Understanding globally important biogeochemical processes from a study of a hypersaline Antarctic lake

Yau S, Lauro FM, DeMaere MZ, Brown MV, ???, Cavicchioli R

# Potential Aims and Scope

Microbial population and community ecology, Integrated genomics and post-genomics approaches in microbial ecology, **Microbial ecology and functional diversity of natural habitats**. ORIGINAL ARTICLE 5 000 word limit. TITLE: 50 characters.

# Abstract (250 word limit)

Organic Lake is a shallow (6.75 m deep) marine-derived hypersaline lake in the Vestfold Hills, East Antarctica with a high concentration of the cloud-forming gas dimethylsulfide (DMS). During sampling, it was vertically stratified into an aerobic mixed zone and a suboxic deep zone, with a peak of C, S and ammonia below the oxycline (6.5 m). Environmental DNA from size fractionated samples (3.0, 0.8 and 0.1 µm) along the depth profile was sequenced and the taxonomic composition and functional diversity determined. The microbial composition resembles other hypersaline and cold environments indicating a strong selection for species. Eucaryotic phytoflagellates related to *Dunaliella* and *Pseudopedinella* are the dominant photosynthetic organisms and presumably contribute the bulk of primary production.Bacterioplankton throughout the water column was dominated by *Marinobacter*, *Roseovarius* and *Psychroflexus* relatives*.* Candidate division RF3, *Halomonas* and *Psychromonas* were overrepresented at 6.5 m and associated with high potential for fermentation of particulate matter and amino acids. The bottom samples were abundant in candidate divisions OD1 and TM7 and were similarly linked to anaerobic processes. A diverse set of functional genes were assigned to the *Marinobacter* and *Roseovarius* clades including rhodopsin, DMSP lyase (*dddD*, *dddL* and *dddP*),Calvin cycle, anaerobic respiration and CO oxidation genes indicating they are metabolically versatile generalists that may contribute to primary production. This study has allowed a rigorous description of the microbial community within a natural environment and sheds light on globally relevant biogeogemical processes such as DMS generation, lithoheterotrophy, chemolithoautotrophy and photoheterotrophy. It shows potential for diverse strategies in may be the rule, not the exception for marine-derived surface water *Gammaproteobacteria* and *Roseobacter* with implications untangling microbial roles in production.

# Introduction

Life in the Antarctic is constrained by extremes of temperature and salinity under a polar light cycle. In the frozen desert, ice-free regions containing liquid water in lakes and ponds are rare oases for life. The Vestfold Hills, located on the eastern shore of the Prydz Bay, East Antarctica (Figure S1), is one such region where hundreds of lakes are found. The lakes were formed from seawater, trapped approximately 10 000 BP when the continental ice-sheet receded from the coast and the land rose above sea-level (Zwartz *et al*., 1998; Gibson, 1999). Life in these lakes can be entirely microbial and of reduced diversity (Bowman *et al.*, 2000b). Differing local conditions has lead each lake to develop unique physical and chemical properties, making them fitting sites to study microbial ecology, biogeochemistry and evolution. The Vestfold Hills contains the highest density of meromictic (permanently stratified) water bodies in Antarctica (Gibson, 1999). They are advantageous study sites as environmental gradients exist within a single, largely closed system allowing species to be more easily related to abiotic factors. By using molecular techniques, a large proportion of the species diversity and gene content can be covered allowing better inference of the functional roles for the taxa present (Laybourn-Parry & Pearce, 2007).

A metagenomic approach, complemented with metaproteomics, has been successfully applied to two lakes in the Vestfold Hills (Ng *et al.*, 2010; Lauro *et al.*,2011; Yau *et al.*, 2011). The first of these was Ace Lake, where a comprehensive description of the community structure, biogeochemical fluxes and responses to resource limitation was achieved (Lauro *et al.*, 2011). The metabolism of the abundant green sulfur bacteria (Ng *et al.*, 2010)was found to play a central role in nutrient cycling and a mathematical model was developed that showed its dominance was dependent on synchronicity with the polar light cycle leading to absence of phage predation (Lauro *et al*., 2011). In the surface water of the second lake, Organic Lake, a member of the virophage virus family was discovered that potentially regulates microbial loop dynamics (Yau *et al*., 2011). Virophage require a helper virus to replicate but are detrimental to their helper (La Scola *et al.*, 2008). The Organic Lake virophage (OLV) likely depends on phycodnaviruses, which infect eucaryotic algae. The presence of OLV would reduce infective phycodnaviruses leading to increased algal blooms and thus carbon flux (Yau *et al.*, 2011). These studies have achieved exceptional insight into Antarctic lakes but are also relevant to other aquatic systems. For example, OLV-like sequences were found in coastal marine, hypersaline and freshwater metagenomes indicating virophage have a wider ecological role (Yau *et al.*, 2011).

This study extends the previous metagenomic analysis of the Organic Lake surface water to examine the entire microbial community along a depth profile from a whole ecosystem perspective. The bottom waters of Organic Lake are unusual due to the high concentration of the volatile gas dimethylsulfide (DMS) (Deprez *et al*., 1986;Franzmann *et al.*, 1987; Gibson *et al.*, 1991; Roberts & Burton 1993a; Roberts *et al.*, 1993b) as well as polysulfides (Roberts & Burton 1993a; Roberts *et al.*, 1993b). Aerosols derived from atmospheric DMS act as cloud condensation nuclei and are hypothesized to affect climate (Charlson *et al.*, 1987). Although the importance of DMS was proposed forty years ago (Lovelock & Maggs, 1972) the first enzymes involved in DMS production were only identified in the last five years (Todd *et al.*, 2007). Concentrations of DMS as high as 5000 nM have been recorded in Organic Lake (Gibson *et al*., 1991), 100 times the maximum concentration recorded from seawater in the adjacent Prydz Bay and at least 1000 times that of the open Southern Ocean (Curran & Jones, 1998). This makes it an ideal location to identify the microbes involved and potentially the basis for DMS accumulation. The bottom waters of Organic Lake were found to be anoxic, but not sulfidic or methanogenic (Franzmann *et al.*, 1987b; Gibson *et al.*, 1991). Although sulfates and organic acids have been recorded (Franzmann *et al.*, 1987b; Gibson *et al.*, 1994), the cold and salinity six times that of seawater, appears to preclude the establishment of sulfate-reducing bacteria, and thus phototrophic sulfur bacteria (Burke & Burton, 1988) indicating other microbes are involved in the unusual sulfur chemistry. This study sought to determine the composition and the functional potential of Organic Lake microbiota and, in conjunction with historic and contemporary physico-chemical data, generate and an integrative understanding of the whole lake ecosystem.

# Materials and Methods

## Sample collection and preparation

Water was collected from Organic Lake on 10 November 2008 (68º27'22.15"S, 78º11'23.95"E) through a 30 cm hole in the 0.8 m thick ice cover above the deepest point in the lake. Samples were collected at 1.7, 4.2, 5.7, 6.5 and 6.7 m depths. Lake water was passed through a 20 µm pore size pre-filter, and then microbial biomass was captured by sequential filtration onto 3.0 µm, 0.8 µm and 0.1 µm pore size membrane filters. Between 1–2 L of lake water was sufficient to clog the filters. DNA was extracted from the filters as previously described (Ng *et al.*, 2010; Lauro *et al.*,2011). DNA from all samples was sequenced using the Roche GS-FLX titanium sequencer. Reads were processed to remove low quality bases, assembled and annotated as previously described (Lauro *et al.*, 2011). Water was also collected for microscopic and chemical analysis at the same sample depths and frozen −80ºC.

## Physical and chemical analyses

An *in situ* profile of pH, conductivity, turbidity, dissolved oxygen (DO) and pressure was measured using a submersible probe (YSI sonde model V6600). A temperature profile was measured using a minimum-maximum mercury thermometer. The 5.7 m sample corresponded to the turbidity maximum and the 6.5 m sample to the turbidity minimum. Conductivity at *in situ* temperature was converted to conductivity at 15ºC according to the relation described by Gibson (1999). The adjusted conductivity brings the temperature within the acceptable range to estimate practical salinity by the formula of Fofonoff and Millard (1983). However, salinity was likely underestimated as Organic Lake salinity is higher than the practical salinity range of 2–42 for which the conductivity to salinity relation holds. Density was calculated from the *in situ* conductivity and temperature using the equations described by Gibson *et al.* (1990) and expressed at temperature T as:

σT = (1000–density) kg/m3

Ammonia, nitrate, nitrite, total nitrogen (TN), total dissolved nitrogen (TDN), dissolved reactive phosphorus (DRP), total phosphorus (TP), total dissolved phosphorus (TDP), total organic carbon (TOC), total dissolved carbon (DOC), total sulfur (TS) and total dissolved sulfur (TDS) were determined by American Public Health Associations Standard Methods at the Analytical Services Tasmania. Values for dissolved nutrients and inorganic N were measured after filtration through 0.1 µm pore size membrane. All other nutrients were measured from water collected after filtration through 20 µm pore size membrane. Ammonia, nitrate, nitrite, DRP, TN, TDN, TP and TDP were measured in a Flow Injection Analyser (Lachat Instruments, Colorado, USA). TOC and DOC were determined in the San++ Segmented Flow Analyser (Skalar, Breda, Netherlands). TS and TDS were analyzed in the 730ES Inductively Coupled Plasma – Atomic Emission Spectrometer (Agilent Technologies, California, USA).

