Microbial Ecology and Biogeography of the Southern Ocean

David Wilkins

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

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

April 2013

Contents

Abstract v

Acknowledgements vii

Publications ix

List of Figures xv

List of Tables xvii

List of Acronyms xix

1 Introduction 1
- 1.1 Physical oceanography of the Southern Ocean . . . 1
  - 1.1.1 Fronts and zones . . . 1
  - 1.1.2 Water masses and circulation . . . 3
  - 1.1.3 Effect of climate change . . . 4
- 1.2 Microbial ecology of the Southern Ocean . . . 4
  - 1.2.1 Bacteria . . . 5
    - 1.2.1.1 Alphaproteobacteria . . . 5
      - 1.2.1.1.1 Roseobacter clade . . . 5
      - 1.2.1.1.2 SAR11 . . . 6
      - 1.2.1.1.3 SAR116 . . . 7
    - 1.2.1.2 Betaproteobacteria . . . 8
    - 1.2.1.3 Gammaproteobacteria . . . 8
      - 1.2.1.3.1 SAR86 . . . 8
      - 1.2.1.3.2 OMG group . . . 9
      - 1.2.1.3.3 Ant4D3 . . . 9
      - 1.2.1.3.4 GSO-EOSA-1 . . . 10
    - 1.2.1.4 Deltaproteobacteria . . . 10
    - 1.2.1.5 CFB . . . 10
    - 1.2.1.6 Cyanobacteria . . . 12
    - 1.2.1.7 Verrucomicrobia . . . 12
    - 1.2.1.8 Other bacteria . . . 12
  - 1.2.2 Archaea . . . 13
  - 1.2.3 Virioplankton . . . 14
- 1.3 Project aims . . . 15
  - 1.3.1 The Polar Front . . . 15
    - 1.3.1.1 Biogeographic role . . . 15
    - 1.3.1.2 Differences in community composition . . . 15
    - 1.3.1.3 Differences in functional potential . . . 15
  - 1.3.2 The role of advection in microbial biogeography . . . 15

2 MINSPEC **17**
2.1 Abstract . . . 17
2.2 Introduction . . . 17
2.2.1 Metagenomic analysis of microbial assemblages . . . 17
2.2.2 The maximum parsimony approach . . . 19
2.3 Methods . . . 20
2.3.1 Implementation of MINSPEC . . . 20
2.3.2 Validation of MINSPEC . . . 21
2.4 Results . . . 22
2.5 Discussion . . . 22
2.6 Conclusions . . . 24

3 The Polar Front **25**
3.1 Abstract . . . 25
3.2 Introduction . . . 26
3.3 Methods . . . 27
3.3.1 Sampling and metagenomic sequencing . . . 27
3.3.2 Phylogenetic analysis of metagenomic data . . . 30
3.3.2.1 BLAST comparison to RefSeq database . . . 30
3.3.2.2 OTU abundances and variance between zones . . . 30
3.3.2.3 Fragment recruitment to verify Operational Taxonomic Unit (OTU) identification . . . 31
3.3.2.4 Additional samples to test "polynya hypothesis" . . . 32
3.3.3 Functional analysis of metagenomic data . . . 32
3.3.3.1 BLAST comparison to KEGG database . . . 32
3.3.3.2 Analysis of functional potential . . . 33
3.3.3.3 Taxonomic decomposition . . . 33
3.4 Results . . . 33
3.4.1 Metagenomic sequencing . . . 33
3.4.2 Phylogenetic analysis of metagenomic data . . . 34
3.4.3 Fragment recruitment to verify OTU identification . . . 39
3.4.4 Additional samples to test alternative "polynya hypothesis" . . . 39
3.4.5 Functional analysis of metagenomic data . . . 39
3.5 Discussion . . . 43
3.5.1 Taxonomic groups differentiating the zones . . . 43
3.5.1.1 GSO-EOSA-1 . . . 43
3.5.1.2 Ammonia-oxidizing Crenarchaeota . . . 45
3.5.1.3 Cyanobacteria . . . 45
3.5.1.4 SAR11 and SAR116 clades . . . 46
3.5.1.5 Bacteroidetes . . . 47
3.5.1.6 Rhodobacterales . . . 47
3.5.1.7 Alteromonadales . . . 48
3.5.1.8 Verrucomicrobia . . . 48
3.5.2 Functional capacities differentiating the zones . . . 48
3.5.2.1 Conclusions . . . 52

4 Deep water formation **53**
4.1 Abstract . . . 53
4.2 Introduction . . . 53
4.3 Methods . . . 54
4.4 Results and Discussion . . . 54

