Organic Lake

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

# Introduction

Antarctic biota live at the extremes of temperature and salinity under a polar light regime. Antarctic lakes are a rare source of liquid water and an oasis for life in the polar desert. They are ideal locations to study questions of microbial biogeography, evolution and are a potential source of novel taxa and genes. The ability to encapsulate a large proportion of the species diversity allows us to infer which taxa may be mediating particular biological processes. For example, a member of the newly discovered virophage family was discovered in Organic Lake that may influence ecosystem stability and carbon flux (Yau *et al*., 2010). Meromictic lakes are ideal systems to link species to microbial processes as abiotic variables exist along a spatial gradient allowing for comparative analysis within a single relatively closed system (Lauro *et al*., 2011).

Organic Lake is a shallow ( ≈7 m) lake located on Long Peninsula in the Vestfold Hills, an ice-free region on the eastern shore of the Prydz Bay, East Antarctica (figure: Vestfold\_map). It consists of remnant seawater that was trapped approximately 10 000 BP when the continental ice-shelf retreated and isostatic rebound caused the land to rise above sea-level (Zwartz *et al*., 1998; Gibson, 1999). Complete separation from the ocean occurred approximately 3 000 BP (Bird *et al*., 1991) and the water has since concentrated to approximately six times the salinity of seawater. 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, likely due to the absence of sulfate reducing bacteria (Gibson *et al.*, 1991). However, dimethyl sulfide (DMS) has been recorded at high concentration in the bottom waters (Deprez *et al*., 1986;Franzmann *et al.*, 1987; Gibson *et al.*, 1991; Roberts & Burton 1993a; Roberts *et al.*, 1993b). DMS concentrations vary throughout the year (Roberts *et al.*, 1993b) indicating active turnover. The high DMS concentration was hypothesized to originate from DMSP breakdown and/or anaerobic DMS production. One pathways of anaerobic generation is methylation of methanethiol (methylmercaptan), however, methanethiol has not been detected in Organic Lake (Roberts *et al*., 1993b). Phototrophic sulfur oxidizing bacteria are also absent (Burke & Burton 1988) indicating sulfur cycling is not mediated by sulfur bacteria but rather other bacteria are mediating the unusual sulfur chemistry.

Organic Lake is sensitive to changes in water level. Between 1989 and 1994 a drop in water level of 0.81 m caused monimolimnion temperatures to fall to −12 ºC and deeper penetration of the mixolimnion down the water column (\*Figure X)(Gibson 1996).

In this study we 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

Organic Lake water was collected on 10 November 2008 (68º27'22.15"S, 78º11'23.95"E) through a 30 cm hole in the 0.8 m ice cover at the deepest point. Lake water was passed through a 20 µm pore size pre-filter then microbial biomass was captured by sequential filtration onto 3.0 µm, 0.8 µm and 0.1 µm pore size membrane filters as described previously (Ng *et al.*, 2010; Lauro *et al*., 2011). Between 1–2 L of lake water was sufficient to clog the filters. The samples were collected at 1.7, 4.2, 5.7, 6.5 and 6.7 m depths. 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 as previously described (Lauro *et al.*, 2011). \*Total protein was extracted from the membrane filters as previously described (Ng *et al.*, 2010; 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* vertical 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).

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 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 analysis

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 able 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 \*family/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 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). The CLUSTER analysis groups samples at successively smaller number of clusters at decreasing thresholds of similarity. 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 cellular composition.