Principal Component Analysis (PCA) was performed using the PRIMER Version 6 statistical package (Clarke & Gorley, 2006) on the normalized physical and chemical parameters to visualize how abiotic factors varied with depth. Inorganic N and dissolved nutrients were not included in the PCA analysis as the values were missing for those variables at 4.2 m, but PCA performed excluding 4.2 m sample and including those parameters showed similar separation of samples.

## Epifluorescence microscopy

Water samples collected for microscopy were preserved in formaldehyde (1% v/v). Cells and virus-like particles (VLPs) were vacuum filtered onto 25 mm polycarbonate 0.015 µm pore-size membrane filters (Nuclepore Track-etched, Whatman, GE Healthcare, USA) with a 0.45 µm pore-size backing filter. The 0.015 µm filter was mounted onto a glass slide with ProLong® Gold anti fade reagent (Invitrogen, Life Technologies, NY, USA) and 2 µl (25 × dilution in sterile filtered milliQ water <0.015 µm) SYBR® Gold nucleic acid stain (Invitrogen, Life Technologies, NY, USA). Prepared slides were visualized in an epifluorescence microscope (Olympus BX61, Hamburg, Germany) under excitation with blue light (460–495 nm, emission 510–550 nm). Cell and VLP counts were performed on the same filter over 30 random fields of view.

## Biological diversity analyses

### Cellular diversity

Diversity of cellular life was assessed using ribosomal small subunit (SSU) gene sequences. Metagenomic reads that matched the 16S and 18S rRNA gene were retrieved using Metaxa (Bengtsson *et al.*, 2011). This software implements hidden markov model based searches to retrieve 12S/16S/18S sequences and trims off regions outside of the SSU gene. Only sequences longer than 200 bp were accepted for downstream analysis. The Quantitative Insights Into Microbial Ecology (QIIME) pipeline (version 1.4.0) (Caporaso *et al*., 2010) implementing UCLUST, was used to group SSU sequences into operational taxonomic units (OTUs) at 97% percent identity against the SILVA SSU reference database (release 108). SSU sequences that did not cluster with sequences from SILVA were allowed to form new OTUs (no suppression). QIIME was then used to choose a representative sequence from each OTU and classify the representative set to the \*genus level using the RDP classifier (Wang *et al*., 2007) trained against SILVA (release 108) sequences (www.arb-silva.de). Assignments were accepted to the highest taxonomic rank with bootstrap value ≥85%. This prevented low confidence matches contributing to counts of high-confidence phylogenetic groups while avoiding grouping all the unclassified taxa together. \*QIIME was used to calculate alpha diversity indices: Chao1, Simpson, Shannon and observed species.

To allow comparison of the relative abundance of taxa between samples, the number of SSU matches per sample filter was normalised to the average number of reads (403 577) obtained for each sample filter. Statistical analysis on the relative SSU gene abundances was performed using the PRIMER Version.6 package (Clarke & Gorley, 2006). The SSU gene counts of each sample filter were square root transformed to reduce the contribution of highly abundant taxa. The Bray-Curtis similarity of the community composition from each sample was computed. Patterns in the resulting similarity matrix were visualized using hierarchical clustering (CLUSTER) and non-parametric Multidimensional Scaling (MDS) routines (Clarke, 1993). Statistical significance of the clusters was determined by the ‘similarity profile’ (SIMPROF) permutation test. To determine if physical and chemical parameters and the patterns in cellular composition were correlated, BEST analysis was performed considering following abiotic variables: conductivity, temperature, turbidity, DO, pH, TOC, TN, TP, TS, total C:N, total C:P, total N:P, cell counts and VLP counts. The Bio-Env procedure in BEST looks at all the abiotic variables in combination and finds a subset sufficient to best explain the biotic structure. A heat map with biclustering dendogram was generated using R and the package ‘seriation’ (Hahsler *et al*., 2008) of the SSU composition.

### Viral diversity

## Analysis of Functional potential

Open reading frames (ORFs) were predicted from quality trimmed metagenomic reads using MetaGene (Noguchi *et al*., 2006). Those ORFs longer than 90 bp were selected for downstream analyses. ORFs were translated into amino acid sequences using the standard bacterial/plastid translation table. Translated ORFs were compared to protein sequences from the Kyoto Encyclopedia of Genes and Genomes (KEGG) GENES database (release 58) using the Basic Local Alignment Search Tool (BLAST) (Altschul *et al.*, 1990). KEGG GENES is a collection of genes from all complete genomes from public resources, primarily NCBI RefSeq. The BLAST output was processed using KEGG Orthology Based Annotation System (KOBAS) version 2.0 (Xie *et al.*, 2011) accepting assignments to KEGG Orthologs with expectation value below 1e−05 and rank greater than 5. Assignments from each sample to KEGG orthologs that matched to marker enzymes in the carbon, nitrogen and sulfur cycles were counted. Normalized frequencies of enzymes from the same pathway were averaged. Genetic potential for chemical conversion via different pathways were summed.

Marker genes that were not well represented by KEGG orthologs were retrieved via alternative strategies depending on their representation in sequence databases. Organic Lake rhodopsin homologs were retrieved if they had a top BLAST match to any in a list of 139 entries in the KEGG GENES database affiliated with bacteriorhodopsin, xanthorhodopsin, halorhodopsin or proteorhodopsin. The DMSP lyases were retrieved from National Center for Biotechnology Information (NCBI) ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) sequence databases. Sequences with experimentally confirmed function were used to query a BLAST database of the translated ORFs predicted from the Organic Lake metagenomic reads (\*table: functional\_genes). Matches were examined if e-value was <1e−10 and accepted if the sequence identity was within the range for related enzymes that putatively had the same function as the query sequence.

## Phylogenetic analyses

Marker gene sequences for phylogenetic analysis were clustered using CD-HIT (\*ref) at 90% global amino acid identity. The longest sequence of the CD-HIT cluster was used as the representative sequence in a BLASTp query against the NCBI non-redundant (NR) database to retrieve full-length sequenced homologs from bacterial isolates. They were included in phylogenetic analysis along with the Organic Lake representative sequences that resided within a desired conserved region, for example the spectral tuning motif of rhodopsin. Phylogenetic analyses were performed in MEGA 5.05 (Tamura *et al.* 2011). Sequences were aligned with MUSCLE (Robert, 2004) using default parameters (gap opening penalty: -2.9, gap extension penalty: 0). Neighbor-joining was used to compute the phylogenies with Poisson substitution model, uniform rates of change and complete deletion of alignment gaps. Node support was tested with bootstrap analysis (500 replicates).

# Results and Discussion

## Abiotic properties and water column structure of Organic Lake

During sampling on November 2008, Organic Lake had a maximum depth of 6.75 m and the surface measured 3.87 m above mean sea level. *In situ* physico-chemical profiles (Figure S2) were measured over the deepest point in the lake (Figure S3) to evaluate the water column properties and structure. Two distinct zones were apparent: a mixed zone above 5.7 m and a suboxic deep zone below (Figure 1A). The separation of these two zones was indicated by a pycnocline starting at 5.7 m. The presence of an oxycline at the same depth indicates depletion of DO due to respiration in the stagnant waters. The pH also decreased with DO, likely due to accumulated fermentation products such as acetic, formic and lactic acids that have been previously recorded in the bottom waters (Franzmann *et al*., 1987b; Gibson *et al*., 1994). The deep zone was not completely anoxic as has been recorded in the past (Franzmann *et al.*, 1987b; Gibson *et al.*, 1991) indicating DO had entered the bottom waters in the last 13 years. Oxygen may be episodically introduced with currents of cold dense water that are generated during ice-formation in the lake shallows (Ferris *et al*., 1999). Samples were collected from the mixed (1.7, 4.2 and 5.7 m) and deep (6.5 m and 6.7 m) zones to determine their nutrient content and microbiology.

All nutrients, except for nitrate and nitrite, reached maximum concentrations at 6.5 m (Table 1) suggestive of a layer of high biological activity above the lake bottom (Figure 1B). Consistent with this, cell and VLP counts were highest at 6.5 m. However, turbidity was lowest at this depth demonstrating turbidity was not principally determined by cell density. Microscopy images do not show a shift in cell morphology that could account for the large drop in turbidity (Figure S3), which suggests particulate matter primarily contributed to turbidity readings. (\*Trophic status. Check how TOC compares to other lakes/ocean). The low turbidity and peak in cell counts and nutrients at 6.5 m suggest increased degradation of particulate matter by the microbial community via processes enriched in the microaerophilic environment. This is supported by the high concentrations of dissolved organic and free amino acids found in the deep zone (Gibson *et al.*, 1994) indicative of breakdown of high molecular weight compounds such as carbohydrates, lipids and proteins.

The C:N and C:P ratios were high compared to the Redfield ratio (Redfield *et al.*, 1963; \*others?) except at 6.5 m indicating this was the only depth where N and P were not relatively limited (Table 1). PCA analysis of physico-chemical parameters showed all samples except the 6.5 m sample separated with depth along the PC1 axis (Figure S4). Accordingly, turbidity, TS and cell density were the strongest explanatory variables for the separation of the 6.5 m sample from the other deep sample. (\*what does this mean?)