5 The advection effect **59**
5.1 Abstract . . . 59
5.2 Introduction . . . 59
5.2.1 Distance and environment effects in microbial biogeography . . . 59
5.2.2 Water mass endemicity and advection of marine microorganisms . . . 60
5.2.3 Aims and approach of this study . . . 61
5.3 Methods . . . 62
5.3.1 Sampling . . . 62
5.3.2 DNA extraction . . . 63
5.3.3 Sequencing . . . 65
5.3.4 Taxonomic assignment . . . 65
5.3.5 Physicochemical and spatial distances . . . 65
5.3.6 Generation of advection distance matrix . . . 66
5.3.7 Ordination of distance matrices and comparison to water masses . . . 67
5.3.8 Testing of advection effect . . . 67
5.3.9 Differential influence of advection on OTU subsets . . . 68
5.4 Results . . . 69
5.4.1 Sequencing and taxonomic assignment . . . 69
5.4.2 Environment and distance effects . . . 74
5.4.3 Testing the advection effect . . . 76
5.4.4 Testing advection effect mechanisms . . . 80
5.4.5 Differential effect of advection on OTU subsets . . . 81
5.5 Discussion . . . 81
5.5.1 Taxonomic resolution . . . 82
5.5.2 Differential influence of advection on OTU subsets . . . 83
5.5.3 Future work . . . 83

6 General discussion **85**
6.1 Contributions of this thesis . . . 85
6.1.1 The Polar Front . . . 85
6.1.1.1 Biogeographic role of the Polar Front . . . 86
6.1.1.2 The Polar Front and climate change . . . 87
6.1.1.3 Future work . . . 88
6.1.2 The advection effect . . . 88
6.1.2.1 Future work . . . 90
6.1.3 MINSPEC . . . 90
6.2 The microbial species concept in the "omics" age . . . 91
6.3 Conclusions . . . 93

References **94**

List of Figures

1.1 Major fronts and water masses of the Southern Ocean 2

2.1 Results of MINSPEC validation 23

3.1 Map showing sites of seawater samples used in the Polar Front study 28
3.2 Summing relative abundances across size fractions 31
3.3 Rank-abundance curves for OTUs in each zone and size fraction 36
3.4 Contribution of OTUs to variance between the North and South zones 38
3.5 Read recruitment to reference genomes 40
3.6 Tree of GSO-EOSA-1 related 16S rRNA genes 44
3.7 Taxonomic decomposition of KEGG modules 50

4.1 Map showing sites of preliminary Antarctic Bottom Water (AABW) samples 55
4.2 Non-Metric Multidimensional Scaling (nMDS) of AABW, North Zone (NZ) and South Zone (SZ) samples 57

5.1 Map showing sites of samples used in the advection study 62
5.2 OTU assignments in the advection study. 73
5.3 nMDS of advective distances between samples. 74
5.4 nMDS of advective distances between samples. 75
5.5 dbRDA ordination of relationship between environment and community. . . 76
5.6 Encounter times for all samples in advection model. 77
5.7 nMDS of advective distances between samples. 79
5.8 Advection effect at different taxonomic resolutions 80

6.1 Biogeographic effect sizes 89

List of Tables

2.1 Examples of spurious OTU identifications 19

3.1 Details of samples used in Polar Front study 29
3.2 Additional samples used to test polynya hypothesis 32
3.3 Twenty most abundant OTUs 35
3.4 Highest-contributing OTUs to the difference between the North and South zones 37
3.5 Contributions of KEGG modules to variance between the North and South zones 41
3.6 Contributions of KEGG ortholog groups to variance between the North and South zones 42

4.1 AABW samples used in the preliminary analysis 54
4.2 Twenty most abundant OTUs in preliminary AABW samples 56

5.1 Full sample data for advection study 70
5.1 (cont.) Full sample data for advection study. 71
5.1 (cont.) Full sample data for advection study. 72
5.2 Correlations between dbRDA axes and physicochemical variables 75
5.3 Results of BVSTEP 82

List of Acronyms

AABW Antarctic Bottom Water.

GAAS Genome relative Abundance and Average Size.

IP Integer Programming.

ITS Internal Transcribed Spacer.

KEGG Kyoto Encyclopedia of Genes and Genomes.

LP Linear Programming.

MEGAN Metagenome Analyzer.

nMDS Non-Metric Multidimensional Scaling.

NZ North Zone.

OTU Operational Taxonomic Unit.

SO Southern Ocean.

SZ South Zone.

Chapter 2

MINSPEC, a bioinformatic tool for metagenomics

Sections of this chapter have been previously published in Wilkins D., Lauro F. M., Williams T. J., DeMaere M. Z., Brown M. V., Hoffman J. M., Andrews-Pfannkoch C., McQuaid J. B., Riddle M. J., Rintoul S. R., and Cavicchioli R. (2013). Biogeographic partitioning of Southern Ocean microorganisms revealed by metagenomics. *Environmental Microbiology*, 15(5):1318–1333.

