# Microbial Ecology and Biogeography

— of the —

# Southern Ocean

*David Wilkins*

*Submitted in fulfillment of the requirements for the Degree of Doctor of Philosophy.*

School of Biotechnology and Biomolecular Sciences
University of New South Wales, Sydney

**March 2013**

# Contents

# List of Figures

# List of Tables

# List of Acronyms

GAAS Genome relative Abundance and Average Size.

**AABW** Antarctic Bottom Water.

**AAD** Australian Antarctic Division.

**AAIW** Antarctic Intermediate Water.

**AAP** Aerobic Anoxygenic Phototrophic.

**AC** Antarctic Convergence.

**ACC** Antarctic Circumpolar Current.

**anammox** Anaerobic Ammonium Oxidation.

**ANOSIM** Analysis of Similarities.

**AOA** Ammonia-Oxidizing Archaea.

**AOB** Ammonia-Oxidizing Bacteria.

**AP** Antarctic Peninsula.

**ASW** Antarctic Surface Water.

**AZ** Antarctic Zone.

**BW** Bottom Water.

**CASO** Climate of Antarctica and the Southern Ocean.

**CDW** Circumpolar Deep Water.

**CEAMARC** Collaborative East Antarctic Marine Census.

**CFB** Cytophaga-Flavobacterium-Bacteroides.

**CTD** Conductivity, Temperature and Depth.

**DCM** Deep Chlorophyll Maximum.

**DFAA** Dissolved Free Amino Acids.

**DGGE** Denaturing Gradient Gel Electrophoresis.

**DMSP** dimethylsulfoniopropionate.

**DOC** Dissolved Organic Carbon.

**DOM** Dissolved Organic Matter.

**FISH** Fluorescence *In Situ* Hybridization.

**GOS** Global Ocean Sampling.

**HMW** High Molecular Weight.

**HNLC** High Nutrient, Low Chlorophyll.

**IP** Integer Programming.

**IPCC** Intergovernmental Panel on Climate Change.

**ITS** Internal Transcribed Spacer.

**KEGG** Kyoto Encyclopedia of Genes and Genomes.

**KEOPS** Kerguelen Ocean and Plateau Compared Study.

**LCDW** Lower Circumpolar Deep Water.

**LP** Linear Programming.

**LWM** Low Molecular Weight.

**MGI** Marine Group I Crenarchaeota.

**MMPA** methylmercaptopropionate.

**NADW** North Atlantic Deep Water.

**NZ** North Zone.

**OMG** Oligotrophic Marine Gammaproteobacteria.

**OTU** Operational Taxonomic Unit.

**PF** Polar Front.

**PFZ** Polar Frontal Zone.

**POM** Particulate Organic Matter.

**RCA** Roseobacter Clade Affiliated.

**SACCF** Southern Antarctic Circumpolar Current Front.

**SAF** Subantarctic Front.

**SAM** Southern Annular Mode.

**SAMW** Subantarctic Mode Water.

**SAZ** Subantarctic Zone.

**SB** Southern Boundary of the Antarctic Circumpolar Current.

**SIMPER** Similarity Percentages.

**SO** Southern Ocean.

**SSU** Small Subunit.

**STF** Subtropical Front.

**SZ** South Zone.

**THC** Thermohaline Circulation.

**UCDW** Upper Circumpolar Deep Water.

# Acknowledgements

# Abstract

# Introduction

Sections of this chapter have been previously published in Wilkins D., Yau S., Williams T. J., Allen M. A., Brown M. V., DeMaere M. Z., Lauro F. M., and Cavicchioli R. (2012). Key microbial drivers in Antarctic aquatic environments. *FEMS microbiology reviews*, pages n/a–n/a.

## Physical Oceanography of the Southern Ocean

The Southern Ocean (SO) is large ($\sim$ 36,000,000 km$^2$), oceanographically complex and an important part of the world's hydro- and biospheres. It drives global Thermohaline Circulation (THC): Antarctic Bottom Water (AABW) formed off the Antarctic Coast is the major source of the World Ocean's Bottom Water (BW) (Jacobs, 2004), the sinking of which is one of two major engines for the THC "global conveyor belt" (the other being BW formation in the North Atlantic). The SO also supports a large fraction of global marine primary production: the upwelling of Circumpolar Deep Water (CDW) south of the Polar Front (PF) returns nutrients transported to the deep ocean by sinking particulate matter (Rath *et al.*, 1998) to the surface.

These important functions are closely linked to the SO's unique oceanography. Like the Arctic Ocean, the SO is circumpolar, entailing physical features such as low surface water temperatures, strong seasonal cycles in temperature and solar irradiation, the seasonal formation of sea ice and exposure to surface sheer forces from strong high-latitude winds. Unlike the Arctic Ocean, however, the SO has broad interfaces with the tropical oceans and circumpolar circulation uninterrupted by any major land mass, and in Antarctic coastal waters experiences powerful katabatic winds off the Antarctic ice cap. These properties shape the SO's unique physical oceanography.

### Fronts and zones

Definitions of the SO's extent vary. Features commonly used to define its northern boundary include the 60$^{\text{th}}$ parallel south and the Antarctic Convergence (AC), while most Australian cartographic and governmental bodies consider the SO to begin at Australia's southern coastline, or approximately the 45$^{\text{th}}$ parallel south. The Australian definition will be used in this thesis.

The surface of the SO is composed of several distinct zones, separated by circumpolar fronts (Figure 1). Step transitions in the temperature and density of surface waters define the locations and extent of these fronts (Sokolov and Rintoul, 2002; Orsi *et al.*, 1995). The northernmost front is the Subtropical Front (STF), which lies at $\sim$ 40–45° S, separating the Subantarctic Zone (SAZ) from the warmer and saltier tropical oceans to its north (Sokolov and Rintoul, 2002). Across the STF, potential temperature at 150 m depth decreases from > 12 to < 10 °C.

Moving southwards, the southern extent of the SAZ is defined by the Subantarctic Front (SAF). The SAF is the northernmost and primary current core of the multiply-branched Antarctic Circumpolar Current (ACC) (Sokolov and Rintoul, 2009), and its position is thus defined by that of the ACC which varies considerably with longitude (Moore *et al.*, 1999). As with the STF, there is a drop in potential temperature of 2–4 °C across the front (Sokolov and Rintoul, 2002). The SAF also marks the northern boundary of the Polar Frontal Zone (PFZ), where the SAZ and colder Antarctic Zone (AZ) waters meet and mix. Although both the SAF and PF represent large step changes in surface characteristics, the PFZ itself is relatively constant (Whitworth III and Nowlin Jr., 1987). This zone, and particularly its bounding fronts, are regions of high primary productivity (e.g. Laubscher *et al.*, 1993; Abell and Bowman, 2005).

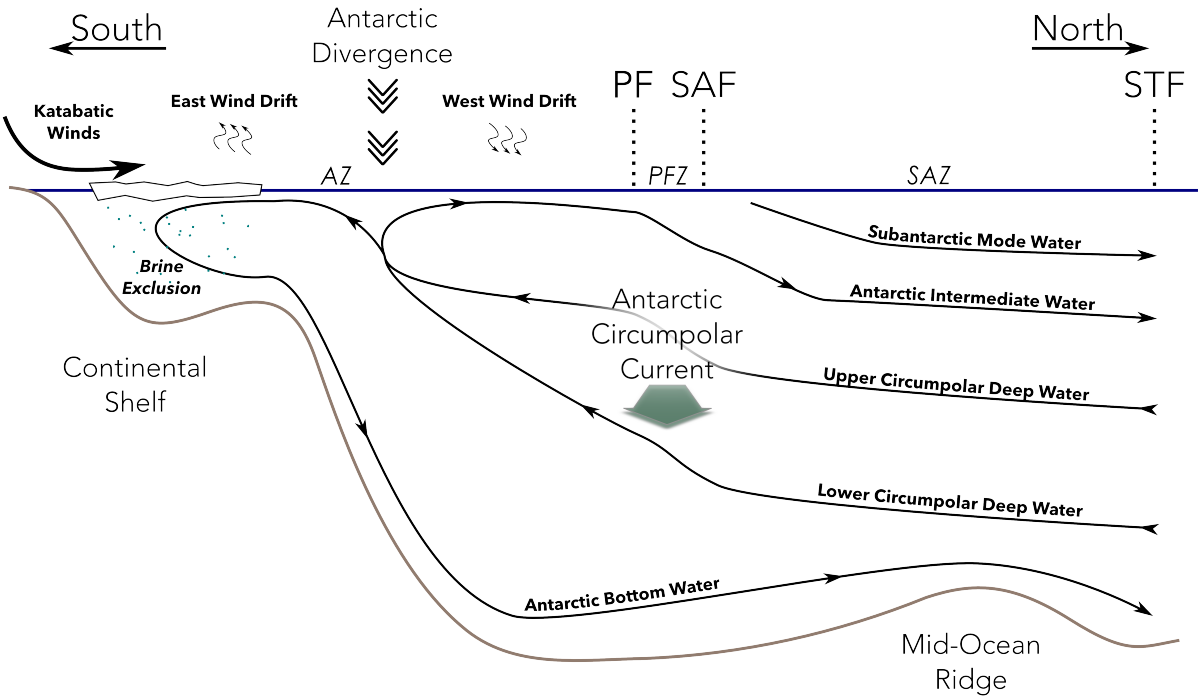



**Figure 1:** North–South cross-section of the Southern Ocean, showing major fronts and water masses. This map is schematic only and not to scale. Acronyms are as follows: Subtropical Front (STF); Subantarctic Zone (SAF); Subantarctic Front (SAF); Polar Frontal Zone (PFZ); Polar Front (PF); Antarctic Zone (AZ).

The southern boundary of the PFZ is the PF, also a current core of the ACC, and associated with a potential temperature drop of $\sim$ 1–1.5 °C (Moore *et al.*, 1999). The waters south of the PF constitute the AZ, the southernmost and coldest (< 2 °C, Sokolov and Rintoul (2002)) major zone. The AZ can be further subdivided by several minor features, including the Southern Antarctic Circumpolar Current Front (SACCF) and Southern Boundary of the Antarctic Circumpolar Current (SB), both weaker southern branches of the ACC. However, these do not represent significant step transitions in physical properties.

## Water masses and circulation

In addition to these surface features, the SO comprises several distinct water masses (Figure 1), the circulation of which forms a major component of global THC. The most extensive of these is CDW, which consists of two layers. The Upper Circumpolar Deep Water (UCDW), characterised by a nutrient maximum and oxygen minimum, originates in the western Indian and south-eastern Pacific oceans (Orsi *et al.*, 1995). The Lower Circumpolar Deep Water (LCDW), characterised by a salinity maximum, originates as sinking North Atlantic Deep Water (NADW) (Whitworth III and Nowlin Jr., 1987). Both layers of CDW shoal southwards across the ACC.

South of the PF is the Antarctic Divergence, a region of transition between dominant easterly and westerly winds. As a consequence of Ekman flow generated by these winds, surface currents are divergent, with those to the north driven further northwards by the westerlies (the West Wind Drift) while those to the south are forced southwards by the easterlies (the East Wind Drift) (Foldvik and Gammelsrød, 1988). This generates a region of upwelling, where the UCDW meet and interact with the upper ocean layers and atmosphere.

Between the divergence and the PF, the surface layer ($\sim$ 100–300 m depth) consists of Antarctic Surface Water (ASW), which is colder, less saline and better ventilated (i.e. more oxygenated) than the CDW. Driven by Ekman transport in the West Wind Drift, the ASW moves northwards towards the PF, with isopycnals (surfaces of constant potential density) sloping gently downwards towards the north. At the PF, the ASW sinks rapidly to form the Antarctic Intermediate Water (AAIW) ($\sim$ 500–1500 m depth), a layer of low-salinity water which underlies SAZ and, moving northwards, contributes to the

intermediate water of the subtropical oceans (Foldvik and Gammelsrød, 1988). (Note that both the SAF and PF can be defined as the locations where the temperature minimum associated with AAIW rapidly decreases in depth (Whitworth III and Nowlin Jr., 1987).) Overlying the AAIW in the SAZ is the Subantarctic Mode Water (SAMW), which forms the surface layer north of the SAF (Speer *et al.*, 2000). Although the SAF is nominally the southern boundary of the SAMW, the surface discontinuity may sometimes occur several degrees to the south of the sub-surface front (e.g. Deacon, 1982; Orsi *et al.*, 1995)

Surface and CDW waters south of the Antarctic Divergence do not move northwards to form AAIW, but instead are driven southwards by the East Wind Drift. The region between the Antarctic Divergence and the coast is the site of dense, cold AABW formation. Katabatic winds from the Antarctic continent form polynyas and cool the surface waters, while brine exclusion during sea ice formation increases the waters' density. This newly formed cold and dense AABW sinks rapidly and flows down the continental shelf and margin to form an abyssal layer beneath the entire SO (Orsi *et al.*, 1999; Foldvik and Gammelsrød, 1988). AABW formation does not occur along the entire continental margin; rather, it is concentrated in the Weddell and Ross seas, and to a lesser extent the D'Urville sea off the Adélie Land coast. AABW is a major source of abyssal water to the World Ocean, and its formation drives global THC. Because AABW is ventilated in the SO before sinking rapidly, AABW is relatively enriched in oxygen compared to other deep layers of the World Ocean, which are oxygen-depleted due to the heterotrophic oxidation of sinking organic matter.

### Effect of climate change

Anthropogenic climate change is having a significant effect on the ACC and the water masses it defines. Changes in the Southern Annular Mode (SAM), a regular pattern of Southern Hemisphere atmospheric circulation characteristics, are leading to an intensification of the westerly winds (Thompson and Solomon, 2002) which drive the ACC. As a consequence of this and other climate change-related effects, the mean annual path of the ACC and its associated fronts and isopycnals has moved $\sim$ 50 km southwards since the 1950s (Gille, 2002). Waters on the poleward side of the ACC have become warmer and more saline, while those to the north cooler and fresher (Böning *et al.*, 2008). The ACC itself is warming and freshening (Böning *et al.*, 2008), from the surface to 900 m depth (Aoki *et al.*, 2003).

Fyfe and Saenko (2005), using a wind-driven model of the southward shift of the ACC, predicted that with conservative assumptions about future anthropogenic greenhouse gas emissions (Intergovernmental Panel on Climate Change (IPCC) A2 and B2 scenarios), the ACC can be expected to move $\sim 1.4^\circ$ southwards by the year 2100. They note that this is equivalent to reducing the volume of the SO south of the ACC by about $16 \times 10^6$ km$^3$, or approximately the volume of the Arctic Ocean.

Aside from the oceanographic effects, these changes can be expected to effect the biology of the SO. In particular, even neglecting the changes in temperature and salinity of the water masses defined by the ACC, the southward migration of the ACC and concomitant change in the relative volumes (and surface areas) of the water masses it defines are a significant change in the size of the microbial habitats these masses represent. Predicting the effects this will have on SO ecosystems and ecosystem functions requires an understanding of the microbial ecology of the SO, and its biogeography relative to the ACC and its associate fronts and water masses. The following section gives an overview of the current state of knowledge.

## Microbial ecology of the Southern Ocean

TODO - Microbes perform key ecosystem functions - Microbes are partitioned by fronts/zones (summarise here, more detail is in individual divisions) - Focus only on Picoplankton (but add gloss on eukarya?) - Focus only on molecular studies

## Bacteria

### Alphaproteobacteria

**Roseobacter** The Roseobacter clade is an abundant and ecologically significant group of marine bacteria, found at high (> 15%) abundance in most marine surface environments (Buchan *et al.*, 2005, and references therein). Unlike some other major proteobacterial groups which are strongly associated with a particular ecological niche (e.g. the SAR11 clade), roseobacters have diverse metabolic abilities, with members capable (for example) of aerobic anoxygenic phototrophy (Biebl *et al.*, 2005; Béjà *et al.*, 2002), degradation of dimethylsulfoniopropionate (DMSP) by at least two pathways (Miller and Belas, 2004; Moran *et al.*, 2003), carbon monoxide oxidation (King, 2003) and heterotrophic utilisation of a broad range of substrates (reviewed in Brinkhoff *et al.*, 2008). Roseobacters are found in the planktonic fraction as well as in commensal association with phytoplankton and metazoans (reviewed in Buchan *et al.*, 2005).

Several 16S rDNA-based studies have identified the Roseobacter Clade Affiliated (RCA) subgroup as ubiquitous and abundant in SO surface waters and to a depth of at least 2200 m, composing $\sim$ 10–30% of surface bacteria (and the majority of Roseobacters) in the Subantarctic and Antarctic zones (Giebel *et al.*, 2009; Murray and Grzymski, 2007; Ghiglione and Murray, 2011) and a major fraction of the population in coastal waters (Murray and Grzymski, 2007; Koh *et al.*, 2011). Two major RCA phylotypes appear to be present in the SO and form the majority of the Roseobacter population. The phylotypes are strictly segregated by the PF, coexisting only within the PFZ (Selje *et al.*, 2004; Giebel *et al.*, 2009) where they may outnumber even the SAR11 clade. There is some evidence that the AZ RCA phylotype originates from the North Atlantic; Giebel *et al.* (2009) noted CDW at 2200 m in the South Zone (SZ) that the waters had an identical temperature-salinity signature to NADW. NADW is formed by the sinking of dense, saline waters in the surface north Atlantic, and is transported to the SO via global thermohaline circulation to become CDW (Callahan, 1972). Consistent with the upwelling of CDW in the AZ south of the PF, (Selje *et al.*, 2004) reported in a global study of RCA 16S rDNA gene fragments that the surface phylotype south of the PF was identical to one found in the Arctic Ocean, while differing by 3 bp from that north of the PF.

Little is known about the functional capabilities of RCA as only two isolated representatives have been described to date. (Giebel *et al.*, 2010) isolated "*Candidatus* Planktomarina temperata" from the North Sea, where it was the dominant phylotype. The authors' identification of the pufM gene encoding a bacteriochlorophyll a subunit suggests at least this member of the RCA is capable of performing aerobic anoxygenic photosynthesis, a function of potentially large ecological significance. (Mayali *et al.*, 2008) isolated an apparently heterotrophic RCA member from subtropical waters, and found *in vitro* evidence that they colonised and increased mortality in blooming dinoflagellates, but did not investigate photosynthetic potential.

Roseobacters, and particularly the RCA, have been strongly associated with phytoplankton blooms in the SO. Two separate 16S rDNA-based studies of a naturally fertilised bloom in the Kerguelen islands region (West *et al.*, 2008; Obernosterer *et al.*, 2011) found that RCA and the Roseobacter NAC11-7 and NAC11-6 clusters were dominant bacterial Operational Taxonomic Units (OTUs) in the bloom patch, suggesting they play a role in heterotrophic degradation of bloom products. Unlike the other clusters, however, RCA representatives were also relatively abundant and metabolically active outside of the patch. Both (Giebel *et al.*, 2009) and (Obernosterer *et al.*, 2011) found that in SO vertical profiles RCA abundances often peaked at the deep chlorophyll maximum, again suggesting an association with phytoplankton.

RCA abundance may follow a seasonal cycle in the SO. (Giebel *et al.*, 2009) found that RCA phylotypes were at maximum 8% of all bacterial 16S rDNA genes during winter but up to 36% in the coastal current and Weddell sea during autumn, while (Ghiglione and Murray, 2011) found the proportion to peak in January in coastal waters off the Antarctic Peninsula and in February off the Kerguelen islands.

A metagenomic study of SOwaters off West Antarctica found that Roseobacter clade Small Subunit (SSU) rRNA sequences were much more abundant in summer than in winter, with *Sulfitobacter* sequences the most abundant within this clade (Grzymski *et al.*, 2012). This is consistent with the association of Roseobacters with phytoplankton (Moran *et al.*, 2003). Nevertheless, Roseobacter clade representatives in these polar waters are metabolically active in both seasons, with an emphasis on high-affinity uptake systems (ABC, TRAP) for capturing labile nutrients such as sugars, polyamines,

amino acids, and oligopeptides (Williams *et al.*, 2012).

**SAR11** The SAR11 clade of Alphaproteobacteria is probably the most abundant class of marine microorganisms worldwide (Morris *et al.*, 2002). "*Candidatus* Pelagibacter ubique" strain HTCC1062, the first and most intensively studied SAR11 isolate, has one of the smallest genomes and gene complements of any known free-living cell as well as a very small cell volume (Giovannoni *et al.*, 2005). The small cell volume, streamlined genome and high proportion of ABC nutrient-uptake transporter genes are all consistent with an oligotrophic lifestyle, scavenging a wide range of substrates using high-affinity, broad-specificity transporters (Giovannoni *et al.*, 2005; Lauro *et al.*, 2009; Sowell *et al.*, 2009). SAR11 cells probably preferentially utilise low over high molecular weight Dissolved Organic Matter (DOM) (Malmstrom *et al.*, 2005) and their relative contribution to uptake of DOM may decrease as substrate concentration increases (Alonso and Pernthaler, 2006). A consequence of this oligotrophic strategy is that SAR11 members are probably unable to take advantage of sudden nutrient influxes, such as during phytoplankton blooms, to rapidly increase cell density (Tripp *et al.*, 2008).

SAR11 has been consistently detected at high abundances in molecular surveys of the SO, in all open ocean regions as well as at depth and in coastal waters, and is usually the dominant alphaproteobacterial, if not bacterial, group (Giebel *et al.*, 2009; Murray and Grzymski, 2007; López-García *et al.*, 2001; Straza *et al.*, 2010; Jamieson *et al.*, 2012; García-Martínez and Rodríguez-Valera, 2000; Ghiglione and Murray, 2011; Murray *et al.*, 2011; Piquet *et al.*, 2011). It is probably more abundant in the epipelagic than at depth (Giebel *et al.*, 2009).

As with the other major bacterial groups, SAR11 seems to exhibit biogeographic partitioning in the SO, and is probably represented by two major ecotypes with a temperature-driven boundary in the region of the PF (Brown *et al.*, 2012). It is probably more abundant in the Subantarctic and polar frontal zones than in the Antarctic Zone (Giebel *et al.*, 2009; Ghiglione and Murray, 2011). This may be related to a competitive advantage of the oligotrophic SAR11 in the High Nutrient, Low Chlorophyll (HNLC) Subantarctic relative to the AZ, where blooming phytoplankton lead to increased concentrations of High Molecular Weight (HMW) DOM and Particulate Organic Matter (POM). Straza *et al.* (2010) found SAR11 accounted for the largest fraction of leucine uptake among all bacterial groups in continental shelf waters off the West Antarctic Peninsula, but a comparatively small fraction of protein uptake, consistent with a role as a Low Molecular Weight (LWM) DOM specialist. West *et al.* (2008), examining 16S rDNA profiles in and out of a natural phytoplankton bloom on the Kerguelen Plateau (Subantarctic), found SAR11 to be a dominant group in HNLC waters outside the bloom patch but relatively less abundant in it. A separate study of the same bloom found SAR11 had a markedly smaller relative contribution to bulk leucine incorporation in the patch than out, suggesting it was not a major contributor to DOM degradation (Obernosterer *et al.*, 2011). Interestingly, SAR11 did dominate in abundance and leucine incorporation at an additional site where a recent and transient phytoplankton bloom had taken place, implying a time lag in the succession between the baseline HNLC and bloom populations. The authors additionally noted that SAR11 abundances at the bloom station began to climb towards non-bloom levels once the bloom had peaked and begun declining. An Antarctic Peninsula SAR11 metaproteome was dominated by ABC transport proteins for the capture and uptake of labile substrates, especially taurine, polyamines and amino acids, and also included DMSP demethylase (Williams *et al.*, 2012). Finally, despite an apparently negative correlation between SAR11 and blooming phytoplankton, Ghiglione and Murray (2011) found only small seasonal changes in abundance during an annual cycle at the Antarctic Peninsula (AP) and Kerguelen Island. These studies are all consistent with the view of SAR11 as a typically non-opportunistic oligotroph specialising in LWM Dissolved Organic Carbon (DOC).

One of the most interesting physiological features of SAR11 representatives is their expression of the retinal-binding pigment proteorhodopsin, which has been shown to act as a proton pump when exposed to light (Beja *et al.*, 2000) and has therefore been implicated in photoheterotrophy. Surprisingly, given very low light levels in Antarctic waters during austral winter, SAR11 proteorhodopsin is present throughout the annual cycle (Williams *et al.*, 2012). This may be consistent with the observation that many marine proteorhodopsins do not appear tuned to maximise energy conversion from available light, which has led Fuhrman *et al.* (2008) to propose at least some proteorhodopsins may perform non-energetic functions such as photoregulatory sensing. Alternatively, constitutive expression of proteorhodopsin for light harvesting in SAR11 may facilitate the ability to immediately respond to cellular energy deficits caused by carbon starvation (Steindler *et al.*, 2011).

**SAR116** The SAR116 clade of Alphaproteobacteria have been detected throughout the world ocean, and in molecular studies of the SO (West *et al.*, 2008; Topping *et al.*, 2006). Topping *et al.* (2006) using Fluorescence *In Situ* Hybridization (FISH) estimated it composed 13.1% ±8.6 to 31.9% ±13.7 of bacterioplankton in the West and East regions of the Scotia Sea respectively.

The only isolated SAR116 representative, "*Candidatus* Puniceispirillum marinum", has been reported to have a versatile repertoire of genes for aerobic CO fixation, C1 metabolism and dimethylsulfoniopropionate degradation, suggesting it may occupy a "marine generalist" niche similar to that of SAR11 and some Roseobacters (Oh *et al.*, 2010). Proteins for ABC and TRAP transport and C1 metabolism with high matches to SAR116 bacteria were detected in both the summer and winter metaproteomes of coastal waters of the Antarctic Peninsula, consistent with a preference for labile compounds and C1 substrates (Williams *et al.*, 2012).

### Betaproteobacteria

The Betaproteobacteria are a large and cosmopolitan class with a range of ecological roles in the World Ocean (reviewed in Kirchman, 2008). While not found at high abundance (Gentile *et al.*, 2006; Ghiglione and Murray, 2011; Jamieson *et al.*, 2012), there is evidence that Betaproteobacteria perform significant ecological functions. Most known Ammonia-Oxidizing Bacteria (AOB) belong to the Betaproteobacteria (Head *et al.*, 1993; Teske *et al.*, 1994). Hollibaugh *et al.* (2002) detected *Nitrosospira*-like 16S rRNA sequences in Ross Sea and Antarctic Peninsula surface waters, and noted that the ribotype appeared similar to one found in the Arctic. Metagenomic and metaproteomic analyses of surface coastal waters off the Antarctic Peninsula show evidence of Calvin cycle carbon fixation and ammonia oxidation in winter performed by ammonia-oxidizing Betaproteobacteria (Grzymski *et al.*, 2012; Williams *et al.*, 2012).

The OM43 clade of Betaproteobacteria has been associated with coastal phytoplankton blooms (Morris *et al.*, 2006) and shown to be an obligate methylotroph capable of utilising methanol and formaldehyde as carbon and energy sources (Giovannoni *et al.*, 2008). As it has the smallest reported genome for a free-living cell, OM43 seems to be highly specialized for this unusual niche (the "genome streamlining;' hypothesis, Mira *et al.* (2001)). OM43 has been detected in a 16S rDNA library in a naturally fertilised bloom in the SAZ (West *et al.*, 2008), where it was the only betaproteobacterial representative, and in a metaproteomic survey of coastal waters on the AP, where methanol dehydrogenase from OM43 was detected (Williams *et al.*, 2012). Although the source of methanol in the marine environment is not yet clear, it may be a byproduct of phytoplankton growth (Heikes *et al.*, 2002) which would be consistent with OM43's observed association with coastal blooms. This possibility suggests OM43, and perhaps other C1 specialists, play an underexplored role in the marine microbial loop. Alternative sources are atmospheric deposition (Sinha *et al.*, 2007) or photochemical degradation of organic material (Dixon *et al.*, 2011). The latter is of particular interest in Antarctic waters, given the high levels of solar irradiation during the austral summer.

