

Supporting Information

Figure S1 Ace Lake and surrounds in the Vestfold Hills. A) Schematic of the Vestfold Hills adapted from Gibson (Gibson, 1999). The Vestfold Hills region is a coastal, ice-free, Antarctic oasis approximately 400 km² in area that is remarkable for the presence of hundreds of lakes, ranging in salinity from fresh to hypersaline (Gibson, 1999). Most of the saline lakes were originally pockets of seawater and retain a marine-derived biota (Gibson, 1999; Roberts & McMinn, 1999; Zwartz *et al.*, 1998; Rankin *et al.*, 1999; Crommer *et al.*, 2005; Andr assy & Gibson, 2007). B) Aerial photograph of Ace Lake and surrounds. Ace Lake (68.473°S, 78.189°E) was formed at the end of the Quaternary ~12,000 yr BP, and was initially fresh (Rankin *et al.*, 1999; Crommer *et al.*, 2005; Coolen *et al.*, 2004; Coolen *et al.*, 2006). Seawater invaded the lake basin during an early Holocene sea level highstand (~7,000 yr BP) that resulted from Holocene sea level rise outstripping isostatic rebound resulting from reduction of mass of the polar ice cap through melting. The lake was reformed after subsequent sea level fall (due to continued isostatic rebound after stabilization of the global sea level) after ~5,000 yr BP, and has remained saline since (Rankin *et al.*, 1999; Crommer *et al.*, 2005; Coolen *et al.*, 2004; Coolen *et al.*, 2006). The lake is exposed to air temperatures ranging annually from approximately -40 to +10°C, and the lake surface is ice covered for ~11 months of the year (some years not completely melting). Below the ice, the mixolimnion is mixed by brine exclusion during ice formation in winter, and wind turbulence occurring after the ice has melted may also contribute to mixing. The lake is cold ranging from ~0 °C in the mixolimnion to <3.5 °C in the monimolimnion. Supersaturated levels of oxygen are seasonally generated in the mixolimnion. The anoxic waters (14-23 m) support stable increasing gradients of salt (up to 4.3%), methane (saturated below 20 m, ~5mM) and H₂S (up to 8mM), and decreasing gradients of sulfate (essentially depleted below 19 m). The surface of Ace Lake is ~9 m above sea level, it has a maximum depth of 25 m, and is approximately 150 m from the nearest marine environment, Long Fjord. The drainage basin of the lake is small (35.5 ha), and consists of low rocky hills comprising Archaean gneisses with limited soil development. Ace Lake receives water from melt of snowbanks that develop in the lee of hills and boulders in winter, and direct capture of precipitation in summer. There is no evidence of groundwater inflow, and no surface or groundwater outflow, although the lake surface has been within ca 1 m of the sill separating the lake from Long Fjord since the 1980s. Few if any lichens or mosses occur within the drainage basin, and higher plants are absent. Adelie Penguins visit the area during summer, although none breed in the drainage basin. Snow Petrels and Wilson's storm-petrels have been recorded breeding, but only in low numbers (<10 pairs). C) Sampling site on Ace Lake, December 2006. D) Datasets, annotation and assembly metrics. E-H) Ace Lake depth-profiles December 2006. A Sonde probe (YSI model 6600) shown in Fig. S1C was used to record: (E) temperature (yellow, Celsius), pH (blue, pH units); (F) salinity (green, ppt); (G) dissolved oxygen (pink, % saturation); (H) turbidity (brown, NTU).

Figure S2 16S rRNA gene sequence analysis. The percentage composition of each clade in each sample is displayed as barplots in iTOL (Letunic & Bork, 2006) for the 0.1 µm (red), 0.8 µm (green) and 3.0 µm (blue) fractions. A) 5 m; B) 11.5 m; C) 12.7 m; D) 14 m; E) 18 m and F) 23 m. The proportion of each clade has been normalized for each size

fraction from each depth. The height of the bars represents the relative contribution of the clade and is comparable only within each size fraction from each depth. The SAR11 clade is the major contributor to the *Rhizobiales*.

