Organic Lake

# Potential Aims and Scope

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

# Abstract (250)

Organic Lake is a shallow (6.75 m deep) hypersaline lake in the Vestfold Hills, East Antarctica, which contains only microbial life. This study utilised a metagenomic approach to describe Organic Lake from a whole ecosystem perspective by integrating information of the biotic composition and genetic potential with physico-chemical factors. Physical and chemical parameters along a depth profile showed Organic Lake is vertically stratified, but not to the extent that it was in the past. The lake was divided into an aerobic mixed zone (mixolimnion) and a deeper stagnant zone (monimolimnion) with distinctive peak of dissolved carbon and ammonia at the oxycline (6.5 m), suggestive of higher microbial production potentially fuelled by nitrate reduction. The cellular and viral population structure was similarly vertically stratified. Taxa that were locally overrepresented at 6.5 m included candidate division Rf3 and *Roseovarius*. Phytoeucarya are the main primary producers.

# Key words

Metagenomics/Antarctic lakes/Virophage/Vestfold Hills

# Introduction

Antarctic biota live at the extremes of temperature and salinity under a polar light regime. The lakes are a rare source of liquid water and an oasis for life. They are ideal location to study questions of biogeography, evolution and a potential source of novel life and genes. The anoxic sediments of meromictic lakes are a record of climate change. The presence of exclusively microbial biota of reduced richness makes Organic Lake an ideal site to study microbial diversity and dynamics and its relationship with ecosystem function. Organic Lake is known to have high levels of reduced sulfur compounds and a member of the newly discovered virophage family that may be mediate ecosystem stability. The ability to encapsulate a large proportion of the species diversity allows us to infer which species 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.

In this study we aimed to:

1. Determine the microbial composition population structure of Organic Lake. Determine if microbial composition varies along the depth gradient and if so, which taxa differ.
2. Describe the functional capacity of the microorganisms in the lake.
3. Link microbial processes to phylogentic groups or to physical-chemical parameters.
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 6 × the salinity of seawater. Organic Lake was recorded to have an extremely high concentration of dimethyl sulfide (DMS) (Franzmann *et al.*, 1987) and other polysulfides (Roberts *et al*., 1993) in the bottom waters.

## 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 30cm 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. 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. Filters were placed in a storage buffer (\*composition) and kept frozen at -80ºC. Water was also collected at the same sample depths for microscopic and chemical analysis and frozen -80ºC. Samples for microscopy were preserved in formaldehyde (1% v/v).

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. Nitrate, nitrite, ammonia, total nitrogen, dissolved reactive phosphorus, total phosphate, total carbon, total dissolved carbon and total sulfur were determined by the Analytical Centre\* (Tasmania). Principal Component Analysis (PCA) was performed using the Primer V.6 statistical package (\*ref) on the (\*ranked?) normalized physical and chemical parameters to visualize how abiotic factors varied with depth.

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

Fixed water samples were vacuum filtered onto a 0.015 µm pore-size filter (Whatman\*) with a 0.45 µm pore-size backing filter. The 0.015 µm filter was mounted onto a glass slide with ProLong Gold (\*brand) and cells and virus-like particles (VLPs) stained with SYBR gold (\*brand) by adding 2 µl m (25 × dilution) into the mountant. All water used in filtration was autoclaved and filtered to <0.015 µm. Prepared slides were visualized in an epifluorescence microscope (Olympus BX61) under excitation with blue light (\* nm, filter set 2). Cell and VLP counts were performed on the same filter with over 30\* random fields of view. Cell density was calculated by \*.

## Diversity analyses

### Cellular diversity analysis

Cellular diversity 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 the 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 normalized 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 1993, Clarke 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. 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 from the November 2008 depth profile and the patterns in bacterial and chloroplast assemblages were correlated, BEST and LINKTREE analyses were performed considering following abiotic variables: salinity, temperature, turbidity, percent dissolved oxygen saturation, pH, nitrate, nitrite, ammonia, dissolved reactive phosphate, total carbon, total dissolved carbon, total nitrogen, total phosphorus and total sulfur. 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 (see above).

