

Microbial Ecology and Biogeography OF THE Southern Ocean

David Wilkins

November 5, 2012

Contents

List of Figures	iii
List of Tables	v
List of Acronyms	viii
Acknowledgements	ix
Abstract	xi
Introduction	1
Microbial ecology of the Southern Ocean	1
Bacteria	1
Alphaproteobacteria	1
Roseobacter	1
Oceanography of the Southern Ocean	10
Water masses and fronts	10
Effect of climate change	10
Role of the Polar Front in biogeography	10
Project questions and hypotheses	10
The Polar Front as a major biogeographic boundary in the Southern Ocean	11
Summary	11
Introduction	11
Methods	13
Sampling and metagenomic sequencing	13
Phylogenetic analysis of metagenomic data	16
BLAST comparison to RefSeq database	16

Identification of minimal species sets with MINSPEC	16
Operational Taxonomic Unit (OTU) abundances and variance between zones	17
Functional analysis of metagenomic data	18
BLAST comparison to Kyoto Encyclopedia of Genes and Genomes (KEGG) database	18
Analysis of functional potential	18
Results	19
Metagenomic sequencing	19
Phylogenetic analysis of metagenomic data	19
Validation of MINSPEC	24
Functional analysis of metagenomic data	26
Discussion	29
Taxonomic groups differentiating the zones	29
GSO-EOSA-1	29
Ammonia-oxidizing Crenarchaeota	31
Cyanobacteria	31
SAR11 and SAR116 clades	32
Bacteroidetes	33
Rhodobacterales	33
Alteromonadales	34
Verrucomicrobia	34
Functional capacities differentiating the zones	36
Conclusions: Biogeographic role of the Polar Front	38
Meso-scale biogeographic drivers of planktonic diversity	41
Conclusions	43

List of Figures

1	Map showing sites of seawater samples used in the Polar Front study	14
2	Rank-abundance curves for OTUs in each zone and size fraction	21
3	Contribution of OTUs to variance between the North and South zones	23
4	Results of MINSPEC validate	25
5	Tree of GSO-EOSA-1 related 16S rRNA genes	30
6	Taxonomic decomposition of KEGG modules	35

List of Tables

1	Details of samples used in Polar Front study	15
2	Twenty most abundant OTUs	20
3	Highest-contributing OTUs to the difference between the North and South zones	22
4	Contributions of KEGG modules to variance between the North and South zones	27
5	Contributions of KEGG ortholog groups to variance between the North and South zones	28

Acronyms

AAP Aerobic Anoxygenic Phototrophic.

ACC Antarctic Circumpolar Current.

ANOSIM Analysis of SIMilarities.

AOA Ammonia Oxidizing Archaea.

AZ Antarctic Zone.

CASO Climate of Antarctica and the Southern Ocean.

CDW Circumpolar Deep Water.

CEAMARC Collaborative East Antarctic Marine Census.

CTD Conductivity, Temperature and Depth.

DFAA Dissolved Free Amino Acids.

DMSP dimethylsulfoniopropionate.

DOC Dissolved Organic Carbon.

DOM Dissolved Organic Matter.

GOS Global Ocean Sampling.

HMW High Molecular Weight.

HNLC High Nutrient, Low Chlorophyll.

IPY International Polar Year.

KEGG Kyoto Encyclopedia of Genes and Genomes.

MGI Marine Group I Crenarchaeota.

MMPA methylmercaptopropionate.

NZ North Zone.

OTU Operational Taxonomic Unit.

PF Polar Front.

PFZ Polar Frontal Zone.

POM Particulate Organic Matter.

SACCF Southern Antarctic Circumpolar Current Front.

SAF Subantarctic Front.

SAZ Subantarctic Zone.

SIMPER SIMilarity PERcentages.

SO Southern Ocean.

STF Subtropical Front.

SZ South Zone.

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.

Microbial ecology of the Southern Ocean

TODO working on this

TODO begin pasted block

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; BÄLJÄÄ *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 Southern Ocean surface waters and to a depth of at least 2200 m, composing 10-30

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 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 phylogenotypes were at maximum 8

A metagenomic study of Southern Ocean waters off West Antarctica found that Roseobacter clade SSU rRNA sequences were much more abundant in summer than in winter, with Sulfitobacter sequences the most abundant within this clade (Murray et al., 2012). This is consistent with the association of Roseobacters with phytoplankton (Moran et al., 2005). 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).

The SAR11 clade of alphaproteobacteria is probably the most abundant class of marine microorganisms worldwide (Morris et al. 2002). *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 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 Southern Ocean, in all open ocean regions as well as at depth and in coastal waters, and is usually the dominant alpha-proteobacterial, if not bacterial, group (Giebel et al. 2009; Murray and Grzymski 2007; Láspez-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. 2010; Piquet et al. 2011). It is probably more abundant in the epipelagic than at depth (Giebel et al. 2009; Wilkins et al. ???).

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. ???). It is probably more abundant in the Subantarctic and polar frontal zones than in the Antarctic Zone (Giebel et al. 2009; Wilkins et al. ???; Ghiglione and Murray 2011). This may be related to a competitive advantage of the oligotrophic SAR11 in the HNLC Subantarctic relative to the AZ, where blooming phytoplankton lead to increased concentrations of HMW DOM and 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 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, Ghiglioni and Murray (2011) found only small seasonal changes in abundance during an annual cycle at the AP and Kerguelen Island. These studies are all consistent with the view of SAR11 as a typically non-opportunistic oligotroph specialising in LMW 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 (BÄl'jÄä 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).

The SAR116 clade of alphaproteobacteria have been detected throughout the world ocean. SAR116 has been detected in molecular studies of the SO, in both the SAZ (West et al. 2008; Topping et al. 2006) and at lower abundance in the AZ (Wilkins et al. ????). Estimates of its relative abundance vary depending on the method used; Topping et al. (2006) using FISH estimated it composed 13.1 +- 8.6

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 dimethylsulfonio-propionate 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).

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 ammonium 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. A metagenomic survey also detected *Nitrosomonas europaea*, *Nitrosomonas eutropha* and *Nitrosospira multiformis* strains in all SO surface waters, and at significantly higher abundances in Antarctic Bottom Water (Wilkins et al. ????). However, ammonia-oxidizing archaea outnumbered ammonia-oxidizing bacteria at most sites, consistent with the view that the former are the major nitrifiers in the marine environment (Wuchter et al. 2006). 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.

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 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

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. 2005). 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 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 be 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).

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 Southern Ocean. In waters off the western Antarctic Peninsula, Ant4D3 was reported to compose 10

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

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) estimated from rRNA abundances that between 18 and 37

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. 2010; Ducklow et al. 2011; Jamieson et al. 2012; Wilkins et al. ????). 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, LÃşpez-GarcÃa 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).

Bacteria of the group Cytophaga-Flavobacterium-*Bacteroides* (CFB) are cosmopolitan and abundant in the world ocean (GlÃückner 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 (Pinhasi 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 remineralization of primary production products. Of the CFB, the class Flavobacteria seem to be in the majority worldwide in both freshwater and ma-

rine environments (O'Sullivan et al. 2004; Cottrell et al. 2005) including the Southern Ocean (Abell and Bowman 2005a).

