Investigation of the biodiversity and biogeography of marine microbes in Australian tropical waters

Taotao Huang

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Department of Chemistry and Biomolecular Sciences

Macquarie University, NSW, Australia

November 2016

Table of Contents

List of figures

List of Tables

Acknowledgements

Firstly, I would like to thank my supervisor, Professor Ian Paulsen for all of his guidance and scientific advice during my PhD. Thank you for always welcoming discussions and providing solutions to my research problems. I am also very appreciative of him for supporting my conference travel to the Ocean Science Meeting in Spain, and helping me to successfully receive Postgraduate Research Funding. I feel very fortunate to be able to work with him.

I would like to thank Dr. Karl Hassan for his useful suggestions on writing my thesis and also allowing me to contribute to his recent publication in mBio.

I would also like to extend my heartiest gratitude to my colleagues and friends, Alaska, Amaranta, Brodie, Fallen, Hasinika, Sheemal, Silas, Qi, Liping and Varsha. They have provided me with constant support, suggestions and ideas. I am so glad to have had the opportunity to work with these talented and amazing future scientists.

I would like to especially thank Sheemal and Hasinika for their infinite support and help during my PhD journey. I will always cherish the wonderful moments we shared.

I am also thankful to have met members of the SynBio group - Heinrich, Tom, Hugh, Elizabeth and Xin. They all have constantly helped and encouraged me while writing my thesis.

I would like to thank Dr. Martin Ostrowski, Postdoctoral researchers and staff members of the Paulsen research group for helping me with my project and administrative work.

I wish to thank the CSC (Chinese Scholarship Council) scholarship and iMQRES (International Macquarie University Research Excellent Scholarship) for providing living allowance and tuition fee.

Finally, I would like to express my deepest gratitude to my family members, first and foremost, my parents, my brother and sisters for their unconditional financial and emotional support. They have always been my shelter and rock when confronted by any obstacles. Most importantly, I would like to thank my wife, Fan Yang, for her endless love, care and support during my PhD candidature.

Declaration

I declare the work presented in this thesis was conducted by me under the direct supervision of Professor Ian Paulsen. None of the work presented has been previously submitted for any other degree.

Taotao Huang

Publications

Hassan KA, Cain AK, **Huang TT**, Liu Q, Elbourne LDH, Boinett CJ, Brzoska AJ, Li L, Ostrowski M, Nhu NTK, Nhu TDH, Baker S, Parkhill J, Paulsen IT (2016). Fluorescence-based flow sorting in parallel with transposon insertion site sequencing identifies multidrug efflux systems in *Acinetobacter baumanniii*. **mBio 7(5):** e01200-16.

Publications in preparation

Huang TT, Ostrowski M, Mazard S, Kumar SS, Brown VM, Messer L, Seymour J, Paulsen IT. Insights into the diversity and biogeography of surface sea prokaryotes in Australian tropical waters. For *ISME Journal*

Huang TT, Ostrowski M, Hasinika K.A.H. Gamage, Brown VM, Messer L, Seymour J, Paulsen IT. Investigating the microbial eukaryotic communities in the surface waters of the Arafura Sea and the Coral Sea. *Submitted to Environmental Microbiology*

Huang TT, Ostrowski M, Brown VM, Messer L, Seymour J, Paulsen IT. Oceanographic structure drives the marine prokaryotes and eukaryotes diversity in the euphotic zone of Australian Tropical Waters. *For ISME Journal*

Conference posters

Huang TT, Ostrowski M, Mazard S, Paulsen IT. Metagenomics of marine microbial community in Northern Australian tropical waters. *Aquatic Sciences Meeting*, Granada, Spain, February 2015

Huang TT, Ostrowski M, Mazard S, Paulsen IT. Single-cell analysis of marine Photosynthetic picoeukaryotes. *3rd Single Cell Analysis Congress*, London, UK, November 2016

Huang TT, Ostrowski M, Mazard S, Paulsen IT. Spatial and temporal variation in prokaryotic community in Australian Tropical Ocean. *Ocean Sciences Meeting*, New Orleans, USA, February 2016

Contributions

Chapter 2:

Insights into the diversity and biogeography of surface sea prokaryotes in Australian tropical waters

This work was conceived by Paulsen, Huang and Ostrowski. The sample collection and DNA extraction was performed by Ostrowski, Brown and Seymour. All experimental work and data analysis were conducted by Huang. The manuscript was written by Huang with contributions from Ostrowski, Kumar, Brown, Seymour and Paulsen.

Chapter 3:

Investigating the microbial eukaryotic communities in the surface waters of the Arafura Sea and the Coral Sea

This work was conceived by Paulsen, Huang and Ostrowski. The sample collection and DNA extraction was performed by Ostrowski, Brown, Messer and Seymour. All experiment work including 18S rDNA library preparation was conducted by Huang. All statistical data analyses were performed by Huang. This work was written by Huang with contributions from Ostrowski, Gamage, Brown, Messer, Seymour and Paulsen.

Chapter 4:

Oceanographic structure drives the marine prokaryotes and eukaryotes diversity in the euphotic zone of Australian tropical waters

This work was conceived by Paulsen, Huang and Ostrowski. The sample collection and DNA extraction was performed by Ostrowski, Brown, Messer and Seymour. All experiment work including 16S and 18S rDNA library preparation was conducted by Huang. All statistical data analyses were performed by Huang. This work was written by Huang with contributions from Paulsen.

Abstract

Marine microbes (including bacteria, archaea, protists, fungi and viruses) play a fundamental role in natural systems. They are not only responsible for almost half of global primary production but also process about one-half of the global biogeochemical flux of biologically important elements, such as carbon, nitrogen, phosphorus, sulphur and iron. Although they are ubiquitous and diverse in the ocean, it is difficult to characterize them due to their small size. In recent years, the development of molecular analysis and high-throughput sequencing has unveiled a large amount of novel diverse assemblages of marine microbial communities and how they are distributed over space and time.

Our seawater samples were collected from Darwin to Cairns across the Arafura Sea, Torres Strait and Coral Sea during an oceanographic transect through this region in the austral spring of 2012. These basins are fully tropical and frequently hit by tropical cyclones. The Arafura Sea and Torres Strait have shallow seafloor depth, and have been identified as one of the most pristine marine environments, least impacted by human activities. This basin has been classified as part of the highly productive North Australian Large Marine Ecosystem. The Coral Sea is an open sea, which includes the world's largest coral reef system. The nutrient concentrations and primary production are typically very low in the euphotic zone of this basin. Although these Australian tropical basins have peculiar environmental variables and marine ecosystems, the biodiversity and biogeography of their marine microbial communities have not been investigated systematically using molecular techniques.

In this PhD thesis, we determined the abundance of autotrophic unicellular phytoplankton in the surface seawaters by flow cytometry using the fluorescence of natural photosynthetic pigments (such as chlorophyll and phycoerythrin). Generally, plastidic protist abundances were consistently lower than those of Cyanobacteria across all sampling sites. *Synechococcus* cell

abundances ranged from 1×10^4 to 3×10^5 cells mL^{-1} in the Arafura Sea/Torres Strait and from 1×10^3 to 6×10^4 cells mL^{-1} in the Coral Sea. *Prochlorococcus* cell numbers were much higher in the Coral Sea ($1\text{-}4.5\times10^5$ cells mL^{-1}) than *Synechococcus*.

Secondly, we reported the first investigation of marine microbial community composition at the surface of Northern Australian tropical waters using 16S and 18S rRNA gene amplicon sequencing. Throughout all of the samples, the dominant representative groups of prokaryotes included *Synechococcus*, *Prochlorococcus* and the SAR11 clade. The eukaryotic assemblages were found to be dominated by SAR supergroups (Stramenopiles, Alveolates and Rhizaria) and Archaeplastida across all sampling sites. Microbial abundances analysis showed distinct biogeographic patterns, for example, *Synechococcus* was the most abundant group in the Arafura Sea/Torres Strait but *Prochlorococcus* dominated in the Coral Sea; diatoms had a much higher abundance in the Arafura Sea/Torres Strait, whereas the Syndiniales and Cnidaria showed an opposite trend with higher abundances in the Coral Sea.

Thirdly, we surveyed the diversity of marine prokaryotic and protistan community within the photic zone of the Northern Australian tropics from depths between 25 and 150 metres. The dominant groups of marine microbial community in the euphotic zone of these sampling areas have similar patterns with their structure in the surface seawaters. However, their abundances showed different vertical distribution. Generally, the phototropic plankton dominated at the upper surface layer of sampling sites and their abundance slowly decreased with depth, but the relative abundance of heterotrophic, mixotrophic or parasitic marine microbes showed an opposite trend.

In addition, multivariate statistics suggested that the distribution patterns of microbial community structure in the surface waters was strongly driven by salinity, temperature and nutrient availability, with phosphate, nitrogen and silicate concentrations being particularly

important. Furthermore, besides the environmental drivers, the seafloor depth in the Arafura Sea/Torres Strait (shallow basin) and the sampling depth in the Coral Sea (deep basin) were also the important factors influencing marine microbial community composition in the euphotic layer of Australian tropical waters.

Overall, this study provides critical insights into patterns of local distribution of abundant microbial producers, which will provide information of great relevance to our understanding of what drives primary productivity in Australian oceans. Our research also addressed the factors that determine why different genetic groups are abundant in one location but not another. The findings help us to answer the fundamental question of how organisms adapt to a particular environmental niche. This knowledge will not only enhance our capacity to predict the resilience of ocean ecosystems and their response to climate-change, but also will help provide clarity on investment choices for a sustainable ecosystem/environment and increase healthy outcomes from activities involving human-ocean interactions such as recreation, food production, fisheries and tourism in Australian waters.

Abbreviations

ABC	ATP-binding cassette
ASTS	Arafura Sea/Torres Strait
BD	Becton Dickinson
CCA	Canonical Correspondence Analysis
Chl *a*	Chlorophyll *a*
cpeB	R-Phycoerythrin beta subunit
CS	Coral Sea
CTD	Conductivity, Temperature and Depth
DGGE	Denaturing Gradient Gel Electrophoresis
dmdA	Dimethylsulfoniopropionate demethylase
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide
DOC	Dissolved Organic Carbon
DOM	Dissolved Organic Matter
FISH	Fluorescence in Situ Hybridization
GBR	Great Barrier Reef
GOE	Great Oxidation Event
HL	High-Light
hzo	Hydrazine Oxidoreductase
ITS	Internal Transcribed Spacer
LDA	Linear Discriminant Analysis
LEfSe	Linear Discriminant Analysis Effect Size
LL	Low-Light
MALV	Marine Alveolate
MAST	Marine Stramenopiles

MDA	Multiple Displacement Amplification
mRNA	Messenger Ribonucleic Acid
ncRNA	None-coding ribonucleic acid
nifH	Nitrogen fixation H gene
NirK	Nitrite Reductase K gene
NMDS	Non-metric Multidimensional Scaling
ntcA	Nitrogen Regulator
OTUs	Operational Taxonomic Units
PCB	Phycocyanobilin
PEB	Phycoerythrobilin
petB	Cytochrome b6
PNG	Papua New Guinea
POM	Particular Organic Matter
PPEs	Phototrophic Picoeukaryotes
PR2	Protist Ribosomal Reference
psbA	Photosystem II protein D1
PSU	Practical Salinity Unit
PUB	Phycourobilin
qPCR	Quantitative Real Time Polymerase Chain Reaction
rbcL	Ribulose-bisphosphate Carboxylase Gene
rRNA	Ribosomal RNA
SAR	Stramenopiles, Alveolates, Rhizaria
SSU	Small Subunit
UCYN-A1	Unicellular Cyanobacteria Group A
Yfr	Cyanobacterial Functional RNA

Chapter 1:

Introduction

1.1 Evolution of photosynthesis

Life on Earth has been shaped and powered by photosynthesis. Photosynthesis is an ancient process on the Earth, which may well have been established at least 3.5 billion years ago (Blankenship 1992, Davis 2004) (Figure 1.1). It provides a constant flux of energy for life to persist and proliferate. Early photosynthetic systems, such as those from purple bacteria, green-sulphur bacteria, green non-sulphur bacteria and Heliobacteria (Xiong et al 2000), are thought to have been anoxygenic. This has been supported by a range of observations from the Archean (>2.5 billion years) sedimentary record. For example, the earliest evidence for phototrophy is found in the 3.416 billion years Buck Reef Chert from the Barberton Greenstone Belt, South Africa (Tice and Lowe 2004). Anoxygenic photosynthesis utilised various inorganic and organic compounds as electrons donors for redox cycling, such as Fe^{2+}, H_2, S^0, HS^-, $S_2O_3^{2-}$, NO^{2-}, AsO_3^{3-}, and various organic central metabolism intermediates (Fischer et al 2016), and even today many types anoxygenic photosynthetic organisms still use them as electron source.

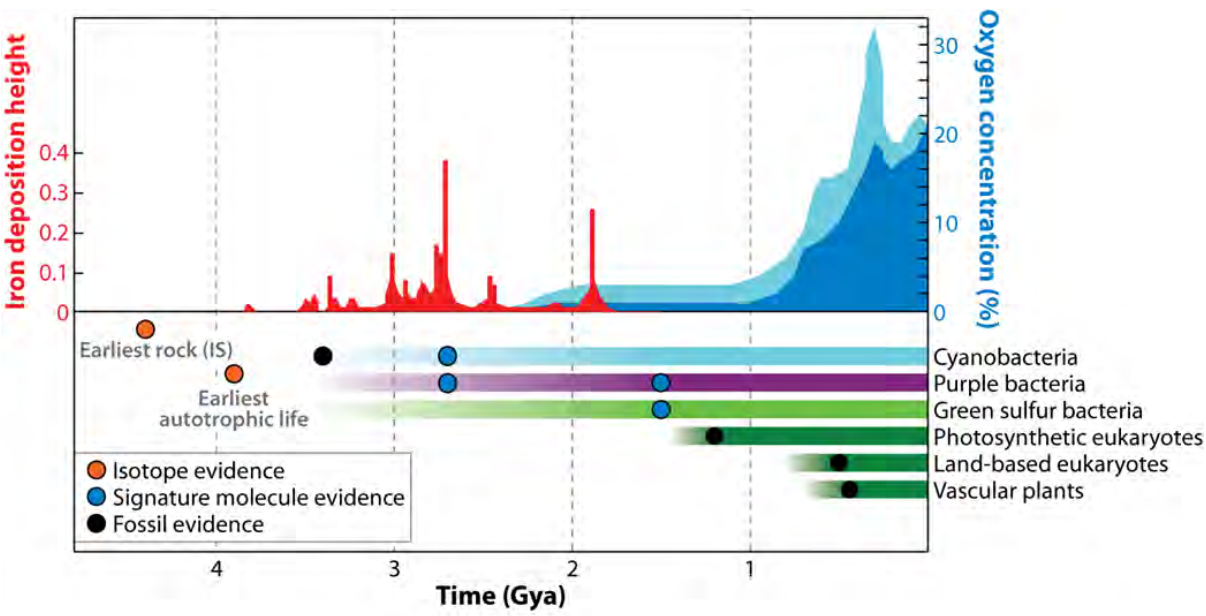

Figure 1.1 Evolution of life and photosynthesis in geological context, highlighting the emergence of groups of photosynthetic organisms. Minimum and maximum estimates for oxygen concentration are indicated by dark blue and light blue areas, respectively. This figure is reprinted from Hohmann-Marriott et al, Annu Rev Plant Biol, 2011.

Oxygenic photosynthetic apparatus likely evolved at around 2.5 billion years ago that allowed life to generate energy and reducing power directly from sunlight and water, freeing it from limited resources of geochemically derived reductants. This process generates O_2 as an end product of water oxidation and fix CO_2 into biomass: $nCO_2 + H_2O + \text{light} \rightarrow (CH_2O)_n + O_2$. Cyanobacteria are widely considered to be responsible for oxygenation of the earth's atmosphere. A permanent rise to high and sustained concentrations of O_2 in the atmosphere, called the "Great Oxidation Event" (GOE), occurred some time between 2.4 and 2.1 billion years ago, that is thought hundreds of millions of years after oxygenic photosynthesis developed. The reasons for this putative delay in accumulation of oxygen in the atmosphere may have be caused by low levels of oxygen generation at the beginning and the role of environmental compounds, such as ferrous iron, that reacted with O_2, thus providing an environmental O_2 sink (Nowicka and Kruk 2016) (Figure 1.1).

The rise in atmospheric O_2 dramatically changed the composition of the Earth's atmosphere and oceans and permanently altered all biogeochemical cycles, which also created a niche for the development of the aerobic respiration and novel biosynthetic pathways (Dismukes et al 2001). Furthermore, a protective ozone layer was established due to oxygen in the atmosphere interacting with ultraviolet solar radiation, which enabled the colonisation of terrestrial environments. It is believed that oxygenic photosynthesis evolved from bacterial anoxygenic photosynthesis. Heliobacteria were the closest relatives to those organisms undertaking oxygenic photosynthesis, supporting by the studies of photosynthetic genes (Xiong et al 2000). Highly efficient aerobic process led to the evolution of complex multicellular organisms (Grula 2005). The engulfment of a cyanobacterium by an unknown heterotrophic protist about 1.8 billion years ago results in the origin of photosynthetic plastids in eukaryotes (Tirichine and Bowler 2011).

1.2 Carbon fixation in the ocean

The marine carbon cycle plays a critical role in Earth's habitability for humans and another large fauna. It includes several processes (Figure 1.2), which are largely determined by various biological and environmental factors. The phototrophic microbes including cyanobacteria like *Prochlorococcus* and *Synechococcus* and a diverse assemblage of eukaryotic phytoplankton use sunlight to convert inorganic carbon (such as CO_2) into Dissolved Organic Matter (DOM). These phototrophic microbes capture nearly 7×10^{16} g carbon annually, corresponding to the conservation of 2.8×10^{18} kJ of energy (Berg 2011). The released dissolved organic matter and Particulate Organic Matter (POM) has several fates. Phytoplankton-derived carbon can either be transferred directly up the marine food web as bacteria succumb to predation by organisms at higher trophic levels (such as zooplankton) or it can enter the microbial loop, whereby diverse heterotrophic bacteria (including flavobacteria and roseobacters) rapidly absorb Dissolved Organic Carbon (DOC) from the water column and are subsequently consumed by bacterivorous protists, which are then consumed by larger zooplankton (Azam et al 1983). CO_2 is released back to the atmosphere via respiration during the catabolism of organic matter by heterotrophic bacteria.

The small grazers and heterotrophic bacteria also regenerate nutrients that are associated with phytoplankton organic matter, particularly nitrogen and phosphorus, which support up to 90% of total primary production (Azam et al 1983). In addition, up to 50% of fixed carbon is exuded by phytoplankton back into the surrounding seawater in the form of DOC (Vieira and Teixeira 1982). A significant fraction of transformed carbon resists further degradation and is transformed into recalcitrant DOC and exported from the surface of the oceans to deep ocean sediments via microbial carbon pump (Figure 1.2).

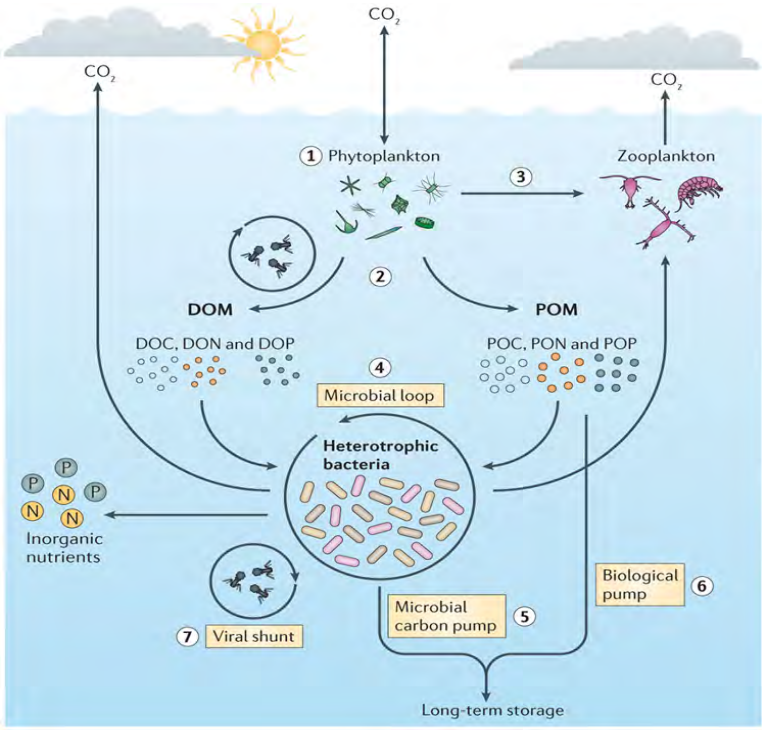

Figure 1.2 Bacterial transformation of phytoplankton-derived organic matter. Reprinted from Buchan et al, Nature Reviews Microbiology, 2014.

The POM sinks out of the photic zone via the biological pump (Figure 1.2), which leads to the sequestration of up to 300 million tons of carbon to the deep sea each year (Seymour 2014). The carbon becomes sequestered in the deep sea and can store there for thousands of years before it makes its way back into the atmosphere. The bottom of the ocean acting as a carbon sink plays a very significant role in regulating our planet's carbon cycle. Viral lysis of heterotrophic bacteria and phytoplankton also plays an important role in the release of both DOM and POM into the ocean, which redirects carbon and nutrients away from higher tropical levels and towards the microbial realm (Wilhelm and Suttle 1999).

1.3 Primary production in the marine system

Primary productivity is the process by which the inorganic form of carbon is converted by photosynthetic and chemosynthetic autotrophs to organic compounds. Marine primary production plays an essential role in food web dynamics, in global biogeochemical cycles and in marine fisheries (Chassot et al 2010, Passow and Carlson 2012). The unicellular

phytoplankton dominate primary production in the ocean, which are responsible for more than 45% of our planet's annual net primary production (Falkowski et al 2004, Field et al 1998). Phytoplankton serve as the base of marine food web, including two main types, cyanobacteria and eukaryotic algae.

Marine Cyanobacteria constitute about 10% of the total ocean marine picoplankton and 25% of ocean net primary productivity in the photic zone (Flombaum et al 2013). The cyanobacteria, *Prochlorococcus* and *Synechococcus* account for a substantial fraction of total marine primary production. *Prochlorococcus* is the most abundant photosynthetic organism on Earth. It has been found to comprise 9 % of gross primary production in the Eastern equatorial Pacific, 39% in the Western equatorial Pacific and up to 82 % in the subtropical North Pacific (Liu et al 1997). *Synechococcus* also contributes significantly to global marine primary production. It has been suggested to contribute up to 16.7% of net primary production in the ocean (Flombaum et al 2013). Other genera within the cyanobacteria such as *Trichodesmium*, *Nostoc* and *Richelia* play a critical role in fixing nitrogen from the atmosphere, thereby increasing the availability of organic nitrogen in the ocean (Leal et al 2009).

The eukaryotic algae are a diverse polyphyletic group, including both unicellular and multicellular organisms. Most multicellular primary producers, such as seagrasses and kelps, attach to substrates for growth. Thus, they are normally restricted to the shallow coastal waters with attachment sites and sufficient light for photosynthesis. Unicellular eukaryotic phytoplankton includes diatoms, photosynthetic dinoflagellates, green algae and prymnesiophytes. It is estimated that diatoms account for 40-45% of net oceanic productivity (Sarthou et al 2005). Dinoflagellates often dominate surface stratified waters. Some dinoflagellates can form harmful blooms or red tides, resulting in fish kills and human health problems (Magana et al 2003). In addition to phytoplankton, chemosynthesis and other primary

producers, such as mangroves, macroalgae, salt marshes and symbiotic algae also contribute to ocean primary production (<10%) (Duarte and Cebrian 1996).

1.4 Cyanobacteria

Cyanobacteria, also known as "blue-green algae", are monophyletic photosynthetic prokaryotes. These utilise sunlight as a source of energy using chlorophyll *a* and various accessory pigments. Although the exact timing of the existence of the first cyanobacteria-like microbes on Earth is still unclear, fossil evidence suggests cyanobacteria originated approximately 2.6-3.5 billion years ago (Garcia-Pichel 2009, Hedges et al 2001, Vincent 2009). Cyanobacteria are morphologically diverse, occurring in different forms including unicellular, filamentous, planktonic and colonial ones (Burja et al 2001). They can thrive in nearly all ecosystems, ranging from marine, freshwater, to terrestrial environments.

Cyanobacteria play an important role in the biogeochemical cycling of nutrients including carbon, nitrogen and phosphorus. They form the base of the food web, serve as the dominant primary producers in many extreme environments, and contribute to a significant proportion of the primary production of the oceans. Their global biomass has been estimated to be 10^{15} g (3×10^{14} g C) of wet biomass (Garcia-Pichel et al 2003), most of which is contributed by marine picocyanobacteria (*Prochlorococcus* and *Synechococcus*). Many cyanobacteria have the ability to fix atmospheric nitrogen in anaerobic conditions by means of the specialized cells called heterocysts. They comprise one of the largest global suppliers of fixed nitrogen in the environment, with the genus *Trichodesmium* alone responsible for 42% of global nitrogen fixation (Latysheva et al 2012). In addition, some species of cyanobacteria also form symbiotic associations with more complex biota and supply reduced nitrogen to the host (Vincent 2009).

Cyanobacteria are one of the most taxonomically challenging groups due to their long and arguably complex evolutionary history (Perkerson et al 2011). Based on the classification in

the second edition of the Bergey's Manual of Systematic Bacteriology, cyanobacteria are currently divided into five broadly recognized subsections (Table 1.1), I (=Chroococcales), II (=Pleurocapsales), III (=Oscillatoriales), IV (=Nostocales) and V (=Stigonematales) (Castenholz 2015). Generally, strains in subsections I and II are unicellular, however, cyanobacteria of subsection I divide by binary fission or budding, whilst subsection II divide by multiple fission. Subsection III strains are filamentous, non-heterocystous cyanobacteria reproducing through trichome breakage. Subsections IV and V are composed exclusively of heterocystous cyanobacteria, which are filamentous strains reproduced by hormogonia formation and have the ability to develop akinetes and heterocysts (Henson et al 2004).

Table 1.1 Diagnostic key to the subsections of the cyanobacteria, according to the *Bergey's Manual of Determinative Bacteriology*

Subsection (*traditional order*)	*Definition criteria*
Subsection I (Chroococcales)	Unicellular, nonfilamentous. Cells occurring singly or in aggregates. Cell division by binary fission in 1, 2, or 3 planes, symmetric or asymmetric, or by budding
Subsection II (Pleurocapsales)	Unicellular, nonfilamentous. Cells occurring singly or in aggregates. Reproduction by multiple fission without growth, yielding beaocytes (cells smaller than the parent cell), or by binary and multiple fission
Subsection III (Oscillatoriales)	Filamentous, binary fission in one plane, yielding unise-riate trichomes without true branching. No heterocysts or akinetes formed
Subsection IV (Nostocales)	Filamentous, division occurring only in one plane to yield uniseriate trichomes without true branching. Heterocysts formed when combined nitrogen is low
Subsection V (Stigonematales)	Filamentous, division occurring periodically or commonly in more than one plane, yielding multiseriate trichomes, truly branched trichomes, or both. Heterocysts formed when combined nitrogen is low

1.4.1 *Prochlorococcus*

Prochlorococcus was discovered in 1985 (Chisholm et al 1988). It is the smallest (the cell diameter is 0.5-0.7 µm) and the most abundant photosynthetic organism on Earth (Morel et al 1993). It is known for its high population density (normally exceed 10^5 cells mL^{-1}) and its wide horizontal and vertical habitat range in the ocean (Campbell et al 1997, DuRand et al 2001, Partensky et al 1999). *Prochlorococcus* has the smallest and highly streamlined genome of any

free-living photosynthetic prokaryote, some species have genomes as small as 1.65 Mb, with only ~1,700 genes (Rocap et al 2003). It possesses a unique photosynthetic pigment complement including divinyl derivatives of chlorophyll *a* (Chl *a*) and Chl *b* in its main antenna complexes (Goericke and Repeta 1992), which causes a slight red shift in its absorption spectra. This unusual pigmentation makes it possible to determine that *Prochlorococcus* comprises nearly 50% of the total chlorophyll in the surface oceans (Partensky et al 1999, Partensky and Garczarek 2010).

The annual mean global abundances of *Prochlorococcus* are estimated of $2.9 \pm 0.1 \times 10^{27}$ cells. *Prochlorococcus* thrives throughout the 40°S to 40°N latitudinal band of oceans (Figure 1.3A) (Flombaum et al 2013). It dominates in the warmer oligotrophic waters, especially the Indian and Western Pacific Ocean subtropical gyres (maximum = 2.5×10^5 and 2.1×10^5 cell mL^{-1}). It is thought to be absent at temperatures below 15 °C (Johnson et al 2006) and outcompeted by other phytoplankton in high-nutrient waters (Jiao et al 2005). *Prochlorococcus* cell concentrations are often at high density at the surface down to depths of 200 m, and much lower in coastal than in offshore areas (Bouman et al 2006, Garczarek et al 2007).

There are two distinct major lineages of *Prochlorococcus*, the High-Light (HL) adapted and Low-Light (LL) adapted ecotypes, which is supported by the phylogenetic studies of the 16S rRNA gene and other marker genes, such as the Internal Transcribed Spacer (ITS) region between 16S and 23S rRNA (Jameson et al 2010, Johnson et al 2006, Martiny et al 2009, Moore and Chisholm 1999, Partensky et al 1999, Rocap et al 2002). This ability enables *Prochlorococcus* occupy the entire euphotic zone and makes their taxonomic diversity uniform on a horizontal scale but differs vertically in the water column. HL-adapted cells are orders of magnitude more abundant in the upper surface layer but are outnumbered by LL-adapted cells at the base of the euphotic zone (Johnson et al 2006, West and Scanlan 1999). HL-adapted stains are monophyletic and further subdivided into at least six clades (HLI-HLVI) (Moore et

al 1998), whereas the LL-adapted stains are polyphyletic and partition into at least six clades (LLI-LLVII) (Malmstrom et al 2013). These clades have different physiological and ecological distinctions, for example, HLI and HLII are distinguished by their temperature optima (Johnson et al 2006), however, HLIII, HLIV and HLV clades flourish in waters that are rich in high nitrogen and phosphorus, but have low iron availability (West et al 2011); LLI are more abundant closer to the surface and during deep mixing events whereas LLII/III and LLIV are more restricted to the lower euphotic zone and have decreased abundance during mixing events (Malmstrom et al 2010, Zinser et al 2007).

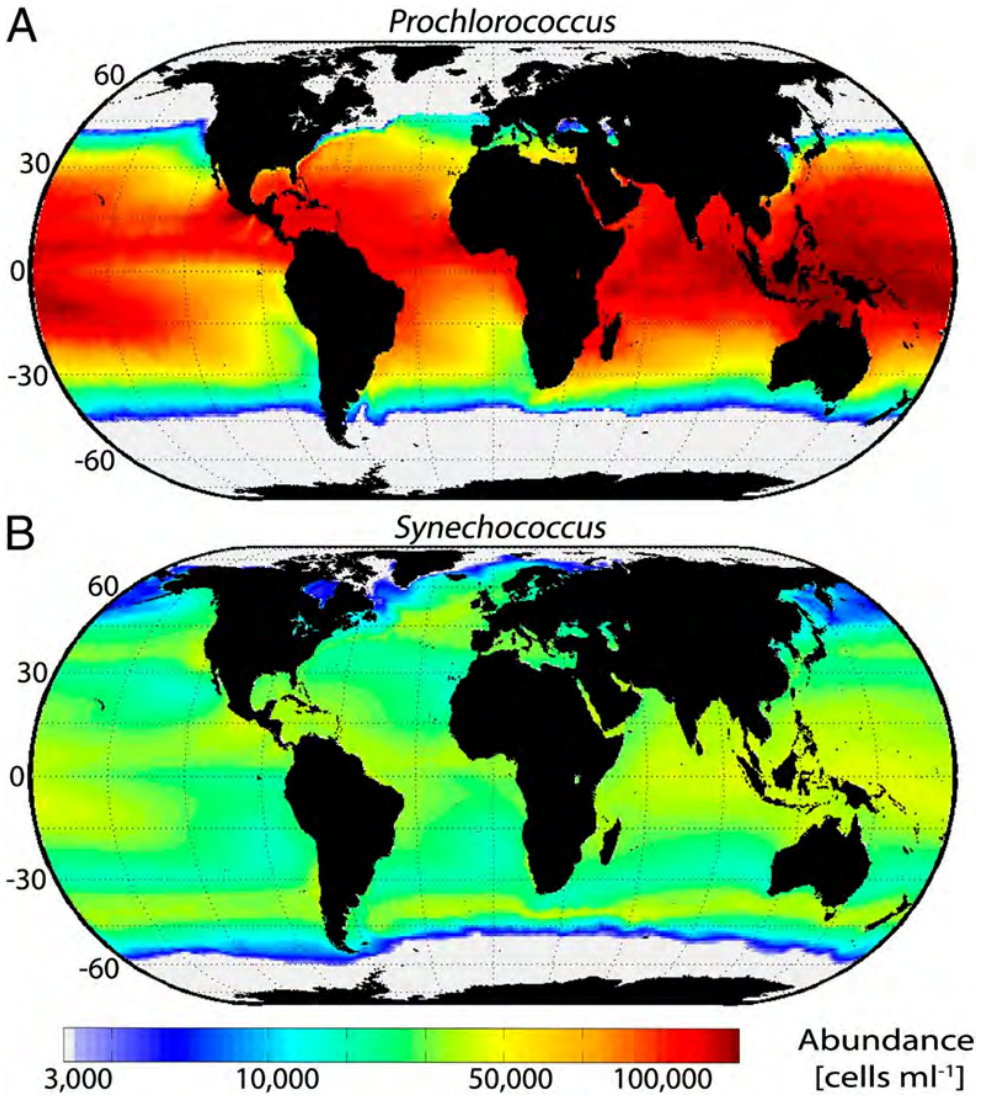

Figure 1.3 Present estimated global distribution of *Prochlorococcus* and *Synechococcus* abundance. Reprinted from Flombaum et al. PNAS 2013.

1.4.2 *Synechococcus*

Synechococcus is closely related to *Prochlorococcus* but is more ancient and genetically diverse, and was first described in 1979 (Waterbury et al 1979). Its cell sizes are normally between 0.6 µm and 1.6 µm. Their cells abundance range from 10^3 cells mL^{-1} in oligotrophic waters and up to 10^5-10^6 cells mL^{-1} in nutrient rich waters (Zwirglmaier et al 2008). The genome of *Synechococcus* consists of a single chromosome with size ranging from 2.2 to 2.86 Mb, and the number of genes ranges from 1,716 to 3,022 (Scanlan et al 2009). The major photosynthetic pigment in *Synechococcus* is chlorophyll *a*, but its main accessory pigments are phycobilliproteins, which is attached to the surface of the photosynthetic membranes (Grossman et al 1993). There are several well-recognized phycobilins: phycocyanobilin (PCB), phycoerythrobilin (PEB), and phycourobilin (PUB) (Ong and Glazer 1991).

The annual global mean abundances of *Synechococcus* are estimated at $7.0 \pm 0.3 \times 10^{26}$ cells (Flombaum et al 2013). *Synechococcus* is much more abundant in nutrient rich environments and the surface of euphotic zone, with very low numbers found down to depths of 200 m. It has a ubiquitous oceanic distribution that includes both polar regions (Figure 1.3B) (Flombaum et al 2013). It is very abundant in the Indian and Western Pacific Oceans. Its abundance peaks at mid-latitudes, and declines in cold current. There is a strong seasonal variation in the distribution and abundance of *Synechococcus*. Its highest/lowest global monthly mean abundance were in March/July.

The marine *Synechococcus* lineage was initially divided into three sub-clusters (A, B, C) based on the specific range of genomic G+C content (Fuller et al 2003). Under the current classification, there are five clusters (1-5) based on morphology, physiology and genetic traits (Castenholz 2015). Cluster 5 strains are truly marine *Synechococcus* comprising three sub-clusters 5.1, 5.2 and 5.3. Sub-cluster 5.1 is the largest and most diverse lineages. These are distinguished with sub-cluster 5.2 by containing phycoerythrin. Sub-cluster 5.2 possesses

phycocyanin as a major pigment in phycobilisomes. Sub-cluster 5.3 was previously within sub-cluster 5.1, however, phylogenetic analysis demonstrates their divergence prior to the differentiation of other *Synechococcus* sub-clusters. Most marine *Synechococcus* strains belong to marine cluster 5.1, including the four most abundant lineages, clades I to IV. Clades I and IV are generally limited to high latitudes (above 30°N and below 30°S) and are dominant in the coastal boundary zone which experience fluxes in nitrate and phosphate concentrations (Zwirglmaier et al 2008). Clade II are abundant in subtropical/tropical latitudes between 30°S and 30°N. Clade III appears to be confined to a small range of nitrate and phosphate concentrations suggesting members of this clade are oligotrophs (Zwirglmaier et al 2008).

1.5 Protist phylogeny

Protists are unicellular eukaryotes that are not categorized as animals, plants or fungi. They cover a wide spectrum of cell sizes and shapes ranging from a bacterial-size for the smallest known species of chlorophytes to meters in length for the largest colonies of radiolaria. They display complex interactions with other protists, metazoan, bacteria, archaea and viruses. Protists are abundant and ubiquitous members of nearly all known ecosystems, and together they represent the majority of eukaryotic biodiversity (Pawlowski 2013, Worden et al 2015). Moreover, protists play a pivotal ecological role in marine ecosystems as primary producers, consumers, decomposers and parasites (Sherr et al 2007). They are particularity important primary producers in coastal waters and highly productive ecosystems that support the world's most important fisheries. Protists are also recognized as important contributors to biogeochemical cycles in the ocean (Sherr et al 2007).

Molecular approaches provide unique opportunities for exploring the phylogenetic and functional diversity of marine protists. This has revealed a wealth of information on marine microbes that have not been cultivable in laboratory settings. In particular, recent efforts on a global scale by Tara Oceans and Malaspina expedition provides new insights into marine protist

diversity and distribution (de Vargas et al 2015, Pernice et al 2016). There is a recent phylogenetic scheme proposed for the major eukaryotic lineages with the presence of well-known, free-living, marine taxa in those supergroups (Caron et al 2012) (Figure 1.4). Although the configuration of supergroups varies, the general consensus includes unikonts (opisthoknots + amoebozoans), archaeplastida, SAR (stramenopiles + alveolates + rhizaria) and excavates.

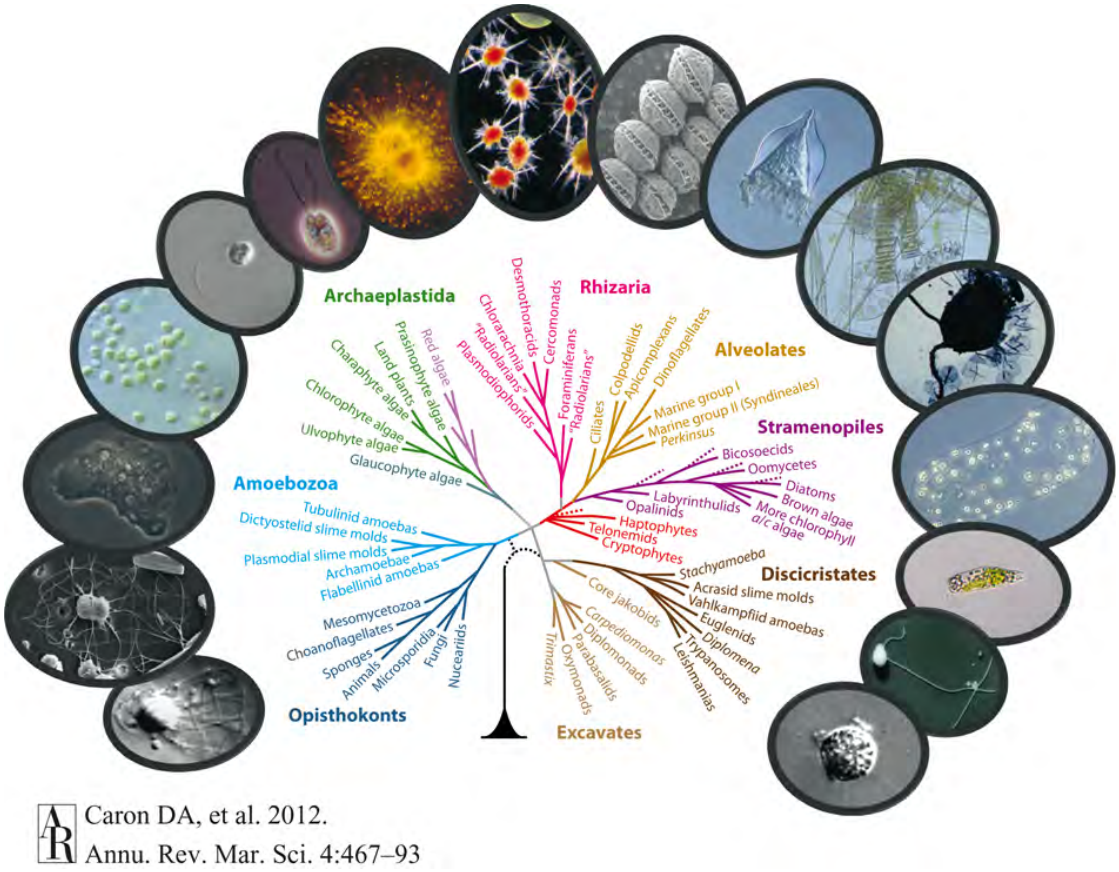

Figure 1.4 Phylogenetic breadths among protists. Reprinted from Caron et al, Annu. Rev. Mar. Sci. 2012

1.5.1 Opisthokonta

Opisthokonts are a super-group of eukaryotes including two main and diversified multicellular lineages: the Holomycota or Nucletmycea (fungi and their unicellular relatives, such as the Nucleariids and *Fonticula alba*) and Holozoa (Metazoa and their unicellular relatives, such as Choanoflagellata and Ichthyosporea) (Brown et al 2009, Caron et al 2012, Mendoza et al 2002,

Torruella et al 2012). They are exclusively heterotrophic and none of the taxa have chloroplasts. but some species harbor symbiotic or kleptoplastidic algae (e.g., lichens, reef-building corals, sacoglossan sea slugs). Molecular phylogenies sometimes suggest a close relationship between Opisthokonta and Amoebozoa, collectively named unikonts (Richards and Cavalier-Smith 2005). However, this relationship is controversial because the root of eukaryotes tree is not resolved nowadays (Roger and Simpson 2009).

Choanoflagellates are a group of small free-living and single-celled flagellates, found in both marine and fresh waters. They are common in marine plankton, where they can be important bacterivorous species, especially in polar ecosystems (Marchant 1985). Choanoflagellates have a close relationship with Metazoa because their cell architecture is very similar to the choanocytes (collar cells) of the most primitive metazoans, sponges. This relationship has been supported by multigene phylogenetic analyses (Torruella et al 2012). Each choanoflagellate has a single flagellum, surrounded by a collar composed of many fine “tentacles” (microvilli) supported by actin microfilaments. The flagellum generates a current draws water through the collar microvilli and ingest particles that adhere to the outer surface of the collar (Leadbeater 2008). The principle particles they prey are bacteria and a range of picoplanktonic cells, including single-celled cyanobacteria, prochlorophytes, pico-eukaryotes and detritus. Members of Acanthoecid choanoflagellates are usually restricted to marine environments habitats and considered as important components in the microbial food web (Leakey et al 2002).

Fungi represent a significant part of the microbial diversity on Earth and perform important functions as decomposers, driving nutrient cycles in detritus environments, and as parasites and symbionts (Hibbett et al 2007, Richards et al 2012). Fungi are globally distributed and grow in both terrestrial environments and aquatic habitats, including extreme environments, such as hypersaline or ionizing radiation areas and deep-sea sediments (Dadachova et al 2007, Raghukumar and Raghukumar 1998). However, fungi are both non-diverse and low abundant

in the marine environment (Burgaud et al 2009), suggesting that the majority of evolutionary diversification of fungi occurred on the land not in the sea. Because of their chitin-rich cell walls and osmotrophic feeding strategies, they flourish in nutritionally rich environments where they can attach to substrates, secrete enzymes, break down complex biological polymers, and take up nutrients. These ecological characteristics also partly explain why marine fungi appear to be rare in many upper and surface marine water column samples (Massana and Pedros-Alio 2008, Richards and Bass 2005) as many pelagic and surface water environments often have low nutrient concentrations and contain a large amount of free-floating or swimming single-celled organisms performing primary production and/or phagotrophic grazing that cannot provide larger physical substrates for attachment and osmotrophy. Marine fungi are considered to be a key contributor to the decomposition of woody and herbaceous substrates and dead animals in coastal and surface marine environments (Mann 1988, Newell 1996). Some marine fungi cause diseases of marine animals and plants, but others form mutualistic symbiotic relationships with other organisms (Hyde et al 1998).

1.5.2 Amoebozoa

The Amoebozoan belong to one of the three evolutionarily distinct major groups of amoeboid eukaryotes, the other two are Rhizaria in corticates and Percolozoa in Eozoa (Cavalier-Smith et al 2015). Molecular genetic analysis supports Amoebozoa as a monoplyletic clade and the sister group to Opisthokonta. They are exclusively heterotrophic and extremely common in marine benthic community but are sometimes found in plankton samples (Moran et al 2007). They have highly flexible cells and use pseudopodia for locomotion and feeding via phagocytosis. The pseudopodia are characteristically exhibited include extensions, which can be tube-like or flat lobes, rather than the hair-like pseudopodia of rhizarian amoeba. Although the majority of amoebozoan species are unicellular, the group also includes the well-studied slime molds, which have a macroscopic, multicellular stage of life. The slime molds were once thought to be fungi because they possess hyphae and their fruiting bodies. They feed on

bacteria, yeasts and fungi, and comprise two groups distinguished by their unique life cycles: Plasmodial and Cellular slime molds. An increasing number of studies have indicated that many diverse amoebae play an important ecological role and act as opportunistic pathogens of animals in aquatic environments. For instance, they may be primary or secondary invading pathogens of crustacea or vertebrates, including fish (Dykova and Lom 2004).

1.5.3 Archaeplastida

This supergroup of obligate phototrophs is a major group of eukaryotes and includes red algae, glaucophytes, green algae and land plants. There is a general consensus that their plastids are of prokaryotic origin, the result of endosymbiosis between a heterotrophic eukaryotic host and a photosynthetic cyanobacterium (Gould et al 2008, Keeling 2010, Reyes-Prieto et al 2007). The red algae includes about 5,000-6,000 species, with relatively few single-celled taxa and a large assemblage of multicellular algae, mainly marine inhabitants (Thomas 2002). Glaucophytes comprise a small group of freshwater unicellular algae comprising only 13 species. The green clade constitutes both green algae and land plants, with about 350,000 species (Mackiewicz and Gagat 2014). Recent microscopical and molecular studies have re-highlighted the diversity and importance of minute chlorophytes in marine ecosystems, particularly among the prasinophytes (Worden 2006). The Prasinophytes are a class of unicellular green algae, mainly include marine picoplanktonic genera, such as *Micromonas* and *Ostreococcus*.

Micromonas is a genus of green algae, with a characteristic swimming behaviour, and a single chloroplast and mitochondrion. It thrives in ecosystems ranging from tropics to poles and could serve as sentinel organisms for biogeochemical fluxes of modern oceans during climate change (Worden et al 2009). It contains the only described species, *Micromonas pusilla*, which is easily cultivable and comprises a large portion of photosynthetic picoeukaryote in several marine ecosystems, such as the Western English Channel (Not et al 2004) and central California waters

(Thomsen and Buck 1998). Compared with the genome size of *Ostreococcus*, *Micromonas* strains have a larger genome (20 Mb and 10,000 genes), which provides a higher ecological flexibility with more genes for nutrient transport or chemical protection (Massana 2011). *Ostreococcus* are known as the smallest free-living eukaryotic species to date, with an average size of 0.8 µm. It has a cosmopolitan distribution, having been found from coastal to oligotrophic waters, and seems more abundant at the deep chlorophyll maximum, with higher nutrient concentrations (Countway and Caron 2006). *Ostreococcus* has been suggested as an ideal model organism for research on eukaryotic genome evolution because of its remarkable simplicity (a naked, nonflagellated cell possessing a single mitochondrion and chloroplast), small size and ease in culturing. As its small cellular and genome sizes (13 Mb and 8,000 genes), it may reveal the "bare limits" of life as a free-living photosynthetic eukaryote, having a high gene density and intergenic reductions (Derelle et al 2006, Massana 2011). It has rapid growth rates (Fouilland et al 2004), and has caused dramatic blooms off the coasts of Long Island and California (Countway and Caron 2006).

1.5.4 Alveolata

The Alveolata (meaning "with cavities") are a monophyletic supergroup of primarily single-celled eukaryotes that have adopted extremely diverse modes of nutrition, such as predation, photoautotrophy and intracellular parasitism. There are three major groups within the alveolates: ciliates, apicomplexans, and dinoflagellates. Despite the large morphological differences among these groups, alveolates share several morphological features including alveoli (a system of abutting membranous positioned beneath the plasma membrane), extrusive organelles (e.g. trichocysts), closed mitosis, tubular mitochondrial cristae and distinct micro-pores through the cell surface that function in pinocytosis (Leander and Keeling 2004, Patterson 1999). The alveoli are not a part of other endomembrane systems, and can be empty (e.g. colpodellids and apicomplexans) or filled with cellulosic material (e.g. thecate dinoflagellates and some ciliates).

The ciliates are one of the most homogeneous of protozoan groups, containing over 8,000 morphological species with about two-thirds of these being free-living and the remainder symbiotic. Symbiotic ciliates can be commensals, mutualists or parasites. They are common almost everywhere with water such as accumulate-lakes, ponds, seawaters, freshwaters and soil (Finlay et al 1998). Ciliate species range in size from as small as 10 µm (e.g. *Strombidium* and *Strobilidium*) to as much as 4,500 µm (some benthic karyorelicteans). This typically large size suggests that the ciliates are commonly on 'top' of the microbial food web. They feed on bacteria, flagellates and phytoplankton, and are now themselves a prey to metazoans (animals), jellyfish and small fish (Lynn 2001).

The ciliates are characterized by three major features (Katz 2001). First, they exhibit nuclear dimorphism, typically with two types of nuclei in their cytoplasm. The macronucleus contains many copies of genome and divided by amitosis. It is the physiologically active nucleus, synthesizing messenger ribonucleic acid (mRNA) that controls the functions of the cell. The micronucleus is a typical diploid protistan nucleus and divides by an endomitosis. It is probably very rarely involved in transcription. Secondly, their sexual process is known as conjugation. This is a temporary fusion of two ciliates during which the partners exchange gametic nuclei. The gametic nuclei are derived by meiosis from the micronucleus and are thus haploid. Thirdly, ciliates typically have a distinctive cytoskeleton comprised of numerous short flagella (cilia) and associated root systems.

Apicomplexans are intracellular, largely nonflagellated obligate parasites of animals. Most of them possess a novel cell invasion apparatus called the ''apical complex'' (Sibley 2010). This "apical complex" consists of a tubulin-based (closed) "conoid" that serves as a spring-like scaffolding for extrusive organelles called "rhoptries", and around which there are two rings, one circling the distal end and the other circling the proximal end of the conoid. The apical

complex is used to attach and penetrate the host cell, and the rhoptries are applied to release products that stimulate the host cell to invaginate and draw in the parasite. They infect animal cells, ranging from epithelial cells and marine invertebrates to the humans (e.g. the infamous *Plasmodium* causes malaria). Molecular phylogenies strongly indicate that dinoflagellates and apicomplexans are more closely related to each other than to ciliates, forming the Myzozoa (Fast et al 2002, Leander and Keeling 2004).

The dinoflagellates are a largely planktonic division of motile unicellular microalgae that have a distinctive flagellar apparatus consisting of a coiled transverse flagellum within a cingular groove or girdle and a posterior flagellum within a sulcal groove (Fensome et al 1999). They form one of the largest groups of marine eukaryotes, containing about 6,000 described species. Most of them are characterized by a dinokaryon. Although classified as eukaryotes, they are the only eukaryotes that have permanently condensed chromosomes thought to lack typical eukaryotic histones and nucleosomes but contain specific DNA-binding basic proteins. The plasmodial genus *Syndinium* is the only known exception in this group, which has base histone-like nuclear proteins. Their populations are distributed depending on temperature, salinity, or depth. Generally, dinoflagellates are a more significant component of the phytoplankton in warmer waters. They often dominate surface stratified waters; and in temperate zones there may be a succession from diatoms to dinoflagellates as the relatively nutrient rich, well mixed water column of spring stabilizes to form a stratified water column with relatively warm, nutrient poor surface waters (Taylor et al 2008).

Dinoflagellates can be found in both freshwater and marine environments, where they present a great diversity including autotrophic, heterotrophic, mixotrophic, parasitic, and symbiotic species. About half of all dinoflagellates are photosynthetic, they typically possess a chloroplast surrounded by three membranes and pigmented with a highly modified carotenoid, peridinin (Hofmann et al 1996). This type of chloroplast is considered to be derived from secondary

endosymbiosis with a red alga. Many photosynthetic dinoflagellates are also capable of heterotrophic feeding (mixotrophic) and can rapidly shift lifestyles from a primary producer to that of a predator (Li et al 1996). Many of them are obligate heterotrophic and act as important eukaryovores in marine ecosystems (Sherr and Sherr 2007). Many extant dinoflagellates are endoparasites of various hosts include other protists (including dinoflagellates), appendicularias, annelids, siphonophores, radiolarians, marine crustaceans and their eggs, and fish and their eggs. The studies on environmental DNA clone analyses revealed that two novel lineages of marine alveolates (MALV) group I (renamed as Syndiniales Groups I) and MALV group II (renamed as Syndiniales Groups II) within the order Syndiniales are dominant components of these parasitic dinoflagellates (Guillou et al 2008). One widely known mutualistic association formed by dinoflagellates (called zooxanthellae, most of them are members within the genus *Symbiodinium*) is their symbiotic relationship with reef-building cnidarians. Corals provide dinoflagellates with a relatively safe refuge from predators and fluctuating environmental conditions. In return, photosynthetic dinoflagellates provide the chief source of food (photosynthate or fixed carbon) for coral-building cnidarians (Davy et al 2012).

Dinoflagellates have a patchy distribution and may bloom to form red tides, or “harmful algal blooms”, when they reproduce rapidly and copiously due to the abundant nutrient concentrations in the water. Depending on the pigments present in these dinoflagellates, these tides actually appear brown, red, orange, or yellow. During major algal blooms, they release toxins into the water, killing many aquatic animals and accumulating in filter feeders such as shellfish, and may poison human upon consumption, with neurotoxins that attack the human nervous system. One red tide organism, *Pfiesteria*, is unusual in that it uses toxins specifically to kill fish, stunning them and then feeding on their tissues (Rublee et al 2005). More than 18 dinoflagellate genera have the ability to bioluminesce and most of them emit a blue-green light (Roenneberg and Taylor 2000). They contain the compound luciferin, which is the same chemical that makes fireflies glow.

