Organic Lake, a lacucosm for studying globally important processes

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)

Organic Lake is a shallow (6.75 m deep) hypersaline lake in the Vestfold Hills, East Antarctica with a high concentration of the cloud-forming gas dimethylsulfide (DMS)(\*ref). 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). By utilizing a metagenomic approach we were able link taxonomic composition with functional diversity and identify ecosystem level processes. Primary production was generated in the surface waters by the eucaryotic phytoflagellates *Dunaliella* and *Pseudopedinella* relatives. This supported heterotrophic bacteria, mainly *Marinobacter*, *Roseovarius* and *Psychroflexus* throughout the water column*.* Diverse and abundant rhodopsin homologs linked to these major bacterial lineages suggest photoheterotrophy as an adaptive strategy. Over abundance of DMSP lyase genes *dddD*, *dddL* and *dddP*,likely encoded by Gammaproteobacteria and Alphaproteobacteria, indicated DMSP hydrolysis to be the origin of high DMS in the bottom waters. 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 sample was dominated by candidate divisions OD1 and TM7. This study has allowed a rigorous description of microbial taxa within a natural habitat and sheds light on globally relevant biogeogemical processes such as DMS generation, lithoheterotrophy and photoheterotrophy.

# Introduction

Life in the Antarctic is constrained by extremes of temperature and salinity under a polar light regime. Within the polar 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: Vestfold\_map) is one such region where hundreds of lakes are found. The lakes were formed from seawater trapped approximately 10 000 years before present when the continental ice-shelf retreated and isostatic rebound caused the land to rise above sea-level (Zwartz *et al*., 1998; Gibson, 1999). Differing local conditions has lead each lake to develop unique physical and chemical properties. The large array of lake systems, with biota that is often entirely microbial, makes them fitting sites to study biogeography and biogeochemistry. The ability to encompass a large proportion of the species diversity using molecular techniques within a relatively closed, stratified system of reduced diversity allows us to better infer functional roles for the taxa present (Laybourn-Parry & Pearce, 2007).

A metagenomic approach, complemented with metaproteomics, has been successfully applied to lakes in the Vestfold Hills: Ace Lake and Organic Lake (Ng *et al.*, 2010; Lauro *et al.*,2011; Yau *et al.*, 2011). A comprehensive description of the Ace Lake ecosystem was achieved that delineated the community structure, biogeochemical fluxes and identified key responses to resource limitation (Lauro *et al.*, 2011). The metabolism of the dominant green sulfur bacteria (Ng *et al.*, 2010)was found to play a central role in C, N and S cycling (Lauro *et al*., 2011). Mathematical modeling showed its dominance was dependent on synchronicity with the polar light cycle and the absence of phage predation (Lauro *et al*., 2011). In contrast, a member of the virophage virus family was discovered in Organic Lake that potentially regulates microbial loop dynamics (Yau *et al*., 2011). Virophage were named for their detrimental effect on the larger helper virus they require to replicate (La Scola *et al.*, 2008). The Organic Lake virophage (OLV) likely depends on phycodnaviruses whose probable hosts are prasinophyte microalgae. The reduction of infective phycodnaviruses by OLV “predation” would lead to increased algal blooms and thus carbon flux (Yau *et al.*, 2011). These studies have gained unprecedented insight into the microbial diversity and function of these remarkable lake environments. Moreover, these findings may have broader relevance to other aquatic systems. For example, OLV-like sequences were found in coastal marine, freshwater and hypersaline lagoon metagenomes indicating a wider ecological role (Yau *et al.*, 2011).

