

ORIGINAL ARTICLE

Metaproteomic analysis of a winter to spring succession in coastal northwest Atlantic Ocean microbial plankton

Anna A Georges¹, Heba El-Swais¹, Susanne E Craig², William KW Li² and David A Walsh¹
¹Department of Biology, Concordia University, Montreal, Quebec, Canada and ²Department of Fisheries and Oceans, Bedford Institute of Oceanography, Dartmouth, Nova Scotia, Canada

In this study, we used comparative metaproteomics to investigate the metabolic activity of microbial plankton inhabiting a seasonally hypoxic basin in the Northwest Atlantic Ocean (Bedford Basin). From winter to spring, we observed a seasonal increase in high-affinity membrane transport proteins involved in scavenging of organic substrates; Rhodobacterales transporters were strongly associated with the spring phytoplankton bloom, whereas SAR11 transporters were abundant in the underlying waters. A diverse array of transporters for organic compounds were similar to the SAR324 clade, revealing an active heterotrophic lifestyle in coastal waters. Proteins involved in methanol oxidation (from the OM43 clade) and carbon monoxide (from a wide variety of bacteria) were identified throughout Bedford Basin. Metabolic niche partitioning between the SUP05 and ARCTIC96BD-19 clades, which together comprise the Gamma-proteobacterial sulfur oxidizers group was apparent. ARCTIC96BD-19 proteins involved in the transport of organic compounds indicated that in productive coastal waters this lineage tends toward a heterotrophic metabolism. In contrast, the identification of sulfur oxidation proteins from SUP05 indicated the use of reduced sulfur as an energy source in hypoxic bottom water. We identified an abundance of Marine Group I Thaumarchaeota proteins in the hypoxic deep layer, including proteins for nitrification and carbon fixation. No transporters for organic compounds were detected among the thaumarchaeal proteins, suggesting a reliance on autotrophic carbon assimilation. In summary, our analyses revealed the spatiotemporal structure of numerous metabolic activities in the coastal ocean that are central to carbon, nitrogen and sulfur cycling in the sea.

The ISME Journal advance online publication, 9 January 2014; doi:10.1038/ismej.2013.234

Subject Category: Microbial ecology and functional diversity of natural habitats

Keywords: microbial oceanography; proteomics; SUP05; SAR324; marine Thaumarchaeota

Introduction

Coastal marine ecosystems support metabolically diverse microbial plankton that have essential roles in energy and nutrient cycling, including the processing of dissolved organic matter from both marine and terrestrial sources (Hedges *et al.*, 1997; Kujawinski, 2011). Heterotrophic marine microbes exhibit various levels of specialization for the uptake and processing of specific components of the dissolved organic matter pool (Pomeroy *et al.*, 2007; Poretsky *et al.*, 2010; Teeling *et al.*, 2012). Moreover, recent studies have uncovered novel metabolic capabilities in marine bacteria and archaea, including rhodopsin-based phototrophy

(Béjà *et al.*, 2000), aerobic anoxygenic phototrophy (Béjà *et al.*, 2002), CO₂ fixation (Newton *et al.*, 2010) and chemolithotrophic oxidation of inorganic compounds such as ammonia (Walker *et al.*, 2010), sulfur (Walsh *et al.*, 2009) and hydrogen (Anantharaman *et al.*, 2013). Metagenomics and single-cell amplified genomes have provided insight into the diversity and metabolic potential of microbial communities in the ocean (DeLong *et al.*, 2006; Yooseph *et al.*, 2007; Woyke *et al.*, 2009; Swan *et al.*, 2011; Grzyski *et al.*, 2012; Rinke *et al.*, 2013).

Functional approaches, such as metatranscriptomics and metaproteomics, are now required to assess gene expression and microbial activity in the environment. Although RNA transcript abundance can provide insight into the metabolic activities of organisms, mass spectrometry (MS)-based metaproteomics approaches are particularly attractive as they provide a direct assessment of *in situ* protein expression (VerBerkmoes *et al.*, 2009).

Correspondence: DA Walsh, Department of Biology, Concordia University, 7141 Sherbrooke Street West, Montreal H4B 1R6, Canada.

E-mail: david.walsh@concordia.ca

Received 20 August 2013; revised 26 October 2013; accepted 23 November 2013

A current limitation of metaproteomics is the requirement of a representative protein sequence database for peptide sequence identification. Nevertheless, a number of community metaproteomics studies have been recently performed on a variety of marine habitats, including the Sargasso Sea (Sowell *et al.*, 2009), along a productivity gradient in the South Atlantic (Morris *et al.*, 2010), a site along the Oregon coast (Sowell *et al.*, 2011), coastal waters of the Antarctic Peninsula (Williams *et al.*, 2012) and a phytoplankton bloom in the North Sea (Teeling *et al.*, 2012).

In this study, we used comparative metaproteomics to investigate the metabolic activities of microbial plankton inhabiting the stratified waters of Bedford Basin, a temperate northwest Atlantic inlet located in Nova Scotia, Canada. Bedford Basin has a maximum depth of 71 m and is connected to the adjoining continental shelf through a narrow (400 m) and shallow (20 m) sill. Importantly, Bedford Basin is the location of a >20-year microbial plankton monitoring program, focused on deciphering long-term plankton dynamics (Li and Dickie, 2001; Li and Harrison, 2008). Time series studies have shown that the annual cycle of phytoplankton cell abundances in Bedford Basin is coherent with the adjoining northwest Atlantic Ocean (Li *et al.*, 2006). In the spring, a bloom is induced due to stabilization of the nutrient-rich water column; phytoplankton are maintained during the summer by regenerated nutrients; in autumn, a secondary bloom is driven by nutrients transported upward as water column stability is eroded; in winter, the water column is mixed and phytoplankton are at a minimum. It is noteworthy that during stratification the deep water is isolated by the sill and periodically becomes hypoxic and occasionally anoxic. During these oxygen minima, the occurrence of nitrification, denitrification and nitrous oxide production have been observed (Punshon and Moore, 2004). Therefore, the well-understood annual cycle in Bedford Basin makes this coastal inlet an attractive system for investigating gene expression patterns and metabolic activities of marine microbes in response to variable environmental conditions. Moreover, a metagenomic database was previously generated for Bedford Basin and the surrounding waters during the Global Ocean Sampling (GOS) expedition (Rusch *et al.*, 2007). Here we used this metagenomic database and a collection of 205 reference genomes from marine bacteria and archaea to identify expressed proteins in the surface and deep waters of Bedford Basin during two distinct seasons (winter and spring). Through comparative metaproteomic analyses, we identified a wide variety of metabolic pathways operating in the cold winter surface and bottom hypoxic waters, as well as surface and bottom waters associated with the Northwest Atlantic spring phytoplankton bloom.

Materials and methods

Sample collection

Seawater samples were collected from the compass buoy station (44°41'37"N, 63°38'25"W) in Bedford Basin in conjunction with the Bedford Basin Monitoring Program. Details of sampling and collection of physicochemical data can be found in Li and Dickie (2001). Water samples (10 l) for metaproteomic and 16S ribosomal RNA (rRNA) analysis were collected from 5 and 60 m during the winter (11 January 2011) and spring (25 May 2011). Seawater was prefiltered through a Whatman GF/D filter (2.7 µm cutoff) and cells were collected on a 0.22 µm Sterivex filter. After filtration, 1.8 ml of sucrose-based lysis buffer was added and filters were stored at -80 °C.