### Gammaproteobacteria

**SAR86** The gammaproteobacterial SAR86 clade is an abundant group in the surface ocean, being e.g. the most abundant genome for an uncultured organism in the Global Ocean Sampling (GOS) dataset (Dupont *et al.*, 2011). While it has been detected in the Southern Ocean (Abell and Bowman, 2005; Topping *et al.*, 2006; West *et al.*, 2008; Obernosterer *et al.*, 2011), little is known of its distribution or ecological role. (Topping *et al.*, 2006) estimated on the basis of FISH activity that SAR86 cells composed 7.8% ±8.2 and 18.3% ±17.0 of total bacterioplankton in the western and eastern Scotia Sea respectively, suggesting that at least in the SAZ it is a major component of the surface community. Genomic analysis of partial SAR86 genomes assembled from metagenomes found the clade have streamlined genomes and are specialized for utilizing lipids and carbohydrates, suggesting minimal competition between SAR86 and SAR11 for DOC (Dupont *et al.*, 2011). This may be reflected by the simultaneous high abundance and activity of SAR11 and SAR86 in the HNLC waters of the SAZ (Obernosterer *et al.*, 2011)

**OMG group** The term Oligotrophic Marine Gammaproteobacteria (OMG) was named for a group of physiologically diverse heterotrophs that belong to previously detected environmental rRNA clades

(OM60, BD1-7, KI89A, OM182, SAR92) (Cho and Giovannoni, 2004). Cultured OMG isolates have been shown to be obligately oligotrophic (Cho and Giovannoni, 2004). Nevertheless, SAR92 is associated with nutrient-rich waters with high phytoplankton abundances (Stingl *et al.*, 2007; Pinhassi *et al.*, 2004). Reports of SAR92 in the SO corroborate this ecology: both West *et al.* (2008) and Obernosterer *et al.* (2011) found SAR92-affiliated OTUs to be far more abundant inside the Kerguelen Ocean and Plateau Compared Study (KEOPS) phytoplankton bloom patch than in typically HNLC SAZ waters outside of it, with abundance declining as the bloom aged. This, combined with the observation that SAR92 growth is highly carbon-limited (Stingl *et al.*, 2007), suggests the clade plays an important role in degradation of organic carbon produced by phytoplankton blooms. It has also been detected in coastal AP and Kerguelen Islands waters (Ghiglione and Murray, 2011). Metaproteomic and metagenomic surveys of coastal waters at Palmer station found OMG to me more abundant in the summer than winter (Williams *et al.*, 2012). TonB-dependent receptor systems from OMG were highly abundant in the metaproteome, indicating that this is the preferred uptake system of ambient substrates (Williams *et al.*, 2012). Certain OMG strains encode proteorhodopsin (HTCC2207, Stingl *et al.* (2007); HTCC2143, Oh *et al.* (2010)), also indicated in the metaproteomic study in both seasons (Williams *et al.*, 2012).

**Ant4D3** In a study of six fosmids from nearshore waters at Palmer station, (Grzymski *et al.*, 2006) identified a uncultured gammaproteobacterium, Ant4D3. It has since been reported as one of the dominant proteobacterial groups in the SO. In waters off the western Antarctic Peninsula, Ant4D3 was reported to compose 10% of the total community and half the gammaproteobacterial community, and 68% of cells incorporating amino acids (Straza *et al.*, 2010). The authors also reported that the clade appears to have low diversity, based on detected rDNA sequences. Like SAR86, Ant4D3 cells were more active in HNLC than bloom conditions on the Kerguelen Plateau (West *et al.*, 2008). However, (Ghiglione and Murray, 2011) reported that 16.5% of tag-pyrosequenced 16S Denaturing Gradient Gel Electrophoresis (DGGE) bands from summer AP waters were affiliated to Ant4D3, dominating the Gammaproteobacteria and outnumbering winter and Kerguelen Island waters. (Murray *et al.*, 2011) similarly found Ant4D3 clones to be highly abundant in a 16S library from waters in the vicinity of Antarctic icebergs. Little is known about the group's function or ecological position, although it has been detected in Arctic waters where it appeared to occupy a DOM utilisation niche different from that of other major heterotrophs e.g. SAR11 (Nikrad *et al.*, 2012).

**GSO-EOSA-1** The GSO-EOSA-1 cluster of sulfur-oxidizing Gammaproteobacteria, which includes the uncultivated ARCTIC96BD-19 and SUP05 lineages as well as cultivated chemoautotrophic clam symbionts, has been reported in global mesopelagic waters (Swan *et al.*, 2011) and oxygen minimum zones (Walsh *et al.*, 2009; Canfield *et al.*, 2010). Three studies have recently identified GSO-EOSA-1 representatives at high abundance in coastal and AZ waters. A metagenomic survey of coastal waters at Palmer station found GSO-EOSA-1 winter bacterioplankton were dominated by Gammaproteobacteria (19.7% of the winter library compared to 2.7% of the summer library) falling into 5 closely-related 0.03 distance bins that were affiliated with the GSO-EOSA-1 complex. (Grzymski *et al.*, 2012), in a metaproteomic survey of coastal waters at Palmer station, found GSO-EOSA-1 proteins composed a large fraction of all gammaproteobacterial proteins detected and were significantly more abundant in winter than summer (20% vs 3%). In a companion metaproteomic analysis of the same sites, (Williams *et al.*, 2012) confirmed this high abundance and seasonal pattern, although GSO-EOSA-1 appeared to be metabolically active at the surface in both summer and winter.

Genomic and metaproteomic analyses of GSO-EOSA-1 representatives, particularly SUP05, have revealed the potential for carbon fixation via the Calvin cycle and sulfur oxidation, even in well-oxygenated waters (Walsh *et al.*, 2009; Swan *et al.*, 2011; Grzymski *et al.*, 2012). (Grzymski *et al.*, 2012) estimated from rRNA abundances that between 18 and 37% of the winter bacterioplankton community comprises OTUs with the potential for chemolithoautotrophy, including GSO-EOSA-1, suggesting winter chemolithoautotrophy may contribute significantly to SO carbon fixation.

### Deltaproteobacteria

Deltaproteobacteria are rarely detected at abundance in global surface waters (see e.g. Venter *et al.*, 2004), and this pattern appears to hold in the Southern Ocean (Murray and Grzymski, 2007; West *et al.*, 2008; Ghiglione and Murray, 2011; Murray *et al.*, 2011; Ducklow *et al.*, 2011; Jamieson *et al.*, 2012).

However, they may increase in abundance in mesopelagic waters (Wright *et al.*, 1997; Pham *et al.*, 2008; Zaballos *et al.*, 2006). At a 3000 m deep site at the PF in the Drake Passage, López-García *et al.* (2001) detected several deltaproteobacterial 16S rDNA sequences, all of which clustered with the marine deltaproteobacterial clade SAR324 previously identified in the mesopelagic Sargasso Sea (Wright *et al.*, 1997). Whole-genome analysis of SAR324 indicates an ecology that includes carbon fixation via the Calvin cycle and sulfur oxidation, as well as oxidation of methylated compounds (Swan *et al.*, 2011). SAR324 may therefore be significant contributors to chemoautotrophy in the dark ocean (Swan *et al.*, 2011).

**CFB**

Bacteria of the group Cytophaga-Flavobacterium-Bacteroides (CFB) are cosmopolitan and abundant in the world ocean (Glöckner *et al.*, 1999). While the CFB often form a major fraction of planktonic taxa (Fandino *et al.*, 2001), they are particularly prevalent in particle-attached communities (DeLong *et al.*, 1993) and are associated with blooming phytoplankton (Pinhassi *et al.*, 2004). Isolated CFB representatives have a well-described aptitude for the degradation of HMW DOM, particularly biopolymers which may be recalcitrant to utilisation by other bacterial heterotrophs (reviewed in Kirchman, 2002), suggesting they play an important role in remineralisation of primary production products. Of the CFB, the class Flavobacteria seem to be in the majority worldwide in both freshwater and marine environments (O'Sullivan *et al.*, 2004; Cottrell *et al.*, 2005) including the Southern Ocean (Abell and Bowman, 2005).

CFB in the SO are strongly biogeographically partitioned. Abell and Bowman (2005), utilising DGGE and 16S sequencing with Flavobacteria-specific primers, found significantly higher abundance and diversity of particle-attached Flavobacteria in the nutrient- and phytoplankton-rich waters south of the PF relative to the HNLC waters of the Subantarctic. This difference in abundance may be largely attributable to the low iron availability in the Subantarctic, which probably limits primary production (Boyd *et al.*, 2007). Both natural and artificial iron fertilization events in the Subantarctic have resulted in high abundances of bacterial heterotrophs (Christaki *et al.*, 2008; Oliver *et al.*, 2004); West *et al.* (2008), identified the CFB as a major component of the bacterial response to blooms induced by natural iron input on the Kerguelen plateau. The higher abundance of CFB in the AZ may also relate to their prevalence in sea ice (Brinkmeyer *et al.*, 2003; Brown and Bowman, 2001), from which they would be released into AZ waters during seasonal melting. Two groups, the uncultured agg58 cluster and the genus *Polaribacter*, appear to dominate CFB populations and activity in the SO (Murray and Grzymski, 2007; Abell and Bowman, 2005; **?**; Obernosterer *et al.*, 2011; West *et al.*, 2008; Ghiglione and Murray, 2011; Ducklow *et al.*, 2011; Straza *et al.*, 2010).

There is some evidence that planktonic and particle-attached CFB, rather than being an integrated population with cells opportunistically shifting between phases, may comprise at least partially distinct groups of phylotypes. In a mesocosm experiment examining colonisation of diatom detritus in SO seawater, 16S DGGE and sequencing analysis showed a large proportion of flavobacterial phylotypes present in the planktonic phase failed to colonise detrital particles during the course of the experiment (**?**). The authors suggest these phylotypes may be slower-growing, perhaps comprising a secondary group of colonisers which only come to dominate when the more accessible detrital nutrients have been exhausted and the primary colonisers have secreted useful secondary metabolites. Alternatively, some flavobacterial groups may not use particle attachment as a primary strategy. Questions around the relationship between particle-attached and free-living microbial communities emphasise the usefulness of size fractionation in molecular studies of marine microbial communities.

Kirchman (2002) suggests 16S clone libraries may systematically underestimate the abundance of CFB in environmental samples, noting that in two studies where both FISH and 16S analysis were employed at the same site there were common discrepancies in CFB abundance estimates between the two methods (Cottrell and Kirchman, 2000; Eilers *et al.*, 2000). Additionally, both FISH and PCR based methods may underestimate CFB abundance relative to metagenomic surveys, due to probe specificity biased against Bacteroidetes 16S rDNA (Cottrell *et al.*, 2005; O'Sullivan *et al.*, 2004).

*Polaribacter* is a gas-vacuolated, proteorhodopsin-expressing flavobacterial genus prevalent in Arctic and Antarctic seawater, and the genome indicates a preference for polymers obtained from algal detritus rather than labile exudates (e.g. taurine, polyamines) (González *et al.*, 2008). An Antarctic Peninsula coastal metagenome found *Polaribacter*-related sequences to be dominant in summer, con-

sistent with an association with phytoplankton blooms and/or being seeded from melting sea-ice (Grzymski *et al.*, 2012). Flavobacterial proteins (including those with the best matches to *Polaribacter* spp.) were similarly much more abundant in the summer versus winter metaproteome from the same sites, with components of TonB-dependent receptor systems predominating (Williams *et al.*, 2012).

### Cyanobacteria

Cyanobacteria, dominated by the genera *Prochlorococcus* and *Synechococcus*, are the most abundant photosynthetic organisms on Earth (Scanlan *et al.*, 2009, and references therein), but little molecular research has been performed on their role in SO ecosystems. This may be because it has been generally accepted that there are no cyanobacteria in AZ waters (Ghiglione and Murray, 2011; Zubkov *et al.*, 1998; Evans *et al.*, 2011), although recent and metaproteomic results (Williams *et al.*, 2012) challenge this assumption. It is not infeasible that cyanobacteria survive at Antarctic temperatures, as (apparently psychrophilic or psychrotolerant) *Synechococcus* and *Prochlorococcus* strains have been identified in several marine-derived Antarctic lakes, including at sub-zero water temperatures (Bowman *et al.*, 2000; Powell *et al.*, 2005; Lauro *et al.*, 2011). Regardless, it is clear that cyanobacteria, if present in the AZ, are at very low abundance and probably of little ecological significance. Cyanobacteria also appear to be at low abundance in the SAZ (Abell and Bowman, 2005; Topping *et al.*, 2006).

### Verrucomicrobia

Verrucomicrobia is a recently described phylum that is ubiquitous in the marine environment, and appears to be composed of several physiologically distinct lineages (Freitas *et al.*, 2012). A small number of representatives of Verrucomicrobia have been detected in the SO (Murray *et al.*, 2011; West *et al.*, 2008; Gentile *et al.*, 2006; Murray and Grzymski, 2007), and Ghiglione and Murray (2011) reported a higher abundance of 16S rDNA clones affiliating with the Verrucomicrobia at a Kerguelen Island site relative to a site near Palmer Station on the Antarctic Peninsula. Little else is known regarding their abundance, diversity or ecological role in the SO.

### Other bacteria

Bacteria of the phylum Planctomycetes have been detected at low abundance in SO molecular surveys (Gentile *et al.*, 2006; López-García *et al.*, 2001; Jamieson *et al.*, 2012; Murray *et al.*, 2011; Abell and Bowman, 2005). Planctomycetes is emerging as a group of interest in marine microbial ecology, for example as performers of Anaerobic Ammonium Oxidation (anammox) (Strous *et al.*, 1999), such as indicated in the metaproteome from coastal West Antarctic waters (Williams *et al.*, 2012). The latter study also detected Nitrospirae proteins pertaining to nitrite oxidation and carbon fixation via the reductive tricarboxylic acid cycle. Members of the Nitrospirae and Planctomycetes are therefore implicated in completing nitrification using nitrite generated by AOA and AOB in Antarctic waters.

Other bacterial groups have been reported at low abundance in SO waters, including Actinobacteria (Bowman and McCuaig, 2003; Brinkmeyer *et al.*, 2003; Abell and Bowman, 2005; Gentile *et al.*, 2006; Murray and Grzymski, 2007; Murray *et al.*, 2011; Ghiglione and Murray, 2011; Piquet *et al.*, 2011; Jamieson *et al.*, 2012), Epsilonproteobacteria (Gentile *et al.*, 2006; Murray and Grzymski, 2007), and Firmicutes (Murray and Grzymski, 2007; Lo Giudice *et al.*, 2011; Murray *et al.*, 2011). Little is known about their respective ecological roles, although Actinobacteria have been associated with marine aggregates (Grossart *et al.*, 2004); interestingly, their terrestrial counterparts have diverse HMW substrate degradation capabilities (reviewed in Kirchman, 2008). A strong negative correlation has been reported between actinobacterial abundance and latitude in a global survey using 16S rDNA clone libraries (Pommier *et al.*, 2007), with higher abundances in tropical and subtropical waters, as for Cyanobacteria.

## Archaea

DeLong *et al.* (1994) first reported the high abundance (up to 34%) of archaea in Antarctic coastal surface waters, a surprising discovery at a time when archaea were generally considered a rare group of strict extremophiles. The majority of rDNA clones they identified were affiliated with the Marine

Group I Crenarchaeota (MGI), while the remainder represented the Group II Euryarchaeota. Subsequent rRNA-based studies are likewise in agreement that MGI are the dominant group of archaea in surface waters of coastal Antarctica, followed by Group II Euryarchaeota (Gerlache Strait, Massana *et al.* (1998); near Anvers Island, Murray *et al.* (1998)1998). Further rRNA-based analysis showed the widespread distribution of Antarctic marine archaea both longitudinally as well north and south of the polar front (Murray *et al.*, 1999; Topping *et al.*, 2006; Jamieson *et al.*, 2012), and the identification of a significant MGI community in benthic sediments on the Antarctic coast (Bowman and McCuaig, 2003).

For MGI, ammonia-oxidizing chemolithoautotrophy is likely the dominant metabolic lifestyle (Ingalls *et al.*, 2006; Berg *et al.*, 2007), suggesting they play major roles in nitrification and carbon fixation in the SO. In a winter coastal Antarctic metaproteome, MGI proteins made up 30% of all identified proteins from bacteria or archaea; no MGI proteins were detected in the summer metaproteome (Williams *et al.*, 2012). The winter metaproteome included MGI proteins pertaining to the 3-hydroxypropionate/4-hydroxybutyrate cycle, the pathway used by ammonia-oxidizing MGI for carbon fixation, and for ammonium transport and oxidation, supporting a nitrification role for Southern Ocean MGI (Williams *et al.*, 2012). The complementary metagenomic analysis of Grzymski *et al.* (2012) proposed chemolithoautotrophy carried out by ammonia-oxidizing MGI and sulfur-oxidizing Gammaproteobacteria (see GSO-EOSA-1, above) to be the major drivers of winter carbon fixation in AZ waters. In summer autotrophic carbon assimilation is driven by algal-driven oxygenic photoautotrophy, consistent with high light availability and intensity, whereas in the polar winter "dark" chemoautotrophy by archaea and bacteria plays a major role in carbon fixation.

Murray *et al.* (1998) found that total archaeal rRNA levels decreased during summer, and noted a negative correlation between archaeal rRNA levels and chlorophyll *a* concentration. Massana *et al.* (1998) also observed a decline in archaeal rRNA levels during spring. Church *et al.* (2003) found a significantly higher (44% increase) abundance of MGI in surface waters in winter compared to summer. Ammonia-oxidizing MGI have been shown to be especially sensitive to photoinhibition (Merbt *et al.*, 2012), which might account for their decline during periods of extended illumination. It has also been speculated that the decline of archaea during spring/summer represents competition with non-archaeal microbes during phytoplankton blooming (Massana *et al.*, 1998), or that the majority of MGI were chemoautotrophic and therefore more competitive compared to heterotrophs during carbon-scarce winter conditions (Murray *et al.*, 1998). However, based on genomic evidence, Marine Group II Euryarchaeota are motile, proteorhodopsin-expressing photoheterotrophs that specialize in protein and lipid degradation (Iverson *et al.*, 2012). This is consistent with results that this group, in contrast to MGI, was more relatively abundant at the surface than at depth (Massana *et al.*, 1998). Murray *et al.* (1998) noted an increase in Group II Euryarchaeota in autumn in waters off Anvers Island; but otherwise the seasonal distribution of this group in Antarctic waters is not well understood.

The numerical dominance of MGI over other archaeal groups in surface and photic zone waters has also become well established (e.g. DeLong *et al.*, 1994; Massana *et al.*, 2000), although not in aphotic waters; López-García *et al.* (2001) detected only euryarchaeaotal sequences in a sample from 3000 m depth at the PF in the Drake Passage, including marine groups II and III and a novel marine group IV. However, these studies also illustrated a potential hazard of probe-dependant methods, namely the high variability of abundance estimates depending on probe design. For example, (Simon *et al.*, 1999) did not detect any DAPI-positive archaeal cells in a summer transect between the polar front and ice edge using the archaea-specific probes ARCH334 and ARCH915, in contrast to the results of several other studies reviewed herein. López-García *et al.* (2001) detected only one archaeal phylotype (an euryarchaeon) in their initial clone library constructed with one archaeal primer pair. This prompted the authors to design an additional five primer pairs, resulting in both a higher number and greater diversity of clones.

## Virioplankton

The "viral shunt", by which nutrients are released via lysis from marine microorganisms and returned to the dissolved and particulate pools, may mediate the flux of a quarter of all organic matter in the microbial loop (Wilhelm and Suttle, 1999) and the viral release of iron from bacterioplankton may be crucial for phytoplanktonic growth (Poorvin *et al.*, 2004). Viral production, and by inference the viral shunt, has been shown to be highly active in HNLC Subantarctic (Evans *et al.*, 2009), iron-

fertilized Subantarctic (Weinbauer *et al.*, 2009) and coastal waters, where viral-mediated carbon flux may account for 50–100% of all heterotrophic production (Guixa-Boixereu *et al.*, 2002). Despite this crucial ecosystem role, however, molecular analysis of the diversity and function of SO virioplankton has been sparse. Two studies (Short and Suttle, 2002, 2005) used probes with specificity to algal virus and cyanophage marker genes respectively, and succeeded in detecting both in SO waters. (Williams *et al.*, 2012) and (Grzymski *et al.*, 2012), in complementary metaproteomic and metagenomic studies of sites near Palmer station on the Antarctic Peninsula, also detected cyanophage (cyanobacterial virus) genes and proteins as well as a single major capsid protein from *Phaeocystis pouchetii* virus PpV01. While these studies are only preliminary, they suggest that the more abundant viruses are phytoplanktonic predators. An extensive molecular survey of SO virioplankton (in the nature of Angly *et al.*, 2006) would clearly be of great value.

# Project questions and hypotheses

## Question 1: Is the PF a major boundary in the biogeography of SO picoplankton?

The PF is a major current core of the ACC, and represents the surface transition between CDW which dominates the AZ and SAMW which dominates the SAF and northwards. Previous studies (e.g. Abell and Bowman, 2005; Giebel *et al.*, 2009; Selje *et al.*, 2004) have suggested the PF is a biogeographical barrier for certain species of picoplankton, but this has not yet been established on the community level, a necessary prerequisite for a comparative study of the two zones.

**Hypothesis:** the picoplanktonic communities to the north and south of the PF are significantly different.

## Question 2: How do the picoplanktonic communities on either side of the PF differ?

In order to predict the effects of climate change-driven changes in the location and properties of the ACC on the SO's microbial ecosystems, the picoplanktonic communities to the north and south must be characterised.

## Question 3: How do the ecosystem functions performed by picoplankton on either side of the PF differ?

Picoplankton in the SO perform numerous important ecosystem functions, for example primary production and heterotrophic utilisation of recalcitrant organic compounds. This project will seek to characterise those ecosystem functions, and in particular the ways in which those functions differ between the two zones.

## Question 4: To what relative extents do water circulation and physicochemical properties define picoplanktonic biogeography?

Findings flowing from the three previous questions will be useful in predicting the effects of climate change-driven change in the physical oceanography of the SO. However, an underlying assumption has been that the distribution and abundance of picoplankton is determined solely by the physical properties of an environment; in other words, that "*everything is everywhere*, but, *the environment selects*" (quotation from Baas Becking (1934); English translation from de Wit and Bouvier (2006)). It is possible, however, that the physical transport of picoplankton through circulation also plays a role. The final question of this project seeks to distinguish between three hypotheses:

**Hypothesis 1:** The distribution of picoplankton in the SO is determined only by the physicochemical properties of the water they inhabit; "*everything is everywhere*, but, *the environment selects*".

**Hypothesis 2:** The distribution of picoplankton in the SO is determined only by the circulation of the water they inhabit; "wherever you go, that's where you are" (Bissett *et al.*, 2010).

**Hypothesis 3:** Both physical transport and environmental selection determine the distribution of SO picoplankton.

# MINSPEC, a bioinformatic tool for metagenomics

Sections of this chapter have been previously published in Wilkins D., Lauro F. M., Williams T. J., DeMaere M. Z., Brown M. V., Hoffman J. M., Andrews-Pfannkoch C., McQuaid J. B., Riddle M. J., Rintoul S. R., and Cavicchioli R. (2012). Biogeographic partitioning of Southern Ocean picoplankton revealed by metagenomics. *Molecular Ecology*.

## Summary

## Introduction

### Metagenomic analysis of microbial assemblages

The identification of the species or Operational Taxonomic Units (OTUs) that compose a microbial community is a primary aim of metagenomics. Typically this is achieved using one of two methods.

The first method is the identification, using a search and alignment algorithm such as BLAST, of specific marker genes or other sequences which are diagnostic for a particular species or OTU. Common targets in microbial ecology are the 16S or other ribosomal subunit rDNA sequences, and the Internal Transcribed Spacer (ITS) regions between 16S–23S rDNA sequences (e.g. Brown *et al.*, 2012). This method provides several advantages. The selected regions are usually highly conserved, and through cultivation and full-genome sequencing have been reliably associated with a particular OTU, allowing very accurate identification and analysis of diversity down to the ecotype level (e.g. Brown *et al.*, 2012). If the copy number of the gene or region is well known, this method also allows for accurate estimations of cell abundance from metagenomes. However, a disadvantage of this method is that the large majority of metagenomic reads which do not happen to cover the region of interest will contribute nothing to the analysis and essentially be wasted. Low-abundance OTUs will therefore be missed, as even if they generate a small number of reads, those reads are unlikely to cover the region of interest.

The second method is to compare assembled or unassembled metagenomic reads to a reference database, using an algorithm such as BLAST, then use probabilistic methods to assign identifications and abundances with varying degrees of confidence. Most commonly, the reads are compared to a database of full genomes (e.g. Lauro *et al.*, 2011; Qin *et al.*, 2010). This method makes more efficient use of metagenomic data compared to the first, as any read can potentially yield a BLAST match and thus contribute to the identification of an OTU. However, interpretation of the results, and particularly calculation of abundances, is more complex. For example, the software tool Genome relative Abundance and Average Size (GAAS) makes use of BLAST match quality, number of matches and estimated genome size to estimate the relative abundances of OTUs in a sample (Angly *et al.*, 2009).

Such estimates are confounded by the presence of multiple OTUs which can generate high-quality BLAST matches ("hits") to a given read. Multiple high-quality hits to a single read are the norm, rather than the exception, in metagenomic studies for several reasons. A microbial assemblage will often include a number of closely-related OTUs (e.g. congeners) which share large sections of highly similar or identical genomic sequence. If several such OTUs are present in the reference database, a metagenomic read from one will yield high-quality BLAST hits to them all. Further, even distantly

**Table 1:** Selected examples of species identified in a marine metagenome using the naïve method. These species were identified in a single sample from the SO (sample 346; see "*The Polar Front as a major biogeographic boundary in the Southern Ocean*"). The sample was compared to the RefSeq database of full genomes using TBLASTX with an E-value maximum of $1.0 \times 10^{-3}$, i.e. only high-quality hits were included. Relative abundances were calculated using GAAS (Angly *et al.*, 2009).

| **Species** | **Relative Abundance (%)** | **Notes** |
|---|---|---|
| Encephalomyocarditis virus | 1.98 | Human pathogen. |
| Marek's disease virus type 1 | 1.49 | Chicken pathogen. |
| Marek's disease virus type 2 | 0.85 | Chicken pathogen. |
| *Francisella philomiragia* | 0.041 | Human and animal pathogen. |
| *Agrobacterium vitis* | 0.040 | Plant and opportunistic human pathogen. |
| *Brucella suis* | 0.011 | Human and swine pathogen (causes brucellosis). |
| *Enterobacter* sp. 638 | 0.0085 | Animal commensal/pathogen. |
| *Bordetella parapertussis* | 0.0075 | Mammalian pathogen (causes mild form of whooping cough). |
| *Neisseria meningitidis* | 0.0074 | Human pathogen. |
| *Yersinia pestis* | 0.0060 | Human/animal pathogen (causes bubonic plague). |

related OTUs are likely to share large regions of identity, and the selection of hit quality thresholds to discriminate between them (for example, a minimum bit score or maximum expectation value) is effectively arbitrary. Thus, while metagenomic studies using whole-genome comparisons almost always use such thresholds as the sole discriminators between OTUs, this method (hereafter the "naïve" method, after Ye and Doak (2009)) will almost inevitably result in the identification of OTUs which are not present in the assemblage, skewing the relative abundance estimates of those which are truly present.