1.5.5 Stramenopiles

Stamenopiles (Latin, *stramen* − straw + *pila* − hairs) constitute one of the major and highly diverse monophyletic group of eukaryotes, currently including at least 21 major classes with five nonphotosynthetic groups (Yoon et al 2009), branching with the Alveolata and Rhizaria within the SAR "supergroup" (Burki 2014). Many stamenopiles are unicellular flagellates, and most others produce flagellate cells at some point in their life cycle. All of those flagellate cells typically possess two different types of flagella (Figure 1.5). A long anteriorly directed immature flagellum is covered with one or two rows of characteristic tripartite (base, shaft and terminal hairs) tubular hairs, while the posteriorly directed mature flagellum is usually whip-like, shorter and smooth. Functionally, anterior flagellum propulsive force through waveform bending to power the cell forward swimming motility and the posterior flagellum presents rapid lateral beating to steer the swimming direction (Geller and Muller 1981, Matsunaga et al 2010). Because the two flagella act differently, the term heterokont was used as an alternative name for the stramenopiles. Many flagellar proteins have been identified which were associated with cell motility, signal transduction and various metabolic activities (Fu et al 2014). The flagellum axoneme typically has the "9+2" microtubular arrangement comprising nine outer doublet microtubules and central pair of microtubules. Stramenopiles typically possess four microtubule roots (R1-R4) in a distinctive pattern. In phagotrophic species, preys are usually drawn to the cytostome formed between the splitting R2 by the tubular flagellar hairs reverse the water current produced by normal flagellar beats.

The Stramenopiles contain more than 100,000 species including very diverse life forms from single cells to large plasmodia to complex multicellular thalli. They comprise a photosynthetic group (Ochrophyta, Heterokontophyta or stramenochromes), which are the predominate eukaryotes in most aquatic environments, and many other notable heterotrophic lineages such as fungi-like organisms (Oomycetes, Hyphochytridiomycetes, Labyrinthulea), heliozoan-like protists (actinophryds), intestine parasites (Opalozoa) and free-living flagellates (Romain

Derelle 2016). Currently, there are sixteen photosynthetic classes of stramenopiles that have been described. The phylogenetic relationships among them are still largely unresolved. The phylum ochrophyta is a group of mostly photosynthetic heterokonts with chloroplasts originated by the secondary endosymbiosis with a red alga. It includes Bacillariophyceae (diatoms), Phaeophyceae (brown algae), Chrysophyceae (golden algae) and Xanthophyceae (yellow-green algae), Dictyochophyceae, Pelagophyceae and Pinguiophyceae. Some members of these classes are heterotrophic, especially amongst the Dictyochophyceae and Chrysophyceae.

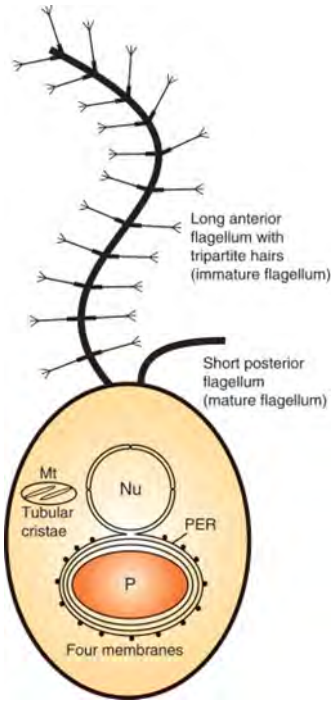

Figure 1.5 A simplified Stramenopile cell. Reprinted from Yoon et al, Encyclopedia of Microbiology (Third Edition), 2009

The bacillariophyceae (diatoms) are the most species-rich class of stramenopiles, comprising more than 200 genera. Diatoms are a photosynthetic group with worldwide distribution and thrive in almost every aquatic environment including fresh and marine waters, soils, in fact almost anywhere moist. They are unicellular and normally between 2-200 µm in diameter or length, although sometimes they can be up to 2 mm long. A unique feature of diatom cells is that they are enclosed within a silica cell wall known as a frustule. This consists of two halves

of identical structure called thecae, although one half (epitheca) is slightly bigger and overlaps the other half (hypotheca). Based on the shape and symmetry of the frustule, they are classified into radial centrics, polar centrics and pennate diatoms (Kroger and Poulsen 2008). This classification is now challenged by recent molecular phylogenetic studies. Their evolutionary success is largely due to their silica cell walls, because biosynthesis of a cell wall made of silica rather than organic material is likely associated with lower energy expenditure of the cells (Raven 1983), and the silica cell wall may serve as a protective armor against phytoplankton predators (Smetacek 1999). They utilize silicic acid to construct their cell walls and are controlled by its availability and distribution.

Diatoms possess a complete urea cycle like animals, inherited from the heterotrophic host of the secondary endosymbiosis, which links them evolutionarily to metazoans (Gross 2012). Planktonic diatoms live a “bloom and bust” life cycle in freshwater and marine environments. This life cycle is dependent on the availability of nutrients (such as, nitrate, phosphate and silicate) required for growth. They grow quickly and become dominant phytoplankton communities in the euphotic layer (“bloom”) when conditions are favorable. However, when nutrients become depleted in the water column, they increase in sinking rate and exit the upper mixed layer (“bust”) and re-enter this zone when nutrients are replenished. This lifestyle makes diatoms play an important role in the export of carbon from oceanic surface waters and the regulation of the biogeochemical cycles of silicon (Dugdale and Wilkerson 1998, Smetacek 1985). They are one of the major contributors to global carbon fixation, and account for 45% of total eukaryotic marine primary production (Yool and Tyrrell 2003). Diatoms carry out one-fifth of the photosynthesis on Earth and generate about as much organic carbon as all the terrestrial rainforests combined (Field et al 1998, Nelson et al 1995). The organic carbon is consumed quickly and serves as a base for marine food webs. Diatoms not only support our most productive fisheries in coastal waters but also deliver a large proportion of organic matter from the surface to the deep water in open ocean, which sink in the form of large and quickly

setting flocs of marine snow amorphous aggregates 0.5 mm or greater in diameter (Alldredge and Silver 1988, Smetacek 1985). Diatoms are also a popular tool for monitoring past and present environmental conditions, and are also commonly used in water quality assessment, as different species are characteristic of different trophic conditions.

The golden algae (Chrysophyceae) are characterized by the presence of lateral filaments on the flagellar hairs. Most photosynthetic members are freshwater flagellates, but heterotrophic members are ubiquitous in both marine and freshwater (Medlin et al 1997). The brown algae (Phaeophyceae) are a major group within the stramenopiles that are mostly found in marine habitats, and distributed worldwide. They include 16 orders with about 1,500 – 2,000 species (Fogg 1996). Morphologies of brown algae range from simple microscopic filaments to giant kelps that may reach 60 m in length. They also play an important role in marine environments, for example, giant kelps produce a large biomass with high growth rates and form prominent underwater forests in coastal marine waters. Many of them are also important as human food sources, and aquaculture of kelp species is very popular in Asia. There are two visible features that distinguish them from all the other algae, one is they are multicellular, and they possess a characteristic color that ranges from an olive green to various shades of brown. They consist of a variety of differentiated multicellular structures including a holdfast, stipe, and blades.

The Dictyochophyceae are mostly unicellular flagellate and some amoeboid algae (members of the Rhizochromulinales). They usually inhabit marine settings, but some freshwater species have also been reported. The dictyochophyceans are typically photoautotrophic but there are many mixotrophic (e.g. Pedinella) and colorless heterotrophic (e.g. Pteridomonas, Ciliophrys) species (Edvardsen et al 2007). This group is named after the silicoflagellates, which are planktonic marine chromists that are both photosynthetic and heterotrophic. Silicoflagellate skeletons usually comprise 1-2% of the siliceous component of marine sediments, which are much less abundant than diatoms. The Pelagophyceae are predominately marine microalgae,

and Pelagomonas and Pelagococcus are important components in the open-ocean picophytoplankton. Most species are non-motile but some species are flagellate. The Pinguiophyceae are a marine microalgal group famous for consisting of a high content of omega-3 fatty acids.

Bicosoecids are one of the most common unicellular heterotrophic flagellates in the aquatic environment. This organism usually feeds on bacteria in the oceanic environment and may have a direct effect on deep-sea marine fluxes of nutrients. Some species (e.g. *Caecitellus* and *Cafeteria*) are amongst the commonest flagellates in coastal waters (del Campo and Massana 2011). Bicosoecides are usually biflagellates with anterior flagellum with tubular mastigonemes. However, some species (*Siluania*) lack posterior flagella and others (*Adriamonas*, *Pseudodendromoans*) have two nearly equal flagella with no hairs. The flagellar transitional region usually lacks a transitional helix. They ingest particles through the R2 microtubules arranged in an L-shape (O'Kelly and Patterson 1996).

In addition, recent environmental surveys have revealed that several new lineages of the previously poorly characterized marine stramenopiles (MAST cells). In marine samples, most of which are heterotrophic organisms (Massana et al 2004b, Massana et al 2014). These unicellular organisms form 18 MAST groups amongst the basal heterotrophic stramenopiles so far, and only the organisms in MAST-3 (Nanomonadea) have been cultured (Cavalier-Smith and Scoble 2013). They were not detected until recently, possibly as they are small and heterotrophic species that possess relatively few morphologically distinctive features. Taxonomic and phylogenetic studies on MAST are necessary to clarify the diversity and evolution of stramenopiles, and understand their ecological roles in marine ecosystem.

1.5.6 Rhizaria

The Rhizaria are a supergroup of mostly unicellular protists containing radiolarians, foraminifers, plasmodiophorids, chlorarachnids, euglyphids and many heterotrophic flagellates. Several phylogenomic analyses suggested the evolutionary affinity of Rhizaria is close to two of the "chromalveolate" groups, stramenopiles and alveolates (the SAR group) (Cavalier-Smith 2010, Hackett et al 2007). The monophyly of the Rhizaria is strongly supported by molecular rDNA studies. They have no clear morphological distinguishing characters (synapomorphies), but most of them possess filose, reticulose, or microtubule-supported pseudopods. The Rhizaria are composed of three highly diverse and possibly monophyletic groups: Foraminifera, Cercozoa and Radiolaria.

Foraminiferans, or forams, are primarily marine organisms, the majority of which live on or within the seafloor sediment, but some have been found in brackish, freshwater and terrestrial habitats (Lejzerowicz et al 2010). The forams typically produce a shell or "test" that are built from various materials (commonly calcium carbonate or agglutinated sediment particles) and constructed in diverse forms. The tests may house diverse unicellular algae as endosymbionts, such as green algae, red algae, diatoms and dinoflagellates (Pflaumann and Jian 1999). A large number of dying planktonic Foraminifera sink down on the seafloor, their mineralized shells preserved as fossils in the accumulating sediment. The Radiolaria produce intricate exteriors of glassy silica with radial or bilateral symmetry, typically with a central capsule dividing the cell into endoplasm and ectoplasm (Yuasa et al 2005). They have many needle-like pseudopods that are supported by bundles of microtubules which radiate outward from the cell bodies of these protists and function to catch food particles. They are considered as zooplankton throughout the ocean, and can often contain symbiotic algae, especially zooxanthellae.

The radiolarians include three groups: Acantharea, Polycystinea and Taxopodida. The Acantharea prossess strontium sulfate skeletons and is closely related to a peculiar genus,

Sticholonche (Taxopodida), which lacks an internal skeleton (Decelle et al 2012). Polycystinea produce skeletons of siliceous material (Krabberod et al 2011). The skeletons of dead radiolarians rain down to the ocean floor, where they make up a large part of the cover of seafloor as siliceous ooze.

The cercozoa are a large protistan group including most amoeboids and flagellates that feed by means of filose pseudopods. This group was established on the basis of molecular phylogenetic studies of rRNA and actin or polyubiquitin as they lack shared morphological characteristics at the microscopic level (Cavalier-Smith 1998, Cavalier-Smith and Chao 2003). Members within this group thrive in soil, freshwater and marine environments, and are very important components in the microbial food web (Bass and Cavalier-Smith 2004, Nikolaev et al 2004). Many cercozoans are bacterivorous (e.g., *Neoheteromita*) and eukaryovorous (e.g., *Metromonas*), thus playing vital roles in nutrient cycles by grazing on microeukaryotes, bacteria and other organic particulates (Myl'nikov and Karpov 2004). However, some lineages are photosynthetic, such as the Chlorarachnea (McFadden et al 1997). The chlorarachniophytes are filose green amoebae that only occur in marine environments where they acquired their plastids from siphonous green algae prey cells through secondary endosymbiosis (Keeling 2004). This initiated great interest from researchers to study the endosymbiotic origins of organelles. Cercozoa are sometimes divided into two subphyla, the Filosa and Endomyxa. The Filosa include many heterotropic flagellates together with some “amoebae”, such as the best-known euglyphids, filose amoebae with shells of siliceous scales or plates. The Endiomyxa includes fungi-like parasites (phytomyxids) and sporozoan parasites (haplosporids) but flagellate members are not found so far (Archibald and Keeling 2004). Although these cercozoan flagellates could be classified into some classes (Metromonadea, Sarcomonadea, Thecofilosea, and Imbricatea), the molecular support is weak and the phenotypic apomorphy is not recognized for each class.

1.5.7 Excavata

Excavates contain parasites of global importance and organisms regarded previously as the oldest members of flagellated organisms (Dawson and Paredez 2013). Although molecular phylogenetic support for this grouping is still controversial, a few clades under this supergroup are well represented as free-living taxa of marine ecosystems, such as well-known heterotrophic flagellates kinetoplastids and diplonemids. The members of euglenozoa clade are very diverse including predators, heterotrophs, photosynthetic autotrophs and parasites. They are common in shallow planktonic and benthic ecosystems, but also appear to inhabit some deep-sea sediments (Buck et al 2000).

1.6 Main drivers of marine microbial community structure

Oceanic ecosystems change in physical, chemical and biological characteristics across spatial and temporal scales. Their structure results from the complex interaction between inhabitant organisms and their environment. The distribution and composition of oceanic plankton (including viruses, prokaryotes, microbial eukaryotes, phytoplankton, and zooplankton) are governed by abiotic and biotic factors. The abiotic factors, such as environmental conditions and nutrient availability have been considered to have a stronger effect (Smetacek 2012). The biotic factors (grazing, pathogenicity, and parasitism) also have a strong effect, and this is has been displayed by an accumulating number of studies (Rohwer and Thurber 2009, Verity and Smetacek 1996).

1.6.1 Abiotic factors

Oceanic temperature is a fundamental environmental factor with geographical, seasonal and temporal variation. It has significant effects on other properties such as biogeochemical cycling of carbon and nitrogen, and on phytoplankton processes such as growth, photo-physiology, and calcification (Hare et al 2007, Rose et al 2009). Temperature in the surface ocean is a

fundamental control on marine eukaryotic phytoplankton community structure and metabolic processes that sets the biogeographical boundaries or biomes of major phytoplankton groups (Moisan et al 2002, Needoba et al 2007, Raven and Geider 1988). In the Tara Ocean global studies, the authors found temperature rather than other environmental factors or geography is the main environmental factor shaping taxonomic and functional microbial community composition in the photic open ocean (Sunagawa et al 2015). These findings also have wide-ranging implications for potential climate change-related effects.

Salinity is a major factor controlling the distribution of biota in aquatic systems. It has been recognized as a principal factor on the survival, growth and development of marine plankton by impacting their cell response to metabolic or osmoregulation changes (Estudillo et al 2000). Increase in the salinity correlates with reduced richness of microbial community diversity (Benlloch et al 2002, Foti et al 2008). It was found to be the key environmental selective force controlling microbial community composition in the Baltic Sea (Herlemann et al 2011). In a large-scale meta-analysis, salinity has been suggested as the major determinant of microbial community (including benthic and pelagic organisms) composition and distribution across many (including ocean) ecosystems and to even exceed the influence of temperature, pH, or other physical and chemical factors (Lozupone and Knight 2007, Nemergut et al 2011, Zinger et al 2011).

Light has strong effects on the microbial composition and distribution in marine environment, especially in the euphotic layer of the ocean (Herlemann et al 2011, Johnson et al 2006, Raes et al 2011). It can directly affect the growth of several microbial groups that are able to sustain, or supplement their energy demands by harvesting photons (Bryant and Frigaard 2006), typically on oxygenic photosynthetic organisms such as cyanobacteria and eukaryotic phytoplanktons. For example, the partitioning of the two main ecotypes of *Prochlorococcus* is mostly influenced by light conditions. These ecotypes (high- and low-light adapted) have

distinct physiology and genetic makeup, which enable them to occupy distinct niches down the water column (Farrant et al 2016, Fuller et al 2003, Morel et al 1993, Partensky et al 1999). Some bacterial phototrophs are capable of anaerobic anoxygenic photosynthesis or proteorhodosin-mediated energy harvesting (e.g. *Rhodobacteraceae*, *Polaribacter*), SAR11 and some Gammaproteobacteria, which are also effected by light density (Cottrell and Kirchman 2009, Giovannoni et al 2005, Stingl et al 2007).

Nutrient availability are important limiting factors to growth, diversity and composition of marine microbes (Smetacek 2012). They vary in concentration and bioavailability across oceanic regions. Important macronutrients in marine ecosystems include carbon, nitrogen, phosphorus, silicon, sulfur, potassium, and sodium. Micronutrients include iron, zinc, copper, manganese and some vitamins. The availability of nutrients in the upper ocean frequently limits the activity and abundance of phytoplankton. Previous studies suggested that nitrogen and phosphorus are the most common limiting factors for marine phytoplankton productivity and metabolism throughout much of the low-latitude ocean surface (Pernice et al 2016, Salazar et al 2016, Villar et al 2015). Surface concentrations of bioavailable nitrogen and phosphorus can be at nanomolar levels in oligotrophic open ocean regions and is rapidly turned over, where picophytoplankton have high abundance (Not et al 2009, Scanlan et al 2009).

1.6.2 Biotic factors

Marine viruses act as major components in marine food web and affect bacteria, archaea and eukaryotic organisms by modulating microbial population size, diversity, metabolic outputs, and gene flow (Brum et al 2015). Viruses are considered to serve as gene reservoirs that change the ecological niches of the host (Sullivan et al 2005). Marine viral cyanophages are able to infect cyanobacteria, and typically these phage genomes can carry genes involved in photosynthesis. Therefore, cyanophages potentially generate and maintain genetic diversity amongst marine picocyanobacteria (Hambly and Suttle 2005, Huang et al 2015).

Grazing is another biotic factor contributing to the structure and abundance of marine microbes. Protists, such as heterotrophic nanoflagellates, ciliates and dinoflagellates, can graze bacteria (e.g. cyanobacteria) (Bouman et al 2006). Larger protists species can also graze smaller protists (Rocke et al 2015). Protists such as heterotrophic dinoflagellates can feed on a diverse array of pray species, ditoms, phototrophic nanoflagellates, mixotrophic dinoflagellates, ciliates, and the eggs, and adult forms of metazoans (Jeong et al 2008).

1.7 Tools for studying marine photosynthetic microorganisms

1.7.1 Cultivation methods

Metabolic and physiological characteristics were applied to describe and classify microorganisms before the development of molecular methods (Staley 2006). It is very difficult to isolate and culture marine microbes because they are small and easily destroyed during the experiments, which make estimates of species abundance using cultivation success impractical. The classical serial dilution method was used to find some of the more abundant organisms, for example, *Micromonas pusilla* and *Hillea marina* in British coastal waters were found using this technique (Knight-Jones 1951). Recent estimates indicate that less than 1% of the actual community in the field have been cultured in the laboratory (Little et al 2008). However, isolation and culture studies provide a chance to investigate the individual characteristics of organisms in respect to their pigment and genetic information, and their optimal growth conditions. In addition, these findings can help to track their natural abundances, and adjust analysis techniques, such as pigment calculations and molecular probes (Massana et al 2004a, Vaulot et al 2008).

1.7.2 Microscopy

Traditional light microscopy can only observe and describe picoplankton organisms. Epifluorescence microscopy measures the emission of light by cellular compounds (e.g.

chlorophyll, phycoerythrin) or by dyes specific compounds (e.g. DNA stained by SYBR green). This method led to the recognition of the importance of picophytoplankton in all marine systems and enumerated different types of cells based on their pigment content or the number and shape of their chloroplasts (Murphy and Haugen 1985). Electron microscopy was developed in the 1960s, which was able to reveal morphological details, such as flagella hairs and species-specific scales (Backe-Hansen and Throndsen 2002). The details of body scales allow unambiguous determination of picoplanktonic species such as *Bathycoccus prasinos* or *Imantonia rotunda* (Eikrem and Throndsen 1990). Scanning electron microscopy has better resolution of surface features, which is useful for some tiny species that have body ornamentation such as organic, silicified or calcified scales. In combination with culturing, these methods can provide detailed information of single organisms and thus improve the understanding of its physiological characteristics and ecological functions (Backe-Hansen and Throndsen 2002). However, the limitations of these techniques is loss of cells in the preparation, clumping of cells or shrinkage and low signal-to-background rations (Vaulot et al 2008).

1.7.3 Flow cytometry

Flow cytometry allows for easy, fast and accurate estimates of picoplankton abundance using the light-scattering properties of each individual cell (a function of cell size and refractive index) and fluorescence from pigments (such as, chlorophyll or phycoerythrin). With its application in oceanography, many important picoplanktonic organisms such as *Prochlorococcus* and *Ostreococcus* have been discovered and quantified (Chisholm et al 1988, Courties et al 1994). It has also allowed the detection of more species, especially strains without fluorescent pigments (such as heterotrophs), which can be stained with fluorescent makers able to bind to specific cell compounds (Simon et al 1994). For example, the use of DNA stains can provide an estimation of genome size and allow separation of species or strains with closely related cell properties (Simon et al 1994). In addition, it also provides a way to sort cells of

interest physically and cultivate pure natural assemblages or perform further investigation on specific groups (Li 1994).

1.7.4 Molecular approaches

In recent years, molecular methods have provided a new approach to study of microbial genetic diversity, spatial-temporal distribution patterns and ecology in natural systems (Biegala et al 2003, Not et al 2009). It allows for the discovery and quantification of the novel uncultivated groups in natural environments. Molecular approaches can successfully identify microorganisms with high taxonomic resolution at a global scale (de Vargas et al 2015). The most commonly used maker is the small subunit (SSU) ribosomal RNA encoding genes (16S rRNA for prokaryotes and 18S rRNA for eukaryotes). These genes are well conserved, slowly evolving and highly abundant, which make them popular choice in investigation of community diversity (Amann and Fuchs 2008). There are also many other phylogenetic markers for discriminate microbes at lower taxonomical levels, such as 16S-23S internal transcribed spaces (ITS), plastid genes, *rbcL* (a large subunit of RuBisCO), *psbA* (photosystem II), *rpoC1* (RNA polymerase subunit), *cpeB* (phycoerythrin), *ntcA* (nitrogen regulator) and *petB* (cytochrome b6 subunit) genes (Ahlgren and Rocap 2012, Penno et al 2006, Zeidner et al 2003). Techniques making use of these genetic markers can either assess the diversity and composition of overall community (cloning libraries, DGGE) or quantify the abundance of specific groups (FISH, qPCR). However, there are several limitations to use of gene sequencing approach when investigating diversity, such as, biases and errors in PCR amplification and production of chimeric sequences. Some groups contain more than one copy of the SSU rRNA gene and some also amplify better than others, which are underrepresented in sequencing library despite their dominance in the natural environment (Not et al 2004, Romari and Vaulot 2004).

1.7.5 Metagenomics

Community metagenomic studies provide powerful tools to construct the genomes from environmentally important microorganisms, bypassing the need for isolation or cultivation of microorganisms (Handelsman 2004). For example, in the Sargasso Sea study, a total of 1.045 billion base pairs of non-redundant sequence were generated, and estimated to derive from at least 1800 genomic species including 148 previously unknown bacterial phylotypes, and identified more than 1.2 million novel genes in these samples (Venter et al 2004). Metagenomics is able to explore and compare the ecology and metabolic profiling of complex environmental microbial communities, and also identify which organisms are undertaking functional roles in the environment (Biddle et al 2008, DeLong et al 2006). Based on sequencing strategies, it can be characterized as unselective (shotgun analysis and next-generation sequencing) and targeted (function-driven and sequence-driven studies) metagenomics (Suenaga 2012). The sequence-based metagenomic analyses of marine microbes have attempted to answer the question of; 'who is there?'; 'what is their role'; 'who is doing what?' and 'what evolutionary processes determine these parameters?' (Kennedy et al 2010). Functional metagenomics is a powerful experimental approach for studying gene function, which can be used to annotate genomes and serve as a complement to sequence-based metagenomics.

1.7.6 Single-cell analysis

Single-cell sequencing has been developed as a very powerful method to obtain coherent data from individual lineages (Ghylin et al 2014, Martinez-Garcia et al 2012), which provides new views to our understanding of genetics by bringing the study of genomes to cellular level. It allows researchers to study the rare and uncultured organisms and dissect the contributions or interactions (such as infections, symbioses and predation) of individual cells to the biology of ecosystems. Previous studies using this approach have revealed a complete genome of a novel

nanovirus from a putatively infected, uncultivated Picozoan protist cell (Yoon et al 2011). Another study sequenced 127 single amplified genomes of uncultured Gammaproteobacteria and indicated that a third of the cells were infected with viruses (Labonte et al 2015). In this approach, microbial cells are first sorted using flow cytometry and collected as a single cell. Then, DNA is isolated and multiple displacement amplification (MDA) is the most widely used technique to amplify the entire genome (Dean et al 2002). However, single cell genomics has many drawbacks and technical challenges, for example, MDA can result in highly uneven genome coverage and chimera formation, single-amplified genomes typically have low genome coverage, and require a highly specialized laboratory facility (Gawad et al 2016).

1.8 Aims of this work

Australia's ocean territory is the third largest on Earth, and roughly 80% of population now lives by the sea. As such, the ocean provides crucial economic, environmental, nutritional and recreational resources central to the nation's identity and iconic way of life. Yet, despite the obvious importance of the ocean, we have only a few understanding of the chief biotic determinant of ocean health and function – the microscopic plankton in Australian oceans. Marine microbes form the foundation of the marine food web and are the engines that drive the chemical cycles that ultimately control the global climate. Therefore, it is important to identify the critical "keystone species" that are most pivotal in governing ocean biogeochemistry and productivity.

The overall aim of this work was to investigate the cellular abundance of cyanobacteria and photosynthetic pico-eukaryotes across temperature and environmental gradients in the Arafura Sea, Torres Strait and Coral Sea during the *RV* Southern expedition in October 2012, undertake molecular analysis of marine prokaryotes and pico-eukaryotes community composition in the surface and euphotic layer, and investigation of the main environmental factors that influence the distribution of natural population of marine microbes in Northern Australian tropical

waters. These findings will provide ideas on the resilience of ocean ecosystems and their response to climate change, and also give suggestions on activities involving human-ocean interactions such as recreation, fisheries and tourism in Australian oceans.

Specific aims:

a. To examine the distribution patterns of marine picophytoplankton across temperature and environmental gradients, and determine the taxonomic composition of prokaryotic microbial communities in the surface seawater samples, and the main environmental drivers on their community composition (Chapter 3)

b. To investigate the diversity of marine microbial eukaryotes in the surface water samples and the major environmental drivers impacting their community composition (Chapter 4)

c. To study the community structure of marine bacteria and protistians within the euphotic zone (200 m) of Arafura Sea/Torres Strait and Coral Sea, and the strong environmental variables on their community structure (Chapter 5)

1.9 References

Ahlgren NA, Rocap G (2012). Diversity and distribution of marine *Synechococcus*: multiple gene phylogenies for consensus classification and development of qPCR assays for sensitive measurement of clades in the ocean. *Front Microbiol* **3**.

Alldredge AL, Silver MW (1988). Characteristics, dynamics and significance of marine snow. *Prog Oceanogr* **20:** 41-82.

Amann R, Fuchs BM (2008). Single-cell identification in microbial communities by improved fluorescence in situ hybridization techniques. *Nat Rev Microbiol* **6:** 339-348.

Archibald JM, Keeling PJ (2004). Actin and ubiquitin protein sequences support a cercozoan/foraminiferan ancestry for the plasmodiophorid plant pathogens. *J Eukaryot Microbiol* **51:** 113-118.

Azam F, Fenchel T, Field JG, Gray JS, Meyerreil LA, Thingstad F (1983). The ecological role of water-column microbes in the sea. *Mar Ecol Prog Ser* **10:** 257-263.

Backe-Hansen P, Throndsen J (2002). Pico- and nanoplankton from the inner Oslofjord, eastern Norway, including description of two new species of *Luffisphaera* (incerta sedis). *Sarsia* **87:** 55-63.

Bass D, Cavalier-Smith T (2004). Phylum-specific environmental DNA analysis reveals remarkably high global biodiversity of Cercozoa (Protozoa). *Int J Syst Evol Micr* **54:** 2393-2404.

Benlloch S, Lopez-Lopez A, Casamayor EO, Ovreas L, Goddard V, Daae FL *et al* (2002). Prokaryotic genetic diversity throughout the salinity gradient of a coastal solar saltern. *Environ Microbiol* **4:** 349-360.

Berg IA (2011). Ecological aspects of the distribution of different autotrophic CO_2 fixation pathways. *Appl Environ Microb* **77:** 1925-1936.

Biddle JF, Fitz-Gibbon S, Schuster SC, Brenchley JE, House CH (2008). Metagenomic signatures of the Peru Margin subseafloor biosphere show a genetically distinct environment.

Proceedings of the National Academy of Sciences of the United States of America **105:** 10583-10588.

Biegala IC, Not F, Vaulot D, Simon N (2003). Quantitative assessment of picoeukaryotes in the natural environment by using taxon-specific oligonucleotide probes in association with tyramide signal amplification-fluorescence in situ hybridization and flow cytometry. *Appl Environ Microb* **69:** 5519-5529.

Blankenship RE (1992). Origin and early evolution of photosynthesis. *Photosynth Res* **33:** 91-111.

Bouman HA, Ulloa O, Scanlan DJ, Zwirglmaier K, Li WKW, Platt T *et al* (2006). Oceanographic basis of the global surface distribution of *Prochlorococcus* ecotypes. *Science* **312:** 918-921.

Brown MW, Spiegel FW, Silberman JD (2009). Phylogeny of the "forgotten" cellular slime mold, *Fonticula alba*, reveals a key evolutionary branch within Opisthokonta. *Mol Biol Evol* **26:** 2699-2709.

Brum JR, Ignacio-Espinoza JC, Roux S, Doulcier G, Acinas SG, Alberti A *et al* (2015). Patterns and ecological drivers of ocean viral communities. *Science* **348**.

Bryant DA, Frigaard NU (2006). Prokaryotic photosynthesis and phototrophy illuminated. *Trends Microbiol* **14:** 488-496.

Buck KR, Barry JP, Simpson AGB (2000). Monterey Bay cold seep biota: euglenozoa with chemoautotrophic bacterial epibionts. *Eur J Protistol* **36:** 117-126.

Burgaud G, Le Calvez T, Arzur D, Vandenkoornhuyse P, Barbier G (2009). Diversity of culturable marine filamentous fungi from deep-sea hydrothermal vents. *Environ Microbiol* **11:** 1588-1600.

Burja AM, Banaigs B, Abou-Mansour E, Burgess JG, Wright PC (2001). Marine cyanobacteria - a prolific source of natural products. *Tetrahedron* **57:** 9347-9377.

Burki F (2014). The eukaryotic tree of life from a global phylogenomic perspective. *Csh Perspect Biol* **6**.

Campbell L, Liu HB, Nolla HA, Vaulot D (1997). Annual variability of phytoplankton and bacteria in the subtropical North Pacific Ocean at Station ALOHA during the 1991-1994 ENSO event. *Deep-Sea Res Pt I* **44:** 167-192.

Caron DA, Countway PD, Jones AC, Kim DY, Schnetzer A (2012). Marine protistan diversity. *Annu Rev Mar Sci* **4:** 467-493.

Castenholz RW (2015). Cyanobacteria. *Bergey's Manual of Systematics of Archaea and Bacteria*. John Wiley & Sons, Ltd.

Cavalier-Smith T (1998). A revised six-kingdom system of life. *Biol Rev* **73:** 203-266.

Cavalier-Smith T, Chao EEY (2003). Phylogeny and classification of phylum Cercozoa (Protozoa). *Protist* **154:** 341-358.

Cavalier-Smith T (2010). Kingdoms Protozoa and Chromista and the eozoan root of the eukaryotic tree. *Biol Letters* **6:** 342-345.

Cavalier-Smith T, Scoble JM (2013). Phylogeny of Heterokonta: incisomonas marina, a uniciliate gliding opalozoan related to *Solenicola* (Nanomonadea), and evidence that Actinophryida evolved from raphidophytes. *Eur J Protistol* **49:** 328-353.

Cavalier-Smith T, Chao EE, Lewis R (2015). Multiple origins of Heliozoa from flagellate ancestors: new cryptist subphylum Corbihelia, superclass Corbistoma, and monophyly of Haptista, Cryptista, Hacrobia and Chromista. *Mol Phylogenet Evol* **93:** 331-362.

Chassot E, Bonhommeau S, Dulvy NK, Melin F, Watson R, Gascuel D *et al* (2010). Global marine primary production constrains fisheries catches. *Ecol Lett* **13:** 495-505.

Chisholm SW, Olson RJ, Zettler ER, Goericke R, Waterbury JB, Welschmeyer NA (1988). A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nature* **334:** 340-343.

Cottrell MT, Kirchman DL (2009). Photoheterotrophic microbes in the Arctic Ocean in summer and winter. *Appl Environ Microb* **75:** 4958-4966.

Countway PD, Caron DA (2006). Abundance and distribution of *Ostreococcus* sp. in the San Pedro Channel, California, as revealed by quantitative PCR. *Appl Environ Microb* **72:** 2496-2506.

Courties C, Vaquer A, Troussellier M, Lautier J, Chretiennotdinet MJ, Neveux J *et al* (1994). Smallest eukaryotic organism. *Nature* **370:** 255-255.

Dadachova E, Bryan RA, Huang XC, Moadel T, Schweitzer AD, Aisen P *et al* (2007). Ionizing radiation changes the electronic properties of melanin and enhances the growth of melanized fungi. *Plos One* **2**.

Davis K (2004). Photosynthesis got a really early start. *New Sci* **184:** 14-14.

Davy SK, Allemand D, Weis VM (2012). Cell biology of cnidarian-dinoflagellate symbiosis. *Microbiol Mol Biol R* **76:** 229-261.

Dawson SC, Paredez AR (2013). Alternative cytoskeletal landscapes: cytoskeletal novelty and evolution in basal excavate protists. *Curr Opin Cell Biol* **25:** 134-141.

de Vargas C, Audic S, Henry N, Decelle J, Mahe F, Logares R *et al* (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**.

Dean FB, Hosono S, Fang LH, Wu XH, Faruqi AF, Bray-Ward P *et al* (2002). Comprehensive human genome amplification using multiple displacement amplification. *Proceedings of the National Academy of Sciences of the United States of America* **99:** 5261-5266.

Decelle J, Suzuki N, Mahe F, de Vargas C, Not F (2012). Molecular phylogeny and morphological evolution of the Acantharia (Radiolaria). *Protist* **163:** 435-450.

del Campo J, Massana R (2011). Emerging diversity within chrysophytes, choanoflagellates and bicosoecids based on molecular surveys. *Protist* **162:** 435-448.

DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard NU *et al* (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311:** 496-503.

Derelle E, Ferraz C, Rombauts S, Rouze P, Worden AZ, Robbens S *et al* (2006). Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proceedings of the National Academy of Sciences of the United States of America* **103:** 11647-11652.

Dismukes GC, Klimov VV, Baranov SV, Kozlov YN, DasGupta J, Tyryshkin A (2001). The origin of atmospheric oxygen on Earth: The innovation of oxygenic photosynthesis. *Proceedings of the National Academy of Sciences of the United States of America* **98:** 2170-2175.

Duarte CM, Cebrian J (1996). The fate of marine autotrophic production. *Limnol Oceanogr* **41:** 1758-1766.

Dugdale RC, Wilkerson FP (1998). Silicate regulation of new production in the equatorial Pacific upwelling. *Nature* **391:** 270-273.

DuRand MD, Olson RJ, Chisholm SW (2001). Phytoplankton population dynamics at the Bermuda Atlantic Time-series station in the Sargasso Sea. *Deep-Sea Res Pt Ii* **48:** 1983-2003.

Dykova I, Lom J (2004). Advances in the knowledge of amphizoic amoebae infecting fish. *Folia Parasit* **51:** 81-97.

Edvardsen B, Eikrem W, Shalchian-Tabrizi K, Riisberg I, Johnsen G, Naustvoll L *et al* (2007). *Verrucophora farcimen* gen. et sp nov (Dictyochophyceae, Heterokonta) - A bloom-forming ichthyotoxic flagellate from the Skagerrak, Norway. *J Phycol* **43:** 1054-1070.

Eikrem W, Throndsen J (1990). The ultrastructure of *Bathycoccus* gen. nov. and *Bathycoccus*. *Prasinos* sp. nov., a nonmotile picoplanktonic alga (Chlorophyta, Prasinophyceae) from the Mediterranean and Atlantic. *Phycologia* **29:** 344-350.

Estudillo CB, Duray MN, Marasigan ET, Emata AC (2000). Salinity tolerance of larvae of the mangrove red snapper (*Lutjanus argentimaculatus*) during ontogeny. *Aquaculture* **190:** 155-167.

Falkowski PG, Katz ME, Knoll AH, Quigg A, Raven JA, Schofield O *et al* (2004). The evolution of modern eukaryotic phytoplankton. *Science* **305:** 354-360.

Farrant GK, Dore H, Cornejo-Castillo FM, Partensky F, Ratin M, Ostrowski M *et al* (2016). Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America* **113:** E3365-E3374.

Fast NM, Xue LR, Bingham S, Keeling PJ (2002). Re-examining alveolate evolution using multiple protein molecular phylogenies. *J Eukaryot Microbiol* **49:** 30-37.

Fensome RA, Saldarriaga JF, Taylor FJRM (1999). Dinoflagellate phylogeny revisited: reconciling morphological and molecular based phylogenies. *Grana* **38:** 66-80.

Field CB, Behrenfeld MJ, Randerson JT, Falkowski P (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281:** 237-240.

Finlay BJ, Esteban GF, Fenchel T (1998). Protozoan diversity: converging estimates of the global number of free-living ciliate species. *Protist* **149:** 29-37.

Fischer WW, Hemp J, Johnson JE (2016). Evolution of oxygenic photosynthesis. *Annu Rev Earth Pl Sc* **44:** 647-683.

Flombaum P, Gallegos JL, Gordillo RA, Rincon J, Zabala LL, Jiao NAZ *et al* (2013). Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proceedings of the National Academy of Sciences of the United States of America* **110:** 9824-9829.

Fogg GE (1996). Algae: an introduction to phycology. *Nature* **381:** 660-660.

Foti MJ, Sorokin DY, Zacharova EE, Pimenov NV, Kuenen JG, Muyzer G (2008). Bacterial diversity and activity along a salinity gradient in soda lakes of the Kulunda Steppe (Altai, Russia). *Extremophiles* **12:** 133-145.

Fouilland E, Descolas-Gros C, Courties C, Collos Y, Vaquer A, Gasc A (2004). Productivity and growth of a natural population of the smallest free-living eukaryote under nitrogen deficiency and sufficiency. *Microb Ecol* **48:** 103-110.

Fu G, Nagasato C, Oka S, Cock JM, Motomura T (2014). Proteomics analysis of heterogeneous flagella in brown algae (Stramenopiles). *Protist* **165:** 662-675.

Fuller NJ, Marie D, Partensky F, Vaulot D, Post AF, Scanlan DJ (2003). Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus* clade throughout a stratified water column in the Red Sea. *Appl Environ Microb* **69:** 2430-2443.

Garcia-Pichel F, Belnap J, Neuer S, Schanz F (2003). Estimates of global cyanobacterial biomass and its distribution. *Algological Studies* **109:** 213-227.

Garcia-Pichel F (2009). Cyanobacteria. In: Schaechter M (ed). *Encyclopedia of Microbiology (Third Edition)*. Academic Press: Oxford. pp 107-124.

Garczarek L, Dufresne A, Rousvoal S, West NJ, Mazard S, Marie D *et al* (2007). High vertical and low horizontal diversity of *Prochlorococcus* ecotypes in the Mediterranean Sea in summer. *Fems Microbiol Ecol* **60:** 189-206.

Gawad C, Koh W, Quake SR (2016). Single-cell genome sequencing: current state of the science. *Nat Rev Genet* **17:** 175-188.

Geller A, Muller DG (1981). Analysis of the flagellar beat pattern of male *Ectocarpus-siliculosus* gametes (Phaeophyta) in relation to chemotactic stimulation by female cells. *J Exp Biol* **92:** 53-66.

Ghylin TW, Garcia SL, Moya F, Oyserman BO, Schwientek P, Forest KT *et al* (2014). Comparative single-cell genomics reveals potential ecological niches for the freshwater acl Actinobacteria lineage. *ISME Journal* **8:** 2503-2516.

Giovannoni SJ, Bibbs L, Cho JC, Stapels MD, Desiderio R, Vergin KL *et al* (2005). Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature* **438:** 82-85.

Goericke R, Repeta DJ (1992). The pigments of *Prochlorococcus marinus*: the presence of divinyl chlorophyll *a* and chlorophyll *b* in a marine prokaryote. *Limnol Oceanogr* **37:** 425-433.

Gould SB, Waller RR, McFadden GI (2008). Plastid evolution. *Annu Rev Plant Biol* **59:** 491-517.

Gross M (2012). The mysteries of the diatoms. *Curr Biol* **22:** R581-R585.

Grossman AR, Schaefer MR, Chiang GG, Collier JL (1993). The phycobilisome, a light-harvesting complex responsive to environmental-conditions. *Microbiol Rev* **57:** 725-749.

Grula JW (2005). Evolution of photosynthesis and biospheric oxygenation contingent upon nitrogen fixation? *Int J Astrobiol* **4:** 251-257.

Guillou L, Viprey M, Chambouvet A, Welsh RM, Kirkham AR, Massana R *et al* (2008). Widespread occurrence and genetic diversity of marine parasitoids belonging to Syndiniales (Alveolata). *Environ Microbiol* **10:** 3349-3365.

Hackett JD, Yoon HS, Li S, Reyes-Prieto A, Rummele SE, Bhattacharya D (2007). Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of rhizaria with chromalveolates. *Mol Biol Evol* **24:** 1702-1713.

Hambly E, Suttle CA (2005). The viriosphere, diversity, and genetic exchange within phage communities. *Curr Opin Microbiol* **8:** 444-450.

Handelsman J (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol R* **68:** 669-685.

Hare CE, Leblanc K, DiTullio GR, Kudela RM, Zhang Y, Lee PA *et al* (2007). Consequences of increased temperature and CO_2 for phytoplankton community structure in the Bering Sea. *Mar Ecol Prog Ser* **352:** 9-16.

Hedges SB, Chen H, Kumar S, Wang DYC, Thompson AS, Watanabe H (2001). A genomic timescale for the origin of eukaryotes. *Bmc Evol Biol* **1**.

Henson BJ, Hesselbrock SM, Watson LE, Barnum SR (2004). Molecular phylogeny of the heterocystous cyanobacteria (subsections IV and V) based on *nifD*. *Int J Syst Evol Micr* **54:** 493-497.

Herlemann DPR, Labrenz M, Jurgens K, Bertilsson S, Waniek JJ, Andersson AF (2011). Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME Journal* **5:** 1571-1579.

Hibbett DS, Binder M, Bischoff JF, Blackwell M, Cannon PF, Eriksson OE *et al* (2007). A higher-level phylogenetic classification of the fungi. *Mycol Res* **111:** 509-547.

Hofmann E, Wrench PM, Sharples FP, Hiller RG, Welte W, Diederichs K (1996). Structural basis of light harvesting by carotenoids: peridinin-chlorophyll-protein from *Amphidinium carterae*. *Science* **272:** 1788-1791.

Huang SJ, Zhang S, Jiao NZ, Chen F (2015). Marine cyanophages demonstrate biogeographic patterns throughout the global ocean. *Appl Environ Microb* **81:** 441-452.

Hyde KD, Jones EBG, Leano E, Pointing SB, Poonyth AD, Vrijmoed LLP (1998). Role of fungi in marine ecosystems. *Biodivers Conserv* **7:** 1147-1161.

Jameson E, Joint I, Mann NH, Muhling M (2010). Detailed analysis of the microdiversity of *Prochlorococcus* populations along a North-South Atlantic Ocean transect. *Environ Microbiol* **12:** 156-171.

Jeong HJ, Seong KA, Du Yoo Y, Kim TH, Kang NS, Kim S *et al* (2008). Feeding and grazing impact by small marine heterotrophic dinoflagellates on heterotrophic bacteria. *J Eukaryot Microbiol* **55:** 271-288.

Jiao NZ, Yang YH, Hong N, Ma Y, Harada S, Koshikawa H *et al* (2005). Dynamics of autotrophic picoplankton and heterotrophic bacteria in the East China Sea. *Cont Shelf Res* **25:** 1265-1279.

Johnson ZI, Zinser ER, Coe A, McNulty NP, Woodward EMS, Chisholm SW (2006). Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* **311:** 1737-1740.

Katz LA (2001). Evolution of nuclear dualism in ciliates: a reanalysis in light of recent molecular data. *Int J Syst Evol Micr* **51:** 1587-1592.

Keeling PJ (2004). Diversity and evolutionary history of plastids and their hosts. *Am J Bot* **91:** 1481-1493.

Keeling PJ (2010). The endosymbiotic origin, diversification and fate of plastids. *Philos T R Soc B* **365:** 729-748.

Kennedy J, Flemer B, Jackson SA, Lejon DPH, Morrissey JP, O'Gara F *et al* (2010). Marine metagenomics: new tools for the study and exploitation of marine microbial metabolism. *Mar Drugs* **8:** 608-628.

Knight-Jones EW (1951). Preliminary studies of nanoplankton and ultraplankton systematics and abundance by a quantitative culture method. *Journal du Conseil* **17:** 140-155.

Krabberod AK, Brate J, Dolven JK, Ose RF, Klaveness D, Kristensen T *et al* (2011). Radiolaria divided into polycystina and spasmaria in combined 18S and 28S rDNA phylogeny. *Plos One* **6**.

Kroger N, Poulsen N (2008). Diatoms-from cell wall biogenesis to nanotechnology. *Annu Rev Genet* **42:** 83-107.

Labonte JM, Swan BK, Poulos B, Luo HW, Koren S, Hallam SJ *et al* (2015). Single-cell genomics-based analysis of virus-host interactions in marine surface bacterioplankton. *ISME Journal* **9:** 2386-2399.

Latysheva N, Junker VL, Palmer WJ, Codd GA, Barker D (2012). The evolution of nitrogen fixation in cyanobacteria. *Bioinformatics* **28:** 603-606.

Leadbeater B (2008). Choanoflagellate evolution: the morphological perspective. *Protistology* **5:** 256-267.

Leakey RJG, Leadbeater BSC, Mitchell E, McCready SMM, Murray AWA (2002). The abundance and biomass of choanoflagellates and other nanoflagellates in waters of contrasting

temperature to the north-west of South Georgia in the Southern Ocean. *Eur J Protistol* **38:** 333-350.

Leal MC, Sa C, Nordez S, Brotas V, Paula J (2009). Distribution and vertical dynamics of planktonic communities at Sofala Bank, Mozambique. *Estuar Coast Shelf S* **84:** 605-616.

Leander BS, Keeling PJ (2004). Early evolutionary history of dinoflagellates and Apicomplexans (Alveolata) as inferred from HSP90 and actin phylogenies. *J Phycol* **40:** 341-350.

Lejzerowicz F, Pawlowski J, Fraissinet-Tachet L, Marmeisse R (2010). Molecular evidence for widespread occurrence of Foraminifera in soils. *Environ Microbiol* **12:** 2518-2525.

Li AS, Stoecker DK, Coats DW, Adam EJ (1996). Ingestion of fluorescently labeled and phycoerythrin-containing prey by mixotrophic dinoflagellates. *Aquat Microb Ecol* **10:** 139-147.

Li WKW (1994). Primary production of prochlorophytes, cyanobacteria, and eukaryotic ultraphytoplankton: measurements from flow cytometric sorting. *Limnol Oceanogr* **39:** 169-175.

Little AEF, Robinson CJ, Peterson SB, Raffa KE, Handelsman J (2008). Rules of engagement: interspecies interactions that regulate microbial communities. *Annual Review of Microbiology* **62:** 375-401.

Liu HB, Nolla HA, Campbell L (1997). *Prochlorococcus* growth rate and contribution to primary production in the equatorial and subtropical north Pacific Ocean. *Aquat Microb Ecol* **12:** 39-47.

Lozupone CA, Knight R (2007). Global patterns in bacterial diversity. *Proceedings of the National Academy of Sciences of the United States of America* **104:** 11436-11440.

Lynn DH (2001). Ciliophora. *eLS*. John Wiley & Sons, Ltd.

Mackiewicz P, Gagat P (2014). Monophyly of Archaeplastida supergroup and relationships among its lineages in the light of phylogenetic and phylogenomic studies. Are we close to a consensus? *Acta Soc Bot Pol* **83:** 263-280.

Magana HA, Contreras C, Villareal TA (2003). A historical assessment of Karenia brevis in the western Gulf of Mexico. *Harmful Algae* **2:** 163-171.

Malmstrom RR, Coe A, Kettler GC, Martiny AC, Frias-Lopez J, Zinser ER *et al* (2010). Temporal dynamics of *Prochlorococcus* ecotypes in the Atlantic and Pacific oceans. *ISME Journal* **4:** 1252-1264.

Malmstrom RR, Rodrigue S, Huang KH, Kelly L, Kern SE, Thompson A *et al* (2013). Ecology of uncultured *Prochlorococcus* clades revealed through single-cell genomics and biogeographic analysis. *ISME Journal* **7:** 184-198.

Mann KH (1988). Production and use of detritus in various fresh-water, estuarine, and coastal marine ecosystems. *Limnol Oceanogr* **33:** 910-930.

Marchant HJ (1985). Choanoflagellates in the Antarctic marine food chain. *Antarctic Nutrient Cycles and Food Webs***:** 271-276.

Martinez-Garcia M, Swan BK, Poulton NJ, Gomez ML, Masland D, Sieracki ME *et al* (2012). High-throughput single-cell sequencing identifies photoheterotrophs and chemoautotrophs in freshwater bacterioplankton. *ISME Journal* **6:** 113-123.

Martiny AC, Tai APK, Veneziano D, Primeau F, Chisholm SW (2009). Taxonomic resolution, ecotypes and the biogeography of *Prochlorococcus*. *Environ Microbiol* **11:** 823-832.

Massana R, Balague V, Guillou L, Pedros-Alio C (2004a). Picoeukaryotic diversity in an oligotrophic coastal site studied by molecular and culturing approaches. *Fems Microbiol Ecol* **50:** 231-243.

Massana R, Castresana J, Balague V, Guillou L, Romari K, Groisillier A *et al* (2004b). Phylogenetic and ecological analysis of novel marine stramenopiles. *Appl Environ Microb* **70:** 3528-3534.

Massana R, Pedros-Alio C (2008). Unveiling new microbial eukaryotes in the surface ocean. *Curr Opin Microbiol* **11:** 213-218.

Massana R (2011). Eukaryotic picoplankton in surface oceans. *Annual Review of Microbiology* **65:** 91-110.

Massana R, del Campo J, Sieracki ME, Audic S, Logares R (2014). Exploring the uncultured microeukaryote majority in the oceans: reevaluation of ribogroups within stramenopiles. *ISME Journal* **8:** 854-866.

Matsunaga S, Uchida H, Iseki M, Watanabe M, Murakami A (2010). Flagellar motions in phototactic steering in a brown algal swarmer. *Photochem Photobiol* **86:** 374-381.

McFadden GI, Gilson PR, Hofmann CJB (1997). Division chlorarachniophyta. *Plant Syst Evol***:** 175-185.

Medlin LK, Kooistra WHCF, Potter D, Saunders GW, Andersen RA (1997). Phylogenetic relationships of the 'Golden algae' (haptophytes, heterokont chromophytes) and their plastids. *Plant Syst Evol***:** 187-219.

Mendoza L, Taylor JW, Ajello L (2002). The class mesomycetozoea: a group of microorganisms at the animal-fungal boundary. *Annual Review of Microbiology* **56:** 315-344.

Moisan JR, Moisan TA, Abbott MR (2002). Modelling the effect of temperature on the maximum growth rates of phytoplankton populations. *Ecol Model* **153:** 197-215.

Moore LR, Rocap G, Chisholm SW (1998). Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* **393:** 464-467.

Moore LR, Chisholm SW (1999). Photophysiology of the marine cyanobacterium *Prochlorococcus*: ecotypic differences among cultured isolates. *Limnol Oceanogr* **44:** 628-638.

Moran DM, Anderson OR, Dennett MR, Caron DA, Gast RJ (2007). A description of seven Antarctic marine Gymnamoebae including a new subspecies, two new species and a new genus: *Neoparamoeba aestuarina antarctica* n. subsp., *Platyamoeba oblongata* n. sp., *Platyamoeba contorta* n. sp and *Vermistella antarctica* n. gen. n. sp. *J Eukaryot Microbiol* **54:** 169-183.

Morel A, Ahn YH, Partensky F, Vaulot D, Claustre H (1993). *Prochlorococcus* and *Synechococcus*: a comparative-study of their optical-properties in relation to their size and pigmentation. *J Mar Res* **51:** 617-649.

Murphy LS, Haugen EM (1985). The distribution and abundance of phototrophic ultraplankton in the North-Atlantic. *Limnol Oceanogr* **30:** 47-58.

Myl'nikov AP, Karpov SA (2004). Review of diversity and taxonomy of cercomonads. *Protistology* **3:** 201-217.

Needoba JA, Foster RA, Sakamoto C, Zehr JP, Johnson KS (2007). Nitrogen fixation by unicellular diazotrophic cyanobacteria in the temperate oligotrophic North Pacific Ocean. *Limnol Oceanogr* **52:** 1317-1327.

Nelson DM, Treguer P, Brzezinski MA, Leynaert A, Queguiner B (1995). Production and dissolution of biogenic silica in the ocean: revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Global Biogeochem Cy* **9:** 359-372.

Nemergut DR, Costello EK, Hamady M, Lozupone C, Jiang L, Schmidt SK *et al* (2011). Global patterns in the biogeography of bacterial taxa. *Environ Microbiol* **13:** 135-144.

Newell SY (1996). Established and potential impacts of eukaryotic mycelial decomposers in marine/terrestrial ecotones. *J Exp Mar Biol Ecol* **200:** 187-206.

