Materials and methods

Characteristics of the lake and samples taken

The water level of Organic Lake was measured by surveying as +1.886 m relative to the survey mark (NMV / S / 53) located at 68°27′23.4″S, 78° 11′ 22.6″E (Supplementary Figures S1). Water was collected from Organic Lake on 10 November 2008 through a 30 cm hole in the 0.8 m thick ice cover above the deepest point in the lake. The sampling hole (68°27′22.2″S, 78°11′23.9″E) was established following bathymetry measurements constructed on a metric grid. Samples were collected for metagenomics, microscopy and chemical analyses at 1.7, 4.2, 5.7, 6.5 and 6.7 m depths (maximum lake depth 6.8 m). For metagenomics, lake water was passed through a 20 μm pore size pre-filter, and microbial biomass captured by sequential filtration onto 3.0 μm, 0.8 μm and 0.1 μm pore size 293 mm polyethersulfone membrane filters, and samples immediately preserved in buffer and cryogenically frozen in liquid nitrogen, as described previously (Ng *et al.*, 2010; Lauro *et al.*, 2011). Between 1–2 L of lake water was sufficient to saturate the holding capacity of the filters. DNA was extracted from the filters, samples sequenced using the Roche GS-FLX titanium sequencer, and reads processed to remove low quality bases, assembled and annotated, as previously described (Ng *et al.*, 2010; Lauro *et al.*, 2011). A summary of the 2.4 Gbp of metagenomic data is provided in Supplementary Table S1.

Physical and chemical analyses

An *in situ* profile of pH, conductivity, turbidity, dissolved oxygen (DO) and pressure was measured using a submersible probe (YSI sonde model V6600). A temperature profile was measured using a maximum-minimum mercury thermometer as the YSI probe did not have a capacity to record temperature below −10 ºC. The 5.7 m sample corresponded to the turbidity maximum and the 6.5 m sample to the turbidity minimum. Conductivity at *in situ* temperature was converted to conductivity at 15 ºC as described previously (Gibson, 1999). The adjusted conductivity brings the temperature to within a range suitable for estimating practical salinity using the formula of Fofonoff and Millard (1983). Salinity was likely to have been underestimated as it is higher than the range (2–42) for which the conductivity–salinity relation holds. However, the relative difference in salinity between the samples would be accurate. Density was calculated from the *in situ* conductivity and temperature using the equations described by Gibson *et al.* (1990) and expressed at temperature T as:

$$\sigma_T = (1000 - \text{density})\ \text{kg/m}^3$$

Ammonia, nitrate, nitrite, total nitrogen (TN), total dissolved nitrogen (TDN), dissolved reactive phosphorus (DRP), total phosphorus (TP), total dissolved phosphorus (TDP), total organic carbon (TOC), total dissolved carbon (DOC), total sulfur (TS) and total dissolved sulfur (TDS) were determined by American Public Health Associations Standard Methods at the Analytical Services, Tasmania. Values for dissolved nutrients were measured after filtration through a 0.1 μm pore size membrane filter. All other nutrients were measured from water collected after filtration through the on-site 20 μm pore size pre-filter. Ammonia, nitrate, nitrite, DRP, TN, TDN, TP and TDP were measured in a Flow Injection Analyser (Lachat Instruments, Colorado, USA). TOC and DOC were determined in the San++ Segmented Flow Analyser (Skalar, Breda, Netherlands). TS and TDS were analyzed in the 730ES Inductively Coupled Plasma–Atomic Emission Spectrometer (Agilent Technologies, California, USA). Principal Component Analysis was performed using the PRIMER Version 6 statistical package (Clarke and Gorley, 2006) on the normalized physical and chemical parameters.

Epifluorescence microscopy

Water samples collected for microscopy were preserved in formaldehyde (1% v/v). Cells and virus-like particles (VLPs) were vacuum filtered onto 25 mm polycarbonate 0.015 μm pore-size membrane filters (Nuclepore Track-etched, Whatman, GE Healthcare, USA) with a 0.45 μm pore-size backing filter. The 0.015 μm filter was mounted onto a glass slide with ProLong® Gold anti fade reagent (Invitrogen, Life Technologies, NY, USA) and 2 μl (25 × dilution in sterile filtered milliQ water <0.015 μm) SYBR® Gold nucleic acid stain (Invitrogen, Life Technologies, NY, USA). Prepared slides were visualized in an epifluorescence microscope (Olympus BX61, Hamburg, Germany) under excitation with blue light (460–495 nm, emission 510–550 nm). Cell and VLP counts were performed on the same filter over 30 random fields of view.