### Viral diversity

## Functional potential of Organic Lake

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 did not have entries in KEGG orthology such as 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. 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

Phylogenetic analyses of protein coding sequences (\*rhodopsin, DMSP lyases) 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

## Physical and chemical properties of Organic Lake

Organic Lake had a maximum depth of 6.75 m (figure: bathymetry) and the surface measured 3.874 m above mean sea level during sampling on November 2008. *In situ* physico-chemical profiles and estimated practical salinity are shown in (figure: probe\_profiles). Organic Lake was vertically stratified into distinct two zones: a mixolimnion at 0–5.7 m and a monimolimnion below 5.7 m. The separation of these two zones was indicated by a steep pycnocline (density gradient) starting at 5.7 m that would \*prevent mixing of the bottom meter of water with the water above. Furthermore, physical and chemical properties were similar in the mixolimnion indicating mixing occurs between these depths. The sharp decrease in DO starting at 5 m is consistent with respiration occurring in isolation from gas exchange leading to oxygen depletion. The pH also decreases with DO, likely due to accumulation of organic acids from increased fermentation in the low oxygen environment (see below). The monimolimnion was not completely anoxic as it was in the past, but suboxic implying oxygen had invaded the bottom waters sometime in the last 13 years. Stratification has since re-established, \*albeit to a diminished degree. A separation between the mixolimnion and monimolimnion was supported by PCA analysis, which showed samples separated with depth along the PC1 axis (74.3% of variation) (figure: PCA).

All nutrients, except for nitrate and nitrite, as well as cell and VLP counts were at a maximum at 6.5 m (table: nutrients\_orglake) revealing this depth to be chemically distinct due to biological activity. Separation of the 6.5 m sample from the other samples occurred along PC2 axis (14.7% of variation) and was driven primarily by turbidity, sulfur, cell counts and TOC/TN. Interestingly, below 5 m, turbidity varied inversely to cell and VLP counts suggesting turbidity was not principally determined by cell density. Since the cell density, fermentation and anaerobic carbon fixation maxima coincide with the turbidity minimum, this suggests turbidity was due to particulate substrates that were degraded at 6.5 m to a much greater extent than the rest of the water column.

\*The C:N ratio was high compared to the Redfield ratio (\*ref) in both dissolved and particulate samples indicating N limitation relative to carbon throughout the water column. N limitation was most pronounced at 6.5 m. Dissolved P was depleted relative to C except at 6.5 m. The dissolved C:N:P ratio was different to the particulate ratios (table: nutrient\_orglake) reflecting differences in rates of synthesis/uptake compared to excretion/breakdown with N and P conserved within the particulate fraction. (\*Research Redfield ratios for lakes)

## Cellular diversity and distribution

A total of 3 959 reads matching to SSU were retrieved from the November 2008 profile. These grouped into 983 operational taxonomic units (OTUs). Bacteria were numerically dominant comprising 77.2% of all SSU sequences. 15.8% of sequences were assigned as Eucarya and 6.9% of SSU sequences could not be classified. Only 2 reads were classified as Archaea suggesting they were rare in Organic Lake. 61% of sequences were classified to a family level. (\*figure microscopy images).

**Bacteria**

The most abundant bacterial classes detected were Gammaproteobacteria, Alphaproteobacteria and Flavobacteria and all of these were found on each filter size and each depth (figure: QIIME\_class). These were the most abundant classes in all samples except the 0.1 µm monimolinion and 0.8 µm 6.5 m sample where candidate division OD1 and RF3 were also dominant. Each class comprised one dominant genus which was *Marinobacter*, *Roseovarius* and *Psychroflexus* for Gammaproteobacteria, Alphaproteobacteria and Flavobacteria respectively. Much of the functional genes identified could be mapped back to these genera (see below). Processes confined to the monimolimnion could then be attributed to genera that were specifically enriched in these strata. *Marinobacter* have been detected in a 16S PCR survey of Organic Lake sediment (Bowman *et al.*, 2000b) (\*which year) and was later cultured from microbial mats (Van Trappen *et al.*, 2002). *Halomonas* has been previously detected in Organic Lake including the species *H. subglaciescola* (ACAM 12) and *H. meridiana* (Franzman *et al*., 1987a; James *et al*., 1990; James *et al*. 1994). The majority of Flavobacteria were of the family Flavobacteriaceae primarily *Psychroflexus.* This is consistent with previous work where strains from that family were isolated, including the type strains *Psychroflexus gondwanense* (ACAM 44) and *Salegentibacter salegens* (ACAM 48) (Franzmann *et al*., 1987b; Dobson *et al*. 1991). *Psychroflexus gondwanense* could comprise up to 10% of the summer bacterial population in the surface (James *et al*. 1994).The detection of these same bacteria shows continuity in some members of the bacterial population over time.

