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 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 ideal sites to study questions of biogeography, biogeochemistry and evolution.

The ability to encapsulate a large proportion of the species diversity within a relatively closed system using shotgun environmental sequencing (metagenomics) allows us to infer which taxa may be mediating particular biological processes. A metagenomic approach, complemented with identification of proteins expressed in the environment (metaproteomics) has been successfully applied to two contrasting lakes from 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 short-circuiting of biogeochemical cycles as a key response to resource limitation (Lauro *et al.*, 2011). Furthermore, the metabolism of the dominant green sulfur bacteria (Ng *et al.*, 2010)was found to play a central role in carbon, nitrogen and sulfur cycling that maintained the entire lake in its current state. Mathematical modeling showed its dominance was dependent synchronicity with a polar light cycle and the absence of phage predation making it highly specialized and the entire system potentially fragile (Lauro *et al*., 2011). By contrast, in Organic Lake, a member of the virophage virus family the Organic Lake virophage (OLV) was discovered that potentially controls carbon flux (Yau *et al*., 2011). Virophage were named for their detrimental effect on the larger “helper” virus it requires to complete its life cycle (La Scola *et al.*, 2008). OLV likely depends upon phycodnaviruses that infect the prasinophyte microalga *Pyramimonas*. The reduction of infective phycodnaviruses from OLV “predation” would lead increased frequency of algal blooms and flux of carbon into the microbial loop (Yau *et al.*, 2011). These studies have gained unprecedented insight into the microbial diversity and function into these remarkable lake environments. Moreover, they have shown 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 (Yau *et al.*, 2011).

Organic Lake is also unusual due to the high concentration of the cloud-forming gas dimethylsulfide (DMS), as well as other polysulfides that have been detected in the bottom waters (Deprez *et al*., 1986;Franzmann *et al.*, 1987; Gibson *et al.*, 1991; Roberts & Burton 1993a; Roberts *et al.*, 1993b). DMS is ecologically relevant as it is part of the CLAW hypothesis (\*ref). This is potentially the highest concentration recorded in a natural body of water (\*ref). DMS concentrations vary throughout the year (Roberts *et al.*, 1993b) indicating active turnover. Organic Lake is a shallow lake (~7 m) that has approximately six times the salinity of seawater (\*ref). 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. The bottom waters were anoxic, but not sulfidic as sulfate reducing bacteria are absent (Gibson *et al.*, 1991). Phototrophic sulfur oxidizing bacteria are also absent (Burke & Burton 1988) indicating other bacteria are mediating the unusual sulfur chemistry. This study focused on understanding Organic Lake from a whole ecosystem perspective in order to place the astonishing virus-virus-host interaction in context and uncover the mechanism for its unusual sulfur chemistry.

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).

This study aimed to:

1. Determine the microbial population structure of Organic Lake along the depth gradient.
2. Describe the functional capacity of the microorganisms.
3. Link microbial processes to lake physico-chemistry.
4. Examine possible microbe-microbe interactions (link microbial groups with each other).

# 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

Nitrate, nitrite, ammonia, 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 standard methods at the Analytical Centre\* (Tasmania). Values for dissolved nutrients and inorganic N were measured from the 0.1 µm filtrate. All other nutrients were measured from water collected after pre-filtration through 20 µm pore size filter.

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

## Physical and chemical properties of Organic Lake

During sampling on November 2008, Organic Lake had a maximum depth of 6.75 m and the surface measured 3.874 m above mean sea level. *In situ* physico-chemical profiles were measured over the deepest point in the lake (Figure S1) to evaluate the water column properties and structure (Figure S2). 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. This is consistent with the presence of an oxycline at the same depth indicating depletion of DO due to respiration in the stagnant waters. The pH also decreased with DO due to accumulation of organic acids from fermentation in the suboxic environment (\*see below). 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 due to processes enriched in the microaerophilic environment (\*see below). The C:N and C:P ratios were high compared to the Redfield ratio (\*ref) 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.

