# Structure of the microbial communities in marine hotspots for climate change

Amaranta Focardi

This thesis is presented in partial fulfilment to the requirements for a

Doctorate of Philosophy

Department of Molecular Sciences

Macquarie University, NSW, Australia

November 2019

MACQUARIE University

# Table of Contents

# List of figures

**Chapter 1: Introduction**



**Chapter 2: Change in the relative contribution of picophytoplankton, bacteria and viruses to the carbon biomass in a model western boundary current**

**Chapter 4: Characterization of the microbial community of the Dalton polynya, an Antarctic hotspot for climate change**

**Chapter 5: General conclusion and future direction**

# List of tables

**Chapter 2:**

# List of supplementary information

**Chapter 2**

SF 2.1 Carbon calibration curve

SF 2.2 Gating strategies photosynthetic communities

SF 2.3 Gating strategies viruses and bacteria

SF 2.4 Contribution to carbon biomass distance from the coast

ST 2.1 Environmental variable

ST. 2.2 Count and carbon quota

**Chapter 3**

ST 3.1 Contigs count

ST 3.2 Metadata

ST 3.3 Contigs information

ST 3.4 Pfam count in EAC, Tasman and aphotic

**Chapter 4**

ST 4.1 Metadata

ST 4.2 Normalized 16S rRNA ASV

ST 4.3 Normalised 18S rRNA ASV

ST 4.4 Cytometry count

For the supplementary material please refers to the following link:
https://www.dropbox.com/sh/o5ans1b40adulhb/AADpDYerd9Ym5L2yAXulqNrYa?dl=0

## Declaration

I declare that this work has not previously been submitted for a degree or diploma in any university. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

Amaranta Focardi

## Acknowledgments

I thank my supervisors, Prof. Ian Paulsen and Dr. Martin Ostrowski, for their support, guidance, and the opportunity to work in a great team.

Particular thank goes to Martin Ostrowski for his incredible patience over the past three years and for his guidance on bioinformatics.

Thanks to all the Paulsen group members.

Lastly, thanks to my family and to Giuseppe, for their endless support and for believing in me.

## Conference proceedings

### Oral presentations

- Focardi A., Ostrowski M., Paulsen I., Assessing the role of viruses on the carbon export in a warming ocean. ACOMO, October 2018, Canberra

- Presentation to Palenik lab at Scripps Institution of Oceanography, San Diego, February 2018

- Focardi A., Viruses identification in sediment: Potential use of high Throughput Flow sorting and meta-genomic analysis. Workshop - Assessing the potential of ancient DNA in marine sediments- Sydney, October 2017

### Poster presentations

- Focardi A., Ostrowski M., Gooseen K., Brown M., Paulsen I., Investigation of the bacteriophage diversity along a western boundary current. AusME, Perth, February 2019

- Focardi A., Ostrowski M., Paulsen I., Cyanobacteria and the carbon flux: the impact of viral mortality in a warming ocean, JAMS, Sydney 2018

- Focardi A., Ostrowski M., Paulsen I., Grazing- and Viral- induced mortality on pico-cyanobacteria population along the East Australian current compared with the Tasman Sea. Ocean Science Meeting, Portland 2018

## Achievements

- Awarded the Postgraduate Research Fund for conference travel from Macquarie University, Australia
- Awarded Macquarie University Research Excellence Scholarship (MQRES) to undertake the doctoral degree

## Research expeditions

- IN2017_V01 RV Investigator - Interactions of the Totten Glacier with the Southern Ocean through multiple glacial cycles – (52 days) January-March 2017
- IN2016_V04 on board the RV Investigator – Influence of temperature and nutrient supply on the biogeochemical function and diversity of ocean microbes – (24 days) August -September 2016.

## Biosafety approvals

5201401141 Handling low level opportunistic bacterial pathogens and GMOs carrying genes from these organisms

520170100114 Molecular ecology of the Southern Ocean and Antarctica

# Contributions

**CHAPTER 2:**

Work was conceptualized by Focardi, Paulsen and Ostrowski. Samples were collected by Focardi, with the help of Ostrowski and colleague Deepa Varkey. Flow cytometry was performed by Focardi, flow sorting by Focardi with help from Ostrowski. Analysis for the carbon content was done at the Bigelow lab for Ocean science in Maine, USA. All data analysis was performed by Focardi. Manuscript was written by Focardi with contribution from Ostrowski and Paulsen.

**CHAPTER 3:**

Work was conceptualized by Focardi, Paulsen and Ostrowski. Samples were collected by Focardi, Gooseen and Brown, with the help of Ostrowski. DNA extraction was performed by Focardi and Gooseen. DNA was sequenced at the Ramaciotti centre for genomics. Sequence assembly was performed by Focardi and Ostrowski. All data analysis was performed by Focardi. Manuscript was written by Focardi with contribution from Ostrowski and Paulsen.

**CHAPTER 4:**

Work was conceptualized by Focardi, Paulsen and Ostrowski. Samples were collected by Focardi with help from colleagues on board the RV investigator. Chlorophyll extraction was performed by Peta Vine. DNA extraction was performed by Focardi. Amplification and DNA sequencing was performed at the Ramaciotti centre for genomics. Sequences assembly was performed by Focardi. All data analysis was performed by Focardi. Manuscript was written by Focardi with contribution from Ostrowski and Paulsen.

## Abstract

Marine microbes (viruses, bacteria and eukaryotes) produce and consume a major portion of organic matter in the ocean. Their metabolic activity contributes to a complex carbon cycle that sustains life in the Ocean. Advances in high-throughput molecular techniques have shed light on the enormous genetic diversity and metabolic potential of microbes that underpin ocean productivity. However, our ability to translate this knowledge into a quantitative understanding of how marine communities function, and how they will respond to environmental change is limited.

This thesis focusses on the integration of quantitative methods with molecular techniques to elucidate the mechanisms controlling the transfer of energy from primary producers to others trophic levels. This work was carried out in the context of two oceanographic voyages encompassing contrasting ocean regions, the East Australian Current (EAC) System, and the Dalton polynya on the Antarctic coast. The physical effects of changing climatic conditions are well documented in both regions, although the ecological consequences are poorly understood.

This study provides the first systematic description of microbial communities in the EAC system and Dalton polynya. Changes in microbial community structure across environmental gradients were determined using high resolution flow cytometry. The contribution of functional groups to the elemental Carbon (C) budget was experimentally determined. Phytoplankton rates of growth and pathways of C and energy transfer were quantified using predator-prey interaction experiments. These studies highlight the microbial contribution to the C budget in each system, with cyanobacteria representing major contributors (75%) to the photosynthetic biomass in the EAC. In contrast heterotrophic bacteria likely play a more prominent role as recyclers in the Antarctic. Viruses were an important source of mortality in the EAC, and likely represent an important driver of carbon flux in the EAC region.

Quantitative analyses were combined with metagenomics studies to also investigate the previously uncharted DNA virus diversity within the EAC system, and to provide benchmark data on the diversity of bacterial and eukaryotic community that inhabit the remote area of the Dalton polynya.

This work integrates comprehensive datasets that provide a baseline on the abundance, genetic diversity of microbial groups and infection rates to define the forces structuring microbial community in two contrasting hotspots for climate change. And provide a baseline to predict the future of production in these two regional ecosystems.

.

## Abbreviations

| | |
|---|---|
| AABW | Antarctic bottom water |
| AMGs | auxiliary metabolic genes |
| ANCOVA | analysis of covariance |
| ANOVA | analysis of variance |
| ASVs | amplicon sequence variants |
| ATP | adenosine triphosphate |
| BCP | biological carbon pump |
| BP | bathypelagic |
| BSP | below the surface polynya |
| C/P | carbon -phosphorus |
| Chl | chlorophyll |
| ChlM | chlorophyll maximum |
| ChlM_O | chorophyll maximum offshore |
| ChlM_P | chorophyll maximum polynya |
| $CO_2$ | carbon dioxide |
| CTD | conductivity - temperature - depth |
| DCM | deep chlorophyll maximum |
| dia. | diameter |
| DMSP | dimethylsulfoniopropionate |
| DNA | deoxyribonucleic acid |
| DOC | dissolved organic carbon |
| DOM | dissolved organic matter |
| DSW | dense shelf water |
| EAC | East Australian current |
| EDTA | ethylenediaminetetraacetic acid |
| FCM | flow cytometry |
| fgC | femtogram of carbon |
| FSC | forward scatter |
| G | mortality due to grazing |
| HDNA | high deoxyribonucleic acid |
| HGT | horizontal gene transfer |
| HL | high light |
| *hli* | high light inducible |
| HNLC | high nutrient low chlorophyll |
| hsp | heat shock protein |
| IPCC | Intergovernmental panel for climate change |
| Kbp | kilobase pair |
| L | litre |
| LDNA | low deoxyribonucleic acid |
| LL | low light |

| | |
|---|---|
| MCP | microbial carbon pump |
| µg | microgram |
| mgC | microgram of carbon |
| ml | millilitre |
| µl | microlitre |
| MLD | Mixed layer depth |
| MODIS | Moderate resolution imaging spectroradiometer |
| MP | mesopelagic |
| mW | milliWatt |
| NCBI | national centre for biotechnology information |
| NEuk | nano eukaryotes |
| nm | nanometer |
| NOx | nitrite/nitrate |
| nr | non-redundant |
| nt | nucleotide |
| ODV | Ocean data view |
| OMZ | oxygen minimum zone |
| ORF | open reading frame |
| OTUs | operational taxonomic units |
| PCA | principal component analysis |
| PE-Euk | phycoerythrin rich eukaryotes |
| pfam | protein family |
| $PO_4$ | phosphate |
| POC | particulate organic carbon |
| POM | particulate organic matter |
| PPE | photosynthetic picoeukaryotes |
| ppGpp | Guanosine pentaphosphate |
| PPP | pentose phosphate pathway |
| Pro | *Prochlorococcus* |
| Qc | carbon quota |
| RDOC | recalcitrant dissolved organic matter |
| rRNA | ribosomal ribonucleic acid |
| RV | research vessel |
| SAMW | subantarctic mode water |
| SBPs | substrate binding proteins |
| SIZ | seasonal ice zone |
| SO | surface offshore |
| SO_BS | surface offshore - below the surface |
| SP | surface polynya |
| Sqrt | square root |
| SSC | side scatter |

| | |
|---|---|
| SST | surface seawater temperature |
| Syn | *Synechococcus* |
| TE | tris EDTA buffer |
| tr_M | transect middle |
| tr_N | transect north |
| TS | Tasman Sea |
| µm | micrometre |
| V | mortality due to viral lysis |
| VPF | viral protein family |
| VPR | viral to prokaryotes ratio |
| WBC | western boundary current |
| µ | instantaneous growth rate |
| µ net | net growth rate |

# Chapter 1:

# Introduction

# INTRODUCTION

## 1.1 The role of the Ocean in the carbon cycle

The marine environment covers roughly 70% of the Earth surface and has a crucial role in the global carbon cycle, it stores 50 times the amount of carbon currently in the atmosphere, and the anthropogenically driven increase in $CO_2$ over the last century has escalated the uptake of $CO_2$ to the Ocean interior (IPCC). During the decade 2004-2013, the Ocean has contributed to the absorption of 2.6 billions tonnes of carbon per year (Bopp *et al.*, 2015), since the industrial era the Ocean helped to reduce roughly 25% of global $CO_2$ emissions (Le Quéré *et al.*, 2018), and thus plays a primary role in Earth habitability.

The major contribution of the Ocean to the reduction of the atmospheric concentration of $CO_2$ has not come without a price, and it has resulted in a constant increase in ocean acidification (Raven *et al.*, 2005). Moreover, the increased concentration of $CO_2$ in the atmosphere is responsible for a global increase in atmospheric temperature, that in return leads to an increase in the sea surface temperature (SST) and ocean stratification. In the marine environment two main mechanisms help in mitigating the accumulation of $CO_2$ in the atmosphere: the physical and the biological carbon pumps (Hülse *et al.*, 2017): I) The physical carbon pump, via the ocean's overturning circulation, is responsible for the dissolution of $CO_2$ and its downward propagation within the cold and dense water masses (Falkowski *et al.*, 2000), and II) the biological carbon pump (BCP) refers to all the biological processes and interactions that are responsible for the transport of carbon from the photic to the aphotic layer of the Ocean, against a chemical gradient (Falkowski *et al.*, 2011) (Fig. 1.2).

Understanding the resilience of the biological carbon pump in light of future climate change, against an increase in $CO_2$ concentration and consequently SST, is a central problem in order to be able to predict the future role of the Ocean in mitigating the atmospheric concentration of $CO_2$ (Passow and Carlson, 2012).

### *1.1.1 The biological carbon pump*

Net primary production is a critical factor in determining the amount of carbon that gets sequestrated from the euphotic zone to the aphotic one; as a consequence primary producers govern how the marine carbon pump helps regulate the atmospheric concentration of $CO_2$ (Falkowski, Laws, Barber and J. W. J. Murray, 2003). Other important components of the biological carbon pump are phytoplankton consumers, viruses and heterotrophic bacteria.

Phytoplanktonic organisms, via photosynthesis, are responsible for the conversion of inorganic carbon to an organic form. Despite their small size, they account for half of the global net primary production (Falkowski, Laws, Barber and J. W. Murray, 2003; Behrenfeld *et al.*, 2006, 2016). Once the carbon is fixed by phytoplankton, it can have different fates. It can enter what is termed the grazing food chain, passing upwards through multiple trophic levels towards the top of the marine food chain (e.g. fish, dolphins, whales), starting with the ingestion of primary producers by heterotrophic zooplankton. Particulate Organic Carbon (POC) processed in this way by the consumers, may be exported from the photic zone through the excretion of fecal pellets, organic aggregates, death or the vertical migration of zooplankton. Via an alternative route a small portion of the organic carbon (OC) reaches the depths the of the ocean, indeed, of the total amount of carbon that is assimilated every year by phytoplankton current estimates state that a minor fraction (~5-10%) is exported to the deep ocean (Henson *et al.*, 2011; Giering *et al.*, 2014). Another import route for the carbon flux is through the viral shunt. Once viruses infect primary producers, they usually burst open the cell to release the new viral progenies, contributing to the release of both POC and Dissolved Organic Carbon (DOC) (Fuhrman, 1999; Suttle, 2007; Danovaro *et al.*, 2011) While POC can aggregate and sink toward the aphotic zone, where it could eventually get stored, DOC can have multiple fates. DOC can enter the microbial loop where it can be (re-)utilised by planktonic organisms (e.g. heterotrophic bacteria). Thanks to the presence of specific transporters, bacteria are indeed capable of the uptake of DOM and DOC to use as an energy source. Within the microbial loop organic matter (OM) taken up by bacteria or

archaea, can find its way up again into the trophic chain, trough the ingestion of microorganism by phagotrophic protists. There is also a more recalcitrant fraction of the DOC (RDOC) that cannot be readily remineralized by the local microbiota and can be exported to deep Ocean. RDOC can have an important role in the storage of $CO_2$ , because its rate of turnover is very slow and can persist in the ocean for thousands of years. The term microbial carbon pump (MCP) has been introduced by Jiao and collegues (Jiao *et al.*, 2010), where they proposed that microorganism (bacteria, archaea and viruses) are at the basis for the production of the RDOC (Jiao *et al.*, 2010; Jiao and Zheng, 2011). The efficiency of DOC export from the surface of the ocean varies on a regional scale, but usually peaks in the oligotrophic subtropical Ocean (Roshan and DeVries, 2017), where it can constitute up to the 50% of carbon exported.

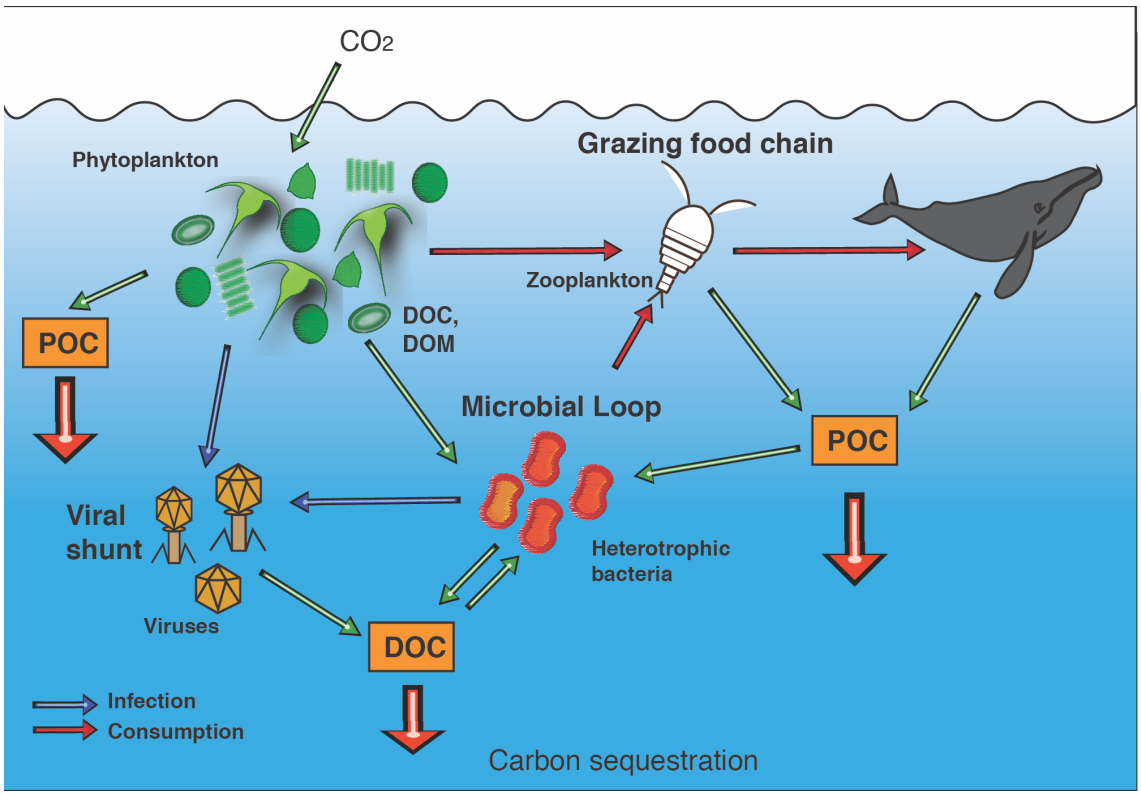



**Figure 1.1** *Schematic representation of the biological carbon pump, the main routes of carbon sequestration in the Ocean and the primary interaction between organisms (lysis and grazing) that causes both a release of DOC and POC.*

## 1.2 Picophytoplankton and their primary role in the carbon pump in a warming environment

The efficiency of the carbon pump, seen as the amount of primary production that is exported from the euphotic zone and effectively sequestered in the deep sea, depends on a variety of factors, from community structure to microbial interaction, grazing and viral infection, but primarily BCP efficiency is correlated with the diversity of primary producers and their relative elemental composition (Boyd and Newton, 1995; Guidi *et al.*, 2009), that in turn are influenced by the condition of the Ocean. Roughly 80% of the total marine primary production occurs in open ocean, where marine picophytoplankton (< 3.0 μm in dia.), including different taxa of picoeukaryotic organisms and the oxyphototrophic picocyanobacteria *Synechococcus* and *Prochlorococcus* usually, represent the main primary producers, especially for the warm and nutrient depleted subtropical gyres. (Zeidner *et al.*, 2003; Worden, Nolan and Palenik, 2004; Grob *et al.*, 2007, Jardillier *et al.*, 2010; 2013; Worden *et al.*, 2015; Guidi *et al.*, 2016). In contrast, primary production in the nutrient rich waters of the Southern Ocean and Antarctic is mainly dominated by diatoms (Tréguer *et al.*, 2017) (Fig. 1.2).

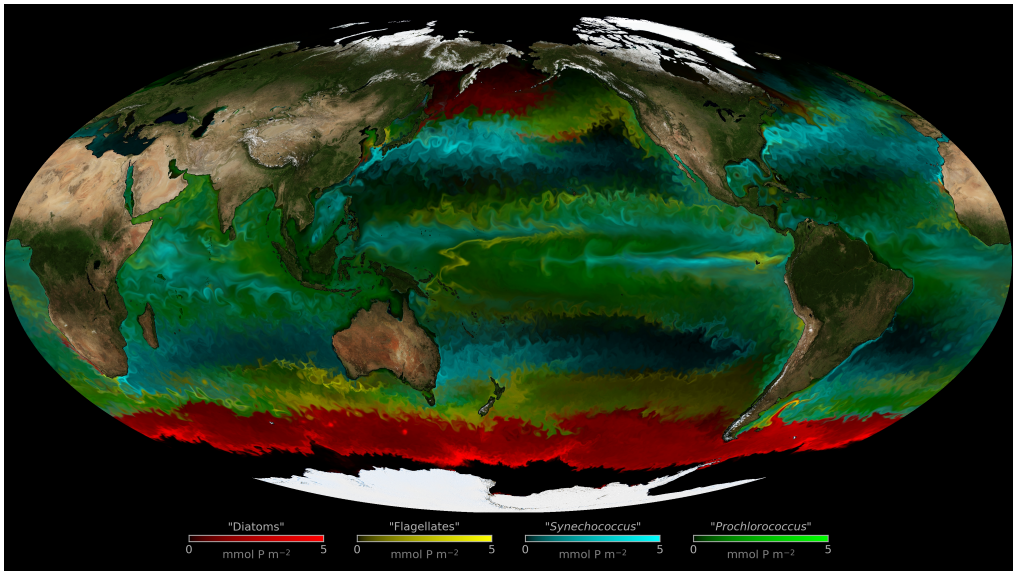



**Figure 1.2** *Visual representation of the distribution of the dominant phytoplankton primary producers built upon a high-resolution ocean circulation model. Image is a reprint from a model produced by the MIT Darwin Project, ECCO2, MITgcm.*

Despite their extremely high numbers in the ocean, picophytoplankton have been overlooked for years in the carbon export analysis. Due to their small size (<10 μm dia.), low Reynolds number and slow sinking rate, it was believed that these organisms contributed much less to the carbon export, in comparison to larger phytoplankton such as Diatoms (Tréguer *et al.*, 2017). However, recent findings have prompted a re-evaluation of the picoplankton contribution to carbon export and highlighted their importance in the export of DOC (Roshan and DeVries, 2017; Zhao *et al.*, 2017) especially in oligotrophic areas, where pico-sized cells dominate the phytoplankton communities. In a re-evaluation of the *Tara* Ocean expedition dataset, where genomics data were correlated with carbon export to detect significant association and try to pin point the main groups contributing to the carbon flux (Guidi *et al.*, 2016), on the global scale *Synechococcus* and its phages were found to be within the most important group. Picoeukaryotes has also been found to be an important player in the carbon biomass in the South Pacific gyre (Grob *et al.*, 2007). With a predicted increase of SST over the next century, the impact of picophytoplankton in the carbon cycle is predicted to further increase. Hence, it has become of critical importance to understand the distribution of picophytoplankton, protist and viruses and their role in the global carbon cycle.

### *1.2.1 Picocyanobacteria, diversity and niche partitioning.*

The marine picocyanobacteria, *Synechococcus* (Waterbury *et al.*, 1979) and *Prochlorococcus* (Chisholm *et al.*, 1988; Partensky, Hess and Vaulot, 1999), are the most abundant phototrophs in the marine environment. In the warm and oligotrophic regions of the Ocean, they can contribute up to 25% of carbon fixation and oxygen production (Jardillier *et al.*, 2010; Flombaum *et al.*, 2013).

Those two picocyanobacteria occupy partially overlapping niches (Zwirglmaier *et al.*, 2008). *Prochlorococcus* thrives better in warm and nutrient depleted areas, where it can reach cellular concentrations up to $10^5$ cells per ml. In contrast, *Synechococcus* presents a widespread distribution in surface marine habitats, from the equator to higher latitudes with a defined boundary in the vicinity of the

polar front (~60˚ S) in the cold Antarctic Ocean (Flombaum *et al.*, 2013; Biller *et al.*, 2015) (Fig. 1.3).

*Synechococcus* and *Prochlorococcus* share a common ancestor and although they are different in size, one clear difference is their light harvesting methods. *Synechococcus,* similar to other cyanobacteria, presents a membrane extrinsic light harvesting antenna, the phycobilisome. The phycobilisome core comprises a cohort of phycobiliproteins (phycoerytrin, phycocyanin, etc.) binding to different chromophores, as phycoerythribilin and phycourobilin (Scanlan *et al.*, 2009; Grébert *et al.*, 2018).

The main light harvesting apparatus for *Prochlorococcus* instead, is prochlorophyte chlorophyll binding protein, that binds a divinyl $Chl_a$ and $Chl_b$ ($Chl_{a2}$, $Chl_{b2}$), adapted to absorb specific wavelength of the blue light (460 nm-480 nm) that are prevalent especially at depth, and lack phycobilisome (Chisholm *et al.*, 1988; Partensky, Hess and Vaulot, 1999; Biller *et al.*, 2015). Chlorophyll-based light harvesting complex of *Prochlorococcus* are built within the membrane, allowing the thylakoids to be highly stacked, providing the capacity for *Prochlorococcus* to maximise light-harvesting capacity without decreasing its cell surface to volume ratio (Partensky, Hess and Vaulot, 1999).

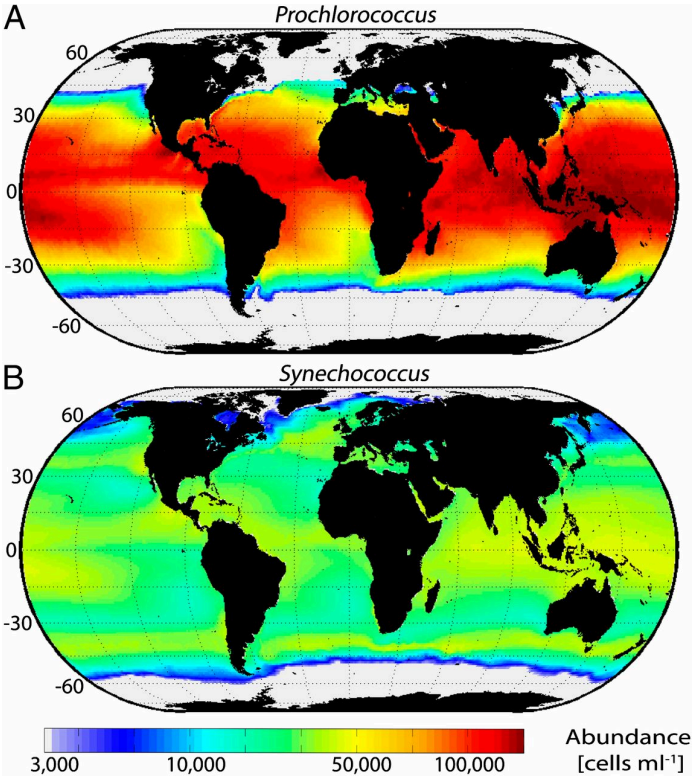



**Figure 1.3** *Mean estimated annual global distribution of marine Prochlorococcus and Synechococcus. The image is reprinted from Flombaum et al. (2013) PNAS (copyright 2013, National Academy of Sciences, USA).*

Numerous phylogenetic and genomic studies on laboratory isolates and environmental samples have shed light on the global biogeography and depth partitioning of the several lineages of the genera *Prochlorococcus* and *Synechococcus* (Fig. 1.4) (Palenik *et al.*, 2006; Zwirglmaier *et al.*, 2008; Scanlan *et al.*, 2009; Mazard *et al.*, 2012; Farrant *et al.*, 2016). The recognized genetic diversity between picocyanobacterial lineages (sometimes referred to as ecotypes) underpins specific adaptations to different niches. The different picocyanobacteria lineages display a specific genetic repertoire, such as the presence of transporters, SBPs (substrate binding proteins) or *hli* (high light inducible) genes, that allows them to optimally adapt to particular environments. The partitioning of ecotypes is believed to be mostly controlled by environmental variability (Farrant *et al.*, 2016), however, the top-down control, of grazers and

lytic viruses , is also recognised but few studies exist that quantify their impact on the genetic structure of cyanobacterial communities.

*Prochlorococcus* presents a clear depth partitioning with the high light adapted (HL) clade dominating the surface layer and the low light adapted (LL) predominant at lower depth (<50m) (Moore, Rocap and Chisholm, 2002). Phylogenetic analysis revealed how those two broad groups resolved into a number of clades , at least 6 for both HL and LL adapted. Some clade shows a latitudinal partitioning based on a specific temperature gradient (e.g. HLI versus HLII) (Zackary I. Johnson *et al.*, 2006). While recent studies have also underlined a specific adaptation related to nutrient bioavailability. For example, LL-II/III display a positive ecological correlation with nitrite concentration (Berube *et al.*, 2019), and *Prochlorococcus* HLIII-HLIV are associated with iron limited, high nutrient low-chlorophyll (HNLC) waters (West *et al.*, 2011).

The biogeography of different *Synechococcus* clades shows a clear latitudinal partitioning, while a depth partitioning is not really evident (Ferris and Palenik, 1998; Zwirglmaier *et al.*, 2008; Huang *et al.*, 2012) with the exception of cases where the sub-surface watermass is moving in a different direction. Water temperature is recognised as one of the main drivers of *Synechococcus* partitioning, with clade I and IV usually found in cold and temperate water, while clade II is adapted to tropical water (Zwirglmaier *et al.*, 2008; Huang *et al.*, 2012; Mazard *et al.*, 2012). Clade III thrives in oligotrophic water where the low nutrient availability is thought to be a limiting factor for primary production. Under the current classification *Synechococcus* presents 17 main clades that have in turn multiple subclades, for example a recent analysis of the *Synechococcus* lineages based on the ecological partitioning of *petB* sequences 8 subclades belonging to clade II have been proposed (Farrant *et al.*, 2016) (Fig. 1.4).

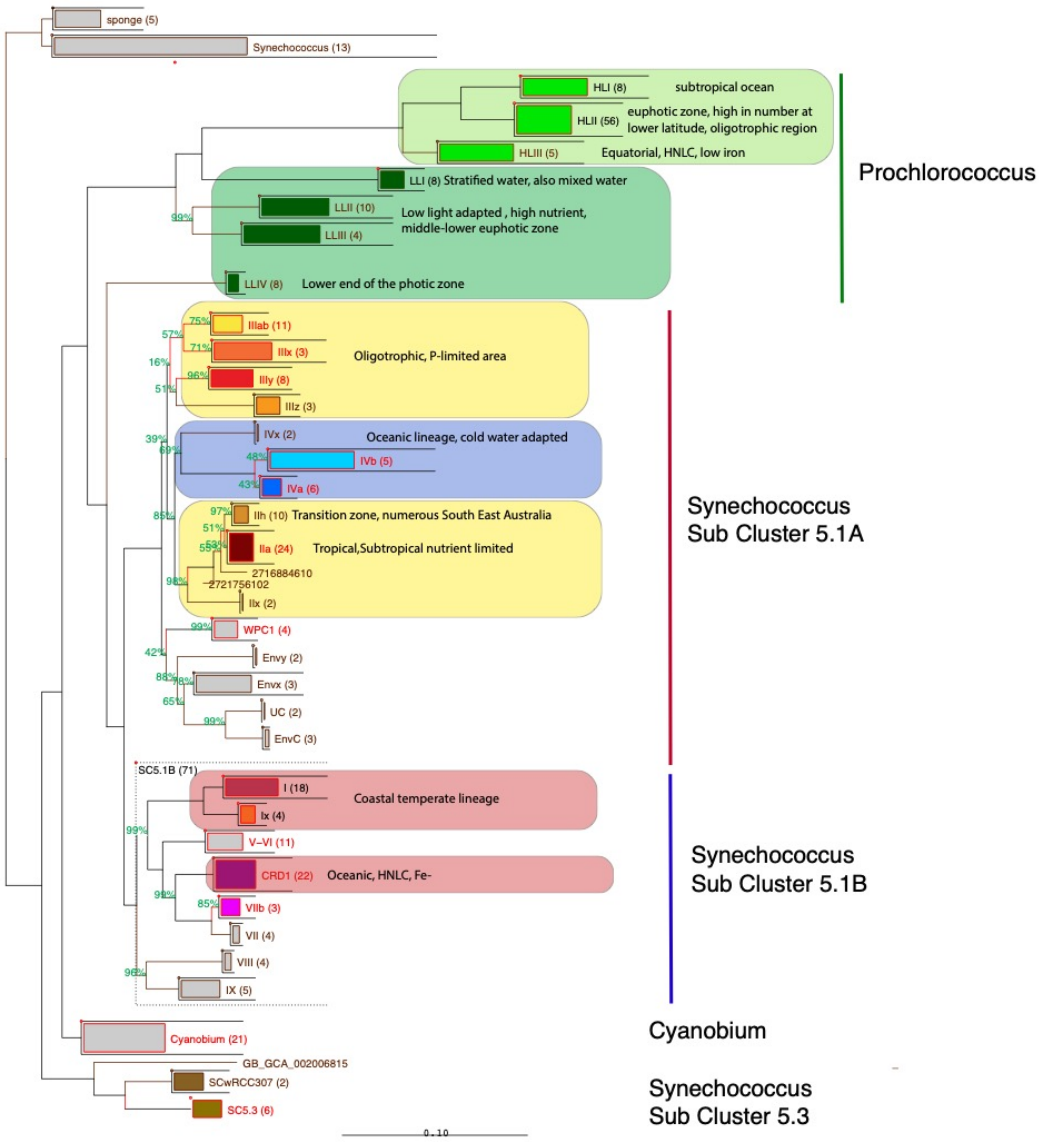



**Figure 1.4** *Picocyanobacteria Genome tree compiled including almost 300 complete, or near complete cyanobacteria genomes available in Genbank, 25 pre-release marine genomes courtesy of the University of Warwick and Station Biologique de Roscoff, four near-complete genomes of local isolates, and 22 Metagenome-assembled genomes from local marine cyanobacteria (using data from the Marine Microbes Project). Alignment was performed in MAFFT and phylogenetic tree constructed in ARB using a maximum parsimony algorithm. Each leaf of the tree represents a different clade, the number in brackets represents the number of genomes included in that leaf. Coloured box groups clade that thrives in similar environment. The tree was compiled with the Genome Tree Database.*

The accumulated wisdom in a growing body of ecological studies, as outlined above, is that environmental selection (temperature, light, nutrients) plays an important role in the spatial partitioning of genetically and phenotypically distinct cyanobacteria. However, there is a growing appreciation for the role of biotic factors, such as predation by grazers and cyanophage in shaping natural populations. It is also important to note that cyanobacterial genetic diversity is influenced by phage-mediated HGT (horizontal gene transfer) (Lindell *et al.*, 2007). The vast body of knowledge in relation to marine cyanobacteria is centred on the Northern Hemisphere and there is still a paucity of information about their abundance, diversity and importance in the carbon cycle for the South Pacific.

## 1.3 Bacteriophage as hidden players of biogeochemical cycle.

Viruses represent the most abundant entities in the Ocean, with a concentration that can surpass $10^7$ particles per ml of seawater (Wommack and Colwell, 2000; Suttle, 2005). They vary in size, usually between 20-400nm, even though the latest discovered giant viruses can be up to 500nm in dia. and 1000nm in length, and are thought to be able to infect every domain of life (Wommack and Colwell, 2000). Since bacterial cells outnumber all of the other domains, bacteriophage (viruses that infect bacteria) are expected to be the predominant viruses in the marine environment.

Bacteriophage have an integral role in the marine ecosystem by infecting, reprogramming and killing bacteria that underpin ocean biogeochemical cycles (Fuhrman, 1999; Suttle, 2005, 2007; Danovaro *et al.*, 2011; Breitbart *et al.*, 2018), but also more specifically as both a driver-of evolution and vector for genetic material to expand the host genetic diversity, as demonstrated by the transfer of *hli* genes from *Prochlorococcus* to phage and back to *Prochlorococcus* again (Lindell *et al.*, 2004). The role of phage in biogeochemical cycles is mostly related to their lytic cycle of infection. During infection the virus hijacks the host metabolic machinery to redirect energy and biosynthetic output towards the replication and assembly of the new phage progeny. Once the virus particles are assembled the new progeny are released via the lysis of the host cell. However lytic infection is

not the only lifestyle of marine bacteriophage. Temperate phage have the ability to integrate their DNA into the genome of their host and be maintained as prophage until an external stress or a change into the environmental condition, in which the host live, provides a cue for the prophage to switch to a lytic style.

Upon lysis of the host and release of mature virions, new dissolved organic matter is released back into the pool of dissolved nutrients in ocean, resupplying potentially limiting nutrient (phosphate, nitrogen, iron, etc.) (Jover *et al.*, 2014), yet also decreasing the amount of energy and carbon available for higher trophic levels.

Bacteriophage that inhabit “extreme” environments or habitats where the paucity of nutrients does not promote an active bacteria metabolism, such as the bottom of the ocean or deep sediment are thought to be responsible for up to 100% of bacterial mortality (Danovaro *et al.*, 2005, 2008, 2011), while in the euphotic zone viral lysis has been estimated to kill between 5-100% of the population each day (Fuhrman, 1999; Evans *et al.*, 2003; Suttle, 2007; Weinbauer *et al.*, 2007; Tijdens *et al.*, 2008; Pasulka, Samo and Landry, 2015; Mojica *et al.*, 2016)

Bacteriophage also have an effect on the diversity and functionality of microbial communities through the expression of Auxiliary Metabolic Genes (AMGs). These are phage encoded metabolic genes that confer and advantage for phage production during the infection cycle. For example, genes that encode for various photosynthesis-related proteins were found in cyanophage that infect *Synechococcus* and *Prochlorococcus* (Mann *et al.*, 2003; Lindell *et al.*, 2004; Puxty *et al.*, 2015). Other AMGs are involved in a variety of metabolic functions including carbon metabolism, nitrogen metabolism, phosphate scavenging, nucleotide metabolism and stress responses (Lindell *et al.*, 2004, 2007; Dammeyer *et al.*, 2008; Millard *et al.*, 2009; Sullivan *et al.*, 2010; Puxty *et al.*, 2015, 2018; Hurwitz and U’Ren, 2016).

To date (January 2019) there are just 5,633 complete bacteriophage genomes in the RefSeq database, of which just 926 are from environmental samples. The majority of those bacteriophage were isolated thanks to the presence of their

specific host. In the marine environment the majority of the known or characterised complete bacteriophage genomes belong the order of Caudovirales (tailed bacteriophage), to which belong all the known phage thought to infect picocyanobacteria (fig. 1.5).

There is only very limited information about the diversity of bacteriophage in the marine environment (Brum *et al.*, 2015, 2016; Hurwitz, Brum and Sullivan, 2015; Roux *et al.*, 2016; López-Pérez *et al.*, 2017; Luo *et al.*, 2017; Needham, Sachdeva and Fuhrman, 2017; Breitbart *et al.*, 2018). While bacteriophage distribution and taxonomic diversity are thought to be regulated by host density (Roux *et al.*, 2016), there is still not a clear understanding of the distribution of AMGs in phage genomes adapted to infect hosts with different environmental preferences. The paucity of data biases our ability to capture a global biogeographical perspective of bacteriophage distribution and especially about their role in global biogeochemical cycles thorough host mortality and expression of different AMGs.

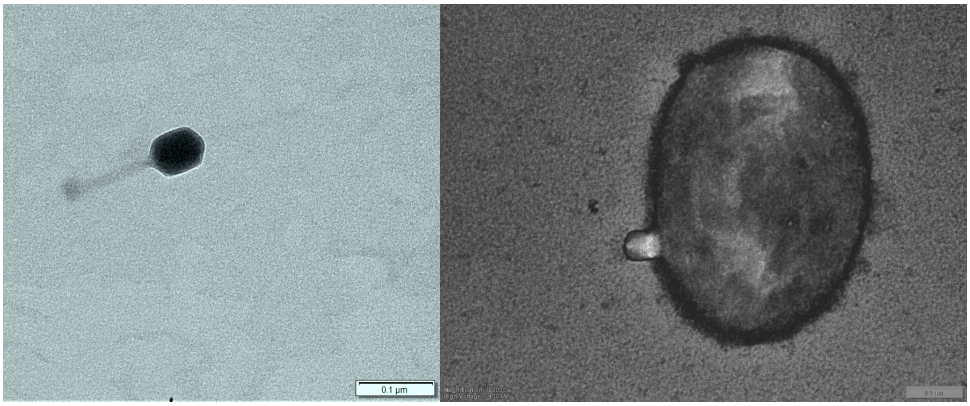



**Figure 1.5** *(a) TEM micrograph of a tailed bacteriophage sampled from along the EAC (East Australian Current) during an oceanographic expedition on board RV Investigator September 2016. (b) Micrograph of a tailed bacteriophage infecting a local isolate of Synechococcus. The scale is presented on the right corner of the micrometre (0.1 µm), 92000 V, high voltage: 100 kV*

## 1.4 High-throughput techniques to characterise microbial communities

### *1.4.1 The use of 'omics techniques to taxonomically and functionally characterise microbial communities.*

Understanding which organisms populate a specific environment, what mechanisms they use to adapt, which biochemical transformations they undertake and what other species they interact with have always been within the most critical questions in the field of microbial ecology. In the last two decades high-throughput sequencing have circumvented the limitations of culturability of environmental microbes, and in combination with advances in bioinformatic analysis have expanded our knowledge of microbial diversity and reshaped the tree of life with the addition of previously uncharacterised microbial lineages (Hug *et al.*, 2016).

Global-scale marine metagenomic efforts, such as the Sargasso Sea metagenome, the Sorcerer global ocean survey (GOS), *Tara* Oceans expedition, OSD (Ocean sampling day) and the establishment of sustained ocean-time series sites near Hawaii ( HOTS, Hawaii Ocean Time Series), San Pedro Ocean (SPOTS, San Pedro Ocean Time Series), Bermuda (BATS, Bermuda Atlantic Time Series) and the Australian National reference station ( Marine Microbes) (Venter *et al.*, 2004; DeLong, 2005; DeLong *et al.*, 2006; Rusch *et al.*, 2007; Williamson *et al.*, 2008; Karl and Church, 2014; Sunagawa *et al.*, 2015; Brum *et al.*, 2016; Roux *et al.*, 2016; Brown *et al.*, 2018) have contributed towards an expanding catalogue of the diversity and functional potential of marine communities both at the surface and throughout the water column. The deep sequencing of highly conserved markers genes, such as bacterial 16S rRNA and eukaryote 18S rRNA, have helped to refine the ecological distribution of specific bacterial and eukaryotic lineages, as well as define distinct patterns in the assembly and composition of microbial communities across oceanic provinces (Bartram *et al.*, 2011; Sunagawa *et al.*, 2015; Farrant *et al.*, 2016; Brown *et al.*, 2018; Tragin and Vaulot, 2018). From a functional perspective they have provided an insight into the importance of microbial communities for biogeochemical cycles.

In contrast to bacteria, eukaryotes and archaea, viruses lack the presence of a universally conserved marker gene that can be utilized to study their taxonomic diversity (Edwards and Rohwer, 2005; Brum and Sullivan, 2015). Metagenomics and metaproteomics have circumvented the need for a marker gene to study viral diversity (Breitbart *et al.*, 2002). Viral metagenomics (viromics), has been a stepping stone in the previously uncharted diversity of viruses in the marine environment. Moreover, untargeted whole genome sequencing approaches have helped unveil the functional potential and metabolic versatility of microbial and viral species; driving the discovery of abundant and novel proteins in the marine environment, highlighting genes that are important for species adaptation to different environments and identifying critical genes in the different biogeochemical cycles (Tyson *et al.*, 2004; DeLong, 2005; Temperton and Giovannoni, 2012; Swan *et al.*, 2013; Hurwitz, Brum and Sullivan, 2015; Brum *et al.*, 2016; Guidi *et al.*, 2016; Roux *et al.*, 2016).

The newer sequencing platforms allow for enough sequencing coverage to be able to confidently reconstruct genome from metagenomic sequences (MAGs = metagenome assembled genome) (Tyson *et al.*, 2004; Albertsen *et al.*, 2013; Delmont *et al.*, 2015; Parks *et al.*, 2017; Tully, Graham and Heidelberg, 2017). The reconstruction of microbial genome directly from metagenomic data has potential to improve the classification of microbial species (Parks *et al.*, 2017) and for the discovery of new metabolic pathways and to reveal how those pathways are partitioned within the active members of a microbial community (Delmont *et al.*, 2015).

### *1.4.2 Flow cytometry and fluorescence activated cell sorting as a tool to quantitatify and characterise functional microbial groups*

One of the limitations of metagenomics lies in the fact that it provides a relative abundance of the contribution of an organism's DNA to a complex sample but no information on cellular abundance. Flow cytometry instead, created an avenue for an accurate counting of populations. Flow cytometry is a fast, scalable and accurate technique that utilized laser-based technology to count and sort cells based on specific cells properties, such as light scatter and specific fluorescence. Cells in a liquid sample are hydrodynamically aligned in a stream of single cells before passing through a series of laser beams. When the cell passes through each beam, it scatters light (Marie et al 2005), that is then detected by multiple detectors. FSC (forward scatter) detector is a photodiode situated in front of the light emitter, and it has been utilised to provide an indication of cell size (Collier, 2000). Orthogonally to this first detector, is the SSC (side scatter) detector, that could indicate the internal structure of the cell. Cytometers are also equipped with multiple optical filters to detect specific fluorescence wavelengths emitted by the cells.

Flow cytometry was firstly applied in oceanography in 1985 (Olson, Vaulot and Chisholm, 1985) and led to important discoveries in the marine environment, such as *Prochlorococcus* (Chisholm *et al.*, 1988) and *Ostreococcus* (Courties *et al.*, 1994). Thanks to the presence of fluorescence pigments, like chlorophyll, phycoerythrin and phycocyanin, picocophytoplankton are easily identified and distinguished based on their specific autofluorescence and light scatter properties (Marie *et al.*, 1997), without the need of applying any fluorescent stain to the samples (Fig 1.6).

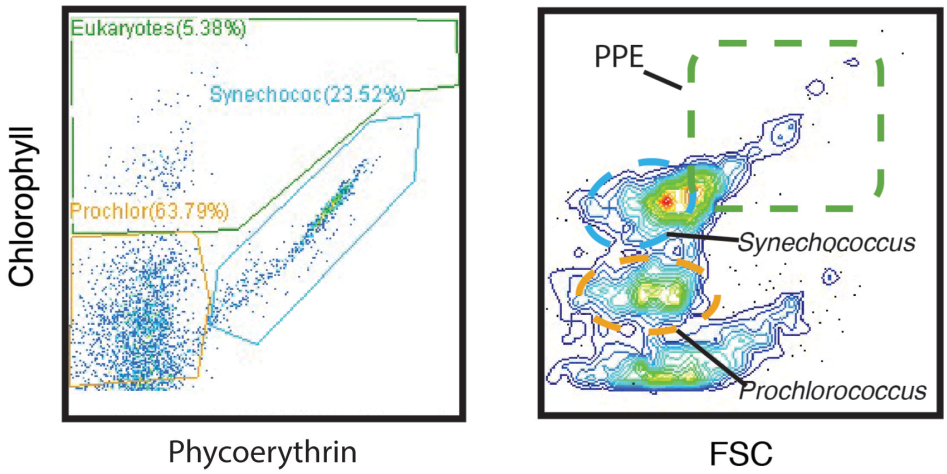



**Figure 1.6** *Cytogram of three different picophytoplankton populations from a seawater sample analysed on a Cytoflex S (Beckman Coulter) flow cytometer. PPE, Synechococcus and Prochlorococcus population were identified based on the fluorescence intensity of their specific pigment and different value of FSC.*

Flow cytometry has revolutionized the way marine samples are counted, while traditional microscopic analyses are still a reliable method in term of phytoplankton species identification, picophytoplankton, and microbes in general, are too small to be easily counted and distinguished under light microscopy. Moreover, the development of newer and more sensitive flow cytometry machines has opened the door to the possibility of not just counting picophytoplankton, but also smaller particles. Using appropriate samples preparation and staining with dyes such as SYBR Green I, it is now possible to also identify different populations of heterotrophic bacteria, and multiple virus populations (Marie *et al.*, 1999; Brussaard, 2004) (Fig 1.7).