Bacterial 16S rRNA gene sequence analysis

DNA was extracted from filters as described in Zaikova *et al.* (2010). The V5 region of the 16S rRNA gene was PCR-amplified using the primers DW786F (5'-GATTAGATACCCTSGTAG-3') and DW926R (5'-CCGTCAATTCMTTTRAGT-3'). These primers were modified from Baker *et al.* (2003) to eliminate bias against marine Alpha-proteobacteria. PCR were performed in 50-µl volumes containing 0.5 µM of each primer, 1X Phire Reaction Buffer (1.5 mM MgCl₂), 0.20 mM of dNTPs, 1.0 µl of Phire Hot Start II DNA Polymerase (Finnzymes Thermofisher Scientific, Mississauga, ON, Canada) and ~10 ng of DNA. Cycling conditions involved an initial 3-min denaturing step at 98 °C, followed by 30 cycles of 5 s at 98 °C, 5 s at 49 °C and 10 s at 72 °C, and a final elongation step of 1 min at 72 °C. Each reverse primer was barcoded with a specific IonXpress sequence to identify samples. PCR amplicons were purified using QIAquick Gel Extraction Kit (Qiagen, Toronto, ON, Canada), pooled and sequenced using an Ion Torrent PGM system and a 316 chip with the Ion Sequencing 200 kit as described in Yergeau *et al.* (2012). V5 amplicons were analyzed using the Mothur pipeline (Schloss *et al.*, 2009). Sequences with average quality <17, length <100 bp or that did not match the IonXpress barcode sequence and PCR reverse primer sequence were discarded. The V5 data set comprised 25 149 reads (5388–7464 reads per sample). Sequences were assigned to taxonomic groups using the Wang method and a bootstrap value cutoff of >60%. Similarity between surface and deep water communities was calculated using the Bray–Curtis similarity between taxonomic profiles.

Metaproteomics

Protein extraction was performed using a combination of a guanidine-HCl lysis method optimized for low biomass (Thompson *et al.*, 2008) and BugBuster (Novagen, Gibbstown, NJ, USA) treatment (full protocol available as Supplementary Online

Material). In gel trypsin digestion of extracted protein was performed as previously described (Shevchenko *et al.*, 2006). Liquid chromatography-MS was performed using a Linear Trap Quadrupole Orbitrap Velos (Thermo Fisher Scientific, Mississauga, ON, Canada) mass spectrometer and digested peptides were previously separated using EASY-nLC (Proxeon, Mississauga, ON, Canada) nano high-performance liquid chromatography. Samples were directly injected onto a nano column (75 μm inner diameter by 15 cm) containing C18 media (Jupiter 5 μm , 300 Å, Phenomenex, Torrance, CA, USA). Separation was achieved by applying a 1% to 27% acetonitrile gradient in 0.1% formic acid over 80 min at 400 nl min^{-1} then followed by a linear gradient of 27% to 52% over 20 min at 400 nl min^{-1} . Electrospray ionization was enabled through applying a voltage of 2.0 kV through a PEEK junction at the inlet of the nano column operated in positive mode with data-dependent acquisition mode. A survey scan was obtained with the Orbitrap mass detector at m/z of 360–2000 with a resolution of 60 000 at m/z 400. The automatic gain control settings were 3×10^6 ions and 2.5×10^5 ions for survey scan. Ions were selected for fragmentation on the linear trap when their intensity was >500 counts. The 10 most intense ions were isolated for ion trap CID with charge states of 2 to 4 and sequentially isolated for fragmentation within the linear ion trap using collision induced dissociation energy set at 35% with q activation set at 0.25 and an activation time of 10 ms at a target value of 30 000 ions. Ions selected for MS/MS were dynamically excluded for 30 s.

Protein identification

To identify peptide sequences, spectra were compared with a protein database consisting of protein-coding sequences from GOS samples collected from and near Bedford Basin (GS02-13) (Rusch *et al.*, 2007) and 205 bacterial and archaeal genomes (see Supplementary Table 1). All spectra were searched using SEQUEST (Eng *et al.*, 1994) implemented in Proteome Discoverer (Thermo Scientific, Mississauga, ON, Canada) with the settings: enzyme type, trypsin; maximum missed cleavage sites, 2; maximum modification for peptide, 4; static modifications, +57 Da (iodoacetamide modification of cysteine) and +16 Da (oxidation of methionine). Percolator, a supervised machine learning tool to discriminate between correct and decoy spectrum (Kall *et al.*, 2007) was used to increase the rate and confidence of peptide identification. Percolator settings were as follows: maximum delta Cn, 0.05; Target False Discover Rate (strict), 0.01; and validation based on q -values. Using these stringent settings, we included peptides identified only once in a sample. The search input consisted of 592 815 spectra. In Proteome Discoverer, we used the 'Enable Protein Group' setting to

group proteins that share sets of identified peptides (that is, redundant proteins) and select the highest scoring protein in the group to serve as the representative proteins (that is, master protein). Protein expression levels were estimated using the spectral counts method by summing all the spectra for the collection of peptides matching each protein (VerBerkmoes *et al.*, 2009).

Taxonomic and functional annotation of proteins

Each master protein was searched against the RefSeq (v52) protein database using BLASTP and the top 10 hits were recorded. The BLASTP search results were loaded into MEGAN and taxonomic assignment was performed using the lowest common ancestor algorithm (bit score >80) (Huson *et al.*, 2011). Proteins were also queried against the Clusters of Orthologous Genes (COG) database to identify a probable function. Only proteins with high sequence similarity (bit score of >80) to a COG were annotated with the corresponding function. The RefSeq analysis were used to further aid in assigning function to the proteins.

Results

Site description and physicochemical environment

In this study, we performed metaproteomic analyses on samples collected during winter (11 January 2011) and spring (25 May 2011) from surface (5 m) and bottom (60 m) waters of Bedford Basin. Over this period, the water column exhibited the typical seasonal stratification cycle (Figure 1). Although density stratification was minimal in January, chemical stratification was apparent and the bottom water was hypoxic ($1.61 \text{ ml l}^{-1} \text{ O}_2$). Most nutrients were at higher concentration in the deep ($21.41 \mu\text{M NO}_3^-$, $45.53 \mu\text{M SiO}_3^{2-}$, $5.36 \mu\text{M PO}_4^{3-}$) than at the surface ($6.83 \mu\text{M NO}_3^-$, $8.87 \mu\text{M SiO}_3^{2-}$, $0.96 \mu\text{M PO}_4^{3-}$). However, lower concentrations of ammonium were observed in the deep ($0.44 \mu\text{M NH}_4^+$), compared with the surface ($2.68 \mu\text{M NH}_4^+$) layer. Primary production was low in the winter ($1.13 \text{ mg m}^{-3} \text{ Chl } a$). Surface waters warmed through the spring and two spring phytoplankton blooms occurred, as evidenced by peaks in surface fluorescence (Figure 1). Samples were collected at the peak of the second phytoplankton bloom ($12.7 \text{ mg m}^{-3} \text{ Chl } a$). At this time, nutrient concentrations were lower at the surface ($0.22 \mu\text{M NO}_3^-$, $0.28 \mu\text{M SiO}_3^{2-}$, $0.39 \mu\text{M PO}_4^{3-}$, $0.65 \mu\text{M NH}_4^+$) than in bottom water ($7.32 \mu\text{M NO}_3^-$, $15.21 \mu\text{M SiO}_3^{2-}$, $2.05 \mu\text{M PO}_4^{3-}$, $13.28 \mu\text{M NH}_4^+$), reflecting the uptake of inorganic nutrient by plankton in surface waters and remineralization of organic matter at depth.

