Combining metagenomics and metaproteomics reveals metabolic pathways used by an uncultivated marine methanol utiliser

Carolina Grob[1], Martin Taubert[1], Alexandra M. Howat[1], Oliver J. Burns[2], Joanna L. Dixon[3], Hans H. Richnow[4], Nico Jehmlich[5], Martin von Bergen[5,6,7], Yin Chen[8], and J. Colin Murrell[1]

[1]School of Environmental Sciences and [2]School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK

[3]Plymouth Marine Laboratory, Prospect Place, West Hoe, Plymouth, Devon, PL1 3DH, UK

[4]Department of Isotope Biogeochemistry, [5]Department of Proteomics and [6]Department of Metabolomics, Helmholtz-Centre for Environmental Research - UFZ, Leipzig, Germany

[7]Department of Chemistry and Bioscience, University of Aalborg, Fredrik Bajers Vej 7H, 9220 Aalborg East, Denmark

[8]School of Life Sciences, University of Warwick, Coventry, CV4 7AL, UK

Running title: metagenomics, metaproteomics and methylotrophy

Corresponding author: J. Colin Murrell, j.c.murrell@uea.ac.uk

Summary

In order to better understand how the activity of microbial organisms influences and regulates all major biogeochemical cycles, a great deal of effort has been put into characterizing the physiology and metabolism of key representatives available in culture. Such effort is often limited by the difficuties associated with isolating novel microorganisms from the environment and cultivating them in the laboratory. To overcome this problem, a variety of culture-independent techniques have been developed that can be used in conjunction with the above to investigate natural microbial populations. In this study, we combined DNA-stable isotope probing with metagenomics and metaproteomics to characterize an as yet uncultivated marine methylotroph that actively incorporated carbon from ^{13}C-labeled methanol into biomass. By metagenomic sequencing of the heavy DNA, we retrieved virtually the whole genome of this bacterium and identified through protein-stable isotope probing the metabolic pathways used to assimilate methanol. This proof-of-concept study is the first in which both DNA- and protein-stable isotope probing has been used to characterize the metabolism of an uncultivated bacterium from the marine environment and thus provides a powerful approach to access the genome and proteome of uncultivated microbes involved in key processes in the environment.

Introduction

One of the main challenges in microbial ecology is to directly access the genomes and understand the metabolism of key microbes involved in biogeochemical cycling. An ideal scenario is to isolate model organisms and then characterize them using conventional physiological and biochemical techniques. However, since many microbes are difficult to cultivate in the laboratory, focussed cultivation-independent techniques are also required. To address these challenges, DNA-Stable Isotope Probing (DNA-SIP), involving the use of ^{13}C-labeled substrates that are incorporated into the biomass of active microbes, has been developed (Radajewski *et al.*, 2000, Dumont and Murrell, 2005, Neufeld *et al.*, 2007b). This technique enables the separation of heavy (^{13}C-labeled) from light (unlabeled) DNA, thus allowing the isolation of ^{13}C-DNA from microbes that have assimilated the target ^{13}C-labeled substrate from those that have not. The identity of active cells can then be determined by interrogating the heavy DNA via PCR using 16S rRNA or genes encoding key steps in microbial metabolism.

Further information regarding the metabolic potential and the metabolic pathways actually being used by target microorganisms during SIP incubations can be obtained by complementary metagenome analysis of the heavy DNA combined with metaproteome analysis, i.e., protein-SIP. This approach yields quantitative data on the incorporation of heavy isotopes of carbon, nitrogen or sulfur into peptides (Seifert *et al.*, 2012) and their sequence analysis results on information on both the phylogeny and physiology of microbial organisms (von Bergen *et al.*, 2013).

In this study, we provide proof-of-concept experiments to show that DNA- and protein-SIP can be combined with metagenomics to characterize the metabolism of an as yet uncultivated marine bacterium. We chose to use methanol and methylotrophic bacteria in the marine environment to develop these techniques because in our previous studies we showed that

marine methylotrophs of the genus *Methylophaga* were present and active in coastal seawater environments (station L4 of the Western Channel Observatory, Plymouth, UK) and that we could use DNA-SIP to recover genes involved in methanol oxidation from this uncultivated *Methylophaga* species (Neufeld *et al.*, 2007a, Neufeld *et al.*, 2008b, Neufeld *et al.*, 2008a). We previously showed that methanol is metabolized *in situ* at station L4 (Dixon *et al.*, 2011, Sargeant, 2013) and therefore in this study we used methylotrophy as a model system with which to combine for the first time DNA- and protein-SIP approaches to access the metabolism of a marine *Methylophaga* which we have failed to isolate and cultivate in the laboratory.

A key enzyme involved in methanol metabolism by methylotrophs, methanol dehydrogenase (MDH), catalyses the conversion of methanol to formaldehyde (Anthony, 1982, Chistoserdova, 2011). The gene coding for the large subunit of the classical MDH, *mxaF*, has been well characterized (Anthony *et al.*, 1994). A homologue of *mxaF* gene, *xoxF*, which can also be involved in methanol metabolism, is present in all known methylotrophs and several non-methylotrophic organisms (Chistoserdova and Lidstrom, 1997, Chistoserdova *et al.*, 2009, Chistoserdova, 2011). Multiple *xoxF* genes, sometimes belonging to more than one of the five distinct *xoxF* clades that have been described (Chistoserdova, 2011, Keltjens *et al.*, 2014), can often be found in a single methylotroph genome, making it difficult to unequivocally assign a functional role to this gene (Chistoserdova, 2011). Based on sequencing data, specific PCR primer sets have been designed to target *mxaF* (McDonald and Murrell, 1997, Neufeld et al., 2007a) and *xoxF* genes (Taubert *et al., in revision*) and thus determine the distribution and diversity of methylotrophic bacteria in the environment. The presence of these functional biomarkers alone does not however imply that they are metabolically active.

In this study, a combination of DNA-SIP and protein-SIP, metagenomics and metaproteomics, 16S rRNA gene, *mxaF* and *xoxF* functional gene amplicon sequencing was used to identify the phylogenetic affiliation and methanol utilization pathways of a marine methylotroph. We estimate that more than 90% of the genome of a marine *Methylophaga* species from the English Channel was obtained and concomitant metaproteomics analysis revealed the pathways of carbon assimilation used by this uncultivated methylotroph.