Figure S3 18S rRNA and 16S-23S rRNA ITS sequence analysis. A) ARB neighbour joining phylogenetic reconstruction identifying the positions of 18S rRNA gene sequences retrieved from Ace Lake metagenome data. Number of gene copies relating to each organisms and the depth at which genes were identified are shown to the right. Microsporidia were used as the outgroup. "Ace" sequences shown in bold are from this study. All except four of the 18S rRNA gene sequences were retrieved from the mixolimnion; the four were from the oxycline and represented a novel clade of the Euglenozoa, the Symbiontida, which have been identified from marine anoxic environments, and contain epibiotic bacteria (Yubuki *et al.*, 2009). The proportion of eucarya in the mixolimnion estimated from reads and absolute numbers of rRNA genes was ~5% of the microbial population. This may over-estimate the number of eucarya as the rRNA counts have not been normalized for copy number which will be as low as a single copy for oligotrophic bacteria (*e.g.* SAR11 clade) compared with >10 copies for *Mesopedinella arctica* and >10,000 copies for dinoflagellates (*e.g.* *Akashiwo sanguinea*) (Zhu *et al.*, 2005). Eucaryal sequences in the mixolimnion represented eight different obligately or potentially phototrophic organisms of marine origin. These included three stramenopiles, two chlorophytes, the potentially karyokleptic coupling of alveolate *Myrionecta rubra* and the cryptophyte *Cryptophyta cryophila* (Johnson *et al.*, 2007), and the obligately heterotrophic metazoan, *Paralabidocera antarctica*. The data indicate that bacteria, and not eucarya, dominated the cellular population in the mixolimnion at the time of sampling. B) Association of SAR11 clade representatives determined from sequence variation of the 16S-23S rRNA ITS region. Ace lake contains ITS sequences associated with three distinct SAR11 clusters: SAR11 surface 1, SAR11 surface 3 Antarctic, and SAR11 deep (Brown & Fuhrman, 2005). SAR11 deep, contains sequences retrieved from deep (>50m – 3000m) marine samples (Field *et al.*, 1997). Sequences in purple have been retrieved from polar regions, while sequences in red have been retrieved from tropical regions.

Figure S4 Microbial cells and virus-like particles counted by epifluorescence microscopy. A) Counts were made for whole water samples of Ace-Lake as a function of depth. Error bars represent two standard deviations. No VLPs were observed at 12.7 m depth and the value reported represents the detection limit of the counting procedure (*i.e.* one cell detected in one field of view); the true number is likely to be lower than this. B) Microscopic analysis of different filter-size fractions of water samples from Ace Lake highlighting the efficacy of the size-fractionation protocol. The bar in each panel is 20 μm long. The absence of very small fluorescing VLPs in contrast to the presence of large numbers of small microbial cells of a single morphotype in all size fractions from the 12.7m depth is consistent with the absence of viral predation in this zone. The absence of larger filamentous cells in the smaller size fractions is evidence of the effectiveness of the size fractionation.

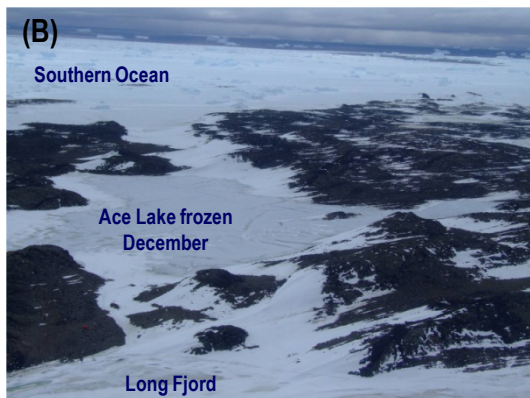
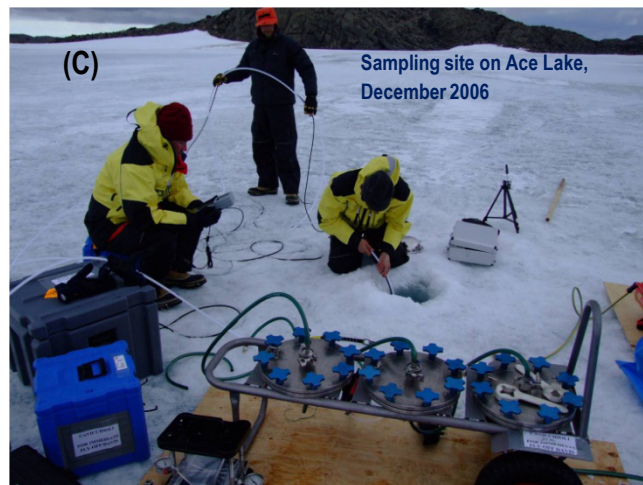
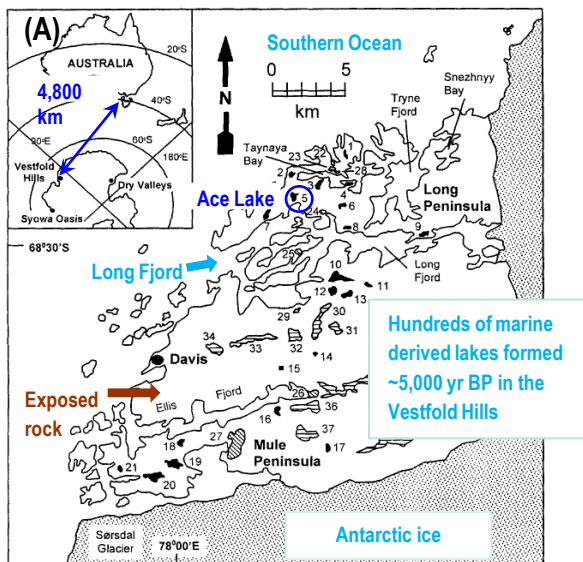
Figure S5 Analysis of trophic strategy according to the model of Lauro *et al.* (2009). The cluster location of the Ace Lake samples and of representative sequenced genomes are: Green (oligotrophic): *Pelagibacter ubique*, *Sphingopyxis alaskensis*, *Chlorobium chlorochromatii*, *Chlorobium phaeobacteroides*, *Chlorobium tepidum*, Ace Lake 12.7 m (Sanger, 454 0.1 μm , 0.8 μm); Blue (oligotrophic): *Planctomyces maris*, *Rhodopirellula baltica*, Ace Lake 23m (Sanger, 454 0.1 μm), Ace Lake 18 m (Sanger, 454 0.1 μm), Ace Lake 14 m (Sanger, 454 0.1 μm), Ace Lake 11.5 m (Sanger, 454 0.1 μm), Ace Lake 5 m (Sanger, 454 0.1 μm); Yellow (copiotrophic): *Flavobacterium johnsoniae*, *Kordia algicida*, *Pseudoalteromonas atlantica*, Ace Lake 23 m (454 0.8 μm , 3.0 μm), Ace Lake 18 m (454 0.8 μm , 3.0 μm), Ace Lake 14 m (454 0.8 μm , 3.0 μm), Ace Lake 12.7 m (454 3.0 μm), Ace Lake 11.5 m (454 0.8 μm , 3.0 μm), Ace Lake 5 m (454 0.8 μm , 3.0 μm); Cyan (copiotrophic): *Shewanella frigidimarina*, *Shewanella denitrificans*, *Marinomonas* sp. MWYL1; Red (copiotrophic): *Photobacterium angustum*, *Vibrio cholerae*, *Vibrio fischeri*, *Vibrio splendidus*.

