Organic Lake

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

# Potential Aims and Scope

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

# Abstract (250)

Organic Lake is a meromictic (permanently stratified) hypersaline lake in the Vestfold Hills, East Antarctica that contains only microbial life and the highest concentration of dimethyl sulfide recorded in a natural body of water. The relatively reduced complexity of this ecosystem makes it an ideal site to link microbial taxa and functional diversity with the unusual lake chemistry. This study utilized an integrated metagenomic approach to describe Organic Lake from a whole ecosystem perspective. The microbial community was vertically stratified with Chlorophyte algae, Flavobacteriaceae, Rhodobacteraceae and Gammaproteobacteria abundant in the mixolimnion and Halomonadaceae, Psychromonadaceae, candidate divisions OD1 and RF3 were overrepresented in the monimolimnion. The functional potential for C, N and S cycling was also partitioned by depth. Potential for mineralization and assimilation of C and N were much more prevalent than fixation, consistent with the high abundance of heterotrophic bacteria and indicative of exogenous nutrient input. Rhodopsin genes, related to *Marinobacter* and *Psychroflexus* implies a potential for phototrophy in the most abundant heterotrophic genera in the lake. Organic Lake is characterized by the absence certain pathways such as nitrification, methanogenesis, sulfur oxidation and sulfate reduction reflecting strong environmental constraints that likely accounts for the accumulation metabolic products at depth, including dimethyl sulfide.

# Key words

Metagenomics/Antarctic lakes/Virophage/Vestfold Hills

# 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 location to study questions of biogeography, evolution and a potential source of novel life and genes. Organic Lake is known to have high levels of reduced sulfur compounds and a member of the newly discovered virophage family that may mediate ecosystem stability. The ability to encapsulate a large proportion of the species diversity allows us to infer which taxa may be mediating particular biological processes. 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.

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

## Background on sample site

Organic Lake is located on Long Peninsula in the Vestfold Hills, an ice-free oasis on the eastern shore of the Prydz Bay, Princess Elizabeth Land, 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 years before present (Bird *et al*., 1991) and the water has since concentrated to approximately six times the salinity of seawater. Hydrogen sulfide is absent in the anoxic bottom waters, likely due to the absence of sulfate reducing bacteria (Gibson *et al.*, 1991). However, DMS, DMSP and other polysulfides have been recorded at high concentration in the bottom waters (Deprez *et al*., 1986;Franzmann *et al.*, 1987; Gibson *et al.*, 1991; Roberts & Burton 1993). DMS and DMSP concentrations vary greatly throughout the year (Gibson *et al.*, 1991; Roberts & Burton 1993) potentially indicating active turnover. Phototrophic sulfur oxidising bacteria are also absent (Burke & Burton 1988) indicating sulfur cycling is not mediated by sulfur bacteria.

## Sample collection and preparation

Water was collected on 10 November 2008 along a depth profile when the lake was covered by a 0.8 m layer of ice through a 30 cm hole above 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. Water was also collected at the same sample depths for microscopic and chemical analysis and frozen −80ºC. An *in situ* vertical profile of pH, conductivity, turbidity, dissolved oxygen (DO), pressure and depth 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 depth sample corresponded to the turbidity maximum and 6.5 m 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 using the formula described by Fofonoff and Millard (1983). However, salinity was likely underestimated as the conductivity to salinity relation holds for practical salinity between 2 and 42 and Organic Lake salinity was much higher. \*Density was estimated from the practical salinity using the equation of Millero *et al.* (1980).

Nitrate, nitrite, ammonia, total nitrogen, dissolved nitrogen, dissolved reactive phosphorus (DRP), total phosphorus, dissolved phosphorus, total organic carbon (TOC), total dissolved carbon (DOC), total sulfur and dissolved sulfur were determined by the Analytical Centre\* (Tasmania).

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. Nitrate, nitrite, ammonia, DRP, DOC, dissolved N, dissolved P and dissolved sulfur 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.

DNA was extracted from the membrane 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 processed to remove low quality bases as previously describes (Lauro *et al.*, 2011). \*Total protein was extracted from the membrane filters as previously described (Ng *et al.*, 2010; Lauro *et al.*,2011).