CFB in the SO are strongly biogeographically partitioned. Abell and Bowman (2005a), 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. Likewise, a large-scale metagenomic analysis of the SO which identified the PF as a major biogeographic boundary found CFB contribute a large fraction of the variance between the zones, with higher abundance south of the PF (Wilkins et al. ????). 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 2005a; Abell and Bowman 2005b; 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 (Abell and Bowman 2005b). 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) (Gonzalez et al., 2008). An Antarctic Peninsula coastal metagenome found Polaribacter-related sequences to be dominant in summer, consistent 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, 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 Southern Ocean 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 metagenomic (Wilkins et al. ?????) and metaproteomic (Williams et al. 2012) results 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 2005a; Wilkins et al. ????; Topping 2006).

The 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. 2010; West et al. 2008; Gentile et al. 2006; Murray and Grzymski 2007). 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. However, a metagenomic survey found higher levels of an OTU with best match to *verrucomicrobium Coraliomargarita akajimensis* in AZ waters than SAZ (Wilkins et al. ????).

Bacteria of the phylum Planctomycetes have been detected at low abundance in SO molecular surveys (Gentile et al. 2006; LÁspez-GarcÃa et al. 2001; Jamieson et al. 2012; Murray et al. 2010; 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 (Williams et al., 2012). 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 Southern Ocean waters, including Actinobacteria (Bowman et al. 2003; Brinkmeyer et al. 2003; Abell and Bowman 2005; Gentile et al. 2006; Murray and Grzynski 2007; Murray et al. 2010; Ghiglione and Murray 2011; Bolhuis et al. 2011; Jamieson et al. 2012), Epsilonproteobacteria (Gentile et al. 2006; Murray and Grzynski 2007), and Firmicutes (Murray and Grzynski 2007; Guidice et al. 2011; Murray et al. 2010). 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.

DeLong et al. (1994) first reported the high abundance (up to 34

For MGI, ammonium-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 Southern Ocean. In a winter coastal Antarctic metaproteome, MGI proteins made up 30

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

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. 1998; Massana et al. 2000), although not in aphotic waters; LÁspez-GarcÁa et al. (2001) detected only euryarchaeotal 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. LÁspez-GarcÁa 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.

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

ZZZ))))

TODO end pasted block

Oceanography of the Southern Ocean

Water masses and fronts

Effect of climate change

Role of the Polar Front in biogeography

Project questions and hypotheses

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

TODO working on this section

TODO begin pasted block

The Southern Ocean (SO) plays a critical role in sustaining marine life around the globe. Upwelling of nutrient-rich Circumpolar Deep Water (CDW) returns nutrients transported to the deep ocean by the sinking of organic matter (Rath *et al.*, 1998) and supports 75% of global ocean primary production north of 30°S. Surface waters at high southern latitudes remain cold (< 3°C) year-round but undergo extreme seasonal variations in sea-ice cover, light levels and day length. Primary production and biomass are high in summer and very low in winter. Bacteria are abundant in the SO despite the low temperatures and seasonal variability in productivity and are a major route for carbon flow (Hessen *et al.*, 2004).

The SO is composed of several zones separated by circumpolar fronts, the locations of which vary temporally and with longitude (Whitworth, 1980; Orsi *et al.*, 1995; Sokolov and Rintoul, 2002). The

fronts separate regions with different physiochemical properties, such as density, salinity, temperature and nutrient concentrations (Sokolov and Rintoul, 2002). Hydrographic, bathythermographic and satellite altimetry data have been used to determine the frontal structure of the Antarctic Circumpolar Current (ACC) south of Australia (Sokolov and Rintoul, 2002). From north to south, the major fronts are the Subtropical Front (STF), the Subantarctic Front (SAF), the Polar Front (PF) and the Southern Antarctic Circumpolar Current Front (SACCF). Each of these fronts consists of multiple branches (Sokolov and Rintoul, 2002, 2009a,b). The Subantarctic Zone (SAZ) lies between the STF and SAF, the Polar Frontal Zone (PFZ) lies between the SAF and the PF, and the Antarctic Zone (AZ) lies between the PF and the Antarctic continent.

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). However, the microbial assemblages that characterize Antarctic waters are generally poorly understood, and their diversity and functional capacity are not well characterized (reviewed in Murray and Grzymski, 2007; Wilkins *et al.*, 2012b). Large scale metagenome surveys have not previously been performed.

Recent anthropogenic climate change may be driving the warming and freshening of the ACC (Böning *et al.*, 2008) and shifting it and its fronts poleward (Fyfe and Saenko, 2005; Biastoch *et al.*, 2009). A community-level understanding is required to effectively understand the main components and dynamics of the microbial food web in the SO and thereby predict the effects of a shifting ACC on the distribution and abundance of plankton. The oceanic changes may have global ecological significance as the SO performs many ecosystem functions, including significant sequestration of anthropogenic CO₂ (Sabine *et al.*, 2004; Mikaloff Fletcher *et al.*, 2006) through both physiochemical processes and the “biological pump” of CO₂ fixation (Thomalla *et al.*, 2011).

In the austral summer of 2006 we initiated a metagenome program based on the sampling design of the Global Ocean Sampling (GOS) expedition (Rusch *et al.*, 2007), aimed at providing a baseline to monitor microbial communities in the Australian section of the SO. To date we have sampled SO water from 73° E to 150° E and 44° S to 68° S at depths from the surface to ~ 3700 m. In this study, we present data for plankton assemblages passed through a 20 µm prefilter and captured onto sequential 0.1 µm, 0.8 µm and 3.0 µm filters. This fractionation approach was originally adopted by the GOS expedition (Rusch *et al.*, 2007) and enables deeper sequencing of a greater representation of the marine microbial community thereby improving the identification of low-abundance taxa. By adopting this approach the metagenome datasets generated from the SO can be directly compared to available GOS data (e.g. Brown *et al.*, 2012). In studies of the SO and marine derived Antarctic lakes, the approach has

also proven useful for learning about the microbial ecosystem in the context of resource partitioning, including particle attachment, trophic status and virus–host interactions (Ng *et al.*, 2010; Lauro *et al.*, 2011; Yau *et al.*, 2011; Williams *et al.*, 2012b).

The samples in our SO study were collected in summer 2007/2008 on the SR3 transect (Sokolov and Rintoul, 2002) of the Climate of Antarctica and the Southern Ocean (CASO), and Collaborative East Antarctic Marine Census (CEAMARC) projects during the International Polar Year (IPY) program. We assessed the taxonomic and functional profiles of microbial communities from either side of the PF, thereby contributing important new information about the microbial ecology of the SO and defining the microbial communities most influenced by the effects of the PF forming a biogeographical barrier.

TODO end pasted block

Methods

Sampling and metagenomic sequencing

Sampling¹ was conducted on board the RSV *Aurora Australis* during cruise V3 CEAMARC / CASO from 13 December 2007 – 26 January 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 1).

