan: Community Ecology Package_. R package version 2.6-2, https://cran.r-project.org/package=vegan

Salazar, G., Paoli, L., Alberti, A., Huerta-Cepas, J., Ruscheweyh, H.-J., Cuenca, M., Field, C.M., Coelho, L.P., Cruaud, C., Engelen, et al. (2019). Gene Expression Changes and Community Turnover Differentially Shape the Global Ocean Metatranscriptome. *Cell* 179, 1068-1083. e21. https://doi.org/10.1016/j.cell.2019.10.014

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

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

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., Xu, Y., (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* 40, W445-51. https://doi.org/10.1093/nar/gks479

Yutin, N., Suzuki, M.T., Béjà, O., (2005). Novel Primers Reveal Wider Diversity among Marine Aerobic Anoxygenic Phototrophs. *Applied and Environmental Microbiology* 7, 12, p. 8958-8962. doi:10.1128/AEM.71.12.8958–8962.2005

CHAPTER III

CHAPTER III

Diversity and community structure of AAP communities across the Atlantic Ocean epipelagic waters

Carlota R. Gazulla, Isabel Ferrera, Clara Arboleda-Baena, Vanessa Balagué, Aleix Obiol, Carolina Marín-Vindas, Aba González-Vega, José Escánez-Pérez, Ramon Massana, Eugenio Fraile-Nuez, Jesús M. Arrieta, Josep M. Gasol, Olga Sánchez

A manuscript in preparation

Abstract:

The epipelagic zone in the ocean is subjected to strong environmental gradients, influencing the composition of the microbial organisms inhabiting this area. Among them, the variability of aerobic anoxygenic phototrophic (AAP) bacterial communities across depth have been barely studied. AAP bacteria are a functional group of phototrophic prokaryotes that can obtain energy from light due to their bacteriochlorophyll *a* based reaction centers. Different studies analyzing their abundances and pigment distribution have postulated that they might be coupled to the phytoplankton. We hypothesized that the diversity and community structure of AAP bacteria will vary across the vertical gradient, related to changes in the environmental variables and the phytoplankton, following the deep chlorophyll maximum (DCM) profile. To test it, we have studied the composition of AAP communities at different depths along the DCM structure, in the South and Mid Atlantic Ocean, by means of amplicon sequencing of the *pufM* gene. Our study shows significant differences in the richness, community structure and taxonomic composition of samples of different layers, highlight the dependance of AAP bacteria on the DCM structure. Remarkably, we show that the main components of AAP bacterial assemblages belong to uncultured representatives from which there was very little information. This study provides a detailed description of these groups, focusing on their horizontal and vertical distribution and their ecological patterns.

3.1 Introduction

The pelagic sunlit zone in the ocean exhibits pronounced gradients of temperature, salinity, and light irradiance, undergoing relevant changes across tens of meters in the vertical scale. With increasing depth, there is a decline in light and a rise in nutrient concentrations. In stratified oceanic waters, the zone corresponding to the upper nutricline where there is still sufficient light to support photosynthesis harbors one of the most noteworthy features in phytoplankton distribution: the Deep Chlorophyll Maximum (DCM) (Cullen 1982, 2015). The DCM appears as a distinctive vertical structure observed in stratified waters, situated above the pycnocline and intricately linked to the nutricline (Estrada et al., 1993; Sharples et al., 2001). This dynamic layer exhibits variability across both space and time as discussed by Cullen (2015), and concentrates a substantial fraction of ocean primary productivity (Lorenzo et al., 2004; Weston et al., 2005; Marañón et al., 2021). Detailed characterization of the fine vertical distribution of phytoplankton groups within the DCM has uncovered its non-uniform structure, revealing the preference of specific phytoplankton groups towards distinct ecological niches (Cabello et al., 2016; Latasa et al., 2017). Considering that heterotrophic bacteria rely on phytoplankton-derived organic matter, the structure of the DCM could also influence the vertical distribution of different heterotrophic groups. Many studies have compared the microbiota in surface and DCM layers and showed differences in cell abundances, as well as in taxonomic and functional richness (Sunagawa et al., 2015; Mestre et al., 2018; Ruiz-González et al., 2019). However, fine vertical profiling of bacterioplankton along the DCM is limited, with only minimal studies conducted so far (Walsh et al., 2015; Marin, 2023; Haro-Moreno et al., 2018).

Prokaryotic ocean microbiomes include photoheterotrophs, i.e. proteorhodopsin (PR)-containing bacteria and aerobic anoxygenic phototrophic (AAPs) bacteria. These two groups exhibit different distributions in surface waters: PR-containing bacteria are more abundant in oligotrophic systems (Gómez-Consarnau et al., 2007, 2010; Morris et al., 2010) and AAP bacteria thrive better in more eutrophic waters (e.g. Cottrell et al. 2006; Jiao et al., 2007; Hojerova et al., 2011). A recent study examined the vertical distribution of pigments associated with these two energy-transducing mechanisms, as well as that of chlorophyll *a* (Chl *a*) in the Mediterranean Sea and the Eastern Atlantic Ocean (Gómez-Consarnau et al., 2019). Interestingly, while the geographical distribution of rhodopsins and Chl *a* revealed contrasting trends, bacteriochlorophyll *a* (BChl *a*)–the photosynthetic pigment of AAP bacteria– and Chl *a* exhibited similar niches along various trophic

gradients (Gomez-Consarnau et al., 2019), suggesting a potential association between AAP bacteria and phytoplankton.

On the other hand, the diversity and taxonomic composition of AAP communities display variability across the surface ocean, with these changes primarily linked to factors such as temperature, salinity, and trophic status (Jiao et al., 2007; Lehours et al., 2018; Gazulla et al., 2022). Besides, seasonal shifts in AAP assemblages in surface waters are also associated with fluctuations in temperature, light availability, or the presence of the spring phytoplankton bloom (Ferrera et al., 2014; Auladell et al., 2019). Despite the wealth of information amassed on their diversity and distribution in the surface ocean, the vertical dimension, that display a high environmental variability along few tens of meters, has largely been overlook. A previous study in the Mediterranean Sea indicated that AAP communities in the DCM or below differed from those above it, with light identified as the primary explanatory variable (Lehours et al., 2010). However, the specificities of these communities along the vertical structure of the DCM remain unknown, despite the fact that the robust vertical gradient of Chl *a* and other physicochemical variables likely act as influential drivers for AAP community assembly. Profiling the vertical distribution of AAP taxa along the DCM would contribute to our understanding of their niche preferences and provide valuable insights for investigating their relationship with phytoplankton communities.

In this chapter, we aimed to evaluate the composition of AAP communities along the DCM structure, in areas characterized by varying productivity levels, by collecting samples at different depths along a latitudinal transect spanning the South and Mid Atlantic Ocean, following the chlorophyll variations indicated by the CTD-fluorometer. The transect encompassed DCM structures of diverse nature, ranging from cold waters with shallow DCMs to warm waters with DCMs located below 100 m. We conducted an analysis of the AAP community diversity and structure by means of amplicon sequencing of the *pufM* gene, the genetic marker for this functional group. We hypothesize that the ecological patterns of AAP communities will vary across the vertical gradient, following the variability of the main environmental variables and, ultimately, related with changes in phytoplankton composition. To do so, we also analyzed the community composition of bacterioplankton and phytoplankton for a subset of samples, and explored similarities and co-occurrences between functional groups. An important feature of this study is the use of primers that have a high phylogenetic coverage that avoids major primer biases, and allows a good representation of all components of the AAP communities (Gazulla et al., 2022). Therefore,

we explored whether AAP phylogroups missed in previous amplicon-based approaches are an important fraction of the AAP communities here described.

3.2 Methods

Sample collection

The Poseidon Expedition, conducted aboard the research vessel *Sarmiento de Gamboa* between March and April 2019, spanned approximately 9,000 km along a latitudinal transect (48°S/26°N) in the Atlantic Ocean. Seawater was systematically collected at various depths in 27 stations, to capture the structure of the DCM, which was delineated using a CTD profiler SBE 911plus, with samples taken at different depths: surface, above DCM, DCM, below DCM, and beyond the end of the fluorescence signal. Additionally, six stations were sampled at a higher resolution throughout the DCM (10 sampling depths). All samples underwent prefiltration through a 200 µm mesh to remove large plankton. The CTD was used to describe a comprehensive suite of variables including temperature, salinity, conductivity, fluorescence, dissolved oxygen, and others. A total of 112 samples were collected from depths ranging between 5 m and 280 m. Flow cytometry was used to assess the abundance of cyanobacteria, nanoeukaryotes, picoeukaryotes, and heterotrophic bacteria with high and low nucleic acid content (HNA and LNA), as described in Gasol and Moran (2015). The concentration of inorganic nutrients, specifically nitrate, nitrite, phosphate, and silicate were determined following the procedures outlined in Catalá et al. (2016). Chl *a* and BChl *a* concentrations were measured using high-performance liquid chromatography (for details, see Chapter IV).

DNA extraction, amplification and sequencing of the *pufM* gene

Approximately 2 L of seawater were filtered through 20 µm and 3 µm prefilters onto a 0.2 µm 47 mm polycarbonate filter. Samples were subsequently stored at -80°C until further processing. DNA was extracted from the 0.2 µm filter (0.2–3 µm fraction) using the phenol–chloroform protocol described in Massana et al., (1997). Partial amplification of the *pufM* gene (~180 bp fragment) was carried out in 12.5 µL reactions using primers pufM_uniF (GGNAAYYTNTWYTAYAAYCCNTTYCA) and pufM_UniR (YCCATNGTCCANCKCCARAA), from Yutin et al. (2005). The amplification followed the conditions detailed in Gazulla et al. (2023). In brief, PCR conditions were as follows: an initial denaturation step at 95 °C for 5 min, 35 cycles at 95 °C (30s), 48°C (45s), 72 °C (45s), and a final elongation step at 72 °C for 7 min. DNA sequencing was conducted on an Illumina MiSeq sequencer by AllGenetics & Biology SL (www.allgenetics.eu).

Noteworthy, certain samples, particularly some from the DCM and below the DCM, could not be successfully amplified, likely due to the low abundance of AAP bacteria. As a result, we were able to obtain high-quality sequencing results from 84 samples.

Amplicon sequence variants generation and taxonomic assignation

We used cutadapt v3.4 (Martin, 2013) for the removal of primers, and DADA2 v1.26 (Callahan et al., 2016) to discern amplicon sequence variants (ASVs) and eliminate chimeras. To infer the phylogeny of these amplicon sequence variants, we utilized phylogenetic placement with the Evolutionary Placement Algorithm v0.3.5 (Barbera et al., 2019) and a custom made *pufM* database (Gazulla et al., 2023). We defined a total of 15 different taxonomic groups based in the classification proposed by Yutin et al. (2007), and encompassing different orders from the Gamma- and Alpha-proteobacteria. Sequences that could not be further assigned were collectively grouped as 'Unclassified'.

Data analyses

All analyses were conducted in R v4.2.0 (R Core Team 2022). The resulting ASV table underwent rarefaction down to 6576 reads using the *vegan* package (Oksanen et al., 2022). We estimated alpha diversity indexes (Chao1, Shannon, and Evenness) and categorized samples into different groups according to their position along the DCM profile and their belonging to the sampled Longhurst provinces (Longhurst, 2007). Differences between groups were assessed by analyses of variance with the *aov* function in the R Stats Package *stats*, (version 3.6.2) and the post-hoc Tukey's 'Honest Significant Difference' method with function *TukeyHDS* from the same package. Correlation between changes in various taxonomic groups and environmental variables were estimated with the *rcorr* function (*Hmisc* package, Harrell and Dupont, 2023), with Bonferroni correction applied. Ordination tests were based on the Bray-Curtis distance index and calculated with the *vegdist* function. Additionally, for a small subset of samples (8 samples from stations 10 and 18), we analyzed the diversity of the bacterioplankton community and that of the eukaryotic community for the micro- (20–200 µm), nano- (3–20 µm), and picoplankton (0.2–3 µm) fractions (Supplementary Information). We compared the community structure of AAP bacteria with that of bacterioplankton and eukaryotic plankton using Mantel correlations with the *mantel* function and 999 permutations. The correlation between the community of AAP bacteria and the bulk bacterioplankton is valid since AAP bacteria only account for a very low percentage of the total bacterial community (1.5% ± 1.1%, Chapter IV). Finally, we normalized the databases and constructed co-occurrence networks using maximal information coefficient (MICe) analysis with the software MICtools, that

identify relevant associations and assess their strength and their statistical significance (Albanese et al. 2018). The null distribution was computed with 200,000 permutations. Subsequently, empirical *p*-values were calculated for all variable pairs. The *p*-values were corrected for multiplicity to control for the false discovery rate (FDR), using the default method. The strength of the significant relationships were then estimated using MICe, which range between 0 (not strong relationship) and 1 (very strong relationship). The network obtained had 161,819 number of co-occurrences (with $p<0.05$) that were filtered and compared considering several MICe threshold (0.4, 0.6, 0.7 and 0.8) and group associations: phytoplankton (phototrophic eukaryotes and cyanobacteria), AAP bacteria and bacterioplankton (removing cyanobacteria). Phototrophic eukaryotic groups were assigned according to common assumptions in the literature. The code of the analyses performed at R 4.2.0 (R Core Team 2022) can be found at https://gitlab.com/ crgazulla/aaps_atlantic_ocean. Amplicon sequences have been deposited in the NCBI Sequence Read Archive (SRA) under BioProject PRJNA1049819.

3.3 Results and Discussion

Oceanographic context

The Poseidon Expedition consisted on a latitudinal transect spanning the South and Mid Atlantic Ocean, encompassing regions of varying productivity (Figure 1, Figure S1). The cruise extended from the eutrophic waters in the South Subtropical Convergence zone to the highly oligotrophic subtropical gyre, the equatorial upwelling, and part of the North tropical Atlantic Ocean. Beginning from the south, the transect started in a high productivity region in the South Atlantic, characterized by low temperatures (mean 9.3 °C) and lower salinity values (mean 34.2), influenced by the proximity to the Southern Ocean (stations 1 and 2, beyond 40° S). These southernmost stations exhibited elevated Chl *a* and nutrient concentrations, particularly silicate (mean 5.53 µM in this area vs. 2.15 µM in the whole transect), that frequently promote diatom blooms in this area (Malviya et al., 2016). Noteworthy, the HPLC analyses from the cruise (data not shown) revealed high values of fucoxanthin, a characteristic pigment of diatoms, in samples from these stations. Moving northwards, temperature and salinity increased up to the maximum values recorded in the transect (temp. >28.5 °C and salinity >37 in stations 10 to 14). These stations, located within the South Atlantic oligotrophic gyre (Figure 1), exhibited low surface fluorescence, and the DCMs appeared below 120 m depths (Figure S1). Oligotrophic gyres, often defined as “deserts” (Irwin & Oliver, 2009) represent stressful environments for growth with minimal nutrients concentrations. In fact, the lowest values

of nitrite, nitrate, phosphate, and silicate were observed in this area (Figure S2). Notably, two stations in this area displayed a two-peak DCM structure (station 10 and station 16). In these cases, the shallowest peak was located at ~40 m and ~25 m depth, respectively, while the deeper one, with the highest values of fluorescence, was found at depths around 100 and 80 m (Figure S1). As the expedition progressed towards the equator, DCMs became shallower (between 50–100 m deep) and were characterized by higher fluorescence and nutrient concentrations, while temperature and salinity decreased. Finally, beyond the equator, seawater temperatures remained at around 20 °C, but salinity increased as we were approaching the North Atlantic oligotrophic gyre. Stations 23 to 28 presented low nutrient levels and strong vertical mixing due to the upwelling waters off the coast of Mauritania, being the mixed layer depth (MLD) below 100 meters in several of them. In this last part of the transect, the DCM depths were around the 100 m depth. The environmental variability across the transect is reflected in the Longhurst provinces classification (Longhurst, 2007) of the stations, which can be assigned to the four provinces in the study area: the South Subtropical Convergence zone (SSTC), the South Atlantic gyre (SATL), the western tropical Atlantic (WTRA), and the North Atlantic Tropical gyre (NATR) (Figure 1B).

AAP richness displays strong horizontal and vertical patterns

We identified a total of 3,418 Amplicon Sequence Variants (ASVs) in the 84 samples analyzed, making the highest number of *pufM* amplicon variants retrieved to date. This number surpasses previous datasets, even those based on a larger number of samples (e.g. Auladell et al., 2019; Gazulla et al., 2022). While our methodology shared similarities with these studies, here we employed the primers proposed in Yutin et al. (2005) with broader phylogenetic coverage, enabling the capture of a greater number of genetic variants. Although these primers were designed almost two decades ago, they haven't been commonly used in marine samples. A recent comparison of these primers and others traditionally employed in AAP studies showed that the use of UniF and UniR (Yutin et al., 2005) yielded higher alpha diversity values and a similar taxonomic composition to that observed from metagenomics (Gazulla et al., 2023). Furthermore, our study included samples from deeper layers in the euphotic zone, which contributed to the high diversity observed, as compared to the abovementioned studies based mostly on surface samples. In this regard, a total of 1,676 ASVs (49%) were exclusively present below the surface. This indicates that a significant portion of the AAP diversity is located in this zone. Studies concentrated exclusively on the surface, ignoring the ocean vertical dimension, are thus probably missing a high fraction of AAP diversity.

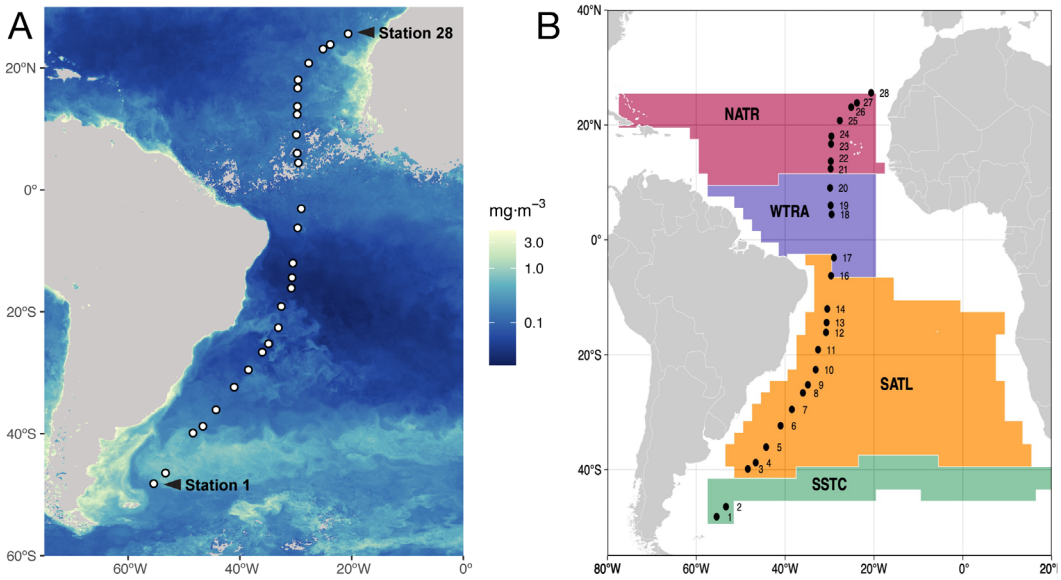

Figure 1. A) Near-surface chlorophyll concentration of the Poseidon Expedition transect along the South and Mid Atlantic Ocean. The colors represent near-surface chlorophyll concentration ($mg{\cdot}m^{-3}$) recorded in March 2019. The data was extracted from the SNPP-VIIRS instrument and obtained from the NASA Ocean color website (https://oceancolor.gsfc.nasa.gov/l3/). Each dot on the transect represents one station, numbered from 1 to 28. Note that station 15 does not exist. B) Partition of the transect into four Longhurst provinces: SSTC, South Subtropical Convergence zone; SATL, South Atlantic gyre, WTRA, Western tropical Atlantic, and NATR, North Atlantic Tropical gyre.

To examine alpha diversity variations horizontally and vertically, we clustered samples based on their Longhurst provinces and their position along the DCM structure, as detailed in the Methods section. Alpha diversity values remarkably exceeded those reported in previous studies with different primer choices (e.g., Bibiloni-Isaksson et al., 2016; Auladell et al., 2019; or Gazulla et al., 2022). Richness values varied between 103 and 339 (mean 214 ± 58), a fair representation of the diversity, as seen by the rarefaction curves (Figure S3), yet surpassing observations in the same area during the Malaspina Expedition (Gazulla et al., 2022). We observed a notable latitudinal pattern of increasing richness from south to north (Figure 2A, Pearson correlation with latitude, n=84, R=0.29, $p<0.01$). For centuries, a pervasive pattern of richness increasing from poles towards the equator has been described for different species of plants and animals (Hillebrand, 2004). The first global ocean approaches based on plankton also showed a latitudinal diversity gradient highly related to temperature (Fuhrman et al., 2008; Tittensor et al., 2010; Sunagawa et al., 2015). However, in these studies, highest richness values were observed in intermediate latitudinal ranges. In any case, there is no clear biogeographical

pattern of microbial diversity, since other studies reported a negative correlation of diversity and latitude, with richness being highest at lowest latitude (Sul et al., 2013), or no correlation at all when only considering samples from high latitudes (Ghiglione et al., 2012). Although our dataset covers a long latitudinal transect (from 48° S to 25° N), we can only observe part of this global pattern in the case of AAPs. It seems that the diversity of AAP bacteria increases towards low latitudes, although in our study this correlation is not related with temperature (n=84, R=0.01, p=0.89). Additionally, we calculated evenness (Figure 2C and D) and Shannon indexes (Figure S4). Samples from the SSTC province, highly influenced by their proximity to the Southern Ocean, exhibited extreme lowest values of temperature and salinity, and high values of Chl *a*, and were excluded from further analyses as they masked some of the observed patterns. After that, there were no significant changes across the latitudinal (horizonal) axis, and samples from SATL, WTRA, and NATR had high and similar mean values of both evenness and Shannon. These values indicate that the abundance of species is well distributed among samples and that the increase in richness with latitude is not a consequence of few species with higher abundance, but rather with an increase in the complexity of communities. Only samples from the SSTC province had significantly lower values of evenness (Figure 2C, Figure S4, Tukey test, p<0.0005). This pattern aligns well with the global diversity trends observed for the prokaryotic plankton (Ibarbalz et al., 2019), where Shannon diversity sharply decreased at high latitudes (above 40°).

Along depth, richness variations followed chlorophyll levels, generally peaking at the Chl *a* maximum (Pearson correlation with Chl *a*=0.30, p<0.01, after removing samples from SSTC province). Previous studies have also shown that bacterioplankton diversity increases from the surface towards deeper waters in the epipelagic (Haro-Moreno et al., 2018; Marín, 2023), even when extending to bathypelagic depths (Sebastian et al., 2021). AAP bacteria are restricted to the epipelagic zone of the ocean, and it seems rather than a gradient of richness with depth, richness follows the distribution of Chl *a* at stations with very different DCM structures. AAP species were evenly distributed in all samples, as no significant differences were found in evenness across depth. Few samples with lower evenness values had a relatively high contribution of some specific ASVs.

Contrary to previous studies describing a negative correlation between AAP diversity and trophic status (Jiao et al., 2007; Jeanthon et al., 2011; Gazulla et al., 2022), we did not observe that trend when considering all samples together (Pearson correlation with Chl *a* concentration, n=84, R =–0.1, p=0.36). When exclusively using samples from the

surface, a negative, albeit not statistically significant, correlation emerged (n=23, R=–0.39, *p*=0.07). All in all, by expanding our sampling efforts beyond the surface to deeper waters across the epipelagic, we were able to record the highest diversity values for AAP bacteria. Furthermore, we observed that the patterns of alpha diversity variation observed at the surface do not necessarily apply throughout the euphotic zone. Instead, the structure of the DCM appears to play an important role in in determining the richness of AAP communities.

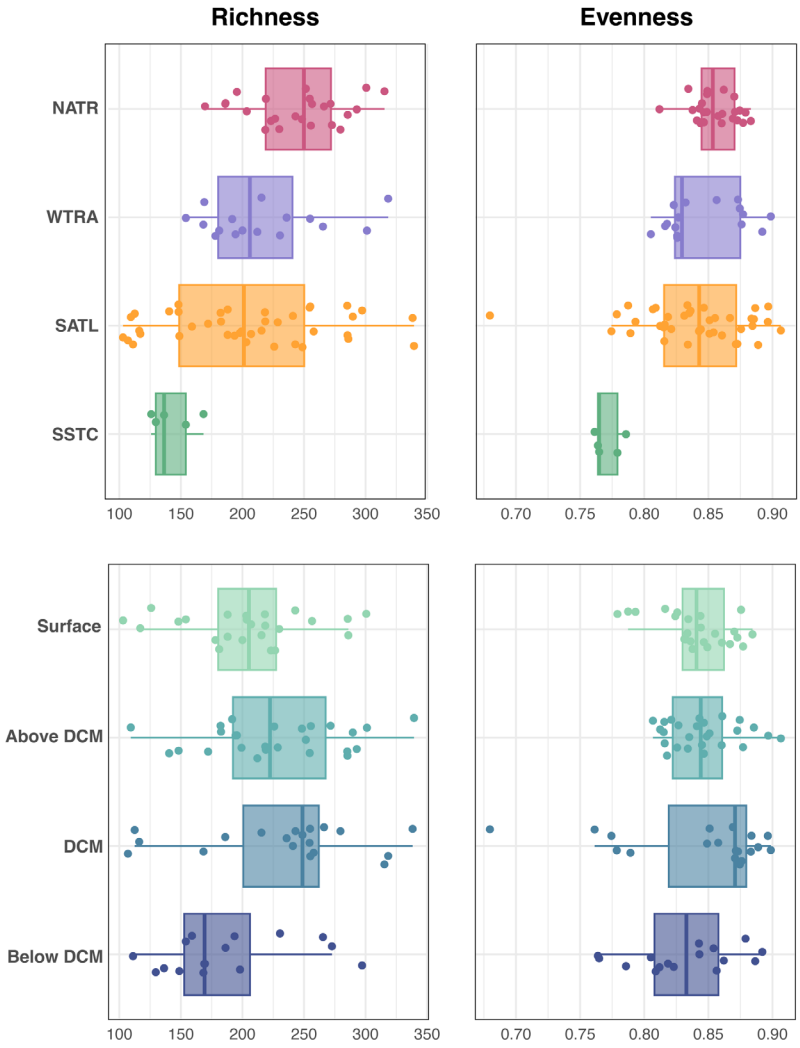

Figure 2. Alpha diversity variation measured across latitude (horizontal scale, above) and vertically through the deep chlorophyll maximum (DCM) structure (below).