## Organic Lake microbial community composition and distribution

### Overall cellular diversity

Metagenomic sequences were obtained from size fractionated (3.0 µm, 0.8 µm and 0.1 µm) microbial biomass from each sample depth (see Table S1 for summary of metagenomic data). To determine the microbial composition, a total of 3 959 reads matching to the SSU gene were retrieved from the metagenomic sequences. These grouped into 983 OTUs. *Bacteria* were numerically dominant comprising 76.2% of all SSU sequences. 16.3% of sequences were assigned as *Eucarya* and 7.5% of SSU sequences could not be classified (\*what are these?). Only 2 reads, assigned to a deep sea hydrothermal clade of *Halobacteriales*, were classified as *Archaea* (data not shown) indicating they were rare in Organic Lake. Proportions of SSU genes may not necessarily reflect the number of cells in the environment because of potential SSU copy number, DNA extraction and sequencing biases. In terms of error from copy number, archaeal SSU gene copies only range from 1–4 (Lee *et al.*, 2009) and nearest sequenced relatives of Bacteria present in this study ranges from 1–6 (data not shown). Thus, it is expected *Archaea* were truly scarce and estimates for bacterial abundance are accurate within this margin of error. (\*mention GAAS here\*recA comparison).

Overall microbial diversity was fairly low, with 15 bacterial phyla and 6 eucaryotic superkingdom divisions in total. Of these, only 7 bacterial phyla and 4 eucaryotic phyla were predominant. (\*diversity indices). Bacterial and eucaryotic classes were generally represented by a single dominant genus (Table S2) indicating there is little intra-division complexity.

### Selective pressure for psychrophilic and halophilic *Eucarya* and *Bacteria*

Details of the composition of each phylum are shown in Table S2 (\*Do euk). Three bacterial classes, *Gammaproteobacteria*, *Alphaproteobacteria* and *Flavobacteria*, were the most abundant and were found on all filter sizes at all depths (Figure 2A). Each of these three classes consisted of one dominant genus (at least 64% of sequences from that class) which was *Marinobacter*, *Roseovarius* and *Psychroflexus* respectively (Table S2). Moderately abundant bacterial phyla were *Actinobacteria* and candidate divisions OD1 and RF3. Lower abundance clades included the *Spirochaetes*, *Lentisphaera*, TM7, *Verrucomicrobia*, Bhi80-139, Bd1-5, SR1 and *Chlamydiae* (Figure 2A).

The dominant *Eucarya* were chlorophyte (green algae) anddictyochophyte (silicoflagellate) algae, which had the same distribution as chloroplasts (Figure 2B). Chlorophytes were principally of the genus *Dunaliella* and dictyochophytes were of the order *Pedinellales*, related to *Pseudopedinella.* Lower abundance Eucarya included *Bacillariophyta* (diatoms), *Dinophyceae* (dinoflagellates), *Fungi* and heterotrophic choanoflagellates. *Bacillariophyta* were related to *Chaetoceros* and would contribute to primary production (\*ref Donna’s paper about diatom distribution). The dinoflagellates? Choanoflagellates have been described in Organic Lake and was the first description of a choanoflagellate in a hypersaline environment (\*Van den hoff).

Cultured relatives of taxa detected in Organic Lake are known to be halophilic and/or psychrophilic (Gauthier *et al.*, 1992; Dobson *et al.*, 1991; Labrenz *et al*., 1999\*\*\*\*). Similarly, uncultured Organic Lake taxa have highest identity to SSU sequences from saline and/or cold environments (\*OD1 Mosier *et al*., 2007; RF3 Demergasso *et al.*, 2010; Aquiluna\*Kang *et al*., 2012) (Table S2). The consistent association with phylotypes from similar environments indicates selection for particular taxonomic groups in Organic Lake. This is further supported by the persistence of the same dominant taxa in Organic Lake over time such as *Dunaliella*, *Psychroflexus*,(Franzmann *et al*., 1987b)  *Marinobacter*, *Halomonas* and *Roseovarius* (Bowman *et al*., 2000b). (\*Persistence may be due to isolation, not selection. But proportions change over time and those members are not excluded)(\*Note that much greater depth of diversity was detected than compared to Bowman in the sediments)

Compared to other Antarctic lakes, the 16S composition is most like that of Ekho Lake, in the Vestfold Hills (Bowman *et al.*, 2000b) and Lake Bonney in the McMurdo Dry Valleys (Glatz *et al*., 2006) that are of similar salinity (15–18%). It must be noted that composition was determined in those studies by PCR amplification of 16S rRNA gene and so can only be compared in terms of broad divisions. These lakes are characterized by the apparent lack of *Cyanobacteria* or phototrophic sulfur bacteria. Antarctic *Synechococcus* relatives appear to only inhabit lakes of close to marine salinity (Powell *et al.*, 2005). Similarly the upper salinity limit for the most halophilic of the phototrophic sulfur bacteria *Rhodospirillaceae* is 15% (Burke & Burton, 1988). The near-absence of *Archaea* is in contrast to Deep Lake, in the Vestfold Hills that has salinity of 35% and is populated almost solely by haloarchaea (Bowman *et al.*, 2000b). These findings correspond well to a recent study of solar salterns along a salinity gradient that showed haloarchaea only dominated above a salinity of 19% (Ghai *et al.*, 2011). The Organic Lake community composition appears reside at the point in a continuum of salinity where entire groups of microorganisms are excluded such as obligate photolithoautotrophic bacteria and the “hyperhalophilic” *Archaea*.

(\*What about the eucaryotic populations? Is there selection on salinity?)

### Distribution of *Eucarya* and *Bacteria* varied according to size and depth

Seriation analysis showed the cellular community composition clustered according to size fraction and depth (Figure 3) and identified taxa that were differentially distributed between, or within the zones. A significant difference in genus-level cellular composition between mixed and deep zone samples was supported by ANOSIM analysis (Rho: 0.53, significance: 0.1%). This shows that most taxonomic groups were more prevalent in a particular zone in Organic Lake with some restricted to a specific niche.

Most of the bacterial lineages with cultured relatives are known to be heterotrophic aerobes (Dobson *et al.*, 1991; Gauthier *et al.*, 1992; Labrenz *et al*., 1999; Hahn *et al.*, 2004; \*). Their predominance implies the suboxic environment precludes the establishment of high numbers of strictly anaerobic bacteria. Only low numbers of strictly anaerobic bacteria were detected including *Clostridia* (primarily *Halanaerobium*) and sulfate-reducing *Deltaproteobacteria*. Known facultative anaerobes included sulfur oxidizing *Epsilonproteobacteria* that may be chemolithoautotrophic. Clearly, if the deep zone of Organic Lake is episodically oxygenated, anaerobes must have some degree of aerotolerance or form spores to endure these events. (\*check amino acid utilization).

#### 20–3.0 µm fraction

The mixed zone samples had a relatively high abundance of *Dunaliella* chloroplasts and chlorophyte algae consistent with large phototrophic cells concentrating near surface light. *Dunaliella* have been previously isolated from Organic Lake and were reported to be the dominant eucaryotic alga (Franzman *et al.*, 1987b). Signatures of algae found at the bottom of the lake are likely due to sedimentation of dead cells or resting cysts as the biflagellated adult cells would be able to control their location in the water column. (\*genome by JGI, tarchive files available).

*Psychroflexus* were enriched on the 3.0 µm samples (Figure 3), although they were also present on the smaller filter sizes (Figure 2). *Psychroflexus gondwanensis* (ACAM 44) (previously *Flavobacterium*), along with several other related Flavobacteria strains, have been isolated from Organic Lake (Franzmann *et al*., 1987b). *P gondwanensis* isolates range in length from approximately 1.5–11.5 µm (Dobson *et al*., 1991), which would account for their enrichment on the 3.0 µm size fraction. *Psychroflexus* from the 3.0 µm fraction was more abundant in the surface and the 6.7 m sample. *Flavobacteria* have been associated with phytoplankton blooms in the Southern Ocean (Abell & Bowman 2005a; Abell & Bowman 2005b), which is hypothesized to be related to their ability to degrade high molecular weight carbon from algal exudates and detritus (reviewed in Kirchman *et al.*, 2009). Likely, Organic Lake *Psychroflexus* fills a similar ecological role. In support of this, *Psychroflexus* clusters with *Dunaliella* chloroplasts in the seriation analysis (Figure 3) and *P. gondwanese* abundance in Organic Lake correlates with average hours of sunshine per day (James *et al.*, 1994). Furthermore, cultured *P. gondwanense* cannot utilize a wide range of labile substrates such as amino acids or monosaccharides but can degrade starch and DNA (Dobson *et al*., 1993). Its presence in the deep zone could also be due to sedimentation as *P. gondwanense* is non-motile and strictly aerobic (Dobson *et al.*, 1991).