Results and discussion

SIP incubations were carried out in duplicate over three days using surface seawater from station L4 in the English Channel, with ^{13}C-labeled and unlabeled (^{12}C) methanol (control) as substrate. The purpose of this study was to use SIP to access the genome and proteome of an uncultivated marine methylotroph, so we chose a substrate concentration that we could confidently measure to monitor methanol consumption throughout the experiment, i.e., 100 µM. Given that surface methanol concentrations at station L4 are in the 16-78 nM range (Beale *et al.*, 2015), it could be argued that this concentration is not environmentally relevant. However, we have previously shown that even 1 µM stimulates the activity of *Methylophaga* from station L4 (Neufeld et al., 2008a). We therefore believe that the methanol concentration used here is suitable for the purpose of the proof-of-concept study presented here.

Total DNA was extracted from seawater at the beginning and end of the experiment and used to determine bacterial diversity. To assess the metabolic potential of the bacterial community, protein was also extracted from SIP incubations after three days. After separating ^{13}C-DNA from ^{12}C-DNA, the former was used to determine the diversity of active methylotrophs in DNA-SIP incubations by isolation and analysis of 16S rRNA, *mxaF* and *xoxF* gene sequences (Supplementary Table S1, for the total number of sequences retrieved from each sample). ^{13}C-DNA was also amplified using multiple displacement amplification to generate sufficient

material for metagenome sequencing and analysis of DNA of the dominant methylotroph in duplicate SIP incubations.

Combining DNA- and Protein-SIP. Bacterial community composition at the beginning of the incubations (T0) was determined by analysis of 16S rRNA gene sequences from duplicate DNA samples. Contributions of different bacterial groups to the total 16S rRNA gene sequences retrieved from each sample (Supplementary Table S1) were virtually identical (not shown), so average values are given below. 16S rRNA gene sequence analysis showed that at T0, the bacterial community was mainly dominated by Alphaproteobacteria (e.g., *Candidatus* Pelagibacter constituted 21% of all 16S rRNA gene sequences retrieved), Betaproteobacteria (e.g., *Achromobacter*, 15%), and Flavobacteria (e.g., *Formosa*, 15%) (Figure 1A). These results agree with previous studies carried out at station L4 where Alphaproteobacteria 16S rRNA gene sequences, particularly those belonging to the SAR11 clade (Pelagibacteraceae) are predominant throughout most of the year followed by Flavobacteria, with Beta- and Gammaproteobacteria also being present (Gilbert *et al.*, 2009, Gilbert *et al.*, 2012, Sargeant, 2013). Eighty different bacterial genera that constituted less than 5% of the total number of sequences could also be identified at the beginning of the SIP incubations (combined under "Others" in Figure 1A), with potential methanol utilizers, such as *Methylophaga*, *Ruegeria* and *Roseovarius* representing <0.5% of the total 16S rRNA gene sequences analyzed.

After incubating for three days with 100 μM methanol, changes in community composition were assessed based on 16S rRNA gene sequences retrieved from unfractionated DNA obtained from duplicate ^{13}C and ^{12}C methanol incubations (Supplementary Table S1). Compared with T0, the bacterial community in all four experiments was significantly enriched in Gammaproteobacteria (75-87% of the sequences; Supplementary Figure S1). At the methanol concentration used in these SIP incubations, 84% of the 16S rRNA gene sequences retrieved from the unfractionated ^{13}C-labeled samples belonged to the genus

Methylophaga (Figure 1A), which only represented 0.01% of the sequences at T0 (included in "Others" in Figure 1A). Other bacteria present at the start of the SIP incubations, such as *Candidatus* Pelagibacter (5%) and *Owenweeksia* (3%), were also present in the unfractionated DNA (Figure 1A).

CsCl density gradient centrifugation was used to separate heavy (^{13}C-labeled) from light (^{12}C, unlabeled) DNA extracted from ^{13}C methanol incubations, following the protocol described in Neufeld et al. (Neufeld et al., 2007b). ^{13}C-DNA was subsequently used to determine the phylogenetic affiliation of active methylotrophs by targeting 16S rRNA gene, as well as *mxaF* and *xoxF* functional genes. Most of the 16S rRNA gene sequences present in ^{13}C-DNA samples belonged to *Methylophaga*, thus indicating rapid incorporation of ^{13}C from methanol into *Methylophaga* biomass (Figure 1B). *mxaF* and *xoxF* gene sequences obtained from ^{13}C-DNA 454 data (not shown) confirmed that the enriched group was most closely related to *Methylophaga thiooxydans* DMS010 (Schäfer, 2007, Boden *et al.*, 2010). The amplicon sequencing results presented above agree with previous DNA-SIP experiments carried out at station L4 using methanol and other C_1 substrates, such as mono- and dimethylamine, dimethylsulfide and methyl bromide, that showed that *Methylophaga* spp. present in the marine environment are capable of metabolizing these compounds (Neufeld et al., 2007a, Neufeld et al., 2008a, Neufeld et al., 2008b). In ^{12}C-DNA samples, only 16% of the total 16S rRNA gene sequences retrieved from duplicate incubations were affiliated to *Methylophaga* whereas *Candidatus* Pelagibacter dominated (60%), despite being present at only 0.1% in the ^{13}C-DNA fraction (Figure 1B). Although this might seem high, the proportion of *Candidatus* Pelagibacter sequences at the end of the incubations (unfractionated DNA in Figure 1A) was lower than at the beginning (T0 in Figure 1A).