Uncultured and barely-known species of AAP bacteria prevail along epipelagic waters in the open ocean.

We asigned the *pufM* sequences to specific taxonomic groups based on their placement in a phylogenetic tree (Figure S5). This classification used the phylogroups defined in Yutin et al. (2007) complemented with the GTDB taxonomy, considering different orders of the Alphaproteobacteria (Sphingomonadales, Rhodobacterales and Rhizobiales) and Gammaproteobacteria (Pseudomonadales and Burkholderiales) classes. In total, we classified the ASVs into 15 distinct taxonomic groups, as illustrated in the phylogenetic tree. Sequences that could not be further assigned (1.25%) were grouped under the category 'Unclassified'. The taxonomic group labeled 'uncultured_G1' encompasses sequences retrieved from various oceanographic expeditions, including the Tara Oceans (Sunagawa et al., 2015) and the Malaspina Expeditions (Sánchez et al., 2023; Acinas et al., 2021), and the Blanes Bay Microbial Observatory (BBMO) time series (Auladell et al., 2019), that formed a differentiated clade phylogenetically related to the previously defined phylogroups C and D (Figure S5). Although ASVs in this category only represented 3.5% of total ASVs, they were relatively abundant in several samples in the DCM and below the DCM layers (Figure 3).

Phylogroup C consistently dominated along all samples and depths (mean relative abundance ± s.d. 41.49% ± 12.9), followed by phylogroup A (13.52% ± 7.65), phylogroup D (12.24% ± 6.4), phylogroup K within the Halieaceae family (Pseudomonadales order, 10.1% ± 6.7), phylogroup B (8.35% ± 9.81) and phylogroup G in the Rhodobacterales (6.94% ± 5.7) (Figure 3). The other taxonomic groups exhibited a mean relative abundance below 3%. Contrary to previous assumptions, our results challenge the prevailing notion that the Pseudomonadales (phylogroup K) and Rhodobacterales (specially phylogroup G) dominate AAP assemblages in the open ocean (Jiao et al., 2007; Lehours et al., 2010; Bibiloni-Isaksson et al., 2016; Gazulla et al., 2022). On the other hand, even though they were first described almost two decades ago (Yutin et al., 2007), phylogroups A, B, C, and D had rarely been reported in subsequent studies, particularly those based on amplicon sequencing (e.g. Jiao et al., 2007; Lehours et al., 2010; Ferrera et al., 2014; Bibiloni-Isaksson et al., 2016; Auladell et al., 2019; Gazulla et al., 2022). This scarcity is attributed to biases in the commonly used *pufM* primers. A recent reassessment of these primers in samples from the Atlantic, Indian, and Pacific Oceans revealed that phylogroups A, B, C, and D are indeed significant components of AAP communities in the open ocean (Gazulla et al., 2023). Our findings further indicate the prominence of these phylogroups as the primary components of AAP communities in the Atlantic Ocean.

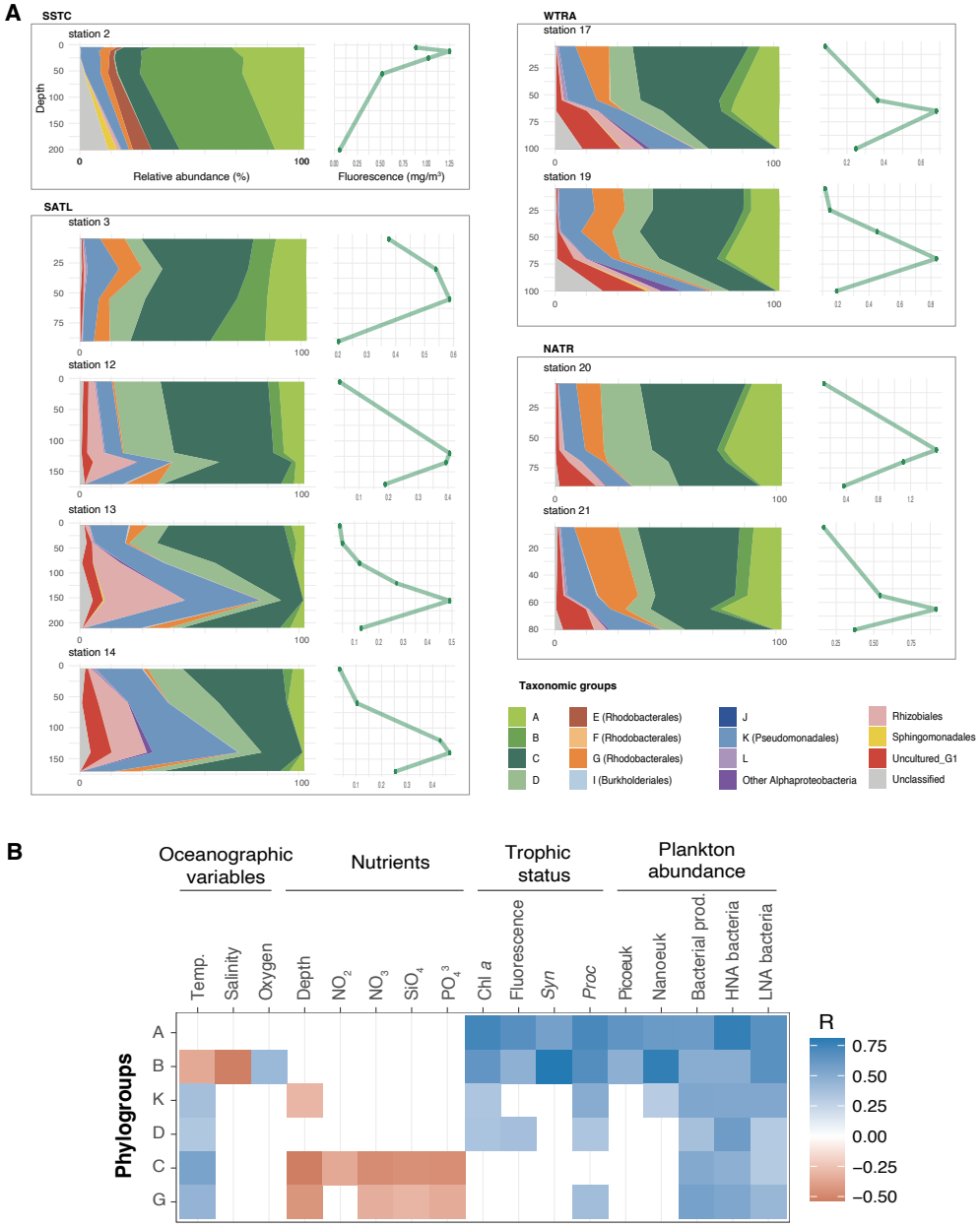

Figure 3. A) Taxonomic composition of AAP bacterial communities (left of each panel) and fluorescence variation (right of each panel) across depths. Stations from different Longhurst provinces are delimited within different boxes. A selection of nine stations are shown. For futher information, see Figure S6. B) Pearson correlation coefficients between the relative abundance of the most prevalent taxonomic groups and environmental variables. The plot only display statistically significant correlations (p<0.05). Temp: Temperature; *Syn*: Abundance of *Synechococcus*; *Pro*: Abundance of *Prochlorococcus*; Bacterial prod.: Heterotrophic bacterial production; HNA bacteria: high-nucleic acid–containing bacteria; LNA bacteria: low–nucleic acid–containing bacteria.

Interestingly, there is very little ecological information about these groups and no clues about their physiology, since there are no representatives isolates. Although typically studied collectively as a cohesive microbial guild, AAP bacteria comprise very diverse metabolisms and life strategies, ranging from generalists to specialists that thrive on different ranges of carbon sources (Yurkov and Csotonyi, 2009), and with distinct habitat preferences (Auladell et al., 2019). This study, as it covered a wide environmental heterogeneity, sheds some light into their ecological preferences (Figure 3B). Phylogroups C and G exhibited a decrease in relative abundance with increasing depth and nutrient concentrations. Conversely, the contribution of phylogroups A and B increased in samples located at the DCM, and correlated with variables associated to the trophic status, such as fluorescence, Chl *a* concentration, and the abundance of *Synechococcus* and *Prochlorococcus* (Figure 3B). Besides, the higher abundance of phylogroup B in samples from the SSTC province, characterized by low temperature and salinity, suggests that this group may be better adapted to cold environments.

On the contrary, other groups that have been commonly described in previous studies, are scarce in this study. For instance, phylogroups E, F, and G representing Rhodobacterales-like bacteria, are typically associated with nutrient rich-conditions and phytoplankton blooms in coastal areas (Lehours et al., 2010, Ritchie and Johnson, 2012, Ferrera et al., 2014, Auladell et al., 2019). As the Poseidon Expedition traversed an open ocean area, the abundance of this group was very low (below 1% for phylogroups E and F). Members of the Burkholderiales, order Gammaproteobacteria (formerly known as Betaproteobacteria), are frequently found in regions characterized by low salinity such as the Arctic Ocean (Boeuf et al., 2013). Accordingly, we found very few sequences of Burkholderiales in our dataset. Noteworthy, the DCM samples from stations 13 and 14 attracted our attention for the high relative abundance of few sequences from phylogroup K and Rhizobiales, the former being notably rare in the rest of the dataset. These samples, located in the South Atlantic gyre at depths of 155 m and 140 m, respectively, likely represent species adapted to regions characterized by low nutrient concentration and reduced irradiance.

This study expands our knowledge on the different taxonomic groups of AAP bacteria, and their distribution across different oceanic regions and address, for the first time, the fine-vertical taxonomic composition of AAP communities along the epipelagic zone. According to their taxonomic composition, there is a higher variability across the vertical scale than across the latitudinal axis (Figure S6). Also, this variability seems to be more

related to the DCM structure than to changes with depth. Samples at the DCM harboured higher alpha diversity, particularly from several groups that were absent from surface samples. Remarkably, around 50% of the ASVs identified in the DCM and below were exclusive to these epipelagic layers, especially in samples collected around the equator (stations 17 to 22). All in all, our results show that the DCM represent an ecotone where we can find strong shifts in the community composition of AAP bacteria, as recently reported for the prokaryotic and eukaryotic communities (Marin, 2023).

The structure of AAP communities varies across Longhurst provinces and along the DCM structure

To assess the biogeographical pattern of AAP communities across the vertical and horizontal scales, we computed Bray-Curtis (BC) dissimilarities between samples and visualized them in a non-metric multidimensional (nMDS) scaling plot (Figure 4). In addition, we complemented this analysis with a PERMANOVA test; salinity (p<0.01) and oxygen (p<0.05) emerged as the variables that better explained the community structure of AAP bacteria (Figure S7), in concordance with other studies evaluating AAP communities across the surface ocean (Lehours et al., 2018; Bibiloni-Isaksson et al., 2016; Gazulla et al., 2022). Samples within the SSTC and NATR provinces exhibited greater similarities among them, clustering together (Figure 4A) and displaying low mean BC dissimilarities among their samples (Figure 4B). The deepest sample in SSTC, obtained from 200 m deep, appears separated from the SSTC cluster, indicating that its composition differs from the rest of samples from this province. These differences can be further appreciated by the high BC dissimilarity values among this specific sample and the others (Figure 4B). All samples from the SSTC have very low values of salinity, which might influence their community composition as they appear distantly located from other provinces in the plot. Samples from the NATR and WTRA provinces are more or less dispersed, and both the oxygen variation and the vertical component probably explain the differences between communities. Finally, the SATL province comprises a vast area within the oligotrophic gyre, likely serving as a bridge between two highly productive areas (SSTC and WTRA), which may account for the considered variability observed within samples from this province. Moreover, samples from these stations covered a broader depth range, as the region is characterized by deeper DCMs, prompting us to extend the sampling down to 280 m depth.

The Longhurst provinces demarcate regions characterized by similar environmental oceanic variables, a classification proven to be effective in explaining the biogeographic

structure of various microbial communities (Friedline et al., 2012; Frank et al., 2016; Milici et al., 2016; Logares et al., 2020; Ruiz-González et al., 2020), including those of AAP bacteria (Gazulla et al., 2022). In contrast, the effects of the vertical component in the distribution of AAP communities have been barely explored. It is widely known that the environmental variability associated to depth is a major driver of change in the composition of prokaryotic and eukaryotic communities (Acinas et al., 1999; Sunagawa et al., 2015; Cabello et al., 2016; Salazar et al., 2016; Latasa et al., 2017; Giner et al., 2020; Sánchez et al., 2023; Junger et al., 2023). To illustrate the vertical variability in our dataset, we colored samples by their position along the DCM profile (Figure 4A) and we could observe a clear ordination based on these layers, indicating that the structure of communities is clearly influenced by their position along the DCM structure. A few exceptions to this ordination correspond to samples from stations 10, 12, and 13 (Figure 4A, samples shown with red border) situated in the oligotrophic gyre and characterized by deeper and weaker DCMs. In these stations, salinity, oxygen, and nutrient concentrations (Figure S1 and S2) remain similar along the first 150 m, which might explain why the composition of communities is similar despite the depth difference. In other stations across the oligotrophic gyre (SATL) and the equator (WTRA), samples appear organized following the DCM structure, probably due to the strong environmental changes in this area (Figures S1 and S2). All in all, changes in the community composition of AAP bacteria increased with depth (Figure 4C), as the dissimilarities between one layer and the one just below increased from the surface to layers below the DCM. A previous study based on *pufM* clone libraries in the Mediterranean Sea showed differences between communities from above and below the DCM (Lehours et al., 2010). In this study, we have expanded the vertical range by considering different layers along the DCM structure and our results show that, besides the influence that can be attributed to the Longhurst provinces, the DCM structure influence the composition of AAP communities in the epipelagic zone.

AAP bacteria show higher significant co-occurrences with phytoplankton than the bulk heterotrophic bacterioplankton.

In the precedent analyses, we focused on abiotic variables as potential drivers of the diversity and community structure of AAP bacteria. However, compelling evidence suggests a potential coupling between AAP bacteria and phytoplankton. Firstly, AAP bacteria are more abundant in eutrophic sites (Cottrell et al., 2006; Jiao et al., 2007; Hojerová et al., 2011). Additionally, the pigments of both groups, Chl *a* and Bchl *a* occupy the same regions along the water column (Gómez-Consarnau et al., 2019). A plausible explanation for this observation could be the

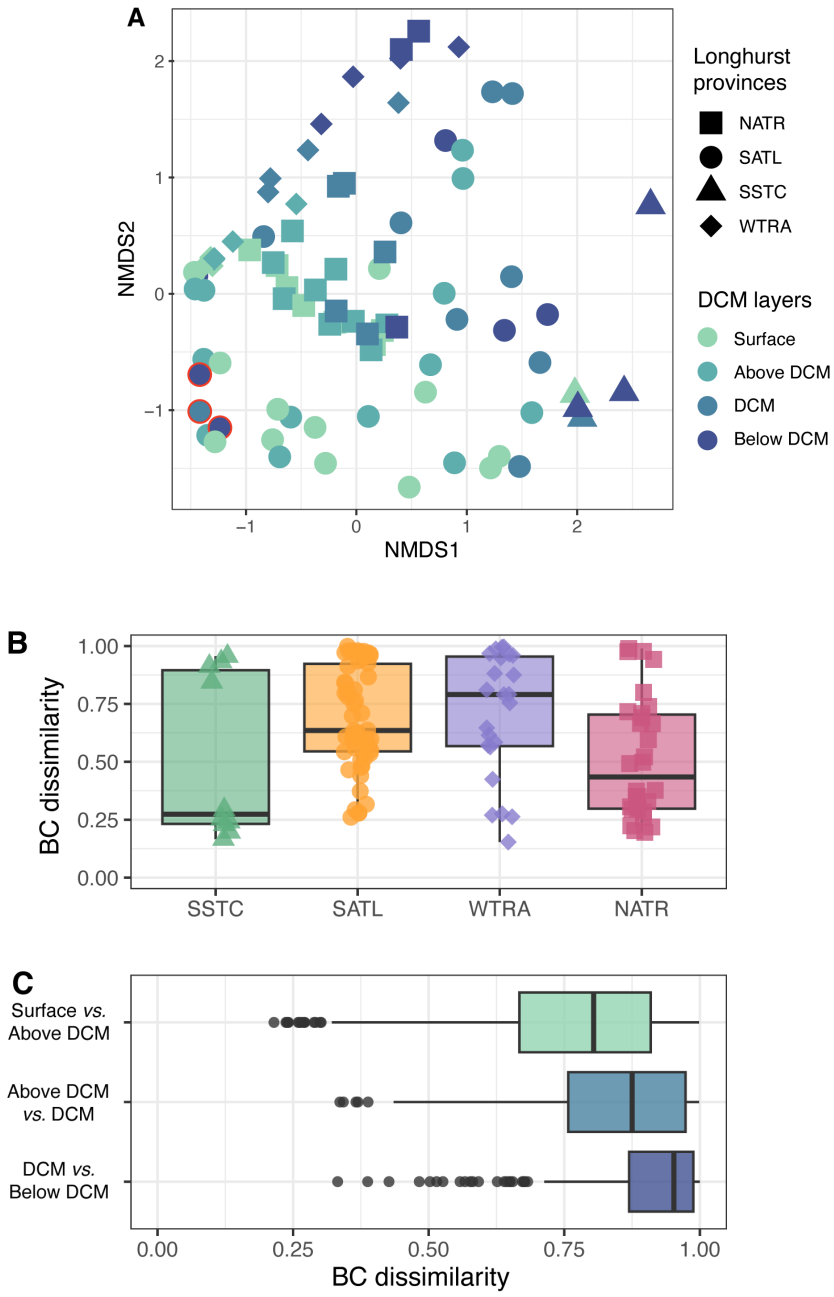

Figure 4. A) Non-metrical multidimensional (nMDS) plot based on the Bray-Curtis (BC) dissimilarities between AAP communities. Samples from the distinct DCM layers and Longhurst provinces are denoted by different color and shape respectively. The red border highlight three samples from deep waters that are clustering with surface samples. B) BC dissimilarity across samples from the same Longhurst province. C) BC dissimilarity between samples from one layer versus samples from the layer below.

dependency of AAP bacteria on phytoplankton derived-dissolved organic matter (DOM). Given the relatively larger size of AAP bacteria compared to the average bacteria (Sieracki et al., 2006), regions with higher DOM availability could sustain increased abundances of AAP bacteria. In this context, considering the variability within AAPs, it is reasonable to anticipate that AAP communities could covary with distinct phytoplanktonic communities. Although evidence supporting this hypothesis is primarily drawn from abundance and pigment patterns, to our knowledge, a comprehensive analysis comparing the composition of phytoplankton and AAP bacterial assemblages has not yet been undertaken. To address this gap, we compared the composition of AAP assemblages with that of prokaryotic and eukaryotic assemblages for a subset of samples within our dataset. Mantel test correlations, based on BC similarities, revealed a robust and statistically significant correlations of AAP bacteria community structure with that of eukaryotes in the three studied size fraction and with the bulk prokaryotic structure associated to the microplankton (20–200 µm) (Table S1, Supplementary Information). In addition, AAP bacteria and the micro-phytoplankton (cyanobacteria and eukaryotic phytoplankton from the microplankton communities) also had a significant positive correlation with AAP communities (Table S1). The microplankton fraction encompass large eukaryotes (both phototrophic and non-phototrophic) and particles that can contain small eukaryotes and prokaryotes. The composition of free-living bacteria and particle attached bacteria is very different (Milici et al., 2017; Suzuki et al., 2017; Grossart, 2010; Crespo et al., 2013; Rieck et al., 2015; Salazar et al., 2015; Mestre et al., 2017a, 2017b), which could explain the contrasting values obtained for the different prokaryotic size fractions. As these correlations are based on dissimilarity values, these results suggest that the patterns observed in the AAP assemblages mirror those of phytoplankton (cyanobacteria and eukaryotic phototrophs), which could be attributed to the dependency of AAP on these phytoplanktonic communities or to the similar responses of both groups to the environmental conditions in those samples.

To gain further insights, we explored the co-occurrence networks of phytoplankton vs. AAP bacteria and phytoplankton vs. bacterioplankton. Co-ocurrence networks allow to make predictions on species interactions (Faust and Raes, 2012), and go beyond environmental factors, which often prove incomplete predictors of community structure configurations in the photic zone (Lima-Méndez et al., 2015; Yeh & Fuhrman, 2022). We observed that among all possible ASV co-occurrences between phytoplankton and AAP bacteria or heterotrophic bacterioplankton, the percentage of significant co-occurrences was higher in the case of AAPs with phytoplankton (0.33% of observed co-occurrences

in AAP bacteria vs 0.22% of observed co-occurrences with bacterioplankton), when considering strength values over 0.4 (MICe>0.4, Table S2). The higher percentage in the case of AAPs persisted even when considering only the most strong co-occurrences detected (MICe over 0.6, 0.7, and 0.8, Table S2). While we refrain from assuming co-occurrences as interactions between species, the observed results suggest that AAP species are more tightly linked to phytoplankton than the bulk bacterioplankton, thus reinforcing the idea of a strong coupling between these two groups. While this trend represents a preliminary insight, further experimental data will be crucial to understand the dependance of AAPs on phytoplankton communities.

3.4 Conclusions

This study provides the first analysis of AAP bacteria communities across fine-scale vertical gradients. While there has been a growing knowledge on the ecological patterns of AAP bacteria in the surface ocean, the vertical component has been generally overlooked. We show that communities vary along the DCM profile, with a high percentage of AAP species being exclusive from layers below the surface. A pervasive pattern of high richness at the DCM was observed across areas of contrasting productivity, which indicates that regardless of the oceanic region, AAP communities are more diverse at this layer. All in all, this study show that the DCM structure influences the diversity, taxonomic composition, and community structure of AAP bacteria. The use of primers with a high phylogenetic coverage allowed to retrieve several phylogroups of AAP bacteria that had been missed in previous approaches and that dominate AAP assemblages at least in the open ocean. The analysis across the horizonal and vertical axis reveals that these groups have contrasting distribution and habitat preferences. Finally, we provide insights of the possible coupling between AAP bacteria and phytoplankton communities, based on their communities composition, which should be further evaluated.

3.5 Acknowledgments

We thank all the crew of the R/V *Sarmiento de Gamboa*, as well as the scientific team of the Poseidon cruise. We would also like to thank Anna Águila Griful, Jose Duch Hurtado, Paula Sabaté, and Oriol Puig for help with DNA extraction and 16S rRNA data generation. This work was supported by grants ECLIPSE (PID2019-110128RB-I00/AEI/10.13039/501100011033) to IF, MICOLOR to JMG (PID2021-125469NB-C31) and OS (PID2021-125469NB-C32), and POSEIDON (CTM2017-84735-R) to JMA, funded by the

Agencia Estatal de Investigación from the Spanish Ministry of Science and Innovation. Authors affiliated to the Institut de Ciències del Mar received the institutional support of the 'Severo Ochoa Centre of Excellence' accreditation (CEX2019-000928-S). CRG was supported by a PIF fellowship from the Universitat Autònoma de Barcelona.

3.6 References

Acinas, S.G., Antón, J., & Rodríguez-Valera, F. (1999) Diversity of free-living and attached bacteria in offshore western Mediterranean waters as depicted by analysis of genes encoding 16S rRNA. *Appl. Environ. Microbiol.* 65, 514–522. https://doi.org/10.1128/aem.65.2.514-522.