Bacterial community structure

During winter, surface and deep water communities were taxonomically indistinguishable from one

another (Bray–Curtis community similarity = 0.96; Figure 2). The winter community comprised of Alpha-proteobacteria (36%), Gamma-proteobacteria (25%) and Flavobacteriales (20%). Within the Alpha-proteobacteria, the SAR11 clade and the

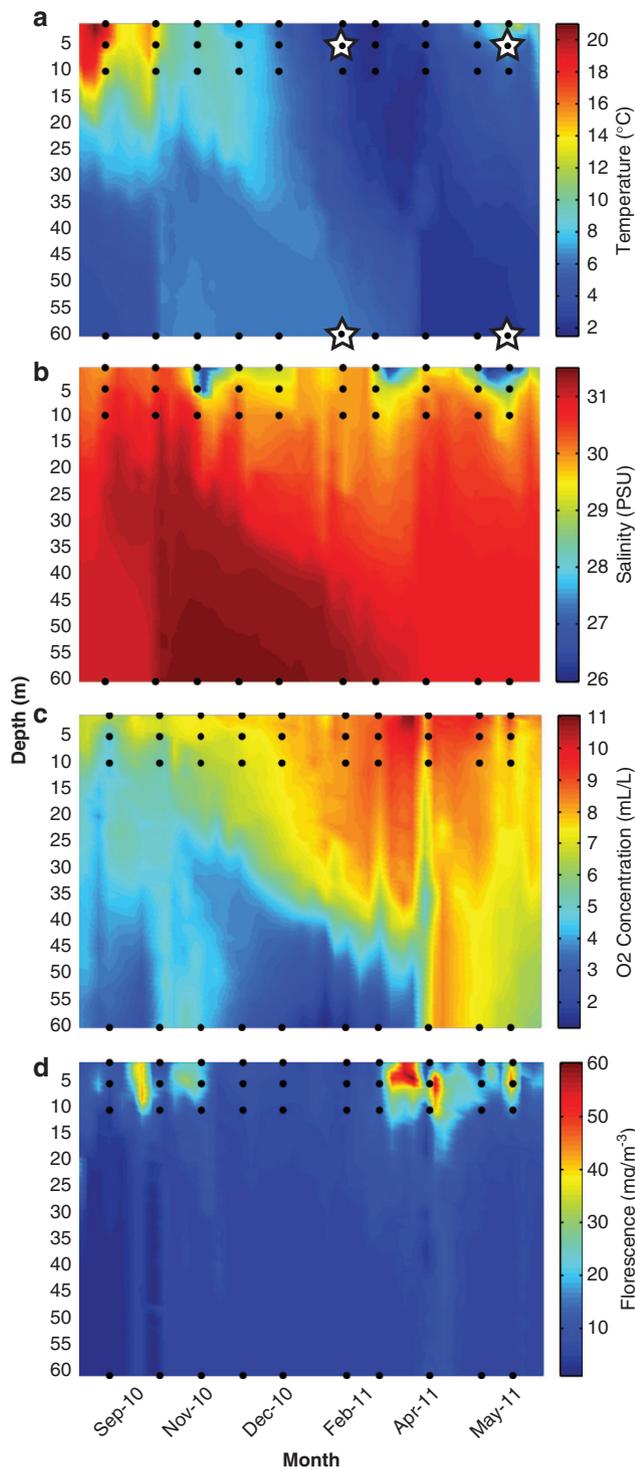


Figure 1 Physicochemical conditions in Bedford Basin over the time-frame relevant to this study. (a) Temperature, (b) salinity (PSU, Practical Salinity Units), (c) oxygen and (d) fluorescence. Stars represent the location of the samples used for metaproteomic analysis.

Rhodobacterales comprised 17% and 13% of sequences, respectively, whereas the OM38 clade was represented by 3% of sequences. The Gamma-proteobacteria were largely Alteromonadales (11%), SAR86 (2%), ZA2333c (2%) and the Gamma-proteobacterial sulfur oxidizers (GSOs; 5%) clade. Beta-proteobacteria (5%), specifically the OM43 clade (3%), were also detected in winter. Vertical structure of the community was observed in spring (Bray–Curtis community similarity = 0.46; Figure 2). Bacteria associated with the surface phytoplankton bloom comprised of Flavobacteriales (47%), Gamma-proteobacteria (20%) and Alpha-proteobacteria (19%). The Flavobacteria comprised of Cytophaga (33%) and Polaribacter (6%), whereas the Alpha-proteobacteria were mostly Rhodobacterales (10%) and the Gamma-proteobacteria were mostly SAR92 (4%), Alteromonadales (4%) and the ZA2333c group (6%). In the deep water, the relative abundance of the SAR11 clade (16%) and Rhodobacterales (10%) was similar to that observed in the winter. However, there was a significant increase in Gamma-proteobacteria (42%), particularly the GSO clade (18%). Nitrospina (1%) and the Delta-proteobacteria SAR324 (4%) clade were also detected in the spring deep layer. The strong vertical partitioning of bacterial taxa associated with the spring phytoplankton bloom suggests distinct metabolic strategies within the surface and bottom water communities.

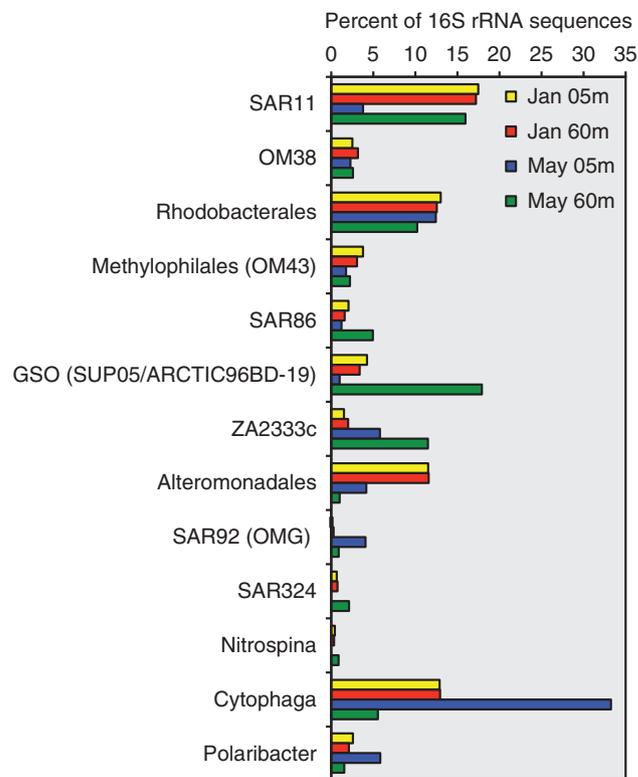


Figure 2 Taxonomic composition of the Bedford Basin bacterial community based on 16S rRNA gene sequences.

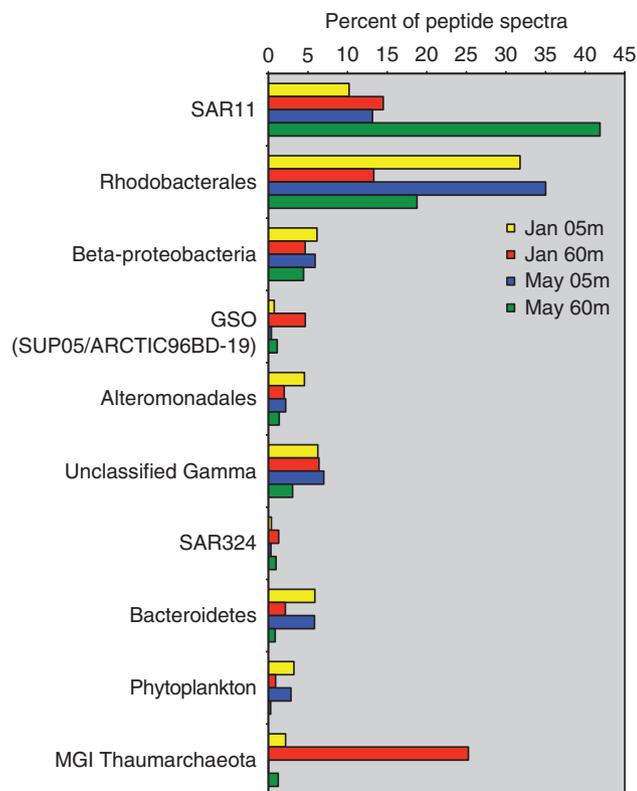


Figure 3 Taxonomic composition of identified proteins in Bedford Basin metaproteomes based on spectral counts.