To complement the DNA-SIP data, proteins were extracted from ^{13}C-labeled and unlabeled (controls) methanol SIP incubations after three days. Tryptic peptides were measured using a

high resolution Orbitrap mass spectrometer and further identified using the OpenMS pipeline (Kohlbacher *et al.*, 2007, Sturm *et al.*, 2008) via the OMSSA search engine (Geer *et al.*, 2004). Protein identity and taxonomic affiliation were determined using a customised NCBInr database (Supplementary Information). 79% of the peptides identified from these protein samples were assigned to *Methylophaga* species and most of the peptides affiliated with this group showed ^{13}C incorporation from methanol (Figure 1C; Supplementary Dataset S1), with an average relative isotope abundance of 88.8% +/- 2.8%. No unlabeled *Methylophaga* peptides were detected in ^{13}C-labeled incubations, showing that the majority of *Methylophaga* biomass (>99% based on the detection limit of the instrument, not shown) was produced after the addition of labeled methanol. This confirms that *Methylophaga* constituted only a minor fraction of the bacterial community at the beginning of the experiment, as observed with 16S rRNA gene sequences data. Finally, ~13% of all peptides found to have ^{13}C incorporation patterns related to crossfeeding were identified as SAR11 peptides (not shown). This suggests that SAR11 cells were still active during the three day incubation with 100 μM methanol, having incorporated ^{13}C-labeled carbon into their peptides through crossfeeding but not by the direct use of this substrate as a carbon source. This is consistent with previous reports showing that members of the SAR11 clade to which *Candidatus* Pelagibacter found in ^{12}C-DNA belongs can oxidize methanol to CO_2, but do not seem to use it as a carbon source (Sun *et al.*, 2011).

Metagenomics of ^{13}C-labeled DNA. In order to investigate the metabolic potential of the organisms that were actively incorporating methanol into their biomass, heavy DNA from the ^{13}C-labeled experiment from two biological replicates was amplified using multiple displacement amplification and the amplified ^{13}C-DNA was used for metagenome sequencing on the Illumina MiSeq DNA sequencing platform (Supplementary Information).

16S rRNA gene sequences retrieved from ^{13}C-DNA metagenome datasets again showed the dominance of *Methylophaga* species, with >50% of them (in both replicates) having been assigned to *M. thiooxydans* using the Ribosomal Database Project (RDP) database (Supplementary Figure S2). Metagenome data were assembled in BaseSpace using SPAdes Genome Assembler v3.0 and the assembled contigs were annotated in RAST, followed by manual correction. The assembly of the whole metagenome dataset yielded 8 large contigs clearly belonging to *Methylophaga* with a total length of 2.60 Mb, an average coverage of 98x and a GC content of 45.7% (Supplementary Dataset S2). Of the remaining reads, assembled into 5,557 much smaller contigs (4.28 Mb and 13x coverage in total), less than 1% belonged to *Methylophaga*. The *Methylophaga* L4 genome derived from the SIP metagenome dataset was most closely related to the genome of *Methylophaga thiooxydans* DMS010 (Schäfer, 2007, Boden et al., 2010), a strain originally isolated from an enrichment culture of the coccolithophore *Emiliana huxleyi* (Schäfer, 2007). First described as *Methylophaga* sp. strain DMS010, this species grows on dimethylsulfide (DMS) and a variety of other C_1 compounds (Schäfer, 2007). *Methylophaga* sp. strain DMS010 was renamed *M. thiooxydans* DMS010 after a new pathway of DMS metabolism (DMS oxidation to tetrathionate) was discovered in this organism (Boden et al., 2010).

Although *M. thiooxydans* DMS010 has a genome size of 3.05 Mb (Boden *et al.*, 2011), closer investigation of this genome uncovered a series of identical phage-like regions that when removed left a core genome of 2.59 Mb. Hence, the 8 contigs of the *M. thiooxydans* strain L4 genome assembled from the metagenome very likely cover > 90% of the genome of this SIP-enriched *Methylophaga* species (Supplementary Figure S3). This SIP-metagenome-derived genome of *M. thiooxydans* strain L4 is also comparable in size with the genomes of *Methylophaga* strains *M. nitratireducenticrescens* JAM1 (3.1 Mb) and *M. frappieri* JAM7 (2.7 Mb) (Villeneuve *et al.*, 2012, Villeneuve *et al.*, 2013). Finally, given the high level of

coverage of the genome, the use of DNA-SIP targeted metagenomic would seem to be a useful complementary molecular ecology technique to single-cell whole genome sequencing.

Analysis of the *M. thiooxydans* strain L4 genome assembled here revealed the presence of key genes involved in one-carbon metabolism (Supplementary Dataset S3). The entire gene cluster coding for the small (*mxaI*) and large (*mxaF*) subunits of the pyrroloquinoline quinone (PQQ)-dependent methanol dehydrogenase and accessory genes was found (Figure 2). The full length *mxaF* gene retrieved from the assembled genome was 98% and 99% percent identical to the *mxaF* gene sequence and the derived amino acid sequence of *M. thiooxydans* DMS010, respectively (Supplementary Figure S4A). Four copies of the *mxaF* homolog *xoxF* were also identified (Supplementary Figure S5). Their sequences were 93-97% identical to *M. thiooxydans* DMS010 (Supplementary Figure S4B).

Two gene clusters containing genes involved in the conversion of methylamine to formaldehyde via two different pathways were also identified. One gene cluster, *mauFBEDAGLMN*, contains genes encoding the large (*mauB*) and small (*mauA*) subunit of methylamine dehydrogenase, a TTQ-dependent dehydrogenase, as well as further accessory genes required for the activity of this enzyme (Supplementary Figure S6; Anthony, 1982). Methylamine dehydrogenase catalyses the direct oxidation of methylamine to formaldehyde and ammonium and is typically found in Proteobacteria that use methylamine as a carbon source (Wischer *et al.*, 2014). The second gene cluster encodes genes for gamma-glutamylmethylamide (GMA) synthetase (*gmaS*), N-methylglutamate (NMG) synthase (*mgsABC*) and NMG dehydrogenase (*mgdABCD*). In this pathway, methylamine is oxidised in a stepwise conversion via the methylated amino acids GMA and NMG to 5,10-methylenetetrahydrofolate (Latypova *et al.*, 2010, Chen *et al.*, 2010; Supplementary Figure S6). This pathway is found in methylotrophs as well as non-methylotrophs that use methylamine as nitrogen source (Wischer et al., 2014).