Cellular diversity analyses

Diversity of *Bacteria, Archaea* and *Eucarya* was assessed using ribosomal small subunit (SSU) gene sequences. Metagenomic reads that matched the 16S and 18S rRNA genes were retrieved using Metaxa (Bengtsson *et al.*, 2011). 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) (www.arb-silva.de). SSU sequences that did not cluster with sequences from SILVA were allowed to form new OTUs (no suppression). A representative sequence from each OTU was chosen and classified to the genus level using QIIME implementing the RDP classifier (Wang *et al*., 2007) trained against SILVA. Assignments were accepted to the lowest taxonomic rank with bootstrap value ≥85%. To allow comparison of the relative abundance of taxa, the number of SSU matches per sample filter was normalized to the average number of reads (403 577). Statistical analysis on the relative SSU abundances was performed using the PRIMER Version 6 package (Clarke and Gorley, 2006). The SSU counts of each sample filter were aggregated to the genus level and square root transformed to reduce the contribution of highly abundant taxa. A resemblance matrix was computed using Bray-Curtis similarity. The upper mixed zone (1.7, 4.2 and 5.7 m) and deep zone (6.5 and 6.7 m) samples were designated as separate groups and an analysis of similarity (ANOSIM) performed to test for difference between the two groups. BEST analysis was performed with the abiotic variables: conductivity, temperature, turbidity, DO, pH, TOC, TN, TP, TS, total C:N, total C:P, total N:P, cell counts and VLP counts. The Bio-Env procedure in BEST looks at all the abiotic variables in combination and finds a subset sufficient to best explain the biotic structure. A heat map with bi-clustering dendogram was generated using R and the package 'seriation' (Hahsler *et al*., 2008) on the normalized square-root transformed SSU counts.

Analysis of functional potential

The relative abundance and taxonomic origin of functional marker genes was used to determine the potential for carbon, nitrogen and sulfur conversions. Open reading frames (ORFs) were predicted from trimmed metagenomic reads using MetaGene (Noguchi *et al*., 2006) accepting those >90 bp in length. ORFs were translated using the standard bacterial/plastid translation table and 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). The BLAST output was processed using KEGG Orthology Based Annotation System (KOBAS) version 2.0 (Xie *et al.*, 2011) accepting assignments to KEGG Orthology (KO) groups with e-value <1e−05 and rank >5. KO groups used as functional markers are listed in Supplementary Table S2. Marker enzymes were assigned to taxonomic groups based on the species of origin of the best KEGG GENES BLASTp match. Marker genes not represented by a KO group were assessed by BLASTp queries of marker gene sequences with experimentally confirmed function (Supplementary Table S3) against a database of translated ORFs predicted from metagenomic reads. Matches were accepted if the e-value was <1e−10 and sequence identity was within the range shared by homologs of the query sequence(s) (Supplementary Table S3). Matches to marker genes were normalized to 100 Mbp per sample and counted. Normalized frequencies of markers from the same pathway were averaged and those from different pathways were summed.

The same marker genes and BLAST procedure was used to compare the DMSP catabolism and photoheterotrophy potential of Organic Lake with nearby Ace Lake (Lauro *et al.*, 2011), Southern Ocean (Wilkins *et al.*, 2012b) and GOS metagenomes (Rusch *et al.*, 2007). Counts of single copy gene *recA* were also determined to estimate the percentage of genomes containing each marker gene (percentage of marker genes relative to *recA*). Matches to *recA* were accepted with e-value <1e-20 (Howard *et al.,* 2008). For GOS samples, the BLAST database was generated from peptide sequences retrieved from CAMERA (camera.calit2.net) while the other BLAST databases were produced as for Organic Lake. The total number of trimmed base pairs for GOS samples was estimated by multiplying the number of reads from each sample by the average read length (822 bp) (Rusch *et al.*, 2007).