Organic Lake is also 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 other polysulfides in the bottom waters (Roberts & Burton 1993a; Roberts *et al.*, 1993b). Atmospheric DMS is a precursor for cloud condensation nuclei making it important in cloud climate regulation (\*ref). Concentrations of DMS as high as 5000 nM have been recorded in Organic Lake (\*what’s the conc in the ocean?), potentially the highest recorded in a natural body of water (Gibson *et al*., 1991) and undergoes annual variation indicating active turnover (\*ref). Unusually, 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.*, 1991), the cold and salinity, six times that of seawater, appears to preclude the establishment of sulfate reducing bacteria (Gibson *et al.*, 1991) (\*what about methanogens). Phototrophic sulfur oxidizing bacteria are also absent (Burke & Burton, 1988) indicating other bacteria mediate the unusual sulfur chemistry. Determining the means of DMS production in Organic Lake may provide unique insight into global processes. In order to gain an understanding of the unusual sulfur chemistry and the microbial community context for the astonishing virus-virus-host interaction in Organic Lake, it was chosen as a model site for examination from a whole ecosystem perspective using metagenomic analysis.

# 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.

## Overall diversity of Organic Lake microbial community

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 indicating they were rare in Organic Lake.

Details of the composition of each bacterial phylum are shown in Table S2. 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 divisions 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 anddictyochophyte algae, which had the same distribution as chloroplasts (Figure 2B). Chlorophytes were principally of the genus *Dunaliella*, halophilic green algae, previously recorded as the dominant alga in Organic Lake (Franzmann *et al*., 1987b). Dictyochophytes (sillicoflagellates) were related to the genus *Pseudopedinella.* (\*lifestyle?) They have only been previously detected in Antarctic lakes by sequencing of the 18S rRNA gene (Unrein *et al.*, 2005\*Bielewics?, Lauro *et al.*, 2011; Yau *et al.*, 2011). (\*how similar where Unreins?). Lower abundance eucarya included Bacillariophyta (diatoms), Dinophyceae, Fungi and heterotrophic choanoflagellates. Bacillariophyta were related to *Chaetoceros* and would contribute to primary production (\*what sort? \*ref Donna’s paper about their 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). (\*comment on why they are adapted to Organic Lake)(\*how much of an impact are the choanoflagellates having on Organic Lake?) Bacterial and eucaryal classes were generally represented by a single dominant genus (Table S2\*check) indicating overall complexity of the Organic Lake is low. (diversity indices\*).

Although the proportions of SSU genes may not necessarily reflect the number of cells in the environment because of potential SSU copy number bias, archaeal SSU gene copies only range from 1–4 (Lee *et al.*, 2009). Thus, it is expected Archaea were truly scarce. Genome information is not available for all taxa, especially Eucarya, to apply a correction for SSU gene copy number. 16S gene copy numbers for the nearest sequenced relatives for the bacteria present ranges from 1–6 (data not shown), consequently estimates for abundance are accurate within this margin of error. Certainly changes in abundance of SSU gene composition between samples are indicative of true relative differences in microbial population. (\*mention GAAS here\*recA comparison).

The microbial composition is similar to other hypersaline cold or high altitude aquatic systems (lakes, salterns) with the dominance of Gammaproteobacteria, Alphaproteobacteria and Flavobacteria as well as the presence of eucaryal algae *Dunaliella* as the main primary producers. 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). Uncultured Organic Lake taxa have highest identity to SSU sequences from saline and/or cold environments (Aquiluna\*Kang *et al*., 2012). The consistent association with phylotypes from hypersaline and cold environments indicates these factors are selecting for particular taxonomic groups in Organic Lake. This is further supported by the detection of the same taxa in Organic Lake over time which implies they are highly adapted to the Organic Lake environment.

### Bacteria and Eucarya diversity and inferred functional group

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; \*).

Only low numbers of strictly anaerobic taxa including Clostridia (primarily *Halanaerobium*) and sulfate-reducing Deltaproteobacteria were detected. RF3 is an uncultured division affiliated with Firmicutes that was first detected in bovine rumen (Tajima *et al.*, 1999). The majority of Organic Lake RF3 sequences were most closely related to environmental sequences from cold hypersaline lakes (\*Table S2) including Laguna Lejía, Chilean Altiplano (Demergasso *et al.*, 2010), Lake Shangmatala, Inner Mongolia (Pagaling *et al*., 2009) and Ekho Lake, Vestfold Hills, East Antarctica (Bowman *et al.*, 2000b). RF3 most likely has an anaerobic lifestyle as most RF3 sequences to date are 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?)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. The predominance of heterotrophic aerobes implies the suboxic environment precludes the establishment of high numbers of strictly anaerobic bacteria.