Metaproteomics

We used MS/MS-based proteomics to examine protein expression patterns in Bedford Basin. We identified 11 668 unique peptide sequences from 5627 distinct proteins (between 2463 and 3018 per sample) (Supplementary Tables S2 and S3). The phylogenetic identity of proteins demonstrated that bacterial and archaeal protein abundances (based on spectral counts) exhibited seasonal and vertical partitioning in the water column (Figure 3). Irrespective of season, Alpha-proteobacteria were represented by 55–60% of peptide mass spectra from surface water, and were comprised mostly of Rhodobacterales (32–35%) and to a lesser extent SAR11 proteins (10–13%). In bottom waters, a similar abundance of Rhodobacterales proteins was detected in winter (13%) and spring (18%). However, SAR11 proteins in the bottom water increased markedly in abundance from winter (15%) to spring (42%), revealing a highly active SAR11 population underlying the spring phytoplankton bloom. Gamma-proteobacteria proteins ranged between 12% and 21% of spectra. We detected proteins seeming to originate from the sulfur-oxidizing SUP05 (4%) clade in the winter hypoxic layer, whereas proteins from the closely related ARCTIC96BD-19 clade were less abundant (~1%), but identified in all samples. Proteins from the Alteromonadales (4%) were particularly abundant in winter surface water, but also found

throughout Bedford Basin. Proteins from Beta-proteobacteria (4%–6%) and the SAR324 clade (1%–3%) were also ubiquitous in our samples.

For the most part, the taxonomic composition of the proteomes and the 16S rRNA gene data were in agreement. However, there was a notable exception. Proteins most closely related to Bacteroidetes represented ~6% of spectra in surface waters, although the Flavobacteria comprised nearly half of the 16S rRNA gene sequences associated with the phytoplankton bloom. This discrepancy may reflect inadequate representation of Bedford Basin Flavobacteria in the protein database used to identify peptides.

High-affinity transport proteins

In addition to ribosomal subunits and chaperonins, the most prevalent proteins we identified were components of high-affinity ATP-binding cassette (ABC) transporters (751 proteins) and tripartite ATP-independent periplasmic (TRAP) transporters (202 proteins; Figure 4). The majority of both types of transporters were represented by periplasmic-binding proteins (PBPs). The high representation of soluble PBPs is commonly observed in metaproteomic studies (Sowell *et al.*, 2009, 2011; Williams *et al.*, 2012) and is likely due to a combined effect of high expression level, which enhances the frequency of substrate capture, as well as the inherent difficulty in extracting membrane proteins from microbial cells compared with soluble proteins such as PBPs (Morris *et al.*, 2010).

ABC and TRAP transporter proteins were identified in all samples, but were consistently more abundant in spring than winter. The winter to spring succession in transporter protein abundance was characterized by a large increase in transporters specific for organic substrates (that is, amino acids, dipeptides, sugars, glycine betaine and taurine; Figure 5), suggesting competition within the community for the acquisition of organic compounds associated with phytoplankton production. We observed a greater than twofold (12% to 28% of all spectra) seasonal increase in transporter abundance in the surface layer and a greater than threefold increase in the deep layer (17% to 57% of all spectra). The vast majority (91% of transporter spectra) were assigned to Alpha-proteobacteria, particularly to members of the SAR11 and Rhodobacterales clades. Surface transporters were comprised largely of Rhodobacterales proteins (winter, 5%; spring, 12%) and to a slightly lesser extent SAR11 (winter, 4%; spring, 9%) whereas the bottom layer were dominated by SAR11 (winter, 11%; spring, 33%) and to a much lesser extent to Rhodobacterales (winter, 3%; spring, 10%).

Membrane transporters for inorganic nutrients were not as frequently observed as those for the uptake of organic compounds, yet still showed a near fourfold seasonal increase (0.26% to 0.9%) in expression level from winter to spring (Figure 5).

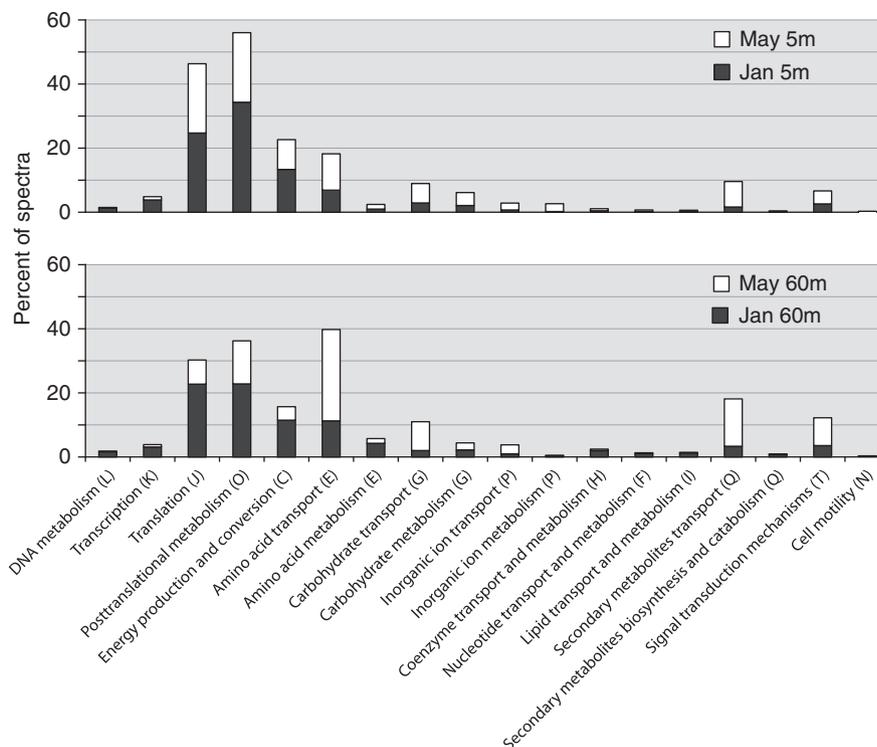


Figure 4 Functional composition of identified proteins in Bedford metaproteomes based on spectral counts. Proteins within the transporter COG categories (E, G, P and Q) were differentiated into those specifically involved in transport function and those involved in the intracellular metabolism of transported solutes.

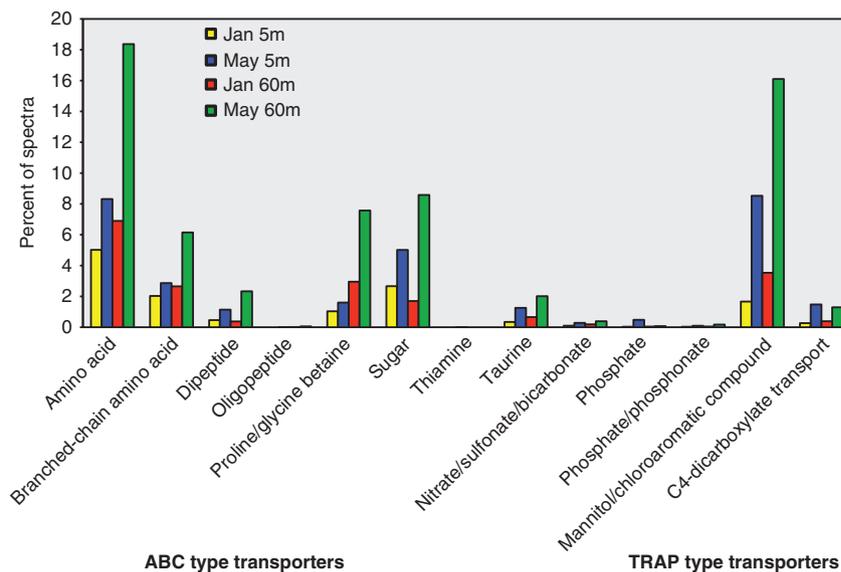


Figure 5 The relative abundance of solute transport proteins identified in Bedford Basin metaproteomes based on spectra counts.

Nine of the transport proteins were identified specifically as TonB-dependent receptors, whereas the remainders were ABC-type transporters for iron, nitrate/sulfonate/bicarbonate, phosphate and phosphonate. In contrast to organic substrates, we found little evidence for uptake of inorganic nutrients by SAR11. In fact, only a single SAR11 transporter

(phosphate) was identified, whereas the remaining were assigned to Rhodobacterales, as well as Beta-, Gamma-proteobacteria and SAR324. This suggests that in spring coastal waters, SAR11 invests more resources in the acquisition of organic compounds rather than inorganic nutrients. This is in contrast to SAR11 expression patterns in the oligotrophic

Sargasso Sea, where phosphate transporters were the most abundant detected proteins (Sowell *et al.*, 2009).