Nikolaev SI, Berney C, Fahrni JF, Bolivar I, Polet S, Mylnikov AP *et al* (2004). The twilight of heliozoa and rise of rhizaria, an emerging supergroup of amoeboid eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America* **101:** 8066-8071.

Not F, Latasa M, Marie D, Cariou T, Vaulot D, Simon N (2004). A single species, *Micromonas pusilla* (Prasinophyceae), dominates the eukaryotic picoplankton in the western English channel. *Appl Environ Microb* **70:** 4064-4072.

Not F, del Campo J, Balague V, de Vargas C, Massana R (2009). New insights into the diversity of marine picoeukaryotes. *Plos One* **4**.

Nowicka B, Kruk J (2016). Powered by light: phototrophy and photosynthesis in prokaryotes and its evolution. *Microbiol Res* **186:** 99-118.

O'Kelly CJ, Patterson DJ (1996). The flagellar apparatus of *Cafeteria roenbergensis* Fenchel & Patterson, 1988 (Bicosoecales equals Bicosoecida). *Eur J Protistol* **32:** 216-226.

Ong LJ, Glazer AN (1991). Phycoerythrins of marine unicellular cyanobacteria. I. Bilin types and locations and energy-transfer pathways in *Synechococcus* spp. phycoerythrins. *J Biol Chem* **266:** 9515-9527.

Partensky F, Hess WR, Vaulot D (1999). *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol R* **63:** 106-127.

Partensky F, Garczarek L (2010). *Prochlorococcus*: advantages and limits of minimalism. *Annual Review of Marine Science* **2:** 305-331.

Passow U, Carlson CA (2012). The biological pump in a high CO_2 world. *Mar Ecol Prog Ser* **470:** 249-271.

Patterson DJ (1999). The diversity of eukaryotes. *Am Nat* **154:** S96-S124.

Pawlowski J (2013). The new micro-kingdoms of eukaryotes. *Bmc Biol* **11**.

Penno S, Lindell D, Post AF (2006). Diversity of *Synechococcus* and *Prochlorococcus* populations determined from DNA sequences of the N-regulatory gene *ntcA*. *Environ Microbiol* **8:** 1200-1211.

Perkerson RB, Johansen JR, Kovacik L, Brand J, Kastovsky J, Casamatta DA (2011). A unique pseudanabaenalean (cyanobacteria) genus *nodosilinea* gen. nov based on morphological and molecular data. *J Phycol* **47:** 1397-1412.

Pernice MC, Giner CR, Logares R, Perera-Bel J, Acinas SG, Duarte CM *et al* (2016). Large variability of bathypelagic microbial eukaryotic communities across the world's oceans. *ISME Journal* **10:** 945-958.

Pflaumann U, Jian ZM (1999). Modern distribution patterns of planktonic foraminifera in the south China Sea and western Pacific: a new transfer technique to estimate regional sea-surface temperatures. *Mar Geol* **156:** 41-83.

Raes J, Letunic I, Yamada T, Jensen LJ, Bork P (2011). Toward molecular trait-based ecology through integration of biogeochemical, geographical and metagenomic data. *Mol Syst Biol* **7**.

Raghukumar C, Raghukumar S (1998). Barotolerance of fungi isolated from deep-sea sediments of the Indian Ocean. *Aquat Microb Ecol* **15:** 153-163.

Raven JA (1983). The transport and function of silicon in plants. *Biological Reviews of the Cambridge Philosophical Society* **58:** 179-207.

Raven JA, Geider RJ (1988). Temperature and algal growth. *New Phytol* **110:** 441-461.

Reyes-Prieto A, Weber APM, Bhattacharya D (2007). The origin and establishment of the plastid in algae and plants. *Annu Rev Genet* **41:** 147-168.

Richards TA, Bass D (2005). Molecular screening of free-living microbial eukaryotes: diversity and distribution using a meta-analysis. *Curr Opin Microbiol* **8:** 240-252.

Richards TA, Cavalier-Smith T (2005). Myosin domain evolution and the primary divergence of eukaryotes. *Nature* **436:** 1113-1118.

Richards TA, Jones MDM, Leonard G, Bass D (2012). Marine fungi: their ecology and molecular diversity. *Annu Rev Mar Sci* **4:** 495-522.

Rocap G, Distel DL, Waterbury JB, Chisholm SW (2002). Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Appl Environ Microb* **68:** 1180-1191.

Rocap G, Larimer FW, Lamerdin J, Malfatti S, Chain P, Ahlgren NA *et al* (2003). Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424:** 1042-1047.

Rocke E, Pachiadaki MG, Cobban A, Kujawinski EB, Edgcomb VP (2015). Protist community grazing on prokaryotic prey in deep ocean water masses. *Plos One* **10**.

Roenneberg T, Taylor W (2000). Automated recordings of bioluminescence with special reference to the analysis of circadian rhythms. *Method Enzymol* **305:** 104-119.

Roger AJ, Simpson AGB (2009). Evolution: revisiting the root of the eukaryote tree. *Curr Biol* **19:** R165-R167.

Rohwer F, Thurber RV (2009). Viruses manipulate the marine environment. *Nature* **459:** 207-212.

Romain Derelle PL-G, Hélène Timpano and David Moreira (2016). A phylogenomic framework to study the diversity and evolution of stramenopiles (=heterokonts). *Mol Biol Evol* **33**.

Romari K, Vaulot D (2004). Composition and temporal variability of picoeukaryote communities at a coastal site of the English Channel from 18S rDNA sequences. *Limnol Oceanogr* **49:** 784-798.

Rose JM, Feng Y, DiTullio GR, Dunbar RB, Hare CE, Lee PA *et al* (2009). Synergistic effects of iron and temperature on Antarctic phytoplankton and microzooplankton assemblages. *Biogeosciences* **6:** 3131-3147.

Rublee PA, Remington DL, Schaefer EF, Marshall MM (2005). Detection of the dinozoans *Pfiesteria piscicida* and *P.shumwayae*: a review of detection methods and geographic distribution. *J Eukaryot Microbiol* **52:** 83-89.

Salazar G, Cornejo-Castillo FM, Benitez-Barrios V, Fraile-Nuez E, Alvarez-Salgado XA, Duarte CM *et al* (2016). Global diversity and biogeography of deep-sea pelagic prokaryotes. *ISME Journal* **10:** 596-608.

Sarthou G, Timmermans KR, Blain S, Treguer P (2005). Growth physiology and fate of diatoms in the ocean: a review. *J Sea Res* **53:** 25-42.

Scanlan DJ, Ostrowski M, Mazard S, Dufresne A, Garczarek L, Hess WR *et al* (2009). Ecological genomics of marine picocyanobacteria. *Microbiol Mol Biol R* **73:** 249-299.

Seymour JR (2014). A sea of microbes: the diversity and activity of marine microorganisms. *Microbiology Australia*.

Sherr BF, Sherr EB, Caron DA, Vaulot D, Worden AZ (2007). Oceanic protists. *Oceanography* **20:** 130-134.

Sherr EB, Sherr BF (2007). Heterotrophic dinoflagellates: a significant component of microzooplankton biomass and major grazers of diatoms in the sea. *Mar Ecol Prog Ser* **352:** 187-197.

Sibley LD (2010). How apicomplexan parasites move in and out of cells. *Curr Opin Biotech* **21:** 592-598.

Simon N, Barlow RG, Marie D, Partensky F, Vaulot D (1994). Characterization of oceanic photosynthetic picoeukaryotes by flow cytometry. *J Phycol* **30:** 922-935.

Smetacek V (1999). Diatoms and the ocean carbon cycle. *Protist* **150:** 25-32.

Smetacek V (2012). Making sense of ocean biota: how evolution and biodiversity of land organisms differ from that of the plankton. *J Biosciences* **37:** 589-607.

Smetacek VS (1985). Role of sinking in diatom life-history cycles: ecological, evolutionary and geological significance. *Mar Biol* **84:** 239-251.

Staley JT (2006). The bacterial species dilemma and the genomic-phylogenetic species concept. *Philos T R Soc B* **361:** 1899-1909.

Stingl U, Desiderio RA, Cho JC, Vergin KL, Giovannoni SJ (2007). The SAR92 clade: an abundant coastal clade of culturable marine bacteria possessing proteorhodopsin. *Appl Environ Microb* **73:** 2290-2296.

Suenaga H (2012). Targeted metagenomics: a high-resolution metagenomics approach for specific gene clusters in complex microbial communities. *Environ Microbiol* **14:** 13-22.

Sullivan MB, Coleman ML, Weigele P, Rohwer F, Chisholm SW (2005). Three *Prochlorococcus* cyanophage genomes: signature features and ecological interpretations. *Plos Biol* **3:** 790-806.

Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G *et al* (2015). Structure and function of the global ocean microbiome. *Science* **348**.

Taylor FJR, Hoppenrath M, Saldarriaga JF (2008). Dinoflagellate diversity and distribution. *Biodivers Conserv* **17:** 407-418.

Thomas DN (2002). Seaweeds. *London: Natural History Museum*.

Thomsen HA, Buck KR (1998). Nanoflagellates of the central California waters: taxonomy, biogeography and abundance of primitive, green flagellates (pedinophyceae, prasinophyceae). *Deep-Sea Res Pt Ii* **45:** 1687-1707.

Tice MM, Lowe DR (2004). Photosynthetic microbial mats in the 3,416-Myr-old ocean. *Nature* **431:** 549-552.

Tirichine L, Bowler C (2011). Decoding algal genomes: tracing back the history of photosynthetic life on Earth. *Plant J* **66:** 45-57.

Torruella G, Derelle R, Paps J, Lang BF, Roger AJ, Shalchian-Tabrizi K *et al* (2012). Phylogenetic relationships within the opisthokonta based on phylogenomic analyses of conserved single-copy protein domains. *Mol Biol Evol* **29:** 531-544.

Vaulot D, Eikrem W, Viprey M, Moreau H (2008). The diversity of small eukaryotic phytoplankton in marine ecosystems. *Fems Microbiol Rev* **32:** 795-820.

Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA *et al* (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304:** 66-74.

Verity PG, Smetacek V (1996). Organism life cycles, predation, and the structure of marine pelagic ecosystems. *Mar Ecol Prog Ser* **130:** 277-293.

Vieira AAH, Teixeira C (1982). Excretion of dissolved organic-matter by natural-populations of marine-phytoplankton at different time lengths during the photoperiod. *Rev Microbiol* **13:** 206-210.

Villar E, Farrant GK, Follows M, Garczarek L, Speich S, Audic S *et al* (2015). Environmental characteristics of Agulhas rings affect interocean plankton transport. *Science* **348**.

Vincent WF (2009). Cyanobacteria. In: Likens GE (ed). *Encyclopedia of Inland Waters*. Academic Press: Oxford. pp 226-232.

Waterbury JB, Watson SW, Guillard RRL, Brand LE (1979). Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium. *Nature* **277:** 293-294.

West NJ, Scanlan DJ (1999). Niche-partitioning of *Prochlorococcus* populations in a stratified water column in the eastern North Atlantic Ocean. *Appl Environ Microb* **65:** 2585-2591.

West NJ, Lebaron P, Strutton PG, Suzuki MT (2011). A novel clade of *Prochlorococcus* found in high nutrient low chlorophyll waters in the South and Equatorial Pacific Ocean. *ISME Journal* **5:** 933-944.

Wilhelm SW, Suttle CA (1999). Viruses and nutrient cycles in the sea. *Bioscience* **49:** 781-788.

Worden AZ (2006). Picoeukaryote diversity in coastal waters of the Pacific Ocean. *Aquat Microb Ecol* **43:** 165-175.

Worden AZ, Lee JH, Mock T, Rouze P, Simmons MP, Aerts AL *et al* (2009). Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *micromonas*. *Science* **324:** 268-272.

Worden AZ, Follows MJ, Giovannoni SJ, Wilken S, Zimmerman AE, Keeling PJ (2015). Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science* **347**.

Xiong J, Fischer WM, Inoue K, Nakahara M, Bauer CE (2000). Molecular evidence for the early evolution of photosynthesis. *Science* **289:** 1724-1730.

Yool A, Tyrrell T (2003). Role of diatoms in regulating the ocean's silicon cycle. *Global Biogeochem Cy* **17**.

Yoon HS, Andersen RA, Boo SM, Bhattacharya D (2009). Stramenopiles *Encyclopedia of Microbiology (Third Edition)*. Academic Press: Oxford. pp 721-731.

Yoon HS, Price DC, Stepanauskas R, Rajah VD, Sieracki ME, Wilson WH *et al* (2011). Single-cell genomics reveals organismal interactions in uncultivated marine protists. *Science* **332:** 714-717.

Yuasa T, Takahashi O, Honda D, Mayama S (2005). Phylogenetic analyses of the polycystine radiolaria based on the 18s rDNA sequences of the spumellarida and the nassellarida. *Eur J Protistol* **41:** 287-298.

Zeidner G, Preston CM, Delong EF, Massana R, Post AF, Scanlan DJ *et al* (2003). Molecular diversity among marine picophytoplankton as revealed by *psbA* analyses. *Environ Microbiol* **5:** 212-216.

Zinger L, Amaral-Zettler LA, Fuhrman JA, Horner-Devine MC, Huse SM, Welch DBM *et al* (2011). Global patterns of bacterial beta-diversity in seafloor and seawater ecosystems. *Plos One* **6**.

Zinser ER, Johnson ZI, Coe A, Karaca E, Veneziano D, Chisholm SW (2007). Influence of light and temperature on *Prochlorococcus* ecotype distributions in the Atlantic Ocean. *Limnol Oceanogr* **52:** 2205-2220.

Zwirglmaier K, Jardillier L, Ostrowski M, Mazard S, Garczarek L, Vaulot D *et al* (2008). Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environ Microbiol* **10:** 147-161.

Chapter 2:

Insights into the diversity and biogeography of surface sea prokaryotes in Australian tropical waters

Insights into the diversity and biogeography of surface sea prokaryotes in Australian tropical waters

Taotao Huang[1], Martin Ostrowski[1], Sheemal S. Kumar[1], Mark V. Brown[2], Justin Seymour[3], Ian T. Paulsen[1]

1. Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, Australia
2. School of Biotechnology and Biomolecular Science, University of New South Wales, Sydney, Australia
3. Climate Change Cluster, University of Technology Sydney, Sydney, Australia

Corresponding Author: Ian Paulsen, Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, New South Wales, 2109, Australia.

Email: ian.paulsen@mq.edu.au

Abstract

The Arafura Sea, Torres Strait and Coral Sea are considered as the high productive basins with the complex hydrological marine environments. However, little is known about the diversity and biogeography of bacterial communities inhabiting surface seawaters in the Australian tropics. Here we report the first survey of prokaryotic communities in these waters using 16S rRNA gene amplicon sequencing. This work identified 4,208 operational taxonomic units (OTUs) with Cyanobacteria and Proteobacteria representing the numerically dominant phyla through all samples. However, distinct subgroups within these phyla show remarkable correlation with different geographical regions. We observed that the prokaryotic community composition is homogeneous in the Coral Sea but spatially heterogeneous in the Arafura Sea/Torres Strait. Based on cell counts the Arafura Sea and Torres Strait region was enriched with cyanobacterial primary producers, the abundance of *Synechococcus* often exceeded 10^5 cells mL^{-1}. In contrast, the Coral Sea had a higher abundance of *Prochlorococcus* (up to 4×10^5 cells mL^{-1}), and SAR11 clade heterotrophic bacteria. While all of the sites sampled in waters across the north of Australia demonstrated high productivity (Chl *a* > 0.08 mg m^3) ordination analysis identified specific environmental variables, including lower nutrient concentrations and higher salinity in the Coral Sea that correlated with the differences in prokaryotic community structure observed in the Arafura Sea and Torres Strait.

Introduction

Prokaryotes play a fundamental role in microbial food webs and biogeochemical cycles in marine ecosystems (Copley 2002, Ducklow and Carlson 1992, Nemergut et al 2011, Pomeroy et al 2007). They constitute up to 70% and 75% of the total biomass in surface and deep waters, respectively (Fuhrman et al 1989, Gasol et al 1997). There are many environmental factors shaping microbial assemblages whose impacts can be observed along on a gradient of local, regional and global scales.

Biogeographic patterns in microbial communities are shaped by the interaction of environmental selection and historical effects (Hanson et al 2012, Martiny et al 2011, O'Malley 2008). According to the classic assumption that 'everything is everywhere, but the environment selects' (Baas-Becking 1934), many studies have found significant correlation between measured environmental conditions and microbial composition (Campbell et al 2011, Monier et al 2015, Salazar et al 2016, Zhang et al 2014). Historical processes have also been considered as an influence on the phylogenetic composition of communities in recent studies (Green and Bohannan 2006, Papke et al 2003, Ramette and Tiedje 2007a). These historical processes include a distance effect and past selection, along with dispersal limitation, which generate a distance decay relationship in similarities of microbial communities, suggesting that the variation in microbial composition increases with increasing geographical distance (Hanson et al 2012). Here we examined microbial assemblages inhabiting distinct water masses in the Northern tropical waters of Australia to examine the various roles of environment and distance in shaping assemblage structure.

The Arafura Sea is a shallow (50-80 m) semi-enclosed continental shelf basin, between the Northern Australian and Indonesian landmasses, covering about 650,000 km^2. This basin is part of the highly productive (>300 g C m^{-2} y^{-1}) North Australian Large Marine Ecosystem

(Sherman 2008). It is fully tropical, experiencing relatively stable trade winds during part of the year and fitful monsoons typically between November and April (Jongsma 1974). While offshore conditions may be more oligotrophic, nutrient enrichment in the western Arafura Sea is thought to be maintained by deepwater undercurrent priming from the Banda Sea and tidal mixing (Kämpf 2015). The very shallow (7 to 15 m) (Harris 1988) Torres Strait lies between the Gulf of Papua and the continental shelf of the Great Barrier Reef (GBR), linking the Arafura Sea to the Coral Sea, providing the only alternative pathway to the Indonesian Throughflow for exchange of tropical waters between the Pacific and Indian Oceans (Tomczak 2003). Together, the Arafura Sea and Torres Strait marine ecosystems have been labelled as one of the least impacted by human activities, despite being a productive fishery and busy international shipping lanes (Halpern et al 2008).

The Coral Sea is a marginal sea of the South Pacific off the northeast coast of Australia. The euphotic zone in the Coral Sea is highly oligotrophic and nutrient concentrations and annual primary production are low (Condie and Dunn 2006). The surface currents in the Coral Sea are split into two branches upon meeting the continental shelf edge, one flowing north and the other flowing south which contributes to the formation of the East Australian Current (Andrews and Clegg 1989). Tropical cyclones are generated frequently in this region along with heavy monsoon rains in the Coral Sea and Arafura Sea during summer (December to March).

Although these Australian tropical basins have peculiar environmental variables and marine ecosystems, the broad composition and biogeography of microbial communities have not been studied thoroughly. It has been reported, based on pigments and microscopy methods, that large diatoms and cyanobacteria are the dominant phytoplankton in the Northern Australian waters (Hallegraeff 1984), and that cyanobacterial nitrogen fixation may play an important role in supplying additional N for primary production offshore (Burford et al 2009). There has been recordings of extremely high rates of nitrogen fixation in this region (Messer et al 2016,

Montoya et al 2004), which has been linked to high numbers of nitrogenase (*nifH*) gene sequences from filamentous cyanobacteria *Trichodesmia* in spring, as well, the heterotrophic unicellular diazotrophic bacteria in the Arafura Sea during winter. The unicellular cyanobacteria (UCYN-A1) are however absent from the Arafura Sea but consist of a significant proportion of *nifH* sequences in the Coral Sea during winter and spring (Messer et al., 2016), suggesting distinct differences in drivers of nitrogen fixation in each water mass.

In this study, we collected surface water samples from the Arafura Sea, Torres Strait and Coral Sea on a transect from Darwin to Cairns during the *RV* Southern Surveyor expedition in October 2012. Surface microbial communities were examined at high spatial resolution (~21 nautical mile separation) by flow cytometry and 16S rRNA amplicon sequencing. Moreover, we investigated the influence of environmental heterogeneity and geographic position on biogeography of microbial communities in these contrasting water bodies.

Materials and methods

Sample collection and environmental parameters measurement

Samples were collected during a research cruise of the *RV* Southern Surveyor in October 2012 undertaken prior to the start of the tropical wet season (Figure 2.1). A total of 84 water samples were collected from the surface underway seawater supply (5 m depth) or using 10 L Niskin bottles during 17 CTD rosette casts with associated conductivity, temperature, pressure, chlorophyll fluorescence and dissolved oxygen measurements of water samples. For each sample, 4 mL of seawater was fixed with 1% paraformaldehyde (final concentration) (Marie et al 1997) for flow cytometry and 8 L was filtered through 0.22 µm pore-size white polyethylesulfone filters for DNA extraction. Both sample types were immediately frozen in liquid nitrogen and stored at -80 °C. The concentrations of inorganic nutrients were determined from depth profile samples collected from CTD stations and processed according to

standardized protocols. Instrument details and data are available via the Australian Ocean Data Network https://researchdata.ands.org.au/southern-surveyor-voyage-ss2012t07hydrology

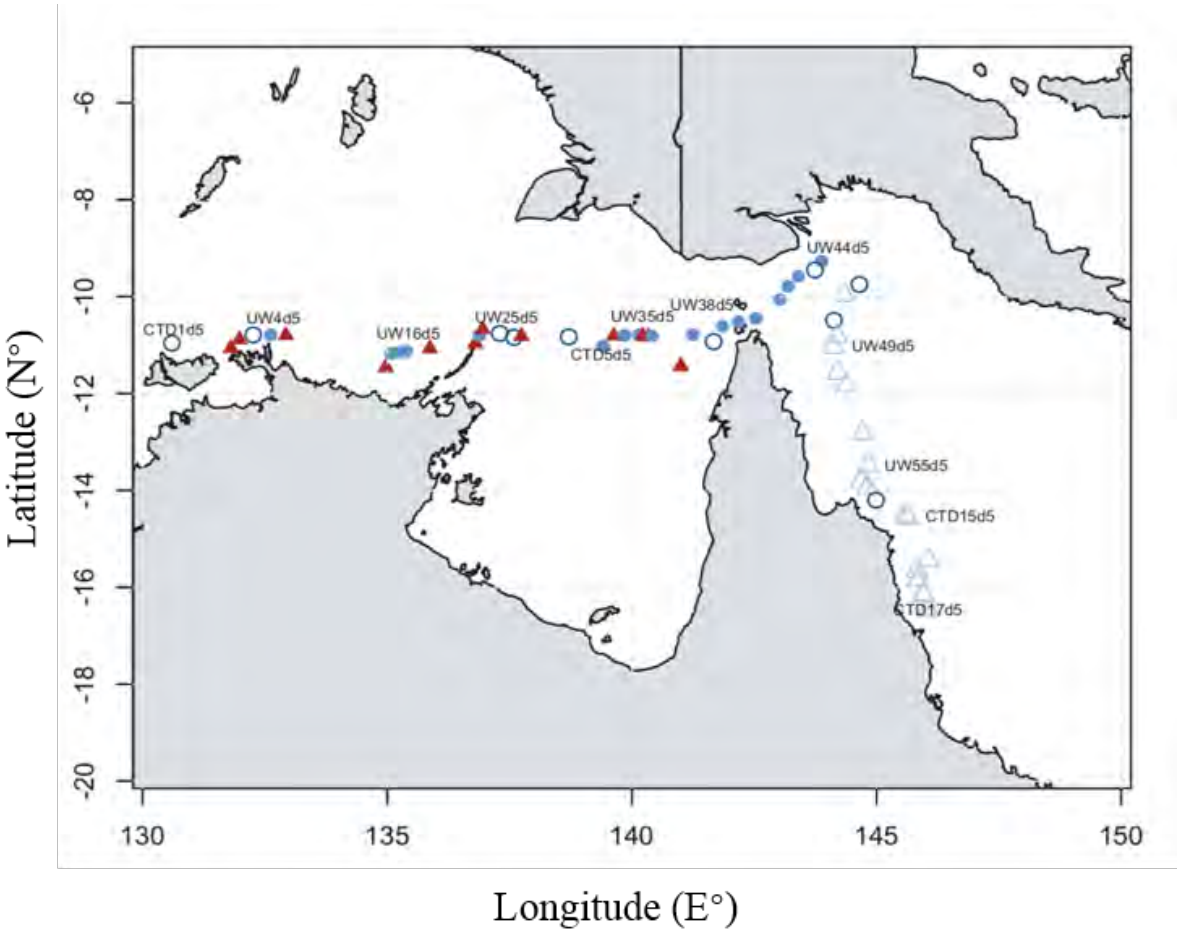

Figure 2.1 Schematic representation of the sampling sites across the Arafura Sea, Torres Strait and Coral Sea during an oceanographic transect in the austral spring of 2012. Samples are grouped into 4 clusters according to simprof analysis: Arafura Sea/Torres Strait (ASTS) 1 (circle), ASTS2 (solid triangle), ASTS3 (solid circle) and Coral Sea (CS, triangle).

Cell abundance measurements

The abundance of cyanobacteria *Synechococcus* and *Prochlorococcus* cells were determined in unstained samples by flow cytometry with a BD INFLUX equipped with a forward scatter PMT (Becton Dickinson, San Jose, CA, USA) using the fluorescence of natural photosynthetic pigments and illuminated by a 200 mW 488 nm laser (Marie et al 1997). Preserved samples were thawed at room temperature and run through the flow cytometer. The trigger was set on the red fluorescence for chlorophyll (690/40 nm bandpass filter). *Synechococcus* were separated based on the presence of both red chlorophyll fluorescence and orange phycoerythrin fluorescence (580/30 bandpass filter). *Prochlorococcus* were distinguished by their small size

and the presence of red chlorophyll fluorescence on forward scatter vs red fluorescence collected with a 670/30 nm band pass filter. Yellow-green beads (0.5 micron; Polysciences, Warrington, PA, USA) were added to each sample and used as an internal standard for calculating absolute cell abundances.

DNA extraction and 16S rRNA gene amplicon sequencing

DNA extractions were performed using the MoBio Power water kit (MoBio Laboratories, Carlsbad, CA, USA), according to the manufacturer's instructions. Prokaryotic diversity was assessed by sequencing of the V1-V3 region of 16S rRNA amplicons with the Illumina MiSeq platform using paired end reads (2×250 bp) and primer set 27F (5'-AGAGTTTGATCMTGGC TCAG-3') and 519R (5'- GWATTACCGCGGCKGCTG-3') (Lane et al 1985, Winsley et al 2012). The barcodes and specific primers for Illumina sequencing using the Nextera Index Kit (Illumina) were added to the 5' ends of each primer pair. The final concentrations in each 50 µl PCR reaction were 0.5 mM each primer, 2.5 mM $MgCl_2$, 1× Buffer (Promega, Wisconsin, United States), 0.2 mM dNTP, 2.5 U Taq DNA polymerase. The amplification conditions comprised an initial denaturation step at 95 °C for 10 min, followed by 35 cycles of 95 °C for 30 s, 55 °C for 10 s and 72 °C for 45 s, with a final extension at 72 °C for 5 min. PCR products were run on a 1.5% agarose gel to check amplicon lengths and were quantified by Quant-IT PicoGreen assay (ThermoFisher, Massachusetts, United States). The PCR products were pooled and purified using AMPure XP beads (Beckman Coulter, Lane Cove, Australia) and sequenced at the Ramaciotti Centre for Genomics.

Sequencing process and microbial community analysis

Sequence processing and analyses were carried out according to the amplicon sequence analysis workflow for the Australian Marine Microbial Biodiversity Initiative (https://gigascience.biomedcentral.com/articles/10.1186/s13742-016-0126-5). Briefly, raw paired ends data were merged by FLASH (Magoc and Salzberg 2011) before using Fastx tools to exclude reads less

than 200 nt and low quality scores (<20). Chimeric sequences were identified during clustering with USEARCH 64bit v 8.1.1756 (Edgar et al 2011). Operational taxonomic units (OTUs) formed by sequences within the quality filtered, trimmed and joined sequences were assigned against a combined AMMBI/BASE reference OTU database (16 June 2016) at a 97% sequence similarity cutoff. Reference OTUs were classified using Mothur classify.seqs (Kozich et al 2013) against the Silva123 database (Quast et al 2013) using default settings. Any OTUs classified as 'Cyanobacteria; Chloroplasts' were further classified against the curated PhytoREF database release of May 2014 (http://www.ncbi.nlm.nih.gov/pubmed/25740460). Community composition information was summarized from filtered biom files.

Statistical data analysis

Statistical analyses were carried out with the R platform (version 3.2.1) using the ecodist and vegan packages. Bray-Curtis dissimilarity was used to determine the differences in community structure between the samples after comparing the samples rarefaction to 7,320 sequences. Non-metric Multidimensional Scaling (NMDS) was applied to the Bray-Curtis dissimilarity matrix to visualize the community structure. Significant clusters were defined using hierarchal clustering with simprof from the clustsig package using the Ward algorithm and default parameters. Environmental variables were fitted to the NMDS plot using envfit (vegan) and similarity percentage analysis (simper) was used to define the contribution of each species to the dissimilarity matrix. Geographic distances were obtained between sampling locations, computed using a least-cost distance strategy as implemented in the 'gdistance' R package. A Mantel correlogram was used to assess the relationship between community composition similarity and environmental variables or geographic distance matrices, and the significance was assessed using 999 permutations. Distance classes of 3,500 km were used. A Linear discriminant analysis (LDA) effect size (LEfSe) pipeline (Segata et al 2011) used to identify differences in taxonomy composition between Arafura Sea/Torres Strait and Coral Sea. Nonparametric factorial Kruskal–Wallis rank sum test was applied to detect taxa with

significant differential abundances. Alpha values of 0.05 were used for the KW rank sum test, and a threshold of 2.0 on logarithmic LDA scores was applied for discriminative features.

Results

Environmental parameters

The Arafura Sea/Torres Strait waters displayed higher temperature, lower salinities and greater nutrient concentrations than the Coral Sea waters during our study period (Figure 2.2).

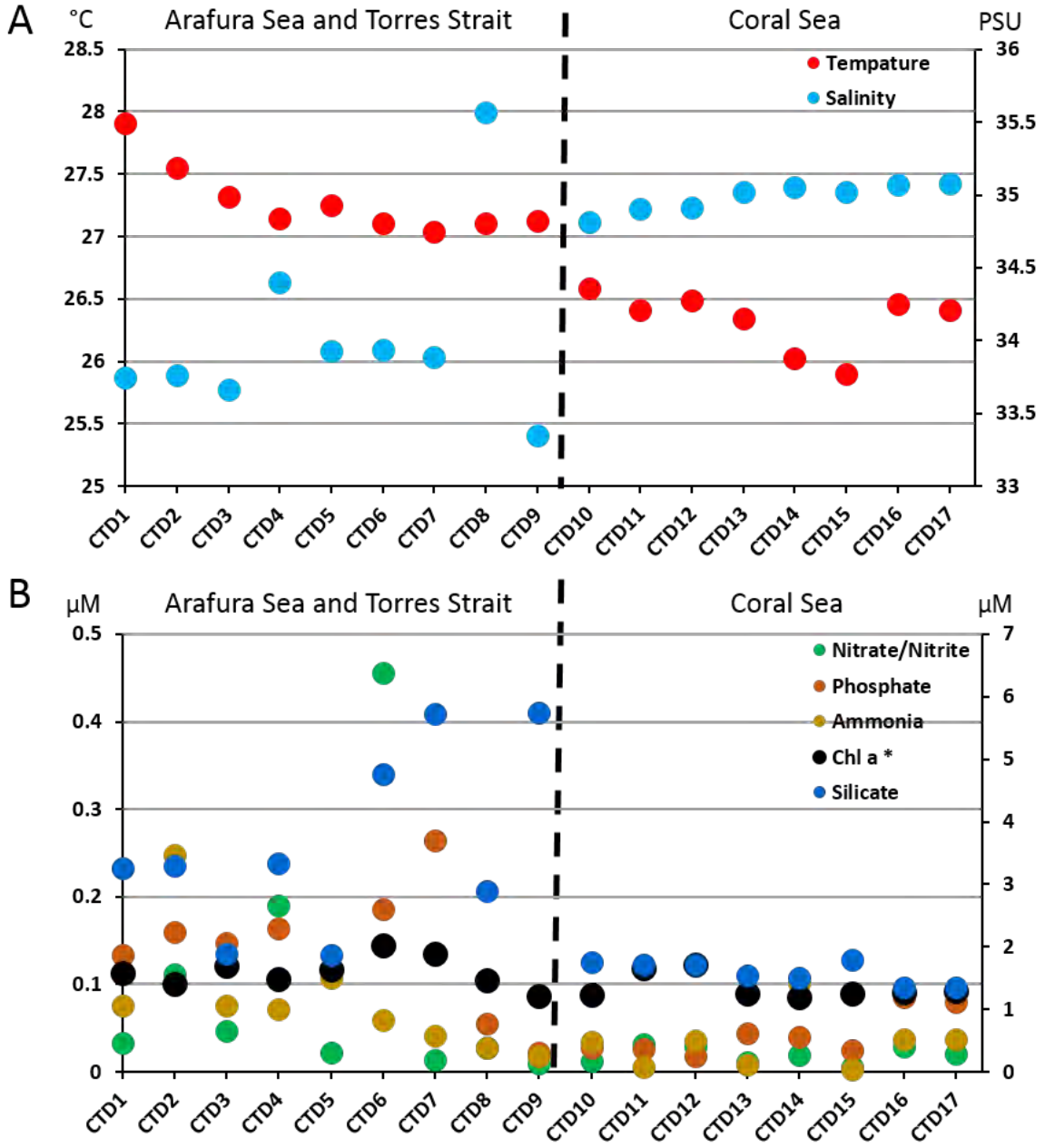

Figure 2.2 Distribution of physical and chemical measurements along the sampling stations. Temperature (red, left axis) and salinity (sky blue, right axis) is shown in A. Nutrients concentrations is shown in B. The right axis is the concentration for silicate (dark blue). The left axis is the concentration for nitrate, phosphate and ammonia. The values of Chl *a* (black, *nb. units are µg/L) is on the right axis.

In the Arafura Sea/Torres Strait, surface temperature and salinity ranged from 27.1 °C to 28.9 °C and from 33.35-34.4 PSU, respectively. In the Coral Sea, the salinity was relatively stable (~35 PSU) in all sampling stations and the average surface temperature ranged from 25.9 °C to 26.6 °C. The nutrient concentrations fluctuated significantly in the Arafura Sea/Torres Strait waters but were more stable in the Coral Sea (Figure 2.2B). At the CTD6 sampling site (Arafura Sea), inorganic N was almost 28-fold higher than in the Coral Sea, reaching 0.456 µM and 0.016 µM, respectively. NH^{4+} concentration peaked at 0.25 µM at the CTD2 sampling site and decreased to around 0.03 µM in the Coral Sea. PO_4^{3-} levels were 0.13–0.26 µM in the Arafura Sea/Torres Strait and 0.02–0.08 µM in Coral Sea. SiO_4^{4-} concentrations displayed the same overall patterns as PO_4^{3-} with the exception of the CTD9 sampling site, which showed a much higher silicate concentration (5.74 µM). The average chlorophyll concentration was slightly higher in Arafura Sea/Torres Strait than Coral Sea, with the value of 0.11 and 0.09 mg/m^3, respectively.

Unicellular phytoplankton cell abundances along transects

Autotrophic unicellular phytoplankton (*Prochlorococcus*, *Synechococcus* and photosynthetic picoeukaryotes (PPE)) are important primary producers that constitute the base of the marine food web. We determined their abundance by flow cytometry using the fluorescence of natural photosynthetic pigments (chlorophyll, phycoerythrin) (Figure 2.3 and 2.4). Generally, PPE abundances were consistently slightly lower in the Coral Sea than the Arafura Sea/Torres Strait, and consistently lower than those of *Synechococcus*, representing, on average, just 4% of *Synechococcus* counts throughout the transect.

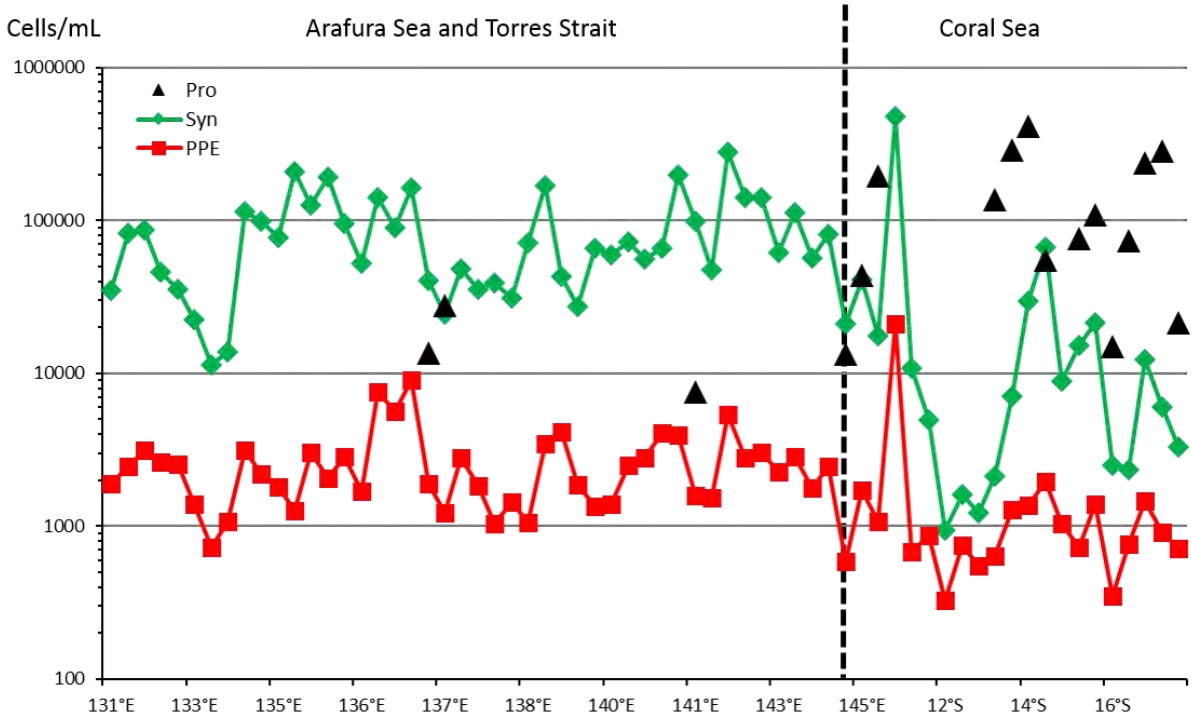

Figure 2.3 Abundance of photosynthetic phytoplankton (red, PPEs), *Synechococcus* (green, Syn) and *Prochlorococcus* (black, Pro) cells in the surface water of the Arafura Sea/Torres Strait and the Coral Sea.

We observed a transition between *Prochlorococcus* and *Synechococcus* cell abundances from the Arafura Sea/Torres Strait to the Coral Sea. *Synechococcus* ranged from 1×10^4 to 3×10^5 cells mL^{-1} in the Arafura Sea/Torres Strait and from 1×10^3 to 6×10^4 cells mL^{-1} in the Coral Sea. While *Prochlorococcus* cells were difficult to enumerate, particularly within the Arafura Sea/Torres Strait (potentially due to their small size, low abundances or water turbidity), we were able to determine *Prochlorococcus* abundances at 17 of the sampled sites (Figure 2.3). This data suggested that *Prochlorococcus* cell numbers were much higher in the Coral Sea (1-4.5×10^5 cells mL^{-1}) than in the Arafura Sea/Torres Strait (8×10^3-2.8×10^4 cells mL^{-1}). A simple one-factor T-test also supports the significant differences between *Prochlorococcus* (t=0.08) and *Synechococcus* (t=0.01) concentrations from the Arafura Sea and the Coral Sea region.

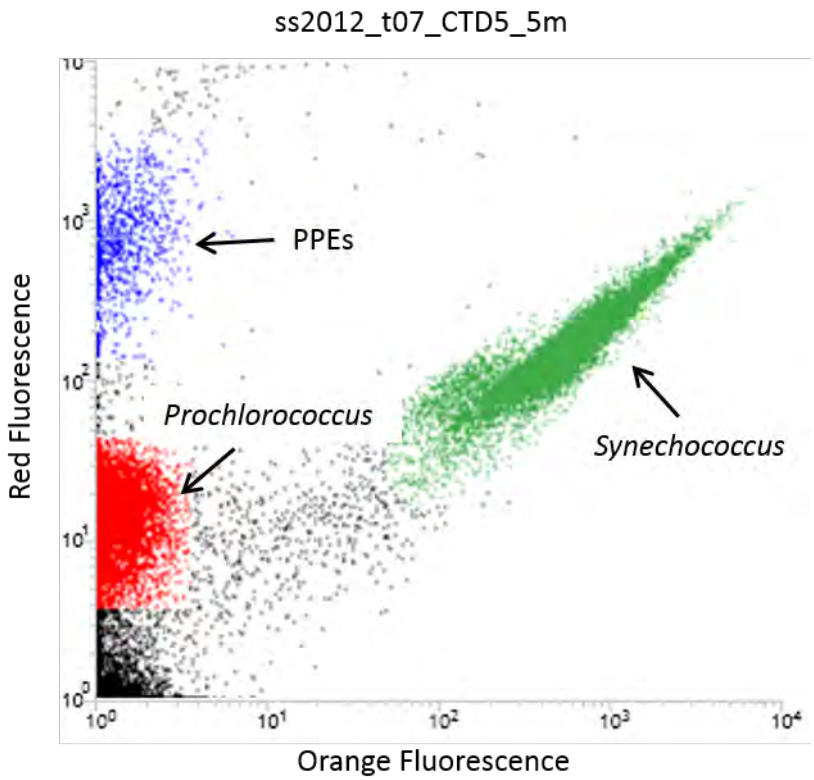

Figure 2.4 An exemplary cytogram illustrating the cell abundance of pico-phytoplankton counted by flow cytometry using the surface seawater sample in CTD5. The *Synechococcus* population is discriminated from other phytoplanktons with the use of orange fluorescence (from the presence of phycoerythrin). *Prochlorococcus* are distinguished from photosynthetic picoeukaryotes (PPEs) by its small cell size and limited chlorophyll content.

Taxonomic abundance and richness in different basins

A total of ~0.5 million V1-V3 16S rRNA sequences were obtained for classification after quality checking and chimera removal. These sequences were clustered into 4,208 operational taxonomic units (OTUs) at 97% similarity resulting in 234 lineages after classification. Rarefaction curves indicated that the sequencing coverage provided a good representation of the microbial species richness at both local and regional scales (Figure 2.5A). Regionally integrated prokaryotic community diversity estimated by the Shannon Index (Figure 2.5B) was higher in the Coral Sea in comparison to the Arafura Sea/Torres Strait basin, which was supported by a simple one-factor T-test (t=0.01).

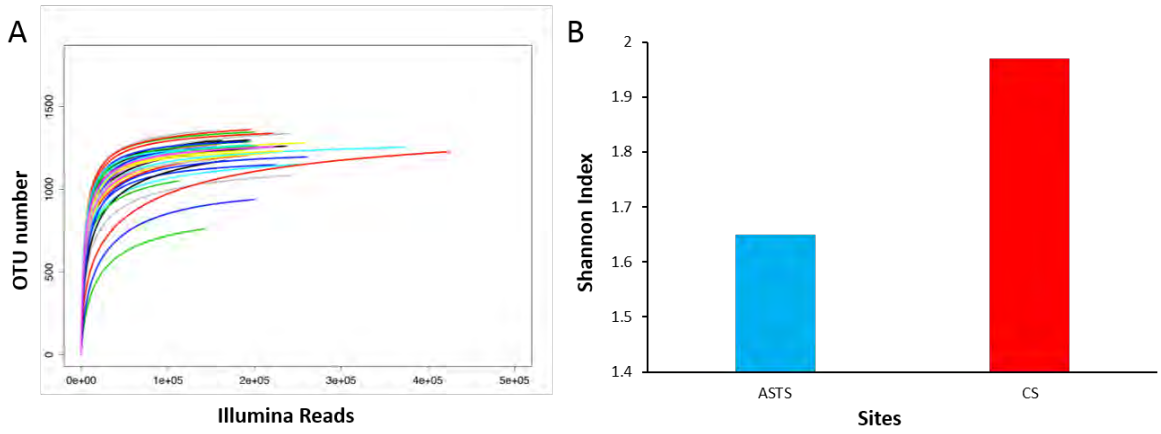

Figure 2.5 Rarefaction curves of the variable region V1-V3 16S rRNA OTUs (97% identity) for individual samples (A) and overall diversity (B, Shannon index) for samples collected in the Arafura Sea/Torres Strait (ASTS) and the Coral Sea (CS).

The two marine cyanobacteria clades, along with the Alphaproteobacteria SAR11 clade, were the three major taxa in the Northern Australian waters. Further, significant proportions of OM1 clade Candidatus *Actinomarina*, Rhodobacteriaceae (Roseobacter), SAR86 clade, SAR116 clade and Flavobacteria marine clades NS5, NS4 and NS2b were observed across the two different regions (Figure 2.6).

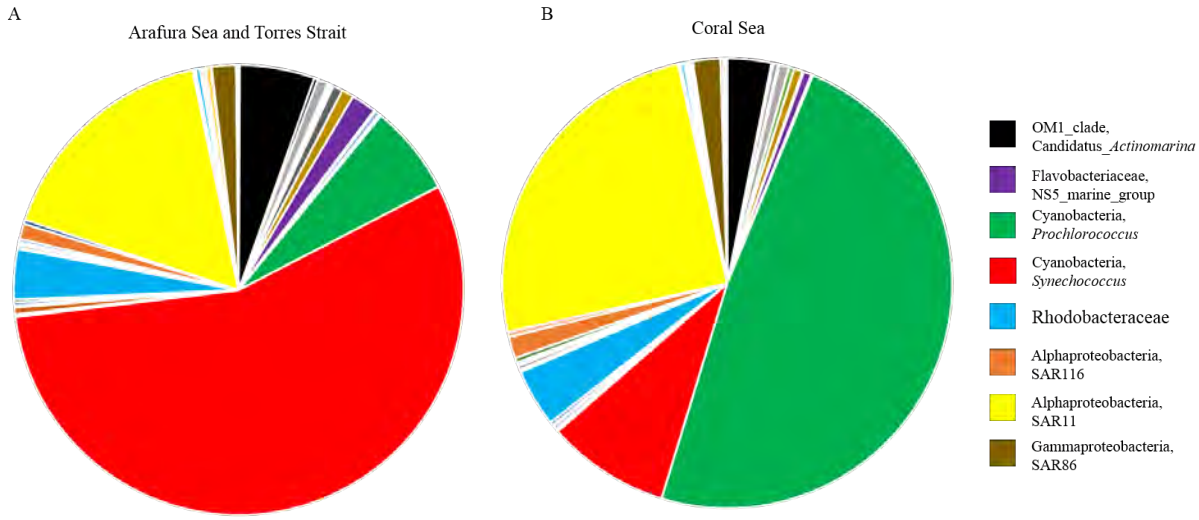

Figure 2.6 Overview of the relative abundance of microbial prokaryote lineages in the surface waters of the Arafura Sea/Torres Strait (A) and the Coral Sea (B).

Overall, the proportion of plastid 16S rRNA gene sequences (belonging to 21 different lineages), provide an indication of the relative abundance of eukaryotic phytoplankton at each site was low (<2%) with maxima up to 7% at some sites (e.g. CTDs 8, 10, 11, and UW34)

(Figure 2.8). Cyanobacteria in both basins accounted for 59% and 53% of total reads, respectively. In line with flow cytometry abundances, *Synechococcus* sequences dominated the Arafura Sea/Torres Strait (up to 88% of reads in some samples) and *Prochlorococcus* dominated the Coral Sea (up to 70% of reads) (Figures 2.6 and 2.8). The percentage of the second largest group, Proteobacteria, was higher in the Coral Sea (41%) than in the Arafura Sea/Torres Strait (31%). Alphaprobacteria were the most abundant Proteobacteria, making up more than 92% of total Proteobacteria sequence reads. The Alphaprobacteria SAR11 clade had a lower relative abundance in Arafura Sea/Torres Strait (21%) than in the Coral Sea (30%). Candidatus *Actinomarina*, within the OM1 clade of the *Actinobacteria*, composed 4.7% of the microbial sequences in the Arafura Sea/Torres Strait and 2.7% in the Coral Sea. Most of the remaining OTUs (abundance < 1%) displayed similar abundances in the two basins.

Beta-diversity patterns of prokaryotic communities

Taxonomic compositional similarity (Bray–Curtis dissimilarities) of 57 surface samples was analyzed by non-matric multidimensional scaling (NMDS) (Figure 2.7). Simprof analysis revealed at least 4 significant clusters, one cluster gathered sites from the Coral Sea, and the other three clustered different samples from the Arafura Sea and Torres Strait. Each cluster displayed a distinct community structure, with the relative abundance of *Synechococcus*, *Prochlorococcus*, SAR11, SAR86, OM1 and *Rhodobacteriaceae* contributing to the dissimilarities between the clusters, as determined by simper (Fig 2.7B). Samples from Coral Sea were tightly clustered, however, the distribution of the remaining clusters was patchy along the Arafura Sea and Torres Strait transect (Figure 2.8). The largest cluster of sites (*n*=20) from this region was characterized by a very high proportion of *Synechococcus* (average 69%) and a decreased proportion of heterotrophic sequences relative to all other sites. The second cluster displayed equivalent proportions of heterotrophs and phototrophs but *Synechococcus* were the dominant phototrophs (average 38%). The third cluster was characterized by equal numbers of *Synechococcus* and *Prochlorococcus* sequences.

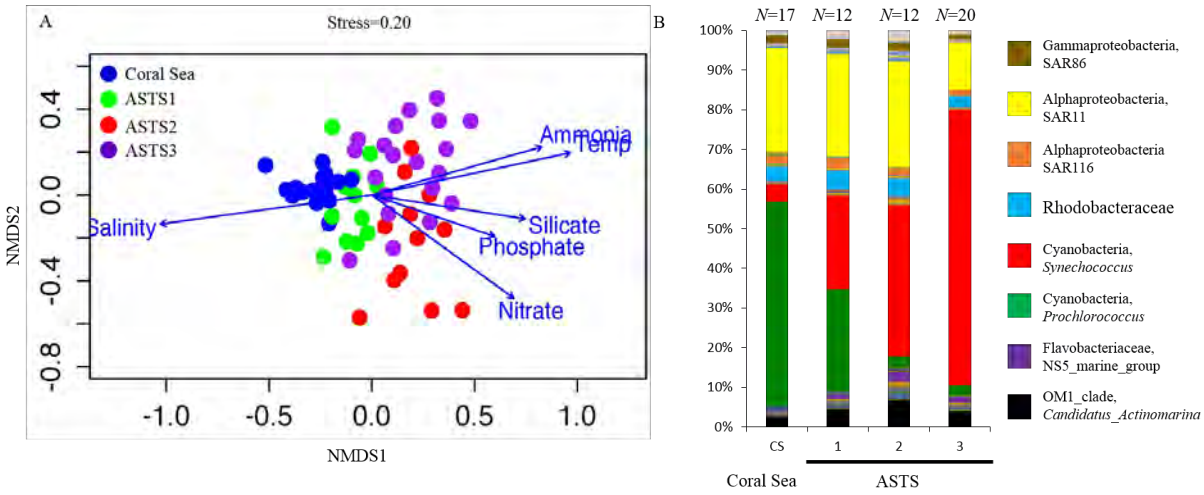

Figure 2.7 Non-metric multidimensional scaling (NMDS) plot (A) based on Bray-Curtis displaying community similarity. Samples are grouped into 4 clusters according to simprof analysis: Coral Sea (CS, blue), Arafura Sea/Torres Strait 1 (ASTS1, green), ASTS2 (red), ASTS3 (purple). Significant environmental variables were correlated with the prokaryote community composition and plotted on NMDS plot. Stacked column graph (B) representing the relative distribution of the dominant groups in the 4 significant clusters. The number of samples in each cluster is given on the top of the columns.

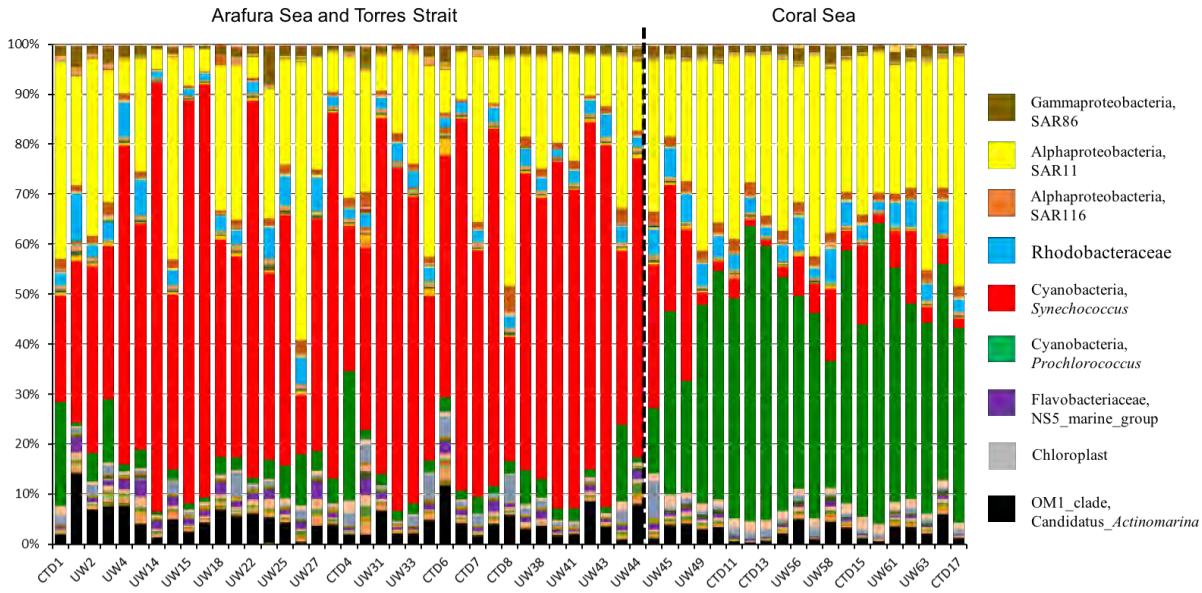

Figure 2.8 Relative abundance of major prokaryotic taxa in each sample across the sampled stations.

Environmental factors drive the prokaryotic community composition

Environmental factors were plotted on the NMDS plot in order to observe the relationship between physico-chemical parameters and community structure (Figure 2.7A). A Pearson correlation was used to test the importance of these environmental variables for prokaryotic

community structure (Table 2.1). The results showed that temperature, salinity and nutrient availabilities were the statistically significant variables, suggesting that they play a role determining the bacterial community composition patterns. Salinity is strongly positively correlated with community structure in the Coral Sea ($P<0.05$). Temperature, nitrate, phosphate and Chl *a* concentrations were strongly positively correlated with the community structure in the Arafura Sea/Torres Strait ($P<0.01$).