2.1 Abstract

Incorrect assignment of taxonomic identity to sequencing reads is a source of error in microbial metagenomic studies. Microbial genomes naturally share large amounts of very similar or identical genomic sequence as a result of common ancestry, horizontal gene transfer or convergent evolution. Thus, when a read is similar to more than one Operational Taxonomic Unit (OTU) in a reference database of microbial genomes, nucleotide identity alone is insufficient to determine the correct taxonomic assignment. This chapter presents a novel method and software tool, MINSPEC, which determines the smallest set of OTUs that explains a given set of matches between metagenomic reads and a reference database. By removing OTUs not in this "maximum parsimony" set, MINSPEC reduces spurious OTU assignments (false positives) and thus increases the accuracy of relative abundance estimates. MINSPEC was validated against simulated metagenomic experiments.

2.2 Introduction

2.2.1 Metagenomic analysis of microbial assemblages

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

The first method is the identification, using a search and alignment algorithm such as BLAST, of specific marker genes or other sequences which are diagnostic for a particular OTU. Common targets in microbial ecology are the 16S or other ribosomal subunit rDNA

sequences, and the Internal Transcribed Spacer (ITS) regions between 16S–23S rDNA sequences (e.g. Brown *et al.*, 2012). This method provides several advantages. The selected regions are usually highly conserved, and through cultivation and full-genome sequencing have been reliably associated with a particular OTU, allowing very accurate identification and analysis of diversity down to the ecotype level (e.g. Brown *et al.*, 2012). If the copy number of the gene or region is well known, this method also allows for accurate estimations of cell abundance from metagenomes. However, a disadvantage of this method is that the large majority of metagenomic reads will not cover the region of interest, and will contribute nothing to the analysis. Low-abundance OTUs will therefore be missed, as the region of interest is unlikely to have been sequenced.

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

Such relative abundance estimates are confounded by the presence of multiple OTUs which can generate high-quality BLAST matches ("hits") to a given read. Multiple high-quality hits to a single read are the norm, rather than the exception, in metagenomic studies for several reasons. A microbial assemblage will often include a number of closely-related OTUs (e.g. congeners) which share large sections of highly similar or identical genomic sequence. If several such OTUs are present in the reference database, a metagenomic read from one will yield high-quality BLAST hits to them all. Further, even distantly related OTUs are likely to share large regions of identity, and the selection of hit quality thresholds to discriminate between them (for example, a minimum bit score or maximum expectation value) is effectively arbitrary. Thus, while metagenomic studies using whole-genome comparisons almost always use such thresholds as the sole discriminators between OTUs, this method (hereafter the "naïve" method, after Ye and Doak (2009)) will almost inevitably result in the identification of OTUs which are not present in the assemblage, skewing the relative abundance estimates of those which are truly present.

This problem is compounded by a systematic overrepresentation within full genome databases of of taxa of particular interest to humans, such as human and agricultural pathogens. Environmental OTUs are comparatively underrepresented. For example, Table 2.1 gives examples of terrestrial plant and animal pathogens, *a priori* unlikely to be truly present, which were identified in an open ocean metagenome with the naïve method.

One commonly used software tool to address this problem, Metagenome Analyzer (MEGAN), aggregates reads with hits to many OTUs to the most recent common ancestor of those OTUs, represented by a higher taxonomic rank e.g. family (Huson *et al.*, 2007). This approach increases the fidelity of the results, but comes at the cost of reduced taxonomic resolution. Particularly in marine assemblages where even fine genomic differences can

Table 2.1: Selected examples of OTUs identified in a marine metagenome using the naïve method. These OTUs were identified in a single sample from the SO (sample 346; see Chapter 3). The sample was compared to the RefSeq database of full genomes using TBLASTX with an E-value maximum of 1.0×10^{-3}, i.e. only high-quality hits were included. Relative abundances were calculated using GAAS (Angly *et al.*, 2009).

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

represent distinct ecological functions (e.g. Brown *et al.*, 2012), a tool which reduces spurious identifications without compromising taxonomic resolution would clearly be valuable.

2.2.2 The maximum parsimony approach

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

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

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

while the constraint was that every high-quality protein annotation had to be represented at least once in the annotated pathways. Validation and testing of MinPath showed it was successful in eliminating spuriously annotated pathways while retaining those genuinely present.

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

2.3 Methods

2.3.1 Implementation of MINSPEC

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

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

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

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

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

where $M_{ij} = 1$ if read i has a mapping (i.e. a high-quality BLAST hit) to OTU j, 0 if not, and $[1, n]$ is the set of all reads. In other words, the constraint function fails if any read does not have at least one of its high-quality BLAST hits represented in the assemblage.

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

It was noted that in some special cases, it may be desirable to include an OTU in the assemblage even if it is not part of the minimal set, if that OTU generated a very large number of BLAST hits. An example of such a situation might be if the sample was known with certainty to contain two very closely related OTUs at roughly equal abundance. In

[1] MINSPEC and the associated metagenomic simulation and validation scripts are open source and available at `https://github.com/wilkox/minspec`.

such a case, it would be expected that almost all metagenomic reads generated by each of these OTUs would also attract BLAST hits to the other, and MINSPEC would thus probably eliminate whichever happened to generate slightly fewer hits. To allow for this, an option was added to prevent MINSPEC from eliminating OTUs which attract a specified number of high-quality hits.