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 1). These data were used to locate the PFZ based on a surface temperature gradient of ~ 1.35 °C across a distance of 45–65 km, placing the PF at approximately –59.70° of latitude, consistant 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 1). The North Zone (NZ) represents waters from the Subtropical, Subantarctic and PFZ regions, while the South Zone (SZ) represents the AZ.

At each station, ~ 250–560 L of seawater was pumped from ~ 1.5–2.5 m below the sea surface into drums stored at ambient temperature on deck. In the case of deep samples, ~ 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 (Port Washington, USA), and immediately stored at –20 °C (Rusch *et al.*, 2007; Ng *et al.*, 2010).

¹Sampling was performed by Jeffrey M. Hoffman and Jeffrey B. McQuaid

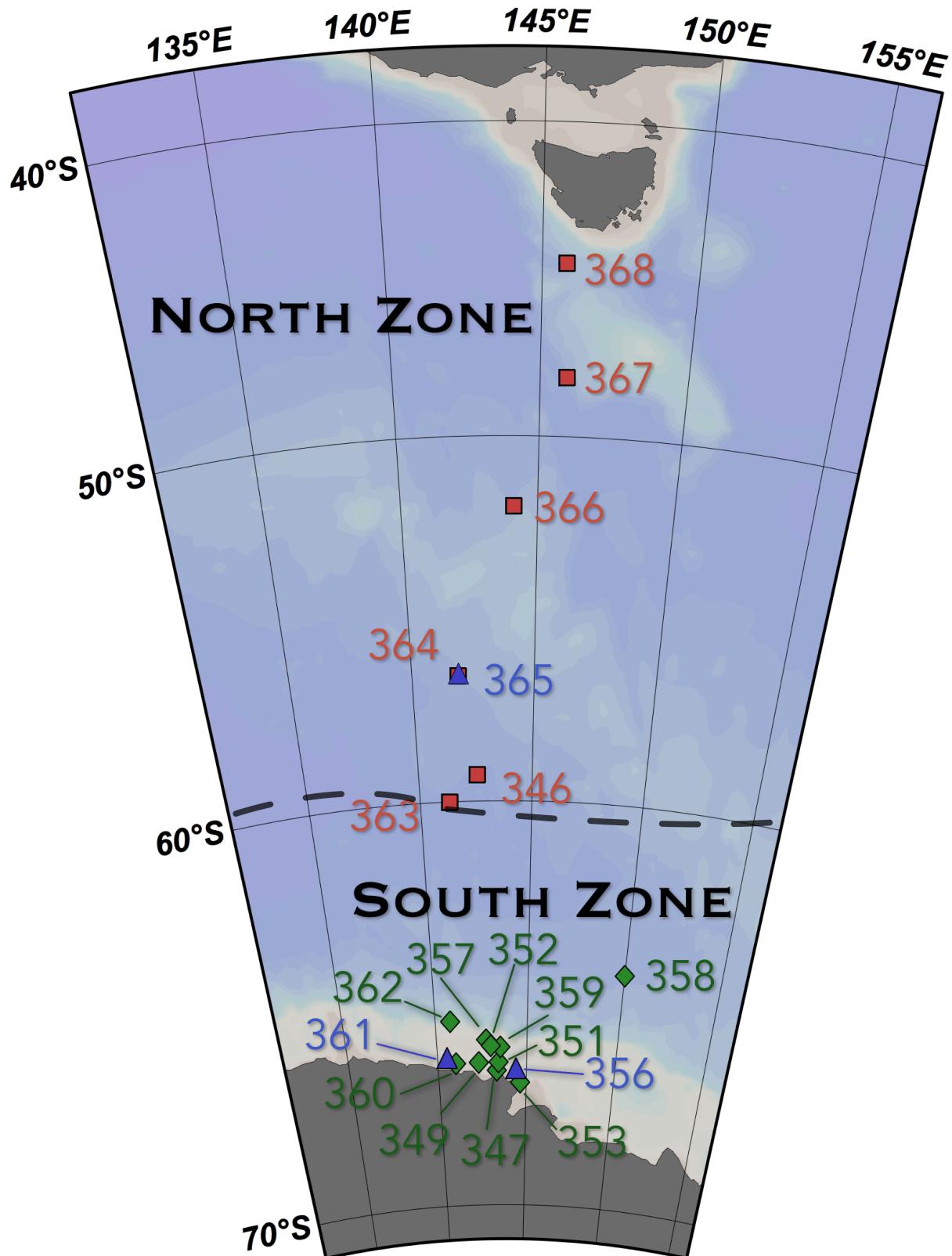


Figure 1: Sites of seawater samples used in this study. Red squares indicate surface samples from the North Zone; green diamonds samples from the South Zone; and blue triangles indicate deep samples. The dashed line gives the approximate location of the Polar Front.

Table 1: Sampling time, location and physiochemical 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 ($\mu\text{g L}^{-1}$)	Volume filtered (L)
346	North	20/12/07	-59.3120	142.5949	4294	2	2.9	33.75	0.3	500
347	South	23/12/07	-66.0213	142.7380	450	2	0.6	34.20	4.0	250
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
362	South	19/01/08	-65.5367	140.8287	1064	2	0.7	32.20	0.5	500
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

DNA extraction² 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 this database was not expected to contain representative genomes for every species present, 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 1.0×10^{-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 (see section following), and hits not belonging to the minimal sets were removed.

Identification of minimal species sets with MINSPEC

A computational method to minimise false OTU identifications and increase the accuracy of OTU abundance estimates (MINSPEC) was developed and implemented in PERL³. 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 is framed as a linear programming problem and solved with GLPSOL (The GNU Linear Programming/MIP solver) (Free Software Foundation, Boston). 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.

To validate this approach and estimate error rates, simulated microbial assemblages were gener-

²DNA extraction was performed by Cynthia Andrews-Pfannkoch and others at the J. Craig Venter Institute

³MINSPEC and the associated metagenomic simulation and validation scripts are open source and available at <https://github.com/wilcox/minspec>; a copy has also been provided in the supplementary information.

ated and simulated metagenomic sampling and **BLAST** search was performed on each assemblage. To simulate sequence identity between taxa, each simulated taxon went through up to fifty rounds in which another taxon was selected at random and deemed to have sequence identity with the first. After each round, the this process was terminated with a 10% probability to simulate an exponential curve of interrelatedness between taxa. A random subset of the simulated taxa were then selected to form the simulated assemblage. Combined with the simulated seqence identity between taxa, this caused some taxa in the assemblage to have identity to taxa outside it. A simulated metagenomic sampling was then performed, in which a taxon was selected at random to generate a read. To simulate a natural rank-abundance curve, the randomly selected taxon would be rejected with probability $1 - \frac{1}{\ln(x)+1}$, where x is the taxon's rank. Simulated **BLAST** matches to the taxon were generated for the remaining reads. Each time a taxon was selected to produce a read, other taxa with simulated sequence identity were also randomly selected to produce **BLAST** matches for that read, simulating the problem of a single read producing multiple matches to closely related taxa.