Table 2.1 The contribution of individual environmental factors on prokaryote communities in the surface waters of the Arafura Sea/Torres Strait 1 (ASTS1), ASTS2, ASTS3 and the Coral Sea (CS). The significance of each factor was determined independently by Pearson correlation.

Variables	ASTS1		ASTS2		ASTS3		CS		ALL	
	r^2	*p*	r^2	*p*	r^2	*p*	r^2	*p*	r^2	*p*
Temperature	0.218	0.147	0.384	0.052	0.303	0.021*	0.137	0.372	0.565	0.001**
Salinity	0.301	0.249	0.197	0.312	0.247	0.384	0.449	0.026*	0.621	0.001**
Nitrate	0.241	0.001**	0.256	0.065	0.256	0.002**	0.224	0.182	0.674	0.001**
Phosphate	0.202	0.064	0.215	0.034*	0.183	0.041*	0.118	0.437	0.577	0.001**
Silicate	0.278	0.328	0.248	0.029*	0.327	0.015*	0.268	0.093	0.534	0.001**
Ammonia	0.198	0.452	0.265	0.652	0.218	0.328	0.014	0.906	0.204	0.018*
Chl *a*	0.224	0.002**	0.284	0.021*	0.336	0.004**	0.100	0.467	0.286	0.002**

Significate codes: '***' <0.001; '**' 0.001-0.01; '*' 0.01-0.05

In order to understand the factors influencing differences between distinct community clusters in the Arafura Sea/Torres Strait in greater detail we summarized the metadata variables for each cluster (Figure 2.9). The Coral Sea was clearly oligotrophic compared to the other sites, displaying the lowest nutrient concentrations, lowest temperature and the highest salinity. The two Arafura Sea/Torres Strait clusters defined by high *Synechococcus* displayed the highest nutrient concentrations, with the one defined by very high *Synechococcus* sequence abundance also displaying the highest *Synechococcus* flow cytometry counts, although the differences between other variables was not significant. The third Arafura Sea/Torres Strait cluster displayed relatively low nutrient concentrations but intermediate salinity, temperature and phytoplankton counts.

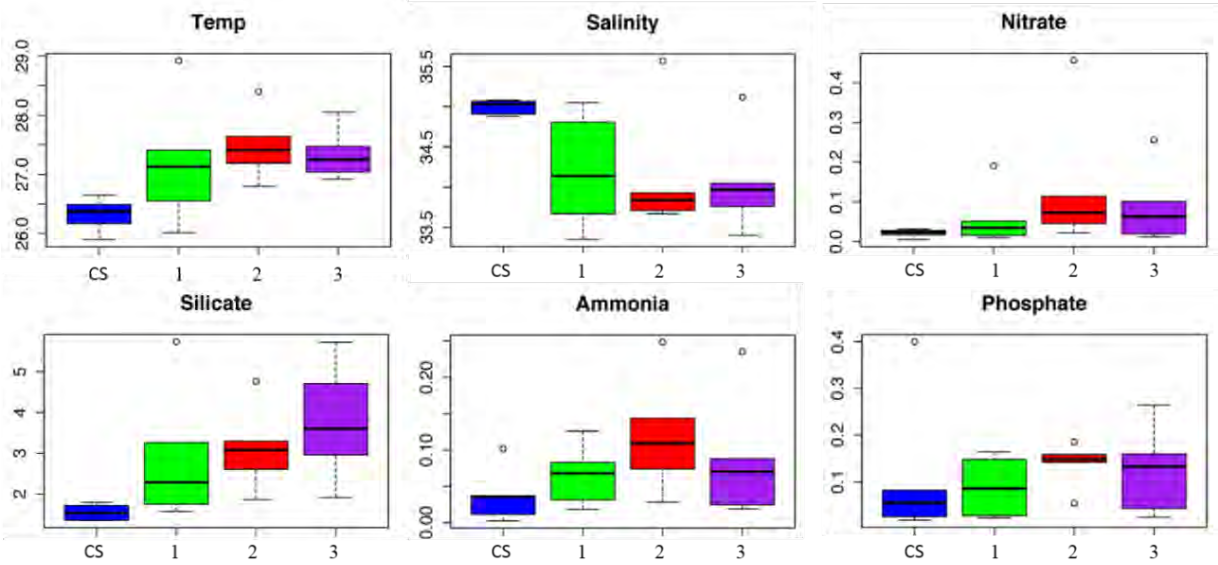

Figure 2.9 The environmental factors influence differences between the four clusters. Samples are classified into 4 clusters according to the simprof analysis: Coral Sea (CS, blue), Arafura Sea/Torres Strait 1 (ASTS1, green), ASTS2 (red), ASTS3 (purple).

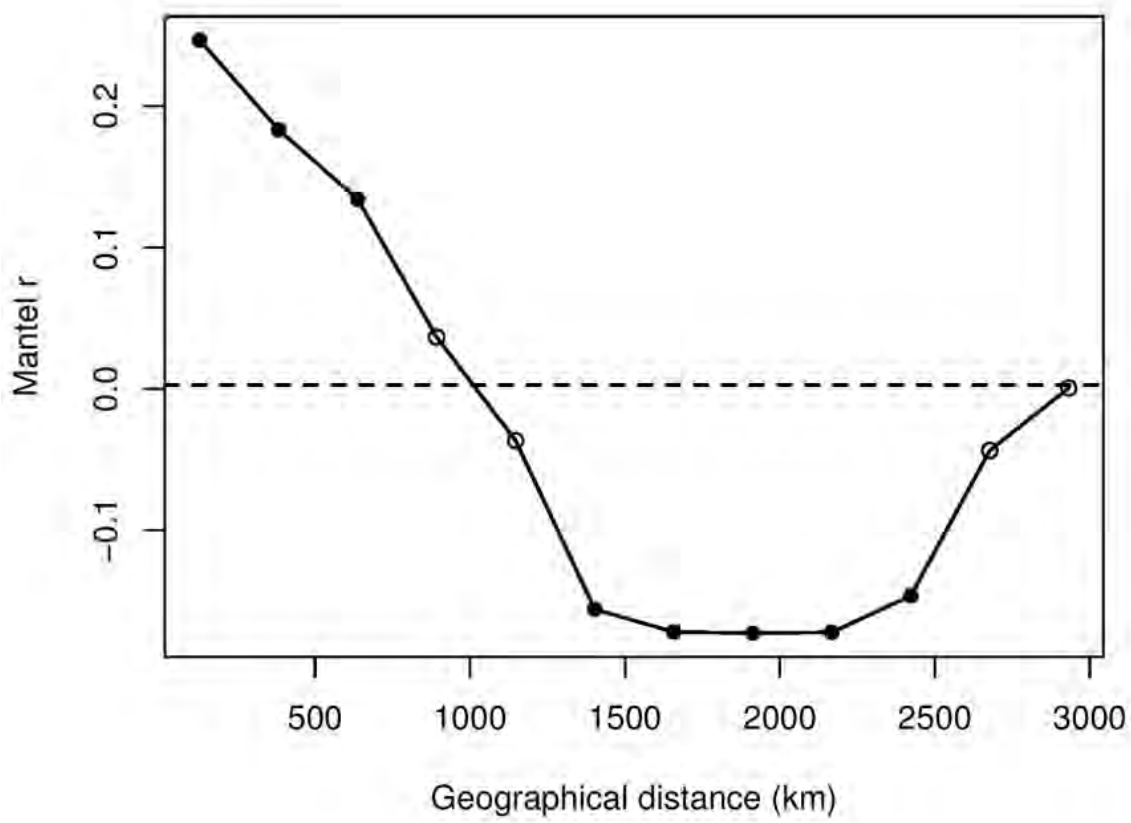

Figure 2.10 Mantel correlogram between community composition (Bray–Curtis dissimilarities) and geographic distance matrices using geographical distance classes set at 500 km. The x-axis is the distance class index and the y-axis is the Mantel *r* (Pearson correlation coefficient) statistic. Filled points represent significant mantel statistical values, and over dashed lines refer to an expected positive correlation but below dashed lines represent a negative correlation. Mantel correlograms were run up to a maximal distance of 3,000 km.

Relation between geographic distance and community composition

We tested the relationship between genetic and geographic distances across space using Mantel correlogram (Figure 2.10). The profile of Mantel correlogram showed a decrease in genetic similarity as geographical distances increased. A significant Mantel correlation was only observed at 600 km or less, showing that there is only significant spatial autocorrelation between space and community composition at this distance and that microbial communities are partially structured by geography according to the basin of origin (P = 0.001 in all cases, r^2 = 0.33).

Identification of enriched taxa based on the LEfSe pipeline

The LEfSe pipeline was used to analyze the abundances of individual bacterial taxa based on LDA effect size. The analysis identified bacterial taxa displaying significantly different abundances between Arafura Sea/Torres Strait and Coral Sea which were then ranked according to LDA scores. A cladogram representing the taxonomic differences between Arafura Sea/Torres Strait and Coral Sea is shown in Figure 2.11. In the Arafura Sea/Torres Strait, there were 12 enriched bacterial groups (Figure 2.12) including *Synechococcus*, Actinobacteria (from phylum to genus), *Flavobacteriia* and *Sphingobacteriia* (from phylum to genus), *Rhodopirellula* (within Planctomycetacia), *Rhizobiales* and *Rhodospirillales* (within Alphaproteobacteria), *Alcaligenaceae* and *Hydrogenophilaceae* (within Betaproteobacteria), *Alteromonadaceae* and *Oceanospirillaceae* and *Pseudomonadaceae* (within Gammaproteobacteria), and MB11C04 marine group (within Verrucomicrobia). In the Coral Sea, the enriched bacterial groups were *Prochlorococcus*, NS9 marine group (within Bacteroidetes), *Rickettsiales* and SAR11 clade (within Alphaproteobacteria), *Bdellovibrionaceae* and Sh765B_TzT_29 (within Deltaproteobacteria), and 6 genea under Gammaproteobacteria, including SAR92, KI89A, SAR86, *Coxiella*, *Marinicella* and *Acinetobacter* (Figure 2.12).

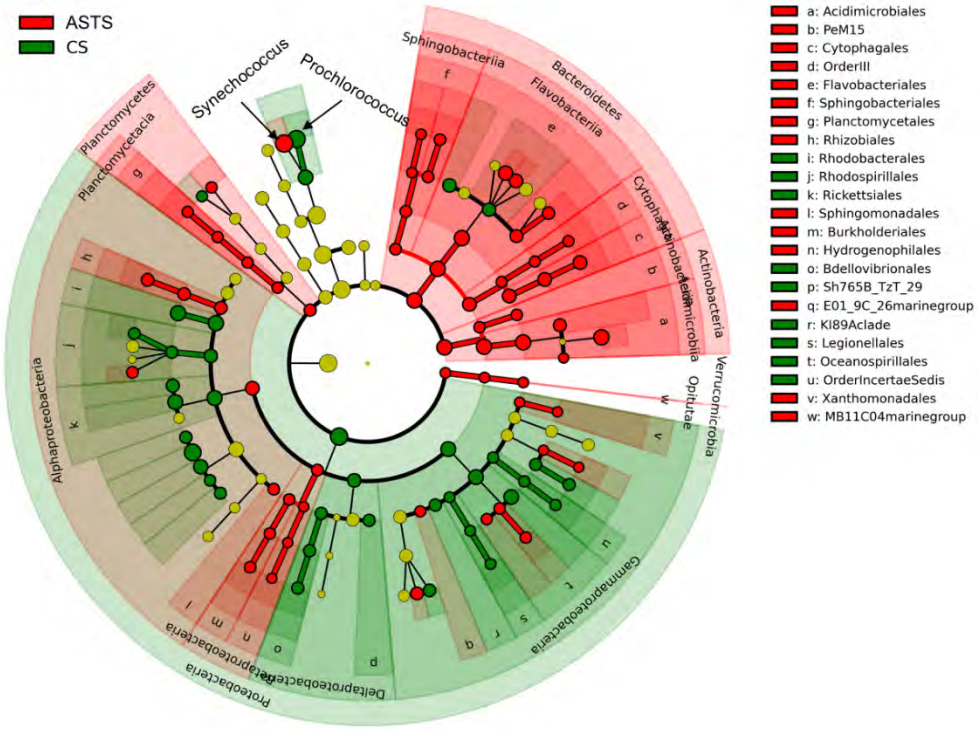

Figure 2.11 LEfSe cladogram indicating the taxonomic distribution differences between sampling sites based on 16S rRNA amplicon sequence data. Lineages with LDA values 2.0 or higher determined by LEfSe are displayed. Red circles and shading indicate lineages enriched in the Arafura Sea and Torres Strait (ASTS); green circles and shading indicate lineages enriched in the Coral Sea (CS). Each shading layer represents one level of taxonomy. The abbreviate labels (a-w) only represent the enriched lineages at taxonomy level of 6. The nodes in yellow indicate taxa that were not significantly differentially represented.

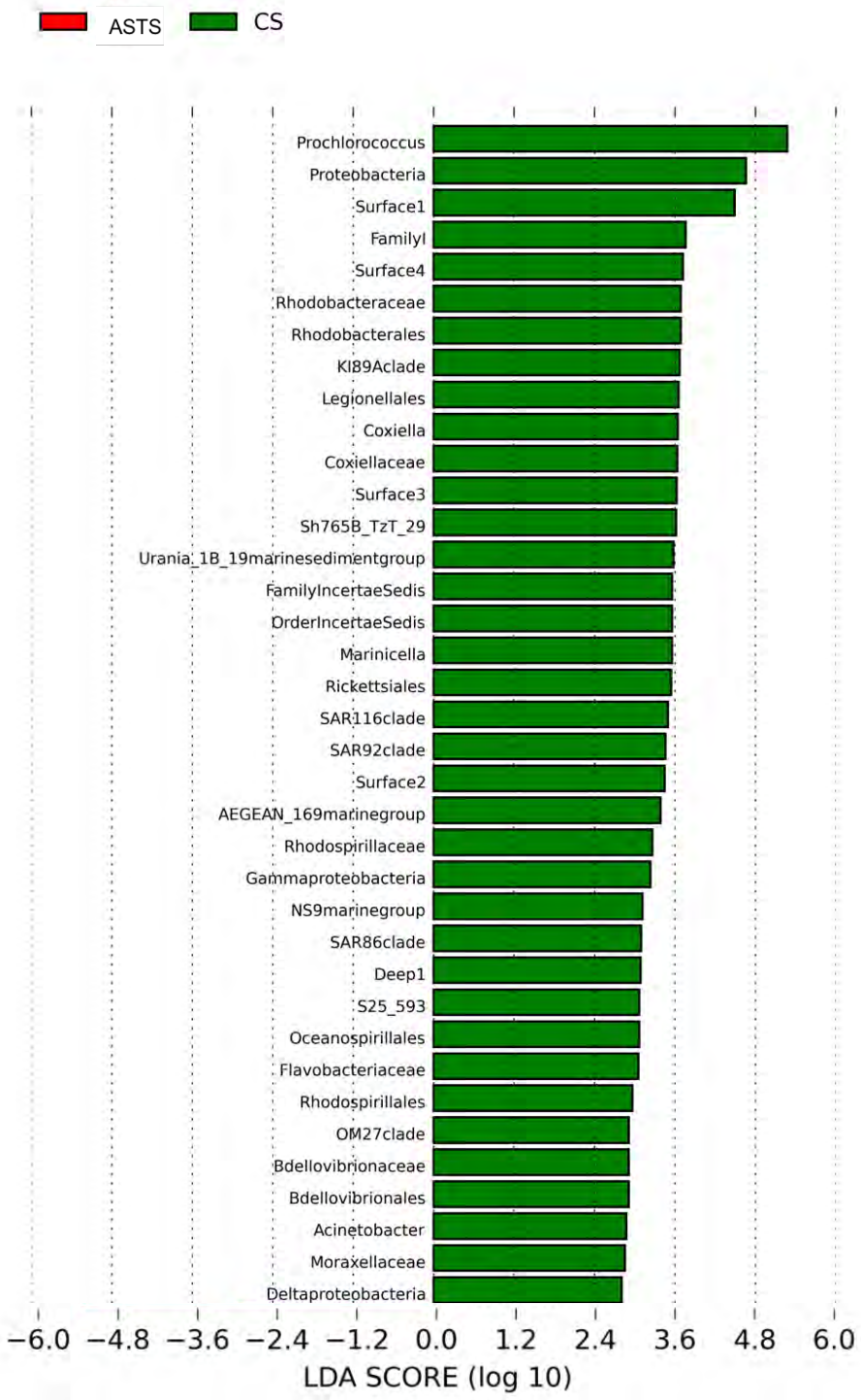

Figure 2.12 Histogram of the LDA scores computed for features differentially abundant between the Arafura Sea and Torres Strait (ASTS, Red) and the Coral Sea (CS, Green).

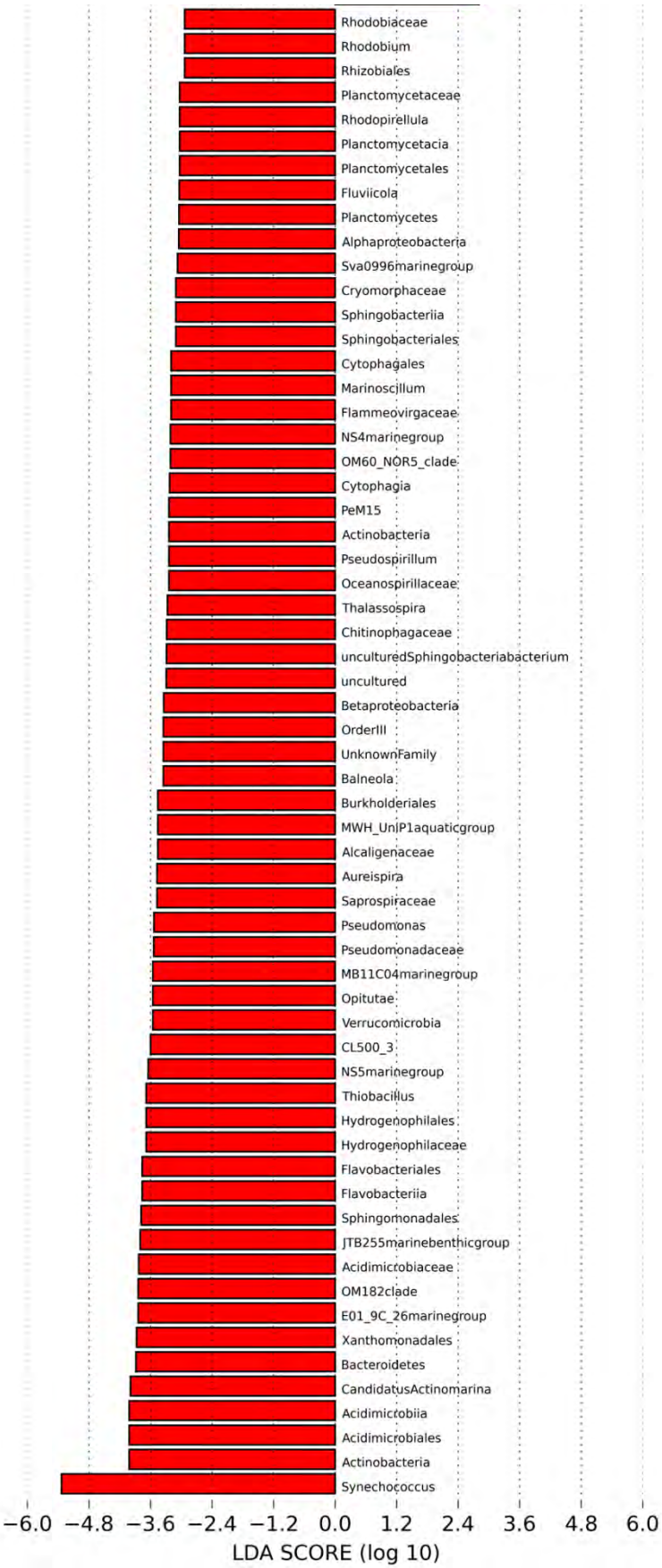

Figure 2.12 To be continued.

Discussion

This high resolution spatial analysis of microbial assemblages spanning tropical Northern Australian waters has revealed the importance of cyanobacterial photosynthesis in these regions. In our study, cyanobacteria accounted for nearly 55% of the sequences retrieved in the surface waters. These photosynthetic phytoplanktons provide the primary food source for zooplanktons, however, together they form the base of the oceanic food chain, which has long been recognized as foundational to fisheries (Worden et al 2015). Hence, the highly productive fisheries and aquaculture in this region may be supported in large by the primary production of a large cyanobacterial population. There was a very distinct distribution pattern of cyanobacteria. *Synechococcus* dominated in the Arafura Sea/Torres Strait region, representing 54.2% of the total prokaryotic community by sequencing and flow cytometry cell counts of up to $3x10^5$ cell mL^{-1}, which is one of the highest levels ever recorded in a marine setting (Flombaum et al 2013, Tarran et al 1999). *Synechococcus* sequence abundance was much lower in the Coral Sea (14%). *Prochlorococcus* displayed an inverse trend, where it was the dominant phototroph in the Coral Sea (38% of sequence abundance) and up to $4.5x10^5$ cells mL^{-1} cell counts.

Differences in nutrient concentration, particularly nitrate, between these water masses are likely key drivers in the abundance of these phototrophs. *Synechococcus* typically has a relatively high abundance in coastal and high dissolved inorganic nitrogen concentration regions (Zwirglmaier et al 2008), and has been shown to grow over a wide range of light intensities and spectral quality (Kana and Glibert 1987) and to utilize a wide variety of nitrogen (N) resources, such as NH_4^+, NO_3^-, NO_2^-, urea, and amino acids (Blanchot et al 1992, Blanchot et al 1997, Tarran et al 1999). Conversely, *Prochlorococcus* is more abundant in offshore and oligotrophic waters (Flombaum et al 2013), generally their presence is associated with regions that have very low dissolved inorganic nitrogen concentrations because they lack the NO_3^- utilization genes (Blanchot and Rodier 1996, Blanchot et al 1997, Campbell et al 1998). This trend can be

observed in our study of the Coral Sea (Figure 2.2B). In a recent assessment of global cyanobacterial cell numbers, the highest average abundance of *Synechococcus* and *Prochlorococcus* was 3.4×10^4 cells mL^{-1}, with 2.5×10^5 cells mL^{-1} described to be present in the India Ocean (Flombaum et al 2013). In our study, the mean abundance of *Synechococcus* declined from 2×10^5 cells mL^{-1} in the Arafura Sea and Torres Strait to only 3×10^4 cells mL^{-1} in the Coral Sea, revealing the Arafura Sea and Torres Strait as a hotspot for *Synechococcus* productivity.

Beyond just the cyanobacterial clades, we identified four assemblage types or clusters from NMDS analysis that were distributed throughout the sampled region (Figure 2.7). While cluster 1 represented a highly similar set of samples, found in the Coral Sea, clusters 2-4 were displayed a patchy distribution in the Arafura Sea and Timor Sea regions. Heterogeneity in the community composition of the Arafura Sea/Torres Strait may be explained by the high microbial standing stock and dynamic oceanographic and weather features in this basin. The Arafura Sea is a semi-enclosed continental shelf basin (50~80 m) where undercurrent priming from the Banda Sea may play an important role in supplying nutrients (Kämpf 2015). Tidal mixing and monsoon storms coupled with rapid biological drawdown may contribute to nutrients, temperature and other factors fluctuating to a higher degree than in the adjacent Coral Sea (Figure 2.2). Indeed, the filamentous diazotroph *Trichodesmium* has been observed in high abundance on numerous occasions (Burford et al 2009, Messer et al 2016, Montoya et al 2004) although peaks in abundance do not always correspond to high *in-situ* rates of nitrogen fixation, an observation that may suggest that *Trichodesmium* populations were established during previous episodes of N-limitation. Nitrogen fixation by the natural population of *Trichodesmium* is highly correlated to the phosphorus levels in the sea and is enhanced at higher irradiance (Sanudo-Wilhelmy SA et al 2001).

We investigated the environmental drivers and geographical distance between sampling locations to understand the bacterial composition in our data set. The strongest environmental driver for the Coral Sea surface samples was salinity ($P<0.01$, explaining 45% of the variance). Salinity has been shown to have a predominant impact on the distribution patterns of microbial species, including benthic and pelagic organisms (Ojaveer et al 2010, Telesh and Khlebovich 2010, Ysebaert et al 2003). It also has been found to be the major determinant of microbial community composition, even exceeding the influence of temperature and pH (Lozupone and Knight 2007, Wu et al 2006). In contrast, the best subgroup of environmental drivers for the Arafura Sea and Torres Strait's surface samples were temperature, nitrate ($P=0.001$) and phosphate concentrations ($P=0.001$).

Although most microbial biogeography studies have demonstrated the contribution of environmental factors and dispersal limitation, historical events also affect species distribution (Leibold et al 2010, Ricklefs 2007). To identify the geographical distance driving the beta-diversity among bacterial communities, we also investigated the relationship between genetic and geographic distances using Mantel Correlograms. The decrease in genetic similarity with geographic distance was found in our study, which is a universal biogeographic pattern observed in communities from all domains of life (Green et al 2004, Martiny et al 2011, Nekola and White 1999, Ramette and Tiedje 2007b). The significant autocorrelation found at the first 600 km in our study suggests that prokaryotes were more homogeneous in taxonomic composition within short distances (Figure 2.10). These relatively short distances, considering the whole sampling distance (~3,500 km), are consistent with the basin-specificity of prokaryotic community composition as described above (Figure 2.7 and 2.8). In fact, the mean and maximal distance between all the samples belonging to the same basin is 37 and 131 km, respectively.

We used the LEfSe tool to detect differentially abundant taxonomic groups within the prokaryote communities in the different sampling basin areas. The shallow and semi-enclosed waters of the Arafura Sea and Torres Strait seem to share similarities in both microbial community composition and higher nutrient concentrations with coastal pelagic environments. The LEfSe results indicated that *Synechococcus* was enriched in the Arafura Sea/Torres Strait and *Prochlorococcus* was enriched in the Coral Sea (Figure 2.11), as discussed above.

The lineages of Actinobacteria, Bacteroidetes (including the classes *Cytophagia*, *Flavobacteriia*, *Sphingobacteriia*) and Planctomycetes were also enriched in the Arafura Sea/Torres Strait (Figure 2.11). Candidatus *Actinomarina* is the most abundant genus within the Actinobacteria in our study. Recent studies revealed that this group is widely distributed in the photic zone of the ocean, both in the tropical and temperate belt, and their abundance ranges from 1 to 10% (Ghai et al 2013), which is similar to our results (Figure 2.8).

Members of the phylum Bacteroidetes (including the classes *Sphingobacteriia*, *Flavobacteriia*, and *Cytophagia*) are an abundant group of marine bacterioplankton, especially in coastal pelagic habitats (Gonzalez et al 2008) where they represent between 10%-30% of the prokaryotes in seawater (Alonso-Saez and Gasol 2007, Cottrell and Kirchman 2000).

Proteobacteria was the second largest group in both basins, but different trends were observed among its classes. Most of the groups in alphaproteobacteria (e.g. SAR11, SAR116) and all groups in deltaproteobacteria (e.g. SAR324) were enriched in the Coral Sea. However, the betaproteobacteria were enriched in the Arafura Sea/Torres Strait, previous work has suggested that there is a strong correlation between organic matter supply and betaproteobacterial growth (Tada et al 2011). Different lineages within the gammaproteobacteria showed distinct distribution patterns, for example K189A were enriched in the Coral Sea, and SAR86 was

enriched in the Arafura Sea/Torres Strait. SAR86 proliferation has been reported with higher concentrations of terrestrial dissolved organic matter (Lindh et al 2015).

This study represents our first view of prokaryotic community composition in surface seawater across the Arafura Sea, Torres Strait and the Coral Sea. The Arafura Sea and Torres Strait appear to be quite distinct from the Coral Sea. The Coral Sea community is more homogenous and is dominated by *Prochlorococcus*, SAR11 and other ecotypes that prefer oligotrophic waters. Microbial populations within the Arafura Sea/Torres Strait show greater variability, but have one of the highest reported concentrations of marine *Synechococcus*. Lineages often associated with coastal pelagic environments appear to have higher abundances within these waters. Nutrient concentrations are significantly higher than in the Coral Sea, but are also highly variable, suggesting that these waters are much more dynamic, probably due to a combination of upwelling events, tropical storms, and terrestrial input.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

We thank the crew of the *RV* Southern Surveyor and Associate Professor Martina Doblin, the Voyage leader for SS2012_t07. This work was supported by ARC DP110102718, ARC DP150102326, ARC FL140100021, CSC scholarship to T. Huang.

Author Contributions:

TH, MO, MB, JS and IP conceived the study, TH, MO and SK conducted experiments and analyzed data, MO, MB and JS collected samples and extracted DNA, TH, MO and IP wrote the manuscript with contributions from all other authors.

References

Alonso-Saez L, Gasol JM (2007). Seasonal variations in the contributions of different bacterial groups to the uptake of low-molecular-weight compounds in northwestern Mediterranean coastal waters. *Applied and environmental microbiology* **73:** 3528-3535.

Andrews JC, Clegg S (1989). Coral Sea circulation and transport deduced from modal information models. *Deep-Sea Res* **36:** 957-974.

Baas-Becking LGM (1934). Geobiologie of inleiding tot de milieukunde. *WP Van Stockum and Zoom (in Dutch): The Hague, the Netherlands*.

Blanchot J, Rodier M, Lebouteiller A (1992). Effect of El-Nino southern oscillation events on the distribution and abundance of phytoplankton in the western Pacific tropical ocean along 165E. *J Plankton Res* **14:** 137-156.

Blanchot J, Rodier M (1996). Picophytoplankton abundance and biomass in the western tropical Pacific Ocean during the 1992 El Nino year: results from flow cytometry. *Deep-Sea Res Pt I* **43:** 877-895.

Blanchot J, Andre JM, Navarette C, Neveux J (1997). Picophytoplankton dynamics in the equatorial Pacific: diel cycling from flow-cytometer observations. *Cr Acad Sci Iii-Vie* **320:** 925-931.

Burford MA, Rothlisberg PC, Revill AT (2009). Sources of nutrients driving production in the Gulf of Carpentaria, Australia: a shallow tropical shelf system. *Mar Freshwater Res* **60:** 1044-1053.

Campbell BJ, Yu LY, Heidelberg JF, Kirchman DL (2011). Activity of abundant and rare bacteria in a coastal ocean. *Proceedings of the National Academy of Sciences of the United States of America* **108:** 12776-12781.

Campbell L, Landry MR, Constantinou J, Nolla HA, Brown SL, Liu H *et al* (1998). Response of microbial community structure to environmental forcing in the Arabian Sea. *Deep-Sea Res Pt Ii* **45:** 2301-2325.

Condie SA, Dunn JR (2006). Seasonal characteristics of the surface mixed layer in the Australasian region: implications for primary production regimes and biogeography. *Mar Freshwater Res* **57:** 569-590.

Copley J (2002). All at sea. *Nature* **415:** 572-574.

Cottrell MT, Kirchman DL (2000). Community composition of marine bacterioplankton determined by 16S rRNA gene clone libraries and fluorescence in situ hybridization. *Applied and environmental microbiology* **66:** 5116-5122.

Ducklow HW, Carlson CA (1992). Oceanic bacterial production. *Advances in Microbial Ecology* **12:** 113-181.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27:** 2194-2200.

Flombaum P, Gallegos JL, Gordillo RA, Rincon J, Zabala LL, Jiao NAZ *et al* (2013). Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and

Synechococcus. *Proceedings of the National Academy of Sciences of the United States of America* **110:** 9824-9829.

Fuhrman JA, Sleeter TD, Carlson CA, Proctor LM (1989). Dominance of bacterial biomass in the Sargasso Sea and its ecological implications. *Mar Ecol Prog Ser* **57:** 207-217.

Gasol JM, del Giorgio PA, Duarte CM (1997). Biomass distribution in marine planktonic communities. *Limnol Oceanogr* **42:** 1353-1363.

Ghai R, Mizuno CM, Picazo A, Camacho A, Rodriguez-Valera F (2013). Metagenomics uncovers a new group of low GC and ultra-small marine Actinobacteria. *Scientific reports* **3:** 2471.

Gonzalez JM, Fernandez-Gomez B, Fernandez-Guerra A, Gomez-Consarnau L, Sanchez O, Coll-Llado M *et al* (2008). Genome analysis of the proteorhodopsin-containing marine bacterium *Polaribacter* sp. MED152 (Flavobacteria). *Proceedings of the National Academy of Sciences of the United States of America* **105:** 8724-8729.

Green J, Bohannan BJM (2006). Spatial scaling of microbial biodiversity. *Trends Ecol Evol* **21:** 501-507.

Green JL, Holmes AJ, Westoby M, Oliver I, Briscoe D, Dangerfield M *et al* (2004). Spatial scaling of microbial eukaryote diversity. *Nature* **432:** 747-750.

Hallegraeff GM, Jeffrey SW (1984). Tropical phytoplankton species and pigments of continental shelf waters of North and North-West Australia. *Mar Ecol Prog Ser* **20:** 59-74.

Halpern BS, Walbridge S, Selkoe KA, Kappel CV, Micheli F, D'Agrosa C *et al* (2008). A global map of human impact on marine ecosystems. *Science* **319:** 948-952.

Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH (2012). Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* **10:** 497-506.

Harris PT (1988). Sediments, bedforms and bedload transport pathways on the continental-shelf adjacent to Torres Strait, Australia Papua New Guinea. *Cont Shelf Res* **8:** 979-1003.

Jongsma D (1974). Marine geology of the Arafura Sea. *Bureau of Mineral Resources, Geology and Geophysics, Canberra* **157:** 1-73.

Kämpf J (2015). Undercurrent-driven upwelling in the northwestern Arafura Sea. *Geophys Res Lett* **42:** 9362-9368.

Kana TM, Glibert PM (1987). Effect of irradiances up to 2000 μE m^{-2} sec^{-1} on marine *Synechococcus* WH7803. I. growth, pigmentation, and cell composition. *Deep-Sea Res* **34**: 479-495.

Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microb* **79:** 5112-5120.

Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR (1985). Rapid-determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proceedings of the National Academy of Sciences of the United States of America* **82:** 6955-6959.

Leibold MA, Economo EP, Peres-Neto P (2010). Metacommunity phylogenetics: separating the roles of environmental filters and historical biogeography. *Ecol Lett* **13:** 1290-1299.

Lindh MV, Lefebure R, Degerman R, Lundin D, Andersson A, Pinhassi J (2015). Consequences of increased terrestrial dissolved organic matter and temperature on bacterioplankton community composition during a Baltic Sea mesocosm experiment. *Ambio* **44:** S402-S412.

Lozupone CA, Knight R (2007). Global patterns in bacterial diversity. *Proceedings of the National Academy of Sciences of the United States of America* **104:** 11436-11440.

Magoc T, Salzberg SL (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27:** 2957-2963.

Marie D, Partensky F, Jacquet S, Vaulot D (1997). Enumeration and cell cycle analysis of natural populations of marine picoplankton by flow cytometry using the nucleic acid stain SYBR Green I. *Appl Environ Microb* **63:** 186-193.

Martiny JB, Eisen JA, Penn K, Allison SD, Horner-Devine MC (2011). Drivers of bacterial beta-diversity depend on spatial scale. *Proceedings of the National Academy of Sciences of the United States of America* **108:** 7850-7854.

Messer LF, Mahaffey C, Robinson CM, Jeffries TC, Baker KG, Isaksson JB *et al* (2016). High levels of heterogeneity in diazotroph diversity and activity within a putative hotspot for marine nitrogen fixation. *ISME Journal* **10:** 1499-1513.

Monier A, Comte J, Babin M, Forest A, Matsuoka A, Lovejoy C (2015). Oceanographic structure drives the assembly processes of microbial eukaryotic communities. *ISME Journal* **9:** 990-1002.

Montoya JP, Holl CM, Zehr JP, Hansen A, Villareal TA, Capone DG (2004). High rates of N_2 fixation by unicellular diazotrophs in the oligotrophic Pacific Ocean. *Nature* **430:** 1027-1031.

Nekola JC, White PS (1999). The distance decay of similarity in biogeography and ecology. *J Biogeogr* **26:** 867-878.

Nemergut DR, Costello EK, Hamady M, Lozupone C, Jiang L, Schmidt SK *et al* (2011). Global patterns in the biogeography of bacterial taxa. *Environ Microbiol* **13:** 135-144.

O'Malley MA (2008). 'Everything is everywhere: but the environment selects': ubiquitous distribution and ecological determinism in microbial biogeography. *Studies in history and philosophy of biological and biomedical sciences* **39:** 314-325.

Ojaveer H, Jaanus A, MacKenzie BR, Martin G, Olenin S, Radziejewska T *et al* (2010). Status of biodiversity in the Baltic Sea. *Plos One* **5**.

Papke RT, Ramsing NB, Bateson MM, Ward DM (2003). Geographical isolation in hot spring cyanobacteria. *Environ Microbiol* **5:** 650-659.

Pomeroy LR, Williams PJI, Azam F, Hobbie JE (2007). The microbial loop. *Oceanography* **20:** 28-33.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P *et al* (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41:** D590-D596.

Ramette A, Tiedje JM (2007a). Biogeography: an emerging cornerstone for understanding prokaryotic diversity, ecology, and evolution. *Microb Ecol* **53:** 197-207.

Ramette A, Tiedje JM (2007b). Multiscale responses of microbial life to spatial distance and environmental heterogeneity in a patchy ecosystem. *Proceedings of the National Academy of Sciences of the United States of America* **104:** 2761-2766.

Ricklefs RE (2007). History and diversity: explorations at the intersection of ecology and evolution. *Am Nat* **170:** S56-S70.

Salazar G, Cornejo-Castillo FM, Benitez-Barrios V, Fraile-Nuez E, Alvarez-Salgado XA, Duarte CM *et al* (2016). Global diversity and biogeography of deep-sea pelagic prokaryotes. *ISME Journal* **10:** 596-608.

Sanudo-Wilhelmy SA, Kustka AB, Gobler CJ, Hutchins DA, Yang M, Lwiza K et al (2001). Phosphorus limitation of nitrogen fixation by *Trichodesmium* in the central Atlantic Ocean. *Nature* **411:** 66-69.

Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS *et al* (2011). Metagenomic biomarker discovery and explanation. *Genome biology* **12:** R60.

Sherman K, Hempel, G. (2008). The UNEP large marine ecosystem report: a perspective on changing conditions in LMEs of the world's regional seas. *UNEP Regional Seas Report and Studies No 182 United Nations Environment Programme Nairobi, Kenya*

Tada Y, Taniguchi A, Nagao I, Miki T, Uematsu M, Tsuda A *et al* (2011). Differing growth responses of major phylogenetic groups of marine bacteria to natural phytoplankton blooms in the western North Pacific Ocean. *Appl Environ Microb* **77:** 4055-4065.

Tarran GA, Burkill PH, Edwards ES, Woodward EMS (1999). Phytoplankton community structure in the Arabian Sea during and after the SW monsoon, 1994. *Deep-Sea Res Pt Ii* **46:** 655-676.

Telesh IV, Khlebovich VV (2010). Principal processes within the estuarine salinity gradient: a review. *Mar Pollut Bull* **61:** 149-155.

Tomczak MJSG (2003). Regional oceanography: an introduction. *Pergamon Press*.

Winsley T, van Dorst JM, Brown MV, Ferrari BC (2012). Capturing greater 16S rRNA gene sequence diversity within the domain bacteria. *Appl Environ Microb* **78:** 5938-5941.

Worden AZ, Follows MJ, Giovannoni SJ, Wilken S, Zimmerman AE, Keeling PJ (2015). Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science* **347**.

Wu QL, Zwart G, Schauer M, Kamst-van Agterveld MP, Hahn MW (2006). Bacterioplankton community composition along a salinity gradient of sixteen high-mountain lakes located on the Tibetan Plateau, China. *Appl Environ Microb* **72:** 5478-5485.

Ysebaert T, Herman PMJ, Meire P, Craeymeersch J, Verbeek H, Heip CHR (2003). Large-scale spatial patterns in estuaries: estuarine macrobenthic communities in the Schelde estuary, NW Europe. *Estuar Coast Shelf S* **57:** 335-355.

Zhang Y, Zhao ZH, Dai MH, Jiao NZ, Herndl GJ (2014). Drivers shaping the diversity and biogeography of total and active bacterial communities in the South China Sea. *Mol Ecol* **23:** 2260-2274.

Zwirglmaier K, Jardillier L, Ostrowski M, Mazard S, Garczarek L, Vaulot D *et al* (2008). Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environ Microbiol* **10:** 147-161.

Chapter 3:

Investigating the microbial eukaryotic communities in the surface waters of the Arafura Sea and Coral Sea

Investigating the microbial eukaryotic communities in the surface waters of the Arafura Sea and Coral Sea

Taotao Huang[1], Martin Ostrowski[1], Hasinika K.A.H. Gamage[1], Mark V. Brown[2], Lauren Messer[3,4], Justin Seymour[3], Ian T. Paulsen[1]

1. Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, Australia
2. School of Biotechnology and Biomolecular Science, University of New South Wales, Sydney, Australia
3. Climate Change Cluster, University of Technology Sydney, Sydney, Australia
4. Current address: Australian Centre for Ecogenomics, University of Queensland, Brisbane, Australia

Corresponding Author: Ian Paulsen, Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, New South Wales, 2109, Australia.

Email: ian.paulsen@mq.edu.au

Abstract

We investigated the diversity of microbial eukaryotes in the surface waters of the Arafura Sea, Torres Strait and Coral Sea, using flow cytometry and high-throughput sequencing of the V9 region of the 18S rRNA gene. A total of 8.6 million sequence reads were clustered into 6,129 Operational Taxonomic Units (OTUs). Rarefaction curves suggested that the sequence data well represented the eukaryote diversity in these areas. The eukaryotic assemblages appeared to be dominated by the SAR supergroups (Stramenopiles, Alveolata and Rhizaria), Metazoa and Archaeplastida both in Arafura Sea/Torres Strait and Coral Sea, together representing ~90% and 68% of all reads and OTUs, respectively. Organismal abundance showed distinct biogeographic patterns, with the Bilateria and diatoms having much higher abundances in the Arafura Sea/Torres Strait compared with the Coral Sea. In contrast, the Syndiniales and Cnidaria showed the opposite with higher abundances in the Coral Sea. Analysis of the environmental metadata indicated that salinity, temperature, phosphate, and silicate concentrations were the strongest drivers influencing eukaryotic community composition. This study provides the first snapshot of microbial marine eukaryote diversity and abundances in tropical surface waters in Northern Australia.

Introduction

Marine microbes play an essential role in natural systems. They are present in large abundances and are central in biogeochemical processes (Massana and Pedros-Alio 2008). They are responsible for almost half of global primary production, mostly by planktonic microorganisms that account for only 0.2% of global primary producer biomass (Field et al 1998). Characterizing microbes in the environment has proven difficult due to their small size, but the development of molecular ecology approaches and high-throughput sequencing has unveiled a large amount of novel diverse assemblages in marine bacterial, archaea and picoeukayotes (Delong 1992, Giovannoni et al 1990, Yooseph et al 2010). Marine eukaryotes are now known to be ubiquitous and diverse in surface oceans (Jürgens 2008), including algae, fungi, protozoa and small metazoa which are important components of marine food webs (Anjusha et al 2013). The photosynthetic eukaryotes account for a significant fraction of primary production in the ocean, especially in oligotrophic conditions.

Marine eukaryotes display diverse lifestyles in marine ecosystems, from autotrophy (photosynthesizers) to heterotrophy (predators, decomposers, parasites), or mixotrophy (Zubkov and Tarran 2008). Their biodiversity contributes to marine ecosystem stability, resilience and function (Loreau et al 2001, Wardle et al 2004). The understanding of marine microbial biodiversity and biogeography patterns has been improved enormously in the past decade, either in regional or global scales, surface or deep-sea or sediments (Green et al 2004, Kirkham et al 2013, Martiny et al 2006, Monier et al 2015, Salazar et al 2015, Salazar et al 2016). This study focuses on the biogeography of eukaryotic microbes in a region of Australian tropical waters extending from the Arafura Sea to the Coral Sea.

The Arafura Sea is a semi-enclosed continental and shallow (50-80 m) shelf basin between Australia and Indonesian New Guinea. It has been identified as one of the most pristine marine environments on the planet (Condie 2011) and classified as a large marine ecosystem of the

highest productivity > 300 g C m^2 yr^{-1} (Sherman 2008). It is subject to seasonal upwelling induced seafloor currents and typical hemipelagic carbonate rich sedimentation during the southeast monsoon (June to November) (Jongsma 1974, Kämpf 2015). The Torres Strait is a shallow (7-10 m) water body (Harris 1988) connecting the Arafura Sea to the continental shelf of the Great Barrier Reef since it was flooded by the rising sea about 8,000 years ago. It is the site at which tropical waters are exchanged between the Pacific and Indian Oceans, in a process called the Indonesian Throughflow (Saint-Cast 2006, Tomczak 2003). During winter (April to November), there are strong prevailing southeast monsoon winds, helping the westward flows from the Coral Sea through to the Torres Strait and Arafura Sea. The Fly River in Papua New Guinea empties a large amount of terrestrial input into the Gulf of Papua (next to Torres Strait and Coral Sea) annually due to rainfalls in the highlands, which influence the local marine environment and hydrology (Saint-Cast 2006). Torres Strait is culturally, ecologically and economically significant, it supports important artisanal and commercial fisheries, and is also an important international shipping lane (Wolanski et al 2013).

The Coral Sea is a marginal sea off the northeast coast of Australia in the South Pacific. In this region, the upper 100 m is primarily warm, with fresh tropical surface waters overlying higher salinity subtropical lower waters (Condie 2003). The circulation patterns of these waters are driven by the westward flow of the south equatorial current. This current splits into two branches on meeting the continental shelf edge; one flowing north along the edge of the Great Barrier Reef from Torres Strait to Arafura Sea, and the other flowing south and contributing to the East Australian Current (Andrews and Clegg 1989, Lyne V 2005). Nutrient concentrations in the surface water in this region are typically very low. The community composition and biogeography of microbial eukaryotes in these areas have not been investigated systematically using molecular techniques.

Various approaches have been used to investigate the marine microbial eukaryotic biodiversity of other regions, including fluorescence *in situ* hybridization staining (Morgan-Smith et al 2013), 18S rRNA amplicon sequencing (Thomas et al 2012), and metagenomics (de Vargas et al 2015). Sequencing-based approaches have unveiled new groups of phagotrophs (Massana et al 2004), parasites (Guillou et al 2008), and phototrophs (Liu et al 2009). Most marine microbial eukaryote studies focused on water column, deep-sea subsurface (Lopez-Garcia et al 2001, Pernice et al 2016) and surface sediments (Salani et al 2012) or in hydrothermal vents (Morgan-Smith et al 2013) at a regional scale or global scale (de Vargas et al 2015, Kirkham et al 2013). In our study, we used high-throughput sequencing of the V9 region of 18S rRNA amplicons to investigate taxonomic composition of microbial eukaryote communities in surface seawater samples which were collected from the Arafura Sea/Torres Strait and Coral Sea during the *RV* Southern expedition in October 2012.

Materials and Methods

Sampling of microbial eukaryotes and oceanographic data analysis

Surface seawater samples were collected from Darwin to Cairns across the Arafura Sea, Torres Strait and Coral Sea during the on-board *RV* Southern Surveyor expedition in October 2012 (Figure 3.1). A total of 61 water samples were collected from the surface underway seawater supply (5m depth) or using 10 L Niskin bottles during 17 CTD casts with associated a rosette sampling system equipped with a Seabird SBE-911+ CTD probe, including temperature, salinity and oxygen sensors. For each sample, 4 mL of seawater was fixed with 1% paraformaldehyde (final concentration) (Marie et al 1997) for abundance analysis. For eukaryotic plankton community structure analysis, 8L of seawater was filtered through 0.22 µm pore-size white polycarbonate filters. Filters were frozen in liquid nitrogen and then stored at -80°C until DNA extraction. The concentrations of inorganic nutrients were determined from depth profile samples collected from CTD stations and processed according to standardized protocols. Instrument details and data are available through the Australian Ocean Data Network

(https://researchdata.ands.org.au/southern-surveyor-voyage-ss2012t07-hydrology/436463?source =suggested_datasets).

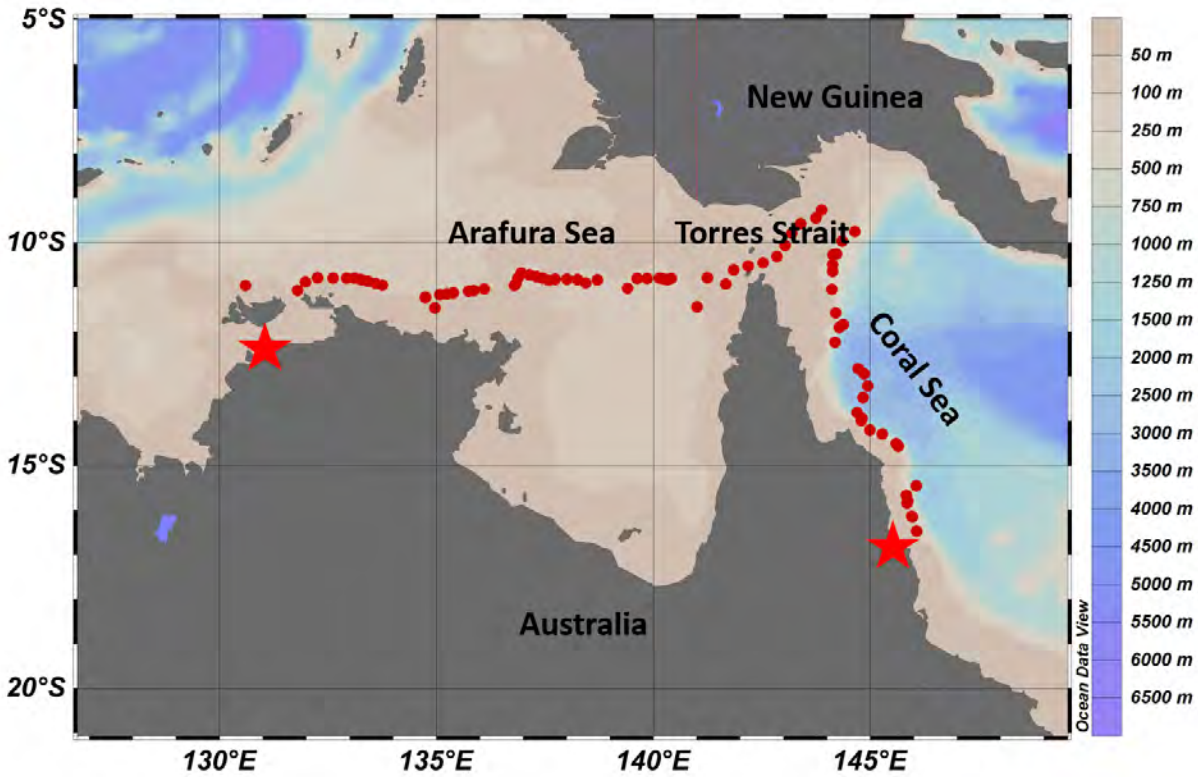

Figure 3.1 Schematic representation of the sampling sites (red dots) from Darwin (left star) to Cairns (right star) in October 2012. Colours bars on the right indicate values for the depths. Reprinted from Chapter 2.

Cell abundance measurements

Preserved samples were quickly thawed at room temperature and transferred into flow tubes. Then, the non-stained fixed samples were processed with BD INFLUX flow cytometry (Becton Dickinson, San Jose, CA, USA) with a blue laser emitting at 488 nm using the trigger setting on red fluorescence from chlorophyll. Pico-eukaryotes were discriminated from *Synechococcus* by the lack of the orange fluorescence and distinguished from *Prochlorococcus* based on their larger cell size. The beads were added into samples and used as an internal standard.

DNA extraction and sequencing of 18S V9 rRNA genes

DNA was extracted using the MoBio Power water kit (MoBio Laboratories, Carlsbad, CA, USA), according to the manufacturer's instructions. Eukaryote diversity was assessed by

amplicon sequencing of the V9 region of 18S rRNA gene with an Illumina Miseq platform using paired end reads (2×150 bp). PCR amplifications were performed with Taq polymerase and primer set 1380F (5'-TTGTACACACCGCCC-3') and 1510R (5'-CCTTCYGCAGGTTC ACCTAC-3') (Amaral-Zettler et al 2009). The barcodes and specific primers for Illumina sequencing using the Nextera Index Kit (Illumina) were added to the 5' ends of primer pair. The 50 µl PCR mixtures contained 0.5 mM each primer, 2.5 mM $MgCl_2$, 1× Buffer (Promega), 0.2 mM dNTP, 2.5 U Taq DNA polymerase. The process consisted of an initial denaturation at 94°C for 3 min before 30 cycles of denaturation at 94 °C for 30 s, annealing at 57 °C for 60 s, extension at 72 °C for 90 s and final extension at 72 °C for 10 min. The reaction product was then stored at 4 °C. PCR products were run on a 1.0% agarose gel to check amplicon lengths and were quantified by Quant-IT PicoGreen assay (ThermoFisher). The amplicons were pooled and purified using AMPure XP beads (Invitrogen) and quantified by Nanodrop ND-2000 (ThermoFisher) prior to sequencing at the Ramaciotti Centre for Genomics.

Processing 18S rRNA sequencing reads

The raw pair end reads were joined by FLASH (Magoc and Salzberg 2011). Fastx software was used to filter merged reads shorter than 70 bp in length and quality scores less than 20. The chimera sequences were checked and removed with the chimera search module of the USEARCH 64bit program v8.1 (Edgar et al 2011). The remaining reads were then clustered into Operational Taxonomy Units (OTUs) using a 97% sequence identify cutoff followed by taxonomic assignment against the latest PR2 database (Guillou et al 2013) using Mothur classify.seqs with default parameters (Kozich et al 2013). The OTU table was filtered to eliminate of spurious OTUs by discarding those OTUs with <0.005% of the total number of sequences (Bokulich et al 2013). Hereafter, the filtered OTU table was used for downstream analysis.

Statistical data analysis

Statistical analysis was carried out with R software (version 3.2.1) using the Vegan package version 2.3-0 (Oksanen et al 2015). The filtered OTU-abundance table obtained from the sequence clustering was sampled down to the smallest sample size (21,558 reads per sample) using rrarefy and dissimilarities between all samples were calculated using a Bray-Curtis dissimilarity matrix for further beta-diversity. We studied the rarefaction curves and estimated taxonomic diversity (Shannon index) at the local (individual sample) and regional (Arafura Sea/Torres Strait and Coral Sea) scales using the rarefied OTU table. The dissimilarity matrix was used to perform a Non-metric Multidimensional Scaling (NMDS) plot (Minchin 1987) to visualize the community composition similarity of samples from Arafura Sea/Torres Strait and Coral Sea. Canonical Correspondence Analysis (CCA) was performed with envfit approach to reveal which local environmental variables correlate with changes in eukaryote community structure (Palmer 1993). BIOENV analysis was used to find the best subset of environmental drivers based on Spearman's rank correlation method (Clarke and Ainsworth 1993). Mantel test was used to test the significance of BIOENV results and the importance of individual environmental variables on community biodiversity.