\*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

# Results

## Physical and chemical properties of Organic Lake

During the November 2008 sampling Organic Lake had a maximum depth of 6.75 m (figure: bathymetry) and measured 3.874 m above mean sea level. Water temperature was below zero (average of -13 ºC) and the estimated maximum practical salinity was 188 (figure:profiles\*cut off graphs to 1.5 m). From 1.5 m to approximately 5 m the physical and chemical parameters remained fairly constant indicating mixing in this zone. Below 5.7 m dissolved oxygen, salinity, temperature and pH varied continuously with depth. However, cell density, VLP density, ammonia, dissolved reactive phosphate, total C, total N, total P and total S were at a maximum at 6.5 m while turbidity, nitrate and nitrite were at a minimum (figure: profiles) indicating some stratification below 5.7 m. We designated 0–5.7 m to be the mixolimnion, as that appears to be the deepest extent of mixing ,and below to be the monimolimnion (\*density converter stability of stratification calculation). The separation between the mixolimnion and monimolimnion is evident through PCA analysis (\*redo PCA with nutrient data. figure: PCA physical), which shows the samples in the mixolimnion grouped together while the other samples separated along the PC1 axis with depth. PC1 described 75.1% of the variation with a roughly equal contribution from DO, salinity, conductivity, and pH. The stratification between 6.5 m and 6.7 m is also reflected in the PCA (\*)

Overall, nutrient analysis showed that Organic Lake appears oligotrophic (\*reference \*table of the nutrients and ratios from other lakes). The C:N:P ratio in the mixolimnion is approximately 60:5:1 and peaks in the monimolimnion at 22:3: 1 (table: sample enviroparams). This shows that nitrogen is limiting throughout the water column (compared to the ocean) but is least limited at 6.5 m indicating relatively high microbial production. 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 section 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 (\*so may represent relict activity). 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 (see below) and accumulation of sulfur compounds compared to similar, but less saline systems.

## Cellular diversity and distribution

A total of 4,359 reads matching to SSU were retrieved from the November 2008 profile. Of those reads, 4,075 could be classified to the domain level (\*make a table of the reads classified for each depth). Of the 16S rDNA sequences, 3,426 reads 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 3:microscopy images) were taken to be bacteria.

Diversity indices (table or figure: diversity indices) 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.

**Bacteria**

The majority of bacteria were from oxidative 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: *Marinobacter*. Alphaproteobacteria were also fairly abundant, especially in the 3.0 µm fraction, and were primarily of the genus *Roseovarius*. (\*complete table of taxa in all sample depths.)

Other bacterial classes present in all samples include Actinobacteria, chloroplasts, Cytophagia, Sphingobacteria, Opitutae, unclassified cyanobacteria, OD1 and Rf3.

The bacterial composition does not vary greatly down the water column. However, the relative abundance of some groups varied vertically (figure:hierarcical clustering). Verrucomicrobia, Actinobacteria and Candidate divisions OD1, TM7 and SR1 were most abundant in the 6.7 m sample. Alphaproteobacteria and Rf3 peaked at 6.5 m depth. Conversely, chloroplasts were largely restricted to 1.6 to 5.7 m. (\*Table of those taxa higher abundance in particular depths.)

(\*look up what these genera do, what metabolic class? Linked to sulphur? Linked to ice and algae?).

**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. Eucaryal abundance, more than composition, varied down the water column with the highest abundance restricted to the mixolimnion. This is consistent with the majority of eucarya being phototrophs or aerobic heterotrophs and thus localizing to regions of high light and oxygen respectively. Signatures in the monimolimnion may represent sedimenting cells.

(\*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??)