## 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 with 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 from Bacteria, Eucarya, Archaea and plastids 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). QIIME was then used to choose a representative sequence from each set of OTUs and classify the representative set to the family level by 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 16S assemblage between each sample filter 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 and \*LINKTREE analyses were performed considering following abiotic variables: conductivity, temperature, turbidity, DO percent saturation, pH, total organic carbon, total nitrogen, total phosphorus, total sulphur, 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. \*LINKTREE takes the subset of variables that explains the biotic structure and finds the thresholds that maximizes the separation of the assemblage into groups by binary separation. Significance of the LINKTREE separations tested with SIMPROF. A heat map with biclustering dendogram was generated using R and the package ‘seriation’ (Hahsler *et al*., 2008) of the cellular composition.

\*Reads with an RDP classification with bootstrap value below 85 % at the phylum level were flagged for manual phylogenetic classification as they likely belong to novel lineages. 16S rDNA gene sequences were aligned using the NAST aligner (\*ref), their 7,682 bp alignments were imported into the Greengenes ARB database (release date\*) while 18S rDNA were fitted to a custom ARB tree. The phylogenetic affiliation was determined by position within the phylogenetic tree created in ARB by neighbor-joining (Ludwig 2004).

### 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. BLAST table 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 dimethylsulfonioproprionate lyases were retrieved from National Center for Biotechnogy Information (NCBI) ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) sequence databases. Only 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.

# Results

## Physical and chemical properties of Organic Lake

Organic Lake had a maximum depth of 6.75 m (figure: bathymetry) and measured 3.874 m above mean sea level during sampling on November 2008. Abiotic parameters measured *in situ* and estimated practical salinity are shown in figure \*X (figure: chemical\_profiles). 0–5.7 m was designated as the mixolimnion as physico-chemical properties were similar between these depths indicating mixing. Below 5.7 m was designated as the monimolimnion as DO decreased while salinity and temperature increased indicative of a pycnocline preventing mixing. A division between the mixolimnion and monimolimnion was supported by PCA analysis, which described 89 % of the variation in physico-chemistry of the samples (figure: PCA). Samples separated along the PC1 axis (74.3% of variation) with depth. Curiously, 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 possibly due to microbial processes. The 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.

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. \*Read about Redfield ratios for lakes

## Cellular diversity and distribution

A total of 3 959 reads matching to SSU were retrieved from the November 2008 profile. Of those reads, \*X were classified as bacterial or chloroplast compared to a 2 archaeal sequences indicating Bacteria were numerically dominant in Organic Lake. Direct microscopic counts of cells in the size range typical of Bacteria or Archaea (figure: microscopy images) were taken to be Bacteria.

**Bacteria**

The majority of bacteria were aerobic heterotrophic groups except for a very small number of cyanobacteria detected. The dominant bacterial classes were Gammaproteobacteria and Flavobacteria.Flavobacteria consisted primarily of *Psychroflexus*, \*more? Previously, several strains of *Psychroflexus* have been isolated from Organic Lake. These are known to be\*. Gammaproteobacteria were primarily of the genus *Marinobacter*. Alphaproteobacteria were also fairly abundant, especially in the 3.0 µm fraction, and were primarily of the family Rhodobacteraceae. Other bacterial classes present in all samples include Actinobacteria, chloroplasts, Cytophagia, Sphingobacteria, Opitutae, unclassified cyanobacteria, OD1 and RF3 (\*table: bacteria\_all\_profile.).

**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) (\*Check the choanoflagellate in Organic Lake taxonomy) 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 6.8 m.

**Vertical stratification of the cellular community.** The cellular community composition varied according to both size fraction and depth (figure:family\_profile3\_sr\_hmap). Samples from the same size fraction and the same stratum formed clusters. The 3.0 µm samples were overrepresented in Flavobacteriaceae and Rhodobacteraceae. This cluster was further divided into mixolimnion and monimolimnion groups where the mixolimnion showed a greater abundance of *Dunaliella* sp. chloroplasts and Chlamydomonadales. This is consistent with the larger size fraction containing mostly large phototrophic algal species and thus localizing to the area of higher light and associated heterotrophic bacteria that metabolise algal exudates. The 0.8 µm monimolimnion samples were compositionally the most different from the rest of the samples. This was due to an increased abundance of candidate division RF3, Halomonadaceae and Psychromonadaceae. In comparison, the 0.8 µm fraction of the mixolimnion was overrepresented in Gammaproteobacteria and Oceanospirillales. The 0.1 µm monimolimnion cluster was mainly distinguished by the presence of candidate division OD1.