Results

Oceanographic context and Phototrophic picoeukaryotes (PPEs) cell abundance

The Arafura Sea/Torres Strait and Coral Sea have different environmental parameters and nutrient variables due to their different geographic locations and hydrology (Figure 2.2, pg74). The Arafura Sea and Torres Strait are a shallow and semi-closed basin, however, the Coral Sea is an open sea with deep-water. Generally, Arafura Sea and Torres Strait waters had a higher surface temperature than Coral Sea waters in October 2012, with average temperatures of 27.4 °C and 26.3 °C, respectively. Salinity showed an opposite trend. In the Arafura Sea, the salinity increased slightly from 33.8 PSU to 33.9 PSU, however, we recorded the highest (35.6 PSU) and lowest (33.4 PSU) salinity in Torres Strait. In the Coral Sea, salinity remained relatively

stable around 35 PSU in all sampling stations. Nutrient concentrations were generally higher and showed greater fluctuation in the Arafura Sea and Torres Strait compared with the Coral Sea. In the Arafura Sea and Torres Strait, the average concentration of inorganic N, PO_4^{3-}, NH_4^+ and SiO_4^{4-} were 5, 2.6, 2.3 and 2.3-fold higher than in Coral Sea, respectively. It is interesting to observe that the highest concentration of nitrate, phosphate, ammonia and silicate occurs at the junction of Torres Strait and the Arafura Sea.

We determined the cell abundance of plastidic protists by flow cytometry using the fluorescence of natural photosynthetic pigments (Figure 2.3, pg76). The average plastidic protist cell density in surface water was highest in the Torres Strait (1.8~4.2×10^3 cells mL^{-1}), lowest in the Coral Sea (554~1,465 cells mL^{-1}), and intermediate in the Arafura Sea (1~3.6×10^3 cells mL^{-1}). The cell densities of the picocyanobacteria *Synechococcus* and *Prochlorococcus* were reported in Chapter 2, and were generally higher by 1-2 orders of magnitude.

Alpha diversity of microbial eukaryotic communities

We examined microbial eukaryote diversity from 47 of the 61 samples collected across the Arafura Sea, Torres Strait and Coral Sea using sequencing of the V9 region of the 18S rRNA gene. A total of ~ 9.6 million raw sequences were generated, and ~ 8.6 million sequences were obtained after quality control and chimera removal that clustered into 6,129 operational taxonomic units (OTUs) at 97% similarity. Nearly 20% of the OTUs (1,246) in our dataset did not match any References barcode in the database, however, they represented only 4.5% of the total reads. The remaining assignable OTUs (4,883) were classified into 1,014 deep-branching lineages amongst the seven recognized supergroups of eukaryotes. Rarefaction analysis indicated that all 18S rRNA amplicon libraries provided good representation of the eukaryotic microbial communities at both local (individual sample) and global (all samples) scales (Figure 3.2). The species accumulation curve was plotted by randomly accumulating an increasing number of sites sampled with the cumulative number of species recorded (Figure 3.3A). The

results showed that the curve rose rapidly at the first 10 samples and flatted into a plateau at ~ 6,000 OTUs once ~ 25 samples were considered. This indicated that the species abundance distribution is more even in our samples. The global pool of OTU abundances showed a good fit to the truncated Preston Lognormal distribution (Preston 1948), and a total eukaryotic plankton richness of ~ 6,155 OTUs was extrapolated, of which only 26 OTUs were not found in our study (Figure 3.3B). Eukaryotic taxonomic diversity in each sample was estimated by Shannon index, ranged from 1.83 to 4.83 with a mean value of 4.16 (Figure 3.2). The alpha diversity of samples did not show significant differences between Arafura Sea/Torres Strait and Coral Sea, average 4.13 and 4.19, respectively.

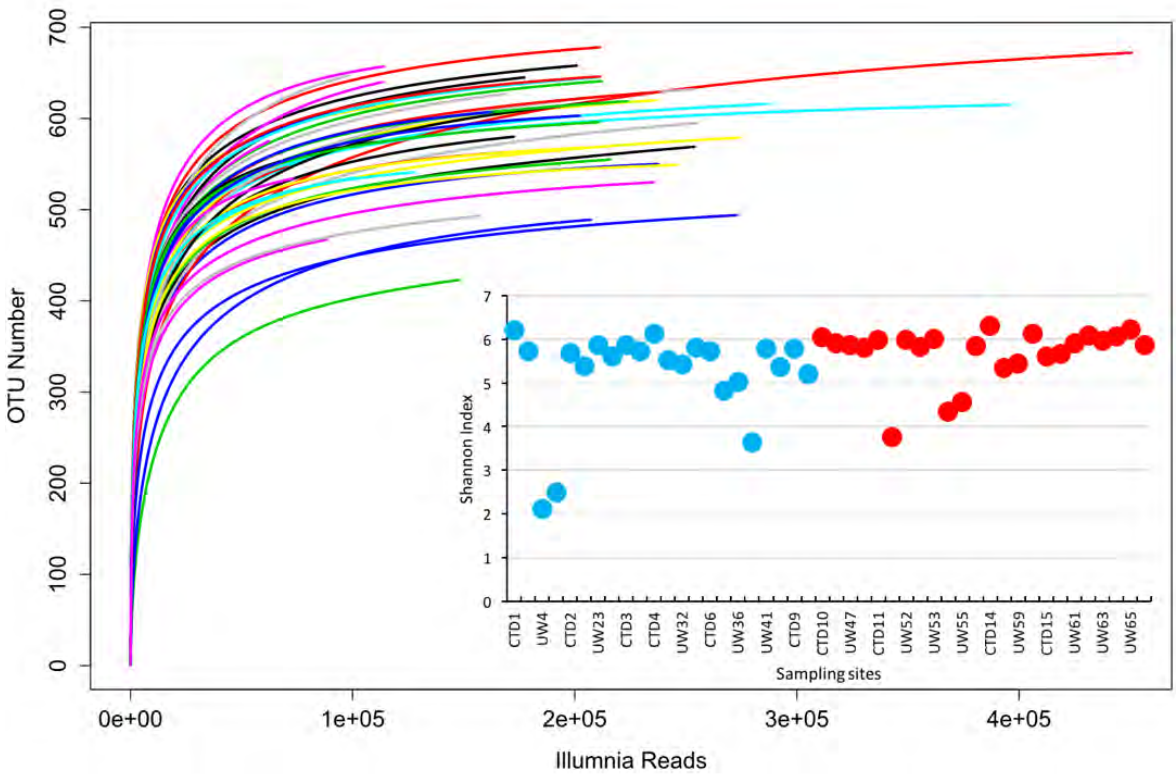

Figure 3.2 Rarefaction curves and OTUs diversity (Shannon Index, inset) for each sample. Blue colour represents samples from the Arafura Sea/Torres Strait and red colour represents samples from the Coral Sea.

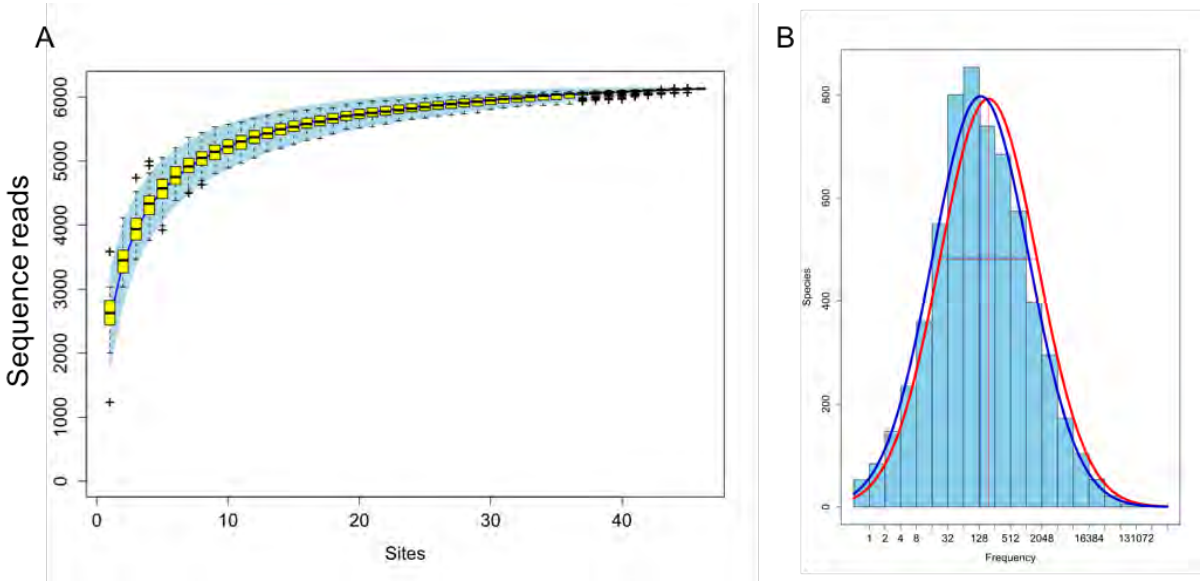

Figure 3.3 (A) Species accumulation curves indicated that species abundance distribution in our dataset. (B) Abundance distribution of eukaryotic species with expected normal curve. Quasi-Poisson fit to octaves (red curve) and maximized likelihood to $\log_2$ abundances (blue curve) approximations were used to fit the OTU abundance distribution to the Preston Lognormal model.

Taxonomic composition of microbial eukaryotes

Taxonomic distribution was analyzed at local (individual), regional (Arafura Sea/Torres Strait and Coral Sea) and global (all samples) scales. The proportions of sequence reads represented the relative abundance of every phylum or genus. Analysis of all samples together at superphylum level showed that the Alveolata (44.8%) and Metazoa (28.9%), Stramenopiles (10.2%), Rhizaria (4.0%) and Archaeplastida (1.7%) had the highest 18S rRNA gene abundances (Figure 3.4A). Unassigned reads represented 4.5% of total reads. Despite this, the analysis of OTU richness showed very different patterns (Figure 3.4B). Alveolata represented 47.0% of OTUs, followed by Unassigned OTUs (25.2%) and Stramenopiles (9.6%), whereas Metazoa only had 3.5%. In terms of 18S rRNA gene abundance at a regional scale (Figure 3.4C), majority of the eukaryotes belonged to the Alveolata in both basins, with 39.3% in Arafura Sea/Torres Strait and nearly 50% in Coral Sea. The Metazoa accounted for 35.0% of all samples in Arafura Sea/Torres Strait, higher than was seen in the Coral Sea (23.3%). The Stramenopiles and Archaeplastida, showed a similar trend with rRNA gene abundances of 11.5% and 2.0% in the Arafura Sea/Torres Strait and 8.9% and 1.2% in the Coral Sea, respectively. Haptophyta represented 1.7% of the eukaryote community in the Arafura

Sea/Torres Strait samples, but 2.9% in the Coral Sea. Fungi (within Opisthokonta) comprised only 0.2% of total reads in both the Arafura Sea/Torres Strait and Coral Sea basin.

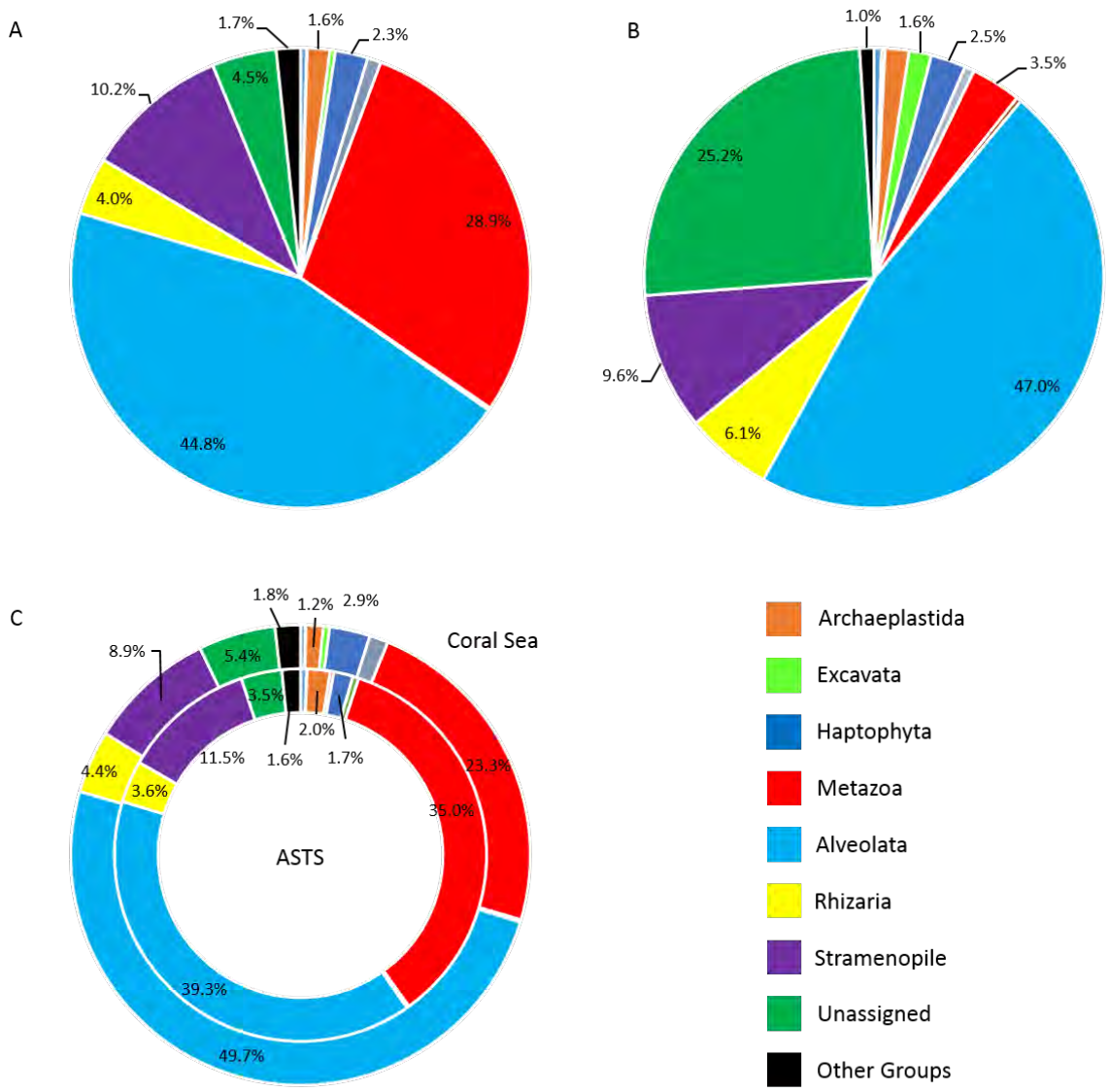

Figure 3.4 Overview of the relative OTUs abundance (A) and richness (B) of different eukaryotic supergroups in Australian tropical waters. The diversity of microbial eukaryotes at the supergroup taxonomic level in Arafura Sea/Torres Strait (ASTS) (C, inner) and Coral Sea (C, outer).

At a deeper taxonomic level, the same 9 groups represented the most abundant 18S rRNA sequences in the Arafura Sea/Torres Strait (Figure 3.5), making up more than 70% of total reads. Their taxa compositions, however, exhibited a significant correlation with geographical region. Bilateria (23%, within Metazoa) had the highest abundance of 18S rRNA sequence in the Arafura Sea/Torres Strait, which was three times higher than seen in the Coral Sea. Diatoms

represented 5.1% of the 18S rRNA abundance in the Arafura Sea/Torres Strait, compared to only 0.9% in Coral Sea. Whereas Dinoflagellata (22%, within the Alveolata) was the most represented 18S rRNA group in the Coral Sea, followed by Syndiniales (15.2%, within the Alveolata) and Cnidaria (14.5%, within the Metazoa), and had lower 18S rRNA abundances (19%, 9.8% and 10.3%, respectively) in the Arafura Sea/Torres Strait. Most of the remaining OTUs (abundance < 1%) showed similar 18S rRNA abundances in these two basins.

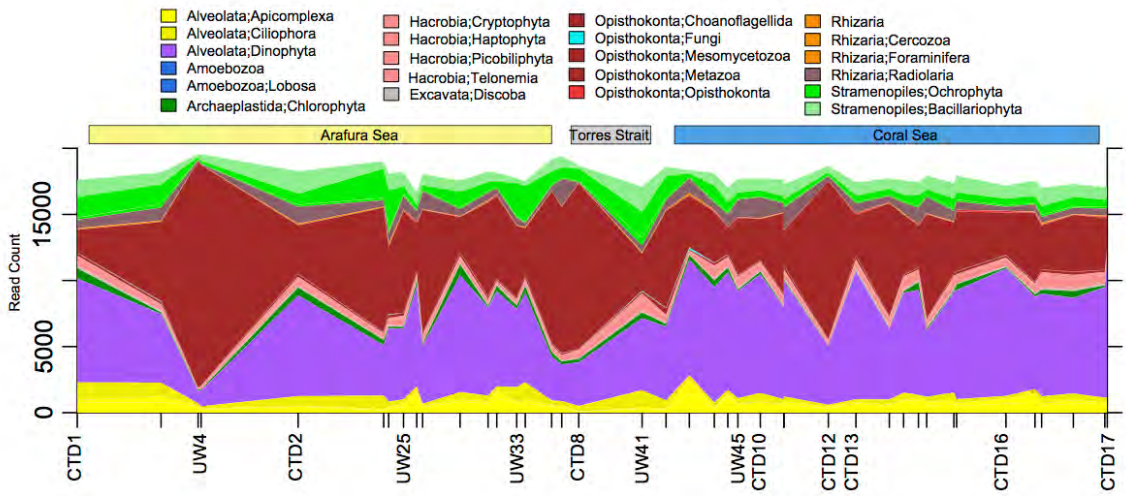

Figure 3.5 Spatial distribution and relative abundance of major eukaryote 18S V9 classes along the Arafura Sea, the Torres Strait and the Coral Sea in October 2012. Taxonomic information was extracted from the PR2 database and summarized at the class level.

Eukaryotic community structures and environmental drivers

Non-metric multidimensional scaling (NMDS) plot was applied to compare the similarity in the eukaryotic community compositions of surface water samples collected from the Arafura Sea/Torres Strait and Coral Sea, based on Bray-Curtis dissimilarities. In the combined NMDS plots of both basins, the Coral Sea samples formed a distinct cluster and were clearly separated from all Arafura Sea/Torres Strait samples (Figure 3.6A), which was supported by the analysis of multivariate homogeneity of group dispersions (p=0.002) using betadisper and anova methods in Implements Marti Anderson's PERMDISP2 procedure. The Arafura Sea/Torres Strait samples showed a greater spread pattern that of the Coral Sea, indicating that there is

greater community diversity between samples in the Arafura Sea/Torres Strait. To reveal which local environmental variables were drivers of eukaryote community structure, we investigated the relationship between the set of factors and eukaryotic community variation by Canonical Correspondence Analysis (CCA) with the envfit function from the Vegan Package (Figure 3.6B).

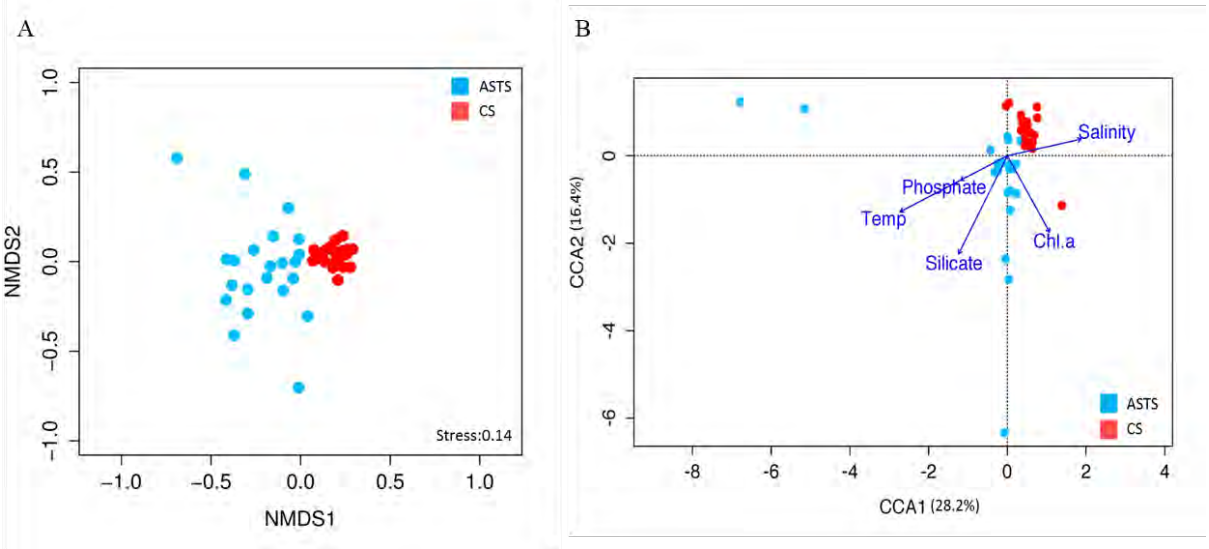

Figure 3.6 (A) Non-metric Multidimensional Scaling (NMDS) plot showing the eukaryotic community similarity between Arafura Sea/Torres Strait (ASTS, blue) and Coral Sea (CS, red) samples. (B) Canonical Correspondence Analysis (CCA) showing the eukaryotic community composition in relation to the environmental variables. Filled circles represent sampling sites from Arafura Sea/Torres Strait and Coral Sea. Arrows represent environmental variables that have statistically significant (P<0.05).

The results indicated that temperature, salinity, phosphate, silicate and Chl *a* have a statistically significant (P<0.05) relationship to the community structure. The axes 1 and 2 explained 28.2% and 16.4% of the community variability, respectively. The BIOENV approach was applied to find the best subset of the environmental variables based on finding maximum correlations between Bray-Curtis dissimilarity and the Euclidean distances of scaled environmental variables. The best subsets of environmental drivers for all samples were temperature, salinity, nitrate and silicate. In contrast, the BIOENV analysis showed that temperature, salinity and phosphate explained differences in community composition of samples in Arafura Sea/Torres Strait, and only temperature and salinity in Coral Sea, respectively. The importance of these

environmental variables and best BIOENV model were also verified for samples from Arafura Sea/Torres Strait and Coral Sea independently and for all samples combined using a Mantel test (Table 3.1).

Table 3.1 The contribution of environmental variables for marine eukaryotes community composition in the Arafura Sea/Torres Strait (ASTS) and Coral Sea (CS).

Environmental variables	All Samples		ASTS		CS	
	Mantel r	*P-value*	*Mantel r*	*P-value*	*Mantel r*	*P-value*
Temperature	0.345	0.001**	0.257	0.048*	0.656	0.001**
Salinity	0.296	0.001**	0.231	0.128	0.483	0.001**
Nitrate	0.079	0.221	0.103	0.646	0.205	0.013*
Phosphate	0.10	0.141	0.064	0.294	0.375	0.001**
Silicate	0.323	0.004**	-0.111	0.796	0.62	0.001**
Ammonia	0.025	0.311	0.174	0.162	0.29	0.01*
Chl a	0.181	0.026*	0.176	0.049*	0.037	0.342
Best BIOENV model	0.364	0.002**	0.345	0.046*	0.783	0.001**

Significate codes: '***' <0.001; '**' 0.001-0.01; '*' 0.01-0.05

Discussion

Marine microbial eukaryotes play a vital role in the marine food webs and biogeochemical cycles (Anjusha et al 2013). Picoeukaryotes, with a cell diameter less than 3 µm (Massana 2011), are distributed widely in marine environments. Although they have relatively low abundance, compared to cyanobacteria, they contribute significantly to marine plankton biomass and productivity due to their greater biovolume (Li 1994, Man-Aharonovich et al 2010, Worden et al 2004). Flow cytometry has typically been used to estimate the abundance of photosynthetic microbes based on their natural fluorescent pigment (Olson et al 1985, Vaulot and Marie 1999). Previous studies indicate that photosynthetic picoeukaryote abundances are typically lower in oligotrophic systems than in coastal and nutrient-rich regions, with an average density between 10^2 and 10^5 cells mL^{-1} in the photic zone (Li 2009). High photosynthetic picoeukaryote concentration were also found in high chlorophyll *a* surface

waters in the Northern Atlantic Ocean and Arctic Ocean (Kirkham et al 2013). Our study found similar results, with the lowest abundance of plastidic protists (~979 cells mL^{-1}) observed in the oligotrophic waters of the Coral Sea, and higher abundance seen in the Arafura Sea (~2.3×10^3 cells mL^{-1}) and Torres Strait (~3×10^3 cells mL^{-1}), along with higher nutrient and chlorophyll *a* concentrations. A few studies have also used flow cytometry to quantify the number of small heterotrophic eukaryotes (Zubkov and Tarran 2008). They account for 20% to 30% of eukaryotes, with a range of 3×10^2 to 3×10^3 cells mL^{-1} (Massana et al 2006). This was not investigated in our study.

The eukaryote assemblage composition in the surface layer of Australian tropical waters was investigated by high-throughput sequencing of the V9 region 18S rRNA gene. This barcode can characterize the micro-eukaryotic diversity over extensive taxonomic and ecological scales with the following advantages: first, it has simple secondary structure and is universally conserved in length, thus allowing relatively unbiased PCR amplification across eukaryotic lineages before Illumina sequencing; second, it includes both highly variable and stable nucleotide positions over evolutionary time frames, allowing discrimination of taxa over a substantial phylogenetic depth; third, it is highly represented in public reference databases across the eukaryotic tree of life, allowing taxonomic assignment among all known eukaryotic lineages (de Vargas et al 2015).

A total of 6,129 OTUs were obtained and classified into 1,014 deep-branching lineages. Rarefaction curves approached saturation and OTU abundances displayed a good fit to the truncated Preston Lognoral distribution, suggesting that the current sequencing effort well represented eukaryote richness at both local and global scales. These indicate that the Australian tropical waters contain a relatively modest number of eukaryotic phylotypes, likely on the order of a few thousand. Shannon index analysis showed that the diversity of eukaryotes was similar

between the Arafura Sea/Torres Strait and the Coral Sea, with an average of 4.13 and 4.19, respectively.

Considered at a supergroup level, the eukaryotic assemblages were dominated by the SAR supergroups (Stramenopiles, Alveolata and Rhizaria), Metazoa and Archaeplastida, together representing ~90% of all reads and 68% of all OTUs. All 18S rRNA data needs to be interpreted cautiously, however, as the relative abundance of rDNA in libraries may not reflect the organismal abundance in nature. This is due to large variations in the rDNA copy numbers between different marine eukaryotic groups, which can lead to an over- or under-estimation of OTU abundances. For example, organisms such as dinoflagellates and cilicates have up to tens or hundreds of thousands of copies of the 18S rRNA gene per cell (Gong et al 2015), but fungi have only 60-220 copies per cell (Gong et al 2013).

The Alveolata, one of the major eukaryotic lineages, occupied 45% of total sequences in our study. Dinoflagellates (Alveolata) include autotrophs, mixtotrophs and heterotrophs that feed via osmotrophy and phagotrophy (Burkholder et al 2008, Sahu et al 2014). Half of the dinoflagellate groups are photosynthetic, with some known to be associated with “red tides” and other monospecific blooms (Taylor et al 2008). Some dinoflagellates (for example, *Symbiodinium*) have particular ecological significance on tropical coral reefs. These dinoflagellates are symbiotes of coral reefs and provide metabolites from photosynthesis to support coral metabolism, growth, reproduction and skeletogenesis (Davies 1991). In return, the dinoflagellates obtain inorganic nutrients and CO_2 from coral’s waste products and increased protection from grazers (Davy et al 2012). Dinoflagellates were the most abundant (21%) 18S rRNA group in the Coral Sea, which is likely due to the neighboring presence of the world’s largest reef system, the Great Barrier Reef.

Syndiniales are a parasitic order within the Alveolata (Guillou et al 2008), including Marine Alveolate (MALV) I and MALV II, both of which were abundant in our survey. This order contains several examples of parasites, such as *Amoebophrya*, that are parasitic on dinoflagellates. The elevated abundance of their host organisms in the Coral Sea compared to the Arafura Sea/Torres Strait may explain, in part, the similar trend observed for members of the Syndiniales. Previous studies suggested MALV II prefer waters with higher light availability and dominate in sunlit marine surface waters (Guillou et al 2008). This is consistent with our results with lower MALV II numbers in the more turbid waters of the Arafura Sea/Torres Strait. In our study, nearly 25% of sequences within the Syndiniales could not be clearly assigned to a specific clade, suggesting the possibility of more diverse lineages within this order. Although Cnidaria and Bilateria were the two phyla with the highest 18S rRNA amplicon abundances amongst the small metazoans in our study, their 18S rDNA relative richness is not a good indicator of their real abundance in nature, due to their high rRNA copy numbers and multicellarity.

Marine planktonic diatoms (Stramenopiles) play a significant global role in biogeochemical cycling and the functioning of aquatic food webs (Armbrust 2009, Falkowski 2002, Smetacek 1998), and contribute a large component of aquatic biomass and nearly 20% of the total primary production on Earth (Falkowski et al 1998, Field et al 1998, Shruti Malviyaa 2016). They are common in nutrient-rich regions, as well as coastal waters and upwelling areas, or during seasonal blooms in the open oceans. Typically, however, they have low abundances in open ocean oligotrophic areas (Armbrust 2009, Bopp et al 2005, Smetacek 2012). This is consistent with the abundance of diatoms in our data set, which showed significant different geographic distribution, representing 5.1% of total reads in the nutrient-rich Arafura Sea/Torres Strait with higher concentration of silicate (average 3.64 μM), while only 0.9% of total reads in the oligotrophic Coral Sea with only average 1.59 μM of silicate. In our data, there is a strong correlation between silicate concentrations in the water column and diatom abundance as

assessed by 18S rRNA abundance, which is consistent with their unique silica-containing cell walls. For example, diatoms occupied 22% of total reads at CTD6, where it had one of the highest silicate concentration (4.77 μM), however, it represented only 1% of total reads at CTD17 with the lowest silicate concentration (1.35 μM). Our survey also unveiled some important members with abundances less than 1%. Fungi represented only 0.2% of total reads, and overall its diversity appears dominated by the Basidiomycota and Ascomycota in the surface waters in our study. Previous studies have suggested that fungi are typically non-diverse and of low abundance in many upper and surface marine water column samples (Massana and Pedros-Alio 2008, Richards and Bass 2005), but have much higher abundances in deep subsurface sediments (Bass et al 2007, Edgcomb et al 2011, Orsi et al 2013) and deep seawaters (Pernice et al 2016).

Collodaria (Rhizaria) includes species with colonial lifestyles and without silicification. They are potentially the most important plankton in the oligotrophic ocean due to its high carbon fixation and primary production rates (Ishitani et al 2012). In our study, the abundance of Collodaria only represented 0.01% of total reads in the Arafura Sea/Torres Strait but was significantly higher (0.4% of total reads) in the oligotrophic Coral Sea waters. Collodaria bear photosynthetic dinoflagellate endosymbionts (Stoecker et al 2009), and it should be noted that both Collodaria and dinoflagellates had increased abundances in the Coral Sea relative to the Arafura Sea/Torres Strait. Mamiellales, an order of prasinophyte (Archaeplastida), contains several widespread marine photosynthetic picophytoplankton taxa, including *Ostreococcus* and *Micromonas*. The former has been recognized as the smallest free-living eukaryote with a 12.56 Mb nuclear genome and rapid growth rates (Derelle et al 2006). The later harbors transporters for ammonium and nitrate (McDonald et al 2010) and has been reported as the dominant photosynthetic picoeukaryote in several oceanic and coastal regions and nutrient-rich environments (Not et al 2004, Thomsen and Buck 1998, Worden et al 2009), this supports our

results, as *Micromonas* is at its highest abundance in the Arafura Sea/Torres Strait where higher nitrogen and ammonia concentrations have been observed.

An NMDS plot showed clear clustering of the marine eukaryotic communities according to geographical region, with the Coral Sea communities showing a very tight clustering (Figure 3.8A). It has been proposed that biogeographical patterns can be explained by four general processes: selection, drift, dispersal and mutation (Hanson et al 2012). These four processes can be evaluated by testing correlation between microbial composition and measured environmental variables across sampled locations (environment effect) and testing whether geographical distance explained the remaining variation in microbial composition (historical processes) (Hanson et al 2012).

The CCA analysis indicated that the eukaryotic community structure strongly correlated with five environmental variables (temperature, salinity, silicate, phosphate and Chl *a*). Utilizing a BIOENV analysis and Mantel test, we identified temperature and salinity as the two major environmental drivers of eukaryotic community structure in Australian tropical waters. Temperature in the surface ocean is a fundamental control on marine eukaryotic phytoplankton community structure and metabolic processes that sets the biogeographical boundaries or biomes of major phytoplankton groups (Moisan et al 2002, Needoba et al 2007, Raven and Geider 1988). It has been reported that temperature has significant effects on other properties such as biogeochemical cycling of carbon and nitrogen, and on phytoplankton processes such as growth, photo-physiology, and calcification (Hare et al 2007, Rose et al 2009). Salinity has been identified as a predominant factor influencing the survival, growth and development of marine microeukaryotic plankton by impacting their cell response to metabolic or osmoregulatory changes (Estudillo et al 2000).

In addition to temperature and salinity, nutrient concentrations were also important influences on the eukaryotic community structure. Previous investigation shows that nutrient variables had a significant impact on marine eukaryotic communities because they are well adapted to preferable nutritional conditions (Ferrierpages and Rassoulzadegan 1994). Nutrients can control the photosynthesis of phytoplankton directly, while heterotrophic eukaryotes can be impacted by nutrients through their effect on phytoplankton growth (Hecky and Kilham 1988). The Mantel test found that inorganic nutrients contributed to variation of the eukaryotic community composition in the Coral Sea, for example, nitrate (P=0.013), phosphate (P=0.001), silicate (P=0.001) and ammonia (P=0.01).

In conclusion, we used amplicon sequences of the V9 region of the 18S rRNA gene to investigate the taxonomy, biodiversity and ecological structure of eukaryotes in surface seawaters across the Arafura Sea/Torres Strait and Coral Sea. We found that the SAR supergroups (Stramenopiles, Alveolata and Rhizaria), Metazoa and Archaeplastida dominating the 18S rRNA read counts in all samples but their relative proportions differed across each sample. Although both Arafura Sea/Torres Strait and Coral Sea shared the same 9 most abundant groups, the relative abundances of some of these groups significantly correlated with geographic region. For example, the Bilateria and diatoms were three times and five times more abundant in the Arafura Sea/Torres Strait than the Coral Sea. In contrast, the Syndiniales were 1.5 times more abundant in the Coral Sea compared with the Arafura Sea/Torres Strait. Community composition differences were largely explained by the environmental variables (temperature, salinity, phosphate, silicate and Chl *a*). These results allowed us better understanding of the biogeographic patterns of eukaryotic communities and their underlying environmental determinants in Australian Tropical Seawater. Further study of the abundance and expression of functional genes is needed to investigate the relationship between eukaryotic microbial communities and ecosystem function and understand how eukaryotes have adapted to the different and changing living conditions in tropical waters of Northern Australia.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We thank the crew of the *RV* Southern Surveyor and Associate Professor Martina Doblin, the Voyage leader for SS2012_t07. This work was supported by ARC DP110102718, ARC DP150102326, ARC FL140100021, CSC scholarship to T. Huang.

Author Contributions:

TH and IP conceived the study, TH conducted experiments, TH and HG analyzed data, MO, MB, LM and JS collected samples and extracted DNA, TH and IP wrote the manuscript with contributions from all other authors.

References

Amaral-Zettler LA, McCliment EA, Ducklow HW, Huse SM (2009). A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *Plos One* **4**.

Andrews JC, Clegg S (1989). Coral Sea circulation and transport deduced from modal information models. *Deep-Sea Res* **36:** 957-974.

Anjusha A, Jyothibabu R, Jagadeesan L, Mohan AP, Sudheesh K, Krishna K *et al* (2013). Trophic efficiency of plankton food webs: observations from the Gulf of Mannar and the Palk Bay, southeast coast of India. *J Marine Syst* **115:** 40-61.

Armbrust EV (2009). The life of diatoms in the world's oceans. *Nature* **459:** 185-192.

Bass D, Howe A, Brown N, Barton H, Demidova M, Michelle H *et al* (2007). Yeast forms dominate fungal diversity in the deep oceans. *P Roy Soc B-Biol Sci* **274:** 3069-3077.

Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R *et al* (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat Methods* **10:** 57-U11.

Bopp L, Aumont O, Cadule P, Alvain S, Gehlen M (2005). Response of diatoms distribution to global warming and potential implications: a global model study. *Geophys Res Lett* **32**.

Burkholder JM, Glibert PM, Skelton HM (2008). Mixotrophy, a major mode of nutrition for harmful algal species in eutrophic waters. *Harmful Algae* **8:** 77-93.

Clarke KR, Ainsworth M (1993). A method of linking multivariate community structure to environmental variables. *Mar Ecol Prog Ser* **92:** 205-219.

Condie S, Ridgeway, K, Griffiths, FB, Rintoul, S, & Dunn, J . (2003). National oceanographic description and information review for national bioregionalisation. *Unpublished report to the National Oceans Office, Hobart*

Condie SA (2011). Modeling seasonal circulation, upwelling and tidal mixing in the Arafura and Timor Seas. *Cont Shelf Res* **31:** 1427-1436.

Davies PS (1991). Effect of daylight variations on the energy budgets of shallow-water corals. *Mar Biol* **108:** 137-144.

Davy SK, Allemand D, Weis VM (2012). Cell biology of cnidarian-dinoflagellate symbiosis. *Microbiol Mol Biol R* **76:** 229-261.

de Vargas C, Audic S, Henry N, Decelle J, Mahe F, Logares R *et al* (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**.

Delong EF (1992). Archaea in coastal marine environments. *Proceedings of the National Academy of Sciences of the United States of America* **89:** 5685-5689.

Derelle E, Ferraz C, Rombauts S, Rouze P, Worden AZ, Robbens S *et al* (2006). Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proceedings of the National Academy of Sciences of the United States of America* **103:** 11647-11652.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27:** 2194-2200.

Edgcomb VP, Beaudoin D, Gast R, Biddle JF, Teske A (2011). Marine subsurface eukaryotes: the fungal majority. *Environ Microbiol* **13:** 172-183.

Estudillo CB, Duray MN, Marasigan ET, Emata AC (2000). Salinity tolerance of larvae of the mangrove red snapper (*Lutjanus argentimaculatus*) during ontogeny. *Aquaculture* **190:** 155-167.

Falkowski PG, Barber RT, Smetacek V (1998). Biogeochemical controls and feedbacks on ocean primary production. *Science* **281:** 200-206.

Falkowski PG (2002). The ocean's invisible forest - Marine phytoplankton play a critical role in regulating the earth's climate. Could they also be used to combat global warming. *Sci Am* **287:** 54-61.

Ferrierpages C, Rassoulzadegan F (1994). Seasonal impact of the microzooplankton on picoplankton and nanoplankton growth-rates in the northwest Mediterranean-Sea. *Mar Ecol Prog Ser* **108:** 283-294.

Field CB, Behrenfeld MJ, Randerson JT, Falkowski P (1998). Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281:** 237-240.

Giovannoni SJ, Britschgi TB, Moyer CL, Field KG (1990). Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **345:** 60-63.

Gong J, Dong J, Liu XH, Massana R (2013). Extremely high copy numbers and polymorphisms of the rDNA operon estimated from single-cell analysis of oligotrich and peritrich ciliates. *Protist* **164:** 369-379.

Gong J, Shi F, Ma B, Dong J, Pachiadaki M, Zhang XL *et al* (2015). Depth shapes alpha- and beta-diversities of microbial eukaryotes in surficial sediments of coastal ecosystems. *Environ Microbiol* **17:** 3722-3737.

Green JL, Holmes AJ, Westoby M, Oliver I, Briscoe D, Dangerfield M *et al* (2004). Spatial scaling of microbial eukaryote diversity. *Nature* **432:** 747-750.

Guillou L, Viprey M, Chambouvet A, Welsh RM, Kirkham AR, Massana R *et al* (2008). Widespread occurrence and genetic diversity of marine parasitoids belonging to Syndiniales (Alveolata). *Environ Microbiol* **10:** 3349-3365.

Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner 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 Res* **41:** D597-D604.

Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH (2012). Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* **10:** 497-506.

Hare CE, Leblanc K, DiTullio GR, Kudela RM, Zhang Y, Lee PA *et al* (2007). Consequences of increased temperature and CO_2 for phytoplankton community structure in the Bering Sea. *Mar Ecol Prog Ser* **352:** 9-16.

Harris PT (1988). Sediments, bedforms and bedload transport pathways on the continental-shelf adjacent to Torres Strait, Australia Papua New Guinea. *Cont Shelf Res* **8:** 979-1003.

Hecky RE, Kilham P (1988). Nutrient limitation of phytoplankton in fresh-water and marine environments: a review of recent-evidence on the effects of enrichment. *Limnol Oceanogr* **33:** 796-822.

Ishitani Y, Ujiie Y, de Vargas C, Not F, Takahashi K (2012). Phylogenetic relationships and evolutionary patterns of the order collodaria (radiolaria). *Plos One* **7**.

Jongsma D (1974). Marine geology of the Arafura Sea. *Bureau of Mineral Resources, Geology and Geophysics, Canberra* **157:** 1-73.

Jürgens K, Massana, R (2008). Protist grazing on marine bacterioplankton. In: Kirchman DL (ed). *Microbial Ecology of the Oceans*, second edn. Willey, New York. pp 383-441.

Kämpf J (2015). Undercurrent-driven upwelling in the northwestern Arafura Sea. *Geophys Res Lett* **42:** 9362-9368.

Kirkham AR, Lepere C, Jardillier LE, Not F, Bouman H, Mead A *et al* (2013). A global perspective on marine photosynthetic picoeukaryote community structure. *ISME Journal* **7:** 922-936.

Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microb* **79:** 5112-5120.

Li WKW (1994). Primary production of prochlorophytes, cyanobacteria, and eukaryotic ultraphytoplankton: measurements from flow cytometric sorting. *Limnol Oceanogr* **39:** 169-175.

Li WKW (2009). From cytometry to macroecology: a quarter century quest in microbial oceanography. *Aquat Microb Ecol* **57:** 239-251.

Liu H, Probert I, Uitz J, Claustre H, Aris-Brosou S, Frada M *et al* (2009). Extreme diversity in noncalcifying haptophytes explains a major pigment paradox in open oceans. *Proceedings of the National Academy of Sciences of the United States of America* **106:** 12803-12808.

Lopez-Garcia P, Rodriguez-Valera F, Pedros-Alio C, Moreira D (2001). Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. *Nature* **409:** 603-607.

Loreau M, Naeem S, Inchausti P, Bengtsson J, Grime JP, Hector A *et al* (2001). Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science* **294:** 804-808.

Lyne V HD (2005). Pelagic regionalisation: national marine bioregionalisation intgration project *CSIRO Marine and Atmospheric Research, Hobart*.

Magoc T, Salzberg SL (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27:** 2957-2963.

Man-Aharonovich D, Philosof A, Kirkup BC, Le Gall F, Yogev T, Berman-Frank I *et al* (2010). Diversity of active marine picoeukaryotes in the Eastern Mediterranean Sea unveiled using photosystem-II *psbA* transcripts. *ISME Journal* **4:** 1044-1052.

Marie D, Partensky F, Jacquet S, Vaulot D (1997). Enumeration and cell cycle analysis of natural populations of marine picoplankton by flow cytometry using the nucleic acid stain SYBR Green I. *Appl Environ Microb* **63:** 186-193.

Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL *et al* (2006). Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* **4:** 102-112.

Massana R, Castresana J, Balague V, Guillou L, Romari K, Groisillier A *et al* (2004). Phylogenetic and ecological analysis of novel marine stramenopiles. *Appl Environ Microb* **70:** 3528-3534.

Massana R, Terrado R, Forn I, Lovejoy C, Pedros-Alio C (2006). Distribution and abundance of uncultured heterotrophic flagellates in the world oceans. *Environ Microbiol* **8:** 1515-1522.

Massana R, Pedros-Alio C (2008). Unveiling new microbial eukaryotes in the surface ocean. *Curr Opin Microbiol* **11:** 213-218.

Massana R (2011). Eukaryotic picoplankton in surface oceans. *Annual Review of Microbiology* **65:** 91-110.

McDonald SM, Plant JN, Worden AZ (2010). The mixed lineage nature of nitrogen transport and assimilation in marine eukaryotic phytoplankton: a case study of micromonas. *Mol Biol Evol* **27:** 2268-2283.

Minchin PR (1987). An evaluation of the relative robustness of techniques for ecological ordination. *Vegetatio* **69:** 89-107.

Moisan JR, Moisan TA, Abbott MR (2002). Modelling the effect of temperature on the maximum growth rates of phytoplankton populations. *Ecol Model* **153:** 197-215.

Monier A, Comte J, Babin M, Forest A, Matsuoka A, Lovejoy C (2015). Oceanographic structure drives the assembly processes of microbial eukaryotic communities. *ISME Journal* **9:** 990-1002.

Morgan-Smith D, Clouse MA, Herndl GJ, Bochdansky AB (2013). Diversity and distribution of microbial eukaryotes in the deep tropical and subtropical North Atlantic Ocean. *Deep-Sea Res Pt I* **78:** 58-69.

Needoba JA, Foster RA, Sakamoto C, Zehr JP, Johnson KS (2007). Nitrogen fixation by unicellular diazotrophic cyanobacteria in the temperate oligotrophic North Pacific Ocean. *Limnol Oceanogr* **52:** 1317-1327.

Not F, Latasa M, Marie D, Cariou T, Vaulot D, Simon N (2004). A single species, *Micromonas pusilla* (Prasinophyceae), dominates the eukaryotic picoplankton in the western English channel. *Appl Environ Microb* **70:** 4064-4072.

Oksanen J, Blanchet FG, Kindit R, Legendre P, Minchin PR, O'Hara RB *et al* (2015). Vegan: community ecology package.

Olson RJ, Vaulot D, Chisholm SW (1985). Marine phytoplankton distributions measured using shipboard flow-cytometry. *Deep-Sea Res* **32:** 1273-1280.

Orsi W, Biddle JF, Edgcomb V (2013). Deep sequencing of subseafloor eukaryotic rRNA reveals active fungi across marine subsurface provinces. *Plos One* **8**.

Palmer MW (1993). Putting things in even better order: the advantages of canonical correspondence analysis. *Ecology* **74:** 2215-2230.

Pernice MC, Giner CR, Logares R, Perera-Bel J, Acinas SG, Duarte CM *et al* (2016). Large variability of bathypelagic microbial eukaryotic communities across the world's oceans. *ISME Journal* **10:** 945-958.

Preston FW (1948). The commonness, and rarity, of species. *Ecology* **29:** 254-283.

Raven JA, Geider RJ (1988). Temperature and algal growth. *New Phytol* **110:** 441-461.

Richards TA, Bass D (2005). Molecular screening of free-living microbial eukaryotes: diversity and distribution using a meta-analysis. *Curr Opin Microbiol* **8:** 240-252.

Rose JM, Feng Y, DiTullio GR, Dunbar RB, Hare CE, Lee PA *et al* (2009). Synergistic effects of iron and temperature on Antarctic phytoplankton and microzooplankton assemblages. *Biogeosciences* **6:** 3131-3147.

Sahu G, Mohanty AK, Samantara MK, Satpathy KK (2014). Seasonality in the distribution of dinoflagellates with special reference to harmful algal species in tropical coastal environment, Bay of Bengal. *Environ Monit Assess* **186:** 6627-6644.

Saint-Cast FC, Scott. (2006). Circulation modelling in Torres Strait. *Geoscience Australia, Record 2006/18, 82pp*.

Salani FS, Arndt H, Hausmann K, Nitsche F, Scheckenbach F (2012). Analysis of the community structure of abyssal kinetoplastids revealed similar communities at larger spatial scales. *ISME Journal* **6:** 713-723.

Salazar G, Cornejo-Castillo FM, Benitez-Barrios V, Fraile-Nuez E, Alvarez-Salgado XA, Duarte CM *et al* (2016). Global diversity and biogeography of deep-sea pelagic prokaryotes. *ISME Journal* **10:** 596-608.

Sherman K, Hempel, G. (2008). The UNEP large marine ecosystem report: a perspective on changing conditions in LMEs of the world's regional seas. *UNEP Regional Seas Report and Studies No 182 United Nations Environment Programme Nairobi, Kenya*

Shruti Malviyaa ES, Stéphane Audicc, Flora Vincenta, Alaguraj Veluchamya, Julie Poulaind, Patrick Winckerd, Daniele Iudiconeb, Colomban de Vargasc, Lucie Bittnera, Adriana Zingoneb, Chris Bowlera, (2016). Insights into global diatom distribution and diversity in the world's ocean. *PNAS* **113:** 1516-1525.

Smetacek V (1998). Biological oceanography: diatoms and the silicate factor. *Nature* **391:** 224-225.

Smetacek V (2012). Making sense of ocean biota: how evolution and biodiversity of land organisms differ from that of the plankton. *J Biosciences* **37:** 589-607.

Stoecker DK, Johnson MD, de Vargas C, Not F (2009). Acquired phototrophy in aquatic protists. *Aquat Microb Ecol* **57:** 279-310.

Taylor FJR, Hoppenrath M, Saldarriaga JF (2008). Dinoflagellate diversity and distribution. *Biodivers Conserv* **17:** 407-418.

Thomas MC, Selinger LB, Inglis GD (2012). Seasonal diversity of planktonic protists in southwestern Alberta Rivers over a 1-year period as revealed by terminal restriction fragment length polymorphism and 18S rRNA gene library analyses. *Appl Environ Microb* **78:** 5653-5660.

Thomsen HA, Buck KR (1998). Nanoflagellates of the central California waters: taxonomy, biogeography and abundance of primitive, green flagellates (Pedinophyceae, Prasinophyceae). *Deep-Sea Res Pt Ii* **45:** 1687-1707.

Tomczak MJSG (2003). Regional oceanography: an introduction. *Pergamon Press*.

Vaulot D, Marie D (1999). Diel variability of photosynthetic picoplankton in the equatorial Pacific. *J Geophys Res-Oceans* **104:** 3297-3310.

Wardle DA, Bardgett RD, Klironomos JN, Setala H, van der Putten WH, Wall DH (2004). Ecological linkages between aboveground and belowground biota. *Science* **304:** 1629-1633.

Wolanski E, Lambrechts J, Thomas C, Deleersnijder E (2013). The net water circulation through Torres strait. *Cont Shelf Res* **64:** 66-74.

Worden AZ, Nolan JK, Palenik B (2004). Assessing the dynamics and ecology of marine picophytoplankton: the importance of the eukaryotic component. *Limnol Oceanogr* **49:** 168-179.

Worden AZ, Lee JH, Mock T, Rouze P, Simmons MP, Aerts AL *et al* (2009). Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *micromonas*. *Science* **324:** 268-272.

Yooseph S, Nealson KH, Rusch DB, McCrow JP, Dupont CL, Kim M *et al* (2010). Genomic and functional adaptation in surface ocean planktonic prokaryotes. *Nature* **468:** 60-66.

Zubkov MV, Tarran GA (2008). High bacterivory by the smallest phytoplankton in the North Atlantic Ocean. *Nature* **455:** 224-248.

Chapter 4:

Oceanographic structure drives the marine prokaryote and eukaryote diversity in the euphotic zone of Australian tropical waters

Oceanographic structure drives the marine prokaryote and eukaryote diversity in the euphotic zone of Australian tropical waters

Taotao Huang[1], Martin Ostrowski[1], Mark V. Brown[2], Lauren Messer[3,4], Justin Seymour[3], Ian T. Paulsen[1]

1. Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, Australia
2. School of Biotechnology and Biomolecular Science, University of New South Wales, Sydney, Australia
3. Climate Change Cluster, University of Technology Sydney, Sydney, Australia
4. Current address: Australian Centre for Ecogenomics, University of Queensland, Brisbane, Australia

Corresponding Author: Ian Paulsen, Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, New South Wales, 2109, Australia.

Email: ian.paulsen@mq.edu.au

Abstract

The Arafura Sea, Torres Strait and Coral Sea are highly productive tropical waters in Northern Australia. Here, we provide the first report of the diversity patterns and community structure of the prokaryotes and protists that inhabit the euphotic zone of these regions using high-throughput sequencing of the 16S and 18S rRNA genes. Rarefaction curves suggested that the amplicon sequences provided a good representation of the marine microbial epipelagic diversity in all samples and the Shannon index indicated that their diversity increased with increasing sampling depth. The prokaryote community was dominated by cyanobacteria and proteobacteria in all samples, together representing 95% and 91% of all reads and Operational Taxonomic Units (OTUs) respectively. The relative abundance of cyanobacteria decreased with depth, however, the heterotrophic proteobacteria showed an opposite trend. The Alveolata, Stramenopiles and Rhizara supergroups dominated the protist community structure across all samples. Their abundances showed different geographic and spatial distribution, for instance, diatoms were more abundant in the silicate-rich basin with the amount dropping significantly when sampling depth exceeded 100 m. While marine microbial community composition was significantly different between basins, a simple Mantel test indicated that local environmental variables (salinity, temperature and nutrients) strongly shaped their β-diversities on a regional scale. In addition, we found that the seafloor depth in the Arafura Sea/Torres Strait and sampling depth in the Coral Sea was a significant influence on the bacterial and protist community structure. Altogether, our study sheds new light on the microbial community composition, diversity patterns and underling mechanisms of their distribution in the upper surface layer of Australian tropical waters.

Introduction

Marine microbes consist of a diverse range of organisms, including microalgae, archaea, bacteria, protozoa, fungi and viruses. They are pervasive and play many essential roles in the marine environment in terms of their biomass, diversity and ecosystem functioning (Azam and Malfatti 2007, Caron et al 2012, Diez et al 2004, Falkowski et al 2008, Foulon et al 2008). Due to their critical role in the function and health of marine ecosystems, it is crucial to understand the distribution of marine microbes and how they are influenced by external physicochemical parameters. This information should help us to predict the responses of marine ecosystems to future environmental changes. However, most marine microbes exist in complex, dynamic and interactive communities, and the organisms within these communities are difficult to characterize due to their small size and our limited capacity to culture them in a laboratory setting (Nemergut et al 2011, Westgate et al 2014). We have only rudimentary knowledge of total prokaryote and protistan species diversity and community structure in most natural environments, and how these features relate to ecosystem function (Schnetzer et al 2011). During the past decade, with the application of molecular ecology and high-throughput sequencing, many previously unsuspected diverse assemblages of marine bacteria, archaea and picoeukaryotes have been unveiled (de Vargas et al 2015, Giovannoni et al 1990, Kirkham et al 2013, Logares et al 2014, Martiny et al 2011, Nemergut et al 2011, Pernice et al 2016, Salazar et al 2015, Yooseph et al 2010).

