

ORIGINAL ARTICLE

Metaproteogenomic analysis of a dominant green sulfur bacterium from Ace Lake, Antarctica

Charmaine Ng¹, Matthew Z DeMaere¹, Timothy J Williams¹, Federico M Lauro¹, Mark Raftery², John AE Gibson³, Cynthia Andrews-Pfannkoch⁴, Matt Lewis⁴, Jeffrey M Hoffman⁴, Torsten Thomas^{1,5} and Ricardo Cavicchioli¹

¹School of Biotechnology and Biomolecular Sciences, The University of New South Wales, Sydney, New South Wales, Australia; ²Bioanalytical Mass Spectrometry Facility, The University of New South Wales, Sydney, New South Wales, Australia; ³Marine Research Laboratories, Tasmanian Aquaculture and Fisheries Institute, University of Tasmania, Hobart, Tasmania, Australia; ⁴J Craig Venter Institute, Rockville, MD, USA and ⁵Centre for Marine Bio-Innovation, The University of New South Wales, Sydney, New South Wales, Australia

Green sulfur bacteria (GSB) (*Chlorobiaceae*) are primary producers that are important in global carbon and sulfur cycling in natural environments. An almost complete genome sequence for a single, dominant GSB species ('C-Ace') was assembled from shotgun sequence data of an environmental sample taken from the O₂-H₂S interface of the water column of Ace Lake, Antarctica. Approximately 34 Mb of DNA sequence data were assembled into nine scaffolds totaling 1.79 Mb, representing approximately 19-fold coverage for the C-Ace composite genome. A high level (~31%) of metaproteomic coverage was achieved using matched biomass. The metaproteogenomic approach provided unique insight into the protein complement required for dominating the microbial community under cold, nutrient-limited, oxygen-limited and extremely varied annual light conditions. C-Ace shows physiological traits that promote its ability to compete very effectively with other GSB and gain dominance (for example, specific bacteriochlorophylls, mechanisms of cold adaptation) as well as a syntrophic relationship with sulfate-reducing bacteria that provides a mechanism for the exchange of sulfur compounds. As a result we are able to propose an explanation of the active biological processes promoted by cold-adapted GSB and the adaptive strategies they use to thrive under the severe physiochemical conditions prevailing in polar environments.

The ISME Journal (2010) 4, 1002–1019; doi:10.1038/ismej.2010.28; published online 18 March 2010

Subject Category: Integrated genomic and post-genomic approaches in microbial ecology

Keywords: green sulfur bacterium; metagenomics; metaproteomics; cold adaptation; Antarctic meromictic lake; ecological dominance

Introduction

Green sulfur bacteria (GSB) (*Chlorobiaceae*) are primary producers that are important in global carbon and sulfur cycling in natural environments (Pfennig, 1975; Overmann, 2006). The GSB are anaerobic photoautotrophs carrying out photosynthesis using light-harvesting chlorosomes and carbon fixation through the reverse tricarboxylic acid (rTCA) cycle (Eisen *et al.*, 2002). Chlorosomes are unique, light-harvesting antennas that are packed with bacteriochlorophyll (BChl) *c*, *d*, *e* and *a*, which enable photosynthesis to occur at low light intensities (Blankenship *et al.*, 1995; Bryant and

Frigaard, 2006; Frigaard and Bryant, 2006). GSB can grow mixotrophically through acetate assimilation in the presence of light and CO₂. For these photo-assimilatory reactions, most GSB use sulfide as an electron donor, although some GSB species are able to oxidize other inorganic reduced sulfur compounds (for example, sulfur, polysulfides, thio-sulfate and tetrathionate), as well as hydrogen and/or ferrous iron (Frigaard and Bryant, 2008). The oxidation of sulfide results in the production of sulfate, and in some instances, the accumulation of sulfur deposits outside the cells (Frigaard and Bryant, 2008). In some anaerobic aquatic systems, syntrophic interactions take place between GSB and sulfate-reducing bacteria (SRB). When this occurs, the sulfate produced from the activity of GSB is used as a terminal electron acceptor by SRB, which in turn is converted to biogenic sulfide. GSB therefore contribute to the sulfur cycle by oxidizing sulfur compounds, and when they coexist with SRB they

Correspondence: R Cavicchioli, School of Biotechnology and Biomolecular Sciences, The University of New South Wales, BABS, Sydney, New South Wales 2052, Australia.
E-mail: r.cavicchioli@unsw.edu.au

Received 25 November 2009; revised 2 February 2010; accepted 11 February 2010; published online 18 March 2010

cooperatively create a complete sulfur cycle, a matter of importance in closed lake systems (Pfennig, 1978; Biebl and Pfennig, 1978).

Green sulfur bacteria are typically found thriving in freshwater and estuarine environments where light is limited, and the waters are anoxic and rich in sulfide. Their habitats include sediments, microbial mats, sulfide-rich hot springs and within a discrete zone of the water column of stratified lakes where light is able to penetrate to the anoxic zone (Wahlund *et al.*, 1991; Ward *et al.*, 1998; Vila *et al.*, 2002; Mallorqui *et al.*, 2005; Martínez-Alonso *et al.*, 2005; Musat *et al.*, 2008). GSB have also been identified in marine environments where light penetration is very limited, at the chemocline of the Black Sea (Manske *et al.*, 2005), and in the benthos near deep-sea hydrothermal vents in the Pacific Ocean (Beatty *et al.*, 2005).

In East Antarctica, an area called the Vestfold Hills contains numerous marine-derived saline lakes that were formed when the ocean receded approximately 5700 years ago (Gibson, 1999; Rankin *et al.*, 1999; Cavicchioli, 2006). Fjords connected to the ocean also cut across the Vestfold Hills. Ellis Fjord is 10 km long, up to 100 m deep and has become a stratified system due to its restricted opening (less than 4 m deep) to the ocean (Burke and Burton, 1988b). In more than a dozen of the stratified systems that have anoxic bottom waters (lakes, marine basins and fjords), members of the Chlorobiaceae (that is, *Chlorobium limicola* and *Chlorobium vibrioforme*) have been reported to be the dominant type of GSB (Burke and Burton, 1988a,b). In Ellis Fjord, the O₂-H₂S interface where GSB have been observed is at a depth of ~40 m beneath the surface where measurable quantities of light (0.1 μE m⁻² s⁻¹) have been detected during summer. *Chlorobium* species have been grown from samples taken at a depth of 100 m below the surface, and based on rates of sedimentation, the GSB in Ellis Fjord were predicted to have survived in the dark for >4.5 years. The capacity of the *Chlorobium* species to dominate has been speculated to be due to a superior ability to harvest the intensity and wavelength of available light during the Antarctic summer (for example, compared to members of the Chromatiaceae), and to remain viable during winter (Burke and Burton, 1988a,b). Isolated *Chlorobium* species have been noted for their ability to grow well in hypersaline water at -2 °C, and to grow at very low light intensities (<1 μE m⁻² s⁻¹; Burke and Burton, 1988a,b).

Ace Lake (in the Vestfold Hills) is the most thoroughly studied meromictic lake in Antarctica (Rankin *et al.*, 1999), and it is the subject of a comprehensive environmental genomic/proteomic program. The lake surface remains covered in approximately 2 m of ice for about 10–11 months (February/March–December) of the year and usually (but not always) melts out during summer (January) (Gibson, 1999; Rankin *et al.*, 1999). Mixing in the

upper waters of the lake is caused by exclusion of salt from the forming ice cover in winter, and by wind mixing when the lake is ice free. A strong mid-water halocline prevents penetration of mixing to deeper waters. The oxygen content in surface waters of the lake is high, and rapidly decreases at the chemocline where a pronounced O₂-H₂S interface exists. Phytoplankton species (for example, *Pyramimonas gelicola*, *Mesodinium rubrum*), cyanobacteria and photosynthetic microorganisms in microbial mats (for example, cyanobacteria, diatoms, algae) are likely to contribute to primary production in the oxic zone of the lake. Primary productivity measurements of Ace Lake indicate that the autotrophic ciliate, *M. rubrum*, which has also been found in Antarctic waters is a major contributor to carbon fixation (15–40%) during its bloom period over summer (Labourn-Parry and Perriss, 1995; Rankin *et al.*, 1999). Autotrophic primary producers convert inorganic carbon to biomass (0.16–0.68 mgC m⁻³ h⁻¹), and in the process release dissolved organic carbon by the action of predation and viral lysis (dissolved organic carbon: 6–11.5 mgC l⁻¹) (Rankin *et al.*, 1999). Owing to physical stability, dissolved organic carbon is likely to cycle within the aerobic zone. Dead cell material from all organisms and fecal pellets (*P. antarctica*) forms particulate organic carbon that gradually sinks (particulate organic carbon: 0.22–0.73 mgC l⁻¹) (Rankin *et al.*, 1999).

At the interface, a brief period (~6 weeks) of high light intensity occurs after the ice cover has melted during summer allowing maximum light penetration (maximum recorded ~100 μmol of photons m² s⁻¹) and enabling the most rapid growth of photosynthetic populations (Burke and Burton, 1988b). The lake experiences a longer period of intermediate light intensity (<30 μmol of photons m² s⁻¹) over summer when the ice cover has not melted, and a period (~6 weeks) of almost complete darkness during winter (Burke and Burton, 1988b). *Chlorobium* species of GSB have been identified at this O₂-H₂S interface (Burke and Burton, 1988b; Coolen *et al.*, 2006), and sulfate reduction by SRB has been reported in the region below the chemocline (Franzmann *et al.*, 1988; Rankin *et al.*, 1999). No axenic cultures of GSB have been obtained from Ace Lake or any of the systems in the Vestfold Hills, limiting physiological and molecular analyses. In contrast, mesophilic and thermophilic counterparts are available, including the complete genome sequences of 12 GSB species that range in size from 1.97 Mb for *Prosthechochloris vibrioformis* DSM 265 to 3.3 Mb for *Chloroherpeton thalassium* ATCC 35110. The first and most comprehensively characterized genome sequence is for the thermophile, *Chlorobaculum tepidum* TLS (Eisen *et al.*, 2002), which was isolated from a hot spring in New Zealand (Wahlund *et al.*, 1991).

In this study, microbial biomass was sampled from the turbidity peak coinciding with the halocline and

oxycline in Ace Lake, and analyzed by metagenomic sequencing and metaproteomic mass spectrometry. The sample was highly enriched for a single, dominant bacterium (referred to as 'C-Ace') enabling a large assembly for a composite genome of the bacterium and high coverage of the proteome to be achieved. As a result we have been able to assemble a hypothetical model of the active biological processes mediated by C-Ace and propose a number of adaptive strategies it may use in response to physiochemical conditions prevailing in the lake, including seasonal changes in solar radiation.

Materials and methods

Antarctic samples

Water samples were collected from Ace Lake (68° 28.33' S, 78° 11.29' E), Vestfold Hills, Antarctica on 22 December 2006. A 2 m hole positioned above the deepest point (25 m depth) of the lake was drilled through the ice cover of Ace Lake to enable sampling of lake water. A volume of 1 and 10 liters was pumped directly from the lake and collected by sequential size fractionation through a 20 µm pre-filter onto filters (3, 0.8 and 0.1 µm pore-sized, 293 cm polyethersulfone membrane filters; Rusch *et al.*, 2007) from depths of 12.7 m (Supplementary Figure S10) and 14 m, respectively. Two independent sets of filters were obtained. Filters were placed into a tube containing a solution of 2.5 mM EGTA, 2.5 mM EDTA, 0.1 mM Tris-EDTA (pH 8), 1 mM PMSF (freshly prepared), 50 µl of protease inhibitor cocktail VI (Calbiochem, San Diego, CA, USA) and the tubes placed into liquid N₂ before storage at -80 °C. The protease cocktail inhibitor has a broad specificity for inhibition of serine, cysteine, aspartic and metalloproteases within protein extracts.