ANOSIM analysis showed a statistical difference in cellular composition (Rho: 0.434, significance: 0.2%) between mixolimnion and the monimolimnion. SIMPER analysis (\*figure: SIMPER\_community) identified the groups that contributed to variance in the mixolimnion were:chlorophyte algae, choanoflagellates, dinoflagellates, silicoflagellates, Flavobacteriaceae, Oceanospirallales, Alteromonadaceae and Gammaproteobacteria. The taxa overrepresented in the monimolimnion included Rhodobacteraceae, RF3, OD1, Psychromonadceae, Halomonadaceae, Desulfobacteraceae and TM7. (\*look up what these genera do, what metabolic class? Linked to sulphur? Linked to ice and algae?).

**Vertical taxonomic variation is correlated with abiotic factors:** Variation in the cellular population structure between samples was significantly correlated (0.519 R-value, 0.3% significance) with the abiotic paramters: DO%, temperature, total sulfur and total nitrogen. This indicates the bacterial community structure was to some extent determined by these factors. (\*RELATE to the species composition?).

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). (\*Calculate percentage of diversity sampled?) 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.

## Whole system functions

**What can we piece together about nutrient cycles in Organic Lake from the taxa we know to be present?**

The main production is mediated by photosynthetic eucarya. There are few cyanobacteria, and no anoxygenic photosynthetic bacteria. Most bacteria are heterotrophic aerobes. eg. Flavobacteria and Gammaproteobacteria. Futhermore 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. Flavobacteria are associated with the eucaryotic algae as they consume algal exudates. The Gammaproteobacteria consume more labile products. The overrepresentation of Rhodobacteraceae, RF3, OD1, Psychromonadceae, Halomonadaceae, Desulfobacteraceae and TM7 in the monimolimnion means they are likely linked to processes connected to the deeper waters. These include: production or lack of degradation of ammonia (eg. ammonification) or fixation. In the carbon cycle these taxa are likely related to fermentation, anaerobic carbon fixation and CO oxidation (\*link those processes with taxa). With regards to the sulfur cycle, they are likely involved in DMSP lysis that produces DMS.

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. Look for nitrate reductase or DMSO reductases.

**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. Similarly,

**Proteorhodopsin** The majority of both these taxa have proteorhodopsin-like genes which may allow a certain extent of phototrophy.

**Stickland fermentation**

**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. DMS is removed from water due to \*evaporation or biological degradation. 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. 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.

(\*Look for nitrate reductase, assimilatory NAS, respiratory NAR and periplasmic Nap? bacteriochlorophyll vs chlorophyll).

**What can we piece together about nutrient cycles from the Proteome?**

**Do any taxa co-vary over depth? Does the viral composition vary with and cellular taxa?**

## Viral diversity and distribution

**What viruses are present in Organic Lake profile?**

**Does the viral composition vary with depth?**

**Is the viral composition correlated with any physical or chemical parameters?**

# Discussion

**Physicochemistry**:

**stratification stability** The limnology of Organic Lake has varied over the 30 years during which it has been monitored. Changes in temperature, salinity and water column structure have been related to variations in water level (Gibson, 1995; Gibson, 1996). When the limnology of Organic Lake was first monitored 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. An oxycline was recorded at 3.5 m. Between 1989 and 1994 the water level dropped 0.81 m, temperatures fell to −8 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 but the mixolimnion has moved even deeper and water temperature is even colder.

**Peak activity at 6.5 m**The highest concentration of DMS has been found to occur just below the oxycline but above the sediment (Deprez *et al.*, 1986;Franzmann *et al.* 1987; Gibson *et al.*, 1991). This was not the case with a study by Roberts & Burton (1993) where DMS was found to occur at highest concentration in the sediments. Although this study did not measure DMS, all other 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 being a region unusual chemically due to relatively higher biological activity. This zone may represent a region where DO and redox potential have decreased while temperature and salinity have increased to levels such that anaerobic processes to dominate. These processes include anaerobic carbon fixation, fermentation (including Stickland fermentation), CO oxidation, \*methane oxidation??? 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.