2.3.2 Validation of MINSPEC

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

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

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

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

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

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

The resulting simulated BLAST outputs were processed with MINSPEC, and the false positive (percentage of OTUs not in the assemblage which nevertheless survived MINSPEC filtering) and false negative (percentage of OTUs present in the assemblage which were not present after MINSPEC filtering) rates calculated. Because a high false negative rate can arise from undersampling, a problem in metagenomic studies both real and simulated, an additional "false negative (MINSPEC)" metric was calculated, which excluded OTUs which were present in the assemblage but through random chance did not generate any reads, the equivalent of "unsampled rare taxa". This rate thus represented only false

negatives attributable to MINSPEC itself. Finally, as a measure of MINSPEC's usefulness, the proportion of "false OTUs" — OTUs that generated BLAST matches but were not part of the assemblage — successfully removed by MINSPEC was calculated.

2.4 Results

Repeated simulated metagenomic experiments with a wide range of permutations of parameters showed that MINSPEC was reliable and able to substantially reduce the rate of false positive OTU identifications, although its effectiveness varied with the parameters of the assemblage and metagenomic experiment (Fig. 2.1).

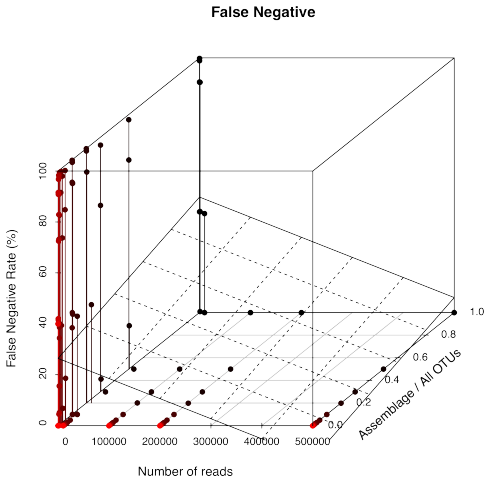
2.5 Discussion

The false negative rate, or percentage of OTUs in the simulated assemblage which were absent from the BLAST results following MINSPEC processing, was generally high, ranging from ∼20% under ideal conditions (a low assemblage / all OTUs ratio, and 500,000-read metagenomic sample) to ∼90% in the worst case (a high assemblage / all OTUs ratio and a small metagenomic sample) (Fig. 2.1a). The assemblage / all OTUs ratio (hereafter referred to as "assemblage ratio") indicates the proportion of simulated OTUs ("all OTUs") that were chosen to form the simulated assemblage. A higher ratio means that any OTU is more likely on average to be part of the assemblage, and thus that any individual failure to detect a OTU is an error. This problem is mitigated with increasing the number of reads, as this makes it less likely that a given OTU would go unsampled. The extreme false negative rates, in some cases 100%, represent extreme simulated scenarios (e.g. an assemblage of 1 OTU drawn from a pool of 100,000), and thus do not reflect real metagenomic studies.

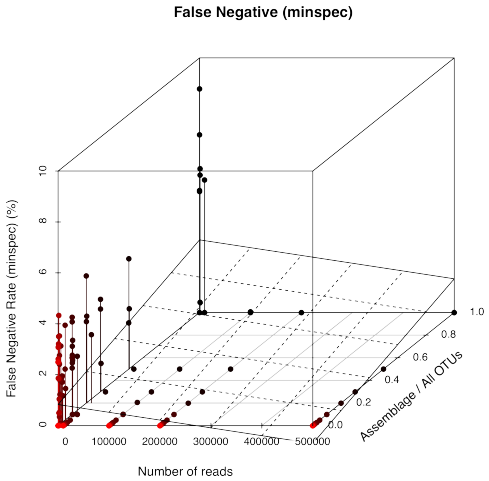
Because the majority of false negatives are attributable to undersampling and failure of OTUs to generate BLAST hits — properties the simulated metagenomic experiments share with real ones — a second metric, the false negative (MINSPEC) rate, was calculated (Fig. 2.1b). This is the proportion of OTUs in the assemblage that generated BLAST hits, but were incorrectly removed by MINSPEC. This rate thus represents error attributable only to MINSPEC. The false negative (MINSPEC) rate was generally low, ranging from ∼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 suggests MINSPEC is more affected by the degree of similarities between OTUs than by undersampling.

The false positive rate, or percentage of OTUs not in the assemblage which nevertheless generated high-quality BLAST matches that were not identified and removed by MINSPEC, was generally ∼0–5% except for extremely small read sets and low assemblage ratios, where it reached as high as 60% (Fig. 2.1c). 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.