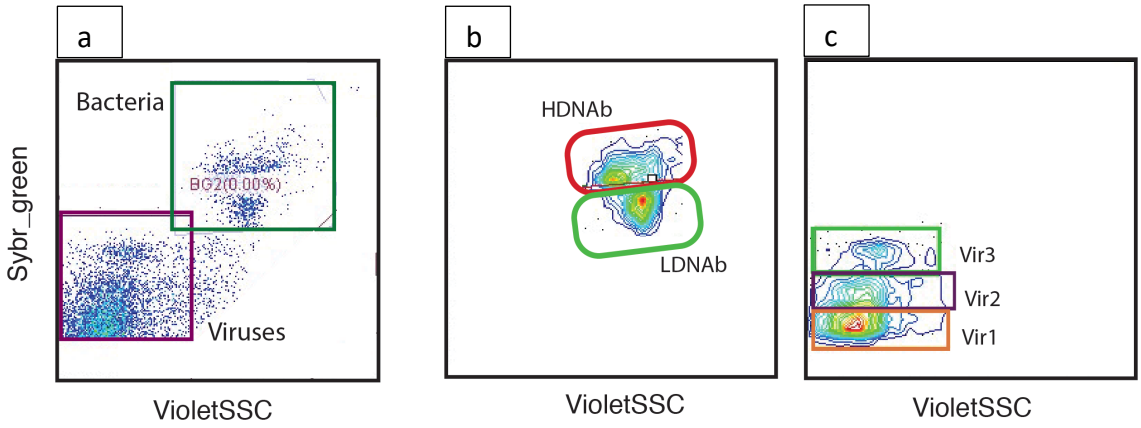



**Figure 1.7** *(a)Cytogram of bacteria (green) and viruses (purple) populations from a seawater sample analysed on a Cytoflex S (Beckman Coulter) flow cytometer after staining with SYBR Green (Invitrogen). (b) Different bacteria populations: HDNAb (high DNA bacteria), LDNAb (low DNA bacteria) identified based on different intensity of SYBR green and different SSC values. (c) Three different viruses populations (Vir1, Vir2, Vir3) identified based on SYBR green intensity against the value of violet SSC.*

In the marine environment flow sorting has been widely used for a range of diverse applications, from the isolation and culturing of novel bacterial species, e.g. sorting of *Synechococcus* or *Prochlorococcus* populations, to the sorting of *Synechococcus* culture tagged with fluorescently labeled marine viruses to genomically characterise viruses that infect *Synechococcus* (Deng *et al.*, 2014). Moreover, high-throughput sorting can be applied to actively sort different photosynthetic organism to specifically analysed the elemental composition of each population (Graff, Milligan and Behrenfeld, 2012). This type of analysis will help to understand the role of picophotosynthetic communities in the carbon cycle both at a regional and global perspective.

## 1.5 Biotic and abiotic factors and their effect on community composition and carbon flux.

The efficiency of the carbon pump is a function of the microbial community composition and its biotic interactions, that are regulated by the physical and chemical status of the Ocean. Temperature and nutrient concentration are within the environmental factors that mostly influence the distribution and stoichiometry of picophytoplankton (Moreno and Martiny, 2018).

The specific elemental composition of the main picophytoplankton groups have a pivotal importance for understanding the partitioning of carbon biomass. Moreover, the elemental quota determine the palatability and nutritional quality of primary producers for consumers and higher trophic levels and ultimately influence the amount of energy that is transferred through the marine food chain (Kwiatkowski *et al.*, 2018). It also represents the value that is integrated into ocean biogeochemical model to assess the future carbon dynamics and budget for the marine ecosystem.

In light of current and future climate change, it is therefore important to understand how a shift in temperature and nutrient bioavailability will affect primary producers, at the cellular level, in order to predict the broader consequences at the ecosystem level.

### *1.5.1 Grazing and viral infection and their effect on community composition and the carbon flux.*

Microbial interactions like grazing and viral infection can affect ecosystem at multiple levels: community composition, the carbon pump and eventually on the flux of all nutrients in the ocean food chain. Active feeding from larger zooplankton on primary producers controls the movement of nutrients to higher trophic levels and eventually sustain life in the Ocean. However, grazing can also have an effect on community composition. Some studies on *Synechococcus* and *Prochlorococcus* have underlined a preferential size-related feeding by phagotrophic protists. This type of preferential feeding could ultimately create a bottleneck on species distribution and succession whether specific clade or

smaller organisms have less selective pressure. Other studies on *Synechococcus* isolates, differences in preferential grazing has been attributed to specific cell surface properties (Simkovsky *et al.*, 2012) and grazer evasion strategies (Scanlan *et al.*, 2009; Apple *et al.*, 2011). Viral infection on the other end hand could have a dual effect on the efficiency of the carbon pump. It could either reduce the efficiency through the lysis of their host and the release of newly formed DOM, that will fuel the microbial loop and hence increase respiration. While some studies suggests that viruses could potentially enhance particle aggregation and as a consequence increase the efficiency of the biological carbon pump via the sinking of heavier particles (Suttle, 2007; Danovaro *et al.*, 2011).

A recent study that integrated data of carbon flux with genomic data found a positive correlation of *Synechococcus* and their phage with increase in carbon export in the oligotrophic areas (Guidi *et al.*, 2016), shifting the attention once again to the importance of small primary producers in a warming ocean.

### *1.5.2 Temperature*

Temperature plays a key role in regulating biochemical process, it can actively affect the growth and metabolisms of an organism and can also have an influence on the elemental composition of different taxa (Moreno and Martiny, 2018). Responses to temperature shifts are really species-specific. While temperature increases can positively affect metabolism, each organism has a temperature maximum that when overcome causes a sharp decline in their metabolism. *Synechococcus* and *Prochlorococcus* ecotypes have shown different distribution and response based on temperature tolerance (Huang *et al.*, 2012). For example, *Synechococcus* clade II, that presents an optimum growth between 22˚C-26˚C, but displays a sharp reduction in growth when temperature was 18˚C or lower. (Varkey *et al.*, 2016).

The predicted increase in SST will lead to prolonged periods of stratification and fewer mixing events, leading to a limitation of nutrients available for phytoplankton (Fig. 1.8) and expanding the oligotrophic area of the Ocean. This magnifies the importance of picophytoplankton, their viruses and the microbial

loop in relation to the carbon flux in a warming environment and the need to fully understand the role, diversity and adaptability of microbial communities, especially picocyanobacteria and bacteriophage.

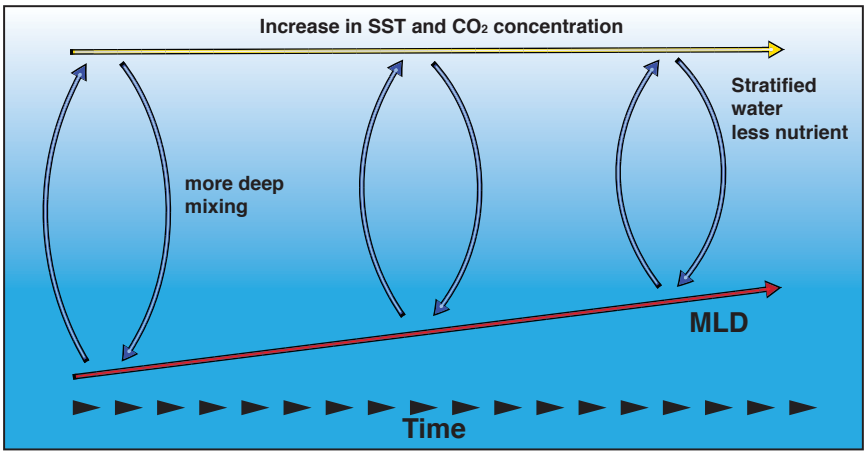



**Figure 1.8** *Effect of climate change on the stability of the water column. Predicted SST increase in the future will increase the thermal stratification decreasing the depth of the mixed layer (MLD = mixed layer depth), resulting in more stratified water with less nutrient bioavailability. Figure adapted from Basu et al., 2018*

### *1.5.3 Nutrients*

Nutrient availability represents one of the main limiting factors to phytoplankton growth and productivity in the euphotic zone. There is a strong latitudinal difference in nutrient concentration within the ocean environment. With high latitude water masses characterised by low temperature and high nutrient load , and low latitude by warm and nutrient limited water (Martiny *et al.*, 2013). Nitrogen and phosphorus are the building blocks of the majority of macromolecules, such as nucleic acid, chlorophyll and macromolecule that make up bacteria wall, and are found at different concentration and active forms in seawater.

For example, two main forms of nitrogen can be utilized by phytoplankton, the more stable $NO_3^-$ and $NH_4^+$, but in some cases also small nitrogen containing organic molecules. Nitrate is present at variable concentrations in the Ocean (fig.1.9), with polar regions displaying a much higher concentration compared to tropical regions, where the concentration can fall below 0.5 μmol/l. In tropical

regions the concentration of nitrate rises just below the euphotic zone, where nitrification is supposed to occur (Rogato *et al.*, 2015). Phosphate shows a similar pattern with high concentrations in polar regions and almost below the limit of detection in tropical regions.

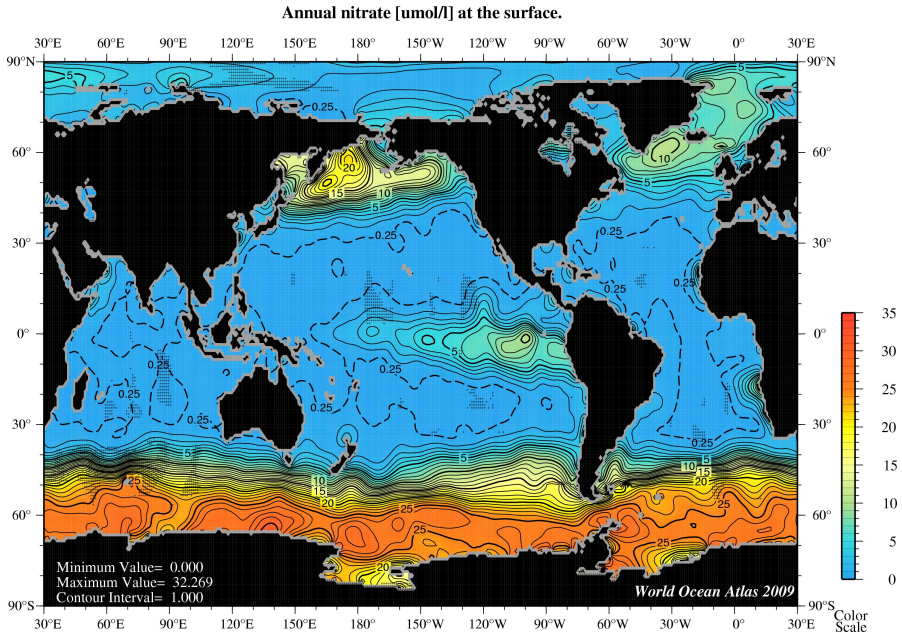



**Figure 1.9** *Composite of the annual global concentration of Nitrate at surface. Data was obtained by the World Ocean Atlas data 2009*

The availability of trace metals, such as iron, zinc, copper, and manganese among others, also has a fundamental role in phytoplankton productivity. Iron, in particular, is a key nutrient, e.g., in the Southern Ocean, where there is no shortage of phosphate and nitrate, iron bioavailability is one of the primary drivers of phytoplankton productivity (Boyd, 2002). Iron is an essential cofactor in macromolecules of photosynthetic organisms, such as chlorophyll, but also electron transfer proteins, such as Fe-S clusters (Watson et al., 2000). Hydrothermal vent and wind-driven dust from the continent are thought to be within the main supplier of iron in the Ocean. However, recent studies have also identified iceberg and melting glacier as a possible local source of iron, and that would explain the higher productivity associated with these areas (Alderkamp *et al.*, 2012, Wadley, Jickells and Heywood, 2014).

Photosynthetic organisms have evolved different strategies to cope with low nutrient concentrations. Picocyanobacteria, for example, can vary their elemental composition in response to a limited bioavailability of nitrogen or phosphorus (Mouginot *et al.*, 2015; Martiny *et al.*, 2016). Another adaptive strategy is the presence of substrate binding proteins or high affinity transporters that are present in some lineages, where the presence of that specific transporter could confer a fitness benefit. For example, HL-adapted *Prochlorococcus* that thrives in the upper euphotic zone lack a nitrate transporter. Instead, genes encoding nitrate transporters are sporadically found in LL-III ecotypes, that are found usually below 50 m depth, where the concentration nitrate increase (Berube *et al.*, 2019). In contrast, *Synechococcus* presents a extensive genetic repertoire for the assimilation of both nitrate and phosphate (Scanlan *et al.*, 2009). Larger phytoplankton, such as diatoms, that usually have higher constraints on nutrient availability, have been found to have up to 8 different transporters for both ammonia and nitrate (Rogato *et al.*, 2015).

Apart from the polar region, the majority of vast oceanographic areas are nutrient-poor oligotrophic waters. *Prochlorococcus* is known to thrive in oligotrophic regions of the ocean where nutrients can be a limiting factor for other primary producers. In contrast, its concentration drops almost below detection in eutrophic regions of the ocean (Partensky, Hess and Vaulot, 1999). *Synechococcus* instead can be found in a wide variety of nutrient concentrations, where its ecotypes distribution and their difference in genomic content well correlate with nutrient bioavailability (Palenik *et al.*, 2006; Zwirglmaier *et al.*, 2008; Scanlan *et al.*, 2009; Farrant *et al.*, 2016).

### *1.5.4 A possible future scenario in the marine food web and carbon cycle*

The 4$^{th}$ IPCC report has predicted increase an increase in SST of almost 2˚C in the upcoming future (IPCC, 2014) this increase will increase stratification limiting the number of mixing events, leading to a lower nutrient available for

phytoplankton to grow in the photic zone and to an expansion of the oligotrophic area of the Ocean.

This change in oceanic condition could results in a shift in the composition of the main primary producers in different region, going from a dominance of larger phytoplankton to dominance of picophytoplankton with shift towards picocyanobacterial (Fig 1.10) and a greater importance of the microbial loop and viral shunt in the marine carbon cycle. Previous studies on modelling the picocyanobacterial distribution have suggested an increase in both *Synechococcus* and *Prochlorococcu*s numbers and expansion of their niche distribution (Flombaum *et al.*, 2013).

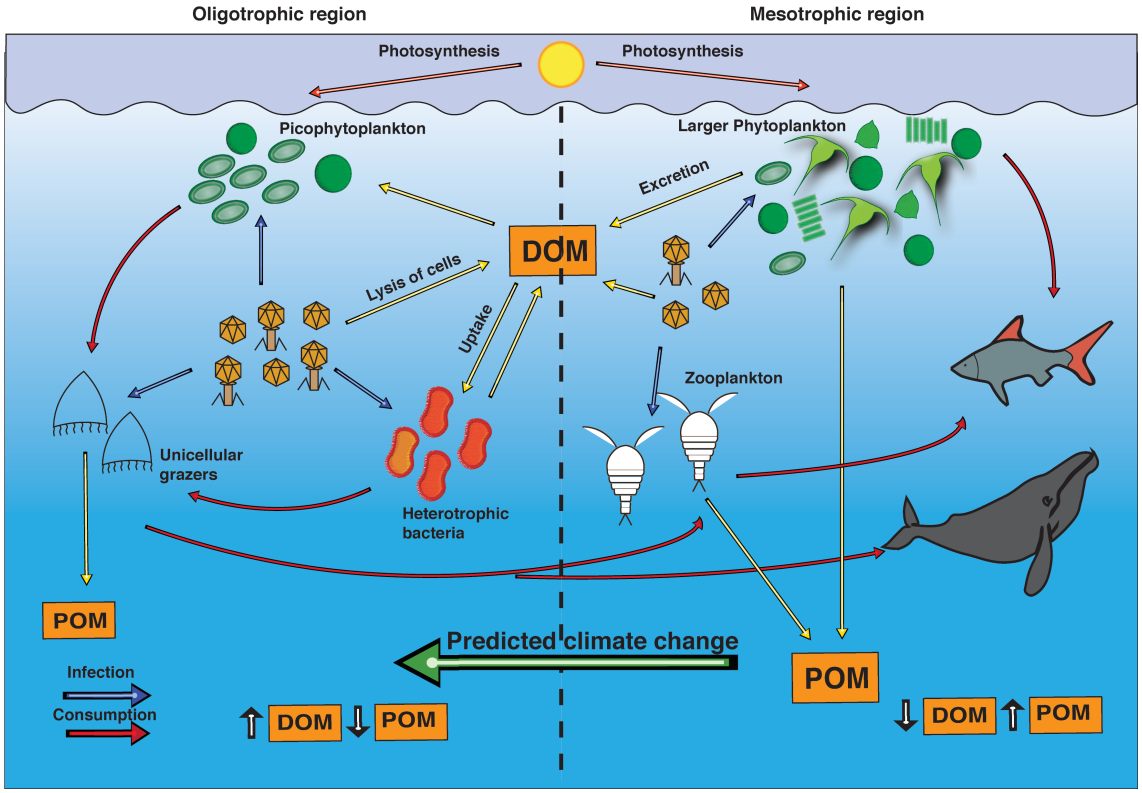



**Figure 1.10** *Schematic representation of the marine food web comparing oligotrophic and mesotrophic regions of the Ocean with implication for future climate change. Primary production is carried out by both picophytoplankton and larger phytoplankton (e.g. diatoms). Phytoplankton are favoured in mesotrophic region where nutrient are not a limiting factor, picophytoplankton are the main photosynthetic organisms in oligotrophic region. Carbon flux in oligotrophic region is controlled primarily by picophytoplankton. Predicted SST increases will lead to a higher water column stratification with fewer nutrient available for phytoplankton growth, eventually shifting the primary producers composition from a diatoms dominated one to a cyanobacteria dominated. Figure re-adapted from Breitbart 2018* (Breitbart *et al.*, 2018)

## 1.6 Rationale of the dissertation

Advances in genome sequencing and genetic techniques have led to numerous significant discoveries and detailed the expansive genetic diversity and metabolic potential of microbes and viruses. But what does all this diversity mean? And how does our understanding of this diversity translate into a better capacity to describe how marine communities function now, and especially how they are going to function in the future? Predicted climate change is going to affect the stability of the water column with a possible expansion of the oligotrophic regions of the Ocean. In order to determine the future of ocean processes that rely on microbes we need a better quantitative understanding of how diverse microbial species contribute to key processes, such as the carbon cycle that is related to the ability of the Ocean to act as a carbon sink.

This thesis focusses on the application of molecular and quantitative methods to elucidate the diversity and role of picoplankton and viruses into the context of primary production and carbon cycling. To gain insight into the possible implications of temperature rising and increased in the water column stratification this thesis focuses on two model ocean regions, the East Australian Current (EAC) System, including the adjacent Tasman Sea and coastal shelf region, and the Dalton polynya on the Antarctic coast, in relation to the Southern Ocean. These two highly dynamic regions are linked by the fact that they are ecosystems undergoing well documented physical changes and therefore represent open water laboratories to start to formulate and answer questions about the effect that climate change will have on microbial communities.

The EAC, a model Western boundary current, originates in the Coral Sea from where it transports warm and nutrient poor water masses into the Tasman Sea. The Dalton polynya is one of the most productive coastal areas of East Antarctica and remains still relatively unexplored from the perspective of microbiology.

### *1.6.1 Specific questions addressed:*

1) What is the abundance of picophytoplankton, and their contribution to the carbon biomass in the East Australian Current compared to the Tasman Sea. (Chapter 2)

2) What are the relative impacts of grazing mortality and viral lysis on picophytoplankton communities in the EAC compared to the Tasman Sea? (Chapter 2)

3) Is there any difference in the main viral communities that inhabit the EAC compared to the Tasman Sea? (Chapter 3)

4) What is the community composition and dynamics of the main picophytoplankton group in the Dalton polynya? (Chapter 4)

***Dissertation format:***

**Second chapter - The importance of picocyanobacteria and viruses for the carbon cycle in the East Australian Current.** Analysis of the specific carbon content quota for picophytoplankton and count of heterotrophic bacteria and viruses from 64 samples collected from sites within the EAC system, the Tasman Sea and 2 transect shore to offshore. To assess the impact of grazing and viral mortality on the carbon flux on this important region dilution experiment were also performed in samples from the EAC and Tasman Sea.

**Third chapter - Bacteriophage diversity associated with the EAC, a temperate hotspot for climate change** – taxonomic classification of the main viral groups retrieved from 64 metagenomic samples collected along the EAC, the Tasman Sea and two transect shore to offshore. Functional characterization of the viral fraction and analysis of specific AMGs in relation to different oceanic provinces. To better elucidate the role of bacteriophage in the biogeochemical cycle of this region.

**Fourth chapter - Microbial diversity in an Antarctic hotspot for climate change**. Molecular analysis and quantitative study of the Primary producers and bacterial diversity from 25 samples collected within the Dalton polynya area in East Antarctica. This data represents benchmark information on the microbial communities that inhabit a highly productive area of the Sabrina Coast in East Antarctica and relatively unexplored.

## 1.7 References:

Alderkamp, A. C. *et al.* (2012) 'Iron from melting glaciers fuels phytoplankton blooms in the Amundsen Sea (Southern Ocean): Phytoplankton characteristics and productivity', *Deep-Sea,* 71-76, pp. 32-48. doi: 10.1016/j.dsr2.2012.03.005

Albertsen, M. *et al.* (2013) 'Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes', *Nature Biotechnology*, 31(6), pp. 533–538. doi: 10.1038/nbt.2579.

Apple, J. K. *et al.* (2011) 'Variability in protist grazing and growth on different marine Synechococcus isolates', *Applied and Environmental Microbiology*, 77(9), pp. 3074–3084. doi: 10.1128/AEM.02241-10.

Bartram, A. K. *et al.* (2011) 'Generation of Multimillion-Sequence 16S rRNA Gene Libraries from Complex Microbial Communities by Assembling Paired-End Illumina Reads', *Applied and Environmental Microbiology*, 77(11), pp. 3846–3852. doi: 10.1128/AEM.02772-10.

Behrenfeld, M. J. *et al.* (2006) 'Climate-driven trends in contemporary ocean productivity', *Nature*, 444(7120), pp. 752–755. doi: 10.1038/nature05317.

Behrenfeld, M. J. *et al.* (2016) 'Revaluating ocean warming impacts on global phytoplankton', *Nature Climate Change*, 6(3), pp. 323–330. doi: 10.1038/nclimate2838.

Berube, P. M. *et al.* (2019) 'Emergence of trait variability through the lens of nitrogen assimilation in Prochlorococcus', *eLife*, 8, pp. 1–28. doi: 10.7554/elife.41043.

Biller, S. J. *et al.* (2015) 'Prochlorococcus: The structure and function of collective diversity', *Nature Reviews Microbiology*. 13(1), pp. 13–27. doi: 10.1038/nrmicro3378.

Bopp, L. *et al.* (2015) *The Ocean a carbon pump*, *ocean-climate*.

Boyd, P. and Newton, P. (1995) ‘Evidence of the potential influence of planktonic community structure on the interannual variability of particulate organic carbon flux’, *Deep Sea Research Part I: Oceanographic Research Papers*, 42(5), pp. 619–639. doi: 10.1016/0967-0637(95)00017-Z.

Breitbart, M. *et al.* (2002) ‘Genomic analysis of uncultured marine viral communities’, *Proceedings of the National Academy of Sciences*, 99(22), pp. 14250–14255. doi: 10.1073/pnas.202488399.

Breitbart, M. *et al.* (2018) ‘Phage puppet masters of the marine microbial realm’, *Nature Microbiology*. Springer US, 3(7), pp. 754–766. doi: 10.1038/s41564-018-0166-y.

Brown, M. V. *et al.* (2018) ‘Data Descriptor: Systematic, continental scale temporal monitoring of marine pelagic microbiota by the Australian Marine Microbial Biodiversity Initiative’, *Scientific Data*, 5(April), pp. 1–10. doi: 10.1038/sdata.2018.130.

Brum, J. R. *et al.* (2015) ‘Ocean Viral Communities’, *Science*, 348(6237), pp. 1261498-1–11. doi: 10.1126/science.1261498.

Brum, J. R. *et al.* (2016) ‘Patterns and ecological drivers of ocean viral communities’, *Science*, 348(6237).

Brum, J. R. and Sullivan, M. B. (2015) ‘Rising to the challenge: Accelerated pace of discovery transforms marine virology’, *Nature Reviews Microbiology*. Nature Publishing Group, 13(3), pp. 147–159. doi: 10.1038/nrmicro3404.

Brussaard, C. P. D. (2004) ‘Optimization of procedures for counting viruses by flow cytometry’, *Applied and environmental microbiology*. American Society for Microbiology, 70(3), pp. 1506–1513. doi: 10.1128/AEM.70.3.1506-1513.2004.

Chisholm, S. W. *et al.* (1988) ‘A novel free-living prochlorophyte abundant in the oceanic euphotic zone Sallie’, *Nature*, 334.

Collier, J. L. (2000) 'FLOW CYTOMETRY AND THE SINGLE CELL IN PHYCOLOGY', *Journal of Phycology*. John Wiley & Sons, Ltd (10.1111), 36(4), pp. 628–644. doi: 10.1046/j.1529-8817.2000.99215.x.

Courties, C. *et al.* (1994) 'Smallest eukaryotic organism', *Nature*, 370(6487), pp. 255–255. doi: 10.1038/370255a0.

Dammeyer, T. *et al.* (2008) 'Efficient Phage-Mediated Pigment Biosynthesis in Oceanic Cyanobacteria', *Current Biology*, 18(6), pp. 442–448. doi: 10.1016/j.cub.2008.02.067.

Danovaro, R. *et al.* (2005) 'Viruses, prokaryotes and DNA in the sediments of a deep-hypersaline anoxic basin (DHAB) of the Mediterranean Sea', *Environmental Microbiology*, 7(4), pp. 586–592. doi: 10.1111/j.1462-2920.2005.00727.x.

Danovaro, R. *et al.* (2008) 'Major viral impact on the functioning of benthic deep-sea ecosystems', *Nature*, 454(7208), pp. 1084–1087. doi: 10.1038/nature07268.

Danovaro, R. *et al.* (2011) 'Marine viruses and global climate change', *FEMS Microbiology Reviews*, 35(6), pp. 993–1034. doi: 10.1111/j.1574-6976.2010.00258.x.

Delmont, T. O. *et al.* (2015) 'Genome reconstructions indicate the partitioning of ecological functions inside a phytoplankton bloom in the Amundsen Sea, Antarctica', *Frontiers in Microbiology*, 6(OCT), pp. 1–19. doi: 10.3389/fmicb.2015.01090.

DeLong, E. F. (2005) 'Microbial community genomics in the ocean', *Nature Reviews Microbiology*, 3(6), pp. 459–469. doi: 10.1038/nrmicro1158.

DeLong, E. F. *et al.* (2006) 'Community genomics among stratified microbial assemblages in the ocean's interior', *Science*, 311(5760), pp. 496–503. doi: 10.1126/science.1120250.

Deng, L. *et al.* (2014) 'Viral tagging reveals discrete populations in Synechococcus viral genome sequence space', *Nature*. Nature Publishing Group, 513(7517), pp. 242–245. doi: 10.1038/nature13459.

Edwards, R. A. and Rohwer, F. (2005) 'Viral metagenomics', *Nature*, 3(June), pp. 504–510. doi: 10.1002/rmv.532.

Evans, C. *et al.* (2003) 'Direct estimates of the contribution of viral lysis and microzooplankton grazing to the decline of a Micromonas spp. population', *Aquatic Microbial Ecology*, 30, pp. 207–219.

Falkowski, P., Laws, E. A., Barber, R. T. and Murray, J. W. J. (2003) 'Phytoplankton and Their Role in Primary, New, and Export Production', *Ocean Biogeochemistry*, (March 2014), pp. 99–121. doi: 10.1007/978-3-642-55844-3.

Falkowski, P., Laws, E. A., Barber, R. T. and Murray, J. W. (2003) 'Phytoplankton and Their Role in Primary, New, and Export Production', *Ocean Biogeochemistry*, pp. 99–121. doi: 10.1007/978-3-642-55844-3_5.

Falkowski, P. *et al.* (2011) 'The Global Carbon Cycle : A Test of Our Knowledge of Earth as a System', *Science*, 291(2000), pp. 290–296. doi: 10.1126/science.290.5490.291.

Farrant, G. K. *et al.* (2016) 'Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria', *Proceedings of the National Academy of Sciences*. National Academy of Sciences, 113(24), pp. E3365–E3374. doi: 10.1073/PNAS.1524865113.

Ferris, M. J. and Palenik, B. (1998) 'Niche adaptation in ocean cyanobacteria', *Nature*, 396, pp. 226–228. doi: 10.1111/mms.12278.

Flombaum, P. *et al.* (2013) 'Present and future global distributions of the marine Cyanobacteria Prochlorococcus and Synechococcus', *Proceedings of the National Academy of Sciences*, 110(24). doi: 10.1073/pnas.1307701110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1307701110.

Fuhrman, J. A. (1999) 'Marine viruses and their biogeochemical and ecological effects', *Nature*, 399, pp. 1–8. doi: 10.1038/21119.

Giering, S. L. C. *et al.* (2014) 'Reconciliation of the carbon budget in the ocean's twilight zone', *Nature*. Nature Publishing Group, 507(7493), pp. 480–483. doi: 10.1038/nature13123.

Graff, J. R., Milligan, A. J. and Behrenfeld, M. J. (2012) 'The measurement of phytoplankton biomass using flowcytometric sorting and elemental analysis of carbon', *Limnology and Oceanography: Methods*, 10(NOVEMBER), pp. 910–920. doi: 10.4319/lom.2012.10.910.

Grébert, T. *et al.* (2018) ' Light color acclimation is a key process in the global ocean distribution of Synechococcus cyanobacteria ', *Proceedings of the National Academy of Sciences*, 115(9), pp. E2010–E2019. doi: 10.1073/pnas.1717069115.

Grob, C. *et al.* (2007) 'Contribution of picoplankton to the total particulate organic carbon concentration in the eastern South Pacific', *Biogeosciences*, 4(5), pp. 837–852. doi: 10.5194/bg-4-837-2007.

Grob, C. *et al.* (2013) 'Elemental composition of natural populations of key microbial groups in Atlantic waters', *Environmental Microbiology*, 15(11), pp. 3054–3064. doi: 10.1111/1462-2920.12145.

Guidi, L. *et al.* (2009) 'Effects of phytoplankton community on production, size and export of large aggregates: A world-ocean analysis', *Limnology and Oceanography*, 54(6), pp. 1951–1963. doi: 10.4319/lo.2009.54.6.1951.

Guidi, L. *et al.* (2016) 'Plankton networks driving carbon export in the oligotrophic ocean', *Nature*. Nature Publishing Group, 532(7600), pp. 465–470. doi: 10.1038/nature16942.

Henson, S. A. *et al.* (2011) 'A reduced estimate of the strength of the ocean's biological carbon pump', *Geophysical Research Letters*, 38(4), pp. 10–14. doi: 10.1029/2011GL046735.

Huang, S. *et al.* (2012) 'Novel lineages of prochlorococcus and synechococcus in the global oceans', *ISME Journal*. Nature Publishing Group, 6(2), pp. 285–297. doi: 10.1038/ismej.2011.106.

Hug, L. A. *et al.* (2016) 'A new view of the tree of life', *Nature Microbiology*, 1(5), pp. 1–6. doi: 10.1038/nmicrobiol.2016.48.

Hülse, D. *et al.* (2017) 'Understanding the causes and consequences of past marine carbon cycling variability through models', *Earth-Science Reviews*. Elsevier, 171(September 2016), pp. 349–382. doi: 10.1016/j.earscirev.2017.06.004.

Hurwitz, B. L., Brum, J. R. and Sullivan, M. B. (2015) 'Depth-stratified functional and taxonomic niche specialization in the " core " and " flexible " Pacific Ocean Virome', *ISME Journal*, pp. 472–484. doi: 10.1038/ismej.2014.143.

Hurwitz, B. L. and U'Ren, J. M. (2016) 'Viral metabolic reprogramming in marine ecosystems', *Current Opinion in Microbiology*. Elsevier Ltd, 31, pp. 161–168. doi: 10.1016/j.mib.2016.04.002.

IPCC (2014) *IPCC,2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)].*

Jiao, N. *et al.* (2010) 'Microbial production of recalcitrant dissolved organic matter: Long-term carbon storage in the global ocean', *Nature Reviews Microbiology*. Nature Publishing Group, 8(8), pp. 593–599. doi: 10.1038/nrmicro2386.

Jiao, N. and Zheng, Q. (2011) ‘MINIREVIEW The Microbial Carbon Pump : from Genes to Ecosystems □’, *Applied and Environmental Microbiology*, 77(21), pp. 7439–7444. doi: 10.1128/AEM.05640-11.

Jover, L. F. *et al.* (2014) ‘The elemental composition of virus particles: Implications for marine biogeochemical cycles’, *Nature Reviews Microbiology*. Nature Publishing Group, 12(7), pp. 519–528. doi: 10.1038/nrmicro3289.

Karl, D. M. and Church, M. J. (2014) ‘Microbial oceanography and the Hawaii Ocean Time-series programme’, *Nature Reviews Microbiology*. Nature Publishing Group, 12(10), pp. 699–713. doi: 10.1038/nrmicro3333.

Kwiatkowski, L. *et al.* (2018) ‘The Impact of Variable Phytoplankton Stoichiometry on Projections of Primary Production, Food Quality, and Carbon Uptake in the Global Ocean’, *Global Biogeochemical Cycles*, 32(4), pp. 516–528. doi: 10.1002/2017GB005799.

Lindell, D. *et al.* (2004) ‘Transfer of photosynthesis genes to and from Prochlorococcus viruses’, *Proceedings of the National Academy of Sciences*, 101(30), pp. 11013–11018.

Lindell, D. *et al.* (2007) ‘Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution’, *Nature*, 449(7158), pp. 83–86. doi: 10.1038/nature06130.

López-Pérez, M. *et al.* (2017) ‘Genome diversity of marine phages recovered from Mediterranean metagenomes: Size matters’, *PLoS Genetics*, 13(9), pp. 1–23. doi: 10.1371/journal.pgen.1007018.

Luo, E. *et al.* (2017) ‘Bacteriophage Distributions and Temporal Variability in the Ocean’s Interior’, *mBio*. Edited by M. A. Moran, 8(6), pp. 1–13. doi: 10.1128/mBio.01903-17.

Mann, N. H. *et al.* (2003) ‘Bacterial photosynthesis genes in a virus’, *Nature*, 424(August), pp. 741–742.

Marie, D. *et al.* (1997) 'Enumeration and Cell Cycle Analysis of Natural Populations of Marine Picoplankton by Flow Cytometry …', *Applied and Environmental Microbiology*, 63(1), pp. 186–193.

Marie, D. *et al.* (1999) 'Enumeration of Phytoplankton, Bacteria, and Viruses in Marine Samples', *Current Protocols in Cytometry*. John Wiley & Sons, Ltd, 10(1), pp. 11.11.1-11.11.15. doi: 10.1002/0471142956.cy1111s10.

Martiny, A. C. *et al.* (2013) 'Strong latitudinal patterns in the elemental ratios of marine plankton and organic matter', *Nature Geoscience*. Nature Publishing Group, 6(4), pp. 279–283. doi: 10.1038/ngeo1757.

Martiny, A. C. *et al.* (2016) 'Interactions between thermal acclimation, growth rate, and phylogeny influence prochlorococcus elemental stoichiometry', *PLoS ONE*, 11(12), pp. 1–12. doi: 10.1371/journal.pone.0168291.

Mazard, S. *et al.* (2012) 'Multi-locus sequence analysis, taxonomic resolution and biogeography of marine Synechococcus', *Environmental Microbiology*. John Wiley & Sons, Ltd (10.1111), 14(2), pp. 372–386. doi: 10.1111/j.1462-2920.2011.02514.x.

Millard, A. D. *et al.* (2009) 'Comparative genomics of marine cyanomyoviruses reveals the widespread occurrence of Synechococcus host genes localized to a hyperplastic region: Implications for mechanisms of cyanophage evolution', *Environmental Microbiology*, 11(9), pp. 2370–2387. doi: 10.1111/j.1462-2920.2009.01966.x.

Mojica, K. D. A. *et al.* (2016) 'Latitudinal variation in virus-induced mortality of phytoplankton across the North Atlantic Ocean', *ISME Journal*. Nature Publishing Group, 10(2), pp. 500–513. doi: 10.1038/ismej.2015.130.

Moore, L. R., Rocap, G. and Chisholm, S. W. (2002) 'Physiology and molecular phylogeny of coexisting Prochlorococcus ecotypes', *Nature*, 393(6684), pp. 464–467. doi: 10.1038/30965.

Moreno, A. R. and Martiny, A. C. (2018) ‘Ecological Stoichiometry of Ocean Plankton’, *Annual Review of Marine Science*, 10(1), pp. 43–69. doi: 10.1146/annurev-marine-121916-063126.

Mouginot, C. *et al.* (2015) ‘Resource allocation by the marine cyanobacterium Synechococcus WH8102 in response to different nutrient supply ratios’, *Limnology and Oceanography*, 60(5), pp. 1634–1641. doi: 10.1002/lno.10123.

Needham, D. M., Sachdeva, R. and Fuhrman, J. A. (2017) ‘Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters’, *ISME Journal*. Nature Publishing Group, 11(7), pp. 1614–1629. doi: 10.1038/ismej.2017.29.

Olson, R. J., Vaulot, D. and Chisholm, S. W. (1985) ‘Marine phytoplankton distributions measured using shipboard flow cytometry’, *Deep Sea Research Part A, Oceanographic Research Papers*, 32(10), pp. 1273–1280. doi: 10.1016/0198-0149(85)90009-3.

Palenik, B. *et al.* (2006) ‘Genome sequence of Synechococcus CC9311: Insights into adaptation to a coastal environment’, *Proceedings of the National Academy of Sciences*, 103(36), pp. 13555–13559. doi: 10.1073/pnas.0602963103.

Parks, D. H. *et al.* (2017) ‘Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life’, *Nature Microbiology*. Springer US, 2(11), pp. 1533–1542. doi: 10.1038/s41564-017-0012-7.

Partensky, F., Hess, W. R. and Vaulot, D. (1999) ‘Prochlorococcus, a marine photosynthetic prokaryote of global significance.’, *Microbiology and molecular biology reviews : MMBR*, 63(1), pp. 106–127. doi: doi:1092-2172/99/$04.00.

Passow, U. and Carlson, C. A. (2012) ‘The biological pump in a high CO2world’, *Marine Ecology Progress Series*, 470(2), pp. 249–271. doi: 10.3354/meps09985.

Pasulka, A. L., Samo, T. J. and Landry, M. R. (2015) 'Grazer and viral impacts on microbial growth and mortality in the southern California Current Ecosystem', *Journal of Plankton Research*, 37(2), pp. 320–336. doi: 10.1093/plankt/fbv011.

Puxty, R. J. *et al.* (2015) 'Shedding new light on viral photosynthesis', *Photosynthesis Research*. Springer Netherlands, pp. 71–97. doi: 10.1007/s11120-014-0057-x.

Puxty, R. J. *et al.* (2018) 'Energy limitation of cyanophage development: Implications for marine carbon cycling', *ISME Journal*. Springer US, 12(5), pp. 1273–1286. doi: 10.1038/s41396-017-0043-3.

Le Quéré, C. *et al.* (2018) 'Carbon cycle-Global Carbon Budget 2017', *Earth Science System Data*, 10, pp. 405–448. doi: 10.5194/essd-10-405-2018.

Raven, J. *et al.* (2005) *Ocean acidification due to increasing*, *The royal society of London*.

Rogato, A. *et al.* (2015) 'The diatom molecular toolkit to handle nitrogen uptake', *Marine Genomics*. Elsevier B.V., 24, pp. 95–108. doi: 10.1016/j.margen.2015.05.018.

Roshan, S. and DeVries, T. (2017) 'Efficient dissolved organic carbon production and export in the oligotrophic ocean', *Nature communications*. Springer US, 8(1), p. 2036. doi: 10.1038/s41467-017-02227-3.

Roux, S. *et al.* (2016) 'Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses', *Nature*. Nature Publishing Group, 537(7622), pp. 689–693. doi: 10.1038/nature19366.

Rusch, D. B. *et al.* (2007) 'The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through eastern tropical Pacific', *PLoS Biology*, 5(3), pp. 0398–0431. doi: 10.1371/journal.pbio.0050077.

Scanlan, D. J. *et al.* (2009) 'Ecological Genomics of Marine Picocyanobacteria', *Microbiology and Molecular Biology Reviews*, 73(2), pp. 249–299. doi: 10.1128/MMBR.00035-08.

Simkovsky, R. *et al.* (2012) 'Impairment of O-antigen production confers resistance to grazing in a model amoeba-cyanobacterium predator-prey system', *Proceedings of the National Academy of Sciences*, 109(41), pp. 16678–16683. doi: 10.1073/pnas.1214904109.

Sullivan, M. B. *et al.* (2010) 'Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments', *Environmental Microbiology*, 12(11), pp. 3035–3056. doi: 10.1111/j.1462-2920.2010.02280.x.

Sunagawa, S. *et al.* (2015) 'Structure and function of the global ocean microbiome', *Science*, 348(6237), p. 1261359. doi: 10.1126/science.1261359.

Suttle, C. A. (2005) 'Viruses in the sea', *Nature*, 437. doi: 10.1111/j.1574-6976.2008.00132.x.

Suttle, C. A. (2007) 'Marine viruses - Major players in the global ecosystem', *Nature Reviews Microbiology*, 5(10), pp. 801–812. doi: 10.1038/nrmicro1750.

Swan, B. K. *et al.* (2013) 'Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean', *Proceedings of the National Academy of Sciences*, 110(28), pp. 11463–11468. doi: 10.1073/pnas.1304246110.

Temperton, B. and Giovannoni, S. J. (2012) 'Metagenomics: Microbial diversity through a scratched lens', *Current Opinion in Microbiology*. Elsevier Ltd, 15(5), pp. 605–612. doi: 10.1016/j.mib.2012.07.001.

Tijdens, M. *et al.* (2008) 'Estimates of bacterial and phytoplankton mortality caused by viral lysis and microzooplankton grazing in a shallow eutrophic lake', *Freshwater Biology*, 53(6), pp. 1126–1141. doi: 10.1111/j.1365-2427.2008.01958.x.

Tragin, M. and Vaulot, D. (2018) 'Green microalgae in marine coastal waters: The Ocean Sampling Day (OSD) dataset', *Scientific Reports*. Springer US, 8(1), pp. 1–12. doi: 10.1038/s41598-018-32338-w.

Tréguer, P. *et al.* (2017) 'Influence of diatom diversity on the ocean biological carbon pump', *Nature Geoscience*. Springer US, 11(1), pp. 27–37. doi: 10.1038/s41561-017-0028-x.

Tully, B. J., Graham, E. D. and Heidelberg, J. F. (2017) 'The Reconstruction of 2,631 Draft Metagenome-Assembled Genomes from the Global Oceans', 4(4), pp. 516–525.

Tyson, G. W. *et al.* (2004) 'Community structure and metabolism through reconstruction of microbial genomes from the environment', *Nature*, 428(6978), pp. 37–43. doi: 10.1038/nature02340.

Varkey, D. *et al.* (2016) 'Effects of low temperature on tropical and temperate isolates of marine Synechococcus', *ISME Journal*. Nature Publishing Group, 10(5), pp. 1252–1263. doi: 10.1038/ismej.2015.179.

Venter, J. C. *et al.* (2004) 'Sequencing of the Sargasso Sea', *Science (New York, N.Y.)*, 304(April), pp. 66–74. doi: 10.1126/science.1093857.

Waterbury, J. B. *et al.* (1979) 'Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium', *Nature*, 277(January), pp. 293–294.

Weinbauer, M. G. *et al.* (2007) 'Synergistic and antagonistic effects of viral lysis and protistan grazing on bacterial biomass, production and diversity', *Environmental Microbiology*, 9(3), pp. 777–788. doi: 10.1111/j.1462-2920.2006.01200.x.

West, N. J. *et al.* (2011) 'A novel clade of Prochlorococcus found in high nutrient low chlorophyll waters in the South and Equatorial Pacific Ocean', *ISME Journal*. Nature Publishing Group, 5(6), pp. 933–944. doi: 10.1038/ismej.2010.186.

Williamson, S. J. *et al.* (2008) 'The Sorcerer II Global Ocean Sampling Expedition : Metagenomic Characterization of Viruses within Aquatic Microbial Samples', *PLoS ONE*, (1). doi: 10.1371/journal.pone.0001456.

Wommack, K. E. and Colwell, R. R. (2000) 'Virioplankton : Viruses in Aquatic Ecosystems †', *Microbiology and Molecular Biology Reviews*, 64(1), pp. 69–114.

Worden, A. *et al.* (2015) 'Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes', *Science*, 347(6223). doi: 10.1126/science.1257594.

Worden, A. Z., Nolan, J. K. and Palenik, B. (2004) 'Assessing the dynamic and ecology of marine picoplankton : the importance of eukaryotic component', *Limnol. Oceanogr.*, 49(1), pp. 168–79.

Zackary I. Johnson *et al.* (2006) 'Niche Partitioning Among Prochlorococcus Ecotypes Along Ocean-Scale Environmental Gradients', *Science*, 311(5768), pp. 1737–1740. doi: 10.1126/science.1118052.

Zeidner, G. *et al.* (2003) 'Molecular diversity among marine picophytoplankton as revealed by psbA analyses', *Environmental Microbiology*, 5(3), pp. 212–216. doi: 10.1046/j.1462-2920.2003.00403.x.

Zhao, Z. *et al.* (2017) 'Picocyanobacteria and deep-ocean fluorescent dissolved organic matter share similar optical properties', *Nature Communications*. The Author(s), 8, p. 15284.

Zwirglmaier, K. *et al.* (2008) 'Global phylogeography of marine Synechococcus and Prochlorococcus reveals a distinct partitioning of lineages among oceanic biomes', *Environmental Microbiology*, 10(1), pp. 147–161. doi: 10.1111/j.1462-2920.2007.01440.x.

# Chapter 2:

# Change in the relative contributions of picophytoplankton, bacteria and viruses to the carbon biomass in a model western boundary current

# Change in the relative contribution of picophytoplankton, bacteria and viruses to the carbon biomass in a model boundary current.

Amaranta Focardi[1], Martin Ostrowski[1] and Ian Paulsen[1]

[1]Department of Molecular Sciences, Macquarie University, North Ryde, 2109, NSW, Australia

## 2.1 Abstract

In the context of current, and future, climate-induced variability in ocean ecosystem it is of crucial importance to accurately quantify the relative contribution of organisms that underpin the biological carbon pump in distinct region of the Ocean. Marine picoplankton (< 3 μm) contribute significantly to the biological carbon cycle in many ocean basins. However, there is still limited data on the relative contribution of microbial groups and dynamics of carbon flow from the base of marine food web. This is particularly true for the Southern hemisphere, and this knowledge gap impairs our ability to understand the factors affecting the present and future of regional and global productivity.

Here we present the first quantitative study on the contribution of microbial groups (< 10 μm dia.) to the carbon (C) biomass of the East Australian Current (EAC) system. Cytometry derived counts of picophytoplankton, bacteria and viruses, coupled with a direct estimation of the C quota for picophytoplankton, show that picocyanobacteria play an important role in the C biomass within the EAC, while Tasman Sea biomass was dominated by picoeukaryotic phytoplankton. Results highlighted a high plasticity in picocyanobacteria C quota primarily correlated with

temperature and a combinatory effect of temperature and nutrient concentration. A series of dilution experiments determined that photosynthetically-fixed C had distinctly different fate in the EAC compared with adjacent water masses. Estimated rates of grazing, representing transfer into the classical food chain, were consistent across the region. In stark contrast, the *in-situ* rates of viral lysis were significantly higher in the EAC. Overall these results highlight a potential shift towards a higher importance of picocyanobacteria for C biomass and a decrease in the relative flux of C from picophytoplankton to higher trophic levels in this biome undergoing significant change.