Marine microbial diversity has not previously been examined in the euphotic zone in the tropical waters of Northern Australia. In this study, we collected samples from the Arafura Sea, Torres Strait and the Coral Sea from depths of 25-150 meters. The Arafura Sea lies between Australia, Indonesia and New Guinea, with depths of 50-80 meters. Climatically, this sea is fully tropical and experiences the relatively stable trade winds during part of the year and intermittent monsoonal flows during the austral summer periods (Jongsma 1974). The benthic boundary layer in this area is turbid (Burford and Rothlisberg 1999, Margvelashvili et al 2008)

due to the relatively high currents at the sea floor (Wolanski 1993). The adjacent land masses are sparsely populated and the region encompassing the Arafura Sea has been identified as having one of the lowest levels of human impact (Halpern et al 2008).

The Torres Strait links the Arafura Sea to the West with the Coral Sea in the East, with very shallow water depth (7 to 15 meters) (Harris 1988). It is an important international and national shipping sea lane, and contains valuable fisheries for prawns and crayfish (Hemer et al 2004). It defines a biological barrier between the Arafura and Coral Seas, and is a geological mixing zone of terrigenous and calcareous sediments (Harris 1999). The Indonesian Throughflow is a warm-water current that transports large amounts of relatively warm and fresh water from the western equatorial Pacific Ocean to the southwest Indian Ocean through the Arafura Sea and Torres Strait, which has a substantial influence on the climate of the entire region (Tomczak 2003). The Fly River in Papua New Guinea (PNG) empties around 7,500 $m^3 s^{-1}$ of fresh water into the Gulf of Papua (adjoining Torres Strait) as rainfalls in the highlands of PNG are 10-13 m annually. This impacts the local marine environment, for instance by lowering temperature and salinity (Alongi et al 1992, Wolanski et al 1995).

The Coral Sea is a sub-region of the South-Western Pacific Ocean, incorporating the Western extremity of the Pacific South Equatorial Current. It is recognized for its rich biodiversity and important heritage value, and harbors the world's largest Coral Reef system, the Great Barrier Reef (GBR). In this oligotrophic region, primary production is low and likely to be nitrate-limited throughout the year (Condie and Dunn 2006). Although this sea contains diverse plankton communities, nanoplankton and picoplankton contain a high proportion of the total chlorophyll (70-95%) and a large percentage of total productivity is likely based on recycled nutrients (Condie and Dunn 2006).

These basins are considered to be a region of high ecological and economical value (Condie 2011), and represent a very productive fishery zone in Australia, providing thousands of tons of shrimps, lobsters, crabs and snapper every year. Marine microorganisms underpin the entire marine food web and thus subsequently drive marine productivity, aquaculture and fishery yields (Azam et al 1983). Therefore, it is critical to investigate the diversity, distributional dynamics and environmental interactions of marine micro-organisms in the euphotic zone of Australia's tropical waters. In this study, we investigated diversity, taxonomic composition and biogeographical distribution of both prokaryotic and protists communities in the euphotic zone of Northern Australian tropical waters using high-throughput sequencing of the 16S and 18S rRNA genes. Samples were taken during the *RV* Southern expedition in October 2012. Moreover, we specifically tested the relative importance of depth and selected environmental factors on microbial diversity and community composition.

Materials and Methods

Study area, sample collection and oceanographic data

Samples were collected in Australian tropical waters across the Arafura Sea/Torres Strait and Coral Sea during a research cruise on-board the *RV* Southern Surveyor in October 2012 (Figure 4.1). A total of 34 seawater samples were collected using 10 L Niskin bottles that included 17 surface samples and 17 multiple depth samples (Table 1), with temperature and salinity measured by a rosette sampling system equipped with a Seabird SEB-911+ CTD (conductivity/temperature/depth) probe. For each sample 8 L of seawater was sequentially filtered through 0.22 μm pore-size white polycarbonate filters (Millipore, Massachusetts, United States). The filters were then frozen in liquid nitrogen immediately and stored at -80 °C until DNA extraction. The concentrations of inorganic nutrients were determined from depth profile samples collected from CTD stations and processed according to standardized protocols. Instrument details and data are available through the Australian Ocean Data Network

(https://researchdata.ands.org.au/southern-surveyor-voyage-ss2012t07-hydrology/436463?source=suggested_datasets).

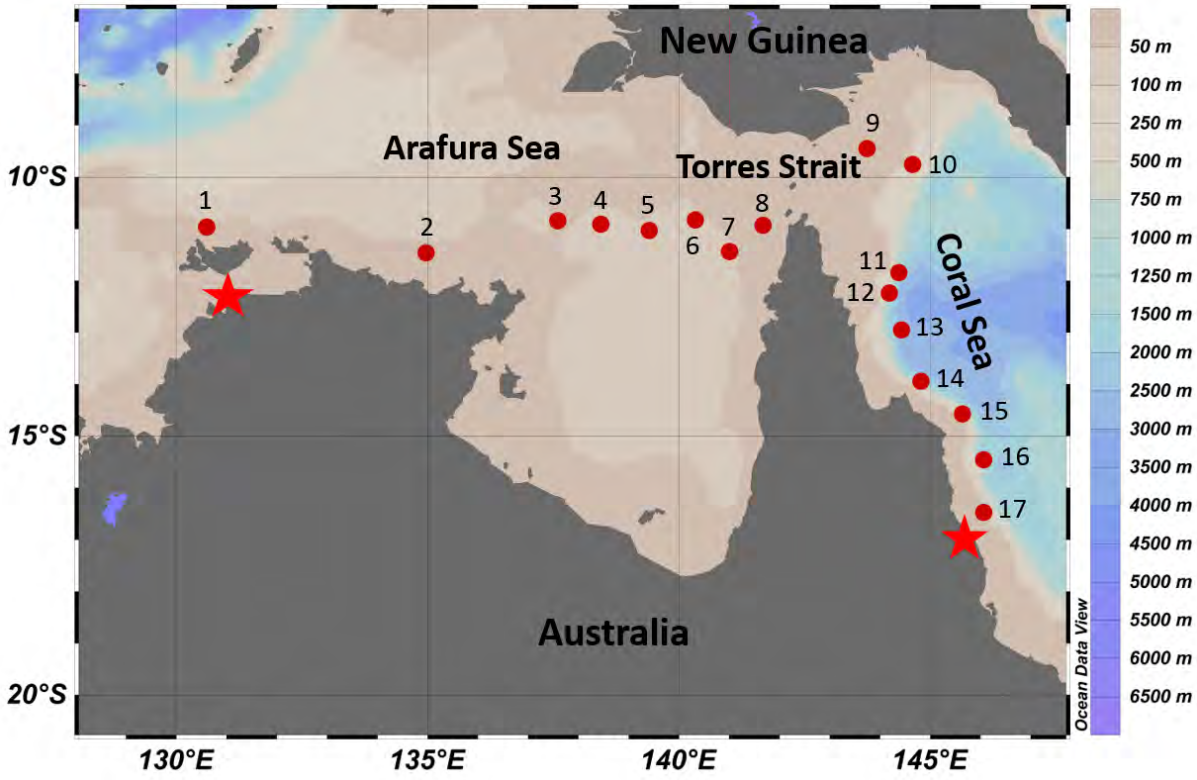

Figure 4.1 Schematic representation of the sampling sites (red dots) from Darwin (left star) to Cairns (right star) in October 2012. Colour bars on the right indicate depths.

DNA extraction and 16S and 18S rRNA gene amplicon sequencing

DNA was extracted using the MoBio PowerWater kit (MoBio Laboratories, Carlsbad, CA, USA), according to the manufacturer's instructions. The V1-V3 region of 16S rRNA gene was amplified using primer set 27F (5'-AGAGTTTGATCMTGGCTCAG-3') and 519R (5'-GWATTACCGCGGCKGCTG-3') (Lane et al 1985, Winsley et al 2012). The V9 region of 18S rRNA gene was amplified using primer set 1380F (5'-TTGTACACACCGCCC-3') and 1510R (5'-CCTTCYGCAGGTTCACCTAC-3') (Amaral-Zettler et al 2009). The barcodes and specific primers for Illumina sequencing using the Nextera Index Kit (Illumina) were added to the 5' ends of primer pair. Polymerase chain reactions (PCR) were conducted in 50 µl reactions mixtures containing 0.5 mM each primer, 2.5 mM $MgCl_2$, 1× Buffer (Promega), 0.2 mM dNTP, 2.5 U Taq DNA polymerase. The following cycling parameters were used for prokaryotes:

initial denaturation step at 95 °C for 10 min, and 35 cycles of 95 °C for 30 s, 55 °C for 10 s, and 72 °C for 45 s, with a final extension at 72 °C for 5 min. For eukaryotes the cycling parameters were: initial denaturation at 94°C for 3 min before 30 cycles of denaturation at 94 °C for 30 s, annealing at 57 °C for 60 s, extension at 72 °C for 90 s and final extension at 72 °C for 10 min. The amplicons were pooled and purified using AMPure XP beads (Illumina) and quantified by Nanodrop ND-2000. The barcoded PCR products were then sequenced at Ramaciotti Centre for Genomics using paired end reads.

Table 4.1 Seafloor and sampling depths at each CTD station.

	Sites	Seafloor depth (m)	Sampling depth (m)		
ASTS	CTD1	45	5	25	-
	CTD2	28	5	-	-
	CTD3	52	5	42	-
	CTD4	58	5	25	-
	CTD5	54	5	37	-
	CTD6	60	5	25	-
	CTD7	40	5	23	-
	CTD8	16	5	-	-
	CTD9	43	5	28	-
CS	CTD10	102	5	56	-
	CTD11	760	5	45	150
	CTD12	1211	5	75	150
	CTD13	2925	5	90	-
	CTD14	1239	5	120	-
	CTD15	789	5	90	-
	CTD16	1282	5	100	-
	CTD17	540	5	75	-

ASTS: Arafura Sea and Torres Strait
CS: Coral Sea

Processing of sequencing data

Paired-end reads were joined by FLASH (Magoc and Salzberg, 2011). Merged reads shorter than 70 bp and quality scores less than 20 were removed. Chimeric sequences were removed de novo with USEARCH 64bit v8.1 (Edgar et al., 2011). The remaining reads were then clustered into Operational Taxonomy Units (OTUs) using a 97% sequence identify cut-off followed by taxonomic assignments. Nuclear 18S OTUs were assigned against the PR2 database (Guillou et al., 2013), while 384 16S OTUs identified as chloroplasts were assigned against the PhytoRef database (Decelle et al., 2015) using Mothur classify.seqs with default settings (wang, cut-off = 80) (Kozich et al., 2013). OTUs with 4 or fewer reads were removed.

Statistical data analysis

All statistical analysis was carried out with R software (Version 3.2.1) using the Vegan Package version 2.3-0 (Oksanen et al 2015). The filtered OTU-abundance table obtained from the sequence clustering was rarefied down to the smallest sample size (7,497 reads per sample for prokaryotic libraries and 62,949 per sample for protistan libraries) using rrarefy function. Alpha diversity, including richness estimation (Shannon Index) and rarefaction curves, were studied at a local scale (individual samples) using the rarefied OTU table. A Bray-Curtis dissimilarity matrix was used on the prokaryote and protist community composition to infer the variation of their assemblages in space and along environmental gradients (β-diversity). A Non-metric Multidimensional Scaling (NMDS) plot (Minchin 1987) was computed to visualize the community composition similarity of samples from different sampling regions and depths using the dissimilarities matrix. BIOENV analysis (Clarke and Ainsworth 1993) was applied to investigate the best subset of environmental variables that impact on biotic community composition using Spearman’s rank correlation method. Mantel tests were used to test the significance of individual environmental factors, and identify the best subset of environmental drivers and spatial factors that explain microbial community biodiversity.

Results

Environmental parameters

The depth of collected samples ranged from 23 to 42 m in the Arafura Sea and Torres Strait which have shallow bottom depths, and ranged from 56 to 150 m in the deeper waters of the Coral Sea. Environmental parameters and nutrient variables showed clear spatial patterns and correlated with the depth of the water. Briefly, the seawaters in the Arafura Sea/Torres Strait had a higher temperature, lower salinity and greater but more fluctuant nutrient concentrations than in the Coral Sea in October 2012 (Figure 4.2).

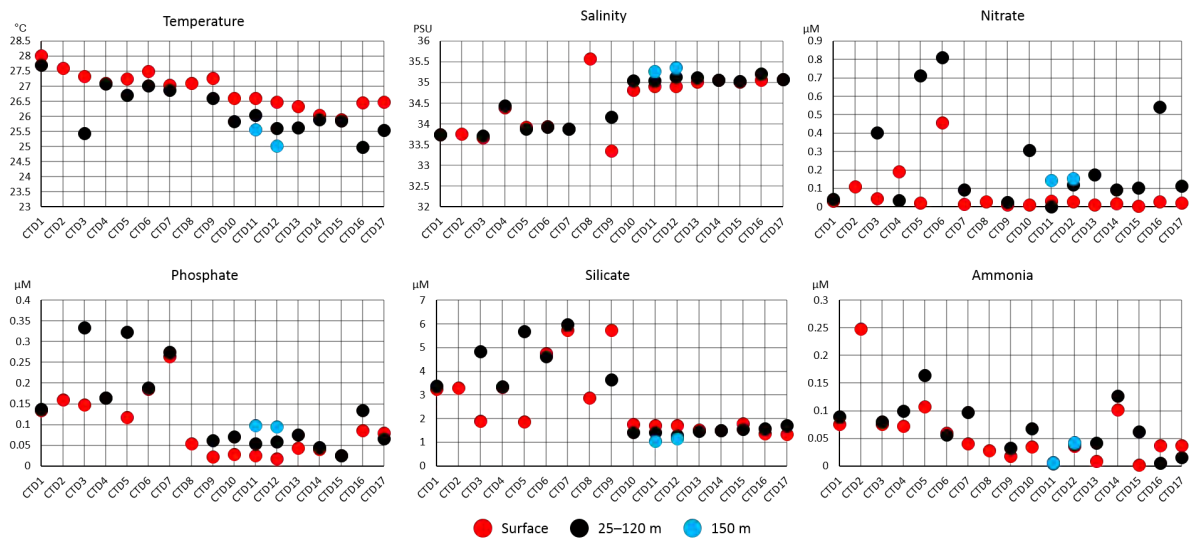

Figure 4.2 Physical parameters and nutrient concentrations at each CTD sampling station. Samples are coloured by depth: surface (red), 25 -120 m (black) and 150 m (blue).

In all sampling sites, water temperature decreased as depth increased, especially in the Coral Sea, average water temperature declined from 26.4 °C at the surface to 25.5 °C at depths from 45-150 m. In contrast, salinity did not change significantly with depth, except in CTD9, where it increased from 33.4 at the surface to 34.2 PSU at 28 m. This is probably caused by the large amount of fresh water emptied into this area from the Fly River in Papua New Guinea (Saint-Cast 2006). It is interesting to observe that the nutrient concentrations increased with depth in most sampling sites, especially in the Arafura Sea and Torres Strait. For example, at the CTD5 sampling site, the concentration of NO_3^-, PO_4^{3-}, SiO_4^{4-} and NH_4^+ at a depth of 37 m (the bottom

depth of 54 m) were 77, 2.8, 3.1 and 1.5-folder higher than at the surface, respectively. One exception to the observed higher nutrient concentrations at deeper depths, were the measurements of SiO_4^{4-} concentrations in the Coral Sea, which showed little change between the surface and deeper waters.

Alpha diversity of prokaryotic and eukaryotic communities

We examined microbial prokaryote diversity using sequencing of the V1-V3 region of the 16S rRNA gene and eukaryote diversity using the V9 region of the 18S rRNA gene from 34 samples, including 17 surface samples and 17 samples from greater depths. After quality checking and chimera removal, a total of 3.1 million 16S rRNA sequences were obtained that clustered into 1,037 operational taxonomic units (OTUs) and 12 million 18S rRNA sequences for protists and fungi, after removal of metazoan reads (1.6 million), that clustered into 5,449 OTUs at 97% similarity. Only 18 OTUs in the 16S OTU table and nearly 22.4% of the OTUs in the 18S OTU table did not match any reference sequences in the database, however, the unassigned OTUs only represented 0.3% and 5% of total reads, respectively.

Rarefaction curves were approaching plateaus indicating that the prokaryotic and protistan libraries provided a good representation of the microbial communities in each of the 34 samples (Figure 4.3A and C). The Shannon index was used to estimate the alpha diversity in each sample. Our data showed prokaryotic taxonomy diversity to range from 3.4 to 5.7, which is known to be generally higher in deeper waters than in the surface samples, except at the CTD3, CTD9 and CTD10 sampling sites which had the opposite trend with the Shannon indices decreasing as the depth increased (Figure 4.3B). For protistan diversity in each sample, the Shannon index ranged from 3.7 to 4.5 (Figure 4.3D). We observed that the alpha diversity of protists generally increased with depth compared to the surface samples. The deepest samples in this survey at the CTD11 and CTD12 sampling sites showed an unusual pattern. At a depth

of 150 m at CTD11, protistan diversity showed the lowest Shannon index value (3.7) in this study, whereas the prokaryote diversity showed the highest Shannon index value (5.7).

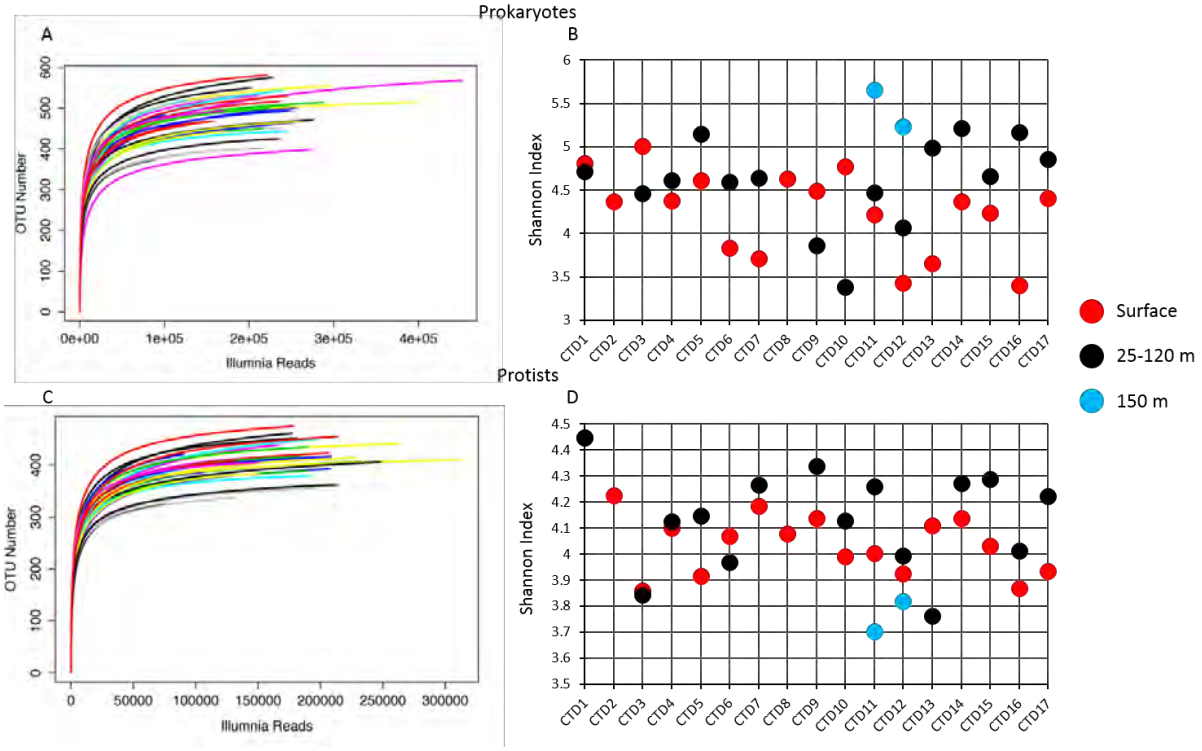

Figure 4.3 Rarefaction curves and Shannon index represent the diversity of the prokaryote (A, B) and protist (C, D) communities in each sample. In B and D, samples are coloured by depth: surface (red), 25 -120 m (black) and 150 m (blue).

Community compositions of prokaryotes and eukaryotes

The taxonomic distributions of the 16S rRNA sequences obtained in the seawater samples are shown in Figure 4.4A. Overall, the most abundant phylum of bacteria were the Cyanobacteria, which constituted 51.5% of the sequence reads in the surface samples and 44.7% of the sequence reads at greater depths (Figure 4.5A). The major taxa compositions in this group exhibited a remarkable correlation with geographical region. At surface and deeper waters, *Synechococcus* was the most abundant group in the Arafura Sea/Torres Strait, representing 30.3% of the total 16S rRNA gene sequences reads, but decreasing significantly in the Coral Sea (10% of the sequences reads), however, *Prochlorococcus* dominated in the Coral Sea (34% of the sequences reads) but not in the Arafura Sea/Torres Strait (10% of the sequences reads) (Figure 4.4A). As the sample depth increased, the percentage of cyanobacteria declined

dramatically, especially when the depth exceeded 100 m. For example, the cynaobacteria comprised only 7.4% of the sequence reads at the depth of 150 m in sampling site CTD15. Interestingly, we observed that the amount of *Synechococcus* in the Coral Sea increased with sampling depths between 25 to 100m. For example, in CTD12, it increased from 1.2% at the surface to 12.2% at a depth of 75 m.

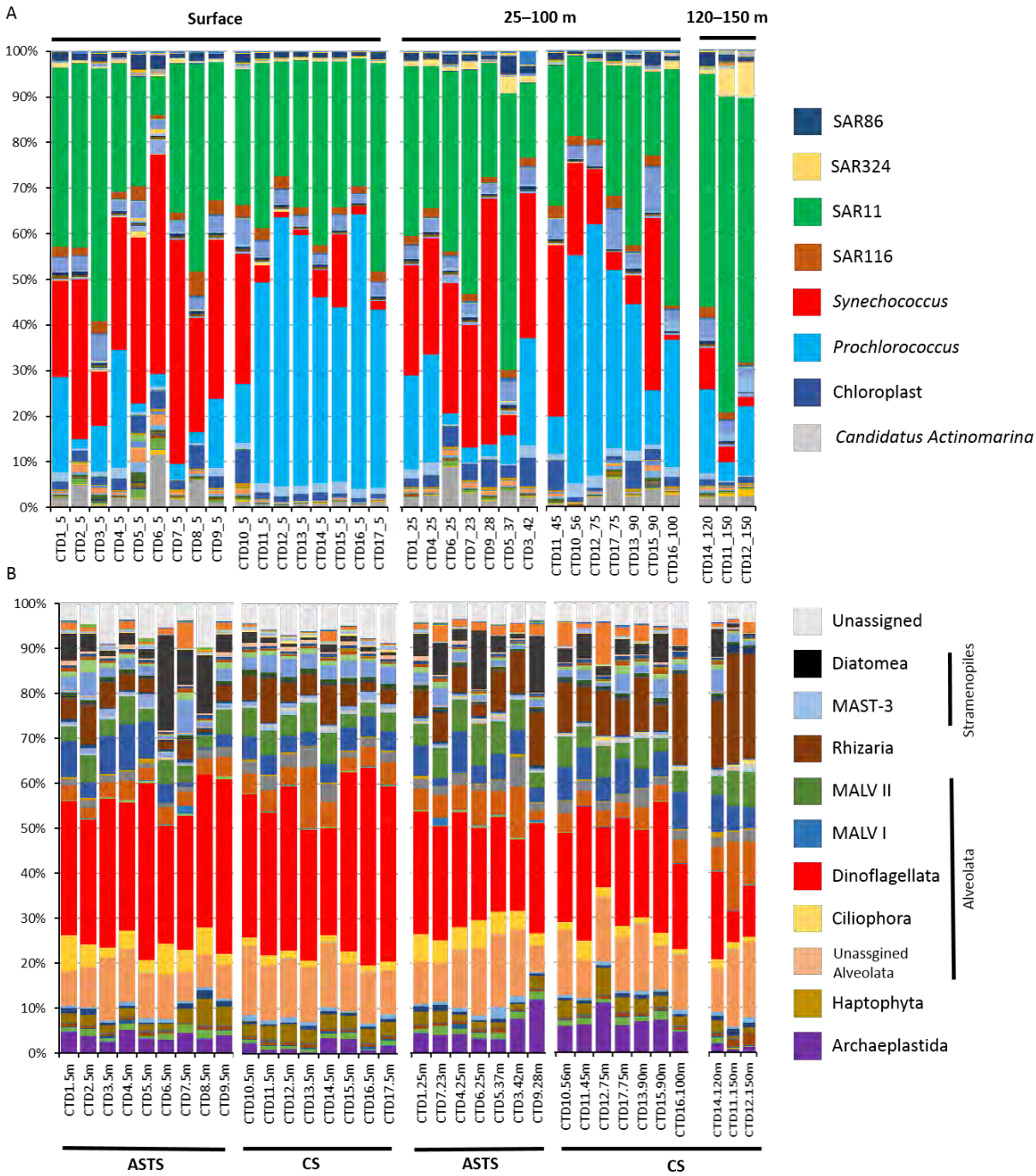

Figure 4.4 Prokaryote (A) and protist (B) community composition in the euphotic layer of the Arafura Sea/Torres Strait (ASTS) and Coral Sea (CS). Based on the sampling depths, samples were grouped into surface, 25-100 m and 120-150 m groupings.

The other dominant phylum of prokaryotes was the Proteobacteria, which represented 44% of the sequence reads in the surface samples, and 50.1% of the sequence reads in the deeper samples. The Alphaproteobacteria was the most abundant class of Proteobacteria, making up more than 92.5% of total Proteobacteria sequence reads and 21% of the total sequence reads. The SAR11 clade in the Alphaproteobacteria was one of the largest groups in all samples, especially when the depth was over 100 m. SAR324, an order of Deltaproteobacteria, showed increased abundance at depths of 150 m. There was much greater diversity within the Proteobacteria than the Cyanobacteria, with the former representing 67.8% of all prokaryote OTUs, compared with 18.6% for the Cyanobacteria (Figure 4.5B). The taxonomic distributions of the 18S rRNA sequences obtained in the seawater samples are shown in Figure 4.4B. At a superphylum level, the Australian tropical waters were dominated by Alveolata (62.9%) and Stramenopiles (14%), followed by Rhizaria (8.0%) and Archaeplastida (1.7%) (Figure 4.5C). The relative abundance of these superphylums were similar between surface samples and the depth samples, except the Rhizaria and Archaeplastida that increased from 5% and 3% at the surface to 11.3% and 5.4% at greater depths (Figure 4.5C). Within the Alveolata superphylum, Dinoflagellates were the most abundant group, accounting for 43% of Alveolata sequences, and their relative abundance decreased as the sampling depth increased. Syndiniales include Marine Alveolate (MALV) I and MALV II, which were also abundant in our survey.

The phototrophic taxa diatoms within Stramenopile supergroup showed significantly different geographic and spatial distribution (Figure 4.6). Generally, the diatom relative abundance was higher in the Arafura Sea/Torres Strait than in the Coral Sea, and higher at lower depths than at the surface. However, it dropped significantly when depths dropped below 100 m. The proportion of fungal sequences was lower than 0.6% in all samples except at a depths of 25 m in CTD6 samples, where it represented 0.93% of sequence reads (Figure 4.6).

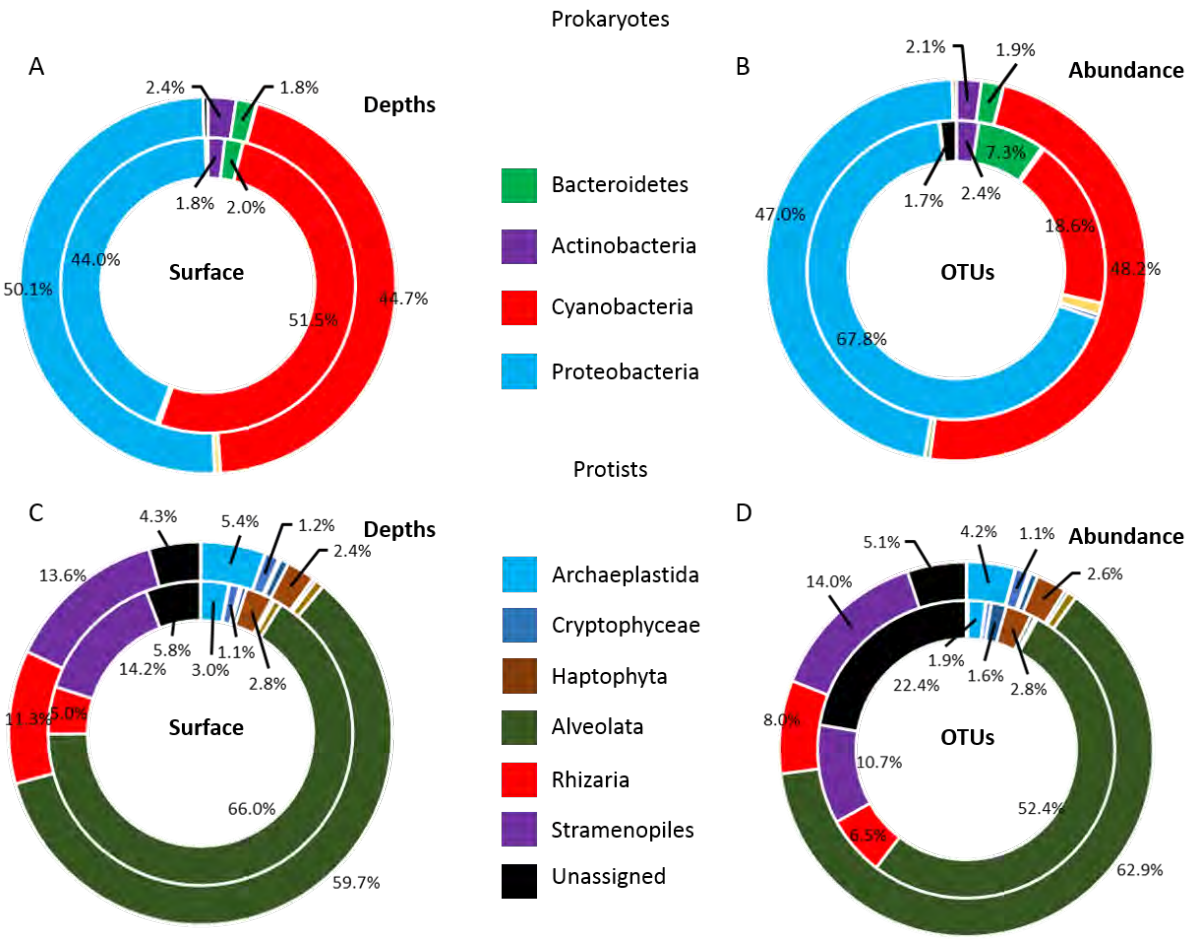

Figure 4.5 Diversity of prokaryote (A, at phylum level) and protist (C, at supergroup level) communities in surface (inner) and depths (outer) samples. The relative OTU abundance (outer) and richness (inner) of prokaryote (B) and protist (D) communities in Northern Australian tropic waters.

Beta-diversity patterns of prokaryotic and eukaryotic communities

To investigate bacterial and protistan community composition among the surface and depth samples collected from the Arafura Sea/Torres Strait and the Coral Sea, we applied a Non-metric multidimensional scaling (NMDS) analysis, based on Bray-Curtis dissimilarities calculated from the subsampled 16S OTU table of 7,497 sequences per sample and 18S OTU table of 62,949 sequences per sample. Generally, the NMDS results revealed that samples from the Arafura Sea/Torres Strait and the Coral Sea were apparently separated along the first axis (Figure 4.7). The adnois analyses also indicate that bacterial (p=0.014) and protists (p=0.002) communities in the Arafura Sea/Torres Strait significantly differ from the Coral Sea.

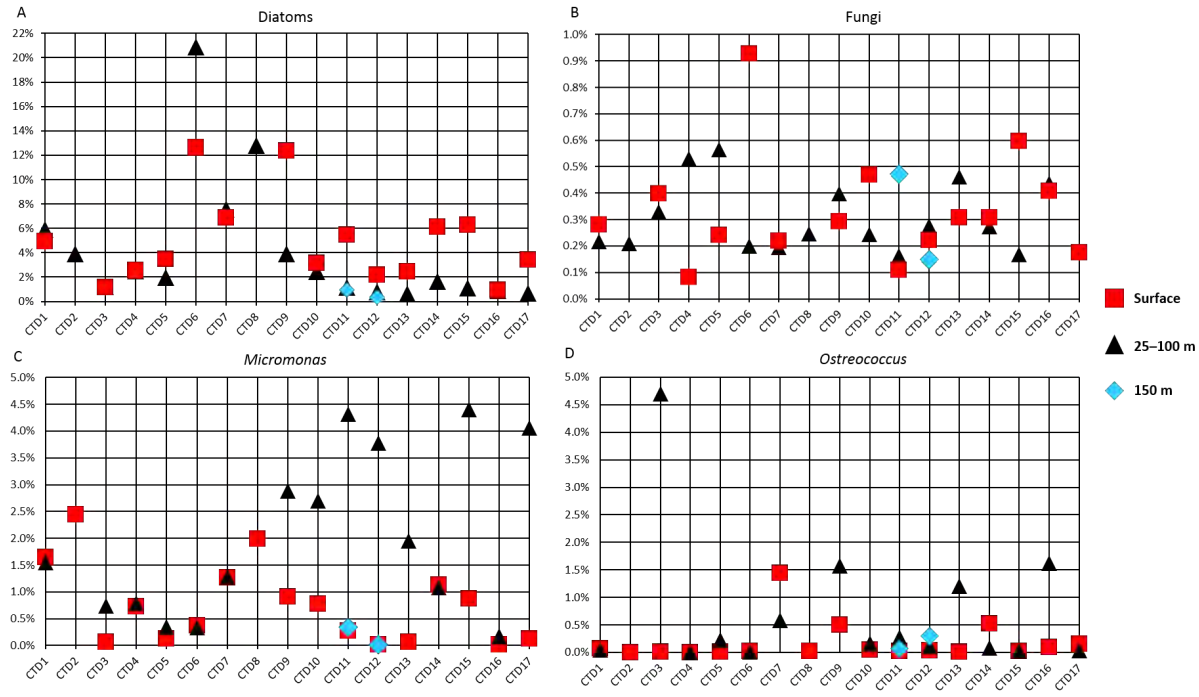

Figure 4.6 The relative abundance of diatoms (A), fungi (B), *Micromonas* (C) and *Ostreococcus* (D) in each CTD sample at surface (red), 25-120 m (black) and 150 m (blue).

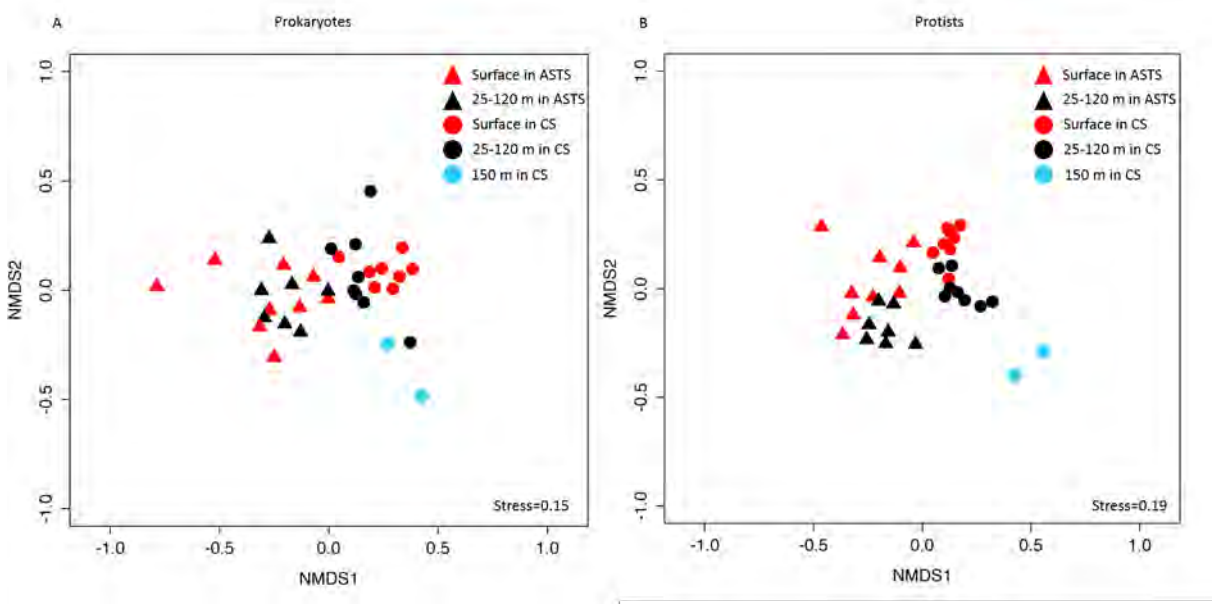

Figure 4.7 Non-metric Multidimensional Scaling (NMDS) plot showing the prokaryote (A) and protist (B) communities similarity in the upper surface layer between the Arafura Sea/Torres Strait (ASTS) and the Coral Sea (CS). The surface (5 m), 25-120 m and 150 m samples are represented in red, black and blue colour, respectively. The solid triangle represents samples from the Arafura Sea/Torres Strait. The solid circle represents samples from the Coral Sea.

In the Coral Sea, the surface samples tend to have similar prokaryotic and protistan community composition, clustering together and separating from the depth samples in the NMDS plot. For

both prokaryotes and protists, the 150 m depth samples formed a distinct outlying group, which in the case of the prokaryotes also clustered with a single sample from a depth of 120 m.

In the NMDS plot, the Arafura Sea and Torres Strait prokaryotic and protistan communities formed a distinct group separate from the Coral Sea. For the prokaryotes, the surface and depth samples from the Arafura Sea/Torres Strait were intermingled and not clearly separated. In contrast, the protistan communities from the surface and depth samples were more clearly separated on the NMDS plot.

Environmental variables, depth and geographic distance effect on community composition

We used a simple Mantel test to detect the effects of environmental variables, horizontal and vertical spatial factors on prokaryotic and protistan community variation (Table 2). The BIOENV approach was also used to find the best subset of the environmental factors based on finding maximum correlations between the Euclidean distances of scaled environmental variables and Bray-Curtis dissimilarity. Overall, salinity, phosphate and silicate were the most important environmental drivers impacting both the prokaryote and protistan community composition. However, there were regional differences. In the Arafura Sea/Torres Strait, we found the strongest subset of environmental factors impacting the prokaryote community were temperature and salinity, and for the eukaryote community this included salinity, nitrate and phosphate presence. For the samples from the Coral Sea, the best subset of environmental drivers for prokaryote community composition were salinity, nitrate and phosphate, and for the eukaryote community were temperature, phosphate and silicate.

The simple Mantel test results revealed that the prokaryotic community structure was significantly correlated with whole environment factors (r=0.33, P=0.001), best subset of environmental drivers (r=0.37, P=0.001), and the sea floor depth (r=0.23, P=0.009). The protistan community structure was only significantly correlated with the whole environment

factors (*r*=0.29, *P*=0.002) and the best subset of environmental drivers (*r*=0.32, *P*=0.001). The importance of individual environmental variables was verified by a simple Mantel test. It indicated that temperature, salinity, phosphate and silicate concentrations played a statistically significant influence on prokaryote and protistan community composition. For the Arafura Sea/Torres Strait samples, the results indicated that sea floor depth was the predominant factor (*r*=0.43 and 0.51, *P*=0.001) shaping the β-diversity of microbial prokaryotes and protistans (Table 4.2).

Table 4.2 Simple Mantel tests for the correlation between individual environmental variables, seafloor depth, geographic distance, sample depth, best subset of the environmental factors (Bestenv) and the prokaryote or protist community composition in the Arafura Sea/Torres Strait (ASTS) and the Coral Sea (CS).

	Prokaryotes						Protists					
	All		ASTS		CS		All		ASTS		CS	
	r	*P*	*r*	*P*	*r*	*p*	*r*	*P*	*r*	*P*	*r*	*P*
Temperature	0.268	**0.001**	0.046	0.357	0.484	**0.001**	0.292	**0.001**	0.231	0.118	0.197	0.059
Salinity	0.327	**0.001**	-0.020	0.481	0.559	**0.001**	0.284	**0.001**	0.319	0.081	0.377	**0.011**
Nitrate	0.161	0.081	0.244	0.078	0.300	0.072	0.004	0.383	0.056	0.361	0.045	0.279
Phosphate	0.362	**0.002**	0.071	0.302	0.521	**0.002**	0.312	**0.029**	0.362	**0.006**	0.370	**0.050**
Silicate	0.288	**0.001**	0.069	0.244	0.255	**0.044**	0.264	**0.003**	0.194	0.048	0.552	**0.001**
Ammonia	0.061	0.230	-0.021	0.506	-0.212	0.974	-0.006	0.459	0.040	0.382	-0.068	0.620
Seafloor Depth	0.229	**0.009**	0.430	**0.001**	-0.123	0.748	0.052	0.259	0.506	**0.001**	-0.146	0.856
Distance	0.141	**0.032**	-0.028	0.601	-0.047	0.660	0.066	0.195	0.035	0.371	0.014	0.424
Sample Depth	0.158	**0.05**	0.082	0.227	0.303	**0.019**	0.506	**0.001**	0.192	0.087	0.663	**0.001**
All env	0.333	**0.001**	0.234	**0.034**	0.634	**0.001**	0.294	**0.002**	0.361	**0.002**	0.322	**0.013**
Bestenv	0.368	**0.001**	0.117	**0.011**	0.586	**0.001**	0.319	**0.001**	0.356	**0.002**	0.358	**0.018**

ASTS, Arafura Sea and Torres Strait; CS, Coral Sea; All env, all environmental variables;
Bestenv, the best subset of the environmental drivers for prokaryotic and protistan (salinity, phosphate and silicate) community structure using the BIOENV approach.
Siginificant *P*-values (<0.05) are indicated in bold.

Discussion

The euphotic layer extends roughly from only a few centimetres in highly turbid waters to less than 200 m depth in most open oceanic areas. Within this zone lives nearly 90% of all marine life and almost all of marine photosynthesis occurs. We collected seawater samples from the photic zone of the Arafura Sea/Torres Strait and the Coral Sea. These basins have complex

hydrological marine environments, with events such as upwelling, coastal run-off and vertical mixing providing significant influences on the microbial community composition (Condie and Dunn 2006). We studied the prokaryotic community diversity by high-throughput sequencing of the 16S rRNA amplicons. The Shannon index results indicated that prokaryote species richness generally increases with depth in the photic zone, with the highest species richness observed at 150 m, the deepest samples in our study (Figure 3.3B).

Generally, the Australian tropical waters were dominated by cyanobacteria and proteobacteria, representing 95% of total reads (Figure 5). The marine cyanobacteria *Prochlorococcus* and *Synechococcus* are the most abundant genera of marine phytoplankton. They make a significant contribution to global chlorophyll biomass and primary production, accounting for an estimated 25% of global carbon fixation in marine ecosystems (Farrant et al 2016). In our survey, cyanobacteria comprised nearly 50% of total prokaryotic rRNA amplicons sequenced, suggesting that they play a key role in primary production in Northern Australian tropical waters, and hence provide a critical underpinning for the highly productive fisheries in this area. Cyanobacterial *Synechococcus* are generally much more abundant in low salinity and high nutrient environments (Saito et al 2005, Zwirglmaier et al 2008). This is consistent with the abundance of *Synechococcus* in our dataset, representing 32% of total 16S rRNA reads in the nutrient-rich Arafura Sea and Torres Strait, while only 9% in the oligotrophic Coral Sea. In contrast, *Prochlorococcus* was the most abundant organism in the Coral Sea, which has lower nutrient concentrations and higher salinity. Previous studies have found that *Prochlorococcus* abundance is often much higher in offshore and oligotrophic waters (Flombaum et al 2013, Partensky et al 1999). We also observed that the cyanobacterial abundance declined dramatically when sampling depths exceeded 100 m in the Coral Sea, due to the lowered levels of sunlight penetrating to those depths.

SAR11 (alphaproteobacteria) is the most abundant and ubiquitous clade of heterotrophic marine bacteria throughout the world's oceans. It accounts for a third of the cells present in surface waters and half of global prokaryotic cells located in the euphotic zone (Morris et al 2002). The SAR11 group averaged 36% of the total sequence reads across all of the sampled sites, indicating that is also an abundant member of the prokaryotic community across the Northern Australian tropical waters. It represented an extremely numerically abundant group in some specific sites, for instance, at 150 m depth at CTD11 and CTD12, it comprised 69% and 58% of the sequenced rRNA amplicons (Figure 4.4). As the dominant heterotrophs in the surface ocean, SAR11 play an important role in the marine carbon cycles, as the flux of organic carbon through bacterioplankton is comparable to daily primary production (Carlson et al 2009). They have significant ecological advantages in oligotrophic environments which stem from adaptations to nutrient poor waters including light-dependent proton pumps (Giovannoni et al 2005) and streamlined genomes (Grote et al 2012).

The SAR324 group within the deltaproteobacteria displayed a very distinct depth profile, with its highest observed abundance at a depth of 150 m (6.5% of total prokaryotic rRNA sequence reads). This group has been reported to have greater abundances in deeper waters (Bryant et al 2016) and appears to be highly metabolically flexible with the ability to utilize diverse electron acceptors such as sulfur, hydrocarbons, C_1 compounds, other organic carbon molecules and is possibly capable of using a range of alkanes as carbon sources (Sheik et al 2014).

Protists are unicellular eukaryotes that are abundant and ubiquitously distributed in oceanic waters (Sherr et al 2007). They play vital ecological roles as primary producers and consumers at and near the base of marine food webs (Caron et al 2012, Thomas et al 2012). We investigated protist and fungal assemblage compositions by using high-throughput sequencing of the V9 region of the 18S rRNA gene. Although nearly 22.4% of OTUs did not match any reference sequences in the database, our sequencing effort provided a good representation of eukaryote

species richness at both local and global scales based on rarefaction curves (Figure 4.3C). Shannon index results indicated that eukaryote diversity, like prokaryote diversity, generally increased with depth, with the exception that eukaryote diversity plummeted in samples from depths of 150m (Figure 4.3D).

In our study, the dominant protists in Australian tropical waters were the Alveolata (62.9%), Stramenopiles (14%) and Rhizara (8%) (Figure 4.5). However, this data must be interpreted cautiously as the rRNA gene copy number in different protists can vary from a few copies to thousands of copies per cell and is generally correlated with genome size (Prokopowich et al 2003). For instance, it has been reported that diatoms have around 37,000 copies of the 18S rRNA gene per cell (Godhe et al 2008), dinoflagallates have more than 12,000 copies per cell (Zhu et al 2005) and ciliates have approximately 310,000 copies number per cell (Gong et al 2013). In contrast, *Micromonas* and *Ostreococcus* have only 4 and 3 copies of the 18S rRNA gene per cell, respectively (Zhu et al 2005).

Dinoflagellates within the Alveolata showed the greatest abundance of 18S rRNA sequence reads among all surface and depth samples in our study. They are a large group of microplankton in marine and freshwater environments. Half of the dinoflagellate groups are autotrophic, while the other half of the group are heterotrophic and mixotrophs that feed upon diatoms, bacteria and flagellates (Dale 2009, Naustvoll 2000). Our results showed that the amount of dinoflagellates decreased as depth increased, potentially linked to lower sea temperature and light radiation (Godhe et al 2001). Marine Alveolate (MALV) I and II within the Syndiniales were also abundant in our results. MALV-II has a wider host spectrum, including radiolarians, cilates, copepods, fish and crabs (Massana 2011), but MALV-I seems to only parasitize dinoflagellates, such as *Amoebophyra* spp (Pernice et al 2016). Another major group was Rhizara, including amoeboid protists within three main groups, cercozoa, radiolarian and formaminifera. Most rhizarian sequences (95%) belonged to radiolaria, including the novel

clades RAD A and RAD B. We observed that rhizarian abundance increased with depth, showing its highest relative abundance at 150 m (Figure 4.4B). This is consistent with previous reports that they are more prevalent in subsurface and deep waters (Not et al 2008).

The phototrophic taxa diatoms are a major component of algal communities, and play a globally significant role in biogeochemical cycles and the functioning of marine food webs (Treguer et al 1995, Yool and Tyrrell 2003). They are estimated to contribute up to 45% of the total oceanic primary production of organic material (Yool and Tyrrell 2003). A unique feature of diatom cells is that their cell wall is composed of silica. So, their abundances are related to the concentration of silicate in the ocean. Some studies indicated that diatoms are common in nutrient-rich regions but have low abundance in open ocean oligotrophic areas (Armbrust 2009, Smetacek 2012). In our study, diatom abundance correlated strongly with silicate concentrations, which were highest in the Torres Strait and deeper waters of the Arafura Sea (Figure 4.6A, 4.2).

Fungi are both non-diverse and have a low abundance in many surface marine ecosystems (Le Calvez et al 2009, Lovejoy et al 2007, Massana and Pedros-Alio 2008), but have much greater abundance in deep subsurface sediments (Orsi et al 2013). They comprised only 0.3% of the total marine 18S rDNA sequences in our study (Figure 4.6B). However, fungi plays an important and diverse ecological roles in marine ecosystems, such as in detritus processing and lignocellulose degradation (Hyde et al 1998, Mann 1988).

Micromonas is a genus of small green algae with single long flagella, mitochondria and chloroplast (Manton and Parke 1960). It has been reported as the dominant photosynthetic picoeukaryote in several oceanic and coastal regions and nutrient-rich environments (Not et al 2004, Thomsen and Buck 1998, Worden et al 2009). In our study, *Micromonas* represented

more than 30% of the Chlorophyta sequences, and its relative abundance increased dramatically at depths (<100m) samples in the Coral Sea (Figure 4.6C).

Ostreococcus has been recognized as an important member of the global phytoplankton community since it was discovered in 1994 (Courties et al 1994). It is cosmopolitan in distribution and notable for its rapid growth rates. The genus includes the smallest free-living eukaryote known to date, *Ostreococcus tauri*, which can be used as an ideal model system for research on protistan genome evolution (Derelle et al 2006). *Ostreococcus* showed a very patchy distribution across the sample sites in this study, with specific sites appearing to support localized blooms of *Ostreococcus*. It averaged 0.29% of the total 18S rRNA reads across all samples, but represented 4.7% of the sequence reads at 42 m in CTD3 and 1.6% at 28 m in CTD9 (Figure 4.6D).

Based on NMDS plot results, it is interesting to find that prokaryotic and protistan community compositions were clearly separated according to the water masses and sampling depths (Figure 4.7). We investigated the relationship between the marine microbial community composition and environmental variables with a simple Mantel test (Table 4.2). Some previous studies found that temperature is the main environmental driver shaping microbial composition in the euphotic ocean layer (Sunagawa et al 2015). We found that temperature was a significant driver (P=0.001) for epipelagic microbial community composition in Australian tropical waters. The strongest association with temperature was seen in the prokaryotic community in the Coral Sea. Other studies have suggested that salinity is the major environmental determinant of microbial community composition which acts by impacting cell responses to metabolic or osmoregulation changes (Estudillo et al 2000, Lozupone and Knight 2007). We also found that salinity was a strong and significant factor influencing prokaryote (r=0.33, P=0.001) and eukaryote (r=0.28, P=0.001) community composition, again, the strongest association was seen with the prokaryotic community in the Coral Sea.

Significant changes in prokaryotic (r=0.30, P=0.019) and protistan (r=0.66, P=0.001) community composition were observed with sampling depths in the Coral Sea. An important factor is that the surface mixed layer depths in this region is typically around 60 m (within the euphotic zone) (Condie and Dunn 2006). The mixed layer produces a distinctive water mass of density-defined layers in the upper water column and its depth determines the physicochemical parameters, such as temperature, nutrients and light, resulting in a vertical stratification of microbial taxa (Hamilton et al 2008, Han et al 2014, Sunagawa et al 2015).

Nutrient availability is also correlated with the microbial community composition in our study. Phosphate was identified as a significant environmental driver on bacteria and protisan community composition by a BIOENV approach and simple Mantel test (Table 4.2). Phosphate limitation is likely an important factor in cyanobacterial and phototropic protist distribution in these regions. Silicate concentration was also an important environmental driver in our study, and is strongly associated with relative abundance of diatoms (Figure 4.6A, 4.2). The silicate compound is essential for diatoms as their cell wall is made of silica. Diatom abundance can subsequently impact the number and distribution of heterotrophic protists, such as dinoflagellates and ciliates.

Interestingly, the seafloor depths of sampling sites in the Arafura Sea/Torres Strait was identified as an important factor influencing marine microbial community structure, especially for eukaryote community composition (r=0.51, P=0.001). The shallow seafloor in the Arafura Sea/Torres Strait probably influences microbial community composition through episodic events causing higher nutrient fluxes. For example, there is a seasonal coastal upwelling from the Banda Sea slope during the southeast monsoon (June to November) that persistently pumps high-nutrient water mass to this area (Moore et al 2003) and induces undercurrents running opposite to the wind-induced surface flow, resulting in turbidity at the benthic boundary layer

(from the seafloor to 30 m above the bottom) across the entire region (Kämpf 2015), which suggest that the distance from the seafloor to the sampling depth might also contribute to the difference in community composition. Such nutrient fluxes support a rich, highly productive marine ecosystem in the Arafura Sea with an estimated productivity of >300 C m^2 yr^{-1} (Sherman 2008).

In conclusion, this study provides the first insight into general diversity patterns of marine prokaryote and protist communities in the eutrophic layer of the Arafura Sea/Torres Strait and the Coral Sea using high-throughput sequencing of the 16S and 18S rRNA genes. The results showed cyanobacteria and proteobacteria dominated the microbial prokaryotic communities in all samples, and few groups in the SAR (Alveolata, Stramenopiles and Rhizara) superphylums dominated protist community. Generally, the relative abundance of photosynthetic taxa (e.g. *Synechococcus*, *Prochlorococcus* and diatoms) decreased, whereas the abundance of heterotrophic groups (e.g. SAR11, SAR324 and rhizaria) increased with depth, especially when sampling depths exceeded 100 m. We found that the local environmental variables (such as salinity, temperature and nutrient availability) have a significant impact on marine protist and prokaryote community composition in the upper layers of Northern Australian tropical waters. The shallow seafloor depths and turbid seawaters in the Arafura Sea/Torres Strait were identified as factors strongly influencing the marine microbial community composition. In addition, the surface mixed layer depths in the Coral Sea also contributed to the vertical distribution of marine microbial taxa. Our findings provide the first view of marine microbial diversity and biogeographic patterns in Australian tropical epipelagic waters.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We thank the crew of the *RV* Southern Surveyor and Associate Professor Martina Doblin, the Voyage leader for SS2012_t07. This work was supported by ARC DP110102718, ARC DP150102326, ARC FL140100021, CSC scholarship to T. Huang.

Author Contributions:

TH and IP conceived the study, TH conducted experiments and analyzed data, MO, MB, LM and JS collected samples and extracted DNA, TH and IP wrote the manuscript with contributions from all other authors.