Metagenomic sequencing, assembly and database construction

DNA extraction and Sanger sequencing of 0.1 µm-sized filters was performed at the J Craig Venter Institute in Rockville, MD, USA (Rusch *et al.*, 2007). The scaffolds and annotations will be available via CAMERA and public sequence repositories such as NCBI and the reads will be available via the NCBI Trace Archive. 54 282 reads totaling 36 688 915 bases were generated. Assembly was performed using Celera WGS-Assembler v5.3 (<http://sourceforge.net/projects/wgs-assembler/files/>) with a unitigger error rate of 3% and an estimated genome size of 2.2 Mb. Isolation of the dominant organism scaffolds was performed by reference to a Self Organising Map created with Synapse (Peltarion, Stockholm, Sweden), where the per-scaffold feature-space vectors were composed of GC content, assembly read depth and the set of 64 normalized trinucleotide frequencies as calculated by Tetra (Teeling *et al.*, 2004). Clear cluster separation identified nine dominant organism scaffolds with

a total extent of 1.79 Mb and average GC content of 52.2%. These nine selected scaffolds contained 92.5% of all scaffolded reads and 77% of all reads in the experiment. Coverage for these scaffolds represented 18.95-fold.

Annotation and analysis of predicted protein sequences

An internally developed annotation pipeline was used, where open reading frame (ORF) prediction was performed using Metagene (<http://metagene.cb.k.u-tokyo.ac.jp/metagene/download.html>) (Noguchi *et al.*, 2006) and Glimmer (<http://glimmer.sourceforge.net/>) (Delcher *et al.*, 1999), whereas subsequent functional assignments were derived from searches against the COG, KEGG, NR, SWISS-PROT and TIGRFAM databases by use of blast and hmmer. The COG category distribution was used for assigning general functional predictions to proteins with no known functional information against the other databases. The significant differences in COG composition between the composite genome and the metaproteome were detected with custom Perl scripts based on a bootstrapping resampling method (Rodriguez-Brito *et al.*, 2006). A confidence interval of 99% was established using 10 000 subsamples of 750 proteins. The KEGG category assignments and the Metacyc database (<http://metacyc.org/>) were used as a source of reference for reconstructing metabolic pathways. A TMHMM tool (<http://www.cbs.dtu.dk/services/TMHMM/>) and SignalP 3.0 (<http://www.cbs.dtu.dk/services/SignalP/>) were used to determine the proportion of proteins with transmembrane domains and signal peptides within the proteomic data set. MEGAN (<http://www-ab.informatik.uni-tuebingen.de/software/megan>) was used to analyze the database of all predicted protein sequences (including those that did not belong to the nine dominant scaffolds) from Metagene to distinguish between proteins taxonomically conserved within sequenced species of the Chlorobiaceae family, and those that fell outside this group. A blastp comparison (e-value <1E-5) was performed on all predicted proteins against the NR database and MEGAN was used with default parameters to generate a phylo profile for all predicted proteins, enabling binning of sequences conserved in *Chlorobiaceae* and those which were novel to this taxon.

Protein extraction from filter membranes

A combination of different physical (sonication/freezing-thaw) and chemical (urea/SDS containing buffers, TCA precipitation) extraction techniques was tested on filtered ocean water samples to develop a method to maximize recovery of whole-cell protein extracts from total biomass on membrane filters. The 0.1 µm membrane filters (12.7 and 14 m samples) were removed from their storage

buffer and cut into quarters using aseptic procedures. Separate extractions were performed on three of the filter quarters. Approximately 1 mg of protein was extracted from $\frac{3}{4}$ of the 0.1 μm membrane in which a total volume of 1 liter of water was filtered. The filters were suspended in a lysis buffer containing 10 mM Tris-EDTA (pH 8.0; Univar, Sydney, New South Wales, Australia), 20 μl of PI VI (Calbiochem), 0.1% sodium dodecyl sulfate (SDS; Univar) and 1 mM dithiothreitol (Sigma-Aldrich, Sydney, New South Wales, Australia). Filters were subjected to three freeze–thaw cycles in liquid N_2 to release cells from the membrane. The buffer containing the cells was disrupted on ice by sonication with a Branson Sonifier for five cycles of 40 s on a 30% duty cycle at a power setting of 3. The supernatant containing the extracted proteins was centrifuged at 5000 g for 25 min at 4 °C to remove cell debris. To remove particles that did not pellet during the centrifugation step, we filtered the protein suspension through a 0.22 μm syringe filter (Millipore, Sydney, New South Wales, Australia) and transferred into a 5 kDa cutoff Amicon Ultra-15 filter unit (Millipore). Proteins were desalted by performing a buffer exchange with 30 ml of 10 mM Tris-EDTA (pH 8.0) followed by concentration to a smaller volume (~ 1 ml). Using a bicinchoinic acid protein assay kit (Sigma-Aldrich), we determined final protein concentrations to be between 100 and 500 μg per filter.

One-dimensional SDS–PAGE and in gel trypsin digestion

Protein samples were suspended in appropriate volumes of SDS–PAGE sample buffer and resolved on a 12% SDS gel using a Mini-PROTEAN system (Bio-Rad, Sydney, New South Wales, Australia) under conditions described previously (Saunders *et al.*, 2006). The gels were stained in silver (Blum *et al.*, 1987) and profile images were acquired using a UMAX PowerLook 1000 flat-bed scanner (Fujifilm, Berthold, Melbourne, Australia). Whole-gel lanes were excised using a sterile, clean scalpel and were cut into individual slices. The gel slices were further cut into smaller gel pieces and washed twice in sterile Milli-Q water followed by washing with acetonitrile. The gel pieces were treated through a series of reduction, alkylation and dehydration steps. *In gel* enzymatic digestion was performed by rehydrating the gel pieces in a buffer containing ~ 200 ng of trypsin (Promega, Sydney, New South Wales, Australia) in 10 mM of NH_4HCO_3 at 37 °C for 14 h, and peptides were extracted using acetonitrile and dried *in vacuo*.

Liquid chromatography and mass spectrometry

Peptide digests were rehydrated in a buffer containing 1% formic acid and 0.05% heptafluorobutyric acid. Digested peptides were separated by nano-LC using an Ultimate 3000 HPLC and autosampler

system (Dionex, Amsterdam, the Netherlands). Samples (2.5 μl) were concentrated and desalted onto a micro C18 precolumn (500 $\mu\text{m} \times 2$ mm; Michrom Bioresources, Auburn, CA, USA) with $\text{H}_2\text{O}:\text{CH}_3\text{CN}$ (98:2, 0.05% heptafluorobutyric acid) at 20 $\mu\text{l min}^{-1}$. After a 4 min wash the precolumn was switched (Valco 10 port valve; Dionex) into line with a fritless nano column (75 $\mu\text{m} \times \sim 10$ cm) containing C18 media (5 μm , 200 Å Magic; Michrom) manufactured according to Gatlin (Gatlin *et al.*, 1998). Peptides were eluted using a linear gradient of $\text{H}_2\text{O}:\text{CH}_3\text{CN}$ (98:2, 0.1% formic acid) to $\text{H}_2\text{O}:\text{CH}_3\text{CN}$ (64:36, 0.1% formic acid) at 350 nl min^{-1} over 30 min. High voltage (1800 V) was applied to the low volume tee (Upchurch Scientific, Oak Harbor, WA, USA) and the column tip was positioned ~ 0.5 cm from the heated capillary ($T = 250$ °C) of an LTQ FT Ultra (Thermo Electron, Bremen, Germany) mass spectrometer. Positive ions were generated by electrospray and the LTQ FT Ultra was operated in data-dependent acquisition mode. A survey scan m/z 350–1750 was acquired in the FT ICR cell (resolution = 1 00 000 at m/z 400, with an initial accumulation target value of 1 000 000 ions in the linear ion trap). Up to the six most abundant ions (> 3000 counts) with charge states of +2, +3 or +4 were sequentially isolated and fragmented within the linear ion trap using collision-induced dissociation with an activation $q = 0.25$ and activation time of 30 ms at a target value of 30 000 ions. m/z ratios selected for mass spectrometry–mass spectrometry (MS/MS) were dynamically excluded for 30 s. Peak lists were generated using Mascot Daemon/extract_msn (Matrix Science, Thermo, London, UK) using the default parameters, and submitted to the database search program Mascot (version 2.1; Matrix Science).

MS/MS data analysis and validation of protein identifications

Peptide identification was performed by database searching using Mascot. Mascot distiller (Applied Biosystems, Foster City, CA, USA) was used as the data import filter with the following criteria applied to the MS/MS ion search: a maximum of once missed cleavage for trypsin, peptide mass tolerance of ± 4 p.p.m., a fragment mass tolerance of ± 0.6 Da and variable modifications of acrylamide, carbamidomethyl and oxidation. The spectra generated from the 0.1 μm protein extraction of the 12.7 m sample were searched against the total assembled sequence database and the 14 m sample was searched against the NR database (January 2008). To discriminate between false-positive and confident peptide matches, validated candidate peptides against a decoy database to calculate the false discovery rate (FDR) in the Mascot searches. The decoy database consisted of shuffle sequences from all the peptides in the C-Ace and NR databases. The same sets of data were searched against their respective shuffle

databases, consisting of randomized amino-acid sequences to estimate false-positive protein matches. A cutoff score was established by adjusting the ion score to a value that, when the following formula was applied $2(n_{\text{shuf}}/(n_{\text{shuf}} + n_{\text{real}}))$, resulted in a peptide FDR of $\leq 5\%$ (n_{shuf} = no. of peptides identified against shuffle database, n_{real} = no. of peptides identified against real database). A second tier of data validation was applied where protein matches were accepted only if they were identified by a minimum of two peptides on the basis that one of the peptides had to be unique (that is, not matched against any other protein).

Results and Discussion

Overview of the proteogenomic data set

A metagenomic assembly derived from samples taken from the 12.7 m zone of Ace Lake revealed large scaffolds with a high level of similarity to genome sequences of GSB, in particular, *P. vibrioformis* DSM 265. As a first indication, of all ORF predictions derived from scaffolds greater than 10 kb, 76% were assigned first to *P. vibrioformis* DSM 265 when searched against the NCBI NR database (e-value $< 1E-5$). Further, the 10 most frequently assigned organisms were all from the Chlorobiaceae family, accounting for 85% of all assignments. A composite genome of the most abundant organism was constructed after assembly by the identification of scaffolds that clustered tightly together in feature-space (see Materials and methods). This resultant composite genome was composed of nine scaffolds totaling 1.79 Mb, similar in size (1.97 Mb) for *P. vibrioformis* DSM 265. Aligned using Nucmer (Kurtz et al., 2004), this composite genome covered 68.3% of the inferred reference *P. vibrioformis* DSM 265 at 86.0% average nucleotide identity. Metagene and MEGAN were used to predict ORFs and make taxonomic assignments, respectively (see Materials and methods). A total of 1631 genes (Supplementary Table S1) were identified in scaffolds of which 1560 were assigned to the Chlorobiaceae family, and 71 to other taxa (Supplementary Figure S1; Supplementary Table S2). Owing to the high number of matches to GSB in the Chlorobiaceae family, the organismal population represented by the composite genome is referred to as C-Ace. In addition to the C-Ace genome, phylogenetic profiling of the metagenome data that did not fall within the scaffolds was used for identifying specific genes that might also belong to C-Ace and could complete a biological process or pathway. To predict ORFs from the nine scaffolds, in addition to Metagene, we used Glimmer (trained against the largest contig of C-Ace). This resulted in a prediction of 17 *P. vibrioformis* DSM 265-related genes (not detected by Metagene), in addition to the 1631 genes predicted by Metagene (Supplementary Table S1: genes with 'Glimmer' under 'Gene no.' heading).