## 2.2 Introduction

Marine picophytoplankton (< 3 µm) play a significant role in the biological carbon pump across in vast expanses of the Oceans, especially the warm oligotrophic gyres (Alvain *et al.*, 2005). Together with heterotrophic bacteria and viruses, they play a pivotal role in the oceanic biogeochemical cycle (Guidi *et al.*, 2016; Moran *et al.*, 2016) and significantly contribute to the export of carbon from surface to the deep Ocean via the biological carbon pump (Azam, 1998; Field *et al.*, 1998; Jiao *et al.*, 2010; Poloczanska, 2013; Giering *et al.*, 2014; Bopp *et al.*, 2015; Zhang *et al.*, 2018).

Although they present a variety of genetically diverse organisms, picophyotoplankton are mainly composed of three distinct groups: photosynthetic pico-eukaryotes (PPE) and the oxyphototrophic *Synechococcus* and *Prochlorococcus* cyanobacteria. These organisms are characterized by a growth rate between 0.4-1.4 a day (Kirchman, 2016), that depends on their physiological status and also a reasonably fast turnover controlled by microzooplankton grazing and viral lysis. With a global distribution, those tiny primary producers are responsible for a substantial fraction of the primary production (Buitenhuis *et al.*, 2012; Worden *et al.*, 2015), with prokaryotic *Prochlorococcus* mostly thriving in tropical and oligotrophic water compared to *Synechococcus* that has a broader distribution (Zwirglmaier *et al.*, 2008; Flombaum *et al.*, 2013; Biller *et al.*, 2015;

Farrant *et al.*, 2016). While, PPE instead are found to be more abundant in mesotrophic waters (Worden, Nolan and Palenik, 2004; Vaulot *et al.*, 2008). Because of their global distribution, and importance for the marine food web, characterizing the contribution of picophytoplankton groups to the C biomass in different oceanic regions is of crucial importance to predict the impact of future climate change on C flow in the marine ecosystems.

Cellular carbon quota (Qc) directly affects the picophytoplankton contribution to the carbon stock and is expected to adjust in response to different physiological stressors (Bertilsson *et al.*, 2003; Heldal *et al.*, 2003; Moreno and Martiny, 2018). However, due to the difficulty of field measurement, direct estimates of specific carbon composition of picophytoplankton in different marine environments are still quite limited (Zubkov *et al.*, 2000; Shalapyonok, Olson and Shalapyonok, 2001; Gundersen *et al.*, 2002; Graff, Milligan and Behrenfeld, 2012; Casey *et al.*, 2013; Grob *et al.*, 2013; Martiny *et al.*, 2013; Graff *et al.*, 2015; Martiny, Talarmin, *et al.*, 2016; Baer *et al.*, 2017). In the absence of direct measurements, typically estimates of the group-specific contribution to C biomass are estimated from count data (e.g. from analytical flow cytometry) converted to C estimates using conversion factors derived either from average Qc of laboratory cultures (Bertilsson *et al.*, 2003; Heldal *et al.*, 2003; Lopez *et al.*, 2016; Martiny, Ma, *et al.*, 2016), or from field experiments. The use of conversion factors based on laboratory cultures has the potential to introduce large errors in the estimates of the oceanic carbon budget, because the growth conditions of clonal isolates in batch cultures rarely reflect the true physiological state of diverse (pico)phytoplankton *in-situ*. Climate-induced variability in species distribution has the potential to alter the biological C pump, connectivity of the food chain and capability of the ocean to act as a C sink. Limited (global) information on the *in-situ* plasticity of picophytoplankton cellular Qc bias our ability to develop models that accurately predict their response to ecosystem variation in the different oceanic regions.

Here we present the first detailed quantitative study of picophytoplankton, their *in-situ* Qc and relative contribution to the carbon biomass for the euphotic zone of East Australian current and the adjacent Tasman Sea. The East Australian

current (EAC), a model western boundary current (WBC), is an hotspot for ocean warming (Wu *et al.*, 2012), with an important role in transporting warm and low nutrient water masses from the tropics region to higher latitudes (Tilburg *et al.*, 2001; Suthers *et al.*, 2011; Archer *et al.*, 2017; Behrenfeld and Boss, 2018). It represents an ideal model site to investigate the possible ecosystem consequence of climate change because of its accessibility from the East Coast of Australia, the availability of long term oceanographic observations in the region (IMOS Integrated Marine Observing System), and the recent establishment of a microbial monitoring program (Brown *et al.*, 2018). The EAC presents an interannual variability (Suthers *et al.*, 2011; Archer *et al.*, 2017), with an average annual southward flow of 1.83 $ms^{-1}$ in its core. After the separation from the continental slope at approximately (30-37 ˚S) (Kane, Oke and Sandery, 2011) it flows eastward while anticyclonic or cyclonic eddies usually form from the south end and continue southward and eastward, reaching sometimes the coast of Tasmania (Ridgway, 1997; Kane, Oke and Sandery, 2011; Suthers *et al.*, 2011), this eddy field is also known as the southward extension of the EAC (Everett *et al.*, 2012).

In order to better understand how the microbial fraction (< 10 µm ) contributes to the carbon budget in the context of the EAC and the Tasman Sea, picophytoplankton heterotrophic bacteria and viruses counts, coupled with direct estimation of the carbon quota per picophytoplankton, were transformed into their relative contribution to the C stock. We also applied a series of dilution experiment to assess the importance of protist grazing and viral lysis on picophytoplankton mortality and their related impact on the carbon flux within the EAC system and the adjacent Tasman sea. The number of sites sampled from different biomes of the East Coast of Australia highlights the variability in cellular Qc in response to different environmental variables and moreover, the importance of picocyanobacteria and viruses for the carbon flux within the EAC system.

## 2.2 Methods

***Study area, samples collection, and environmental parameters***

Samples were collected during an oceanographic research expedition on board the RV Investigator (IN2016v04, CSIRO, MNF Australia) in September 2016 (austral spring) on the East coast of Australia (Fig. 2. 1).

A total of 80 samples were collected between 0 to 110m depths along three different shore-normal transects, a lagrangian drift along the EAC from 30.36°S to 33.04°S and three coastal sites on the continental shelf (bottom depth ~100 m). The three transects covered different oceanic regions at latitude 36.22°S, 32.47°S and 27.99°S. The transect South was offshore in the Tasman Sea water with influence of the southward extension of the EAC in two sites. While the transect middle and transect North had five different stations each, between the coast and EAC (shore), within the EAC and offshore. Samples along each transect were collected sequentially, within 24h, while samples in along the drift were collected every 4h. Coastal samples were not collected sequentially.

The water was collected from depth profiles undertaken with a rosette (SBE11) equipped with 12L x 24 Niskin bottles (OceanTest Equipment Inc. Florida) and coupled with SBE911plus CTD fitted with additional multiples sensors (Fluorescence, oxygen, turbidity). Samples for nutrients ($NO_x$, $PO_4$, Si), dissolved oxygen, salinity were collected concomitantly and analysed on board.

Data and methodology followed for the physical and chemical analyses were made available through the CSIRO Ocean and Atmosphere data trawler repository.
(https://www.cmar.csiro.au/data/trawler/survey_details.cfm?survey=IN2016_V04 ).

### *Characterization of Microbial communities*

A total of 80 samples were collected from 34 CTD deployments for the analysis of microbial communities at different depths, from the surface (~2 m depth) to 110 m below the surface Four ml of unfiltered seawater was fixed with paraformaldehyde at a final concentration of 0.5% (w/v) for 15 min in the dark, flash frozen in liquid nitrogen and stored at -80˚C until further analysis on shore.

Samples were thawed at 34˚C for 10 minutes then, picophotosynthetic communities were analysed and enumerated on a Cytoflex S flow cytometer (Beckman Coulter), equipped with 3 different lasers (violet 405 nm, yellow 532 nm and blue 488 nm) and their related detectors. The distinctive intensity of the specific photosynthetic pigments, chlorophyll and phycoerythrin, were used against side scatter on both the violet and blue laser and forward scatter, from the blue laser, to distinguish *Prochlorococcus*, photosynthetic Picoeukaryotes, and *Synechococcus* populations (Marie *et al.*, 1997). The flow rate was kept stable at 30 µl/min and each sample run for 2 min. The flow rate and cell number estimates were cross referenced against counts made on an Influx (Becton Dickinson) using volumetric determination of the flow rate.

To enumerate heterotrophic bacteria, samples were diluted 1/5 with TE buffer and stained with SYBR green I (Invitrogen), both pre-filtered with a 0.02 µm filter (GE healthcare), for 10 minutes in the dark and then loaded on the same cytometer and two different populations (LDNA, HDNA) were identified. The two populations were identified based on signal intensity from SYBR fluorescence, LDNA (low DNA bacteria) and HDNA (High DNA bacteria).

The viruses were analysed following a standard protocol (Brussaard, 2004). Briefly, thawed samples were diluted with TE buffer and then incubated at 80°C for 10 min with 1X SYBR green I (Invitrogen), both pre-filtered as per bacteria protocol, cooled for 5 min in the dark, prior to counting on the Cytoflex S flow cytometer.

Consistent gating strategies were applied to all the samples to discriminate at least 3 different population of viruses (Virus1, Virus2, Virus3) and two populations of heterotrophic bacteria (LDNA, HDNA) (Suppl. Fig. 2.3). The detector voltages and gates were fine-tuned at the beginning, and middle of each run, to ensure that the populations were well resolved from background noise.

For the photosynthetic communities two distinct *Prochlorococcus* (HL, LL), *Synechococcus* and 3 different groups of Eukaryotic algae, picoEukaryotes (PPE), nano-Eukaryotes (Nano) and phycoerythrin rich eukaryotes (PE-Euk), were identified based on different Chlorophyll (Chl), forward scatter (FSC) and side scatter (SSC) biplots (Suppl. Fig. 2.2).

***Cellular carbon content determination for picophotosynthetic functional groups and specific contribution to carbon biomass***

Triplicate samples of *Synechococcus, Prochlorococcus*, and picoeukaryotes were flow-sorted from six samples collected at contrasting stations, three along the EAC, two along the Tasman Sea and one from a coastal site. Cells were sorted using the drop-pure selector on an Influx BC flow-sorter equipped with three different lasers. Based on expected carbon content a minimum of $5 \times 10^5$, $3 \times 10^6$, $6 \times 10^6$ cells were sorted for picoeukaryotes, *Synechococcus* and *Prochlorococcus* respectively. Sort rates kept between 3,000 events/s and 8,000 event/s for a sort purity higher than 95%. Organisms were distinguished based on specific autofluorescence as per cell counting. Sorted cells were then deposited on a pre-combusted (450 ˚C for 8 h) GFF filter using a vacuum pump and flash frozen in liquid $N_2$. Sort purity and yield were verified at regular intervals by reanalysing an aliquot of the sorted population. Sorted populations with < 90% yield, or > 5% difference from the expected purity and yield were discarded. Analyses for the carbon elemental quota were carried out at the Bigelow Laboratory for Ocean study in Maine. Carbon estimation were corrected against multiple blank measurements of the pre-combusted filter, and sheath fluid before proceeding with further analysis.

A calibration curve was performed between the mean forward scatter of each picophytoplankton population sorted and the relative carbon quota derived from

the analysis of the same population on an elemental analyser (Casey *et al.*, 2013). Both natural populations of *Synechococcus* and natural populations and cultured isolates of *Prochlorococcus* were included in the calibration curve. Due to the variability in mean FSC and relative carbon content, natural populations of picoeukaryotes were excluded from the calibration curve. The calibration curve was used to convert mean FSC in the relative amount of carbon for both *Synechococcus* and *Prochlorococcus* populations from all the 80 samples analysed. The specific carbon quota for each *Synechococcus, Prochlorococcus*, population was used to convert the group count in units of carbon biomass expressed in $mgC/m^3$ of seawater.

For picoeukaryotes, even though mean FSC was marginally greater than cyanobacteria, the Qc was up to 10 times higher, and it was not possible to establish a direct correlation between the Qc and the mean FSC per populations ($r^2$=0.45). We decided then to use those fixed values obtained from the elemental analysis to convert the PPE cell counts into standing stock of carbon. Using the cellular quota obtained from Tasman, Coastal and EAC, respectively to estimate the proportion of PPE carbon from each region. Water properties from the offshore sites in the north and middle transects were similar to the EAC than sites from the Tasman Sea. Because samples from the north and middle transects encompass different regions (shore, EAC and offshore), estimates from the coastal site were used for the near-shore site, and estimates from the coastal site were used from the EAC were used for both the EAC and offshore samples.

### *Heterotrophic bacteria and Viruses*

Heterotrophic bacteria and viruses were converted to carbon stock using a fixed, average value from the literature of 12 fgC $cell^{-1}$ and 0.2 fgC $virus^{-1}$ respectively (Suttle, 2005; Caron *et al.*, 2017).

### *Dilution experiment*

Dilution experiments were performed following a modification of Landry protocol (Landry and Hassett, 1982; Landry, Kirshtein and Constantinou, 1995; Evans,

Pearce and Brussaard, 2009) in different oceanic region. Six experiments were conducted in the EAC, three in the Tasman Sea and three in Coastal sites. For all the experiments five dilutions were performed in triplicate. Firstly, the water collected from the surface (~2 m) was gently pre-filtered, directly from the Niskin bottle, through a 60 µm mesh (Millipore) to remove all the macro-zooplankton grazers. The water was then passed through a 0.2 µm filter (Millipore) and distributed into 500 ml acid-clean Polycarbonate bottles (Nalgene) at a different volumes. Subsequently, 60 µm-filtered water was gently added, to a final relative amount of pre-filtered seawater equivalent to 100%, 75%, 50%, 25%, 10% (v/v).

Rates of viral infection were estimated in a parallel experiment, following the same procedure but using 'virus free' seawater, i.e. filtered through a 30 kDa cassette (Millipore) as diluent. Nutrient supplements were not performed as an amendment of nutrient could possibly affect the bacterial growth and bias the experiment. The experiment bottles were then incubated in a 500 L deck incubator, constantly cooled with surface seawater and at the same incident light at the sea surface in order to replicate the conditions *in-situ*.

An aliquot (5 mL) of water was collected from each bottle and fixed with 0.1 % (w/v) final concentration of freshly-prepared paraformaldehyde, at two-time points: as soon as the dilution was performed and after 24 hours of incubation. After 15 min of fixation at 4˚C in the dark, the samples were flash frozen in liquid nitrogen and stored in -80˚C until further analysis. Grazing mortality, viral mortality and population turnover were inferred from regression curves of the instantaneous growth vs. dilution applied, as per the equation:

$\mu net = \mu - (G+V)*D$ (0.2 µm series)

$\mu net = (\mu - V) - G * D$ (30 kDa series)

Where µ net represents the net growth rate, µ is the instantaneous growth rate in absence of "predator", G = mortality due to grazing, V = mortality due to viruses and D represents the dilution applied to the experiments. An analysis of covariance was used to test the difference between the two regression

coefficients, a significant p-value would indicate a significant viral lysis or grazing rate.

### *Statistical analysis*

All the statistical analysis was performed in R (R Core Development Team, 2008), unless stated otherwise. Linear models followed by an analysis of variance were used to infer the significance of the correlation between mean FSC and cellular carbon content. The role of environmental factors on both the microbial distribution and their contribution to the carbon biomass was analysed by principal component analysis and Spearman correlation analysis of (the square root transformed) data presented as a Bray-Curtis similarity matrix.

Lastly, a generalized linear mixed model (lme4, R) was used to infer the specific response of picocyanobacteria cellular carbon biomass to the different environmental factors. Environmental factors were scaled and used as independent variables, while both depth range and the individual transect were used as grouping factors.

## 2.4 Results

The samples analysed were collected at the beginning of the Austral spring, during a research expedition on-board the RV *Investigator* off the East coast of Australia, in the region encompassed by the EAC and Tasman Sea. The sites sampled have a diverse range of physicochemical parameters (Suppl. Table 2.3), based on geographical location and depth at which the samples were collected. Results are reported for three depth intervals: 0m, 10-50 m, and 60-110m, respectively: surface, between the surface and the MLD, and the deep chlorophyll max (DCM).

Briefly, the surface seawater temperature decreased southwards, from 23°C to 17°C, and was higher in the EAC compared to adjacent sites nearer to the coast and offshore, with surface water sampled inside the EAC (CTD27-40) demonstrating the highest SST (23.43˚C). The concentration of macronutrients in surface waters generally increased southward, with a higher nutrient

concentration in the Tasman Sea compared to samples from the Coastal or EAC stations. Samples collected from the chlorophyll max layer (45-110 m) at near-shore stations of both the Transect North (TN) and middle (TM) displayed relatively high nutrient concentrations in comparison to other stations at the same depth (Suppl.Table. 2.1). This phenomenon is in line with other observations where wind driven exchange of water masses produces a deep upwelling and high nutrient load in near-shore waters, in regions encompassed by the EAC (Roughan and Middleton, 2002, 2004; Macdonald, Baird and Middleton, 2009).

Fluorescence was also high at depths between 10-40 and 50-110 m from TM and TN, supporting the hypothesis that these two specific stations had undergone a subsurface wind-driven upwelling event before sampling. Overall, sites in the Tasman Sea or Transect South (TS) presented the highest fluorescence and dissolved oxygen levels.

### *Picocyanobacteria specific Qc and variability across different oceanic regions*

In order to understand the specific contribution of each picophytoplankton group to the carbon biomass we analysed the cellular elemental carbon composition for each group. Cellular carbon quota was estimated for *Synechococcus*, *Prochlorococcus* and PPE populations sorted from six samples (1 coastal, 3 EAC and 2 Tasman) using an Influx flow cytometer, with a sort purity always > 95%. The six samples in triplicate yielded a total of approximately 150 million cells (*Synechococcus, Prochlorococcus* and PPE).

Picocyanobacteria Qc was highly correlated with the mean FSC ($r^2$= 0.90, $p < 0.01$) (Suppl. Fig 1), this significant correlation allowed us to use the coefficient of the regression curve to convert the mean FSC of each cyanobacteria population into their relative carbon quota.

While overall, the mean Qc (fg C $cell^{-1}$ ) followed a log-normal distribution (Fig. 2.2), environmental populations of both *Synechococcus* and *Prochlorococcus* displayed a broad spectrum of mean cellular carbon contents, which varied

according to their source. Cellular Qc values ranged from 170-590 fg C $cell^{-1}$and 46-400 fg C $cell^{-1}$ for *Synechococcus* and *Prochlorococcus* respectively.

To better understand how environmental variables, influence the plasticity of the cyanobacteria carbon quota a generalized linear mixed model was applied (lme4 in R) considering both the oceanic region and depth as grouping factors. The influence of temperature was significant for both *Synechococcus* and *Prochlorococcus* but with opposite effects. Temperature had a negative correlation with *Synechococcus* Qc ($E=-0.35$, $p < 0.005$) (Fig. 2.3), and a weak positive correlation with *Prochlorococcus* Qc (Table 2.1b).

*Synechococcus* Qc did not show any significant correlation with the concentration of different nutrients. However, *Prochlorococcus* was positively correlated with nitrate ($p < 0.008$) (Fig. 2.4). Depth had no significant effect on *Synechococcus* Qc but was positively correlated with *Prochlorococcus,* indicating that deeper populations of *Prochlorococcus* have higher carbon quota than those at the surface ($p < 0.005$) (Fig. 2.5).

The model applied revealed that changes in carbon quota for both *Synechococcus* and *Prochlorococcus* could be predicted by different combinations of parameters, NOx + Temperature + Phosphate for *Synechococcus* and depth and Temperature + Phosphate, NOx+ nitrite for *Prochlorococcus* (Table 2.1).

Because it was not possible to establish a direct correlation between the Qc and the mean FSC of populations of picoeukaryotes, a fixed value was used to convert the PPE cell numbers into carbon biomass. The fixed value used was based the elemental analysis of populations from different regions: 2276.67 +/- 332 fgC $cell^{-1}$, 3064.04 +/- 451 fgC $cell^{-1}$ and 2940.86 +/- 474.75 fgC $cell^{-1}$ for samples sorted from coastal, EAC and Tasman Sea respectively.

### *Picophytoplankton, bacteria, and viruses*

Within the subset of phytoplankton analysed in this study, i.e. those that could be analysed by flow cytometry (< 10 µm in dia.), cyanobacteria were always the most abundant phototrophs in terms of cell numbers and relative proportion of carbon in almost all the samples, from surface to DCM (Fig. 2.7).

*Prochlorococcus* dominated the surface phototrophic communities in warmer waters and lower latitudes, both along the EAC, the TN and TM, with an abundance up to an order of magnitude higher than in the Tasman Sea. *Synechococcus* was the most abundant phototrophs across TS and Tasman Sea (Suppl. Table 2.2).

Three distinct PPE groups were identified almost in all the samples, photosynthetic pico-eukaryotes (PPE), nano-eukaryotes (NEuk), and phycoerythrin rich eukaryotes (PE-Euk). PPE were always the most abundant small photosynthetic eukaryotes, followed by NEuk and lastly PE-Euk. The abundance of all three groups typically peaked at the local chlorophyll maximum depth. While temperature explains the distribution and abundance patterns of oxyphototrophic cyanobacteria, the cellular abundance of all three picoeukaryote groups correlated with higher concentrations of nitrite ($p < 0.05$), and negatively correlated with temperature ($p < 0.05$) (Fig. 2.11). Two groups of heterotrophic bacteria were resolved by cytometry of DNA-stained (SYBR Green I) samples, LDNA and HDNA bacteria showed a regional distribution with HDNA bacteria dominating the colder and nutrient rich water of TS, while the EAC and the stations from the TN and TM, were characterised by a higher number of LDNA bacteria. In coastal stations, the number of HDNA and LDNA cells were comparable at each depth. Virus populations also displayed distinct distribution patterns across the region; three viral populations were usually identified by flow cytometry (V1, V2, V3). V1, that represents the population of viruses with the lowest fluorescence intensity, likely representing the smaller viruses, was always the most abundant. The population peaked at coastal samples and also below the surface at the shore station on the transect North. V3, that represents the population with the highest intensity, in which are usually found larger viruses,

but also viruses that infect photosynthetic organism (Brussaard *et al.,* 2004), was usually one magnitude lower in number than V1 (Fig. 2.8).

### *The contribution of distinct microbial groups to carbon biomass*

Carbon quota per cell was used to convert the abundance of each different microbial group into an estimate to the carbon biomass. The total microbial carbon biomass was higher at surface for coastal samples and just below surface for both transect Middle and transect North, reaching almost 25mgC per $m^3$ of seawater (Fig. 2.10). Integrated depth (from 0-110m) results highlighted the relatively high contribution of the heterotrophic bacteria fraction to the carbon biomass throughout the water column. While in the EAC picocyanobacteria represented the main contributor to microbial carbon biomass, representing almost 42% of the biomass. Eukaryotic picophytoplankton represented the largest portion (41%) in the transect south. Even though they are not comparable in size, viruses also had a substantial role in contribution to carbon biomass (2-4%).

### *Grazing and Viral mortality on a WBC*

Rates of grazing and viral infection were inferred from regression analysis in the different dilution experiments. While both viral infection and grazing rates were highly variable within sites, both *Synechococcus* and *Prochlorococcus* showed a clear trend for higher viral infection along the EAC at this time of the year. *Prochlorococcus* mortality due to viral infection went from 15% to 199% of the standing stock in the EAC, while for samples collected outside the EAC, infection rates were relatively low, apart from one coastal site that recorded 114% of the standing stock lysed. *Synechococcus* instead showed more stable rates with an average of 2% of the standing stock lysed by viruses in samples from non EAC water and up to 85% lysed in samples from inside the EAC water.

Grazing rates were comparable between samples collected in the EAC or non-EAC region with rates for *Synechococcus* on average higher in the EAC. The rates of grazing on picoeukaryotes rates showed a similar trend and were similar

in the two water masses, on average 13% of the population grazed in the EAC and 16% in non-EAC. While grazing mortality was present at every site, viral mortality displayed higher variability, with values up to 140% in the EAC, and an average of 43% of the population infected in the EAC compared to the 30% in non-EAC. (Table 2.2)

## 2.5 Discussion

In light of predicted future climate change and increase in SST, it is of crucial importance to understand the contribution of picoplankton to the carbon biomass, to be able to predict the future productivity of specific oceanic regions. In the context of Australian Ocean, the ongoing strengthening of the EAC and its poleward extension towards the Southern Ocean is already responsible for the shift in distribution of pelagic fish (Last *et al.*, 2011), while the consequences that this trend has and will have for the organisms at the base of the food web is still not clear. In this study we showed that cyanobacteria are characterized by a variable cellular carbon quota, that ranged from 40 fgC to 590 fgC per cell. Cellular C quota varied according to depth and the region in which the samples were collected impacted on the relative contribution of these groups to the total carbon stock. This study highlights the importance of prokaryotic groups as a significant portion of carbon biomass in warm marine environments, such as the EAC, compared to more temperate regions, such as the Tasman Sea. Moreover, the dilution experiment emphasized the role of viruses for the carbon turnover, especially in the EAC, through a higher infection of primary producers. Suggesting the importance of the viral shunt, instead of the canonical grazing food web, for the nutrient recycling in the context of the EAC system.

### *Plasticity of carbon quota for picocyanobacteria*

*Synechococcus* and *Prochlorococcus* presented an average surface values for carbon quota, 278.08 ± 53.52 fgC cell$^{-1}$ and 89.02 ± 17.13 fgC cell$^{-1}$ respectively, comparable with previous studies in the North Atlantic (Casey *et al.*, 2013; Baer *et al.*, 2017). The significant increase of the Qc for *Prochlorococcus* with depth

(Fig. 2.6), is in line with the niche preference of different ecotypes. LL-adapted *Prochlorococcus*, usually found below 50m, have in general higher carbon content than the HL counterpart that thrives in surface water. In this study environmental variability seems to play a critical role in regulating the plasticity of picocyanobacterial Qc.

*Prochlorococcus* Qc seemed to respond positively to temperature increase (Martiny, Ma, *et al.*, 2016) but also to the concentration of nitrite. Concentrations of nitrite were low at surface but increase in depth, suggesting that *Prochlorococcus* carbon quota could be positively correlated with the depth of the nitracline (Durand, Olson and Chisholm, 2001). This result could be explained also by the niche preference of the different *Prochlorococcus* clade, nitrite can indeed be assimilated by all the LLI clade, and usually, their depth distribution is associated with a peak in nitrite concentration (Berube *et al.*, 2019). *Synechococcus* Qc was instead significantly higher in samples collected from the Tasman Sea (Fig. 2.6), and contrary to what reported in other study was negatively correlated with temperature ($p < 0.05$). Temperature has a central role in regulating the metabolisms and the biochemical processes of a cell. However, from the current knowledge, it is difficult to discern a consistent effect that temperature has on the elemental composition of picocyanobacteria (Moreno and Martiny, 2018). However, the registered difference in carbon content could be partially explained by the biogeography of the different clades of *Synechococcus*. Concomitant study on the spatial distribution of the main cyanobacteria ecotype (in prep.) highlighted the positive correlation of the distribution of *Synechococcus* clade I and clade IV ($p < 0.05$), known to be better adapted to cold temperature and coastal ecosystem, with higher carbon quota. In contrast, clade II, that is usually found in oligotrophic waters and is the predominant clade in the EAC, reported a negative correlation with an increase in carbon quota ($p < 0.05$).

All together these results suggest that physiological stressors, as change in environmental parameters, influence the elemental composition(Bertilsson *et al.*, 2003; Grob *et al.*, 2013; Garcia, Bonachela and Martiny, 2016; Lopez *et al.*, 2016; Martiny, Talarmin, *et al.*, 2016; Baer *et al.*, 2017; Moreno and Martiny, 2018) but,

that specific clade itself contributes differently to the variability of the carbon stock and eventually to the carbon cycle.

***Microbial distribution and their relative contribution to the carbon biomass: the importance of cyanobacteria in a warming ocean.***

This is the first comprehensive analysis of picoplankton for SE Australia, targeting not just the photosynthetic organism but also multiple groups of heterotrophic bacteria (LNA, HNA) and viruses (V1, V2, V3) that are likely to play an important role in the carbon cycle. The differential distribution of these groups across the EAC system reflected the physicochemical properties and depth of the region sampled. Cyanobacteria are known to be the dominant photosynthetic organisms in tropical/oligotrophic environments (Zwirglmaier *et al.*, 2008; Farrant *et al.*, 2016). In this study samples from the transect North and the EAC presented cyanobacteria counts consistent with similar data from the Coral Sea, an oligotrophic region of the NE Australia Ocean where the EAC develops (Moore *et al.*, 2019).

*Prochlorococcus,* that was the most abundant photosynthetic organism in almost all the samples, was an order of magnitude higher in the warmer station of the EAC compared to the Southern transect, respectively $2.5 \times 10^4 \pm 0.4$ cells/ml and $5.8 \times 10^3 \pm 1.6$ cells/ml ($p<0.05$). A coastal sample, in proximity to Sydney, showed the highest *Prochlorococcus* abundance for the whole water column (5-6 $\times 10^4$ cells/ml), even if the temperature was lower than the EAC. Previous work on the California Current has found a similar result, where a peak in *Prochlorococcus* count was linked to a temporally limited increase in nutrient concentration (Martiny, Talarmin, *et al.*, 2016). HDNA bacteria were more abundant than the LDNA in the colder water of the Tasman Sea compared to warmer water (Fig. 2.8) suggesting a shift in the main heterotrophic bacteria assemblages with an increase in temperature. Previous study formulated hypothesis of a shift to smaller bacteria population with increase in temperature (Moran *et al.*, 2015).

Understanding how absolute population numbers can translate into contribution to carbon biomass is of crucial importance in order to better interpret and be able to model the role of picoplankton in the marine carbon cycle. The choice of carbon conversion factor affects the carbon stock results. It is, therefore, important to accurately choose or calculate a carbon conversion factor that reflects the local population. Here we used direct estimation of the cellular carbon elemental quota for distinct functional groups to convert microbial numbers into a community estimate for the microbial fraction. There was a clear transition in the groups that contribute the most to microbial carbon biomass when comparing the EAC and the Tasman Sea, but also from near-shore to off-shore (Suppl. Fig.2.4). Picoeukaryotic algae contributes up to 50% of the carbon biomass in surface samples from the TrS and roughly 40% of the shore samples from TrM and TrN (Fig. 2.10). Prokaryotic cyanobacteria contributed more than 50% to the microbial biomass in surface samples collected within the EAC system (Suppl. Fig. 2.4b), and for offshore samples of the TrN. Similar importance of picoeukaryotes for nutrient rich coastal sites has been shown for the south coast of Australia and in the South Pacific (Patten, van Ruth and Rodriguez, 2018). While prokaryotic groups are the dominant contributors to the carbon biomass in oligotrophic areas of the Ocean.

### *Viral mortality in a WBC*

In the context of future productivity for the Australian Ocean it is important to understand the main pathway of energy transfer from primary producers through different links in the trophic chain. Here we used dilution experiments to understand the mortality rates due to grazing or viral lysis for picophytosynthetic communities of both the EAC and the Tasman Sea.

While the rates were not stable, there was an increase in mortality due to viral lysis in the EAC for picocyanobacteria. A similar trend has been observed in the North Atlantic Ocean with increased rates of viral lysis observed in tropical regions compared to higher latitudes (Mojica *et al.*, 2016). Viral lysis and grazing contribute in two different ways in the context of the food chain and eventually the carbon pump. Protistan grazers of primary producers act as trophic transfer

link to higher consumers (Pasulka, Samo and Landry, 2015). Viruses instead, though the viral-shunt, divert this pool of energy towards bacteria and hence, promote the remineralisation of potentially limiting nutrients (Fuhrman, 1999; Suttle, 2007) as well as bacteria biomass and respiration. A shift towards a predominant viral lysis pathway, due to a poleward expansion of the EAC towards the Tasman Sea, could have great consequences for the amount of carbon exported to the Ocean interior, but also reduce the amount of energy available for higher trophic level, and eventually, a change in the energy flow through the marine food web could impact the productivity of the region itself.

## 2.6 Conclusion

Future climate-induced changes in the marine environment are expected to affect species range distributions and abundance (McMahon *et al.*, 2015), and therefore to influence primary productivity, the connectivity of the food chain and associated energy and carbon flow (Moran *et al.*, 2016). Specific carbon elemental quota of the main primary producers directly impacts the trophic efficiency of the food web, and overall affect the sustainability of marine life. Here we showed that natural populations of *Synechococcus* and *Prochlorococcus* presented a high plasticity in their specific carbon content. Carbon quota changed both spatially and vertically over almost an order of magnitude. Numerous studies that looked at the contribution of picocyanobacteria to the carbon stock utilized a fixed value, extrapolated from literature, to convert their cyanobacteria count into biomass. While this approach would give a good basic understanding of the carbon stock, our results highlighted how, the use of fixed abundance to carbon conversion factors, could introduce error in term of underestimation/overestimation of cyanobacteria role in the total carbon biomass. Eventually, this error could led to an incorrect model prediction of the role and importance of picocyanobacteria for the carbon pump in a future climate change scenarios (Kwiatkowski *et al.*, 2018). Moreover, this work highlighted the importance of prokaryotic primary producers in a warmer environment, concomitantly with a shift from larger bacteria predominant in the Tasman Sea to

smaller bacteria in the warmer EAC. The East coast of Australia has undergoing significant warming over the past few decades, as it has been reported from a fifty years SST monitoring. SST warming and a potential strengthening of the EAC could lead to a restructuring of the marine microbial food web, with a shift from a picoeukaryote dominated primary production, typical of the Tasman Sea, to a photosynthetic prokaryotic primary production with the increasing importance of the viral shunt for energy flux.

## 2.7 Acknowledgements

This work was supported by Australian Research Council awards FL140100021 to IP. This work was also supported by the Marine National Facility and we thank the Chief Scientist Martina Doblin and officers and crew of the *R/V Investigator* during cruise IN16_v04. We would like to thank the numerous crew members and scientists involved in the generation, curation, and deposition of IMOS oceanographic data. We thank the Bigelow laboratory for the analysis of the carbon content.

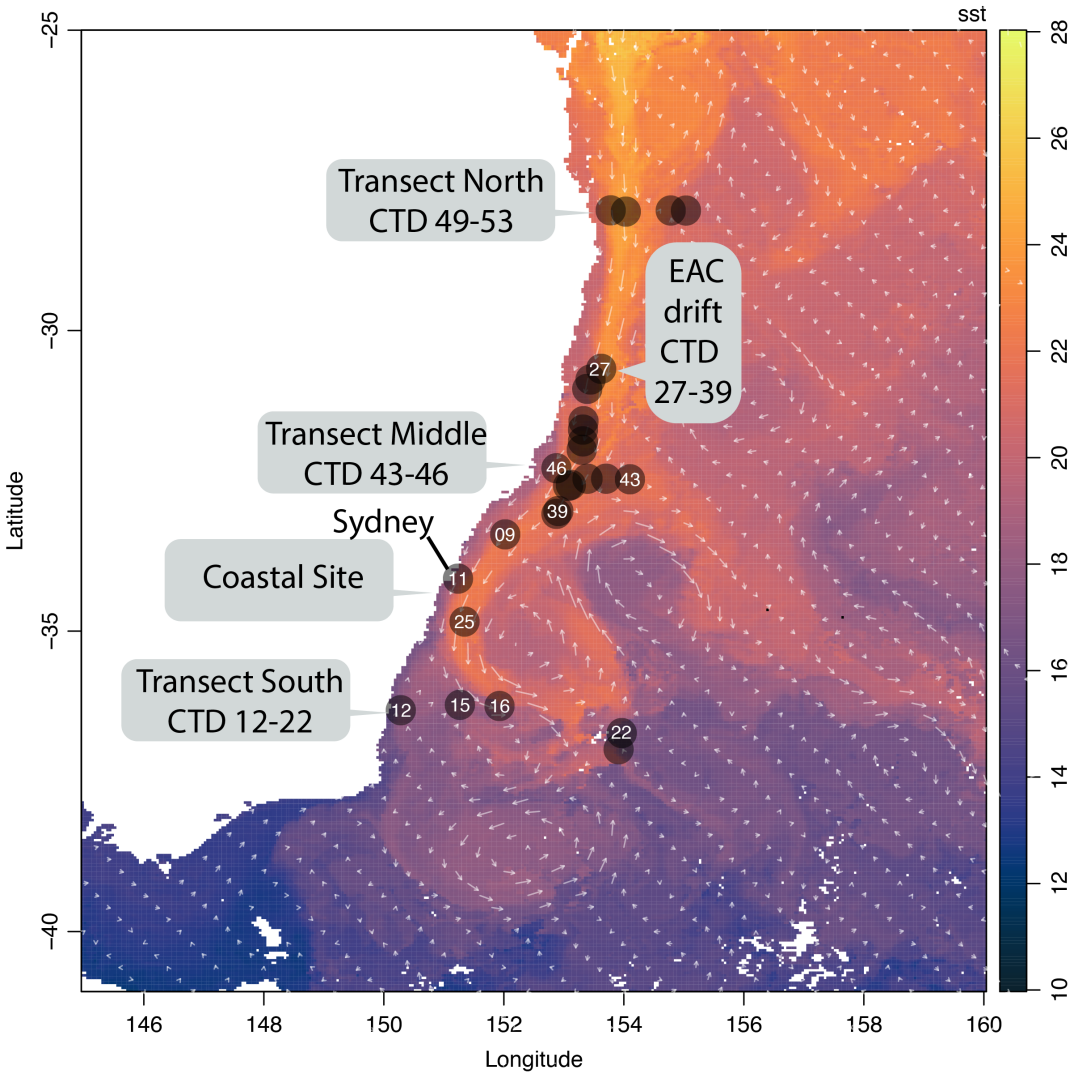



**Figure 2.1**. *Map of the different transects and sampling location. Seawater colour is relative to the monthly composite average surface seawater temperature on a 4km resolution. Data was obtained from the AODN portal for the month of September 2016. White arrows are relative to the main surface current direction. Dots are relative to the sampling sites, CTD deployment number is highlighted in some of the samples and in the transect box.*

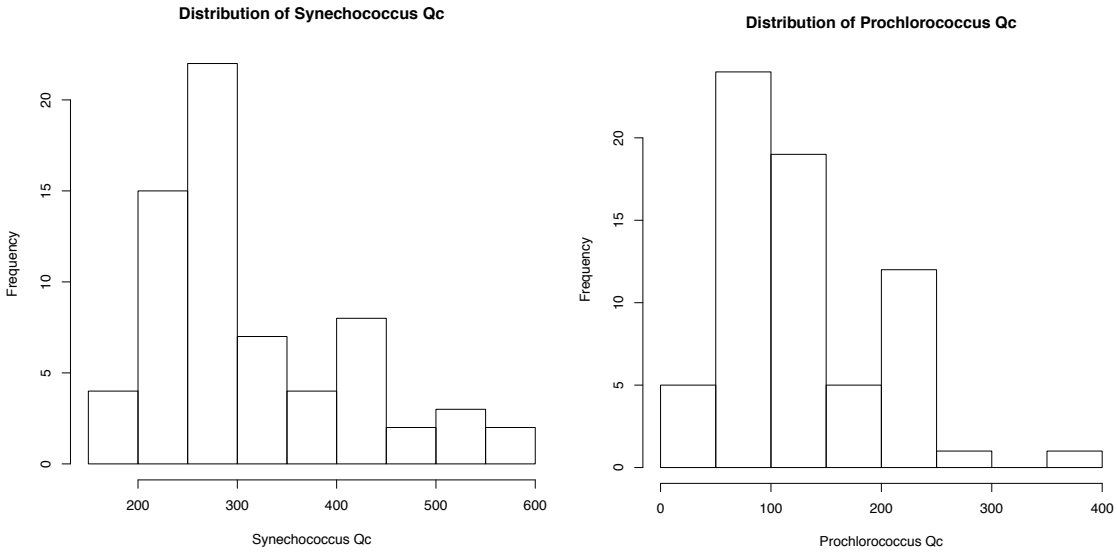



**Figure .2.2** *Frequency distribution of Synechococcus (left) and Prochlorococcus (right) relative carbon quota per cell (FgC/cell).*

a)

*Qc_Synechococcus*

| Predictors | CI | p | |
|---|---|---|---|
| Temperature_s | 2.17 – 3.26 | <0.001 | *** |
| Nitrite_s | 0.04 – 0.15 | <0.001 | *** |
| NOx | 0.01 – 0.09 | <0.001 | *** |
| Temperature_s:Nitrite_s | 0.10 – 0.21 | <0.001 | *** |
| Temperature_s:Phosphate | 0.25 – 0.77 | 0.004 | ** |
| Nitrite_s:Phosphate | 0.02 – 0.14 | <0.001 | *** |
| Temperature_s:Nox | 0.01 – 0.09 | <0.001 | *** |
| Nitrate_s:Nox | 0.00 – 0.07 | <0.001 | *** |
| Temperature_s:Nitrite_s:Phosphate | 0.16 – 0.86 | 0.021 | ** |
| Temperature_s:Nitrite_s:Nitrate_s | 0.19 – 0.50 | <0.001 | *** |
| Nitrite_s:Phosphate:Nitrate_s | 0.04 – 0.62 | 0.008 | ** |
| Temperature_s:Phosphate:Nox | 2.28 – 5.29 | <0.001 | *** |
| Temperature_s:Nitrate_s:Nox | 0.12 – 0.87 | 0.025 | ** |
| Temperature_s:Nitrite_s:Nitrate_s:Nox | 0.11 – 0.65 | 0.003 | ** |
| Temperature_s:Phosphate:Nitrate_s:Nox | 0.69 – 0.97 | 0.021 | ** |
| Nitrite_s:Phosphate:Nitrate_s:Nox | 0.49 – 0.86 | 0.002 | ** |
| Temperature_s:Nitrite_s:Phosphate:Nitrate_s:Nox | 0.73 – 0.88 | <0.001 | *** |

b)

*Qc_Prochlorococcus*

| Predictors | CI | p | |
|---|---|---|---|
| Phosphate | 0.16 – 0.82 | 0.015 | ** |
| Nitrite_s | 0.17 – 1.00 | 0.049 | * |
| Nitrate_s | 4.44 – 732.90 | 0.002 | ** |
| Phosphate:Temperature_s | 2.82 – 15.15 | <0.001 | *** |
| NOx:Nitrate_s | 0.01 – 0.51 | 0.009 | ** |
| Temperature_s:Nitrate_s | 0.02 – 0.77 | 0.026 | * |
| Phosphate:NOx:Temperature_s | 0.18 – 0.85 | 0.018 | ** |
| Phosphate:NOx:Nitrate_s | 0.36 – 0.86 | 0.008 | ** |
| Phosphate:Temperature_s:Nitrate_s | 1.11 – 109.15 | 0.041 | * |
| Phosphate:Nitrite_s:Nitrate_s | 3.15 – 209.25 | 0.002 | ** |
| NOx:Nitrite_s:Nitrate_s | 0.00 – 0.15 | <0.001 | *** |
| Phosphate:NOx:Nitrite_s:Nitrate_s | 1.36 – 3.29 | 0.001 | ** |
| Phosphate:NOx:Temperature_s:Nitrite_s:Nitrate_s | 1.03 – 1.43 | 0.024 | * |

**Table 2.1** *Most important predictors for Synechococcus (a), Prochlorococcus (b) carbon quota with the confidence interval for each predictor and the relative p value calculated with a generalized mixed model analysis in R (lme4).*

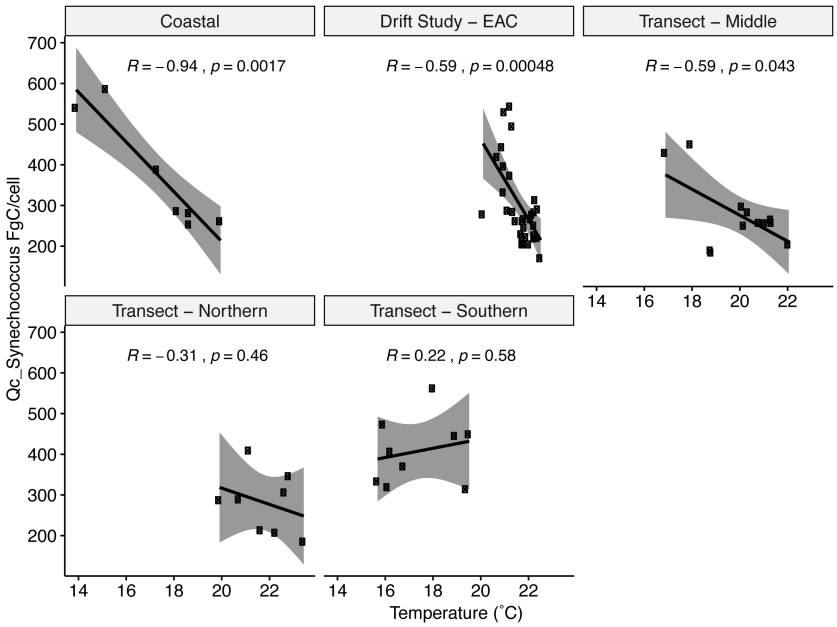



**Figure 2.3** *Linear regression of Synechococcus Qc at different temperature in the different provinces. Coefficient of the regression and relative p value on the top corner of each plot.*

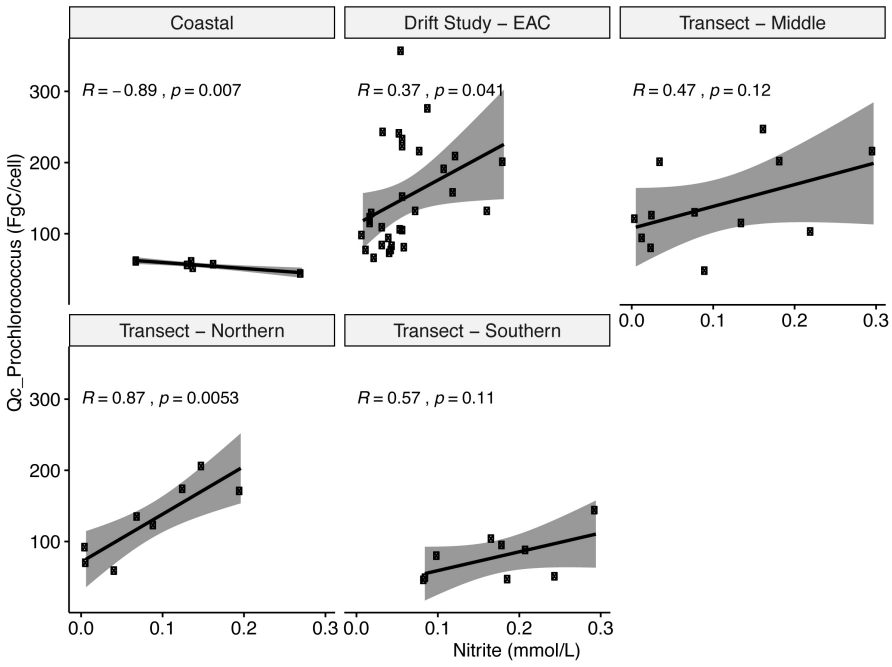



**Figure 2.4** *Linear regression of Prochlorococcus Qc at different concentration of nitrite in the different provinces. Coefficient of the regression and relative p value on the top corner of each plot.*