References

Alongi DM, Christoffersen P, Tirendi F, Robertson AI (1992). The influence of fresh-water and material export on sedimentary facies and benthic processes within the Fly Delta and adjacent Gulf of Papua (Papua New Guinea). *Cont Shelf Res* **12:** 287-326.

Amaral-Zettler LA, McCliment EA, Ducklow HW, Huse SM (2009). A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *Plos One* **4**.

Armbrust EV (2009). The life of diatoms in the world's oceans. *Nature* **459:** 185-192.

Azam F, Fenchel T, Field JG, Gray JS, Meyerreil LA, Thingstad F (1983). The ecological role of water-column microbes in the sea. *Mar Ecol Prog Ser* **10:** 257-263.

Azam F, Malfatti F (2007). Microbial structuring of marine ecosystems. *Nat Rev Microbiol* **5:** 782-791.

Baas-Becking LGM (1934). Geobiologie of inleiding tot de milieukunde. *WP Van Stockum and Zoom (in Dutch): The Hague, the Netherlands*.

Bryant JA, Aylward FO, Eppley JM, Karl DM, Church MJ, DeLong EF (2016). Wind and sunlight shape microbial diversity in surface waters of the North Pacific Subtropical Gyre. *ISME Journal* **10:** 1308-1322.

Burford MA, Rothlisberg PC (1999). Factors limiting phytoplankton production in a tropical continental shelf ecosystem. *Estuar Coast Shelf S* **48:** 541-549.

Carlson CA, Morris R, Parsons R, Treusch AH, Giovannoni SJ, Vergin K (2009). Seasonal dynamics of SAR11 populations in the euphotic and mesopelagic zones of the northwestern Sargasso Sea. *ISME Journal* **3:** 283-295.

Caron DA, Countway PD, Jones AC, Kim DY, Schnetzer A (2012). Marine protistan diversity. *Annu Rev Mar Sci* **4:** 467-493.

Chase JM, Myers JA (2011). Disentangling the importance of ecological niches from stochastic processes across scales. *Philos T R Soc B* **366:** 2351-2363.

Clarke KR, Ainsworth M (1993). A method of linking multivariate community structure to environmental variables. *Mar Ecol Prog Ser* **92:** 205-219.

Condie SA, Dunn JR (2006). Seasonal characteristics of the surface mixed layer in the Australasian region: implications for primary production regimes and biogeography. *Mar Freshwater Res* **57:** 569-590.

Condie SA (2011). Modeling seasonal circulation, upwelling and tidal mixing in the Arafura and Timor Seas. *Cont Shelf Res* **31:** 1427-1436.

Courties C, Vaquer A, Troussellier M, Lautier J, Chretiennotdinet MJ, Neveux J *et al* (1994). Smallest eukaryotic organism. *Nature* **370:** 255-255.

Dale B (2009). Eutrophication signals in the sedimentary record of dinoflagellate cysts in coastal waters. *J Sea Res* **61:** 103-113.

de Vargas C, Audic S, Henry N, Decelle J, Mahe F, Logares R *et al* (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**.

Derelle E, Ferraz C, Rombauts S, Rouze P, Worden AZ, Robbens S *et al* (2006). Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proceedings of the National Academy of Sciences of the United States of America* **103:** 11647-11652.

Diez B, Massana R, Estrada M, Pedros-Alio C (2004). Distribution of eukaryotic picoplankton assemblages across hydrographic fronts in the Southern Ocean, studied by denaturing gradient gel electrophoresis. *Limnol Oceanogr* **49:** 1022-1034.

Dumbrell AJ, Nelson M, Helgason T, Dytham C, Fitter AH (2010). Relative roles of niche and neutral processes in structuring a soil microbial community. *ISME Journal* **4:** 337-345.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27:** 2194-2200.

Estudillo CB, Duray MN, Marasigan ET, Emata AC (2000). Salinity tolerance of larvae of the mangrove red snapper (*Lutjanus argentimaculatus*) during ontogeny. *Aquaculture* **190:** 155-167.

Falkowski PG, Fenchel T, Delong EF (2008). The microbial engines that drive Earth's biogeochemical cycles. *Science* **320:** 1034-1039.

Farrant GK, Dore H, Cornejo-Castillo FM, Partensky F, Ratin M, Ostrowski M *et al* (2016). Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America* **113:** E3365-E3374.

Flombaum P, Gallegos JL, Gordillo RA, Rincon J, Zabala LL, Jiao NAZ *et al* (2013). Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proceedings of the National Academy of Sciences of the United States of America* **110:** 9824-9829.

Foulon E, Not F, Jalabert F, Cariou T, Massana R, Simon N (2008). Ecological niche partitioning in the picoplanktonic green alga Micromonas pusilla: evidence from environmental surveys using phylogenetic probes. *Environ Microbiol* **10:** 2433-2443.

Giovannoni SJ, Britschgi TB, Moyer CL, Field KG (1990). Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **345:** 60-63.

Giovannoni SJ, Bibbs L, Cho JC, Stapels MD, Desiderio R, Vergin KL *et al* (2005). Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature* **438:** 82-85.

Godhe A, Noren F, Kuylenstierna M, Ekberg C, Karlson B (2001). Relationship between planktonic dinoflagellate abundance, cysts recovered in sediment traps and environmental factors in the Gullmar Fjord, Sweden. *J Plankton Res* **23:** 923-938.

Godhe A, Asplund ME, Harnstrom K, Saravanan V, Tyagi A, Karunasagar I (2008). Quantification of diatom and dinoflagellate biomasses in coastal marine seawater samples by real-time PCR. *Appl Environ Microb* **74:** 7174-7182.

Gong J, Dong J, Liu XH, Massana R (2013). Extremely high copy numbers and polymorphisms of the rdna operon estimated from single cell analysis of oligotrich and peritrich ciliates. *Protist* **164:** 369-379.

Grote J, Thrash JC, Huggett MJ, Landry ZC, Carini P, Giovannoni SJ *et al* (2012). Streamlining and core genome conservation among highly divergent members of the SAR11 clade. *Mbio* **3**.

Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner 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 Res* **41:** D597-D604.

Halpern BS, Walbridge S, Selkoe KA, Kappel CV, Micheli F, D'Agrosa C *et al* (2008). A global map of human impact on marine ecosystems. *Science* **319:** 948-952.

Hamilton AK, Lovejoy C, Galand PE, Ingram RG (2008). Water masses and biogeography of picoeukaryote assemblages in a cold hydrographically complex system. *Limnol Oceanogr* **53:** 922-935.

Han D, Kang I, Ha HK, Kim HC, Kim OS, Lee BY *et al* (2014). Bacterial communities of surface mixed layer in the pacific sector of the western Arctic Ocean during sea-ice melting. *Plos One* **9**.

Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH (2012). Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* **10:** 497-506.

Harris PT (1988). Sediments, bedforms and bedload transport pathways on the continental-shelf adjacent to Torres Strait, Australia Papua New Guinea. *Cont Shelf Res* **8:** 979-1003.

Harris PT (1999). Environmental management of Torres Strait: a marine geologist's perspective. In: Gostin VA (ed). *Gondwana to Greenhouse: Environmental Geoscience—an Australian perspective*. pp 149-160.

Hemer MA, Harris PT, Coleman D, Hunter J (2004). Sediment mobility due to currents and waves in the Torres Strait Gulf of Papua region. *Cont Shelf Res* **24:** 2297-2316.

Hyde KD, Jones EBG, Leano E, Pointing SB, Poonyth AD, Vrijmoed LLP (1998). Role of fungi in marine ecosystems. *Biodivers Conserv* **7:** 1147-1161.

Jongsma D (1974). Marine geology of the Arafura Sea. *Bureau of Mineral Resources, Geology and Geophysics, Canberra* **157:** 1-73.

Kämpf J (2015). Undercurrent-driven upwelling in the northwestern Arafura Sea. *Geophys Res Lett* **42:** 9362-9368.

Kirkham AR, Lepere C, Jardillier LE, Not F, Bouman H, Mead A *et al* (2013). A global perspective on marine photosynthetic picoeukaryote community structure. *ISME Journal* **7:** 922-936.

Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microb* **79:** 5112-5120.

Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR (1985). Rapid-determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proceedings of the National Academy of Sciences of the United States of America* **82:** 6955-6959.

Le Calvez T, Burgaud G, Mahe S, Barbier G, Vandenkoornhuyse P (2009). Fungal diversity in deep-sea hydrothermal ecosystems. *Appl Environ Microb* **75:** 6415-6421.

Logares R, Audic S, Bass D, Bittner L, Boutte C, Christen R *et al* (2014). Patterns of rare and abundant marine microbial eukaryotes. *Curr Biol* **24:** 813-821.

Lovejoy C, Vincent WF, Bonilla S, Roy S, Martineau MJ, Terrado R *et al* (2007). Distribution, phylogeny, and growth of cold-adapted picoprasinophytes in Arctic Seas. *J Phycol* **43:** 78-89.

Lozupone CA, Knight R (2007). Global patterns in bacterial diversity. *Proceedings of the National Academy of Sciences of the United States of America* **104:** 11436-11440.

Mann KH (1988). Production and use of detritus in various fresh-water, estuarine, and coastal marine ecosystems. *Limnol Oceanogr* **33:** 910-930.

Manton I, Parke M (1960). Further observations on small green flagellates with special reference to possible relatives of Chromulina pusilla Butcher. *Journal of the Marine Biological Association of the United Kingdom* **39:** 275-298.

Margvelashvili N, Saint-Cast F, Condie S (2008). Numerical modelling of the suspended sediment transport in Torres Strait. *Cont Shelf Res* **28:** 2241-2256.

Martiny JB, Eisen JA, Penn K, Allison SD, Horner-Devine MC (2011). Drivers of bacterial beta-diversity depend on spatial scale. *Proceedings of the National Academy of Sciences of the United States of America* **108:** 7850-7854.

Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL *et al* (2006). Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* **4:** 102-112.

Massana R, Pedros-Alio C (2008). Unveiling new microbial eukaryotes in the surface ocean. *Curr Opin Microbiol* **11:** 213-218.

Massana R (2011). Eukaryotic picoplankton in surface oceans. *Annual Review of Microbiology* **65:** 91-110.

Minchin PR (1987). An evaluation of the relative robustness of techniques for ecological ordination. *Vegetatio* **69:** 89-107.

Monier A, Comte J, Babin M, Forest A, Matsuoka A, Lovejoy C (2015). Oceanographic structure drives the assembly processes of microbial eukaryotic communities. *ISME Journal* **9:** 990-1002.

Moore TS, Marra J, Alkatiri A (2003). Response of the Banda Sea to the southeast monsoon. *Mar Ecol Prog Ser* **261:** 41-49.

Morris RM, Rappe MS, Connon SA, Vergin KL, Siebold WA, Carlson CA *et al* (2002). SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420:** 806-810.

Naustvoll LJ (2000). Prey size spectra and food preferences in thecate heterotrophic dinoflagellates. *Phycologia* **39:** 187-198.

Nemergut DR, Costello EK, Hamady M, Lozupone C, Jiang L, Schmidt SK *et al* (2011). Global patterns in the biogeography of bacterial taxa. *Environ Microbiol* **13:** 135-144.

Not F, Latasa M, Marie D, Cariou T, Vaulot D, Simon N (2004). A single species, *Micromonas pusilla* (Prasinophyceae), dominates the eukaryotic picoplankton in the western English channel. *Appl Environ Microb* **70:** 4064-4072.

Not F, Latasa M, Scharek R, Viprey M, Karleskind P, Balague V *et al* (2008). Protistan assemblages across the Indian Ocean, with a specific emphasis on the picoeukaryotes. *Deep-Sea Res Pt I* **55:** 1456-1473.

Oksanen J, Blanchet FG, Kindit R, Legendre P, Minchin PR, O'Hara RB *et al* (2015). Vegan: community ecology package.

Orsi W, Biddle JF, Edgcomb V (2013). Deep sequencing of subseafloor eukaryotic rRNA reveals active fungi across marine subsurface provinces. *Plos One* **8**.

Partensky F, Hess WR, Vaulot D (1999). *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol R* **63:** 106-127.

Pernice MC, Giner CR, Logares R, Perera-Bel J, Acinas SG, Duarte CM *et al* (2016). Large variability of bathypelagic microbial eukaryotic communities across the world's oceans. *ISME Journal* **10:** 945-958.

Prokopowich CD, Gregory TR, Crease TJ (2003). The correlation between rDNA copy number and genome size in eukaryotes. *Genome* **46:** 48-50.

Saito MA, Rocap G, Moffett JW (2005). Production of cobalt binding ligands in a *Synechococcus* feature at the Costa Rica upwelling dome. *Limnol Oceanogr* **50:** 279-290.

Salazar G, Cornejo-Castillo FM, Benitez-Barrios V, Fraile-Nuez E, Alvarez-Salgado XA, Duarte CM *et al* (2015). Global diversity and biogeography of deep-sea pelagic prokaryotes. *ISME journal*.

Schnetzer A, Moorthi SD, Countway PD, Gast RJ, Gilg IC, Caron DA (2011). Depth matters: microbial eukaryote diversity and community structure in the eastern North Pacific revealed through environmental gene libraries. *Deep-Sea Res Pt I* **58:** 16-26.

Sheik CS, Jain S, Dick GJ (2014). Metabolic flexibility of enigmatic SAR324 revealed through metagenomics and metatranscriptomics. *Environ Microbiol* **16:** 304-317.

Sherman K, Hempel, G. (2008). The UNEP large marine ecosystem report: a perspective on changing conditions in LMEs of the world's regional seas. *UNEP Regional Seas Report and Studies No 182 United Nations Environment Programme Nairobi, Kenya*

Sherr BF, Sherr EB, Caron DA, Vaulot D, Worden AZ (2007). Oceanic protists. *Oceanography* **20:** 130-134.

Sloan WT, Lunn M, Woodcock S, Head IM, Nee S, Curtis TP (2006). Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environ Microbiol* **8:** 732-740.

Smetacek V (2012). Making sense of ocean biota: how evolution and biodiversity of land organisms differ from that of the plankton. *J Biosciences* **37:** 589-607.

Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G *et al* (2015). Structure and function of the global ocean microbiome. *Science* **348**.

Thomas MC, Selinger LB, Inglis GD (2012). Seasonal diversity of planktonic protists in southwestern Alberta Rivers over a 1-year period as revealed by terminal restriction fragment length polymorphism and 18S rRNA gene library analyses. *Appl Environ Microb* **78:** 5653-5660.

Thomsen HA, Buck KR (1998). Nanoflagellates of the central California waters: taxonomy, biogeography and abundance of primitive, green flagellates (Pedinophyceae, Prasinophyceae). *Deep-Sea Res Pt Ii* **45:** 1687-1707.

Tomczak MJSG (2003). Regional oceanography: an introduction. *Pergamon Press*.

Treguer P, Nelson DM, Vanbennekom AJ, Demaster DJ, Leynaert A, Queguiner B (1995). The silica balance in the world ocean: a reestimate. *Science* **268:** 375-379.

Westgate MJ, Barton PS, Lane PW, Lindenmayer DB (2014). Global meta-analysis reveals low consistency of biodiversity congruence relationships. *Nat Commun* **5**.

Winsley T, van Dorst JM, Brown MV, Ferrari BC (2012). Capturing greater 16S rRNA gene sequence diversity within the domain bacteria. *Appl Environ Microb* **78:** 5938-5941.

Wolanski E (1993). Water circulation in the Gulf of Carpentaria. *J Marine Syst* **4:** 401-420.

Wolanski E, King B, Galloway D (1995). Dynamics of the turbidity maximum in the Fly River estuary, Papua New Guinea. *Estuar Coast Shelf S* **40:** 321-337.

Worden AZ, Lee JH, Mock T, Rouze P, Simmons MP, Aerts AL *et al* (2009). Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *micromonas*. *Science* **324:** 268-272.

Yool A, Tyrrell T (2003). Role of diatoms in regulating the ocean's silicon cycle. *Global Biogeochem Cy* **17**.

Yooseph S, Nealson KH, Rusch DB, McCrow JP, Dupont CL, Kim M *et al* (2010). Genomic and functional adaptation in surface ocean planktonic prokaryotes. *Nature* **468:** 60-66.

Zhu F, Massana R, Not F, Marie D, Vaulot D (2005). Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *Fems Microbiol Ecol* **52:** 79-92.

Zwirglmaier K, Jardillier L, Ostrowski M, Mazard S, Garczarek L, Vaulot D *et al* (2008). Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environ Microbiol* **10:** 147-161.

Chapter 5:

General discussion and future directions

5.1 General discussion from this study

The tropical marine environments of northern Australia present a diverse range of geomorphological and oceanographic conditions, ranging from the well mixed and nutrient rich shallow waters of the Arafura Sea and Torres Strait, through to the open ocean oligotrophic waters of the Coral Sea. These waters can be highly productive, and have been identified as a hotspot for nitrogen fixation, however, the diversity of marine microbes inhabiting these regions has not been investigated systematically using molecular techniques. In this thesis, we collected samples from the surface and euphotic zone of this region during an oceanographic transect in October 2012 prior to the start of the tropical wet season. The Arafura Sea is shallow (50-80 m) semi-enclosed continental shelf basin, between the Northern Australian and Indonesian landmass. It is fully tropical and experiences relatively stable trade winds during part of the year and fitful monsoons (Jongsma 1974). It has been classified as one of the most pristine marine environments on the planet (Condie 2011). The Torres Strait (7-10 m) is an important international shipping lane, which links the Arafura Sea and the Coral Sea. The Coral Sea is an open sea, which includes the world's largest coral reef system. The nutrient concentrations and primary production are typically very low in the euphotic zone of this basin (Condie and Dunn 2006).

In chapter 2, we determined the abundance of marine cyanobacteria *Synechococcus* and *Prochlorococcus* in the sampling regions by flow cytometry using the fluorescence of natural photosynthetic pigments (chlorophyll, phycoerythrin). The results suggested that *Synechococcus* cell abundances were higher in the Arafura Sea/Torres Strait (ranging from 1×10^4 to 3×10^5 cells mL^{-1}) than that of Coral Sea (ranging from 1×10^3 to 6×10^4 cells mL^{-1}). It is difficult to count *Prochlorococcus* cells from the Arafura Sea/Torres Strait, perhaps due to its small cell size, low abundance and turbidity of the water. Based on the obtained data, *Prochlorococcus* cell numbers were much higher in the Coral Sea ($1\text{-}4.5\times10^5$ cells mL^{-1}) and lower in the Arafura Sea/Torres Strait ($8\times10^3\text{-}2.8\times10^4$ cells mL^{-1}). These findings were

consistent with previous studies, that *Synechococcus* is abundant in nutrient-rich regions but *Prochlorococcus* was most abundant in the warm oligotrophic waters (Blanchot et al 1997, Flombaum et al 2013, Tarran et al 1999). The parametric models predicted that the highest average abundance of *Synechococcus* and *Prochlorococcus* was 3.4×10^4 cells mL^{-1} and 2.5×10^5 cells mL^{-1} in the Indian Ocean (Flombaum et al 2013). The observed *Synechococcus* abundances in the Arafura Sea were generally higher than predicted from the modelling, with some locations consisting of highest cell counts ever reported for *Synechococcus*. The high abundance of cyanobacteria in the Northern Australian tropical waters presumably plays a key role in supporting the productive fisheries and aquaculture thus providing thousands of tons of shrimps, lobsters, crabs and snapper every year.

We report the first investigation into the prokaryote diversity in the Arafura Sea/Torres Strait and the Coral Sea. Cyanobacteria and Proteobacteria were the two most abundant groups in this study, together accounting for nearly 95% of the total 16S rRNA sequencing reads. However, their distinct subgroups showed significant correlation within different geographical regions. For example, the cyanobacterium *Synechococcus* was the largest group in the Arafura Sea/Torres Strait and the abundance decreased dramatically in the Coral Sea, which was consistent with our observed cell counts based on flow cytometry. In contrast, *Prochlorococcus* were numerically dominant in the Coral Sea, representing up to 60% of total 16S rRNA sequencing reads in some samples, such as CTD12 and CTD16. Our on-field studies showed *Synechococcus* representing the dominant phototrophic bacterium in the Arafura Sea/Torres Strait, however, *Prochlorococcus* cell numbers were much higher in the Coral Sea than in the Arafura Sea/Torres Strait based on the obtained data using flow cytometry. In some studies, the heterotrophic SAR11 clade (within Alphaproteobacteria) has been considered as the most abundant bacterial group (Morris et al 2002), and it averaged nearly 35% of the total 16S rRNA sequencing reads in the Coral Sea.

Previous studies suggest that the bacterial diversity is higher in oligotrophic waters than in eutrophic waters (Villaescusa et al 2010). We have similar results, prokaryotic community diversity showed higher diversity within the oligotrophic Coral Sea compared with the more nutrient-rich Arafura Sea/Torres Strait waters. The NMDS ordination illustrates the split between the Arafura Sea/Torres Strait samples and the Coral Sea samples. The surface samples in the Arafura Sea/Torres Strait appeared to have a similar community structure, while samples in the Coral Sea showed a higher variability. Salinity was one of the strong environmental drivers affecting the prokaryotic community composition in the Coral Sea. *Prochlorococcus* were numerically dominant in this area, and their abundance has previously been positively correlated with salinity (Crosbie and Furnas 2001). Temperature, nitrate, phosphate and Chl *a* concentration were strongly positively ($P<0.01$) correlated with the prokaryotic community structure in the Arafura Sea/Torres Strait, which has a shallow seafloor depth. Nutrient availability in shallow or coastal waters has been shown to have a major impact on phytoplankton composition patterns (Ling et al 2015).

In chapter 3, we investigated the cellular abundances of plastidic protists by flow cytometry using the fluorescence of natural photosynthetic pigments. In previous marine studies, their typical abundances are $1\text{-}3\times10^3$ cells mL^{-1} in oligotrophic systems and up to 10^5 cells mL^{-1} in coastal and nutrient-rich regions (Li 2009, Sanders et al 2000). High photosynthetic picoeukaryote concentrations were positively correlated with high chlorophyll *a* in surface waters in the north Atlantic Ocean and Arctic Ocean (Kirkham et al 2013). In this chapter, similar results were reported, with the average plastidic protist cell concentration in surface water being considerably higher in the nutrient-rich Arafura Sea/Torres Strait (up to 4.2×10^3 cells mL^{-1}) than in the oligotrophic Coral Sea (up to 1,465 cells mL^{-1}). The cell abundances of the picocyanobacteria *Synechococcus* and *Prochlorococcus* were generally higher by 1-2 orders of magnitude.

Eukaryotic assemblages were dominated by the SAR supergroups (Stramenopiles, Alveolata and Rhizaria), small Metazoa and Archaeplastida, together representing ~90% of total 18S rRNA sequencing reads in the surface water of the Northern Australian trophic waters. However, 18S rRNA data needs to be interpreted cautiously, as the relative abundance of rRNA gene in libraries may not reflect the organismal abundance in nature due to rRNA gene copy numbers varying over a very broad range in different marine eukaryotic groups, which can lead to over- or under-estimation of their abundance. Although the Arafura Sea/Torres Strait and the Coral Sea shared the same 9 most abundant groups, the relative abundances of some groups showed distinct biogeographic patterns, such as diatoms representing 5% of total 18S rRNA sequencing reads in the nutrient-rich Arafura Sea/Torres Strait, while only 1% of total 18S rRNA sequencing reads in the oligotrophic Coral Sea. Dinoflagellates (Alveolata) were the most abundant group in the Coral Sea, occupying 21% of the total 18S rRNA sequencing reads. This is possibly because the Coral Sea harbors the world's largest reef system, the Great Barrier Reef, and some dinoflagellates (e.g. *Symbiodinium*) are symbiotes of coral reefs and provide metabolites from photosynthesis to support coral metabolism, growth, reproduction and skeletogenesis (Davies 1991).

The differences in marine eukaryotic community composition in the surface water were largely explained by temperature, salinity and nutrient concentrations. Salinity and temperature have been reported as significant drivers for the marine microeukayotic plankton (Lima-Mendez et al 2015, Sunagawa et al 2015). Nutrient variables had a significant impact on marine eukaryotic communities because they are well adapted to preferable nutritional conditions (Ferrierpages and Rassoulzadegan 1994). Silicate concentration is a limitation for diatom abundance due to their unique silica-containing cell walls. We established that their abundance has a strong correlation with silicate concentration in the Arafura Sea/Torres Strait.

In chapter 4, we investigated the marine microbial community diversity at various depths within the euphotic zone (<200 m) of the Northern Australian tropics. Not surprisingly, at depths from 25-100m, the marine microbial community composition had similar patterns to their structure on the surface. Throughout all the samples *Synechococcus*, *Prochlorococcus* and SAR11 clade dominated prokaryote assemblages. Specific groups within the SAR supergroups, such as dinoflagellates, Marine Alveolate (MALV) I, MALV II, retaria and diatoms, dominated the protist community composition based on 18S rRNA read counts. However, their abundances showed different geographic and vertical distribution. For example, cyanobacteria abundance dropped significantly and the heterotrophic SAR11 group gradually dominated the prokaryotic community with respect to an increase in sampling depth, especially at depths over 100 m in the Coral Sea. This very likely correlates with light radiation and nutrient availability with increasing water depth in the ocean. The amount of dinoflagellates decreased as depth increased, this is potentially linked to low sea temperature and light radiation (Godhe et al 2001). Generally, the phototropic plankton dominated at the upper surface layer of sampling sites and their abundance slowly decreased with depth, but the relative abundance of heterotrophic, mixotrophic or parasitic marine microbes showed an opposite trend.

Statistical analysis suggested that the prokaryotic and eukaryotic community compositions were clearly separated based on the water masses and sampling depth, and strongly shaped by local environmental variables (salinity, temperature and nutrient availabilities). In addition, the seafloor depths of sampling sites in the Arafura Sea/Torres Strait were identified as an important factor influencing marine microbial community structure, especially the community composition of eukaryotes. This may be explained by the particular hydrodynamics in this basin and the sampling depths being close to the seafloor. High-nutrient water mass was pumped into this area by the seasonal coastal upwelling of Banda sea slope during the southeast monsoon season (June to September) (Moore et al 2003). This causes undercurrents on the seafloor, increasing turbidity and nutrient concentrations in seawaters in this area. This phenomena likely

significantly impacts the microbial community composition, especially the heterotrophic eukaryotic community diversity (Condie and Dunn 2006).

5.2 Future directions

5.2.1 Further sampling in the Northern Australian tropical waters

One of the limitations of this work was that we only investigated the marine microbial community composition in the Arafura Sea, Torres Strait and Coral Sea at a single time point (October 2012).

These sampling areas are fully tropical and frequently hit by tropical cyclones, and experiencing relatively stable trade winds during part of the year and fitful monsoons typically between November and April (Jongsma 1974). Further sampling of seawaters across the seasons in these regions will help us understand how seasonal changes affect the marine microbial community structure, and how these important microbial communities adapt to the local complex hydrological marine environments (Adamczyk and Shurin 2015, Vague et al 2008).

Long-term time series sampling is crucial for studying the differentiation of important marine microbial communities over years, as well as the changes in local environmental parameters. Information on the diversity, distributional dynamics and environmental interactions of marine microorganisms can help us to monitor the responses of local marine ecosystems and thus predict future climate changes, and provide suggestions for the management of local fisheries and aquaculture in the Northern Australian tropical waters.

The Coral Sea harbours the world's largest reef system, the Great Barrier Reef (GBR), which is a popular destination for tourists, contributing $3 billion per year. Climate change is considered as the greatest threat to the GBR, causing coral bleaching due to substantially

warmer ocean temperatures, which will inevitably cause loss of reef habitat and biodiversity (Brown 1997, Hoegh-Guldberg et al 2007). Future investigation on the interactions between corals and local marine microbial communities, and the relationships between corals and environment variables (such as nutrient availability and temperature) in this area is becoming necessary and urgent. It is also important to understand the effects of increasing seawater temperature and bleaching on the functional roles of the coral microbial communities and their holobiont, including *Symbiodinium*, bacteria, viruses, fungi, archaea and endolithic algae (Littman et al 2011, Reshef et al 2006). This will help us to better understand the coral ecosystems and provide advice on the protection of coral heath during thermal stress.

5.2.2 Culturing representative members of the marine microbial community

Culture-based studies of marine microbial communities provide valuable information for predictive models. Flow cytometry can be applied to sort the representative groups of marine microbial communities from the fresh seawaters based on the fluorescence from their photosynthetic pigment content or from stained dyes. The sorted populations can be cultured in liquid medium for further investigation, such as, analysing the particular clades, ecotypes or strains within a species. We can also sort single-cell targeted marine microbial community (e.g. *Synechococcus*, *Prochlorococcus* or photosynthetic picoeukaryotes) from natural seawater samples onto agar plate. The single colony growing on the plate represents one pure species, which can be subcultured in liquid medium for future analysis.

Additional studies on the representative cultured members of microbial communities, such as utilization of “omics” technologies, will allow us to reassess microbial ecology theories by linking genetic and functional properties of microbial communities, and relating taxonomic and functional diversity to comprehend ecosystem stability (Schneider and Riedel 2010). The cultured representative groups in our lab can be used to explore the influence of different temperatures, salinity or nutrient limitations on their gene expression by proteomics and

transcriptomics. For example, previous proteomics studies suggested that the growth and photosynthesis in *Synechococcus* strain WH8102 was significantly decreased at low temperatures (Varkey et al 2016). This work will help us understand their responses to environmental pressures and acclimation strategies in natural conditions.

5.2.3 Exploring the ecology of marine microbial community using 'omics' based-tools

5.2.3.1 Metagenomics

Although we investigated the diversity of marine microbes in the Northern Australian tropical waters, the relationship between microbial communities and ecosystem function, and how microbes and/or their genes interact in various environmental niches have not been surveyed. Metagenomics is a powerful experimental approach to explore and compare the ecology and metabolic profiling of the complex microbial communities by direct sequencing of environmental genomic DNA (Simon and Daniel, 2011).

Studying metagenomics of marine phytoplankton has revealed habitat-dependent distribution of taxa and gene families, in part shaped by the biogeochemical dynamics characterizing each environment (Biddle et al 2008, Dinsdale et al 2008, Ghai et al 2010). The sequence-based strategy has found many genes encoding novel enzymes, such as dimethylsulfoniopropionate-degrading enzymes (*dmdA*), nitrite reductases (*NirK*), chitinases, glycerol dehydratases, and hydrazine oxidoreductases (*hzo*) (Bartossek et al 2010, Knietsch et al 2003, Li et al 2010, Simon and Daniel 2011).

Further metagenomic studies on marine microbial communities in the Australian tropical waters should provide insights into the functional potential of these micro-organisms in the

environment and identify which genes are playing a fundamental role during the ecological processes in the environment, which helps us understand the metabolic strategies they use in distinct ecological niches.

5.2.3.2 Single-cell genomics

Single-cell genomics provides new views to our understanding of genetics by allowing the study of individual cells using omics approaches. At present, it allows us to identify and assemble the genomes of microorganisms, and dissect the contributions of individual cells to the biology of ecosystems (Gawad et al 2016). Previous studies revealed the distinct organismal interactions of uncultivated individual marine microbe cell by single-cell genomics. For example, the globally abundant marine cyanobacterium *Prochlorococcus* are composed of hundreds of subpopulations with distinct "genomic backbones," each backbone consisting of a different set of core gene alleles linked to a small distinctive set of flexible genes (Yoon et al 2011) .

I have established single-cell genomics technique in our group. Single photosynthetic picoeukaryotes (PPEs) were sorted into 96 well-plate by flow cytometry with an INFLUX cell sorter (BD Biosciences). Multiple displacement amplification (MDA) was performed followed by the thermal cell lysis. 18S rRNA gene sequencing was undertaken on successful whole genome amplification products. A high number of single cells related to *Nannochloris* spp. were identified using this approach. Whole genome sequencing was performed on successful MDA products. However, only very short contigs were obtained following de novo assembly, making further analyses problematic.

Further single-cell genomics research would not only allow us to generate significant genome data from individual photosynthetic picoeukaryotes, but could also reveal complex biotic interactions among previously uncharacterized marine microorganisms, with each cell

undergoing distinct types of interaction. It also provides the chance to reconstruct the eukaryotic tree of life, using the identified uncultivable taxa isolated directly from natural environment.

5.2.3.3 Metatranscriptomics

Metatranscriptomics provides an opportunity to gain insights into the functionality of microbial communities by analysing the community transcripts, which isolated directly from the environment. This method should reach beyond investigating the genomic potential of the community and allow for the correlation of in situ activity (function) with the specific environmental conditions (Chistoserdova 2010, Gosalbes et al 2011, Helbling et al 2012).

Previous metatranscriptomic studies found that bacterioplankton express a high abundance of transcripts encoding ammonia and phosphate transporters at a coastal site off the south-eastern USA. This is consistent with the observation of elevated concentrations of ammonia and phosphate at the survey site, highlighting the ability to correlate the expression of genes with the local environmental conditions (Gifford et al 2011). Metatranscriptomic studies have also been used to study specific metabolic pathways in the environment. These studies observed the increases in phosphate two component systems and phosphate ATP-binding cassette (ABC) transporters with enrichment of samples with dissolved organic matters (DOM) (McCarren et al 2010).

Non-coding RNAs (ncRNAs) mostly function as regulators, and in model organisms are known to control significant environmental processes such as amino acid biosynthesis and photosynthesis (Steglich et al 2008, Vogel et al 2003). Although not many ncRNAs have been functionally characterized in cyanobacteria, selected examples indicate their relevance for regulation and stress adaptation in this microbial group, such as Yfr (*c*yanobacterial *f*unctional

*R*NA) 1 to -7 in three isolations of *Prochlorococcus* and in *Synechococcus* sp. strain WH8102 (Axmann et al 2005).

Metatranscriptomics could help us to establish the foundation for comparative assessments of seasonal and annual changes in the gene expressions of prokaryotic and eukaryotic communities. This will provide insights into the regulation of biogeochemical processes and what is driving the distribution of marine microbial community in the Northern Australian tropical ocean.

5.2.3.4 Metaproteomics

In addition to the transcriptional level, gene expression can be regulated at the translational and post-transcriptional levels, thus studying the expression changes at the protein level is critical. Metaproteomics is a new approach that enables the direct observation of community protein profiles in mixed microbial assemblages (Maron et al 2007, Simon and Daniel 2011).

In our study, we found that cyanobacterium *Synechococcus* dominated in the nutrient-rich Arafura Sea/Torres Strait, while *Prochlorococcus* dominated in the nutrient-deplete Coral Sea. A comprehensive study of the Sargasso Sea surface metaproteome indicated a high abundance of proteins involved in photosynthesis, carbon fixation (e.g. enzyme ribulose 1, 5 bisphosphate carboxylase/oxygenase) and nitrogen metabolism (e.g. nitrate and nitrite transporters, cyanate hydratase, and the nitrogen regulatory protein P-II) were related to the cyanobacteria *Synechococcus* and *Prochlorococcus*. High abundance of SAR11 periplasmic substrate-binding proteins and transporters were also found, suggesting that cells attempt to maximize nutrient uptake activity and thus gain a competitive advantage in oligotrophic environments (Schneider and Riedel 2010, Sowell et al 2009).

Our previous results revealed that the abundance of distinct marine microbial communities have different distributions in the Northern Australian tropical waters. Characterization of their metaproteome in contrasted local complex environmental conditions should allow detection of proteins preferentially associated with specific local stresses (e.g. salinity, temperature or nutrient concentrations) and identification of key functional genes and metabolic pathways involved in local adaptations.

Together with the studies on the metagenome and the metatranscriptome, these approaches could provide valuable insights into the structure and physiology of how different phylogenetic groups inhabit a specific environment and their contribution to functioning of the ecosystem, which may provide suggestions for sustainable management of our environment.

5.3 References

Adamczyk EM, Shurin JB (2015). Seasonal Changes in Plankton Food Web Structure and Carbon Dioxide Flux from Southern California Reservoirs. *Plos One* **10**.

Axmann IM, Kensche P, Vogel J, Kohl S, Herzel H, Hess WR (2005). Identification of cyanobacterial non-coding RNAs by comparative genome analysis. *Genome biology* **6**.

Bartossek R, Nicol GW, Lanzen A, Klenk HP, Schleper C (2010). Homologues of nitrite reductases in ammonia-oxidizing archaea: diversity and genomic context. *Environ Microbiol* **12:** 1075-1088.

Biddle JF, Fitz-Gibbon S, Schuster SC, Brenchley JE, House CH (2008). Metagenomic signatures of the Peru Margin subseafloor biosphere show a genetically distinct environment. *Proceedings of the National Academy of Sciences of the United States of America* **105:** 10583-10588.

Blanchot J, Andre JM, Navarette C, Neveux J (1997). Picophytoplankton dynamics in the equatorial Pacific: diel cycling from flow-cytometer observations. *Cr Acad Sci Iii-Vie* **320:** 925-931.

Brown BE (1997). Coral bleaching: causes and consequences. *Coral Reefs* **16:** S129-S138.

Chistoserdova L (2010). Recent progress and new challenges in metagenomics for biotechnology. *Biotechnol Lett* **32:** 1351-1359.

Condie SA, Dunn JR (2006). Seasonal characteristics of the surface mixed layer in the Australasian region: implications for primary production regimes and biogeography. *Mar Freshwater Res* **57:** 569-590.

Condie SA (2011). Modeling seasonal circulation, upwelling and tidal mixing in the Arafura and Timor Seas. *Cont Shelf Res* **31:** 1427-1436.

Crosbie ND, Furnas MJ (2001). Abundance distribution and flow-cytometric characterization of picophytoprokaryote populations in central (17 degrees S) and southern (20 degrees S) shelf waters of the Great Barrier Reef. *J Plankton Res* **23:** 809-828.

Davies PS (1991). Effect of daylight variations on the energy budgets of shallow-water corals. *Mar Biol* **108:** 137-144.

Dinsdale EA, Edwards RA, Hall D, Angly F, Breitbart M, Brulc JM *et al* (2008). Functional metagenomic profiling of nine biomes. *Nature* **455:** 830-830.

Ferrierpages C, Rassoulzadegan F (1994). Seasonal impact of the microzooplankton on picoplankton and nanoplankton growth-rates in the northwest Mediterranean-Sea. *Mar Ecol Prog Ser* **108:** 283-294.

Flombaum P, Gallegos JL, Gordillo RA, Rincon J, Zabala LL, Jiao NAZ *et al* (2013). Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proceedings of the National Academy of Sciences of the United States of America* **110:** 9824-9829.

Gawad C, Koh W, Quake SR (2016). Single-cell genome sequencing: current state of the science. *Nat Rev Genet* **17:** 175-188.

Ghai R, Martin-Cuadrado AB, Molto AG, Heredia IG, Cabrera R, Martin J *et al* (2010). Metagenome of the Mediterranean deep chlorophyll maximum studied by direct and fosmid library 454 pyrosequencing. *ISME Journal* **4:** 1154-1166.

Gifford SM, Sharma S, Rinta-Kanto JM, Moran MA (2011). Quantitative analysis of a deeply sequenced marine microbial metatranscriptome. *ISME Journal* **5:** 461-472.

Godhe A, Noren F, Kuylenstierna M, Ekberg C, Karlson B (2001). Relationship between planktonic dinoflagellate abundance, cysts recovered in sediment traps and environmental factors in the Gullmar Fjord, Sweden. *J Plankton Res* **23:** 923-938.

Gosalbes MJ, Durban A, Pignatelli M, Abellan JJ, Jimenez-Hernandez N, Perez-Cobas AE *et al* (2011). Metatranscriptomic approach to analyze the functional human gut microbiota. *Plos One* **6**.

Helbling DE, Ackermann M, Fenner K, Kohler HPE, Johnson DR (2012). The activity level of a microbial community function can be predicted from its metatranscriptome. *ISME Journal* **6:** 902-904.

Hoegh-Guldberg O, Mumby PJ, Hooten AJ, Steneck RS, Greenfield P, Gomez E *et al* (2007). Coral reefs under rapid climate change and ocean acidification. *Science* **318:** 1737-1742.

Jongsma D (1974). Marine geology of the Arafura Sea. *Bureau of Mineral Resources, Geology and Geophysics, Canberra* **157:** 1-73.

Kirkham AR, Lepere C, Jardillier LE, Not F, Bouman H, Mead A *et al* (2013). A global perspective on marine photosynthetic picoeukaryote community structure. *ISME Journal* **7:** 922-936.

Knietsch A, Bowien S, Whited G, Gottschalk G, Daniel R (2003). Identification and characterization of coenzyme B-12-dependent glycerol dehydratase- and diol dehydratase-encoding genes from metagenomic DNA libraries derived from enrichment cultures. *Appl Environ Microb* **69:** 3048-3060.

Li M, Hong YG, Klotz MG, Gu JD (2010). A comparison of primer sets for detecting 16S rRNA and hydrazine oxidoreductase genes of anaerobic ammonium-oxidizing bacteria in marine sediments. *Appl Microbiol Biot* **86:** 781-790.

Li WKW (2009). From cytometry to macroecology: a quarter century quest in microbial oceanography. *Aquat Microb Ecol* **57:** 239-251.

Lima-Mendez G, Faust K, Henry N, Decelle J, Colin S, Carcillo F *et al* (2015). Determinants of community structure in the global plankton interactome. *Science* **348**.

Ling J, Zhang YY, Dong JD, Wang YS, Feng JB, Zhou WH (2015). Spatial variations of bacterial community and its relationship with water chemistry in Sanya Bay, South China Sea as determined by DGGE fingerprinting and multivariate analysis. *Ecotoxicology* **24:** 1486-1497.

Littman R, Willis BL, Bourne DG (2011). Metagenomic analysis of the coral holobiont during a natural bleaching event on the Great Barrier Reef. *Env Microbiol Rep* **3:** 651-660.

Maron PA, Ranjard L, Mougel C, Lemanceau P (2007). Metaproteomics: a new approach for studying functional microbial ecology. *Microb Ecol* **53:** 486-493.

McCarren J, Becker JW, Repeta DJ, Shi YM, Young CR, Malmstrom RR *et al* (2010). Microbial community transcriptomes reveal microbes and metabolic pathways associated with dissolved organic matter turnover in the sea. *Proceedings of the National Academy of Sciences of the United States of America* **107:** 16420-16427.

Moore TS, Marra J, Alkatiri A (2003). Response of the Banda Sea to the southeast monsoon. *Mar Ecol Prog Ser* **261:** 41-49.

Morris RM, Rappe MS, Connon SA, Vergin KL, Siebold WA, Carlson CA *et al* (2002). SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420:** 806-810.

Reshef L, Koren O, Loya Y, Zilber-Rosenberg I, Rosenberg E (2006). The coral probiotic hypothesis. *Environ Microbiol* **8:** 2068-2073.

Sanders RW, Berninger UG, Lim EL, Kemp PF, Caron DA (2000). Heterotrophic and mixotrophic nanoplankton predation on picoplankton in the Sargasso Sea and on Georges Bank. *Mar Ecol Prog Ser* **192:** 103-118.

Schneider T, Riedel K (2010). Environmental proteomics: analysis of structure and function of microbial communities. *Proteomics* **10:** 785-798.

Simon C, Daniel R (2011). Metagenomic analyses: past and future trends. *Appl Environ Microb* **77:** 1153-1161.

Sowell SM, Wilhelm LJ, Norbeck AD, Lipton MS, Nicora CD, Barofsky DF *et al* (2009). Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. *ISME Journal* **3:** 93-105.

Steglich C, Futschik ME, Lindell D, Voss B, Chisholm SW, Hess WR (2008). The challenge of regulation in a minimal photoautotroph: non-coding RNAs in *Prochlorococcus*. *Plos Genet* **4**.

Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G *et al* (2015). Structure and function of the global ocean microbiome. *Science* **348**.

Tarran GA, Burkill PH, Edwards ES, Woodward EMS (1999). Phytoplankton community structure in the Arabian Sea during and after the SW monsoon, 1994. *Deep-Sea Res Pt Ii* **46:** 655-676.

Varkey D, Mazard S, Ostrowski M, Tetu SG, Haynes P, Paulsen IT (2016). Effects of low temperature on tropical and temperate isolates of marine *Synechococcus*. *ISME Journal* **10:** 1252-1263.

Vaque D, Guadayol O, Peters F, Felipe J, Angel-Ripoll L, Terrado R *et al* (2008). Seasonal changes in planktonic bacterivory rates under the ice-covered coastal Arctic Ocean. *Limnology Oceanogragy* **53**: 2427-2438.

Villaescusa JA, Casamayor EO, Rochera C, Velazquez D, Chicote A, Quesada A *et al* (2010). A close link between bacterial community composition and environmental heterogeneity in maritime Antarctic lakes. *Int Microbiol* **13:** 67-77.

Vogel J, Bartels V, Tang TH, Churakov G, Slagter-Jager JG, Huttenhofer A *et al* (2003). RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res* **31:** 6435-6443.

Yoon HS, Price DC, Stepanauskas R, Rajah VD, Sieracki ME, Wilson WH *et al* (2011). Single-cell genomics reveals organismal interactions in uncultivated marine protists. *Science* **332:** 714-717.

Appendix

Appendix I: Biosafety approval letter

Office of the Deputy Vice-Chancellor (Research)

Research Office
C5C Research HUB East, Level 3, Room 324
MACQUARIE UNIVERSITY NSW 2109 AUSTRALIA

Phone **+61 (0)2 9850 4194**
Fax +61 (0)2 9850 4465
Email biosafety@mq.edu.au

6 February 2015

Professor Ian Paulsen
Department of Chemistry and Biomolecular Sciences
Faculty of Science and Engineering
Macquarie University

Dear Professor Paulsen,

Re: "Characterizing bacterial transporter, regulator and metabolic genes" (Ref: 5201401141)

NOTIFICATION OF A NOTIFIABLE LOW RISK DEALING (NLRD)

The above application has been reviewed by the Institutional Biosafety Committee (IBC) and has been approved as an NLRD, effective 6 February 2015.

This approval is subject to the following standard conditions:

- The NLRD is conducted by persons with appropriate training and experience, within a facility certified to either Physical Containment level 1 (PC1) or PC2.
- Working requiring Quarantine Containment level 2 (QC2) does not commence until the facility has been certified
- Only persons who have been *assessed by the IBC* as having appropriate training and experience may conduct the dealing. This includes persons involved in all parts of the dealing e.g. researchers, couriers and waste contractors. A copy of the IBC's record of assessment must be retained by the project supervisor.
- NLRDs classified under Part 1 of Schedule 3 of the Regulations must be conducted in a facility (or class of facilities) that is certified to at least a PC1 level **and** that is mentioned in the IBC record of assessment as being appropriate for the dealing;
- NLRDs classified under Part 2.1 of Schedule 3 of the Regulations must be conducted in a facility (or class of facilities) that is certified to at least a PC2 level **and** that is mentioned in the IBC record of assessment as being appropriate for the dealing;
- Any transport of the GMO must be conducted in accordance with the Regulator's *Guidelines for the Transport of GMOs* available at: http://www.ogtr.gov.au/internet/ogtr/publishing.nsf/Content/transport-guide-1
- A copy of the IBC's record of assessment has been attached to this approval.
- The record of assessment must be kept by the person or organisation for 8 years after the date of assessment by the IBC (regulation 13 C of the *Gene Technology Regulations 2001*).
- All NLRDs undertaken by Macquarie University will be reported to the OGTR at the end of every financial year.
- If the dealing involves organisms that may produce disease in humans, the NLRDs must be conducted in accordance with the vaccination requirements set out in the Australian Standard AS/NZS 2243.3:2010

1

- The Chief Investigator must inform the Biosafety Committee if the work with GMOs is completed or abandoned.

Please note the following standard requirements of approval:

1. Approval will be for a period of 5 years subject to the provision of annual reports. If, at the end of this period the project has been completed, abandoned, discontinued or not commenced for any reason, you are required to submit a Final Report. If you complete the work earlier than you had planned you must submit a Final Report as soon as the work is completed. Please contact the Committee Secretary at biosafety@mq.edu.au for a copy of the annual report.

A Progress/Final Report for this study will be due on: 1 February 2016

2. If you will be applying for or have applied for internal or external funding for the above project it is your responsibility to provide the Macquarie University's Research Grants Management Assistant with a copy of this email as soon as possible. Internal and External funding agencies will not be informed that you have final approval for your project and funds will not be released until the Research Grants Management Assistant has received a copy of this email.

If you need to provide a hard copy letter of approval to an external organisation as evidence that you have approval, please do not hesitate to contact the Committee Secretary at the address below.

Please retain a copy of this email as this is your formal notification of final Biosafety approval.

Yours Sincerely

Associate Professor Subramanyam Vemulpad

Chair, Macquarie University Institutional Biosafety Committee

Encl. Copy of record submitted by Macquarie University to the OGTR.

Biosafety Secretariat
Research Office
Level 3, Research Hub, Building C5C East
Macquarie University
NSW 2109 Australia

T: +61 2 9850 6848
F: +61 2 9850 4465

http://www.mq.edu.au/research

MACQUARIE UNIVERSITY
SYDNEY ~ AUSTRALIA
CRICOS Provider Number 00002J

2

Appendix II: Additional work on other publications. Contributed in flow sorting and data analysis in the development of the TraDIS methodology.

AMERICAN SOCIETY FOR MICROBIOLOGY | mBio®

OBSERVATION

Fluorescence-Based Flow Sorting in Parallel with Transposon Insertion Site Sequencing Identifies Multidrug Efflux Systems in *Acinetobacter baumannii*

Karl A. Hassan,[a] Amy K. Cain,[b,c] TaoTao Huang,[a] Qi Liu,[a] Liam D. H. Elbourne,[a] Christine J. Boinett,[b] Anthony J. Brzoska,[a] Liping Li,[a] Martin Ostrowski,[a] Nguyen Thi Khanh Nhu,[d,e] Tran Do Hoang Nhu,[d] Stephen Baker,[d,f] Julian Parkhill,[b] Ian T. Paulsen[a]

Department of Chemistry and Biomolecular Sciences, Macquarie University, Sydney, NSW, Australia[a]; Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom[b]; Liverpool School of Tropical Medicine, Malawi-Liverpool-Wellcome Trust Clinical Research Programme, Blantyre, Malawi[c]; The Hospital for Tropical Diseases, Wellcome Trust Major Overseas Programme, Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam[d]; School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Queensland, Australia[e]; Centre for Tropical Medicine, Nuffield Department of Clinical Medicine, Oxford University, Oxford, United Kingdom[f]

ABSTRACT **Multidrug efflux pumps provide clinically significant levels of drug resistance in a number of Gram-negative hospital-acquired pathogens. These pathogens frequently carry dozens of genes encoding putative multidrug efflux pumps. However, it can be difficult to determine how many of these pumps actually mediate antimicrobial efflux, and it can be even more challenging to identify the regulatory proteins that control expression of these pumps. In this study, we developed an innovative high-throughput screening method, combining transposon insertion sequencing and cell sorting methods (TraDISort), to identify the genes encoding major multidrug efflux pumps, regulators, and other factors that may affect the permeation of antimicrobials, using the nosocomial pathogen *Acinetobacter baumannii*. A dense library of more than 100,000 unique transposon insertion mutants was treated with ethidium bromide, a common substrate of multidrug efflux pumps that is differentially fluorescent inside and outside the bacterial cytoplasm. Populations of cells displaying aberrant accumulations of ethidium were physically enriched using fluorescence-activated cell sorting, and the genomic locations of transposon insertions within these strains were determined using transposon-directed insertion sequencing. The relative abundance of mutants in the input pool compared to the selected mutant pools indicated that the AdeABC, AdeIJK, and AmvA efflux pumps are the major ethidium efflux systems in *A. baumannii*. Furthermore, the method identified a new transcriptional regulator that controls expression of *amvA*. In addition to the identification of efflux pumps and their regulators, TraDISort identified genes that are likely to control cell division, cell morphology, or aggregation in *A. baumannii*.**

IMPORTANCE **Transposon-directed insertion sequencing (TraDIS) and related technologies have emerged as powerful methods to identify genes required for bacterial survival or competitive fitness under various selective conditions. We applied fluorescence-activated cell sorting (FACS) to physically enrich for phenotypes of interest within a mutant population prior to TraDIS. To our knowledge, this is the first time that a physical selection method has been applied in parallel with TraDIS rather than a fitness-induced selection. The results demonstrate the feasibility of this combined approach to generate significant results and highlight the major multidrug efflux pumps encoded in an important pathogen. This FACS-based approach, TraDISort, could have a range of future applications, including the characterization of efflux pump inhibitors, the identification of regulatory factors controlling gene or protein expression using fluorescent reporters, and the identification of genes involved in cell replication, morphology, and aggregation.**

Received 16 July 2016 **Accepted** 12 August 2016 **Published** 6 September 2016

Citation Hassan KA, Cain AK, Huang T, Liu Q, Elbourne LDH, Boinett CJ, Brzoska AJ, Li L, Ostrowski M, Nhu NTK, Nhu TDH, Baker S, Parkhill J, Paulsen IT. 2016. Fluorescence-based flow sorting in parallel with transposon insertion site sequencing identifies multidrug efflux systems in *Acinetobacter baumannii*. mBio 7(5):e01200-16. doi:10.1128/mBio.01200-16.

Editor Claire M. Fraser, University of Maryland School of Medicine

Address correspondence to Karl A. Hassan, karl.hassan@mq.edu.au.

To be effective in killing or stalling the growth of bacterial cells, antimicrobials must reach their cellular targets. For the majority of antimicrobials, these targets are in the cytoplasm, meaning that they must cross the cell envelope to induce their effects. The cell envelope is a particularly important factor for antimicrobial resistance in Gram-negative bacteria, since it includes two membrane permeability barriers with different surface chemistries, presenting significant potential to limit the accumulation of chemically diverse antimicrobial compounds (1).

In addition to preventing accumulation of antimicrobials, all bacteria employ sets of efflux pumps that mediate the active expulsion of these compounds should they cross a biological membrane (2). Many antimicrobial efflux pumps in bacteria have multidrug recognition profiles. Therefore, the increased expression of a single pump can result in resistance to a broad spectrum of antimicrobial classes. In Gram-negative bacteria, efflux pump overexpression has been shown to promote clinically significant levels of antimicrobial resistance (3). Genes encoding efflux

pumps have been identified in all bacterial genomes sequenced to date and can be found in large numbers (4). For example, strains of the opportunistic human pathogen *Acinetobacter baumannii* typically encode more than 50 putative efflux pumps, accounting for approximately 1.5% of their protein coding potential (5).

Despite their abundance, only a few transporters resembling drug efflux pumps have been experimentally characterized in most bacterial species. It can be difficult to discern which, if any, of the uncharacterized pumps could play an active role in protecting the cell against cytotoxic compounds without conducting labor-intensive experimental investigations. Furthermore, it can be even more challenging to identify the regulatory proteins that control expression of active multidrug efflux pumps. In this study, we sought to identify these proteins in *A. baumannii* by directly assessing drug accumulation within a population of more than 100,000 random transposon mutants. To this end, we applied fluorescence-activated cell sorting (FACS) in parallel with transposon-directed insertion sequencing (TraDIS) (6, 7). This novel approach, which we have named "TraDISort," was able to identify genes in *A. baumannii* that are associated with increased or decreased accumulation of ethidium bromide, a cationic quaternary ammonium derivative and a common substrate of multidrug efflux pumps.