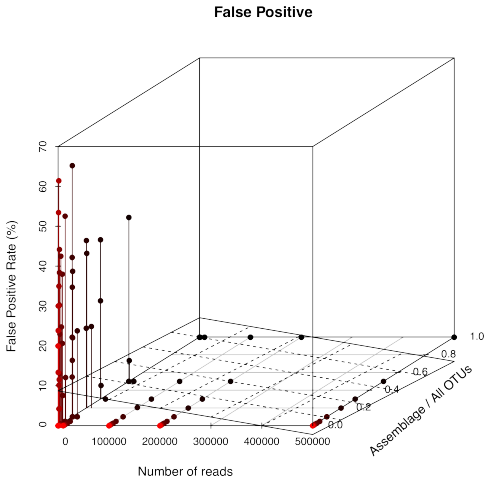
The proportion of false OTUs removed was calculated to measure MINSPEC's efficacy in identifying and eliminating OTU which are not part of the sampled assemblage yet generate high-quality BLAST matches. This rate varied from 0–1 depending on the param-



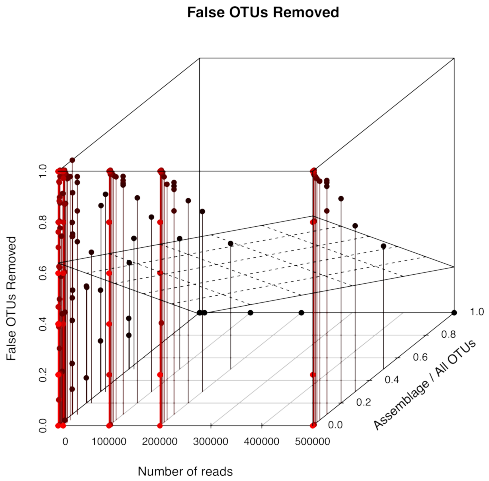
(a) False negative rate (%) — the percentage of OTUs in the assemblage that were absent from the BLAST results following MINSPEC processing.



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



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



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

Figure 2.1: Results of repeated trials of MINSPEC on simulated metagenomic studies with multiple permutations of parameters (number of reads, number of simulated OTUs, size of simulated assemblage). The number of simulated OTUs and size of simulated assemblage are represented as a ratio on the z-axis ("assemblage / all OTUs"). Each permutation was repeated five times. A plane representing a linear regression has been overlaid on each plot to indicate the trend. Points have been tinted to aid the perception of depth; colour is not otherwise meaningful.

eters of the assemblage (**??**). For simulations with a low assemblage ratio, the proportion was generally high (> 0.6), although there were simulated experiments with a low ratio where the proportion was low or zero. However, in all simulations with an assemblage ratio of 1, the proportion was 0, and the regression indicated a generally inverse relationship between the ratio and the proportion of false OTUs removed. This is likely because in assemblages with a higher assemblage ratio, there are fewer false OTUs to remove; in assemblages with a ratio of 1, there are none. The high proportion of false OTUs correctly identified in simulations with a low assemblage ratio is thus a good indication that MINSPEC is effective at identifying and removing false OTUs, especially as this proportion far exceeds the false positive and false negative (MINSPEC) rates for comparable experiments. As expected, increasing the number of reads improved MINSPEC's accuracy.

2.6 Conclusions

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 OTU 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.

At the time of writing, MINSPEC has been used in two published projects: Wilkins *et al.* (2013a) and Williams *et al.* (2013).

References

Abell G. C. J. and Bowman J. P. (2005). Colonization and community dynamics of class *Flavobacteria* on diatom detritus in experimental mesocosms based on Southern Ocean seawater. *FEMS Microbiology Ecology*, 53(3):379–391.

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.

Agogué H., Lamy D., Neal P. R., Sogin M. L., and Herndl G. J. (2011). Water mass-specificity of bacterial communities in the North Atlantic revealed by massively parallel sequencing. *Molecular Ecology*, 20(2):258–274.

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

Anderson L. A. and Sarmiento J. L. (1994). Redfield ratios of remineralization determined by nutrient data analysis. *Global Biogeochemical Cycles*, 8(1):65–80.

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

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

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

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

Baas Becking L. G. M. *Geobiologie Of Inleiding Tot De Milieukunde*. W.P. Van Stockum & Zoon, The Hague, 1934.

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

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

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

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

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

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

Blainey P. C. (2013). The future is now: single-cell genomics of bacteria and archaea. *FEMS Microbiology Reviews*, 37(3):407–427.

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

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

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

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

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

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

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

Brown M. V., Lauro F. M., DeMaere M. Z., Muir L., Wilkins D., Thomas T., Riddle M. J., Fuhrman J. A., Andrews-Pfannkoch C., Hoffman J. M., McQuaid J. B., Allen A., Rintoul S. R., and Cavicchioli R. (2012). Global biogeography of SAR11 marine bacteria. *Molecular systems biology*, 8:595.

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

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

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

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

Caporaso J. G., Kuczynski J., Stombaugh J., Bittinger K., Bushman F. D., Costello E. K., Fierer N., Pena A. G., Goodrich J. K., and Gordon J. I. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature methods*, 7(5):335–336.

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.

Caro-Quintero A. and Konstantinidis K. T. (2011). Bacterial species may exist, metagenomics reveal. *Environmental Microbiology*, 14(2):347–355.