This problem is compounded by a systematic overrepresentation within databases of reference genomes of species of interest to humans, such as human and agricultural pathogens. Environmental OTUs are comparatively underrepresented. For example, Table 1 gives examples of terrestrial plant and animal pathogens, *a priori* unlikely to be truly present, which were identified in an open ocean metagenome using the naïve method.

## The maximum parsimony approach

Ye and Doak (2009) identified an analogous problem in the annotation of biochemical pathways in genomes and metagenomes. They noted that a common method is to annotate a pathway as present if a single protein within that pathway attracts at least one high-quality BLAST hit. However, because many proteins are shared by multiple pathways, and databases of orthologous genes are often incomplete, this method has resulted in many clearly spurious annotations, such as an ascorbic acid synthesis pathway in the human genome (humans require dietary vitamin C) and a mitochondrial pathway in *Escherichia coli* (annotated in the Kyoto Encyclopedia of Genes and Genomes (KEGG) PATHWAY database).

The authors developed a software tool, MINPATH, to combat this problem and increase the accuracy and fidelity of pathway annotations. MINPATH computes the smallest possible set of pathways ("maximum parsimony") sufficient to explain a set of annotated proteins. As a simple example, if a genome is annotated with all the proteins which belong to pathway A, and one of those proteins also happens to belong to pathway B — that is, it is shared by both pathways — the naïve approach would annotate both pathways as present. However, the most parsimonious explanation is that pathway A is present, and B is not.

MINPATH was implemented by framing the construction of a maximum parsimony pathway set as an Integer Programming (IP) problem. IP is a subset of algorithms for solving Linear Programming (LP) problems, which seek to maximise the value of a linear function (the objective function) within a set of constraints. In this case, the objective function was maximised by decreasing the number of annotated biochemical pathways, while the constraint was that every high-quality protein annotation

had to be represented at least once in the annotated pathways. Validation and testing of MinPath showed it was successful in eliminating spuriously annotated pathways while retaining those genuinely present.

It was noted that this as this problem is essential isomorphic with that of spurious annotations in microbial metagenomes, the "maximum parsimony" method would likely also work in this domain. The aim of the project described in this chapter was thus to develop and test a software tool, MINSPEC, which would find the most parsimonious set of species necessary to explain a set of observed BLAST hits generated by a metagenome, using the approach of (Ye and Doak, 2009) as a model.

# Methods

## Implementation of MINSPEC

A computational method to minimise false OTU identifications and increase the accuracy of OTU abundance estimates (MINSPEC) was developed and implemented in PERL[1]. Following the approach of Ye and Doak (2009) to the parsimonious reconstruction of biochemical pathways (MinPath), MINSPEC computes the smallest set of OTUs sufficient to explain a set of observed high-quality hits against RefSeq (or any other sequence database). The minimal set computation was framed as a IP problem and solved with GLPSOL (The GNU Linear Programming/MIP solver) (Free Software Foundation, Boston).

The objective function for the IP problem was constructed as follows (adapted from Ye and Doak, 2009):

$$\min \sum_{j=1}^{s} A_j$$

where $s$ is the number of OTUs in the assemblage, and $A_j = 1$ if OTU $j$ is in the assemblage, 0 if not. In other words, the objective function is satisfied by minimising the number of OTUs in the assemblage. The constraint function was constructed as follows (adapted from Ye and Doak, 2009):

$$\sum_{j=1}^{s} M_{ij} A_j \geq 1 \quad \forall i \in [1, n]$$

where $M_{ij} = 1$ if read $i$ has a mapping (i.e. a high-quality BLAST hit) to OTU $j$, and $[1, n]$ is the set of all reads. In other words, the constraint function fails if any read does not have at least one of its high-quality BLAST hits represented in the assemblage.

This approach eliminates many of the spurious OTU identifications which result from reads with strong identity to more than one OTU. The "minimal species set" is liable to exclude some low-abundance OTUs, but gives more faithful abundance estimates and eliminates many false positives.

It was noted that in some special cases, it may be desirable to include an OTU in the assemblage even if it is not part of the minimal set, if that OTU generated a very large number of BLAST hits. An example of such a situation might be if the sample was known with certainty to contain a two very closely related OTUs at roughly equal abundance. In such a case, it would be expected that almost all metagenomic reads generated by each of these OTUs would also attract BLAST hits to the other, and MINSPEC would thus probably eliminate whichever happened to generate slightly fewer hits. To allow for this, an option was added such that MINSPEC will not eliminate OTUs to which a specified threshold number of reads attract high-quality hits.

## Validation of MINSPEC

To establish the usefulness of MINSPEC, a validation method was devised to experimentally determine its error rates and efficacy (i.e. number of spurious OTUs identified and removed).

A set of simulated microbial OTUs was generated. To simulate genomic sequence identity between OTUs, each simulated OTU went through up to fifty rounds in which another OTU was selected at

[1] MINSPEC and the associated metagenomic simulation and validation scripts are open source and available at `https://github.com/wilkox/minspec`; a copy has also been provided in the supplementary information.

random and marked as having sequence identity with the first. This process was terminated with a 10% probability at each round, simulating an exponential curve of interrelatedness between OTUs. A random subset of the simulated OTUs were then selected to form a simulated microbial assemblage. Because of the previously established simulated sequence identity between OTUs, some OTUs in the assemblage would be marked as having identity to other OTUs both within the assemblage and outside of it.

A simulated metagenomic sampling was then performed. In each round, an OTU was selected at random. To produce a natural rank-abundance curve of OTU abundance within the assemblage, the probability that the selected OTU yielded a read was

$$\frac{1}{ln(x)+1}$$

where $x$ is the OTU's rank. Simulated BLAST matches to the OTU were generated for the read. These matches would include accurate high-quality "genuine" hits to the OTU that produced the read, as well as to other randomly selected OTUs both within and out of the assemblage which had been previously marked as having sequence identity to the "genuine" OTU.

To fully explore the limits and reliability of MINSPEC, the simulated metagenomic experiment described above was performed with all possible permutations of the following parameters: number of simulated OTUs [100; 1,000; 10,000; 50,000; 100,000]; size of simulated assemblage [1; 10; 100; 300; 500; 1,000; 10,000]; number of simulated metagenomic reads [10; 100; 1,000; 10,000; 100,000; 200,000; 500,000]. Each permutation was repeated five times, except for those where the size of the assemblage would exceed the number of OTUs simulated.

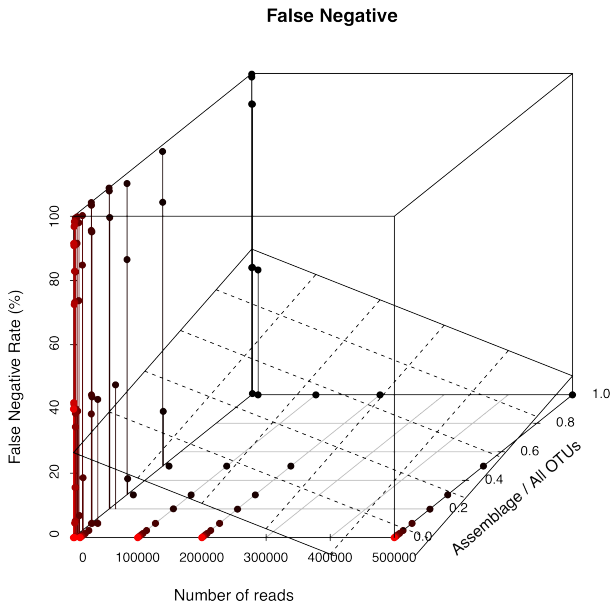
The resulting simulated BLAST outputs were processed with MINSPEC, and the false positive (percentage of OTUs not in the assemblage which nevertheless survived MINSPEC filtering) and false negative (percentage of OTUs present in the assemblage which were not present after MINSPEC filtering) rates calculated. Because a high false negative rate can arise from undersampling, a problem in metagenomic studies both real and simulated, an additional "false negative (MINSPEC)" metric was calculated, which excluded OTUs which were present in the assemblage but through random chance did not generate any reads, the equivalent of "unsampled rare taxa". This rate thus represented only false negatives attributable to MINSPEC itself. Finally, as a measure of MINSPEC's usefulness, the proportion of "false OTUs" — OTUs that generated BLAST matches but were not part of the assemblage — successfully removed by MINSPEC was calculated.

# Results

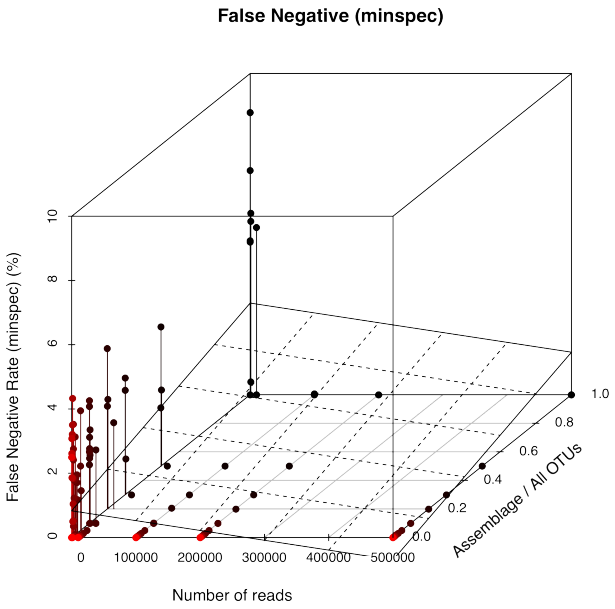
Repeated simulated metagenomic experiments with a wide range of permutations of parameters showed that MINSPEC was reliable and able to substantially reduce the rate of false positive OTU identifications, although its effectiveness varied with the parameters of the assemblage and metagenomic experiment (Figure 2).

# Discussion

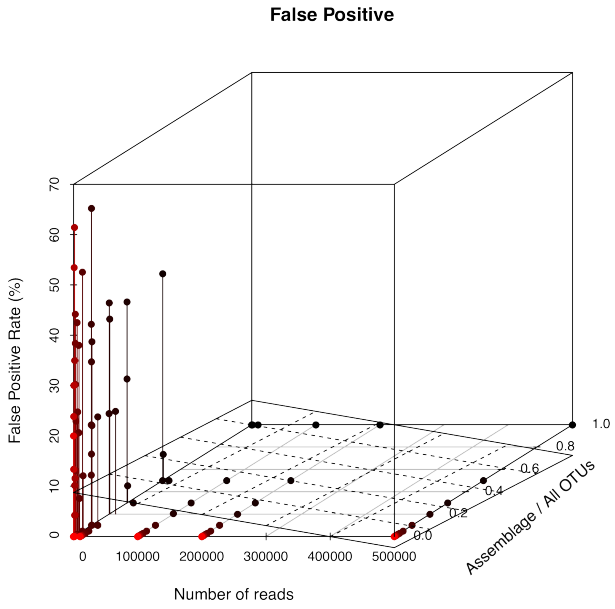
The false negative rate, or percentage of OTUs in the simulated assemblage which were absent from the BLAST results following MINSPEC processing, was generally high, ranging from $\sim$ 20% under ideal conditions (a low assemblage / all OTUs ratio, and 500,000-read metagenomic sample) to $\sim$ 90% in the worst case (a high assemblage / all OTUs ratio and a small metagenomic sample) (Figure 2a). The assemblage / all OTUs ratio (hereafter referred to as "assemblage ratio") indicates the proportion of simulated OTUs ("all OTUs") that were chosen to form the simulated assemblage. A higher ratio means that any OTU is more likely on average to be part of the assemblage, and thus that any individual failure to detect a OTU is an error. This problem is mitigated with increasing the number of reads, as this makes it less likely that a given OTU would go unsampled. The extreme false negative rates, in some cases 100%, represent extreme simulated scenarios (e.g. an assemblage of 1 OTU drawn from a pool of 100,000), and thus do not reflect real metagenomic studies.



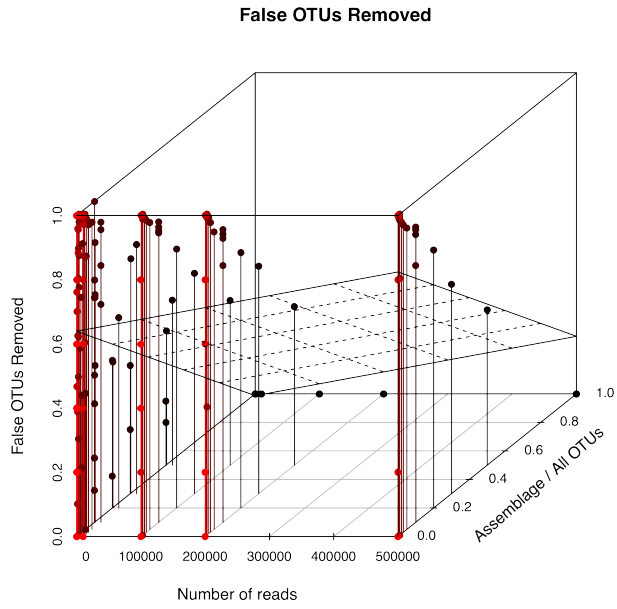

**(a)** False negative rate (%) — the percentage of OTUs in the assemblage that were absent from the BLAST results following MINSPEC processing.




**(b)** MINSPEC-attributable false negative rate (%) — the percentage of OTUs not in the assemblage that generated BLAST hits but were incorrectly removed by MINSPEC.




**(c)** False positive rate (%) — the percentage of OTUs not in the assemblage that were present in the BLAST results following MINSPEC processing.




**(d)** Proportion of false OTUs — OTUs that were not part of the simulated assemblage but which generated hits due to simulated sequence identity — that were correctly identified and removed by MINSPEC.

**Figure 2:** Results of repeated trials of MINSPEC on simulated metagenomic studies with multiple permutations of parameters (number of reads, number of simulated OTUs, size of simulated assemblage). The number of simulated OTUs and size of simulated assemblage are represented as a ratio on the z-axis ("assemblage / all OTUs"). Each permutation was repeated five times. A plane representing a linear regression has been overlaid on each plot to indicate the trend. Points have been tinted to aid the perception of depth; colour is not otherwise meaningful.

Because the majority of false negatives are attributable to undersampling and failure of OTUs to generate BLAST hits — properties the simulated metagenomic experiments share with real ones — a second metric, the false negative (MINSPEC) rate, was calculated (Figure 2b). This is the proportion of OTUs in the assemblage that generated BLAST hits, but were incorrectly removed by MINSPEC. This rate thus represents error attributable only to MINSPEC. The false negative (MINSPEC) rate was generally low, ranging from $\sim$ 0–1% for low assemblage ratios, to $\sim$ 15–20% under high ratios. Surprisingly, increasing the number of reads only slightly decreased the rate, at both low and high assemblage ratios. This suggests MINSPEC is more affected by the degree of similarities between OTUs than by undersampling.

The false positive rate, or percentage of OTUs not in the assemblage which nevertheless generated high-quality BLAST matches that were not identified and removed by MINSPEC, was generally $\sim$ 0–5% except for extremely small read sets and low assemblage ratios, where it reached as high as 60% (Figure 2c). These results reinforce the value of larger read sets, and show that once a modest metagenome size is reached ($\sim$ 100,000 reads) very few false positives can be expected.

The proportion of false OTUs removed was calculated to measure MINSPEC's efficacy in identifying and eliminating OTU which are not part of the sampled assemblage yet generate high-quality BLAST matches. This rate varied from 0–1 depending on the parameters of the assemblage (**??**). For simulations with a low assemblage ratio, the proportion was generally high ($> 0.6$), although there were simulated experiments with a low ratio where the proportion was low or zero. However, in all simulations with an assemblage ratio of 1, the proportion was 0, and the regression indicated a generally inverse relationship between the ratio and the proportion of false OTUs removed. This is likely because in assemblages with a higher assemblage ratio, there are fewer false OTUs to remove; in assemblages with a ratio of 1, there are none. The high proportion of false OTUs correctly identified in simulations with a low assemblage ratio is thus a good indication that MINSPEC is effective at identifying and removing false OTUs, especially as this proportion far exceeds the false positive and false negative (MINSPEC) rates for comparable experiments. As expected, increasing the number of reads improved MINSPEC's accuracy.

Overall, the simulated experiments validated both the accuracy and usefulness of MINSPEC as a tool for reducing error in metagenomic studies. It is worth noting that the assemblage ratio is not an inherent property of an assemblage, although it is limited by the assemblage's species richness. Rather, it can be decreased, and thus the accuracy of the metagenomic experiment improved, by performing BLAST searches against larger databases with finer taxonomic resolution. These results thus reinforce the value of both large read sets and comprehensive reference databases in obtaining high-quality metagenomic results.

# The Polar Front as a major biogeographic boundary in the Southern Ocean

Sections of this chapter have been previously published in Wilkins D., Lauro F. M., Williams T. J., DeMaere M. Z., Brown M. V., Hoffman J. M., Andrews-Pfannkoch C., McQuaid J. B., Riddle M. J., Rintoul S. R., and Cavicchioli R. (2012). Biogeographic partitioning of Southern Ocean picoplankton revealed by metagenomics. *Molecular Ecology*.

## Summary

## Introduction

The Southern Ocean (SO) consists of zones separated by circumpolar fronts, the locations of which vary with time and longitude (Whitworth, 1980; Orsi *et al.*, 1995; Sokolov and Rintoul, 2002). From north to south, the major fronts are the Subtropical Front (STF), the Subantarctic Front (SAF) and the Polar Front (PF) (Figure 1). The SAF and PF are associated with the Antarctic Circumpolar Current (ACC), the world's largest ocean current and a defining oceanographic feature of the SO. Anthropogenic climate change may be driving the warming and freshening of the ACC (Böning *et al.*, 2008). It may also be shifting the ACC and associated fronts poleward; the mean path of the current has moved $\sim$ 50 km south since the 1950s (Gille, 2002). If this trend continues, by 2100 this migration will have reduced the volume of water south of the ACC by approximately the equivalent of the entire Arctic Ocean (Fyfe and Saenko, 2005).

Climate change may also be affecting the zones separated by these fronts. The major zones of the SO are the Subantarctic Zone (SAZ), between the STF and SAF; the Polar Frontal Zone (PFZ), between the SAF and the PF; and the Antarctic Zone (AZ), between the PF and the Antarctic continent (Figure 1). These zones have different physicochemical properties, such as density, salinity, temperature and nutrient concentrations (Sokolov and Rintoul, 2002), and the fronts represent stepwise transitions in these properties (Whitworth III and Nowlin Jr., 1987). However, as a consequence of climate change, waters on the poleward side of the ACC have become warmer and more saline, while those to the north cooler and fresher (Böning *et al.*, 2008).

The PF has been suggested to be a major biogeographical boundary in the distribution and abundance of both zooplankton (Chiba *et al.*, 2001; Hunt *et al.*, 2001; Esper and Zonneveld, 2002; Ward *et al.*, 2003) and bacterioplankton (Selje *et al.*, 2004; Abell and Bowman, 2005; Giebel *et al.*, 2009; Weber and Deutsch, 2010). The effects of climate change on the physical oceanography of the SO may therefore have global ecological significance, particularly as the SO is a major site for sequestration of anthropogenic $CO_2$ (Sabine *et al.*, 2004; Mikaloff Fletcher *et al.*, 2006) through both physicochemical processes and the microbe-driven "biological pump" of $CO_2$ fixation (Thomalla *et al.*, 2011), forming a potential feedback loop (Cox *et al.*, 2000). However, the microbial assemblages that inhabit the SO are generally poorly understood, and their diversity and functional capacity poorly characterised (reviewed in Murray and Grzymski, 2007). A community-level understanding is needed to predict the effects of a shifting ACC on the distribution, abundance and ecosystem functions of plankton.

Large scale metagenome surveys have not previously been performed in the SO. Compared to traditional environmental microbiological methods such as culture-based or DGGE surveys, metagenomic studies are able to offer both deeper (capture of sequence from rare taxa) and broader (larger sample sizes with greater statistical significance) data on the taxonomic makeup and functional capacities of marine microbial communities (e.g. Rusch *et al.*, 2007; Angly *et al.*, 2006; Dinsdale *et al.*, 2008). Metagenome sampling is also better suited to large-scale biogeographical studies, as it allows large numbers of samples to be taken across broad spatial and temporal scales with uniform sampling, storage and processing, compared to e.g. culture-based methods where technical replicability is more difficult.

This project is part of a metagenomic survey of the SO begun in the austral summer of 2006, based on the sampling design of the Global Ocean Sampling (GOS) expedition (Rusch *et al.*, 2007). The GOS sampling strategy involves size fractionation of plankton assemblages by passing seawater through a 20 µm prefilter and capturing planktonic biomass on sequential 3.0 µm, 0.8 µm and 0.1 µm filter membranes. As well as allowing deeper sequencing of metagenomes and thus better representation of low-abundance taxa (Rusch *et al.*, 2007), this approach allows comparison to other samples collected using the GOS methods (e.g. Brown *et al.*, 2012). Because of this size fractionation, this study will focus on the picoplanktonic component of the sampled microbial assemblages.

# Methods

## Sampling and metagenomic sequencing

Sampling[2] was conducted on board the RSV *Aurora Australis* during Australian Antarctic Division (AAD) cruise V3 Collaborative East Antarctic Marine Census (CEAMARC) / Climate of Antarctica and the Southern Ocean (CASO) from December 13th 2007–January 26th 2008. This cruise occupied the SR3 latitudinal transect from Hobart, Australia (44° S) to the Mertz Glacier, Antarctica (67° S) within a longitudinal range of 140–150° E. Nineteen samples (16 surface, 3 deep) were obtained along almost the entire latitudinal range (Figure 3).

A range of data were recorded by integrated instruments on the RSV *Aurora Australis* including location, water column depth, water temperature, salinity, fluorescence and meteorological data (Table 2). These data were used to locate the PFZ based on a surface temperature gradient of $\sim 1.35$ °C across a distance of 45–65 km, placing the PF at approximately $-59.70°$ of latitude, consistent with previous descriptions (Moore *et al.*, 1999; Sokolov and Rintoul, 2002). Samples were accordingly grouped into "North" and "South" zones, while the three deep samples composed a "Deep" zone (Table 2). The North Zone (NZ) represents waters from the Subtropical, Subantarctic and PFZ regions, while the South Zone (SZ) represents the AZ.

At each station, $\sim$ 250–560 L of seawater was pumped from $\sim$ 1.5–2.5 m below the sea surface into drums stored at ambient temperature on deck. In the case of deep samples, $\sim$ 225–230 L of seawater was collected from Niskin bottles attached to a CTD (SeaBird, Bellevue, USA). Seawater samples were prefiltered through a 20 µm plankton net, then filtrate was captured on sequential 3.0 µm 0.8 µm and 0.1 µm 293 mm polyethersulfone membrane filters (Pall, Port Washington, USA), and immediately stored at $-20$ °C (Rusch *et al.*, 2007; Ng *et al.*, 2010).

DNA extraction[3] was performed at the J. Craig Venter Institute (Rockville, USA) as described in Rusch *et al.* (2007). Pyrosequencing was performed on a GS20 FLX Titanium instrument (Roche, Branford, USA) also at the J. Craig Venter Institute as described in Lauro *et al.* (2011). Duplicate reads and reads with many pyrosequencing errors were removed as described in Lauro *et al.* (2011).

## Phylogenetic analysis of metagenomic data

### BLAST comparison to RefSeq database

A subset of the RefSeq microbial (bacterial and archaeal) genome database (release 41, retrieved May 31 2012 from `ftp://ftp.ncbi.nih.gov/refseq/release/`) was prepared by excluding sequences with the words "shotgun", "contig", "partial", "end" or "part" in their headers (Angly *et al.*, 2009). Because

[2]Sampling was performed by Jeffrey M. Hoffman and Jeffrey B. McQuaid.

[3]DNA extraction was performed by Cynthia Andrews-Pfannkoch and others at the J. Craig Venter Institute.

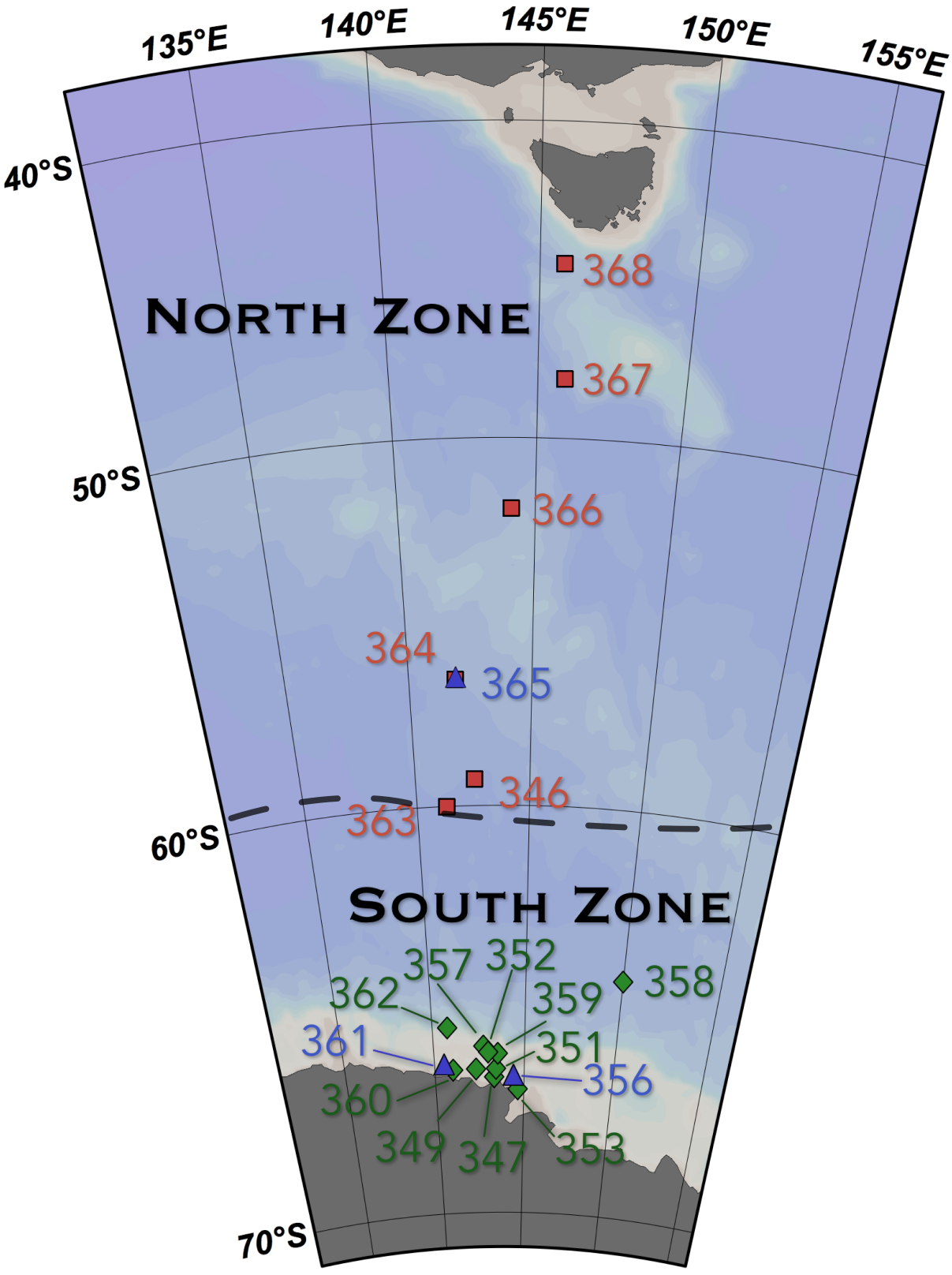



**Figure 3:** Sites of seawater samples used in this study. Red squares indicate surface samples from the North Zone; green diamonds samples from the South Zone; and blue triangles indicate deep samples. The dashed line gives the approximate location of the Polar Front.