Fluorescence-activated cell sorting to enrich for mutants displaying aberrant accumulation of ethidium. Ethidium readily intercalates into nucleic acids, whereupon its fluorescence intensity increases significantly. Consequently, ethidium is differentially fluorescent inside and outside cells, and cellular fluorescence can be used as a proxy for its cytoplasmic concentration (8). We hypothesized that when cells are treated with a subinhibitory concentration of ethidium, the ethidium concentrations in the cytoplasm of cells with defective multidrug efflux machinery should be higher than the concentration in wild-type cells at equilibrium, and conversely, the concentration in cells with overactive efflux machinery should be below that in wild-type cells To test this hypothesis, we examined populations of three isogenic strains of *A. baumannii* AB5057-UW (9) that differentially expressed AdeIJK, a major multidrug efflux pump in *A. baumannii*, which recognizes ethidium as a substrate (10, 11): (i) wild-type AB5075-UW, (ii) a mutant containing a transposon insertion in *adeJ*, and (iii) a mutant containing a transposon insertion in *adeN*, which encodes a negative regulator of *adeIJK* expression (9). When examined by flow cytometry, populations of the different cell types displayed distinct but partially overlapping fluorescence profiles that were in agreement with our predictions, i.e., the average fluorescence of the *adeJ* and *adeN* mutant populations was above and below that of the wild-type population, respectively (see Fig. S1A in the supplemental material). We repeated this experiment, using equivalent isogenic strains of *Acinetobacter baylyi* ADP1 (5), and made the same observations (see Fig. S1B). Based on these experiments, we predicted that it would be possible to use FACS to enrich cells from a large mutant pool that display differential ethidium accumulation or efflux based on their fluorescence intensity.

A mutant library containing more than 100,000 unique insertion mutants of *A. baumannii* BAL062 was generated using a Tn*5*-based custom transposon, and the insertion sites in the mutant pool were mapped by TraDIS (7). This library was treated with 40 μM ethidium bromide (1/16× MIC of the parental strain) and subjected to FACS to collect cells containing the highest concentrations of ethidium (i.e., the 2% most fluorescent cells) and cells containing the lowest concentrations of ethidium (i.e., the 2% least fluorescent cells). DNA was isolated from the selected pools of cells, and TraDIS was used to identify the chromosomal locations of the Tn*5* insertion sites in these cells (7). Transposon insertions were significantly (>2-fold change; *Q* value, <0.05) less abundant in 162 genes and more abundant in 24 genes in the low-fluorescence population and less abundant in 159 genes and more abundant in 24 genes in the high-fluorescence population compared to the input pool (see Data Set S1 in the supplemental material).

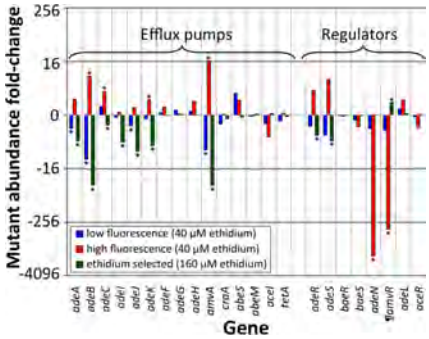

FIG 1 Selection of *A. baumannii* mutants carrying insertions in genes encoding the characterized efflux pumps AdeABC (12), AdeIJK (10), AdeFGH (20), AmvA (13, 14), CraA (21), AbeS (22), AbeM (23), and AceI (24, 25) and regulators AdeRS and BaeRS, which control expression of *adeABC* (15, 26); AdeN, which controls *adeIJK* (16); AdeL, which controls *adeFGH* (20); and AceR, which controls *aceI* (27). Bars represent the fold change in mutant abundance in cells selected for low ethidium fluorescence (blue), high ethidium fluorescence (red), or growth in 62.5 μg/ml (approximately 158 μM) ethidium bromide (hatched green; 1/4× MIC) compared to the starting mutant pool. Positive values indicate higher mutant abundance in the selected pool, whereas negative values indicate lower abundance. Asterisks indicate values supported by a *Q* value of 0.05 or below. ¶, the gene named here as *amvR* encodes a TetR family regulator that represses *amvA* gene expression (see text for details).

FACS in parallel with TraDIS identifies the active ethidium efflux pumps encoded by *A. baumannii* and core efflux pump regulators. Following the experiments with targeted mutants, we hypothesized that many cells containing the highest concentrations of ethidium would have transposon insertions in genes encoding efflux pumps or activators of efflux pumps, and conversely, cells containing the lowest concentrations of ethidium would have insertions in genes encoding negative regulators of efflux pumps. Comparisons of the insertion sites in the mutant input pool with those in the high- and low-fluorescence pools supported this proposal (Fig. 1; see also Data Set S1 in the supplemental material). Mutants carrying insertions in genes encoding several multidrug efflux pumps, particularly *adeABC* (12), *adeIJK* (10), and *amvA* (13, 14), and genes encoding the *adeABC* activator, *adeRS* (15), were overrepresented in the highly fluorescent populations (Fig. 1). Inactivation of these genes is likely to reduce the rate of efflux and thus result in a higher cytosolic concentration of ethidium. In contrast, inactivated mutants of these genes were less abundant in the low-fluorescence populations (Fig. 1), since the efflux pumps encoded or regulated by these genes help to lower the concentration of ethidium in the cell. We used the

Transporter Automated Annotation Pipeline (http://www.membranetransport.org/) to search for genes encoding novel efflux pumps in the *A. baumannii* BAL062 genome. We identified 56 genes that are likely to encode novel efflux pumps, or components of novel efflux pumps, based on their primary sequence characteristics (see Table S1 in the supplemental material). These efflux pumps are likely to recognize small-molecule substrates, but our data did not suggest that any of these efflux pumps have a significant *in vivo* role in ethidium efflux, since none were significantly differentially selected by our fluorescence-based selection (see Table S1).

Some of the most highly differentially selected genes in the flow-sorted samples were genes that encode transcriptional repressors known or predicted to control expression of multidrug efflux systems. For example, mutants carrying insertions in *adeN*, which controls expression of *adeIJK* (16), were 1,469-fold less abundant in the highly fluorescent output pool compared to the input pool (Fig. 1). Additionally, mutants carrying insertions in BAL062_01495, which encodes a TetR family regulator, were 371-fold less abundant in the highly fluorescent output pool compared to the input pool (Fig. 1). BAL062_01495 is adjacent to and divergently transcribed from *amvA* in the BAL062 chromosome. To test whether the TetR family protein encoded by BAL062_01495 was able to regulate expression of *amvA*, we compared *amvA* expression levels in the *A. baumannii* AB5075-UW parental strain and a strain harboring a transposon insertion in the gene orthologous to BAL062_01495. The level of *amvA* expression measured by reverse transcription-quantitative PCR (qRT-PCR) (5) in the mutant strain was 5.7- ± 1.9-fold higher than that in the parental strain during late exponential phase, indicating that the TetR family regulator controls expression of *amvA*. Consequently, we have tentatively named this novel regulator AmvR.

To confirm the specific involvement of different multidrug efflux pumps and their regulators in controlling the accumulation of ethidium in *A. baumannii*, we conducted flow cytometry on targeted mutants of *adeB*, *adeR*, *adeJ*, *adeN*, *adeG*, *adeL*, *amvA*, *amvR*, *craA*, *abeS*, and *abeM*. These mutant strains were loaded with 40 μM ethidium bromide, and the fluorescence of 10,000 cells was determined by flow cytometry (Fig. 2). The TraDISort method identified the AdeABC, AdeIJK, and AmvA efflux systems and their regulators, AdeRS, AdeN, and AmvR, as playing a role in ethidium accumulation. The fluorescence profiles of the specific mutant populations closely reflected these findings. As seen in our preliminary experiments (see Fig. S1 in the supplemental material), the average fluorescence of the *adeJ* and *adeN* mutant populations was above and below that of the parental cell population, respectively (Fig. 2B). The *amvA* and *amvR* mutant cells showed fluorescence profiles very similar to those of *adeJ* and *adeN* mutants, respectively (Fig. 2D), in line with the function of AmvR as a repressor of *amvA* expression. The average fluorescence of the *adeB* and *adeR* mutant cell populations was a similar degree higher than that of the parental population, highlighting the role of AdeB in ethidium efflux and of AdeR in controlling the expression of *adeABC* (Fig. 2A). The fluorescence profiles of mutant populations of other multidrug efflux systems, which were not identified using the TraDISort approach, were very similar to that of the parental strain (Fig. 2).

TraDIS following fitness-induced selection using ethidium bromide. In addition to FACS to enrich for cells displaying aberrant accumulation of ethidium, we cultured the mutant library in the presence of ethidium bromide. This experiment used a higher concentration of ethidium bromide (1/4× MIC of the parental strain) than that used in the FACS analyses to impose a chemical selection that would allow us to identify mutants with a fitness advantage or defect in the presence of ethidium by TraDIS. In the ethidium bromide-selected mutant pools, transposon insertions were less abundant in 63 genes and more abundant in eight genes compared with the input control pools. This suggests that gene loss generally results in a fitness defect, rather than advantage, under ethidium selection, which is in keeping with general evolutionary theory. Mutants containing transposon insertions in efflux pump genes and their regulators were the most highly differentially selected by ethidium bromide. The pattern of selection among these mutants overlapped with the selection pattern in the low-ethidium-fluorescence FACS experiment (Fig. 1). For example, mutants carrying transposon insertions in the *adeABC*, *adeIJK*, *amvA*, and *adeRS* genes were less abundant in the ethidium-

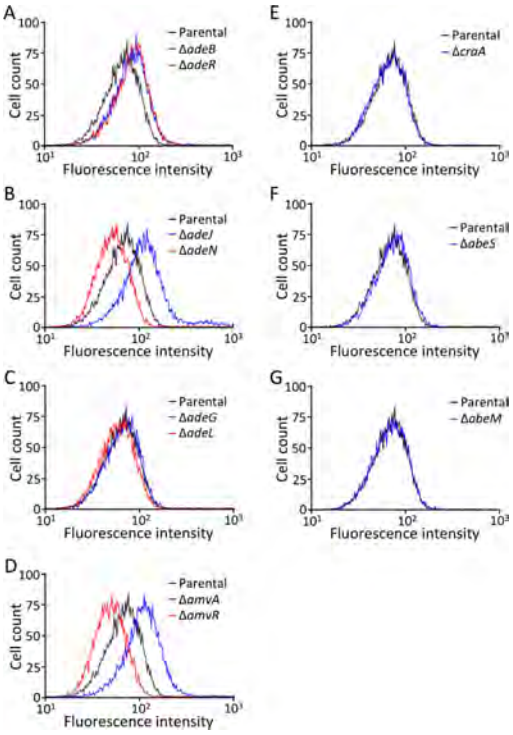

FIG 2 Flow cytometric analysis of *Acinetobacter baumannii* AB5075-UW parental strain (black), inactivated efflux pump mutants (blue), and inactivated efflux regulatory mutants (red). The fluorescence profile of the parental population is shown in all panels and is overlaid with the profiles of Δ*adeB* and Δ*adeR* (A), Δ*adeJ* and Δ*adeN* (B), Δ*adeG* and Δ*adeL* (C), Δ*amvA* and Δ*amvR* (D), Δ*craA* (E), Δ*abeS* (F), and Δ*abeM* (G). Cell populations were exposed to 40 μM ethidium bromide, and each curve shows the fluorescence intensity for 10,000 cells. The cell populations show distinct fluorescence profiles based on the concentration of ethidium in the cell cytoplasm.

selected output pool (Fig. 1), confirming the role of these multidrug efflux pumps and regulators in resistance to ethidium.

Similarly to efflux pump genes and their regulators, mutations in the DNA modification methylase gene, BAL062_03687, were significantly negatively selected by ethidium bromide and less abundant in the low-fluorescence samples compared to the input pool. Methylation mediated by BAL062_03687 could protect DNA from ethidium intercalation and thereby reduce fluorescence and provide resistance to ethidium-induced mutation.

Several genes controlling the composition of the cell membranes, cell wall, or capsule were also negatively selected by ethidium bromide (BAL062_00585, BAL062_00596, BAL062_01038, BAL062_03374, BAL062_03418, BAL062_03480, BAL062_03481, BAL062_03674, and BAL062_03869 [see Data Set S1 in the supplemental material]). These genes may help to reduce uptake of ethidium. Some of these genes were significantly negatively selected in both the low- and high-fluorescence FACS-selected samples and could thus play a role in controlling cell morphology or size (see below). In contrast, several capsule biosynthesis genes (BAL062_03853, BAL062_03857, and BAL062_03858) were positively selected by the ethidium treatment. This highlights the influence that the sugar composition of the capsule could play in regulating the accumulation of amphipathic small molecules into the cell.

FACS in parallel with TraDIS identifies genes involved in cell division and aggregation. In conducting FACS to enrich for mutants displaying aberrant accumulation of ethidium in *A. baumannii*, we gated to target cells with uniform forward and side scatter and limited the collection of dead or aggregated cells that may complicate downstream analyses (see Fig. S2 in the supplemental material). As a consequence of this gating, we identified a number of mutants that are likely to have cell division defects or enhanced aggregation properties. These mutants were negatively selected in both the low- and high-fluorescence FACS-selected pools relative to the input pool, and included 80 (49.4 to 50.3%) of the significantly selected genes in these pools. For example, mutants carrying insertions in the *mreBCD* gene cluster (BAL062_00713 to BAL062_00715), *rlpA* (BAL062_01224), *rodA* (BAL062_01226), and *ftsI* (BAL062_02811), which are likely to function in cell division, were in very low abundance in each of the flow-sorted mutant pools relative to the input pool (see Fig. S3). Mutants carrying insertions in biotin biosynthesis genes were also significantly less abundant in the FACS-selected pools than in the input pool and, to a lesser extent, in the ethidium-selected pools. The role of biotin in ethidium resistance, cell structure, or aggregation is at present unknown but may be related to its function as a cofactor in fatty acid synthesis. Two capsular polysaccharide biosynthetic genes were significantly less abundant in both of the flow-sorted mutant pools than in the input pool. These mutants may have a higher tendency toward aggregation or different cell morphologies or may display light-scattering properties different from those of other mutant cells (see Fig. S3). Approximately 20% of the inactivated genes in mutants negatively selected by FACS were annotated as hypothetical proteins, and many more had been assigned only putative functions. These genes could be targeted in future investigations exploring cell division and aggregation/biofilm formation in *A. baumannii*.

While insertions in genes implicated in cell replication and increased aggregation were negatively selected by the flow sorting, there appeared to be enrichment for mutants that are less likely to aggregate in culture. The majority of these mutants harbored transposon insertions in the *csu* type I pilus biosynthesis and regulatory gene cluster (BAL062_01328 to BAL062_01334 [see Fig. S4 in the supplemental material]). These genes are likely to function in biotic or abiotic cell adherence/aggregation and biofilm formation (17). Therefore, we suspect that the strains carrying mutations in these genes are less likely to aggregate, leading to their enrichment in our flow-sorted samples.

Conclusions. In this study, we identified the genes that control accumulation of the antimicrobial dye ethidium into the Gram-negative hospital-associated pathogen *A. baumannii*. We exploited the differential fluorescence of ethidium inside and outside the cell to enrich for mutants showing aberrant accumulation of ethidium by FACS and used TraDIS to identify the transposon insertion sites within the enriched mutants. This work highlighted the importance of three multidrug efflux systems, AdeABC, AdeIJK, and AmvA, in reducing ethidium accumulation and promoting resistance. We also confirmed the importance of two regulatory systems, AdeRS and AdeN, that control expression of two of these pumps and identified the first known regulator for the AmvA efflux pump, which we have called AmvR. These results demonstrate the utility of the TraDISort method in identifying bacterial multidrug resistance efflux pumps and will be particularly useful when studying bacterial species for which little is known with respect to the major efflux systems. In addition to the core efflux pumps, the TraDISort method identified a large number of novel genes that are likely to be involved in cell division and/or aggregation. This application considerably expands the scope of utility for this method.

To our knowledge, this study represents the first time that FACS or any other physical selection method has been applied in parallel with TraDIS to physically enrich for phenotypes of interest in mutant populations prior to sequencing. The results demonstrate the feasibility of this combined approach to generate statistically significant results and avoid potential false positives that can arise in traditional fluorescent screening approaches, where individual strains are isolated and studied. In addition to those applications described above, we anticipate that FACS applied in parallel with TraDIS could have a range of additional applications in microbiological research: for example, to rapidly screen saturation mutant libraries carrying fluorescent reporters for genes involved in regulation, to identify the efflux pumps inhibited by novel efflux inhibitors, and to inform *in vitro* evolution studies with fluorescent reporters to identify mutants with improved metabolic productivity (18).

Ethidium accumulation in isogenic *Acinetobacter* mutants measured by flow cytometry. *Acinetobacter baumannii* AB5075-UW and Tn*26* insertion mutants of *adeB* (ABUW_1975-150::T26), *adeR* (ABUW_1973-195::T26), *adeJ* (ABUW_0843-122::T26), *adeN* (ABUW_1731-148::T26) *adeG* (ABUW_1335-195::T26), *adeL* (ABUW_1338-193::T26), *amvA* (ABUW_1679-169::T26), *amvR* (ABUW_1678-136::T26), *craA* (ABUW_0337-173::T26), *abeS* (ABUW_1343-187::T101), and *abeM* (ABUW_3486-184::T26) were obtained from the Manoil lab collection (9). The strains were grown in Mueller-Hinton (MH; Oxoid) broth with shaking overnight, diluted 1:100 in fresh MH broth, grown to late exponential phase, and diluted to an optical density at 600 nm (OD_{600}) of 0.6 in MH broth containing 40 μM ethidium bromide (Sigma-Aldrich), approximately 1/16 of the MIC for the parental strain (250 μg/ml). This concentration is below the MIC for all

strains tested and provided good fluorescent resolution between cells differentially expressing an efflux pump. The cells were incubated at room temperature for 20 min and then further diluted to an OD_{600} of 0.018 in MH broth containing 40 μM ethidium bromide for flow cytometric analyses. The ethidium fluorescence of 10,000 cells from each population was examined on a BD Influx flow cytometer using a 200-mW 488-nm laser (Coherent Sapphire) equipped with a small particle forward scatter detector. Ethidium bromide fluorescence was detected using a 580/30 bandpass filter. The cells were counted from within populations gated by forward scatter versus forward scatter pulse width, to discriminate against aggregated cells, followed by forward and side scatter to ensure that only living cells of uniform size were examined (see Fig. S2 in the supplemental material). *Acinetobacter baylyi* ADP1 wild type and *adeJ* and *adeN* mutants, generated in our previous studies (5), were examined according to the same method, except that 15 μM ethidium bromide was used due to the higher susceptibility of this strain to ethidium.

Transposon mutant library generation and verification by TraDIS. A dense Tn5 mutant library was constructed in *A. baumannii* BAL062, a global clone II isolate (ENA accession numbers LT594095 to LT594096), as previously described (6, 7). Briefly, a custom transposome that included a kanamycin resistance cassette amplified from the pUT-km1 plasmid was generated using the EZ-Tn5 custom transposome construction kit (Epicentre). The custom transposome was electroporated into BAL062, and the cells were plated on kanamycin selective medium (10 mg/liter). More than 100,000 mutants were collected and stored as glycerol stocks at −80°C. Aliquots of stock containing approximately 10^9 cells were grown overnight in MH broth. Genomic DNA was isolated from the cultures, and the transposon insertion sites were sequenced across four lanes of the Illumina HiSeq sequencing system. The insertion sites were mapped and analyzed statistically using protocols and bioinformatic tools in the TraDIS toolkit (7). The number of insertions per gene, as a factor of gene size (insertion index), was calculated for cells grown in MH broth to illustrate the evenness of transposon insertions across the genome and to show that the library was sufficiently saturated for experimental analyses. Insertions at the extreme 3′ end (last 10%) of each gene were filtered since they may not inactivate the gene. When the data were plotted against frequency, we observed a bimodal distribution of insertion indexes in the BAL062 library, with the peaks correlating with genes that tolerate or do not tolerate insertions when cultured under permissive growth conditions (see Fig. S5 in the supplemental material) (19). Using the method described in reference 20, as executed through the TraDIS toolbox (7), essential genes were identified as those with an insertion index below 0.0047 (n = 475) and were excluded from later analyses (see Fig. S5). On average, among the nonessential genes (n = 3,362) there were 35.9 unique insertions per kb of gene sequence (see Fig. S5).

FACS to enrich for *A. baumannii* mutants showing aberrant accumulation of ethidium. An aliquot of BAL062 mutant library stock containing approximately 10^9 cells was grown overnight in MH broth. The overnight culture was diluted 1:100 and grown to late exponential phase (OD_{600} of 5.5). The cells were diluted to an OD_{600} of 0.6 in MH broth containing 40 μM ethidium bromide (approximately 1/16 of the MIC of the parental strain) and then further diluted 1:100 in 40 μM ethidium bromide for FACS. This concentration of ethidium bromide was used because it provided excellent differentiation between mutants known to differentially accumulate ethidium (see Fig. S1 in the supplemental material) and was well below the MIC of these mutants, so that it would not cause changes to the mutant ratios because of cell death during the sorting procedure. Cells were sorted using a BD Influx flow cytometer on the basis of ethidium fluorescence (as described above) using the highest purity mode (1 drop single). Single cells with uniform forward and side scatter were gated, and pools of the most highly and weakly fluorescent cells (2% of total single cells) within this gate were collected in separate tubes containing fresh MH broth (150,000 to 175,000 cells across four replicates [see Table S3 in the supplemental material]). The cells collected were grown overnight, DNA was isolated, and insertion sites were mapped by TraDIS as described above. Comparisons between ratios of insertion sites in the control and experimental mutant pools were made using the statistical comparison scripts in the TraDIS toolbox (7). Genes with fewer than 10 mapped reads in any data set being compared were excluded from the analyses. Genes described as being significantly differentially selected between the control and experimental samples were those showing a greater than 2-fold change in mutant abundance with a Q value below 0.05.

For comparison to the FACS-enriched mutants, we also selected mutants based on their competitive fitness in ethidium bromide. An aliquot of BAL062 mutant library stock containing approximately 10^9 cells was grown overnight in MH broth. The overnight culture was diluted 1:100 and grown overnight in 62.5 μg/ml (158.5 μM) of ethidium bromide (equivalent to 1/4 of the MIC for the parental strain) to impose a chemical selection that would allow us to identify mutants with a fitness advantage or defect in the presence of ethidium bromide. Genomic DNA was isolated, and the insertion sites were determined by TraDIS. A replicate experiment with no ethidium was used as the reference in these experiments.

Accession number(s). The TraDIS sequence data files were deposited into the European Nucleotide Archive under accession numbers listed in Table S2 in the supplemental material.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.01200-16/-/DCSupplemental.

Figure S1, JPG file, 0.1 MB.
Figure S2, JPG file, 0.2 MB.
Figure S3, JPG file, 0.3 MB.
Figure S4, JPG file, 0.2 MB.
Figure S5, JPG file, 0.1 MB.
Data Set S1, XLSX file, 0.5 MB.
Table S1, DOCX file, 0.2 MB.
Table S2, DOCX file, 0.1 MB.
Table S3, DOCX file, 0.1 MB.

FUNDING INFORMATION

This work, including the efforts of Julian Parkhill, was funded by Wellcome Trust (098051). This work, including the efforts of Stephen Baker, was funded by Wellcome Trust (100087/Z/12/Z). This work, including the efforts of Amy K Cain and Christine Boinett, was funded by Medical Research Council (MRC) (G1100100/1). This work, including the efforts of Stephen Baker, was funded by Royal Society (100087/Z/12/Z). This work, including the efforts of Karl A Hassan and Ian T Paulsen, was funded by Department of Health | National Health and Medical Research

Council (NHMRC) (1060895). This work, including the efforts of Karl A Hassan and Liam D.H. Elbourne, was funded by Macquarie University (9201401563).

This work was supported by National Health and Medical Research Council (Australia) Project Grant (1060895) to ITP and KAH, a Macquarie University Research Development Grant (9201401563) to KAH and LDHE, and a Wellcome Trust grant (number 098051) to JP. AKC and CJB were supported by the Medical Research Council [grant number G1100100/1]. SB is a Sir Henry Dale Fellow, jointly funded by the Wellcome Trust and the Royal Society (100087/Z/12/Z).

REFERENCES

1. **Schweizer HP**. 2012. Understanding efflux in gram-negative bacteria: opportunities for drug discovery. Expert Opin Drug Discov **7**:633–642. http://dx.doi.org/10.1517/17460441.2012.688949.
2. **Paulsen IT, Chen J, Nelson KE, Saier MH, Jr.** 2001. Comparative genomics of microbial drug efflux systems. J Mol Microbiol Biotechnol **3**:145–150.
3. **Piddock LJ**. 2006. Clinically relevant chromosomally encoded multidrug resistance efflux pumps in bacteria. Clin Microbiol Rev **19**:382–402. http://dx.doi.org/10.1128/CMR.19.2.382-402.2006.
4. **Ren Q, Paulsen IT**. 2007. Large-scale comparative genomic analyses of cytoplasmic membrane transport systems in prokaryotes. J Mol Microbiol Biotechnol **12**:165–179. http://dx.doi.org/10.1159/000099639.
5. **Brzoska AJ, Hassan KA, de Leon EJ, Paulsen IT, Lewis PJ**. 2013. Single-step selection of drug resistant Acinetobacter baylyi ADP1 mutants reveals a functional redundancy in the recruitment of multidrug efflux systems. PLoS One **8**:e56090. http://dx.doi.org/10.1371/journal.pone.0056090.
6. **Langridge GC, Phan MD, Turner DJ, Perkins TT, Parts L, Haase J, Charles I, Maskell DJ, Peters SE, Dougan G, Wain J, Parkhill J, Turner AK**. 2009. Simultaneous assay of every Salmonella typhi gene using one million transposon mutants. Genome Res **19**:2308–2316. http://dx.doi.org/10.1101/gr.097097.109.
7. **Barquist L, Mayho M, Cummins C, Cain AK, Boinett CJ, Page AJ, Langridge GC, Quail MA, Keane JA, Parkhill J**. 2016. The TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries. Bioinformatics **32**:1109–1111. http://dx.doi.org/10.1093/bioinformatics/btw022.
8. **Lambert B, Le Pecq JB**. 1984. Effect of mutation, electric membrane potential, and metabolic inhibitors on the accessibility of nucleic acids to ethidium bromide in *Escherichia coli* cells. Biochemistry **23**:166–176. http://dx.doi.org/10.1021/bi00296a027.
9. **Gallagher LA, Ramage E, Weiss EJ, Radey M, Hayden HS, Held KG, Huse HK, Zurawski DV, Brittnacher MJ, Manoil C**. 2015. Resources for genetic and genomic analysis of emerging pathogen Acinetobacter baumannii. J Bacteriol **197**:2027–2035. http://dx.doi.org/10.1128/JB.00131-15.
10. **Damier-Piolle L, Magnet S, Brémont S, Lambert T, Courvalin P**. 2008. AdeIJK, a resistance-nodulation-cell division pump effluxing multiple antibiotics in Acinetobacter baumannii. Antimicrob Agents Chemother **52**:557–562. http://dx.doi.org/10.1128/AAC.00732-07.
11. **Rajamohan G, Srinivasan VB, Gebreyes WA**. 2010. Novel role of Acinetobacter baumannii RND efflux transporters in mediating decreased susceptibility to biocides. J Antimicrob Chemother **65**:228–232. http://dx.doi.org/10.1093/jac/dkp427.
12. **Magnet S, Courvalin P, Lambert T**. 2001. Resistance-nodulation-cell division-type efflux pump involved in aminoglycoside resistance in Acinetobacter baumannii strain BM4454. Antimicrob Agents Chemother **45**:3375–3380. http://dx.doi.org/10.1128/AAC.45.12.3375-3380.2001.
13. **Rajamohan G, Srinivasan VB, Gebreyes WA**. 2010. Molecular and functional characterization of a novel efflux pump, AmvA, mediating antimicrobial and disinfectant resistance in Acinetobacter baumannii. J Antimicrob Chemother **65**:1919–1925. http://dx.doi.org/10.1093/jac/dkq195.
14. **Hassan KA, Brzoska AJ, Wilson NL, Eijkelkamp BA, Brown MH, Paulsen IT**. 2011. Roles of DHA2 family transporters in drug resistance and iron homeostasis in Acinetobacter spp. J Mol Microbiol Biotechnol **20**:116–124. http://dx.doi.org/10.1159/000325367.
15. **Marchand I, Damier-Piolle L, Courvalin P, Lambert T**. 2004. Expression of the RND-type efflux pump AdeABC in *Acinetobacter baumannii* is regulated by the AdeRS two-component system. Antimicrob Agents Chemother **48**:3298–3304. http://dx.doi.org/10.1128/AAC.48.9.3298-3304.2004.
16. **Rosenfeld N, Bouchier C, Courvalin P, Périchon B**. 2012. Expression of the resistance-nodulation-cell division pump AdeIJK in Acinetobacter baumannii is regulated by AdeN, a TetR-type regulator. Antimicrob Agents Chemother **56**:2504–2510. http://dx.doi.org/10.1128/AAC.06422-11.
17. **Eijkelkamp BA, Stroeher UH, Hassan KA, Paulsen IT, Brown MH**. 2014. Comparative analysis of surface-exposed virulence factors of Acinetobacter baumannii. BMC Genomics **15**:1020. http://dx.doi.org/10.1186/1471-2164-15-1020.
18. **Williams TC, Pretorius IS, Paulsen IT**. 2016. Synthetic evolution of metabolic productivity using biosensors. Trends Biotechnol **34**:371–381. http://dx.doi.org/10.1016/j.tibtech.2016.02.002.
19. **Barquist L, Langridge GC, Turner DJ, Phan MD, Turner AK, Bateman A, Parkhill J, Wain J, Gardner PP**. 2013. A comparison of dense transposon insertion libraries in the salmonella serovars Typhi and Typhimurium. Nucleic Acids Res **41**:4549–4564. http://dx.doi.org/10.1093/nar/gkt148.
20. **Coyne S, Rosenfeld N, Lambert T, Courvalin P, Périchon B**. 2010. Overexpression of resistance-nodulation-cell division pump AdeFGH confers multidrug resistance in Acinetobacter baumannii. Antimicrob Agents Chemother **54**:4389–4393. http://dx.doi.org/10.1128/AAC.00155-10.
21. **Roca I, Marti S, Espinal P, Martínez P, Gibert I, Vila J**. 2009. CraA, a major facilitator superfamily efflux pump associated with chloramphenicol resistance in Acinetobacter baumannii. Antimicrob Agents Chemother **53**:4013–4014. http://dx.doi.org/10.1128/AAC.00584-09.
22. **Srinivasan VB, Rajamohan G, Gebreyes WA**. 2009. Role of AbeS, a novel efflux pump of the SMR family of transporters, in resistance to antimicrobial agents in Acinetobacter baumannii. Antimicrob Agents Chemother **53**:5312–5316. http://dx.doi.org/10.1128/AAC.00748-09.
23. **Su XZ, Chen J, Mizushima T, Kuroda T, Tsuchiya T**. 2005. AbeM, an H^+-coupled *Acinetobacter baumannii* multidrug efflux pump belonging to the MATE family of transporters. Antimicrob Agents Chemother **49**:4362–4364. http://dx.doi.org/10.1128/AAC.49.10.4362-4364.2005.
24. **Hassan KA, Jackson SM, Penesyan A, Patching SG, Tetu SG, Eijkelkamp BA, Brown MH, Henderson PJ, Paulsen IT**. 2013. Transcriptomic and biochemical analyses identify a family of chlorhexidine efflux proteins. Proc Natl Acad Sci U S A **110**:20254–20259. http://dx.doi.org/10.1073/pnas.1317052110.
25. **Hassan KA, Liu Q, Henderson PJ, Paulsen IT**. 2015. Homologs of the *Acinetobacter baumannii* AceI transporter represent a new family of bacterial multidrug efflux systems. mBio **6**:e01982-14. http://dx.doi.org/10.1128/mBio.01982-14.
26. **Lin MF, Lin YY, Yeh HW, Lan CY**. 2014. Role of the BaeSR two-component system in the regulation of Acinetobacter baumannii adeAB genes and its correlation with tigecycline susceptibility. BMC Microbiol **14**:119. http://dx.doi.org/10.1186/1471-2180-14-119.
27. **Hassan KA, Elbourne LD, Li L, Gamage HK, Liu Q, Jackson SM, Sharples D, Kolstø AB, Henderson PJ, Paulsen IT**. 2015. An ace up their sleeve: a transcriptomic approach exposes the AceI efflux protein of Acinetobacter baumannii and reveals the drug efflux potential hidden in many microbial pathogens. Front Microbiol **6**:333. http://dx.doi.org/10.3389/fmicb.2015.00333.

Appendix III: Sampling locations, cell abundances of the autotrophic unicellular phytoplankton, and the measurements of environment parameters in the surface water of the Arafura Sea, Torres Strait and Coral Sea

Sample sites	Latitude (S)	Longitude (E)	Cell abundance (cells/mL)[a]			Temperature (°C)	Salinity (PSU)	Nutrient concentration (μM)				Chl *a* (μg/L)
			Syn	Pro	PPEs			Nitrate	Phosphate	Silicate	Ammonia	
CTD1	10.96111111	130.6013889	34701	-	1888	27.91	33.745	0.033374	0.13426	3.259441	0.075401	0.1125
UW1	11.07445	131.80135	82394	-	2455	28.405	33.76	0.045128	0.13918	3.265421	0.098451	0.1319
UW2	10.88531667	131.9782833	86251	-	3123	27.921	33.731	0.046254	0.142546	3.271548	0.123125	0.0886
UW3	10.78983333	132.2585667	45961	-	2636	28.92	33.743	0.051426	0.1432156	3.295486	0.125874	0.078
UW4	10.79398333	132.61785	35367	-	2529	28.05	33.768	0.053211	0.1478952	3.289456	0.135482	0.075
UW5	10.79638333	132.9271667	22634	-	1385	27.95	33.711	0.059864	0.145615	3.291547	0.140149	0.072
UW6	10.79791667	133.1073833	11277	-	722	27.79	33.725	0.0621458	0.1460524	3.284125	0.153246	0.065
UW7	10.82748333	133.2656667	-	-	-	27.69	33.774	0.069148	0.146845	3.286542	0.172546	0.068
UW8	10.86926667	133.4231	13716	-	1067	27.65	33.784	0.072145	0.147124	3.2886452	0.189745	0.066
UW9	10.91461667	133.5941833	-	-	-	27.63	33.731	0.079355	0.149856	3.293546	0.196524	0.071
UW10	10.95625	133.7513833	-	-	-	27.54	33.722	0.0865421	0.150264	3.296451	0.198452	0.078
UW11	11.23213333	134.7446333	113832	-	3132	27.42	33.748	0.092451	0.153648	3.295418	0.201245	0.105
UW12	11.22463333	134.7467	99420	-	2199	27.48	33.755	0.095412	0.155648	3.294587	0.2301245	0.098
UW13	11.22513333	134.7490833	77035	-	1816	27.51	33.763	0.098457	0.154325	3.297149	0.225648	0.105
UW14	11.17251667	135.0789833	209170	-	1270	27.78	33.758	0.104526	0.159846	3.314545	0.235621	0.099
CTD2	11.46166667	134.9658333	126211	-	3048	27.55	33.76	0.110996	0.16014	3.303668	0.248532	0.101
UW15	11.15308333	135.2311	191312	-	2066	27.71	33.771	0.1000546	0.16524	3.001242	0.224518	0.103
UW16	11.13086667	135.39365	-	-	-	27.56	33.755	0.0984512	0.159812	2.894512	0.204518	0.112
UW17	11.08783333	135.74155	95738	-	2857	27.43	33.724	0.112654	0.15246	2.9354812	0.214587	0.101
UW18	11.07206667	135.8652	52436	-	1699	27.4	33.695	0.082541	0.157891	2.651254	0.192458	0.1
UW19	10.96121667	136.7998	90262	-	5613	26.48	33.739	0.073524	0.1510245	2.812546	0.145218	0.105
UW20	10.96645	136.7974167	164718	-	9038	26.8	33.704	0.0698451	0.152874	2.778452	0.154142	0.115
UW21	10.9305	136.8298667	40534	13677	1909	27.14	33.658	0.063845	0.154735	2.698457	0.099854	0.107
UW22	10.79848333	136.8772833	24222	27796	1217	27.23	33.679	0.060124	0.154512	2.66654	0.084512	0.11
UW23	10.68295	136.9509667	48060	-	2805	27.43	33.673	0.052684	0.153478	2.484513	0.091458	0.0782
UW24	10.72645	137.1399167	35331	-	1843	27.48	33.671	0.054251	0.151045	2.214584	0.0875148	0.107

UW25	10.7641	137.30425	39018	-	1034	27.39	33.661	0.0518745	0.153201	2.178951	0.0821547	0.137
UW26	10.79601667	137.4414833	-	-	-	27.37	33.659	0.049855	0.149257	1.915426	0.084512	0.147
CTD3	10.84027778	137.5886111	31026	-	1440	27.32	33.668	0.04642	0.147479	1.898488	0.07555	0.121
UW27	10.82113333	137.7358	71240	-	1051	27.33	33.746	0.05426	0.149885	1.998451	0.077514	0.142
UW28	10.82331667	138.00675	-	-	-	27.46	33.986	0.09221	0.157415	2.245781	0.074578	0.132
UW29	10.82538333	138.25	169503	-	3472	27.16	34.221	0.152364	0.162451	2.984512	0.0724581	0.11
CTD4	10.83215	138.71365	42869	-	4109	27.15	34.402	0.190694	0.164136	3.334361	0.07199	0.106
UW30	10.90972222	138.4427778	27128	-	1856	27.07	34.157	0.08456	0.124587	2.458123	0.094578	0.104
CTD5	10.80733333	139.6267333	65398	-	1352	27.25	33.932	0.022272	0.117462	1.865164	0.107584	0.117
UW31	10.80066667	139.8548333	59605	-	1389	27.13	33.958	0.082451	0.127458	2.2658451	0.0954212	0.116
UW32	11.02916667	139.4069444	72056	-	2508	27.15	33.991	0.108742	0.146625	2.678951	0.085412	0.124
UW33	10.79245	140.1393333	56281	-	2785	27.03	33.974	0.256614	0.1598621	3.356217	0.0751526	0.143
UW34	10.81446667	140.2226833	-	-	-	27.22	33.947	0.345121	0.174583	4.487513	0.062154	0.121
CTD6	10.82555556	140.3266667	65715	-	4080	27.11	33.934	0.456492	0.185604	4.769445	0.059699	0.145
UW35	10.80745	140.4082333	-	-	-	27.53	33.901	0.254876	0.228745	5.248546	0.0502456	0.161
CTD7	11.4375	141.0036111	199981	-	3962	27.04	33.884	0.013467	0.264012	5.72523	0.040972	0.135
UW36	10.78686667	141.2413667	98414	7547	1594	27.03	34.725	0.024576	0.105481	3.984125	0.035684	0.112
CTD8	10.92777778	141.6708333	47398	-	1527	27.11	35.57	0.027563	0.054551	2.888043	0.027905	0.105
UW37	10.609	141.8488667	279920	-	5393	26.98	35.124	0.0271452	0.048561	3.514521	0.0245125	0.096
UW38	10.5247	142.1826667	141312	-	2822	26.92	34.813	0.025561	0.044128	3.986542	0.0241254	0.927
UW39	10.45403333	142.5287167	142333	-	3044	26.99	34.425	0.020215	0.040158	4.236547	0.022654	0.887
UW40	10.30738333	142.8457167	62027	-	2255	27.08	34.254	0.0198475	0.035841	4.356821	0.0210241	0.0913
UW41	10.06245	143.0313167	111602	-	2826	27.29	33.987	0.017854	0.0302654	4.652154	0.020458	0.0849
UW42	9.792066667	143.1883167	-	-	-	27.45	33.512	0.014253	0.027854	5.158462	0.0190251	0.0886
UW43	9.579016667	143.3976667	57269	-	1791	27.12	33.398	0.012843	0.024512	5.584625	0.0187543	0.087
CTD9	9.4475	143.7427778	81274	-	2454	27.13	33.353	0.010254	0.022578	5.73947	0.018078	0.0874
UW44	9.266916667	143.8799667	20970	13268	583	27.39	33.985	0.011853	0.0254511	4.87952	0.025894	0.088
CTD10	9.752777778	144.6475	40826	43550	1733	26.59	34.811	0.012072	0.028001	1.760863	0.034994	0.0877
UW45	9.971283333	144.3509333	17631	194828	1077	26.56	34.714	0.014562	0.029451	1.751273	0.031542	0.0879
UW46	10.25198333	144.21675	-	-	-	26.51	34.845	0.017952	0.027485	1.746832	0.030478	0.0835
UW47	10.49335	144.1322667	480998	1366610	21127	26.45	34.822	0.018951	0.026481	1.740123	0.025982	0.0885

UW48	10.64528333	144.12885	10812	-	686	26.58	34.854	0.021832	0.026002	1.739834	0.021005	0.0836
UW49	11.05153333	144.10935	-	-	-	26.65	34.877	0.022567	0.025874	1.725185	0.0192141	0.113
UW50	10.2688	144.1344167	-	-	-	26.59	34.899	0.0258933	0.025419	1.729845	0.011213	0.115
UW51	11.57626667	144.2018	-	-	-	26.51	34.911	0.029478	0.025587	1.720054	0.007584	0.116
CTD11	12.2375	144.1866667	4950	-	870	26.41	34.905	0.031606	0.025613	1.717635	0.005978	0.119
CTD12	11.84222222	144.3755556	939	-	327	26.49	34.912	0.028395	0.017999	1.723871	0.036252	0.122
UW52	11.90206667	144.2850333	-	-	-	26.48	35.029	0.019875	0.035554	1.645102	0.021305	0.125
CTD13	12.82416667	144.715	-	-	-	26.34	35.022	0.011373	0.043896	1.538209	0.008571	0.0894
UW53	12.9452	144.8553667	1614	-	745	26.43	35.029	0.012784	0.045876	1.533425	0.024734	0.0764
UW54	13.21863333	144.9398833	1227	-	554	26.33	35.034	0.014512	0.042154	1.529361	0.044713	0.0767
UW55	13.47865	144.8316667	2136	137624	635	26.17	35.046	0.015783	0.040664	1.522714	0.078412	0.081
UW56	13.81381667	144.6944	7120	288118	1285	26.1	35.052	0.017685	0.399945	1.514854	0.098745	0.078
CTD14	13.94111111	144.8119444	29835	-	1362	26.03	35.055	0.01937	0.039765	1.512461	0.101501	0.0854
UW57	14	144.7900167	66459	414697	1957	26.01	35.056	0.018452	0.034512	1.529874	0.075412	0.089
UW58	14.20256667	144.9884167	8836	54586	1042	26.02	35.049	0.016369	0.032145	1.574325	0.084512	0.082
UW59	14.2889	145.2766667	15191	75909	731	25.94	35.032	0.010245	0.028415	1.641185	0.057849	0.09
UW60	14.50896667	145.5971	-	-	-	25.98	35.039	0.008793	0.027014	1.720184	0.025103	0.0853
CTD15	14.57027778	145.6383333	21510	108127	1394	25.9	35.017	0.005186	0.025519	1.797248	0.002246	0.0895
CTD16	15.67271667	145.8309333	2497	15042	351	26.46	35.067	0.029327	0.085333	1.357489	0.036899	0.091
UW61	15.45361111	146.0608333	2354	73822	760	26.41	35.068	0.027899	0.084212	1.354412	0.0345872	0.093
UW62	15.84823333	145.8475833	12400	235841	1465	26.37	35.074	0.024897	0.083984	1.356781	0.0356982	0.095
UW63	16.15543333	145.9692167	-	-	-	26.48	35.072	0.026367	0.082113	1.354741	0.036114	0.094
UW64	16.1418	145.95145	5973	284887	918	26.53	35.075	0.022547	0.081114	1.350143	0.036897	0.091
UW65	15.7943	145.8605667	-	-	-	26.48	35.076	0.021698	0.0806541	1.350741	0.374891	0.091
CTD17	16.47472222	146.0641667	3277	21386	710	26.41	35.077	0.020662	0.080254	1.350287	0.037603	0.092

a. Syn represents *Synechococcus*, Pro represents *Prochlorococcus*, PPEs represents photosynthetic pico-eukaryotes.

Appendix IV: Data processing statistics of the 16S rRNA sequence reads

Sampling sites	Sampling depth (m)	16S rRNA raw reads	Join paired ends reads	Quality trimmed (>20)	Length trimmed (>350 bp)	After chime checked	Reads assigned
CTD1	25	218509	132555	127734	127733	108857	107892
CTD1	5	106934	80387	77002	77002	64600	64121
CTD10	5	148759	108169	103483	103483	85208	84039
CTD10	56	242566	177735	172158	172158	150167	149668
CTD11	150	229203	169814	161108	161106	121968	119155
CTD11	45	278957	207805	199085	199079	165825	162789
CTD11	5	182400	135518	130652	130651	111775	111118
CTD12	150	146443	112401	107131	107131	84288	83131
CTD12	5	168275	165288	159866	159866	138503	137906
CTD12	75	163770	107090	101691	101691	88158	87932
CTD13	5	216482	154183	149303	149303	131107	130753
CTD13	90	193341	145688	140170	140170	117144	115175
CTD14	120	141597	111048	106602	106602	88607	87915
CTD14	5	509993	388463	375712	375706	324584	322950
CTD15	5	214204	167988	162444	162442	141214	140672
CTD15	90	161165	119361	114760	114760	96919	95716
CTD16	100	197400	149939	143780	143780	119038	117735
CTD16	5	165131	130496	126693	126692	112039	111751
CTD17	5	175246	139257	134781	134771	116933	116427
CTD17	75	379590	299436	288094	288093	244241	240992
CTD2	5	229525	172813	164481	164479	135792	134700
CTD3	42	74212	54258	51668	51668	42212	41611
CTD3	5	71224	53968	51498	51497	42305	41868
CTD4	25	20809	15627	14984	14984	12552	12407
CTD4	5	13747	10066	9623	9623	8057	7956
CTD5	37	73821	54850	51961	51961	40335	39543
CTD5	5	136705	89216	83593	83592	62574	61558

CTD6	25	178201	133644	126785	126785	102355	100790
CTD6	5	89715	61777	57791	57791	43541	42887
CTD7	23	172439	132074	126125	126125	104859	103660
CTD7	5	89911	70098	67112	67110	57562	57231
CTD8	5	142168	105492	100821	100821	83800	82610
CTD9	28	91984	70031	67074	67073	56224	55502
CTD9	5	107378	80371	77157	77157	64873	64441
UW1	5	190752	126791	119882	119878	89509	88467
UW14	5	194612	136004	131335	131335	114100	113943
UW15	5	141858	99273	95647	95647	82157	82011
UW16	5	276043	202778	195739	195739	170248	169946
UW18	5	83992	26761	25534	25534	21128	21061
UW2	5	109010	71807	68949	68949	57828	57400
UW20	5	248216	175525	167750	167748	135792	133982
UW22	5	223592	154218	147773	147772	123127	122667
UW23	5	190227	123193	117004	117004	89627	88828
UW25	5	538998	371393	354298	354283	279816	277430
UW27	5	25181	18377	17548	17548	14017	13949
UW29	5	174712	121940	117477	117476	101043	100811
UW3	5	154994	90671	86403	86401	73137	72285
UW31	5	171561	116913	111902	111902	91769	91366
UW32	5	259384	197152	190238	190238	166348	166050
UW33	5	169566	122654	118204	118204	103046	102699
UW34	5	353236	240519	229535	229535	186535	183888
UW35	5	189350	134983	129839	129837	110855	110429
UW36	5	163734	110296	105791	105791	88750	88370
UW37	5	153047	108117	103881	103881	87960	87301
UW38	5	288951	193481	185762	185762	155659	154467
UW39	5	201929	147856	142395	142395	123614	123224
UW4	5	130665	95081	90739	90739	73791	73470
UW41	5	268134	210757	203036	203035	175207	174808

UW42	5	205411	139248	133178	133178	111650	111322
UW43	5	232672	179287	172438	172438	147703	147452
UW44	5	436795	299987	285404	285404	227402	226275
UW45	5	235028	158084	151828	151828	125647	124974
UW47	5	235928	38080	36220	36219	30740	30593
UW49	5	157270	89536	85883	85883	72907	72561
UW5	5	127098	93724	89275	89273	71642	71257
UW51	5	71025	10673	10187	10187	8859	8838
UW55	5	126921	42511	40884	40884	35523	35382
UW56	5	119147	16111	15349	15349	12999	12933
UW58	5	95273	7962	7516	7514	6349	6320
UW60	5	138896	83479	80312	80311	68142	67800
UW61	5	149518	54555	52103	52103	43947	43793
UW62	5	134262	56700	50442	50442	42936	42696
UW63	5	186509	90303	81802	81802	69886	69522
UW64	5	164735	60607	52340	52340	44110	43897

Appendix V: Data processing statistics of the 18S rRNA sequence reads

Sampling sites	Sampling depth (m)	18S rRNA raw reads	Join paired ends reads	Quality trimmed (>20)	Length trimmed (>350 bp)	After chime checked	Reads assigned
CTD1	25	239428	235283	224667	223975	223023	201018
CTD1	5	218216	213860	204693	203600	202606	184923
CTD10	56	169724	166964	159996	159104	158280	144160
CTD10	5	291925	287955	277186	276020	274856	257037
CTD11	150	264954	259633	248148	247493	246643	203214
CTD11	45	266770	262205	251448	250562	249679	231546
CTD11	5	232521	227430	218319	217200	216318	201835
CTD12	150	244951	241046	230968	228834	227711	173469
CTD12	5	286823	283733	272839	271043	269994	260189
CTD12	75	281284	276443	265194	263828	262645	234365
CTD13	5	94277	92719	89149	88606	88170	81949
CTD13	90	159728	157831	151644	151443	150916	132705
CTD14	120	108392	105753	100745	100380	99830	90812
CTD14	5	240630	237395	217355	215956	226277	205508
CTD15	5	262422	258909	248629	247321	246228	232173
CTD15	90	181478	179522	172754	171647	170781	154668
CTD16	100	185452	183128	176024	175152	174443	148897
CTD16	5	167279	165079	159298	158884	158366	148496
CTD17	5	249959	247154	238415	237187	236206	221265
CTD17	75	267147	260670	248492	245436	244137	225637
CTD2	5	306392	301974	288755	287849	286848	267380
CTD3	42	198500	195623	187757	187331	186742	167298
CTD3	5	271545	267343	256169	254636	253489	237109
CTD4	25	316354	309846	296348	296004	294996	266976
CTD4	5	227394	224048	214891	214183	213299	194773
CTD5	37	246243	241802	232040	231582	230796	204791
CTD5	5	256486	251494	240699	238974	237806	222679

CTD6	25	266333	261741	250412	249610	248604	227766
CTD6	5	254649	250465	240075	236963	235793	220678
CTD7	23	136227	133279	127501	127207	125525	115865
CTD7	5	271580	267202	255293	254896	254098	242094
CTD8	5	477533	470291	450318	449918	448639	420622
CTD9	28	221945	218524	209432	208957	208266	189140
CTD9	5	422467	416280	399064	397442	396139	370334
UW1	5	225547	223034	214564	214081	213322	190904
UW22	5	120741	118908	113606	113420	113064	104608
UW23	5	271857	266764	254523	253915	252922	235473
UW25	5	191684	188770	180188	179847	179114	168244
UW27	5	248525	243608	232434	232249	231550	220466
UW32	5	25543	23501	21418	21363	21013	19121
UW33	5	120639	119450	114389	114218	113857	105031
UW36	5	187434	185807	178870	178305	177707	165331
UW4	5	155078	152487	144963	144584	1168787	141887
UW41	5	130445	128151	122756	122573	122179	110835
UW43	5	116868	115189	111069	109792	109188	101525
UW44	5	66014	65106	62593	61979	61682	57748
UW45	5	180249	176654	168739	168694	168144	157109
UW47	5	181146	178963	171986	171863	171305	161541
UW49	5	237510	233550	223219	223119	222438	206284
UW5	5	217486	214052	203773	203439	1571757	198462
UW52	5	107088	105997	102119	101682	101281	92922
UW53	5	217125	213719	205242	204720	204005	186979
UW54	5	267179	263733	252527	252266	251480	239391
UW55	5	217030	214175	205752	205337	204676	192835
UW56	5	252173	248631	238937	238184	237328	220841
UW58	5	172332	169469	162741	162140	161464	152499
UW59	5	222231	219175	210062	209429	208707	196812
UW60	5	77933	77194	74375	73624	73241	66954

UW61	5	118763	116459	111639	110832	110354	101657
UW62	5	227095	222704	212600	211779	210967	194658
UW63	5	226523	222127	212160	211717	210978	194236
UW64	5	215642	211821	202398	202256	201595	187147
UW65	5	137940	135656	129972	128692	128083	116207

Appendix VI: Seafloor depth, sampling depth and measurements of environmental parameters in the Arafura Sea, Torres Strait and Coral Sea

Sampling site	Seafloor depth (m)	Sampling depth (m)	Temperature (°C)	Salinity (PSU)	Nutrient concentrations (µM)			
					Nitrate	Phosphate	Silicate	Ammonia
CTD1	46	25	27.703	33.741	0.043379	0.137561	3.374221	0.08912
CTD3	52	42	25.442	33.718	0.403088	0.333969	4.83388	0.080554
CTD4	58	25	27.079	34.446	0.03435	0.163894	3.349317	0.100049
CTD5	54	37	26.7	33.884	1.712205	0.322726	5.672398	0.164446
CTD6	60	25	27.016	33.923	0.809472	0.188263	4.632895	0.055839
CTD7	40	23	26.879	33.879	0.091732	0.274326	5.973033	0.097231
CTD9	43	28	26.6	34.168	0.023619	0.061837	3.638031	0.032711
CTD10	102	56	25.84	35.041	0.307272	0.070228	1.41867	0.068123
CTD11	760	45	26.042	35.039	0.001574	0.053661	1.4136	0.004084
CTD11	760	150	25.564	35.269	0.145614	0.098321	1.0541	0.006412
CTD12	1211	75	25.616	35.143	0.118704	0.059138	1.287514	0.038835
CTD12	1211	150	25.024	35.364	0.154301	0.094624	1.143723	0.043351
CTD13	2925	90	25.624	35.118	0.176021	0.074472	1.475581	0.041446
CTD14	1239	120	25.898	35.064	0.092541	0.044659	1.508329	0.127048
CTD15	789	90	25.854	35.028	0.103088	0.025813	1.546737	0.061641
CTD16	1282	100	24.972	35.209	0.542225	0.134583	1.582598	0.005123
CTD17	539	75	25.544	35.078	0.112371	0.066166	1.720425	-0.015607