Photoautotrophic bacteria were represented by a few cyanobacteria sequences, however these sequences were not assigned to free-living cyanobacteria (\*check) and may be chloroplast. *Roseovarius tolerans* strains are known to produce bacteriochlorophyll *a* in the dark (Labrenz *et al*., 1999) and could be aerobic anoxygenic phototrophs.. (\*find out from the environments candidates if they are associated with).

### Distribution of microbes occurs according to size and depth

(**Description of size and depth distributions, speculation of why they might have that distribution. Generate hypotheseses of metabolism and physiology)**

Seriation analysis showed the cellular community composition clustered according to size fraction and depth (Figure 3) and identified taxa 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 indicates that taxonomic groups were adapted to specific niches within the lake.

The mixed zone samples from the 3.0 µm size fraction 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, but at highest abundance at 4 m (Franzman *et al.*, 1987b). Signatures of algae found at the bottom of the lake are likely due to sedimentation as they are biflagellated so would be able to control their location in the water column. (\*find out about Dunaliella and the genome).

*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 previously isolated from Organic Lake (Franzmann *et al*., 1987b). These Flavobacteria isolates are rods ranging in length from approximately 1.5–11.5 µm (Dobson *et al*., 1991). Flavobacteria have also been linked with a particle attached lifestyle (\*ref), which would be consistent with enrichment on the 3.0 µmfilter. Like all other characterized *Psychroflexus* species to date (Donachie *et al.*, 2004; Chen *et al.*, 2009; Yoon *et al*., 2009;Zhang *et al*., 2010), *P. gondwanensis* is an orange-pigmented (from carotenoid pigments) aerobic chemoheterotroph (Dobson *et al.*, 1993; Bowman *et al.*, 1998).

*Psychroflexus* from the 3.0 µm samples were 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 (\*ref). Likely, Organic Lake *Psychroflexus* have a similar function as the isolate *P. gondwanense* can degrade starch, displays beta-glucosidase activity (Zhang *et al.*, 2010) and shows no growth enhancement from labile substrates such as monosaccharides and amino acids (Dobson *et al*., 1993) (\*check). Immunofluoresence staining has shown *P. gondwanese* to be most abundant in surface waters and strongly correlated with average hours of sunshine per day (James *et al.*, 1994). This would be consistent with growth response to increased primary production. Its presence in the deep zone could be due to sedimentation as *P. gondwanense* is non-motile and a strict aerobe. However, some species are capable to nitrate reduction (Zhang *et al*., 2010) so it is possible Organic Lake *Psychroflexus* species may be metabolically active in the deep zone (\*check nitrate reductase). The varied size and depth distribution of this one genus implies some variation in Organic Lake Flavobacteria such as substrate preference or particle attached vs planktonic lifestyle.

*Roseovarius* was also principally found on the 3.0 µm filter and was enriched at 6.5m and 4.2 m. (\*are they large cells or particle attached?) As a member of the Roseobacter clade, which is known to have diverse metabolic capabilities such DMSP degradation and aerobic anoxygenic photosynthesis (reviewed in Wagner-Döbler & Biebl, 2006), this distribution suggests *Roseovarius* was occupying multiple niches in Organic Lake. The population at 6.5 m may be contributing to the unusual chemistry at that depth (\*How is it found in other hypersaline lakes?) (\*check bacteriochlorophyll A from Roseovarius).

*Marinobacter* dominated the 0.8 µm size fraction but were less abundant in the 6.5 m sample. The concentration of *Marinobacter* on this size fraction is consistent with the cell size of isolates (\*ref) reflective of planktonic cells. They are aerobic heterotrophs originally isolated on hydrocarbons (\*ref), and generally prefer labile substrates such as sugars, amino acids and organic acids (\*ref) potentially made available from breakdown of high molecular weight organic matter by Flavobacteria. *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 allowed for their presence throughout the water column (\*see below). 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 (\*see below).