Lower abundance Gammaproteobacteria included unclassified Alteromonadales, *Saccharospirillum*, *Halomonas* and *Psychromonas* (\*see table for listings). Other Alphaproteobacteria were also roseobacters including unclassified Rhodobacterales, *Loktanella* and *Albimonas*. Other bacterial classes present in all samples included Deltaproteobacteria, Epsilon Proteobacteria, Actinobacteria, Cytophagia, Sphingobacteria, and Opitutae (\*table: bacteria\_all\_profile).

Proteobacteria was the most abundant phylum consisting of Gammaproteobacteria and Alphaproteobacteria which comprised 66% and 27% of proteobacteria sequences respectively. Deltaproteobacteria and Epsilonproteobacteria each comprised approximately 1% of the proteobacterial sequences and the remaining 4% could not be classified to a class level.The vast majority of Gammaproteobacteria were *Marinobacter* (64.5% of Gammaproteobacteria), followed by *Saccharospirillum* (7%), *Halomonas* (4.5%) and *Psychromonas* (2.3%). A sizable proportion of SSU sequences were unclassified Alteromonadales and were more abundant in the monimolimnion.The most abundant Alphaproteobacteria were *Roseovarius* (76%), *Loktanella* (5.7%) and *Albimonas* (1.5%). \*have these been detected before?Deltaproteobacteria genera were *Desulfotignum* (38%), *Desulfopila* (20%), *Peredibacter* (8.5%) and *Bacteriovorax* (8.5%).Epsilonproteobacteria were *Sulfurimonas* (75.5%),  *Sulfurospirillum* (8.1%)and *Arcobacter* (7%)

**Eucarya**

The dominant eucarya were the photosynthetic flagellates from the families Chlorophyceae (green algae) andDictyochophyceae (silicoflagelates) predominantly of the genera *Dunaliella* and *Pseudopedinella* respectively. Bacillariophyceae (diatoms), Dinophyceae (dinoflagellates) (\*photosynthetic dinos?) and heterotrophic choanoflagellates (class Codonosigidae) were present at low abundances throughout the water column. Two groups showed highly localized distributions: fungi and ciliates. Fungi were restricted to the 0.1 µm fraction of the 1.7 m sample while ciliate signatures (Intramaculata; Spirotrichea) were detected only the 0.1 µm fraction on the 6.7 m. Strangely, the smallest size fraction had proportionally more Eucaryal SSU sequences, specifically from silicoflagellates and choanoflagellates. Chloroplast sequences seemed to co-vary with their hosts.

**Vertical stratification of the cellular community.**

\*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 cellular community composition varied according to both size fraction and depth with samples from the same size fraction and the same depth formed clusters. (figure: profile3\_genus\_heatmap). The 3.0 µm samples were overrepresented in *Psychroflexus* and *Roseovarius*. This cluster was further divided into mixolimnion and monimolimnion groups where the mixolimnion showed a greater abundance of *Dunaliella* chloroplasts and Chlorophyte algae. This is consistent with the larger size fraction containing mostly large phototrophic algal species that localize to surface light and being associated with epiphytic *Psychroflexus* which metabolize algal exudates. Signatures of chlorophyte algae and their chloroplasts were also found at the bottom of the lake, likely due to sedimentation. *Roseovarius* was found throughout the water column but were enriched in the monimolimnion. Some *Roseovarius* are symbionts with dinoflagellates, *R. tolerans* the type species of the genus produces bacteriochlorophyll A where they degrade the DMSP produced. Potentially there are several *Roseovarius* species in Organic Lake performing different functions. However, the dominant type is associated with processes in the monimolimnion, and appears over abundant at 6.5 m.