Metabolism of the ARCTIC96BD-19 clade

On the basis of 16S rRNA analysis, the GSO clade was relatively abundant in the spring deep layer (Figure 2). GSO comprises two closely related subgroups: SUP05 and ARCTIC96BD-19 (Walsh *et al.*, 2009). Although ARCTIC96BD-19 is widespread in the surface (Marshall and Morris, 2013) and deep ocean (Swan *et al.*, 2011), SUP05 is believed to be restricted to hydrothermal vents (Sunamura *et al.*, 2004; Anantharaman *et al.*, 2013) and oxygen-deficient pelagic habitats (Stevens and Ulloa, 2008; Lavik *et al.*, 2009; Zaikova *et al.*, 2010; Glaubitz *et al.*, 2013). Unfortunately, the V5 region of the 16S rRNA gene cannot distinguish between the SUP05 and ARCTIC96BD-19 lineages. However, we detected numerous transporter proteins that appeared to be derived from single-cell amplified genomes of the ARCTIC96BD-19 clade (Swan *et al.*, 2011), yet we did not identify any protein in the spring deep layer derived from SUP05 genomes. Specific ARCTIC96BD-19 transporters included those for amino acids, dipeptides, sugars and taurine (Supplementary Table S4). On the basis of spectral counts, transporters for amino acids were the most abundant. Previous genome analysis of the ARCTIC96BD-19 clade indicated the capacity for sulfur oxidation and CO₂ fixation (Swan *et al.*, 2011). We did not detect any ARCTIC96BD-19 proteins involved in sulfur oxidation or CO₂ fixation, which suggests that when this lineage is inhabiting productive coastal waters it tends toward a heterotrophic metabolism.

Metabolism of the SUP05 clade

We identified peptides that matched to 92 proteins from the SUP05 clade, almost all (90) of which

were only detected in the hypoxic bottom layer. Metagenomic analysis of SUP05 indicates the potential for a chemolithoautotrophic metabolism using sulfur compounds coupled to either aerobic respiration (Anantharaman *et al.*, 2013) or denitrification (Walsh *et al.*, 2009). Here, we identified SUP05 proteins for the sulfur oxidation pathway (SoxAB), the reverse dissimilatory sulfite reductase (DsrABCEK) and adenosine phosphosulfate reductase (Table 1). In total, we identified 21 peptides that matched to 8 SUP05 sulfur oxidation proteins and all were also unique, matching to no other proteins in the database. We further analyzed the peptide sequences to validate that they were indeed specific for SUP05 proteins (Supplementary Table S5). Peptide sequence from SUP05-derived Dsr proteins were used as query in similarity searches against the NCBI nr-protein database. The only sequences that were 100% identical to the peptides were from the SUP05-related sulfur-oxidizing symbionts of deep-sea clams and mussels (Kuwahara *et al.*, 2007; Newton *et al.*, 2007) and Dsr sequences recovered from an oxygen minimum zone (Lavik *et al.*, 2009). Five of the six peptides from SUP05-related SoxA and SoxB were also specific to SUP05 proteins. This single nonspecific peptide was also 100% identical to SoxB proteins from other marine Gamma-proteobacteria, which explains its detection in spring surface waters. The identification of this suite of proteins suggests that SUP05 is using reduced sulfur compounds such as thiosulfate or elemental sulfur (S⁰) as an energy source in the hypoxic deep layer. In sulfidic waters, SUP05 can potentially use sulfide as an energy source using either sulfide:quinone oxidoreductase (Sqr) or flavocytochrome *c*/sulfide dehydrogenase (Fcc) (Walsh *et al.*, 2009). However, neither complex was detected here, consistent with the expected absence of a sulfide source in this hypoxic water column. We did not detect SUP05 proteins for denitrification, aerobic respiration or carbon fixation and further study is required to identify the

Table 1 SUP05 proteins inferred to function in sulfur oxidation pathways

SUP05 homolog	No. of unique peptides	No. of peptides	No. of spectra				
			January 05 m	January 60 m	May 05 m	May 60 m	
<i>Dissimilatory sulfite reductase</i>							
DsrA	Sup05_0722/0723	5	5	0	33	0	0
DsrB	Sup05_0724	3	3	0	18	0	0
DsrC	Sup05_0728	2	2	0	9	0	0
DsrE	Sup05_0725	2	2	0	17	0	0
DsrK	Sup05_0730	1	1	0	3	0	0
<i>Sox systems</i>							
SoxA	Sup05_0516	2	2	7	31	0	0
SoxB	Sup05_0028	4	4	0	8	0	2
APS reductase	Sup05_0147	1	1	0	6	0	0
Total		21	21	7	127	0	2

Abbreviation: APS, adenosine phosphosulfate.

Table 2 Periplasmic-binding proteins of high-affinity transporters inferred to be from the SAR324 clade

	SAR324 homolog	No. of unique peptides	No. of peptides	No. of spectra			
				January 05 m	January 60 m	May 05 m	May 60 m
<i>Organic solutes</i>							
Amino acids	AFIA_0198	3	4	1	15	3	13
Amino acids	AFIA_0212	3	3	0	2	0	4
Dipeptide	AFIB_0487	3	4	0	10	0	7
Dipeptide	AFIA_0556	7	8	2	14	3	34
Sugar	AFIA_1706	5	6	0	10	4	17
Mannitol/chloroaromatic	AFIB_0724	3	4	0	2	0	3
Spermidine/putrescine	AFIA_1712	2	2	0	2	1	4
<i>Inorganic solutes</i>							
Nitrate/sulfonate/bicarbonate	AFIB_0217	3	4	0	15	4	5
Phosphate/phosphonate	AFIB_1101	2	3	0	9	3	4
Iron (Fe ³⁺)	AFIA_1002	2	2	0	2	0	2

role these bacteria have in carbon, nitrogen and sulfur cycling in hypoxic environments of the North Atlantic.

Metabolism of the SAR324 clade

SAR324 is a deep-branching clade within the Delta-proteobacteria (Wright *et al.*, 1997) that inhabits a wide range of marine settings, including the surface (DeLorenzo *et al.*, 2012) and deep ocean (Sheik *et al.*, 2013) and oxygen minimum zones (Wright *et al.*, 2012). Despite its ubiquity, little is known about SAR324 metabolism and physiology. Single-cell genome sequencing of SAR324 from the mesopelagic zone revealed capabilities for sulfur oxidation, autotrophic CO₂ fixation, and C1 metabolism (CO and methane), as well as multiple ABC transporters for sugars, amino acids and oligopeptides (Swan *et al.*, 2011). Hence SAR324, at least in the deep sea, exhibits a mixotrophic metabolic potential. By including these single-cell genomes in our analysis, we were able to gain insight into the metabolism of SAR324 in the coastal North Atlantic. Peptide spectra matching to SAR324 proteins were identified in all samples and strongly suggested a heterotrophic lifestyle (Table 2, Supplementary Table S6). A diverse array of PBPs for sugars, amino acids, peptides and polyamines were present in the metaproteomes. SAR324 proteins catalyzing the oxidation of methane were not identified, however, we did detect subunits for carbon monoxide oxidation (Supplementary Table S7). PBPs were specific for inorganic nutrients including nitrate/sulfonate/bicarbonate, phosphate/phosphonate and iron (Table 2, Supplementary Table S6). Transcriptome analysis of SAR324 populations associated with a hydrothermal vent plume also identified transporters for carbon and amino-acid uptake (Sheik *et al.*, 2013), suggesting that although this ubiquitous group has the genetic potential to grow autotrophically, it uses heterotrophic growth in a wide range of marine habitats.