**Table 2:** Sampling time, location and physicochemical properties of samples used in this study. All data were retrieved from underway instruments aboard the RSV *Aurora Australis*, with the exception of temperature, salinity and fluorescence data for the three deep samples, which was obtained from the CTD (SeaBird, Bellevue, USA) instrument used to collect the samples.

| Sample | Zone | Date | Latitude | Longitude | Water Column Depth (m) | Sample Depth (m) | Temperature (°C) | Salinity (PSU) | Fluorescence ($\mu gL^{-1}$) | Volume filtered (L) |
|---|---|---|---|---|---|---|---|---|---|---|
| 346 | North | 20/12/07 | −59.3120 | 142.5949 | 4294 | 2 | 2.9 | 33.75 | 0.3 | 500 |
| 349 | South | 27/12/07 | −66.5662 | 142.3169 | 370 | 1.5 | −1.3 | 34.40 | 2.3 | 250 |
| 351 | South | 28/12/07 | −66.5587 | 143.4303 | 823 | 1.5 | −0.6 | 34.30 | 1.3 | 500 |
| 352 | South | 29/12/07 | −66.7650 | 143.3240 | 164 | 2.5 | −0.8 | 34.30 | 3.1 | 500 |
| 353 | South | 30/12/07 | −67.0521 | 144.6786 | 180 | 2 | −1.8 | 34.40 | 0.3 | 500 |
| 356 | Deep | 03/01/08 | −66.7617 | 144.4138 | 920 | 920 | −1.9 | 34.69 | 0.1 | 230 |
| 357 | South | 05/01/08 | −66.1719 | 143.0193 | 580 | 2 | −0.4 | 34.15 | 2.5 | 500 |
| 358 | South | 09/01/08 | −64.3001 | 150.0306 | 3550 | 2 | 0 | 33.55 | 0.5 | 500 |
| 359 | South | 12/01/08 | −66.1903 | 143.5292 | 540 | 2 | −0.2 | 34.21 | 2.5 | 500 |
| 360 | South | 13/01/08 | −66.5817 | 141.0211 | 316 | 2 | −0.7 | 34.04 | 6.2 | 500 |
| 361 | Deep | 14/01/08 | −66.4727 | 140.5572 | 1203 | 1170 | −1.8 | 34.56 | 0.1 | 225 |
| 363 | North | 22/01/08 | −60.0001 | 141.3094 | 4473 | 2 | 3.3 | 33.77 | 0.1 | 500 |
| 364 | North | 23/01/08 | −56.6953 | 141.8780 | 3693 | 2 | 4 | 33.70 | 0.5 | 500 |
| 365 | Deep | 23/01/08 | −56.6967 | 141.9125 | 3693 | 3693 | 0.5 | 34.69 | 0.1 | 230 |
| 366 | North | 24/01/08 | −52.0233 | 144.1362 | 3180 | 2 | 7.6 | 33.84 | 0.3 | 500 |
| 367 | North | 25/01/08 | −48.2487 | 145.9025 | 3490 | 2 | 11 | 34.43 | 0.2 | 500 |
| 368 | North | 26/01/08 | −44.7180 | 145.7775 | 3201 | 2 | 14.8 | 34.96 | 1.3 | 560 |

**Table 3:** Dates and locations of additional samples used to falsify the hypothesis that the Mertz Glacier polynya was responsible for the observed biogeographic partitioning.

| Sample | Date | AAD Voyage Number | Latitude | Longitude |
|---|---|---|---|---|
| 388 | 20/10/08 | V1 2008–09 | −63.8090 | 115.1613 |
| 398 | 22/10/08 | V1 2008–09 | −64.8018 | 112.3730 |
| 390 | 30/10/08 | V1 2008–09 | −64.8152 | 80.7204 |
| 392 | 13/12/08 | V2 2008–09 | −64.1826 | 76.4537 |
| 393 | 15/12/08 | V2 2008–09 | −55.2573 | 74.2531 |

this database was not expected to contain representative genomes for every species present, Operational Taxonomic Units (OTUs) in this study are defined by the best species match to this database, and may for example represent congeners.

The metagenomic reads from each sample were compared against this database using TBLASTX, with default parameters except for: E-value threshold $1.0 \times 10^{-3}$, cost to open gap 11, cost to extend gap 1, masking of query sequence by SEG masking with lookup table only. The outputs of all TBLASTX searches against RefSeq were processed by MINSPEC, and hits not belonging to the minimal sets were removed.

### OTU abundances and variance between zones

The relative OTU abundances for each sample were determined using the PERL script Genome relative Abundance and Average Size (GAAS) (Angly *et al.*, 2009). Briefly, GAAS estimates the relative abundance of OTUs from the number and quality of BLAST matches to each species, taking into account differences in genome size. GAAS was run with the default settings. To normalise for reads which did not yield acceptable matches, the relative abundances for each sample were scaled by that sample's effective BLAST hit rate. An OTU profile was generated for each sample by encoding the scaled relative abundance of each OTU from each size fraction as a separate variable.

To test the hypothesis that the oceanic zones harbour significantly different communities, Analysis of Similarities (ANOSIM) with 999 permutations was performed on a standardised, log-transformed Bray-Curtis resemblance matrix of OTU profiles. Similarity Percentages (SIMPER) analysis was performed to identify the contribution of individual OTUs to differences between the zones. All statistical procedures were performed in PRIMER 6 as described by Clarke and Warwick (2001).

### Additional samples to test alternative "polynya hypothesis"

Because many samples from the SZ were taken in the region of the Mertz Glacier polynya, an alternative hypothesis for any difference observed between the SZ and NZ would be that the difference represented the effect of the polynya, rather than the PF. A set of samples collected[4] from waters south of the PF on different voyages and in a different longitudinal range were used for an additional test to falsify this hypothesis. Sampling, DNA extraction and sequencing were performed as described above, although only the 0.1–0.8 µm size fraction was processed. The resulting metagenomic reads were compared to RefSeq and weighted relative abundances calculated with GAAS as described above.

Two ANOSIM tests were performed on the standardised and log-transformed Bray-Curtis resemblance matrices of the samples' OTU abundance profiles and similarly prepared matrices representing only the 0.1 µm fractions of the SZ and NZ samples. In the first ANOSIM test ("polynya hypothesis"), all samples from the Mertz Glacier polynya region (i.e. all SZ samples apart from 358) were grouped together, while the samples from the open ocean were placed in a separate group. In the second test ("PF hypothesis"), the additional samples were grouped with all the SZ samples.

[4]Collection was performed by Ricardo Cavicchioli, Federico M. Lauro and Mark V. Brown. TODO who else was on these voyages?

## Functional analysis of metagenomic data

### BLAST comparison to Kyoto Encyclopedia of Genes and Genomes (KEGG) database

In order to identify functional differences between the zones, the set of metagenomic reads from each sample was compared against the KEGG GENES database (retrieved July 2 2010 from `ftp://ftp.genome.jp/pub/kegg/genes/fasta/genes.pep`) with BLASTX, with default parameters except for: maximum number of database sequence alignments 10; E-value threshold $1.0 \times 10^{-3}$; gap opening penalty 11; gap extension penalty 1; masking of query sequence by SEG masking for lookup table only.

#### Analysis of functional potential

Genes identified by BLASTX were aggregated to KEGG ortholog groups according to the KEGG Orthology schema (`ftp://ftp.genome.jp/pub/kegg/genes/ko`, retrieved Mar 29 2011), and ortholog group abundances calculated for each sample. Following Coleman and Chisholm (2010), a read was considered a hit to a given ortholog group if the top three hits for that read (or all hits if fewer than three total hits) were to genes from the same ortholog group, and had bit scores > 40. If the bit score difference between any two top hits was greater than 30, only the hits above this difference were considered.

Ortholog group counts were then used to calculate the abundance of KEGG modules. Because many ortholog groups are members of more than one module, the abundance $a_m$ of each module $m$ was calculated as

$$a_m = \sum_{K=1}^{n} \frac{C_K}{M_K}$$

where $n$ is the number of ortholog groups $K$ belonging to module $m$, $C_K$ is the number of hits to ortholog group $K$, and $M_K$ is the total number of modules to which $K$ belongs. To account for differences in sequencing depth between samples, module abundances were scaled to 500,000 reads per sample. To test the hypothesis that the NZ and SZ harbour significantly different functional potential, one-way ANOSIM with 999 permutations was performed as above on a standardised, log-transformed Bray-Curtis distance resemblance matrix of the module and ortholog group profiles. A functional profile was generated for each sample by summing the scaled abundances of each module from all size fractions, and SIMPER performed as above to identify modules which contributed highly to the variation in functional potential between the two zones.

#### Taxonomic decomposition

To link modules with a high contribution to variance or otherwise of interest to specific taxa ("taxonomic decomposition"), a script was written which decomposed each module to its constituent ortholog groups, then each ortholog group to each representative gene sequence in the KEGG database. Because each KEGG sequence is associated with a particular organism, this allowed the abundance of each module to be decomposed into the relative contribution from each of these organisms. This allowed functional contributions to be putatively assigned to genera which were not identified in our taxonomic analysis, as the database included gene sequences for organisms for which a full genome was not available. To aid interpretation of these results, the taxonomic decompositions for some modules were aggregated to higher taxonomic ranks, such as class or family.

# Results

## Metagenomic sequencing

6.6 Gbp of 454 sequence data representing picoplankton in the size range 0.1 – 3.0 µm was obtained from 16 samples. After removal of low-quality reads, 454 sequencing yielded 157,507 – 597,689 reads per sample (mean 354,399) of lengths ranging from 100 to 606 bp (mean 378).

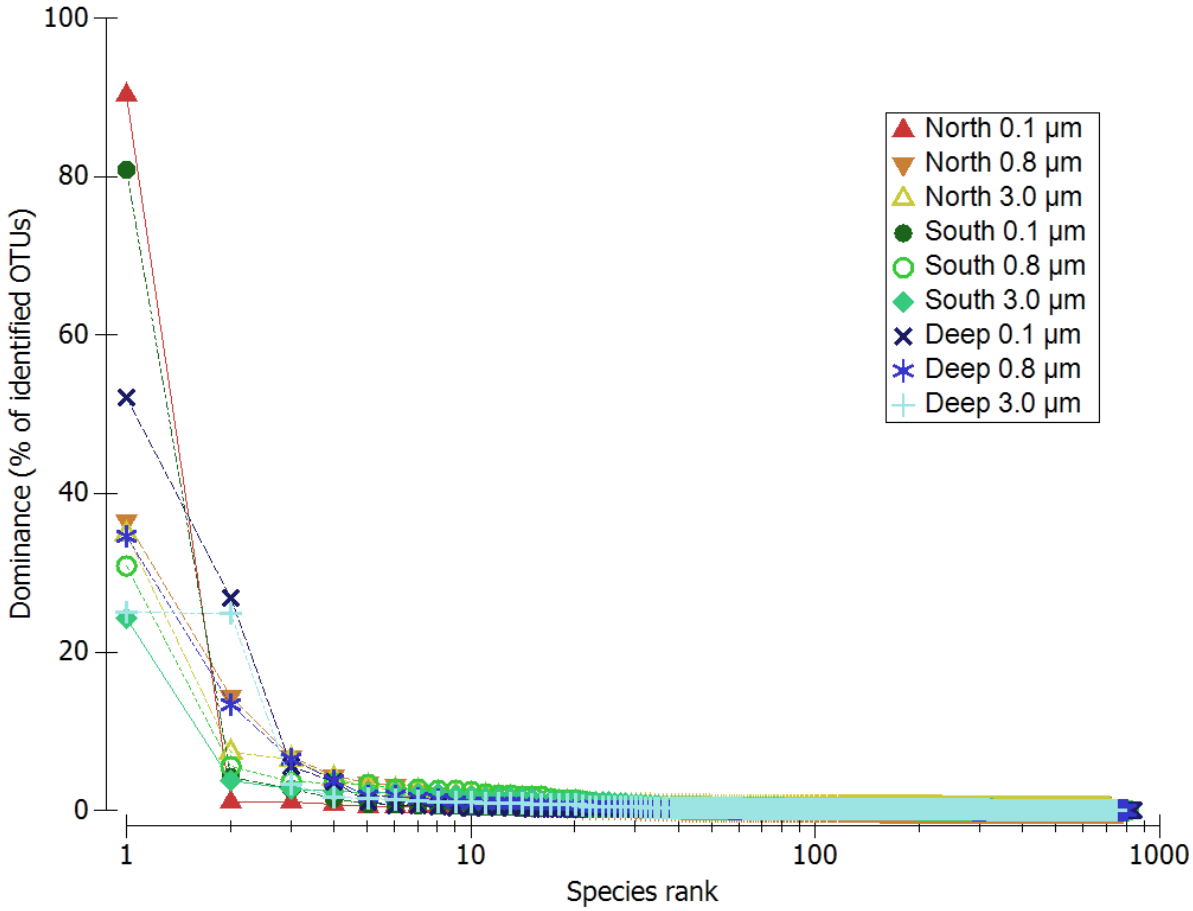



**Figure 4:** Rank-abundance curves for OTUs identified in each zone and size fraction. The dominance of a given OTU is its relative abundance as a percentage of all identified OTUs. The x-axis is scaled logarithmically. Generated using PRIMER 6.

## Phylogenetic analysis of metagenomic data

The proportion of reads in each sample which yielded matches to RefSeq ranged from 25% to 85% (mean 62%). The most abundant OTUs in each sample are given in Table 4 and a full list of OTU abundances in the supplementary material (`PF-all-OTUs.csv`). All samples and size fractions exhibited very low OTU evenness (Figure 4).

ANOSIM analysis showed that the zones harbor significantly different microbial communities ($R = 0.451$, $p < 0.004$). SIMPER was performed in order to identify the contribution of individual OTUs to the difference between the NZ and SZ. The results for the highest contributors are provided in Table 5, and are graphically summarised for all OTUs in Figure 5.

The SIMPER analysis found that no single OTU contributed more than 2.9% of variance and 74% of variance was contributed by OTUs with a contribution less than 1%. There was also a large difference in the contribution to variance of the three size fractions, with approximately 52% of all variance contributed by OTUs from the 3.0 µm fraction, 37% by the 0.8 µm fraction, and 9% by the 0.1 µm fraction. Notably, OTUs within several taxonomic groups that had high contribution to variance covaried in their relative representation in the NZ and SZ. For example, Bacteroidetes and GSO-EOSA-1 representatives were on average more abundant in the SZ; while *Prochlorococcus* and *Synechococcus* spp., SAR11 and SAR116 were on average more abundant in the NZ (Figure 5). Some groups, such as the Alteromonadales, had variable relative representation depending on size fraction.

**Table 4:** Relative abundances (as percentages) of the twenty most abundant OTUs identified in this study, in each zone and size fraction.

| OTU | North | | | South | | | Deep | | |
|---|---|---|---|---|---|---|---|---|---|
| | 0.1 µm | 0.8 µm | 3.0 µm | 0.1 µm | 0.8 µm | 3.0 µm | 0.1 µm | 0.8 µm | 3.0 µm |
| "*Candidatus* Pelagibacter ubique" HTCC1062 | 61.76 | 25.00 | 23.87 | 58.85 | 22.40 | 17.61 | 37.05 | 24.56 | 17.66 |
| *Nitrosopumilus maritimus* SCM1 | 0.01996 | 0.01438 | 0.009508 | 1.076 | 1.309 | 1.210 | 19.09 | 9.463 | 17.77 |
| "*Candidatus* Ruthia magnifica" str. Cm (*Calyptogena magnifica*) | 0.6699 | 0.6458 | 0.5484 | 2.987 | 2.616 | 1.025 | 3.945 | 4.601 | 2.264 |
| *Roseobacter* sp. OCh114 | 0.3125 | 2.932 | 1.588 | 0.4477 | 3.994 | 2.657 | 0.1259 | 1.228 | 0.6792 |
| *Synechococcus* sp. CC9902 | 0.1081 | 9.837 | 4.973 | 0.0007484 | 0.004156 | 0.09733 | 0.002846 | 0.01502 | 0.01058 |
| *Silicibacter pomeroyi* DSS-3 | 0.2578 | 2.286 | 1.154 | 0.3070 | 2.505 | 1.576 | 0.1224 | 0.9417 | 0.4988 |
| *Gramella forsetii* strain KT0803 | 0.2412 | 1.210 | 1.755 | 0.4993 | 2.347 | 1.890 | 0.2078 | 0.6179 | 0.5173 |
| "*Candidatus* Vesicomyosocius okutanii" strain HA | 0.4634 | 0.4642 | 0.2078 | 1.970 | 1.807 | 0.2174 | 2.480 | 2.662 | 1.167 |
| *Robiginitalea biformata* strain HTCC2501 | 0.2751 | 1.099 | 1.297 | 0.4722 | 1.878 | 1.405 | 0.2265 | 0.6188 | 0.6946 |
| *Flavobacterium psychrophilum* strain JIP02/86 | 0.1718 | 0.8409 | 1.224 | 0.4316 | 1.960 | 1.598 | 0.1599 | 0.4744 | 0.6001 |
| *Synechococcus* sp. CC9311 | 0.03014 | 4.624 | 4.409 | 0.0007221 | 0.002778 | 0.02764 | 0.001580 | 0.002863 | 0.009241 |
| "*Candidatus* Puniceispirillum marinum" IMCC1322 | 0.6444 | 2.077 | 1.267 | 0.3586 | 1.377 | 0.7109 | 0.3425 | 1.062 | 0.5345 |
| *Silicibacter* sp. TM1040 | 0.2274 | 1.652 | 0.8738 | 0.2709 | 1.803 | 1.233 | 0.07665 | 0.5890 | 0.2957 |
| *Jannaschia* sp. DFL-12 | 0.1776 | 1.378 | 0.7350 | 0.2443 | 1.692 | 0.8009 | 0.07338 | 0.6515 | 0.3078 |
| *Zunongwangia profunda* strain SM-A87 | 0.1522 | 0.7487 | 1.059 | 0.2968 | 1.410 | 1.204 | 0.1353 | 0.3478 | 0.4971 |
| *Colwellia* sp. 34H | 0.02345 | 0.3636 | 2.736 | 0.05207 | 0.5140 | 1.041 | 0.05137 | 0.4687 | 0.8013 |
| *Coraliomargarita akajimensis* strain DSM 45221 | 0.03698 | 0.07573 | 0.1197 | 0.1154 | 1.543 | 1.680 | 0.02614 | 0.3040 | 0.2740 |
| *Jannaschia* sp. CCS1 | 0.1173 | 0.9344 | 0.4784 | 0.1711 | 1.230 | 0.8239 | 0.05865 | 0.4462 | 0.2118 |
| *Pseudoalteromonas atlantica* strain T6c | 0.01251 | 0.4772 | 1.993 | 0.02270 | 0.4089 | 1.132 | 0.02634 | 0.2143 | 0.7459 |
| *Saccharophagus degradans* strain 2-40 | 0.06532 | 0.4325 | 0.5429 | 0.1289 | 1.072 | 0.8663 | 0.07798 | 0.2844 | 0.3165 |
| *Flavobacterium johnsoniae* strain UW101 | 0.08822 | 0.4220 | 0.6141 | 0.2034 | 0.9389 | 0.8578 | 0.07545 | 0.2255 | 0.3300 |
| *Capnocytophaga ochracea* strain DSM 7271 | 0.1143 | 0.4830 | 0.5399 | 0.2314 | 0.8815 | 0.6814 | 0.08964 | 0.2840 | 0.5043 |
| *Marinomonas* sp. MWYL1 | 0.03777 | 0.2529 | 0.3026 | 0.1514 | 1.300 | 0.7006 | 0.07393 | 0.2439 | 0.2155 |
| *Cellvibrio japonicus* strain Ueda107 | 0.05884 | 0.3080 | 0.3231 | 0.1155 | 0.9917 | 0.4713 | 0.06774 | 0.2981 | 0.2549 |
| *Marinobacter hydrocarbonoclasticus* VT8 | 0.04093 | 0.2889 | 0.3883 | 0.08418 | 0.7195 | 0.3848 | 0.1250 | 0.6667 | 1.066 |
| *Pseudoalteromonas haloplanktis* strain TAC125 | 0.01389 | 0.2505 | 0.8896 | 0.03427 | 0.3561 | 0.6530 | 0.1092 | 1.203 | 0.1503 |
| *Teredinibacter turnerae* strain T7901 | 0.05665 | 0.3051 | 0.3081 | 0.1138 | 0.9174 | 0.5127 | 0.06558 | 0.2649 | 0.1885 |
| *Acinetobacter baumannii* strain SDF | 0.004886 | 0.007187 | 0.4073 | 0.006260 | 0.04218 | 1.459 | 0.004285 | 0.01229 | 0.3155 |

**Table 5:** The thirty OTUs with the highest contributions to the difference between the NZ and SZ. Abundances are zonal averages and have been standardised and log-transformed. As each OTU on each size fraction was encoded as a separate variable in the SIMPER analysis, the size fraction is given after each OTU name.

| OTU | Abundance South | Abundance North | Contribution to variance (%) |
|---|---|---|---|
| *Synechococcus* sp. CC9311 0.8 µm | 0.00 | 1.08 | 2.88 |
| *Synechococcus* sp. CC9902 0.8 µm | 0.00 | 1.04 | 2.81 |
| *Synechococcus* sp. CC9311 3.0 µm | 0.01 | 0.98 | 2.59 |
| *Synechococcus* sp. CC9902 3.0 µm | 0.04 | 0.76 | 2.03 |
| "*Candidatus* Pelagibacter ubique" HTCC1062 3.0 µm | 1.97 | 2.40 | 1.97 |
| "*Candidatus* Ruthia magnifica" str. Cm (*Calyptogena magnifica*) 0.1 µm | 0.82 | 0.25 | 1.57 |
| *Colwellia* sp. 34H 3.0 µm | 0.34 | 0.66 | 1.32 |
| "*Candidatus* Ruthia magnifica" str. Cm (*Calyptogena magnifica*) 0.8 µm | 0.74 | 0.25 | 1.32 |
| "*Candidatus* Pelagibacter ubique" HTCC1062 0.8 µm | 2.32 | 2.48 | 1.32 |
| "*Candidatus* Vesicomyosocius okutanii" strain HA 0.1 µm | 0.62 | 0.18 | 1.20 |
| *Coraliomargarita akajimensis* strain DSM 45221 0.8 µm | 0.48 | 0.04 | 1.13 |
| *Coraliomargarita akajimensis* strain DSM 45221 3.0 µm | 0.49 | 0.06 | 1.10 |
| *Roseobacter* sp. OCh114 0.8 µm | 1.01 | 0.81 | 1.08 |
| *Pseudoalteromonas atlantica* strain T6c 3.0 µm | 0.38 | 0.54 | 1.08 |
| "*Candidatus* Vesicomyosocius okutanii" strain HA 0.8 µm | 0.57 | 0.19 | 1.04 |
| *Acinetobacter baumannii* strain SDF 3.0 µm | 0.45 | 0.18 | 0.95 |
| *Gramella forsetii* strain KT0803 0.8 µm | 0.72 | 0.43 | 0.94 |
| *Marinomonas* sp. MWYL1 0.8 µm | 0.46 | 0.11 | 0.92 |
| *Roseobacter* sp. OCh114 3.0 µm | 0.76 | 0.54 | 0.91 |
| *Flavobacterium psychrophilum* strain JIP02/86 0.8 µm | 0.63 | 0.32 | 0.89 |
| *Silicibacter pomeroyi* DSS-3 0.8 µm | 0.75 | 0.69 | 0.86 |
| *Brachyspira hyodysenteriae* strain WA1 3.0 µm | 0.47 | 0.19 | 0.84 |
| "*Candidatus* Ruthia magnifica" str. Cm (*Calyptogena magnifica*) 3.0 µm | 0.34 | 0.21 | 0.82 |
| *Pseudoalteromonas haloplanktis* strain TAC125 3.0 µm | 0.22 | 0.33 | 0.77 |
| *Robiginitalea biformata* strain HTCC2501 0.8 µm | 0.61 | 0.40 | 0.74 |
| *Nitrosopumilus maritimus* SCM1 0.1 µm | 0.27 | 0.01 | 0.72 |
| *Gramella forsetii* strain KT0803 3.0 µm | 0.59 | 0.59 | 0.71 |
| *Lysinibacillus sphaericus* strain C3-41 3.0 µm | 0.29 | 0.02 | 0.71 |
| *Nitrosopumilus maritimus* SCM1 0.8 µm | 0.25 | 0.01 | 0.70 |
| *Silicibacter* sp. TM1040 0.8 µm | 0.59 | 0.55 | 0.69 |

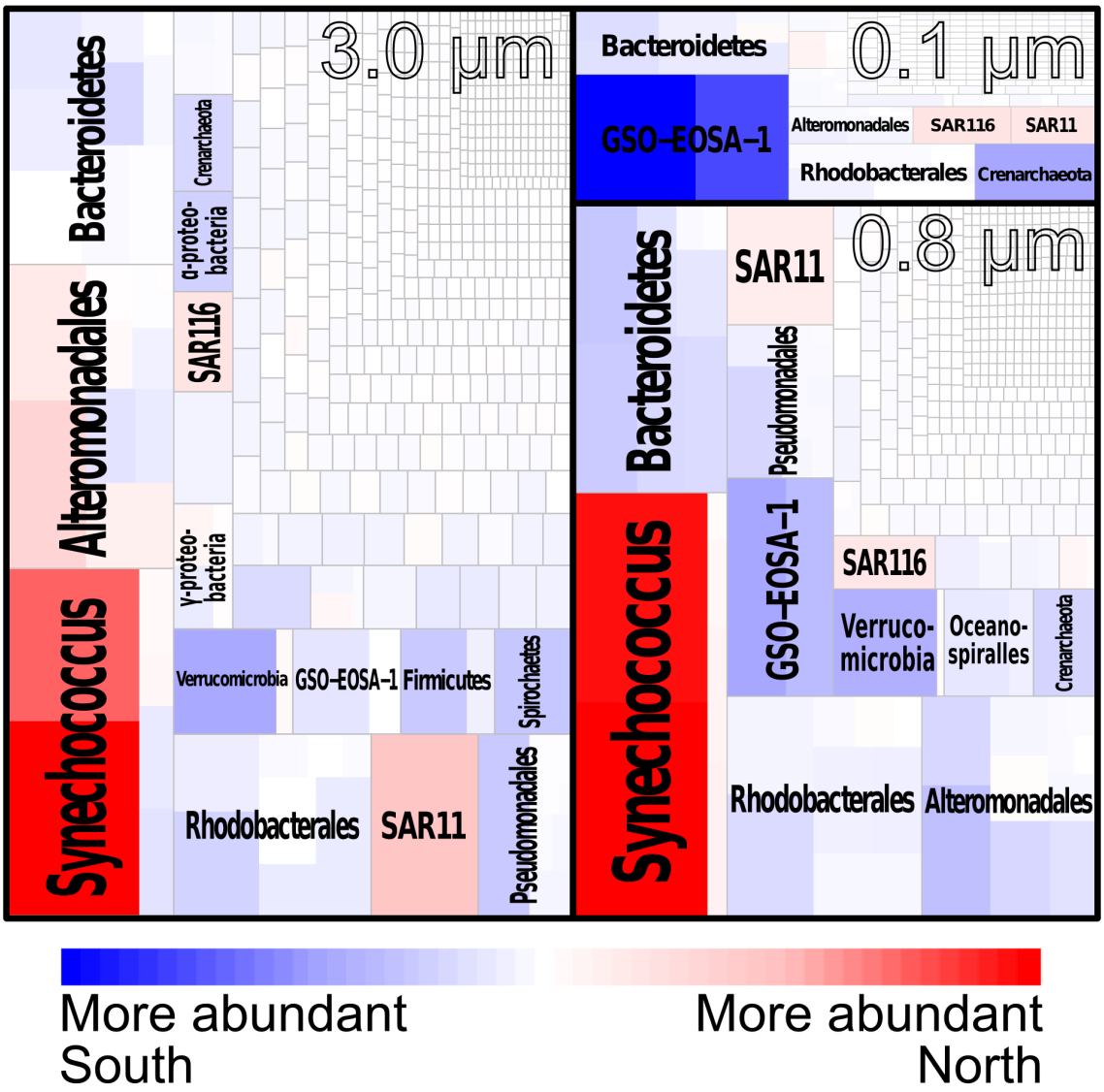



**Figure 5:** Contribution of OTUs to variance between the North and South zones, and differential abundance of OTUs from each size fraction between the two zones. Each coloured (red or blue) rectangle represents an OTU identified through analysis of BLAST matches between SO metagenome data and the RefSeq database. The area of each rectangle as a proportion of the total plot area corresponds to that OTU's contribution to the total variance between the two zones. The colour of each rectangle corresponds to difference in relative abundance of that OTU between the zones, with blue indicating a higher relative abundance south of the PF, and red a higher abundance north of the PF. OTUs from clades or taxonomic ranks of interest have been grouped, with labels in bold and groups separated by gray lines. Groups and OTUs with a low contribution to variance which were not grouped are unlabeled. OTUs from each size fraction have also been grouped, with labels in black outline and size fractions separated by thick black lines. The total contribution to variance of each size fraction is given as a percentage. The data used to generate this figure are given in the supplementary material (`PF-OTUs-SIMPER.csv`).