For metaproteomics, the biomass from 12.7 m collected on a 0.1 μm filter was extracted and processed through 12 liquid chromatography MS/MS runs (see Materials and methods) generating approximately 100 000 MS/MS spectra. A total of 689 proteins ($\sim 42\%$ of the 1631 predicted proteins) were identified with initial searches against the C-Ace genome, and refined to 504 ($\sim 31\%$ of predicted proteins) confident protein identifications (FDR $< 5\%$; minimum of two peptide matches; total 3970 unique peptides). Only 2 of the 504 proteins in the metaproteome matched to genes from the group of 71 genes that were assigned to taxa distinct from the Chlorobiaceae family; neither of these genes (159801926 and 159801879) could be assigned functions.

Eighty-nine percent of the total proteins from metaproteomics could be assigned to a COG category, with the largest number represented by Translation, ribosomal structure and biogenesis (J), Energy production and conversion (C) and amino-acid transport and metabolism (E) (Supplementary Figure S2). The Translation, ribosomal structure and biogenesis (J) and Energy production and conversion (C) COG categories were statistically overrepresented in the metaproteome versus the composite genome. In contrast, Cell envelope biogenesis, outer membrane (M), Defence mechanisms (V), Inorganic ion transport and metabolism (P), DNA replication, recombination and repair (L) and General function prediction only (R) categories were underrepresented in the metaproteome. The high representation and relative abundance of proteins from the J and C categories is likely to reflect their absolute abundance in the cell and the relative importance they have in enabling growth. In contrast, the reduced representation of proteins in categories M and P may relate to the membrane localization of many of the proteins in these categories, and the reduced efficiency of solubilizing these types of hydrophobic proteins during protein extraction.

A total of 45 proteins were predicted to have leader peptides, and 55 to have one or more transmembrane domain (Supplementary Figure S3). The majority of these were annotated with functions in proton/electron, phosphate, sodium, iron and protein transport. A number of others were predicted to be membrane-bound proteases/peptidases, chlorosome proteins or hypothetical proteins.

Major cellular pathways and microbial processes likely to be functional at the time of sampling were reconstructed from metaproteomic (and relevant metagenomic) data, using KEGG and Metacyc databases as references. In addition to the 12.7 m sample, metaproteomic data were generated from the 14 m zone (with matches against NR) to examine processes taking place in the anaerobic zone immediately below the oxycline; in particular, the sulfur cycle and possible interactions between GSB and SRB (see Sulfur cycle).

DNA RNA and protein processing

Fundamental processes related to transcription, translation, DNA replication and repair, cell division, protein folding and secretion were represented in the metaproteome (Supplementary Table S3). Protein secretion in *C-Ace* is likely to occur by mechanisms that include the Sec translocation system. Many components of this system were present in the metaproteome, including the signal recognition particle (159800974), which recognizes the leader peptide sequence of the protein and targets it for transport through the inner membrane of the cell, and the membrane-associated Sec translocase protein subunits, SecA (159800433), SecD (159803634), SecE (159803636), YajC (159800617) and YidC (159803616) (Supplementary Table S3). Several transcriptional regulators were identified (Supplementary Table S3) and one of these, ArsR, may be involved in metal ion resistance and ModE in molybdenum uptake. This may be an important regulator as many metals in the lake are at significantly higher levels than the ocean (Rankin *et al.*, 1999). Moreover, molybdenum is important as a cofactor in the reaction center (RC) of the light-harvesting apparatus (see Photosynthesis), and in several enzymes involved in sulfide oxidation (see Sulfur cycle).

Several DNA processing and nucleic acid binding proteins that have core functions in the biology of the cell may also be important for cold adaptation, as has been reported for other microorganisms (Grau *et al.*, 1994; Mizushima *et al.*, 1997; Graumann and Marahiel, 1999; Lim *et al.*, 2000; Giangrossi *et al.*, 2002; Cavicchioli, 2006; Rodrigues and Tiedji, 2008). These include, DNA gyrase A (159801931), DNA-binding protein HU- β (159803292), RecA (159801217), RecN (159802659), DrpA (159803132), a DEAD-box RNA helicase (159802095) and a cold-shock protein (159801159) (Supplementary Table S3). In addition, two types of peptidyl-prolyl *cis-trans* isomerases (Supplementary Table S3) may facilitate protein folding/refolding at low temperature.

Photosynthesis

The metaproteomic data are strongly supportive of *C-Ace* synthesizing a chlorosome-based photosynthesis system. Many *Chlorobia*-like chlorosome envelope proteins (CsmA, 159801197; CsmB, 159801265, 159800823, 159803384; CsmC, 159801195; CsmF, 159800523; CsmH, 159802205; CsmJ, 159802353) were detected (Supplementary Table S3; Supplementary Figure S1). In addition, a Fenna-Matthews-Olsen (FMO) protein (159802469) and three photosystem (Psc) proteins (PscA, 159803328; PscC, 159802726; PscD, 159802326) were detected. Chlorosomes are light-harvesting antenna structures packed with thousands of BChl *c*, *d* or *e* oligomers encased in a protein-associated lipid monolayer (Frigaard *et al.*, 1998; Figure 1). Light energy is funneled from the chlorosome via the baseplate through the FMO proteins into the RC, which is

located in the membrane bilayer (Hauska *et al.*, 2001; Hohmann-Marriott and Blankenship, 2007).

The *fmoA* gene is unique to GSB and has been used as a phylogenetic marker to classify species within the Chlorobiaceae family (Imhoff, 2003). Consistent with the overall best matches of the *C-Ace* genome to GSB, phylogeny inferred from the *fmoA* gene places *C-Ace* closest to *P. vibrioformis* (Supplementary Figure S4).

The three *C-Ace* genes annotated as *csmB* each had highest blastp matches to the single *csmB* gene in the genomes of three *Chlorobium* species (CT2054, Plut_2005 and Cpha266_0202). The genome sequence of *P. vibrioformis* DSM 265 is annotated with three *csmB* genes, and blastp matches to *csmD* and *csmI* from *C. tepidum* could not be identified. It is possible that *C-Ace* encodes multiple *csmB* genes and lacks *csmD* and *csmI* genes as does *P. vibrioformis*, or that *csmD* and *csmI* are present but were not sequenced. A *csmE* gene (but not protein) was identified in the metagenome data but was not the part of the composite *C-Ace* genome assembly (see Supplementary Table S3: data for *csm* gene functions). Although expression of *csmA* is essential for cell viability under photoautotrophic conditions in *C. tepidum* (Chung *et al.*, 1998), expression of all *csm* genes may not be necessary for the stability, or affect the biogenesis and light-harvesting function of chlorosomes (Frigaard *et al.*, 2004). Comparisons of GSB *csm* genes have classified chlorosome envelope proteins into four motif families: CsmA/CsmE, CsmB/CsmF (CsmH), CsmC/CsmD (CsmH) and CsmI/CsmJ/CsmX (Li and Bryant, 2009). The high sequence similarity and functional relatedness of Csm proteins within their groups is indicative of a series of gene duplication events and subsequent divergence (Li and Bryant, 2009). Representatives of each of the four Csm groups were identified in the *C-Ace* metaproteome. Owing to the functional redundancy within Csm groups, the identified proteins may represent the full complement expressed in *C-Ace*.

RC complex is composed of two copies of the integral membrane protein PscA, one copy of the peripheral protein PscB, two copies of the cytochrome *c* protein PscC and two copies of PscD (Hauska *et al.*, 2001). PscA binds the primary electron donor P840, the primary electron acceptor A₀ and the FeS cluster, F_X, whereas PscB attaches to the two terminal FeS clusters, F_A and F_B. Although PscB was not detected in the metaproteome, a *pscB* gene was identified in the metagenome data but was not part of the composite *C-Ace* genome assembly.

Our data indicate that *C-Ace* expresses the necessary chlorosome, FMO and RC proteins for cell viability and absorption of light energy for driving photosynthesis. The seasonal variation in light intensity at the depth where *C-Ace* dominates varies markedly (0.01–10 μmol of photons $\text{m}^2 \text{s}^{-1}$) (Burke and Burton, 1988b; Rankin *et al.*, 1999). Short periods of relatively high light intensity (for example, 100 μmol photon $\text{m}^2 \text{s}^{-1}$ during ice melt out)

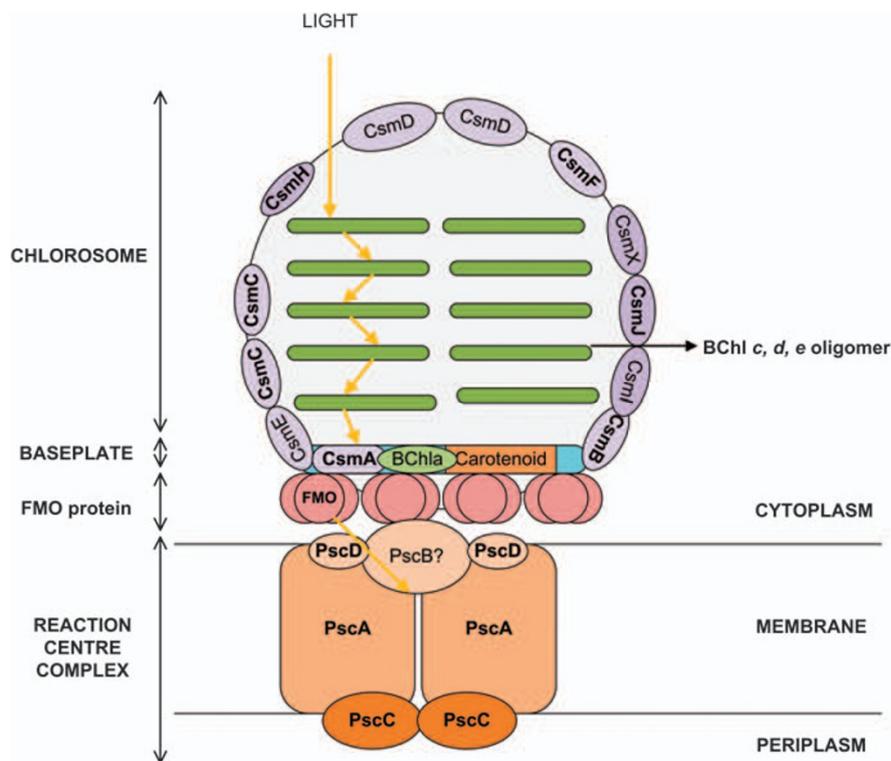


Figure 1 Photosynthetic apparatus. The photosynthetic unit comprises of the chlorosome containing the BChls encased in a monolipid-protein envelope, the Fenna–Matthews–Olson (FMO) protein and a reaction center (RC) complex. Proteins involved in chlorosome and envelop formation include CsmA, B, C, D, E, F, H, I, J and X. The BChls *c*, *d* or *e* are represented as rodlike structures absorbing light energy that gets funneled to the RC via the baseplate and FMO. The FMO protein is part of the photosynthetic apparatus and is organized as trimers and binds to BChl *a* in the baseplate. The FMO and baseplate mediates the transfer of energy between the chlorosome and RC, which is found in the inner membrane. The core RC of the complex is composed of two copies of integral membrane proteins, PscA and one copy of the peripheral protein PscB, which carries the FeS centers. Two copies of PscC cytochrome *c* proteins and one copy of PscD make up the rest of the RC complex. Proteins identified in the metaproteome are shown in bold. The diagram is adapted from Hauska *et al.* (2001) and Bryant and Frigaard (2006).

have also been recorded (Burke and Burton, 1988a,b). Clearly, it would be an advantage to be able to control the synthesis of the photosynthetic proteins (Morgan-Kiss *et al.*, 2009), perhaps by increasing the ratio of BChl *c* to BChl *a* within the chlorosomes, as observed in other GSB as light intensity decreases (Borrego and Garcia-Gil, 1994; Borrego *et al.*, 1999).