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 taxa [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 taxa simulated. The resulting simulated **BLAST** outputs were processed with **MINSPEC**, and the false positive (percentage of taxa not in the assemblage which nevertheless survived **MINSPEC** filtering) and false negative (percentage of taxa present in the assemblage which were not present after **minsperc** 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 taxa which were present in the assemblage but through random chance did not generate any reads, the equivilant 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” taxa — those which generated **BLAST** matches but were not part of the assemblage — that were succesfully removed by **MINSPEC** was calulated.

OTU abundances and variance between zones

The relative OTU abundances for each sample were determined using the PERL script **GAAS** (Genome relative Abundance and Average Size)(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).

Functional analysis of metagenomic data

BLAST comparison to 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 1.0×10^{-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. Modules with a high contribution to variance or otherwise of interest were then linked to taxonomy (“taxonomic decomposition”) by noting the genus of the organism associated with each gene in the KEGG GENES database and thus calculating the relative contribution of each genus to each module’s abundance. 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.

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).

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 2 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 2).

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 3, and are graphically summarised for all OTUs in Figure 3.

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*

Table 2: 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
<i>Nitrosopumilus maritimus</i> SCM1	0.01996	0.01438	0.009508	1.076	1.309	1.210	19.09	9.463	17.77
"Candidatus Ruthia magnifica" str. Cm (<i>Calyptogena magnifica</i>)	0.6699	0.6458	0.5484	2.987	2.616	1.025	3.945	4.601	2.264
<i>Roseobacter</i> sp. OCh114	0.3125	2.932	1.588	0.4477	3.994	2.657	0.1259	1.228	0.6792
<i>Synechococcus</i> sp. CC9902	0.1081	9.837	4.973	0.007484	0.004156	0.09733	0.002846	0.01502	0.01058
<i>Silicibacter pomeroyi</i> DSS-3	0.2578	2.286	1.154	0.3070	2.505	1.576	0.1224	0.9417	0.4988
<i>Gramella forsetii</i> 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	0.2078	1.970	1.807	0.2174	2.480	2.662	1.167	
<i>Robiginitalea biformata</i> strain HTCC2501	0.2751	1.099	1.297	0.4722	1.878	1.405	0.2265	0.6188	0.6946
<i>Flavobacterium psychrophilum</i> strain JIP02/86	0.1718	0.8409	1.224	0.4316	1.960	1.598	0.1599	0.4744	0.6001
"Candidatus Punciepirillum marinum"	0.03014	4.624	4.409	0.007221	0.002778	0.02764	0.001580	0.002863	0.009241
IMCC1322	0.6444	2.077	1.267	0.3586	1.377	0.7109	0.3425	1.062	0.5345
<i>Silicibacter</i> sp. TM1040	0.2274	1.652	0.8738	0.2709	1.803	1.233	0.07665	0.5890	0.2957
<i>Jannaschia</i> sp. DFL-12	0.1776	1.378	0.7350	0.2443	1.692	0.8009	0.07338	0.6515	0.3078
<i>Zunongwangia profunda</i> strain SM-A87	0.1522	0.7487	1.059	0.2968	1.410	1.204	0.1353	0.3478	0.4971
<i>Cotovelitia</i> sp. 34H	0.02345	0.3636	2.736	0.05207	0.5140	1.041	0.05137	0.4687	0.8013
<i>Coraliomargarita akajimensis</i> strain DSM 45221	0.03698	0.07573	0.1197	0.1154	1.543	1.680	0.02614	0.3040	0.2740
<i>Jannaschina</i> sp. CCS1	0.1173	0.9344	0.4784	0.1711	1.230	0.8239	0.05865	0.4462	0.2118
<i>Pseudoalteromonas atlantica</i> strain T6c	0.01251	0.4772	1.993	0.02270	0.4089	1.132	0.02634	0.2143	0.7459
<i>Saccharophagus degradans</i> strain 2-40	0.06532	0.4325	0.5429	0.1289	1.072	0.8663	0.07798	0.2844	0.3165
<i>Flavobacterium johnsoniae</i> strain UW101	0.08822	0.4220	0.6141	0.2034	0.9389	0.8578	0.07545	0.225	0.3300
<i>Capnocytophaga ochracea</i> strain DSM 7271	0.1143	0.4830	0.5399	0.2314	0.8815	0.6814	0.08964	0.2840	0.5043
<i>Marinomonas</i> sp. MWYL1	0.03777	0.2529	0.3026	0.1514	1.300	0.7006	0.07393	0.2439	0.2155
<i>Cellvibrio japonicus</i> strain Ueda107	0.05884	0.3080	0.3231	0.1155	0.9917	0.4713	0.06774	0.2981	0.2549
<i>Marinobacter hydrocarbonoclasticus</i> VT8	0.04093	0.2889	0.3883	0.08418	0.7195	0.3848	0.1250	0.6667	1.066
<i>Pseudoalteromonas haloplanktis</i> strain TAC125	0.01389	0.2505	0.8896	0.03427	0.3561	0.6530	0.1092	1.203	0.1503
<i>Teredinibacter turnerae</i> strain T7901	0.05665	0.3051	0.3081	0.1138	0.9174	0.5127	0.06558	0.2649	0.1885
<i>Acinetobacter baumannii</i> strain SDF	0.004886	0.007187	0.4073	0.006260	0.04218	1.459	0.004285	0.01229	0.3155