### Additional samples to test alternative "polynya hypothesis"

The ANOSIM analysis strongly supported the "PF hypothesis" (R = 0.435, p = 0.002), i.e. that the PF was primarily responsible for the difference observed between the NZ and SZ, over the "polynya hypothesis" (R = 0.29, p = 0.005) that the influence of the Mertz Glacier polynya was responsible. This also provides evidence that the PF effect is robust over a longitudinal range and over time.

### Functional analysis of metagenomic data

ANOSIM analysis of the samples' KEGG ortholog group and module profiles revealed that the zones had significantly different functional potential (ortholog group: R = 0.642, p < 0.001; module: R = 0.871, p < 0.001). SIMPER was performed on the profiles in order to identify the specific functional differences between the zones. The highest-contributing modules are given in Table 6, and a complete list in the supplementary material (`PF-modules-SIMPER.csv`). The highest-contributing ortholog groups are given in Table 7, and a complete list in the supplementary material (`PF-ortholog-groups-SIMPER.csv`). No single ortholog group or module contributed more than 2.2% of the variance, indicating a complex and diverse pattern of functional differences. There was a strong trend for ortholog groups and modules with higher contributions to variance to be overrepresented in the NZ in the 3.0 µm fraction but the SZ in the smaller fractions, indicating that the functional diversity of each zone was strongly segregated by size fraction.

## Discussion

### Taxonomic groups differentiating the zones

#### GSO-EOSA-1

The Gammaproteobacterial Sulfur Oxidizer-EOSA-1 (GSO-EOSA-1) cluster, represented in RefSeq by the OTUs "*Candidatus* Vesicomyosocius okutanii" strain HA and "*Candidatus* Ruthia magnifica" strain Cm. (*Calyptogena magnifica*) (Walsh *et al.*, 2009), made a large contribution to variance between the NZ and SZ, with higher abundance in the SZ: relative abundances of GSO-EOSA-1 in the SZ were 5.2%, 3.4% and 0.25% in the 0.1, 0.8 and 3.0 µm size fractions respectively, compared to 1.1%, 0.84% and 0.30% in the NZ (Table 4). The contribution to variance of this group was highest in the 0.1 µm size fraction, followed by 0.8 µm and 3.0 µm (Table 5). This pattern most likely represents a small cell size and lack of association with particulate matter.

"*Ca.* R. magnifica" and "*Ca.* V. okutanii" are chemoautotrophic endosymbionts of deep-sea bivalves (Kuwahara *et al.*, 2007; Newton *et al.*, 2007) and are thus unlikely to be present in open ocean surface waters. However, GSO-EOSA-1 representative ARCTIC96BD-19 has recently been reported at high abundance in Antarctic coastal waters (Ghiglione and Murray, 2011; Grzymski *et al.*, 2012). The majority of 16S rRNA genes from this metagenome with best BLASTN matches to "*Ca.* R. magnifica" and "*Ca.* V. okutanii" clustered with ARTIC96BD-19 in a neighbour-joining phylogenetic tree (Figure 6), indicating this is the dominant GSO-EOSA-1 representative. Single-cell genomic analysis of ARCTIC96BD-19 from global mesopelagic waters indicates the lineage is probably mixotrophic, able to couple carbon fixation to oxidation of reduced sulphur compounds as well as assimilate organic carbon (Swan *et al.*, 2011). GSO-EOSA-1 cytochrome C oxidase (CoxII) has been identified in a winter metaproteome of Antarctic Peninsula coastal waters, suggesting the capacity for aerobic respiration (Williams *et al.*, 2012). Taken together, this evidence suggests the GSO-EOSA-1 representative in Antarctic coastal waters is a versatile chemolithoautotroph capable of aerobic respiration.

It has been proposed that during the winter months, chemolithoautotrophy is dominant over photoautotrophy as the major carbon fixation input in AZ waters due to the lack of available light, both from seasonal darkness and ice cover (Grzymski *et al.*, 2012). The high relative abundance of GSO-EOSA-1 we detected in SZ compared to NZ waters may therefore represent the remnants of an annual winter increase in population in the marginal ice zone which does not occur in the open ocean.

**Table 6:** The thirty KEGG modules with the highest contributions to the difference between the NZ and SZ. Abundances are zonal averages and have been standardised and log-transformed.

| KEGG module | Abundance South | Abundance North | Contribution to variance (%) |
|---|---|---|---|
| Photosystem II | 0.42 | 0.57 | 2.21 |
| Complex I (NADH dehydrogenase), NADH dehydrogenase I/diaphorase subunit of the bidirectional hydrogenase | 0.01 | 0.24 | 1.80 |
| Photosystem I | 0.43 | 0.34 | 1.70 |
| Pyrimidine deoxyribonucleotide biosynthesis, CDP/CTP → dCDP/dCTP,dTDP/dTTP | 0.51 | 0.66 | 1.16 |
| Histidine degradation, histidine → N-formiminoglutamate → glutamate | 0.42 | 0.31 | 1.14 |
| Methionine salvage pathway | 0.29 | 0.43 | 1.14 |
| sn-Glycerol 3-phosphate transport system | 0.29 | 0.16 | 1.11 |
| Complex I (NADH dehydrogenase), NADH dehydrogenase I | 1.08 | 1.05 | 1.06 |
| Branched-chain amino acid transport system | 0.79 | 0.83 | 0.96 |
| Dipeptide transport system | 0.14 | 0.02 | 0.95 |
| Adenine nucleotide biosynthesis, IMP → ADP/dADP,ATP/dATP | 0.62 | 0.74 | 0.95 |
| Glycine betaine/proline transport system | 0.66 | 0.56 | 0.94 |
| Sulfur reduction, sulfate → H2S | 0.54 | 0.44 | 0.91 |
| Simple sugar transport system | 0.46 | 0.39 | 0.90 |
| Peptides/nickel transport system | 0.99 | 0.98 | 0.89 |
| Ribosome, eukaryotes | 0.26 | 0.27 | 0.89 |
| Multiple sugar transport system | 0.55 | 0.55 | 0.86 |
| Type II general secretion system | 0.21 | 0.21 | 0.82 |
| Sulfonate/nitrate/taurine transport system | 0.45 | 0.37 | 0.82 |
| Guanine nucleotide biosynthesis, IMP → GDP/dGDP,GTP/dGTP | 0.72 | 0.82 | 0.81 |
| RNA polymerase II, eukaryotes | 0.11 | 0.20 | 0.76 |
| Histidine biosynthesis, PRPP → histidine | 0.94 | 0.86 | 0.76 |
| Putrescine transport system | 0.18 | 0.09 | 0.72 |
| Leucine biosynthesis, pyruvate → 2-oxoisovalerate → leucine | 1.29 | 1.37 | 0.71 |
| C5 isoprenoid biosynthesis, non-mevalonate pathway | 0.70 | 0.77 | 0.71 |
| Leucine degradation, leucine → acetoacetate + acetyl-CoA | 0.64 | 0.59 | 0.71 |
| Thiamine transport system | 0.13 | 0.05 | 0.69 |
| Spliceosome, 35S U5-snRNP | 0.18 | 0.20 | 0.68 |
| Cytochrome b6f complex | 0.14 | 0.12 | 0.67 |
| Menaquinone biosynthesis, chorismate → menaquinone | 0.25 | 0.27 | 0.66 |

**Table 7:** The thirty KEGG ortholog groups with the highest contribution to the difference between the NZ and SZ. Abundances are zonal averages and have been standardised and log-transformed. As each ortholog group on each size fraction was encoded as a separate variable in the SIMPER analysis, the size fraction is given after each ortholog group name.

| KEGG ortholog group | Abundance South | Abundance North | Contribution to variance (%) |
|---|---|---|---|
| Hypothetical protein 3.0 µm | 0.11 | 0.24 | 0.26 |
| Hypothetical protein 0.8 µm | 0.68 | 0.57 | 0.24 |
| Ribonucleoside-diphosphate reductase alpha chain [EC:1.17.4.1] 0.8 µm | 0.17 | 0.24 | 0.15 |
| DNA polymerase III subunit alpha [EC:2.7.7.7] 0.8 µm | 0.25 | 0.19 | 0.14 |
| Hypothetical protein 0.1 µm | 0.26 | 0.24 | 0.12 |
| Proline dehydrogenase / delta 1-pyrroline-5-carboxylate 0.8 µm | 0.10 | 0.04 | 0.12 |
| Aminomethyltransferase [EC:2.1.2.10] 0.8 µm | 0.25 | 0.19 | 0.12 |
| Ribonucleoside-diphosphate reductase alpha chain [EC:1.17.4.1] 3.0 µm | 0.02 | 0.08 | 0.12 |
| Sarcosine oxidase, subunit alpha [EC:1.5.3.1] 0.8 µm | 0.22 | 0.17 | 0.12 |
| Integrator complex subunit 6 3.0 µm | 0.07 | 0.05 | 0.11 |
| Multicomponent $Na^+$:$H^+$ antiporter subunit D 0.8 µm | 0.11 | 0.05 | 0.11 |
| Glutamine synthetase [EC:6.3.1.2] 0.8 µm | 0.24 | 0.19 | 0.11 |
| Pyruvate dehydrogenase E1 component [EC:1.2.4.1] 0.8 µm | 0.15 | 0.10 | 0.11 |
| Cobaltochelatase CobN [EC:6.6.1.2] 0.8 µm | 0.11 | 0.06 | 0.11 |
| Formate dehydrogenase, alpha subunit [EC:1.2.1.2] 0.8 µm | 0.15 | 0.10 | 0.11 |
| DNA-directed RNA polymerase subunit beta [EC:2.7.7.6] 3.0 µm | 0.03 | 0.08 | 0.11 |
| Glutamate synthase (NADPH/NADH) large chain [EC:1.4.1.13 1.4.1.14] 0.8 µm | 0.25 | 0.22 | 0.11 |
| Dimethylglycine dehydrogenase [EC:1.5.99.2] 0.8 µm | 0.17 | 0.14 | 0.11 |
| Flagellin 0.8 µm | 0.06 | 0.10 | 0.10 |
| DNA-directed RNA polymerase subunit beta [EC:2.7.7.6] 3.0 µm[a] | 0.03 | 0.08 | 0.10 |
| Photosystem II PsbA protein 0.8 µm | 0.01 | 0.06 | 0.09 |
| Aldehyde dehydrogenase (NAD+) [EC:1.2.1.3] 0.8 µm | 0.17 | 0.13 | 0.09 |
| Glutamate synthase (NADPH/NADH) large chain [EC:1.4.1.13 1.4.1.14] 3.0 µm | 0.02 | 0.07 | 0.09 |
| Thymidylate synthase (FAD) [EC:2.1.1.148] 0.8 µm | 0.02 | 0.06 | 0.09 |
| Topoisomerase IV subunit A [EC:5.99.1.-] 0.8 µm | 0.11 | 0.07 | 0.09 |
| DNA mismatch repair protein MutS 0.8 µm | 0.13 | 0.08 | 0.09 |
| Glutamate dehydrogenase [EC:1.4.1.2] 0.8 µm | 0.07 | 0.03 | 0.09 |
| DNA polymerase I [EC:2.7.7.7] 0.1 µm | 0.12 | 0.11 | 0.09 |
| GTP-binding protein 0.8 µm | 0.26 | 0.21 | 0.09 |
| GTP-binding protein 3.0 µm | 0.03 | 0.07 | 0.09 |

[a] Due to an error in the KEGG database, this module is encoded twice.

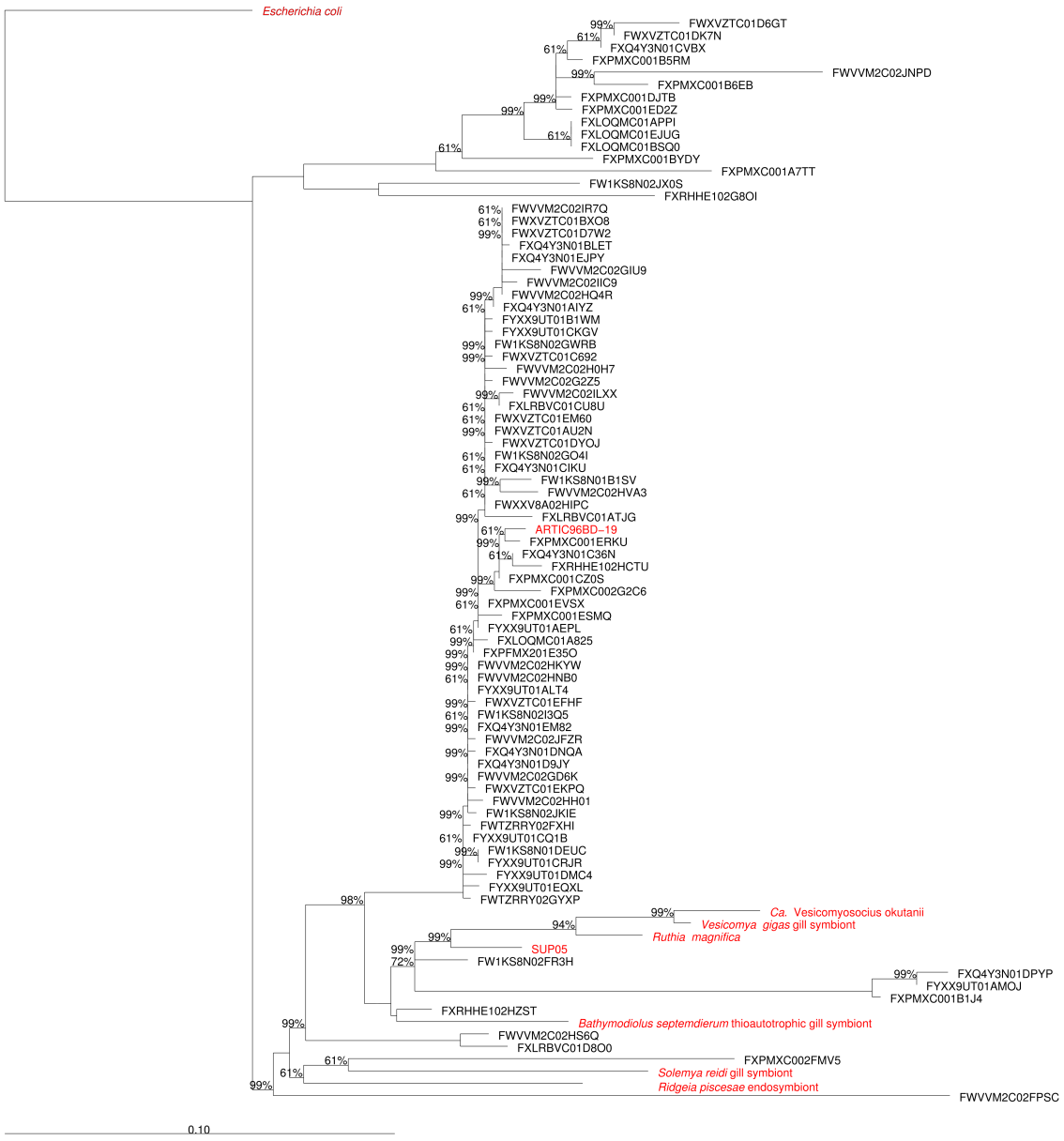



**Figure 6:** Neighbour-joining tree of GSO-EOSA-1-like 16S rRNA gene sequences from the metagenomes in this study. Sequences labeled in black text are reads from the metagenomes. Red labels are 16S rRNA gene sequences from Gammaproteobacterial Sulfur Oxidizers (GSO) and other Gammaproteobacteria. The tree was constructed using ARB (Ludwig *et al.*, 2004).

**Ammonia-oxidizing Crenarchaeota**

*Nitrosopumilus maritimus* SCM1 and *Cenarchaeum symbiosum* are chemolithoautotrophic, nitrifying members of the Marine Group I Crenarchaeota (MGI) (Preston *et al.*, 1996; Walker *et al.*, 2010) and are the only representatives in the reference database of the Ammonia-Oxidizing Archaea (AOA). The contribution of OTUs of *C. symbiosum* to the AOA signature was low. As *C. symbiosum* is a sponge symbiont (Preston *et al.*, 1996) and given the poor representation of AOA in RefSeq, it is likely this OTU has attracted sequences originating from planktonic AOA and *C. symbiosum* itself is not present. AOA were moderate contributors to variance between the NZ and SZ, and were overrepresented in the SZ in all size fractions (Figure 5). As with the GSO-EOSA-1 cluster, MGI have been proposed to be abundant chemolithoautotrophs and therefore major drivers of winter carbon fixation in Antarctic coastal waters (Grzymski *et al.*, 2012; Williams *et al.*, 2012).

Sample 353 had a particularly high relative abundance of *N. maritimus* OTUs (7.5% of the 0.1 µm fraction; 0.8 µm: 11%; 3.0 µm: 12%). This sample was taken closer to the Antarctic continent (3.7 km) than any other, in relatively shallow (180 m) waters 17.6 km from the Mertz Glacier. The high abundance of ammonia oxidizers may reflect an input of ammonia from terrestrial sources (e.g. penguin guano), or resuspension of benthic sediments in which MGI are abundant (Bowman and McCuaig, 2003) by near-shore turbulence and iceberg scouring. Breakdown of water column stratification has been previously suggested as a cause of increased AOA abundance in Antarctic coastal surface waters (Kalanetra *et al.*, 2009).

**Cyanobacteria**

OTUs of the cyanobacterial genera *Prochlorococcus* and *Synechococcus* were overrepresented in the NZ in all size fractions (Figure 5). The mean relative abundance of cyanobacteria in samples 367 and 368, the two northernmost samples, was strikingly higher than the mean abundance across all other samples in the NZ. *Synechococcus* sp. CC9902 alone composed greater than 22% of the 0.8 µm fraction in these samples, consistent with *Synechococcus* species' average cell diameter of approximately 0.9 µm. The high abundance of both cyanobacterial genera on the 3.0 µm fraction has previously been reported (Lauro *et al.*, 2011) and may be attributable to aggregation (Lomas and Moran, 2011).

Samples 367 and 368 were separated from the other samples north of the PF by the STF. While the STF was not a significant boundary on the assemblage level, it may mark a significant biogeographical boundary for these cyanobacteria. *Synechococcus* and *Prochlorococcus* together represent a large proportion of both phytoplankton abundance and carbon fixation in temperate and tropical waters, in many regions contributing more than half of total primary production (Liu *et al.*, 1997, 1998; André *et al.*, 1999). The role of the STF in determining the latitudinal range of *Synechococcus* and *Prochlorococcus* is therefore important, as it will affect models of ocean productivity under changing climactic conditions, and warrants further investigation. Despite the high abundance of cyanobacteria north of the STF, they were also a significant feature of the SAZ; for example, *Synechococcus* sp. CC9902 composed 3–5% of the 0.8 µm fraction in SAZ samples.

These results extend the latitudinal distribution of both *Prochlorococcus* and *Synechococcus* to include presence at very low abundance as far south as the Antarctic coast. *Prochlorococcus* have been reported to be restricted to tropical and subtropical waters within 40° of latitude (Partensky *et al.*, 1999), and to be a negligible (Ghiglione and Murray, 2011) or undetectable (Grzymski *et al.*, 2012) component of marine picoplankton in Antarctic waters. However, these findings are consistent with findings of a logarithmic relationship of cyanobacterial numbers with temperature, where cyanobacteria were found at concentrations of 103 – 104 cells per litre even in the coldest waters, approximately four orders of magnitude less than in waters around Tasmania (Marchant *et al.*, 1987). Cyanophage proteins have also been detected in a metaproteomic analysis of Antarctic Peninsula coastal surface waters (Williams *et al.*, 2012).

**SAR11 and SAR116 clades**

"*Ca.* P. ubique" HTCC1062 is a good representative of total SAR11 abundance in this study, as it is a member of the SAR11 phylotype which is most abundant in SO waters (Brown *et al.*, 2012). "*Ca.* P. ubique" HTCC1062 was the most abundant OTU across all samples and fractions (NZ average: 62%, 25% and 24% of the 0.1 µm, 0.8 µm and 3.0 µm fractions respectively; SZ: 59%, 22% and 18%) and

one of the most significant contributors to variance between the NZ and SZ. The high abundance of SAR11 in the 0.1 µm fraction is consistent with the small size of SAR11 cells (Rappé *et al.*, 2002). The higher representation in the NZ may reflect the competitiveness of SAR11 members in regions with low Dissolved Organic Carbon (DOC) concentrations due to low primary productivity (Giovannoni *et al.*, 2005; Alonso and Pernthaler, 2006), such as the High Nutrient, Low Chlorophyll (HNLC) SAZ. Overall, these findings are consistent with reports that SAR11 is ubiquitous in the world's oceans (Mary *et al.*, 2006; Carlson *et al.*, 2009) and more abundant north of the ACC (Giebel *et al.*, 2009).

OTUs of "*Ca.* P. marinum" from the SAR116 clade were a moderate contributor to variance between the NZ and SZ with higher abundance in the NZ (Figure 5). A genomic analysis reported "*Ca.* P. marinum" IMCC1322 to be a metabolic generalist with genes for aerobic CO fixation, C1 metabolism and a "*Ca.* P. ubique"-like dimethylsulfoniopropionate (DMSP) demethylase, suggesting SAR116 and SAR11 occupy similar ecological niches (Oh *et al.*, 2010). In the Scotia Sea, SAR116 abundance (determined using Fluorescence *In Situ* Hybridization (FISH)) was reported to be higher in more productive waters where SAR11 numbers were lower (Topping *et al.*, 2006). However, this analysis across an extended latitudinal transect indicates that overall SAR11 and SAR116 have similar biogeographic distributions.

**Bacteroidetes**

OTUs of the phylum Bacteroidetes, in particular members of the class Flavobacteria, were found to be abundant (NZ average: 1.2%, 5.0% and 6.9% of the 0.1 µm, 0.8 µm and 3.0 µm fractions respectively; SZ: 2.3%, 9.8% and 9.1%) and significant contributors to variance between the NZ and SZ (Figure 5). Flavobacteria have been previously reported to compose the majority of both Bacteroidetes (Murray and Grzymski, 2007) and total planktonic biomass (Abell and Bowman, 2005) in the SO, as well as being abundant in sea ice (Brown and Bowman, 2001). As heterotrophic degraders of High Molecular Weight (HMW) compounds in the form of both Dissolved Organic Matter (DOM) and Particulate Organic Matter (POM) (Kirchman, 2002), marine Flavobacteria are major components of marine aggregates (Rath *et al.*, 1998; Crump *et al.*, 1999; Zhang *et al.*, 2007). The higher abundance of Flavobacteria OTUs on the 0.8 µm and 3.0 µm fractions indicates their association with particulate matter. Similar size partitioning of SO Flavobacteria has previously been reported (Abell and Bowman, 2005).

The higher abundance of OTUs of Flavobacteria in the SZ may reflect an input of cells from melting sea ice (Brown and Bowman, 2001), the higher rates of primary productivity in the south, and the role of the Flavobacteria as degraders of HMW DOM. Because deposition of marine snow is a major route for sequestration of fixed carbon in the ocean (e.g. Hessen *et al.*, 2004), the Flavobacteria that associate with this particulate matter represent a remineralising shunt, which would decrease carbon sequestration by this route.

**Rhodobacterales**

Members of the order Rhodobacterales were abundant (NZ average: 1.2%, 10% and 5.5% of the 0.1 µm, 0.8 µm and 3.0 µm fractions respectively; SZ: 1.6%, 13% and 7.9%) and high contributors to variance, overrepresented in the SZ on all size fractions. As several members of the Roseobacter clade have been shown to have symbiotic relationships with marine eukaryotic algae (**?**Wagner-Döbler and Biebl, 2006), and their abundance in the SO has previously been linked to phytoplankton blooms (West *et al.*, 2008; Obernosterer *et al.*, 2011), it is likely that their overrepresentation in the SZ is related to the higher density of phytoplankton in the AZ.

OTUs of *Roseobacter denitrificans* Och114 and *Silicibacter pomeroyi* DSS-3 were consistently the most abundant Roseobacter clade representatives. *R. denitrificans* and *S. pomeroyi* fall within a subclade of Aerobic Anoxygenic Phototrophic (AAP) members of the Roseobacter clade (Swingley *et al.*, 2007). These species have diverse mixotrophic metabolisms, with genomic and experimental evidence of photoheterotrophic respiration of organic carbon, fixation of $CO_2$, oxidation of CO, oxidation of reduced sulfur compounds, and utilization of the abundant marine osmolyte DMSP (King, 2003; Moran *et al.*, 2004; Wagner-Döbler and Biebl, 2006; Swingley *et al.*, 2007; Brinkhoff *et al.*, 2008; Howard *et al.*, 2008). This metabolic diversity suggests a complex ecological role, particularly with respect to the capture and release of climatically active gases ($CO_2$, CO, dimethylsulfide) involved in carbon and sulfur cycling.

#### Alteromonadales

Members of the gammaproteobacterial order Alteromonadales were large contributors to variance. Most OTUs were overrepresented in the SZ but some were overrepresented in the NZ on the 3.0 µm fraction (Figure 5). *Colwellia psychrerythraea* 34H was one of the most abundant OTUs in the Alteromonadales that exhibited this distribution (NZ average: 0.14%, 2.2% and 16% of the 0.1 µm, 0.8 µm and 3.0 µm fractions respectively; SZ: 0.52%, 5.1% and 10%). *C. psychrerythraea* 34H was isolated from Arctic sediment, grows well at low temperatures and secretes extracellular polysaccharides (Huston *et al.*, 2000; Junge *et al.*, 2003; Methé *et al.*, 2005). Similar to other *Colwellia* species grown under laboratory conditions, cells have widths of 0.4–0.8 µm and lengths of 1.5–4.5 µm (Jung *et al.*, 2006). Growth temperature can have a major impact on cell morphology, enzyme secretion and global gene expression in psychrophiles (e.g. Feller and Gerday, 2003; Junge *et al.*, 2003; Williams *et al.*, 2011; Cavicchioli, 2006; Campanaro *et al.*, 2011). Moreover, marine bacteria can alter their cell dimensions in response to nutrient flux (e.g. Kjelleberg *et al.*, 1987). It is therefore possible that the populations of Alteromonadales captured on the 3.0 µm filters (overrepresented in the NZ) had different physiological properties to those on the 0.1 and 0.8 µm filters (overrepresented in the SZ).