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

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

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

Cho J.-C. and Tiedje J. M. (2000). Biogeography and degree of endemicity of fluorescent Pseudomonas strains in soil. *Applied and Environmental Microbiology*, 66(12):5448–5456.

Chou H. H. and Holmes M. H. (2001). DNA sequence quality trimming and vector removal. *Bioinformatics*, 17(12):1093–1104.

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

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

Clarke K. R. and Gorley R. N. *PRIMER v6: User Manual / Tutorial*, 1st edition edition, 2006.

Clarke K. R. and Warwick R. M. (1998). Quantifying structural redundancy in ecological communities. *Oecologia*, 113(2):278–289.

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

Coale K. H., Johnson K. S., Chavez F. P., Buesseler K. O., Barber R. T., Brzezinski M. A., Cochlan W. P., Millero F. J., Falkowski P. G., and Bauer J. E. (2004). Southern Ocean iron enrichment experiment: carbon cycling in high-and low-Si waters. *Science*, 304(5669): 408–414.

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

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

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

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

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

CSIRO . Autonomous, in situ microbial diagnostics and monitoring for microbial oceanography workshop 2012 summary, 2012. URL `http://www.csiro.au/en/Organisation-Structure/Divisions/Marine--Atmospheric-Research/HiResMicroOceanography-Hobart-workshop-2012/HiResMicroOceanography-Hobart-workshop-2012-summary.aspx`.

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.

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

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

Declerck S. A. J., Winter C., Shurin J. B., Suttle C. A., and Matthews B. (2013). Effects of patch connectivity and heterogeneity on metacommunity structure of planktonic bacteria and viruses. *The ISME Journal*, 7(3):533–542.

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

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

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

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

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

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

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

El-Sayed S. Z. (2005). History and evolution of primary productivity studies of the Southern Ocean. *Polar Biology*, 28(6):423–438.

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

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

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

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

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

Feng H., Cochran J. K., and Hirschberg D. J. (1999). ^{234}Th and ^{7}Be as tracers for the transport and dynamics of suspended particles in a partially mixed estuary. *Geochimica et Cosmochimica Acta*, 63(17):2487–2505.

Finlay B. J. (2002). Global Dispersal of Free-Living Microbial Eukaryote Species. *Science*, 296(5570):1061–1063.

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

Franck V. M., Brzezinski M. A., Coale K. H., and Nelson D. M. (2000). Iron and silicic acid concentrations regulate Si uptake north and south of the Polar Frontal Zone in the Pacific Sector of the Southern Ocean. *Current Opinion in Microbiology*, 47(15-16): 3315–3338.

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

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

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

Galand P. E., Potvin M., Casamayor E. O., and Lovejoy C. (2009). Hydrography shapes bacterial biogeography of the deep Arctic Ocean. *Nature*, 4(4):564–576.

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

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

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

Ghiglione J.-F., Galand P. E., Pommier T., Pedrós-Alió C., Maas E. W., Bakker K., Bertilson S., Kirchmanj D. L., Lovejoy C., Yager P. L., and Murray A. E. (2012). Pole-to-pole biogeography of surface and deep marine bacterial communities. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 109(43):17633–17638.

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

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

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

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

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

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

Goldenfeld N. and Woese C. (2007). Biology's next revolution. *Nature*, 445(7126):369–369.

González J. M., Fernández-Gómez B., Fernàndez-Guerra A., Gómez-Consarnau L., Sánchez O., Coll-Lladó M., Del Campo J., Escudero L., Rodríguez-Martínez R., Alonso-Sáez L., Latasa M., Paulsen I., Nedashkovskaya O., Lekunberri I., Pinhassi J., and Pedrós-Alió C. (2008). Genome analysis of the proteorhodopsin-containing marine bacterium *Polaribacter* sp. MED152 (Flavobacteria). *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 105(25):8724–8729.

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

Grossart H.-P., Kiørboe T., Tang K., and Ploug H. (2003). Bacterial colonization of particles: growth and interactions. *Applied and Environmental Microbiology*, 69(6):3500–3509.

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

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

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

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

Hales T. C. (2005). A Proof of the Kepler Conjecture. *Annals of Mathematics*, 162(3):1065–1185.

Hambly E. and Suttle C. A. (2005). The viriosphere, diversity, and genetic exchange within phage communities. *Current Opinion in Microbiology*, 8(4):444–450.

Hamdan L. J., Coffin R. B., Sikaroodi M., Greinert J., Treude T., and Gillevet P. M. (2013). Ocean currents shape the microbiome of Arctic marine sediments. *The ISME Journal*, 7 (4):685–696.

Hamilton A. K., Lovejoy C., Galand P. E., and Ingram R. G. (2008). Water masses and biogeography of picoeukaryote assemblages in a cold hydrographically complex system. *Limnology and Oceanography*, pages 922–935.

Hanson C. A., Fuhrman J. A., Horner-Devine M. C., and Martiny J. B. H. (2012). Beyond biogeographic patterns: processes shaping the microbial landscape. *Nature Reviews Microbiology*, 10(7):497–506.