Marker gene sequences for phylogenetic analysis were clustered using the CD-HIT web server (Huang *et al*., 2010) at 90% global amino acid identity. A representative sequence from the clusters that resided within a desired conserved region and homologs from cultured strains were used in phylogenetic analyses performed in MEGA 5.05 (Tamura *et al.* 2011) implementing MUSCLE using default parameters (gap opening penalty: −2.9, gap extension penalty: 0). Neighbor-joining was used to compute the phylogenies with a Poisson substitution model, uniform rates of change and complete deletion of alignment gaps. Node support was tested with bootstrap analysis (500 replicates).

Figure S1 Organic Lake expedition. **(A)** Schematic of the Vestfold Hills, on the eastern shore of Prydz Bay, East Antarctica, showing the location of Organic Lake, adapted from Gibson (Gibson, 1999). The Vestfold Hills is approximately 400 km^2 in area and contains a remarkable diversity of more than 300 lakes which range in salinity from fresh to hypersaline (Gibson, 1999). Most of the saline lakes were originally pockets of seawater, trapped less than 10 000 BP when the continental ice-sheet receded and the land rose above sea-level, and retain a marine-derived biota (Zwartz *et al*., 1998; Gibson, 1999). Differing local conditions has led each lake to develop unique physical and chemical properties, and life in the lakes tends to be entirely microbial with low levels of diversity (Bowman *et al.*, 2000). The Vestfold Hills contains the highest density of meromictic (permanently stratified) water bodies in Antarctica (Gibson, 1999). The strong physico-chemical stratification within a single, largely closed system, provides the opportunity to investigate how microbial communities and ecosystem processes have evolved in the cold and in response to gradients of nutrients, oxygen, salinity and solar irradiance. **(B)** Aerial photograph of Organic Lake (inset), and surrounds including Ace Lake, Long Fjord and the open Southern Ocean. **(C)** Sampling site at Organic Lake showing mobile work shelters

(MWSs) tethered by ropes and ice screws for protection against strong winds (up to 140 km/h in 2008/09), and sampling equipment, thermometer and YSI probe. The hole in the floor of the MWS enabled direct access to lake water below the surface ice. **(D)** Surveying the water level of Organic Lake and taking lake bathymetry measurements (see Figure S2).

Figure S2 Vertical profiles of physical and chemical parameters of Organic Lake taken *in situ* at the deepest point in the lake on 9 November 2008. σ_T (1000−density) was calculated from temperature and conductivity. *In situ* physico-chemical profiles measured over the deepest point in the lake (Supplementary Figure S3) determined the existence of two zones: an upper mixed zone above 5.7 m and a suboxic deep zone below 5.7 m (Figure 1A). The separation of the two zones was indicated by a pycnocline and oxycline starting at 5.7 m. The pH also decreased with DO, likely due to fermentation products such as acetic, formic and lactic acids that have been reported in the bottom waters (Franzmann *et al*., 1987b; Gibson *et al*., 1994). The deep zone was not completely anoxic (Figure 1A). Oxygen may be episodically introduced to bottom waters as a result of currents of cold dense water sinking during surface ice-formation (Ferris *et al*., 1999). In comparison to meromictic lakes such as Ace Lake that have strong pycnoclines and a steep salt gradient in the anoxic zone, Organic Lake is shallow and has relatively weak stratification (Gibson, 1999). Samples were collected from the upper mixed (1.7, 4.2 and 5.7 m) and deep (6.5 m and 6.7 m) zones.

Figure S3 Bathymetry of Organic Lake 9 November 2008.

Figure S4 Epifluorescence microscopy images of Organic Lake microbiota (<20 μm) filtered onto 0.01 μm polycarbonate membrane and stained with SYBR Gold. (**A**) 1.7 m, (**B**) 4.2 m, (**C**) 5.7 m, (**D**) 6.5 m, (**E**) 6.7 m. Scale bar = 20 μm.

Figure S5 Principle component analysis of physico-chemical parameters and cell/VLP counts of Organic Lake profile. Data points are the sampling depths 1.7, 4.2, 5.7, 6.5 and 6.7 m. The overlaid vector diagram shows the relative contribution of the variables to explaining the difference between samples. PC1 explained 74.3% and PC2 14.7% of the variation between samples. Abbreviations: cond, conductivity; temp, temperature; turb, turbidity.