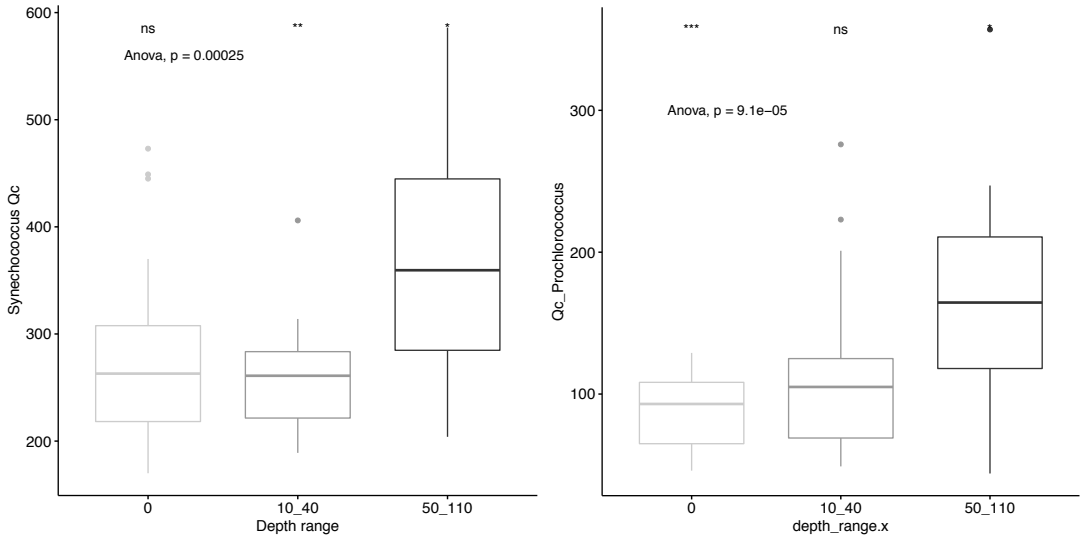



**Figure. 2.5** *Average depth specific carbon quota for Synechococcus (left) and Prochlorococcus (right) (*** p< 0.001, ** p<0.01, * p<0.05). Carbon quota is expressed in FgC/cell*

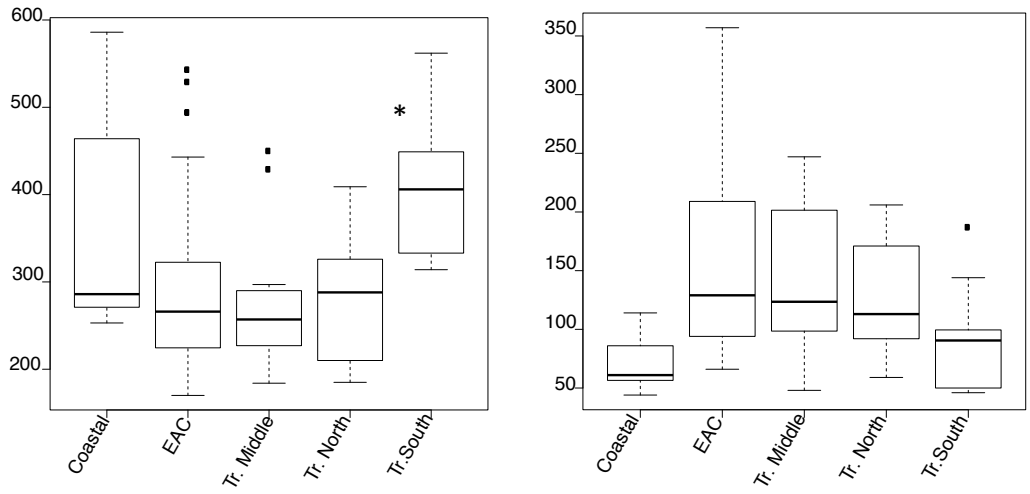



**Figure 2.6** *Integrated average carbon content at the different site for Synechococcus (a) and Prochlorococcus (b). Carbon quota is expressed in FgC/cell*

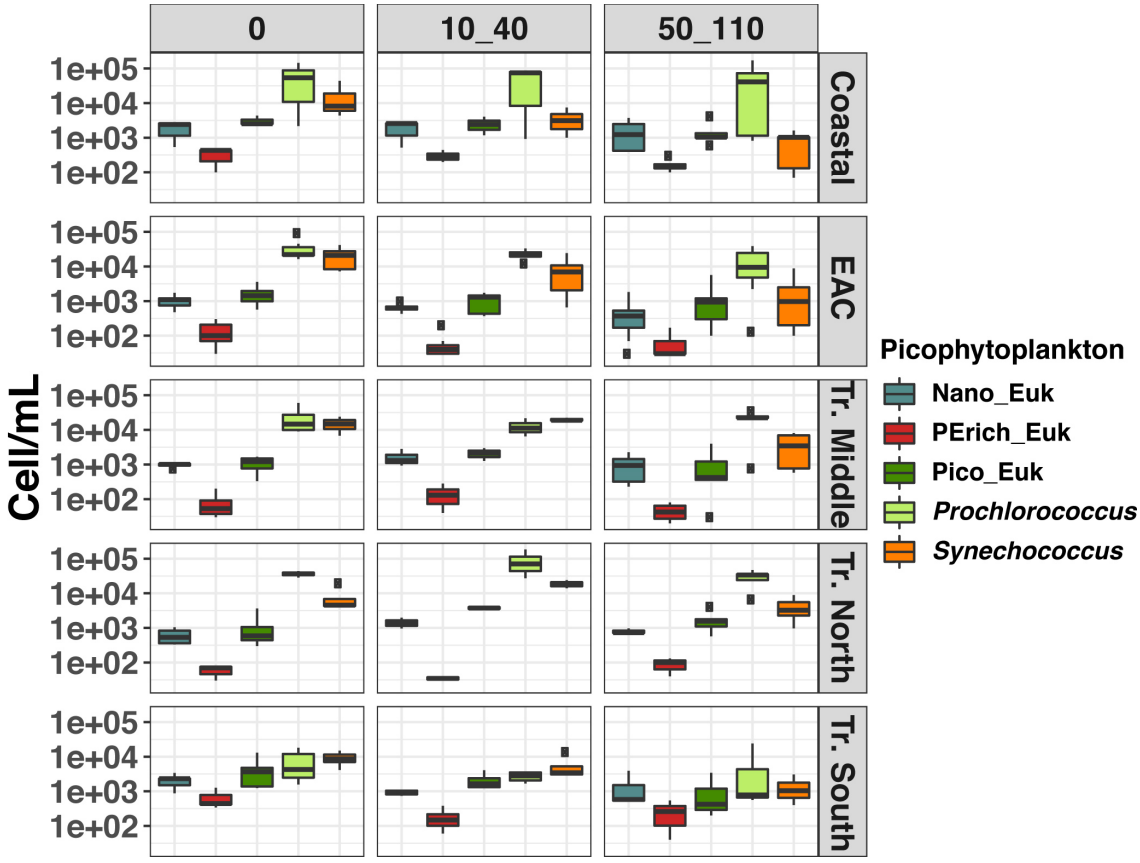



**Figure 2.7** *Relative abundance and distribution of the photosynthetic communities in the different transect at the depth range sampled.*

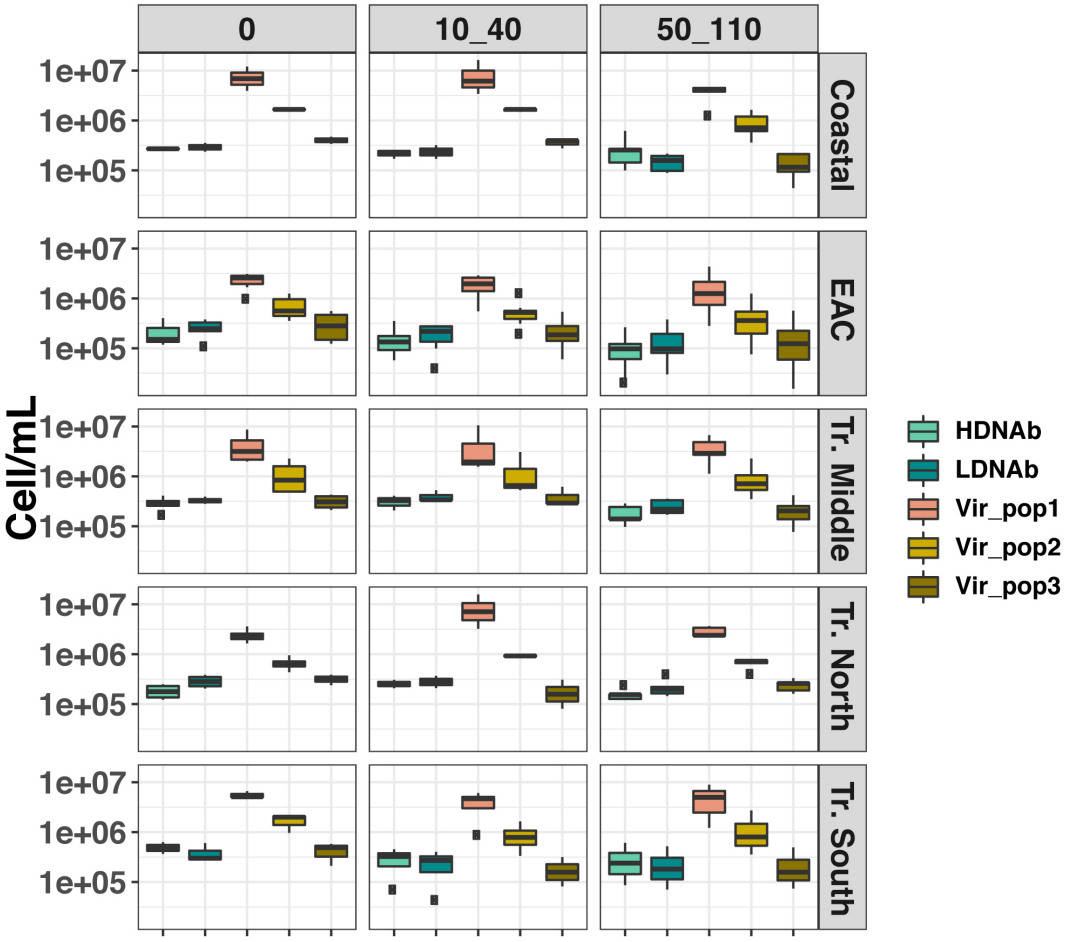



**Figure. 2.8** *Relative abundance of the different viral (Vir_pop1,Vir_pop2,Vir_pop3) and heterotrophic bacteria (HDNAb, LDNAb) groups at different depth along the sampled transect.*

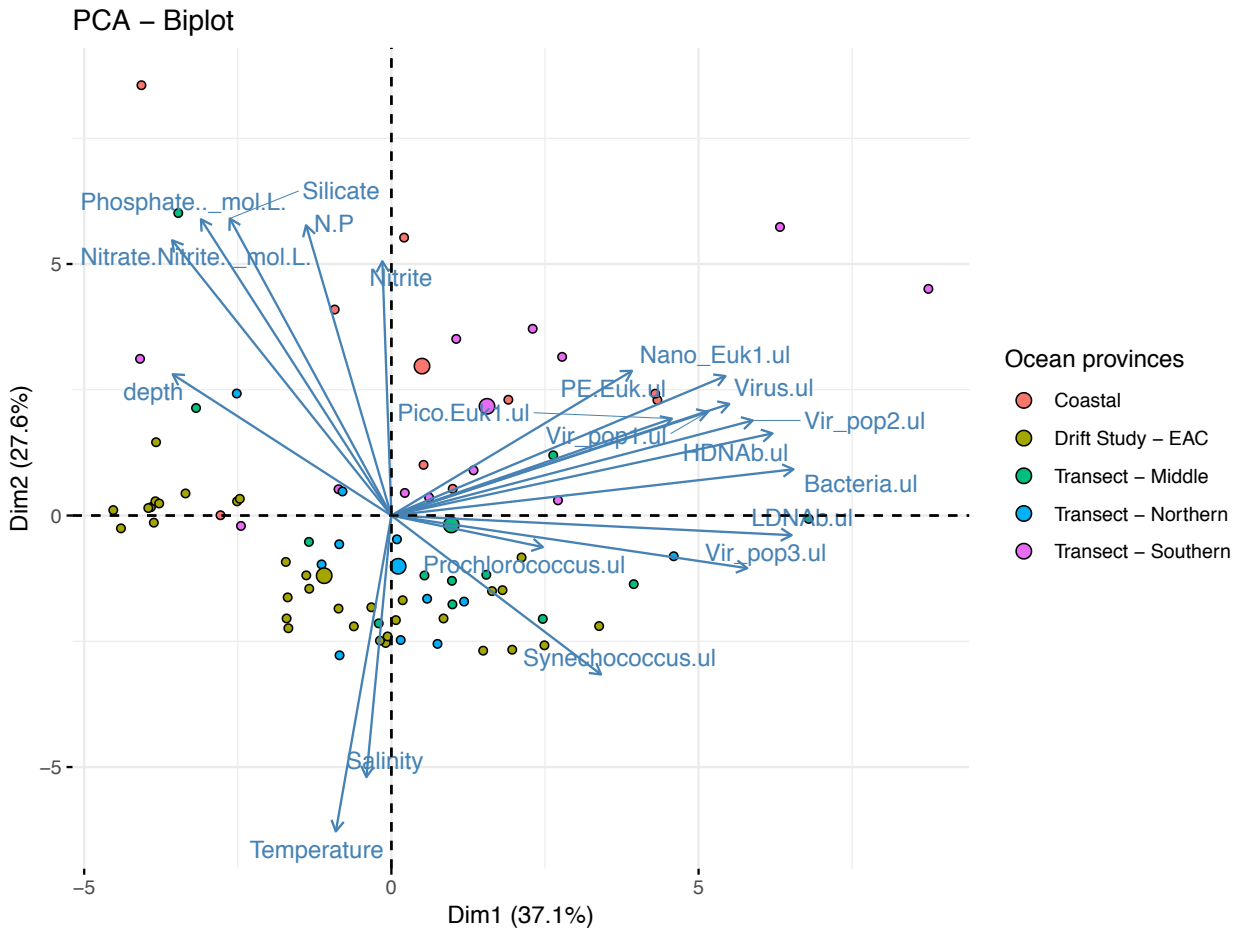



**Figure. 2.9** *Principal component analysis. Biplot of the relationship between photosynthetic groups distribution and environmental parameters. The first axis (x) and second axis (y) respectively explained 37.1% and 27.6% of the results. Samples are colour coded based on Oceanic provinces. Dot size is based on the importance of the specific site in driving the dissimilarity.*

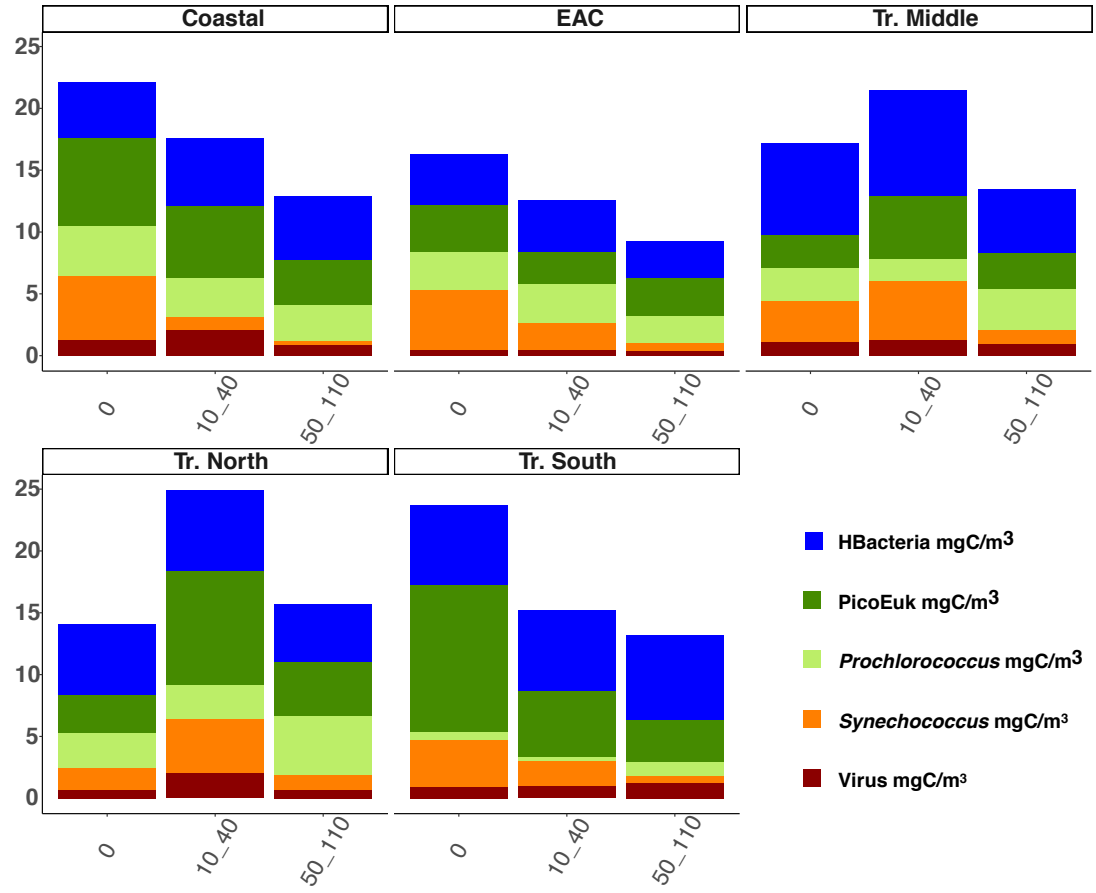



**Figure 2.10** *Relative contribution of the five picoplankton (Hbacteria, Viruses, PicoEuk, Synechococcus and Prochlorococcus) groups to picoplankton carbon biomass expressed in mgC/m$^3$ of seawater in the different transect sampled at different depth range (0, 10-40, 50-110).*

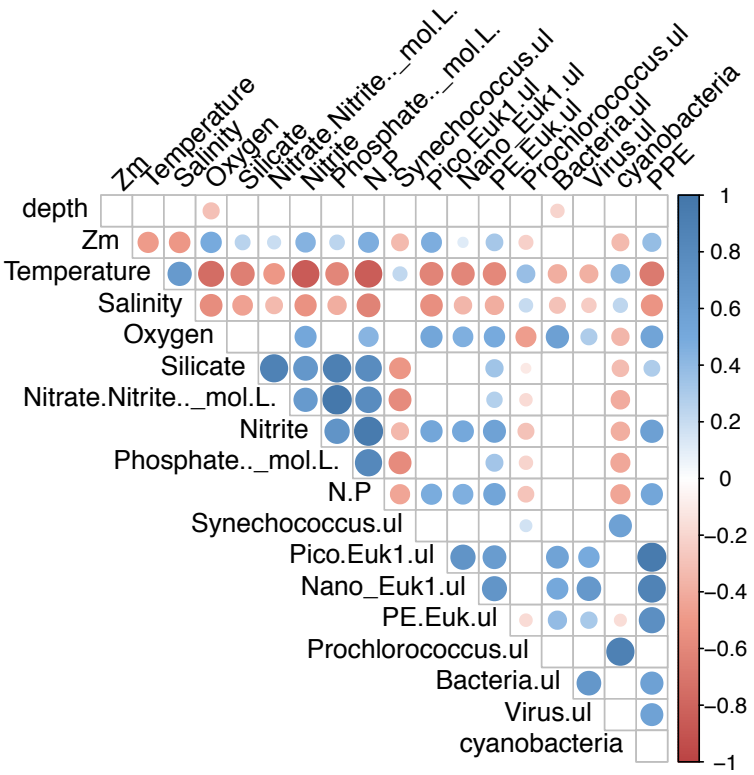



**Figure 2.11** *Plot of the correlation between the environmental variable and the microbial groups based on counts from all the samples calculated on R. The dot diameter is related to the degree of correlation, while the colour is based on a positive (blue) or negative (red) correlation. All the correlation presents are statistically significant ($p < 0.05$).*

| | *Synechococcus* | | | | | *Prochlorococcus* | | | | | Piko Euk | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | cell/ml | μ_tot | Tot. M | G | V | cell/ml | μ_tot | Tot. M | G | V | cell/ml | μ_tot | Tot. M | G | V |
| CTD9 | 149000 | -0.53 | -0.17 | -0.17 | 0 | 37400 | 2.52 | -2.72 | -1.39 | -1.33 | 14150 | 1.46 | -2.91 | -0.211 | -2.699 |
| CTD11 | 13000 | 0.36 | -0.04 | -0.014 | -0.026 | 3700 | 0.0292 | 0 | 0 | 0 | 18740 | -0.33 | 0 | 0 | 0 |
| CTD13 | 35100 | 0.13 | 0 | 0 | 0 | 16700 | -0.4 | 0 | 0 | 0 | 3740 | -0.46 | 0 | 0 | 0 |
| CTD15 | 14600 | 0.28 | -0.06 | -0.06 | 0 | 13600 | -0.28 | -0.62 | 0 | -0.62* | 2207 | 0.052 | -0.123 | -0.123 | 0 |
| CTD22 | 4100 | 0.04 | 0 | 0 | NA | 2540 | 0.11 | -0.23 | -0.23 | NA | 3233 | 0.59 | 0 | 0 | NA |
| CTD27 | 33600 | -0.04 | -0.64 | 0 | -0.64* | 25000 | -0.2 | 0 | 0 | 0 | 3430 | 0.2 | -0.49 | 0 | -0.49 |
| CTD28 | 36700 | -0.25 | -0.209 | -0.209 | NA | 27600 | 0.03 | 0 | 0 | NA | 2690 | -0.37 | -0.44 | -0.44 | NA |
| CTD30 | 29000 | -0.16 | -0.033 | -0.033 | 0 | 97500 | | 0 | 0 | 0 | 2163 | -0.47 | -0.22 | -0.22 | - |
| CTD32 | 39600 | 1.46 | -3.079 | -0.97 | -2.11* | 29600 | 1.07 | -2.66 | -0.14 | -2.52* | 5330 | -0.03 | -1.05 | -1.05 | NA |
| CTD35 | 9760 | 0.51 | -0.02 | -0.02 | 0 | 22000 | -0.39 | 0 | 0 | 0 | 3503 | 0.2 | -0.19 | -0.19 | 0 |
| CTD37 | 42600 | 0.67 | -0.81 | -0.33 | -0.48 | 6350 | 0.61 | -1.86 | -0.17 | -1.69* | 540 | -0.23 | -0.35 | 0 | -0.35 |
| CTD42 | 15300 | 0.33 | -0.2 | 0 | -0.2* | 10800 | 0.5 | -0.68 | 0 | -0.68* | 2433 | 0.093 | -0.04 | -0.04 | 0 |
| CTD48 | 6520 | 0.5 | -0.34 | -0.19 | -0.45* | 29300 | 0.38 | -0.22 | -0.17 | -0.05 | 1120 | 1.27 | -0.61 | -0.015 | -0.59* |

**Table. 2.2** *Grazing (G) and Viral (V) mortality rates for Synechococcus , Prochlorococcus and Pico_Euk. * represents samples for which viral lysis was significantly different from grazing mortality (pvalue ancova < 0.05). u.tot represents the instantaneous growth rate calculated from the viral dilution (30kDa), in 2 experiments pico eukatyotes presented negative growth in the viral dilution and positive in the grazing experiment (yellow).*

## 2.8 References


Alvain, S. *et al.* (2005) 'Remote sensing of phytoplankton groups in case 1 waters from global SeaWiFS imagery', *Deep-Sea Research Part I: Oceanographic Research Papers*, 52(11), pp. 1989–2004. doi: 10.1016/j.dsr.2005.06.015.

Archer, M. R. *et al.* (2017) 'On the Variability of the East Australian Current: Jet Structure, Meandering, and Influence on Shelf Circulation', *Journal of Geophysical Research: Oceans*, 122(11), pp. 8464–8481. doi: 10.1002/2017JC013097.

Azam, F. (1998) 'Microbial Control of Oceanic Carbon Flux: The Plot Thickens', *Science*, 280(5364), pp. 694 LP – 696. doi: 10.1126/science.280.5364.694.

Baer, S. E. *et al.* (2017) 'Stoichiometry of Prochlorococcus, Synechococcus, and small eukaryotic populations in the western North Atlantic Ocean', *Environmental Microbiology*, 19(4), pp. 1568–1583. doi: 10.1111/1462-2920.13672.

Behrenfeld, M. J. and Boss, E. S. (2018) 'Student's tutorial on bloom hypotheses in the context of phytoplankton annual cycles', *Global Change Biology*, 24(1), pp. 55–77. doi: 10.1111/gcb.13858.

Bertilsson, S. *et al.* (2003) 'Elemental composition of marine Prochlorococcus and Synechococcus : Implications for the ecological stoichiometry of the sea', *Limnology and Oceanography*, 48(5), pp. 1721–1731. doi: 10.4319/lo.2003.48.5.1721.

Berube, P. M. *et al.* (2019) 'Emergence of trait variability through the lens of nitrogen assimilation in Prochlorococcus', *eLife*, 8, pp. 1–28. doi: 10.7554/elife.41043.

Biller, S. J. *et al.* (2015) 'Prochlorococcus: The structure and function of collective diversity', *Nature Reviews Microbiology*. Nature Publishing Group, 13(1), pp. 13–27. doi: 10.1038/nrmicro3378.

Bopp, L. *et al.* (2015) *The Ocean a carbon pump*, *ocean-climate*.

Brown, M. V. *et al.* (2018) 'Data Descriptor: Systematic, continental scale temporal monitoring of marine pelagic microbiota by the Australian Marine Microbial Biodiversity Initiative', *Scientific Data*, 5(April), pp. 1–10. doi: 10.1038/sdata.2018.130.

Brussaard, C. P. D. (2004) 'Optimization of procedures for counting viruses by flow cytometry', *Applied and environmental microbiology*. American Society for Microbiology, 70(3), pp. 1506–1513. doi: 10.1128/AEM.70.3.1506-1513.2004.

Buitenhuis, E. T. *et al.* (2012) 'Picophytoplankton biomass distribution in the global ocean', *Earth System Science Data*, 4(1), pp. 37–46. doi: 10.5194/essd-4-37-2012.

Caron, D. A. *et al.* (2017) 'Planktonic food web structure at a coastal time-series site: I. Partitioning of microbial abundances and carbon biomass', *Deep-Sea Research Part I: Oceanographic Research Papers*, 121(January), pp. 14–29. doi: 10.1016/j.dsr.2016.12.013.

Casey, J. R. *et al.* (2013) 'Changes in partitioning of carbon amongst photosynthetic pico- and nano-plankton groups in the Sargasso Sea in response to changes in the North Atlantic Oscillation', *Deep-Sea Research Part II: Topical Studies in Oceanography*. Elsevier, 93, pp. 58–70. doi: 10.1016/j.dsr2.2013.02.002.

Durand, M. D., Olson, R. J. and Chisholm, S. W. (2001) 'Phytoplankton population dynamics at the Bermuda Atlantic Time-series station in the Sargasso Sea', *Deep-Sea Research Part II: Topical Studies in Oceanography*, 48(8–9), pp. 1983–2003. doi: 10.1016/S0967-0645(00)00166-1.

Evans, C., Pearce, I. and Brussaard, C. P. D. (2009) 'Viral-mediated lysis of microbes and carbon release in the sub-Antarctic and polar frontal zones of the Australian Southern Ocean', *Environmental Microbiology*, 11(11), pp. 2924–2934. doi: 10.1111/j.1462-2920.2009.02050.x.

Everett, J. D. *et al.* (2012) 'An avenue of eddies: Quantifying the biophysical properties of mesoscale eddies in the Tasman Sea', *Geophysical Research Letters*, 39(16), pp. 1–5. doi: 10.1029/2012GL053091.

Farrant, G. K. *et al.* (2016) 'Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria', *Proceedings of the National Academy of Sciences*. National Academy of Sciences, 113(24), pp. E3365–E3374. doi: 10.1073/PNAS.1524865113.

Field, C. B. *et al.* (1998) 'Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components', *Science*, 281(5374), pp. 237 LP – 240. doi: 10.1126/science.281.5374.237.

Flombaum, P. *et al.* (2013) 'Present and future global distributions of the marine Cyanobacteria Prochlorococcus and Synechococcus', *Proceedings of the National Academy of Sciences*, 110(24). doi: 10.1073/pnas.1307701110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1307701110.

Fuhrman, J. A. (1999) 'Marine viruses and their biogeochemical and ecological effects', *Nature*, 399, pp. 1–8. doi: 10.1038/21119.

Garcia, N. S., Bonachela, J. A. and Martiny, A. C. (2016) 'Interactions between growth-dependent changes in cell size, nutrient supply and cellular elemental stoichiometry of marine Synechococcus', *ISME Journal*. Nature Publishing Group, 10(11), pp. 2715–2724. doi: 10.1038/ismej.2016.50.

Giering, S. L. C. *et al.* (2014) 'Reconciliation of the carbon budget in the ocean's twilight zone', *Nature*. Nature Publishing Group, 507(7493), pp. 480–483. doi: 10.1038/nature13123.

Graff, J. R. *et al.* (2015) 'Analytical phytoplankton carbon measurements spanning diverse ecosystems', *Deep-Sea Research Part I: Oceanographic Research Papers*. Elsevier, 102, pp. 16–25. doi: 10.1016/j.dsr.2015.04.006.

Graff, J. R., Milligan, A. J. and Behrenfeld, M. J. (2012) 'The measurement of

phytoplankton biomass using flowcytometric sorting and elemental analysis of carbon', *Limnology and Oceanography: Methods*, 10(NOVEMBER), pp. 910–920. doi: 10.4319/lom.2012.10.910.

Grob, C. *et al.* (2013) 'Elemental composition of natural populations of key microbial groups in Atlantic waters', *Environmental Microbiology*, 15(11), pp. 3054–3064. doi: 10.1111/1462-2920.12145.

Guidi, L. *et al.* (2016) 'Plankton networks driving carbon export in the oligotrophic ocean', *Nature*. Nature Publishing Group, 532(7600), pp. 465–470. doi: 10.1038/nature16942.

Gundersen, K. *et al.* (2002) 'Elemental C, N, and P cell content of individual bacteria collected at the Bermuda Atlantic Time-series Study (BATS) site', *Limnology and Oceanography*, 47(5), pp. 1525–1530. doi: 10.4319/lo.2002.47.5.1525.

Heldal, M. *et al.* (2003) 'Elemental composition of single cells of various strains of marine Prochlorococcus and Synechococcus using X-ray microanalysis', *Limnology and Oceanography*, 48(5), pp. 1732–1743.

Jiao, N. *et al.* (2010) 'Microbial production of recalcitrant dissolved organic matter: Long-term carbon storage in the global ocean', *Nature Reviews Microbiology*. Nature Publishing Group, 8(8), pp. 593–599. doi: 10.1038/nrmicro2386.

Kane, T. J. O., Oke, P. R. and Sandery, P. A. (2011) 'Predicting the East Australian Current', *Ocean Modelling*. Elsevier Ltd, 38(3–4), pp. 251–266. doi: 10.1016/j.ocemod.2011.04.003.

Kirchman, D. L. (2016) 'Growth Rates of Microbes in the Oceans', *Annual Review of Marine Science*, 8, pp. 285–309.

Kwiatkowski, L. *et al.* (2018) 'The Impact of Variable Phytoplankton Stoichiometry on Projections of Primary Production, Food Quality, and Carbon Uptake in the Global Ocean', *Global Biogeochemical Cycles*, 32(4), pp. 516–528. doi:

10.1002/2017GB005799.

Landry, M. R. and Hassett, R. P. (1982) 'Landry and Hassett', *Marine biology*, 67, pp. 283–288. doi: 10.1007/BF00397668.

Landry, M. R., Kirshtein, J. and Constantinou, J. (1995) 'A refined dilution technique for measuring the community grazing impact of microzooplankton', *Marine Ecology Progress Series*, 120(1), pp. 53–63. doi: 10.3354/meps120053.

Last, P. R. *et al.* (2011) 'Long-term shifts in abundance and distribution of a temperate fish fauna : A response to climate change and fishing practices Long-term shifts in abundance and', *Global ecology and biogeography*, 20(April), pp. 58–72. doi: 10.1111/j.1466-8238.2010.00575.x.

Lopez, J. S. *et al.* (2016) 'Diel variability in the elemental composition of the marine cyanobacterium synechococcus', *Journal of Plankton Research*, 38(4), pp. 1052–1061. doi: 10.1093/plankt/fbv120.

Macdonald, H. S., Baird, M. E. and Middleton, J. H. (2009) 'Effect of wind on continental shelf carbon fluxes off southeast Australia: A numerical model', *Journal of Geophysical Research: Oceans*, 114(5), pp. 1–17. doi: 10.1029/2008JC004946.

Marie, D. *et al.* (1997) 'Enumeration and Cell Cycle Analysis of Natural Populations of Marine Picoplankton by Flow Cytometry …', *Applied and Environmental Microbiology*, 63(1), pp. 186–193.

Martiny, A. C. *et al.* (2013) 'Strong latitudinal patterns in the elemental ratios of marine plankton and organic matter', *Nature Geoscience*. Nature Publishing Group, 6(4), pp. 279–283. doi: 10.1038/ngeo1757.

Martiny, A. C., Talarmin, A., *et al.* (2016) 'Biogeochemical interactions control a temporal succession in the elemental composition of marine communities', *Limnology and Oceanography*, 61(2), pp. 531–542. doi: 10.1002/lno.10233.

Martiny, A. C., Ma, L., *et al.* (2016) 'Interactions between thermal acclimation,

growth rate, and phylogeny influence prochlorococcus elemental stoichiometry', *PLoS ONE*, 11(12), pp. 1–12. doi: 10.1371/journal.pone.0168291.

McMahon, K. W. *et al.* (2015) 'Millennial-scale plankton regime shifts in the subtropical North Pacific Ocean', *Science*, 350(6267), pp. 1530–1533. doi: 10.1126/science.aaa9942.

Mojica, K. D. A. *et al.* (2016) 'Latitudinal variation in virus-induced mortality of phytoplankton across the North Atlantic Ocean', *ISME Journal*. Nature Publishing Group, 10(2), pp. 500–513. doi: 10.1038/ismej.2015.130.

Moore, L. R. *et al.* (2019) 'Unicellular Cyanobacteria Are Important Components of Phytoplankton Communities in Australia's Northern Oceanic Ecoregions', *Frontiers in Microbiology*, 9(January), pp. 1–16. doi: 10.3389/fmicb.2018.03356.

Moran, A. G. *et al.* (2015) 'More , smaller bacteria in response to ocean ' s warming ?', *Proc. R. Soc*, 282, pp. 1–9. doi: http://dx.doi.org/10.1098/rspb.2015.0371.

Moran, M. A. *et al.* (2016) 'Deciphering ocean carbon in a changing world', *Proceedings of the National Academy of Sciences*, 113(12), pp. 3143–3151. doi: 10.1073/pnas.1514645113.

Moreno, A. R. and Martiny, A. C. (2018) 'Ecological Stoichiometry of Ocean Plankton', *Annual Review of Marine Science*, 10(1), pp. 43–69. doi: 10.1146/annurev-marine-121916-063126.

Pasulka, A. L., Samo, T. J. and Landry, M. R. (2015) 'Grazer and viral impacts on microbial growth and mortality in the southern California Current Ecosystem', *Journal of Plankton Research*, 37(2), pp. 320–336. doi: 10.1093/plankt/fbv011.

Patten, N. L., van Ruth, P. D. and Rodriguez, A. R. (2018) 'Spatial variability in picophytoplankton, bacteria and viruses in waters of the Great Australian Bight (southern Australia)', *Deep-Sea Research Part II: Topical Studies in Oceanography*. Elsevier Ltd, pp. 1–12. doi: 10.1016/j.dsr2.2018.04.009.

Poloczanska, E. S. (2013) ‘Global imprint of climate change on marine life’, *Nature Climate Change*, 3(August), pp. 919–925. doi: 10.1038/nclimate1958.

R Core Development Team (2008) *Computational Many-Particle Physics*. Edited by H. Fehske, R. Schneider, and A. Weiße. Berlin, Heidelberg: Springer Berlin Heidelberg (Lecture Notes in Physics). doi: 10.1007/978-3-540-74686-7.

Ridgway, K. R. (1997) ‘Seasonal cycle of the East Australian Current’, *Journal of Geophysical Research C: Oceans*, 102(C10), pp. 22921–22936. doi: 10.1029/97JC00227.

Roughan, M. and Middleton, J. H. (2002) ‘A comparison of observed upwelling mechanisms off the east coast of Australia’, *Continental Shelf Research*, 22(17), pp. 2551–2572. doi: 10.1016/S0278-4343(02)00101-2.

Roughan, M. and Middleton, J. H. (2004) ‘On the East Australian Current: Variability, encroachment, and upwelling’, *Journal of Geophysical Research C: Oceans*, 109(7). doi: 10.1029/2003JC001833.

Shalapyonok, A., Olson, R. J. and Shalapyonok, L. S. (2001) ‘Arabian Sea phytoplankton during Southwest and Northeast Monsoons 1995: Composition, size structure and biomass from individual cell properties measured by flow cytometry’, *Deep-Sea Research Part II: Topical Studies in Oceanography*, 48(6–7), pp. 1231–1261. doi: 10.1016/S0967-0645(00)00137-5.

Suthers, I. M. *et al.* (2011) ‘The strengthening East Australian Current, its eddies and biological effects - an introduction and overview’, *Deep-Sea Research Part II: Topical Studies in Oceanography*, 58(5), pp. 538–546. doi: 10.1016/j.dsr2.2010.09.029.

Suttle, C. A. (2005) ‘Viruses in the sea’, *Nature*, 437. doi: 10.1111/j.1574-6976.2008.00132.x.

Suttle, C. A. (2007) ‘Marine viruses - Major players in the global ecosystem’, *Nature Reviews Microbiology*, 5(10), pp. 801–812. doi: 10.1038/nrmicro1750.

Tilburg, C. E. *et al.* (2001) 'The Dynamics of the East Australian Current System: The Tasman Front, the East Auckland Current, and the East Cape Current', *Journal of Physical Oceanography*, 31(10), pp. 2917–2943. doi: 10.1175/1520-0485(2001)031<2917:TDOTEA>2.0.CO;2.

Vaulot, D. *et al.* (2008) 'The diversity of small eukaryotic phytoplankton (≤3 μm) in marine ecosystems', *FEMS Microbiology Reviews*, 32(5), pp. 795–820. doi: 10.1111/j.1574-6976.2008.00121.x.

Worden, A. *et al.* (2015) 'Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes', *Science*, 347(6223). doi: 10.1126/science.1257594.

Worden, A. Z., Nolan, J. K. and Palenik, B. (2004) 'Assessing the dynamic and ecology of marine picoplankton : the importance of eukaryotic component', *Limnol. Oceanogr.*, 49(1), pp. 168–79.

Wu, L. *et al.* (2012) 'Enhanced warming over the global subtropical western boundary currents', *Nature Climate Change*. Nature Publishing Group, 2(3), pp. 161–166. doi: 10.1038/nclimate1353.

Zhang, C. *et al.* (2018) 'Evolving paradigms in biological carbon cycling in the ocean', *National Science Review*, 5(4), pp. 481–499. doi: 10.1093/nsr/nwy074.

Zubkov, M. V. *et al.* (2000) 'Picoplankton community structure on the Atlantic Meridional Transect: A comparison between seasons', *Progress in Oceanography*, 45(3–4), pp. 369–386. doi: 10.1016/S0079-6611(00)00008-2.

Zwirglmaier, K. *et al.* (2008) 'Global phylogeography of marine Synechococcus and Prochlorococcus reveals a distinct partitioning of lineages among oceanic biomes', *Environmental Microbiology*, 10(1), pp. 147–161. doi: 10.1111/j.1462-2920.2007.01440.x.

# Chapter 3:

# Investigating the bacteriophage diversity across a temperate hotspot for climate change

# Investigating the bacteriophage diversity across a temperate hotspot for climate change

Amaranta Focardi[1], Martin Ostrowski[1], Mark Brown[2], Kirianne Goossen[3], and Ian Paulsen[1]

[1]Department of Molecular Sciences, Macquarie University, North Ryde, 2109, NSW, Australia

[2]School of Environmental and Life Sciences, University of Newcastle Australia, Callaghan, Australia.

[3]CSIRO Oceans and Atmosphere, Castray Esplanade, Hobart, Australia

## 3.1 Abstract (232)

Virus- and bacteriophage-induced mortality can have a significant impact on marine productivity and alter the flux of nutrients in marine microbial food-webs. Viral mediated horizontal gene transfer can also influence species fitness and community composition. However, there are very few studies of marine viral diversity in the Southern Hemisphere. We carried out the first genetic study of bacteriophage communities within the East Australian Current system encompassing the Tasman Sea, Coral Sea, the EAC jet and neritic coastal sites. Bacteriophage DNA sequences were extracted from 60 assembled metagenomes and six metaviromes obtained from 24 different locations that vary by sample depth and by fractionation method (> 0.2 µm fraction [community metagenomes], < 0.2 µm dia. [metaviromes]). More than 20,000 bacteriophage contigs were recovered from the assembled sequences, with contig sizes ranging

from 2 kbp to 197 kbp. These included 624 large genomic fragments (> 15 kbp), representing near complete bacteriophage genomes. Predicted genes from the entire pool of contigs were used to explore the functional diversity of phage core and auxiliary metabolic genes. Bacteriophage diversity displayed distinct depth and regional patterns. While there was a clear difference in the bacteriophage populations associated with the EAC or the Tasman Sea for the euphotic zone, reflected both at the taxonomic and functional level. In contrast, bathypelagic phage were similar between the oceanic systems suggesting that, differently from the euphotic zone, the host community was comparable.

## 3.2 Introduction

Viruses are the most abundant entity in the marine environment, with a concentration that can reach as high as $10^7 ml^{-1}$ they can outnumber bacteria by more than an order or two of magnitude. In the ocean the majority of viruses infect bacteria and are termed bacteriophages (Wommack and Colwell, 2000).

Bacteriophage and viruses have a considerable influence on the ecology and the biogeochemical cycles of the ocean (Fuhrman, 1999; Suttle, 2007). Viral-induced mortality can influence the flux of nutrients in marine microbial food-webs through the release of newly dissolved organic matter (Breitbart *et al.*, 2018). It is estimated that viruses promote the release of 145 gigatons of carbon per year in the marine environment (Lara *et al.*, 2017) and therefore represents a major route of C -recycling. The DOM released via viral-induced mortality can either fuel microbial growth, through recycling limiting nutrients in the Ocean (Danovaro *et al.*, 2011; Jover *et al.*, 2014), or enhance the efficiency of the carbon pump (Guidi *et al.*, 2016). Moreover, bacteriophage can promote genetic variation of their host via horizontal gene transfer (Lindell *et al.*, 2004), and alter ecosystem productivity through the expression of specific viral-encoded auxiliary metabolic genes (AMGs) (Hurwitz and U'Ren, 2016). AMGs are widespread in bacteriophage and are believed to encode functions that sustain the host metabolism during infection. For example, the core photosystem II genes *psbA* and *psbD* (Mann *et al.*, 2003; Puxty *et al.*, 2015) in cyanophages play a role in preventing photo-inhibition, thus maintaining phototosynthetic electron transport in order to boost phage production; or genes involved in the carbon metabolism, like *cp12* and *talC* (Millard *et al.*, 2009; Sullivan *et al.*, 2010; Puxty *et al.*, 2015) that when expressed inhibit the Calvin cycle and redirect the carbon flux to the pentose phosphate pathway (Thompson *et al.*, 2011) to promote deoxynucleotide biosynthesis. Despite the importance of bacteriophage in the marine environment there is still a paucity of information on the diversity of viral communities in the Southern hemisphere. Here we carried out the first study of the bacteriophage diversity in the region encompassed by the East Australian Current (EAC).

The EAC system is a model western boundary current, and an ideal system to investigate the biological impacts of climate change (Wu *et al.*, 2012; Hobday and Pecl, 2014; Pecl *et al.*, 2014). The poleward extension of the EAC with its warm core eddies (Ridgway, 2007; Suthers *et al.*, 2011) to lower latitudes is thought to play a role in transporting microbial species and shifting the boundaries of microbial provinces (Messer at al., in prep; Varkey et al., in prep) that are usually associated with warmer water to higher latitudes in the Tasman Sea. Previous studies on the microbial assemblages have underlined how those two oceanic provinces present specific microbial assemblages associated with the different characteristic of the provinces (Seymour *et al.*, 2012; Brown *et al.*, 2018). While the balance between microbial dispersal and selection in this dynamic region is complex, we hypothesized viral mortality is playing a pivotal role in community turnover, and hence may be a driver of the emergent patterns of microbial diversity across this region. However, there is no information about the identity and diversity of the main viral groups, and especially those infecting primary producers in this region. This gap in knowledge limits our ability to fully understand the factors that govern microbial community structure, and which influence present and future productivity in a region undergoing significant climate-related change.

Here we present the first study of the genetic diversity of viral communities in both the euphotic and aphotic zone of the EAC and Tasman Sea (0-4,000 m depth). Viral genome fragments were assembled from metagenomic contigs derived from both the microbial (> 0.2 μm) and the viral fractions (< 0.2 μm) and covered a wide latitudinal range (28-37˚ S, 151-154˚E).

## 3.3 Methods

### *Sample collection*

Samples were collected from 26 stations during an oceanographic expedition on board the RV Investigator (IN2016_v04) in September 2016 during the Austral spring (Fig. 3.1). Water from specific depths up to 4,000 m was sampled using a rosette equipped with 24 x 12 L Niskin bottles and a CTD fitted with additional optical sensor (Fluorometer, Oxygen). The Samples were collected on a Lagrangian drift along the East Australian current (30.65˚ – 32.99˚ S), two shore-normal longitudinal transects at the southern (Tr_S: 36.9˚S, 150.9˚ -154.25˚ E), middle (Tr_M: 32.4˚ S, 152.87˚ – 154.10˚ E), and two site in the northern sector of the voyage (Tr_N: 28˚s – 155˚E) (Fig. 3.1). The sampling station of the transect middle were: one coastal station, a station inside the EAC, and two station offshore. Sixty samples were collected for the large 'microbial' fraction (> 0.2 µm) and six samples for the small 'viral' fraction (< 0.2 µm).

For the large fraction 2L of seawater were filtered through a 0.2 µm dia. pore size PES filter (Sterivex, Millipore) and stored at -80˚C (Brown *et al.*, 2018). Water was collected from 4-5 different depths (Suppl. table 3.2) at each station (3 in the euphotic zone and 1-2 in deep samples).

For the small 'viral' fraction, 20 L of the filtrate from the large fraction (< 0.2 µm) was treated with iron chloride for 1 hr (John *et al.*, 2011; Hurwitz *et al.*, 2013) to flocculate the viral particles at room temperature. The flocculate was collected by filtration on a 0.8 µm dia. pore size polycarbonate filter (147 mm dia.) and stored at 4˚C until subsequent processing (within two months of sampling). All small 'viral' samples were collected from the surface (~2 m depth), except for one sample collected at 3,000 m. Two of the surface samples were collected in a warm core eddy (CTD002 and CTD003)

### *DNA Extraction*

For the small viral fraction, DNA was extracted using a modification of the iron chloride flocculation protocol (John *et al.*, 2011). Briefly, the viral particle concentrate was resuspended in ascorbate buffer (5.0 ml) on a rotary shaker overnight at 4˚C and then concentrated using an Amicon column (30kDa) at 4˚C and 1000 x g until reached the volume desired (1 mL). After resuspension, samples were treated with DNAase 100U $ml^{-1}$ (Roche) in a reaction buffer at 37º C for 1 h (Hurwitz *et al.*, 2013). The DNAase was inactivated with an equimolar concentration of EDTA/EGTA (final concentration 0.1M) used to chelate the metal ions. DNA was subsequently extracted using the AMPvir kit (qiagen) following the protocol recommendation. Library preparation for NGS was carried out with using the Nextera XT kit for low DNA yield samples and sequenced on an Illumina Miseq platform at the Ramaciotti Centre for Genomics, UNSW, Australia. DNA for sequencing the large 'microbial' fraction extracted and sequenced as part of a the Australian Microbiome Initiative (formerly the Australian Marine Microbes Project ) following standard protocol (Brown *et al.*, 2018).

### *Metagenome analysis*

The quality of both forward and reverse sequence reads were assessed with FastQC. Adaptors and low quality reads were removed with Trimmomatic-0-3.36 (Bolger, Lohse and Usadel, 2014) using default parameters, and trimmed reads were assembled with Spades (v 3.10.1) using the --careful option (Nurk S, Meleshko D, Korobeynikov A, 2017) in line with the recommendations of Nishimura *et al.,* 2017. The average length of assembled reads spanned from a minimum length of 300 to a max of 198,000 nt.