Further analysis of the *M. thiooxydans* strain L4 genome enabled reconstruction of key steps in carbon and nitrogen metabolism (Table 1). Carbon fixation from formaldehyde could occur via the ribulose monophosphate (RuMP; Johnson and Quayle, 1965) cycle (Entner-Doudoroff variant), as all the necessary genes are present in the genome. Genes encoding key enzymes of other potential carbon fixation pathways, such as the serine and Calvin-Benson-Bassham (CBB) cycles were missing, indicating the likely absence of these pathways in *M. thiooxydans* strain L4 (Table 1). Both Entner-Doudoroff and pentose phosphate pathways were complete, but no gene encoding 6-phosphofructokinase, a key enzyme of the glycolysis pathway, was found. Genes encoding the 2-oxoglutarate dehydrogenase complex of the tricarboxylic acid (TCA) cycle, *sucA* and *sucB*, were also missing (Table 1), while the *lpd* gene encoding the lipoamide dehydrogenase was only found in conjunction with the pyruvate dehydrogenase complex. All other genes of the TCA cycle were detected. These results are consistent with the notion that bacteria using the RuMP cycle as their major carbon assimilation pathway tend to have an incomplete TCA cycle (Anthony, 1982). The resulting incomplete TCA cycle could still channel carbon from pyruvate to all intermediates that may be required for central biosynthetic pathways, including oxoglutarate and succinyl-CoA, but cannot be used to generate energy by oxidising carbon compounds to CO_2. To produce reducing power through the oxidation of formaldehyde to CO_2, two potential pathways were present: the dissimilatory hexulose phosphate (HuP) cycle (Chistoserdova *et al.*, 2000) and the tetrahydromethanopterin (H_4MPT)-dependent oxidation pathway (Vorholt *et al.*, 2000). Regarding nitrogen metabolism, genes for assimilatory nitrate reduction to ammonium, including a assimilatory nitrate reductase and a NAD(P)H-dependent nitrite reductase, were present in the genome of *M. thiooxydans* strain L4, but no genes linked to dissimilatory nitrate reduction were found. The glutamine synthase/glutamine 2-oxoglutarate amidotransferase (GS/GOGAT) pathway for ammonia assimilation (Trotsenko and Murrell, 2008), as well as alanine dehydrogenase

and glutamate dehydrogenase genes, were also present. For sulfur metabolism, all genes required for assimilatory reduction of sulfate to sulfide were present in the genome, and furthermore all genes required for cysteine biosynthesis were found. No genes specific for dissimilatory sulfate reduction were present. Finally, the biosynthetic pathway for ectoine, a common bacterial osmolyte, was found, coinciding with the marine habitat of *M. thiooxydans* strain L4.

Using a similar targeted metagenomics approach, Kalyuzhnaya and collaborators (Kalyuzhnaya *et al.*, 2008) were able to retrieve the near complete genome of a novel methylotroph, *Methylotenera mobilis*, from Lake Washington sediment. By complementing this culture-independent approach with metaproteomic analyses, here we further determined which of the metabolic pathways identified in the reconstructed genome were being expressed by *Methylophaga thiooxydans* strain L4 during growth on methanol, thus gaining a deeper understanding of the metabolism of this organism.

Metaproteomics and metabolic reconstruction. An LC-MS/MS analysis of proteolytic peptides lysates from proteomes isolated after three day SIP incubations was performed and analyzed using the predicted proteins of the *M. thiooxydans* strain L4 genome as reference database. Of the 2,522 protein-encoding genes predicted in the eight contigs of the *M. thiooxydans* strain L4, 737 were identified, accounting for 29% of the proteome of this *Methylophaga* species (Supplementary Dataset S4). Based on these data, metaproteomic reconstruction of the central carbon metabolic pathways of *M. thiooxydans* strain L4 growing on methanol was achieved (Figure 3). The large subunit MxaF of methanol dehydrogenase was detected along with some of the accessory proteins, MxaD, MxaE, MxaJ, MxaG, MxaR and MxaL, indicating expression of the inducible *mxaFJGIRSACKL* operon (Amaratunga *et al.*, 1997, Toyama *et al.*, 1998), as predicted from the metagenome. An additional *mxaRSACKLC* operon (Beck *et al.*, 2014) and at least three of the four alternative methanol

dehydrogenase genes *xoxF* were expressed, together with the associated *xoxJ*. Since four out of the five methanol dehydrogenase genes identified in the metagenome-derived genome of *Methylophaga thiooxydans* strain L4 were expressed, the role of individual *xoxF* genes in methanol oxidation by this methylotroph remains unclear.

In *M. thiooxydans* strain L4, the formaldehyde resulting from oxidation of methanol is assimilated into cell material via the RuMP cycle (Johnson and Quayle, 1965). This was not surprising since all members of the Piscirickettsiaceae family (Chistoserdova and Lidstrom, 2013), including *Methylophaga* spp. (Janvier *et al.*, 1985), have been reported to use this carbon assimilation pathway. All proteins required for the Entner-Doudoroff variant of this cycle and for the transketolase/transaldolase system were detected, including the key enzymes 3-hexulosephosphate synthase and 2-keto-3-dexoy-6-phosphogluconate (KDPG) aldolase (Trotsenko and Murrell, 2008, Chistoserdova, 2011). Formaldehyde is also oxidised to CO_2 via the H_4MPT-dependent pathway to provide this methylotroph with energy for biosynthesis. In contrast, no 6-phosphogluconate dehydrogenase of the oxidative branch of the HuP pathway was detected, possibly indicating preferential use of the direct formaldehyde oxidation pathway via H_4MPT.

The variant of the RuMP cycle detected in the metaproteome fixes three moles of formaldehyde into one mole of pyruvate as the central intermediate (Anthony, 1982). For the conversion of pyruvate, and also phospho*enol*pyruvate (PEP), to further central intermediates such as acetyl-CoA and oxaloacetate, several enzymes were detected that connect to the TCA cycle, including PEP carboxylase, the pyruvate dehydrogenase complex, oxaloacetate decarboxylase and pyruvate carboxyl transferase. Also, all enzymes of the lower part of the Entner-Doudoroff pathway, following KDPG aldolase, were present, enabling the conversion of glyceraldehyde 3-phosphate to pyruvate and vice versa, via PEP synthase, establishing a further connection between carbon fixation and central carbon metabolism in *M. thiooxydans*

strain L4. Finally, the proteins required for oxidation of methylamine, from both the *mau* and the *gma*/*mgs*/*mgd* gene cluster, have not been identified in the proteome data. This is not surprising, as the incubations were carried out with methanol as growth substrate and these pathways are only induced in the presence of methylamine.