Figure S6 Phylogenetic tree of rhodopsin homologs including proteorhodopsin, bacteriorhodopsin, actinorhodopsin and xanthorhodopsin. *Halobacterim salinarum* R1 halorhodopsin was used as an out-group. The tree was computed from a 78 amino acid region spanning the motif involved in 'spectral tuning' using the neighbor-joining algorithm. Organic Lake sequences from this study are shown in red and marked with an asterisk (*). Numbers in parentheses are counts of sequences which clustered with the Organic Lake homolog shown in the tree with 90% amino acid identity. Sequences with confirmed activity are shown in bold. Accession numbers from top to bottom are: EAZ99241, EDP63929, EGF32634, ZP_09955974, AEG32267, EDY76405, EDY88259, YP_445623, ACN42850, EIC91904, ZP_02194911, AAZ21446, AAT38609, AEE49633, EAS71907, sequence from John Bowman (personal communication), EAQ40507, EAQ40925, EAR12394, EHQ04368, EAZ94876, EIA08356, AEE20201, EEG43331, ZP_09501337 and YP_001689404.

Figure S7 Genomic maps of Organic Lake scaffolds containing the OL-R1 rhodopsin homolog. All genes surrounding OL-R1 had best BLAST matches to *Octadecabacter* sequences. The scale

below shows the number of base pairs. The sample depth and filter from which the scaffold was assembled is shown in parentheses beside the scaffold ID.

Figure S8 Phylogenetic tree of DddD DMSP lyase homologs. *E. coli* carnitine coenzyme A transferase was used as an out-group. *Dinoroseobacteria shibae* DFL 12 and *Ruegeria pomeroyi* DSS-3 homologs are non-functional outgroup (Todd *et al.*, 2011). The tree was computed from a 75 amino acid region within the conserved amino-terminal class III coenzyme A domain (CaiB) using the neighbor-joining algorithm. Organic Lake sequences from this study are shown in red and marked with an asterisk (*).Numbers in parentheses are counts of sequences which clustered with the Organic Lake homolog shown in the tree with 90% amino acid identity. Sequences with confirmed DMSP lyase activity are shown in bold. Accession numbers from top to bottom are: EBA01716, AEV37420, ACY01992, ADZ91595, EAQ63474, ABR72937, ACV84065, ACY02894, ABI89851, YP_002822700, EEE36156, ABV95365, AAV94987 and EGB36199.

Figure S9 Genomic maps of Organic Lake scaffolds containing the OL-dddD homolog. DddT and DddA had best BLAST matches to *Halomonas* sp. HTNK1 (*Gammaproteobacteria*) and *Hoeflea phototrophica* DFL-43 (*Alphaproteobacteria*), respectively. The numbers represent base pairs. The sample depth and filter from which the scaffold was assembled is shown in parentheses beside the scaffold ID. The open reading frames annotated at DddT in the 6.5 m scaffold are likely to represent one contiguous DddT gene.

Figure S10 Phylogenetic tree of DddL DMSP lyase homologs from Organic Lake and public databases. The tree was computed from an 84 amino acid N-terminal region using the neighbor-joining algorithm. Organic Lake sequences from this study are shown in red and marked with an asterisk (*). Numbers in parentheses are counts of sequences which clustered with the Organic Lake homolog shown in the tree with 90% amino acid identity. Sequences with confirmed DMSP lyase activity are shown in bold. Accession numbers from top to bottom are: EEB86351, ADK55772, EAQ07081, EEE47811, EAV43167, EAU41122, EAQ10619, ABV95046, EAQ04071, ABA77574 and EHJ04839.

Figure S11 Phylogenetic tree of DddP DMSP lyase homologs from Organic Lake and public databases. The tree was computed from a 129 amino acid C-terminal region including the predicted catalytic sites using the neighbor-joining algorithm. Organic Lake sequences from this study are shown in red and marked with an asterisk (*). Numbers in parentheses are counts of sequences which clustered with the Organic Lake homolog shown in the tree with 90% amino acid identity. Sequences with confirmed DMSP lyase activity are shown in bold. Accession numbers from top to bottom are: ZP_01755203, YP_167522, YP_613011, YP_682809, EAP77700, ZP_01741265, ZP_01036399, ZP_01881042, ZP_05063825, AFO91571, YP_509721, ZP_01448542, AEQ39103, AEQ39091, XP_001823911, XP_389272 and ACF19795.