DNA from each sample was assembled separately and then contigs from all assemblies were pooled together and dereplicated with CD-HIT (Li and Godzik, 2006) using a 98% similarity threshold, a total of 22,701,531 contigs were identified.Putative viral contigs were identified using Virsorter v.1.05 (Roux *et al.*, 2015), 26,560 from just categories 1 and 2 in Virsorter , were retained for further analysis on the bacteriophage assemblages diversity, where category 1 includes

the contigs with the most confident predictions (enrich in viral like genes and >1 viral hallmark gene) and category 2 contigs that have either enrichment in viral like genes or >1 phage hallmark genes. MetaGeneMark2 (Zhu, Lomsadze and Borodovsky, 2010) was used to identify the open reading frame.

### *Taxonomy and functional annotation of Viral contigs*

Taxonomy was assigned with blastn against the NCBI nr database. Viral protein families were calculated with an hmmsearch against the HMM models from the IMG/VR database (Paez-Espino, Chen, *et al.*, 2017). Finally contigs were functionally annotated with Interproscan 5 (Jones *et al.*, 2014) and blastx search against the NCBI nr protein database (released October 2018) using Diamond (> 70% similarity nt, > 60% similarity aa) (Buchfink, Xie and Huson, 2015).

Similarity to previously sequenced viral DNA fragments from metagenomic studies was assessed with a blastn search against the IMG/VR database (2018). Viral fragments were considered similar if they shared >50% of the genes with a >= 50% of nt similarity, this relaxed cutoff was chosen to try to identify distant homologues of the reference phage. Viral contigs were grouped in four categories based on the number of genes shared with their closest homologue: novel bacteriophage (N) (<10% shared genes), low similarity (LS) (10-50% shared genes), similar (S) (50-80% shared genes), known phage (K) (>80% shared genes).

### *Bacteriophage contigs identification.*

In order to better understand bacteriophage distribution in Australian waters we selected just contigs > 15kbp from the total dataset, that likely represent near complete genomes of abundant bacteriophage. These contigs underwent a second quality control step and only those with at least six hits to a viral protein family VPFs (HMM modules from IMG/VR) were retained. This approach selected a final number of 624 contigs that represents almost complete bacteriophage (Suppl. table 3.1).

Filtered Reads from each sample were mapped against the final viral contigs database (624) using bowtie2 (Langmead, 2013) to understand the relative coverage of each contigs in the different sites.

***Analysis of functional profile for each site.***

To characterise the functional diversity, the genome-centric approach was followed by a gene-centric approach. Genes from all the 26,560 contigs were annotated with Interproscan with an evalue cutoff < $10^{-5}$. Function was assigned to a total of almost 37,000 genes. Similar, to the approach used by Coutinho *et al*., 2017, the functional profile for each sample was characterised by summing up the relative abundance of each pfam proportionally to the relative abundance of the contigs in which it was found. For example if contig A contained gene 1 and gene 2 and its abundance was 10 in sample A and 2 in sample B, then the functional profile for samples A was calculated as gene 1 multiplied by ten plus gene 2 multiplied by 10, while for sample B genes were multiplied by 2.

***Statistical analysis***

Statistical analysis was performed in R v3.4.1 (R Core, 2016). To better understand bacteriophage regional distribution, contig count per samples were rarefied and sqrt transformed. Sample sites were clustered based on similarity (Bray-Curtis) of contig abundance with hierarchical clustering analysis in SIMPROF test, significance of the cluster was then tested with ADONIS (Vegan) (Oksanen *et al.*, 2015). Contigs that significantly contributed to the dissimilarity between cluster and Oceanic provinces, were identified using SIMPER (Similarity percentage). We used the unsupervised multiple learning kernel, mixKernel (Mariette and Villa-vialaneix, 2017) in the mixOmics package (Rohart *et al.*, 2017), using the full-UMKL option, to integrate physicochemical and contig abundance dataset into a single exploratory analysis, and similarity results were visualized with a KPCA (Kernel PCA). Multidimensional scaling plot were used to visualize how much both environmental variables and grouping factor (Oceanic provinces) explained the distribution of viral contigs.

## 3.4 Results

Viral diversity was inferred from 66 metagenomic samples collected on 9˚ of latitude (37 -28˚S) across the EAC and the Tasman Sea during September 2016. To better elucidate the diversity of large genomic fragment or nearly complete phage, the analysis focussed on a subset of the assembled contigs (624) ranging from 15 kbp to 197 kbp. The water column depth for each metagenome sample was categorised as followed: Surface (0-30 m), Chl-max (60-130 m), deep (500-4000 m). Distance from the coast was also considered when the samples were analysed, samples were grouped into: shore samples, between the shore and the EAC, EAC and offshore. To better elucidate the possible role of AMGs in phage adaptation to the different provinces sampled we investigated the distribution of the normalized counts of open reading frame for each sample.

### *Viral similarity to previously publish databased*

Genes from the 624 contigs were screened against the IMG/VR (Paez-Espino, Chen, *et al.*, 2017; Paez-Espino, Pavlopoulos, *et al.*, 2017) database using BLAST to assess the degree of similarity of the viral contigs with previously sequenced and assembled viral genomic fragments from other virome or metagenomic studies (Fig. 3.2). Approximately half of the contigs (343) shared little similarity to previously sequenced bacteriophage and were therefore considered as novel bacteriophage. Those novel phages were differentially distributed and the one that shared less than 10%gene contributed up to the 28% of the deep pool of contigs (Fig. 3.2), suggesting that phage that inhabit the aphotic zone are still underrepresented in the main “Oceanic metagenomic effort”. Approximately 15% of the contigs (94) belongs to known phages that have been found also in different Oceanic region (*Tara* Oceans). Those widely distributed phages are likely to play an important role in the microbial assemblages and marine ecosystem in general.

### *Viral diversity in the different oceanic provinces and depth distribution*

The sites sampled spanned across a wide latitudinal range and presented an extensive environmental variability (Suppl. table 3.2). While the transect South (Tasman Sea) was characterised by lower surface seawater temperature (15-17˚ C) compared to the other transects and the drift along the EAC, it was also characterised by a higher nutrient input. Sample collected in the EAC instead presented always a higher surface seawater temperature (21˚- 22˚ C) and relatively low nutrient input. The transect middle the sites were located in a longitudinal range (from coastal site to EAC, to offshore samples) and presented different condition, the coastal had high nutrient input and surface seawater temperature between 19˚- 20˚ C, while samples in the EAC and offshore were characterised by high surface seawater temperature but low nutrient concentration.

A similarity profile test (SIMPROF) was used to define sites cluster based on the relative bacteriophage distribution. The analysis identified 8 significant clusters based on Bray-Curtis similarity of the relative abundance of the viral contigs (Fig. 3.3). Cluster analysis showed a clear separation between samples collected in the euphotic zone compared from the aphotic zone. Mesopelagic (MP_7) and bathypelagic (BP_4) clusters did not present a different distribution in the Oceanic provinces, suggesting that viral populations in deep Ocean are similar in different region, furthermore the majority of the contigs were either unknown viruses or with low similarity to previously sequenced Mediterranean deep viruses.

The euphotic zone instead presented six statistically different clusters. VS_1 grouped the surface samples from the virome analysis, cluster TS_2 and TDCM_3 grouped site from the Tasman Sea. While EAC_5,6 and OF_8 grouped site from either the EAC, Tr_N and Tr_M, suggesting a clear separation of viral communities between the EAC and Tasman Sea ($p$ value < 0.001) (Fig. 3.3c).

### *EAC and Tasman Sea presents different viral communities*

Although the taxonomy of more than half of the phage contigs remained unknown, there was a clear difference in the main groups present in the different provinces (Fig. 3.4). While, contigs that resembled phages sequenced from the Mediterranean Sea seems to be quite abundant in all the different provinces, the EAC was enriched with either *Synechococcus,* or *Prochlorococcus* phage. The Tasman Sea on the contrary, seemed to be enriched with *Puneceispirillum* phage, that is known to infect SAR116, one of the most abundant marine bacteria. Roseobacter viruses and Cellulophaga phage also reached high concentration in the Tasman Sea, while they were practically absent in the EAC.

### *Functional profile and main AMGs in the different region*

The differences showed in the bacteriophage distribution were also reflected in the functional profile (relative abundance of Pfam in the different samples). The functional profile showed a clear depth separation between the photic vs the aphotic zone, but also a distinct separation in the euphotic zone of the two main oceanic provinces (EAC and Tasman Sea) (Fig. 3.6).

### *Viral Functional diversity in the aphotic zone*

For the samples collected between 500 – 4000 m the most abundant auxiliary metabolic genes encoded for putative ABC transporters or transmembrane region of ABC transporter. Followed by genes that encode for enzyme involved in the de novo synthesis of nucleotide, as AIR synthase, and ribonucleotide reductase. While, genes related to phage lysogenic strategies were relatively sporadic in deep samples when compared to samples from the Euphotic zone.

### *Viral functional diversity in the photic zone*

In the photic zone the majority of genes encode for enzymes involved in the synthesis of deoxyribonucleotide. EAC samples were characterised by a high prevalence of AMGs usually encoded by phages that infect cyanobacteria. Genes that encode for proteins involved in photosynthetic activity (*psbA, psbD,* T-type Phycobiliprotein Lyase) and in nitrogen-related stress (2OG-Fe-oxygenase) were

within the most common, followed by transaldolase (PPP), MazG, that is involved in cell survival under nutritional stress conditions, and multiple peptidase family. The functional profile for the samples collected in Tasman Sea was characterised by a prevalence of genes that encode for enzyme related to the radical SAM superfamily ($p<0.01$), 4Fe-4S single cluster domain, but also the RmlD an enzyme involved in the biosynthesis of the precursor of L-Rhamnose. Furthermore, Tasman samples were enriched with different methyltransferases, glycosyltransferases and pyrophosphohydrolases, some of which were almost absent in the EAC (Fig. 3.7).

***Metagenome assembled genome: two novel cyanophage***

Two of the contigs likely represent complete genomes of two novel bacteriophage that putatively infect cyanobacteria. Those two MAGs are similar in size EAC_VIR1 is 192 Kbp and EAC_VIR2 is 195 Kbp, their GC% content is comparable ~ 38% and similar to previously sequenced cyanophages.

EAC_VIR1 had 246 ORF of which just 60 had a function assigned. Of this 10 represents AMGs and included genes involved in the photosynthetic pathway (*psbA, psbD, hli*), carbon metabolisms (*zwf, G6PD, 6PGD*), nucleotide metabolisms (*nrdB, nrdA, td, MazG*)

EAC_VIR2 had 232 ORF of which 137 were functionally characterised (Suppl. Table 3 ), A total of 16 AMGs were identified for this contigs and included genes related to photosynthesis (*hli, psbA, psbD ,petE*), carbon metabolism (*td, gnd, zwf, talC*), nucleotide metabolism (glutareodoxin, *MazG, cobS, nrdA, nrdB*) and heat shock response (*hsp*). tRNA genes were identified using tRNAscan in Geneious, both of the MAGs presented different tRNA types, respectively 9 (Asp,Ala,Thr,Leu,Gly,Pro,Ser,Arg,Val) for EAC_VIR1 and 4 (Arg,Thr,Ser,Pro) for EAC_VIR2. Those two MAGs were highly abundant in warmer water of the EAC and tr_N, tr_M and almost absent in the Tasman sea.

## 3.5 Discussion

Despite their importance in the marine environment little is still known about the main viral communities encompassing the EAC, a hotspot for climate change, and the Tasman Sea, two important regions for the Australian marine ecosystem. It is therefore important to understand the viral diversity and their role in shaping microbial community assemblages in a region undergoing significant climate-related change. Here we carried out the first study of the genetic diversity of bacteriophage community for the region encompassed by the EAC. While the emergence of new generation sequencing technologies has accelerated the discovery of viruses in the marine environment (Williamson *et al.*, 2008; Brum *et al.*, 2016; Roux *et al.*, 2016), in this study the majority of the large genomic contigs still belonged to putative novel bacteriophage. This result emphasizes the gaps that still exist in understanding the diversity of marine bacteriophage, especially for the Southern Hemisphere.

Significant difference, both at the taxonomy and functional level, of the bacteriophage communities along the East coast of Australia, were associated with environmental gradients, photic/aphotic and a temperature and nutrient gradient, as warm and nutrient poor EAC vs. the Tasman Sea (Fig. 3.7). With this study we provide a benchmark dataset to start to understand bacteriophage diversity and their potential role in the microbial patterns along this region.

### *Comparisons between different fraction: how much diversity are we losing?*

Usually, seawater for the analysis of the microbial communities is filtered through different pore size filter to enrich for specific fraction: microbial and attached-fraction (> 0.2 μm) and the concentrated viral fraction (< 0.2 μm). Here, for four different sites (CTD14, CTD16, CTD21, for surface and CTD54 for deep samples), we compared the diversity of the identified bacteriophage contigs across the two sizes fraction. While the different fraction enriched for specific groups, here we found that, differently from previous results the viral fraction

cluster reasonably well with its microbial fraction counterpart collected from the same site (Fig. 3.5b) when studying the distribution of the main bacteriophage community (> 15 kpb). These results suggest that the bacteriophage from the 0.2 μm fraction, possibly representing phage actively infecting bacteria, could provide a good and reliable starting point to understand the bacteriophage diversity.

### *Ecology of mesopelagic and bathypelagic phages*

Bacteriophage from the aphotic zone were almost all classified as unknown viruses and shared the least similarity with previously known phage (Fig. 3.2), suggesting that the diversity of bathypelagic phage is still underrepresented in the data from the main oceanographic campaign effort (Rusch *et al.*, 2007; Brum, Schenck and Sullivan, 2013; Hurwitz, Brum and Sullivan, 2015) and were similar in the different provinces (Fig. 3.5a). While the genomic content revealed a high number of putative transporters and enzyme related to the de novo synthesis of nucleotide (AIR synthase).

Marine deep ecosystems are characterised by cold temperature and limited bioavailability of nutrients for microbial metabolisms, in this scenario phage-induced mortality secure the release of bioavailable organic compounds (Fuhrman, 1999; Suttle, 2007) that can sustain high microbial biomass (Danovaro *et al.*, 2008, 2011). Deep communities indeed, registered a relative decrease in integrase genes and concurrent flow-cytometry analysis of the bacterial and viral communities, also revealed an apparent increase in the VBR (viral to bacteria ratio) with depth, altogether those results suggest a high prevalence of lytic strategies in deep marine environment, however the VBR was highly variable between the samples. This result is in line with the “piggyback the winner/loser” (PBW-PBL) hypothesis (Knowles *at al.,* 2016), for which high host density is correlated to a shift to lysogeny strategies, reflected also in a low VBR. Moreover, the expression of virally-encoded transporter, that represented the most abundant gene in deep marine ecosystem may enhance the uptake of nutrient by the host, to later be used as building blocks for the de novo synthesis of nucleotide and ultimately the assembly of new phage (Coutinho *et al.*, 2017). Those results emphasize one more time the importance of viruses for the nutrient

"scavenging" especially in habitat where nutrients are supposed to be a limiting factor for microbial growth and metabolism.

***Phage adapt their genomic content based on host and environmental gradient.***

The EAC is a highly dynamic marine ecosystem and the jet between the EAC and the Tasman sea represents one of the most productive regions of East Australia. It is therefore important to understand viral diversity and the role of viruses in shaping microbial communities in a region undergoing significant climate-related change. In contrast to deep ecosystems, viral diversity both at the taxonomic and functional level in the euphotic zone presented a marked regional distribution pattern (Fig. 3.5).

Bacteriophage need their host to replicate and because of that their abundance is usually related to the presence and regional distribution of their specific host. Previous work (Seymour *et al.*, 2012) and concurrent genomic (in.prep) and flow cytometry data (Chapter 2), showed that cyanobacteria are highly abundant in the EAC where they represent the main primary producers; not surprisingly phage thought to infect cyanobacteria predominated in the warm water of the samples collected in the EAC (EAC_6,EAC_8). Here the cyanophage exploited the cyanobacteria to boost their own production, samples were in fact enriched in Pfam involved in the photosynthesis pathway (*psbA, psbD*) and phycobiliprotein Lyase (Fig. 3.7). While *psbA* genes are not essential genes in a phage they confer an advantage in high-light conditions, the phage can in fact replace the light damage cyanobacterial photosystem II machinery with the phage derived one (Puxty *et al.*, 2015). On the other hand T-type Phycobiliprotein Lyase could provide an advantage for a fast phyco-biliprotein assembly during phage infection (Gasper *et al.*, 2017).

Nutrient availability (nitrogen, phosphate, iron) is one of the main limiting factors in host growth (Behrenfeld *et al.*, 2006). Western boundary currents, such as the EAC, are known to be relatively poor in nutrient. In this case the EAC was characterised by lower amounts of nitrogen and phosphorus compared to the Tasman Sea. Genes that encode proteins related to nutrient stress in the cell

(2OG-Fe, *MazG, PhoH)* and nutrient scavenging (phosphate binding protein, putative ABC transporter) were highly enriched in the EAC (Fig.3.7). *MazG* a nucleoside triphosphate pyrophosphohydrolase, has a role in the hydrolysis of ppGpp (Guanosine pentaphosphate). When in a starvation state the cell usually increase the concentration of ppGpp, phage encoded *MazG* could play a role in "reactivate" host metabolisms and tricking the cell in "believing" to be in a state of the nutrient-replete environment, and eventually facilitate virus production (Gross, Marianovsky and Glaser, 2006; Warwick-Dugdale *et al.*, 2019). Finally, *talC* , an enzyme presents in almost all cyanophage (Thompson *et al.*, 2011), that divert the carbon from the Calvin cycle to the PP pathway, and ultimately lead to an increase in energy used for phage production was also significantly abundant in the EAC and offshore samples from the north and middle transect (Fig. 3.7). Altogether those results highlight how bacteriophage in the EAC exploit photosynthetic bacterial communities, by modulating the photosynthetic apparatus to cope with higher light intensity, but also expressing carbon cycle genes that divert energy and carbon towards phage production and nutrient scavenging genes to boost the capture of limiting nutrients (Thompson *et al.*, 2011; Coutinho *et al.*, 2017).

Once the phage finishes its lytic cycle, it usually uses a peptidase or endolysin to hydrolase the bacterial membrane and release the phage progeny. A concurrent study with samples collected from the same sites to investigate the impact of viral infection rates for picophytoplankton mortality (Chapter 2), hypothesised that phage play an important role in cyanobacteria mortality in the EAC. The high presence of peptidase genes and endolysin genes sequences in this water mass could support the results from the chapter 2 about the importance of bacteriophage for the turnover of cyanobacteria assemblages in the photic zone of the EAC compared to the Tasman Sea.

While the EAC is usually dominated by prokaryotic primary producers and hence phage that infect them, the colder and more productive Tasman Sea was enriched with eukaryotic primary producers and reported an increase of *Cellulophaga* and *Roseobacter* phage, but also *Puneceispirillum* phage.

Both *Cellulophaga* and *Roseobacter* species, are usually found in association with marine algae, the latter comprise a group of bacteria that have an important role in the oceanic sulphur cycle through the degradation of DMSP produced by marine algae (Dobler and Biebl, 2006). In contrast to EAC bacteriophage, in the Tasman Sea bacteriophage displayed a different pool of AMGs that could support phage production. This marked difference was also reflected in the NMDS analysis of the ORF distribution in the different samples (Fig. 3.6). Temperature and nutrients were significantly correlated to regional distribution of functional genes reflected in the NMDS analysis.

While in the EAC phage encoded a variety of AMGs to better exploit photosynthetic bacteria, in the Tasman Sea bacteriophage the range of AMGs reflected a greater proportion of infection of heterotrophic organisms. To modulate their metabolisms, the main AMGs encoded for Fe-S cluster protein, Radical SAM superfamily and different glycosyltransferases. Iron-sulfur cluster proteins, are widespread in viral contigs in the photic zone of the Ocean (Hurwitz, Brum and Sullivan, 2015) and are involved in a variety of processes that include electron transfer, catalysis and regulatory processes and oxygen-iron sensing. They are also the main prosthetic group of the radical SAM, a superfamily of enzymes that are involved in the biosynthesis of vitamins, coenzymes, antibiotics and also serve as an oxidizing agent in DNA repair (Wang and Frey, 2007). Glycosyltransferase like family genes are usually encoded by lytic phage to protect their DNA during infection for host enzyme , but in lysogenic phage could be involved in the pathogenic of their host (Markine-Goriaynoff *et al.*, 2004).

## 3.6 Conclusion

In conclusion, using both a genome and gene-centric approach we provide a first insight on the distribution and functional diversity of the main bacteriophage encompassing the EAC and the upper Tasman Sea. Meso- and bathypelagic bacteriophages are still mostly unknown, but they possibly represent one of the major players in deep nutrient recycling. Moreover, phage communities from the aphotic zone were similar across a wide latitudinal gradient, but different from the phage found in the euphotic one. The euphotic zone instead presented different

bacteriophage communities associated with either the EAC or Tasman Sea. While the difference in bacteriophage communities could possibly be explained by the regional distribution of their specific host, such as cyanophage in the EAC. The genomic content associated with the different environment suggest how phage adapt their specific repertoire of auxiliary metabolic genes (AMGs) to better exploit their host based on a change in nutrient bioavailability and temperature.

## 3.7 Acknowledgements

This work was supported by Australian Research Council awards FL140100021 to IP and DP150102326 & DP170101925 to MVB, MO. This work was also supported by the Marine National Facility and we thank the Chief Scientist Martina Doblin and officers and crew of the *R/V Investigator* during cruise IN16_v04. *We would like to acknowledge the contribution of the Marine Microbes Project consortium (*https://data.bioplatforms.com/organization/pages/bpa-marine-microbes/consortium*) in the generation of data used in this publication. The Marine Microbes Project is supported by funding from Bioplatforms Australia through the Australian Government National Collaborative Research Infrastructure Strategy (NCRIS) and the* Education Investment Fund (EIF) Super Science Initiative. Sections of these data were sourced as part of the Integrated Marine Observing System (IMOS) - IMOS is also supported by the Australian Government through NCRIS. We would like to thank the numerous crew members and scientists involved in the generation, curation, and deposition of IMOS oceanographic data.

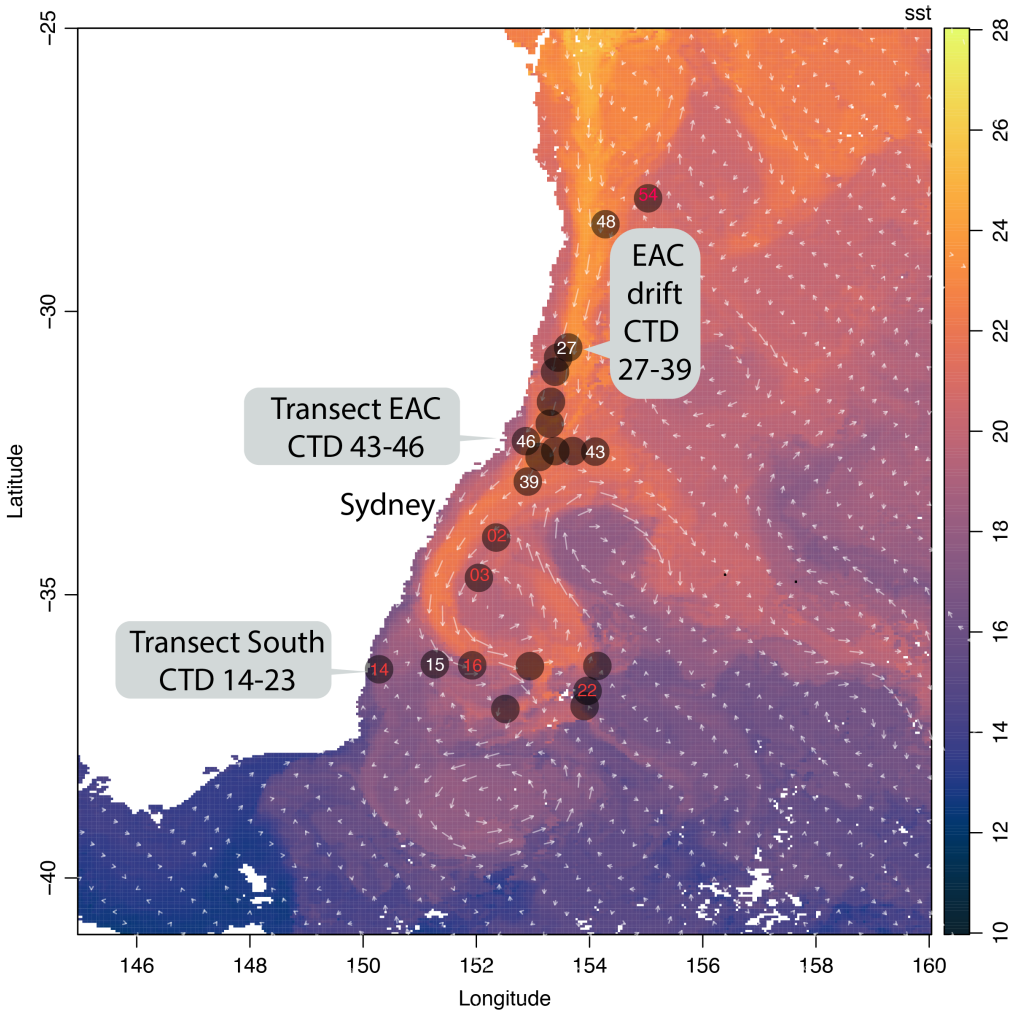



**Figure 3.1** *Location of the sampling sites along the East coast of Australia: Transect South, Transect EAC (middle), the Lagrangian drift along the EAC and the sample North. Data for the 8-day composite average surface seawater temperature on a 4 km resolution was obtained from the AODN for the period September 2016. Sampling sites and transects are highlighted. Colours of the CTD deployment are related to the availability of metagenome for either the microbial fraction (white) (> 0.2 μm) or the microbial (> 0.2 μm) and viral fraction (< 0.2 μm) (red).*

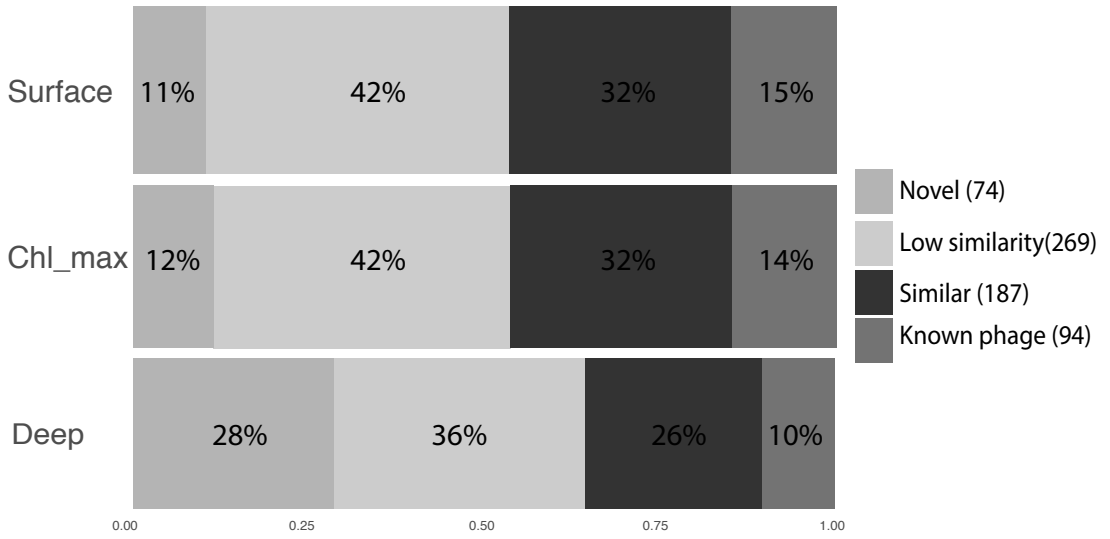
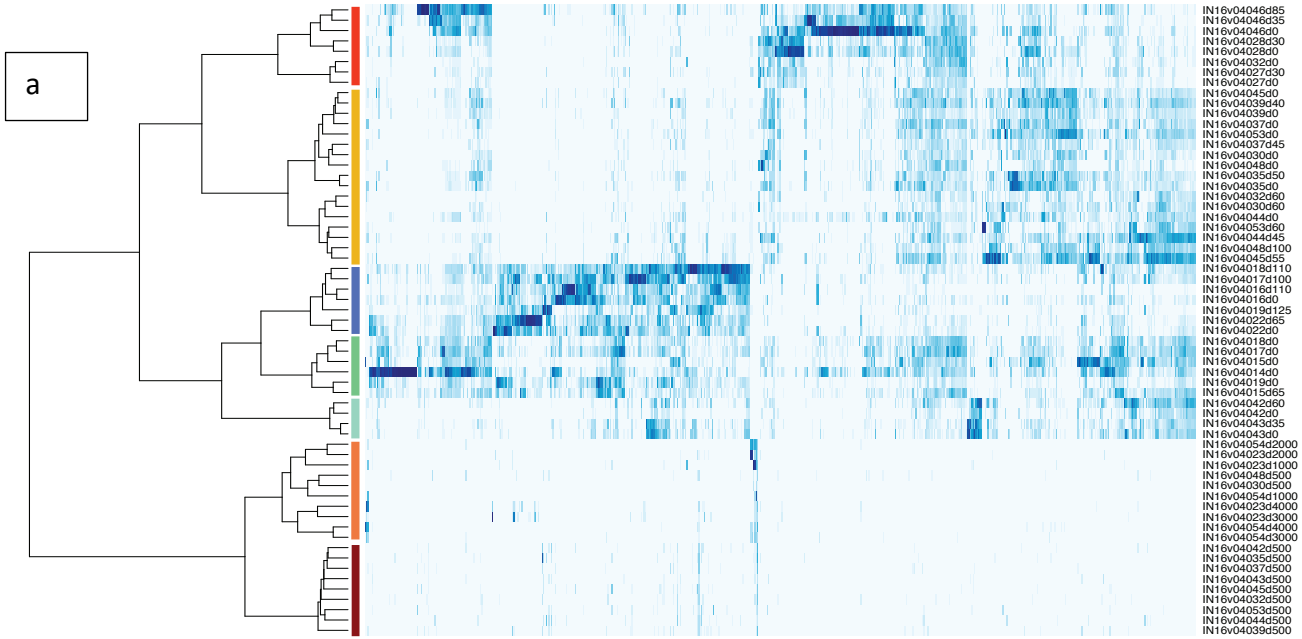
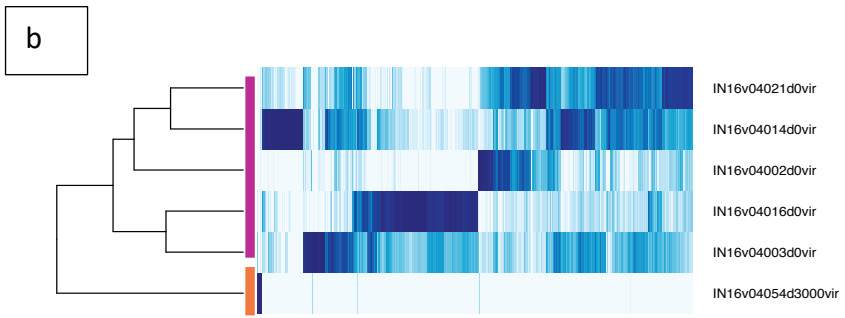



**Figure 3.2** *Relative proportion of novel viral contigs compared with previously sequenced viromes and metagenomes from the IMG/VR database. Contigs were analysed in 3 bins corresponding to broad sample depth ranges.*


a
IN16v04046d85
IN16v04046d35
IN16v04046d0
IN16v04028d30
IN16v04028d0
IN16v04032d0
IN16v04027d30
IN16v04027d0
IN16v04045d0
IN16v04039d40
IN16v04039d0
IN16v04037d0
IN16v04053d0
IN16v04037d45
IN16v04030d0
IN16v04048d0
IN16v04035d50
IN16v04035d0
IN16v04032d60
IN16v04030d60
IN16v04044d0
IN16v04053d60
IN16v04044d45
IN16v04048d100
IN16v04045d55
IN16v04018d110
IN16v04017d100
IN16v04016d110
IN16v04016d0
IN16v04019d125
IN16v04022d65
IN16v04022d0
IN16v04018d0
IN16v04017d0
IN16v04015d0
IN16v04014d0
IN16v04019d0
IN16v04015d65
IN16v04042d60
IN16v04042d0
IN16v04043d35
IN16v04043d0
IN16v04054d2000
IN16v04023d2000
IN16v04023d1000
IN16v04048d500
IN16v04030d500
IN16v04054d1000
IN16v04023d4000
IN16v04023d3000
IN16v04054d4000
IN16v04054d3000
IN16v04042d500
IN16v04035d500
IN16v04037d500
IN16v04043d500
IN16v04045d500
IN16v04032d500
IN16v04053d500
IN16v04044d500
IN16v04039d500


b
IN16v04021d0vir
IN16v04014d0vir
IN16v04002d0vir
IN16v04016d0vir
IN16v04003d0vir
IN16v04054d3000vir

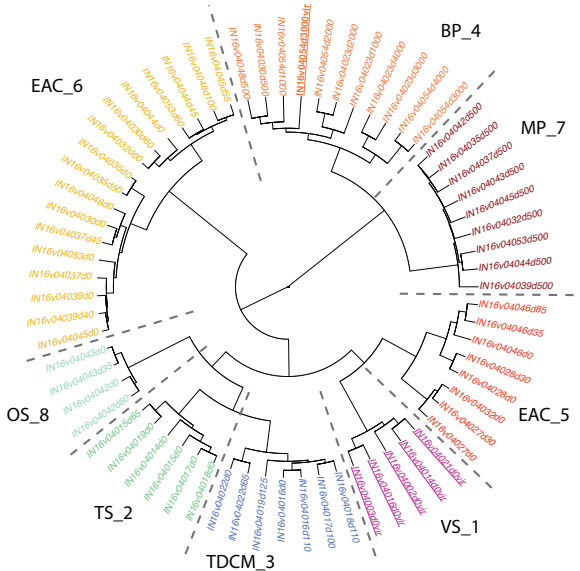

c
BP_4
MP_7
EAC_5
VS_1
TDCM_3
TS_2
OS_8
EAC_6

**Figure 3.3** *Relative abundance of large viral contigs (n = 624). Each row represents a different sample while each column corresponds to a different contig. (a) The relative abundance of viral contigs compared with all samples for the > 0.2 µm fraction; (b) samples from the < 0.2 µm fraction. The mean coverage is square root transformed. The left coloured lines represent the samples grouped into significant clusters (based on the dendrogram shown in (c);), (c) Circular dendrogram showing the relationship between significant clusters, colour-coded based on cluster. VS_1 = Virome samples, TS_2= Tasman Sea surface, TDCM_3 = Tasman Sea DCM, EAC_5,6 = EAC and Tr_N , OS_8 = Coastal and Offshore samples from Tr_M , MP_7 = Mesopelagic samples (500 m), BP_4= Bathypelagic samples and a virome samples.*

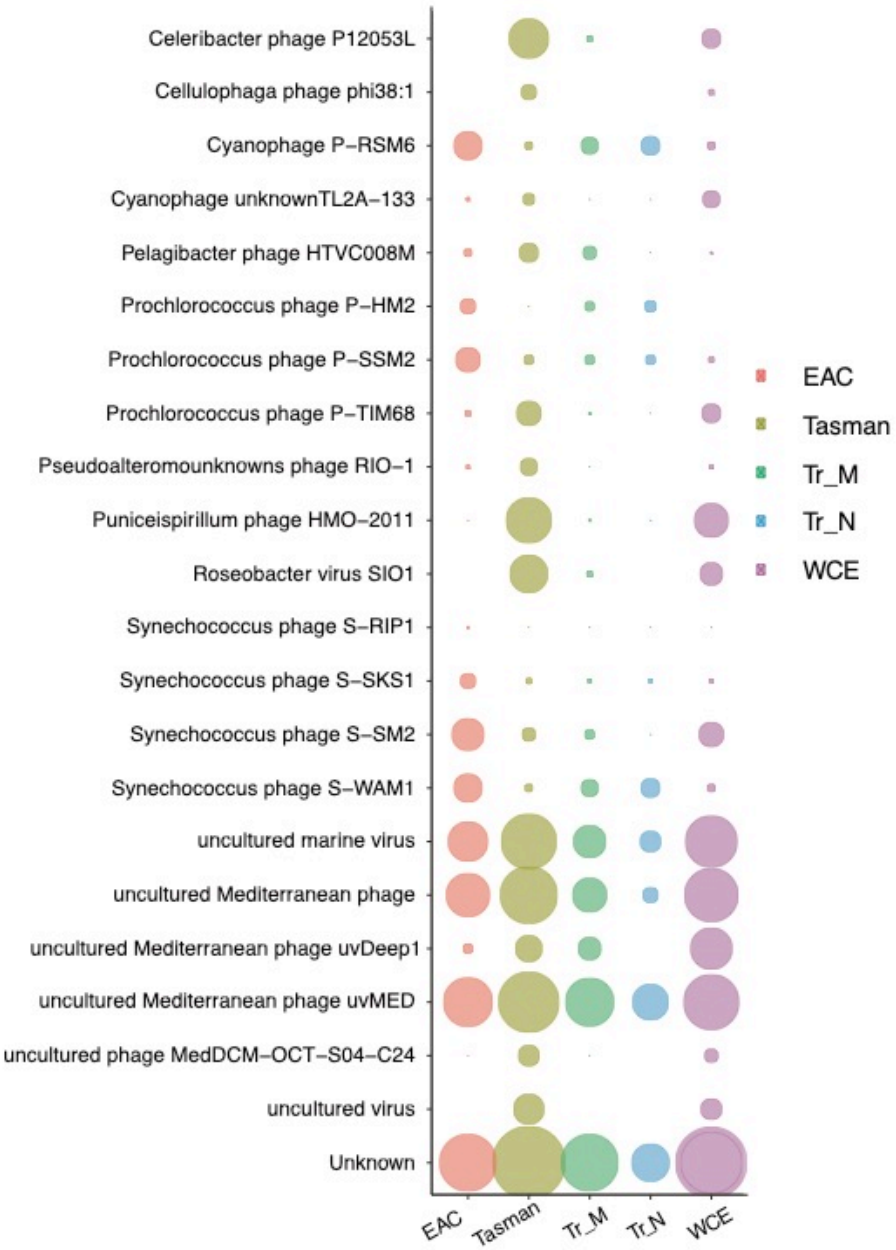



**Figure 3.4** *Distribution of the main bacteriophage species identified by blastN against Refseq (September 2018) in the euphotic (0-130m) zone of the EAC, Tasman Sea, Tr_M, Tr_N and WCE (warm core eddy). Size of the bubble equates to the log transformed relative abundance the main bacteriophage species.*

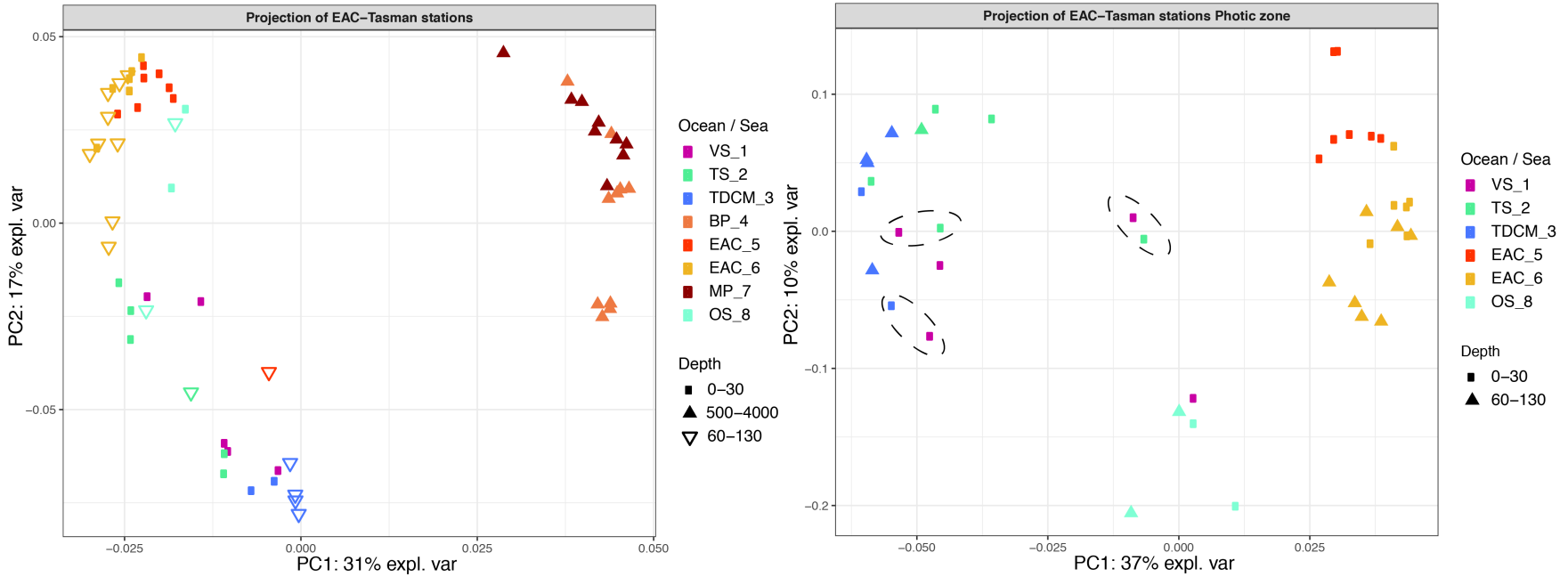



**Figure 3.5** *Ordination plot viral diversity observed in the first two principle component axes for all the stations and depths (left), and for samples within the euphotic zone (right). Samples are colour coded based on hierarchal clusters calculated the SIMPROF test (Clustsig) based on Bray-Curtis similarity (conf. > 95%). Shape are related to different depth. Dashed circles in the plot from the photic zone grouped microbial and viral fraction from the same sampling site.*

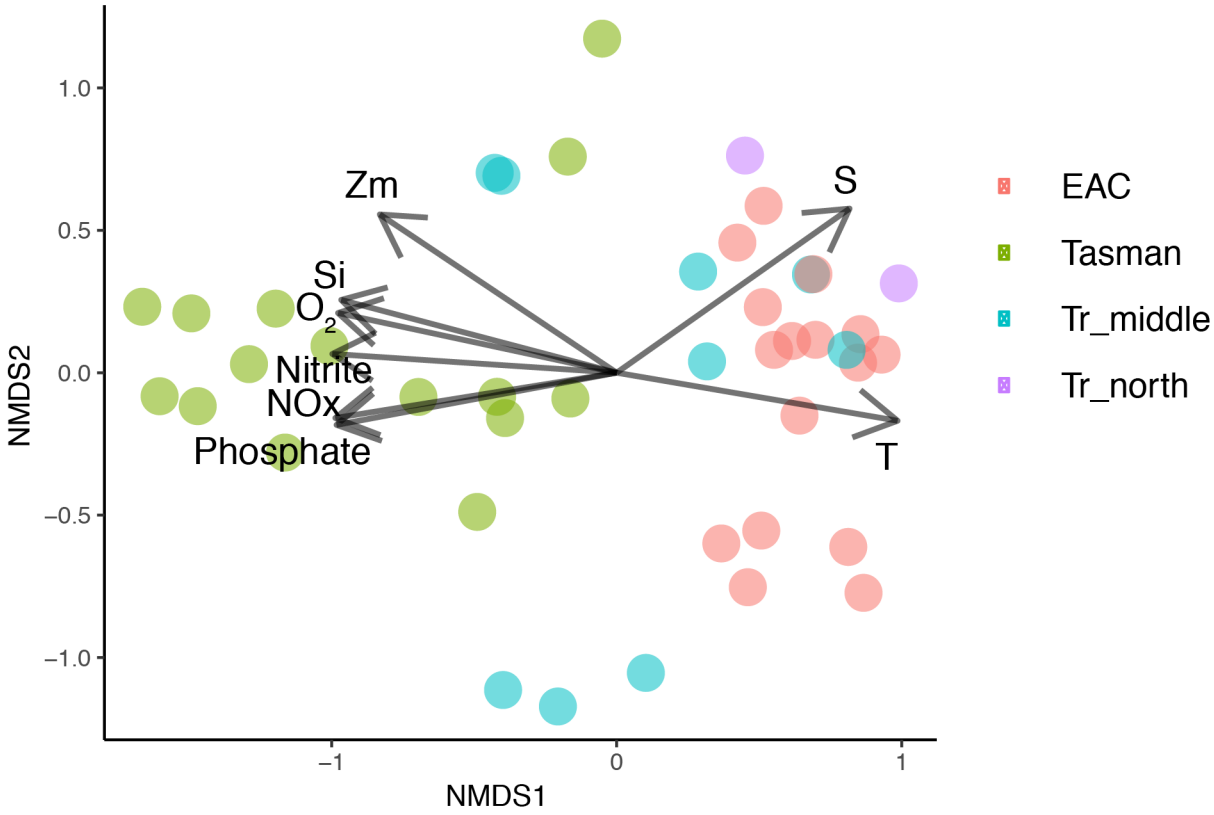



**Figure 3.6** *Non-metric multidimensional scaling (NMDS) of the relationship between sites based on normalised gene count. Significant correlations of environmental variables are shown in black arrows (p<0.05). S= salinity, T= temperature, Si= silicate. The sites are colour coded based on the transect in which they were collected.*


**Figure 3.7** *Distribution of the 80 most abundant gene in the photic zone of the EAC and Tasman Sea. Different colour represents the water different ocean region, either EAC (red) or Tasman Sea (blue). Bubble size indicates to genes relative abundance.*

## 3.7 References

Behrenfeld, M. J. *et al.* (2006) 'Climate-driven trends in contemporary ocean productivity', *Nature*, 444(7120), pp. 752–755. doi: 10.1038/nature05317.

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

Breitbart, M. *et al.* (2018) 'Phage puppet masters of the marine microbial realm', *Nature Microbiology*, 3(7), pp. 754–766. doi: 10.1038/s41564-018-0166-y.

Brown, M. V. *et al.* (2018) 'Data Descriptor: Systematic, continental scale temporal monitoring of marine pelagic microbiota by the Australian Marine Microbial Biodiversity Initiative', *Scientific Data*, 5(April), pp. 1–10. doi: 10.1038/sdata.2018.130.

Brum, J. R. *et al.* (2016) 'Patterns and ecological drivers of ocean viral communities', *Science*, 348(6237).

Brum, J. R., Schenck, R. O. and Sullivan, M. B. (2013) 'Global morphological analysis of marine viruses shows minimal regional variation and dominance of non-tailed viruses', *ISME Journal*. Nature Publishing Group, 7(9), pp. 1738–1751. doi: 10.1038/ismej.2013.67.

Buchfink, B., Xie, C. and Huson, D. H. (2015) 'Fast and sensitive protein alignment using DIAMOND', *Nature Methods*, 12(1), pp. 59–60. doi: 10.1038/nmeth.3176.

Coutinho, F. H. *et al.* (2017) 'Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans', *Nature Communications*. Nature Publishing Group, 8(May), pp. 1–12. doi: 10.1038/ncomms15955.

Danovaro, R. *et al.* (2008) 'Major viral impact on the functioning of benthic deep-sea ecosystems', *Nature*, 454(7208), pp. 1084–1087. doi: 10.1038/nature07268.

Danovaro, R. *et al.* (2011) 'Marine viruses and global climate change', *FEMS Microbiology Reviews*, 35(6), pp. 993–1034. doi: 10.1111/j.1574-6976.2010.00258.x.

Dobler, I. W. and Biebl, H. (2006) 'Environmental Biology of the Marine Roseobacter Lineage', *Annual Review of Microbiology*, 60(3), pp. 255–280. doi: 10.1146/annurev.micro.60.080805.142115.

Fuhrman, J. A. (1999) 'Marine viruses and their biogeochemical and ecological effects', *Nature*, 399, pp. 1–8. doi: 10.1038/21119.