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

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

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

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

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

Huang J., Su Z., and Xu Y. (2005). The evolution of microbial phosphonate degradative pathways. *Journal of Molecular Evolution*, 61(5):682–690.

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

Huntley M. E., Lopez M. D., and Karl D. M. (1991). Top predators in the Southern ocean: a major leak in the biological carbon pump. *Science*, 253(5015):64–66.

Huson D. H., Auch A. F., Qi J., and Schuster S. C. (2007). MEGAN analysis of metagenomic data. *Genome Research*, 17(3):377–386.

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

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

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

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

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

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

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

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

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

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

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

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

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

Knights D., Kuczynski J., Charlson E. S., Zaneveld J., Mozer M. C., Collman R. G., Bushman F. D., Knight R., and Kelley S. T. (2011). Bayesian community-wide culture-independent microbial source tracking. *Nature methods*, 8(9):761–763.

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

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

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

Lauro F. M., Chastain R. A., Blankenship L. E., Yayanos A. A., and Bartlett D. H. (2007). The unique 16S rRNA genes of piezophiles reflect both phylogeny and adaptation. *Applied and Environmental Microbiology*, 73(3):838–845.

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.

Legendre P. and Anderson M. J. (1999). Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs*, 69(1):1–24.

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

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

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

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

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

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

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

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

Martiny J. B. H., Bohannan B. J. M., Brown J. H., Colwell R. K., Fuhrman J. A., Green J. L., Horner-Devine M. C., Kane M., Krumins J. A., Kuske C. R., Morin P. J., Naeem S., Ovreas L., Reysenbach A.-L., Smith V. H., and Staley J. T. (2006). Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology*, 4(2):102–112.

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

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

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

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

Mazloff M. R., Heimbach P., and Wunsch C. (2010). An eddy-permitting Southern Ocean state estimate. *Journal of physical oceanography*, 40:880–899.

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

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

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

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

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

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

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

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

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

Moran M. A., Buchan A., González J. M., Heidelberg J. F., Whitman W. B., Kiene R. P., Henriksen J. R., King G. M., Belas R., Fuqua C., Brinkac L., Lewis M., Johri S., Weaver B., Pai G., Eisen J. A., Rahe E., Sheldon W. M., Ye W., Miller T. R., Carlton J., Rasko D. A., Paulsen I. T., Ren Q., Daugherty S. C., Deboy R. T., Dodson R. J., Durkin A. S.,

Madupu R., Nelson W. C., Sullivan S. A., Rosovitz M. J., Haft D. H., Selengut J., and Ward N. (2004). Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. *Nature*, 432(7019):910–913.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Paris C. B., Helgers J., van Sebille E., and Srinivasan A. (2013). Connectivity Modeling System: A probabilistic modeling tool for the multi-scale tracking of biotic and abiotic variability in the ocean. *Environmental Modelling and Software*, 42(C):47–54.

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

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

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

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

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

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

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

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

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

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

Qin J., Li R., Raes J., Arumugam M., Burgdorf K. S., Manichanh C., Nielsen T., Pons N., Levenez F., Yamada T., Mende D. R., Li J., Xu J., Li S., Li D., Cao J., Wang B., Liang H., Zheng H., Xie Y., Tap J., Lepage P., Bertalan M., Batto J.-M., Hansen T., Le Paslier D., Linneberg A., Nielsen H. B., Pelletier E., Renault P., Sicheritz-Ponten T., Turner K., Zhu H., Yu C., Li S., Jian M., Zhou Y., Li Y., Zhang X., Li S., Qin N., Yang H., Wang J., Brunak S., Doré J., Guarner F., Kristiansen K., Pedersen O., Parkhill J., Weissenbach J., MetaHIT Consortium , Bork P., Ehrlich S. D., and Wang J. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464(7285):59–65.

Quast C., Pruesse E., Yilmaz P., Gerken J., Schweer T., Yarza P., Peplies J., and Glöckner F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, 41(Database issue):D590–6.

Ramette A. and Tiedje J. M. (2007). Multiscale responses of microbial life to spatial distance and environmental heterogeneity in a patchy ecosystem. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 104(8):2761–2766.

Ramette A. and Tiedje J. M. (2006). Biogeography: An Emerging Cornerstone for Understanding Prokaryotic Diversity, Ecology, and Evolution. *Microbial Ecology*, 53(2):197–207.

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.

Rosenberg M. and Rintoul S. R. Aurora Australis Marine Science Cruise AU1203 – Oceanographic Field Measurements and Analysis. Technical report, 2012.

Rosenberg M. S. and Anderson C. D. (2011). PASSaGE: pattern analysis, spatial statistics and geographic exegesis. Version 2. *Methods in Ecology and Evolution*, 2(3):229–232.

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

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

Sauer K., Camper A. K., Ehrlich G. D., Costerton J. W., and Davies D. G. (2002). Pseudomonas aeruginosa displays multiple phenotypes during development as a biofilm. *Journal of Bacteriology*, 184(4):1140–1154.