#### Verrucomicrobia

Two representatives of the phylum Verrucomicrobia, *Coraliomargarita akajimensis* and *Akkermansia* sp. Muc-30, were moderate contributors to variance and overrepresented in the SZ (Figure 5). Surprisingly given the small cell size of *C. akajimensis* (Yoon *et al.*, 2007), its contribution to variance increased with size fraction. A global survey reported a similar fractionation pattern, and suggested marine Verrucomicrobia may be predominantly particle attached (Freitas *et al.*, 2012). However, little else is known about the distribution and ecological roles of marine Verrucomicrobia (Freitas *et al.*, 2012).

### Functional capacities differentiating the zones

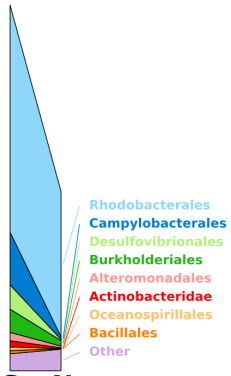
A number of modules with transport functions (sn-glycerol 3-phosphate transport system, dipeptide transport system, peptides/nickel transport system, simple sugar transport system, sulfonate/nitrate/taurine transport system) were overrepresented in the SZ (Table 6). As the genomes of copiotrophic bacteria have evolved to have a higher number of narrow-specificity transporters relative to oligotrophic genomes (Lauro *et al.*, 2009), these differences may reflect the higher nutrient availability and thus a dominance of copiotrophs in the SZ. The taxonomic contributors to these modules were varied, although members of the Rhodobacterales were prominent (Figure 7).

The glycine betaine/proline transport module was also overrepresented in the SZ, though this probably reflects glycine betaine's role as an osmo- and cryoprotectant in the colder SZ waters. This is supported by the major taxonomic contributor to this module, genus *Psychromonas*, which includes several psychrophilic species.

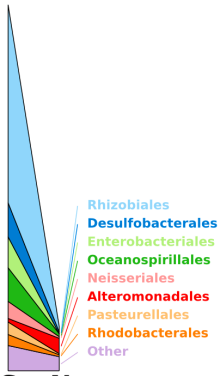
One exception to this pattern was the branched-chain amino acid transport system module, overrepresented in the NZ. The genera *Pelagibacter* and *Puniceispirillum* were major contributors to this module's overabundance in the NZ (Figure 7). As both SAR11 (Giovannoni *et al.*, 2005) and SAR116 (Grote *et al.*, 2011) representatives encode branched-chain amino acid transporters, the abundance of this module is likely to represent taxonomic differences between the zones.

Biosynthesis pathways for all major nucleic acids (pyrimidine deoxyribonucleotide biosynthesis, adenine nucleotide biosynthesis, guanine nucleotide biosynthesis) were consistently high contributors to variance and overabundant in the NZ. This pattern seems inconsistent with the more oligotrophic nature of the NZ, as oligotrophic cells generally have smaller genomes (Lauro *et al.*, 2009) and slower growth rates than copiotrophs, and would therefore be expected to require a lower rate of de novo nucleotide biosynthesis. A possible explanation for this is that SZ cells have higher availability of extracellular DNA as a byproduct of decaying phytoplankton (Lomas and Moran, 2011), which can be imported and salvaged for nucleic acids (Paul *et al.*, 1988) thus reducing the requirement for de novo synthesis. No single taxonomic group contributed a large fraction of the difference in this module (Figure 7), suggesting this is a widespread adaptation.

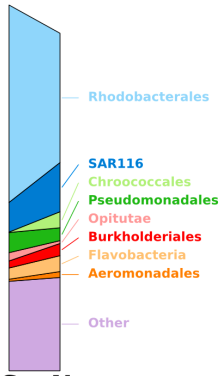
The methionine salvage pathway module had a large contribution to variance between the zones and was overrepresented north of the PF. This may reflect the higher availability of DMSP in the SZ



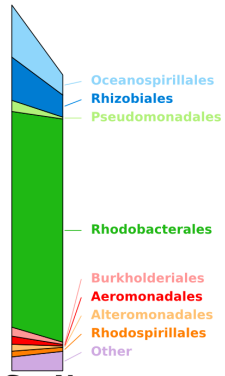

sn−Glycerol 3−phosphate transport system



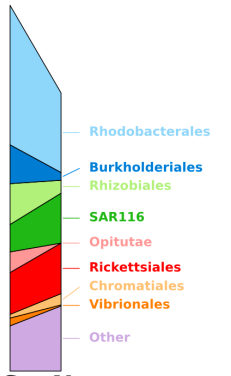

Dipeptide transport system



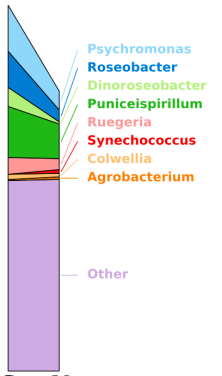

Peptides/nickel transport system



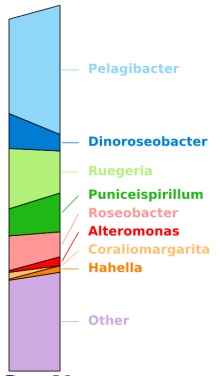

Simple sugar transport system



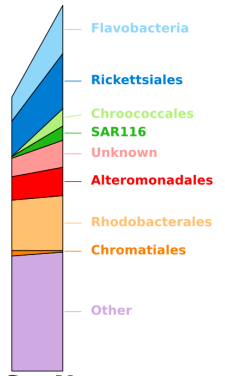

Sulfonate/nitrate/taurine transport system



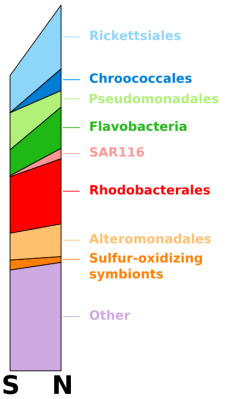

Glycine betaine/proline transport system



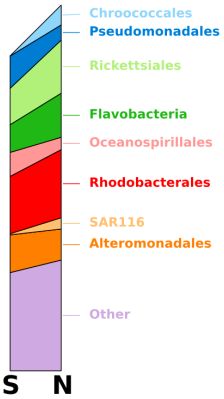

Branched−chain amino acid transport system



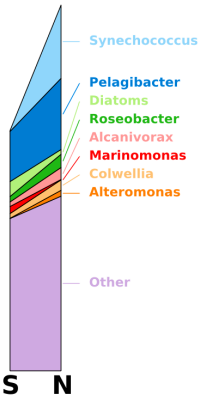

Pyrimidine deoxyribonuleotide biosynthesis



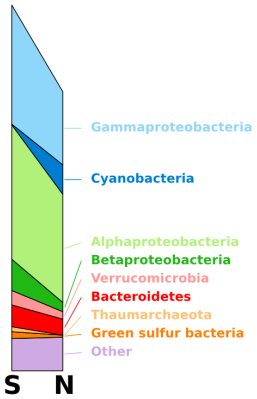

Adenine nucleotide biosynthesis



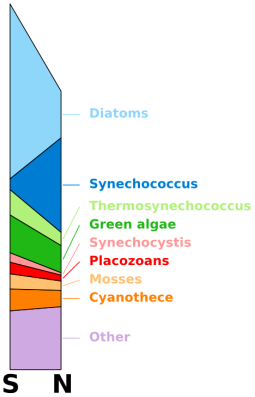

Guanine nucleotide biosynthess



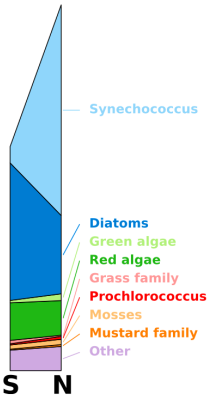

Methionine salvage pathway



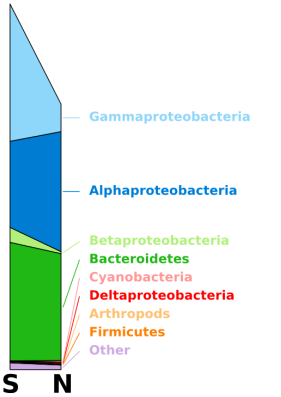

Sulfur reduction, sulfate → H2S



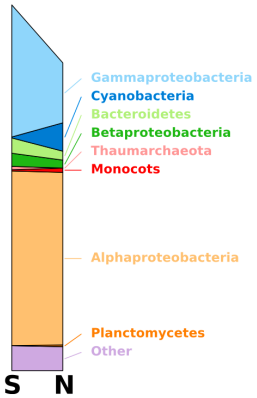

Photosystem I




Photosystem II




Histidine degradation, histidine N−formiminoglutamate → glutamate




Histidine biosynthesis, PRPP → histidine

**Figure 7:** Decomposition of KEGG modules of interest to contributing classes, orders or genera. The left side of each stack (S) indicates the proportion of the module abundance contributed by each class, order or genus in the South Zone, while the right side (N) represents the North Zone. As the contributions are relative and represent unitless module abundances, no axis is given and proportions are not comparable between modules. Contributing classes, orders or genera are arranged in descending order of the difference in the relative contributions between the zones. Only the eight highest contributors for each module are shown, with the remainder collapsed into the "Other" group. The taxonomic ranks to which each module was decomposed are as follows: sn-glycerol 3-phosphate transport, peptide-nickel transport, simple sugar transport and sulfonate/nitrate/taurine transport were decomposed to order; glycine betaine/proline transport and branched-chain amino acid transport to genus; pyrimidine deoxyribonucleotide biosynthesis, adenine nucleotide biosynthesis and guanine nucleotide biosynthesis to order; methionine salvage to genus; sulphur reduction to class; photosystem I and photosystem II to genus; histidine degradation to glutamate and histidine biosynthesis to class.

as a byproduct of blooming eukaryotic algae. DMSP is a major carbon and sulfur source for marine microorganisms, and is commonly assimilated by bacteria through demethylation to methylmercaptopropionate (MMPA), followed by further catabolism to the climatically important compounds dimethylsulfide or methanethiol (review in Curson *et al.*, 2011). However, when DMSP is scarce, MMPA may be derived from methionine through the alternative methionine salvage pathway (Reisch *et al.*, 2011). The genus *Synechococcus*, a noted contributor to marine DMSP uptake and assimilation (Vila-Costa *et al.*, 2006), was a very high contributor to the abundance of this module in the NZ (Figure 7), suggesting *Synechococcus* species may use this route when DMSP is unavailable.

The sulfur reduction module was overrepresented in the SZ, and it is likely that this result is strongly driven by taxonomic differences. While the taxonomic breakdown indicated a large number of genera contributed to the difference, the Gammaproteobacteria were the highest-contributing class (Figure 7). This module also includes the assimilatory sulfate reduction pathway, which is widespread in marine bacteria, but is absent from SAR11, with known representatives reported to lack genes for assimilatory sulfate reduction (cysDNCHIJ) (Tripp *et al.*, 2008). The higher relative abundance of SAR11 in the NZ may therefore contribute to the lower abundance of genes for assimilatory sulfate reduction in that zone.

The sulfur reduction module also included adenylylsulfate reductase (APS reductase, encoded by aprAB), an enzyme implicated in sulfite detoxification during heterotrophic growth on organosulfonates (Meyer and Kuever, 2007) (N.B. in recent KEGG releases, aprA is no longer included in this module). As the GSO-EOSA-1 representative SUP05 has been found to encode APS reductase, the overabundance of this module may reflect sulfur oxidation through the reverse dissimilatory sulfate reduction pathway (Walsh *et al.*, 2009). Also, Roseobacter clade bacteria are involved in the decomposition of abundant organic sulfur compounds (e.g. DMSP, organosulfonates), and hence have been accorded an important role in marine sulfur cycling (Moran *et al.*, 2007).

The photosystem II module was overrepresented in the NZ. Given the underrepresentation of cyanobacterial OTUs in the SZ, this may reflect a dominance of primary production by eukaryotic algae south of the PF and cyanobacteria to the north. Decomposition of the taxonomic affiliations of ortholog groups contributing to this module found OTUs of *Synechococcus* and *Prochlorococcus* to be major contributors to the difference (Figure 7). Variation in the photosystem I module, which was marginally overrepresented in the SZ, could largely be attributed to diatoms and other eukaryotic phytoplankton (Figure 7), again supporting a dominance of eukaryotic phytoplankton in SZ primary production. Diatoms have previously been reported at higher abundance south of the PF, and their distribution is likely to be linked to the higher concentration of dissolved silica in that region (Trull *et al.*, 2001). As both eukaryotic phytoplankton and cyanobacteria would be expected to encode both complete photosystems, the differences in module abundance probably reflect the degree of similarity between the photosystem I and II genes in the KEGG database and those found in the sampled environments.

The histidine degradation to glutamate module, which comprises four ortholog groups mediating the degradation of histidine to glutamate via N-formiminoglutamate, was overrepresented in the SZ. The histidine biosynthesis module was also overrepresented in the SZ. While the concentration of dissolved histidine in the SO is generally low (Kawahata and Ishizuka, 2000), blooming eukaryotic phytoplankton (which are more prevalent in the SZ) may deplete nitrate while releasing Dissolved Free Amino Acids (DFAA). As DFAA become available, they are used by bacteria to sense

the decaying bloom. Histidine may therefore act as a proxy for DFAA to regulate the expression of bacterial aminopeptidases, which are involved in lysing diatoms (Bidle and Azam, 2001). The class Bacteroidetes, while a small contributor to the histidine biosynthesis module in the SZ, was a large contributor to histidine degradation (Figure 7), supporting an association between Bacteroidetes and phytoplanktonic bloom products. It is also possible that uptake and degradation of histidine to glutamate (which generates ammonia as a by-product) may function as a limited nitrogen source.

**Conclusions: Biogeographic role of the Polar Front**

These results show that there are major taxonomic and functional differences across the PF. The differences in functional potential between the NZ and SZ reflect both their taxonomic profiles and fundamental trophic and ecological differences. In particular, they provide genomic support that the NZ is more oligotrophic than the SZ (Pollard *et al.*, 2002; Giovannoni *et al.*, 2005; Alonso and Pernthaler, 2006; Lauro *et al.*, 2009), and are consistent with the observation that primary production is higher south of the PF (Strutton *et al.*, 2000; Williams *et al.*, 2010). Our findings extend previous work in defining the PF as a strong biogeographic boundary which shapes not only the composition, but also the functional capacity of microbial communities in the SO.

A possible alternative hypothesis for the observed separation is that the samples are partitioned by the continental margin, as all but one of the SZ samples were taken in waters over the Antarctic continental shelf and slope in the vicinity of the Mertz glacier polynya. However, ANOSIM analysis of an alternative grouping of the samples into "polynya" and "open ocean" had poorer support ($R = 0.309$, $p < 0.01$) than the grouping based on the PF. Additional taxonomic profiles for samples taken from the region south of the PF in other seasons (austral summers 06/07, 08/09) and in other sectors of the SO (70–115° E) also supported the PF as the major discriminator (data not shown). TODO write up these results Taken together, this evidence strongly supports the hypothesis that the PF is a major biogeographical boundary in the SO independent of a latitudinal gradient or of the effect of the continental margin and Mertz polynya.

These results do not exclude the possibility that other major SO fronts, particularly the STF and SAF, are also significant biogeographic boundaries, as has been reported in some previous reports for specific taxonomic groups (e.g. Abell and Bowman, 2005). While the sampling resolution in this study was not sufficient to resolve the effects of other fronts, there are some indications in the data of further structure within the zones. The two samples north of the STF had significantly larger cyanobacterial populations than the remaining NZ samples (see discussion of *Prochlorococcus* and *Synechococcus*, above). Future sampling across these fronts at higher resolution will provide the data necessary to investigate finer biogeographic patterns.

The nature and function of microbial communities in the SO are of global significance because of the large oceanic expanse that is involved and the importance of the carbon fixation and nutrient cycling that occurs there. Knowledge of these communities and their biogeographic drivers has relevance for understanding and predicting the long-term effects of environmental change in the region. These findings provide a basis for predicting how climate change-driven shifts in the SO may affect microbial communities; in particular, the effects of changes in the nature and location of the ACC on the ecosystem functions of SO microorganisms.

# Mesoscale biogeographic drivers of planktonic diversity

## Introduction

## Methods

### Sampling

Sampling[5] was conducted on board the RSV *Aurora Australis* during cruise V3 from January 25th–February 12th 2012. This cruise occupied two latitudinal transects: one from Hobart, Australia ($\sim$ 44° S) to the Mertz Glacier, Antarctica ($\sim$ 67° S), within a longitudinal range of 140–150° E; the second from waters north of Cape Poinsett, Antarctica ($\sim$ 66° S) to Freemantle, Australia ($\sim$ 32° S) within a longitudinal range of 110–120° E.

TODO need a map of samples (once sequencing results back)

At each station, $\sim$ 250–560 L of seawater was pumped from $\sim$ 1.5–2.5 m below the sea surface into drums stored at ambient temperature on deck. At some stations, an additional sample was taken from the Deep Chlorophyll Maximum (DCM), as determined by chlorophyll fluorescence measurements taken from a Conductivity, Temperature and Depth (CTD) (SeaBird, Bellevue, USA) cast at each sampling station. In the case of deep and intermediate water samples, $\sim$ 120–240 L of seawater was collected from Niskin bottles attached to a CTD. The depths were selected based on temperature, salinity and dissolved oxygen profiles established by CTD casts at each sampling station to capture water from the targeted water mass. Profiles were generated on the CTD downcast, and bottle firings (i.e. sample collection) on the returning upcast at the selected depths.

Seawater samples were prefiltered through a 20 µm plankton net, then filtrate was captured on sequential 3.0 µm 0.8 µm and 0.1 µm 293 mm polyethersulfone membrane filters (Pall, Port Washington, USA), and immediately stored at $-20$ °C (Rusch *et al.*, 2007; Ng *et al.*, 2010).

### DNA extraction

DNA extraction was performed using a modified version of the phenol-chloroform method described in Rusch *et al.* (2007). Samples were thawed in a 37 °C water bath. Half of the storage buffer ($\sim$ 10 mL) was decanted into a clean 50 mL centrifuge tube. If the volume decanted was less than 10 mL, the difference was made with sterile water (Sigma-Aldrich, St. Louis, USA). An equal volume of 50% sucrose lysis buffer (50 mM TRIS-HCl, 40 mM EDTA, 0.75 M Sucrose, pH 8) was added such that the final concentration was 25% sucrose lysis buffer. A small pinch of lysozyme (Sigma-Aldrich, St. Louis, USA) (final concentration $\sim$ 2.5 mg/mL) and 1 mL TRIS-EDTA (10 mM TRIS, 1 mM EDTA, pH 8) was added.

The filter membrane was removed from the storage tube and cut in half aseptically. One half was returned to the storage tube, which was refrozen at $-80$ °C. The remaining half was cut in half again, and one quarter-filter placed atop the other such that the biomass (filtrand) on each piece was facing outwards. Keeping the filters together, they were cut into very fine ($\sim$ 3 mm by 10 mm) strips, which were placed in the 50 mL centrifuge tube containing the buffer and lysozyme mixture. This tube was

[5]Sampling was performed by David Wilkins, Timothy J. Williams and Sheree Yau.

mixed by gentle inversion, then tapped such that all filter strips collected at the bottom of the tube and were covered by lysis buffer. The tube was then incubated in a 37 °C shaking water bath at 275 RPM for 30–60 min.

200 µL of 20 mg/mL Proteinase K (Sigma-Aldrich, St. Louis, USA) was added to the tube, which was mixed by gentle inversion. The tube was gently tapped such that all filter strips collected at the bottom covered by lysis buffer. The tube was then subjected to three freeze-thaw cycles, each cycle consisting of 20–30 min in a −80 °C freezer followed by 20–30 min in a 55 °C water bath. After the final complete thaw, 200 µL of 20 mg/mL Proteinase K and 2 mL of 10% SDS (Sigma-Aldrich, St. Louis, USA) were added to the tube. The tube was mixed by gentle inversion then gently tapped such that all filter strips collected at the bottom covered by lysis buffer. It was then incubated in a 55 °C shaking water bath at 175 RPM for two hours.

The supernatant was pipetted from the tube using a genomic tip and split evenly into two new 50 mL centrifuge tubes. An equal volume of buffer-saturated (10 mM TRIS HCl, 1 mM EDTA, pH 8) phenol (Sigma-Aldrich, St. Louis, USA) was added to each of the tubes, which were mixed by gentle inversion. The mixtures were then fractionated in a fixed-angle rotor centrifuge for 15 min at 3700 RPM at room temperature. The bottom layer of each tube was removed by pipette into a new 50 mL centrifuge tube. Each of these two tubes was then made to 50 mL with sterile water (Sigma-Aldrich, St. Louis, USA). After mixing by gentle inversion, each 50 mL mixture was then split evenly into two new 50 mL centrifuge tubes, resulting in four tubes each containing 25 mL of mixture. These tubes were then made to 50 mL with 1-propanol (Sigma-Aldrich, St. Louis, USA). The mixtures were homogenised by gentle inversion and incubated at 4 °C overnight.

Following incubation, the tubes were centrifuged using a fixed-angle rotor for 30 min at 7500 RPM and room temperature. The majority of the supernatant was removed by decanting, and the tubes left to sit until the remaining supernatant ($\sim$ 1 mL) collected at the bottom over the precipitated pellet. The pellet was then resuspended by gentle pipetting with a genomic tip, and the suspension placed in a new 1.5 mL microcentrifuge tube (four tubes total). These tubes were then centrifuged in a microcentrifuge for 10 minutes at 13,000 RPM and room temperature. The supernatant was removed by pipette and the tubes placed in a 37 °C heat block with the lids opened and covered by a sterile KimWipe (Kimberly-Clark, Irving, USA) for 10 min, or longer if the supernatant did not evaporate completely in that time. 93.75 µL of TRIS-EDTA was added to each tube, and the tubes were incubated at 4 °C for one hour to allow the DNA pellet to redissolve.

After this incubation, the pellets were gently pipetted with a genomic tip to ensure complete resuspension. The suspensions from all four tubes were combined, and an additional 750 µL of TRIS-EDTA added. This was then split evenly into two new 1.5 mL microcentrifuge tubes ($\sim$ 562.5 µL per tube).

750 µL of buffer-saturated phenol was added to each tube, and the tubes mixed gently by inversion until a visible emulsion formed. Phase separation was performed by centrifugation for 5 min at 13,000 RPM and room temperature. The upper (aqueous) phase was removed to a new 1.5 mL microcentrifuge tube using a genomic tip.

750 µL of phenol-chloroform-isoamyl alcohol (25:24:1) mixture (Sigma-Aldrich, St. Louis, USA) was added to each tube, and the tubes mixed by gentle inversion until a visible emulsion formed. Phase separation was performed by centrifugation for 5 min at 13,000 RPM and room temperature. The upper (aqueous) phase was removed to a new 1.5 mL microcentrifuge tube using a genomic tip.

75 µL of 3 M sodium acetate (pH 8) and 750 µL of 1-propanol was added to each tube. The tubes were centrifuged at 13,000 RPM and room temperature for 30 min to precipitate the DNA. The supernatant was removed by pipetting, and 100 µL of 70% ethanol added. The tubes were centrifuged again at 13,000 RPM and room temperature for 5 min. The supernatant was removed by pipetting and the DNA pellet dried in a 37 °C heat block. The DNA was dissolved overnight in 40–200 µL of TRIS-EDTA, depending on the expected yield.

# Results

# Discussion

# References


Abell G. G. J. and Bowman J. P. (2005). Ecological and biogeographic relationships of class Flavobacteria in the Southern Ocean. *FEMS Microbiology Ecology*, 51:265–277.

Alonso C. and Pernthaler J. (2006). Roseobacter and SAR11 dominate microbial glucose uptake in coastal North Sea waters. *Environmental Microbiology*, 8(11):2022–2030.

André J. M., Navarette C., Blanchot J., and Radenac M. H. (1999). Picophytoplankton dynamics in the equatorial Pacific: Growth and grazing rates from cytometric counts. *Journal of Geophysical Research*, 104(C2):3369–3380.

Angly F. E., Felts B., Breitbart M., Salamon P., Edwards R. A., Carlson C., Chan A. M., Haynes M., Kelley S., Liu H., Mahaffy J. M., Mueller J. E., Nulton J., Olson R., Parsons R., Rayhawk S., Suttle C. A., and Rohwer F. (2006). The marine viromes of four oceanic regions. *PLoS Biology*, 4(11):e368.

Angly F. E., Willner D., Prieto-Davó A., Edwards R. A., Schmieder R., Vega-Thurber R., Antonopoulos D. A., Barott K., Cottrell M. T., Desnues C., Dinsdale E. A., Furlan M., Haynes M., Henn M. R., Hu Y., Kirchman D. L., McDole T., McPherson J. D., Meyer F., Miller R. M., Mundt E., Naviaux R. K., Rodriguez-Mueller B., Stevens R., Wegley L., Zhang L., Zhu B., and Rohwer F. (2009). The GAAS Metagenomic Tool and Its Estimations of Viral and Microbial Average Genome Size in Four Major Biomes. *PLoS Computational Biology*, 5(12):e1000593.

Aoki S., Yoritaka M., and Masuyama A. (2003). Multidecadal warming of subsurface temperature in the Indian sector of the Southern Ocean. *Journal of Geophysical Research*, 108(C4):8081–8088.

Baas Becking L. G. M. *Geobiologie of inleiding tot de milieukunde*. W.P. Van Stockum & Zoon, The Hague, 1934.

Beja O., Aravind L., Koonin E. V., Suzuki M. T., Hadd A., Nguyen L. P., Jovanovich S. B., Gates C. M., Feldman R. A., Spudich J. L., Spudich E. N., and DeLong E. F. (2000). Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science*, 289(5486):1902–1906.

Béjà O., Suzuki M. T., Heidelberg J. F., Nelson W. C., Preston C. M., Hamada T., Eisen J. A., Fraser C. M., and DeLong E. F. (2002). Unsuspected diversity among marine aerobic anoxygenic phototrophs. *Nature*, 415(6872):630–633.

Berg I. A., Kockelkorn D., Buckel W., and Fuchs G. (2007). A 3-Hydroxypropionate/4-Hydroxybutyrate Autotrophic Carbon Dioxide Assimilation Pathway in Archaea. *Science*, 318(5857):1782–1786.

Bidle K. D. and Azam F. (2001). Bacterial control of silicon regeneration from diatom detritus: significance of bacterial ectohydrolases and species identity. *Limnology and Oceanography*, 46(7):1606–1623.

Biebl H., Allgaier M., Tindall B. J., Koblížek M., Lünsdorf H., Pukall R., and Wagner-Döbler I. (2005). *Dinoroseobacter shibae* gen. nov., sp. nov., a new aerobic phototrophic bacterium isolated from dinoflagellates. *International Journal of Systematic and Evolutionary Microbiology*, 55(Pt 3):1089–1096.

Bissett A., Richardson A. E., Baker G., Wakelin S., and Thrall P. H. (2010). Life history determines biogeographical patterns of soil bacterial communities over multiple spatial scales. *Molecular Ecology*, 19(19):4315–4327.

Böning C. W., Dispert A., Visbeck M., Rintoul S. R., and Schwarzkopf F. U. (2008). The response of the Antarctic Circumpolar Current to recent climate change. *Nature Geoscience*, 1(12):864–869.

Bowman J. P. and McCuaig R. D. (2003). Biodiversity, community structural shifts, and biogeography of prokaryotes within Antarctic continental shelf sediment. *Applied and Environmental Microbiology*, 69(5):2463–2483.