Gasper, R. *et al.* (2017) 'Distinct features of cyanophage-encoded T-type phycobiliprotein lyase ΦCpeT: The role of auxiliary metabolic genes', *Journal of Biological Chemistry*, 292(8), pp. 3089–3098. doi: 10.1074/jbc.M116.769703.

Gross, M., Marianovsky, I. and Glaser, G. (2006) 'MazG – a regulator of programmed cell death in Escherichia coli', *Molecular Microbiology*, 59(2), pp. 590–601. doi: 10.1111/j.1365-2958.2005.04956.x.

Guidi, L. *et al.* (2016) 'Plankton networks driving carbon export in the oligotrophic ocean', *Nature*. Nature Publishing Group, 532(7600), pp. 465–470. doi: 10.1038/nature16942.

Hobday, A. J. and Pecl, G. T. (2014) 'Identification of global marine hotspots: Sentinels for change and vanguards for adaptation action', *Reviews in Fish Biology and Fisheries*, 24(2), pp. 415–425. doi: 10.1007/s11160-013-9326-6.

Hurwitz, B. L. *et al.* (2013) 'Evaluation of methods to concentrate and purify ocean virus communities through comparative , replicated metagenomics', *Environmental Microbiology*, 15, pp. 1428–1440. doi: 10.1111/j.1462-2920.2012.02836.x.

Hurwitz, B. L., Brum, J. R. and Sullivan, M. B. (2015) 'Depth-stratified functional and taxonomic niche specialization in the " core " and " flexible " Pacific Ocean Virome', *ISME Journal*, pp. 472–484. doi: 10.1038/ismej.2014.143.

Hurwitz, B. L. and U'Ren, J. M. (2016) 'Viral metabolic reprogramming in marine ecosystems', *Current Opinion in Microbiology*. Elsevier Ltd, 31, pp. 161–168. doi: 10.1016/j.mib.2016.04.002.

John, S. G. *et al.* (2011) 'A simple and efficient method for concentration of ocean viruses by chemical flocculation', *Environmental Microbiology Reports*, 3(2), pp. 195–202. doi: 10.1111/j.1758-2229.2010.00208.x.

Jones, P. *et al.* (2014) 'InterProScan 5: Genome-scale protein function classification', *Bioinformatics*, 30(9), pp. 1236–1240. doi: 10.1093/bioinformatics/btu031.

Jover, L. F. *et al.* (2014) 'The elemental composition of virus particles: Implications for marine biogeochemical cycles', *Nature Reviews Microbiology*. Nature Publishing Group, 12(7), pp. 519–528. doi: 10.1038/nrmicro3289.

Langmead (2013) 'Bowtie2', *Nature methods*, 9(4), pp. 357–359. doi: 10.1038/nmeth.1923.Fast.

Lara, E. *et al.* (2017) 'Unveiling the role and life strategies of viruses from the surface to the dark ocean', *Science Advances*, 3(9), p. e1602565. doi: 10.1126/sciadv.1602565.

Li, W. and Godzik, A. (2006) 'Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences', *Bioinformatics*, 22(13), pp. 1658–1659. doi: 10.1093/bioinformatics/btl158.

Lindell, D. *et al.* (2004) 'Transfer of photosynthesis genes to and from Prochlorococcus viruses', *Proceedings of the National Academy of Sciences*, 101(30), pp. 11013–11018.

Mann, N. H. *et al.* (2003) 'Bacterial photosynthesis genes in a virus', *Nature*, 424(August), pp. 741–742.

Mariette, J. and Villa-vialaneix, N. (2017) 'Integrating TARA Oceans datasets using unsupervised multiple kernel learning Methods Unsupervised multiple

kernel learning', pp. 1–16.

Markine-Goriaynoff, N. *et al.* (2004) 'Glycosyltransferases encoded by viruses', *Journal of General Virology*, 85(10), pp. 2741–2754. doi: 10.1099/vir.0.80320-0.

Millard, A. D. *et al.* (2009) 'Comparative genomics of marine cyanomyoviruses reveals the widespread occurrence of Synechococcus host genes localized to a hyperplastic region: Implications for mechanisms of cyanophage evolution', *Environmental Microbiology*, 11(9), pp. 2370–2387. doi: 10.1111/j.1462-2920.2009.01966.x.

Nurk S, Meleshko D, Korobeynikov A, P. P. (2017) 'metaSPAdes: A New Versatile Metagenomic Assembler', *Genome Res.*, 1(27), pp. 30–47. doi: 10.1101/gr.213959.116.4.

Oksanen, A. J. *et al.* (2015) 'Vegan'.

Paez-Espino, D., Chen, I. M. A., *et al.* (2017) 'IMG/VR: A database of cultured and uncultured DNA viruses and retroviruses', *Nucleic Acids Research*, 45(D1), pp. D457–D465. doi: 10.1093/nar/gkw1030.

Paez-Espino, D., Pavlopoulos, G. A., *et al.* (2017) 'Nontargeted virus sequence discovery pipeline and virus clustering for metagenomic data', *Nature Protocols*. Nature Publishing Group, 12(8), pp. 1673–1682. doi: 10.1038/nprot.2017.063.

Pecl, G. T. *et al.* (2014) 'Ocean warming hotspots provide early warning laboratories for climate change impacts', *Reviews in Fish Biology and Fisheries*, 24(2), pp. 409–413. doi: 10.1007/s11160-014-9355-9.

Puxty, R. J. *et al.* (2015) 'Shedding new light on viral photosynthesis', *Photosynthesis Research*. Springer Netherlands, pp. 71–97. doi: 10.1007/s11120-014-0057-x.

R Core, T. (2016) 'Crossdating in dplR', 2, pp. 1–12. doi: 10.1016/j.dendro.2009.12.001.

Ridgway, K. R. (2007) 'Long-term trend and decadal variability of the southward penetration of the East Australian Current', *Geophysical Research Letters*, 34(13), pp. 1–5. doi: 10.1029/2007GL030393.

Rohart, F. *et al.* (2017) 'mixOmics: An R package for 'omics feature selection and multiple data integration', *PLoS Computational Biology*, 13(11), pp. 1–20. doi: 10.1371/journal.pcbi.1005752.

Roux, S. *et al.* (2015) 'VirSorter: mining viral signal from microbial genomic data', *PeerJ*, 3, p. e985. doi: 10.7717/peerj.985.

Roux, S. *et al.* (2016) 'Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses', *Nature*. Nature Publishing Group, 537(7622), pp. 689–693. doi: 10.1038/nature19366.

Rusch, D. B. *et al.* (2007) 'The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through eastern tropical Pacific', *PLoS Biology*, 5(3), pp. 0398–0431. doi: 10.1371/journal.pbio.0050077.

Seymour, J. R. *et al.* (2012) 'Contrasting microbial assemblages in adjacent water masses associated with the East Australian Current', *Environmental Microbiology Reports*, 4(5), pp. 548–555. doi: 10.1111/j.1758-2229.2012.00362.x.

Sullivan, M. B. *et al.* (2010) 'Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments', *Environmental Microbiology*, 12(11), pp. 3035–3056. doi: 10.1111/j.1462-2920.2010.02280.x.

Suthers, I. M. *et al.* (2011) 'The strengthening East Australian Current, its eddies and biological effects - an introduction and overview', *Deep-Sea Research Part II: Topical Studies in Oceanography*, 58(5), pp. 538–546. doi: 10.1016/j.dsr2.2010.09.029.

Suttle, C. A. (2007) 'Marine viruses - Major players in the global ecosystem', *Nature Reviews Microbiology*, 5(10), pp. 801–812. doi: 10.1038/nrmicro1750.

Thompson, L. R. *et al.* (2011) 'Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism', *Proceedings of the National Academy of Sciences*, 108(39), pp. E757–E764. doi: 10.1073/pnas.1102164108.

Wang, S. C. and Frey, P. A. (2007) 'S -adenosylmethionine as an oxidant : the radical SAM superfamily', *Trends in Biochemical Sciences*, 32(3). doi: 10.1016/j.tibs.2007.01.002.

Warwick-Dugdale, J. *et al.* (2019) 'Host-hijacking and planktonic piracy: how phages command the microbial high seas', *Virology Journal*, 16(1), p. 15. doi: 10.1186/s12985-019-1120-1.

Williamson, S. J. *et al.* (2008) 'The Sorcerer II Global Ocean Sampling Expedition : Metagenomic Characterization of Viruses within Aquatic Microbial Samples', *PLoS ONE*, (1). doi: 10.1371/journal.pone.0001456.

Wommack, K. E. and Colwell, R. R. (2000) 'Virioplankton : Viruses in Aquatic Ecosystems †', *Microbiology and Molecular Biology Reviews*, 64(1), pp. 69–114.

Wu, L. *et al.* (2012) 'Enhanced warming over the global subtropical western boundary currents', *Nature Climate Change*. Nature Publishing Group, 2(3), pp. 161–166. doi: 10.1038/nclimate1353.

Zhu, W., Lomsadze, A. and Borodovsky, M. (2010) 'Ab initio gene identification in metagenomic sequences', *Nucleic Acids Research*, 38(12), pp. e132–e132. doi: 10.1093/nar/gkq275.

# Chapter 4:

# Characterization of the microbial community of the Dalton polynya, an Antarctic hotspot for climate change

# Characterization of the microbial community of the Dalton polynya, an Antarctic hotspot for climate change

## 4.1 Abstract

Antarctic coastal polynyas are hotspots for primary productivity and have a central role as a carbon sink for the Southern Ocean. The Dalton polynya in the proximity of the Totten glacier system is within the most active polynyas for East Antarctica. Yet, microbial diversity remains unclear from this region. Here we provide a detailed characterization of both the bacterial and eukaryotic community compositions of the Dalton polynya using both molecular and quantitative approaches. This study showed that the polynya presented a different microbial community compares to the adjacent offshore water, with both *Fragilariopsis* and *Phaeocystis* predominant inside the polynya while different families of dinoflagellata were more abundant outside. Bacteria also showed a spatial partitioning with *Planktomarina* and *Polaribacter* increasing in the Dalton polynya, while SAR11 increased offshore. Together these results provide a comprehensive benchmark dataset of the community of the Dalton polynya.

## 4.2 Introduction

Microbial communities provide a basis for the functioning of the marine ecosystem and play a pivotal role in nutrient cycling. They affect the structure and functioning of the marine food web and are fundamentally the source of carbon to the food web or carbon sinks in deep water (Legendre and Rivkin, 2002; Fuhrman, 2009; Jiao *et al.*, 2010). The Southern Ocean plays a global role in the marine carbon and nutrient cycle (Sarmiento and Le Quéré, 1996). For example, Sub-Antarctic Mode Water (SAMW) that forms below highly productive areas of Antarctica and spreads throughout the Southern Hemisphere at depth is thought to be the main source of nutrient at the thermocline for tropical areas (J L Sarmiento et al. 2004).

Spatial variation of sea-ice coverage and irradiance are, together with nutrient bioavailability, the main controlling factors of productivity in Antarctica. While for most of the year Antarctic pelagic communities are covered by a thick layer of sea ice, during the austral summer the ice retreats and part of the area around the Antarctic continent becomes free from ice. Due to the different sea-ice thickness polynyas are the first area in the Antarctic continent to became ice-free in spring (Mundy and Barber, 2001). Coastal polynyas represent hotspots for primary and secondary production, harbouring the highest biomass of primary producers on the continental shelf (Arrigo, 2003), thus they have a pivotal role as a strong carbon sink for the Southern Ocean (Arrigo, van Dijken and Long, 2008). Primary producers are not the only organisms thriving inside of ice-free polynyas, organisms from higher trophic levels, like zooplankton, whales and seals have been shown to extensively use these areas as active feeding grounds (Arrigo, 2003; Karnovsky, Ainley and Lee, 2007; Labrousse *et al.*, 2018). Moreover, the enhanced flux of carbon, due to an active food chain, provides nutrients for a rich benthic community (Wing *et al.*, 2018).

The Dalton polynya, in the Sabrina coast, is one of the most active polynya, in term of ice formation, of East Antarctica (Tamura *et al.*, 2016). The continuous melting of the Totten glacier offsets the formation of dense shelf water (DSW), the precursor Antarctic bottom water (AABW) (Silvano *et al.*, 2018), that in turn is responsible for an increase in ice melting from the glacier. The continuous input of freshwater creates a highly stratified environment, that will likely affect the composition of the microbial community and primary producers.

While a considerable number of studies have characterised microbial communities in the polynya in west Antarctica, i.e. the Amundsen polynya (DiTullio *et al.*, 2000; Karnovsky, Ainley and Lee, 2007; Alderkamp *et al.*, 2012; Delmont *et al.*, 2015; Jiang *et al.*, 2016; Yang, Jiang and Lee, 2016; Swalethorp *et al.*, 2019) and in some coastal areas in East Antarctica (Grzymski *et al.*, 2012; Williams *et al.*, 2012, 2013, Gionfriddo *et al.*, 2016), the Dalton polynya has received less attention so far.

The aim of this study was to provide a detailed characterization of the composition of the microbial communities in the Dalton polynya, through the combination of 16S rRNA and 18S rRNA gene sequencing, coupled with flow cytometry analysis for a quantitative study of viruses, bacteria and picophytoplankton populations. This dataset provides benchmark information on the taxonomic diversity of the microbial communities and a baseline for experimental, metagenomic and modelling approaches to understand the impact that future climate change scenarios, such as accelerated ice melting from the Totten glacier, will have on the primary productivity of this region and beyond.

## 4.3 Methods

### *4.3.1 Sample collection*

Samples were collected during an oceanographic expedition on board the R/V Investigator along the Sabrina Coast, East Antarctica during the Austral summer (January, February IN2017_V01) of 2017 (Fig. 4.1). The voyage duration was 52 days. The Dalton polynya study was carried out over 3 days in February 2017 and because of the fast sea ice coverage it was not possible to go further inside the polynya during that time of the year .Water was collected from different depths (0-3,000 m) at 7 sites in a transect from inside the Dalton polynya, close to the ice edge, to outside (Fig. 4.1) (65.68˚S-64.9˚S) (~100 km), with a SBE911plus CTD (Conductivity-Temperature-Depth) (Seabird electronics) mounted on a SBE11 rosette holding 24 x twelve litre Niskin bottles (Ocean test Equipment Florida) and fitted with additional sensors (fluorescence, oxygen, turbidity and altimeter).

Four litres of seawater were filtered for microbial composition analysis using a 0.2 µm pore diameter PES sterivex filter (Millipore). Subsequently after filtration the filter was stored at -80˚C until DNA extraction on shore. Two ml of seawater were fixed with a 0.5% final concentration of paraformaldehyde, flash frozen in liquid nitrogen and then stored at -80˚C. For Chlorophyll concentration analysis one litre of seawater was collected from different depths within the euphotic zone, filtered on a GFF filter (Millipore) in the dark and stored at -80˚C before chlorophyll extraction on shore. Samples were collected concomitantly for hydrochemistry

analysis of the water column from all of the depths sampled (Oxygen, Salinity, NOx, Nitrate, Phosphate, Silicate, Ammonium). Samples were analysed on-board and data made available through the data repository of CSIRO (https://www.cmar.csiro.au/data/trawler/) held under survey IN2017_V01 (Armand *et al.*, 2018). Samples for DNA extraction and flow cytometry were imported under quarantine regulations and processed in an approved quarantine facility laboratory at Macquarie University.

### *4.3.2 DNA extraction and amplicon sequencing*

DNA was extracted using the Power Water extraction kit, following the manufacturer's protocol. DNA was assessed for quality (Nanodrop) and quantity with a picogreen assay (Invitrogen) and stored at -80˚C. Bacterial and Eukaryotic diversity was assessed by amplicon sequencing of the V1-V3 16S rRNA gene (27f – 519r) and V4 of the 18S rRNA (18sV4F_18sV4R) gene region, respectively. PCR amplifications for both marker genes were carried out at the Ramaciotti Centre for Genomics (UNSW, Sydney), and library prepared using the Nextera XT index kit (Illumina) as outlined in the Australian Microbiome Marine Microbes methods SOP for 16S amplicons (https://data.bioplatforms.com/organization/pages/australian-microbiome/methods). The prepared libraries were then sequenced on a MiSeq platform (Illumina), using pair-end 300 bp reads for 16S rRNA and pair-ends 250 bp for 18S rRNA.

### *4.3.3 Bioinformatic analysis*

Cutadapt (Martin 2018) was used to remove the adapters and primers. Adapter free reads were then processed with DADA2 (Callahan *et al.*, 2016), following the DADA2 protocol (https://benjjneb.github.io/dada2/tutorial.html) and fine-tuning the parameters according to the type and quality of reads that were analysed. Briefly, quality plots were generated for each pair of fastq files to decide the quality filtering to apply, both forward and reverse reads were trimmed at the end to eliminate bases with a low-quality score, but to maintain enough length for the overlapping to be able to merge forward and reverse reads. At this point DADA2

utilises the learned error rates in base calling to infer the number of error-free unique sequences in each sample (Callahan et al. 2016). Following dereplication and inference of sequence variants reads were merged and submitted to a last quality check to remove chimeric reads. Only ASVs (amplicon sequence variants) that had more than 10 reads were consider for further analysis. Taxonomy was assigned using assignTaxonomy and assignSpecies function from DADA2. assignTaxonomy is based on the naïve Bayesian classifier approach of Wang *et al*., 2007. We used two different databases to taxonomically identify the reads from 16S rRNA and 18S rRNA analysis, the Silva v. 132 database and PR2 database (2018) respectively (Guillou *et al.*, 2013; Quast *et al.*, 2013).

### *4.3.4 Chlorophyll extraction*

Chlorophyll (Chla) was extracted with 99% acetone at 4˚C overnight. After 24h milliQ water was added to reduce the concentration of Acetone to 90%, and pigment concentration analysed on spectrophotometer. Concentration of chlorophyll a was calculated based on Jeffrey et al. 1975. The resulting Chl *a* concentration units are in $\mu$g $ml^{-1}$ (Clementson, 2010). 4.2.5 Flow cytometry analysis of viral, bacterial and photosynthetic communities

Seawater samples were thawed at 34˚C for 5 minutes before flow cytometry analysis on a Cytoflex S instrument (Beckman Coulter) equipped with 3 different lasers. Violet laser (405 nm) was used to better distinguish small particles based on SSC value. Photosynthetic communities were distinguished based on specific autofluorescence intensity of the photosynthetic pigments. Chlorophyll autofluorescence was used against side scatter and forward scatter to distinguish 3 different Picoeukaryotes communities (PPE picoeukaryotes, NEuk nanoeukaryotes, PE-Euk phycoerythrin rich picoeukaryotes (Marie, Simon and Vaulot, 2007) Flow rate was kept stable at 30 µl/min and samples run for two minutes. For both bacteria and viruses, samples were first diluted 1/10 with 0.02 µm filtered TE buffer. For bacteria, diluted samples were stained for 10 minutes in the dark with SYBR green I (Invitrogen) 0.02 µm filtered, prior to the analysis. Viruses instead were prepared as Brussaard, 2014, briefly diluted samples were stained with a final concentration of 1% of Sybr Green I (Invitrogen) 0.02 µm filtered, heated at 80˚C for 10 minutes in the dark and then counted on the cytometer. Virus counts were corrected against a blank consisting of TE (Tris-EDTA) buffer + Sybr Green I.

### *4.3.6 Statistical analysis*

Statistical analysis was performed in R (R Core Development Team, 2008) if not stated otherwise. ASVs counts for both 18S and 16S were normalized in a two step process: first counts were rarefied to an equal sequencing depth per sample (16S 26,428 filtered reads per sample, 18S = 21,116 filtered reads per sample), and then the rarefied tables were square rooted. Diversity indices (Richness, Shannon), for both the 18S rRNA and 16S rRNA, were calculated using the Vegan package (Oksanen *et al.*, 2019). Samples were then clustered (separately) with SIMPROF using a Bray-Curtis similarity index and significance of the cluster calculated with ADONIS for both the bacteria and eukaryote tables. We used an unsupervised multiple kernel clustering analysis, from the package mixKernel (mixOmics), followed by a kernel PCA to further assess the similarity between samples. The use of a multiple kernel analysis allowed the integration of multiple datasets into a single set of explanatory variables.

## 4.4 Results and Discussion

Antarctic polynyas are within the most productive area of Antarctica, characterised by phytoplankton blooms that sustain a dynamic marine ecosystem. Here we present the first study of the microbial community associated with the Dalton polynya. Seawater samples collected on a latitudinal transect (100 km) from the ice-edge inside the polynya to samples offshore, outside the polynya, were used to quantify the number of heterotrophic bacteria, viruses and picophytoplankton via FCM analysis. High throughput sequencing of both 16S rRNA and 18S rRNA phylogenetic markers was used to assess the taxonomic diversity of prokaryotic and eukaryotic microbial communities both at an ice-edge and offshore and at multiple depths down the water column.

### *4.4.1 Environmental variables*

The euphotic zone of the Dalton polynya was characterised by a lower surface seawater temperature (SST) and salinity concentration compared to samples collected outside the polynya (-1˚C to 2˚C), especially for surface samples because of the influence of melting sea-ice (Fig. 4.2). SST inside of the polynya was always below 0˚C, while the temperature below the surface was higher inside the polynya than outside, suggesting a high degree of stratification, due to sea ice melting (Silvano *et al.*, 2018). Overall, temperature was at its minimum at the depth corresponding to the maximum chlorophyll fluorescence (chl max) and then increased at the Oxygen Minimum Zone (OMZ) (Fig. 4.3). The concentration of ammonium was also higher inside the polynya ($p < 0.05$), while nitrate/nitrite (NOx) and silicate were higher outside of the polynya. Apart from ammonium, nutrients were usually higher at the chlorophyll max (Suppl Table S1).

### *4.4.2 Abundance of the microbial communities*

Samples from within the euphotic zone of the Dalton polynya were characterised by a different number of microbes compared to samples collected outside. Cytometry analysis revealed that the total numbers of bacteria decreased along the transect from the ice-edge to offshore ($2.4*10^5$ to $1.9*10^5$cells/ml), while total abundance of viruses remained stable ($1.2 *10^6$ cells/ml) (Fig. 4.4). The higher

number of bacteria inside the polynya could indicate enhanced microbial activity and consequently active recycling of dissolved and particulate organic matter produced or excreted by primary producers. Taking into account Low- and High-DNA containing bacterial populations, which could easily be distinguished, cell numbers of HNA displayed a steep reduction, from $1.5*10^5$ to $0.9*10^5$ cells ml (Fig. 4.5). SAR11 are known to dominate the LDNA bacteria fraction, while HDNA bacteria usually groups a diverse number of phyla as Bacteroidetes and Gammaproteobacteria (Moran *et al.*, 2015). This result reinforces the steep reduction seen for Flavobacteriales (Bacteroidetes) along the transect.

Viruses numbers were comparable inside and outside the polynya, with concentration in the around $1.2 \times 10^6$ viruses/ml for samples in the euphotic zone. This observation is slightly higher than recorded from previous study done in the Kerguelen area (Evans et al. 2009). The number of virus particles dropped to as low as $10^3$ viruses/ml for samples collected from 3,000 m depth.

Virus numbers among the three different viral populations that could be resolved by flow cytometry displayed different trends across the transect (Fig. 4.5).Viral population 1 (Vir1) was always the most abundant one, followed by viral population 2 and 3 (Vir2, Vir3), but while both Vir1 and Vir2 increased in concentration along the transect from the sea-ice to offshore, Vir3 was more numerous close to the sea-ice and was usually at peak below the surface. Virus abundance is thought to follow the host distribution, so the differential distribution of the 3 groups could suggest the presence of viruses adapted to different hosts. It has been reported that viruses that infect autotrophic microorganisms (bacteria and eukaryotes) are usually found in the Vir3 group (high Sybr green) (Brussard at al. 2004). Indeed, here Vir3 positively correlate with Chla concentration, calculated from spectrophotometric analysis, and oxygen ($p<0.05$) (Fig. 4.6). This result suggests that viruses could play an important role in phytoplankton mortality inside the Dalton polynya.

All three photosynthetic eukaryotic populations (< 10 µm in dia.) displayed higher numbers outside of the Polynya, (picoEuk, nanoEuk and PE.Euk) and all

displayed a positive correlation with distance from the sea-ice edge ($p<0.05$). Larger diatoms (> 10µm) and colonies of *Phaeocystis*, which were not analysed by flow cytometry because of their cell size and low cellular abundance, are reported to dominate the primary productivity of Antarctic polynyas, (Arrigo, 2003; Karnovsky, Ainley and Lee, 2007; Raes *et al.*, 2018), while smaller eukaryotes, like chlorophytes (< 10 µm), are usually more numerous in offshore water.

### *4.4.3 Microbial diversity*

The composition of microbial communities was analysed though high throughput sequencing of the 16S rRNA and 18S rRNA genes. The resulting sequence libraries were resolved at the level of single nucleotide variants (ASV) to better elucidate fine scale diversity between and within samples. The resolution of fine scale variation is needed to better understand niche partitioning (Needham, Sachdeva and Fuhrman, 2017; Chafee *et al.*, 2018). As a general trend, richness was higher for eukaryote 18S rRNA genes in comparison to bacterial 16S rRNA genes (Fig. 4.7b). Both datasets showed higher diversity in deeper samples (Fig. 4.7a). This trend was especially marked for bacteria with an increase of 88% in richness going from surface to community from deep samples.

To better elucidate the different structure of microbial communities inside the Dalton polynya (sea-ice) in comparison to samples collected offshore, samples sites were clustered based on both 18S rRNA and 16S rRNA normalized counts, separately, using a Bray-Curtis method in SIMPROF. Hierarchical clustering resolved 7 and 8 different significant clusters for bacteria and eukaryotic community, respectively (Fig. 4.8). There was a clear separation of community profiles from different water column depths, a trend that was evident in both 18S rRNA and 16S rRNA results. Samples collected from the aphotic zone clustered into three main groups (DS, DS_2, DS_3) that were consistent for bacteria and eukaryote analyses, except for a single sample from the oxygen minimum zone (OMZ) that for the Eukaryotes clustered on its own. While the cluster DS includes deep samples from below 1,000 m, DS_2 was composed of samples from above 1,000 m and DS_3 clustered samples from between 200-300 m.

The photic zone showed also similar partitioning results for bacterial and eukaryotic community profiles. Surface communities showed a clear separation between samples collected inside the polynya (SP) and samples collected outside the polynya (SO). With surface samples from outside the polynya clustered with polynya samples from below the surface (30-50 m) (BSP), for eukaryotes there was no difference between below surface samples collected inside and outside the polynya (SO_BS). Moreover, Eukaryotic samples from the chlorophyll max (ChlM) did not show any clear partitioning between location in which they were collected, while for the bacteria there was a clear distinction between samples collected inside the polynya (ChlM_P) and outside (ChlM_O).

### *4.4.5 What is driving the dissimilarity between polynya and offshore samples?*

To understand which factors drive the dissimilarity between samples we used a multiple kernel approach, from the mixKernel package (mixOmics) (Rohart *et al.*, 2017). The ordination plot of the observed dissimilarity of the different samples based on Eukaryotic, Bacterial count showed both a depth partitioning (Fig. 4.9a), but also a partitioning for samples collected inside the polynya and outside especially for the euphotic zone (Fig. 4.9b). Based on the first 2 KPCA axes we extrapolated the 10 most informative environmental variables as well as bacterial and eukaryotic Phyla that mostly contribute to the dissimilarity, using the permute Kernel option (mixKernel) (Mariette and Villa-vialaneix, 2017).

While the concentration of NOx, oxygen and nitrite were the most important environmental factors, Dinoflagellata, Ciliophora and Ochrophyta were the most important eukaryotic phylum. Dinoflagellata sequences were more abundant in offshore samples compared to samples inside the polynya, where Ochrophyta were observed in greater relative abundance.

Proteobacteria and Bacteroidetes were the main phyla for Bacteria, and while Proteobacteria, mainly represented by *Pelagibacter* SAR11 which were highly abundant offshore, Bacteroidetes had the opposite trend.

### *4.4.6 General bacteria diversity*

There was a clear trend in bacterial diversity along the transect going from the site closest to the ice-edge to offshore, especially for samples in the euphotic zone. This trend was characterised by a higher bacterial richness in the polynya compared to offshore in both the surface and deeper samples. More than 4,000 sequence variants (ASVs) with a minimum count of 20 reads across the entire dataset were identified and utilized for further study of the bacterial diversity (1,3 million total reads).

**Euphotic zone**

Within the euphotic zone the dissimilarity between sites was mainly due to the relative distribution of two phyla: Proteobacteria and Bacteriodetes. Going into a finer taxonomic resolution, more than 50% of the single ASVs belong to the class of Alphaproteobacteria, of this especially the Pelagibacter lineage (SAR11), that is the most abundant bacterial species in the Ocean (Brown *et al.*, 2012; Wilkins *et al.*, 2013; Giovannoni, 2016; Delmont *et al.*, 2017), was the one with the highest relative abundance. Even though 570 ASVs that could be taxonomically assigned to four different clades of SAR11 (clade Ia, Ib, II, IV) were present (Fig.4.10), one specific ASV belonging to clade Ia contributed up to 30% of the relative abundance of 16S rRNA sequences, in samples from the photic zone. Due to the ubiquity of the Order and the well described ecotype partitioning linked to temperature, SAR11 has also been proposed as a possible marker species to assess the influx of tropical ecotypes towards the polar region (Brown *et al.*, 2012; Delmont *et al.*, 2017). Here, SAR11 clade Ia (144 ASVs) a cold water adapted clade, known to be prominent in the Southern Ocean (Brown *et al.*, 2012; Delmont *et al.*, 2017), accounted for the highest fraction of the SAR11 clade. The relative abundance of SAR11 clade Ia was higher in surface samples where the temperature was higher than 1˚C and decreased going from offshore to inside of the polynya, SAR11 indeed was positively correlated with temperature (Fig. 4.11).

Rhodobacteraceae (Alphaproteobacteria) are recognised as an important functional group in marine habitats. Members of this Order/Class, especially the

genus Planktomarina, were present in higher relative abundance within the euphotic zone of the polynya. Previous studies have associated this group with metabolic activities that enhance ecosystem productivity. For example, *de novo* genome reconstruction of several Rhodobacteraceae highlighted the presence of genes involved in the biosynthesis of vitamin $B_{12}$ (Delmont *et al.*, 2015), which is an important vitamin required by phytoplankton. Rhodobacteraceae have an important role in the sulphur cycle, with the majority of them having the ability to demethylate DMSP to use it as a source of energy (Delmont *et al.*, 2015). A number of specific algae lineages that are observed in this study, e.g. *Phaeocystis* spp., *Gymnodinium nelson* and some crysophyte (Yoch, 2002) , are known to produce large amounts of DMSP within phytoplankton blooms.

Flavobacteriales, belonging to the Bacteriodetes phylum, are also a class that is known to be abundant in the Southern Ocean (Wilkins *et al.*, 2013; Williams *et al.*, 2013; Raes *et al.*, 2018) and especially in Antarctic water (Gionfriddo *et al.*, 2016). Different families and genera of flavobacteriales were present in this study, with Polaribacter the most abundant inside the polynya (4% of the total reads). All flavobacteriales showed a similar trend and were more abundant close to the ice-shelf than offshore (fig 4.10). Correlation analysis at the genus level showed a negative correlation with temperature and a positive correlation with chlorophyll, nitrate/nitrite and ammonium, all in higher concentration inside the polynya (Fig. 4.11). Positive correlation with chlorophyll highlights a possible relationship with primary producers. Flavobacteria are characterised by the presence of a high number of putative transporters for high molecular weight compounds, suggesting that they have the capacity to utilise complex organic compounds produced by algae. A previous study in Newcombe Bay, East Antarctica, have found that flavobacteria actively utilize alga-derived compounds, and in turn they release simpler compounds (Williams *et al.*, 2013).

**Aphotic zone**

Samples from the aphotic zone were characterised by a high relative abundance of bacteria belonging to the order of SAR324 clade, Thiomicrospirales and

Nitrospinales. Both SAR324 and Thiomicrospinales are chemiolitotrophic bacteria. Both of those bacteria have been found in high abundance in oceanic hydrothermal vent plumes (Cao *et al.*, 2016), where they oxidize compounds such as methane and sulphur containing compounds as source of energy. Nitrite-oxidizing bacteria, such as *Nitrospina*, were also numerous in deep samples and close to the OMZ, and where indeed positively correlated with depth, temperature and negatively with oxygen.

### *4.4.7 General Eukaryotic diversity*

Eukaryotic ASV richness was five times higher than observed in the bacterial data, especially for samples collected in the euphotic zone. A total of 13,554 ASVs with a minimum count of 20 reads, were maintained for further analysis of the general eukaryotic diversity (1.4 million total reads).

The majority of reads belonged to the class of Bacillariophyceae, Dinophyceae and Syndiniales (Fig. 4.12). It is important to acknowledge that the genomes of the majority of eukaryotic phyla, such as Dinophyta and Bacillariophyta, are characterised by a variable copy number of the 18S rRNA marker gene per cell (10-1,000) making it difficult to discern a quantitative description of the Eukaryotic population. However, in remote areas, such the Sabrina Coast, from where no previous data on the total eukaryotic community are available, marker gene surveys based on the 18S rRNA gene provide a good qualitative understanding of which eukaryote taxa are present, and some capacity to observe changes in relative abundance of genera and species within each Class.

There were clear trends in the proportion of major eukaryote classes in the data with Ochrophytes, Haptophytes and Stamenopiles in higher relative abundance in the polynya and Dynoflagellata, Ciliophora and Radiolaria in higher relative abundance offshore (Fig. 4.12).

CIliophora are an important microzooplankton phyla for the food web in the Oceanic environment, they represent the intermediate link between pico-nano primary producers and larger zooplankton (Jiang *et al.*, 2013) . Here Ciliophora were more abundant in the offshore samples. *Leegaardiellidea* , one of the main

genera, was found in similar study to be more present at the transitional area of the polynya rather than close to the sea ice (Jiang *et al.*, 2016). Sequences belonging to Tintinnidae family (Ciliophora, Spirotrichea, Tintinnida) were also numerous in this study. A previous study, that focused on two specific morphologically defined tintinnid species in this area, found an interesting physical association between them and different species of *Fragilariopsis*, especially *Fragilariopsis cylindrus* (Armbrecht *et al.*, 2017).

Syndiniales represents up to the 10% of reads count for offshore samples, compared to the 5% of inside the polynya. The majority of Syndiniales have been observed as obligate parasites and highly abundant worldwide (Lepère *et al.*, 2016; Clarke *et al.*, 2019). Because of their lifestyle they could potentially have an important impact in the marine food web, although the significance of the change in relative abundance observed here requires further investigation.

Deep samples (below 200m) registered a drastic (74%) decreased in number of reads, while samples richness remained similar, or slightly increased, especially in the Dalton polynya, suggesting that the high primary productivity could influence the bathypelagic community (Wing *et al.*, 2018). With Radiolaria and Syndiniales again within the most numerous groups. Both *Fragilariopsis spp.* And *Chaetoceros* spp were present in deep samples where they make up for 4% of the reads. The presence of diatoms at these depths (400 – 1000 m) could suggest a favourable carbon pump with an enhance export of carbon to the aphotic zone of the Ocean (DiTullio *et al.*, 2000).

**Euphotic zone: distribution of the main primary producers**

While Diatoms are known to be the main primary producers of the Southern Ocean and within the most abundant organisms, different genera and species display specific biogeographical distributions, usually in relation with environmental variables (Fiala and Oriol, 1990; Armand *et al.*, 2005; Tréguer *et al.*, 2017; Rigual-Hernández *et al.*, 2018). Here the main diatom species (*Bacillariophyceae*) were differently distributed between the polynya and offshore

sites. Sequences classified as diatoms of the genus *Fragilariopsis* were the most abundant amongst primary producers.

Within the polynya *Fragilariopsis cylindrus*, a relatively small (10 µm dia.) and known sea-ice associated species (Kang and Fryxell, 1992; Quillfeldt, 2004), registered a 58% increase in relative abundance ($p < 0.05$) compared with offshore samples. In contrast, *Fragilariopsis sublineata* did not show any significant increase in the polynya. *Fragilariopsis* spp. sequences were represented the most abundant primary producers in a study of the eukaryotic community from McMurdo Station (Bertrand *et al.*, 2015).

Similar results were registered by *Chaetoceros peruvianus* and *Corethron inerme* with an increase of 90 and 93% respectivly. *Hemiaulus sinensis* followed an opposite trend with a 52% decrease. Both *Chaetoceros* and *Corethron* negatively correlated with temperature (Fig. 4.13) and positively correlated with silicate and nitrite. *Hemiaulus sinensis* was positively correlated with temperature and negatively with Nitrate/Nitrite and silicate concentration.

*Phaeocystis antarctica* (Haptophyta) was more numerous inside the polynya registering the highest increase in relative abundance (110%). *Phaeocystis* is a DMSP producing algae and are recognised as the most important bloom forming primary producers in west Antarctica (DiTullio *et al.*, 2000; Delmont *et al.*, 2015). Because of their C/P ratio they have a greater potential impact in the carbon export in comparison to diatoms (DiTullio *et al.*, 2000), making them a significant species in the carbon cycle especially in light of climate change.

## 4.5 Conclusions

Here we presented a comprehensive quantitative and taxonomic analysis of the microbial community of the Dalton polynya, a highly productive region in East Antarctica. Seawater temperature and nutrient bioavailability seem to play an important role in shaping the microbial community.

Speculatively the lower seawater temperature, coupled with nutrient concentration and possibly a higher iron input due to ice melting inside the polynya, create a suitable habitat for blooms of diatoms such as *Flagilariopsis* spp. and *Chatoceros* and for the bloom forming Haptophyta*, Phaeocystists Antarctica* (Fig. 4.14). These primary producers drive in turn a bacterial community structure characterised by increased populations of “larger” bacteria, including Rhodobacteraceae, as *Planktomarina*, Bacteroidetes, as *Polaribacter,* and the Gammaproteobacteria SAR86, usually found in the HDNAb fraction of FCM analysis. These bacteria are known to be able to uptake HMW-OM produced by algae blooms (*Polaribacter*), or to utilize DMSP produce by algae as a source of energy (Rhodobacteraceae). These intermediate, larger bacteria can break down HMW compounds into simpler one for other bacteria to uptake. Moreover, diatoms are characterised by faster sinking rates, that could resolve in an active transport of POC to the Ocean interior.

While Diatoms remain the main primary producers, other groups seem to play an important role in the warmer offshore samples. Dinoflagellata and especially Gymnodiniales, most of which are known to be photosynthetic organisms, were found to increase in number in offshore samples, as well as the parasite Syndiniales, and the grazers Ciliophora. Because of their parasitic nature Syndiniales are thought to be important players in favouring the export of POC to deep water. The different Eukaryotic diversity was in turn reflected by a majority of smaller bacteria, especially belonging to SAR11 Ia clade (LDNAb). The limit between the photic and aphotic zone registered an increase in the *Nitrospina.* This nitrite oxidizing bacteria play a pivotal role in the nitrogen cycle providing new nitrate that can be assimilated by primary producers to fuel new primary

productivity. Moreover, the export of POC could support a diverse benthic community. But while both difference in nutrient and seawater temperature seems to explain the observed shift in eukaryotic community from the polynya to offshore, the possible biotic interactions and underlying metabolic nutrient cycles that govern the microbial communities need to be investigated further.

This study is the first to integrate molecular data of bacteria and eukaryotes, with high resolution cytometry of the main microbial communities. Moreover, these data will be integrated with metagenomic analysis, that is currently in the process of being analysed to possibly fully understand the microbial dynamics that characterised this important area for primary productivity in East Antarctica. Since future climate change are expected to trigger an increase in melting water in the Antarctic polynyas, with a possible impact on the microbial community composition that still need to be quantified. It is, therefore, fundamental to understand the dynamics of microbial communities, and their underlying metabolism, in such environment, and to incorporate this molecular-level information into models to forecast the impact of climate change on local and global primary productivity.

## 4.6 Acknowledgments

This study was supported by Australian Research Council awards FL140100021 to IP. The authors wish to thank the CSIRO Marine National Facility (MNF) for its support in the form of sea time on RV Investigator, support personnel, scientific equipment and data management. All data and samples acquired on the voyage are made publically available in accordance with MNF Policy. Logistical support was provided through the Australian Antarctic Division's Antarctic Science Support Grant #4333. We thank the Marine National Facility, the IN2017-V01 scientific party-led by the Chief Scientists L.K. Armand and P. O'Brien, MNF support staff and ASP crew members led by Captain M. Watson for their help and support on board the RV Investigator. Thank you to the Ramaciotti Sequencing Centre who performed the library preparation and sequencing.

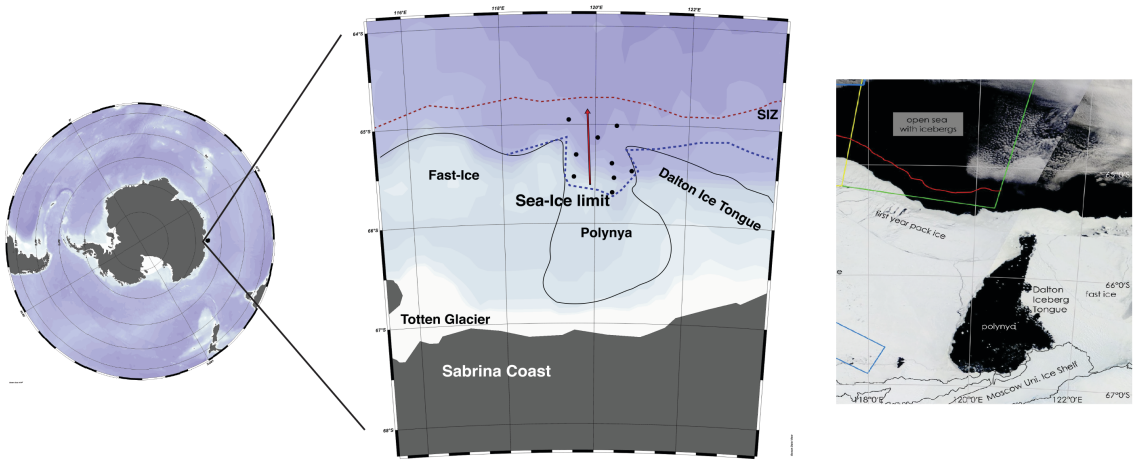



**Figure 4.1** *Map of the sampling site In the Sabrina coast. Red dashed line refers to the seasonal ice zone (SIZ), blue line to the sea ice limit at the time of the sampling. Red arrow indicates the direction of the transect from the site closer to the sea-ice to offshore. Sea-ice report is a composite from the TERRA MODIS (MODerate Resolution Imaging Spectroradiometer) VIS scene provided by NASA and reprinted from Armand at al. 2018. The map was created in OceanData view* (Schlitzer, 2018).