Isoprenoid, carotenoid and tetrapyrrole biosynthesis

Important enzymes in the methylerythritol phosphate (MEP; Supplementary Figure S5), carotenoid biosynthesis (Figure 2) and tetrapyrrole synthesis pathways (Figure 3) were detected in the metaproteome, illustrating that these pathways are functioning to support the growth of *C-Ace* in the lake. The MEP pathway (also known as the non-mevalonate pathway) is used for the biosynthesis of isopentenyl diphosphate and dimethylallyl diphosphate (Supplementary Figure S5; Supplementary Table S3); a pathway that has been identified in some phototrophs (Hunter, 2007).

Carotenoid biosynthesis commences with the synthesis of phytoene, leading to the synthesis of chlorobactene and a group of minor carotenoids

(Figure 2). In addition to converting γ -carotene to chlorobactene, in brown-pigmented GSB, CruU converts β -isorenieratene to isorenieratene. β -Isorenieratene is generated from chlorobactene by CruB (Maresca *et al.*, 2008). The *cruB* gene has been identified only in the genomes of isorenieratene-producing *Chlorobium* species (Maresca and Bryant, 2006). *P. vibrioformis* DSM 265 (green pigmented) encodes *cruA*, but not *cruB*. In the *C-Ace* genome, *cruA* was identified, but a blastp search for *cruB* from three isorenieratene-producing *Chlorobium* species revealed only matches to *cruA*. Because the *C-Ace* genome is incomplete, it is not possible to determine if *cruB* is encoded, and therefore whether it has the capacity to produce the brown pigment isorenieratene. However, the overall similarity between the genomes of *C-Ace* and *P. vibrioformis* DSM 265 suggests it to be green pigmented.

Tetrapyrrole synthesis is essential for the generation of heme, bacteriochlorophyll (BChl), siroheme and vitamin B₁₂ (Yaronskaya and Grimm, 2006). For *C-Ace*, a complete pathway was reconstructed that flows from glutamate metabolism to the synthesis of heme, siroheme, and BChl *a* and *c* (Figure 3). The activity of the BChU C-20 methyltransferase appears to determine the type of BChl accumulated in GSB,

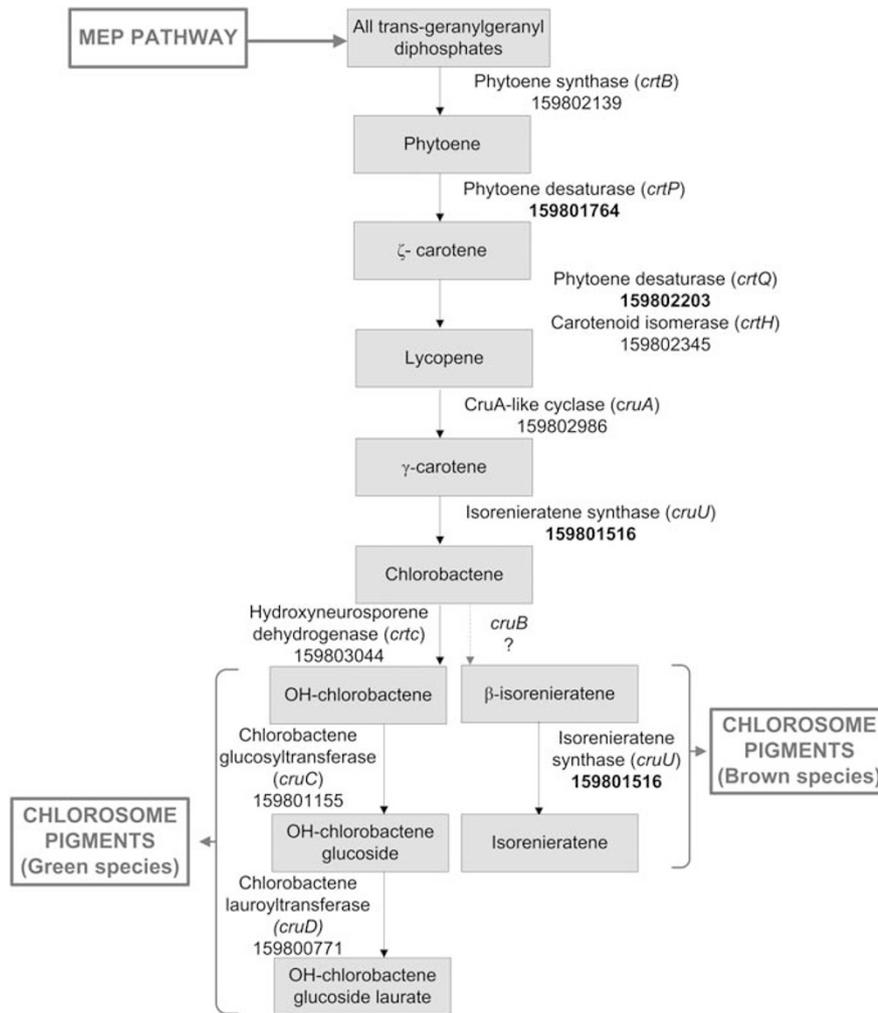


Figure 2 Pigment biosynthesis. Pathways for synthesis of green and brown pigments. Gene numbers corresponding to proteins detected in the metaproteome (bold font) or not detected (normal font) in the metaproteome but identified in the *C-Ace* metagenome are shown. Carotenoid biosynthesis is initiated by the condensation of two geranylgeranyl diphosphate (C_{20}) molecules to phytoene by phytoene synthase, *CrtB* (159802139) (Figure 2; Supplementary Table S3). Phytoene is converted by a series of desaturation, isomerization and cyclization steps to form γ -carotene, which is then converted to OH-chlorobactene by *CrtU* (159801516) and *CrtC* (159803044). *CrtC*, *CruC* (159801155) and *CruD* (15980077) are in turn collectively responsible for conversion of the major green pigment, chlorobactene to a group of minor carotenoids (Frigaard and Bryant, 2004; Maresca and Bryant, 2006).

and *Chlorobium* species unable to carry out this methylation have significantly reduced growth rates at low light intensities (Maresca *et al.*, 2004). The identification of BChU in the metaproteome suggests that *C-Ace* produces BChl *c*. BChl *c*-producing strains of *C. tepidum* have been shown to grow faster at low light intensities ($8 \mu\text{mol}$ of photons $\text{m}^2 \text{s}^{-1}$) compared with isogenic BChl *d* strains, and to be compromised at very high light intensities ($>300 \mu\text{mol}$ of photons $\text{m}^2 \text{s}^{-1}$) (Maresca *et al.*, 2004). In addition to being adapted to low light intensities, BChl *c*-producing strains of GSB are efficient in having an enhanced capacity to absorb light (wavelengths 350–850 nm) in comparison with their BChl *d*-producing counterparts (Maresca *et al.*, 2004). Red light ($\sim 700 \text{ nm}$) is attenuated by the ice cover on Ace Lake whereas green light ($\sim 519 \text{ nm}$) has the greatest penetration

through the water column (Burch, 1988), which falls within the absorption spectra of BChl *c*. If *C-Ace* is a BChl *c*-producing strain (as our data indicate), the light conditions (wavelength and seasonal intensity) that are experienced at the O_2 – H_2S interface in Ace Lake are likely to suit the growth of the organism.

Carbon dioxide fixation

Typical of GSB (Buchanan and Arnon, 1990; Eisen *et al.*, 2002), *C-Ace* appears to use the rTCA cycle for autotrophic fixation of CO_2 (Figure 4). Although most enzymes are common to both the oxidative and the reductive TCA cycles, the latter requires key enzymes to catalyze irreversible steps: pyruvate ferredoxin oxidoreductase (PFOR) and 2-oxoglutarate ferredoxin oxidoreductase (OGFOR), and an

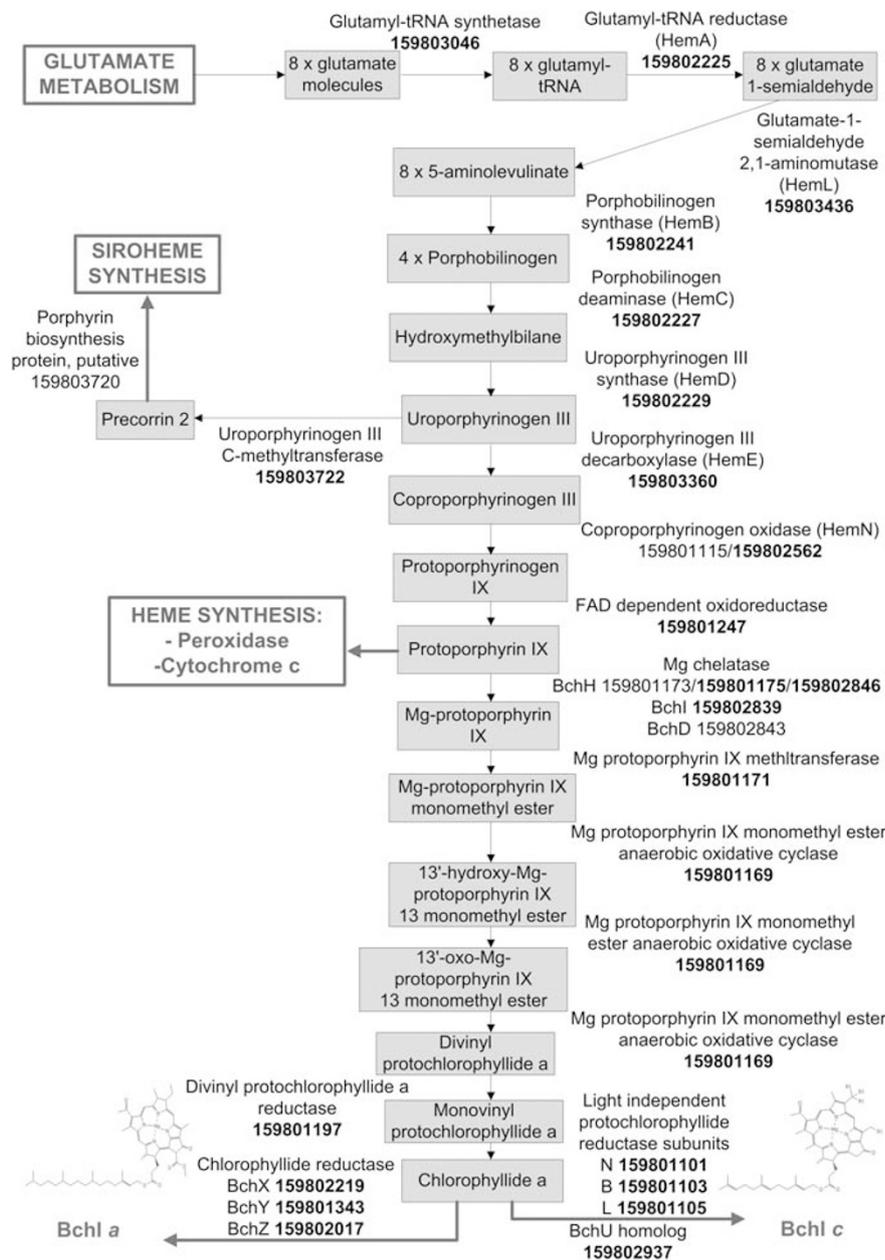


Figure 3 Bacteriochlorophyll biosynthesis. Pathway of bacteriochlorophyll *a/c* biosynthesis and associated enzymes in *C-Ace*. Through a series of reactions, from glutamate and ATP, uroporphyrinogen III is synthesized. The fate of uroporphyrinogen III is then determined by two enzymes, uroporphyrinogen-III C-methyltransferase (159803722) and uroporphyrinogen decarboxylase, HemE (159803360). The methylation of uroporphyrinogen III by uroporphyrinogen-III C-methyltransferase enables the synthesis of siroheme, vitamin B₁₂ and other coenzymes, whereas HemE feeds uroporphyrinogen III into a pathway for heme and BChl synthesis (Phillips *et al.*, 2003). Protoporphyrin IX is the branch point for heme or BChl synthesis, with the BChl pathway requiring the incorporation of Mg²⁺ into protoporphyrin IX to produce Mg-protoporphyrin IX. BChU C-20 methyltransferase (159803570) directs Mg-protoporphyrin IX toward BChl *c* synthesis, while a series of alternative reductions generates chlorophyllide *a* leading to BChl *c* synthesis. Unlike the light-dependent protochlorophyllide oxidoreductase (POR) in cyanobacteria, protochlorophyllide is converted to chlorophyllide in green sulfur bacteria (GSB) by light-independent protochlorophyllide oxidoreductase (DPOR) (Gomez Maqueo Chew *et al.*, 2007), a three subunit (BChB 159801103, BChN 159801101, BChL 159801105) enzyme complex. The light-independent reduction occurs in all anoxygenic photosynthetic bacteria enabling BChls to be synthesized irrespective of the presence of light (Fujita and Bauer, 2000).