One carbon metabolism

Interest in bacteria that grow on one carbon (C1) compounds such as methanol has intensified recently with the discovery that methanol concentrations in the ocean can be high (~400 nM) and turnover time can be as little as 1 day (Dixon *et al.*, 2011). Numerous proteins related to C1 metabolism were identified here (Table 3). Consistent with other coastal studies (Sowell *et al.*, 2011; Williams *et al.*, 2012), we detected the expression of methanol dehydrogenase large subunit proteins derived from the Beta-proteobacteria OM43 clade (Table 3). Methanol dehydrogenase was detected in all samples and exhibited highest similarity to homologs from two methylotrophic strains (HTCC2181 and KB13; Supplementary Table S8; Giovannoni *et al.*, 2008). On the basis of previous genome analysis, OM43 bacteria use the RuMP pathway for the assimilation of carbon from C1 compounds (Giovannoni *et al.*, 2008). Here, we also identified the expression of a key enzyme of the RuMP pathway, namely hexulose-6-phosphate synthase, in all samples (Table 3, Supplementary Table S8), suggesting OM43 may be assimilating carbon derived from methanol. However, interpretation of the functional relevance of this protein must be treated cautiously, as OM43 may also use this enzyme in the dissimilarity cyclic-oxidation-of-formaldehyde pathway (Giovannoni *et al.*, 2008). Little is known about the sources of methanol in the water column. Methanol may be deposited from the atmosphere and produced by phytoplankton (Milne *et al.*, 1995; Heikes *et al.*, 2002) and bacterial transformation of algal carbohydrates (Sieburth and Keller, 1988). OM43 have been associated with phytoplankton blooms (Morris *et al.*, 2006), however, our identification of OM43 methanol dehydrogenase and hexulose-6-phosphate synthase proteins in all samples suggest methylotrophic populations are sustained in the Northwest Atlantic even at times of low phytoplankton abundance.

Table 3 Proteins inferred to function in the metabolism of C1 compounds

	No. of proteins	No. of spectra			
		January 5 m	January 60 m	May 05 m	May 60 m
Methanol dehydrogenase	2	59	31	395	147
Hexulose 6-phosphate synthase	2	6	7	20	7
5,10-Methylene-tetrahydrofolate dehydrogenase	3	1	1	0	2
5,10-Methylenetetrahydrofolate reductase	1	0	0	0	1
Formate tetrahydrofolate ligase	4	0	0	38	3
Formate dehydrogenase	10	34	54	74	170
Serine hydroxymethyltransferase	10	3	23	13	2
Dimethylglycine dehydrogenase	2	0	1	2	7
Glycine cleavage system protein P	4	2	0	3	17
Glycine cleavage system T protein	12	4	2	31	58

Abbreviation: C1, one carbon.

In aquatic environments, significant CO production occurs from photochemical degradation of dissolved organic matter (Valentine and Zepp, 1993) and bacteria have been shown to consume 88% of photochemical CO production (Tolli and Taylor, 2005). In oxic environments, CO oxidation is carried out by an aerobic-type carbon monoxide dehydrogenase (CODH), which produces CO₂ and reducing equivalents (King and Weber, 2007). We identified a wide range of CODH subunits (CoxLMS) in all four metaproteomes. The CODHs were most similar to homologs from Roseobacter and SAR116 clades of Alpha-proteobacteria, Planctomycetes and Delta-proteobacteria. (Supplementary Table S7). It is unclear how the diversity of CODHs translates to CO oxidation activity in the coastal Atlantic as the specificity of CODH complexes for CO has only been characterized in a few species. Moreover, it appears that only species that contain two distinct forms of CoxL are able to oxidize CO (Cunliffe, 2010). In any case, the abundance and diversity of putative CODH proteins warrants further investigation into their functional and ecological importance in the ocean.

Archaeal nitrification and carbon fixation

Marine Group I (MGI) Thaumarchaeota are ubiquitous in the ocean and have an important role in nitrification (Stahl and de la Torre, 2012). We identified an abundance of MGI proteins particularly in the hypoxic deep layer (25% of all spectra; Figure 3). Archaeal nitrification during the winter was indicated by the identification of an archaeal ammonia monooxygenase subunit (AmoB) with peptides best matching *Nitrosopumilus maritimus* SCM1 (Walker *et al.*, 2010; Supplementary Table S9). An absence of bacterial Amo or other nitrification genes supports the role of MGI as the dominant nitrifiers in the hypoxic deep water of this coastal basin. Ammonia-oxidizing archaea (AOA) are believed to outcompete ammonia-oxidizing bacteria at low ammonia concentration because of an extremely high affinity for ammonia (Martens-Habbena *et al.*, 2009). Genes encoding ammonium

transporters were the most highly expressed transcripts in metatranscriptomic studies of MGI associated with an oxygen minimum zone (Stewart *et al.*, 2012) and a deep-sea hydrothermal vent plume (Baker *et al.*, 2012), which supports the hypothesis that the high capacity for ammonium acquisition is a result of overexpression of ammonium transporters. However, we were unable to detect any archaeal ammonium transporter proteins in our metaproteomes, suggesting that high mRNA expression levels may not directly relate to high protein expression (Vogel and Marcotte, 2012). On the other hand, this observation may be because of subcellular localization of proteins and a selection against membrane proteins by our protein extraction methods (Morris *et al.*, 2010).

Ammonia oxidation by AOA and ammonia-oxidizing bacteria produces nitrite, which is subsequently oxidized to nitrate by nitrite-oxidizing bacteria to complete the process of nitrification. Another possibility is nitrite could serve as a substrate for annamox, as has been shown in the chemocline of the Black Sea (Lam *et al.*, 2007) and oxygen minimum zones (Lam *et al.*, 2009). We could not identify any proteins that may have a role in nitrite oxidation. One explanation could be an absence of closely related nitrite-oxidizing or annamox bacterial reference genomes in our protein database.

Fundamental questions remain regarding the physiology of MGI, and whether individual MGI strains are capable of both heterotrophic and autotrophic metabolism. Genome analysis of MGI suggests a potential for mixotrophy, based on the presence of transporters for organic compounds such as amino acids, peptides, taurine and glycerol (Martens-Habbena *et al.*, 2009). Moreover, a recent study has shown that the expression of ammonia monooxygenase does not necessarily indicate autotrophic CO₂ fixation (Mussmann *et al.*, 2011). Our data support that MGI nitrification is linked to autotrophic CO₂ fixation in Bedford Basin. We identified proteins catalyzing three steps of the archaeal 3-hydroxypropionate/4-hydroxybutyrate pathway of CO₂ fixation (Berg *et al.*, 2010), including

acetyl-CoA carboxylase, methylmalonyl-CoA mutase and 4-hydroxybutyryl CoA dehydratase (Supplementary Table S9). No transporters for organic compounds were detected, suggesting a reliance on autotrophic carbon assimilation. We did, however, detect the PBP and permease components of a high-affinity ABC-type phosphate/phosphonate transport system. Representatives of CO₂ fixation proteins, and an absence of organic substrate transporters, in the metaproteome supports the idea that MGI are autotrophic in Bedford Basin.

Ammonia assimilation

Proteins involved in the assimilation of ammonia were common and at highest abundance in the ammonia-depleted winter deep layer (1% of all spectra). Identified proteins included glutamine synthetase (GS), glutamate synthase, glutamate dehydrogenase and nitrogen regulatory protein PII (Supplementary Tables S2 and S10). GS homologs were most similar to a diverse set of bacteria, including Roseobacter, Bacteroidetes and SAR324 clade members in the spring and GSO, unclassified Gamma-proteobacteria, and MGI in the winter. On the basis of a winter metaproteome from Antarctic coastal waters, Williams *et al.* (2012) proposed that bacteria and archaea utilize distinctive ammonia assimilation pathways based on the presence of AOA glutamate dehydrogenase proteins and an absence of AOA GS proteins. Although we detected AOA glutamate dehydrogenase in our metaproteomes, the detection of archaeal GS suggests that under conditions of low ammonia availability the AOA are also able to make use of the high-affinity GS pathway.