*Roseovarius* was also principally found on the 3.0 µm filter and was enriched at 6.5m and 4.2 m suggesting surface and deep adapted types. The type species *R. tolerans*, an isolate from the hypersaline Antarctic Ekho Lake and a close relative of Organic Lake *Roseovarius*,has a large cell size (1.1–2.2 μm long) (Labrenz *et al*., 1999) accounting for its accumulation on this size fraction. The abundance of *Roseovarius* in the deep zone is unexpected as, *R. tolerans* (\*other isolates) is strictly aerobic (Labrenz *et al.*, 1999). *Roseovarius* counts were not highest at the bottom sample where they would be expected to accrue from sedimentation, but rather at 6.5 m indicating metabolic capacity to grow in the microaerophilic environment. *Roseovarius* is a member of the *Roseobacter* clade whose diverse metabolic capabilities include DMSP degradation, aerobic anoxygenic phototrophy (AAnP) and symbiotic relationships with dinoflagellates (reviewed in Wagner-Döbler & Biebl, 2006). Mixed zone *Roseovarius* could conceivably perform any or all of these roles in Organic Lake. *R. tolerans* can produce bacteriochlorophyll *A* (Bchla), however, production is suppressed in the light or constant dim conditions (Labrenz *et al.*, 1999) so whether AAnP could in the polar summer depends on the longevity of Bchla in the cell(\*).

(\*check bacteriochlorophyll A from Roseovarius. likely none is being produced expressed in this sample. Look for the pufLM genes which form part of the reaction centre, not the harvesting complex. Some Roseobacters have pufLM but no Bchl expressed as Bchl expression is dependent upon environmental conditions).

#### 3–0.8 µm size fraction

*Marinobacter* dominated the 0.8 µm size fraction at all depths except 6.5 m. The concentration of *Marinobacter* on this size fraction is consistent with the cell size of isolates (\*ref) reflective of planktonic cells. They are known as aerobic heterotrophs originally isolated on hydrocarbons (\*ref), and generally prefer labile substrates such as sugars, amino acids and organic acids (\*ref). Other isolates are capable of interacting with marine algae (\*ref marinobacter adhaerans) and oxidation of manganese (\*ref Marinobacter manganoxydans MnI7-9) and arsenate (M santoriniensis sp.) making them metabolically versatile. *Marinobacter* are ubiquitous in the marine environment (\*ref) but appear to be enriched in several hypersaline Antarctic lakes due to their halotolerance (Bowman *et al.*, 2000b; Naganuma *et al.*, 2005; Glatz *et al*., 2006;\*). *Marinobacter* isolates from Antarctic lakes are capable of anaerobic respiration using dimethyl sulfoxide (DMSO) (Matsuzaki *et al*., 2006), nitrate (\*ref) which would allow for their presence throughout the water column (\*see below).

The related *Saccharospirillum*, like the name suggests degrades polysaccharides (\*ref) and is likely associated with algal blooms (\*ref).

By contrast, RF3, *Halomonas* and *Psychromonas* were concentrated on the 6.5 m sample and are the most likely candidates for mediating processes confined to that depth. RF3 most likely has an anaerobic lifestyle as most sequences to date are from anaerobic environments including mammalian gut (Tajima *et al.*, 1999; Ley *et al*., 2006; Samsudin *et al*., 2011), sediment (Yanagibayashi *et al.*, 1999; Röske *et al.*, 2012), municipal waste leachate (Huang *et al.*, 2005), anaerobic sludge (Chouari *et al.*, 2005; Goberna *et al*., 2009; Rivière *et al.*, 2009; Tang *et al*., 2011), a subsurface oil well head (Yamane *et al.*, 2011) and the anaerobic zone of saline lakes (Bowman *et al.*, 2000b; Humayoun *et al*., 2003; Schmidtova *et al*., 2009). However, some members have been found in aerobic environments such as surface waters of hypersaline systems (Demergasso *et al*., 2008; Yilmaz *et al.*, 2012), surface of a freshwater lake (Xing *et al.*, 2009) and compost (Partanen *et al.*,2010).

(\*Bacilli? Halomonas? Psychromonas? Clostridia)

#### 0.8–0.1 µm size fraction

There was had a large number of eucaryotic SSU sequences on the 0.1 µm size fraction, specifically from fungi, *Dictyochophyceae*, *Dinophyceae* and choanoflagellates. These taxa were all found on larger size fractions, except fungi which were uniquely present in the 1.7 m 0.1 µm sample and were classified as *Cordyceps* and *Ascomycota*. The presence of these *Eucarya* on the smallest filter may due to small size during particular stages in their life history; such as a cyst formation (\*sizes) or sexual reproduction (\*check), or degraded cellular material.

The mixed zone of the 0.1 µm was dominated by *Pedinellales* and their chloroplast sequences consistent with active phototrophic cells localizing to surface light. *Pedinellales* have only been previously detected in Antarctic lakes from molecular studies (Unrein *et al.*, 2005; Lauro *et al.*, 2011; Yau *et al.*, 2011). (\*Unrein was just one fragment at 95% identity to *Apedinella*)(\*check).

The 0.1 µm deep samples were distinguished by the presence of candidate divisions OD1 and TM7 which were concentrated on the lake bottom. The prevalence of these two divisions almost exclusively on the smallest size fraction is consistent with a small cell size. Another study similarly found OD1 to predominate in <0.2 µm fraction of ground water plankton (Miyoshi *et al.*, 2005). OD1 has a wide distribution and has been consistently associated with anoxic environments implying an anaerobic physiology (Harris *et al*., 2004). In the marine environments, it has also been associated with reduced environments with high sulfur such as sulfate and sulfides (Harris *et al*., 2004; Elshahed *et al.*,2005;\*other ref). Genomic fragments of a member of OD1 from Zodletone Spring, Oklahoma showed oxygen sensitive enzymes related to anaerobic or facultative anaerobic bacteria (Elshahed *et al.*,2005). Thus, the distribution of Organic Lake OD1 is consistent with an anaerobic metabolism.

“*Candidatus* Aquiluna”, in the Luna-1 cluster of Actinobacteria (\*Hahn *et al.*, 2004; Hahn *et al.*, 2009) was most abundant on the 0.1 µm size fraction at 1.7 m depth, however it was also present in the deep zone of the 0.1 and 3.0 µm size fractions. The genus has small cells, <1.2 µm in length (Hahn *et al.*, 2009), consistent with their concentration on the smallest size fraction. One isolated from surface Artic seawater has been genome sequenced and found to contain genes for actinorhodopsin (\*Kang *et al*., 2012). Similarly, actinorhodopsins were expressed in the surface of Ace Lake (Lauro *et al.*, 2011) indicating a potential for photoheterotrophy (see below). Isolates were aerobic chemoheterotrophs (\*ref) but the presence on multiple size fractions implies of both the aerobic and anaerobic zones implies it is facultatively anaerobic or present in the bottom due to sedimentation.

## Ecosystem functions and links to taxonomic composition

Variation in the cellular population structure was significantly correlated (Rho: 0.519, significance: 0.3%) with the abiotic parameters DO, temperature, TS and TN. (\*RELATE to the species composition?) Molecular markers for key C, N and S conversions were retrieved from the metagenomic reads to determine the capacity for nutrient cycling in Organic Lake, especially those that affect these influential factors.

Analysis of nutrient cycling potential showed the absence of major chemical pathways. Genes involved in nitrification (ammonia oxidation AMO genes) were not detected suggesting a limited capacity to re-form bioavailable N and a mechanism for accumulation of ammonia (Figure 4B). Why ammonia oxidation does not occur is unclear (\*check AMO study Dry Valleys). Genes for methanogenesis were also absent (Figure 4A). This is to be expected as methanogenesis usually only occurs when alternate electron acceptors such as sulfate and nitrate depleted in anoxic conditions (\*ref paper SRB vs methanogens)(\*it is probably energetically unfavourable if it is hydrogenoclastic, and limited by competitive exclusion if fermentative). Moreover, methane oxidation enzymes that were detected are not indicative of active methane cycling as they related to alkane hydroxylases and therefore most likely involved in hydrolysis of compounds such as phenol, which has been previously detected in the bottom waters of Organic Lake (Roberts & Burton 1993a; Roberts *et al.*, 1993b). Although Organic Lake is not sulfidic, it is possible for ‘cryptic’ sulfur cycling to occur where there is no chemical signature. Sulfate reduction can be tightly coupled to sulfide oxidation such that there is no detectable sulfide (\*Canfield *et al.*, 2010), or sulfate (Ng *et al.*, 2010; Lauro *et al.*, 2011). However, in Organic Lake, sulfur oxidation genes were undetectable and dissimilatory sulfate reduction extremely limited (Figure 4C) indicating incomplete sulfur cycling. Consistent with this, very few sulfur oxidizers were found (\*check) and sulfate-reducing *Deltaproteobacteria* were present (Figure 2 and Table S2) at very low abundance.

(\*paper in SRB vs methanogens) The factors that exclude these processes may be cold, salinity or low levels of oxygen in the deep zone. Since the deep zone of Organic Lake has been found to be totally anoxic in the past (\*ref), yet neither methane (\*ref) nor methanogens have been detected, this implies their absence was not due to oxygen incursion in the recent past but a longer term inhibition.

The absence of key pathways and the increased potential for mineralization indicates both a mechanism by which metabolic end-products may accumulate and a possibility of net nutrient loss. The potential for respiration, fermentation and CO oxidation was much higher than potential for carbon fixation (Figure 4A). The capacity for N assimilation, mineralization and denitrification was higher than fixation (Figure 4B) indicating a net loss of C and N from the system. We recognize the balance in genetic potential and does not account for expression or activity. The indication that there is a net decline in N is supported by accumulation of ammonia at 6.5 m, which could occur from dissimilative nitrate reduction that is not converted to nitrogen gas, deamination (\*check) or Stickland fermentation of amino acids.