Bowman J. P., Rea S. M., McCammon S. A., and McMeekin T. A. (2000). Diversity and community structure within anoxic sediment from marine salinity meromictic lakes and a coastal meromictic marine basin, Vestfold Hills, Eastern Antarctica. *Environmental Microbiology*, 2(2):227–237.

Boyd P. W., Jickells T., Law C. S., Blain S., Boyle E. A., Buesseler K. O., Coale K. H., Cullen J. J., Baar H. J. W.de, Follows M., Harvey M., Lancelot C., Levasseur M., Owens N. P. J., Pollard R., Rivkin R. B., Sarmiento J., Schoemann V., Smetacek V., Takeda S., Tsuda A., Turner S., and Watson A. J. (2007). Mesoscale Iron Enrichment Experiments 1993-2005: Synthesis and Future Directions. *Science*, 315(5812):612–617.

Brinkhoff T., Giebel H.-A., and Simon M. (2008). Diversity, ecology, and genomics of the Roseobacter clade: a short overview. *Archives of Microbiology*, 189(6):531–539.

Brinkmeyer R., Knittel K., Jürgens J., Weyland H., Amann R., and Helmke E. (2003). Diversity and Structure of Bacterial Communities in Arctic versus Antarctic Pack Ice. *Applied and Environmental Microbiology*, 69(11):6610–6619.

Brown M. V. and Bowman J. P. (2001). A molecular phylogenetic survey of sea-ice microbial communities (SIMCO). *FEMS Microbiology Ecology*, 35(3):267–275.

Brown M. V., Lauro F. M., DeMaere M. Z., Muir L., Wilkins D., Thomas T., Riddle M. J., Fuhrman J. A., Andrews-Pfannkoch C., Hoffman J. M., McQuaid J. B., Allen A., Rintoul S. R., and Cavicchioli R. (2012). Global biogeography of SAR11 marine bacteria. *Molecular systems biology*, 8.

Buchan A., González J. M., and Moran M. A. (2005). Overview of the marine Roseobacter lineage. *Applied and Environmental Microbiology*, 71(10):5665–5677.

Callahan J. E. (1972). The structure and circulation of deep water in the Antarctic. *Deep Sea Research and Oceanographic Abstracts*, 19(8):563–575.

Campanaro S., Williams T. J., Burg D. W., De Francisci D., Treu L., Lauro F. M., and Cavicchioli R. (2011). Temperature-dependent global gene expression in the Antarctic archaeon *Methanococcoides burtonii*. *Environmental Microbiology*, 13(8):2018–2038.

Canfield D. E., Stewart F. J., Thamdrup B., De Brabandere L., Dalsgaard T., DeLong E. F., Revsbech N. P., and Ulloa O. (2010). A Cryptic Sulfur Cycle in Oxygen-Minimum-Zone Waters off the Chilean Coast. *Science*, 330(6009):1375–1378.

Carlson C. A., Morris R., Parsons R., Treusch A. H., Giovannoni S. J., and Vergin K. (2009). Seasonal dynamics of SAR11 populations in the euphotic and mesopelagic zones of the northwestern Sargasso Sea. *The ISME Journal*, 3(3):283–295.

Cavicchioli R. (2006). Cold-adapted archaea. *Nature Reviews Microbiology*, 4(5):331–343.

Chiba S., Ishimaru T., Hosie G. W., and Fukuchi M. (2001). Spatio-temporal variability of zooplankton community structure off east Antarctica (90 to 160°E). *Marine Ecology Progress Series*, 216:95–108.

Cho J. C. and Giovannoni S. J. (2004). Cultivation and Growth Characteristics of a Diverse Group of Oligotrophic Marine Gammaproteobacteria. *Applied and Environmental Microbiology*, 70(1):432–440.

Christaki U., Obernosterer I., Van Wambeke F., Veldhuis M., Garcia N., and Catala P. (2008). Microbial food web structure in a naturally iron-fertilized area in the Southern Ocean (Kerguelen Plateau). *Deep Sea Research Part II: Topical Studies in Oceanography*, 55(5-7):706–719.

Church M. J., DeLong E. F., Ducklow H. W., Karner M. B., Preston C. M., and Karl D. M. (2003). Abundance and distribution of planktonic Archaea and Bacteria in the waters west of the Antarctic Peninsula. *Limnology and Oceanography*, 48(5):1893–1902.

Clarke K. R. and Warwick R. M. *Change in marine communities: an approach to statistical analysis and interpretation*. PRIMER-E, Plymoth, 2nd edition, 2001.

Coleman M. L. M. and Chisholm S. W. S. (2010). Ecosystem-specific selection pressures revealed through comparative population genomics. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 107(43):18634–18639.

Cottrell M. T. and Kirchman D. L. (2000). Community Composition of Marine Bacterioplankton Determined by 16S rRNA Gene Clone Libraries and Fluorescence In Situ Hybridization. *Applied and Environmental Microbiology*, 66(12):5116–5122.

Cottrell M. T., Waidner L. A., Yu L., and Kirchman D. L. (2005). Bacterial diversity of metagenomic and PCR libraries from the Delaware River. *Environmental Microbiology*, 7(12):1883–1895.

Cox P. M., Betts R. A., Jones C. D., Spall S. A., and Totterdell I. J. (2000). Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature*, 408(6809):184–187.

Crump B. C., Armbrust E. V., and Baross J. A. (1999). Phylogenetic analysis of particle-attached and free-living bacterial communities in the Columbia River, its estuary, and the adjacent coastal ocean. *Applied and Environmental Microbiology*, 65(7):3192–3204.

Curson A. R. J., Todd J. D., Sullivan M. J., and Johnston A. W. B. (2011). Catabolism of dimethylsulphoniopropionate: microorganisms, enzymes and genes. *Nature Reviews Microbiology*, 9(12):849–859.

Wit R.de and Bouvier T. (2006). 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? *Environmental Microbiology*, 8(4):755–758.

Deacon G. E. R. (1982). Physical and biological zonation in the Southern Ocean. *Deep Sea Research Part A. Oceanographic Research Papers*, 29(1):1–15.

DeLong E. F., Franks D. G., and Alldredge A. L. (1993). Phylogenetic Diversity of Aggregate-Attached vs. Free-Living Marine Bacterial Assemblages. *Limnology and Oceanography*, 38(5):924–934.

DeLong E. F., Wu K. Y., Prézelin B. B., and Jovine R. V. (1994). High abundance of Archaea in Antarctic marine picoplankton. *Nature*, 371(6499):695–697.

Dinsdale E. A., Edwards R. A., Hall D., Angly F., Breitbart M., Brulc J. M., Furlan M., Desnues C., Haynes M., Li L., McDaniel L., Moran M. A., Nelson K. E., Nilsson C., Olson R., Paul J., Brito B. R., Ruan Y., Swan B. K., Stevens R., Valentine D. L., Thurber R. V., Wegley L., White B. A., and Rohwer F. (2008). Functional metagenomic profiling of nine biomes. *Nature*, 452(7187):629–632.

Dixon J. L., Beale R., and Nightingale P. D. (2011). Rapid biological oxidation of methanol in the tropical Atlantic: significance as a microbial carbon source. *Biogeosciences Discussions*, 8(2):3899–3921.

Ducklow H. W., Myers K., Erickson M., Ghiglione J. F., and Murray A. E. (2011). Response of a summertime Antarctic marine -bacterial community to glucose and ammonium enrichment. *Aquatic Microbial Ecology*, 64(3):205–220.

Dupont C. L., Rusch D. B., Yooseph S., Lombardo M.-J., Richter R. A., Valas R., Novotny M., Yee-Greenbaum J., Selengut J. D., Haft D. H., Halpern A. L., Lasken R. S., Nealson K., Friedman R., and Venter J. C. (2011). Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. pages 1–14.

Eilers H., Pernthaler J., Glöckner F. O., and Amann R. (2000). Culturability and In Situ Abundance of Pelagic Bacteria from the North Sea. *Applied and Environmental Microbiology*, 66(7):3044–3051.

Esper O. and Zonneveld K. A. F. (2002). Distribution of organic-walled dinoflagellate cysts in surface sediments of the Southern Ocean (eastern Atlantic sector) between the Subtropical Front and the Weddell Gyre. *Marine Micropaleontology*, 46(1):177–208.

Evans C., Pearce I., and Brussaard C. P. D. (2009). Viral-mediated lysis of microbes and carbon release in the sub-Antarctic and Polar Frontal zones of the Australian Southern Ocean. *Environmental Microbiology*, 11(11):2924–2934.

Evans C., Thomson P. G., Davidson A. T., Bowie A. R., Enden R.van den, Witte H., and Brussaard C. P. D. (2011). Potential climate change impacts on microbial distribution and carbon cycling in the Australian Southern Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography*, 58(21-22): 2150–2161.

Fandino L. B., Riemann L., Steward G. F., Long R. A., and Azam F. (2001). Variations in bacterial community structure during a dinoflagellate bloom analyzed by DGGE and 16S rDNA sequencing. *Aquatic Microbial Ecology*, 23:119.

Feller G. and Gerday C. (2003). Psychrophilic enzymes: hot topics in cold adaptation. *Nature Reviews Microbiology*, 1(3):200–208.

Foldvik A. and Gammelsrød T. (1988). Notes on Southern Ocean hydrography, sea-ice and bottom water formation. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 67(1-2):3–17.

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

Fuhrman J. A., Schwalbach M. S., and Stingl U. (2008). Proteorhodopsins: an array of physiological roles? *Nature Reviews Microbiology*, 6:488–494.

Fyfe J. C. and Saenko O. A. (2005). Human-induced change in the Antarctic Circumpolar Current. *Journal of Climate*, 18(15):3068–3073.

García-Martínez J. and Rodríguez-Valera F. (2000). Microdiversity of uncultured marine prokaryotes: the SAR11 cluster and the marine Archaea of Group I. *Molecular Ecology*, 9(7):935–948.

Gentile G., Giuliano L., D'Auria G., Smedile F., Azzaro M., De Domenico M., and Yakimov M. M. (2006). Study of bacterial communities in Antarctic coastal waters by a combination of 16S rRNA and 16S rDNA sequencing. *Environmental Microbiology*, 8(12):2150–2161.

Ghiglione J. F. and Murray A. E. (2011). Pronounced summer to winter differences and higher wintertime richness in coastal Antarctic marine bacterioplankton. *Environmental Microbiology*, 14(3): 617–629.

Giebel H.-A., Brinkhoff T., Zwisler W., Selje N., and Simon M. (2009). Distribution of *Roseobacter* RCA and SAR11 lineages and distinct bacterial communities from the subtropics to the Southern Ocean. *Environmental Microbiology*, 11(8):2164–2178.

Giebel H.-A., Kalhoefer D., Lemke A., Thole S., Gahl-Janssen R., Simon M., and Brinkhoff T. (2010). Distribution of *Roseobacter* RCA and SAR11 lineages in the North Sea and characteristics of an abundant RCA isolate. *The ISME Journal*, 5:8–19.

Gille S. T. (2002). Warming of the Southern Ocean Since the 1950s. *Science*, 295(5558):1275–1277.

Giovannoni S. J., Tripp H. J., Givan S., Podar M., Vergin K. L., Baptista D., Bibbs L., Eads J., Richardson T. H., Noordewier M., Rappé M. S., Short J. M., Carrington J. C., and Mathur E. J. (2005). Genome streamlining in a cosmopolitan oceanic bacterium. *Science*, 309(5738):1242–1245.

Giovannoni S. J., Hayakawa D. H., Tripp H. J., Stingl U., Givan S. A., Cho J.-C., Oh H.-M., Kitner J. B., Vergin K. L., and Rappé M. S. (2008). The small genome of an abundant coastal ocean methylotroph. *Environmental Microbiology*, 10(7):1771–1782.

Glöckner F. O., Fuchs B. M., and Amann R. (1999). Bacterioplankton compositions of lakes and oceans: a first comparison based on fluorescence in situ hybridization. *Applied and Environmental Microbiology*, 65(8):3721–3726.

González J. M., Fernández-Gómez B., Fernàndez-Guerra A., Gómez-Consarnau L., Sánchez O., Coll-Lladó M., Del Campo J., Escudero L., Rodríguez-Martínez R., Alonso-Sáez L., Latasa M., Paulsen I., Nedashkovskaya O., Lekunberri I., Pinhassi J., and Pedrós-Alió C. (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(25):8724–8729.

Grossart H. P., Schlingloff A., Bernhard M., Simon M., and Brinkhoff T. (2004). Antagonistic activity of bacteria isolated from organic aggregates of the German Wadden Sea. *FEMS Microbiology Ecology*, 47(3):387–396.

Grote J., Bayindirli C., Bergauer K., Moraes P.Carpintero de, Chen H., D'Ambrosio L., Edwards B., Fernández-Gómez B., Hamisi M., Logares R., Nguyen D., Rii Y. M., Saeck E., Schutte C., Widner B., Church M. J., Steward G. F., Karl D. M., DeLong E. F., Eppley J. M., Schuster S. C., Kyrpides N. C., and Rappé M. S. (2011). Draft genome sequence of strain HIMB100, a cultured representative of the SAR116 clade of marine *Alphaproteobacteria*. *Standards in Genomic Sciences*, 5(3):269–278.

Grzymski J. J., Carter B. J., DeLong E. F., Feldman R. A., Ghadiri A., and Murray A. E. (2006). Comparative Genomics of DNA Fragments from Six Antarctic Marine Planktonic Bacteria. *Applied and Environmental Microbiology*, 72(2):1532–1541.

Grzymski J. J., Riesenfeld C. S., Williams T. J., Dussaq A. M., Ducklow H., Erickson M., Cavicchioli R., and Murray A. E. (2012). A metagenomic assessment of winter and summer bacterioplankton from Antarctica Peninsula coastal surface waters. *The ISME Journal*, 6(10):1901–1915.

Guixa-Boixereu N., Vaqué D., Gasol J. M., Sánchez-Cámara J., and Pedrós-Alió C. (2002). Viral distribution and activity in Antarctic waters. *Deep Sea Research Part II: Topical Studies in Oceanography*, 49 (4):827–845.

Head I. M., Hiorns W. D., Embley T. M., McCarthy A. J., and Saunders J. R. (1993). The phylogeny of autotrophic ammonia-oxidizing bacteria as determined by analysis of 16S ribosomal RNA gene sequences. *Journal of General Microbiology*, 139(6):1147–1153.

Heikes B. G., Chang W., Pilson M. E. Q., Swift E., Singh H. B., Guenther A., Jacob D. J., Field B. D., Fall R., Riemer D., and Brand L. (2002). Atmospheric methanol budget and ocean implication. *Global Biogeochemical Cycles*, 16(4):1133.

Hessen D. O., Ågren G. I., Anderson T. R., Elser J. J., and de Ruiter, P.C. (2004). Carbon sequestration in ecosystems: the role of stoichiometry. *Ecology*, 85(5):1179–1192.

Hollibaugh J. T., Bano N., and Ducklow H. W. (2002). Widespread Distribution in Polar Oceans of a 16S rRNA Gene Sequence with Affinity to *Nitrosospira*-Like Ammonia-Oxidizing Bacteria. *Applied and Environmental Microbiology*, 68(3):1478–1484.

Howard E. C., Sun S., Biers E. J., and Moran M. A. (2008). Abundant and diverse bacteria involved in DMSP degradation in marine surface waters. *Environmental Microbiology*, 10(9):2397–2410.

Hunt B. P. V., Pakhomov E. A., and McQuaid C. D. (2001). Short-term variation and long-term changes inthe oceanographic environment and zooplankton community in the vicinity of a sub-Antarctic archipelago. *Marine Biolog*, 138:369–381.

Huston A. L., Krieger-Brockett B. B., and Deming J. W. (2000). Remarkably low temperature optima for extracellular enzyme activity from Arctic bacteria and sea ice. *Environmental Microbiology*, 2(4): 383–388.

Ingalls A. E., Shah S. R., Hansman R. L., Aluwihare L. I., Santos G. M., Druffel E. R. M., and Pearson A. (2006). Quantifying archaeal community autotrophy in the mesopelagic ocean using natural radiocarbon. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 103 (17):6442–6447.

Iverson V., Morris R. M., Frazar C. D., Berthiaume C. T., Morales R. L., and Armbrust E. V. (2012). Untangling Genomes from Metagenomes: Revealing an Uncultured Class of Marine Euryarchaeota. *Science*, 335(6068):587–590.

Jacobs S. S. (2004). Bottom water production and its links with the thermohaline circulation. *Antarctic Science*, 16(04):427–437.

Jamieson R. E., Rogers A. D., Billett D., Smale D. A., and Pearce D. A. (2012). Patterns of marine bacterioplankton biodiversity in the surface waters of the Scotia Arc, Southern Ocean. *FEMS Microbiology Ecology*, 80:452–468.

Jung S.-Y., Oh T.-K., and Yoon J.-H. (2006). *Colwellia aestuarii* sp. nov., isolated from a tidal flat sediment in Korea. *International Journal of Systematic and Evolutionary Microbiology*, 56(1):33–37.

Junge K., Eicken H., and Deming J. W. (2003). Motility of *Colwellia psychrerythraea* Strain 34H at Subzero Temperatures. *Applied and Environmental Microbiology*, 69(7):4282–4284.

Kalanetra K. M., Bano N., and Hollibaugh J. T. (2009). Ammonia-oxidizing *Archaea* in the Arctic Ocean and Antarctic coastal waters. *Environmental Microbiology*, 11(9):2434–2445.

Kawahata H. and Ishizuka T. (2000). Amino acids in interstitial waters from ODP Sites 689 and 690 on the Maud Rise, Antarctic Ocean. *Geochemical Journal*, 34(4):247–261.

King G. M. (2003). Molecular and Culture-Based Analyses of Aerobic Carbon Monoxide Oxidizer Diversity. *Applied and Environmental Microbiology*, 69(12):7257–7265.

Kirchman D. L. (2002). The ecology of Cytophaga–Flavobacteria in aquatic environments. *FEMS Microbiology Ecology*, 39(2):91–100.

Kirchman D. L. *Microbial ecology of the oceans*. John Wiley & Sons, Inc., Hoboken, New Jersey, second edition, 2008.

Kjelleberg S., Hermansson M., and Mårdén P. (1987). The transient phase between growth and nongrowth of heterotrophic bacteria, with emphasis on the marine environment. *Annual Review of Microbiology*, 41:25–49.

Koh E. Y., Phua W., and Ryan K. G. (2011). Aerobic anoxygenic phototrophic bacteria in Antarctic sea ice and seawater. *Environmental Microbiology Reports*, 3(6):710–716.

Kuwahara H., Yoshida T., Takaki Y., Shimamura S., Nishi S., Harada M., Matsuyama K., Takishita K., Kawato M., Uematsu K., Fujiwara Y., Sato T., Kato C., Kitagawa M., Kato I., and Maruyama T. (2007). Reduced Genome of the Thioautotrophic Intracellular Symbiont in a Deep-Sea Clam, *Calyptogena okutanii*. *Current Biology*, 17(10):881–886.

Laubscher R. K., Perissinotto R., and McQuaid C. D. (1993). Phytoplankton production and biomass at frontal zones in the Atlantic sector of the Southern Ocean. *Polar Biology*, 13(7).

Lauro F. M., McDougald D., Thomas T., Williams T. J., Egan S., Rice S., DeMaere M. Z., Ting L., Ertan H., Johnson J., Ferriera S., Lapidus A., Anderson I., Kyrpides N., Munk A. C., Detter C., Han C. S., Brown M. V., Robb F. T., Kjelleberg S., and Cavicchioli R. (2009). The genomic basis of trophic strategy in marine bacteria. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 106(37):15527–15533.

Lauro F. M., DeMaere M. Z., Yau S., Brown M. V., Ng C., Wilkins D., Raftery M. J., Gibson J. A., Andrews-Pfannkoch C., Lewis M., Hoffman J. M., Thomas T., and Cavicchioli R. (2011). An integrative study of a meromictic lake ecosystem in Antarctica. *The ISME Journal*, 5(5):879–895.

Liu H., Nolla H. A., and Campbell L. (1997). *Prochlorococcus* growth rate and contribution to primary production in the equatorial and subtropical North Pacific Ocean. *Aquatic Microbial Ecology*, 12(1): 39–47.

Liu H., Campbell L., Landry M. R., Nolla H. A., Brown S. L., and Constantinou J. (1998). *Prochlorococcus* and *Synechococcus* growth rates and contributions to production in the Arabian Sea during the 1995 Southwest and Northeast Monsoons. *Deep Sea Research Part II: Topical Studies in Oceanography*, 45 (10-11):2327–2352.

Lo Giudice A., Caruso C., Mangano S., Bruni V., Domenico M., and Michaud L. (2011). Marine Bacterioplankton Diversity and Community Composition in an Antarctic Coastal Environment. *Microbial Ecology*, 63(1):210–223.

Lomas M. W. and Moran S. B. (2011). Evidence for aggregation and export of cyanobacteria and nano-eukaryotes from the Sargasso Sea euphotic zone. *Biogeosciences*, 8(1):203–216.

López-García P., López-López A., Moreira D., and Rodríguez-Valera F. (2001). Diversity of free-living prokaryotes from a deep-sea site at the Antarctic Polar Front. *FEMS Microbiology Ecology*, 36(2-3): 193–202.

Ludwig W., Strunk O., Westram R., Richter L., Meier H., Yadhukumar , Buchner A., Lai T., Steppi S., Jobb G., Förster W., Brettske I., Gerber S., Ginhart A. W., Gross O., Grumann S., Hermann S., Jost R., König A., Liss T., Lüssmann R., May M., Nonhoff B., Reichel B., Strehlow R., Stamatakis A., Stuckmann N., Vilbig A., Lenke M., Ludwig T., Bode A., and Schleifer K.-H. (2004). ARB: a software environment for sequence data. *Nucleic Acids Research*, 32(4):1363–1371.

Malmstrom R. R., Cottrell M. T., Elifantz H., and Kirchman D. L. (2005). Biomass production and assimilation of dissolved organic matter by SAR11 bacteria in the Northwest Atlantic Ocean. *Applied and Environmental Microbiology*, 71(6):2979–2986.

Marchant H. J., Davidson A. T., and Wright S. W. (1987). The distribution and abundance of chroococcoid cyanobacteria in the Southern Ocean. *Proc. NIPR Symp. Polar Biol*, 1:1–9.

Mary I., Heywood J. L., Fuchs B. M., Amann R., Tarran G. A., Burkill P. H., and Zubkov M. V. (2006). SAR11 dominance among metabolically active low nucleic acid bacterioplankton in surface waters along an Atlantic meridional transect. *Aquatic Microbial Ecology*, 45(2):107–113.

Massana R., Taylor L. T., Murray A. E., Wu K. Y., Jeffrey W. H., and DeLong E. F. (1998). Vertical Distribution and Temporal Variation of Marine Planktonic Archaea in the Gerlache Strait, Antarctica, During Early Spring. *Limnology and . . .*, 43(4):607–617.

Massana R., DeLong E. F., and Pedrós-Alió C. (2000). A Few Cosmopolitan Phylotypes Dominate Planktonic Archaeal Assemblages in Widely Different Oceanic Provinces. *Applied and Environmental Microbiology*, 66(5):1777–1787.

Mayali X., Franks P. J. S., and Azam F. (2008). Cultivation and Ecosystem Role of a Marine *Roseobacter* Clade-Affiliated Cluster Bacterium. *Applied and Environmental Microbiology*, 74(9):2595–2603.

Merbt S. N., Stahl D. A., Casamayor E. O., Martí E., Nicol G. W., and Prosser J. I. (2012). Differential photoinhibition of bacterial and archaeal ammonia oxidation. *FEMS Microbiology Letters*, 327(1): 41–46.

Methé B. A., Nelson K. E., Deming J. W., Momen B., Melamud E., Zhang X., Moult J., Madupu R., Nelson W. C., Dodson R. J., Methe B. A., Nelson K. E., Deming J. W., Momen B., Melamud E., Zhang X., Moult J., Madupu R., Nelson W. C., Dodson R. J., Brinkac L. M., Daugherty S. C., Durkin A. S., DeBoy R. T., Kolonay J. F., Sullivan S. A., Zhou L., Davidsen T. M., Wu M., Huston A. L., Lewis M., Weaver B., Weidman J. F., Khouri H., Utterback T. R., Feldblyum T. V., and Fraser C. M. (2005). The psychrophilic lifestyle as revealed by the genome sequence of *Colwellia psychrerythraea* 34H through genomic and proteomic analyses. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 102(31):10913–10918.

Meyer B. and Kuever J. (2007). Molecular Analysis of the Diversity of Sulfate-Reducing and Sulfur-Oxidizing Prokaryotes in the Environment, Using *aprA* as Functional Marker Gene. *Applied and Environmental Microbiology*, 73(23):7664–7679.

Mikaloff Fletcher S. E., Gruber N., Jacobson A. R., Doney S. C., Dutkiewicz S., Gerber M., Follows M., Joos F., Lindsay K., Menemenlis D., Mouchet A., Müller S. A., and Sarmiento J. L. (2006). Inverse estimates of anthropogenic $CO_2$ uptake, transport, and storage by the ocean. *Global Biogeochemical Cycles*, 20(2):GB2002.

Miller T. R. and Belas R. (2004). Dimethylsulfoniopropionate Metabolism by *Pfiesteria*-Associated *Roseobacter* spp. *Applied and Environmental Microbiology*, 70(6):3383–3391.

Mira A., Ochman H., and Moran N. A. (2001). Deletional bias and the evolution of bacterial genomes. *Trends in genetics : TIG*, 17(10):589–596.

Moore J. K., Abbott M. R., and Richman J. G. (1999). Location and dynamics of the Antarctic Polar Front from satellite sea surface temperature data. *Journal of Geophysical Research*, 104:3052–3073.

Moran M. A., Belas R., Schell M. A., González J. M., Sun F., Sun S., Binder B. J., Edmonds J., Ye W., Orcutt B., Howard E. C., Meile C., Palefsky W., Goesmann A., Ren Q., Paulsen I., Ulrich L. E., Thompson L. S., Saunders E., and Buchan A. (2007). Ecological Genomics of Marine Roseobacters. *Applied and Environmental Microbiology*, 73(14):4559–4569.

Moran M. A., González J. M., and Kiene R. P. (2003). Linking a Bacterial Taxon to Sulfur Cycling in the Sea: Studies of the Marine Roseobacter Group. *Geomicrobiology Journal*, 20(4):375–388.

Moran M. A., Buchan A., González J. M., Heidelberg J. F., Whitman W. B., Kiene R. P., Henriksen J. R., King G. M., Belas R., Fuqua C., Brinkac L., Lewis M., Johri S., Weaver B., Pai G., Eisen J. A., Rahe E., Sheldon W. M., Ye W., Miller T. R., Carlton J., Rasko D. A., Paulsen I. T., Ren Q., Daugherty S. C., Deboy R. T., Dodson R. J., Durkin A. S., Madupu R., Nelson W. C., Sullivan S. A., Rosovitz M. J., Haft D. H., Selengut J., and Ward N. (2004). Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. *Nature*, 432(7019):910–913.

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

Morris R. M., Longnecker K., and Giovannoni S. J. (2006). *Pirellula* and OM43 are among the dominant lineages identified in an Oregon coast diatom bloom. *Environmental Microbiology*, 8(8):1361–1370.

Murray A. E. and Grzymski J. J. (2007). Diversity and genomics of Antarctic marine micro-organisms. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1488):2259–2271.

Murray A. E., Wu K. Y., Moyer C. L., Karl D. M., and DeLong E. F. (1999). Evidence for circumpolar distribution of planktonic Archaea in the Southern Ocean. *Aquatic Microbial Ecology*, 18(3):263–273.

Murray A. E. A., Preston C. M. C., Massana R. R., Taylor L. T. L., Blakis A. A., Wu K. K., and DeLong E. F. (1998). Seasonal and spatial variability of bacterial and archaeal assemblages in the coastal waters near Anvers Island, Antarctica. *Applied and Environmental Microbiology*, 64(7):2585–2595.