Figure S12 Phylogenetic tree of DmdA DMSP demethylase homologs from Organic Lake and public databases. The tree was computed from a 128 amino acid region using the neighbor-joining algorithm. Organic Lake sequences from this study are shown in red and marked with an asterisk (*). Numbers in parentheses are counts of sequences which clustered with the Organic Lake homolog shown in the tree with 90% amino acid identity. Sequences with confirmed DMSP lyase activity are shown in bold. Accession numbers from top to bottom are: EDZ60447,

YP_265671, EDZ61098, EAU51039, YP_003550401, EDP61332, EAQ26389, ABV94056, AAV94935, AAV95190, EDY79173, EDY89914, EAW42451, AAV94935 and AAV97197.

Table S1 Summary of metagenomic data for Organic Lake samples.

Table S2 Full list of KEGG Orthologs (KO) involved in carbon, nitrogen and sulfur conversions that were searched for in the Organic Lake metagenome. Abbreviations: rTCA, tricarboxylic acid cycle; WL, Wood-Ljungdahl pathway; AAnP, aerobic anoxygenic phototrophy; DNRA, dissimilatory nitrate reduction to ammonia; anaerobic ammonia oxidation; ASR, assimilatory sulfate reduction; DSR, dissimilatory sulfate reduction; SRB, sulfate-reducing bacteria.

Table S3 Sequences used in this study as BLAST queries for retrieving homologs in the Organic Lake metagenomes. (%ID) is the minimum amino acid sequence identity for matches to be considered a homolog.

Table S4 Microbial taxa detected in the Organic Lake water column profile. Analysis of SSU gene sequences shown in phylum, class and genus ranks as defined by the SILVA taxonomy except RF3 which is placed with the *Firmicutes* according to Tajima *et al.* (1999). SSU gene sequences were classified to the genus level or to the lowest rank with bootstap confidence >85% (see materials and methods). The best BLAST matches to environmental SSU clone sequences are shown for the abundant candidate divisions RF3 and OD1.

References

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

Balashov SP, Imasheva ES, Boichenko VA, Antón J, Wang JM, Lanyi JK. (2005) Xanthorhodopsin: a proton pump with a light-harvesting carotenoid antenna. *Science* **309**: 2061–2064.

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.

Bowman JP, McCammon SA, Rea SM, McMeekin TA. (2000) The microbial composition of three limnologically disparate hypersaline Antarctic lakes. *FEMS Microbiol Lett* **183**: 81–88.

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 and Gorley RN. (2006) PRIMER v6: User Manual/Tutorial. PRIMER-E, Plymouth.

Curson ARJ, Sullivan MJ, Todd JD, Johnston AWB. (2010) Identification of genes for dimethyl sulfide production in bacteria in the gut of Atlantic Herring (*Clupea harengus*). *ISME J* **4**: 144–146.

Curson ARJ, Sullivan MJ, Todd JD, Johnston AWB. (2011) DddY, a periplasmic dimethylsulfoniopropionate lyase found in taxonomically diverse species of Proteobacteria. *ISME J* **5**: 1191–1200.

Ferris JM, Gibson JAE, Burton HR. (1991) Evidence of density currents with the potential to promote meromixis in ice-covered saline lakes. *Palaeogeogr Palaeoclimatol Palaeoecol* **84**: 99–107.

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

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.

Fuchs BM, Spring S, Teeling H, Quast C, Wulf K, Schattenhofer M, Yan Shi *et al.* (2007) Characterization of a marine gammaproteobacterium capable of aerobic anoxygenic photosynthesis. *PNAS* **104**: 2891–2896.

Gibson JAE. (1999) The meromictic lakes and stratified marine basins of the Vestfold Hills, East Antarctica. *Antarct Sci* **11**:175–192.