The 0.8 µm mixolimnion was overrepresented in *Marinobacter*,unidentified Alteramonadales and *Saccharospirillum*. The 0.8 µm monimolimnion samples were distinct from the mixolimnion due to an increased abundance of candidate division RF3, *Halomonas* and *Psychromonas*. *Marinobacter* was abundant in all samples except for the 6.5 m sample. It is a generalist facultative anaerobe that appears to occupy a related niche to *Psychroflexus* as its distribution down the water column is similar. It may be epiphytic or a symbiont of Dinoflagellates. \*Perhaps it would occupy 6.5 m but it is outcompeted by *Halomonas*, *Psychromonas* and RF3.

The 0.1 µm monimolimnion samples were distinguished by the presence of candidate division OD1 whereas the mixolimnion was dominated by silicoflagellates and an unknown chloroplast sequences (\*likely from the silicoflagellate). Fungal sequences were also uniquely present in the 1.7 m sample.

Figure: profile3\_genus\_heatmap also shows groups of organisms that co-vary. *Dunaliella*,Chlorophyte algae, *Psychroflexus*, *Roseovarius* and *Marinobacter* were a high abundance large size cluster. Another group was Alteromonadales, Gammaproteobacteria, Flavobacteriales, *Saccharospirillum*, Proteobacteria, OD1, RF3 and *Halomonas*.

ANOSIM analysis showed a statistical difference in genus level cellular composition (Rho: 0.53, significance: 0.1%) between mixolimnion and the monimolimnion. SIMPER analysis (\*figure: SIMPER\_community) identified the taxa that contributed to variance and supported the results from the seriation plot. Taxa overrepresented in the in the mixolimnion were *Marinobacter*, *Dunaliella*, *Psychroflexus*, Chlorophyte algae, Silicoflagellate algae and *Saccharospirillum*. The taxa overrepresented in the monimolimnion included *Roseovarius*, RF3, OD1, Rhodobacterales, *Halomonas*, *Psychromonas*, Desulfobacteraceae and TM7.

\*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. Furthermore, even if two reads originate from the same SSU gene, some regions are more unique so a read that matches to a less unique region may not cluster with the correct OTU.

## Ecosystem functions are linked to taxonomic composition

Variation in the cellular population structure was significantly correlated (0.519 R-value, 0.3% significance) with the abiotic parameters: DO, temperature, TS and TN. (\*RELATE to the species composition?) (\* Look into nitrogen and sulfur cycling capacity of the taxa). This supports a link between microbial processes and the lake chemistry. Most processes can be putatively attributed to known functions of the taxa present.

Oxygenic photosynthesis is mainly via photosynthetic eucarya. There are few cyanobacteria. Aerobic and anaerobic anoxygenic photosynthesis may 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 in when grown in the dark, but continuous dim light inhibited production (\*Labrenz *et al.*, 1999). \*Could it be doing photosynthesis? Rosovarius is nutritionally diverse, so may be occupying several ecological niches eg. symbiont with dinoflagellates, sulfur conversions, phototrophy via photosynthesis or rhodopsins.

Most bacteria are heterotrophic aerobes. eg. *Psychroflexus* and *Marinobacter*. Furthermore, the conditions are suboxic and some have the capacity for anaerobic respiration, potentially nitrate reduction, DMSO reduction, iron reduction?. There are no sulfur oxidizing bacteria but low abundances of sulfate reducing bacteria were detected. (\*discuss redox potential switching at the oxycline) Flavobacteria are associated with the eucaryotic algae as they consume algal exudates. The *Marinobacter* consume more labile products. \*What are the Roseobacters doing? The overrepresentation of Rhodobacteraceae, RF3, OD1, *Psychromonas*, *Halomonas*, Desulfobacteraceae and TM7 in the monimolimnion means they are likely linked to processes connected to the deeper waters. These include: production of ammonia possibly through nitrate reduction, amino acid fermentation, sulfur cycling \*more? *Halomonas* isolates are capable of converting nitrate to gas \*fermentation? With regards to the sulfur cycle, they are likely involved in DMSP cleavage. Rhodobacteraceae such as marine Roseobacters are linked with DMSP cleavage eg. model species *Ruegeria pomeroyi*.