In this study, we combine for the first time DNA- and protein- stable isotope probing with metagenomics and functional gene amplicon sequencing to investigate the methanol metabolism of an uncultivated marine methylotroph. This focussed metagenomic approach yielded a near complete genome of an uncultivated *Methylophaga* species, *M. thiooxydans* strain L4, and metaproteomics analysis established the pathways of methanol metabolism in this bacterium. This focussed 'omics approach using ^{13}C-labeled substrates will have significant utility in cultivation-independent studies in microbial ecology.

Experimental Procedures

Stable Isotope Probing (SIP) experimental set up. Surface seawater for SIP experiments was collected from station L4 in the English Channel (50°15.0'N; 4°13.0‘W) on September the 2nd, 2013. Four litres of seawater were filtered in duplicate through 0.22 µm Sterivex TM filters (Merck Millipore) using a peristaltic pump (Watson-Marlow 502S, 1 ml min^{-1}), to extract DNA for analysis of the bacterial community composition at the start of the SIP experiment. Four 2 L gas-tight glass bottles were filled with 0.75 L of the same seawater, inoculated with 75 µmol of ^{13}C-labeled (2 bottles) or (^{12}C) unlabeled (2 control bottles) methanol and incubated at 25°C in a shaking incubator (50 rpm) for three days. Methanol concentration in the incubation bottles was measured every day using an Agilent 7890A gas chromatograph (GC) equipped with a 7693A autosampler and a HP-5 column (see Supplementary Information). After incubation for three days, no methanol could be detected, indicating that sufficient ^{13}C had been incorporated during the SIP experiment. Seawater from all four SIP incubations was filtered through Sterivex filters using a 50 ml syringe. All

filters were stored at -20°C before extracting DNA and proteins within two weeks of the start of the experiment.

DNA extraction. DNA was extracted from Sterivex filters by adding 1.6 ml of SET buffer (0.75 M sucrose, 40 mM EDTA, 50 mM Tris-HCl pH 9) and 0.2 ml of 10% (w/v) SDS and incubating with rotation in a hybridization oven (Hybaid, Waltham, MA, USA) at 55°C for 2.5 h. After incubation, two phenol:chloroform:isoamyl alcohol (25:24:1) extractions and a single chloroform:isoamyl extraction were performed before precipitating the DNA overnight at -20°C with a glycogen solution (Roche, Basel, Switzerland), 7.5 M ammonium acetate and ethanol, as previously described by Neufeld and collaborators (Neufeld et al., 2007a; see Supplementary Information). DNA was pelleted by centrifugation at 4,500 x *g* for 30 min before washing twice with 80% (v/v) ethanol, drying for 15 min at room temperature and resuspending in 50 µl of nuclease-free water.

Protein extraction. The phenol phases from the DNA extraction were combined and mixed with a five-fold volume of ice-cold 100 mM ammonium acetate in methanol and also left at -20°C overnight for precipitation. Samples were centrifuged for 30 min at 4,500 x *g* and protein pellets were washed twice with 100 mM ammonium acetate in methanol, twice with ice cold 80% (v/v) acetone and once with ice cold 70% (v/v) ethanol and finally dried at room temperature.

DNA-SIP centrifugation and fractionation. For each sample, 5 µg of DNA extracted from SIP incubations were added to a mixture of 7.163 M CsCl and Gradient Buffer (0.1 M Tris, 0.1 M KCl and 1 mM EDTA) set to a final density of 1.725 g ml^{-1} before centrifugation for 40 hrs at 20°C and 44,100 rpm (~177,000 x *g*) with vacuum, maximum acceleration and no brake, using a Vti 65.2 rotor and a Optima™ LE-80K Ultracentrifuge (Beckman Coulter). Fractionation of CsCl gradients was done using a low-flow peristaltic pump as described in Neufeld and colleagues (Neufeld et al., 2007b). A total of twelve CsCl fractions, each of

425 µl, were obtained, ranging from heavy to light DNA. DNA from all fractions was precipitated by adding 20 µg of linear polyacrylamide (LPA) and 900 µl of PEG-NaCl 6000 solution (30%/1.6M) and left at room temperature overnight before centrifugation at 13,000 x *g* for 30 min. DNA pellets were washed with 500 µl of 70% (v/v) ethanol, centrifuged for another 10 min, air-dried for 15 min and resuspended in 50 µl of TE buffer (10 mM Tris-HCl, 1 mM EDTA).

DNA amplicon sequencing. Primer sets used in this study were: 27Fmod (5′ AGRGTTTGATCMTGGCTCAG 3′) and 519Rmodbio (5′-GTNTTACNGCGGCKGCTG-3') to amplify 16S rRNA gene, 1003F/1555R (Neufeld et al., 2007a) to amplify *mxaF* gene and a primer set targeting clade 5 to amplify *xoxF* gene (Taubert *et al., in revision*). In each case, three independent PCR products were combined. Purified PCR products were sequenced by 454 pyrosequencing (GS FLX Titanium system, MR DNA, Shallowater, TX, USA). 16S rRNA gene data analysis was done according to Dowd *et al.*, 2011) and DeSantis *et al.*, 2006). *mxaF* and *xoxF* functional gene amplicon sequencing data was analyzed using software packages mothur (Schloss *et al.*, 2009) and USEARCH (Edgar, 2013). See Supplementary Information for details.