ATP-dependent citrate lyase. All the enzymes are encoded in the *C-Ace* genome and most of the enzymes (or subunits thereof) for the rTCA cycle-specific and reversible reactions were detected in the metaproteome (Figure 4). 2-Oxoglutarate dehydrogenase was absent from the metagenome;

this enzyme is not present in GSB (Overmann and Garcia-Pichel, 2006), and its absence from *C-Ace* is therefore unlikely to be an artifact of an incomplete genome. Thus, it is likely that *C-Ace* has a partial oxidative TCA cycle that provides 2-oxoglutarate for ammonia assimilation. The detection of citrate

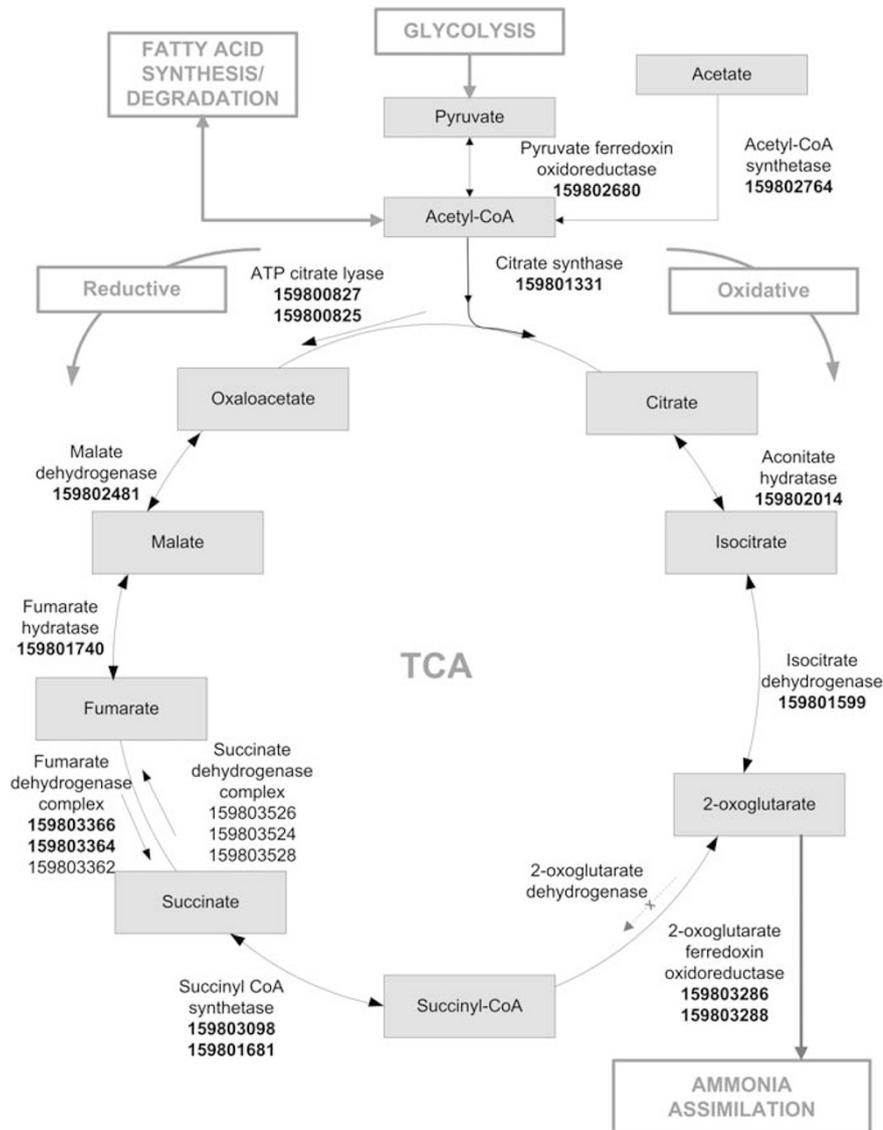


Figure 4 Tricarboxylic acid (TCA) cycle. Partial oxidative and complete reductive TCA cycles in *C-Ace*.

synthase (159801331) and ATP-citrate lyase (subunits 159800827 and 159800825) indicates that the TCA cycle is operating in both the reductive and the oxidative directions, for carbon fixation and ammonia assimilation, respectively. Generation of the essential precursor 2-oxoglutarate by the oxidative pathway requires fewer steps and no ATP compared with the reductive direction. The operation of a full circuit of the rTCA cycle alongside a partial oxidative cycle may indicate that *C-Ace* has sufficient regulatory capacity to avoid futile cycling of carbon flow.

Glycolysis, gluconeogenesis and acetate assimilation
Glycolytic and gluconeogenic pathways were identified in *C-Ace* (Supplementary Figure S6). Acetyl CoA generated from the rTCA cycle or by activation of exogenous acetate by acetyl-CoA synthetase

(159802764) can be reductively carboxylated to pyruvate by PFOR (159802680). The ability to assimilate simple organic carbon compounds such as acetate is advantageous for GSB when growth is impeded by limiting electron-donating substrates (for example, sulfide, thiosulfate) (Overmann and Garcia-Pichel, 2006). Cell yield has been shown to triple when acetate is provided as an additional carbon source in sulfide- and thiosulfate-limited conditions (Overmann and Pfennig, 1989). Decomposition of organic matter by microbial activity in Ace Lake is likely to provide acetate as an additional carbon source for *C-Ace* (Franzmann *et al.*, 1991b). The availability of acetate as an additional carbon source to CO₂ would enable increased biomass formation when conditions are suitable for phototrophic growth. The number of photosynthetic GSB in Ace Lake is lowest in winter (May–August) and highest in summer (December–January: 6×10^7 cells ml⁻¹), and

the capacity of *C-Ace* to reach high cell numbers during periods of optimal light conditions is likely to assist survival through the winter months (Burke and Burton, 1988b; Rankin *et al.*, 1999). As *C-Ace* cannot proliferate without light, cells could be more vulnerable during the winter period to lysis and grazing.

Storage and metabolism of organic carbon

A trehalose synthase (159803420) and an ADP-glucose type glycogen synthase (159801111) were identified in the metaproteome (Supplementary Table S3), illustrating a capacity to synthesize storage material. Accumulation of trehalose can promote salt tolerance (Makihara *et al.*, 2005) and cryoprotection. Members of the *Chlorobiaceae* grow under a range of hypersaline conditions, and the synthesis of trehalose by the enzyme trehalose synthase in *C-Ace* may provide a mechanism for osmoprotection in response to the high salt gradient at the chemocline (20–30 g l⁻¹), in addition to serving as storage material. Under conditions of nutrient imbalance, when carbon substrates and light energy are in excess, but nitrogen and phosphorus substrates are limiting, GSB synthesize glycogen, which is accumulated intracellularly (Overmann, 2006). The data for *C-Ace* are consistent with phototrophic growth in summer leading to the synthesis of glucose and complex sugars. Polymers such as glycogen could be broken down during dark periods (including during the winter months), and serve as a source of energy for cell maintenance. However, these would be unable to serve as a carbon reserve in the dark because the byproducts cannot be re-assimilated without light (Mas and Van Gemerden, 1995). Polymer-degrading enzymes, α -amylases (159803024, 159803422) and an α -glucan phosphorylase (159802594), were also detected in the metaproteome (Supplementary Table S3). The identification of enzymes involved in both polyglucose storage and breakdown might reflect conditions of sulfide flux. Efficient oxidation of sulfide by GSB during summer (see Sulfur cycling) may cause local depletion of this substrate. Under such conditions it is likely that carbon would be stored as glycogen. When sulfide becomes available from SRB, *C-Ace* accesses its polyglucose reserves to generate biomass. Oligotrophic marine bacteria have evolved enhanced capacities to generate storage material, compared with copiotrophs (Lauro *et al.*, 2009). Some species (for example, heterotrophic *Sphingopyxis alaskensis*) use storage pathways even when simpler forms of metabolizable carbon sources are available (Williams *et al.*, 2009). Similar to other GSB (Lauro *et al.*, 2009), *C-Ace* has genomic signatures of an oligotroph.

C-Ace has limited capabilities for the assimilation of organic compounds, which is consistent with other GSB (Eisen *et al.*, 2002). Thus, the partial phosphotransferase system consisting of enzyme I

(159803242), enzyme IIA (159803172) and an HPr kinase (Supplementary Table S3) is likely to serve a regulatory function in nitrogen and carbon metabolism, rather than in transport (Eisen *et al.*, 2002; Commichau *et al.*, 2006).

Collectively, these data indicate that in the nutrient-restricted lake environment, *C-Ace* has a system for gluconeogenesis and glycogen synthesis that can rely on CO₂ and acetate assimilation to synthesize glucose and store glycogen/starch reserves to facilitate growth when they are needed. In addition, trehalose synthesis may facilitate salt adaptation.

Pentose phosphate pathway

Ribose sugars (for nucleotide synthesis) and reducing power (NADPH) are generated by the pentose phosphate pathway. Although a complete set of genes are present in *C-Ace* (Supplementary Figure S7), the only enzyme involved in the oxidative steps that was identified in the metaproteome was glucose 6-phosphate dehydrogenase (159801277). In contrast, all the enzymes in the nonoxidative steps of the pentose phosphate pathway were identified in the metaproteome.

Fatty acid biosynthesis

A fatty acid biosynthesis (*fab*) gene cluster consisting of *fabF*, an acyl carrier protein (ACP), *fabG*, *fabD*, *fabH* and *plsX* (Heath and Rock, 2002; Mansilla *et al.*, 2004) is encoded and expressed in *C-Ace* (Supplementary Figure S8); only the ACP (159803296) was not detected in the metaproteome.

The first step in the anaerobic biosynthesis of straight/saturated chain fatty acids is the conversion of acetyl CoA to malonyl-CoA, catalyzed by a Mn²⁺ requiring acetyl-CoA carboxylase complex (Supplementary Figure S8). The initiation of fatty acid biosynthesis begins with the transfer of malonate group on malonyl-CoA to the ACP (159802037) by S-malonyltransferase (FabD—159802045). 3-Oxoacyl-ACP synthase III (FabH—159802049) catalyzes a condensation reaction between acetyl-CoA and malonyl-ACP to form a four-carbon, acetoacetyl-ACP with the loss of CO₂. The acetoacetyl-ACP proceeds through a cycle of chain elongation steps involving the addition of a two-carbon unit from malonyl-ACP to the growing acetyl-ACP molecule by 3-oxoacyl-ACP synthase II (FabF—159802033); the 3-oxoacyl-ACP synthase I, *fabB* gene that can also perform the same condensation reaction was not detected in the metagenome. Subsequently, ketoester formation (*fabG*—159802041), water removal (*fabZ*—159802776) and formation of a saturated acyl-ACP (*fabI*—159801595) enables further condensation reactions to occur.