**Bacterial composition is vertically stratified. The smaller size fractions separate into mixolimnion and monimolimnion and within the monimolimnion there is additional stratification while the 3.0 fraction is only separated in the monimolimnion** MDS (figure:16sMDS) and CLUSTER (figure:16scluster) (\*try hierarchical clustering?) analysis of the bacterial and eucaryal community composition down the water column of Organic Lake showed samples separated according to both size fraction and depth. The 0.1 µm samples displayed the largest degree of separation with the mixolimnion samples (GS374-GS376) forming one cluster and the two monimolimnion samples (GS377 and GS378) another. However, the MDS (figure:MDS) shows separation between the monimolimnion samples. The 0.8 µm samples were similarly separated similarly to the 0.1 µm fraction but with less difference between the mixolimnion and the monimolimnion. All the 3.0 µm samples clustered together except the 6.5 m sample (GS377-3.0). This indicates the 0.1 and the 0.8 µm bacterial populations are stratified vertically with a further difference between the monimolimnion samples while the 3.0 µm community is only stratified in the monimolimnion..

**Is there a relationship between community assemblage patterns and environmental parameters measured?:** Bacterial assemblages were weakly, but significantly correlated (0.374 correlation, 1.3% significance) with turbidity, temperature, pH and dissolved oxygen. This reflects a correlation between based on depth related parameters (temperature, pH and dissolved oxygen) and turbidity. This indicates the bacterial community structure, is to a small extent, determined by these factors. (\*RELATE or BEST to relate enviro params to the species composition).

## 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?**

## Read-based analysis of diversity vs molecular markers

**How does read-based analysis compare to 16S/18S analysis in what is there? In terms of the relative abundances? and in terms of statistical differences?**

## Whole system functions

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

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

(\*Look for nitrate reductase, assimilatory NAS, respiratory NAR and periplasmic Nap? What about fixation?)

Conditions are suboxic. Was there nitrate reduction back in the day? Or DMSO reduction? Put Marinobacter into a tree and see if they are related to the Bonney or Suribati-Ike types. Look for nitrate reductase or DMSO reductases. Look for DMS cycling genes.

**What can we piece together about nutrient cycles in Organic Lake from the functional genes?**

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

## 

# Discussion

**Physicochemistry**: The mixolimnion is thermally and chemically stratified displaying two zones: one just beneath the ice to 1.5 m and the second from 1.5 m to 5 m. This is typical of meromictic Antarctic lakes in the early summer when mixing due to ice-formation and convection ceases in the mixolimnion (Gibson 1996).

The physiochemistry of Organic Lake has varied greatly over last 30 years during which it has been monitored. These changes, particularly 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 1978–1984 (Franzmann *et al*., 1987b), it was considered meromictic due to the stable temperature (-5 to -6 ºC), increased density and anoxia of water below 5 m. A thermocline and oxycline occurred at approximately 3.5 m. During this time, the water level rose from 6.5 m to 7.5 m and density at 3 m decreased, likely due to dilution of salts (Franzmann *et al.*, 1987b; Gibson, 1995). Between 1989 and 1994 the water level dropped 0.81 m, temperatures fell (minimum annual temperature of -16.5 ºC in the mixolimnion and -8 ºC in the monimolimnion) and the mixolimnion penetrated slightly deeper than 4 m (Figure X)(Gibson 1996). \*Other bathymetry files that you can’t open. \*density? Temperature and salinity profiles from this study are similar to those in the 1990’s so it is possible Organic Lake has remained somewhat stable since that time. However, there is no data available between 1997 and this study to definitively confirm this. 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.

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).

**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).

## References

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**, pp. 117–43.

\*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.

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, 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.

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.

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.

Roberts NJ, Burton HR, Pitson GA (1993) Volatile organic compounds from Organic Lake, an Antarctic hypersaline, meromictic lake. *Antarct Sci* **5**: 361–366.

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.

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.

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?).