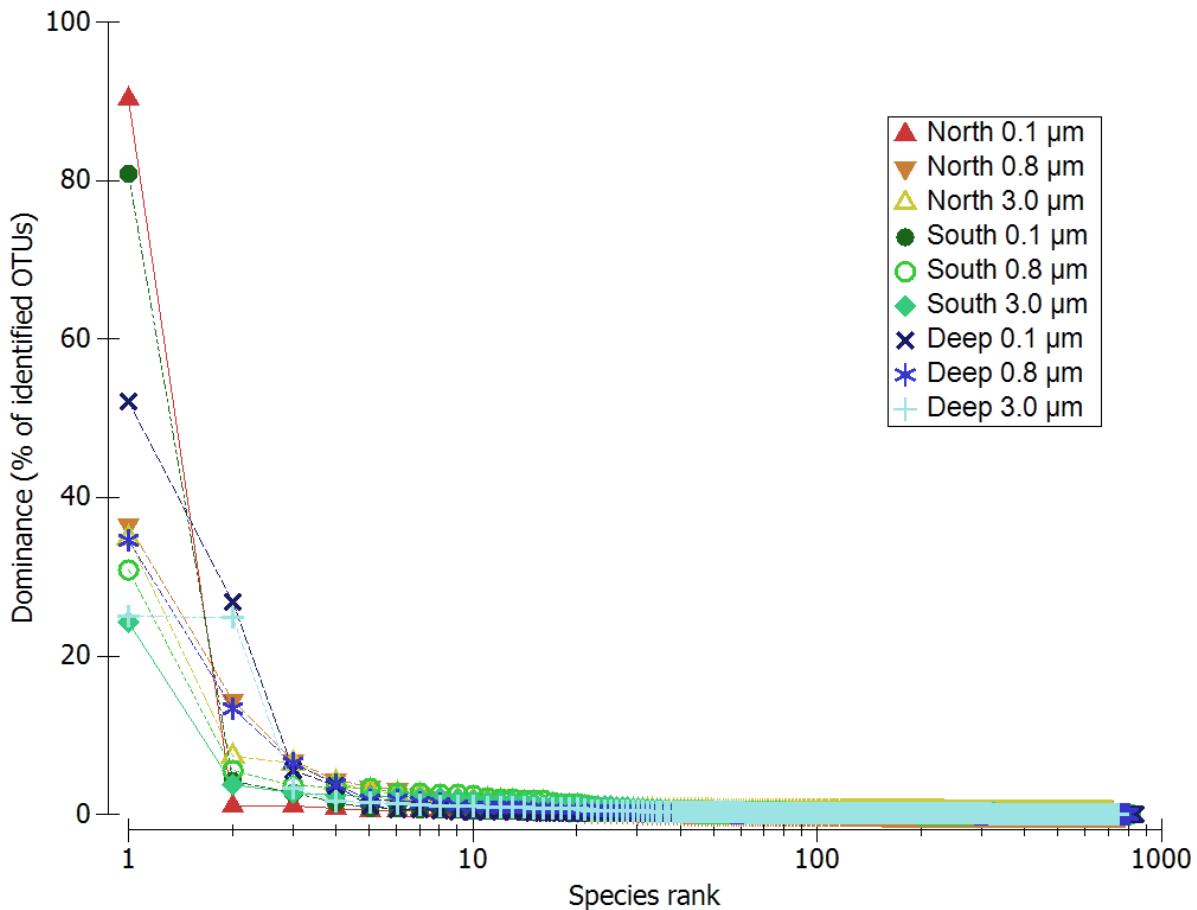


Figure 2: Rank-abundance curves for OTUs identified in each zone and size fraction. The dominance of a given OTU is its relative abundance as a percentage of all identified OTUs. The x-axis is scaled logarithmically. Generated using PRIMER 6.

Table 3: 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 (<i>Calyptogena magnifica</i>) 0.1 µm	0.82	0.25	1.57
<i>Colwellia</i> sp. 34H 3.0 µm	0.34	0.66	1.32
"Ca. Ruthia magnifica" str. Cm (<i>Calyptogena magnifica</i>) 0.8 µm	0.74	0.25	1.32
"Ca. 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
<i>Coraliomargarita akajimensis</i> strain DSM 45221 0.8 µm	0.48	0.04	1.13
<i>Coraliomargarita akajimensis</i> strain DSM 45221 3.0 µm	0.49	0.06	1.10
<i>Roseobacter</i> sp. OCh114 0.8 µm	1.01	0.81	1.08
<i>Pseudoalteromonas atlantica</i> strain T6c 3.0 µm	0.38	0.54	1.08
"Ca. Vesicomyosocius okutanii" strain HA 0.8 µm	0.57	0.19	1.04
<i>Actinobacter baumannii</i> strain SDF 3.0 µm	0.45	0.18	0.95
<i>Gramella forsetii</i> strain KT0803 0.8 µm	0.72	0.43	0.94
<i>Martinomonas</i> sp. MWYL1 0.8 µm	0.46	0.11	0.92
<i>Roseobacter</i> sp. OCh114 3.0 µm	0.76	0.54	0.91
<i>Flavobacterium psychrophilum</i> strain JIP02/86 0.8 µm	0.63	0.32	0.89
<i>Silicibacter pomeroyi</i> DSS-3 0.8 µm	0.75	0.69	0.86
<i>Brachyspira hyodysenteriae</i> strain WA1 3.0 µm	0.47	0.19	0.84
"Ca. Ruthia magnifica" str. Cm (<i>Calyptogena magnifica</i>) 3.0 µm	0.34	0.21	0.82
<i>Pseudoalteromonas haloplanktis</i> strain TAC125 3.0 µm	0.22	0.33	0.77
<i>Robiginitalea biformata</i> strain HTCC2501 0.8 µm	0.61	0.40	0.74
<i>Nitrosopumilus maritimus</i> SCM1 0.1 µm	0.27	0.01	0.72
<i>Gramella forsetii</i> strain KT0803 3.0 µm	0.59	0.59	0.71
<i>Lysinibacillus sphaericus</i> strain C3-41 3.0 µm	0.29	0.02	0.71
<i>Nitrosopumilus maritimus</i> SCM1 0.8 µm	0.25	0.01	0.70
<i>Silicibacter</i> sp. TM1040 0.8 µm	0.59	0.55	0.69