A net loss in essential elements implies that a there may be an influx of exogenous nutrients occurs to sustain the lake system. However, external input, such as from glacial melt-water, could only occur in the summer months when the lake is ice-free. Furthermore, the water column structure is characteristic of a negative water balance (\*Gibson) indicating the Organic Lake system has been largely closed in the recent past. Thus, if external inputs occur, they are episodic and would necessitate interim strategies for C, N and S conservation as was noted for the nearby Ace Lake (Lauro *et al*. 2011).

As was observed in the microbial community composition, the molecular markers were distributed according to size fraction and depth. The majority of the genetic potential for known C, N and S metabolism was restricted to the 0.8 and 3.0 µm size fractions. The lack of ascribed functional genes in the 0.1 µm filter reflects the paucity of cellular life in that fraction and the high representation of candidate divisions, which are likely to have fewer homologs in sequence databases. Aerobic processes such as aerobic respiration and aerobic carbon fixation were more abundant in the mixed zone where DO concentration was highest. Conversely reactions inhibited by oxygen including fermentation, anaerobic carbon fixation, nitrogen fixation, ammonification (\*), anammox and dissimilatory sulfate reduction were more prevalent in the suboxic deep zone. Potential for nitrogen assimilation, denitrification, nitrogen mineralization, assimilatory sulfate reduction and sulfur mineralization were abundant pathways that showed no clear difference with depth indicating they are linked to abundant taxa that were present throughout the lake, or they are genes with a wide distribution among different taxa. (\*test for difference in distribution of genes in mixed and deep zones).

Anaerobic carbon fixation, fermentation and CO oxidation were processes associated with the increased biological activity at 6.5 m. (\*why not at 6.7 m too?) Likewise, genes for Stickland fermentation were detected here which could also contribute to the accumulation of ammonia (\*figure). Conversely, assimilatory sulfate reduction is lowest here, perhaps because sulfur can be assimilated from DMSP/DMS breakdown (\*see below).

Most processes can be attributed to known functions or distribution of the taxa detected (\*map KOs to taxa). Oxygenic photosynthesis was presumably mediated out by phytoflagellates as chloroplasts were abundant (Figure 2). These taxa were the main source of primary production in the mixed zone with potentially some contribution from diatoms and photosynthetic dinoflagellates. However, the vast majority of the marker genes for the Calvin-Benson-Bassham autotrophic carbon fixation cycle, ribulose bisphosphate carboxylase oxygenase (RuBisCO) and phosphoribulose kinase (prKA) were assigned to *Gammaproteobacteria* (\*which gamma?) and not to eucaryotic phytoplankton. This implies the *Gammaproteobacteria*, even the surface heterotrophic lineages, have the capacity for autotrophic carbon fixation. Most likely they have chemolithoautotrophic capacity such as metal oxidation.

### Potential for lithoheterotrophy

CO is an indirect green house gas as it contributes to methane and nitrous oxide concentrations (\*refMoran). It is formed in aquatic environments during photochemical degradation of organic molecules (\*ref). Purely carboxydotrophic bacteria oxidize CO to CO2 using the enzyme CO dehydrogenase and fix a proportion of the CO2 with ribulose 1,5-bisphosphate carboxylase (RuBISCO). Alternatively, it has been proposed that roseobacters may fix CO2 by anaplerotic mechanisms (\*ref Moran 2007). Many roseobacters such as *R. pomeroyi* only possess CO dehydrogenase and are able to oxidize CO at low concentrations as an energy source and assimilate organic carbon for growth thus limiting organic carbon oxidation. (\*link CO oxidation to taxon).

### Diverse proteorhodopsin homologs are linked to most bacterial lineages

The first rhodopsin found in bacteria, termed proteorhodopsin (PR) because of its *Gammaproteobacteria* origin, acts as a light-driven proton pump and was hypothesized to be used for energy generation (Béjà *et al*., 2000). Metagenomic studies have since shown PR are diverse, widely distributed in the surface ocean (Rusch *et al*., 2007) and associated with diverse bacterial clades including *Alphaproteobacteria* (de la Torre *et al.*, 2003) and *Bacteroidetes* (Venter *et al.*, 2004) as well as *Euryarchaeota* (Frigaard *et al.*, 2006). Related to PRs is a clade of rhodopsins linked to non-marine or coastal aquatic environments (Sharma *et al.*, 2008). Within this clade are actinorhodopsins, associated with *Actinobacteria* (Sharma *et al.*, 2008; Sharma *et al.*, 2009) and xanthorhodopsin, characterized from the sphingomonad *Salinibacter ruber* (\*ref), which we will refer to as actino-xanthorhodopsins.

A total of 399 reads matching to rhodopsins were detected in Organic Lake, which formed 124 clusters at 90% amino acid identity. Phylogenetic analysis revealed six well-supported rhodopsin groups named for their taxonomic affiliation: MAR, OL-1, OCT, SAL, AQU and PSY (Figure S8). Only the PSY clustered with the PRs showing most Organic Lake rhodopsin diversity was within the actino-xanthorhodopsin clade. All groups had an L or M residue corresponding to position 105 in the SAR86 PR denoting tuning to surface green light (Man *et al.*, 2003; Gomez-Consarnau *et al.*, 2007), which is consistent with the shallow water in Organic Lake and is characteristic of coastal samples (Rusch *et al.*, 2007). All sequenced homologs that clustered with Organic Lake rhodopsins originated from polar and/or lake species (Gosink *et al.*, 1997;Ward & Priscu, 1997; Bowman *et al.*,1998; Antón *et al.*, 2002;Hahn, 2009; Kang *et al.*, 2012).

The phylogeny of most Organic Lake rhodopsins was consistent with the bacterial groups present. MAR, PSY, OCT and AQU groups clustered with homologs from *Marinobacter*, *Psychroflexus*, *Octadecabacter* and “*Candidatus* Aquiluna rubra” respectively (Figure S8), all of which are genera detected in the lake (Table S2). Xanthorhodopsin was described from the sphingomonad *Salinibacter* *ruber* (\*ref), thus the Organic Lake SAL group likely originates from related *Sphingobacteria* (Table S2). The abundant MAR and PSY rhodopsins has a distribution that agrees with the distribution of *Marinobacter* and *Psychroflexus* (\*Figure) further supporting to their phylogenetic origins. However, the most abundant group, OL-1, had no close homologs from GENBANK. From its high abundance and concentration on the 3.0 µm fraction, OL-1 group most likely originated from the *Rhodobacterales* including *Roseovarius* (\*Figure). Although it is possible that OL-1 was encoded by *Flavobacteria* as they are similarly abundant in the 3.0 µm fraction, all known *Flavobacteria* only possess PRs, not actino-xanthorhodopsins.In contrast, diverse rhodopsins occur in roseobacters: *Octadecabacter articus* and *O. antarcticus* have actino-xanthorhodopsins and *Alphaproteobacteria* HTCC2255 has a PR (\*ref, Moran 2007). However the occurrence of rhodopsins is highly variable with only two of the 42 sequenced roseobacter genomes encoding a rhodopsin. As most of the sequenced roseobacters are marine, apart from *Octadecabacter* it stands to reason that there is selection for rhodopsins in certain polar or coastal bacterial lineages. For example, only *Marinobacter* sp. ELB17, which is an Antarctic lake isolate possesses a rhodopsin gene. It is not present in current genomic sequences for or oil degrading *M. hydrocarbonoclasticus*, dinoflagellate symbiont *M. algicola* or particle associated *M. adhaerens* which are from the ocean surface. However, it is also not present in the Canada Basin, Artic isolate *Marinobacter* sp. BSs20148.

If there are approximately 3 000 bacterial SSU sequences, assuming PR to be single copy and SSU copy number to range from one to ten, 13–100% of Organic Lake bacteria have a PR. This is comparable to the Mediterranean Sea estimates of 13% (Sabehi *et al.*, 2005).(\*use recA or radA, as Sharma *et al.*, 2008 which saw Punta Cormorant has 36% vs open ocean 63%).

\*Proton pump activity positions

Recently, proteorhodopsins of marine Flavobacteria and *Vibrio* have been associated with light-dependent energy generation (Gomez-Consarnau *et al*., 2007), especially under low carbon conditions (\*ref). This is a potential mechanism for conserving carbon for growth and may contribute to the success of PR bearing lineages Organic Lake. Certainly this is likely to be the case for Organic Lake *Psychroflexus* as it is both taxonomically related to *Dokdonia* and has a PR of the same phylotype. This is less clear for the otherrhodopsin groupsthey do not have well characterized relatives.If these proteorhodopsin homologs in Organic Lake add to energy generation, this would indicate mixotrophy is a common strategy in all the dominant bacterial lineages present. This may also allow them to occupy low oxygen environments.

### Aerobic Anoxygenic Photosynthesis

Aerobic and anaerobic anoxygenic photosynthesis may be occurring and would be mediated by the Roseobacters present such as *Roseovarius*. *R. tolerans* is the type species of the genus and was isolated from Ekho Lake, a meromictic hypersaline lake in the Vestfold Hills (\*Labrenz *et al.*, 1999). It was found to produce bacteriochlorophyll A when grown in the dark, but continuous dim light inhibited production (\*Labrenz *et al.*, 1999).