Solution1) just accept the reference OTUs, start with longer seqs, lower identity threshold, remove as much crap as possible from the OTU listings eg mitos. Check by blasting some of the mitos? Solution2) Suppress new OTUs, but then assign the reads to OTUs using RDP. Solution3) If the new OTUs are not being aggregated, use a tiered procedure eg. align to reference, then aggregate the remaining reads. Solution4) De novo assignment of OTUs using UCLUST. This is problematic because you will get artificial inflation of the OTUs that are rare and span different regions of the SSU gene.

(\*Explanations for the differences in the diversity indices: 1.the inclusion of the euks and potentially the mitochondria increases the species richness – try QIIME diversity calculations with just 16S. 2. 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:** Thirty-one Gram-negative rod-shaped bacterial strains were isolated from Organic Lake in 1984: ACAM 1–31 (Franzman *et al*. 1987b). ACAM 3–31 were non-pigmented bacteria similar to *H*. *elongata*. These putative *Halomonas* strains were all catalase and oxidase positive and clustered into three groups, A, B and C on the basis of biochemical properties. Nine of the 11 group C isolates produced 10 µm filaments as well as short rods throughout their growth cycle. Some were motile by polar flagella, some by cell flexing. ACAM 13, 15, 26 could reduce nitrate to gas anaerobically. Nitrate is reduced to nitrite by most strains. Growth of these strains was not stimulated by glucose or other sugars but it was by a number of amino acids (alanine, arginine, lysine, threonine, proline) as well as organic acids succinate and lactate. They could not produce hydrogen sulfide. Growth was possible in salt concentrations from 0.1 to 20% (w/v). The proposed species name is *H. subglaciescola* and the type strain is ACAM 12 (Franzman *et al*., 1987a). Other non-pigmented strains were isolated from Organic Lake and surrounding saline lakes (James *et al*., 1990) and determined by biochemical testing to be a new species, *H. meridiana.* The type strain ACAM 246 (biovar II) was isolated from Burch Lake and the reference strain ACAM 233 (biovar I) from Organic Lake. These strains were rod-shaped, 1.9–4.5µm in length, sometimes occurring in pairs and produced 1–2 lateral flagella. They were catalase and oxidase positive, aerobic and do not reduce nitrate to nitrite. It was described as haloversatile to describe how although optimal growth in media occurred in salt concentrations of ~3%, it can tolerate up to ten times this concentration.

Pigmented bacteria ACAM 1 and 2 isolated in 1984 (Franzmann *et al*., 1987b) and 1986 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 44 ) and *F. gondwanense* (ACAM 48) 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).

**RF3:** RF3 is a candidate division recognised in SILVA that was first identified from a bovine rumen clone library as related to low GC Gram positive (Firmicutes) bacteria (Tajima *et al*., 1999). It has been identified associated with hypersaline waters (Yilmaz *et al*., 2012) out of the Global Ocean Sampling (GOS) Expedition dataset (Rusch *et al*., 2007), hypersaline lake Mono(\*ref), hypersaline lake Tebenquiche (Salar de Atacama) (\*ref), leachate from solid waste landfill (\*ref), compost (\*ref), crude oil (\*ref), anaerobic reactor (\*ref) as well as the gut.

**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.

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, 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.

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.

\*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, Burton HR, Pitson GA. (1993) 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.

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.

## Extra 16S stuff

**All sample MDS:** Patterns in the bacterial composition of all Organic Lake samples filters were visualized using MDS and CLUSTER analyses. A broad division was evident between 0.8 and 3.0 µm bacterial compositions and the 0.1 µm population. The 0.1 µm samples were the most divergent from one another with the December 2006 surface water sample (GS233-0.1 µm) clearly the most divergent. The sample taken in December 2008, although taken at approximately the same time of year as the 2006 sample, tended to cluster more closely with the profile samples (GS374-GS379). Differences in composition could be seen according to depth with the deeper profile samples (GS377-0.1 and GS378-0.8) grouping together but these spatial differences were much smaller. Not much variation within the 0.8 and 3.0 µm cluster was evident except that the 2006 sample (GS233-0.8) was again the most divergent and the GS377-0.8 sample also separated.