Acinas, S.G. Sánchez, P., Salazar, G., Cornejo-Castillo, F.M., Sebastián, M., Logares, R., Royo-Llonch, M., Paoli, L., Sunagawa, S., Hingamp, P., et al. (2021) Deep ocean metagenomes provide insight into the metabolic architecture of bathypelagic microbial communities. *Commun. Biol.* 4, 1–15. https://doi.org/10.1038/s42003-021-02112-2

Albanese, D., Riccadonna, S., Donati, C., & Franceschi, P. (2018) A practical tool for maximal information coefficient analysis. *GigaScience* 7:giy032

Auladell, A., Sánchez, P., Sánchez, O., Gasol, J.M., & Ferrera, I. (2019) Long-term seasonal and interannual variability of marine aerobic anoxygenic photoheterotrophic bacteria. *ISME J* 13: 1975–1987. https://doi.org/10.1038/s41396-019-0401-4

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. *Syst Biol* 68: 365–369. https://doi.org/10.1093/sysbio/syy054

Bibiloni-Isaksson, J., Seymour, J.R., Ingleton, T., van de Kamp, J., Bodrossy, L., & Brown, M.V. (2016) Spatial and temporal variability of aerobic anoxygenic photoheterotrophic bacteria along the East coast of Australia. *Environ Microbiol* 18: 4485–4500. https://doi.org/10.1111/1462-2920.13436

Boeuf, D., Cottrell, M.T., Kirchman, D.L., Lebaron, P., & Jeanthon, C. (2013) Summer community structure of aerobic anoxygenic phototrophic bacteria in the western Arctic Ocean. *FEMS Microbiol Ecol* 85: 417–432. https://doi.org/10.1111/1574-6941.12130

Cabello, A.M., Latasa, M., Forn, I, Morán, X.A.G. & Massan R. (2016) Vertical distribution of major photosynthetic picoeukaryotic groups in stratified marine waters. *Environ. Microbiol.* 18: 1578–1590. doi:10.1111/1462-2920.13285

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. *Nat Methods* 13: 581–583. https://doi.org/10.1038/nmeth.3869.

Catalá, T. S., Álvarez-Salgado, X. A., Otero, J., Iuculano, F., Companys, B., Horstkotte, B., Romera-Castillo, C., Nieto-Cid, M., Latasa, M., Morán X.A.G., et al. (2016) Drivers of fluorescent

dissolved organic matter in the global epipelagic ocean. *Limnol Oceanogr* 61, 1101–1119. https://doi.org/10.1002/lno.10281

Crespo, B.G., Pommier, T., Fernández-Gómez, B., & Pedrós-Alió, C. (2013) Taxonomic composition of the particle-attached and free-living bacterial assemblages in the Northwest Mediterranean Sea analyzed by pyrosequencing of the 16S rRNA. *MicrobiologyOpen* 2:541–552.

Cottrell, M.T., Mannino, A., & Kirchman, D.L. (2006) Aerobic anoxygenic phototrophic bacteria in the Mid-Atlantic Bight and the North Pacific Gyre. *Appl Environ Microbiol* 72: 557–564. doi:10.1128/AEM.72.1.557-564.2006

Cullen, J.J. (1982) The Deep Chlorophyll Maximum: Comparing Vertical Profiles of Chlorophyll-a. *Can. J. Fish. Aquat. Sci.* 39, 791–803. https://doi.org/10.1139/f82-108

Cullen, J.J. (2015) Subsurface chlorophyll maximum layers: Enduring enigma or mystery solved?, *Annu. Rev. Mar. Sci.*, 7, 207–239

Estrada, M., Marrase, C., Latasa, M., Berdalet, E., Delgado, M., & Riera, T. (1993) Variability of deep chlorophyll máximum characteristics in the northwestern Mediterranean. *Mar Ecol Prog Ser* 92: 289–300.

Faust, K., & Raes, J. (2012) Microbial interactions: from networks to models. *Nat Rev Microbiol* 10, 538–550 . https://doi.org/10.1038/nrmicro2832

Ferrera, I., Borrego, C.M., Salazar, G., & Gasol, J.M. (2014) Marked seasonality of aerobic anoxygenic phototrophic bacteria in the coastal NW Mediterranean Sea as revealed by cell abundance, pigment concentration and pyrosequencing of *pufM* gene. *Environ Microbiol* 16: 2953–2965. https://doi.org/10.1038/s41598-018-22413-7

Frank, A.H., Garcia, J.A.L., Herndl, G.J., & Reinthaler, T. (2016) Connectivity between surface and deep waters determines prokaryotic diversity in the North Atlantic deep water. *Environ Microbiol* 18: 2052–2063. https://doi.org/10.1111/1462-2920.13237

Friedline, C.J., Franklin, R.B., McCallister, S.L., & Rivera, M.C. (2012) Bacterial assemblages of the Eastern Atlantic Ocean reveal both vertical and latitudinal biogeographic signatures. *Biogeosciences* 9: 2177–2193. https://doi.org/10.5194/bg-9-2177-2012

Fuhrman JA. Steele, JA., Hewson, I., Schwalbach, M.S., Brown, M.V., Green, J.L., & Brown, J.H. (2008) A latitudinal diversity gradient in planktonic marine bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 105, 7774–7778. doi: 10.1073/pnas.0803070105; pmid: 18509059

Gu, Z., Gu, L., Eils, R., Schlesner, M., & Brors, B., (2014) circlize Implements and enhances circular visualization in R. *Bioinformatics* 30(19), 2811–2812. https://doi.org/10.1093/bioinformatics/btu393

Gasol, J.M., & Morán, X.A.G. (2015). Flow Cytometric Determination of Microbial Abundances and Its Use to Obtain Indices of Community Structure and Relative Activity. In: McGenity, T.J., Timmis, K.N., Nogales, B. (eds) *Hydrocarbon and Lipid Microbiology Protocols*. Springer

Protocols Handbooks. Springer, Berlin, Heidelberg. https://doi.org/10.1007/8623_2015_139

Gazulla, C.R., Auladell, A., Ruiz-González, C., Junger, P.C., Royo-Llonch, M., Duarte, C.M., Gasol, J.M., Sánchez, O., & Ferrera, I. (2022) Global diversity and distribution of aerobic anoxygenic phototrophs in the tropical and subtropical oceans. *Environ Microbiol* 24:2222–2238. https://doi.org/10.1111/1462-2920.15835

Gazulla, C.R., Cabello, A.M., Sánchez, P., Gasol, J.M., Sánchez, O., & Ferrera, I. (2023) A Metagenomic and Amplicon Sequencing Combined Approach Reveals the Best Primers to Study Marine Aerobic Anoxygenic Phototrophs. *Micro Ecol* 86(3), 2161–2172. https://doi.org/10.1007/s00248-023-02220-y

Ghiglione, J.-F., Galand, P.E., Pommier, T., Pedrós-Alió, C., Maas, E.W., Bakker, K., Bertilson, S., Kirchman, D.L., Lovejoy, C., Yager, P.L., et al. (2012). Pole-to-pole biogeography of surface and deep marine bacterial communities. *Proc. Natl. Acad. Sci.* 109, 17633–17638. doi: 10.1073/pnas.1208160109

Gómez-Consarnau, L., J. M. González, M. Coll-Lladó., Gourdon, P., Pascher, T., Neutze, R., Pedrós-Alió, C., & Pinhassi, J. (2007) Light stimulates growth of proteorhodopsin-containing marine Flavobacteria. *Nature* 445: 210–213.

Gómez-Consarnau, L., Raven, J.A., Levine, N.M., Cutter, L.S., Wang, D., Seegers, B., Arístegui, J., Fuhrman, J.A., Gasol, J.M., & Sañudo-Wilhelmy., S.A. (2019) Microbial rhodopsins are major contributors to the solar energy captured in the sea. *Sci Adv* 5:1–8. doi:10.1126/sciadv.aaw8855

Grossart, H.P. (2010) Ecological consequences of bacterioplankton lifestyles: Changes in concepts are needed. *Environ. Microbiol. Rep.* 2:706–714. doi: https://doi.org/10.1111/j.1758-2229.2010.00179.x

Haro-Moreno, J.M., López-Pérez, M., de la Torre, J.R., Picazo, A., Camacho, A., & Rodriguez-Valera, F., (2018) Fine metagenomic profile of the Mediterranean stratified and mixed water columns revealed by assembly and recruitment. *Microbiome* 6, 1–19. https://doi.org/10.1186/s40168-018-0513-5

Hillebrand H. (2004) On the generality of the latitudinal diversity gradient. *Am. Nat.* 163(2), 192–211. https://doi.org/10.1086/381004

Hojerová, E., Mašín, M., Brunet, C., Ferrera, I., Gasol, J.M., & Koblížek, M. (2011) Distribution and growth of aerobic anoxygenic phototrophs in the Mediterranean Sea. *Environ Microbiol* 13: 2717–2725. doi:10.1111/j.1462-597 2920.2011.02540.x

Ibarbalz FM., Henry, N., Brandao, MC., Martini, S., Busseni, G., Byrne, H., Coelho L.P., Endo, H., Gasol, J.M., Gregory, A.C., Mahé, F. et al,. (2019) Global Trends in Marine Plankton Diversity across Kingdoms of Life. *Cell* 179, 1084–1097

Irwin, A.J., & Oliver, M.J. (2009) Are ocean deserts getting larger? *Geophys. Res. Lett.* 36. doi: 10.1029/2009gl039883

Jeanthon, C., Boeuf, D., Dahan, O., le Gall, F., Garczarek, L., Bendif, E.M., & Lehours, A.C. (2011) Diversity of cultivated and metabolically active aerobic anoxygenic phototrophic bacteria along an oligotrophic gradient in the Mediterranean Sea. *Biogeosciences* 8: 1955–1970. https://doi.org/10.5194/bg-8-1955-2011

Jiao, N., Zhang, Y., Zeng, Y., Hong, N., Liu, R., Chen, F., & Wang, P. (2007) Distinct distribution pattern of abundance and diversity of aerobic anoxygenic phototrophic bacteria in the global ocean. *Environ Microbiol* 9: 3091–3099. doi:10.1111/j.1462-2920.2007.01419.x

Junger, PC., Sarmento, H., Giner, C.R., Mestre, M., Sebastián, M., Morán, X.A.G., Arístegui, J., Agustí, S., Duarte C.M., Acinas, S.G. et al., (2023) Global biogeography of the smallest plankton across ocean depths. *Sci. Adv*. 9,eadg9763.DOI:10.1126/sciadv.adg9763

Latasa, M., Cabello, A.M., Morán, X.A.G., Massana, R., & Scharek., R. (2017) Distribution of phytoplankton groups within the deep chlorophyll máximum. *Limnol Oceanogr* 62,2, 665-685. https://doi.org/10.1002/lno.10452

Lehours, A.C., Cottrell, M.T., Dahan, O., Kirchman, D.L., & Jeanthon, C. (2010) Summer distribution and diversity of aerobic anoxygenic phototrophic bacteria in the Mediterranean Sea in relation to environmental variables. *FEMS Microbiol Ecol* 74: 397–409. https://doi.org/10.1111/j.1574-6941.2010.00954.x

Lehours, A.C., Enault, F., Boeuf, D., & Jeanthon, C. (2018) Biogeographic patterns of aerobic anoxygenic phototrophic bacteria reveal an ecological consistency of phylogenetic clades in different oceanic biomes. *Sci Rep* 8: 1–10. https://doi.org/10.1038/s41598-018-22413-7

Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., Chaffron, S., Ignacio-Espinosa, J. C., Roux, S., Vincent, F., et al., (2015) Ocean plankton. Determinants of community structure in the global plankton interactome. *Science* 348(6237), 1262073. https://doi.org/10.1126/science.1262073

Logares, R., Deutschmann, I.M., Junger, P.C., Giner, C.R., Krabberød, A.K., Schmidt, T.S.B., Rubinat-Ripoll, L., Mestre, M., Salazar, G., Ruiz-González, C., et al. (2020) Disentangling the mechanisms shaping the surface ocean microbiota. *Microbiome* 8:55 1–17. https://doi.org/10.1186/s40168-020-00827-8

Longhurst, A. (2007) Ecological Geography of the Sea. 2nd Edition, San Diego, CA: Academic Press. https://doi.org/10.1016/B978-0-12-455521-1.X5000-1

Lorenzo, L.M., Figueiras, F.G., Tilstone, G.H., Arbones, B., & Mirón, I., (2004) Photosynthesis and light regime in the Azores Front region during summer: are light-saturated computations of primary production sufficient? *Deep Sea Res. Part I Oceanogr. Res. Pap.* 51, 1229–1244.

Malviya, S., Scalco, E., Audic, S., Vincent, F., Veluchamy, A., Poulain, J., Wincker, P., Iudicone, D., de Vargas, C., Bittner, L., et al. (2016) Insights into global diatom 640 distribution and diversity in the world's ocean. *Proc Natl Acad Sci U.S.A.* 113(11):E1516-25. doi:10.1073/pnas.1509523113

Marañon E., Wambeke, FV., Uitz, J., Boss, E.S., Dimier, C., Dinasquet, J., Engel, A., Haentjens, N., Pérez-Lorenzo, M., Taillandier, V., & Zancker B. (2021) Deep maxima of phytoplankton biomass, primary production and bacterial production in the Mediterranean Sea. *Biogeosciences* 18, 1749–1767 https://doi.org/10.5194/bg-18-1749-2021

Marin, Carolina. (2023) Patterns of connectivity in marine microbial communities. Doctoral thesis.

Martin, M. (2013) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17: 10. https://doi.org/10.14806/ej.17.1.200

Massana, R., Murray, A.E., Preston, C.M., & DeLong, E.F. (1997) Vertical distribution and phylogenetic characterization of marine planktonic. *Microbiology* 63:50–56. https://doi.org/10.1128/aem.63.1.50-56.1997

Mestre, M., Borrull, E., Sala, M., & Gasol, J.M. (2017a) Patterns of bacterial diversity in the marine planktonic particulate matter continuum. *ISME J.* 11: 999–1010.

Mestre, M., Ferrera, I., Borrull, E., Ortega-Retuerta, E., Mbedi, S., Grossart, H.P. et al. (2017b) Spatial variability of marine bacterial and archaeal communities along the particulate matter continuum. *Mol. Ecol.* 26: 6827–6840.

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. *Proc Natl Acad Sci U.S.A.* 115: E6799–E6807. https://doi.org/10.1073/pnas.1802470115.

Milici, M., Tomasch, J., Wos-Oxley, M.L., Decelle, J., Jáuregui, R., Wang, H., Deng, Z.-L., Plumeier, I., Helge-Ansgar, G., Badewien, T.H., et al. (2016) Bacterioplankton biogeography of the Atlantic Ocean: a case study of the distance-decay relationship. *Front Microbiol* 7: 1–15. https://doi.org/10.3389/fmicb.2016.00590

Milici, M., Vital, M., Tomasch, J., Badewien, T.H., Giebel, H.A., Plumeier, I., Wang, H., Pieper, D.H., Wagner-Döbler, I., & Simonet, M. (2017) Diversity and community composition of particle-associated and free-living bacteria in mesopelagic and bathypelagic Southern Ocean water masses: evidence of dispersal limitation in the Bransfield Strait. *Limnol. Oceanogr.* 62, 1080–1095

Morris, R. M., B. L. Nunn, C. Frazar, Goodlett, D.R., Ting, Y.S., & Rocap, G. (2010) Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. *ISME J.* 4:673–685.

Oksanen, J., Simpson, G., Blanchet, F. et al. (2022) vegan: Community Ecology Package. R package version 2.6–2. https://cran.r-project.org/package=vegan.

Harrell F., & Dupont C. (2023). Hmisc: Harrell Miscellaneous. Available at: https://cran.r-project.org/web/packages/Hmisc/index.html

Rieck, A., Herlemann, D.P.R., Jürgens, K. & Grossart, H.P. (2015) Particle-associated differ from

free-living bacteria in surface waters of the baltic sea. *Front. Microbiol.* 6: 1297.

Ritchie, A.E., Johnson, Z.I. (2012) Abundance and genetic diversity of aerobic anoxygenic phototrophic bacteria of coastal regions of the Pacific Ocean. *Appl Environ Microb* 78: 2858–66.

Ruiz-González, C., Logares, R., Sebastián, M., Mestre, M., Rodríguez-Martínez, R., Galí, M., Sala, M., Acinas, S.G., Duarte, C.M., Gasol, J.M. (2019) Higher contribution of globally rare bacterial taxa reflects environmental transitions across the surface ocean. *Mol Ecol* 28:1930–1945. https://doi.org/10.1111/mec.15026

Ruiz-González, C., Mestre, M., Estrada, M., Sebastián, M. Salazar, G., Agustí, S., Moreno-Ostos, E., Reche, I., Álvarez-Salgado, X.A., Morán, X.A.G., et al. (2020) Major imprint of surface plankton on deep ocean prokaryotic structure and activity. *Mol Ecol* 29: 1820–1838. https://doi.org/10.1111/mec.15454

Salazar, G., Cornejo-Castillo, F.M., Borrull, E., Díez-Vives, C., Lara, E., Vaqué, D. et al. (2015) Particle-association lifestyle is a phylogenetically conserved trait in bathypelagic prokaryotes. *Mol. Ecol.* 24: 5692–5706.

Salazar, G., Cornejo-Castillo, F.M., Benítez-Barrios, V., Fraile-Nuez, E., Álvarez-Salgado, X.A., Duarte, C.M., Gasol, J.M., Acinas, S.G., (2016) Global diversity and biogeography of deep-sea pelagic prokaryotes. *ISME J.* 10, 596–608. https://doi.org/10.1038/ismej.2015.137

Sanchez, P., Sebastián, M., Pernice, M. Rodríguez-Martínez, R., Pesant, S., Agustí, S., Gojobori, T., Logares, R., Sala, M.M. et al. (2023) Marine picoplankton metagenomes from eleven vertical profiles obtained by the Malaspina Expedition in the tropical and subtropical oceans. bioRxiv. https://doi.org/10.1101/2023.02.06.526790

Sebastián, M., Ortega-Retuerta, E., Gómez-Consarnau, L., Zamanillo, M., Álvarez, M., Arístegui, J., Gasol, J.M. (2021) Environmental gradients and physical barriers drive the basin-wide spatial structuring of Mediterranean Sea and adjacent eastern Atlantic Ocean prokaryotic communities. *Limnol. Oceanogr.* 66, 4077–4095. https://doi.org/10.1002/lno.11944

Sharples, J., Moore, C.M., Rippeth, T.P., Holligan, P.M., Hydes, D.J., Fisher, N.R., Simpson, J.H. (2001) Phytoplankton distribution and survival in the thermocline. *Limnol Oceanogr* 46: 486–496.

Sieracki, M.E., Gilg, I.C. Thier, E.C., Poulton, N.J., Goericke, R. (2006) Distribution of planktonic aerobic anoxygenic photoheterotrophic bacteria in the northwest Atlantic. *Limnol Oceanogr* 51: 38–46. doi:10.4319/lo.2006.51.1.0038

Sul, W.J., Oliver, T.A., Ducklow, H.W., Amaral-Zettler, A.L., Sogin, M.L. (2013) Marine bacteria exhibit a bipolar distribution. *Proc. Natl. Acad. Sci. U.S.A.* 110, 2342–2347

Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G., et al. (2015) Structure and function of the global ocean microbiome. *Science* 348: 1–10. https://doi.org/10.1126/science.1261359

Suzuki, S., Kaneko, R., Kodama, T., Hashihama, F., Suwa, S., Tanita, I., Furuya, K., & Hamasaki, K. (2017) Comparison of community structures between particle-associated and free-living prokaryotes in tropical and subtropical Pacific Ocean surface waters. *J Oceanogr* 73: 383–395.

Tittensor, D.P. Mora, C., Jetz, W., Lotze, H.K., Ricard, D., Berghe, E.V., & Worm, B. (2010) Global patterns and predictors of marine biodiversity across taxa. *Nature* 466, 1098–1101 doi: 10.1038/nature09329; pmid: 20668450

Walsh, E.A., Smith, D.C., Sogin, M.L., & D'Hondt, S. (2015) Bacterial and archaeal biogeography of the deep chlorophyll maximum in the South Pacific Gyre. *Aquat. Microb. Ecol.* 75, 1–13. https://doi.org/10.3354/ame01746

Weston, K., Fernand, L., Mills, D.K., Delahunty, R., & Brown, J. (2005) Primary production in the deep chlorophyll maximum of the central North Sea. *J. Plankton Res.* 27, 909–922. https://doi.org/10.1093/plankt/fbi064

Yeh, YC. & Fuhrman, J.A. (2022) Contrasting diversity patterns of prokaryotes and protists over time and depth at the San-Pedro Ocean Time series. *ISME COMMUN.* 2, 36 https://doi.org/10.1038/s43705-022-00121-8

Yutin, N., Suzuki, M.T. & Béjà, O. (2005) Novel primers reveal wider diversity among marine aerobic anoxygenic phototrophs. *Appl Environ Microb* 71: 8958–8962.

Yutin, N., Suzuki, M.T., Teeling, H., Weber, M., Venter, J.C., Rusch, D.B., & Béjà, O. (2007) Assessing diversity and biogeography of aerobic anoxygenic phototrophic bacteria in surface waters of the Atlantic and Pacific oceans using the Global Ocean sampling expedition metagenomes. *Environ Microbiol* 9: 1464–1475. https://doi.org/10.1111/j.1462-2920.2007.01265.x

Yurkov, V.V. & Csotonyi, J.T. (2009) New light on aerobic anoxygenic phototrophs. In: Hunter CN, Daldal F, Thurnauer MC, et al. (eds). *The Purple Phototrophic Bacteria. Advances in Photosynthesis and Respiration*, Vol. 28. Dordrecht, the Netherlands: Springer, 31–55.

3.7 Supplementary figures

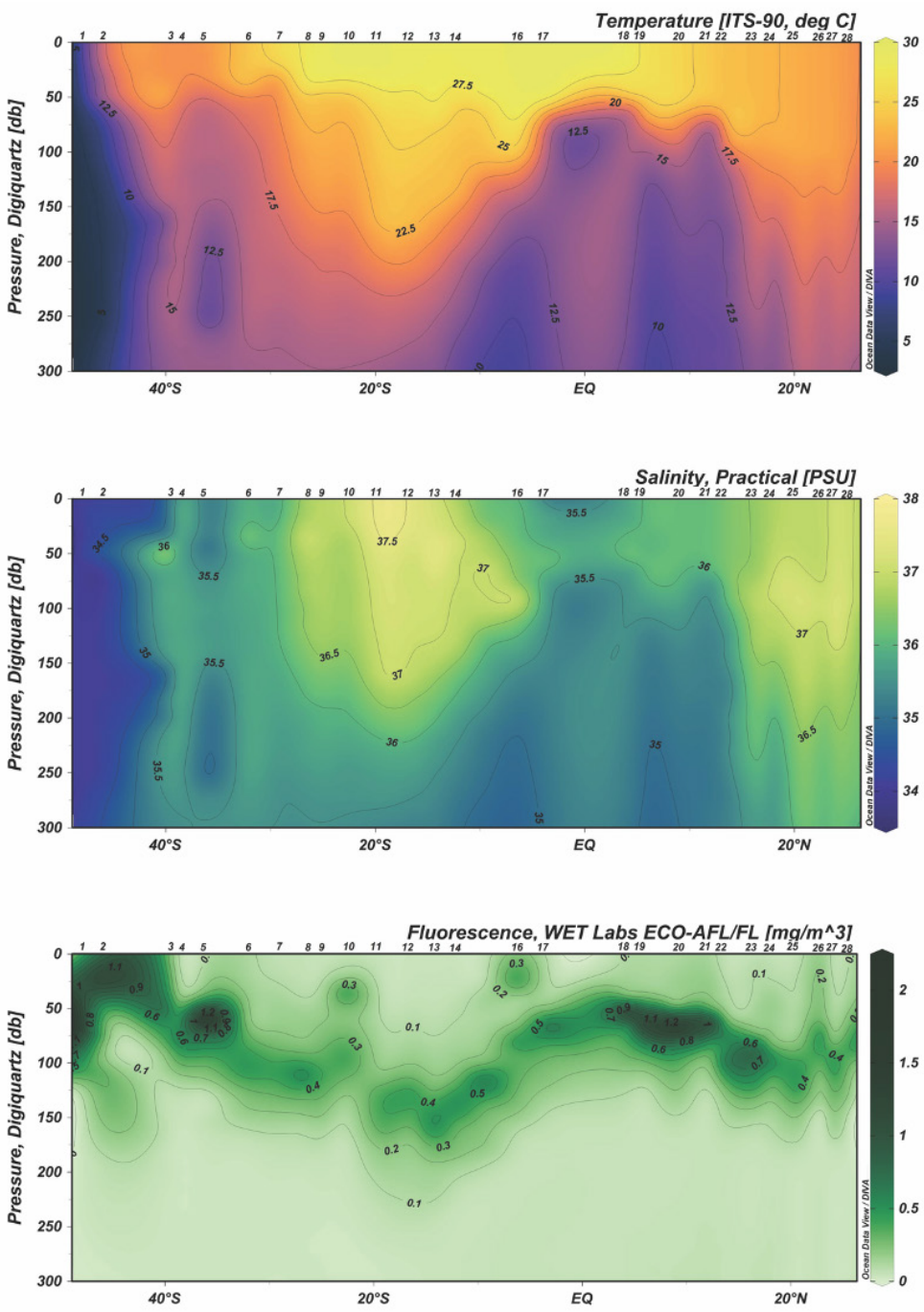

Figure S1. Section distribution of temperature, salinity, and fluorescence along the South And Mid Atlantic Ocean. Data extracted from the CTD SBE 911plus files.