Conclusions

Bedford Basin is a temperate inlet that experiences the seasonal cycle and spring phytoplankton blooms that characterize the Northwest Atlantic Ocean. The deep water also exhibits temporal structure, most notably the formation of hypoxic conditions in the fall-winter. Our metaproteomic analysis has provided insight into the metabolic succession that occurs from winter to spring in Bedford Basin, and demonstrated clear vertical and temporal partitioning of transport and metabolic functions in the water column. During the spring bloom, ABC and TRAP transporters from a diversity of bacterial lineages (including SAR324 and ARCTIC96BD-19) have a role in scavenging of organic nutrients. The OM43 methanol utilization pathway was ubiquitous, suggesting OM43 may be only weakly linked to seasonality, compared with typical heterotrophic bacteria, and therefore its abundance may serve as an indicator for longer-term environmental changes that may be occurring in the coastal ocean.

Although bacterial community composition in surface and deep water was indistinguishable in the winter, the metaproteomes revealed distinct metabolic pathways operating in the deep hypoxic water. In addition to the detection of AOA in winter, we also identified sulfur oxidation proteins from the SUP05 clade, which potentially expands the ecological niche of the SUP05 clade from strictly oxygen-depleted and hydrothermal environments to include moderately hypoxic basins. Further comparisons of 'meta'omics' data sets originating from disparate SUP05 habitats will significantly contribute to our understanding of how this metabolically versatile marine lineage adapts to local environmental conditions and influences marine biogeochemistry.

As a concluding remark, we would like to address the current limitations of metaproteomic studies such as this one. First, ecological conclusions based on the absence of proteins are difficult to make because of technical issues such as cellular localization and protein fractionation by different extraction protocols. Similarly, many peptide spectra were simply not assigned to proteins because the proteins were not present in the GOS and reference genome database used here. Second, our reference genome database did not contain any genomes generated directly from the Bedford Basin community we analyzed. Although we validated the specificity of peptides that appear to have originated from the SUP05, ARCTIC96BD-19 and SAR324 clades, it is possible that some may have originated from different phylogenetic lineages, particularly if the genes encoding these proteins are prone to lateral gene transfer between marine microbes. One promising solution to these two limitations is to collect and analyze metagenomic/transcriptomic and metaproteomic data sets from the same samples, as exemplified by recent studies of Antarctic coastal waters (Grzyski *et al.*, 2012; Williams *et al.*, 2012). Another solution is the increasing number of single-cell amplified genomes that are becoming available for a wide variety of poorly characterized bacteria and archaea (Rinke *et al.*, 2013), which will continually provide more comprehensive databases for peptide matching. Finally, the incorporation of *de novo* peptide prediction tools into metaproteomic workflows should also increase the resolution of metaproteomic data sets by identifying novel peptide sequences that may not be currently represented in protein databases.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

We acknowledge Jeff Anning for his logistical support in collecting samples from Bedford Basin. We also thank Etienne Yergeau and Sylvie Sanschagrin from the NRC-BRI for providing access to their DNA

sequencing facility, as well as Jean-Pierre Falgout and Marcos DiFalco from Concordia University for their help with generating the metaproteomics data sets. This work was supported by NSERC (402214-2011) and CRC (950-221184) research grants and the Bedford Basin Monitoring Program of Fisheries and Oceans Canada.

References

- Anantharaman K, Breier JA, Sheik S, Dick GJ. (2013). Evidence for hydrogen oxidation and metabolic plasticity in widespread deep-sea sulfur-oxidizing bacteria. *Proc Natl Acad Sci USA* **110**: 330–335.
- Baker BJ, Lesniewski RA, Dick GJ. (2012). Genome-enabled transcriptomics reveals archaeal populations that drive nitrification in a deep-sea hydrothermal plume. *ISME J* **6**: 2269–2279.
- Baker GC, Smith JJ, Cowan DA. (2003). Review and re-analysis of domain-specific 16S primers. *J Microbiol Methods* **55**: 541–555.
- Béjà O, Aravind L, Koonin EV, Suzuki MT, Hadd A, Nguyen LP *et al.* (2000). Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**: 1902–1906.
- Béjà O, Suzuki MT, Heidelberg JF, Nelson WC, Preston CM, Hamada T *et al.* (2002). Unsuspected diversity among marine aerobic anoxygenic phototrophs. *Nature* **415**: 630–633.
- Berg IA, Kockelkorn D, Ramos-Vera WH, Say RF, Zarzycki J, Hugler M *et al.* (2010). Autotrophic carbon fixation in archaea. *Nat Rev Microbiol* **8**: 447–460.
- Cunliffe M. (2010). Correlating carbon monoxide oxidation with *cox* genes in the abundant marine Roseobacter clade. *ISME J* **5**: 685–691.
- DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard NU *et al.* (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**: 496–503.
- DeLorenzo S, Brauer SL, Edgmont CA, Herfort L, Tebo BM, Zuber P. (2012). Ubiquitous dissolved inorganic carbon assimilation by marine bacteria in the Pacific Northwest coastal ocean as determined by stable isotope probing. *PLoS One* **7**: 46695.
- Dixon JL, Beale R, Nightingale PD. (2011). Microbial methanol uptake in northeast Atlantic waters. *ISME J* **5**: 704–716.
- Eng JK, McCormack AL, Yates JR III. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **5**: 976–989.
- Giovannoni SJ, Hayakawa DH, Tripp HJ, Stingl U, Givan SA, Cho JC *et al.* (2008). The small genome of an abundant coastal ocean methylotroph. *Environ Microbiol* **10**: 1771–1782.
- Glaubitz S, Kiesslich K, Meeske C, Labrenz M, Jurgens K. (2013). SUP05 dominates the Gammaproteobacterial sulfur oxidizer assemblages in pelagic Redoxclines of the central baltic and black seas. *Appl Environ Microbiol* **79**: 2767–2776.
- Grzymalski JJ, Riesenfeld CS, Williams TJ, Dussaq AM, Ducklow H, Erickson M *et al.* (2012). A metagenomic assessment of winter and summer bacterioplankton from Antarctica Peninsula coastal surface waters. *ISME J* **6**: 1901–1915.
- Hedges JI, Keil RG, Benner R. (1997). What happens to terrestrial organic matter in the ocean? *Org Geochem* **27**: 195–212.
- Heikes BG, Chang W, Pilson MEQ, Swift E, Singh HB, Guenther A *et al.* (2002). Atmospheric methanol budget and ocean implication. *Global Biogeochem Cy* **16**: 1–13.
- Huson D, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC. (2011). Integrative analysis of environmental sequences using MEGAN4. *Genome Res* **9**: 1552–1560.
- Kall L, Canterbury JD, Weston J, Noble WS, MacCoss MJ. (2007). Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods* **4**: 923–925.
- King GM, Weber CF. (2007). Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. *Nat Rev Microbiol* **5**: 107–118.
- Kujawinski EB. (2011). The impact of microbial metabolism on marine dissolved organic matter. *Ann Rev Mar Sci* **3**: 567–599.
- Kuwahara H, Yoshida T, Takaki Y, Shimamura S, Nishi S, Harada M *et al.* (2007). Reduced genome of the thioautotrophic intracellular symbiont in a deep-sea clam, *Calyptogena okutanii*. *Curr Biol* **17**: 881–886.
- Lam P, Jensen MM, Lavik G, McGinnis DF, Müller B, Schubert CJ *et al.* (2007). Linking crenarchaeal and bacterial nitrification to anammox in the Black Sea. *Proc Natl Acad Sci USA* **104**: 7104–7109.
- Lam P, Lavik G, Jensen M, Van de Vossenberg J, Schmid M, Woebken D *et al.* (2009). Revising the nitrogen cycle in the Peruvian oxygen minimum zone. *Proc Natl Acad Sci* **106**: 4752.
- Lavik G, Stuhmann T, Bruchert V, Van der Plas A, Mohrholz V, Lam P *et al.* (2009). Detoxification of sulphidic African shelf waters by blooming chemolithotrophs. *Nature* **457**: 581–586.
- Li WKW, Dickie P. (2001). Monitoring phytoplankton, bacterioplankton, and virioplankton in a coastal inlet (Bedford Basin) by flow cytometry. *Cytometry* **44**: 236–246.
- Li WKW, Harrison WG, Head EJ. (2006). Coherent assembly of phytoplankton communities in diverse temperate ocean ecosystems. *Proc Biol Sci* **273**: 1953–1960.
- Li WKW, Harrison WG. (2008). Propagation of an atmospheric climate signal to phytoplankton in a small marine basin. *Limnol Oceanogr* **53**: 1734–1745.
- Marshall KT, Morris RM. (2013). Isolation of an aerobic sulfur oxidizer from the SUP05/Arctic96BD-19 clade. *ISME J* **7**: 452–455.
- Martens-Habbena W, Berube P, Urakawa H, de La Torre J, Stahl D. (2009). Ammonia oxidation kinetics determine niche separation of nitrifying Archaea and Bacteria. *Nature* **461**: 976–979.
- Milne PJ, Riemer DD, Zika RG, Brand LE. (1995). Measurement of vertical distribution of isoprene in surface seawater, its chemical fate, and its emission from several phytoplankton monocultures. *Mar Chem* **48**: 237–244.
- Morris RM, Longnecker K, Giovannoni SJ. (2006). *Pirellula* and *OM43* are among the dominant lineages identified in an Oregon coast diatom bloom. *Environ Microbiol* **8**: 1361–1370.