**GS233:** Within GS233, community composition on the three size fractions differed taxonomically and in relative abundance. 0.1µm had the lowest hit rate, followed by the 3.0 µm fraction and the 0.8 µm fraction although the 3.0 µm filter was more species rich\*. The 0.1 µm fraction was dominated by candidate division OD1 (\*what?), 0.8 µm by Flavobacteriaceae and the 3.0 µm by Flavobacteriaceae and Rhodobacteraceae. In 0.8 and 3.0 µm most of the Flavobacteriaceae was classified to the genus level as *Psychroflexus* but other (\*amount) of cryptic genera were present indication heterogeneity in the population*.* The abundance of *Psychroflexus* in 0.8 µm was the highest of all the samples. The 0.1 µm fraction also had a relatively high abundance of Flavobacteriaceae, but mainly of the genus *Persicivirga*. The presence of *Persicivirga* was unique this sample. Rhodobacteraceae were of diverse genera, those that could be classified included *Roseovarius*, *Loktanella* and *Sulfitobacter* but again, there were several cryptic genera. *Glaciecola* is characteristic of this membrane size.

**GS374:** In contrast, the surface sample taken during ice cover GS374 was dominated by *Marinobacter, Psychroflexus* (and relatives) and *Roseovarius*. The dominance of these clades varied with filter size with GS374-0.1 and and GS374-0.8 clearly dominated by *Marinobacter* while the 3.0 um filter had a larger proportion of *Psychroflexus* and *Roseovarius*, although *Marinobacter* was still quite abundant.

**GS375:** The dominant genera on GS375 followed a similar pattern to GS374 with the exception of *Roseovarius* being slightly more abundant in the 3.0 um fraction than *Psychroflexus*.

At the bottom of the profile GS378, the dominant 16S reads on the 0.1um fraction are now candidate division OD1. The larger size fractions are again dominated by *Marinobacter* and *Psychroflexus.*

*Psychroflexus* appears to dominate in the samples that are partially ice-free, or totally ice-free. It is by far the dominant bacterium in GS233 and GS379. It is mainly found on the larger size fractions. When it is dominant, it is primarily found on the 0.8 um fraction. *Marinobacter* appears to vary inversely with *Psychroflexus*. It is found to dominate the samples from the profile and are always more numerous on the 0.8 um filter. Chloroplasts? *Psychromonas* occurs in the deeper samples and reach larger populations in the partially ice free resample. Not surprisingly *Bacillariophyta* chloroplasts are mainly associated with surface samples \*up to 3.5 m and on the bottom, probably due to sedimentation. In the surface, *Bacillariophyta* could be inversely proportional to the *Chlorophyta*.

*Roseovarius* is consistently more abundant on the 3.0 um fraction of a sample. Its abundance between samples doesn’t appear to vary much except that it is most dominant in the profile. Chlorophyta seems to vary with depth and doesn’t appear to be the most dominant in the bottom, or when the lake is completely thawed.

**Is total bacterial abundance related to any physical parameter in the profile?** The relative abundance of bacteria estimated from 16S frequencies in the metagenome showed counts from 0.1 µm size fractions were consistently lower than the 0.8 and the 3.0 µm size fractions (Figure 4, Figure S3). 16S counts from the larger size fractions were similar (\*test) to one another except in the December 2006 samples where there the highest counts were obtained from the 0.8 µm size fraction. Microscopic examination revealed the majority of bacteria were filamentous or chain forming (Figure 3), with an average cell size (\*) and were indeed be captured on the larger size fractions. This is consistent with morphologies of bacteria isolated previously from the lake (\*ref). The 16S counts also varied with turbidity in a size dependent fashion (Figure S3). Bacterial abundances on the 0.1 µm fraction varied inversely with turbidity. In the mixolimnion the bacterial abundances on the 0.8 µm and 3.0 µm size fractions varied directly with turbidity, although the 0.8 µm counts peaked at 4 m rather than at 5.6 m while in the monimolimnion only abundances in the 0.8 µm fraction varied directly (Figure S2). This suggests bacteria on the 0.8 µm and 3.0 µm size fractions contribute to turbidity in the mixolimnion but in the monimolimnion the 0.8 µm sized bacteria are the main contributors. Assuming that turbidity is directly related to bacterial density, this suggests bacteria on the 0.1 µm filters do not contribute significantly to cell density while the larger size fractions do. (\*test if cell counts varies with any of the envirodata. This may be related to Organic Lake being relatively copiotrophic? cells are motile and actively search nutrient particles and they have a cell size that varies with food availability, feast famine strategy. Chemical data to show it is copiotrophic?).