Curiously, there was had a large number of Eucarya 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 (dictyochophytes) and their chloroplast sequences consistent with active phototrophic cells localizing to surface light. The 0.1 µm deep samples were distinguished by the presence of candidate divisions OD1 and TM7 which were concentrated on the lake bottom. “*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. A member of this genus isolated from surface Artic seawater has been genome sequenced and found to contain genes for actinorhodopsin (\*Kang *et al*., 2012).

How to divide the taxonomic data?

Current structure:

\*overview of major groups and what they do: size and depth differences

Other ways

\*Taxonomic groups:

Bacteria

**chemoorganoheterotrophic**: Gammaproteobacteria, Flavobacteria/Sphingobacteria, Cytophaga, Vc2.1\_bac22, Sb-1, Alphaproteobacteria, Actinobacteria, Clostridia, Bacilli, Deltaproteobacteria which are the SRB.

photoorganoheterotrophic,

chemolithoautotrophic: : Epsilonproteobacteria (Sulfurimonas) oxidize reduced sulfur compounds such as sulfide, sulfur, DMSO, nitrate (I guess DMS is possible too). Usually they convert sulfide to sulfite and to sulfate by sulfite oxidase, or they may use the reversal of the APS reductase system used by SRB. Deltaproteobacteria which are the SRB.

photolithoautotrophic: some cyanos but really none.

chemoorganoautotrophic,

photoorganoautotrophic)

chemolithoheterotrophic: Epsilonproteobacteria (Sulfurimonas) oxidize reduced sulfur compounds such as sulfide, sulfur, DMSO, nitrate (I guess DMS is possible too). Usually they convert sulfide to sulfite and to sulfate by sulfite oxidase, or they may use the reversal of the APS reductase system used by SRB

photolithoheterotrophic:

Eucarya (photolithoautotrophic, mixotrophic, chemoorganoheterotrophic)

## 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.

C, N and S cycling potential was characterized by net nutrient loss and the absence or restriction of certain pathways (Figure 4). The potential for respiration, fermentation and CO oxidation was much higher than potential for carbon fixation (Figure 4A). Similarly, 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. As genes involved in nitrification were not detected, this suggests a limited capacity to reform bioavailable N, contributing to overall decline of N and accumulation of ammonia (\*short circuit of N via ammonia?). Oxidizing conditions in the deep zone limited metabolic reactions such as methanogenesis and dissimilatory sulfate reduction, supported by lack of these genes. The genes detected for methane oxidation are in the same family as alkane hydroxylases and are 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).

We recognize the balance in genetic potential and does not account for expression or activity. Several processes were supported by the distribution of compounds in the lake. The indication that there is a net decline in N is supported by accumulation of ammonia at 6.5 m. This also suggests a general mechanism by which end-products of metabolism can accumulate such as organic acids from fermentation and DMS. Nitrate reduction by the lack of nitrate in the water column. (\*nitrite?) 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 reflects the paucity of cellular life in that size fraction and the high representation of candidate divisions, which are unlikely to have 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 the most abundant taxa and not subject to the DO or pH gradient within the lake. (\*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 largely carried out by phytoflagellates as chloroplasts from and *Dunaliella* anddictyochophyte algae were abundant and there were few cyanobacteria sequences (Figure 2). These taxa were the main source of primary production in the mixed zone with potentially some contribution from diatoms and photosynthetic dinoflagellates. (\*Check which taxa the RUBISCO and phosphoribulose kinase map to.)

### Nitrate reduction

Ammonia found at higher concentration in the deep zone is consistent with previous studies, however in past studies found the maximum concentration of ammonia was 0.82 mg L-1, approximately 6.5 times lower (Franzmann *et al*., 1987b). Ammonia accumulation in the deep zone was hypothesized to originate from nitrate reductino (Franzmann *et al.*, 1987b).