## Cellular diversity and distribution

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 metagenome data). To determine the microbial composition of the samples, a total of 3 959 reads matching to SSU were retrieved from the metagenomic sequences which 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. Only 2 reads, assigned to a deep sea hydrothermal clade of Halobacteriales, were classified as Archaea revealing that they were rare in Organic Lake.

### Microbial composition

Three bacterial classes, Gammaproteobacteria, Alphaproteobacteria and Flavobacteria, were the most abundant and were found on all filter sizes at all depths (Figure 2A). Each class consisted of one dominant genus (at least 64% of sequences from that class) which was *Marinobacter*, *Roseovarius* and *Psychroflexus* respectively. Details of the composition of each division are shown in Table S2. Chloroplast sequences were also abundant in all samples except the 6.5 m 0.8 µm sample. Chlorophyte anddictyochophyte algae were the dominant Eucarya and had the same distribution as chloroplasts (Figure 2B). Many functional genes could be linked to these abundant taxa (\*see below) which contribute to major metabolic processes in Organic Lake. Moderately abundant bacterial divisions (1–5% of bacterial SSU) were Actinobacteria and candidate divisions OD1 and RF3. Lower abundance eucarya included Bacillariophyta, Dinophyceae, Fungi and heterotrophic choanoflagellates. These taxa had restricted distributions in the water column (Figure 2) and mediate conversions or functions specific to the zone or sample they were found in.

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

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%).

The mixed zone samples from the 3.0 µm size fraction had a greater abundance of *Dunaliella* chloroplasts and Chlorophyte algae consistent with large phototrophic cells concentrating near surface light. Signatures of algae found at the bottom of the lake are likely due to sedimentation. The abundance of *Psychroflexus* on the 3.0 µm sample filters reflects an epiphytic lifestyle involving metabolism of algal exudates and detritus, particularly recalcitrant high molecular weight polymers (\*ref). *Roseovarius* was also principally found on the 3.0 µm filter and was enriched at 6.5m and 4.2 m. As a member of the Roseobacter clade, which is known to have diverse metabolic capabilities such DMSP degradation and aerobic anoxygenic photosynthesis, 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 (\*check bacteriochlorophyll A from Roseovarius).

*Marinobacter* dominated the 0.8 µm size fraction but were less abundant in the 6.5 m sample. They are aerobic heterotrophs preferring 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 also capable of anaerobic respiration using DMSO, nitrate or iron as electron acceptors (\*ref) allowing for their presence throughout the water column. 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. The presence of these Eucarya on the smallest filter may due to small size during particular stages in their life history, such as a sporulation, or degraded cellular material. The mixed zone of the 0.1 µm was dominated by Pedinellales (Dictyochophyceae) 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.

## Ecosystem functions are linked 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, carbon monoxide oxidation, 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.)

### 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 used for light sensing, as chloride pumps in maintaining osmotic balance or proton pumps for ATP production (\*ref). They are abundant in the surface ocean (\*GOS paper). A total of 399 reads matching to rhodopsins were detected in Organic Lake which formed 124 clusters at 90% amino acid identity (\*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 proteorhodopsin and indicates tuning to surface green light (λmax 525nm) (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 (\*ref), especially under low carbon conditions (\*ref). This is a potential mechanism for conserving carbon for growth and may contribute to the success of proteorhodopsin bearing lineages Organic Lake. Certainly this is likely to be the case for Organic Lake *Psychroflexus* which are in the same lineage of proteorhodopsins as *Dokdonia*. 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** 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. However, the deep zone was not completely anoxic as has been recorded in the past (\*ref) indicating oxygen had invaded the bottom waters in the last 13 years. Oxygen may be episodically introduced as cold dense littoral water generated during ice-formation flows down the basin sides (\*Ferris *et al*., 1999).

**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 gondwanense* (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 gondwanense* 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.

## Acknowledgements

## References

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

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.

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.

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.

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.

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

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.

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

\*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, 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.

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

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.

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.

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

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.

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.

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.

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.

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.

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.

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.

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.

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. Ecological structuring of bacterial and archaeal taxa in surface ocean waters. *FEMS Microbiol Ecol* 2012;e-pub ahead of print 4 March 2012, doi:10.1111/j.1574-6941.2012.01357.x

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.