Gibson JAE, Ferris JM, Burton HR. (1990) Temperature density, temperature conductivity and conductivity-density relationships for marine-derived saline lake waters. *ANARE Research Notes*, No. 78.

Gibson JAE, Qiang XL, Franzmann PD, Garrick RC, Burton HR. (1994) Volatile fatty and dissolved free amino acids in Organic Lake, Vestfold Hills, East Antarctica. *Polar Biol* **14**: 545–550.

Gómez-Consarnau L, Akram N, Lindell K, Pedersen A, Neutze R, Milton DL *et al.* (2010) Proteorhodopsin phototrophy promotes survival of marine bacteria during starvation. *PLoS Biol.* **8**: e1000358.

Gómez-Consarnau L, González JM, Coll-Lladó M, Gourdon P, Pascher T, Neutze R *et al*. (2007) Light stimulates growth of proteorhodopsin-containing marine *Flavobacteria*. *Nature* **445**: 210–213.

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

Howard EC, Henriksen JR, Buchan A, Reisch CR, Bürgmann, Welsh R, Ye W *et al.* (2006) Bacterial taxa that limit sulfur flux from the ocean. *Science* **314**: 649–652.

Howard EC, Sun S, Biers EJ, Moran MA. (2008) Abundant and diverse bacteria involved in DMSP degradation in marine surface waters. *Environ Microbiol* **10**: 2397–2410.

Huang Y, Niu B, Gao Y, Fu L, Li W. (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* **26**: 680–682.

Labrenz M, Collins MD, Lawson PA, Tindall BJ, Schumann P, Hirsch P. (1999) *Roseovarius tolerans* gen. nov., sp. nov., a budding bacterium with variable bacteriochlorophyll *a* production from hypersaline Ekho Lake. *Int J Syst Bacteriol* **49**: 137–147.

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.

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.

Rusch DB, Halpern AL, Sutton G, Heidelberg 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.

Scott KM, Sievert SM, Abril FN, Ball LA, Barrett CJ, Blake RA *et al.* (2006) The genome of deep-sea vent chemolithoautotroph *Thiomicrospira crunogena* XCL-2. *PLoS Biol* **4**: e383.

Sievert SM, Scott KM, Klotz MG, Chain PSG, Hauser LJ, Hemp J *et al.* (2008) Genome of the Epsilonproteobacterial chemoautotroph *Sulfurimonas denitrificans*. *Appl Environ Microbiol* **74**: 1145–1156.

Tajima K, Aminov RI, Nagamine T, Ogata K, Nakamura M, Matsui H *et al*. (1999) Rumen bacterial diversity as determined by sequence analysis of 16S rDNA. *FEMS Microbiol Ecol* **29**: 159–169.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**: 2731–2739.

Todd JD, Curson ARJ, Kirkwood M, Sullivan MJ, Green RT, Johnston AWB. (2011) DddQ, a novel, cupin-containing, dimethylsulfoniopropionate lyase in marine roseobacters and in uncultured marine bacteria. *Environ Microbiol* **13** :427–438.

Todd JD, Curson ARJ, Nikolaidou-Kataraidou N, Brearley CA, Watmough NJ, Chan Y *et al.* (2010) Molecular dissection of bacterial acrylate catabolism – unexpected links with dimethylsulfoniopropionate catabolism and dimethyl sulfide production. *Environ Microbiol* **12**: 327–343.

Todd JD, Kirkwood M, Newton-Payne S, Johnston AWB. (2012) DddW, a third DMSP lyase in model Roseobacter marine bacterium, *Ruegeria pomeroyi* DSS-3. *ISME J* **6** :223–226.

Todd JD, Rogers R, Li YG, Wexler M, Bond PL, Sun L *et al.* (2007) Structural and regulatory genes required to make the gas dimethyl sulfide in bacteria. *Science* **315**: 666–669.

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.

Wilkins D, Lauro FM, Williams TJ, DeMaere MZ, Brown MV, Hoffman JM, Andrews-Pfannokoch C *et al.* (2012b) Biogeographic partitioning of Southern Ocean microorganisms revealed by metagenomics. *Environ Microbiol* doi: 10.1111/1462-2920.12035.

Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S *et al*. (2011) KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res* **39**: W316–W322.

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.