Figure S6 Statistical (STAMP) analysis of normalised mass spectra from COG annotated proteins between each zone in Ace Lake. Proteins are shown grouped into COG categories. Only differences with corrected p-value < 0.05 and effect size <5 are displayed. A) Mixolimnion vs interface; B) Mixolimnion vs monimolimnion; C) Interface vs monimolimnion. Blue, mixolimnion; green, interface; orange, monimolimnion. COG category descriptions are: E, Amino acid transport and metabolism; G, Carbohydrate transport and metabolism; J, Translation, ribosomal structure and biogenesis; C, Energy production and conversion; P, Inorganic ion transport and metabolism; H, Co-enzyme transport and metabolism; O, Post-translational modification, protein turnover and chaperones; K, Transcription; L, Replication, recombination and repair; N, Cell motility.

Figure S7 High resolution heatmap plot and functional clustering of KEGG assignments for the predicted ORFs from the metagenomic reads for Ace lake samples. The heat-scale is the percentage of ORFs assigned to each individual KEGG pathway.

- Andrássy I, Gibson JAE. (2007). Nematodes from saline and freshwater lakes of the Vestfold Hills, including description of *Hypodontolaimus antarcticus* sp. n. *Polar Biol* **30**: 669-678.
- Brown MV, Fuhrman JA. (2005). Marine bacterial microdiversity as revealed by internal transcribed spacer analysis. *Aquat Microb Ecol* **41**: 15-23.
- Coolen MJL, Hopmans EC, Rijpstra WIC, Muyzera G, Schoutena S, Volkman JK et al. (2004). Evolution of the methane cycle in Ace Lake (Antarctica) during the Holocene: response of methanogens and methanotrophs to environmental changes. *Org Geochem* **35**: 1151–1167.
- Coolen MJL, Muyzer G, Schouten S, Volkman JK, Sinninghe Damste JS. (2006). Sulfur and methane cycling during the Holocene in Ace Lake (Antarctica) revealed by lipid and DNA stratigraphy. In: Neretin, LN (ed.). *Past and Present Marine Water Column Anoxia (NATO Science Series: IV-Earth and Environmental Sciences)*. Springer: Dordrecht, The Netherlands, pp 41–65.