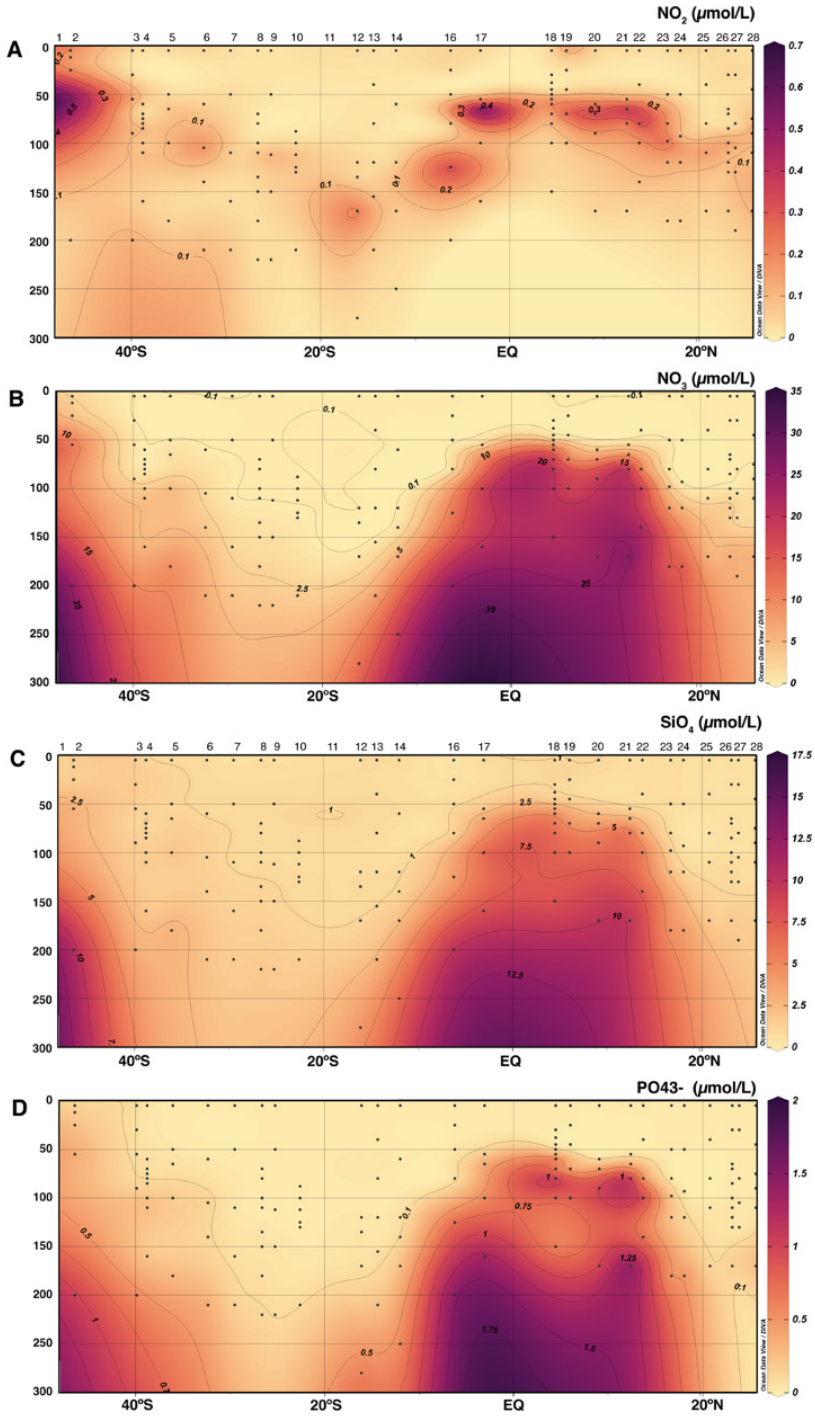

Figure S2. Section distribution of nitrite (NO_2^-), nitrate (NO_3^-), silicate (SiO_4^{4-}) and phosphate (PO_4^{-3}) along the South And Mid Atlantic Ocean. Black dots in all panels indicate the depths of the discrete samples.

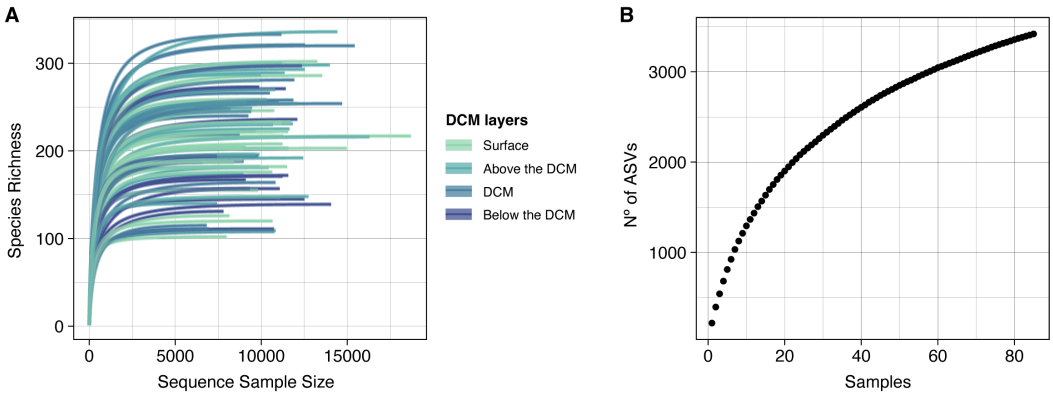

Figure S3. A) Rarefaction curves for each sample, colour-coded by layers along the DCM structure. B) Sample-based rarefaction curve, considering all samples collectively.

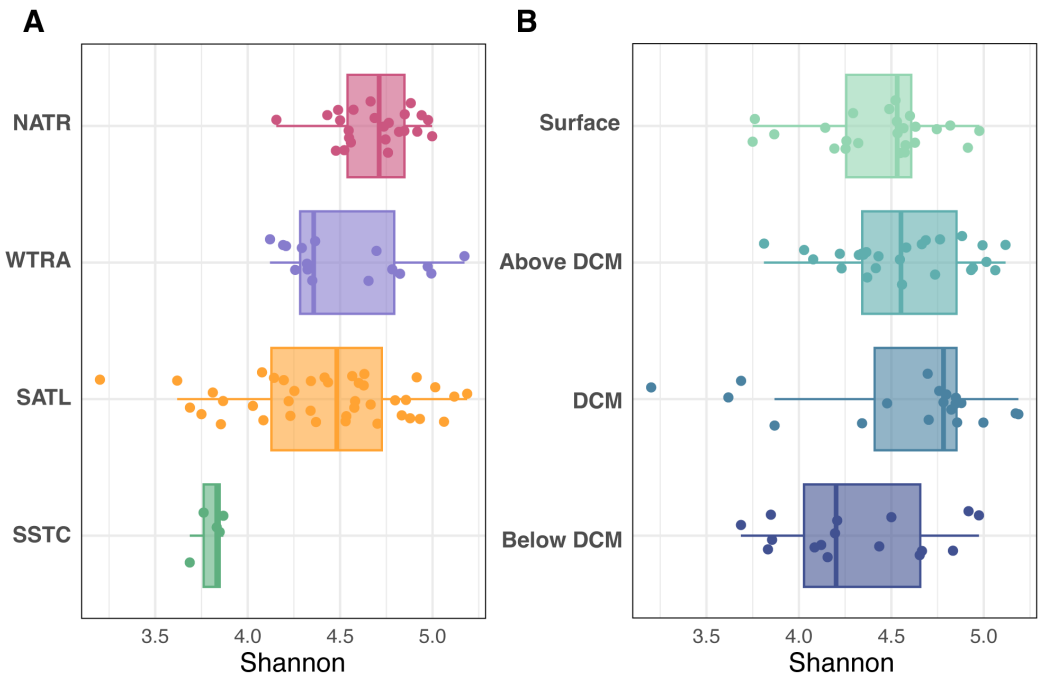

Figure S4. Shannon index value distribution along the horizontal and vertical axis.

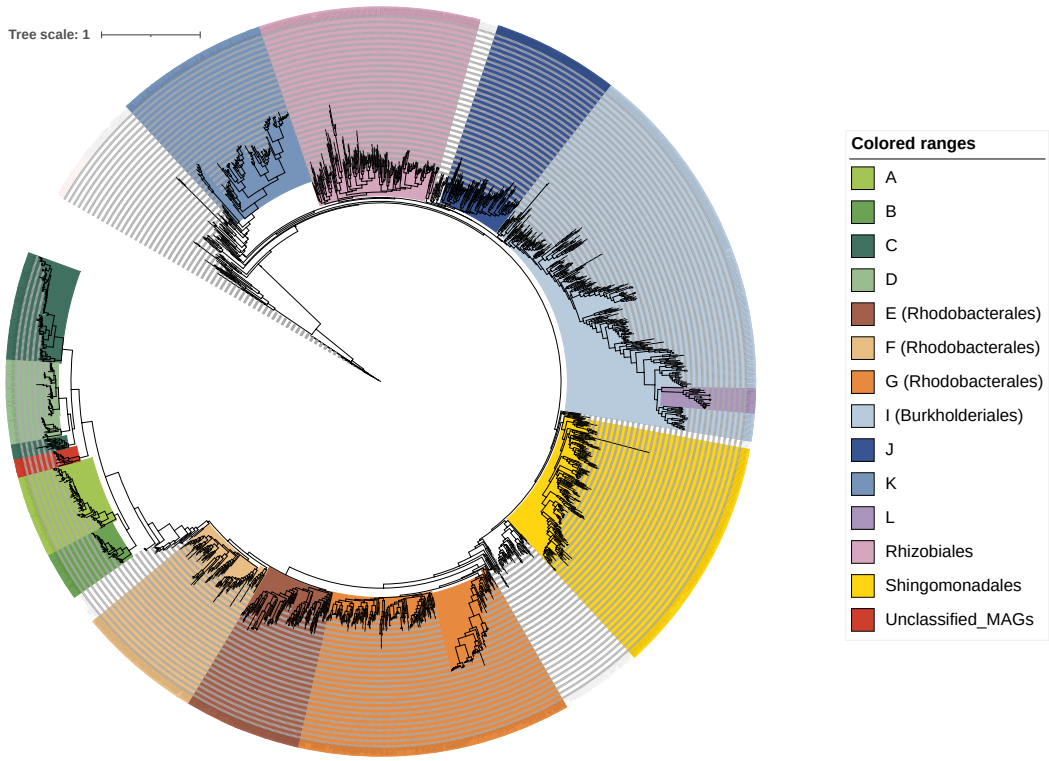

Figure S5. Phylogenetic tree of *pufM* sequences. The tree was constructed with RAxML v8.2 (Stamakis 2014) and visualized using iTOL (Letunic and Bork, 2011). Legend in the left shows the taxonomic groups defined in this study.

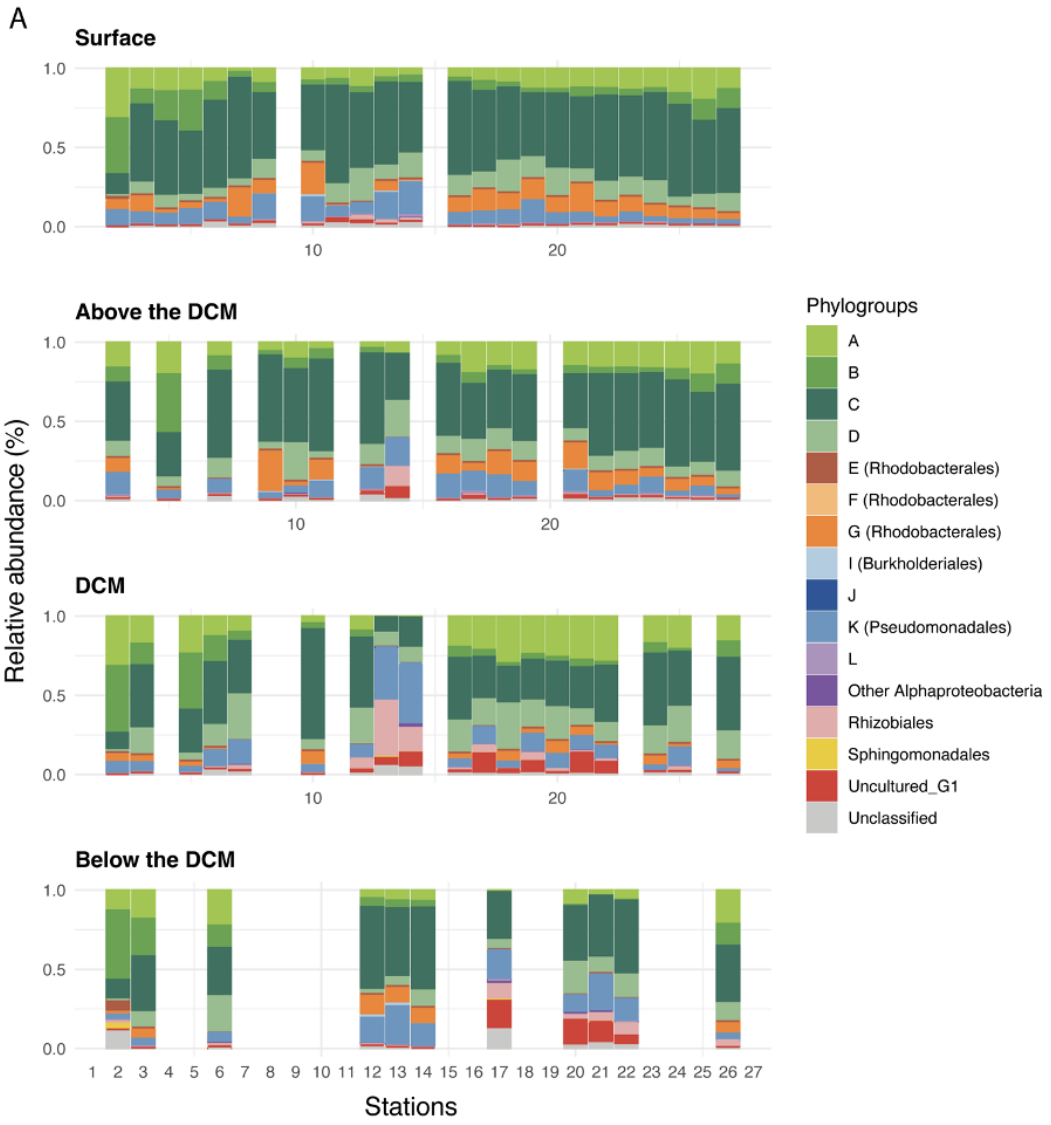

Figure S6. Relative abundance of the main taxonomic groups within the deep chlorophyll maxima (DCM) layers. Absent columns correspond to samples that could not be successfully amplified.

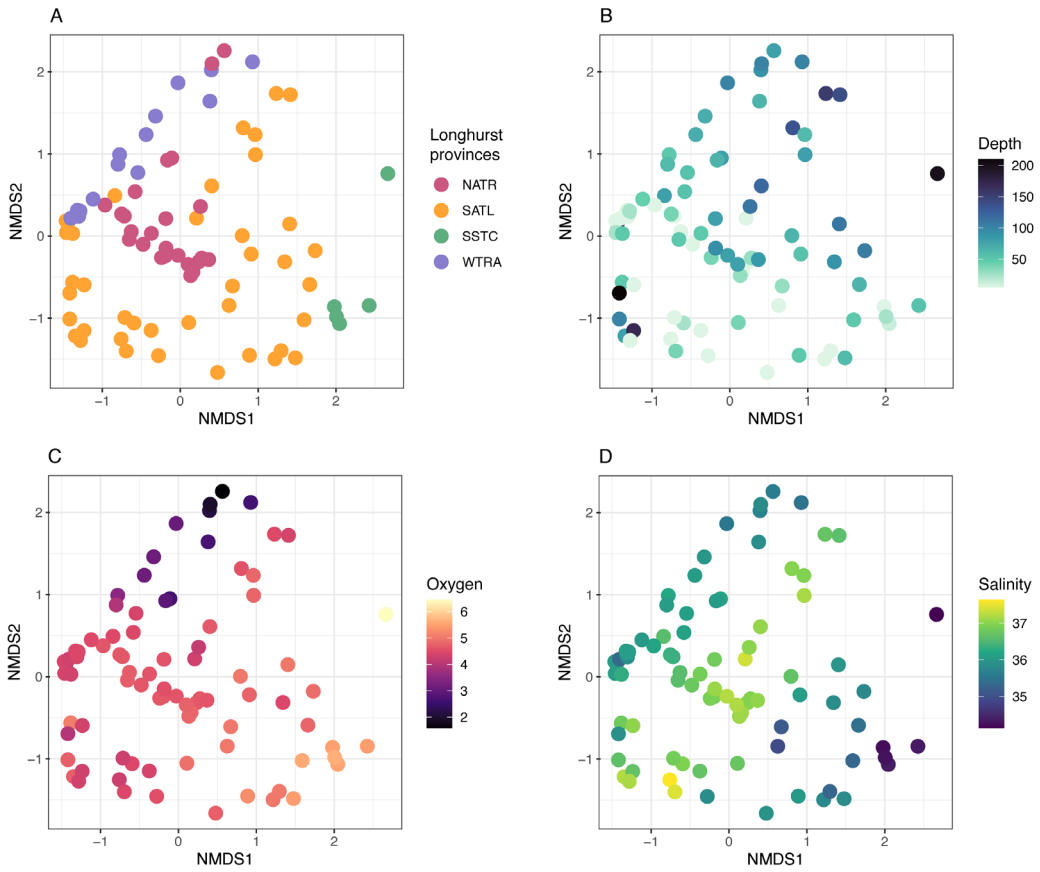

Figure S7. Non-metrical multidimensional (nMDS) plots based on the Bray-Curtis dissimilarities between AAP communities, colored by: A) their belonging to the four different Longhurst provinces, B) sampled depth, C) oxygen concentration (mg/L), and D) salinity values. SSTC, South Subtropical Convergence zone; SATL, the South Atlantic gyre, WTRA, Western tropical Atlantic, NATR, North Atlantic Tropical gyre.

3.8 Supplementary tables

Table S1. Mantel test correlations (Pearson correlations) based on the Bray-Curtis dissimilarity matrices of AAP bacteria and the different microbial groups and size fractions. Significant results (p<0.05) are highlighted in bold. Size fractions are as follows: picoplankton (0.2–3 µm), nanoplankton (3–20 µm), and microplankton (20–200 µm).

AAP bacteria *vs*			
Group	Size fraction	R	p-value
Prokaryotes (16S rRNA)	picoplankton	0.2186	0.1465
	nanoplankton	0.3563	0.0528
	microplankton	**0.4739**	**0.0185**
Cyanobacteria (16S rRNA)	picoplankton	0.1978	0.1652
	nanoplankton	0.0044	0.4858
	microplankton	**0.4854**	**0.0141**
Eukaryotes (18S rRNA)	picoplankton	**0.3699**	**0.047**
	nanoplankton	**0.3911**	**0.0332**
	microplankton	**0.5264**	**0.0074**
Phototrophic eukaryotes (18S rRNA)	picoplankton	0.1846	0.1878
	nanoplankton	0.0381	0.4188
	microplankton	**0.3974**	**0.0258**

Table S2. Summary of significant (p<0.05) co-occurrences between the phytoplanktonic community vs the prokaryotic community and the AAP bacteria community (in bold). MICe: maximal information coefficient.

	MICe	Nº of total significant co-occurrences	% observed interactions
Prokaryotes vs phytoplankton	> 0.4	20122	0.215
AAP bacteria vs phytoplankton		**14641**	**0.328**
Prokaryotes vs phytoplankton	> 0.6	13016	0.139
AAP bacteria vs phytoplankton		**7655**	**0.171**
Prokaryotes vs phytoplankton	> 0.7	5395	0.058
AAP bacteria vs phytoplankton		**3352**	**0.075**
Prokaryotes vs phytoplankton	> 0.8	1860	0.019
AAP bacteria vs phytoplankton		**1146**	**0.025**

3.9 Supplementary Information

The same DNA extracts were used to analyse the 16S rDNA, 18S rDNA and the *pufM* gene. The V4 region of the 18S rRNA gene (~130 bp) was amplified the eukaryotic universal primers TAReukFWD1 and TAReukREV3 (Stoeck et al., 2010), while for the 16S rDNA we used primers 515Y and 926R (Parada et al., 2016). DNA sequencing was conducted on an Illumina MiSeq sequencer by AllGenetics & Biology SL (www.allgenetics.eu).

For the 18S rDNA, we trimmed Illumina raw reads to remove amplification primers using cutadapt v1.16 (Martin, 2011) and processed them with DADA2 v1.12.1 (Callahan et al. 2016). We set DADA2's truncLen to 250,220, maxEE to 2,4 and used pool=TRUE for sample inference. We removed chimeric ASVs with the method *pooled* as implemented in DADA2's function *removeBimeraDenovo*, and removed ASVs shorter than 300 bp. We added group level taxonomy (in general a formal Class) using DADA2's function *assignTaxonomy* using eukaryotesV4 version 2 database (Obiol et al., 2020) with a minimum bootstrap confidence of 50. We assigned trophic mode to eukaryotic taxonomic groups according to common assumptions in the literature (e.g. Bachy et al., 2022). Sequences from the 16S rRNA gene were also processed with cutadapt v1.16 (Martin, 2011) and DADA2 v1.12.1 (Callahan et al. 2016), the taxonomic assignation was done with SILVAr138 (Quast et al., 2013).

References from the Supplementary Information

Bachy, C., Hehenberger, E., Ling, Y.-C., Needham, D. M., Strauss, J., Wilken, S., & Worden, A. Z. (2022). Marine Protists: A Hitchhiker's Guide to their Role in the Marine Microbiome. In L. J. Stal & M. S. Cretoiu (Eds.), *The Marine Microbiome* (pp. 159–241). Springer International Publishing. https://doi.org/10.1007/978-3-030-90383-1_4

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. *Nat Methods* 13: 581–583. https://doi.org/10.1038/nmeth.3869.

Letunic, I., & Bork, P. (2019) Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* 47: W256–W259. https://doi.org/10.1093/nar/gkz239.

Martin, M. (2013) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17: 10. https://doi.org/10.14806/ej.17.1.200

Obiol, A., Giner, C. R., Sánchez, P. Duarte, C.M., Acinas, S.G. & Massana, R. (2020). A metagenomic assessment of microbial eukaryotic diversity in the global ocean. *Mol Ecol Resour* 20(3), 718–731. https://doi.org/10.1111/1755-0998.13147

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. *Environ Microbiol*18(5), 1403–1414. doi: 10.1111/1462-2920.13023

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

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

CHAPTER IV

CHAPTER IV

Vertical distribution and light energy capture by AAP bacteria along contrasted areas of productivity in the Atlantic Ocean

Carlota R. Gazulla, Michal Koblížek, Jesús M. Mercado, Josep M. Gasol, Olga Sánchez, Isabel Ferrera

A manuscript under review in *Limnology & Oceanography*

Abstract:

Aerobic anoxygenic phototrophic (AAP) bacteria are a common part of microbial communities in the sunlit ocean. They contain bacteriochlorophyll *a*-based photosystems that harvest solar energy for their metabolism. Across different oceanic areas and regimes, AAP bacteria seem to be more abundant in eutrophic areas, associated to high chlorophyll concentrations. While most of these studies are based on surface samples, little information is available on their vertical distribution in euphotic zones of the major ocean basins. We hypothesize that AAPs will follow a similar structure to the chlorophyll depth profile across areas with different degrees of stratification. To test this hypothesis, we enumerated AAP cells and determined bacteriochlorophyll *a* concentrations along the photic zone of a latitudinal transect in the South and Mid Atlantic Ocean. We show that the distribution of AAP bacteria is highly correlated to the chlorophyll *a* concentration and the abundance of picophytoplankton across vertical and horizontal gradients. Furthermore, we estimate the light energy captured by AAP bacteria across the water column and find that, while they share a common latitudinal pattern of light capture with the picophytoplankton, they display a unique vertical arrangement with highest photoheterotrophic activity is in the surface ocean.

4.1 Introduction

The sunlit ocean, where there is enough light to sustain primary production, hosts a vast array of microorganisms that drive biogeochemical transformations and sustain life in the global ocean. Phytoplankton and other primary producers provide a continuous supply of dissolved organic matter that supports the growth of heterotrophic bacteria, that incorporate carbon into the base of the microbial food web (Azam et al., 1983). Besides the purely autotrophic and heterotrophic organisms, photoheterotrophic organisms are also key elements of the planktonic food chain since they are capable of harvesting light but rely heterotrophically on organic matter for growth. Alongside proteorhodopsin (PR)-containing bacteria, aerobic anoxygenic phototrophic (AAP) bacteria have attracted the interest of microbial ecologists since they were discovered to be widespread in the surface ocean (Kolber et al., 2000). AAP bacteria usually represent 1 to 7% of total bacteria in the euphotic zone (Koblížek, 2015), but their larger size compared to other bacteria (Sieracki et al., 2006) and higher growth rates (Koblížek et al., 2007; Ferrera et al., 2011; Fecskeová et al., 2021) make them important components of the carbon cycle in the ocean.

AAP bacteria are close relatives of the purple non-sulfur bacteria as both contain bacteriochlorophyll-based reaction centres, although AAP bacteria have a purely aerobic lifestyle and cannot fix carbon. The photoheterotrophic metabolism of AAPs is based on the light harvesting pigment bacteriochlorophyll *a* (Bchl *a*) that facilitates synthesis of ATP, and on organic substrates for growth. This advantageous metabolic combination led to the suggestion that AAP bacteria would be prevalent in oligotrophic ocean regions (Kolber et al., 2000). However, subsequent studies showed that AAP bacteria are actually more abundant in productive regions or shelf seas (Cottrell et al., 2006; Mašín et al., 2006; Sieracki et al., 2006; Hojerová et al., 2011; Ritchie & Johnson 2012; Vrdoljak et al., 2019), coastal lagoons (Lamy et al., 2011), and estuaries (Schwalbach & Fuhrman, 2005; Cottrell et al., 2010), where they can account for up to 12% of the total prokaryotic community. In contrast, in oligotrophic sites such as the Sargasso Sea or the North Pacific gyre their relative contribution is low (<2%, Cottrell et al., 2006; Sieracki et al., 2006; Jiao et al., 2007; Gazulla et al., 2022). While AAPs are present throughout the entire epipelagic zone, most studies have focused on surface samples. The few studies that have dealt with their vertical distribution in the ocean showed that AAP bacteria had abundance trends different than those observed for heterotrophic bacteria or cyanobacteria (Cottrell et al., 2006; Sieracki et al., 2006; Jiao et al., 2007; Lami et al., 2007). Across the Mediterranean Sea, higher AAP abundances (Hojerová et al., 2011) and Bchl *a* concentrations

(Gómez-Consarnau et al., 2019) were observed above and overlapping the deep chlorophyll maxima (DCM). Altogether, these results suggest that AAP bacteria could be associated to phytoplankton not only in the ocean surface but also along its water column.