(\*Put Marinobacter into a tree and see if they are related to the Bonney or Suribati-Ike types to indicate if they are nitrate reducers or DMSO reducers).

**CNS cycles**. The majority of the genetic potential for known C, N and S conversions was restricted to the 0.8 and 3.0 µm size fractions indicating they may perform the main chemical processes in the lake. The lack of ascribed functional genes in the 0.1 µm may also reflect abundance of candidate divisions which likely do not have homologs with known functions in sequence databases. Anaerobic processes, such as fermentation, anaerobic carbon fixation, carbon monoxide oxidation were clearly overrepresentation in the monimolimnion (\*import into PRIMER and do statistical test). Aerobic processes such as aerobic respiration and aerobic carbon fixation were clearly more abundant in the mixolimnion. Notably, Genes for enzymes involved in methanogenesis, nitrification and sulfur oxidation were not detected. Overall genetic potential for assimilation and mineralization was abundant but potential for fixation was scarce. For example, there were many more genes for aerobic respiration, CO oxidation and fermentation than carbon fixation (figure: CNS\_cycles).

**Proteorhodopsin** Diverse rhodopsin genes were detected in Organic Lake that broadly clustered with rhodopsins related to Gamma and Alphaprotebacteria types and Flavobacterial types (figure: rhodopsin\_tree). The most abundant rhodopsin types were Organic proteorhodopsin type I, which did not cluster with any sequences in Genbank non-redundant database; *Marinobacter* sp. ELB17 type, which is an Antarctic strain isolated from the hypersaline Lake Bonney, McMurdo Dry Valleys and Flavobacteria type, most closely related to *Psychroflexus.* There were also three lower abundance types related to *Octadecabacter*, xanthorhodopsin of *Salinibacter* and an actinorhodopsin related to *Aquiluna*. The relative abundance, distribution down the water column and size distribution of the rhodopsin genes agrees with the taxonomic composition of Organic Lake. For example, *Marinobacter* was abundant, concentrated on the 0.8 µm filter and was relatively enriched in the mixolimnion as is the *Marinobacter* related rhodopsin. This suggests that the Organic proteorhodopsin type I originated from *Roseovarius* because it was also an abundant bacterial genus and was concentrated on the 3.0 µm fraction. \*Have *Roseovarius* been reported to have rhodopsins? What about Roseobacters in general? Any other way to link to its origin? eg. nucleotide bias? GC content? BLAST vs enviroNR?

Rhodopsin in Flavobacteria and *Vibrio* has been associated with light dependent energy generation (\*ref), especially under low carbon conditions. If it fulfills a similar role in Organic Lake, this would indicate a certain extent of mixotrophy in the dominant, largely heterotrophic lineages present. Mixotrophy has been hypothesized to be prevalent in Antarctic nanoflagellates (\*ref Laybourne-Parry mixotrophy) but seems to apply to bacteria as well.

**Stickland fermentation** Dissolved N is limited compared to C, but least limited at 6.5 m. There is also a peak of ammonia at 6.5 m which may be due to increased fermentation at that depth. Stickland fermentation is one pathway detected in Organic Lake that could account for the ammonia production in the deeper samples.

**Dimethylsulfide metabolism**

Genes for DMSP lyases *dddD*, *dddL* and *dddP* were detected in Organic Lake of which *dddD* was the most abundant. It was most abundant at the bottom of the lake suggesting the high concentration of DMS that has been detected is due to breakdown of DMSP, likely formed from algal exudates. Like the rhodopsin genes, the putative taxonomic origin of the *dddD* genes is consistent with the taxonomic composition of Organic Lake. The most abundant types were a *Marinobacter* type and one distantly which clustered with *Halomonas* but not with high bootstrap score*.* The diversity of *dddD* genes is not yet well sampled so a link to taxonomy is not as conclusive.