- Morris RM, Nunn BL, Frazar C, Goodlett DR, Ting YS, Rocap G. (2010). Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. *ISME J* **4**: 673–685.
- Mussmann M, Brito I, Pitcher A, Sinninghe Damste JS, Hatzenpichler R, Richter A *et al.* (2011). Thaumarchaeotes abundant in refinery nitrifying sludges express amoA but are not obligate autotrophic ammonia oxidizers. *Proc Natl Acad Sci USA* **108**: 16771–16776.
- Newton ILG, Woyke T, Auchtung TA, Dilly GF, Dutton RJ, Fisher MC *et al.* (2007). The *Calyptogena magnifica* chemoautotrophic symbiont genome. *Science* **315**: 998–1000.
- Newton RJ, Griffin LE, Bowles KM, Meile C, Gifford S, Givens CE *et al.* (2010). Genome characteristics of a generalist marine bacterial lineage. *ISME J* **4**: 784–798.
- Pomeroy RL, Williams PJB, Azam F, Hobbie JE. (2007). The microbial loop. *Oceanography* **20**: 28–33.
- Poretsky R, Sun S, Mou X, Moran M. (2010). Transporter genes expressed by coastal bacterioplankton in response to dissolved organic carbon. *Environ Microbiol* **12**: 616–627.
- Punshon S, Moore R. (2004). Nitrous oxide production and consumption in a eutrophic coastal embayment. *Mar Chem* **91**: 37–51.
- Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson JJ, Cheng JF *et al.* (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**: 431–437.
- Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S *et al.* (2007). The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. *Plos Biol* **5**: 398–431.
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M *et al.* (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537–7541.
- Sheik CS, Jain S, Dick GJ. (2013). Metabolic flexibility of enigmatic SAR324 revealed through metagenomics and metatranscriptomics. *Environ Microbiol*; e-pub ahead of print 3 June 2013; doi:10.1111/1462-2920.12165.
- Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M. (2006). In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* **1**: 2856–2860.
- Sieburth JM, Keller MD. (1988). Methylaminotrophic bacteria in xenic nanoalgal cultures: incidence, significance, and role of methylated algal osmoprotectants. *Biol Oceanogr* **6**: 383–395.
- Sowell SM, Abraham PE, Shah M, Verberkmoes NC, Smith DP, Barofsky DF *et al.* (2011). Environmental proteomics of microbial plankton in a highly productive coastal upwelling system. *ISME J* **5**: 856–865.
- Sowell SM, Wilhelm LJ, Norbeck AD, Lipton MS, Nicora CD, Barofsky DF *et al.* (2009). Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. *ISME J* **3**: 93–105.
- Stahl DA, de la Torre JR. (2012). Physiology and diversity of ammonia-oxidizing archaea. *Annu Rev Microbiol* **66**: 83–101.
- Stevens H, Ulloa O. (2008). Bacterial diversity in the oxygen minimum zone of the eastern tropical South Pacific. *Environ Microbiol* **10**: 1244–1259.
- Stewart FJ, Ulloa O, DeLong EF. (2012). Microbial metatranscriptomics in a permanent marine oxygen minimum zone. *Environ Microbiol* **14**: 23–40.
- Sunamura M, Higashi Y, Miyako C, Ishibashi J, Maruyama A. (2004). Two bacteria phylotypes are predominant in the Suiyo seamount hydrothermal plume. *Appl Environ Microbiol* **70**: 1190–1198.
- Swan BK, Martinez-Garcia M, Preston CM, Sczyrba A, Woyke T, Lamy D *et al.* (2011). Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science* **333**: 1296–1300.
- Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennis CM *et al.* (2012). Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science* **336**: 608–611.
- Thompson MR, Chourey K, Froelich JM, Erickson BK, Verberkmoes NC, Hettich RL. (2008). Experimental approach for deep proteome measurements from small-scale microbial biomass samples. *Anal Chem* **80**: 9517–9525.
- Tolli JD, Taylor CD. (2005). Biological CO oxidation in the Sargasso Sea and in Vineyard Sound, Massachusetts. *Limnol Oceanogr* **50**: 1205–1212.
- Valentine RL, Zepp RG. (1993). Formation of carbon monoxide from the photodegradation of terrestrial dissolved organic carbon in natural waters. *Environ Sci Tech* **27**: 409–412.
- Verberkmoes N, Denev V, Hettich R, Banfield J. (2009). Systems biology: functional analysis of natural microbial consortia using community proteomics. *Nat Rev Microbiol* **7**: 196–205.
- Vogel C, Marcotte EM. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* **13**: 227–232.
- Walker CB, De la Torre JR, Klotz MG, Urakawa H, Pinel N, Arp DJ *et al.* (2010). Nitrosopumilus maritimus genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proc of the Natl Acad Sci* **107**: 8818.
- Walsh D, Zaikova E, Howes C, Song Y, Wright J, Tringe S *et al.* (2009). Metagenome of a versatile Chemolithoautotroph from expanding oceanic dead zones. *Science* **326**: 578–582.
- Williams TJ, Long E, Evans F, Demaree MZ, Lauro FM, Raftery MJ *et al.* (2012). A metaproteomic assessment of winter and summer bacterioplankton from Antarctic Peninsula coastal surface waters. *ISME J* **6**: 1883–1900.
- Woyke T, Xie G, Copeland A, González J, Han C, Kiss H *et al.* (2009). Assembling the marine metagenome, one cell at a time. *PLoS One* **4**: e5299.
- Wright JJ, Konwar KM, Hallam SJ. (2012). Microbial ecology of expanding oxygen minimum zones. *Nat Rev Microbiol* **10**: 381–394.
- Wright TD, Vergin KL, Boyd PW, Giovannoni SJ. (1997). A novel delta-subdivision proteobacterial lineage from the lower ocean surface layer. *Appl Environ Microbiol* **63**: 1441–1448.
- Yergeau E, Lawrence JR, Sanschagrin S, Waiser MJ, Korber DR, Greer CW. (2012). Next-generation sequencing of microbial communities in the Athabasca River and its tributaries in relation to oil sands mining activities. *Appl Environ Microbiol* **78**: 7626–7637.

Yooseph S, Sutton G, Rusch DB, Halpern AL, Williamson SJ, Remington K *et al.* (2007). The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. *PLoS Biol* **5**: e16.

Zaikova E, Walsh DA, Stilwell CP, Mohn WW, Tortell PD, Hallam SJ. (2010). Microbial community dynamics in a seasonally anoxic fjord: Saanich Inlet, British Columbia. *Environ Microbiol* **12**: 172–191.

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