The mechanism for unsaturated fatty acid synthesis in GSB is not yet known. In contrast to the highly conserved nature of the rest of the fatty acid synthesis pathway across bacteria, the enzyme

responsible for insertion of the *cis* double bond of unsaturated fatty acids is highly variable, and has yet to be identified in many anaerobic bacteria. GSB lack FabA, and it is not known if FabZ (a homologue of FabA) can serve as an isomerase as well as a dehydratase in GSB. FabF is likely to regulate the balance of monounsaturated versus saturated fatty acid products in *C-Ace*, as it does in *Escherichia coli* after a downshift in growth temperature (de Mendoza *et al.*, 1992). At low temperature, an increased rate of synthesis and accumulation of unsaturated 18:1 Δ 9 (*cis*-vaccenic acid) has been observed in *E. coli* (de Mendoza *et al.*, 1992). FabF is responsible for the elongation of 16:1 Δ 9 (palmitoleic acid) over saturated fatty acids 16:0 (palmitic acid), and directs substrates toward the synthesis of 18:1 Δ 9 in response to a reduction in temperature (de Mendoza *et al.*, 1992; Russell, 2008). Fatty acid analysis of particulate matter in Ace Lake identified the straight chain saturated 16:0 and monounsaturated 16:1 Δ 9 fatty acids as the most abundant at depths where *C-Ace* predominates (Volkman *et al.*, 1988). Other lower abundance fatty acids included 18:1 Δ 9, 18:0, 14:0 and 15:0. *Chlorobium* species have been attributed to producing the 16:0, 16:1 Δ 9 and 14:0 fatty acids (Volkman *et al.*, 1988).

Overall, the metaproteomic data for *C-Ace* are consistent with FabF performing a key role in the *in situ* synthesis of monounsaturated fatty acids, and *C-Ace* contributing to the abundant fatty acid signatures detected in Ace Lake. The synthesis of unsaturated fatty acids by *C-Ace* is also consistent with the role that unsaturated fatty acids have in enhancing membrane fluidity at low temperatures; a physiological adaptation common in cold-adapted microorganisms (Russell, 2008). Lipid analysis of thermophilic *C. tepidum* detected predominantly saturated fatty acids and low quantities of unsaturated fatty acids (Sørensen *et al.*, 2007).

Ammonia assimilation

With the carbon skeleton provided by 2-oxoglutarate, ammonia/ammonium is assimilated into two central nitrogen intermediates, glutamate and glutamine, which serve as nitrogen donors for the cell's entire biosynthetic needs (Supplementary Figure S9). The ammonia concentration measured in Ace Lake at the depth where *C-Ace* resides (and deeper) was 50–55 μ M, at least 10-fold higher than that of nitrate and nitrite (Burton, 1980). Both glutamine synthetase (159802199) and glutamate synthase (59801694, 159802974, 159801691) were represented in the metaproteomic data (Supplementary Figure S9; Supplementary Table S3). Enzymes required for the fixation of dinitrogen or the assimilation of nitrate were present in the metagenome, but absent from the metaproteomic data (Supplementary Table S3). The data suggest that at the time of sampling *C-Ace* derived its nitrogen from ammonia, a preferred nitrogen source, and would

proceed through the ATP-dependent glutamine synthetase/glutamate synthase pathway (Supplementary Figure S9), the preferred pathway under conditions of low ammonia concentration.

Sulfur cycling

Sulfur-oxidizing GSB require light as the energy source, with hydrogen sulfide (H₂S), elemental sulfur and thiosulfate among potential electron donors for the reduction of CO₂. The ability to use thiosulfate as an electron donor is not found in all GSB, and in the absence of a Sox cluster (*soxXY-ZABW*) in the metagenome data, this may also be the case in *C-Ace*. A pathway for sulfide oxidation was defined for *C-Ace* with many of the relevant proteins detected in the metaproteome (Figure 5 enlarged inset of sulfur cycle; Supplementary Table S3). Sulfide–quinone reductase, SQR (159803450), oxidizes sulfide to zero-valent sulfur, which is released as a soluble polysulfide chain (Griesbeck *et al.*, 2002). Flavocytochrome *c*, FccAB (159803630, 159803632), was also detected; as an alternative mechanism for sulfide oxidation to sulfur, its role in GSB is unclear (Frigaard and Bryant, 2008). The presence of SQR and FccAB in the *C-Ace* metaproteome indicates that both are relevant to sulfide oxidation in Ace Lake, and may contribute to the depletion of this reduced sulfur species from the depth at which *C-Ace* grows (see Storage and metabolism of organic carbon). The dissimilatory sulfite reductase system, DsrNCABLEEFHTMKJOP (159803690–159803718) catalyzes the subsequent conversion of sulfur to sulfite. The most likely candidate for the final oxidation of sulfite to sulfate is a polysulfide-reductase-like complex 3 (159803444, 159803446, 159803448). A sulfate permease, SulB (159802643), is likely to be used to export sulfate produced by the cell, making it available as an electron acceptor for sulfate reduction by SRB. Similar to most GSB, there is no evidence for assimilatory sulfate reduction in *C-Ace*, emphasizing its strict dependence on H₂S generated by SRB. A number of other proteins, whose roles in sulfur metabolism are not clear, were identified in the metaproteome or metagenome, including heterodisulfide reductase subunits AB (159800421, 159800423) and a molybdenum-containing protein (159803678) related to polysulfide reductase (Supplementary Table S3).

The H₂S concentration in Ace Lake (December 1987) increases from 0–1 mM at the depth where *C-Ace* grows, to 8 mM at the deepest point of the lake (Rankin *et al.*, 1999). In contrast, sulfate is undetectable at maximum depth, increases to a peak of 9 mM at a depth immediately below the chemocline and decreases to 3 mM at the surface of the lake. These data are consistent with H₂S released by SRB in the anoxic zones diffusing up the water column and sulfide being fully used by *C-Ace* (and other GSB). The production of sulfate by *C-Ace* would in turn provide a continuous cycling of sulfur

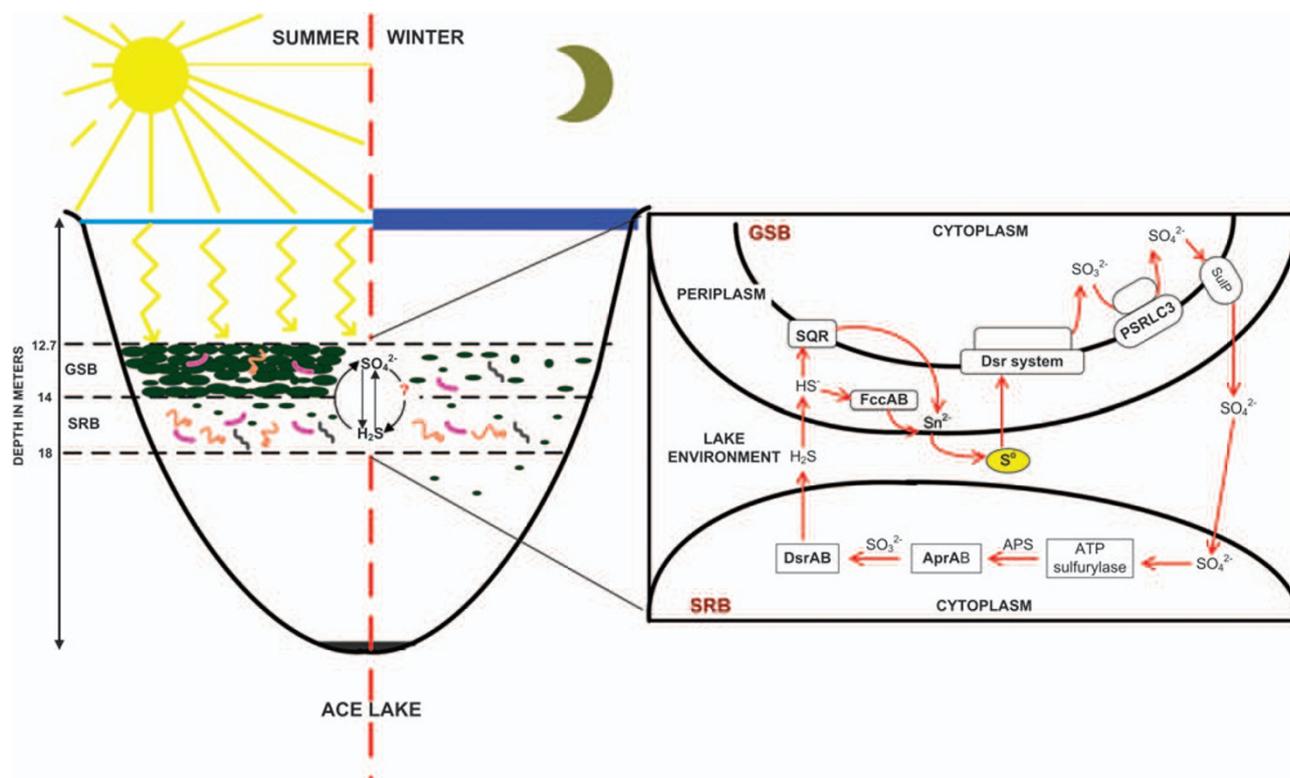


Figure 5 Sulfur cycle. Sulfur cycle for *C*-Ace and sulfate-reducing bacteria (SRB) within Ace Lake. Sulfide in the form of H_2S is generated by the community of SRB residing in the anoxic zone of the lake (below 12.7 m). The exchange of sulfur between the green sulfur bacteria (GSB) and SRB is shown in the inset. During summer, the thick zone of *C*-Ace just below the chemocline (12 m) oxidizes the sulfide produced from the SRB to sulfate. Sulfate then diffuses to neighboring SRB in the zone and down the water column, whereupon it is reduced back to sulfide. The extent of light penetration in summer versus winter is shown, with phototrophic energy generation and cell growth occurring during the summer months. Enzymes identified in the metaproteome are shown in the inset.

compounds between the GSB and the SRB in the lake (Figure 5). Support for this was further obtained from metaproteomic analysis of biomass collected from the 14 m zone directly below the 12.7 m zone of *C*-Ace. Enzymes used by SRB in sulfate reduction that were identified included an adenylylsulfate reductase, AprA subunit, and a dissimilatory sulfite reductase enzyme, DsrA and B subunit (Supplementary Table S3). Preliminary metagenomic analysis of data derived from the 3 μm membrane fraction of the 12.7 m zone also revealed the presence of SRB, indicating that the SRB coexist with the GSB. These SRB had a high proportion of genes involved in motility, chemotaxis and degradation of chlorinated aromatic compounds, indicating the SRB can use recalcitrant compounds (possibly deposited as the remains of zooplankton and phytoplankton from the upper aerobic zone) as a carbon source while cycling sulfate to sulfide.

Oxidative stress

The dissolved O_2 concentration measured at the time of sampling was ~ 0.4 mM at the surface and ~ 0.2 mM at the 12.7 m sampling depth (data not shown). Owing to its position near the steep oxycline in the lake, and its obligate anaerobic

metabolism, *C*-Ace may be prone to oxidative stress by reactive oxygen species generated as a result of the low potential of electron transfers typical of anaerobes (Imlay, 2003). Several oxidative stress proteins that may fulfill roles in oxidative defense were identified in the metaproteome (Supplementary Table S3): a superoxide dismutase (159801535), catalase-peroxidase/hydroperoxidase I (159801779), alkyl hydroperoxide reductase (159802085), thiol peroxidase (159802716), thioredoxins (159801715, 159801555) and thioredoxin reductase (159801713). A number of *C*-Ace proteins contain FeS clusters (for example, chlorosome RC, coenzymes, bacteriochlorophylls), which are extremely sensitive to attack by the superoxide radical, O_2^- . The removal of O_2^- is catalyzed by superoxide dismutase enzymes through the dismutation of O_2^- to hydrogen peroxide, H_2O_2 . H_2O_2 is still potentially toxic to the cell and can be converted to H_2O and O_2 by catalase-peroxidase/hydroperoxidase I. Alkyl hydroperoxide reductase, thiol peroxidase and thioredoxins are able to reduce other organic hydroperoxides and may also catalyze the conversion of H_2O_2 (Farr and Kogoma, 1991; Li *et al.*, 2009). Alkyl hydroperoxide reductase has also been shown to scavenge endogenous H_2O_2 (Li *et al.*, 2009). In *C. tepidum*, chlorobactene has also been shown to function as a

photoprotectant, and specific quinones within the chlorosome have an even larger role in providing a mechanism to quench the excited states of chlorosomes (on exposure to light and O₂) preventing the production of singlet oxygen species (Frigaard and Bryant, 2006; Kim *et al.*, 2007; Li *et al.*, 2009). Similar mechanisms may be used by *C-Ace* to counter oxidative damage if cells are exposed to O₂ during summer.