- Cromer L, Gibson JAE, Swadling KM, Ritz DA. (2005). Faunal microfossils: Indicators for ecological change in an Antarctic saline lake. *Palaeogeogr Palaeoclimatol Palaeoecol* **221**: 83-97.
- Field KG, Gordon D, Wright T, Rappé M, Urback E, Vergin K et al. (1997). Diversity and depth-specific distribution of SAR11 cluster rRNA genes from marine planktonic bacteria. *Appl Environ Microbiol* **63**: 63–70.
- Gibson JAE. (1999). The meromictic lakes and stratified marine basins of the Vestfold Hills, East Antarctica. *Antarctic Sci* **11**: 175–192.
- Johnson MD, Oldach D, Delwiche CF, Stoecker DK. (2007). Retention of transcriptionally active cryptophyte nuclei by the ciliate *Myrionecta rubra*. *Nature* **445**: 426-428.
- Lauro FM, McDougald D, Thomas T, Williams TJ, Egan S, Rice S et al. (2009). The genomic basis of trophic strategy in marine bacteria. *Proc Natl Acad Sci USA* **106**: 15527–15533.
- Letunic I, Bork P. (2006). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**: 127-128.
- Rankin LM, Gibson JAE, Franzmann PD, Burton HR. (1999). The chemical stratification and microbial communities of Ace Lake, Antarctica: a review of the characteristics of a marine-derived meromictic lake. *Polarforschung* **66**: 35–52 (http://epic.awi.de/Publications/Polarforsch1996_1-2_4.pdf)
- Roberts D, McMin A. (1999). A diatom-based palaeosalinity history of Ace Lake, Vestfold Hills, Antarctica. *The Holocene* **9**: 401–408.
- Yubuki N, Edgcomb VP, Bernhard JM, Leander BS. (2009). Ultrastructure and molecular phylogeny of *Calkinsia aureus*: cellular identity of a novel clade of deep-sea euglenozoans with epibiotic bacteria *BMC Microbiol* **9**: 16.
- Zhu F, Massana R, Not F, Marie D, Vaultot D. (2005). Mapping of picoeucaryotes in the marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiol Ecol* **52**: 79–92.
- 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 Sc Lett* **155**: 131–145.



GOS Sample ID	Depth (m)	Size fractions	Trimmed reads	Predicted ORFs (from reads)	ORFs (from reads) with COG/KEGG assignments	Contigs over 5kb (reads in contigs)	Predicted ORFs (from contigs)	Peptide hits
GS232	5	Sanger	281,490	421,252	112,490 / 96,771	*	*	*
		454-0.1µm	539,536	638,757	109,511 / 103,201	2,809 (349,015)	45,281	5,728
		454-0.8µm	468,122	485,021	150,660 / 135,451	296 (66,743)	3,215	*
		454-3.0µm	160,835	138,191	24,920 / 22,240	33 (2,980)	353	*
GS231	11.5	Sanger	283,663	427,889	124,332 / 107,323	*	*	*
		454-0.1µm	523,650	608,671	99,175 / 95,266	2,814 (390,490)	47,987	3,213
		454-0.8µm	474,419	511,909	218,126 / 176,332	174 (161,891)	2,321	*
		454-3.0µm	373,226	307,850	57,961 / 51,915	64 (8,139)	766	*
GS230	12.7	Sanger	54,446	75,576	42,790 / 41,391	*	*	*
		454-0.1µm	442,389	492,995	200,203 / 227,726	88 (282,232)	3,039	12,718
		454-0.8µm	529,711	555,328	209,078 / 234,682	86 (313,550)	2,187	*
		454-3.0µm	208,272	215,741	80,925 / 77,340	75 (49,942)	1,792	*
GS229	14	Sanger	10,042	14,326	5,261 / 4,469	*	*	*
		454-0.1µm	413,992	458,942	100,045 / 88,300	228 (22,556)	2,443	3,427
		454-0.8µm	453,205	435,534	142,743 / 129,403	139 (45,083)	2,118	*
		454-3.0µm	291,065	301,580	105,756 / 99,188	31 (2,422)	262	*
GS228	18	Sanger	9,672	15,077	3,667 / 3,008	*	*	*
		454-0.1µm	362,490	389,077	51,312 / 44,290	29 (1,815)	260	725
		454-0.8µm	544,302	556,243	186,455 / 163,878	154 (14,806)	1,334	*
		454-3.0µm	278,846	287,423	95,876 / 81,161	2 (131)	15	*
GS227	23	Sanger	100,085	160,302	33,462 / 27,302	*	*	*
		454-0.1µm	482,527	547,170	84,257 / 73,074	1,136 (51,163)	12,339	1,602
		454-0.8µm	553,234	611,717	161,973 / 137,632	105 (7,904)	825	*
		454-3.0µm	264,160	270,188	85,596 / 72,206	6 (287)	48	*
TOTAL	---	---	8,103,379	8,926,759	2,487,574 / 2,283,749	8,269 (1,771,149)	126,585	27,413

(D)