Metagenome sequencing. DNA from the heavy fraction of duplicate ^{13}C experiments was amplified using REPLI-g Mini Kit (Qiagen), using 5-10 ng of DNA (or nuclease-free water for negative controls) as starting material, following instructions provided by the manufacturer. The amplified DNA was purified using LPA and PEG-NaCl 6000 solution (see Supplementary Information) and 4 µg from each sample were sent for MiSeq, 2 x 300 bp, Illumina sequencing (2 million reads; MR DNA, Shallowater, TX, USA). The metagenome data received was analyzed using MG-RAST (Meyer *et al.*, 2008) to determine the phylogenetic classification of the reads based on the Ribosomal Database Project (RDP) database (Wang *et al.*, 2007), assembled in BaseSpace (basespace.illumina.com) using

384 SPAdes Genome Assembler v3.0 (Bankevich *et al.*, 2012) and annotated in RAST (Aziz *et*
385 *al.*, 2008).

386 **Sequence data deposition.** Nucleotide sequences from 454 amplicon pyrosequencing
387 obtained in this study were deposited in the GenBank nucleotide sequence database under
388 accession numbers KM657588 (*mxaF*) and KM657641 - KM657644 (*xoxF*). Raw data from
389 454 amplicon pyrosequencing of 16S rRNA and functional gene amplicons have been
390 deposited in the Sequence Read Archive (SRA) of NCBI under accession numbers
391 SRR1576828, SRR1576831, SRR1584480 - SRR1584483, SRR1584485, SRR1584486,
392 SRR1584503, SRR1584504, SRR1584506 and SRR1584507. Annotated genome sequences
393 of *M. thiooxydans* strain L4 are available in the GenBank Whole Genome Shotgun (WGS)
394 database under accession number JRQD01000000. Raw Illumina MiSeq data were deposited
395 at BaseSpace (https://basespace.illumina.com/s/eiGGwUvz6xBP).

396 **Protein-SIP analyses.** Protein extracts were denatured by incubation in SDS sample buffer
397 (62.5 mM Tris/HCl pH 6.8, 10% glycerol (v:v), 2% SDS (w:v), 5% mercaptoethanol (v:v),
398 0.005% bromophenol blue) at 90°C for 10 minutes, followed by centrifugation at 13,000 x *g*
399 for 10 minutes. Supernatants were subjected to one-dimensional SDS polyacrylamide gel
400 electrophoresis for prefractionation as described previously (Taubert *et al.*, 2012). Gel lanes
401 were cut into four bands each. Bands were destained and dehydrated, followed by reduction
402 with 10 mM dithiothreitol for 30 minutes at room temperature and subsequently alkylation
403 with 100 mM iodacetamide for 30 minutes at room temperature. Proteolysis with trypsin was
404 performed overnight at 37°C. Extracted peptides were desalted and concentrated using
405 ZipTip-µC18 (Merck Millipore, Darmstadt, Germany). Solvents were evaporated under
406 vacuum and samples were resuspended in 0.1% formic acid for LC-MS/MS analysis. Mass
407 spectrometry analysis was performed by an Orbitrap Fusion instrument (Thermo Fisher
408 Scientific, Waltham, MA, USA) coupled to a TriVersa NanoMate (Advion, Ltd., Harlow,

UK). In total, 5 µL of the peptide lysates were separated via a Dionex Ultimate 3000 nano-LC system (Dionex/Thermo Fisher Scientific, Idstein, Germany). Raw data files were converted to peak lists and analyzed using TOPPAS v1.11.0 and OpenMS pipeline (Kohlbacher et al., 2007, Sturm et al., 2008), with the OMSSA search algorithm v2.1.8 (Geer et al., 2004). Two databases were used, one consisting of protein sequences obtained from the NCBInr database and one consisting of the predicted protein sequences from the metagenome. Only peptides with a false discovery rate (FDR) <2%, estimated by a decoy database, and peptide rank equal 1 were considered as identified (see Supplementary Information).

Acknowledgments

We thank Dr. A. Crombie for help with GC measurements and the captain and crew of the R.V. Plymouth Quest for collecting the seawater used in this work. The authors are grateful for using the analytical facilities of the Centre for Chemical Microscopy (ProVIS) at the Helmholtz Centre for Environmental Research which is supported by European Regional Development Funds (EFRE – Europe funds Saxony) and the Helmholtz Association. This work was supported by the Gordon and Betty Moore Foundation Marine Microbiology Initiative Grant GBMF3303 to J. Colin Murrell and Yin Chen and through the Earth and Life Systems Alliance, Norwich Research Park, Norwich, UK.

References

Amaratunga, K., Goodwin, P.M., O'Connor, C.D., and Anthony, C. (1997) The methanol oxidation genes *mxaFJGIR(S)ACKLD* in *Methylobacterium extorquens*. *FEMS Microbiol Lett* **146**: 31-38.

Anthony, C. (1982) *Biochemistry of methylotrophs*. London, United Kingdom: Academic Press.

433 Anthony, C., Ghosh, M., and Blake, C.C.F. (1994) The structure and function of methanol
434 dehydrogenase and related quinoproteins containing pyrroloquinoline quinone. *Biochem J*
435 **304**: 665–674.
436 Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A. et al. (2008) The
437 RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**: 75.
438 Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S. et al.
439 (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell
440 sequencing. *J Comput Biol* **19**: 455-477.
441 Beale, R., Dixon, J.L., Smyth, T.J., and Nightingale, P.D. (2015) Annual study of oxygenated
442 volatile organic compounds in UK shelf waters. *Mar Chem* **171**: 96-106.
443 Beck, D.A., McTaggart, T.L., Setboonsarng, U., Vorobev, A., Kalyuzhnaya, M.G., Ivanova,
444 N. et al. (2014) The expanded diversity of Methylophilaceae from Lake Washington through
445 cultivation and genomic sequencing of novel ecotypes. *PloS one* **9**: e102458.
446 Boden, R., Kelly, D.P., Murrell, J.C., and Schäfer, H. (2010) Oxidation of dimethylsulfide to
447 tetrathionate by *Methylophaga thiooxidans* sp. nov.: a new link in the sulfur cycle. *Environ*
448 *Microbiol* **12**: 2688-2699.
449 Boden, R., Ferriera, S., Johnson, J., Kelly, D.P., Murrell, J.C., and Schäfer, H. (2011) Draft
450 genome sequence of the chemolithoheterotrophic, halophilic methylotroph *Methylophaga*
451 *thiooxydans* DMS010. *J Bacteriol* **193**: 3154-3155.
452 Chen, Y., Scanlan, J., Song, L., Crombie, A., Rahman, M.T., Schäfer, H., and Murrell, J.C.
453 (2010) γ-Glutamylmethylamide is an essential intermediate in the metabolism of
454 methylamine by *Methylocella silvestris*. *Appl Environ Microbiol* **76**: 4530-4537.
455 Chistoserdova, L. (2011) Modularity of methylotrophy, revisited. *Environ Microbiol* **13**:
456 2603-2622.