### DMSP and DMS metabolism

Homologs of DMSP lyase genes *dddD*, *dddL* and *dddP*, which catalyse the breakdown of DMSP forming DMS as a by-product, were detected in Organic Lake at levels comparable to other dominant processes such as respiration and fermentation (\*figure:DMS\_cycle). DMSP lyases are from completely unrelated enzyme families and confer the Ddd (DMSP-dependent DMS) phenotype (Curson *et al.*, 2011). The most abundant in Organic Lake, *dddD*, comprised approximately 70% of the DMSP lyase genes. Organic Lake *dddD* homologs clustered with *dddD* genes with confirmed DMSP lyase activity (Figure S6) which supports their putative function as DMSP lyases and were divided into two main *dddD* types. One clade grouped with a *Marinobacter* sp. ELB17 homolog and was enriched on the 0.8 µm fraction, consistent with the distribution of *Gammaproteobacteria* including *Marinobacter,* unclassified *Alteromonadales* and *Saccharospirillum* (Figure S6). The other clade, with high identity (~80%) to *Halomonas* sp. HTNK *dddD*, comprised the majority (75%) of *dddD* homologs and was restricted to the 3.0 µm fraction. However, its distribution did not reflect the location of *Halomonas*, which is concentrated on the 0.8 µm fraction. Thus far, *dddD* genes have predominantly been found in Gammaproteobacteria and in some Alpha and Betaproteobacteria (Curson et al., 2011) and as the *Halomonas* group *dddD* was confined to 3.0 µm this suggests Alphaproteobacteria such as, *Roseovarius*, *Loktanella*, *Albimonas* and other unclassified Rhodobacterales as more likely originators of the most abundant *dddD*.

*dddL* encodes a small polypeptide with unknown functional domains apart from a C-terminal cupin metal-binding pocket. To date, it has been found only in Alphaproteobacteria, predominantly from the Roseobacter clade (Curson *et al.*, 2011). Two *dddL* groups were detected in Organic Lake: a *Sulfitobacter* and a *Marinobacter* group (Figure S7). The former clusters with *dddL* from *Sulfitobacter* sp. EE-36 and other Rhodobacteraceae (\*check if they are from different Roseobacter clades). In *Sulfitobacter* sp. EE-36 *dddL* gene aloneis sufficient for the Ddd phenotype (Curson *et al*., 2008). The latter group and the more abundant *dddL* type, forms a separate clade from the known *dddL* homologs and includes a hypothetical protein from *Marinobacter manganoxydans* MnI7-9, a deep-sea manganese oxidizing bacterium. This finding suggests the *Marinobacter* group *dddL* is an unrecognized member of this enzyme family and is the first report of *dddL* in Gammaproteobacteria, although whether it confers the Ddd phenotype requires further confirmation. The *Sulfitobacter* group is most abundant at 6.5 m depth while the *Marinobacter* group is most abundant at 5.7 m. Both are predominantly located on the 0.8 µm fraction, which suggests the origin of both Organic Lake *dddL* types are Gammaproteobacteria. The *Marinobacter* group *dddL* fits the distribution of *Marinobacter* while the concentration of the *Sulfitobacter* group *dddL* at 6.5 m indicates this homolog originated from other Gammaproteobacteria such as *Psychromonas* or *Halomonas* which are predominant at 6.5 m.

These data suggest *dddD* would mediate the majority of DMSP degradation, followed by *dddL* and *dddP* leading to the high concentration of DMS that has been detected in bottom waters (\*ref). This function was most likely performed by *Rhodobacterales* similar to *Roseovarius* as well as *Marinobacter* and potentially other unclassified *Alteromonadales*.

Usually methanogenic or sulfate reducing bacteria breakdown DMS in anoxic conditions (\*ref). Since only sulfate reducing bacteria were detected but at very low abundance, faster rates of DMSP production than DMS degradation would account for the high concentration in the deep zone. Alternatively, other anaerobic routes of DMS production, eg. via anaerobic breakdown of methionine may account for the DMS in the bottom waters. Reduction of DMSO may be another source of DMS accumulation.

## Viral diversity and distribution

## Acknowledgements

## References

Abell GCJ and Bowman JP. (2005a) Colonization and community dynamics of class *Flavobacteria* on diatom detritus in experimental mesocosm based on Southern Ocean seawater. *FEMS Microbiol Ecol* **53**: 379–391.

Abell GCJ and Bowman JP. (2005b) Ecological and biogeographic relationships of class Flavobacteria in the Southern Ocean. *FEMS Microbiol Ecol* **51**: 265–277.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990) Basic Local Alignment Search Tool. *J Mol Biol* **215**: 403–410.

Antón J, Oren A, Benlloch S, Rodríguez-Valera F, Amann R, Roselló-Mora R. (2002) *Salinibacter ruber* gen. nov., sp. nov., a novel extremely halophilic member of the *Bacteria* from saltern crystallizer ponds. *Int J Syst Evol Microbiol* **52**: 485–491.

Béjà O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP, Jovanovich SB *et al*. (2000) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**: 1902–1906.

Bengtsson K, Eriksson KM, Hartmann M, Wang Z, Shenoy BD, Grelet G-A *et al*. (2011) Metaxa: a software tool for automated detection and discrimination among ribosomal small subunit (12S/16S/18S) sequences of archaea, bacteria, eukaryotes, mitochondria, and chloroplasts in metagenomes and environmental sequencing datasets. *Antonie Van Leeuwenhoek* **100**: 471–475.

Bird MI, Chivas AR, Radnell CJ, Burton HR. (1991) Sedimentological and stable-isotope evolution of lakes in the Vestfold Hills, Antarctica. *Palaeogeogr Palaeoclimatol Palaeoecol* **84**: 109–130.

Bowman JP, McCammon SA, Lewis T, Skerratt JH, Brown JL, Nichols DS, McMeekin TA. (1998) *Psychroflexus torquis* gen. nov., sp. nov., a psychrophilic species from Antarctic sea ice, and reclassification of *Flavobacterium gondwanense* (Dobson et al. 1993) as *Psychroflexus gondwanense* gen. nov., comb. nov. *Microbiology* **144**: 1601–1609.

Bowman JP, McCammon SA, Rea SM, McMeekin TA. (2000b) The microbial composition of three limnologically disparate hypersaline Antarctic lakes. *FEMS Microbiol Lett* **183**: 81–88.

Burke CM and Burton HR. (1988) Photosynthetic bacteria in meromictic lakes a stratified fjords of the Vestfold Hills, Antarctica. *Hydrobiologia* **165**: 13–23.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al*. (2010) QIIME allows analysis of high-throughput community sequence data. *Nat Methods* **7**: 335–336.

Charlson RJ, Lovelock JE, Andreae MO, Warren SG. (1987) Oceanic phytoplankton, atmospheric sulphur, cloud albedo and climate. *Nature* **326**: 655–661.

Chen YG, Cui XL, Wang YX, Tang SK, Zhang YQ, Li WJ, Liu JH *et al*. (2009) *Psychroflexus sediminis* sp. nov., a mesophilic bacterium isolated from salt lake sediment in China. *Int J Syst Evol Microbiol* **59**: 569–573.

Chouari R, Le Paslier D, Daegelen P, Ginestet P, Weissenbach J, Sghir A. (2005) Novel predominant archaeal and bacterial groups revealed by molecular analysis of an anaerobic sludge digester. *Environ Microbiol*  **7**: 1104–1115.

Clarke KR. (1993) Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology* **18**: 117–143.

Clarke KR and Gorley RN. (2006) PRIMER v6: User Manual/Tutorial. PRIMER-E, Plymouth.

Curran MAJ and Jones GB. (1998) Spatial distribution of dimethylsulfide and dimethylsulfonioproprionate in the Australasian sector of the Southern Ocean. *J Geophys Res* **103**: 16 677–16 689.

Curson ARJ, Rogers R, Todd JD, Bearley CA, Johnston AWB (2008) Molecular genetic analysis of a dimethysulfonioproprionate lyase that liberates the climate-changing gas dimethylsulfide in several marine α-proteobacteria and *Rhodobacter sphaeroides*. *Environ Microbiol* **10**: 757–767.

Curson ARJ, Todd JD, Sullivan MJ, Johnston AWB (2011) Catabolism of dimethylsulphonioproprionate: microorganisms, enzymes and genes. *Nat Rev Microbiol* **9**: 849–859.

de la Torre JR, Christianson LM, Béjà O, Suzuki MT, Karl DM, Heidelberg J *et al.* (2003) Proteorhodopsin genes are distributed among divergent bacterial taxa. *PNAS* **100**: 12830–12835.

DeSantis Jr. TZ, Hugenholtz P, Keller K, Brodie EL, Larsen N, Piceno YM *et al*. (2006) NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. *Nucleic Acids Res* **34**:W394–399.

Demergasso C, Escudero L, Casamayor EO, Chong G, Balagué V, Pedrós-Alió. (2008) Novelty and spatio-temporal heterogeneity in the bacterial diversity of hypersaline Lake Tebenquiche (Salar de Atacama). *Extremophiles* **12**: 491–504.