Additionally, and in contrast to the growing knowledge on the distribution of AAP bacteria, there is little information regarding their phototrophic activity (i.e. light capture) in the natural environment. Laboratory studies with AAP cultures showed that light exposure reduced their respiration rate (Koblížek et al., 2010; Hauruseu & Koblízek 2012), and increased the efficiency of carbon metabolism (Piwosz et al., 2018). In result, the cultures grown on light and dark regime accumulated more biomass than cultures kept in the dark (Hauruseu & Koblížek, 2012, Piwosz et al., 2018). Light stimulation of growth of natural populations of AAP bacteria was also demonstrated in field experiments (Ferrera et al., 2017, Sánchez et al., 2020). In addition, theoretical estimates indicate that AAP bacteria harvest more light energy per cell than PR-containing bacteria (Kirchman & Hanson 2013). However, estimates of actual energy fluxes related to light in the marine environment are scarce. By determining AAP abundance and pigment concentration we can determine the light energy captured by AAP cells, providing information on their in situ physiology and, ultimately, allowing the estimation of their potential phototrophic activity. We studied in detail the distribution of AAP bacteria along the water column in areas of contrasting productivity across the South and Mid Atlantic Ocean. We took samples in 27 stations at different depths following the fluorescence variation depicted by the CTD profiler and enumerated the abundance of AAP bacteria, as well as that of other photosynthetic microorganisms. We hypothesize that the distribution of AAP bacteria along the water column will be strongly connected to the DCM structure, and with higher abundances at more productive sites. In addition, we used an improved high-performance liquid chromatography (HPLC) protocol to accurately detect picomolar (pM) concentrations of Bchl *a* in the water column. Based on the microscopic and pigment data, we estimated the light energy captured by both Bchl *a* and chlorophyll *a* (Chl *a*) and described how this varied spatially and along the depth gradient.

4.2 Methods

Sampling

The POSEIDON Expedition took place between March and April 2019, on board the Spanish RV *Sarmiento de Gamboa*. It covered a distance of 9,000 km across a latitudinal transect (48°S/26°N) in the Atlantic Ocean. A total of 27 stations were sampled at different

depths along the epipelagic zone. A SeaBird 911-plus CTD profiler was used to profile temperature, salinity, conductivity, fluorescence, dissolved oxygen, and other variables. The fluorescence values from the CTD allowed us to define the DCM structure along the water column and, according to this structure, in each station we took samples at the surface, above the DCM, at the DCM peak, below the DCM and at the shallowest depth where chlorophyll fluorescence was undetectable. Six of these stations were sampled at a higher resolution (10 depths across the DCM structure). At two stations the DCM structure displayed two peaks of chlorophyll maxima and samples were taken at each peak as well as between peaks (stations 10 and 16). Stations are numbered from 1 to 28, yet there are no samples from station 15 in our dataset. In total, we collected 172 samples from different depths ranging from 5 m to 280 m. All seawater samples were prefiltered by a 200-µm mesh to remove large plankton. The oceanographic context is in Section 3.3 from Chapter III.

Cell abundances

Samples (1.6 mL) were fixed using a solution of 1% paraformaldehyde + 0.05% glutaraldehyde (final concentrations), deep frozen in liquid nitrogen and stored at –80 °C until analysed. Prokaryotic and photosynthetic picoeukaryotic abundances were determined by flow cytometry as described in Gasol and Morán (2015). Additionally, total bacteria (i.e. prokaryotes) and AAP bacteria were counted using infra-red epifluorescence microscopy (Mašín et al., 2006) with recent modifications (Piwosz et al., 2022). Nine mL of seawater were fixed with formaldehyde to a final concentration of 3.7% and filtered onto white 25 mm polycarbonate filters of 0.2 µm pore size (Nuclepore; Whatman). Cells were stained with 4'6-diamidino-2-phenylindole (DAPI) and counted using an epifluorescence Zeiss Axio Imager.D2 microscope equipped with a Plan-Apochromat 63×/1.46 Oil Corr objective and a Collibri LED module illumination system (Carl Zeiss, Jena, Germany). Images for DAPI fluorescence (total bacteria) were taken under blue emission, the autofluorescence of Chl *a* was recorded in the red part of the spectrum and both AAP bacteria and Chl *a*-containing bacteria were recorded in the infrared part of the spectrum. To obtain net AAP bacterial counts, the contribution of Chl *a*-containing organisms to the infrared image was subtracted (Cottrell et al., 2006). We took at least 10 microphotographs for every sample and analysed them with the ACMEtool2.0 software (Bennke et al., 2016).

Pigment analyses

Pigment composition in seawater samples was analysed using an improved high-performance liquid chromatography (HPLC) protocol derived from the method of Goericke

and Repeta (1993). Between 1 and 2 L of seawater were prefiltered through a 200 µm mesh, and then collected onto 25 mm polyethersulfone (PES) 0.45 µm pore size filter discs (Meck Millipore Ltd.) to assure that all AAP cells were captured (Lami et al., 2009). The use of 25 mm PES filters also reduced the volume of filtered seawater, reduced the amount of extraction solvents needed, and simplified the extraction protocol. Filters were conserved at –80°C until pigment extraction which was conducted in the laboratory. Filters were transferred to cryogenic vials with 1 mL of 7:2 acetone:methanol (v/v) mixture and ~30 mg of glass beads, and then disintegrated in a Mini-Beadbeater™ (Biospec Products, USA) during 2 minutes, to break the filters and extract the pigments. After a short centrifugation (3 min, 14.100 rcf), the clear extracts were transferred to HPLC vials for further processing. We used a SCL-40 HPLC (Nexera series, Shimadzu, Japan) equipped with SPD-M40 PDA detector and a UV-VIS detector with a deuterium (D2) lamp. For each sample, 50 µL were injected into the system. Pigments were separated using a heated (40°C) Kinetex® 2.6µm C8 100 Å column (Phenomenex Inc., CA, USA) with binary solvent system consisting of A: 75% methanol + 25% 28 mM ammonium acetate, and B: 100% methanol, with a flow rate of 1 mL·min–1. The gradient was the following: A/B 100/0 (0min), A/B 0/100 (21 min), 0/100 (23 min), 100/0 (24 min), and 100/0 (25 min). The peak for BChl *a* was registered at 770 nm. The pigment concentration in the original samples was calculated from the peak area. The use of Kinetex® column with core shell technology resulted in sharper peaks, which helped to accurately integrate small BChl *a* peaks. The peak of Chl *a* and divinyl-chlorophyll *a* were integrated together at 656 nm, so values of Chl *a* presented in the results include both Chl *a* and divinyl-chlorophyll *a*. The HPLC system was calibrated using 100% methanol extracts of *Synechocystis* sp. PCC6803 for Chl *a* and *Rhodobacter sphaeroides* for BChl *a*, where the concentrations of Chl *a* and Bchl *a* were determined spectroscopically using a Shimadzu UV2600 spectrometer. From the Bchl *a* concentration and AAP cell abundances, we calculated the Bchl *a* cell quota as the quantity of Bchl *a* per cell.

Satellite-derived variables and irradiance per depth calculations.

Given that no in situ PAR data was measured during the cruise, satellite-derived photosynthetic active radiation (PAR) and diffuse attenuation coefficients at 490 nm (Kd) were obtained from the Suomi-NPP/Visible Infrared Imaging Radiometer Suite instrument (SNPP-VIIRS). Eight-day composites at 9.28 km resolution (level 3-mapped data) were downloaded from the NASA Ocean Color Web (https://oceancolor.gsfc.nasa.gov/l3/). Both variables for each station were extracted according to the sampling date and the station coordinates. In case of data gaps due to clouds, the average values of a 10x10

pixel square centred on the station coordinates were calculated. In situ irradiance per depth ($PAR_{(Z)}$) was calculated as:

$$\mathrm{PAR_{(z)}} = \mathrm{PAR_{(0)}} e^{-(\mathrm{K_d})z} \quad \text{eq. 1}$$

where $PAR_{(0)}$ is the subsurface average daily downward irradiance for each station, assuming an average surface reflectance of 6% (Bailey et al., 2008), z is the depth in meters and K_d is the diffuse attenuation coefficient determined for each station. In equations 2 and 3 below, $PAR_{(Z)}$ is denoted as I (irradiance).

Bioenergetic calculations

The cellular daily energy captured per each cell Wwas estimated applying the theoretical calculations of Kirchman and Hanson (2013) for AAP bacteria, assuming a hyperbolic response to light, as it was also estimated in Gómez-Consarnau et al., (2019). First, we calculated the number of photosynthetic units (PSU) per cell, considering that AAP bacteria contain 34 molecules of Bchl *a* (Yurkov & Beatty 1998), and assuming an average number of 300 molecules of Chl *a* in the case of the autotrophic picoplankton as in Gómez-Consarnau et al., (2019). We used AAP abundances obtained through microscopic analyses, and the abundances of the autotrophic picoplankton estimated as the sum of the abundances of *Prochlorococcus*, *Synechococcus*, and picoeukaryotes from the flow cytometry analyses. These calculations were applied to stations 5 to 28, and we decided to remove stations 1, 2, 3 and 4 from the analysis, since we assumed that a large fraction of the phytoplankton in samples from those stations are diatoms and could alter our estimations. As suggested by Kirchman and Hanson (2013), we considered that electron fluxes vary with irradiance in a Michaelis-Menten type relationship as:

$$V = \frac{V_{max} \cdot I}{K_m + I} \quad \text{eq. 2}$$

where V_{max} is the maximum flux and K_m is the half saturation constant, that is, the value of irradiance (I) at which V is half of V_{max}. The values of V_{max} and K_m used in the equations, as well as the energy per photon for Bchl *a* and Chl *a* were extracted from Kirchman and Hanson (2013).

$$Cellular\ daily\ energy\ yield\ (J \cdot cell^{-1} day^{-1}) = \frac{PSU}{cell} \times V \times E_0 \times t \quad \text{eq. 3}$$

The cellular daily energy yield was estimated as in equation 3, where E_0 is the energy per photon and t is 43200 seconds, as we considered 12 h of light per day.

All calculations were performed in the R Core v4.2.0 (R Core Team 2022) environment, using *tidyverse* (Wickham et al., 2019) and *ggplot2* (Wickham 2016) packages for data processing and figure representations. Besides, Ocean Data View v5.6.5 was used to represent sectional distribution images.

4.3 Results

Bacterioplankton abundance variation across a vertical and horizontal gradient

Bacterioplankton (i.e. bacteria and archaea) abundances obtained from the flow cytometry analyses varied between $1.1 \cdot 10^5$ and $2.2 \cdot 10^6$ cells/mL (mean ± standard deviation, s.d., $6.1 \cdot 10^5$ cells/mL ± $4.1 \cdot 10^5$) and were mostly concentrated between 0 and 75 m depth, with a very similar distribution of high nucleic acid-containing (HNA) and low nucleic acid-containing (LNA) prokaryotes in all samples (Figure S1). A high proportion of the bacterioplankton cells were *Prochlorococcus* (~47%), which dominated in almost all stations and depths with values between 10^5 and 10^6 cells/mL. *Synechococcus* oscillated between 10^2 and 10^4 cells/mL and their abundance increased towards the boundary areas of the oligotrophic gyre (Figure S1B). Plastidic picoeukaryotes were more abundant at the DCM, in eutrophic areas, with densities up to 10^4 cells/mL (mean ± s.d., $1.6 \cdot 10^3$ cells/mL ± $2.7 \cdot 10^3$). Altogether, the autotrophic picoplankton —*Synechococcus*, *Prochlorococcus* and picoeukaryotes— ranged between $6 \cdot 10^4$ and $1.2 \cdot 10^6$ cells/mL (mean ± s.d., $3.05 \cdot 10^5$ cells/mL ± $2 \cdot 10^5$, Figure S1B). The AAP abundance ranged between undetectable values below the DCM (in 7% of samples) and $5.8 \cdot 10^4$ cells/mL (mean ± s.d., $1.1 \cdot 10^4$ ± $1 \cdot 10^4$ cells/mL), surpassing the abundances of *Synechococcus* and picoeukaryotes in most samples by an order of magnitude. In some stations within the oligotrophic gyre (stations 11 and 12), with deeper DCMs, we could detect them down to 280 m depth (our deepest sampling). AAP bacteria represented between 0.05% and 8% of total bacterioplankton (mean ± s.d., 1.5 ± 1.1 %). The highest values were observed in the eutrophic southernmost stations and in the equatorial upwelling, corresponding to samples with high concentrations of Chl *a* (Pearson correlation between untransformed AAP abundances and Chl *a*, n=172, R=0.62, $p<0.005$, and between %AAP, n=172, R=0.42, $p<0.005$, Figure 1B). This correlation was especially strong and significant in samples from the DCM and below the DCM (Pearson correlations with Chl *a*, n=27, R=0.69, $p<0.005$ and n=75, R=0.82, $p<0.005$ respectively). Interestingly, in station 10, which had a two peak DCM structure, the AAP abundance was maximal coinciding with both DCMs (Figure S2). All in all, the distribution of AAP bacteria along the water column followed the Chl *a* variation, with maximal abundances generally above the DCM and in

some cases, at the DCM (Figure 1B). In most stations, the maximum AAP abundances coincided with the maximum cell abundances of the autotrophic picoplankton.

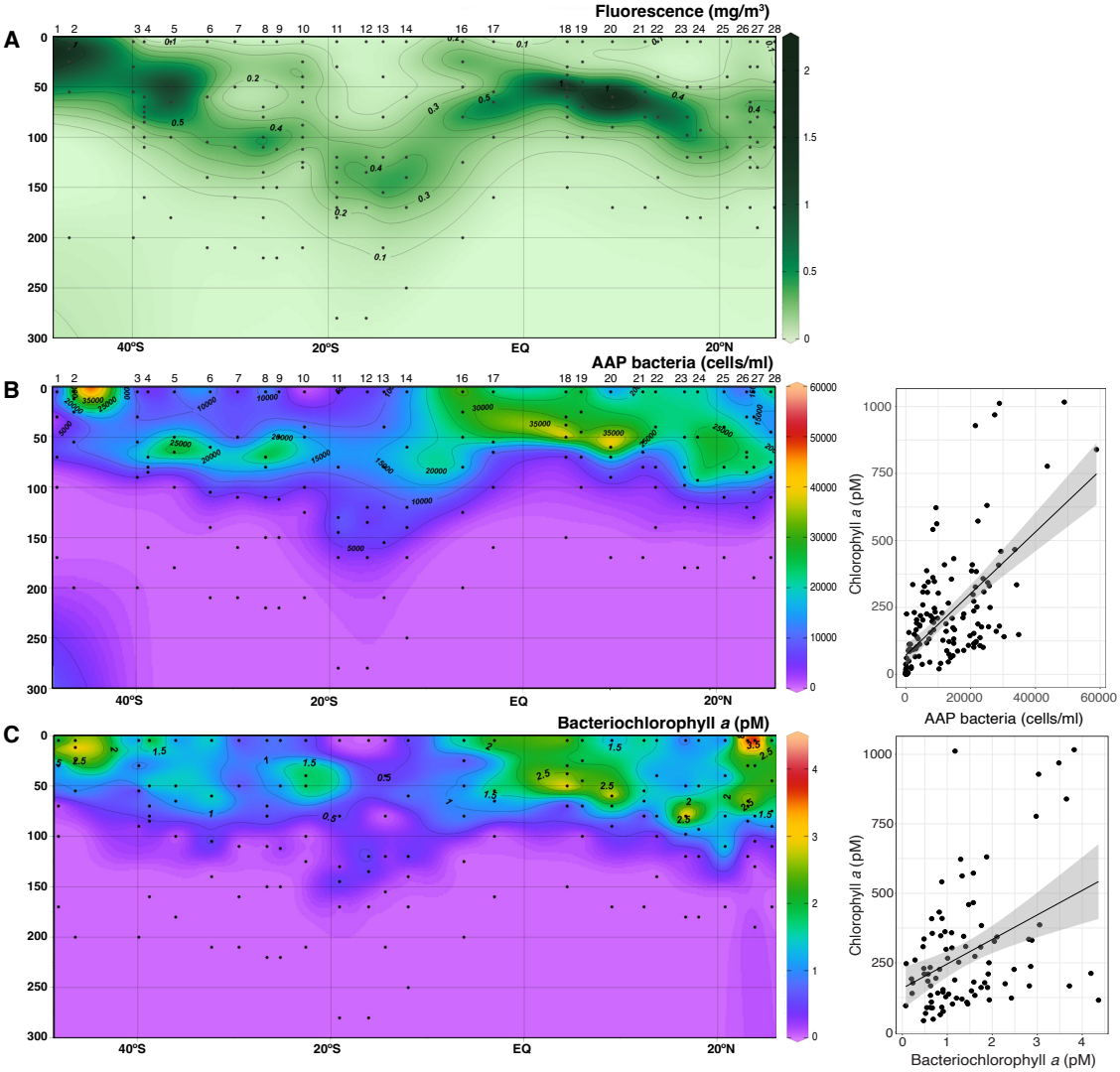

Figure 1. Pigment concentration and abundance distribution of aerobic anoxygenic phototrophic (AAP) bacteria along the South and Mid Atlantic Ocean. A) Sectional distribution of chlorophyll fluorescence (mg/m^3) as obtained from the CTD/Fluorometer data. B) Sectional distribution of AAP abundance (cells/mL) estimated by epifluorescence microscopy (left) and its correlation with the chlorophyll *a* concentration (right). C) Bacteriochlorophyll *a* concentration (pM) estimated with HPLC (left) and its correlation with the chlorophyll *a* concentration (right). Station numbers are indicated above panels A and B. Black dots in the left panels indicate the depths of sampling.

Bacteriochlorophyll *a* distribution

Bchl *a* concentration in the upper 150 m of the water column varied between 0 and 4.36 pM (mean ± s.d., 1.43 pM ± 0.98). The total concentration of Chl *a* was at least two orders of magnitude higher than that of Bchl *a* and varied between 0.89 and 1016 pM (mean ± s.d., 207.6 pM ± 205.3). As shown for AAP abundances, the distribution of Bchl *a* was highly correlated with that of Chl *a* (Pearson correlation between Bchl *a* and Chl *a*, n=172, R=0.56, $p<0.005$). When we considered only the surface samples, the correlation was weaker (n=27, R=0.36, $p=0.09$) due to the high Chl *a* concentration in samples from stations 1, 2, and 3, where we detected a diatom bloom (results not shown). When these stations were removed, Bchl *a* and Chl *a* were strongly and statistically correlated in the surface (n=24, R=0.62, $p<0.005$). In almost all samples, Bchl *a* concentration peaked above or at the DCM (Figure 3), except for a few sites (stations 1, 4, 21, and 27), where Bchl *a* was highest at the surface. The proportion of Bchl *a* in total photosynthetic pigments (Bchl *a*/(Bchl *a* + Chl *a*), ranged from 0% to 3.65%, being higher in the surface and decreasing in samples from deeper DCM layers (Figure S3). The Bchl *a* cellular quota ranged between 8.9 and 809 ag/cell (mean 49.1 ag/cell). The variation of the Bchl *a* quota correlated well with the abundance of the autotrophic picoplankton, specially the abundance of picoeukaryotes and with some inorganic nutrients such as nitrate (Figure S4).

Light energy captures by AAP and phytoplankton

The concentration of Chl *a*-containing photosynthetic units (PSUs) varied between $1.78 \cdot 10^{9}$ to $2.04 \cdot 10^{12}$ PSU/L, exceeding by around two orders of magnitude Bchl *a*-containing PSUs ($1.34 \cdot 10^{9}$ to $7.73 \cdot 10^{10}$ PSU/L). Chl *a*-containing PSUs significantly increased from surface towards deeper layers (Tukey test, $p<0.05$, Figure S4), while the concentration of Bchl *a* containing PSUs was very similar in the surface, above the DCM or in the DCM (Tukey test, $p=0.665$), and only significantly decreased below the DCM, where the pigment concentration was very low. The cellular light energy captured by AAP cells ranged between $3.16 \cdot 10^{-12}$ and $1.56 \cdot 10^{-9}$ J/cell·day (mean $8.01 \cdot 10^{-11}$ J/cell·day), about an order of magnitude lower than that captured by autotrophic picoplankton (mean $3.51 \cdot 10^{-10}$ J/cell·day) (Figure 2A). The cellular light energy captured by autotrophic picoplankton cells increased from the surface to the DCM, consistent with the pattern of increasing concentration of Chl *a*-containing PSUs. On the contrary, the energy captured by AAP cells significantly decreased with depth (Figure 2, Tukey test, $p<0.005$), since while light intensity declined, the concentration of Bchl *a*-containing PSU per cell did not vary across the different epipelagic layers, and was only significantly lower below

the DCM. The divergent pattern of energy captured by both pigments along depth is evident when we represent the ratio of cellular daily energy captured by Bchl *a* and Chl *a* in each layer of the epipelagic (Figure 3). AAP bacteria capture the maximum cellular light energy in the surface, and their contribution rapidly decreases towards deeper waters. At all depths, the values of light energy captured by AAP bacteria exceeds the costs of constructing the proteins and pigments structuring their PSUs (calculated by Kirchman and Hanson (2013) to be ca. $8.29 \cdot 10^{-15}$ J per PSU), with a mean net benefit of $1.47 \cdot 10^{-10}$ J/cell·day. Finally, we estimated the depth-integrated energy potentially captured by AAP bacteria and by autotrophic picoplankton in the water column (Figure 3B). For both pigments, the concentration of PSU per square meter was higher at more productive regions and lower in the oligotrophic gyres. In contrast to the differences between AAP bacteria and picophytoplankton along the vertical gradient, both groups displayed the same horizonal pattern of estimated energy captured, with lowest values in the South Atlantic oligotrophic gyre that increased towards more eutrophic sites (Figure 3B, Pearson correlation of depth-integrated energy captured, n=27, R=0.49, $p<0.05$).

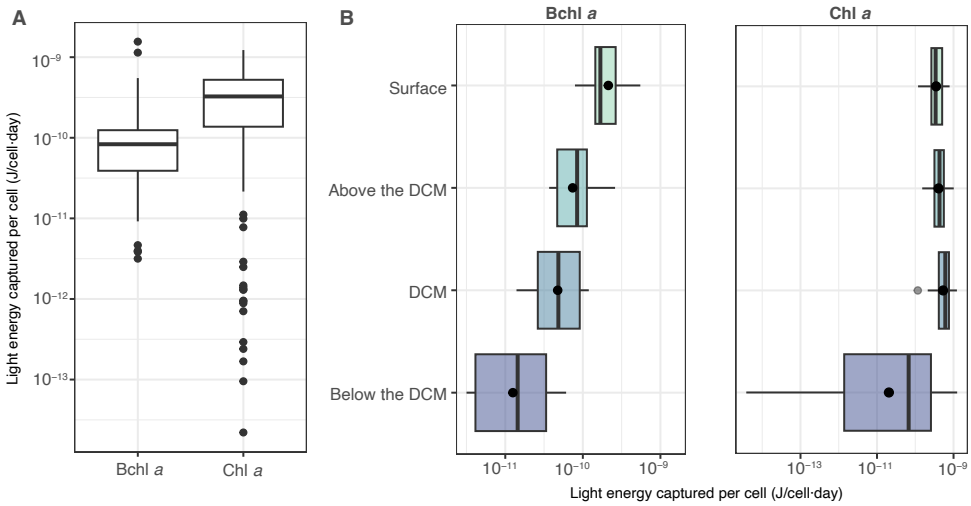

Figure 2. Cellular daily energy captured by bacteriochlorophyll *a* (Bchla *a*) and chlorophyll *a* (Chl *a*) along the South and Mid Atlantic Ocean. A) Light energy captured per cell and day by bacteriochlorophyll *a* and chlorophyll *a*. B) Pattern of light energy captured per cell and day along the epipelagic layers, for each pigment. The dot within the boxplots is the mean, while the thick line represents the median. Outliers are represented with dots outside the boxes.