457 Chistoserdova, L., and Lidstrom, M.E. (1997) Molecular and mutational analysis of a DNA
458 region separating two methylotrophy gene clusters in *Methylobacterium extorquens* AM1.
459 *Microbiology+* **143**: 1729-1736.
460 Chistoserdova, L., and Lidstrom, M. (2013) Aerobic Methylotrophic Prokaryotes. In *The*
461 *Prokaryotes*. Rosenberg, E., DeLong, E., Lory, S., Stackebrandt, E., and Thompson, F. (eds):
462 Springer Berlin Heidelberg, pp. 267-285.
463 Chistoserdova, L., Kalyuzhnaya, M.G., and Lidstrom, M.E. (2009) The expanding world of
464 methylotrophic metabolism. *Annu Rev Microbiol* **63**: 477-499.
465 Chistoserdova, L., Gomelsky, L., Vorholt, J.A., Gomelsky, M., Tsygankov, Y.D., and
466 Lidstrom, M.E. (2000) Analysis of two formaldehyde oxidation pathways in *Methylobacillus*
467 *flagellatus* KT, a ribulose monophosphate cycle methylotroph. *Microbiology+* **146**: 233-238.
468 DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K. et al. (2006)
469 Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with
470 ARB. *Appl Environ Microbiol* **72**: 5069-5072.
471 Dixon, J.L., Beale, R., and Nightingale, P.D. (2011) Microbial methanol uptake in northeast
472 Atlantic waters. *ISME J* **5**: 704-716.
473 Dowd, S., Delton Hanson, J., Rees, E., Wolcott, R., Zischau, A., Sun, Y. et al. (2011) Survey
474 of fungi and yeast in polymicrobial infections in chronic wounds. *J Wound Care* **20**: 40-47.
475 Dumont, M.G., and Murrell, J.C. (2005) Community-level analysis: key genes of aerobic
476 methane oxidation. *Methods Enzymol* **397**: 413-427.
477 Edgar, R.C. (2013) UPARSE: highly accurate OTU sequences from microbial amplicon
478 reads. *Nat Methods* **10**: 996-998.
479 Geer, L.Y., Markey, S.P., Kowalak, J.A., Wagner, L., Xu, M., Maynard, D.M. et al. (2004)
480 Open mass spectrometry search algorithm. *J Proteome Res* **3**: 958-964.

Gilbert, J.A., Field, D., Swift, P., Newbold, L., Oliver, A., Smyth, T. et al. (2009) The seasonal structure of microbial communities in the Western English Channel. *Environ Microbiol* **11**: 3132-3139.

Gilbert, J.A., Steele, J.A., Caporaso, J.G., Steinbruck, L., Reeder, J., Temperton, B. et al. (2012) Defining seasonal marine microbial community dynamics. *ISME J* **6**: 298-308.

Janvier, M., Frehel, C., Grimont, F., and Gasser, F. (1985) *Methylophaga marina* gen. nov., sp. nov. and *Methylophaga thalassica* sp. nov., marine methylotrophs. *Int J Syst Bacteriol* **35**: 131-139.

Johnson, P.A., and Quayle, J. (1965) Microbial growth on C_1 compounds. Synthesis of cell constituents by methane-and methanol-grown *Pseudomonas methanica*. *Biochem J* **95**: 859-867.

Kalyuzhnaya, M.G., Lapidus, A., Ivanova, N., Copeland, A.C., McHardy, A.C., Szeto, E. et al. (2008) High-resolution metagenomics targets specific functional types in complex microbial communities. *Nat Biotechnol* **26**: 1029-1034.

Keltjens, J.T., Pol, A., Reimann, J., and den Camp, H.J.O. (2014) PQQ-dependent methanol dehydrogenases: rare-earth elements make a difference. *Appl Microbiol Biotechnol*: 1-21.

Kohlbacher, O., Reinert, K., Gröpl, C., Lange, E., Pfeifer, N., Schulz-Trieglaff, O., and Sturm, M. (2007) TOPP—the OpenMS proteomics pipeline. *Bioinformatics* **23**: e191-e197.

Latypova, E., Yang, S., Wang, Y.-S., Wang, T., Chavkin, T.A., Hackett, M. et al. (2010) Genetics of the glutamate-mediated methylamine utilization pathway in the facultative methylotrophic beta-proteobacterium *Methyloversatilis universalis* FAM5. *Mol Microbiol* **75**: 426-439.

McDonald, I.R., and Murrell, J.C. (1997) The methanol dehydrogenase structural gene *mxaF* and its use as a functional gene probe for methanotrophs and methylotrophs. *Appl Environ Microbiol* **63**: 3218-3224.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M. et al. (2008) The metagenomics RAST server–a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**: 386.

Neufeld, J.D., Chen, Y., Dumont, M.G., and Murrell, J.C. (2008a) Marine methylotrophs revealed by stable isotope probing, multiple displacement amplification and metagenomics. *Environ Microbiol* **10**: 1526-1535.

Neufeld, J.D., Boden, R., Moussard, H., Schäfer, H., and Murrell, J.C. (2008b) Substrate-specific clades of active marine methylotrophs associated with a phytoplankton bloom in a temperate coastal environment. *Appl Environ Microbiol* **74**: 7321-7328.

Neufeld, J.D., Schäfer, H., Cox, M.J., Boden, R., McDonald, I.R., and Murrell, J.C. (2007a) Stable isotope probing implicates *Methylophaga* spp. and novel Gammaproteobacteria in marine methanol and methylamine metabolism. *ISME J* **1**: 480-491.

Neufeld, J.D., Vohra, J., Dumont, M.G., Lueders, T., Manefield, M., Friedrich, M.W., and Murrell, J.C. (2007b) DNA stable isotope probing. *Nat Protoc* **2**: 860-866.

Radajewski, S., Ineson, P., Parekh, N.R., and Murrell, J.C. (2000) Stable isotope probing as a tool in microbial ecology. *Nature* **403**: 646-649.

Sargeant, S. (2013) Microbial utilisation of methanol in seawater. In *Department of Life Sciences*: University of Warwick, p. 429.