Since the bottom waters do not mix with the surface, physical dispersal would be much hampered and in the absence of biological breakdown, DMS could potentially accumulate. Usually methanogens or sulfate reducers mediate breakdown in anoxic conditions. Since these were not detected, faster rates of DMSP production than DMS degradation may account for the high concentration in the bottom water. Alternatively, other anerobic 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

**Physicochemistry**:

**stratification stability** The water column structure of Organic Lake has varied over the 30 years during which it has been monitored. When first observed between 1978 and 1984 (Deprez *et al.* 1986; Franzmann *et al*., 1987b), it was considered meromictic due to the stable bottom temperatures of approximately −6 ºC, increased density and anoxia of water below 5 m. During that time, the water level was increasing causing a lens of fresher surface water separating effectively insulating the midwaters from contact with ice and allowing a midwater heat-trap. Between 1989 and 1994 the water level dropped 0.81 m, waters temperatures fell to −12 ºC in the monimolimnion, the mixolimnion penetrated to 4 m and the degree of stratification was reduced (Figure X)(Gibson 1996). The water column structure from this study is similar to that of the 1990’s and further shows that the once anoxic bottom waters have become suboxic indicating an oxygenation event sometime in the last 20 years.

**Peak activity at 6.5 m** The pycnocline at 6.5 m precludes mixing of the bottom waters. The pycnocline coincides with a drop in dissolved oxygen content likely depleted due to the activity of aerobic heterotrophic bacteria. This depth also shows a decrease in pH as fermentation can occur in the suboxic waters. Nutrients except for nitrate and nitrite were highest at the oxycline but above the sediment. Cell counts were highest at this depth too, consistent with this particular zone possessing relatively higher biological activity. These processes include anaerobic carbon fixation, fermentation (including Stickland fermentation), CO oxidation, \*methane oxidation??? (hits to methane MO occur at the bottom of the lake, could be another sort of MO such as DMS MO? AMO?) and DMSP lysis. Conversely, assimilatory sulfate reduction is lowest here, perhaps because they can assimilate sulfur directly in the DMSP/DMS or amino acid fermentation metabolism.

Franzmann *et al.* (1987b) were not able to successfully isolate any anaerobic bacteria which may be due to periodic oxygenation or microaerophilic nature of the bottom waters preventing anaerobes from establishing permanently.

**Cellular life** in the water column of Organic Lake is vertically stratified to a limited extend varying in relative abundance rather than in composition. Variation in the bacterial population down the depth profile mainly occurs in the 0.1um size fraction while the composition of the larger size fractions remained fairly homogeneous. Candidate divisions and Alphaproteobacteria are proportionally more abundant below 5.7 m. There is stratification within the monimolimnion with 6.5 m representing a local zone of higher productivity, probably due to nitrate reduction. Eucarya are relatively more abundant in the surface above 6.5 m which is consistent with the high proportion of phototrophs (*Dunaliella*) requiring greater access to light and heterotrophic nanoflagellates requiring more oxygenated waters.

(\*discuss: Franzmann paper that they are Dunaliella are dominant and choanoflagellates are present but no Chaetoceros. Chaetoceros may be transient members of the population from an ice community because (Wright and Burton, 1981) made no mention of them either. Composition is quite different from 2006 samples in the way that there are no prasinophytes like pyramimonas detected and Dunaliella is at much higher numbers. Mention that silicoflagellates were not previously detected in Organic, first report of dictyochophyceae in Antarctic lakes was from Unrein 2005, so may be important in Antarctic but were often missed. Fungi and ciliates being in small size fractions is perplexing. Fungi found in Bielewicz 2010 and Unrein 2005. Discuss the possible succession of eucarya in the lake. Perhaps link to Fedes models of strain cycling due to viral pressures. Also potential link to the polar night transition (Bielewics 2010) that some taxa a more light tolerant. Eucarya occupying a main role as primary producers. How do the heterotrophs survive the low oxygen??)