Genes that do not match known GSB

Among the 71 genes that were assigned to phyla distinct from GSB in the Chlorobiaceae family (Supplementary Figure S1; Supplementary Table S2), a number of them provided clues about possible mechanisms of adaptation. In *C-Ace*, six genes were predicted to be involved in capsular polysaccharide formation and lipopolysaccharide production: putative capsular polysaccharide biosynthesis protein, mannose-1-phosphate guanylyltransferase, UDP-*N*-acetylglucosamine 2-epimerase, UDP-*N*-acetyl-D-mannosamine dehydrogenase, GDP-mannose 4,6-dehydratase and a putative glycosyltransferase (genes 159800457, 159800859, 159800861, 159800863, 159800461 and 159800467 in Supplementary Table S2). The presence of these genes may indicate a requirement for polysaccharide structures distinct from other known members of the Chlorobiaceae as a mechanism for achieving cold adaptation. Genome evolution through lateral gene transfer was reported to be important for cold adaptation of *Methanococcoides burtonii*; a methanogen isolated from the bottom waters of Ace Lake (Allen *et al.*, 2009). Consistent with the analysis of the *C-Ace* genome, the acquisition of polysaccharide biosynthesis genes by lateral gene transfer was linked to the ability of *M. burtonii* to grow at low temperature (Allen *et al.*, 2009).

Genes encoding the three subunits (M, S and R) of a type I DNA restriction and modification (RM) system were arranged in *C-Ace* genome in the order: subunit R (159800462), subunit S (159800469), a gene with a DNA-binding domain (159800473) and subunit M (159800475). In addition, two copies of a *res* gene (159801046 and 159802022), belonging to one of the two subunits (Res and Mod) of a type III restriction enzyme complex were present in the genome. One possibility is that these defense mechanisms help protect *C-Ace* against bacteriophage attack over the dark winter period, when cell replication has effectively ceased. An ABC transporter for the uptake of iron was also encoded in the *C-Ace* genome (159802314, 159802318, 159802322). The ability to import iron is essential for the synthesis of heme groups and the Fe-S clusters of RCs of the light-harvesting units in GSB. Another gene encoding the periplasmic component of an ABC transporter (159802310) involved in nitrate/sulfonate/bicarbonate uptake was also identified. In the *M. burtonii* genome, the Defence mechanisms

COG category was overrepresented compared to genomes from archaeal thermophiles and hyperthermophiles (Allen *et al.*, 2009). The individual COGs were mainly from type I RM systems, and specific ABC transporters. The presence of specific polysaccharide biosynthesis, RM system and ABC transporter genes in *C-Ace* and *M. burtonii*, is consistent with them having a role in cold adaptation in both archaea and bacteria, and argues for further experimental work to be performed to test these inferences.

Conclusion

The ability to perform metaproteomics with matching metagenome data for an environmental community consisting largely of a single GSB, *C-Ace*, has allowed us to construct a working model describing the protein complement required for thriving under cold, nutrient- and O₂-limited, and extreme light conditions (Figure 6). The success of GSB has been attributed to the extremely efficient light capturing abilities of the chlorosomes, which allows growth at extremely low light intensities, and (for those GSB that have a Dsr system) the efficient and complete oxidation of sulfur (Frigaard and Bryant, 2006, 2008). The survival strategy for *C-Ace* during the winter months depends on building up sufficient biomass to remain viable during winter and re-commence growth when sufficient light resumes later in the season. Although this capacity has been inferred (Burke and Burton, 1988a,b), the pathways that facilitate an obligate phototrophic metabolism of GSB, mixotrophic assimilation of simple carbon compounds and sulfide oxidation in Antarctic GSB have been mapped out for the first time, in *C-Ace*.

We regard the metabolic simplicity of *C-Ace* as an asset for dealing with an environment as challenging as Ace Lake: nutrient-depleted (oligotrophic) conditions with minimal light, little or no O₂, but available H₂S. Metabolic conservatism is a hallmark of an oligotrophic physiology (Lauro *et al.*, 2009), and *C-Ace* can generate biomass from a narrow pool of ubiquitous substrates (for example, CO₂, acetate, pyruvate). The metabolic efficiency is further underscored by having CO₂ fixation and the generation of the carbon skeleton for amino-acid biosynthesis being served by a single pathway that can be turned in either reductive or oxidative directions. The rTCA cycle for carbon assimilation has been considered the primordial pathway for CO₂ fixation, and to have its origins in anaerobic environments rich in H₂S (Wächtershäuser, 1990). The metaproteome data for *C-Ace* also indicate that although the 2-oxoglutarate necessary for ammonia assimilation could be obtained from the rTCA cycle (Overmann, 2006), this precursor can be generated through the less energy-intensive oxidative route.

This study identified traits that are likely to be specifically important for adaptation to the cold and the proliferation of *C-Ace* in Ace Lake. This

Division. Mass spectrometric results were obtained at the Bioanalytical Mass Spectrometry Facility within the Analytical Centre of the University of New South Wales. This work was undertaken using infrastructure provided by NSW Government co-investment in the National Collaborative Research Infrastructure Scheme. Subsidized access to this facility is gratefully acknowledged. The work of JCVI members was supported by funding provided from the Gordon and Betty Moore Foundation.

References

- Allen MA, Lauro FM, Williams TJ, Burg D, Siddiqui KS, De Francisci D *et al.* (2009). The genome sequence of the psychrophilic archaeon, *Methanococcoides burtonii*: the role of genome evolution in cold adaptation. *ISME J* **3**: 1012–1035.
- Beatty JT, Overmann J, Lince MT, Manske AK, Lang AS, Blankenship RE *et al.* (2005). An obligately photosynthetic bacterial anaerobe from a deep-sea hydrothermal vent. *Proc Natl Acad Sci USA* **102**: 9306–9310.
- Biebl H, Pfennig N. (1978). Growth yields of green sulfur bacteria in mixed cultures with sulfur and sulfate reducing bacteria. *Arch Microbiol* **117**: 9–16.
- Blankenship RE, Olson JM, Miller M. (1995). Antenna complexes from green sulfur bacteria. In Blankenship RE, Madigan MT, Bauer CE (eds). *Anoxygenic Photosynthetic Bacteria*. Kluwer: Dordrecht, The Netherlands, pp 399–435.
- Blum H, Beier H, Gross HJ. (1987). Silver staining of proteins in polyacrylamide gels. *Electrophoresis* **8**: 93–99.
- Borrego CM, Garcia-Gil LJ. (1994). Rearrangement of light harvesting bacteriochlorophyll homologs as a response of green sulfur bacteria to low light intensities. *Photosynth Res* **45**: 21–30.
- Borrego CM, Gerola PD, Miller M, Cox RP. (1999). Light intensity effects on pigment composition and organization in the green sulfur bacterium *Chlorobium tepidum*. *Photosynth Res* **59**: 159–166.
- Bryant DA, Frigaard NU. (2006). Prokaryotic photosynthesis and phototrophy illuminated. *Trends Microbiol* **14**: 488–496.
- Buchanan BB, Arnon DI. (1990). A reverse KREBS cycle in photosynthesis: consensus at last. *Photosynth Res* **24**: 47–53.
- Burch MD. (1988). Annual cycle of phytoplankton in Ace Lake, an ice covered, saline meromictic lake. *Hydrobiologia* **165**: 59–75.
- Burke CM, Burton HR. (1988a). The ecology of photosynthetic bacteria in Burton Lake, Vestfold Hills, Antarctica. *Hydrobiologia* **165**: 1–11.
- Burke CM, Burton HR. (1988b). Photosynthetic bacteria in meromictic lakes and stratified fjords of the Vestfold Hills, Antarctica. *Hydrobiologia* **165**: 13–23.
- Burton HR. (1980). Methane in a saline Antarctic lake. In: Trudinger PA, Walter MR, Ralph, BJ (eds). *Biogeochemistry of Ancient and Modern Environments*. Australian Academy of Science: Canberra, Australia, pp 243–251.
- Cavicchioli R. (2006). Cold-adapted Archaea. *Nat Rev Microbiol* **5**: 331–343.
- Chung S, Shen G, Ormerod J, Bryant DA. (1998). Insertional inactivation studies of the *csmA* and *csmC* genes of the green sulfur bacterium *Chlorobium vibrioforme* 8327: the chlorosome protein CsmA is required for viability but CsmC is dispensable. *FEMS Microbiol Lett* **164**: 353–361.
- Commichau FM, Forchhammer K, Stülke J. (2006). Regulatory links between carbon and nitrogen metabolism. *Curr Opin Microbiol* **9**: 167–172.
- Coolen MJL, Muyzer G, Schouten S, Volkman JK, Sinninghe Damsté 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.
- de Mendoza D, Garwin JL, Cronan Jr JE. (1992). Overproduction of *cis*-vaccenic acid and altered temperature control of fatty acid synthesis in mutant *Escherichia coli*. *J Bacteriol* **151**: 1608–1611.
- Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. (1999). Improved microbial gene identification with GLIMMER. *Nucleic Acid Res* **27**: 4636–4641.
- Eisen JA, Nelson KE, Paulsen IT, Heidelberg JF, Wu M, Dodson RJ *et al.* (2002). The complete genome sequence of *Chlorobium tepidum* TLS, a photosynthetic, anaerobic, green-sulfur bacterium. *Proc Natl Acad Sci USA* **99**: 9509–9514.
- Farr SB, Kogoma T. (1991). Oxidative stress response in *Escherichia coli* and *Salmonella typhimurium*. *Microbiol Rev* **55**: 561–585.
- Franzmann PD, Roberts NJ, Mancuso CA, Burton HR, McMeekin TA. (1991b). Methane production in meromictic Ace Lake, Antarctica. *Hydrobiologia* **210**: 191–201.
- Franzmann PD, Skyring GW, Burton HR, Deprez PP. (1988). Sulfate reduction rates and some aspects of limnology of four lakes and a fjord in the Vestfold Hills, Antarctica. *Hydrobiologia* **165**: 25–33.
- Frigaard NU, Bryant DA. (2004). Seeing green bacteria in a new light: genomics-enabled studies of the photosynthetic apparatus in green sulfur bacteria and filamentous anoxygenic phototrophic bacteria. *Arch Microbiol* **182**: 265–276.
- Frigaard NU, Bryant DA. (2006). Chlorosomes: antenna organelles in photosynthetic green bacteria. In: Shively JM (ed). *Complex Intracellular Structures in Prokaryotes*. Springer: Berlin, Germany, pp 79–114.
- Frigaard NU, Bryant DA. (2008). Genomic insights into the sulfur metabolism of phototrophic green sulfur bacteria. In: Hell R, Dahl C, Knaff D, Leustek T (eds). *Sulfur Metabolism in Phototrophic Organisms*. Springer: New York, USA, pp 60–76.
- Frigaard NU, Maresca JA, Yunker CE, Jones AD, Bryant DA. (2004). Genetic manipulation of carotenoid biosynthesis in the green sulfur bacterium *Chlorobium tepidum*. *J Bacteriol* **186**: 5210–5220.
- Frigaard NU, Matsuura K, Hirota M, Miller M, Cox RP. (1998). Studies of the location and function of isoprenoid quinones in chlorosomes from green sulfur bacteria. *Photosynth Res* **58**: 81–90.
- Fujita Y, Bauer CE. (2000). Reconstitution of light-independent protochlorophyllide reductase from purified BChl and BChN-BChB subunits. *In vitro* confirmation of nitrogenase-like features of a bacteriochlorophyll biosynthesis enzyme. *J Biol Chem* **275**: 23583–23588.
- Gatlin CL, Kleemann GR, Hayes LG, Link AJ, Yates III JR. (1998). Protein identification at the low femtomole level from silver-stained gels using a new fritless