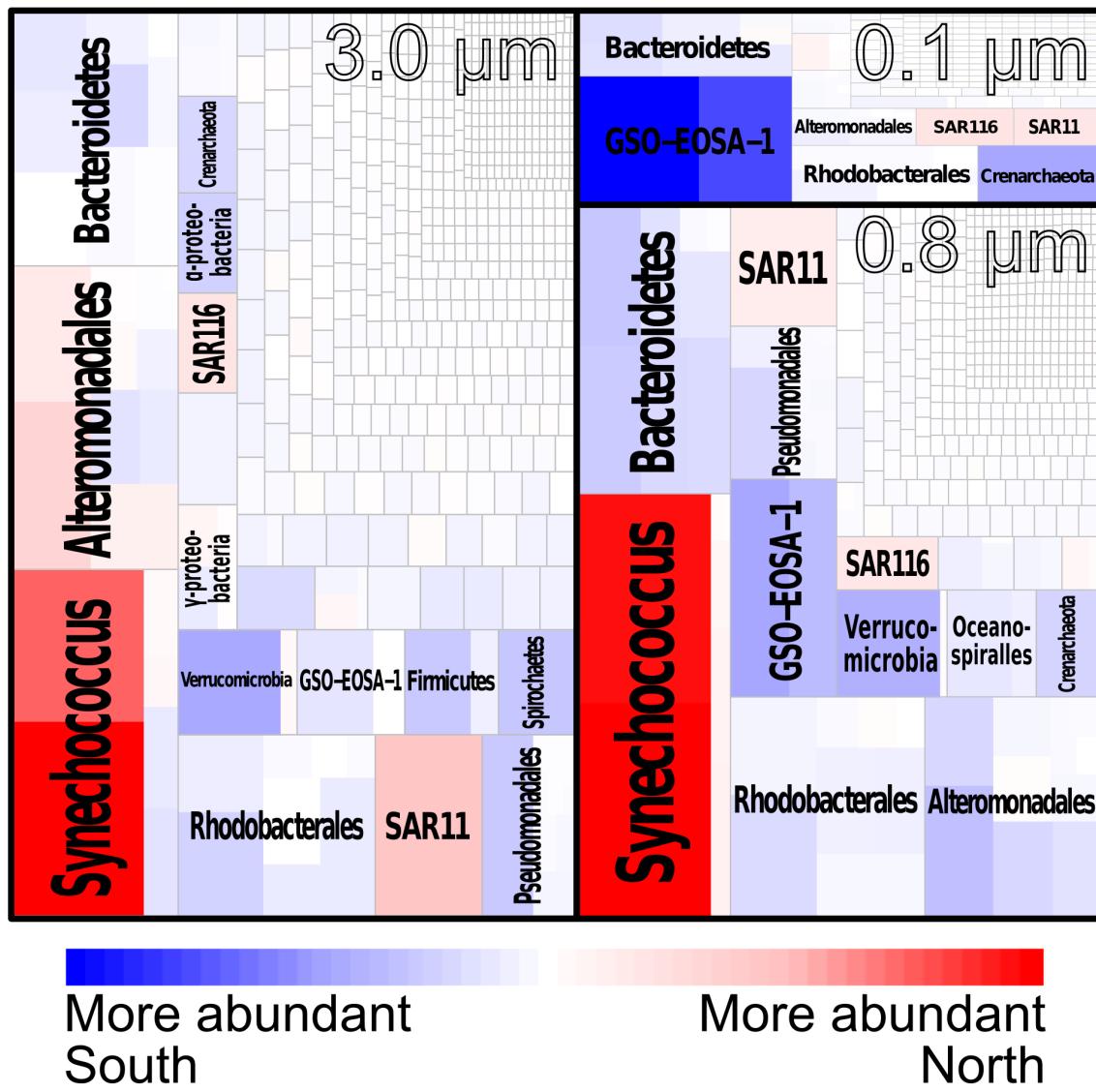


Figure 3: Contribution of OTUs to variance between the North and South zones, and differential abundance of OTUs from each size fraction between the two zones. Each coloured (red or blue) rectangle represents an OTU identified through analysis of BLAST matches between SO metagenome data and the RefSeq 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).

spp., SAR11 and SAR116 were on average more abundant in the NZ (Figure 3). Some groups, such as the Alteromonadales, had variable relative representation depending on size fraction.

Validation of MINSPEC

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.

The false negative rate, or percentage of taxa in the assemblage which were absent from the BLAST results following MINSPEC processing, was generally high, ranging from ~ 20% under ideal conditions (a low assemblage / all taxa ratio, and 500,000-read metagenomic sample) to ~ 90% in the worst case (a high assemblage / all taxa ratio and a small metagenomic sample) (Figure 4a). The assemblage / all taxa ratio (hereafter referred to as “assemblage ratio”) indicates the proportion of simulated taxa (“all taxa”) which was chosen to form the simulated assemblage. A higher ratio means it is more likely on average that any randomly selected taxon is part of the assemblage, and thus that any individual failure to detect a taxon is incorrect. This problem is mitigated with increasing the number of reads, as this makes it less likely that a given taxon would go undetected. The extreme false negative rates, in some cases 100%, represent extreme simulated scenarios (e.g. an assemblage of 1 taxon drawn from a pool of 100,000), and thus are unlikely to reflect real metagenomic studies.

Because the majority of false negatives are attributable to undersampling and failure of taxa to generate BLAST hits — properties the simulated metagenomic experiments share with real ones — a second metric, the false negative (MINSPEC) rate, was calculated (Figure 4b). This is the proportion of taxa in the assemblage which 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 ~ 0–1% for low assemblage ratios, to ~ 15–20% under high ratios. Surprisingly, increasing the number of reads only slightly decreased the rate, at both low and high assemblage ratios. This may be because MINSPEC requires only one read which has identity to a single taxon to ensure that taxon is not removed.

The false positive rate, or percentage of taxa not in the assemblage which nevertheless generated high-quality BLAST matches that were not removed by MINSPEC, was generally ~ 0–5% except for extremely small read sets and low assemblage ratios, where it reached as high as 60% (Figure 4c). These results reinforce the value of larger read sets, and show that once a modest metagenome size is reached (~ 100,000 reads) very few false positives can be expected.

The proportion of false taxa removed was calculated to measure MINSPEC’s success at identifying