Diversity indices show that different sample depths are not that different from one another. However, species richness estimates are much higher than the values calculated from PCR amplification of 16S rDNA gene of Organic Lake sediment (Bowman *et al*., 2000b) and Lake Bonney deep waters (Glatz *et al*., 2006) (table X). This is mainly due to the very high number of OTUs calculated for this study (100s vs 10s) which is likely an artefact of how the OTUs were assigned. This is because the diversity is calculated from OTU table so when you use pick\_otus.py without suppression of those that are non-reference sequences, all the other reads form OTUs, and since these new OTUs are not easily able to assemble into a single type because they span different parts of the SSU gene, this greatly inflates the apparent number of OTUs. Also the inclusion of the euks and potentially the plastids increases the species richness. The primers used in the Bowman 2000b paper were at least biased towards 16S but they certainly excluded the euks –try QIIME alpha diversity calculations without the euks. 3. Using the short sequences makes the diversity apparently higher because two things that would be the same ribotype are now being split into two – try QIIME with a larger size cut-off but there’s not much that can be done about that. All these problems are to do with how species are delineated. The only solution is to attempt to delineate species in exactly the same way as Bowman. However an estimate of total richness should be comparable?\*Check out the phylotypes and see how different they really are).

**nutrient cycling:** The pattern of low nitrate and high ammonia at the oxycline is consistent with nitrate reduction under suboxic conditions, which occurs in other Antarctic lakes such as the west lobe of Lake Bonney (Voytek *et al*., 1999). Several taxa identified in Organic Lake (see cellular diverisity below) are related to Antarctic bacteria capable of nitrate reduction. However, 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, Burton? and Bonney, Vida 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.

**Bacteria isolates from Organic Lake:**

Pigmented bacteria ACAM 1 and 2 isolated in 1984 and in 1986 a further 14 strains were isolated from various depths of the lake (2 – 5 m) (Franzmann *et al*., 1987b). These were taxonomically characterised (Dobson *et al.*, 1991). All were Gram-negative rods, non-motile by flagella, did not produce gas from nitrate, could not grow anaerobically without nitrate and could not hydrolyse chitin. Growth was stimulated by inositol, arabinose, methionine, isoleucine, leucine and valine. The strains clustered into two phena: 1 and 2. Phenon 1 were oxidase and catalase positive, able to reduce nitrate to nitrite, able to hydrolyse starch and esculin. One strain, ACAM 554 could grow anaerobically with nitrate. Salt tolerance range was similar to *Halomonas* species. Cells were 1.2–11.5 µm long. Growth was stimulated by arginine, gluconic acid, pyruvate, maltose, orhithine and lactose ie. amino acids, sugars and organic acids. Phenon 2 were phenotypically more diverse. They were of a similar cell size and salt tolerance range to group 1. All members were able to hydrolyse starch and esculin, were oxidase positive. They were not able to reduce nitrate to nitrite (Dobson *et al*., 1991). Sequencing of the 16S rRNA gene of the two groups identified them as two new species: *F*. *salegens* (ACAM 48 ) and *F. gondwanense* (ACAM 44) which correspond to phenon 1and 2 respectively (Dobson *et al*., 1993). (\*See the paper for full listing of biochemical properties). Subsequent phylogenetic analysis reclassified *F. gondwanense* to a new genus *Psychroflexus* within the *Cytophaga-Flavobacteria-Bacteroidetes* (CFB) group along with sea ice isolate *P.* *torquis* (Bowman *et al*., 1998). *F*. *salegens* was reclassified as *Salegentibacter salegens* (McCammon & Bowman, 2000).

**Carbon fixation**

There are in total 6 autotrophic carbon fixation pathways known.