- electrospray interface for liquid chromatography-microspray and nanospray mass spectrometry. *Anal Biochem* **263**: 93–101.
- Giangrossi M, Giuliadori AM, Gualerzi CO, Pon CL. (2002). Selective expression of the β -subunit of nucleoid-associated protein HU during cold shock in *Escherichia coli*. *Mol Microbiol* **44**: 205–216.
- Gibson JAE. (1999). The meromictic lakes and stratified marine basins of the Vestfold Hills, East Antarctica. *Antarct Sci* **11**: 175–192.
- Gomez Maqueo Chew A, Frigaard NU, Bryant DA. (2007). Bacteriochlorophyllide *c* C-8² and C-12¹ methyltransferases are essential for adaptation to low light in *Chlorobaculum tepidum*. *J Bacteriol* **189**: 6176–6184.
- Grau R, Gardiol D, Glikin GC, de Mendoza D. (1994). DNA supercoiling and thermal regulation of unsaturated fatty acid synthesis in *Bacillus subtilis*. *Mol Microbiol* **11**: 933–941.
- Graumann PL, Marahiel MA. (1999). Cold shock response in *Bacillus subtilis*. *J Mol Microbiol Biotechnol* **2**: 203–209.
- Griesbeck C, Schütz M, Schödl T, Bathe S, Nausch L, Mederer N et al. (2002). Mechanism of sulfide–quinone reductase investigated using site-directed mutagenesis and sulfur analysis. *Biochemistry* **41**: 11552–11565.
- Hauska G, Schoedl T, Remigy H, Tsiotis G. (2001). The reaction centre of green sulfur bacteria. *Biochim Biophys Acta* **1507**: 260–277.
- Heath RJ, Rock CO. (2002). The Claisen condensation in biology. *Nat Prod Rep* **19**: 581–596.
- Hohmann-Marriott M, Blankenship RE. (2007). Variable fluorescence in green sulfur bacteria. *Biochim Biophys Acta* **1769**: 106–113.
- Hunter WN. (2007). The non-mevalonate pathway of isoprenoid precursor biosynthesis. *J Biol Chem* **282**: 21573–21577.
- Imhoff JF. (2003). Phylogenetic taxonomy of the family *Chlorobiaceae* on the basis of 16S rRNA and *fmo* (Fenna–Matthews–Olson protein) gene sequence. *Int J Syst Evol Microbiol* **53**: 941–951.
- Imlay JA. (2003). Pathways of oxidative damage. *Annu Rev Microbiol* **57**: 395–418.
- Kim H, Li H, Maresca JA, Bryant DA, Savikhin S. (2007). Triplet excitation formation as a novel photoprotection mechanism in chlorosomes in *Chlorobium tepidum*. *Biophys J* **93**: 192–201.
- Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol* **5**: R12.
- Lauro FM, McDougald D, Thomas T, Willams TJ, Egan S, Rice S et al. (2009). The genomic basis of trophic strategy in marine bacteria. *Proc Natl Acad Sci* **106**: 15527–15533.
- Labourn-Parry J, Perriss SJ. (1995). The role and distribution of the autotrophic ciliate *Mesodinium rubrum* (*Myrionecta rubra*) in three Antarctic saline lakes. *J Plankton Res* **17**: 1835–1850.
- Li H, Bryant DA. (2009). Envelope proteins of the CsmB/CsmF and CsmC/CsmD motif families influence the size, shape, and composition of chlorosomes in *Chlorobaculum tepidum*. *J Bacteriol* **191**: 7109–7120.
- Li H, Jubelirer S, Costas AMG, Frigaard NU, Bryant DA. (2009). Multiple antioxidant proteins protect *Chlorobaculum tepidum* against oxygen and reactive oxygen species. *Arch Microbiol* **191**: 853–867.
- Lim J, Thomas T, Cavicchioli R. (2000). Low temperature regulated DEAD-box RNA helicase from Antarctic archaeon, *Methanococoides burtonii*. *J Mol Biol* **297**: 553–567.
- Makihara F, Tsuzuki M, Sato K, Masuda S, Nagashima KV, Abo M et al. (2005). Role of trehalose synthesis pathways in salt tolerance mechanism of *Rhodobacter sphaeroides* f. sp. *denitrificans* IL106. *Arch Microbiol* **184**: 56–65.
- Mallorqui N, Arellano JB, Borrego CM, Garcia-Gil LJ. (2005). Signatures of green sulfur bacteria in lower Pleistocene deposits from the Banyoles lacustrine area (Spain). *J Paleolimnol* **34**: 271–280.
- Mansilla MC, Cybulski LE, Albanesi D, de Mendoza D. (2004). Control of membrane lipid fluidity by molecular thermosensors. *J Bacteriol* **186**: 6681–6688.
- Manske AK, Glaeser J, Kuypers MMM, Overmann J. (2005). Physiology and phylogeny of green sulfur bacteria forming a monospecific phototrophic assemblage at a depth of 100 meters in the Black Sea. *Appl Environ Microbiol* **71**: 8049–8060.
- Maresca JA, Bryant DA. (2006). Two genes encoding new carotenoid-modifying enzyme in the green sulfur bacterium *Chlorobium tepidum*. *J Bacteriol* **188**: 6217–6223.
- Maresca JA, Gomez Maqueo Chew A, Ponsati MR, Frigaard NU, Ormerod JG, Bryant DA. (2004). The *bchU* gene of *Chlorobium tepidum* encodes the C-20 methyltransferase in bacteriochlorophyll *c* biosynthesis. *J Bacteriol* **186**: 2558–2566.
- Maresca JA, Romberger SP, Bryant DA. (2008). Isorenieratene biosynthesis in green sulfur bacteria requires the cooperative actions of two carotenoid cyclases. *J Bacteriol* **190**: 6384–6391.
- Martínez-Alonso M, van Bleijswijk J, Gaju N. (2005). Diversity of anoxygenic phototrophic sulfur bacteria in the microbial mats of the Ebro Delta: a combined morphological and molecular approach. *FEMS Microbiol Ecol* **52**: 339–350.
- Mas J, Van Gemerden H. (1995). Storage products in purple and green sulfur bacteria. In: Blankenship RE, Madigan MT, Bauer CE (eds). *Anoxygenic Photosynthetic Bacteria*. Kluwer: Dordrecht, The Netherlands, pp 973–990.
- Mizushima T, Kataoka K, Ogata Y, Inoue R, Sekimizu K. (1997). Increase in negative supercoiling of plasmid DNA in *Escherichia coli* exposed to cold shock. *Mol Microbiol* **23**: 381–386.
- Morgan-Kiss RM, Chan LK, Modla S, Weber TS, Warner M, Czymmek KJ et al. (2009). *Chlorobaculum tepidum* regulates chlorosome structure and function in response to temperature and electron donor availability. *Photosynth Res* **99**: 11–21.
- Musat N, Halm H, Winterholler B, Hoppe P, Peduzzi S, Hillion F et al. (2008). A single-cell view on the ecophysiology of anaerobic phototrophic bacteria. *Proc Natl Acad Sci USA* **105**: 17861–17866.
- Noguchi H, Park J, Takagi T. (2006). MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res* **34**: 5623–5630.
- Overmann J. (2006). The family Chlorobiaceae. In: Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E (eds). *The Prokaryotes: A Handbook on the Biology of Bacteria*. Springer: New York, USA, pp 359–378.
- Overmann J, Garcia-Pichel F. (2006). The phototrophic way of life. In: Dworkin M, Falkow S (eds). *The Prokaryotes: Ecophysiology and Biochemistry*. Springer: New York, USA, pp 32–85.

- Overmann J, Pfennig N. (1989). *Pelodictyon phaeoclathratiforme* sp. nov., a new brown-coloured member of the *Chlorobiaceae* forming net-like colonies. *Arch Microbiol* **152**: 401–406.
- Pfennig N. (1975). The phototrophic bacteria and their role in sulfur cycling. *Plant Soil* **43**: 1–16.
- Pfennig N. (1978). General physiology and ecology of phototrophic bacteria. In: Clayton RK, Sistrom WR (eds). *The Photosynthetic Bacteria*. Plenum Press: New York, USA, pp 3–18.
- Phillips JD, Whitty FG, Kushner JP, Hill CP. (2003). Structural basis for tetrapyrrole coordination by uroporphyrinogen decarboxylase. *EMBO J* **22**: 6225–6233.
- 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.
- Rodrigues DF, Tiedji JM. (2008). Coping with our cold planet. *Appl Environ Microbiol* **74**: 1677–1686.
- Rodriguez-Brito B, Rohwer F, Edwards RA. (2006). An application of statistics to comparative metagenomics. *BMC Bioinformatics* **7**: 162.
- Rusch DB, Halpern AL, Sutton G, Heidelberg K, Williamson S, Yooseph S et al. (2007). The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biol* **5**: e77.
- Russell NJ. (2008). Membrane components and cold sensing. In: Margesin R, Schinner F, Marx JC, Gerday C (eds). *Psychrophiles: From Biodiversity to Biotechnology*. Springer: Berlin, Germany, pp 177–190.
- Saunders NFW, Ng C, Raftery M, Guilhaus M, Goodchild A, Cavicchioli R. (2006). Proteomic and computational analysis of secreted proteins with type I signal peptides from the Antarctic archaeon. *J Proteome Res* **5**: 2457–2464.
- Sørensen PG, Cox RP, Miller M. (2007). Chlorosome lipids from *Chlorobium tepidum*: characterization and quantification of polar lipids and wax esters. *Photosynth Res* **95**: 191–196.
- Teeling H, Meyerdierks A, Bauer M, Amann R, Glöckner FO. (2004). Application of tetranucleotide frequencies for the assignment of genomic fragments. *Environ Microbiol* **6**: 938–947.
- Vila X, Guyoneaud R, Cristina XP, Figueras JB, Abella CA. (2002). Green sulfur bacteria from hypersaline Chiprana Lake (Monegros, Spain): habitat description and phylogenetic relationship of isolated strains. *Photosynth Res* **71**: 165–172.
- Volkman JK, Burton HR, Everitt DA, Allen DI. (1988). Pigment and lipid composition of algae and bacterial communities in Ace Lake, Vestfold Hills, Antarctica. *Hydrobiologica* **165**: 41–57.
- Wächtershäuser G. 1990. Evolution of the first metabolic cycles. *Proc Natl Acad Sci USA* **87**: 200–204.
- Wahlund TM, Woese CR, Castenholz RW, Madigan MT. (1991). A thermophilic green sulfur bacterium from New Zealand hot springs, *Chlorobium tepidum* sp. nov. *Arch Microb* **156**: 81–90.
- Ward DM, Ferris MJ, Nold SC, Bateson MM. (1998). A natural view of microbial diversity within hot spring cyanobacterial mat communities. *Microbiol Mol Biol Rev* **62**: 1353–1370.
- Willams T, Ertan H, Ting L, Cavicchioli R. (2009). Carbon and nitrogen substrate utilization in the marine bacterium *Sphingopyxis alaskensis* strain RB2256. *ISME J* **3**: 1036–1052.
- Yaronskaya E, Grimm B. (2006). The pathway from 5-aminolevulinic acid to protochlorophyllide and protoheme. In: Grimm B, Porra RJ, Rüdiger W, Scheer H (eds). *Chlorophylls and Bacteriochlorophylls: Biochemistry, Biophysics, Functions and Applications*. Springer: New York, USA, pp 173–188.

Supplementary Information accompanies the paper on The ISME Journal website (<http://www.nature.com/ismej>)