Savoye N., Benitez-Nelson C. R., Burd A. B., and Cochran J. K. (2005). ^{234}Th sorption and export models in the water column: a review.

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

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

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

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

Siddiqui K. S., Williams T. J., Wilkins D., Yau S., Allen M. A., Brown M. V., Lauro F. M., and Cavicchioli R. (2013). Psychrophiles. *Annual Review of Earth and Planetary Sciences*, 41(1).

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

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

Sogin M. L., Morrison H. G., Huber J. A., Welch D. M., Huse S. M., Neal P. R., Arrieta J. M., and Herndl G. J. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 103(32):12115–12120.

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

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

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

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

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

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

Storch D. and Sizling A. L. (2008). The concept of taxon invariance in ecology: do diversity patterns vary with changes in taxonomic resolution? *Folia Geobotanica*, 43:329–344.

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

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

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

Sul W. J., Oliver T. A., Ducklow H. W., Amaral-Zettler L. A., and Sogin M. L. (2013). Marine bacteria exhibit a bipolar distribution. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 110(6):2342–2347.

Suttle C. A. (2005). Viruses in the sea. *Nature*, 437(7057):356–361.

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

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

Tamura T., Williams G. D., Fraser A. D., and Ohshima K. I. (2012). Potential regime shift in decreased sea ice production after the Mertz Glacier calving. *Nature Communications*, 3:826–.

Temperton B. and Giovannoni S. J. (2012). Metagenomics: microbial diversity through a scratched lens. *Current Opinion in Microbiology*, 15(5):605–612.

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

Tettelin H., Masignani V., Cieslewicz M. J., Donati C., Medini D., Ward N. L., Angiuoli S. V., Crabtree J., Jones A. L., Durkin A. S., Deboy R. T., Davidsen T. M., Mora M., Scarselli M., Margarit y Ros I., Peterson J. D., Hauser C. R., Sundaram J. P., Nelson W. C., Madupu R., Brinkac L. M., Dodson R. J., Rosovitz M. J., Sullivan S. A., Daugherty S. C., Haft D. H., Selengut J., Gwinn M. L., Zhou L., Zafar N., Khouri H., Radune D., Dimitrov G., Watkins K., O'Connor K. J. B., Smith S., Utterback T. R., White O., Rubens C. E., Grandi G., Madoff L. C., Kasper D. L., Telford J. L., Wessels M. R., Rappuoli R., and Fraser C. M. (2005). Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial "pan-genome". *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 102(39):13950–13955.

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

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

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

Tréguer P., Nelson D. M., Van Bennekom A. J., DeMaster D. J., Leynaert A., and Quéquiner B. (1995). The silica balance in the world ocean: a reestimate. *Science*, 268(5209):375–379.

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.

van Sebille E., Johns W. E., and Beal L. M. (2012). Does the vorticity flux from Agulhas rings control the zonal pathway of NADW across the South Atlantic? *Journal of Geophysical Research*, 117(C5):C05037.

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

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

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

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

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

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

Waterbury J. B., Willey J. M., Franks D. G., Valois F. W., and Watson S. W. (1985). A cyanobacterium capable of swimming motility. *Science*, 230(4721):74–76.

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

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

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

Whitaker R. J., Grogan D. W., and Taylor J. W. (2003). Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science*, 301(5635):976–978.

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

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

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

Wilkins D. Biogeographic partitioning of Southern Ocean picoplankton. In *ISME 13*, Copenhagen, 2012.

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. (2013). Biogeographic partitioning of Southern Ocean microorganisms revealed by metagenomics. *Environmental Microbiology*, 15(5):1318–1333.

Wilkins D., van Sebille E., Rintoul S. R., Lauro F. M., and Cavicchioli R. (2013). Advection shapes Southern Ocean microbial assemblages independent of distance and environment effects. *Submitted.*

Wilkins D., Yau S., Williams T. J., Allen M. A., Brown M. V., DeMaere M. Z., Lauro F. M., and Cavicchioli R. (2013). Key microbial drivers in Antarctic aquatic environments. *FEMS Microbiology Reviews*, 37(3):303–335.

Williams G. D., Bindoff N. L., Marsland S. J., and Rintoul S. R. (2008). Formation and export of dense shelf water from the Adélie Depression, East Antarctica. *Journal of Geophysical Research*, 113(C4):C04039.

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. (2013). The role of planktonic Flavobacteria in processing algal organic matter in coastal East Antarctica revealed using metagenomics and metaproteomics. *Environmental Microbiology*, 15(5):1302–1317.

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

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

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

Youssef N., Sheik C. S., Krumholz L. R., Najar F. Z., Roe B. A., and Elshahed M. S. (2009). Comparison of species richness estimates obtained using nearly complete fragments and simulated pyrosequencing-generated fragments in 16S rRNA gene-based environmental surveys. *Applied and Environmental Microbiology*.

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

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