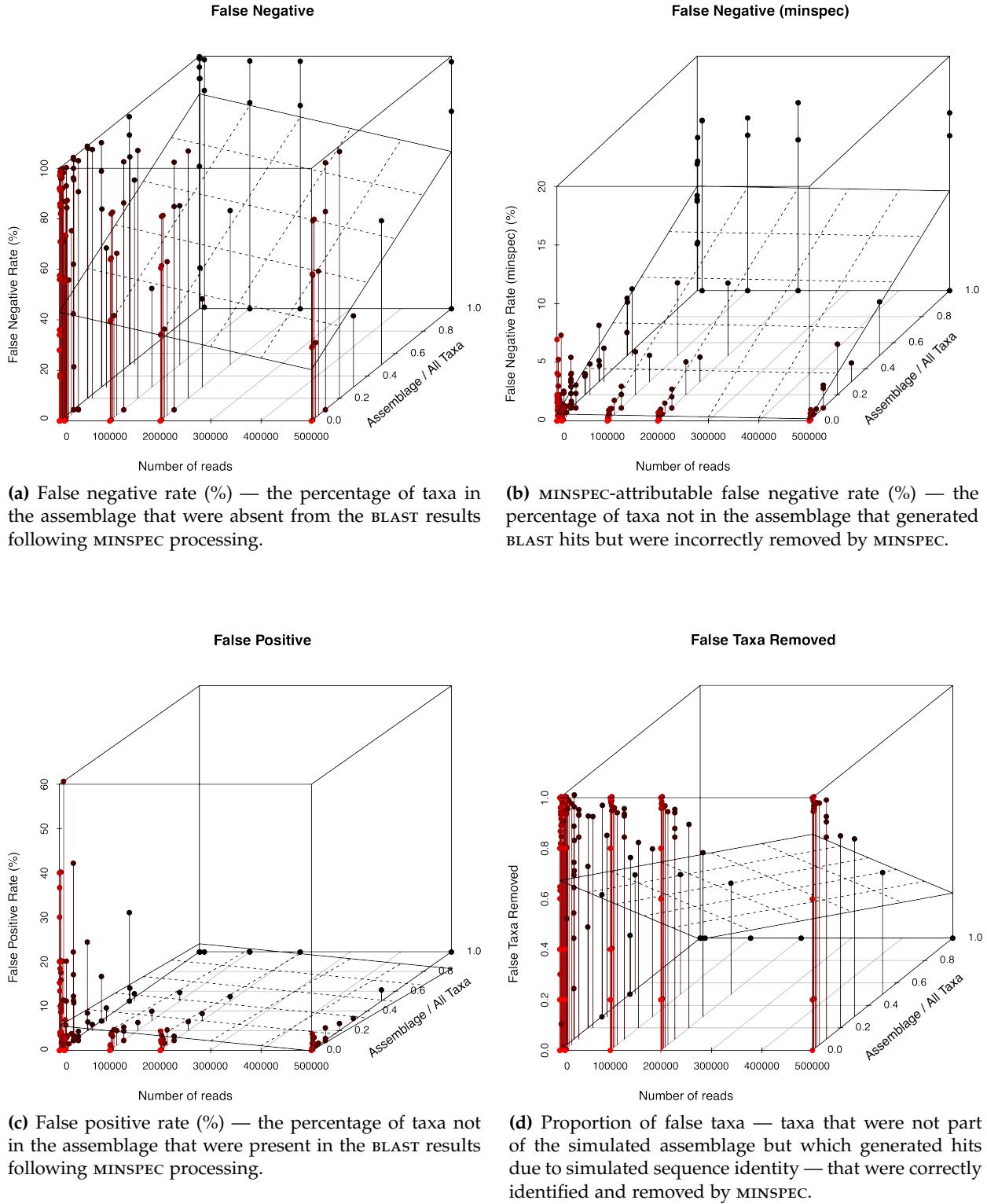


Figure 4: Results of repeated trials of MINSPEC on simulated metagenomic studies with multiple permutations of parameters (number of reads, number of simulated taxa, size of simulated assemblage). The number of simulated taxa and size of simulated assemblage are represented as a ratio on the z-axis (“assemblage / all taxa”). Each permutation was repeated five times. A plane representing a linear regression has been overlayed on each plot to indicate the trend. Points have been tinted to aid the perception of depth; colour is not otherwise meaningful.

and eliminating taxa 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 (Figure 4d). 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 taxa removed. This is likely because in assemblages with a higher assemblage ratio, there are fewer false taxa to remove; in assemblages with a ratio of 1, there are none. The high proportion of false taxa correctly identified in simulations with a low assemblage ratio is thus a good indication that MINSPEC is generally successful at identifying and removing false taxa, 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.

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 4, and a complete list in the supplementary material (PF-modules-SIMPER.csv). The highest-contributing ortholog groups are given in Table 5, 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.

Table 4: 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/dCTPdTDP/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 5: 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 $\text{Na}^+:\text{H}^+$ 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:2.1.1.48] 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

^aDue to an error in the KEGG database, this module is encoded twice.

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 2). 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 3). 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 5), 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.*, 2012a). 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.

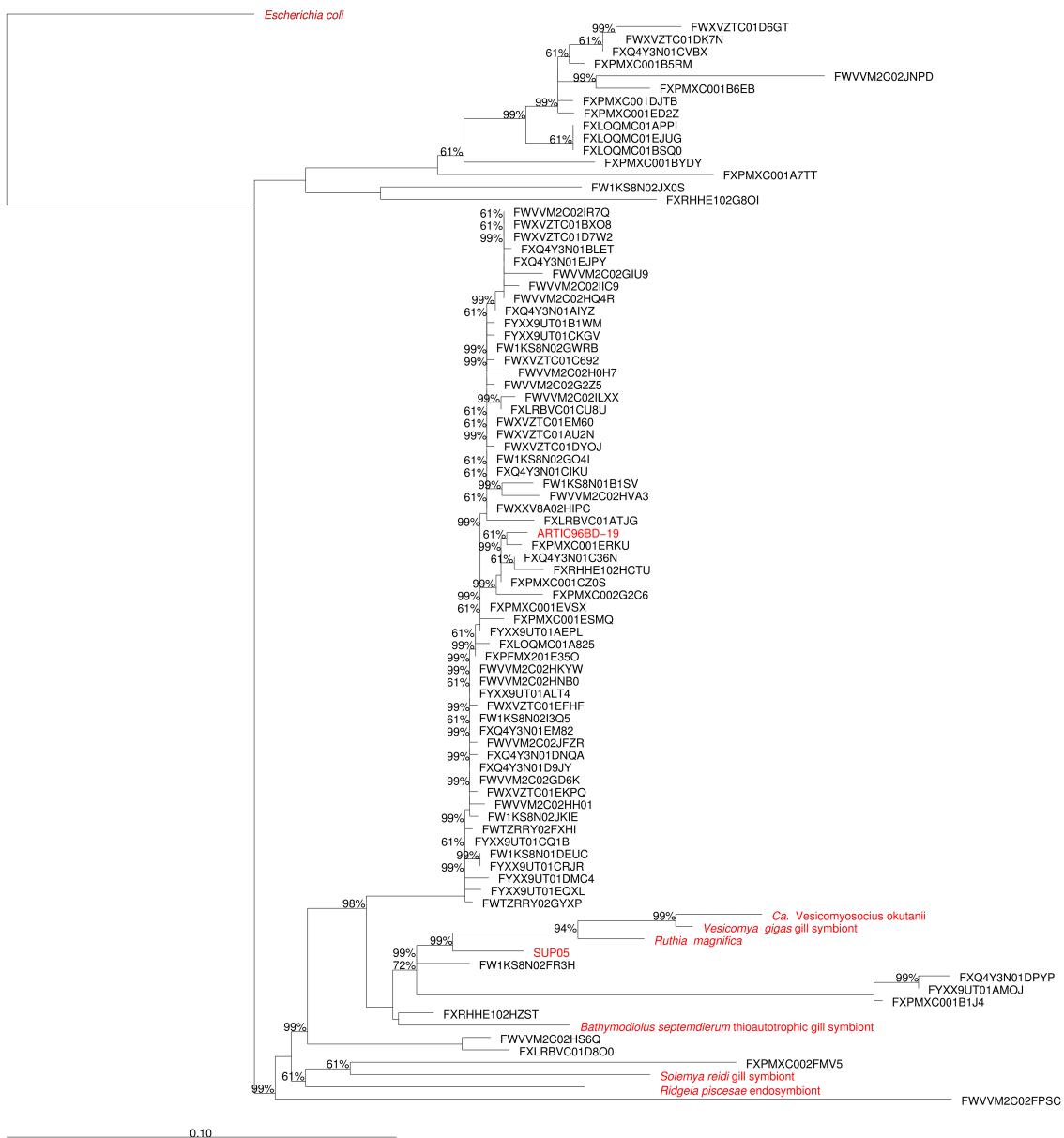


Figure 5: Neighbour-joining tree of GSO-EOSA-1-like 16S rRNA gene sequences from the metagenomes in this study. Sequences labeled in black text are reads from the metagenomes. Red labels are 16S rRNA gene sequences from Gammaproteobacterial Sulphur Oxidisers (GSO) and other gammoproteobacteria. The tree 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 3). 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.*, 2012a).

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 3). 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.*, 2012a).

SAR11 and SAR116 clades

“*Candidatus Pelagibacter 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 “*Candidatus Puniceispirillum marinum*” from the SAR116 clade were a moderate contributor to variance between the NZ and SZ with higher abundance in the NZ (Figure 3). 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 hybridisation) 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 3). 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 remineralizing 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 CO₂, 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 (CO₂, 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 3). *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 3). 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).