Demergasso C, Dorador C, Meneses D, Blamey J, Cabrol N, Escudero L, Chong G. (2010) Prokaryotic diversity pattern in high-altitude ecosystems of the Chilean Altiplano. *J Geophys Res* **115**: G00D09

Deprez PP, Franzmann PD, Burton HR. (1986) Determination of reduced sulfur gases in Antarctic lakes and seawater by gas chromatography after solid adsorbent preconcentration. *J Chromatogr* **362**: 9–21.

Dobson SJ, James SR, Franzmann PD, McMeekin TA. (1991) A numerical taxonomic study of some pigmented bacteria isolated from Organic Lake, an antarctic hypersaline lake. *Arch Microbiol* **156**: 56–61.

Dobson SJ, Colwell RR, McMeekin TA, Franzmann PD. (1993) Direct sequencing of the polymerase chain reaction-amplified 16S rRNA gene of *Flavobacterium gondwanense* sp. nov. and *Flavobacterium salegens* sp. nov., two new species from a hypersaline Antarctic lake. *Int J Syst Bacteriol* **43**: 77–83.

Donachie SP, Bowman JP, Alam M. (2005) *Psychroflexus tropicus* sp. nov., an obligately halophilic *Cytophaga-Flavobacterium-Bacteroides* group bacterium from an Hawaiian hypersaline lake. *Int J Syst Evol Microbiol* **54**: 935–940.

Edgar RC. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nuc Acids Res* **32**: 1792–1797.

Ferris JM, Gibson JAE, Burton HR. (1991) Evidence of density currents with the potential to promote meromixis in the ice-covered saline lakes. *Palaeogeogr Palaeoclimatol Palaeoecol* **84**: 99–107.

Franzmann PD, Burton HR, McMeekin TA. (1987a) *Halomonas subglaciescola*, a new species of halotolerant bacteria isolated from Antarctica. *Int J Syst Bacteriol* **37**: 27–34.

Franzmann PD, Deprez PP, Burton HR, van den Hoff J*.* (1987b) Limnology of Organic Lake, Antarctica, a meromictic lake that contains high concentrations of dimethyl sulfide. *Aust J Mar Freshw Res* **38**:409–417.

Frigaard NU, Martinez A, Mincer TJ, DeLong EF. (2006) Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and Archaea. *Nature* **439**: 847–850.

Fofonoff NP and Millard RC Jr. (1983) Algorithms for computation of fundamental properties of seawater. *UNESCO Technical Papers in Marine Science*, no.**44**.

Fuhrman JA, Schwalbach MS, Stingl U. (2008) Proteorhodopsins: an array of physiological roles? *Nat Rev Microbiol* **6**: 488–494.

Gauthier MJ, Lafay B, Christen R, Fernandez L, Acquaviva M, Bonin P, Betrand JC. (1992) *Marinobacter hydrocarbonoclasticus* gen. nov., sp. nov., a new, extremely halotolerant, hydrocarbon-degrading marine bacterium. *Int J Syst Bacteriol* **42**: 568–576.

Ghai R, Pašić L, Fernández AB, Martin-Cuadrado A-B, Mizuno CM, McMahon KD, Papke RT *et al.* (2011) New abundant microbial groups in aquatic hypersaline environments. *Sci Rep* **1**: srep00135.

\*Gibson JAE *et al.* (1989) Temperature profiles of saline lakes of the Vestfold Hills. *ANARE Research Notes,* No.67, 75pp

Gibson JAE, Ferris JM, Burton HR. (1990) Temperature density, temperature conductivity and conductivity-density relationships for marine-derived saline lake waters. *ANARE Research Notes*, No. 78.

Gibson JAE, Garrick RC, Franzmann PD, Deprez PP, Burton H. (1991) Reduced sulfur gases in saline lakes of the Vestfold Hills, Antarctica. *Palaeogeo Palaeoclimatol Palaeoecol* **84**: 131–140.

Gibson JAE, Qiang XL, Franzmann PD, Garrick RC, Burton HR. (1994) Volatile fatty and dissolved free amino acids in Organic Lake, Vestfold Hills, East Antarctica. *Polar Biol* **14**: 545–550.

Gibson JAE, Burton HR, Gallagher JB. (1995) Meromictic Antarctic lakes as indicators of local water balance: structural changes in Organic Lake, Vestfold Hills 1978–1994.  *ANARE Research Notes*, No.94, 16pp.

\*Gibson JAE *et al.* (1996) Meromictic Antarctic lakes as recorders of climate change: the structures of Ace and Organic Lakes, Vestfold Hills, Antarctica. *Papers and Proceedings of the Royal Society of Tasmania* **130**:73–78.

Gibson JAE. (1999) The meromictic lakes and stratified marine basins of the Vestfold Hills, East Antarctica. *Antarct Sci* **11**: 175–192.

Glatz RE, Lepp PW, Ward BB, Francis CA. (2006) Planktonic microbial community composition across steep physical/chemical gradients in permanently ice-covered Lake Bonney, Antarctica. *Geobiology* **4**: 53–67.

Goberna M, Insam H, Franke-Whittle IH. (2009) Effect of biowaste sludge maturation on the diversity of thermophilic bacteria and archaea in an anaerobic reactor. *Appl Environ Microbiol* **75**: 2566–2572.

Gosink JJ, Herwig RP, Staley JT. (1997) *Octadecabacter articus* gen. nov., sp. nov., and *O. antarcticus*, sp. nov., nonpigmented, psychrophilic gas vacuolate bacteria from polar sea ice and water. *System Appl Microbiol* **20**: 356–365.

Hahn MW, Stadler P, Wu QL, Pöckl. (2004) The filtration–acclimatization method for isolation of an important fraction of the not readily cultivable bacteria. *J Microbiol Methods* **57**: 379–390.

Hahn MW. (2009) Description of seven candidate species affiliated with the phylum *Actinobacteria*, representing planktonic freshwater bacteria. *Int J Syst Evol Microbiol* **59**: 112–117.

Hahsler M, Hornik K, Buchta C. (2008) Getting things in order: an introduction to R package seriation. *J Stat Softw* **25**:1–34.

Huang L, Zhu S, Zhou H, Qu L. (2005) Molecular phylogenetic diversity of bacteria associated with the leachate of a closed municipal solid waste landfill. *FEMS Microbiol Lett* **242**: 297–303.

Humayoun SB, Bano N, Hollibaugh JT. (2003) Depth distribution of microbial diversity in Mono Lake, a meromictic soda lake in California. *Appl Environ Microbiol* **69**: 1030–1042.

James SR, Dobson SJ, Franzmann PD, McMeekin TA. (1990) *Halomonas meridiana*, a new species of extremely halotolerant bacteria from Antarctic saline lakes. *System Appl Microbiol* **13**: 270–278.

James SR, Burton HR, McMeekin TA, Mancuso CA. (1994) Seasonal abundance of *Halomonas meridiana*, *Halomonas subglaciescola*, *Flavobacterium gondwanense* and *Flavobacterium salegens* in four Antarctic Lakes. *Antarctic Sci* **6**: 325–332.

Kang I, Lee K, Yang S-J, Choi A, Kang D, Lee YK, Cho J-C. (2012) Genome sequence of “*Candidatus* Aquiluna” sp. strain IMCC13023, a marine member of the *Actinobacteria* isolated from an Artic Fjord. *J Bacteriol* **194**: 3550–3551.

Kirchman DL. (2002) The ecology of *Cytophaga-Flavobacteria* in aquatic environments. *FEMS Microbiol Ecol* **39**: 91–100.

La Scola B, Desnues C, Pagnier I, Robert C, Barrassi L, Fournous G, Merchat C *et al.* (2008) The virophage as a unique parasite of the giant mimivirus. *Nature* **455**: 100–105.

Labrenz M, Collins MD, Lawson PA, Tindall BJ, Schumann P, Hirsch P. (1999) *Roseovarius tolerans* gen. nov., sp. nov., a budding bacterium with variable bacteriochlorophyll *a* production from hypersaline Ekho Lake. *Int J Syst Bacter* **49**: 137–147.

Lauro FM, DeMaere MZ, Yau S, Brown MV, Ng C, Wilkins D *et al.* (2011) An integrative study of a meromictic lake ecosystem in Antarctica. *ISME J* **5**: 879–895.

Ley RE, Turnbaugh PJ, Klein S, Gordon JI. (2006) Human gut microbes associated with obesity. *Nature* **444**: 1022–1023.

Lovelock JE and Maggs RJ. (1972) Atmospheric dimethyl sulfide and the natural sulphur cycle. *Nature* **237**: 452–453.

\*Ludwig W., *et al.* (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363–1371.

Laybourn-Parry J and Pearce D. (2007) The biodiversity and ecology of Antarctic lakes: models for evolution. *Phil Trans R Soc B* **364**: 2273–2289.

Lee ZM, Bussema C 3rd, Schmidt TM. (2009) rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. *Nucleic Acids Res* **37** (Database issue): D489–D493.

Man D, Wang W, Sabehi G, Aravind L, Post AF, Massana R *et al*. (2003) Diversification and spectral tuning in marine proteorhodopsins.  *EMBO J* **22**: 1725–1731.

Matsuzaki M, Kubota K, Satoh T, Kunugi M, Ban S, Imura S. (2006) Dimethyl sulfoxide-respiring bacteria in Suribati Ike, a hypersaline lake, in Antarctica and the marine environment. *Polar Biosci* **20**: 73–87.