### Potential for lithoheterotrophy

CO is an indirect green house gas as it contributes to methane and nitrous oxide concentrations (\*ref). 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.

### Potential for bacterial photoheterotrophy

Microbial rhodopsins are retinal binding proteins that act as light-driven ion pumps for translocating chloride ions (halorhodopsins), protons (bacteriorhodopsins, proteorhodopsins and xanthorhodopsins) and for light sensing (sensor rhodopsins). The first rhodopsin found in Bacteria, termed proteorhodopsin (PR) because of its Gammaproteobacterial 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 is 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), CFB group (Venter *et al.*, 2004) and \*Actinobacteria as well as Euryarchaea (Frigaard *et al.*, 2006). A total of 399 reads matching to rhodopsins were detected in Organic Lake which formed 124 clusters at 90% amino acid identity. 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 other single copy gene). (\*compare to similar sequencing efforts like GOS. Is this a lot?). Phylogenetic analysis of a representative sequences that spanned the proteorhodopsin spectral tuning region (Man *et al.*, 2003) showed six groups: *Marinobacter*, unknown OL rhodopsin, *Octadecabacter*, Xanthorhodopsin, Actinorhodopsin and *Flavobacteria* groups (Figure S8). All of which have an L or M residue corresponding to position 105 in SAR86 PR and indicating tuning to surface green light (λmax 525 nm) (Man *et al.*, 2003) consistent with the shallow waters in Organic Lake.

The most abundant groups were: the unknown OL rhodopsin; the *Marinobacter* group, which clustered with *Marinobacter* sp. ELB17 isolated from the Antarctic hypersaline Lake Bonney and the *Flavobacteria* group, most closely related to *Psychroflexus* of which *P. gondwanensis* was isolated from Organic Lake (Franzmann *et al*., 1987b)*.* The relative abundance, size and depth distributions of the *Marinobacter*, *Flavobacteria* and Actinorhodopsin homologs agrees with their proposed phylogenetic origin (\*Figure). From its abundance and concentration on the 3.0 µm fraction, the OL rhodopsin group most likely originated from *Roseovarius* or other unclassified Rhodobacterales (\*what about euk? or Flavo origin?).

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* which are in the same clade as *Dokdonia* PR. This is less clear for the *Marinobacter* and OL rhodopsin groups as they do not have well characterized relatives.Thus far, the only known Roseobacter to possess a rhodopsin is the Alphaproteobacteria HTCC2255 isolate (\*ref, Moran 2007)(\*remake tree with alphaproteobacteria homolog)(\*distribution?)(\*what proportion have it? use recA?). If these proteorhodopsin homologs in Organic Lake adds to energy generation, this would indicate mixotrophy is a common strategy in the dominant bacterial lineages present. This may also allow them to occupy low oxygen environments?

The Xanthorhodopsin may play a sensory role in Organic Lake, but by far, the most abundant rhodopsin type was related to proteorhodospins. (\*Taxonomic origin?) Probable role in phototaxis.

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

# Discussion

**stratification stability** When first surveyed between 1978 and 1984 (Deprez *et al.* 1986; Franzmann *et al*., 1987b), it was considered meromictic (permanently stratified) due to the stable bottom temperatures of approximately −6 ºC and a pycnocline between 3–4 m. Organic Lake is sensitive to changes in water level. When water level increases a lens of fresher surface water effectively insulates the mid-waters from contact with ice creating a mid-water heat-trap. Negative water balance, such as the drop of 0.81 m observed between 1989 and 1994, cause bottom temperatures to fall and the mixed zone to descend deeper down the water column (Gibson, 1996). The water column structure from this study is similar to that of the 1990’s.