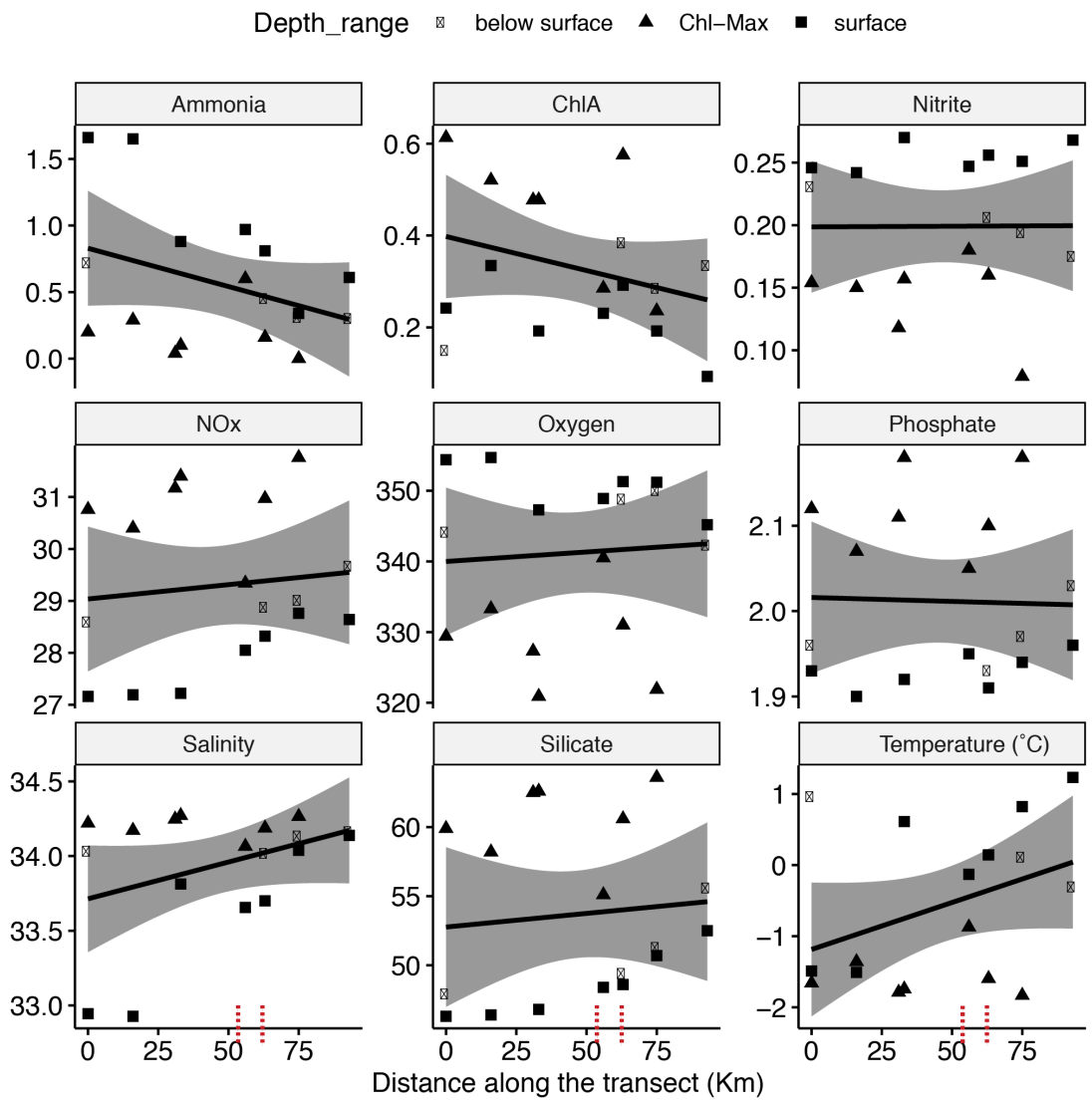



**Figure 4.2** *Chlorophyll, nutrients and physical characteristic of the sites sampled in the Euphotic zone along the transect, the distance was calculated as the distance from site closest to the ice-edge (CTD 25) inside the polynya. All nutrients are expressed in μmol/L, while salinity is expressed in PSU, temperature in ˚C and Chlorophyll in μg/L. Red dashed lines on the x axis correspond to the opening of the polynya. Black circles correspond to the different depth and distance at which the sample was collected.*

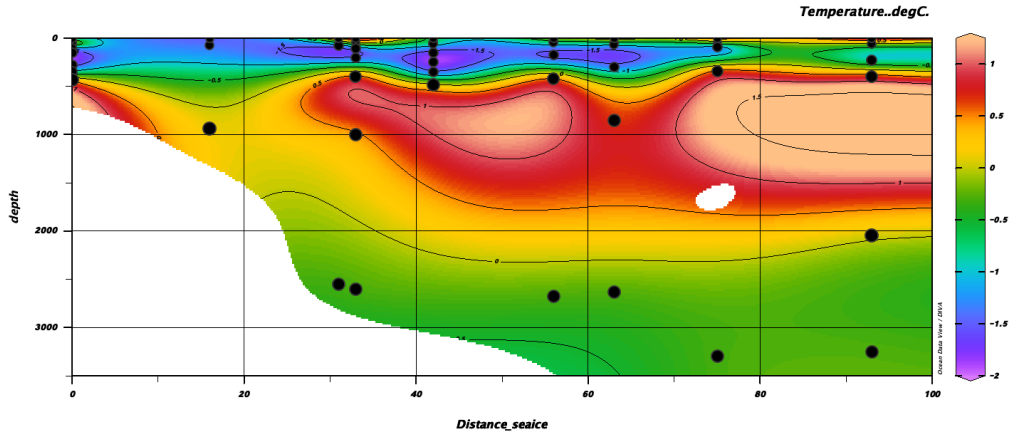



**Figure 4.3** *Integrated temperature profile for the transect (distance from sea-ice) versus depth.*

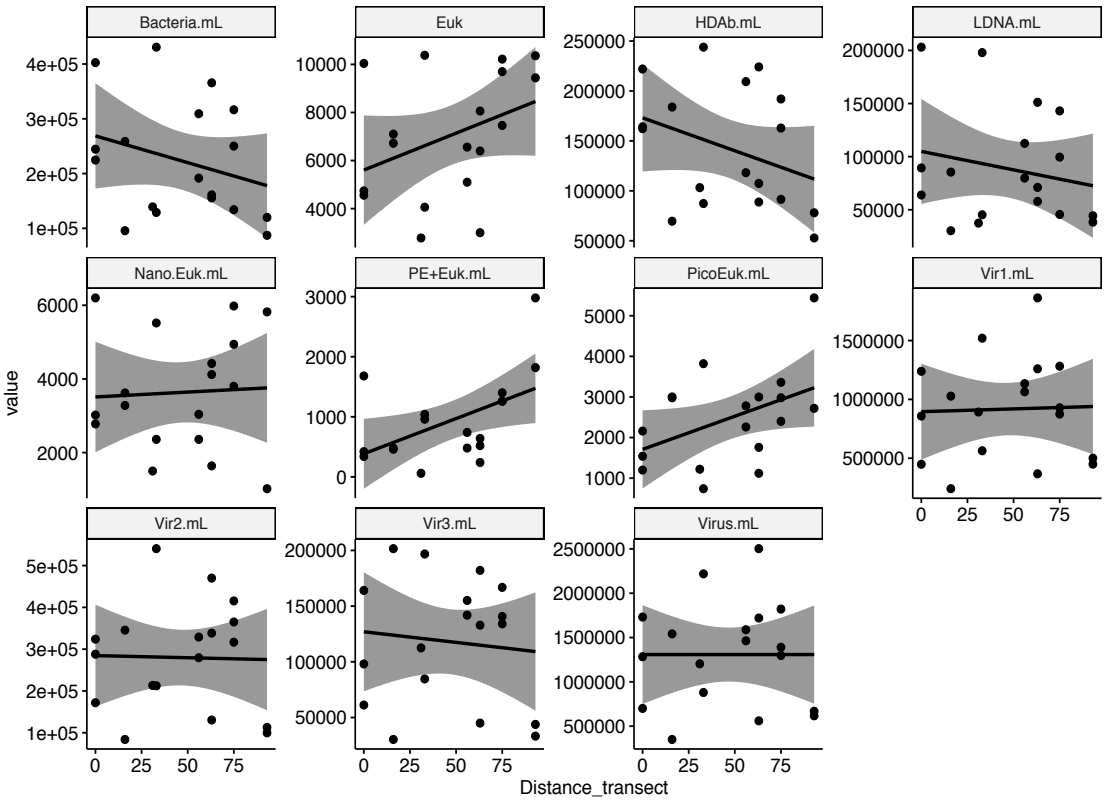



**Figure 4.4** *Integrated distribution of the different microbial groups in the euphotic layers, samples are arrange based on distance from the inner part of the polynya (km). Count are expressed in cell per ml of seawater.*

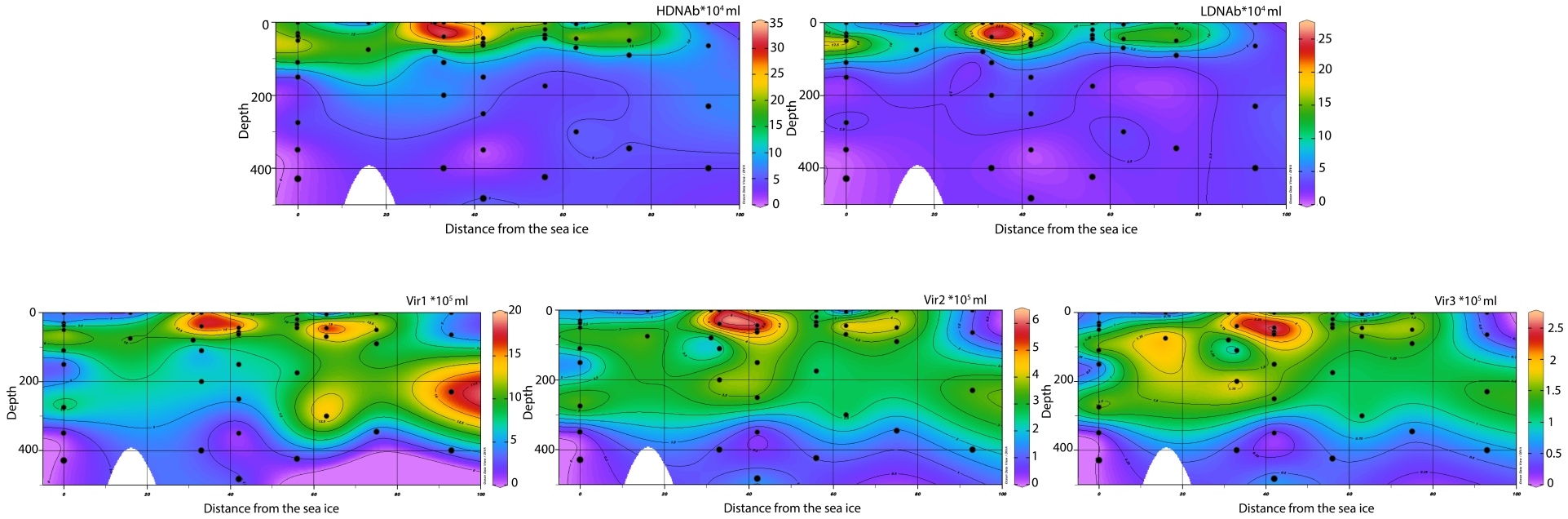



**Figure 4.5** *Integrated distribution of the two bacteria groups (HDNA, LDNA) , and the three different viral groups (Vir1, Vir2, Vir3) identified with FCM analysis. Samples are arrange based on distance from the inner part of the polynya (km) for the first 500m of seawater, count are expressed as number of cell/ml of seawater.*

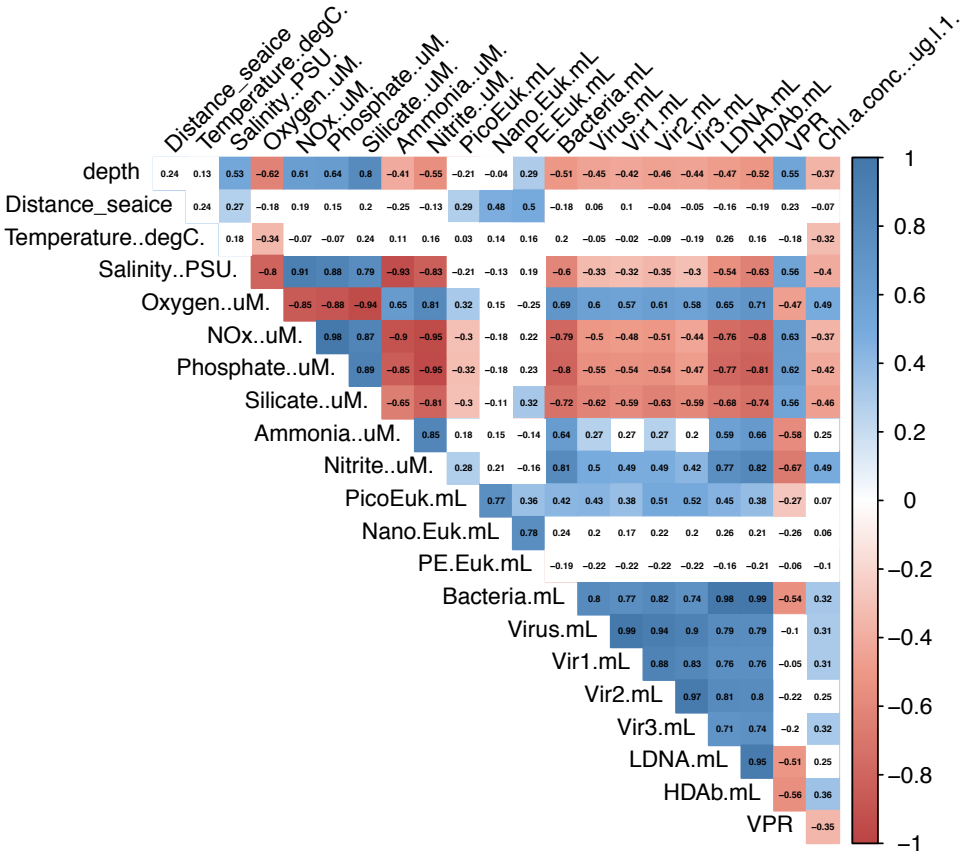



**Figure 4.6** *Spearman's Rho correlation analysis between microbial count obtained with FCM and the environmental parameters. Coloured squares represent significant correlation (p<0.05), while the different colour indicate the degree of correlation red (negative), blue (positive).*

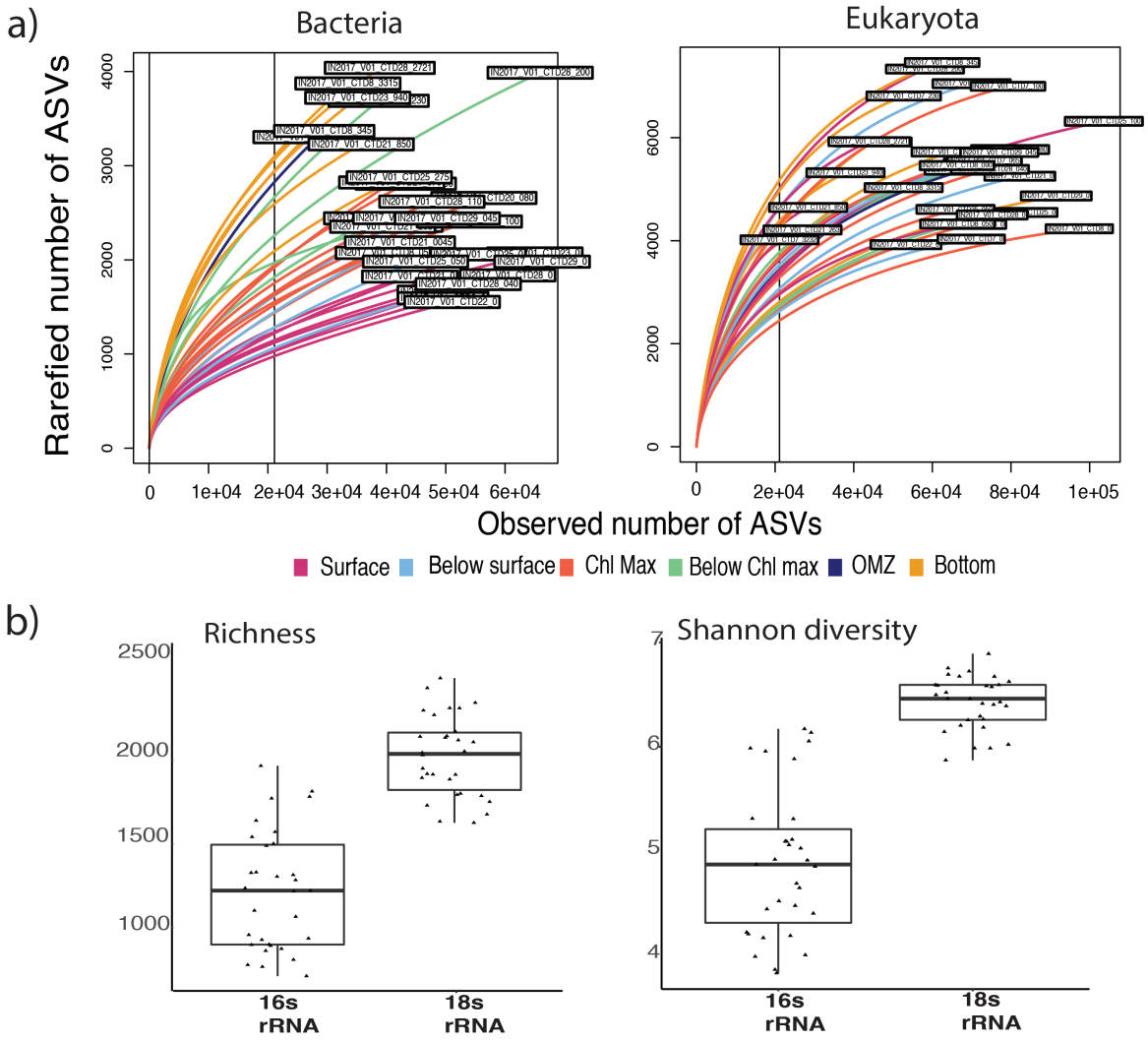



**Figure 4.7** *(a) Rarefication curves of the number of observed 16S rRNA (Bacteria) ASVs versus the rarefied number of ASVs and for the 18S rRNA (Eukaryote) ASVs. Samples were colour coded based on the depths at which the sample were collected. (b) Diversity index plot for bacteria and eukaryote rRNA gene sample Richness (left), Shannon diversity (right)*

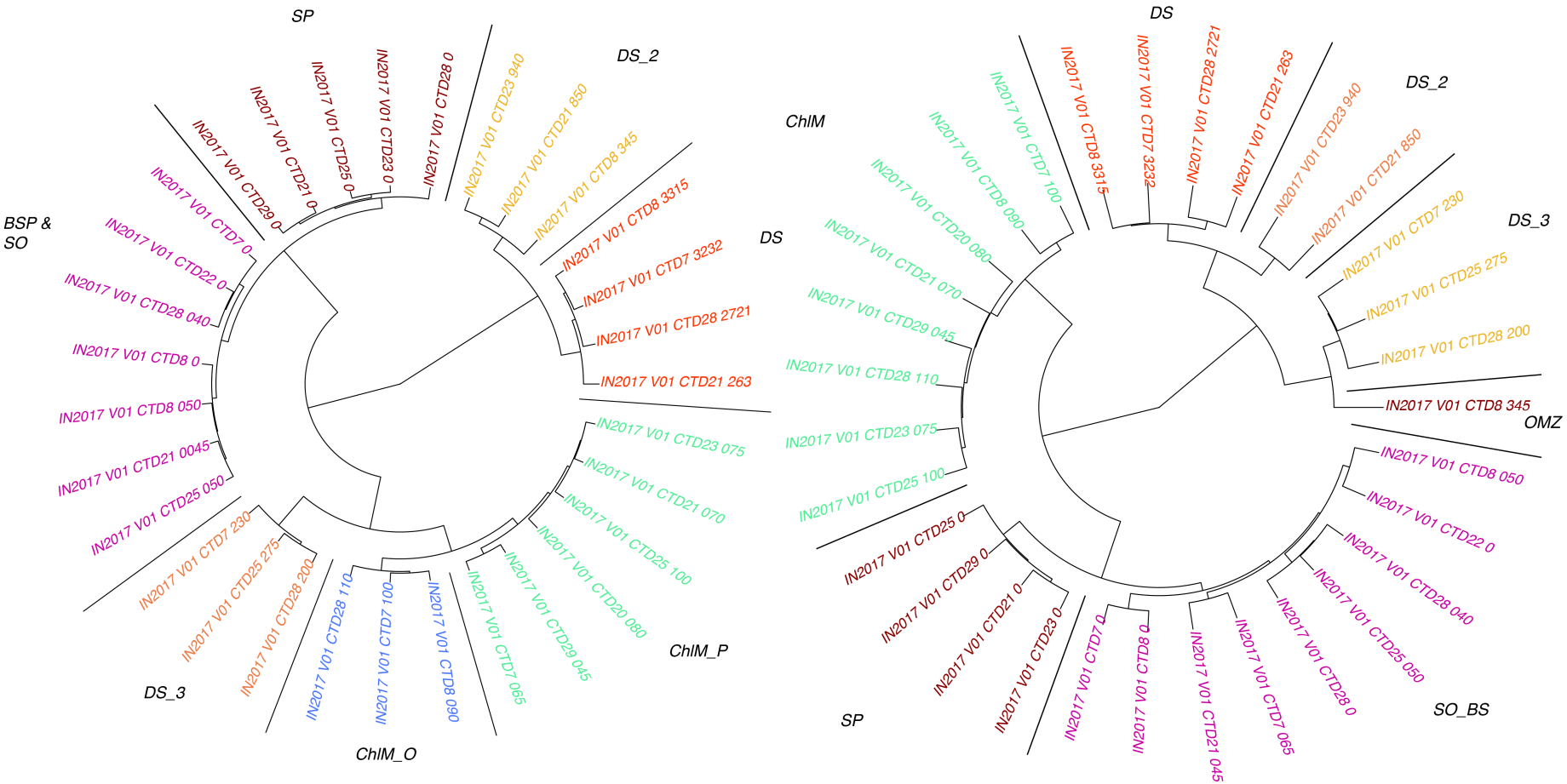

SP
DS_2
BSP &
SO
DS
ChlM_P
DS_3
ChlM_O
ChlM
DS
DS_2
DS_3
OMZ
SP
SO_BS
IN2017 V01 CTD28 0
IN2017 V01 CTD23 940
IN2017 V01 CTD21 850
IN2017 V01 CTD8 345
IN2017 V01 CTD8 3315
IN2017 V01 CTD7 3232
IN2017 V01 CTD28 2721
IN2017 V01 CTD21 263
IN2017 V01 CTD23 075
IN2017 V01 CTD21 070
IN2017 V01 CTD25 100
IN2017 V01 CTD20 080
IN2017 V01 CTD29 045
IN2017 V01 CTD7 065
IN2017 V01 CTD8 090
IN2017 V01 CTD28 110
IN2017 V01 CTD7 100
IN2017 V01 CTD25 0
IN2017 V01 CTD21 0
IN2017 V01 CTD29 0
IN2017 V01 CTD7 0
IN2017 V01 CTD22 0
IN2017 V01 CTD28 040
IN2017 V01 CTD8 0
IN2017 V01 CTD8 050
IN2017 V01 CTD21 0045
IN2017 V01 CTD25 050
IN2017 V01 CTD7 230
IN2017 V01 CTD25 275
IN2017 V01 CTD28 200
IN2017 V01 CTD8 3315
IN2017 V01 CTD7 3232
IN2017 V01 CTD28 2721
IN2017 V01 CTD21 263
IN2017 V01 CTD23 940
IN2017 V01 CTD21 850
IN2017 V01 CTD7 230
IN2017 V01 CTD25 275
IN2017 V01 CTD28 200
IN2017 V01 CTD8 345
IN2017 V01 CTD8 050
IN2017 V01 CTD22 0
IN2017 V01 CTD28 040
IN2017 V01 CTD25 050
IN2017 V01 CTD28 0
IN2017 V01 CTD7 065
IN2017 V01 CTD21 045
IN2017 V01 CTD8 0
IN2017 V01 CTD7 0
IN2017 V01 CTD23 0
IN2017 V01 CTD21 0
IN2017 V01 CTD29 0
IN2017 V01 CTD25 0
IN2017 V01 CTD7 100
IN2017 V01 CTD8 090
IN2017 V01 CTD20 080
IN2017 V01 CTD21 070
IN2017 V01 CTD29 045
IN2017 V01 CTD28 110
IN2017 V01 CTD23 075
IN2017 V01 CTD25 100

**Figure 4.8.** *Circular dendrogram of the significant clusters, for the normalised count of the bacteria 16S rRNA gene sequences (A) and eukaryote 18S rRNA gene sequences (B) samples, identified with SIMPROF analysis, using a Bray-Curtis similarity index. Samples are colour coded based on cluster membership. ChlM= deep chl max, ChlM_P= chl max polynya, ChlM_O= Chl max offshore, SP = Dalton surface , BSP_SO = below surface polynya_surface offshore, SO_BS= Surface offshore_ below surface, DS = deep samples (below 1,000 m), DS_2= deep samples (~1,000 m), DS_3 = deep samples (200-300 m), OMZ = oxygen minimum zone*

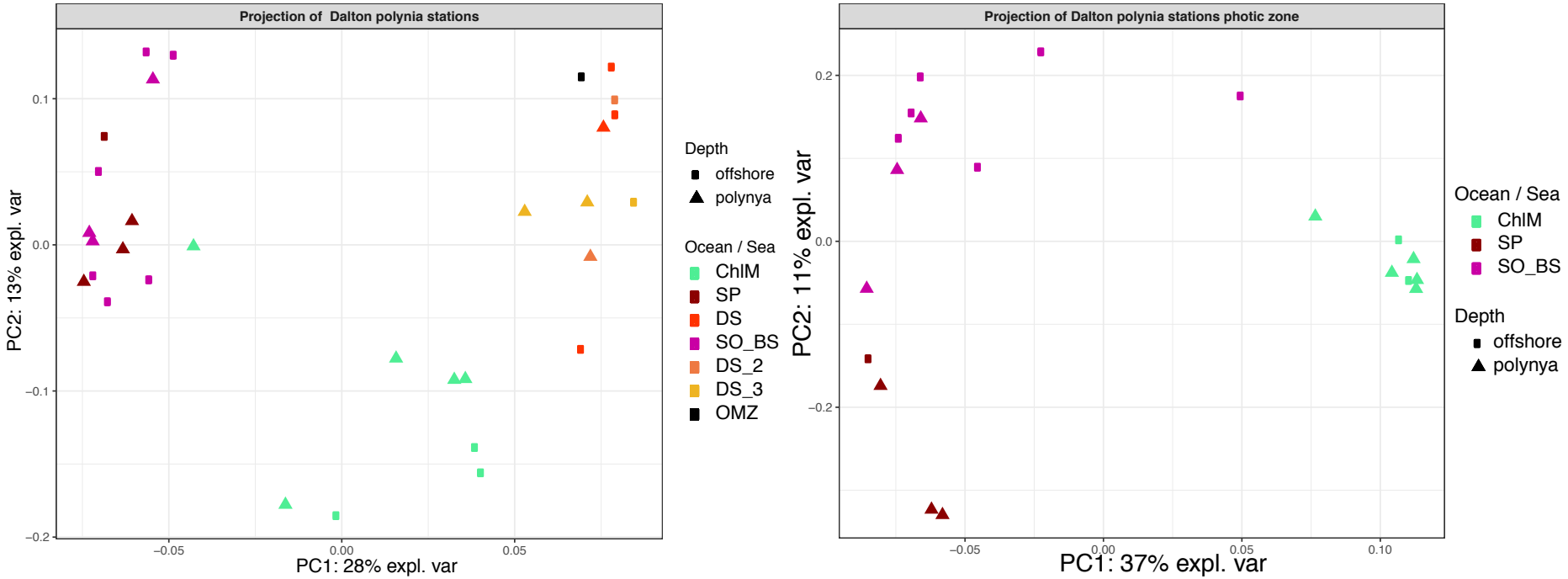
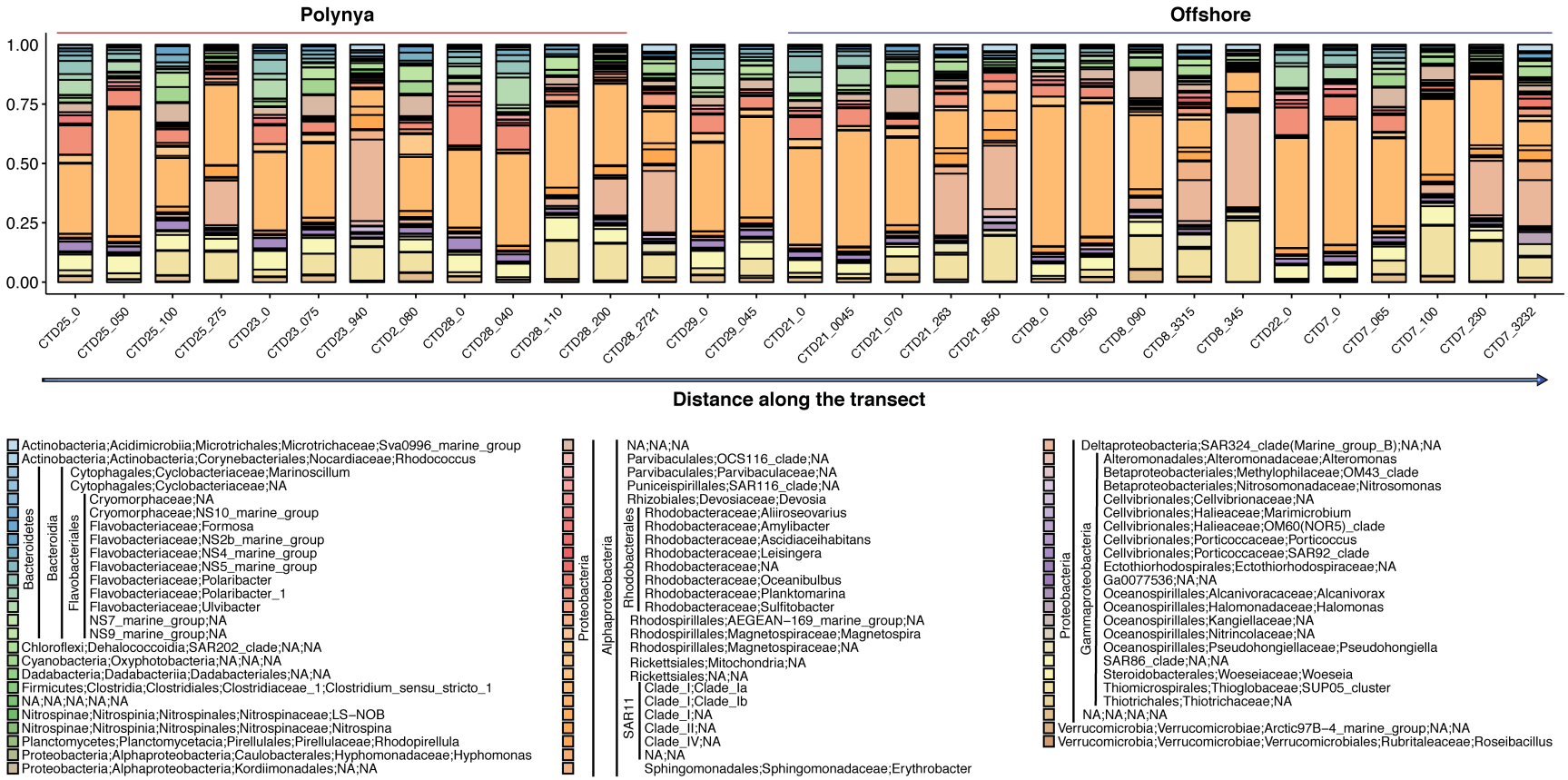



**Figure 4.9** *(a) Ordination plot of the observed dissimilarity in the first two KPCA axes for all sites and depths and. Ordination plot is based on a combination of results from Eukaryotic, Bacterial abundance and environmental parameters calculated using the mixKernel approach (mixOmics). Samples are colour-coded based on hierarchical clusters calculated with SIMPROF test (based on the dendrogram shown in fig 4.4b). Shape are related to either polynya or offshore samples. (b) Ordination plot of the observed dissimilarity in the first two KPCA axes for the euphotic zone (0-200 m) at all sites, based on combination of results from Eukaryotic, Bacterial abundance and environmental parameters calculated using the mixKernel approach (mixOmics).*


Polynya
Offshore
1.00
0.75
0.50
0.25
0.00
CTD25_0
CTD25_050
CTD25_100
CTD25_275
CTD23_0
CTD23_075
CTD23_940
CTD2_080
CTD28_0
CTD28_040
CTD28_110
CTD28_200
CTD28_2721
CTD29_0
CTD29_045
CTD21_0
CTD21_0045
CTD21_070
CTD21_263
CTD21_850
CTD8_0
CTD8_050
CTD8_090
CTD8_3315
CTD8_345
CTD22_0
CTD7_0
CTD7_065
CTD7_100
CTD7_230
CTD7_3232
Distance along the transect
Actinobacteria;Acidimicrobiia;Microtrichales;Microtrichaceae;Sva0996_marine_group
Actinobacteria;Actinobacteria;Corynebacteriales;Nocardiaceae;Rhodococcus
Bacteroidetes
Bacteroidia
Flavobacteriales
Cytophagales;Cyclobacteriaceae;Marinoscillum
Cytophagales;Cyclobacteriaceae;NA
Cryomorphaceae;NA
Cryomorphaceae;NS10_marine_group
Flavobacteriaceae;Formosa
Flavobacteriaceae;NS2b_marine_group
Flavobacteriaceae;NS4_marine_group
Flavobacteriaceae;NS5_marine_group
Flavobacteriaceae;Polaribacter
Flavobacteriaceae;Polaribacter_1
Flavobacteriaceae;Ulvibacter
NS7_marine_group;NA
NS9_marine_group;NA
Chloroflexi;Dehalococcoidia;SAR202_clade;NA;NA
Cyanobacteria;Oxyphotobacteria;NA;NA;NA
Dadabacteria;Dadabacteriia;Dadabacteriales;NA;NA
Firmicutes;Clostridia;Clostridiales;Clostridiaceae_1;Clostridium_sensu_stricto_1
NA;NA;NA;NA;NA
Nitrospinae;Nitrospinia;Nitrospinales;Nitrospinaceae;LS−NOB
Nitrospinae;Nitrospinia;Nitrospinales;Nitrospinaceae;Nitrospina
Planctomycetes;Planctomycetacia;Pirellulales;Pirellulaceae;Rhodopirellula
Proteobacteria;Alphaproteobacteria;Caulobacterales;Hyphomonadaceae;Hyphomonas
Proteobacteria;Alphaproteobacteria;Kordiimonadales;NA;NA
Proteobacteria
Alphaproteobacteria
NA;NA;NA
Parvibaculales;OCS116_clade;NA
Parvibaculales;Parvibaculaceae;NA
Puniceispirillales;SAR116_clade;NA
Rhizobiales;Devosiaceae;Devosia
Rhodobacterales
Rhodobacteraceae;Aliiroseovarius
Rhodobacteraceae;Amylibacter
Rhodobacteraceae;Ascidiaceihabitans
Rhodobacteraceae;Leisingera
Rhodobacteraceae;NA
Rhodobacteraceae;Oceanibulbus
Rhodobacteraceae;Planktomarina
Rhodobacteraceae;Sulfitobacter
Rhodospirillales;AEGEAN−169_marine_group;NA
Rhodospirillales;Magnetospiraceae;Magnetospira
Rhodospirillales;Magnetospiraceae;NA
Rickettsiales;Mitochondria;NA
Rickettsiales;NA;NA
SAR11
Clade_I;Clade_Ia
Clade_I;Clade_Ib
Clade_I;NA
Clade_II;NA
Clade_IV;NA
NA;NA
Sphingomonadales;Sphingomonadaceae;Erythrobacter
Proteobacteria
Deltaproteobacteria;SAR324_clade(Marine_group_B);NA;NA
Gammaproteobacteria
Alteromonadales;Alteromonadaceae;Alteromonas
Betaproteobacteriales;Methylophilaceae;OM43_clade
Betaproteobacteriales;Nitrosomonadaceae;Nitrosomonas
Cellvibrionales;Cellvibrionaceae;NA
Cellvibrionales;Halieaceae;Marimicrobium
Cellvibrionales;Halieaceae;OM60(NOR5)_clade
Cellvibrionales;Porticoccaceae;Porticoccus
Cellvibrionales;Porticoccaceae;SAR92_clade
Ectothiorhodospirales;Ectothiorhodospiraceae;NA
Ga0077536;NA;NA
Oceanospirillales;Alcanivoracaceae;Alcanivorax
Oceanospirillales;Halomonadaceae;Halomonas
Oceanospirillales;Kangiellaceae;NA
Oceanospirillales;Nitrincolaceae;NA
Oceanospirillales;Pseudohongiellaceae;Pseudohongiella
SAR86_clade;NA;NA
Steroidobacterales;Woeseiaceae;Woeseia
Thiomicrospirales;Thioglobaceae;SUP05_cluster
Thiotrichales;Thiotrichaceae;NA
NA;NA;NA;NA
Verrucomicrobia;Verrucomicrobiae;Arctic97B−4_marine_group;NA;NA
Verrucomicrobia;Verrucomicrobiae;Verrucomicrobiales;Rubritaleaceae;Roseibacillus

.

**Figure 4.10** *Bar chart representation of the main Bacterial component in the transect along the Dalton polynya at all the depth samples. Communities composition was obtained by the sequencing of the 16s rRNA V1-V3 region. Here are represented just the ASVs with proportion higher than 1% of the total community. Samples sites are arranged based on distance from the ice-edge, and each deployment is arranged based on sample depth going from surface to deep samples. Red and blue line on top of the bar chart represent samples inside and outside the polynya respectively.*

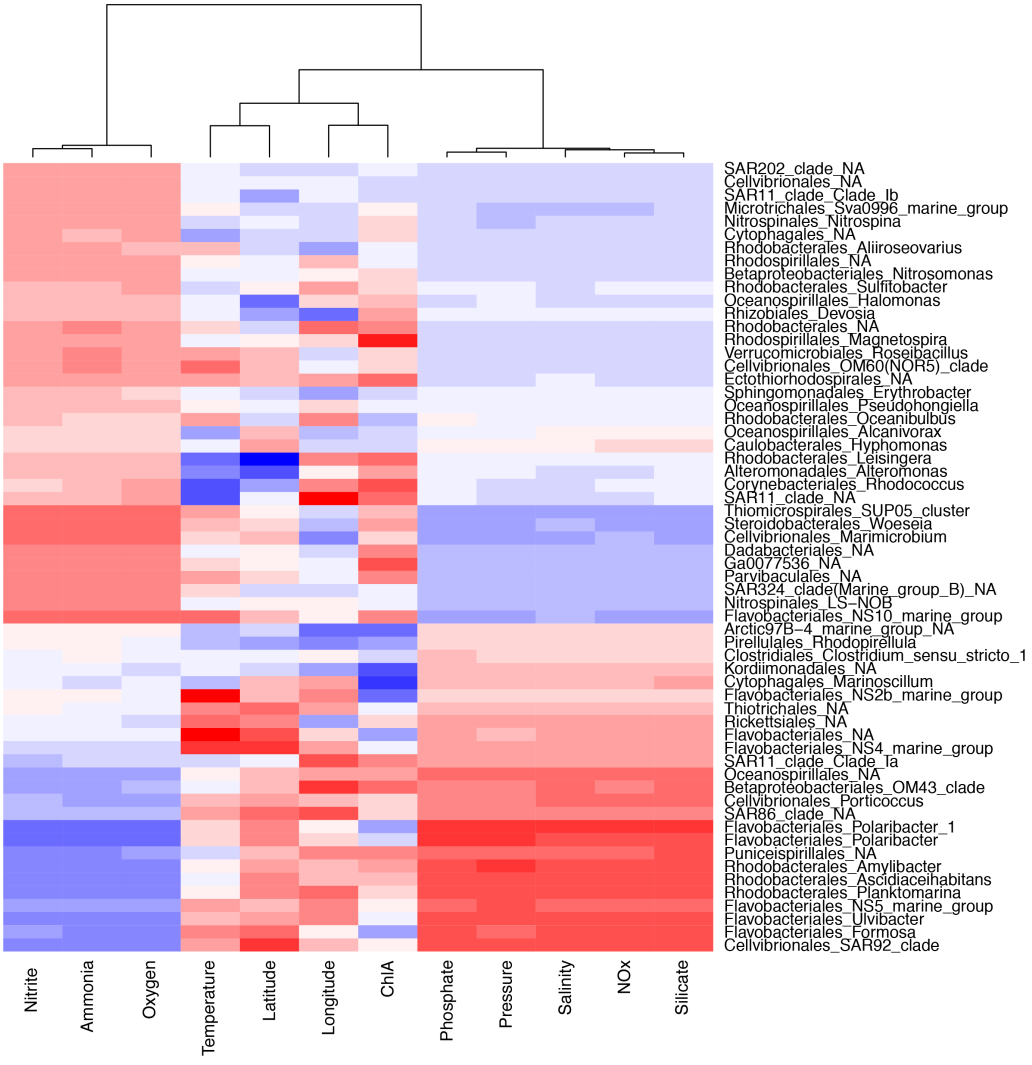



**Figure 4.11** *Heatmap plot of the correlation between abundant bacteria (Genus) and environmetal parameters, just ASVs with an average relative abundance greater than 1% across all samples are shown in the plot. Spearman's correlation coefficients are represented from -1 (red) to 1 (blue).*

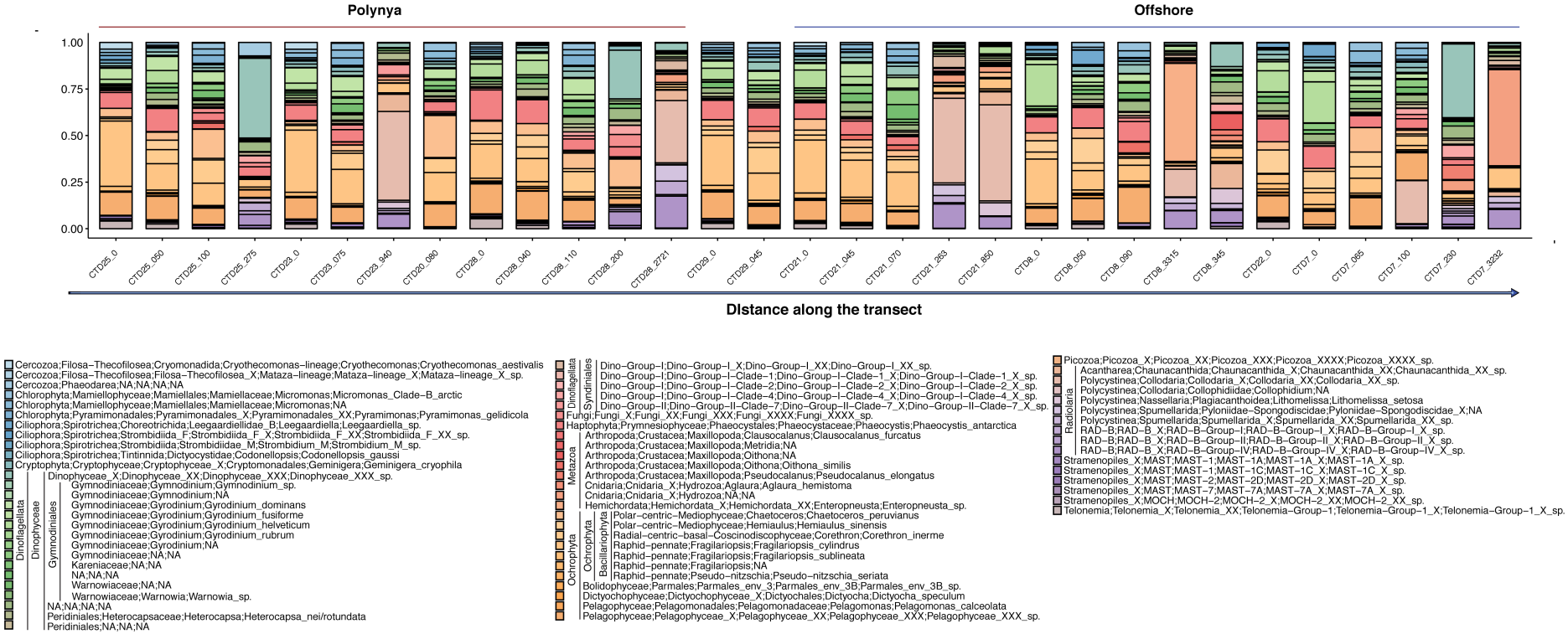



**Figure 4.12** Bar chart representation of the main Bacteria component in the transect along the Dalton polynya at all the depth sampled. Communities composition was obtained by the sequencing of the 16s rRNA V1-V3 region. Here are represented just the ASVs with proportion higher than 1% of the total community. Samples sites are arranged based on distance from the ice-edge and each deployment is arranged based on sample depth going from surface to deep samples. Red and blue line on top of the bar chart represent samples inside and outside the polynya respectively*)*

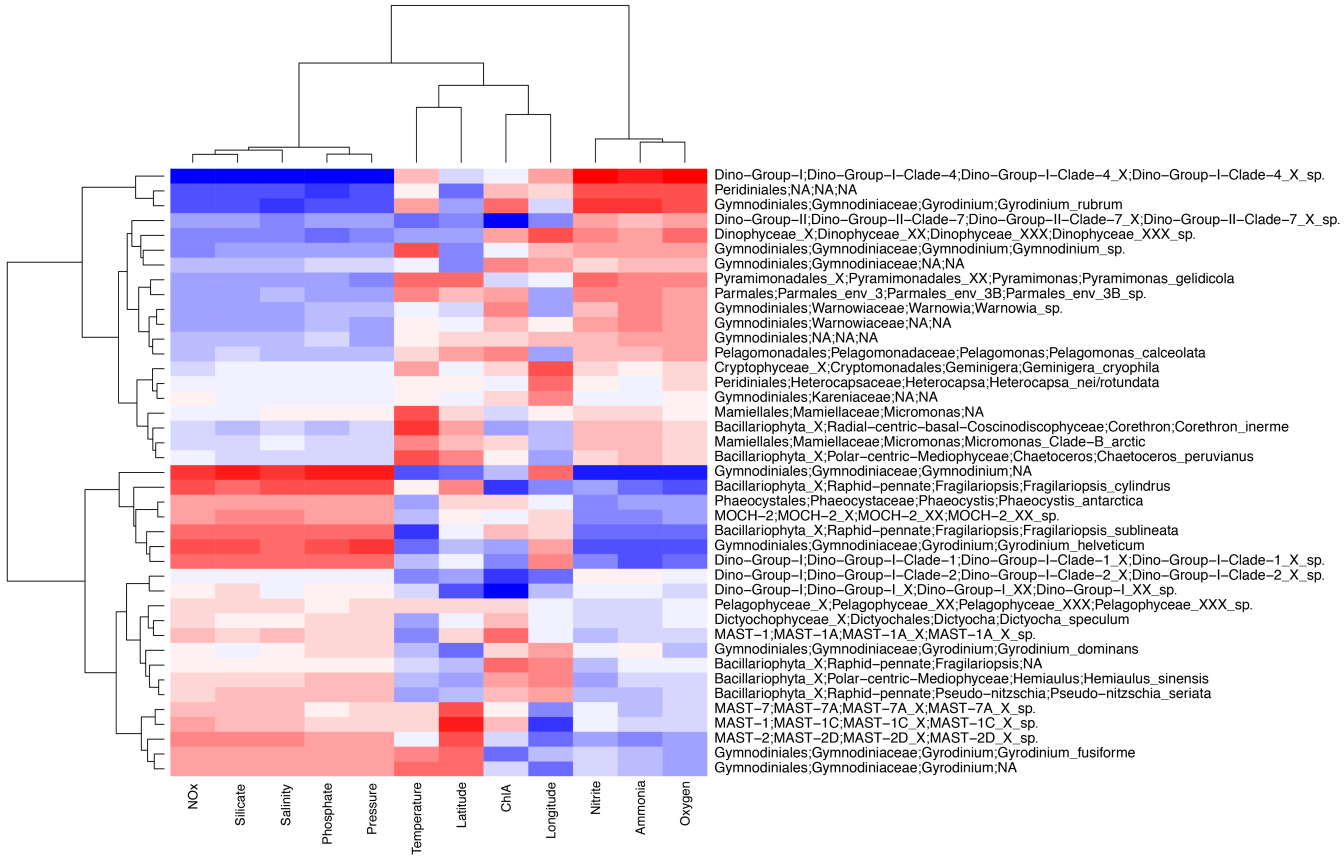



**Figure 4.13** *Heatmap plot of the correlation between the main species of phototrophs, calculated as the one with a concentration higher than 1% in the samples, and the environemtal variables. Correlation coefficient goes from -1 (red) to 1 (blue)*

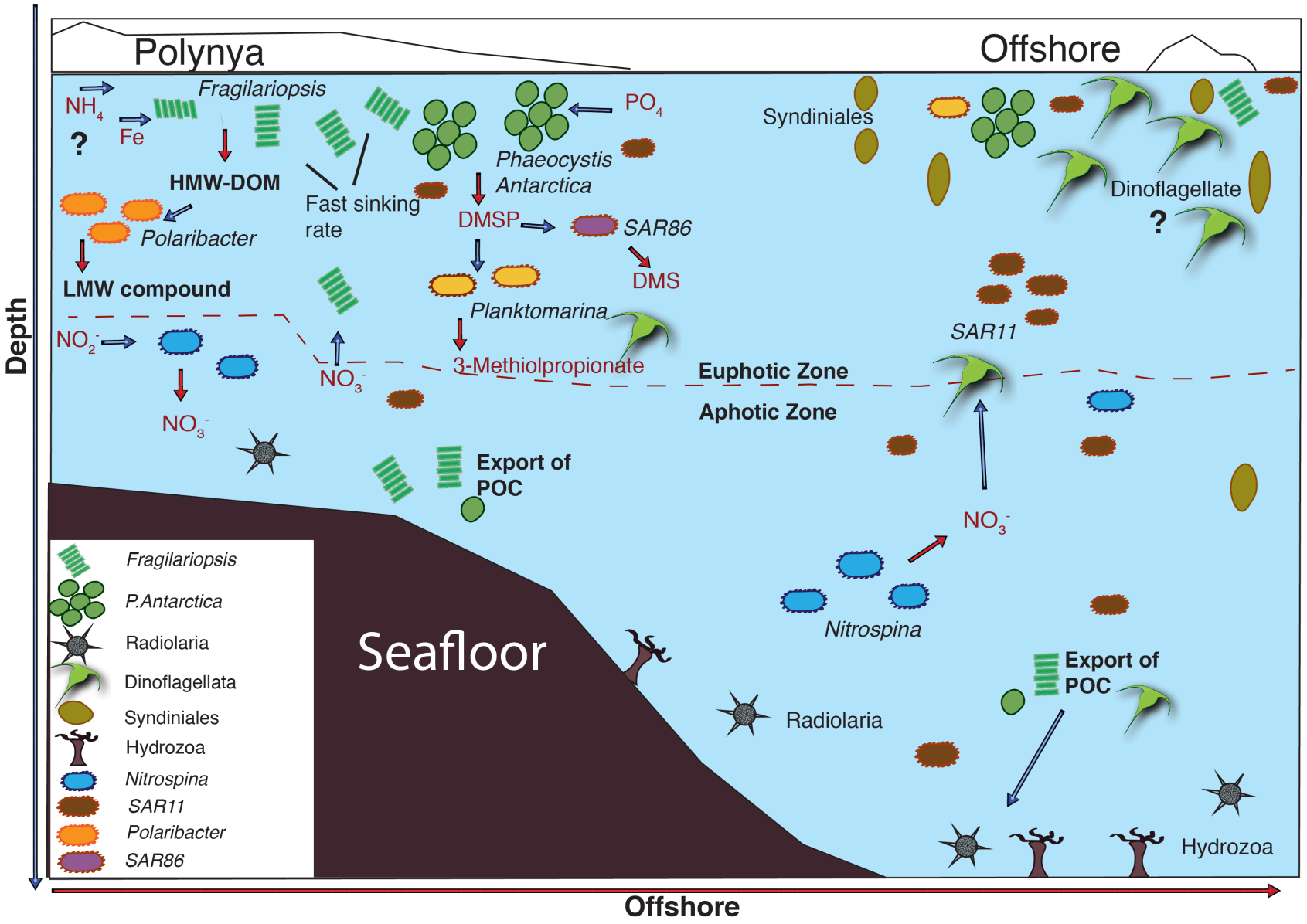

Polynya
Offshore
Depth
Offshore
NH4
Fe
?
Fragilariopsis
HMW-DOM
Polaribacter
LMW compound
NO2-
NO3-
NO3-
Fast sinking rate
PO4
Phaeocystis Antarctica
DMSP
SAR86
DMS
Planktomarina
3-Methiolpropionate
Euphotic Zone
Aphotic Zone
Export of POC
Seafloor
Syndiniales
Dinoflagellate
?
SAR11
NO3-
Nitrospina
Radiolaria
Export of POC
Hydrozoa
Fragilariopsis
P.Antarctica
Radiolaria
Dinoflagellata
Syndiniales
Hydrozoa
Nitrospina
SAR11
Polaribacter
SAR86

**Figure 4.14** *Conceptual model of the main ecological interactions that govern the community inside the Dalton polynya and offshore. The model is oriented on a similar transect from the polynya (left) to offshore (right). Green cells represent all photosynthetic organisms, and the shape is relative to different species. Yellow-orange bacteria are capable of uptaking high molecular weight compounds - dissolved organic matter (HMW-DOM). Blue bacteria are nitrite oxidizing bacteria (NOB). Blue arrow corresponds to uptake and the red arrow to production of particular metabolites. Text in red refers to the main chemicals, while bold black to the main OC-OM. The red-dashed line represents the limit of the photic zone.*

## 4.6 References

Alderkamp, A. C. *et al.* (2012) 'Iron from melting glaciers fuels phytoplankton blooms in the Amundsen Sea (Southern Ocean): Phytoplankton characteristics and productivity', *Deep-Sea Research Part II: Topical Studies in Oceanography*, 71–76, pp. 32–48. doi: 10.1016/j.dsr2.2012.03.005.