McCammon SA and Bowman JP. (2000) Taxonomy of Antarctic *Flavobacterium* species: description of *Flavobacterium gillisiae* sp. nov., *Flavobacterium tegetincola* sp. nov.and *Flavobacterium xanthum* sp.nov., nom. rev. and reclassification of [*Flavobacterium*] *salegens* as *Salegentibacter salegens* gen. nov., comb. nov. *Int J Syst Evol Microbiol* **50**: 1055–1063.

\*Millero FJ, Chen CT, Bradshaw A, Schleicher K. (1980) A new high pressure equation of state for seawater. *Deep Sea Res A* **27**: 255–264.

Miyoshi T, Iwatuski T, Naguma T. (2005) Phylogenetic characterization of 16S rRNA gene clones from deep-groundwater microorganisms that pass through 0.2 µm-pore-size filters. *Appl Environ Microbiol* **71**: 1084–1088.

Naganuma T, Hua PN, Okamoto T, Ban S, Imura S, Kanda H. (2005) Depth distribution of euryhaline halophilic bacteria in Suribati Ike, a meromictic lake in East Antarctica. *Polar Biosci* **28**: 964–970.

Ng C, DeMaere MZ, Williams TJ, Lauro FM, Raftery M, Gibson JAE *et al.* (2010) Metaproteogenomic analysis of a dominant green sulfur bacterium from Ace Lake, Antarctica. *ISME J* **4**:1002–1019.

Noguchi H, Park J, Takagi T. (2006) MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res* **34**: 5623–5630.

Pagaling E, Wang H, Venables M, Wallace A, Grant WD, Cowan DA, Jones BE *et al.* (2009) Microbial biogeography of six salt lakes in Inner Mongolia, China and a Salt Lake in Argentina. *Appl Environ Microbiol* **75**: 5750–5760.

Partanen P, Hultman J, Paulin L, Auvinen P, Romantschuk M. (2010) Bacterial diversity at different stages of the composting process. *BMC Microbiol* **10**: 94.

Powell LM, Bowman JP, Skerratt JH, Franzmann PD, Burton HR. (2005) Ecology of a novel *Synechococcus* clade occurring in dense populations in saline Antarctic lakes. *Mar Ecol Prog Ser* **291**: 65–80.

Redfield AC, Ketchum BH, Richards FA. (1963) The influence of organisms on the composition of seawater, In: Hill MN (ed). The sea. John Wiley and Sons: New York, pp 26–77.

Rivière D, Desvignes V, Pelletier E, Chaussonnerie S, Guermazi S, Weissenbach, Li T *et al.* (2009) Towards the definition of a core of microorganisms involved in anaerobic digestion of sludge. *ISME J* **3**: 700–714.

Roberts NJ and Burton HR. (1993a) Sampling volatile organics from a meromictic Antarctic lake. *Polar Biol* **13**: 359–361.

Roberts NJ, Burton HR, Pitson GA. (1993b) Volatile organic compounds from Organic Lake, an Antarctic hypersaline, meromictic lake. *Polar Biol* **13**: 361–366.

Röske K, Sachse R, Scheerer C, Röske I. (2012) Microbial diversity and composition of the sediment in the drinking water reservoir Saidenbach (Saxonia, Germany). *Syst Appl Microbiol* **35**: 35–44.

Rusch DB, Halpern AL, Sutton G, Heidelbergg KB, Williamson S, Yooseph S *et al.* (2007) The *Sorcerer II* Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol* **5**: 398–431.

Sabehi G, Loy A, Jung K-H, Partha R, Spudich JL, Isaacson T, Hirschberg J *et al*. (2005) New insights into metabolic properties of marine bacteria encoding proteorhodopsins. *PLoS Biol* **3**: e273.

Samsudin AA, Evans PN, Wright AG, Al Jassim R. (2011) Molecular diversity of the foregut bacteria community in the dromedary camel (*Camelus dromedariusi*). *Environ Microbiol* **13**: 3024–3035.

Schmidtova J, Hallam SJ, Baldwin SA. (2009) Phylogenetic diversity of transition and anoxic zone bacterial communities within a near-shore anoxic basin: Nitinat Lake. *Environ Microbiol* **11**: 3233–3251.

Sharma AK, Zhaxybayeva O, Papke RT, Doolittle WF. (2008) Actinorhodopsins: proteorhodopsin-like gene sequences found predominantly in non-marine environments. *Environ Microbiol* **10**: 1039–1056.

Sharma AK, Sommerfeld K, Bullerjahn GS, Matteson AR, Wilhelm SW, Jezbera J, Brandt U *et al*. (2009) Actinorhodopsin genes discovered in diverse freshwater habitats and among cultivated freshwater *Actinobacteria*. *ISME J* **3**: 726–737.

Tajima K, Aminov RI, Nagamine T, Ogata K, Nakamura M, Matsui H *et al*. (1999) Rumen bacterial diversity as determined by sequence analysis of 16S rDNA. *FEMS Microbiol Ecol* **29**: 159–169.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**: 2731–2739.

Tang Y, Ji P, Hayashi J, Koike Y, Wu X, Kida K. (2011) Characteristic microbial community of a dry thermophilic methanogenic digester: its long-term stability and change with feeding. *Appl Microbiol Biotechnol* **91**: 1477–1461.

Tian F, Yu Y, Chen B, Li H, Yao Y-F, Guo X-K. (2009) Bacterial, archaeal and eukaryotic diversity in Artic sediment as revealed by 16S rRNA and 18S rRNA gene clone libraries analysis. *Polar Biol* **32**: 93–103.

Todd JD, Rogers R, Li YG, Wexler M, Bond PL, Sun L, Curson ARJ *et al.* (2007) Structural and regulatory genes required to make the gas dimethyl sulfide in bacteria. *Science* **315**: 666–669.

Unrein F, Izaguirre I, Massana R, Balagué V, Gasol JM. (2005) Nanoplankton assemblages in maritime Antarctic lakes: characterisation and molecular fingerprinting comparison. *Aquat Microb Ecol* **40**: 269–282.

Van Trappen S, Mergaert J, Van Eygen S, Dawyndt P, Cnockaert MC, Swing J. (2002) Diversity of 746 heterotrophic bacteria isolated from microbial mats from ten Antarctic lakes. *System Appl Microbiol* **25**: 603–610.

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

Wagner-Döbler I and Biebl H. (2006) Environmental biology of the marine *Roseobacter* lineage. *Ann Rev Microbiol* **60**: 255–280.

Ward BB and Priscu JC. (1997) Detection and characterization of denitrifying bacteria from a permanently ice-covered Antarctic lake. *Hydrobiologia* **347**: 57–68.

Wang Q, Garrity GM, Tiedje JM, Cole JR. (2007) Naïve Bayesian classifier for rapid assignment of rRNA sequences into new bacterial taxonomy. *Appl Environ Microbiol* **73**: 5261–5267.

\*Wu J, Mao X, Cai T, Luo J, Wei L. (2006) KOBAS server: a web-based platform for automated annotation and pathway identification. *Nucleic Acids Res* **34**: W720–W724.

Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li CY, Wei L. (2011) KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res* **39**: W316–W322.

Xing P, Hahn MW, Wu QL. (2009) Low taxon richness of bacterioplankton in high-altitude lakes of the eastern Tibetan Plateau, with a predominance of *Bacteroidetes* and *Synechoccocus* spp. *Appl Environ Microbiol* **75**: 7017–7025.

Yamane K, Hattori Y, Ohtagaki H, Fujiwara K. (2011) Microbial diversity with dominance of 16S rRNA genes sequences with high GC contents at 74 and 98°C subsurface crude oil deposits in Japan. *FEMS Microbiol Ecol* **76**: 220–235.

Yanagibayashi M, Nogi Y, Li L, Kato C. (1999) Changes in the microbial community in Japan Trench sediment from a depth of 6292 m during cultivation without decompression. *FEMS Microbiol Lett* **170**: 271–279.

Yau S, Lauro FM, DeMaere MZ, Brown MV, Thomas T, Raftery MJ *et al.* (2011) Virophage control of antarctic algal host-virus dynamics. *Proc Natl Acad Sci USA* **108**: 6163­–6168.

Yilmaz P, Iversen MH, Hankeln W, Kottman R, Quast C, Glöckner FO. (2012) Ecological structuring of bacterial and archaeal taxa in surface ocean waters. *FEMS Microbiol Ecol* **81**: 373–385.

Yoon JH, Kang SJ, Jun YT, Oh TK. (2009) *Psychroflexus salinarum* sp. nov., isolated from a marine solar saltern. *Int J Syst Evol Microbiol* **59**: 2404–2407.

Zhang H, Hosoi-Tanabe S, Nagata S, Ban S, Imura S. (2010) *Psychroflexus* lacisalsi sp. nov., a moderate halophilic bacterium isolated from a hypersaline lake (Hunazoko-Ike) in Antarctica. *J Microbiol* **48**: 160­–164.

Zwartz D, Bird M, Stone J, Lambeck K. (1998) Holocene sea-level change and ice-sheet history in the Vestfold Hills, East Antarctica. *Earth Planet Sci Lett* **155**: 131­–145.