**Cellular life** Many of the bacteria identified in this study, including *Marinobacter*, *Roseovarius*, *Psychroflexus* and *Halomonas* have been previously detected in a 16S PCR survey of Organic Lake sediment (Bowman *et al.*, 2000b) showing some continuity in the population over time. *Marinobacter* has been cultured from microbial mats (Van Trappen *et al.*, 2002) and strains of Flavobacteria have been consistently isolated, including *Psychroflexus gondwanensis* (ACAM 44) and *Salegentibacter salegens* (ACAM 48) (Franzmann *et al*., 1987b; Dobson *et al*. 1991). The dominance of *Psychroflexus* is consistent with previous work which found *Psychroflexus gondwanensis* could comprise up to 10% of the summer bacterial population in the surface (James *et al*. 1994). *Halomonas* has been previously cultured including the species *H. subglaciescola* (ACAM 12) and *H. meridiana* (Franzman *et al*., 1987a; James *et al*., 1990; James *et al*. 1994). However, this is the first report of candidate divisions RF3 and OD1 and potentially links them to their functional potential. RF3 was isolated from bovine rumen and is somewhat related to Clostridia. It also co-occurs with Clostridia and Bacilli.

The phytoplankton population appears to undergo succession. For example, genera previously reported such as *Chaetoceros* and *Pyramimonas* (Franzmann *et al*., 1987a) were not detected or at very low abundances in this study. This may be linked to strain cycling due to viral pressures and/or linked with light tolerance during the polar day-night transition (Bielewics *et al.*, 2010).

The first report of dictyochophyceae in Antarctic lakes was from Unrein 2005, so they may be important in Antarctic phytoplankton that have been missed. They are on the 0.1 um so may have a small cell size. Fungi and ciliates being in small size fractions is perplexing. Fungi found in Bielewicz 2010 and Unrein 2005.

Diversity indices (table: diversity\_indices\_hypersaline\_lakes) between sample filters and sample depths were not significantly different from one another indicating diversity is similar throughout the water column. The estimate of total species richness (Chao1) was much higher than previously calculated from a 16S clone library of the sediment (Bowman *et al*., 2000b). This is due to the use of metagenomic reads when forming OTUs inflating the apparent number of OTUs and occurs for several reasons. Non-overlapping reads that cover different sections of the SSU gene will not be grouped as the same OTU if that gene is not present in the SILVA release 108 reference set. A read may match group with a partial sequence in the SILVA reference database, but if a large proportion of the read is outside the reference sequence, it will form its own OTU.

**nutrient cycling:** Bowman *et al.* (2000b) hypothesized that redox potential was too high in Organic Lake for anaerobic respiration to occur. However, Roberts & Burton (1993) proposed the positive redox potential values measured previously were due to leakage of Kemmerer bottles used for sampling as negative values were obtained with modified bottles.

Organic Lake is enriched in sulfur compared to similar Antarctic Lakes (\*table of sulfate in other lakes). Salinity is purportedly too high for sulfate reducing bacteria (Franzmann *et al*., 1987a)or phototrophic sulfur bacteria to occur (Burke & Burton, 1988)(\*check other lakes such as Pendant, Bonney andVida for the presence of sulfate reducers and GSB). This is consistent with the lack of these species in the taxonomic analysis and alternative sulfur chemistry compared to similar, but less saline systems.

The genetic potential of the lake indicates a net loss as certain key steps in the cycle are not present. This could indicate exogenous inputs that are feeding the lake cycle.

Mixotrophy seems to be a prevalent strategy. There is not a one-to-one correspondence of one taxon occupying a single function, but there is some functional overlap.

The high DMS concentration was hypothesized to originate from DMSP breakdown and/or anaerobic DMS production (\*ref). One possible pathway of anaerobic generation is methylation of methanethiol (methylmercaptan), however, methanethiol has not been detected in Organic Lake (Roberts *et al*., 1993b).

Previous work using immunofluorescence staining for *Psychroflexus gondwanese* has shown it to comprise up to 10% of the summer bacterial composition at 2 m (James *et al.*, 1994). This is comparable to this study, where ~8.5% of SSU sequences were *Psychroflexus* (Figure 2) across filter sizes at 1.7 m providing some validation

## Acknowledgements

## References

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

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.

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.

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.

\*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.

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.

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.

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.

\*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.

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.

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.

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.

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.

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.

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.