Murray A. E., Peng V., Tyler C., and Wagh P. (2011). Marine bacterioplankton biomass, activity and community structure in the vicinity of Antarctic icebergs. *Deep Sea Research Part II: Topical Studies in Oceanography*, 58(11-12):1407–1421.

Newton I. L. G., Woyke T., Auchtung T. A., Dilly G. F., Dutton R. J., Fisher M. C., Fontanez K. M., Lau E., Stewart F. J., Richardson P. M., Barry K. W., Saunders E., Detter J. C., Wu D., Eisen J. A., and Cavanaugh C. M. (2007). The *Calyptogena magnifica* Chemoautotrophic Symbiont Genome. *Science*, 315(5814):998–1000.

Ng C., DeMaere M. Z., Williams T. J., Lauro F. M., Raftery M., Gibson J. A., Andrews-Pfannkoch C., Lewis M., Hoffman J. M., Thomas T., and Cavicchioli R. (2010). Metaproteogenomic analysis of a dominant green sulfur bacterium from Ace Lake, Antarctica. *The ISME Journal*, 4(8):1002–1019.

Nikrad M. P., Cottrell M. T., and Kirchman D. L. (2012). Abundance and Single-Cell Activity of Heterotrophic Bacterial Groups in the Western Arctic Ocean in Summer and Winter. *Applied and Environmental Microbiology*, 78(7):2402–2409.

Obernosterer I., Catala P., Lebaron P., and West N. J. (2011). Distinct bacterial groups contribute to carbon cycling during a naturally iron fertilized phytoplankton bloom in the Southern Ocean. *Limnology and Oceanography*, 56(6):2391–2401.

Oh H. M., Kwon K. K., Kang I., Kang S. G., Lee J. H., Kim S. J., and Cho J. C. (2010). Complete Genome Sequence of "*Candidatus* Puniceispirillum marinum" IMCC1322, a Representative of the SAR116 Clade in the *Alphaproteobacteria*. *Journal of Bacteriology*, 192(12):3240–3241.

Oliver J. L., Barber R. T., Smith Jr W. O., and Ducklow H. W. (2004). The heterotrophic bacterial response during the Southern Ocean iron experiment (SOFeX). *Limnology and Oceanography*, 49(6): 2129–2140.

Orsi A. H., Whitworth T., and Nowlin W. D. (1995). On the meridional extent and fronts of the Antarctic Circumpolar Current. *Deep Sea Research Part I: Oceanographic Research Papers*, 42(5):641–673.

Orsi A. H., Johnson G. C., and Bullister J. L. (1999). Circulation, mixing, and production of Antarctic Bottom Water. *Progress in Oceanography*, 43(1):55–109.

O'Sullivan L. A., Fuller K. E., Thomas E. M., Turley C. M., Fry J. C., and Weightman A. J. (2004). Distribution and culturability of the uncultivated 'AGG58 cluster' of the Bacteroidetes phylum in aquatic environments. *FEMS Microbiology Ecology*, 47(3):359–370.

Partensky F., Hess W. R., and Vaulot D. (1999). *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiology and Molecular Biology Reviews*, 63(1):106–127.

Paul J. H., DeFlaun M. F., and Jeffrey W. H. (1988). Mechanisms of DNA utilization by estuarine microbial populations. *Applied and Environmental Microbiology*, 54(7):1682–1688.

Pham V. D., Konstantinidis K. T., Palden T., and DeLong E. F. (2008). Phylogenetic analyses of ribosomal DNA-containing bacterioplankton genome fragments from a 4000 m vertical profile in the North Pacific Subtropical Gyre. *Environmental Microbiology*, 10(9):2313–2330.

Pinhassi J., Sala M. M., Havskum H., Peters F., Guadayol Ò., Malits A., and Marrasé C. (2004). Changes in bacterioplankton composition under different phytoplankton regimens. *Applied and Environmental Microbiology*, 70(11):6753–6766.

Piquet A. M. T., Bolhuis H., Meredith M. P., and Buma A. G. J. (2011). Shifts in coastal Antarctic marine microbial communities during and after melt water-related surface stratification. *FEMS Microbiology Ecology*, 76(3):413–427.

Pollard R. T., Lucas M. I., and Read J. F. (2002). Physical controls on biogeochemical zonation in the Southern Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography*, 49(16):3289–3305.

Pommier T., Canbäck B., Riemann L., Boström K. H., Simu K., Lundberg P., Tunlid A., and Hagström Å. (2007). Global patterns of diversity and community structure in marine bacterioplankton. *Molecular Ecology*, 16(4):867–880.

Poorvin L., Rinta-Kanto J. M., Hutchins D. A., and Wilhelm S. W. (2004). Viral release of iron and its bioavailability to marine plankton. *Limnology and Oceanography*, 49(5):1734–1741.

Powell L. M., Bowman J. P., Skerratt J. H., Franzmann P. D., and Burton H. R. (2005). Ecology of a novel *Synechococcus* clade occurring in dense populations in saline Antarctic lakes. *Marine Ecology Progress Series*, 291(28 April):65–80.

Preston C. M., Wu K. Y., Molinski T. F., and DeLong E. F. (1996). A psychrophilic crenarchaeon inhabits a marine sponge: *Cenarchaeum symbiosum* gen. nov., sp. nov. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 93(13):6241–6246.

Qin J., Li R., Raes J., Arumugam M., Burgdorf K. S., Manichanh C., Nielsen T., Pons N., Levenez F., Yamada T., Mende D. R., Li J., Xu J., Li S., Li D., Cao J., Wang B., Liang H., Zheng H., Xie Y., Tap J., Lepage P., Bertalan M., Batto J.-M., Hansen T., Le Paslier D., Linneberg A., Nielsen H. B., Pelletier E., Renault P., Sicheritz-Ponten T., Turner K., Zhu H., Yu C., Li S., Jian M., Zhou Y., Li Y., Zhang X.,

Li S., Qin N., Yang H., Wang J., Brunak S., Doré J., Guarner F., Kristiansen K., Pedersen O., Parkhill J., Weissenbach J., MetaHIT Consortium , Bork P., Ehrlich S. D., and Wang J. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464(7285):59–65.

Rappé M. S., Connon S. A., Vergin K. L., and Giovannoni S. J. (2002). Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature*, 418(6898):630–633.

Rath J., Wu K. Y., Herndl G. J., and DeLong E. F. (1998). High phylogenetic diversity in a marine-snow-associated bacterial assemblage. *Aquatic Microbial Ecology*, 14(3):261–269.

Reisch C. R., Stoudemayer M. J., Varaljay V. A., Amster I. J., Moran M. A., and Whitman W. B. (2011). Novel pathway for assimilation of dimethylsulphoniopropionate widespread in marine bacteria. *Nature*, 473(7346):208–211.

Rusch D. B., Halpern A. L., Sutton G., Heidelberg K. B., Williamson S., Yooseph S., Wu D., Eisen J. A., Hoffman J. M., Remington K., Beeson K., Tran B., Smith H., Baden-Tillson H., Stewart C., Thorpe J., Freeman J., Andrews-Pfannkoch C., Venter J. E., Li K., Kravitz S., Heidelberg J. F., Utterback T., Rogers Y.-H., Falcón L. I., Souza V., Bonilla-Rosso G., Eguiarte L. E., Karl D. M., Sathyendranath S., Platt T., Bermingham E., Gallardo V., Tamayo-Castillo G., Ferrari M. R., Strausberg R. L., Nealson K., Friedman R., Frazier M., and Venter J. C. (2007). The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biology*, 5(3):e77–e77.

Sabine C. L., Feely R. A., Gruber N., Key R. M., Lee K., Bullister J. L., Wanninkhof R., Wong C. S., Wallace D. W. R., Tilbrook B., Millero F. J., Peng T.-H., Kozyr A., Ono T., and Rios A. F. (2004). The Oceanic Sink for Anthropogenic $CO_2$. *Science*, 305(5682):367–371.

Scanlan D. J., Ostrowski M., Mazard S., Dufresne A., Garczarek L., Hess W. R., Post A. F., Hagemann M., Paulsen I., and Partensky F. (2009). Ecological Genomics of Marine Picocyanobacteria. *Microbiology and Molecular Biology Reviews*, 73(2):249–299.

Selje N. N., Simon M. M., and Brinkhoff T. T. (2004). A newly discovered *Roseobacter* cluster in temperate and polar oceans. *Nature*, 427(6973):445–448.

Short C. M. and Suttle C. A. (2005). Nearly Identical Bacteriophage Structural Gene Sequences Are Widely Distributed in both Marine and Freshwater Environments. *Applied and Environmental Microbiology*, 71(1):480–486.

Short S. M. and Suttle C. A. (2002). Sequence Analysis of Marine Virus Communities Reveals that Groups of Related Algal Viruses Are Widely Distributed in Nature. *Applied and Environmental Microbiology*, 68(3):1290–1296.

Simon M., Glöckner F. O., and Amann R. (1999). Different community structure and temperature optima of heterotrophic picoplankton in various regions of the Southern Ocean. *Aquatic Microbial Ecology*, 18(3):275–284.

Sinha V., Williams J., Meyerhöfer M., Riebesell U., Paulino A. I., and Larsen A. (2007). Air-sea fluxes of methanol, acetone, acetaldehyde, isoprene and DMS from a Norwegian fjord following a phytoplankton bloom in a mesocosm experiment. *Atmospheric Chemistry and Physics*, 7(3):739–755.

Sokolov S. and Rintoul S. R. (2002). Structure of Southern Ocean fronts at 140°E. *Journal of Marine Systems*, 37(1):151–184.

Sokolov S. and Rintoul S. R. (2009). Circumpolar structure and distribution of the Antarctic Circumpolar Current fronts: 1. Mean circumpolar paths. *Journal of Geophysical Research*, 114(C11):C11018.

Sowell S. M., Wilhelm L. J., Norbeck A. D., Lipton M. S., Nicora C. D., Barofsky D. F., Carlson C. A., Smith R. D., and Giovanonni S. J. (2009). Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. *The ISME Journal*, 3(1):93–105.

Speer K., Rintoul S. R., and Sloyan B. (2000). The Diabatic Deacon Cell. *Journal of physical oceanography*, 30(12):3212–3222.

Steindler L., Schwalbach M. S., Smith D. P., Chan F., and Giovannoni S. J. (2011). Energy Starved *Candidatus* Pelagibacter Ubique Substitutes Light-Mediated ATP Production for Endogenous Carbon Respiration. *PLoS ONE*, 6(5):e19725.

Stingl U., Tripp H. J., and Giovannoni S. J. (2007). Improvements of high-throughput culturing yielded novel SAR11 strains and other abundant marine bacteria from the Oregon coast and the Bermuda Atlantic Time Series study site. *The ISME Journal*, 1:361–371.

Straza T. R. A., Ducklow H. W., Murray A. E., and Kirchman D. L. (2010). Abundance and single-cell activity of bacterial groups in Antarctic coastal waters. *Limnology and Oceanography*, 55(6):2526–2536.

Strous M., Fuerst J. A., Kramer E. H. M., Logemann S., Muyzer G., Van De Pas-Schoonen K. T., Webb R., Kuenen J. G., and Jetten M. S. M. (1999). Missing lithotroph identified as new planctomycete. *Nature*, 400(6743):446–449.

Strutton P. G., Griffiths F. B., Waters R. L., Wright S. W., and Bindoff N. L. (2000). Primary productivity off the coast of East Antarctica (80- 150°E): January to March 1996. *Deep Sea Research Part II: Topical Studies in Oceanography*, 47:2327–2362.

Swan B. K., Martinez-Garcia M., Preston C. M., Sczyrba A., Woyke T., Lamy D., Reinthaler T., Poulton N. J., Masland E. D. P., Gomez M. L., Sieracki M. E., DeLong E. F., Herndl G. J., and Stepanauskas R. (2011). Potential for Chemolithoautotrophy Among Ubiquitous Bacteria Lineages in the Dark Ocean. *Science*, 333(6047):1296–1300.

Swingley W. D., Sadekar S., Mastrian S. D., Matthies H. J., Hao J., Ramos H., Acharya C. R., Conrad A. L., Taylor H. L., Dejesa L. C., Shah M. K., O'Huallachain M. E., Lince M. T., Blankenship R. E., Beatty J. T., and Touchman J. W. (2007). The Complete Genome Sequence of *Roseobacter denitrificans* Reveals a Mixotrophic Rather than Photosynthetic Metabolism. *Journal of Bacteriology*, 189(3):683–690.

Teske A., Alm E., Regan J. M., Toze S., Rittmann B. E., and Stahl D. A. (1994). Evolutionary relationships among ammonia- and nitrite-oxidizing bacteria. *Journal of Bacteriology*, 176(21):6623–6630.

Thomalla S. J., Waldron H. N., Lucas M. I., Read J. F., Ansorge I. J., and Pakhomov E. (2011). Phytoplankton distribution and nitrogen dynamics in the southwest indian subtropical gyre and Southern Ocean waters. *Ocean Science*, 7(1):113–127.

Thompson D. W. J. and Solomon S. (2002). Interpretation of Recent Southern Hemisphere Climate Change. *Science*, 296(5569):895–899.

Topping J. N., Heywood J. L., Ward P., and Zubkov M. V. (2006). Bacterioplankton composition in the Scotia Sea, Antarctica, during the austral summer of 2003. *Aquatic Microbial Ecology*, 45(3):229–235.

Tripp H. J., Kitner J. B., Schwalbach M. S., Dacey J. W. H., Wilhelm L. J., and Giovannoni S. J. (2008). SAR11 marine bacteria require exogenous reduced sulphur for growth. *Nature*, 452(7188):741–744.

Trull T., Rintoul S. R., Hadfield M., and Abraham E. R. (2001). Circulation and seasonal evolution of polar waters south of Australia: implications for iron fertilization of the Southern Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography*, 48(11):2439–2466.

Venter J. C., Remington K., Heidelberg J. F., Halpern A. L., Rusch D., Eisen J. A., Wu D., Paulsen I., Nelson K. E., Nelson W., Fouts D. E., Levy S., Knap A. H., Lomas M. W., Nealson K., White O., Peterson J., Hoffman J., Parsons R., Baden-Tillson H., Pfannkoch C., Rogers Y.-H., and Smith H. O. (2004). Environmental Genome Shotgun Sequencing of the Sargasso Sea. *Science*, 304(5667):66–74.

Vila-Costa M., Simó R., Harada H., Gasol J. M., Slezak D., and Kiene R. P. (2006). Dimethylsulfoniopropionate Uptake by Marine Phytoplankton. *Science*, 314(5799):652–654.

Wagner-Döbler I. and Biebl H. (2006). Environmental Biology of the Marine *Roseobacter* Lineage. *Annual Review of Microbiology*, 60(1):255–280.

Walker C. B., Torre J. R.de la, Klotz M. G., Urakawa H., Pinel N., Arp D. J., Brochier-Armanet C., Chain P., Chan P. P., Gollabgir A., Hemp J., Hügler M., Karr E. A., Könekke M., Shin M., Lawton T. J., Lowe T., Martens-Habbena W., Sayavedra-Soto L. A., Langf D., Sievert S. M., Rosenzweig A. C., Manning G., and Stahl D. A. (2010). *Nitrosopumilus maritimus* genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 107(19):8818–8823.

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

Ward P., Whitehouse M., Brandon M., Shreeve R., and Woodd-Walker R. (2003). Mesozooplankton community structure across the Antarctic Circumpolar Current to the north of South Georgia: Southern Ocean. *Marine Biology*, 143(1):121–130.

Weber T. S. and Deutsch C. (2010). Ocean nutrient ratios governed by plankton biogeography. *Nature*, 467(7315):550–554.

Weinbauer M. G., Arrieta J. M., Griebler C., and Herndl G. J. (2009). Enhanced viral production and infection of bacterioplankton during an iron-induced phytoplankton bloom in the Southern Ocean. *Limnol. Oceanogr*, 54(3):774–784.

West N. J., Obernosterer I., Zemb O., and Lebaron P. (2008). Major differences of bacterial diversity and activity inside and outside of a natural iron-fertilized phytoplankton bloom in the Southern Ocean. *Environmental Microbiology*, 10(3):738–756.

Whitworth T. (1980). Zonation and geostrophic flow of the Antarctic Circumpolar Current at Drake Passage. *Deep Sea Research Part I: Oceanographic Research Papers*, 27(7):497–507.

Whitworth III T. and Nowlin Jr. W. D. (1987). Water masses and currents of the Southern Ocean at the Greenwich Meridian. *Journal of Geophysical Research*, 92(C6):6462–6476.

Wilhelm S. W. and Suttle C. A. (1999). Viruses and nutrient cycles in the sea. *BioScience*, 49(10):781–788.

Wilkins D., Lauro F. M., Williams T. J., DeMaere M. Z., Brown M. V., Hoffman J. M., Andrews-Pfannkoch C., McQuaid J. B., Riddle M. J., Rintoul S. R., and Cavicchioli R. (2012). Biogeographic partitioning of Southern Ocean picoplankton revealed by metagenomics. *Molecular Ecology*.

Wilkins D., Yau S., Williams T. J., Allen M. A., Brown M. V., DeMaere M. Z., Lauro F. M., and Cavicchioli R. (2012). Key microbial drivers in Antarctic aquatic environments. *FEMS microbiology reviews*, pages n/a–n/a.

Williams G. D., Nicol S., Aoki S., Meijers A. J. S., Bindoff N. L., Iijima Y., Marsland S. J., and Klocker A. (2010). Surface oceanography of BROKE-West, along the Antarctic margin of the south-west Indian Ocean (30–80°E). *Deep Sea Research Part II: Topical Studies in Oceanography*, 57(9-10):738–757.

Williams T. J., Lauro F. M., Ertan H., Burg D. W., Poljak A., Raftery M. J., and Cavicchioli R. (2011). Defining the response of a microorganism to temperatures that span its complete growth temperature range (-2 °C to 28 °C) using multiplex quantitative proteomics. *Environmental Microbiology*, 13 (8):2186–2203.

Williams T. J., Long E., Evans F., DeMaere M. Z., Lauro F. M., Raftery M. J., Ducklow H., Grzymski J. J., Murray A. E., and Cavicchioli R. (2012). A metaproteomic assessment of winter and summer bacterioplankton from Antarctic Peninsula coastal surface waters. *The ISME Journal*, 6(10):1883–1900.

Wright T. D., Vergin K. L., Boyd P. W., and Giovannoni S. J. (1997). A novel âĹĆ-subdivision proteobacterial lineage from the lower ocean surface layer. *Applied and Environmental Microbiology*, 63 (4):1441–1448.

Ye Y. and Doak T. G. (2009). A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. *PLoS Computational Biology*, 5(8):e1000465.

Yoon J., Yasumoto-Hirose M., Katsuta A., Sekiguchi H., Matsuda S., Kasai H., and Yokota A. (2007). Coraliomargarita akajimensis gen. nov., sp. nov., a novel member of the phylum 'Verrucomicrobia' isolated from seawater in Japan. *International Journal of Systematic and Evolutionary Microbiology*, 57 (5):959–963.

Zaballos M., López-López A., Ovreas L., Bartual S. G., D'Auria G., Alba J. C., Legault B., Pushker R., Daae F. L., and Rodriguez-Valera F. (2006). Comparison of prokaryotic diversity at offshore oceanic locations reveals a different microbiota in the Mediterranean Sea. *FEMS Microbiology Ecology*, 56(3): 389–405.

Zhang R., Liu B., Lau S. C. K., Ki J.-S., and Qian P.-Y. (2007). Particle-attached and free-living bacterial communities in a contrasting marine environment: Victoria Harbor, Hong Kong. *FEMS Microbiology Ecology*, 61(3):496–508.

Zubkov M. V., Sleigh M. A., Tarran G. A., Burkill P. H., and Leakey R. J. G. (1998). Picoplanktonic community structure on an Atlantic transect from 50°N to 50°S. *Deep Sea Research Part I: Oceanographic Research Papers*, 45(8):1339–1355.

# Appendix A

# MINSPEC source code

```
#!/usr/bin/perl

#minspec: determines minimal set of species needed to explain
# species assignments from a dataset of metagenomic reads,
# eliminating spurious species assignments.

#based on the approach of, and borrows heavily from, MinPath:
# Ye, Y. and Doak, T.G.. A parsimony approach to biological pathway
# reconstruction/inference for genomes and metagenomes. PLoS
# Computational Biology 2009, vol 5 num 8
#http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1000465

#written by David Wilkins <david@wilkox.org>
#minspec lives at https://github.com/wilkox/minspec

#this software is released into the public domain. To the extent
# possible under law, all copyright and related or neighboring
# rights are waived and permission is explicitly and irrevocably
# granted to copy, modify, adapt, sell and distribute this software
# in any way you choose.

$USAGE = q/USAGE:

minspec -b <blast hittable>

OPTIONAL:

-g  <filename> Produce a BLAST hit table containing only hits to species present in the
    minimal set, suitable for processing with GAAS. Default filename is <blast hittable>.
    filtered
-l  <filename> Produce a list of species, indicating whether or not they are present in the
    minimal set ('1' = present, '0' = not present). Default filename is <blast hittable>.
    minimal.list
-max  <maximum read count> Set a maximum number of reads with identity to a species, less than
     which the species may still be parsimoniously eliminated but equal to or more than which
    the species will be marked as present regardless of its presence in the minimal set. Set
    to 50 if -m flag is provided without a value.
/;

use Getopt::Long;

GetOptions (
'b=s' => \$blastoutputfile,
'g:s' => \$makegaas,
'l:s' => \$makelist,
'max:s' => \$maxthreshold,
) or die("$USAGE");

#check for required arguments and set defaults

die ("$USAGE\n") if !defined $blastoutputfile;
```

```
45 die ("ERROR - you did not specify any output format, so there's no use in running the script!
       Try using -g or -l\n") unless (defined $makegaas || defined $makelist);
46 $maxthreshold = 50  if (!$maxthreshold && defined $maxthreshold);
47 $makegaas = "$blastoutputfile.filtered" if (!$makegaas && defined $makegaas);
48 $makelist = "$blastoutputfile.minimal.list" if (!$makelist && defined $makelist);
49
50 ##BODY
51 &doLP;
52 &makegaas if defined $makegaas;
53 &makelist if defined $makelist;
54 exit;
55 ##END BODY
56
57 #SUBS
58
59 sub doLP { #set up and run the lp
60
61   #read in blast output
62   die ("ERROR - could not open blast output file $blastoutputfile\n") unless open(BLAST, "<
         $blastoutputfile");
63
64   #set unique ids
65   #the MPS file format has some very strict requirements -
66   # replacing read and OTU names with these IDs ensures these
67   # are never inadvertantly broken
68   $readuid = "AAAAAAA";
69   $specuid = "LLLLLL";
70
71   #get read-species mappings from the BLAST output
72   while ($line = <BLAST>) {
73     chomp $line;
74     unless ($line =~ /^(\S+)\s+(\S+)/) {
75       print ("ERROR - malformed line in blast output file $blastoutputfile, line $.\n");
76       next;
77     }
78     $read = $1;
79     $species = $2;
80
81     #assign unique IDs to read and OTU if
82     # they don't already have them
83     if (!exists ($readuidof{$read})) {
84       $readuidtrans{$readuid} = $read;
85       $readuidof{$read} = $readuid;
86       }
87     unless (exists ($specuidof{$species})) {
88       $specuidtrans{$specuid} = $species;
89       $specuidof{$species} = $specuid;
90       }
91
92     #store read-species mappings
93     push(@{$allreads{$readuidof{$read}}}, $specuidof{$species});
94     push(@{$allspecies{$specuidof{$species}}}, $readuidof{$read});
95     ++$readuid;
96     ++$specuid;
97
98     }
99   close BLAST;
100
101  #produce MPS file
102  die unless open(MPS, ">test.mps");
103  print MPS q/NAME          PATH
104 ROWS
105  N  NUM/;
106
107  #first, list all the reads
108  foreach $read (keys(%allreads)) {
109    print MPS "\n G  $read";
110    }
111
112  #next, list all species with their read mappings
113  print MPS "\nCOLUMNS";
```

```
114   foreach $species (keys(%allspecies)) {
115     $speciesspacerlength = 10 - length($species);
116     print MPS "\n    $species" . " "x$speciesspacerlength . "NUM                1";
117     undef %preventduplicatereads;
118     foreach $read (@{$allspecies{$species}}) {
119       next if exists($preventduplicatereads{$read});
120       $readspacerlength = 19 - length($read);
121       print MPS "\n    $species" . " "x$speciesspacerlength . "$read" . " "x$readspacerlength
              . "1";
122       $preventduplicatereads{$read} = "";
123       }
124     }
125
126   #list all reads (constraint function)
127   print MPS "\nRHS";
128   foreach $read (keys(%allreads)) {
129     $readspacerlength = 17 - length($read);
130     print MPS "\n    RHS1      $read" . " "x$readspacerlength . "1.0";
131     }
132
133   #list all species (not sure why this is needed by GLPSOL complains
134   # without it)
135   print MPS "\nBOUNDS";
136   foreach $species (keys(%allspecies)) {
137     print MPS "\n BV BND1      $species    ";
138     }
139
140   #and done
141   print MPS "\nENDATA\n";
142   close MPS;
143
144   #run the LP
145   die ("ERROR - LP execution with glpsol failed (command: glpsol test.mps -o test.mps.LPout)\n
          ") unless system("glpsol test.mps -o test.mps.LPout") == 0;
146
147   #parse the LP output
148   die ("ERROR - could not open the LP output test.mps.LPout\n") unless open(LP, "<test.mps.
          LPout");
149   while ($line = <LP>) {
150     $reading = 1 if $line =~ /Column name/;
151     next unless $reading == 1;
152     next if $line =~ /Column name/;
153     chomp $line;
154     next if $line =~ /^-/;
155     last if $line eq "";
156     @line = split(/\s+/, $line);
157     $species = @line[2];
158     $speciesname = $specuidtrans{$species};
159     $presence{$speciesname} = @line[4]; #set to 1 if the species is present in the minimal set
              , 0 if not
160
161     if (defined $maxthreshold && $presence{$2} == 0 && @{$allspecies{$specuidof{$speciesname
              }}} >= $maxthreshold) { #if a max threshold is set AND the species presence is set to
              zero AND the count of reads hitting to that species is greater than the max threshold
              ...
162       $presence{$speciesname} = 1; #...set species to present
163       }
164
165     }
166
167   #clean up
168   close LP;
169   system("rm test.mps test.mps.LPout");
170
171 } #end of doLP sub
172
173 sub makegaas { #makes an blast hittable for gaas, identical to the input but with non-
          parsimonious species removed
174
175   die ("ERROR - could not open blast output file $blastoutputfile\n") unless open(BLAST, "<
          $blastoutputfile");
```

```
176   die ("ERROR - could not create filtered blast hittable for gaas at $makegaas\n") unless open
          (GAAS, ">$makegaas");
177
178   while ($line = <BLAST>) {
179     chomp $line;
180     die ("ERROR - malformed blast hittable line at line $. of $blastoutputfile\n") unless
          $line =~ /^(\S+)\s+(\S+)/;
181     print GAAS "$line\n" if $presence{$2} == 1; #print the line if this species is present in
          the minimal set
182     }
183
184   close BLAST;
185   close GAAS;
186
187 } #end of makegaas sub
188
189 sub makelist { #makes a simple list of species, indicating whether or not they are present in
      the minimal set
190
191   die ("ERROR - could not create list of species at $makelist\n") unless open(LIST, ">
        $makelist");
192
193   foreach $species (sort {$a <=> $b} (keys(%presence))) {
194     print LIST "$species\t$presence{$species}\n";
195     }
196
197   close LIST;
198
199 } #end of makelist sub
```