1. Calvin cycle, plants/cyanobacteria, alpha and beta proteobacteria eg. Rhodobacteraceae are purple non-sulfur bacteria while purple sulfur bacteria are gamma proteobacteria Chromatiales.
2. Reductive citric acid cycle aka reverse TCA cycle aka Arnon Buchanan cycle. Anaerobic and microaerophilic bacteria eg. in *Chlorobium*.
3. Reductive acetyl-CoA aka Wood-Ljungdahl. In anaerobic bacteria and archaea such as methanogens and acetate-producing bacteria eg Clostridium like *Moorella thermoacetica*. Carbon dioxide is reduced to carbon monoxide which is converted to acetyl-CoA via CO Dehydrogenase and acetyl-CoA synthase.
4. 3-hydroxypropionate bicycle. Only in green non-sulfur bacteria eg *Chloroflexus.*
5. 3-hydroxylpropionate/4-hydroxylbutyrate cycle. Aerobic archaea *Metallosphaera sedula.*
6. Dicarboxylate/4-hydroxybutyrate cycle. Anaerobic archaea *Ignicoccus hospitalis*.

**Stickland reaction**

Stickland reaction is the fermentation of amino acids as a sole carbon and energy source, generally in protein rich environments but may be used even when protein biosynthesis is impaired. This pathway is the oxidation of one amino acid coupled to the reduction of another. The amino acid acting as the electron donor is oxidised to a carboxylic acid one carbon shorter than the original amino acid. eg. Alanine is converted to acetate. The electron acceptor is reduced to a carboxylic acid of the same length. All amino acids can be donors and acceptors except histidine?

Fonknechten *et al*. (2010) describe in the genome of *Clostridium sticklandii* DSM 519. DSM 519 can oxidize threonine, arginine, lysine and serine while reducing glycine and proline. The arginine can be converted to ornithine which is then disproportionated to act as both oxidant and reductant. In the reductive pathway, ornithine can be reduced directly L-proline ornithine cyclodeaminase and then proceed to 5-aminovaleric acid (aka 5-aminopentanoate) via **D-proline reductase**. The reductive pathway can also proceed via ornithine aminotransferase adn PCA reductase. In the oxidative pathway L-ornithine is converted to D-ornithine by racemase and then oxidized to acetate, D-alanine and ammonia. Glycine can also be oxidised and reduced. In the oxidation pathway glycine is oxidised by NAD+ into methylene–THF, CO2 and ammonia via the glycine cleavage system. In the reduction pathway, it is reduced to acetyl phosphate by **glycine reductase** (Stickland). Amino acids are utilized in a order of preference arginine, serine, threonine, cysteine, proline and glycine are rapidly used. lysine, histidine, asparagine and valine disappear during stationary phase. Aromatic and branched chain amino acids can be degraded by an unknown pathway. Glutamate and alanine are not utilized and are excreted. Threonine can be processed three ways. (i) Threonine dehydratase oxidises threonine into 2-amino-3-ketobutyrate which proceeds to glycine and acetyle-CoA. (ii) Threonine aldolase converts threonine into glycine and acetylaldehyde. (iii) Threonine dehydratase converts threonine into ammonia and 2-ketobutyrate. Arginine proceeds to arginine deiminase pathway to ornithine via citrulline. Other bacteria that encode genes in amino acid degradation are *Alkaliphilus oremlandii* and *A. metalliredigens* as well as other *Clostridium* species. Genes for glycine reduction (grdA,B,C,E and X) were also found in *Photobacterium profundum* chromosome 2 (Vezzi *et al.*, 2005). Selenoproteins are proteins where cysteine is replaced by selenocysteine (sec) encoded by UGA stop-codon. Complex machinery is required for translations insertion. Glycine reductase A was the first selenoprotein discovered. Selenocysteine requires a selenocysteinyl-tRNA. Clostridia are obligate aerobes because oxygen inactivates enzymes such as pyruvate ferredoxin oxidoreductase that have iron-sulfur sites. Small amounts of oxygen can be tolerated by use of oxygen detoxification enzymes such as peroxide repressor (PerR), Mn-superoxide dismutase, superoxide reductase, alkyl hydroperoxide reductase, rubrerythrin, glutathione peroxidases, seleno-peroxiredoxin and thioredoxin-dependent peroxidase.

*Clostridium propionicum* can ferment L-alanine to ammonia, CO2, acetate and proprionate.

## 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.

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.

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.