Schäfer, H. (2007) Isolation of *Methylophaga* spp. from marine dimethylsulfide-degrading enrichment cultures and identification of polypeptides induced during growth on dimethylsulfide. *Appl Environ Microbiol* **73**: 2580-2591.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B. et al. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537-7541.

Seifert, J., Taubert, M., Jehmlich, N., Schmidt, F., Völker, U., Vogt, C. et al. (2012) Protein-based stable isotope probing (protein-SIP) in functional metaproteomics. *Mass Spectrom Rev* **31**: 683-697.

Sturm, M., Bertsch, A., Gröpl, C., Hildebrandt, A., Hussong, R., Lange, E. et al. (2008) OpenMS–an open-source software framework for mass spectrometry. *BMC Bioinformatics* **9**: 163.

Sun, J., Steindler, L., Thrash, J.C., Halsey, K.H., Smith, D.P., Carter, A.E. et al. (2011) One carbon metabolism in SAR11 pelagic marine bacteria. *PLoS One* **6**: e23973.

Taubert, M., Vogt, C., Wubet, T., Kleinsteuber, S., Tarkka, M.T., Harms, H. et al. (2012) Protein-SIP enables time-resolved analysis of the carbon flux in a sulfate-reducing, benzene-degrading microbial consortium. *ISME J* **6**: 2291-2301.

Toyama, H., Anthony, C., and Lidstrom, M.E. (1998) Construction of insertion and deletion mxa mutants of *Methylobacterium extorquens* AM1 by electroporation. *FEMS Microbiol Lett* **166**: 1-7.

Trotsenko, Y.A., and Murrell, J.C. (2008) Metabolic aspects of aerobic obligate methanotrophy. *Adv Appl Microbiol* **63**: 183-229.

Villeneuve, C., Martineau, C., Mauffrey, F., and Villemur, R. (2012) Complete genome sequences of *Methylophaga* sp. strain JAM1 and *Methylophaga* sp. strain JAM7. *J Bacteriol* **194**: 4126-4127.

Villeneuve, C., Martineau, C., Mauffrey, F., and Villemur, R. (2013) *Methylophaga nitratireducenticrescens* sp. nov. and *Methylophaga frappieri* sp. nov., isolated from the biofilm of the methanol-fed denitrification system treating the seawater at the Montreal Biodome. *Int J Syst Evol Microbiol* **63**: 2216-2222.

554 von Bergen, M., Jehmlich, N., Taubert, M., Vogt, C., Bastida, F., Herbst, F.-A. et al. (2013)
555 Insights from quantitative metaproteomics and protein-stable isotope probing into microbial
556 ecology. *ISME J* **7**: 1877-1885.
557 Vorholt, J.A., Marx, C.J., Lidstrom, M.E., and Thauer, R.K. (2000) Novel formaldehyde-
558 activating enzyme in *Methylobacterium extorquens* AM1 required for growth on methanol. *J*
559 *Bacteriol* **182**: 6645-6650.
560 Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007) Naive Bayesian classifier for
561 rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ*
562 *Microbiol* **73**: 5261-5267.
563 Wischer, D., Kumaresan, D., Johnston, A., El Khawand, M., Stephenson, J., Hillebrand-
564 Voiculescu, A.M. et al. (2014) Bacterial metabolism of methylated amines and identification
565 of novel methylotrophs in Movile Cave. *ISME J, 101038/ismej2014102*.

566

567 **Table and figure legends**

568 Table 1. Central pathways for carbon and nitrogen metabolism detected in the metagenome
569 dataset obtained from ^{13}C-DNA of the ^{13}C-labeled incubations. (1) No 6-
570 phosphofructokinase, (2) No 2-oxoglutarate dh complex, (3) Via 2-keto-3-
571 desoxyphosphogluconate aldolase (KDPGA) and transketolase/transaldolase (TK/TA), (4)
572 Via glutamine synthase/ glutamine 2-oxoglutarate aminotransferase (GS/GOGAT).
573 Figure 1. Phylogenetic diversity of the total bacterial community (A) at the beginning (T0)
574 and end (unfractionated DNA) of the Stable Isotope Probing (SIP) experiment, (B) of ^{13}C-
575 labeled (heavy) and unlabeled (light) DNA and (C) in total peptides and peptides labeled by
576 methylotrophy from SIP samples incubated for three days with 100 μM of ^{13}C-labeled
577 methanol. Results presented in A and B are based on 16S rRNA gene pyrosequencing data,
578 whereas in (C) they are based on protein SIP analysis of the ^{13}C samples.

Figure 2. The methanol dehydrogenase gene cluster of *Methylophaga thiooxydans* strain L4 (Genbank accession LP43_0439 to LP43_0425) (*M. t.* L4) retrieved from assembled metagenomic sequences, compared with sequences from the available genomes of three other *Methylophaga* species: *M. thiooxydans* DMS010 (*M. t.* DMS010), *M. nitratireducenticrescens* strain JAM1 (*M. n.* JAM1) and *M. frappieri* strain JAM7 (*M. f.* JAM7). Different colours correspond to different genes. *mxaF* and *mxaI* correspond to the large and small subunit of methanol dehydrogenase, *mxaG* encodes the associated cytochrome and *mxaJ* is a gene of unknown function required for activity. *mxaDE* and *mxaYX* have regulatory functions in gene expression, *mxaRSACKL* are required for maturation and activation of the enzyme.

Figure 3. Key steps in carbon metabolism of *M. thiooxydans* strain L4 identified by combining metagenomics and metaproteomics: coloured arrows indicate pathways that were present in the metagenome and were detected completely (green), partially (yellow) or not detected (red) in the metaproteomics data. Key proteins are: (1) methanol dehydrogenase, (2) formaldehyde activating enzyme, (3) D-arabino-3-hexulose 6-phosphate formaldehyde lyase, (4) 6-phospho-3-hexuloisomerase, (5) 2-keto-3-deoxyphosphogluconate aldolase. PP and ED are pentose phosphate and Entner-Doudoroff pathway, respectively. TCA, tricarboxylic acid cycle; H_4MPT, tetrahydromethanopterin; HuP, dissimilatory hexulose phosphate cycle; RuMP, ribulose monophosphate cycle; GS/GOGAT, glutamine synthase/glutamine 2-oxoglutarate amidotransferase.