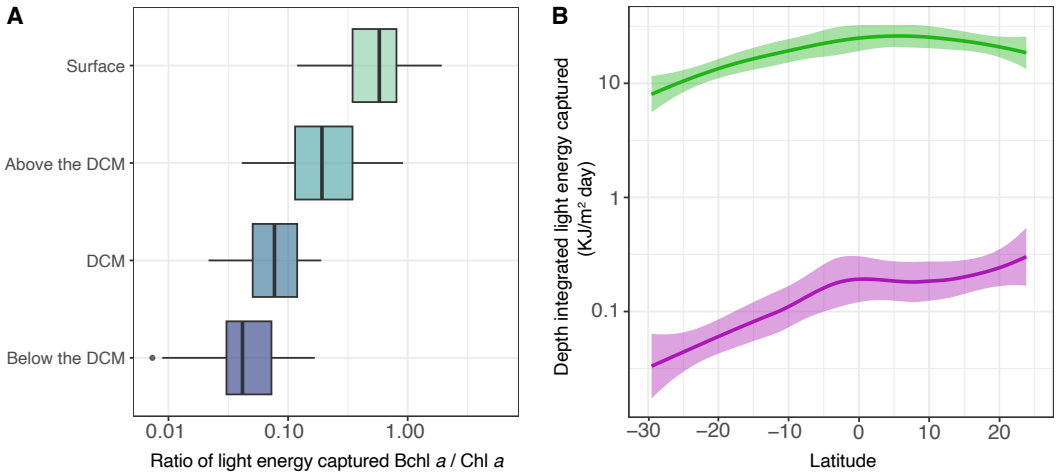

Figure 3. Vertical and horizonal cellular daily energy captured. A) Ratio of the cellular daily energy captured by Bchl *a* and by Chl *a* at each layer within the epipelagic and B) depth integrated light energy captured by phytoplankton (green line), and AAP bacteria (purple line) along the Atlantic Ocean transect.

4.4 Discussion

Aerobic anoxygenic phototrophic (AAP) bacteria are ubiquitous in the sunlit zone of the ocean, where they play an important ecological role in the carbon cycle despite their relatively low abundances (Kolber et al., 2000; Kolber et al., 2001; Koblížek 2015). Although their distribution in the surface waters of different oceanic regimes has been reviewed, little information is known about their vertical distribution patterns along the photic zone. Our sampling took place in the South and Mid Atlantic Ocean, in the South hemisphere fall and Northern hemisphere spring, following a transect of contrasting productivity and nutrient concentrations. The areas with higher productivity corresponded to either the southernmost stations, influenced by the proximity to the Southern Ocean and where frequent diatom blooms are reported (Malviya et al., 2016), and the stations along the equatorial upwelling. Samples from the oligotrophic gyre could be distinguished by their low surface Chl *a* values (Figure 1A), but *Synechococcus* abundances also served as a good biological indicator of the gyre boundaries (Hartmann et al., 2012), as their concentrations decreased around an order of magnitude in samples within the oligotrophic gyre. *Prochlorococcus*, the most abundant cyanobacteria in the South Atlantic Ocean (Zubkov et al., 2000; Carvalho et al., 2020), numerically dominated all stations and samples. These cyanobacteria were the main components of the picophytoplankton, and together with picoeukaryotes, their abundances were in line with previous data

estimated in this region (Visintini et al., 2021). Picoeukaryotes contribute significantly to the total C biomass in the Atlantic Ocean, as seen before by the Atlantic Meridional Transects (Tarran et al., 2006). In our samples, these three groups —*Synechococcus*, *Prochlorococcus* and picoeukaryotes—, were the main components of the phytoplankton in most of the transect, with the exception of stations 1, 2, 3 and 4, where diatoms were also an important fraction of the total phytoplankton. All in all, the sampled transect showed a high variability in terms of trophic conditions which was reflected in the DCM structures along the water column, with some DCM being more shallow and intense in terms of Chl *a* concentrations while others were located below the 100 m depth.

Along these variable vertical profiles, the abundance of AAP bacteria followed the Chl *a* distribution, peaking above or at the DCM level, and in most samples coinciding with the autotrophic picoplankton maximum abundance, confirming the outlined hypothesis that AAP bacteria would be connected to the DCM structure. Across the Atlantic Ocean, we observed the highest abundances in the eutrophic southernmost stations and in the equatorial upwelling, where the input of nutrients from deeper layers supports the productivity of phytoplankton communities. Likewise, the distribution of both Bchl *a* and Chl *a* was very similar and these pigments occupied similar areas along the water column. The concentration of Bchl *a* in our sampling was comparable to that in other marine regions like the Pacific Ocean (Kolber et al., 2001; Goericke 2002; Lami et al., 2007; Jiao et al., 2010), the Atlantic Ocean (Cottrell et al., 2006; Sieracki et al., 2006), the Baltic Sea (Koblížek et al., 2005) and the Mediterranean Sea (Lami et al., 2009; Hojerová et al., 2011; Ferrera et al., 2014; Gómez-Consarnau et al., 2019). It has been extensively documented that AAP bacteria are more abundant in more productive areas (reviewed in Koblížek, 2015), and the overlap between AAP cells and Chl *a* has been reported before in areas from the Atlantic and Pacific Oceans (Kolber et al., 2001; Schwalbach & Fuhrman 2005; Cottrell et al., 2006; Mašín et al., 2006, Jiao et al., 2007, Jiao et al., 2010), or the Mediterranean Sea (Hojerová et al., 2011), yet information along vertical profiles is scarce and mainly focused on the Northern hemisphere (for example, Sieracki et al., 2006, Mašín et al., 2006, Lami et al., 2009). The fact that Bchl *a* concentration mirrors the distribution of Chl *a* throughout depth and along the latitudinal axis, could indicate that the photoheterotrophic activity of AAP bacteria might be favoured or controlled by the same mechanisms that control phytoplankton, but it could also indicate the existence of strong biotic interactions between these two microbial guilds. In fact, co-occurrence networks have shown interactions of some AAP bacteria with phototrophic eukaryotes (Auladell et al., 2019), and various AAP species have been isolated from

phytoplankton, particularly from dinoflagellates (Biebl et al., 2005; Yang et al., 2018). AAP cells with higher Bchl *a* content were associated to the most productive areas, with higher abundances of picophytoplankton or nitrate, while the light intensity across sites did not have any effect on the Bchl *a* quota. Similarly, the number of Bchl *a*-containing PSUs per cell did not vary along the surface, above the DCM and at the DCM, suggesting that no photoacclimatation processes were operating, or at least we could not detect them. In this sense, although a similar regulatory circuit based on the *ppsR* (a transcriptional regulator of photosystems) has been described in several subclasses of Alpha- and Gammaproteobacteria (Liotenberg et al., 2008; Tomasch et al., 2011; Zheng et al., 2011), different AAP strains show variable adaptations to changes in light intensity. For example, *Roseobacter litoralis* decreased three times the number of Bchl *a* molecules in the PSU following increasing irradiance (Selyanin et al., 2016), whereas *Dinoroseobacter shibae* adjusted its electron transport rate (Piwosz et al., 2018). These studies are based on cultivated AAP species and little information regarding their phototrophic activity that can be applied to natural communities exists. Thus, following the approaches by other authors (Kirchman & Hanson 2013; Gómez-Consarnau et al., 2019), we assumed a constant number of molecules per PSU and a constant electron transfer rate in our calculations.

Previous approaches estimating the light energy captured by AAP bacteria in natural environments were theoretical (Kirchman & Hanson 2013), or based solely on pigment data and hypothesized AAP abundances (Gómez-Consarnau et al., 2019). While these studies have constituted a solid base for this study, we estimated the light energy captured by AAP cells and the autotrophic picoplankton, based on in situ cell abundances and pigment concentration data, and provide, for the first time, an accurate picture of the potential phototrophy of AAP bacteria in the open ocean. Our data shows that photoheterotrophy is especially strong at the surface, where the light energy captured by AAP cells is highest (Figure 3). However, this additional source of energy gained from light cannot cover the maintenance energy costs of a large bacterium, estimated as $4 \cdot 10^{-10}$ J (Kirchman & Hanson 2013), but it probably replaces part of the oxidative phosphorylation, as seen in previous experiments (Koblížek et al., 2010; Hauruseu & Koblížek 2012), providing a fraction of the cell metabolic needs and saving substrates that can be further used for biosynthesis (Hauruseu & Koblížek 2012). All in all, the light energy captured by Bchl *a* was about an order of magnitude lower than that captured by Chl *a*, which makes sense since AAP bacteria harvest light to supplement their heterotrophic metabolism and do not rely on phototrophy for growth as phytoplankton do. It has been estimated that about 20% of their energy requirement could be satisfied by photoheterotrophy (Kolber et al., 2001),

and that this reduction in the organic carbon consumption would be equivalent to 5.4% of the primary production in oceanic waters. Taken together, these results suggest that AAP activity could reduce substantially respiration in surface waters, and, in areas where they are abundant, this could have an impact on the global carbon cycle.

In summary, we show that AAP bacteria are widespread throughout the open ocean in areas of contrasting productivity, from surface down to 250 m deep. They are strongly associated with the autotrophic picoplankton, as seen by their abundance distribution and Bchl *a* quota variation. The preference of AAP bacteria for chlorophyll-rich waters could be due to common variables affecting the distribution of both functional groups but could also indicate their dependence on the phytoplankton. AAP cells are larger than average bacteria (Sieracki et al., 2006) and display higher growth rates under different scenarios (Koblížek et al., 2007; Ferrera et al., 2011, 2017). To sustain such a dynamic metabolism, they must rely on large quantities of phytoplankton-derived dissolved organic matter, and therefore they would preferably develop in eutrophic areas. The light harvesting ability provides AAP bacteria with an additional source of energy that can be used to complement their energetic needs, especially in the surface ocean.

Overall, this is the first study that uses both, in situ cell abundances and pigment data to estimate the potential solar energy captured by AAP bacteria in such an extensive area of the open ocean. The extent of our sampling area allowed the detection of large-scale patterns, making predictions on the potential phototrophic activity of AAP bacteria along the epipelagic, contributes to a better understanding of the energy fluxes in the ocean and, ultimately, in the global carbon cycle. Future studies analysing the interaction between AAP bacteria communities and the organic matter derived from different phytoplanktonic groups will shed light on the relationship between these functional groups, providing a better understanding of the ecologic role of AAP bacteria in the marine environment.

4.5 Acknowledgments

We thank the people of the Laboratory of Anoxygenic Phototrophs in Třeboň, where the epifluorescence microscopy and HPLC analyses were performed. We are particularly grateful to Lucija Kanjer MSc. for her help during optimization of our HPLC protocol, Dr. Martí Gali for his help in obtaining the satellite data and Dr. Laura Gómez-Consarnau for her guidance in the bioenergetic calculations. We would also like to thank all the crew members of the R/V *Sarmiento de Gamboa*, as well as the scientific team, particularly Alba

González-Vega, Eugenio Fraile-Nuez, and chief scientist Dr. Jesús M. Arrieta. This work was supported by grants ECLIPSE (PID2019-110128RB-I00/AEI/10.13039/501100011033) to IF and MICOLOR to JMG (PID2021-125469NB-C31) and OS (PID2021-125469NB-C32), funded by the Agencia Estatal de Investigación from the Spanish Ministry of Science and Innovation. Authors affiliated to the Institut de Ciències del Mar received the institutional support of the 'Severo Ochoa Centre of Excellence' accreditation (CEX2019-000928-S). CRG was supported by a PIF fellowship from the Universitat Autònoma de Barcelona.

4.6 References

Azam, F., Fenchel, T., Field, J., Gray, J., Meyer-Reil, L., & Thingstad, F. (1983) The Ecological Role of the Water-Column Microbes in the Sea. *Mar Ecol Prog Ser* 10: 257–263. doi:10.1021/acs.joc.6b00938

Auladell, A., Sánchez, P., Sánchez, O. et al., (2019) Long-term seasonal and interannual variability of marine aerobic anoxygenic photoheterotrophic bacteria. *ISME J* 13, 1975–1987. https://doi.org/10.1038/s41396-019-0401-4

Bennke, C.M., Reintjes, G., Schattenhofer, M., Ellrott, A., Wulf, J., Zeder, M., & Fuchs, B.M. (2016) Modification of a high-throughput automatic microbial cell enumeration system for shipboard analyses. *Appl Environ Microbiol* 82: 3289–3296. doi:10.1128/AEM.03931-15

Biebl, H., Allgaier, M., Tindall, B.J., Koblížek, M., Lünsdorf, H., Pukall, R., et al., (2005) *Dinoroseobacter shibae* gen. nov., sp. nov., a new aerobic phototrophic bacterium isolated from dinoflagellates. *Int J Syst Evol Microbiol*. 55:1089–96.

Carvalho, A.C.O., Kerr, R., Mendes, C.R.B., Azevedo, J.L.L., & Tavano., V.M. (2020) Progress in Oceanography Phytoplankton strengthen CO_2 uptake in the South Atlantic Ocean. *Prog Oceanogr* 102476. doi:10.1016/j.pocean.2020.102476

Cottrell, M. T., A. Mannino, and D. L. Kirchman. 2006. Aerobic anoxygenic phototrophic bacteria in the mid-atlantic bight and the north pacific gyre. Appl Environ Microbiol 72: 557–564. doi:10.1128/AEM.72.1.557-564.2006

Cottrell, M. T., J. Ras, and D. L. Kirchman. 2010. Bacteriochlorophyll and community structure of aerobic anoxygenic phototrophic bacteria in a particle-rich estuary. ISME Journal 4: 945–954. doi:10.1038/ismej.2010.13

Fecskeová LK, Piwosz K, Šantic D, Šestanovic S, Tomaš AV, Hanusová M, Šolic M, Koblížek M. (2021) Lineage-specific growth curves document large differences in response of individual groups of marine bacteria to the top-down and bottom-up controls. *mSystems* 6:e00934-21. https://doi.org/10.1128/mSystems.00934-21.

Ferrera, I., Borrego, C.M., Salazar, G. &Gasol., J.M. (2014) Marked seasonality of aerobic anoxygenic phototrophic bacteria in the coastal NW Mediterranean Sea as revealed by cell

abundance, pigment concentration and pyrosequencing of *pufM* gene. *Environ Microbiol* 16: 2953–2965. doi:10.1111/1462-2920.12278

Ferrera, I., Gasol, J.M., Sebastián, M., Hojerová, E., & Kobížek, M. (2011) Comparison of growth rates of aerobic anoxygenic phototrophic bacteria and other bacterioplankton groups in coastal mediterranean waters. *Appl Environ Microbiol* 77: 7451–7458. doi:10.1128/AEM.00208-11

Ferrera, I., Sánchez, O., Kolářová, E., Koblížek, M., & Gasol, J.M. (2017) Light enhances the growth rates of natural populations of aerobic anoxygenic phototrophic bacteria. *ISME J* 11: 2391–2393. doi:10.1038/ismej.2017.79

Gasol, J.M., & Morán, X.A.G. (2015). Flow Cytometric Determination of Microbial Abundances and Its Use to Obtain Indices of Community Structure and Relative Activity. In: McGenity, T.J., Timmis, K.N., Nogales, B. (eds) *Hydrocarbon and Lipid Microbiology Protocols*. Springer Protocols Handbooks. Springer, Berlin, Heidelberg. https://doi.org/10.1007/8623_2015_139

Gazulla, C.R., Auladell, A., Ruiz-González, C. Junger, P.C., Duarte., C.M., Gasol, J.M., Sánchez, O., & Ferrera, I. (2022) Global diversity and distribution of aerobic anoxygenic phototrophs in the tropical and subtropical oceans. *Environ Microbiol* 24: 2222–2238. doi:10.1111/1462-2920.15835

Goericke, R. & Repeta, D. J. (1993) Chlorophylls a and b and divinyl chlorophylls a and b in the open subtropical North Atlantic Ocean *MEPS* 101:307-313

Goericke, R. (2002) Bacteriochlorophyll *a* in the ocean: Is anoxygenic bacterial photosynthesis important? *Limnol Oceanogr* 47: 290–295. doi:10.4319/lo.2002.47.1.0290

Gómez-Consarnau, L., Raven, J.A., Levine, N.M., et al. (2019) Microbial rhodopsins are major contributors to the solar energy captured in the sea. *Sci Adv* 5: 1–8. doi:10.1126/sciadv.aaw8855

Hartmann, M., Grob, C., Tarran, G.A., Martin, A., Burkill, P., Scanlan, D., & Zubkov. M.V. (2012) Mixotrophic basis of Atlantic oligotrophic ecosystems. *Proc Natl Acad Sci U S A* 109: S756–S760.

Hauruseu, D. & Koblížek, M. (2012) Influence of light on carbon utilization in aerobic anoxygenic phototrophs. *Appl Environ Microbiol* 78: 7414–7419. doi:10.1128/AEM.01747-12

Hojerová, E., Mašín, M., Brunet, C., Ferrera, I., Gasol, J.M., & Koblížek. M. (2011) Distribution and growth of aerobic anoxygenic phototrophs in the Mediterranean Sea. *Environ Microbiol* 13: 2717–2725. doi:10.1111/j.1462-2920.2011.02540.x

Jiao, N., Zhang, Y., Zeng, Y., Hong, N., Liu, R., Chen, F., & Wang. P. (2007) Distinct distribution pattern of abundance and diversity of aerobic anoxygenic phototrophic bacteria in the global ocean. *Environ Microbiol* 9: 3091–3099. doi:10.1111/j.1462-2920.2007.01419.x

Jiao, N., Zhang, F. & Hong, N. (2010) Significant roles of bacteriochlorophyll *a* supplemental to chlorophyll *a* in the ocean. *ISME J* 4, 595–597 (2010). https://doi.org/10.1038/ismej.2009.135

Kirchman, D.L., & Hanson, T.E. (2013) Bioenergetics of photoheterotrophic bacteria in the oceans. *Environ Microbiol Rep* 5: 188–199. doi:10.1111/j.1758-2229.2012.00367.x

Koblížek, M. (2015) Ecology of aerobic anoxygenic phototrophs in aquatic environments. *FEMS Microbiol Rev* 39: 854–870. doi:10.1093/femsre/fuv032

Koblížek, M., Mašín, M., Ras, J., & Poulton, A.J. (2007) Rapid growth rates of aerobic anoxygenic phototrophs in the ocean. *Environ Microbiol* 9: 2401–2406. doi:10.1111/j.1462-2920.2007.01354.x

Koblížek, M., Mlčoušková, J., Kolber, Z., & Kopecký, J. (2010) On the photosynthetic properties of marine bacterium COL2P belonging to *Roseobacter* clade. *Arch Microbiol* 192: 41–49. doi:10.1007/s00203-009-0529-0

Koblížek, M., Stoń-Egiert, J., Sagan, S., & Kolber, Z.S. (2005) Diel changes in bacteriochlorophyll *a* concentration suggest rapid bacterioplankton cycling in the Baltic Sea. *FEMS Microbiol Ecol* 51: 353–361. doi:10.1016/j.femsec.2004.09.016

Kolber, Z.S., Van Dover, C.L., Niederman, R.A., & Falkowski, P.G. (2000) Bacterial photosynthesis in surface waters of the open ocean. *Nature* 407: 177–179. doi:10.1038/35025044

Kolber, Z.S., Plumley, F.G., Lang, A. S. et al., (2001) Contribution of aerobic photoheterotrophic bacteria to the carbon cycle in the ocean. *Science* 292: 2492–2495. doi:10.1126/science.1059707

Lami, R., Cottrell, M. T., Ras, J., Ulloa, O., Obernosterer, I., Claustre, H., Kirchman, D.L., & Lebaron, P. (2007) High abundances of aerobic anoxygenic photosynthetic bacteria in the South Pacific Ocean. *Appl Environ Microbiol* 73: 4198–4205. doi:10.1128/AEM.02652-06

Lami, R., Čuperová, Z., Ras, J., Lebaron, P., & Koblížek, M. (2009) Distribution of free-living and particle-attached aerobic anoxygenic phototrophic bacteria in marine environments. *Aquatic Microb Ecol* 55: 31–38. doi:10.3354/ame01282

Lamy, D., De Carvalho-Maalouf, P., Cottrell, M.T., et al. (2011) Seasonal dynamics of aerobic anoxygenic phototrophs in a Mediterranean coastal lagoon. *Aquat Microb Ecol* 62: 153–163. doi:10.3354/ame01467

Liotenberg, S., Steunou, A., Picaud, M., Reiss-husson, F., Astier, C., Ouchane, S., Paris-sud, F.-U. & Pierre, F.-U. (2008) Organization and expression of photosynthesis genes and operons in anoxygenic photosynthetic proteobacteria. *Environ Microbiol* 10: 2267–2276. doi:10.1111/j.1462-2920.2008.01649.x

Malviya, S., Scalco, E., Audic, S., et al. (2016) Insights into global diatom distribution and diversity in the world's ocean. *Proc Natl Acad Sci U S A* E1516–E1525.

Mašín, M., A. Zdun, J. Stoń-Egiert, M. Nausch, M. Labrenz, V. Moulisová, & M. Koblížek. (2006)

Seasonal changes and diversity of aerobic anoxygenic phototrophs in the Baltic Sea. Aquatic Microbial Ecology 45: 247–254. doi:10.3354/ame045247

Piwosz, K., D. Kaftan, J. Dean, J. Šetlík, & M. Koblížek. (2018) Nonlinear effect of irradiance on photoheterotrophic activity and growth of the aerobic anoxygenic phototrophic bacterium *Dinoroseobacter shibae*. *Environ Microbiol* 20: 724–733. doi:10.1111/1462-2920.14003

Piwosz, K., Villena-Alemany, C. & Mujakić, I. (2022) Photoheterotrophy by aerobic anoxygenic bacteria modulates carbon fluxes in a freshwater lake. *ISME J* 16, 1046–1054 https://doi.org/10.1038/s41396-021-01142-2

Ritchie, A. E., & Z. I. Johnson. (2012) Abundance and genetic diversity of aerobic anoxygenic phototrophic bacteria of coastal regions of the pacific ocean. *Appl Environ Microbiol* 78: 2858–2866. doi:10.1128/AEM.06268-11

Sánchez, O., I. Ferrera, I. Mabrito, et al., (2020) Seasonal impact of grazing, viral mortality, resource availability and light on the group-specific growth rates of coastal Mediterranean bacterioplankton. *Sci Rep* 10. doi:10.1038/s41598-020-76590-5

Sánchez, O., M. Koblížek, J. M. Gasol, & I. Ferrera. (2017) Effects of grazing, phosphorus and light on the growth rates of major bacterioplankton taxa in the coastal NW Mediterranean. *Environ Microbiol Rep* 9: 300–309. doi:10.1111/1758-2229.12535

Schwalbach, M.S. & Fuhrman, J.A. (2005) Wide-ranging abundances of aerobic anoxygenic phototrophic bacteria in the world ocean revealed by epifluorescence microscopy and quantitative PCR. *Limnol Oceanogr* 50: 620–628. doi:10.4319/lo.2005.50.2.0620

Selyanin, V., Hauruseu, D., & Koblížek. M. (2016) The variability of light-harvesting complexes in aerobic anoxygenic phototrophs. *Photosynth Res* 128: 35–43. doi:10.1007/s11120-015-0197-7

Sieracki, M.E., Gilg, I.C., Thier, E.C., Poulton, N. J., & Goericke, R. (2006) Distribution of planktonic aerobic anoxygenic photoheterotrophic bacteria in the northwest Atlantic. *Limnol Oceanogr* 51: 38–46. doi:10.4319/lo.2006.51.1.0038

Tarran, GA., Heywood, JL., & Zubkov, M.V. (2006) Latitudinal changes in the standing stocks of nano- and picoeukaryotic phytoplankton in the Atlantic Ocean. *Deep Sea Res. Part II Top. Stud. Oceanogr.* 53:1516-1529, https://doi.org/10.1016/j.dsr2.2006.05.004.

Tomasch, J., Gohl, R., Bunk, B., Diez, M.S. & Wagner-Döbler, I. (2011) Transcriptional response of the photoheterotrophic marine bacterium *Dinoroseobacter shibae* to changing light regimes. I*SME J* 5: 1957–1968. doi:10.1038/ismej.2011.68

Visintini, N., Martiny, A.C., & Flombaum, P. (2021) *Prochlorococcus* , *Synechococcus* , and picoeukaryotic phytoplankton abundances in the global ocean. *Limnol Oceanogr Let* 207–215. doi:10.1002/lol2.10188

Vrdoljak Tomaš, A. (2019) Dynamics of Aerobic Anoxygenic Phototrophs along the trophic gradi-

ent in the central Adriatic Sea. *Deep Sea Res. Pt II* 164, 112–121. https://doi.org/10.1016/j.dsr2.2019.06.001

Wickham, H. (2016) Elegant Graphics for Data Analysis: ggplot2,.

Wickham, H., M. Averick, J. Bryan, et al. (2019) Welcome to the Tidyverse. *J Open Source Softw* 4: 1686. doi:10.21105/joss.01686

Yang, Q., Jiang, Z.-W., Huang, C.-H., Zhang, R.-N., Li, L.-Z., Yang, G., et al. (2018) *Hoeflea prorocentri* sp. nov., isolated from a culture of the marine dinoflagellate *Prorocentrum mexicanum* PM01. *Antonie Van Leeuwenhoek*. 111:1845–53.

Yurkov, V. V, & J. T. Beatty. (1998) Aerobic anoxygenic phototrophic bacteria. *Microb. Mol. Biol Rev.* 62: 695–724.

Zheng, Q., Zhang, R., Koblížek, M., Boldareva, E.N., Yurkov, V., Yan, S., & Jiao, N. (2011) Diverse Arrangement of Photosynthetic Gene Clusters in Aerobic Anoxygenic Phototrophic Bacteria. *PLoS One* 6: 1–7. doi:10.1371/journal.pone.0025050

Zubkov, M.V., Sleigh, M.A., Burkill, P.H., & Leakey, R.J.G. (2000) Picoplankton community structure on the Atlantic Meridional Transect: a comparison between seasons. *Prog Oceanogr* 45: 369–386.