Armand, L. *et al.* (2018) 'Interactions of the Totten Glacier with the Southern Ocean through multiple glacial cycles (IN2017-V01): Post-survey report', (March). doi: 10.4225/13/5ACEA64C48693.

Armand, L. K. *et al.* (2005) 'The biogeography of major diatom taxa in Southern Ocean sediments: 1. Sea ice related species', *Palaeogeography, Palaeoclimatology, Palaeoecology*, 233(1–2), pp. 93–126.

Armbrecht, L. H. *et al.* (2017) 'First observations of living sea-ice diatom agglomeration to tintinnid loricae in East Antarctica', *Journal of Plankton Research*, 39(5), pp. 795–802. doi: 10.1093/plankt/fbx036.

Arrigo, K. R. (2003) 'Phytoplankton dynamics within 37 Antarctic coastal polynya systems', *Journal of Geophysical Research*, 108(C8). doi: 10.1029/2002jc001739.

Arrigo, K. R., van Dijken, G. and Long, M. (2008) 'Coastal Southern Ocean: A strong anthropogenic CO2 sink', *Geophysical Research Letters*, 35(21), pp. 1–6. doi: 10.1029/2008GL035624.

Bertrand, E. M. *et al.* (2015) 'Phytoplankton–bacterial interactions mediate micronutrient colimitation at the coastal Antarctic sea ice edge', *Proceedings of the National Academy of Sciences*, 112(32), pp. 9938–9943. doi: 10.1073/pnas.1501615112.

Brown, M. V. *et al.* (2012) 'Global biogeography of SAR11 marine bacteria', *Molecular Systems Biology*. Nature Publishing Group, 8(595), pp. 1–13. doi: 10.1038/msb.2012.28.

Callahan, B. J. *et al.* (2016) 'DADA2: High-resolution sample inference from Illumina amplicon data', *Nature Methods*, 13, p. 581.

Cao, H. *et al.* (2016) 'Delta-proteobacterial SAR324 group in hydrothermal plumes on the South Mid-Atlantic Ridge', *Scientific Reports*, 6, p. 22842.

Chafee, M. *et al.* (2018) 'Recurrent patterns of microdiversity in a temperate coastal marine environment', *ISME Journal*, 12(1), pp. 237–252. doi: 10.1038/ismej.2017.165.

Clarke, L. J. *et al.* (2019) 'A globally distributed Syndiniales parasite dominates the Southern Ocean micro-eukaryote community near the sea-ice edge', *The ISME Journal*, 13(3), pp. 734–737. doi: 10.1038/s41396-018-0306-7.

Delmont, T. O. *et al.* (2015) 'Genome reconstructions indicate the partitioning of ecological functions inside a phytoplankton bloom in the Amundsen Sea, Antarctica', *Frontiers in Microbiology*, 6(OCT), pp. 1–19. doi: 10.3389/fmicb.2015.01090.

Delmont, T. O. *et al.* (2017) 'The global biogeography of amino acid variants within a single SAR11 population is governed by natural selection', *bioRxiv*, p. 170639. doi: 10.1101/170639.

DiTullio, G. R. *et al.* (2000) 'Rapid and early export of Phaeocystis antarctica blooms in the Ross Sea, Antarctica', *Nature*, 404(6778), pp. 595–598. doi: 10.1038/35007061.

Fiala, M. and Oriol, L. (1990) 'Light-temperature interactions on the growth of Antarctic diatoms', *Polar Biology*, 10(8), pp. 629–636. doi: 10.1007/BF00239374.

Fuhrman, J. A. (2009) 'Microbial community structure and its functional implications', *Nature*, 459(7244), pp. 193–199. doi: 10.1038/nature08058.

Giovannoni, S. J. (2016) 'SAR11 Bacteria: The Most Abundant Plankton in the Oceans', *Annual Review of Marine Science*, 9(1), pp. 231–255. doi:

10.1146/annurev-marine-010814-015934.

Grzymski, J. J. *et al.* (2012) 'A metagenomic assessment of winter and summer bacterioplankton from Antarctica Peninsula coastal surface waters', *ISME Journal*, 6(10), pp. 1901–1915. doi: 10.1038/ismej.2012.31.

Guillou, L. *et al.* (2013) 'The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy', *Nucleic acids research*. 2012/11/26, 41(Database issue), pp. D597–D604. doi: 10.1093/nar/gks1160.

Jiang, Y. *et al.* (2013) 'Using pelagic ciliated microzooplankton communities as an indicator for monitoring environmental condition under impact of summer sea-ice reduction in western Arctic Ocean', *Ecological Indicators*, 34, pp. 380–390. doi: 10.1016/j.ecolind.2013.05.026.

Jiang, Y. *et al.* (2016) 'Pelagic ciliate communities within the Amundsen Sea polynya and adjacent sea ice zone, Antarctica', *Deep-Sea Research Part II: Topical Studies in Oceanography*, 123, pp. 69–77. doi: 10.1016/j.dsr2.2015.04.015.

Jiao, N. *et al.* (2010) 'Microbial production of recalcitrant dissolved organic matter: Long-term carbon storage in the global ocean', *Nature Reviews Microbiology*, 8(8), pp. 593–599. doi: 10.1038/nrmicro2386.

Kang, S. H. and Fryxell, G. A. (1992) 'Fragilariopsis cylindrus (Grunow) Krieger: The most abundant diatom in water column assemblages of Antarctic marginal ice-edge zones', *Polar Biology*, 12(6–7), pp. 609–627. doi: 10.1007/BF00236984.

Karnovsky, N., Ainley, D. G. and Lee, P. (2007) 'Chapter 12 The Impact and Importance of Production in Polynyas to Top-Trophic Predators: Three Case Histories', in *Elsevier Oceanography Series*, pp. 391–410. doi: 10.1016/S0422-9894(06)74012-0.

Labrousse, S. *et al.* (2018) 'Coastal polynyas: Winter oases for subadult southern

elephant seals in East Antarctica', *Scientific Reports*, 8(1), pp. 1–15. doi: 10.1038/s41598-018-21388-9.

Legendre, L. and Rivkin, R. B. (2002) 'Fluxes of carbon in the upper ocean: Regulation by food-web control nodes', *Marine Ecology Progress Series*, 242, pp. 95–109. doi: 10.3354/meps242095.

Lepère, C. *et al.* (2016) 'In situ associations between marine photosynthetic picoeukaryotes and potential parasites – a role for fungi?', *Environmental Microbiology Reports*, 8(4), pp. 445–451. doi: 10.1111/1758-2229.12339.

Marie, D., Simon, N. and Vaulot, D. (2007) 'Phytoplankton Cell Counting by Flow Cytometry', *Algal Culturing Techniques*, (January), pp. 253–267. doi: 10.1016/b978-012088426-1/50018-4.

Mariette, J. and Villa-vialaneix, N. (2017) 'Integrating TARA Oceans datasets using unsupervised multiple kernel learning Methods Unsupervised multiple kernel learning', *bioRxiv*, pp. 1–16.

Moran, A. G. *et al.* (2015) 'More , smaller bacteria in response to ocean ' s warming ?', *Proc. R. Soc*, 282, pp. 1–9. doi: http://dx.doi.org/10.1098/rspb.2015.0371.

Mundy, C. J. and Barber, D. G. (2001) 'On the relationship between spatial patterns of sea-ice type and the mechanisms which create and maintain the North Water (NOW) polynya', *Atmosphere-Ocean*, 39(3), pp. 327–341. doi: 10.1080/07055900.2001.9649684.

Needham, D. M., Sachdeva, R. and Fuhrman, J. A. (2017) 'Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters', *ISME Journal*, 11(7), pp. 1614–1629. doi: 10.1038/ismej.2017.29.

Oksanen, A. J. *et al.* (2019) 'vegan: Community Ecology Package'.

Quast, C. *et al.* (2013) 'The SILVA ribosomal RNA gene database project: improved data processing and web-based tools', *Nucleic acids research*. 2012/11/27, 41(Database issue), pp. D590–D596. doi: 10.1093/nar/gks1219.

Quillfeldt, C. Von (2004) 'The diatom Fragilariopsis cylindrus and its potential as an indicator species for cold water rather than for sea ice', *Vie et milieu*, 54, pp. 137–143.

R Core Development Team (2008) *Computational Many-Particle Physics*. Edited by H. Fehske, R. Schneider, and A. Weiße. Berlin, Heidelberg: Springer Berlin Heidelberg (Lecture Notes in Physics). doi: 10.1007/978-3-540-74686-7.

Raes, E. J. *et al.* (2018) 'Oceanographic boundaries constrain microbial diversity gradients in the South Pacific Ocean', *Proceedings of the National Academy of Sciences*, 115(35), p. 201719335. doi: 10.1073/pnas.1719335115.

Rigual-Hernández, A. S. *et al.* (2018) 'Diatom species fluxes in the seasonally ice-covered Antarctic Zone: New data from offshore Prydz Bay and comparison with other regions from the eastern Antarctic and western Pacific sectors of the Southern Ocean', *Deep-Sea Research Part II: Topical Studies in Oceanography*, (December 2017), pp. 1–14. doi: 10.1016/j.dsr2.2018.06.005.

Rohart, F. *et al.* (2017) 'mixOmics: An R package for 'omics feature selection and multiple data integration', *PLoS Computational Biology*, 13(11), pp. 1–20. doi: 10.1371/journal.pcbi.1005752.

Sarmiento, J. L. *et al.* (2004) 'High-latitude controls of thermocline nutrients and low latitude biological productivity', *Nature*, 427(6969), pp. 56–60. doi: 10.1038/nature02127.

Sarmiento, J. L. and Le Quéré, C. (1996) 'Oceanic Carbon Dioxide Uptake in a Model of Century-Scale Global Warming', *Science*, 274(5291), pp. 1346 LP – 1350. doi: 10.1126/science.274.5291.1346.

Schlitzer, R. (2018) 'Ocean Data View', *odv.awi.de*.

Silvano, A. *et al.* (2018) 'Freshening by glacial meltwater enhances melting of ice shelves and reduces formation of Antarctic Bottom Water', *Science Advances*, 4(4), pp. 1–12. doi: 10.1126/sciadv.aap9467.

Swalethorp, R. *et al.* (2019) 'Microzooplankton distribution in the Amundsen Sea Polynya (Antarctica) during an extensive Phaeocystis antarctica bloom', *Progress in Oceanography*, 170, pp. 1–10. doi: 10.1016/j.pocean.2018.10.008.

Tamura, T. *et al.* (2016) 'Sea ice production variability in Antarctic coastal polynyas', *Journal of Geophysical Research: Oceans*, 121(5), pp. 2967–2979. doi: 10.1002/2015JC011537.

Tréguer, P. *et al.* (2017) 'Influence of diatom diversity on the ocean biological carbon pump', *Nature Geoscience*. Springer US, 11(1), pp. 27–37. doi: 10.1038/s41561-017-0028-x.

Wilkins, D. *et al.* (2013) 'Biogeographic partitioning of Southern Ocean microorganisms revealed by metagenomics', *Environmental Microbiology*, 15(5), pp. 1318–1333. doi: 10.1111/1462-2920.12035.

Williams, T. J. *et al.* (2012) 'A metaproteomic assessment of winter and summer bacterioplankton from Antarctic Peninsula coastal surface waters', *ISME Journal*, 6(10), pp. 1883–1900. doi: 10.1038/ismej.2012.28.

Williams, T. J. *et al.* (2013) 'The role of planktonic Flavobacteria in processing algal organic matter in coastal East Antarctica revealed using metagenomics and metaproteomics', *Environmental Microbiology*, 15(5), pp. 1302–1317. doi: 10.1111/1462-2920.12017.

Wing, S. R. *et al.* (2018) 'Contribution of sea ice microbial production to Antarctic benthic communities is driven by sea ice dynamics and composition of functional guilds', *Global Change Biology*, 24(8), pp. 3642–3653. doi: 10.1111/gcb.14291.

Yang, E. J., Jiang, Y. and Lee, S. H. (2016) 'Microzooplankton herbivory and community structure in the Amundsen Sea, Antarctica', *Deep-Sea Research Part*

*II: Topical Studies in Oceanography*. Elsevier, 123, pp. 58–68. doi: 10.1016/j.dsr2.2015.06.001.

Yoch, D. C. (2002) 'Dimethylsulfoniopropionate: Its Sources, Role in the Marine Food Web, and Biological Degradation to Dimethylsulfid', *Applied and Environmental Microbiology*, 68(12). doi: 10.1128/AEM.68.12.5804.

# Chapter 5:

# General conclusion and future directions

## 5.1 Summary

Marine microbial communities play a pivotal role in the global carbon cycle. While primary producers are responsible for the fixation of carbon, a network of connections and competitions with other microbes govern the way organic carbon and energy are utilized and transported within the Ocean. Climate-induced variability in species distribution has the potential to alter the biological carbon pump, connectivity of the food chain and eventually the capability of the Ocean to act as a carbon sink. This thesis aimed to elucidate the mechanisms and interactions that control the carbon flow, in two model ocean regions of the Southern Hemisphere, the EAC, and the Dalton polynya. With the use of both quantitative methods and molecular techniques, this work provides baseline datasets to start to formulate hypotheses on the future productivity of these two regional ecosystems.

The second chapter of this dissertation showed how picophytoplankton that inhabit the highly dynamic EAC system, present a plastic carbon quota, that correlated well with a change in environmental parameters. Elemental quota converted in contribution to the carbon biomass, for each microbial group (< 10 m), highlighted the importance of picocyanobacteria in a warming environment. This increase in importance is, however, accompanied by a higher impact of viral lysis and hence, a possible shift in the network of connections that are at the basis of the carbon flux.

The identity and genomic diversity of the viruses inhabiting the EAC were explored in chapter three. Metagenomic analysis of samples collected within the EAC and upper Tasman Sea system provided useful information about the previously uncharted bacteriophage diversity for these regional ecosystems. There was a clear separation in the communities that inhabit the photic zone of these two regions, both at the taxonomic and functional level. The results highlighted the importance of cyanophage, in the context of the carbon cycle, in the warm environment of the EAC.

Baseline datasets on the diversity of microbial communities are essential in order to be able to understand the network of connections that governs different marine ecosystems, and eventually, to be able to predict the resilience of these network under predicted stressor. Chapter four provides the first comprehensive analysis on both the bacteria and eukaryote communities of the Dalton polynya, a highly productive region in East Antarctica, which is undergoing significant climate-related change. Molecular studies on the taxonomic diversity of these communities were coupled with a quantitative analysis of the smaller microbial fraction (< 10 m). This suggested that a different network of interaction drives microbial diversity inside and outside the polynya.

This dissertation provides three new datasets, that can be integrated with on-going research, both on a regional perspective and a broader scale, to better understand and refine the productivity and connectivity of the food chain under future climate change.

## 5.2 Marine microbes in a changing environment

### *5.2.1 How can these data be used?*

The latest report released by the Intergovernmental Panel on Climate Change (IPCC, 2014) predicted a further global increase in temperature of 1.5˚C more before 2025. There is no doubt that this predicted increased will drastically affect the marine environment and thus the microbes that thrives in it. In the Ocean, biodiversity has a crucial role in maintaining ecosystem function on which humans rely on, such as fisheries (Soliveres *et al.*, 2016). Stability of higher trophic levels in the marine environment relies on the diversity of the major player throughout the whole food web (Soliveres *et al.*, 2016). While microbes are the foundation of the food web, they are traditionally overlooked in conservation effort that usually revolved around specific organisms, that have a higher economic impact.

***Regional perspective***

On a regional perspective, data on the carbon quota and rates of growth and mortality from chapter two can be integrated into Ocean model for refining the future productivity of the EAC region. The EAC system encompasses some of the most productive regions, in term of fish stock, for the Australia marine economy (Hobday and Hartmann, 2006; Hobday *et al.*, 2011).

So how are these predicted shifts in the microbial community going to impact fisheries? Will picocyanobacteria production sustain the large fish stock that is found primarily at the jet between the EAC and the Tasman Sea?

Even if more pieces of information are needed to answer these questions, the integration of refined data on carbon biomass and connectivity of the microbial web, with knowledge of fisheries stock, could help in understanding the possible future implication of climate change on fish stock and dynamics. On a more political side, results from such integration will help in decision making for the future maintenance of ecologically sustainable fisheries, however, more efforts to link the growing molecular and microbial data with ecosystem and fisheries monitoring programs is required.

***Broader perspective***

The last decade has seen massive advances in the characterization of microbes and especially viruses in different environments, mostly because of the advancement in the field of metagenomics (metaviromics). Nowadays the primary role of viruses in the functioning of diverse ecosystem has been recognised widely (Suttle, 2007; Sunagawa *et al.*, 2015; Guidi *et al.*, 2016; Roux *et al.*, 2016; Breitbart *et al.*, 2018). The majority of recent breakthroughs in the understanding of the structure and diversity of marine microbial communities, are based on data collected for OSD, *Tara* oceans, BATS, HOT, SPOT(Williamson *et al.*, 2008; Kopf *et al.*, 2015; Sunagawa *et al.*, 2015; Luo *et al.*, 2017; Gregory *et al.*, 2019; Sieradzki *et al.*, 2019) which are or were all based in the Northern Hemisphere. By stark comparison, the Southern hemisphere is chronically under-sampled, for example, of the 191 sampling sites of OSD just 19 were situated in

the Southern Hemisphere. As a result, a significant portion of the ocean, in particular, the Southern Pacific, Indian and Southern Ocean basins remain poorly characterised, at least from a microbial and viral perspective. The southern hemisphere biome presents unique features, such as a chronically iron-poor region, highly dynamic boundary current systems (e.g. EAC), and Antarctica. Generalisations, or models, built upon observations that are heavily biased towards the Northern Hemisphere are potentially inaccurate when it comes to describing the Southern Hemisphere biome, or either not complete when it comes to investigating global responses upon different stressors. The implementation of models, for a generalised response across oceans, requires further ground-truthing, particularly focussing on the under-sampled regions of the Southern Hemisphere, as presented in this thesis.

The datasets presented in this dissertation try to narrow this gap. Chapter three contributed more than 400 novel bacteriophage scaffolds (>15 kbp) that represent complete or almost complete genomes, as well as 26,000 viral contigs (1-190 Kbp), from a region where relatively little or nil is known about the viral diversity. Chapter four presented the first molecular description on the diversity of bacteria and eukaryotes that inhabit a crucial area from Antarctica. While more work needs to be done, the dataset presented in this dissertation can be implemented, together with results from the Northern Hemisphere, into global models.

## 5.3 Future directions

### *5.3.1 Implications for variability in Picocyanobacteria elemental quotas*

One of the main aims of chapter two was to understand the possible implication of temperature increases for the elemental composition of natural populations of picocyanobacteria. In the context of how this information can be converted into a better understanding of the carbon flux within the EAC system.

This study was one of the first to use calculated carbon content, and the first one in the context of the EAC. The results highlighted a highly variable carbon quota that well correlates with a change in environmental parameters (Chapter 2) and, especially temperature for *Synechococcus,* related to a lower Qc in the EAC (Fig. 5.1). What are the implications of this finding on a general scenario?

As temperature is expected to rise in the future, we predict this will lead to an increase in numbers of picocyanobacteria, but also a lower carbon quota for *Synechococcus*. Moreover, the observed shift from HDNA bacteria, that dominated Tasman Sea, in favour of LDNA bacteria in the EAC (Chapter 2), suggests a restructuring of the main microbial groups that are responsible for organic matter recycling. A shift towards LDNA bacteria fraction, that is usually dominated by SAR11 could affect the functioning of the marine ecosystem, towards an enhancement of the microbial carbon pump role in the carbon cycle.

On a regional scale the lower amount of carbon for *Synechococcus*, will results in them being a less nutritional source of food, with possible impacts for higher trophic levels. As highlighted previously, the EAC system is a highly active region that encompasses or influence some of the most important regions for the Australian marine economy, in term of fish stock. Hence, understanding how a change in the main primary producers is going to affect the whole food chain is of fundamental importance.

More work needs to be done to refine the constraints on primary producer carbon quota and how to integrate this into biogeochemical models. While our experiment targeted a “latitudinal” gradient, information on temporal succession and dynamics are also fundamental to have a complete understanding of the contribution of picocyanobacteria to the carbon biomass.

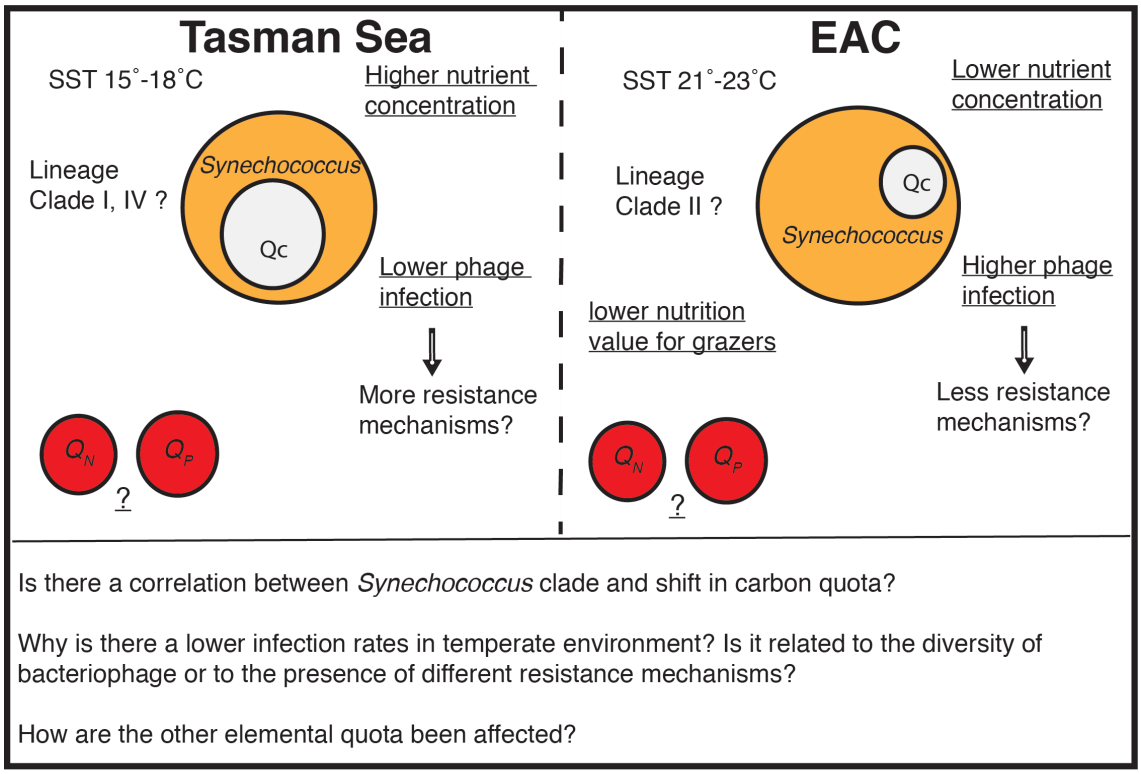



**Figure 5.1** *Schematic representation of the main findings from chapter two and possible future directions. The two sections correspond to the two different regions (Tasman Sea and EAC).* $Q_N$ *(quota of nitrogen)* $Q_P$ *(quota of phosphorus).*

The establishment of a microbial monitoring program across a set of National Reference Station Time series (Brown *et al.*, 2018), that include monthly-based sampling, provide the perfect avenue for future work on the elemental composition. Three of the stations situated on the East Coast of Australia are impacted by the EAC, either directly or by its seasonal southerly extension, NSI (North Stradbroke Island) in front of Brisbane, PHB (Port Hacking) Sydney, and MAI (Maria Island) in Tasmania. Monthly sampling and a seasonal overview of the Qc analysis for picocyanobacteria in these three different regions could help to strengthen our hypothesis on the impact of temperature raising on different elemental composition. It would also help to expand the analysis to a global cellular response, in term of change in $Q_P$ (quota of phosphorus) and $Q_N$ (nitrogen quota).

Based on correlation with genomic analysis from samples collected at the same time, one of the hypotheses of chapter two was that the variability in cyanobacteria Qc is likely a combination of a response to environmental variables and niche distribution of different clades. While, from field data on Qc variability, is it difficult to understand if the Qc change reported is a general response of the total community of *Synechococcus*, or is the response from a specific clade. The integration of field data with laboratory experiments on different clades could help us understand how the different ecotypes adjust their elemental quota in relation to multiple stressors, such as temperature, phosphate stress, nitrogen stress, and iron limitation. Results from experimental procedures could then be validated with the integration of temporal Qc data, from monthly sampling at the NRS, and with results from the succession of the different ecotypes from genomic data.

Data on elemental composition is needed to refine regional biogeochemical models. A better numerical understanding of the plasticity of carbon quota and its response to environmental stressors could be linked to the biogeography of different clades, with the aid of machine learning approaches, to better predict the productivity of a given region, and in this case the EAC system. While in a broader scale these data need to be incorporated into carbon models to better understand the future of the Ocean as a carbon sink.

Picocyanobacteria in the EAC showed an increase in viral lysis as well, that correlated well with a higher number of cyanophage detected in the metagenome from the EAC (Chapter 3). The higher infection rates could either be linked to different bacteriophage community (Chapter 3), but also the different picocyanobacteria clades that thrive in the EAC system could have fewer resistance mechanisms to fight viral infection, compared to the clades found in the Tasman Sea. Due to the streamlined genomes of both *Synechococcus* and *Prochlorococcus,* the fitness cost of having resistance traits against viral infection could be too large. Especially in an area like the EAC, where competition for nutrients is already high, and energy is revolved around expressing transporter for better uptake of essential nutrients. More work could be done to strengthen the results for the degree of viral infection for cyanobacteria in the EAC compared to the Tasman Sea.

### *5.3.2 Future work on viral activity within the EAC system*

AMGs are widespread in viruses and are involved in a variety of metabolic activity. Thus, through the expression of these genes, viruses could potentially alter ecosystem productivity and have a pivotal role in all the biogeochemical cycles (Suttle 2007).

For some of the better-studied genes, such as *psbA*, the viral version and the host one are usually quite discernible (Fig. 5.2) and can be used as a quantitative tool to provide insight into the degree of viral infection (Sieradzki *et al.,* 2019). An integration of metatranscriptomic data with the viral metagenomic results is currently undergoing. While looking at the expression of specific genes in both cyanophage and their cyanobacteria counterpart (*psbA*) will help in refining the results found in chapter 2. The analysis could broaden to capture more information also on the infection of other important bacteria in this regional environment, such as SAR11 for the EAC or *Roseobacter* for the Tasman Sea.

The EAC system is a highly active region that encompasses or influences some of the most important regions for the Australian marine economy. To understand how climate change is going to affect the structure of the marine food web that sustains this highly productive area, a critical element is to integrate knowledge of the viral diversity and activity with studies of the whole microbial ecosystem. This study was the first to look at the viral diversity of the EAC system and it provides a baseline database to start to understand virus (DNA) diversity within the East Coast of Australia. But what can this data tell us? One interesting further development would be to integrate this novel dataset with data on the microbial diversity of this region. Such an integration would help to better understand the role of the different microbial groups in structuring this regional ecosystem.

Ideally, data already available from three years of monthly metagenomic sequencing from the three National reference stations mentioned above (part of the Australian Microbiome framework data initiative) could be investigated to understand temporal variability of the bacteriophage diversity at these three sites. Results from chapter 3 showed that metagenomes from the microbial fraction

(0.2 μm), even though not enriched for DNA viruses, provide a sound basis for an understanding of the main phage (DNA) diversity. Bacteriophage temporal dynamics integrated with bacteria results will help us better understand the community dynamics.

Moreover, as mentioned above the monthly-based sampling, at the seven NRS provides for the perfect avenue for future work. Ideally the on-going monthly sampling for the microbial community could be integrated with a novel virus-focussed sampling effort to target both the DNA and RNA viruses within the Australian Marine Microbial Biodiversity Initiative (AMMBI). This could provide further insight on the total diversity, dynamics and temporal variability of total virus communities around Australia.

Another interesting question is: why are cyanobacteria in the EAC more susceptible to viral infection? As discussed above either cyanobacterial cell numbers and presence/absence of resistance mechanisms for the different picocyanobacteria lineages could play a role in the different infection rates registered. Bacteriophage host specificity could also lead to different infection rates. Isolation of bacteriophage from these two regional environments and screening of purified phages against numerous *Synechococcus* lineages, for example, could provide insight into the differences in bacteriophage host range. Temporal studies of environmental infection rates coupled with environmental and laboratory-based transcriptomic analysis could help understand the mechanisms of resistance utilize by the different *Synechococcus* lineages.

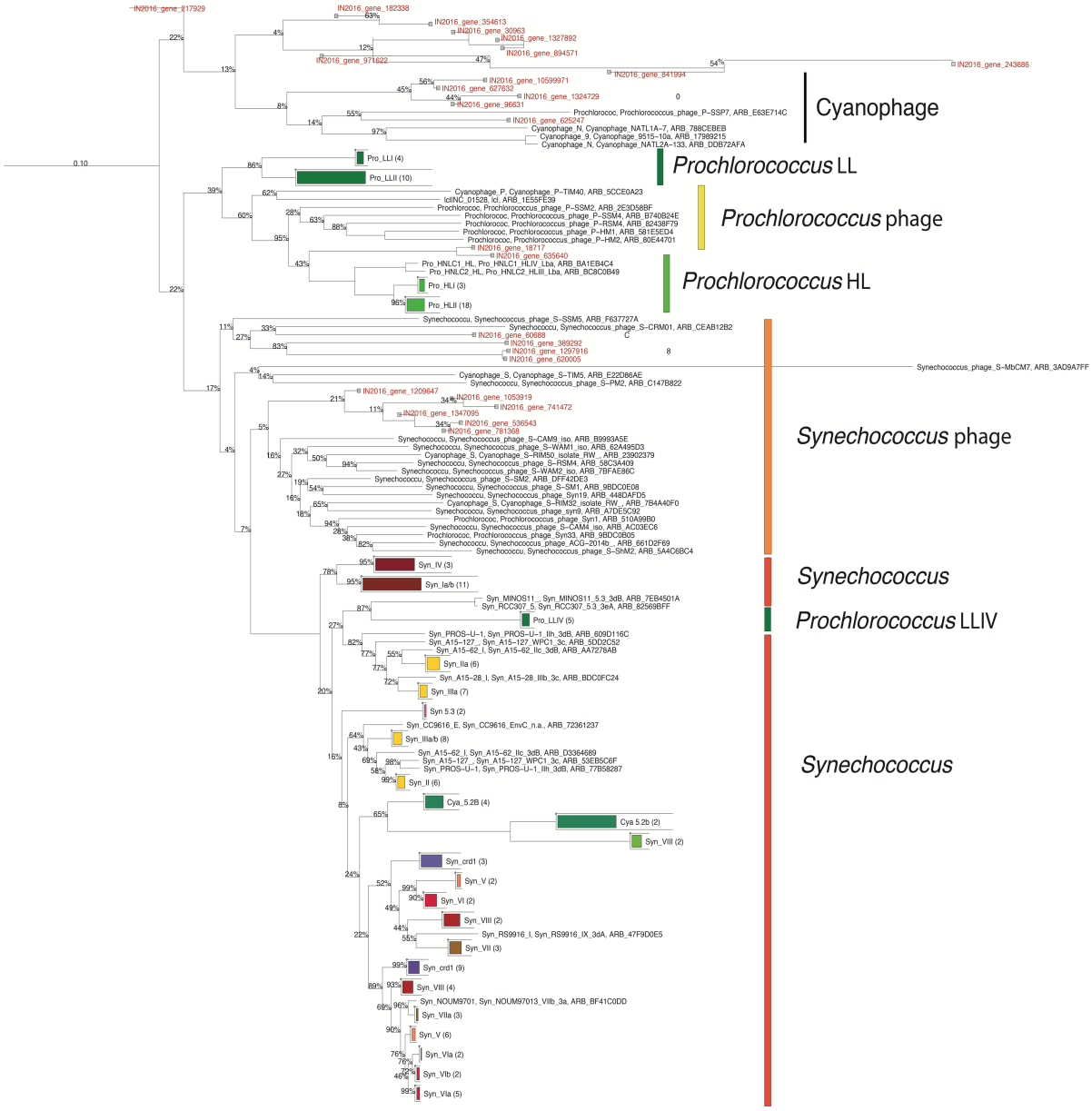



**Figure 5.2** *Phylogenetic tree of psbA sequences from Synechococcus, Prochlorococcus and bacteriophage infecting them. The phylogenetic tree was constructed aligning 90 psbA sequences from picocyanobacteria, 31 phage sequences present in NCBI database and novel psbA sequences (text in red) derived from the data in Chapter 3. Alignment was performed in MAFFT and tree constructed in ARB. Different colours on the tree leaf group lineages that thrive in similar environments. Number in bracket is related to the number of sequences for that leaf. Colour bar on the right groups the different Synechococcus, Prochlorococcus or phage.*

### *5.3.3 Dalton polynya – a benchmark dataset and future work to start to understand microbial network.*

This is the first study that integrates molecular data for bacteria and eukaryotic communities, with high-resolution flow cytometry analysis to quantify the bacteria, viruses and picophotosynthetic abundance in such an important area for primary productivity and the carbon cycle in East Antarctica (Chapter 4). Future climate change will trigger an increase in glacier melting. It is not possible to address any question on how the community is going to respond without a baseline understanding of the microbial community structure in the context of these areas. So, this benchmark dataset provides the basis to investigate further microbial interaction in a region undergoing significant change due to sea-ice melting, where temperature, salinity, and stratification seem to be the main drivers in shaping the community interaction.

Further work is still undergoing to better characterised the network of interaction that structure the diverse microbial community encounter within the polynya and offshore. The data presented here was just a subset of the full dataset of the samples collected during a 52 days expedition. Twenty-four more sites were sampled (5-8 depths each) and 54 virome samples collected along the Sabrina coast. While the analysis of the virome will begin shortly, analysis, for both the genomic and cytometry samples, from the remaining 24 sites has already been done.

Moreover, deep sequencing metagenomic analysis was carried out for the same subset of samples inside the Dalton polynya. Results from the metagenome will provide further inside on the different metabolic pathway used inside and outside the polynya, possibly strengthening the hypothesis of the community functioning highlighted in the discussion of Chapter 4.

One of the main problems when sampling remote regions where relatively little is known about the microbial community diversity and functioning, is the lack of reference or similar species in the database. In this scenario, metagenome assembled genomes provides a new efficient way for: characterizing novel species, but also provide an insight into the genetic makeup of both bacteria and

eukaryotes that are adapted to thrive in such a diverse environment. From preliminary results on the co-assembly of the contigs from the polynya, we retrieved at least 500 almost complete genomes, that includes bacterial, eukaryote, and viral genomes.

***How can microdiversity help us to better characterised community in the polynya and more broadly to understand species adaptation?***

While chapter three provide a general overview about the structure of the community and formulate hypotheses on the network of interactions that are at the basis of the species diversity found. However, the majority of the species (eukaryote) or genus(bacteria) identified showed an extensive microdiversity.

For examples 450 ASVs belonged to the genus *Fragilariopsis,* including 200 *Fragilariopsis cylindrus* and 190 *Fragilariopsis sublineata* ASVs. Sequences alignment of the 69 more abundant *Fragilariopsis* ASVs (count > 5000) identified consistent specific base changes that distinguished between *F.sublineata* and *F.cylyndrus*. Moreover, the consensus phylogenetic tree constructed from the alignment showed that within each *Fragilariopsis* species there were possible distinct lineages. The same could be said for the most abundant bacteria SAR11. More than 144 abundant ASVs (count > 500) were identified for SAR11 clade Ia, where specific lineage subgroups showed differential distribution between sites. But, what can this microdiversity at the species level tell us? Are these base changes a neutral mutation or a specific adaptation to different conditions? More work is currently on-going to better understand the distribution of these different variants.

Ideally results from this study will be integrated with other molecular work done in different locations around Antarctica and the Southern Ocean to better understand : i) a biogeography of the community that inhabit such remote areas, ii) if the specific variants that we encounter in the polynya are also present elsewhere and if yes which are the drivers of their distribution.

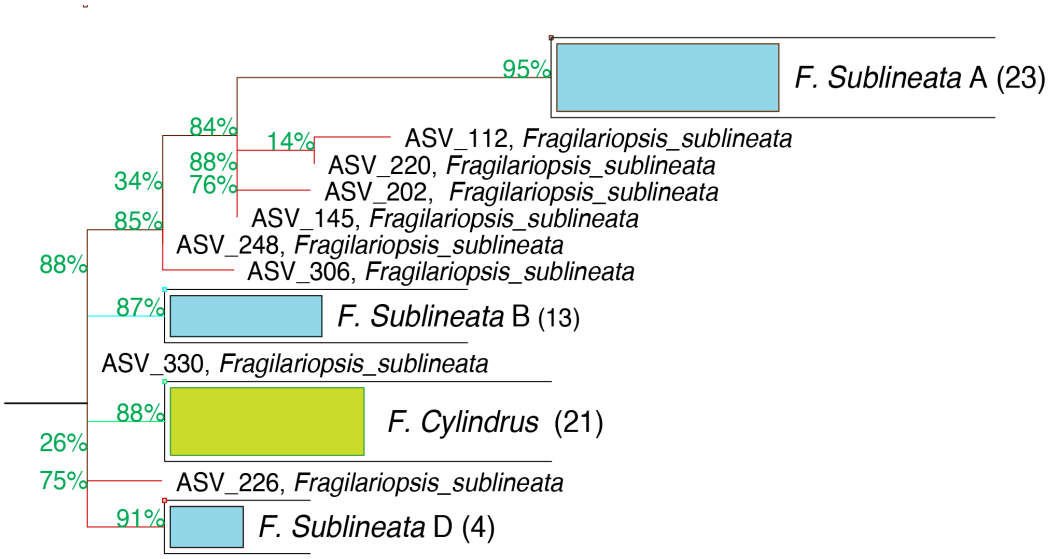



**Figure 5.3** *Fragilariopsis phylogenetic tree. Sixty-nine abundant ASVs belonging the the Fragilariospsis genus were used to construct the phylogenetic tree. Reads were aligned using MAFFT, the alignment was curated in arb, and the consensus phylogenetic tree was constructed using PhyML and Neighbor-joining approaches. Branch labels represent confidence levels calculated by PhyML. The two species are identifed with different colour. Number*

## 5.4 Towards a better quantitative understanding of the microbial communities of the Southern Pacific.

During my candidature I was involved in the analysis of over 1,500 cytometry samples collected during the oceanographic expedition (IN2016_V04, IN2017_V01) and one collaborative project for samples collected on the GO-SHIP expedition (Raes *et al.*, 2018) (Fig. 5.4). Multiple depths (8) are present for almost all location, going from surface to 4000 m depth. A standard protocol was used to counts the different samples. For each sample from three to five groups of viruses were identified, and LDNA, HDNA bacteria. For the phototrophic fraction: two groups of *Prochlorococcus*, two of *Synechococcus* and three different eukaryotes (PEuk, NEuk, PE-Euk).

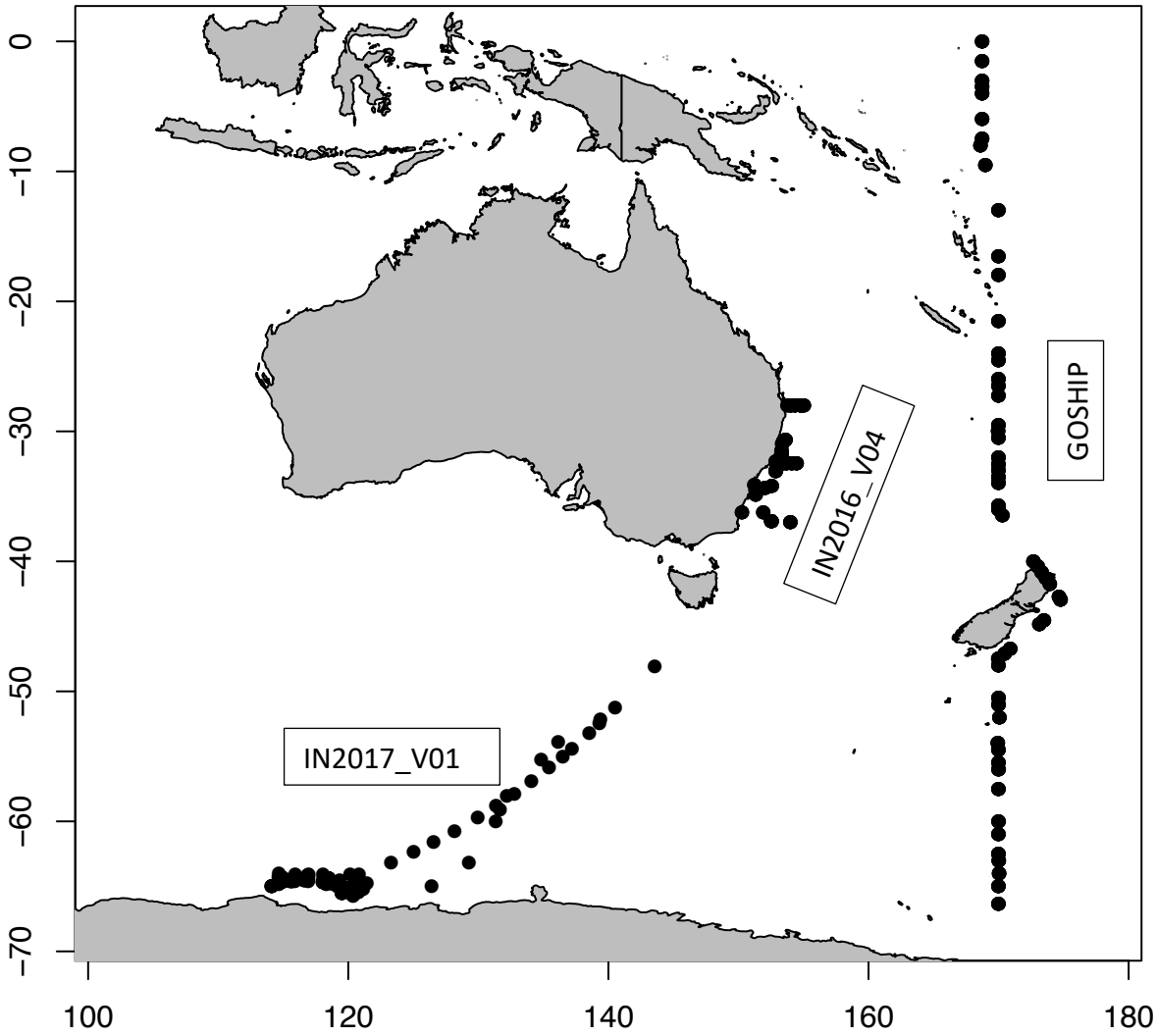



**Figure 5.4** *Map of the cytometry samples analysed. I collected and analysed samples from the IN2016_V04, IN2017_V01. Analysis of the GO-SHIP samples was done in collaboration with Michaela Larsson (UTS)*

## . 5.5 Conclusion

This dissertation presents the first baseline description of abundance, genetic diversity of microbial groups and infection rates (for the EAC) to define the forces structuring microbial community that inhabit the EAC and the Dalton polynya, two contrasting hotspots for climate change. These data demonstrate the importance of picocyanobacteria and viruses for the carbon dynamics in the EAC region. And how the diversity of primary producers drives distinct communities in the Dalton polynya. These results have implications for biomass turnover and community structure for these two regions that are undergoing substantial climate-related change, with finding relevant to both microbial ecologists and possibly policymakers. Finally, this dissertation provides the baseline to predict the future of production in these two regional ecosystems.

## 5.6 References

Breitbart, M. *et al.* (2018) 'Phage puppet masters of the marine microbial realm', *Nature Microbiology*, 3(7), pp. 754–766. doi: 10.1038/s41564-018-0166-y.

Brown, M. V. *et al.* (2018) 'Data Descriptor: Systematic, continental scale temporal monitoring of marine pelagic microbiota by the Australian Marine Microbial Biodiversity Initiative', *Scientific Data*, 5(April), pp. 1–10. doi: 10.1038/sdata.2018.130.

Gregory, A. C. *et al.* (2019) 'Marine DNA Viral Macro- and Microdiversity from Pole to Pole', *Cell*, pp. 1–15. doi: 10.1016/j.cell.2019.03.040.

Guidi, L. *et al.* (2016) 'Plankton networks driving carbon export in the oligotrophic ocean', *Nature*, 532(7600), pp. 465–470. doi: 10.1038/nature16942.

Hobday, A. J. *et al.* (2011) 'Defining dynamic pelagic habitats in oceanic waters off eastern Australia', *Deep-Sea Research Part II: Topical Studies in Oceanography*, 58(5), pp. 734–745. doi: 10.1016/j.dsr2.2010.10.006.

Hobday, A. J. and Hartmann, K. (2006) 'Near real-time spatial management based on habitat predictions for a longline bycatch species', *Fisheries Management and Ecology*, 13(6), pp. 365–380. doi: 10.1111/j.1365-2400.2006.00515.x.

IPCC (2014) *IPCC,2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)].*

Kopf, A. *et al.* (2015) 'The ocean sampling day consortium', *GigaScience*, 4(1), p. 27. doi: 10.1186/s13742-015-0066-5.

Luo, E. *et al.* (2017) ‘Bacteriophage Distributions and Temporal Variability in the Ocean’s Interior’, *mBio*. Edited by M. A. Moran, 8(6), pp. 1–13. doi: 10.1128/mBio.01903-17.

Raes, E. J. *et al.* (2018) ‘Oceanographic boundaries constrain microbial diversity gradients in the South Pacific Ocean’, *Proceedings of the National Academy of Sciences*, 115(35), p. 201719335. doi: 10.1073/pnas.1719335115.

Roux, S. *et al.* (2016) ‘Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses’, *Nature*, 537(7622), pp. 689–693. doi: 10.1038/nature19366.

Sieradzki, E. T. *et al.* (2019) ‘Dynamic marine viral infections and major contribution to photosynthetic processes shown by spatiotemporal picoplankton metatranscriptomes’, *Nature Communications*, 10(1). doi: 10.1038/s41467-019-09106-z.

Soliveres, S. *et al.* (2016) ‘Biodiversity at multiple trophic levels is needed for ecosystem multifunctionality’, *Nature*, 536, p. 456.

Sunagawa, S. *et al.* (2015) ‘Structure and function of the global ocean microbiome’, *Science*, 348(6237), p. 1261359. doi: 10.1126/science.1261359.

Suttle, C. A. (2007) ‘Marine viruses - Major players in the global ecosystem’, *Nature Reviews Microbiology*, 5(10), pp. 801–812. doi: 10.1038/nrmicro1750.

Williamson, S. J. *et al.* (2008) ‘The Sorcerer II Global Ocean Sampling Expedition : Metagenomic Characterization of Viruses within Aquatic Microbial Samples’, *PLoS ONE*, (1). doi: 10.1371/journal.pone.0001456.