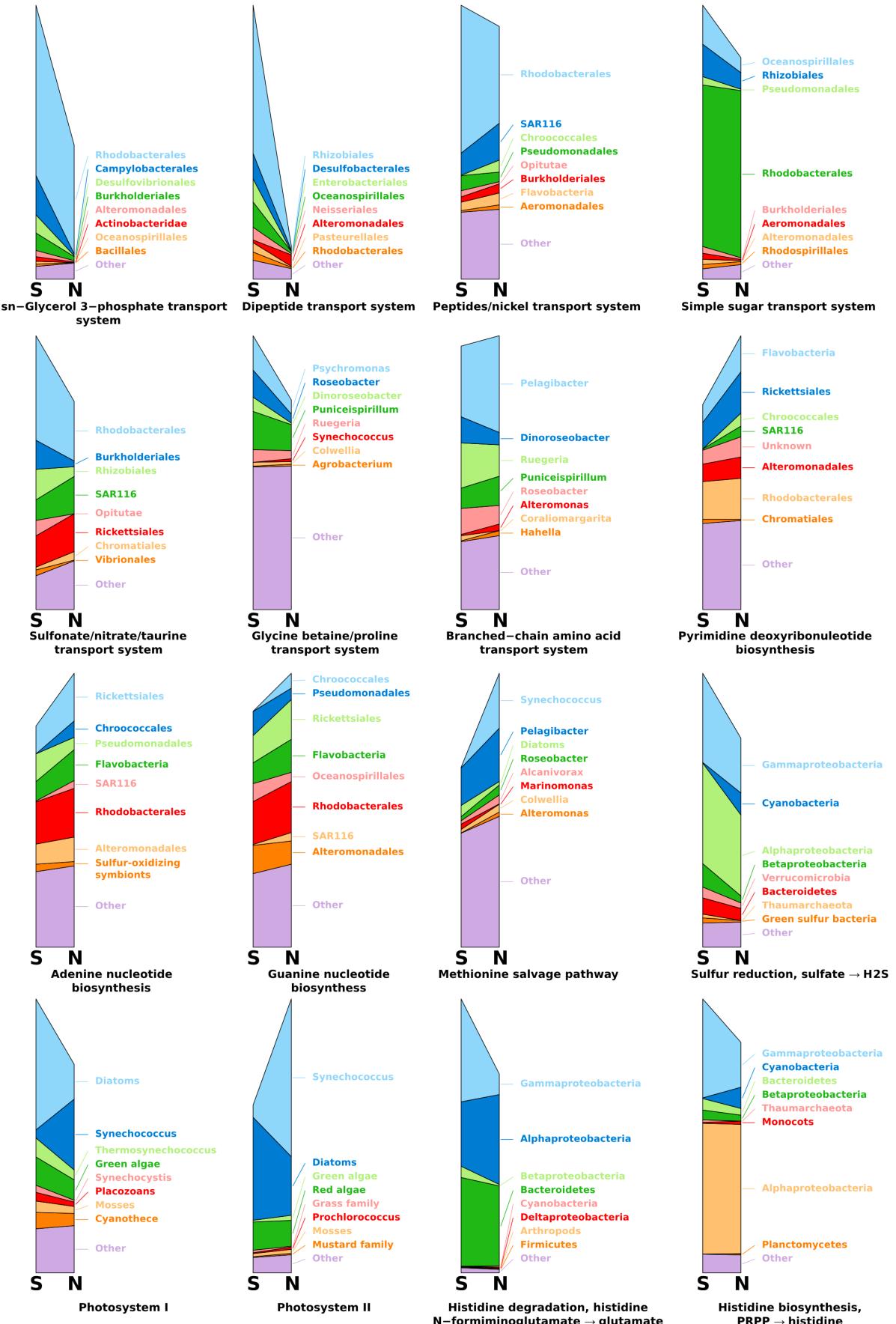


Figure 6: Decomposition of KEGG modules of interest to contributing classes, orders or genera. The left side of each stack (S) indicates the proportion of the module abundance contributed by each class, order or genus in the South Zone, while the right side (N) represents the North Zone. As the contributions 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.

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 4). 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 6). TODO methods for generating this figure TODO does this figure belong in results? 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 6). 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 deoxyribonuleotide 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 6), 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 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 methylmercaptocaptopropionate (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 6), 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 6). 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 6). Variation in the photosystem I module, which was marginally overrepresented in the SZ, could largely be attributed to diatoms and other eukaryotic phytoplankton

(Figure 6), 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 6), 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 ($70\text{--}115^\circ$ 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.

Meso-scale biogeographic drivers of planktonic diversity

Conclusions

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., 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.
- Bastoch A., Böning C. W., Schwarzkopf F. U., and Lutjeharms J. R. E. (2009). Increase in Agulhas leakage due to poleward shift of Southern Hemisphere westerlies. *Nature*, 462(7272):495–498.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Clarke K. R. and Warwick R. M. *Change in marine communities: an approach to statistical analysis and interpretation*. PRIMER-E, Plymouth, 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.
- 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.
- 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.
- Feller G. and Gerday C. (2003). Psychrophilic enzymes: hot topics in cold adaptation. *Nature Reviews Microbiology*, 1(3):200–208.
- 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.
- Fyfe J. C. and Saenko O. A. (2005). Human-induced change in the Antarctic Circumpolar Current. *Journal of Climate*, 18(15):3068–3073.
- 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., Zwissler 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.
- 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.
- 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., 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.
- 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.
- 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 in the oceanographic environment and zooplankton community in the vicinity of a sub-Antarctic archipelago. *Marine Biology*, 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.
- 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.
- Kjelleberg S., Hermansson M., and Mårdén P. (1987). The transient phase between growth and non-growth of heterotrophic bacteria, with emphasis on the marine environment. *Annual Review of Microbiology*, 41:25–49.
- 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.
- 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.
- 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.
- 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.
- Marchant H. J., Davidson A. T., and Wright S. W. (1987). The distribution and abundance of chroococoid 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., Burkhill 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.

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 CO₂ uptake, transport, and storage by the ocean. *Global Biogeochemical Cycles*, 20(2):GB2002.

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., 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.

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.

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.

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.

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.

- 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.
- 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.
- 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.
- 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 CO₂. *Science*, 305(5682):367–371.
- 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.
- 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.
- Sokolov S. and Rintoul S. R. (2009). Circumpolar structure and distribution of the Antarctic Circumpolar Current fronts: 2. Variability and relationship to sea surface height. *Journal of Geophysical Research*, 114(C11):C11019.
- 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., Reinthal 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.

- 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.
- 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.
- Vila-Costa M., Simó R., Harada H., Gasol J. M., Slezak D., and Kiene R. P. (2006). Dimethylsulfonio-propionate 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 Woold-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.
- 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.
- 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.
- Williams T. J., Wilkins D., Long E., Evans F., DeMaere M. Z., Raftery M. J., and Cavicchioli R. (2012). The role of planktonic *Flavobacteria* in processing algal organic matter in coastal East Antarctica revealed using metagenomics and metaproteomics. *Environmental Microbiology*, pages n/a–n/a.
- Yau S., Lauro F. M., DeMaere M. Z., Brown M. V., Thomas T., Raftery M. J., Andrews-Pfannkoch C., Lewis M., Hoffman J. M., Gibson J. A., and Cavicchioli R. (2011). Virophage control of antarctic algal host-virus dynamics. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 108(15):6163–6168.
- 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.
- 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.