4.7 Supplementary figures

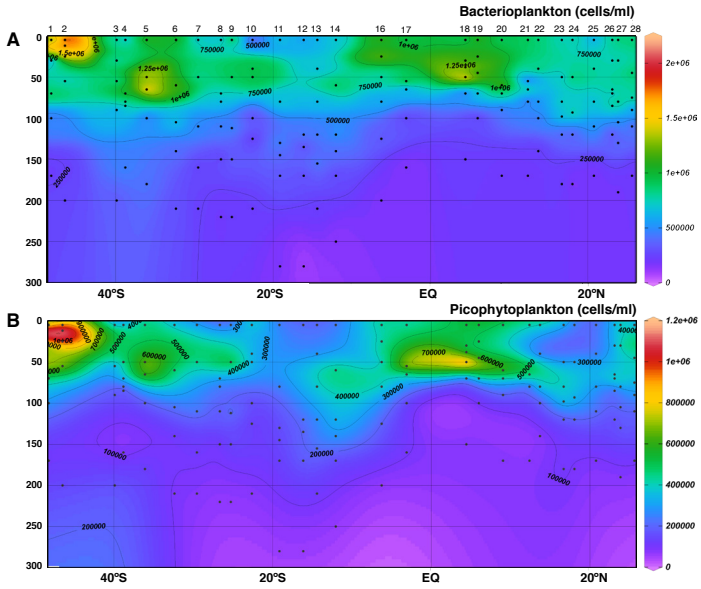

Figure S1. Distribution of bacterioplankton (A) and picophytoplankton (B) along the water column in the south and mid-Atlantic Ocean. Abundances of total prokaryotes are from epifluorescence microscopic analyses, after staining cells with DAPI. Picophytoplankton data is the sum of *Synechococcus*, *Prochlorococcus* and picoeukaryotes determined by flow cytometry.

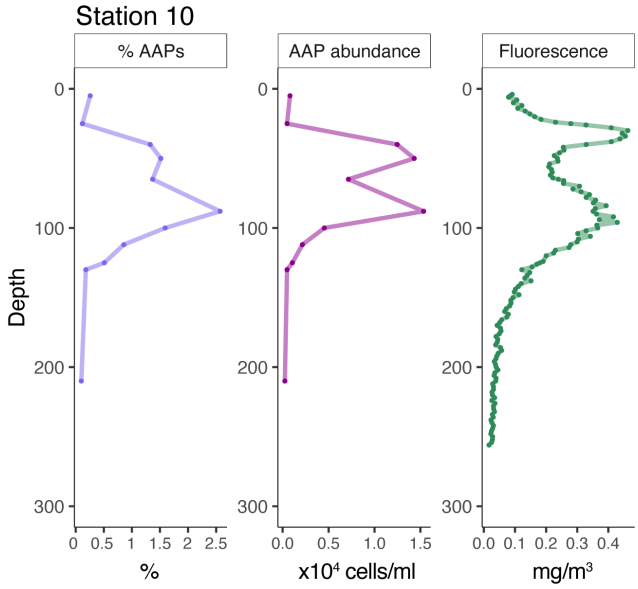

Figure S2. Depth profile of station 10 showing (from left to right) the percentage of AAP bacterial abundances in the bacterioplankton, the concentration of cell abundances of AAPs, and the fluorescence depicted by the CTD profiler.

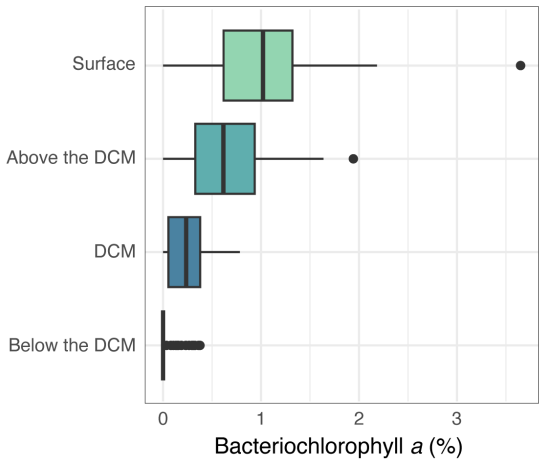

Figure S3. Bacteriochlorophyll *a* proportion at each layer of the epipelagic structure.

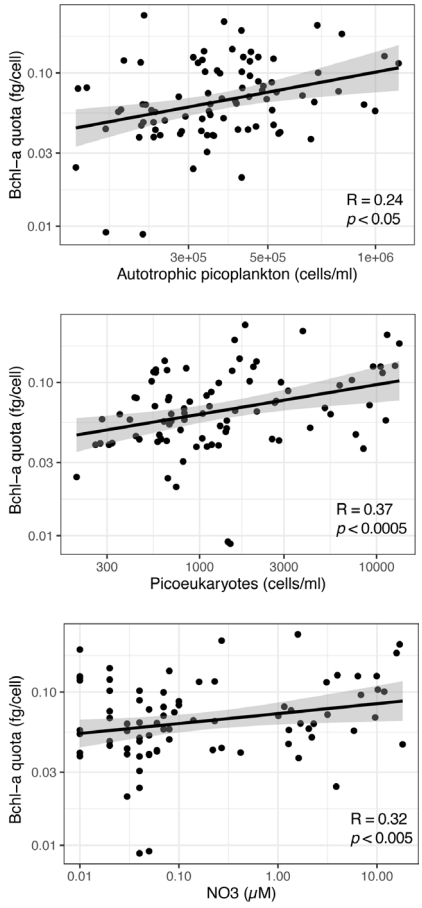

Figure S4. Pearson correlation between the Bchl *a* quota and (from left to right), the picophytoplankton, picoeukaryote abundance, and nitrate (NO_3).

GENERAL DISCUSSION

GENERAL DISCUSSION

This thesis offers comprehensive insights into the global ecology of aerobic anoxygenic phototrophic (AAP) bacteria in the marine environment. Here, I provide the first global description of the diversity of AAP bacteria in the surface ocean, describing their biogeographic patterns and investigating the ecological processes structuring AAP communities across the global ocean (**Chapter I**). I have also tested and evaluated the molecular tools available to study the diversity of this functional group in order to first improve (**Chapter II**) and later successfully apply them (**Chapter III**) to study the horizontal and vertical distributions of AAP bacteria across a vast area characterized by contrasting productivity levels. Finally, while these chapters significantly contribute to the current knowledge on the diversity of AAP bacteria, **Chapter IV** goes beyond diversity by analyzing their distribution in the water column and by estimating their bioenergetics in order to make predictions on their phototrophic activity and, ultimately, on their ecological role in the marine environment. In this general discussion, I integrate the results obtained in the different chapters, focusing on three main points: 1) the molecular tools used to study AAP bacteria and how the technical approach influence our results. To illustrate it, I provide additional data analyses not included in the previous chapters. Then I will discuss 2) the uncultured diversity of AAP bacteria and 3) the ecological role of this group of marine microorganisms, outlining the current knowledge on these topics and setting out future perspectives to increase our understanding on their ecology.

1. The technical approaches to study AAP bacteria

Advances in our understanding of the diversity of AAP bacteria have been consequence of the development of molecular tools and high-throughput sequencing (HTS) techniques that have taken place in the last two decades. The parallel evolution of metagenomics and PCR-based techniques have allowed to perform multiple and complementary approaches to studying the marine microbiota, significantly expanding our understanding on its ecology. This evolution is somehow represented in this thesis, albeit at a smaller scale, focused on a singular bacterial functional group. The objective of **Chapter I** was to delineate the diversity and biogeographic patterns of the surface AAP communities at a global scale. To perform such an extensive sampling, we based our study on a PCR-based amplicon approach, since it allowed to retrieve enough copies of the *pufM* gene to perform a high-resolution ecological study without an excessive cost. We selected the most widely employed primers from previous studies of AAP bacteria in the marine

environment (e.g., Lehours et al., 2010; Jeanthon et al., 2011; Boeuf et al., 2013; Ferrera et al., 2014; Lehours and Jeanthon 2015; Bibiloni-Isaksson et al., 2016; Lehours et al., 2018; Auladell et al., 2019). These primers had proven useful in determining the ecological habitats of some clades of marine AAP bacteria (Lehours et al., 2018) and in assessing the seasonality of AAP assemblages (Ferrera et al., 2014; Bibiloni-Isaksson et al., 2016; Auladell et al., 2019). However, and already discussed in **Chapter I**, concern about potential biases towards some specific groups of AAP bacteria existed (e.g. discussed in Lehours et al., 2010 and Auladell et al., 2019). For this reason, **Chapter II** directly addressed this issue by evaluating the existing primers for the *pufM* gene and designing alternative ones. Unfortunately, our attempts to design primers that would produce longer amplicons were unsuccessful. However, we were able to refine the existing primers while maintaining the original amplicon size. An essential part of this analysis was the use of samples for which both amplicon and metagenomic data were available, which allowed a thorough comparison. **Chapter II** revealed that the use of different primers sets leads to variations in the taxonomic composition of AAP communities. This explains the discrepancies observed in AAP community composition between **Chapter I** and **Chapter III**. In **Chapter I**, Gammaproteobacteria sequences dominate almost all oceanic regions with a mean relative abundance of 58.4% of the AAP community, while in **Chapter III** this group is the fourth most abundant (mean relative abundance 10.1%). The novelty in **Chapters II** and **III** derived from the use of different primers, and is the recovery of phylogroups A, B, C, and D that were missing in **Chapter I**, as well as in previous amplicon-based studies. As seen in **Chapter II** and **Chapter III**, these phylogroups not only seem to prevail in AAP assemblages of the surface ocean, but also throughout the whole water column. Remarkably, despite these compositional differences, the identified ecological trends remain consistent regardless of the used primers.

Another difference between chapters that is worth commenting is the use of a different taxonomic assignation approach. In **Chapter I**, we wanted to provide an overview of the main patterns governing AAP bacteria in the global ocean, focusing on the rare and ubiquitous taxa and the ecological processes controlling their presence, so the Malaspina *pufM* gene dataset was categorized into 6 large groups representing the main order or families observed within the Alpha- and Gammaproteobacteria. In **Chapter II**, we intended to give a detailed evaluation of the phylogenetic coverage of different primers, with a particular focus on the phylogroups defined by Yutin et al. (2007) that were previously absent from amplicon approaches. Therefore, we change the taxonomic assignation between chapters, and broke down the wide groups from **Chapter I** into

smaller clades and phylogroups, keeping this classification in **Chapter III**.

Efforts to improve the performance of primers for AAP bacteria have also been undertaken in freshwater ecosystems (Villena-Alemany et al., 2023). These authors did successfully develop new *pufM* primers that generate longer amplicons thereby improving the accuracy of taxonomic assignment. In their study, that covers the seasonal patterns of AAP bacteria in a freshwater lake, they show a predominance of Gammaproteobacteria. It's worth noting that the composition of AAP communities in marine and freshwater ecosystems differs significantly, as freshwater AAP assemblages commonly consist of different orders of the former Betaproteobacteria, now classified as Gammaproteobacteria (Salka et al., 2011; Caliz and Casamayor 2014; Piwosz et al., 2021; Villena-Alemany et al., 2022). A future goal would be to search for primers suitable for both freshwater and marine ecosystems, producing longer amplicons. Such an approach could be beneficial for studies focusing on interfaces between freshwater and marine environments, such as estuaries or coastal groundwaters.

The Atlantic Ocean: a study case

In **Chapter II** we show that the taxonomic composition of AAP assemblages vary depending on the primer set used, leading to significant discrepancies in the taxonomic composition of AAP assemblages. These differences were further observed when comparing the taxonomic composition of samples from **Chapter I** and **III**, both including samples from the open ocean. In light of these contrasting results, we decided to re-analyze a subset of samples from the Malaspina Expedition (part of Chapter I). We re-sequenced the DNA samples from the Malaspina Expedition with the primer set UniF/UniR –used in **Chapter III**– and compared the results obtained with those in **Chapter I**, in which we employed the primer pair pufMF/pufM_WAW (see Methodology in Annex I of this thesis). The aim of this analysis was to assess whether the ecological trends observed were conserved despite the potential biases associated with the choice of primers. We selected a total of 16 samples from the North Atlantic Ocean, since this region was not covered in **Chapter II**, and the inclusion of these samples could complement the results observed for the surface AAP communities in the South and Mid Atlantic Ocean in **Chapter III**. The selected stations covered three Longhurst provinces (Figure 1A) that exhibited strong gradients of salinity, temperature, and fluorescence, as depicted in Figure S2 of **Chapter I**.

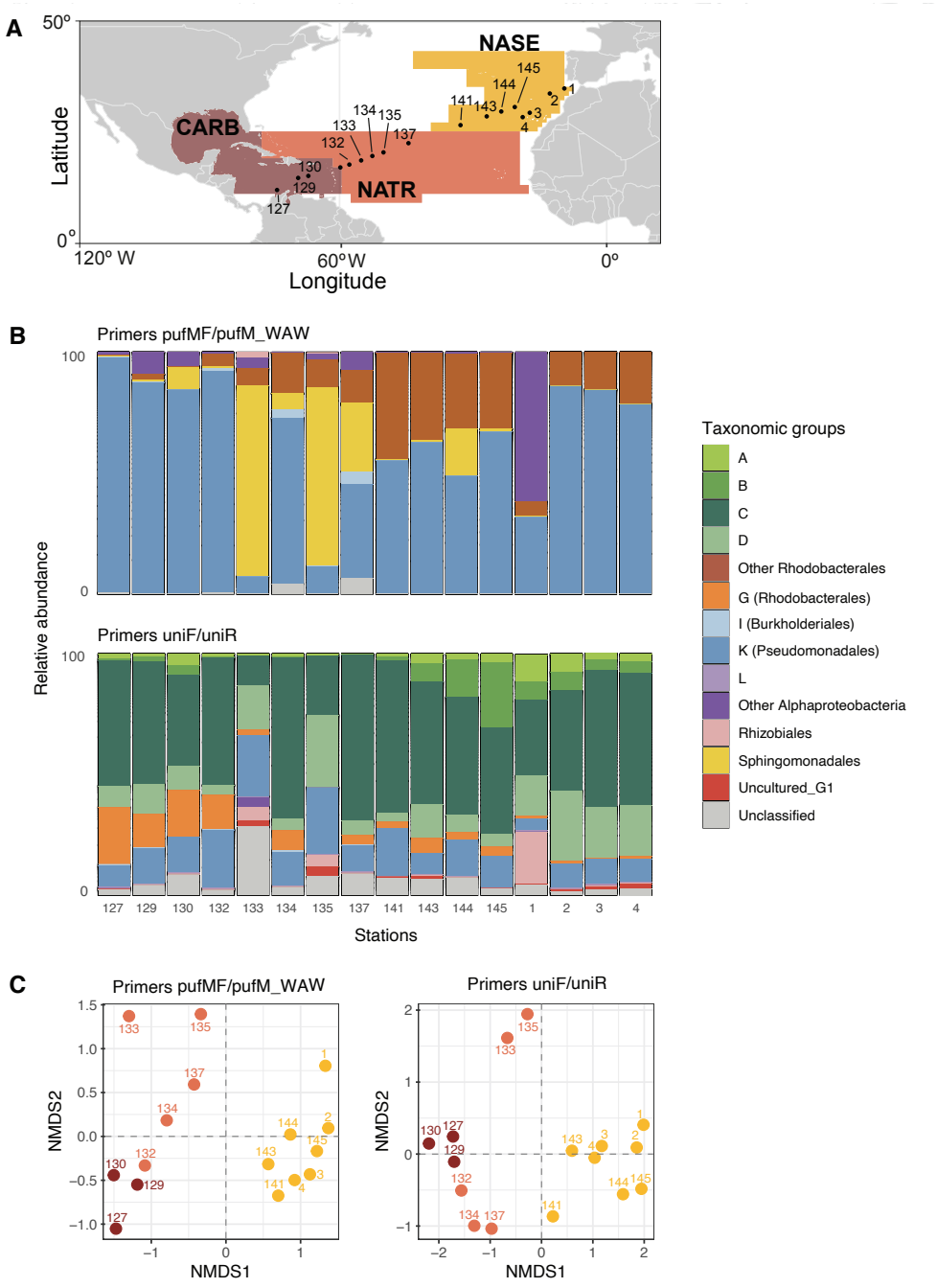

Figure 1. Analysis of North Atlantic Ocean communities with different sets of primers. A) Stations selected for the analysis. Numbers correspond to the station numbering in the Malaspina Expedition located within the following Longhurst regions: CARB; Caribbean Sea, NASE, Northeast Atlantic subtropical gyre, NATR, North Atlantic tropical gyre. B) Community composition of these samples according to the used primer set. C) Non-metric multidimensional scaling (NMDS) of these samples based on Bray-Curtis dissimilarity distances.

The community composition obtained using the different primers is clearly different, as illustrated in Figure 1B. While this observation has been extensively described and discussed in **Chapter II**, it is noteworthy to highlight the differential amplification patterns in the Rhodobacterales order. Sequences associated to Rhodobacterales appear in both amplicon approaches (Figure 1B), however, they belong to different phylogenetic clusters within this order. The Rhodobacterales-like ASVs from the upper panel, amplified with primers pufMF/pufM_WAW, are mainly associated to phylogroups E and F and we classified them as 'Other Rhodobacterales'. This group was abundant in stations within the oligotrophic North Atlantic gyre (stations 141-145). In contrast, Rhodobacterales-like ASVs from the lower panel (Figure 1B), amplified with primers UniF/UniR, were almost exclusively from Phylogroup G (also from the Rhodobacterales order) and peaked in stations within the Caribbean Sea, with high values of fluorescence. Interestingly, phylogroup G includes *Roseobacter*-like bacteria that are usually associated to phytoplankton blooms (Buchan et al., 2005; Wagner-Döbler and Biebl, 2006; Alonso-Sáez et al., 2015). This observation was not apparent in **Chapter II**, primarily due to the limited number of samples tested. Extending the analysis to a larger number of samples and conducting a thorough inspection of the phylogeny within this clade allowed to report these differences. When examining the sequences of phylogroup G, we observed that they contained several mismatches with the pufMF primer (Figure 1 in **Chapter II**), which probably difficulted its amplification in the study of Chapter I, as well as in previous studies (e.g. Auladell et al., 2019). According to these results, Rhodobacterales are not abundant in the open ocean but some subgroups of this order have a high affinity for the pufMF primer, which may result in an overestimation of their relative contribution to AAP assemblages in those areas. Moreover, it could aslo explain the high abundance of Alpha-Rhodobacterales observed in the Malaspina Expedition samples from the oligotrophic gyre (**Chapter I**).

Another worthwhile result to comment is the comparison between the taxonomic composition of the analysis presented, with the communities studied in **Chapter III**, pertaining to the South and Mid Atlantic Ocean, using the same set of primers. Despite the 8 years of difference between both samplings (2010-2011 for the Malaspina Expedition and 2019 for the Poseidon Expedition) and the fact that the cruises took place at different seasons (the Leg 7 of Malaspina Expedition occurred in June and July of 2011, and the Poseidon Expedition took place in March 2019), the composition is comparable (Figure 3, **Chapter III**, Figure 1 in the Discussion), indicating that a similar composition of AAP assemblages might be found across regions. In any case, this should be further tested

since we cannot discard variations among different seasons or differences with other ocean regions.

In terms of community structure, in **Chapter II** we showed that despite primer biases, the clustering of samples belonging to different oceanic regions was significantly consistent, indicating that similarities between samples were conserved between different amplicon approaches. Here, we observed a similar structuring of the samples according to their Longhurst province as seen by their clustering in the plots (Figure 1C) and the high correlation between Bray-Curtis dissimilarity matrices (Mantel correlation, R=0.7419 , *p*=0.001). Samples from CARB and NATR provinces form a separate cluster, while samples from the NASE, within the oligotrophic gyre, are more disperse. These results are in agreement with the observation that the oligotrophic gyres are transition areas and AAP communities in those provinces display higher dissimilarities (Figure 4 in **Chapter III**). The fact that samples from the same Longhurst region clustered together was mainly explained by the high selection occurring in surface ocean communities (**Chapter I**), and it is likely the explanation for this result (Figure 1C). In addition, in both **Chapter I** and in this analysis, stations 133, 135 and 1 (Figure 1B), are more dissimilar than their adjacent stations, evidencing that regardless of the used primers, these patterns are maintained. Altogether, we found enough evidences supporting the hypothesis that the ecological trends observed in Chapter I using primers pufMF/pufM_WAW are factual, even though these primers do not have a good coverage for some groups of AAP bacteria.

2. Shedding light into the unknown diversity of AAP bacteria

This thesis places phylogroups A, B, C, and D into a central position within the diversity of AAP bacteria, as we show that they are significant contributors to AAP communities in the open ocean (**Chapter II** and **Chapter III**). After being described for the first time from the metagenomic Ocean Global Expedition survey (Yutin et al., 2007), these phylogroups were subsequently overlooked in amplicon-based studies and were regarded as minor components of AAP bacterial communities. As a result, the information on the ecology of these groups is scarce. In **Chapter II**, we prove their ubiquity across the surface global ocean, and we detect the seasonality of phylogroups A and D in the Mediterranean Sea, missed in the long-term seasonal study of AAP bacteria conducted by Auladell et al. (2019) at the Blanes Bay Microbial Observatory (BBMO). In **Chapter III**, we document for the first time their horizontal and vertical patterns along the water column, and provide some data on their potential habitat preferences. For example, phylogroup C is especially

abundant in surface waters, it negatively correlates to depth and, contrary to phylogroups A and B, does not seem to be associated with the trophic status of the water. Although being abundant in open ocean samples, this phylogroup is scarcely found in samples from the Mediterranean Sea, even when using the same primer set (BBMO; Chapter II), as well as in the Alboran Sea (Santos-Bruña and Ferrera, unpublished results). In contrast, phylogroup B seems to be more abundant in areas with higher chlorophyll values, and phylogroup A shows a seasonal signal, dominating AAP communities in the NW Mediterranean Sea during winter (**Chapter II**).

Although phylogroups A, B, C, and D exhibit different ecological behavior, they are closely related phylogenetically. The phylogenetic analysis of these groups shows that they form a distinct sister clade to the Rhodobacterales order (Alphaproteobacteria) (Figure 2). A limited number of reconstructed metagenomes-assembled genomes (MAGs) deposited in the Genome Taxonomy Database (GTDB) cluster them within this clade and have also been assigned to the Rhodobacterales. Notably, phylogroup A contains the *pufX* gene in its operon, a gene previously reported only in anaerobic and aerobic Rhodobacterales (Yutin and Beja, 2005). Interestingly, the analyses of the photosynthetic operon in samples from the Tara Ocean dataset granted the discovery of some genomes with potential for anoxygenic phototrophy and carbon fixation via the Calvin-Benson-Bassham cycle (Graham et al., 2018). Due to the novel nature of this clade within the Alphaproteobacteria, they proposed the creation of a family named '*Candidatus* Luxescamonaceae'. The inclusion of these *pufM* sequences in our phylogenetic trees (Figure S5 in **Chapter I** and Figure S5 in **Chapter III**) showed that they cluster within phylogroups C and D. Given the little information available for these groups, caution should be taken before assuming that phylogroups C and D are part of the '*Candidatus* Luxescamonaceae' family or that they also possess the capacity for carbon fixation. Nevertheless, these observations open several intriguing questions related to their functionality and evolutionary history.

In fact, the confirmation of a certain functional ability of a given organism, for example the carbon fixation capacity, cannot be based on genomics alone, particularly if it is done only through the reconstruction of MAGs, since this approach has limitations related to for instance to their completeness or contamination that can lead to potential misinterpretations (Bowers et al., 2017; Parks et al., 2017). Obtaining Single Amplified Genomes (SAGs) of '*Candidatus* Luxescamonaceae' would be instrumental in confirming the presence of RuBisCO in their genomes, as SAGs bioinformatic reconstruction circumvents taxonomic binning issues (Swan et al., 2011). In any case, the ultimate proof to address this question

would lie in obtaining isolates and experimentally testing the ability of these bacteria to incorporate CO_2. Additionally, the coexistence of anoxygenic phototrophy and carbon fixation capacity in bacteria closely related to the Rhodobacterales entangles discussion regarding the evolution in the *Rhodobacteraceae*.

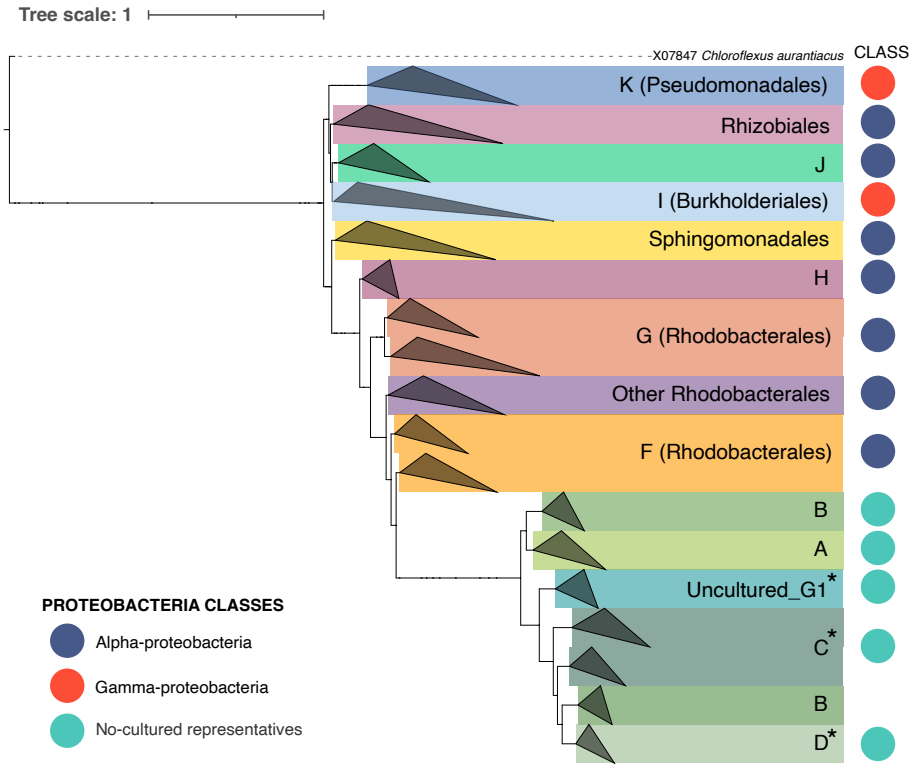

Figure 2. Phylogenetic tree of *pufM* sequences. Clades from the same phylogroup have been collapsed. The tree was constructed with RAxML v8.2 (Stamakis 2014) and visualized and edited using iTOL (Letunic and Bork, 2011). *Asterisks indicate the clades containing sequences from the '*Candidatus* Luxescamonaceae' cluster.

There are two scenarios commonly contemplated in the evolution of anoxygenic phototrophy within *Rhodobacteraceae*: i) phototrophy is an ancestral trait that has been frequently lost, or ii) the photosynthetic capacity was acquired via horizontal gene/operon transfer (HTG or HOT). The loss of photosynthetic genes has been suggested in the case of haloalkaliphilic Rhodobacterales (Kopetjka et al., 2017) or the *Roseobacter* clade (Koblížek et al., 2013). The analysis of 33 concatenated core genes within the photosynthetic gene cluster (PGC) belonging to the Rhodobacteraceae suggested a heterotrophic last common ancestor and the introduction of the PGC by HOT (Brinkmann

et al., 2018). This theory has also been used to explain the acquisition of anoxygenic photosynthesis in the phylum Gemmatimonadota (Zeng et al., 2014). Also, the discovery of a putative PGC encoding a type-II reaction center in the Myxococcota phylum suggests a common phototrophic ancestor (Liu et al., 2023) of this phylum with the Proteobacteria. Both scenarios are not mutually exclusive, and, in fact, the evolution of photosynthesis is probably a combination of vertical evolution, horizontal transfer processes, and PGC losses. Further investigating the function and evolution of marine AAP bacteria along with their putative capacity for carbon fixation could provide valuable insights to this complex topic.

3. Towards defining the ecological role of AAP bacteria in the ocean

The diversity approaches applied in this thesis have provided an accurate picture of the structure and diversity of open ocean AAP communities, unraveling the processes structuring them in the horizontal and vertical scales. These studies significantly contribute to a cumulative two-decade-long research into the distribution of AAP bacteria across different aquatic systems, the composition of their communities, and their ecological role. Although the capability of light harvesting distinguishes this group from other heterotrophic bacteria, there has been limited exploration of the activity and physiological role of AAP bacteria in their natural environments. Culture-based studies showed that AAP species increase their carbon metabolism and decrease their respiration rates when exposed to light (e.g. Hauruseu and Koblížek 2012; Piwosz et al., 2018), and the study of natural populations revealed that light enhances the growth rates of AAP bacteria (Ferrera et al., 2017; Sánchez et al., 2020). In fact, at optimal light availability AAP bacteria exhibited a high growth efficiency in freshwater lakes (Piwosz et al., 2020).

The light harvesting bioenergetics of AAP cells has been investigated mainly based on cultured species (Koblízek et al., 2010; Selyanin et al. 2016; Piwosz et al., 2018) due to the difficulties of making direct measurements and estimates on natural populations. Moreover, Kirchman and Hanson (2013) presented the first theoretical approach in which they showed that the light energy captured by AAP cells was higher than for proteorhodopsin (PR)-containing cells. However, the high abundance of PR-containing bacteria make this group a major contributor to the total solar energy captured (Gómez-Consarnau et al., 2019). In **Chapter IV**, using cell abundances and pigment measurements, we present the first assessment of the light energy capture in the open ocean, enabling predictions of their photoheterotrophic activity. We show that the solar energy capture is highest in more

eutrophic areas, specially in the surface ocean, and that it rapidly decreases with depth. Besides the obvious benefits that the phototrophic activity has for AAP bacteria, the use of solar energy also reduces the demand for organic carbon consumption, impacting carbon cycling in the oceans. It has been speculated that the contribution of Bchl a based phototrophy and the shift of their metabolism form purely heterotrophic to phototrophic could determine whether an oceanic region acts as sink or a source of atmospheric CO_2 (Jiao et al., 2010). In fact, a recent study from freshwater lakes demonstrated that the dark measurements of microbial activity overestimate the respiration rates of AAP bacteria, masking their potential impact on heterotrophic systems (Piwosz et al., 2021). In this thesis, I have provided insights into the diversity and phototrophic activity (i.e. light capture) of AAPs in their natural environment. Future research should try to couple activity and diversity studies, to answer '*who*' is doing '*what*' (Munson-McGee et al., 2002). The first attempts to do so already showed the importance of AAP bacteria in modulating carbon fluxes in lakes and elucidated the contribution of some groups to respiration rates (Piwosz et al., 2021).

An unresolved ecological matter is whether AAP bacteria have a preference for a free-living or a particle-attached lifestyle. While they have mostly been studied as free-living organisms, different studies have shown a variable proportion of total AAP abundance in the particle-attached AAP fraction (10–90%) (Waidner and Kirchman 2007; Lami et al., 2009; Cottrell et al., 2010). We intended to delve deeper into this matter during the Poseidon Expedition, and for this purpose, we collected water for DNA analyses using a serial filtration protocol that resulted in three different size fractions: the small one (0.2–3 µm) representing the picoplankton size fraction, primarily composed of free-living organisms; medium size class (3–20 µm) corresponding to the nanoplankton size fractions and bacteria attached to small particles; and the large size class (20–200 µm) containing the microplankton size fraction and bacteria attached to larger particles. Unfortunately, we were unable to amplify any DNA sample from the medium and large size fractions, likely due to the low concentrations of DNA obtained after extracting these filters. The number of copies of the *pufM* gene in these DNA extracts may have been insufficient for successful amplification. Moving forward, to address this issue requires modification of the sampling protocol. This could involve increasing the volume of seawater filtered from the Niskin bottles, by for example, using in situ pumps, or alternatively, by directly collecting particles using a marine snow catcher. An interesting question would be to test whether phylogroups A, B, C, and D also dominate the AAP communities associated with particles or whether the taxonomic composition varies among different size fractions.

Finally, the theory suggesting that AAPs may have a central role in transforming phytoplankton-derived dissolved organic matter (DOM) has been postulated in the past (Koblížek et al., 2015; Piwosz et al., 2020) and is also being hypothesized in this thesis. In fact, the so-mentioned Roseobacter and Rhodobacter clades are important in recycling DOM produced by associated phytoplankton (Buchan et al. 2014). The results of **Chapter III**, where we analyzed the cooccurrence of AAP bacteria and phytoplankton, revealed stronger interactions between them, as compared with the rest of the bacterioplankton. Additionally, the distribution of cells and pigments along the spatial and vertical scale (**Chapter IV**) evidenced a strong coupling between both groups. Exploring the relationship between phytoplankton and AAP bacteria in more detail could provide valuable information on the trophic interactions of AAP bacteria and their role in the carbon cycle. Studying the composition of AAP communities during blooms of different types of phytoplankton could illustrate how different AAP bacterial groups co-vary with algae during the course of a bloom. Besides, combining these studies with transcriptomics could allow to further identify the response of different AAP lineages over time, as it has been shown previously (Fecskeová et al., 2021). Moreover, experiments growing AAP populations on dissolved organic matter at different concentrations and with different quality, under changing light regimes would shed light on the interplay between AAP bacteria, organic carbon, and light.

A final note

All the environmental data used in this thesis is based on samples collected across oceanographic expeditions and a microbial observatory. While none of these sampling exercises was designed to study AAP bacteria, the dedication and effort of numerous scientists has permitted to obtain valuable data, which in turn has allowed to delineate the global distribution, diversity, and community structure trends of AAP bacteria in the marine environment. Additionally, this collaborative effort has been crucial to improve the molecular techniques used to study them.

In essence, my thesis on the ecology of aerobic anoxygenic phototrophic bacteria contributes to the advancement of knowledge in the field of marine microbial ecology, by delving into the study of this functional group in the global ocean. At the same time, it serves as a starting point to numerous future studies that I have pointed out along the general discussion, which will enhance our understanding of AAP bacteria and, more broadly, marine microbial ecology.

References

Alonso-Sáez L, Díaz-Pérez L, & Morán XAG. (2015) The hidden seasonality of the rare biosphere in coastal marine bacterioplankton. *Environ Microbiol.* 2015;17:3766–80.

Auladell, A., Sánchez, P., Sánchez, O. et al. (2019) Long-term seasonal and interannual variability of marine aerobic anoxygenic photoheterotrophic bacteria. *ISME* 13, 1975–1987. https://doi.org/10.1038/s41396-019-0401-4

Bibiloni-Isaksson, J., Seymour, J.R., Ingleton, T., van de Kamp, J., Bodrossy, L., & Brown, M.V. (2016) Spatial and temporal variability of aerobic anoxygenic photoheterotrophic bacteria along the East coast of Australia. *Environ Microbiol* 18: 4485–4500. https://doi.org/10.1111/1462-2920.13436

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

Brinkmann, H., Göker, M., Koblížek, M. et al. (2018) Horizontal operon transfer, plasmids, and the evolution of photosynthesis in *Rhodobacteraceae*. *ISME* 12, 1994–2010. https://doi.org/10.1038/s41396-018-0150-9

Buchan A, González JM, & Moran MA. (2005) Overview of the marine *Roseobacter* lineage. *Appl Environ Microbiol.* 200571(10):5665-77. doi: 10.1128/AEM.71.10.5665-5677.2005. PMID: 16204474; PMCID: PMC1265941.

Caliz, J., & Casamayor, E.O. (2014) Environmental controls and composition of anoxygenic photoheterotrophs in ultraoligotrophic high-altitude lakes (Central Pyrenees). *Env Microbiol Rep* 6(2), 145–151

Cottrell MT, Ras J, & Kirchman DL. (2010) Bacteriochlorophyll and community structure of aerobic anoxygenic phototrophic bacteria in a particle-rich estuary. *ISME* 4:945–54.

Duarte CM (2015) Seafaring in the 21st century: the Malaspina 2010 circumnavigation expedition. Limnol Oceanogr Bull 24:11–14. https://doi.org/10.1002/lob.10008

Fecskeová, L.K., Piwosz, K., Hanusová, M. et al. (2019) Diel changes and diversity of *pufM* expression in freshwater communities of anoxygenic phototrophic bacteria. *Sci Rep* 9, 18766 (2019). https://doi.org/10.1038/s41598-019-55210-x

Ferrera I, Borrego CM, Salazar G, et al. (2014) Marked seasonality of aerobic anoxygenic phototrophic bacteria in the coastal NW Mediterranean Sea as revealed by cell abundance, pigment concentration and pyrosequencing of pufM gene. *Environ Microbiol* 16:2953–65.

Ferrera I, Sánchez O, Kolárová E, Koblížek M, Gasol JM. (2017) Light enhances the growth rates of natural populations of aerobic anoxygenic phototrophic bacteria. *ISME* 11:2391–2393. https://doi.org/10.1038/ismej.2017.79.

Graham ED, Heidelberg JF, Tully BJ (2018) Potential for primary productivity in a globally-distributed bacterial phototroph. ISME 12:1861–1866. https:// doi. org/ 10. 1038/ s41396- 018- 0091-3

Gómez-Consarnau, L.,Raven, JA., Levine, NM., Cutter, LS., Wang, D., Seegers, B., Arístegui, J., Fuhrman, JA:, Gasol, JM., & Sañudo-Wilhelmy, SA. (2019) Microbial rhodopsins are major contributors to the solar energy captured in the sea. *Science Advances* 10.1126/sciadv. aaw8855, 5, 8.

Hauruseu D, & Koblížek M. (2012) The influence of light on carbon utilization in aerobic anoxygenic phototrophs. *Appl Environ Microb* 2012;78:7414–9.

Koblížek, M., Mlčoušková, J., Kolber, Z., & Kopecký, J. (2010) On the photosynthetic properties of marine bacterium COL2P belonging to *Roseobacter* clade. *Arch Microbiol* 192: 41–49. doi:10.1007/s00203-009-0529-0

Kopejtka K, Tomasch J, Zeng Y, Tichý M, Sorokin DY, Koblížek M. (2007) Genomic analysis of the evolution of phototrophy among haloalkaliphilic Rhodobacterales. *Genome Biol Evol.* 9:1950–62.

Jeanthon, C., Boeuf, D., Dahan, O., le Gall, F., Garczarek, L., Bendif, E.M., & Lehours, A.C. (2011) Diversity of cultivated and metabolically active aerobic anoxygenic phototrophic bacteria along an oligotrophic gradient in the Mediterranean Sea. *Biogeosciences* 8: 1955–1970. https://doi.org/10.5194/bg-8-1955-2011

Jiao, N., Zhang, F. & Hong, N. (2010) Significant roles of bacteriochlorophyll a supplemental to chlorophyll a in the ocean. *ISME* 4, 595–597 https://doi.org/10.1038/ismej.2009.135

Kirchman, D. L., & Hanson T. E. (2013) Bioenergetics of photoheterotrophic bacteria in the oceans. *Environ Microbiol Rep* 5: 188–199. doi:10.1111/j.1758-2229.2012.00367.x

Koblížek, M, Mlčoušková, J., Kolber Z, et al. (2010) On the photosynthetic properties of marine bacterium COL2P belonging to *Roseobacter* clade. *Arch Microbiol* 2010;192:41–9.

Koblížek M, Zeng Y, Horák A, Oborník M. (2013) Regressive evolution of photosynthesis in the *Roseobacter* clade. *Adv Bot Res.* 66:385–405.

Koblížek, M. (2015) Ecology of aerobic anoxygenic phototrophs in aquatic environments. *FEMS Microbiol Rev* 39: 854–870.

Munson-McGee, J.H., Lindsay, M.R., Sintes, E. et al. (2022) Decoupling of respiration rates and abundance in marine prokaryoplankton. *Nature* 612, 764–770 https://doi.org/10.1038/s41586-022-05505-3

Lami R, Čuperová Z, Ras J, et al. (2009) Distribution of free-living versus particle-attached aerobic anoxygenic phototrophic bacteria in marine environments. *Aquat Microbial Ecol* 55:31–8.

Lehours, A., & Jeanthon, C. (2015) The hydrological context determines the beta-diversity of aerobic

anoxygenic phototrophic bacteria in European Arctic seas but does not favor endemism. *Front Microbiol* 6: 638. https://doi.org/10.3389/fmicb.2015.00638

Lehours, A.C., Cottrell, M.T., Dahan, O., Kirchman, D.L., & Jeanthon, C. (2010) Summer distribution and diversity of aerobic anoxygenic phototrophic bacteria in the Mediterranean Sea in relation to environmental variables. *FEMS Microbiol Ecol* 74: 397–409. https://doi.org/10.1111/j.1574-6941.2010.00954.x

Lehours, A.C., Enault, F., Boeuf, D., & Jeanthon, C. (2018) Biogeographic patterns of aerobic anoxygenic phototrophic bacteria reveal an ecological consistency of phylogenetic clades in different oceanic biomes. *Sci Rep* 8: 1–10. https://doi.org/10.1038/s41598-018-22413-7

Parks, D. H., Rinke, C., Chuvochina, M., Chaumeil, P. A., Woodcroft, B. J., Evans, P. N., ... & Hugenholtz, P. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* 2(11), 1533-1542.

Pesant S, Not F, Picheral M, Kandels-Lewis S, Le Bescot N, Gorsky G, et al. (2015) Open science resources for the discovery and analysis of Tara Oceans data. *Sci Data*. 2:150023–16.

Piwosz, K., D. Kaftan, J. Dean, J. Šetlík, & M. Koblížek. (2018) Nonlinear effect of irradiance on photoheterotrophic activity and growth of the aerobic anoxygenic phototrophic bacterium Dinoroseobacter shibae. *Environ Microbiol* 20: 724–733. doi:10.1111/1462-2920.14003

Piwosz K, Vrdoljak A, Frenken T, et al (2020) Light and primary production shape bacterial activity and community composition of aerobic anoxygenic phototrophic bacteria in a microcosm experiment. *mSphere* 5:e00354-20. https://doi.org/10.1128/mSphere.00354-20.

Piwosz, K., Villena-Alemany, C. & Mujakić, I. (2022) Photoheterotrophy by aerobic anoxygenic bacteria modulates carbon fluxes in a freshwater lake. *ISME* 16, 1046–1054. https://doi.org/10.1038/s41396-021-01142-2

Sánchez P, Sebastián M, Pernice M et al. (2023) Marine picoplankton metagenomes from eleven vertical profiles obtained by the Malaspina Expedition in the tropical and subtropical oceans. *bioRxiv*. https://doi.org/10.1101/2023.02.06.526790

Sánchez, O., Ferrera, I., Mabrito, I. et al. (2020) Seasonal impact of grazing, viral mortality, resource availability and light on the group-specific growth rates of coastal Mediterranean bacterioplankton. *Sci Rep* 10, 19773 https://doi.org/10.1038/s41598-020-76590-5

Salka I, Čuperová Z, Mašín M, Koblížek M, Grossart H-P. (2011) Rhodoferax-related *pufM* gene cluster dominates the aerobic anoxygenic phototrophic communities in German freshwater lakes. *Environ Microbiol* 2011; 13: 2865–2875.

Selyanin, V., Hauruseu, D. & Koblížek, M. (2016) The variability of light-harvesting complexes in aerobic anoxygenic phototrophs. *Photosynth Res* 128, 35–43 https://doi.org/10.1007/s11120-015-0197-7

Villena-Alemany, C., Mujakić, I., Porcal, P., Koblížek, M. & Piwosz, K. (2023) Diversity dynamics of

aerobic anoxygenic phototrophic bacteria in a freshwater lake. *Environ Microbiol Rep*15(1), 60–71. Available from: https://doi.org/10.1111/1758-2229.13131

Villena-Alemany, C. et al. (2023) Phenology and ecological role of Aerobic Anoxygenic Phototrophs in fresh Waters *bioRxiv* 11.17.567504; doi: https://doi.org/10.1101/2023.11.17.567504

Waidner, L.A., & Kirchman, D.L. (2007) Aerobic anoxygenic phototrophic bacteria attached to particles in turbid waters of the Delaware and Chesapeake estuaries. *Appl Environ Microb* 2007;73:3936–44.

Wagner-Döbler, I., & Biebl, H. (2006). Environmental biology of the marine *Roseobacter* lineage. *Ann Review of Microbiol* 60, 255-280.

Yutin, N., Suzuki, M., Teeling, H., Weber, M., Venter, J.C., Rusch, D.B., & Béjà, O. (2007) Assessing diversity and biogeography of aerobic anoxygenic phototrophic bacteria in surface waters of the Atlantic and Pacific Oceans using the Global Ocean Sampling expedition metagenomes. *Environ Microbiol* 9:1464–1475. https://doi.org/10.1111/j.1462-2920.2007.01265.x

Yutin, N., & Béjà, O. (2005) Putative novel photosynthetic reaction center organizations in marine aerobic anoxygenic photosynthetic bacteria: insights from environmental genomics and metagenomics. *Environ Microbiol* 7: 2027–2033.

CONCLUSIONS

CONCLUSIONS

i. In the surface tropical and subtropical ocean, AAP bacteria displayed contrasting values of alpha diversity in different areas, with the lowest diversity values observed in the most productive regions. The trophic status, together with salinity and temperature, were the primary variables explaining changes in both diversity and community structure of AAP communities across these regions.

ii. Community dissimilarities among AAP assemblages increased with distance between stations, and this pattern was linked to their positioning within the different Longhurst provinces, particularly at the borders of certain provinces, where there were significant shits in the composition and structure of AAP communities.

iii. The main ecological processes shaping the structure of AAP communities are a combination of homogeneous and heterogeneous selection, with small changes in the environmental conditions that translate into substantial changes in the composition of AAP communities.

iv. The biogeographic patterns observed for the surface AAP communities differ from those observed in the bulk bacterioplankton, indicating that AAP bacteria exhibit their own ecological trends across the surface ocean.

v. The primers traditionally employed in previous amplicon-based studies targeting the *pufM* gene present serious biases towards specific phylogroups. These primers tend to overestimate the relative contribution of Gammaproteobacteria and some Alphaproteobacteria clades, while hindering the amplification of other groups. This has likely resulted in the misrepresentation of the AAP diversity in prior studies employing these primer set.

vi. The combination of metagenomics and amplicon sequencing approaches has proven to be a useful tool to evaluate primer-biases, allowing the design of new primers with broader phylogenetic coverage and the testing of their performance in environmental samples.

vii. A significant fraction of AAP populations in the open ocean consists of understudied species belonging to phylogroups A, B, C, and D. The high number of mismatches in the forward primer region for these groups has difficulted their recovery in previous studies. Using primers UniF/UniR and pufM_F/pufM_WAW avoids these large

primer biases and unveils a good representation of the taxonomic composition of AAP communities.

viii. The application of primers UniF/UniR in samples from the South and Mid Atlantic Ocean illustrate that phylogroups A, B, C, and D are important components of AAP bacterial communities both in surface waters and along the water column. Besides, although they are phylogenetically related, these groups display contrasting ecological trends. The ecological significance of these populations in marine ecosystems is underexplored and merits further investigation.

ix. Throughout the water column in the epipelagic zone, AAP communities exhibit pronounced vertical trends linked to the deep chlorophyll maximum (DCM) profile. The taxonomic composition of samples changes across the different layers of the DCM structure, with highest richness values at the DCM layer. In turn, the latitudinal gradient has a lower impact on the taxonomic composition of AAP communities. However, it does influence alpha diversity, with AAP communities displaying higher richness values around the equator.

x. In the Atlantic Ocean, AAP bacteria are widespread from the surface down to a depth of 250 m, being more abundant in areas with higher chlorophyll a concentration. Along the vertical axis, AAP bacteria follow a similar trend to that observed in phytoplankton, with cell abundances and pigment concentrations reaching highest values at the DCM or just above it.

xi. The correlation observed between phytoplankton and AAP abundances, together with a notable number of species co-occurrences between both groups of organisms, provides valuable insights into a possible coupling between these groups that deserves further investigation.

xii. The estimation of the bioenergetic role of AAP cells in the Atlantic Ocean shows that solar energy capture is higher in more eutrophic areas, especially at the ocean surface, and that it rapidly decreases with depth. While the energy gained from light is lower compared to that captured by the phytoplankton, and while it cannot cover the energetic cost of bacterial maintenance, it likely contributes to some of the oxidative phosphorylation and thus saves substrates that can be derived to biosynthesis.

Annex I

The sample collection and DNA extraction was performed as described in Chapter I (Gazulla et al., 2023), since the 16 samples used in this analysis are from the Malaspina Expedition.

Amplification and sequencing of the* pufM *gene

Partial amplification of the *pufM* gene (~180 bp fragment) was carried out in 12.5 µL reactions using primers pufM_uniF (GGNAAYYTNTWYTAYAAYCCNTTYCA) and pufM_UniR (YCCATNGTCCANCKCCARAA), from Yutin et al. (2005). The amplification followed the conditions detailed in Gazulla et al. (2023). Briefly, PCR conditions were as follows: an initial denaturation step at 95 °C for 5 min and 35 cycles at 95 °C (30s), 48°C (45s), 72 °C (45s), and a final elongation step at 72 °C for 7 min. DNA sequencing was conducted on an Illumina MiSeq sequencer by AllGenetics & Biology SL (www.allgenetics.eu).

Amplicon sequence variants generation and taxonomic assignation

We used cutadapt v3.4 (Martin, 2013) for the removal of primers, and DADA2 v1.26 (Callahan et al., 2016) to discern amplicon sequence variants (ASVs) and eliminate chimeras. To infer the phylogeny of these amplicon sequence variants, we utilized phylogenetic placement with the Evolutionary Placement Algorithm v0.3.5 (Barbera et al., 2019) and a custom made pufM database (Gazulla et al., 2023). The taxonomic assignation was done as described in Gazulla et al., 2023 (Chapter III).

Data analyses

All analyses were conducted in R v4.2.0 (R Core Team 2022), using *phyloseq* (McMurdie and Holmes, 2013) and *tidyverse* (Wickham et al., 2019). Ordination tests were based on the Bray-Curtis distance index and calculated with the *vegdist* function from *vegan* (Oksanen et al., 2022). Figures were done with *ggplot* (Wickham et al., 2016).

'Retxes de sol atravessen blaus marins,
ses algues tornen verdes i brillen ses estrelles,
que ja s'ha fet de nit i es plàncton s'il·lumina
i canten ses balenes a 30.000 quilòmetres d'aquí.'

Antònia Font, *Batiscafo Katiuscas*


UAB
Universitat Autònoma
de Barcelona
Institut
de Ciències
del Mar
INSTITUTO ESPAÑOL DE OCEANOGRAFÍA