

# New insights into the pelagic microorganisms involved in the methane cycle in the meromictic Lake Pavin through metagenomics

Corinne Biderre-Petit, Najwa Taib, Hélène Gardon, Corentin Hochart, Didier Debroas

▶ **To cite this version:**

Corinne Biderre-Petit, Najwa Taib, Hélène Gardon, Corentin Hochart, Didier Debroas. New insights into the pelagic microorganisms involved in the methane cycle in the meromictic Lake Pavin through metagenomics. FEMS Microbiology Ecology, 2018, 10.1093/femsec/fiy183 . hal-01984730

**HAL Id: hal-01984730**
**https://hal.science/hal-01984730**

Submitted on 26 Nov 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# New insights into the pelagic microorganisms involved in the methane cycle in the meromictic Lake Pavin through metagenomics

Biderre-Petit Corinne[1*], Taib Najwa[1§], Gardon Hélène[1], Hochart Corentin[1], Debroas Didier[1]

[1]Université Clermont Auvergne, CNRS, Laboratoire Microorganismes: Génome et Environnement, F-63000 Clermont-Ferrand, France

[§]Present adress: Hub Bioinformatique et Biostatistique, Unité de Biologie évolutive de la cellule microbienne, Département de Microbiologie, Institut Pasteur, Paris, France

*corresponding author: Corinne Biderre-Petit, corinne.petit@uca.fr

Abstract

Advances in metagenomics have given rise to the possibility of obtaining genome sequences from uncultured microorganisms, even for those poorly represented in microbial community, thereby providing important means to study their ecology and evolution. In this study, metagenomic sequencing was carried out at four sampling depths having different oxygen concentrations or environmental conditions in the water column of Lake Pavin. By analyzing the sequenced reads and matching the contigs to the proxy genomes of the closest cultivated relatives, we evaluated the metabolic potential of the dominant planktonic species involved in the methane cycle. We demonstrated that methane-producing communities were dominated by the genus *Methanoregula* while methane-consuming communities were dominated by the genus *Methylobacter*, thus confirming prior observations. Our work allowed the reconstruction of a draft of their core metabolic pathways. Although hydrogenotrophs, the presence of the genes required for acetate activation in the methanogen genome were also detected. Regarding methanotrophy, *Methylobacter* was present in the same areas as the non-methanotrophic, methylotrophic *Methylotenera,* which could suggest a relationship between these two groups. Furthermore, the presence of a large gene inventory for nitrogen metabolism (nitrate transport, denitrification, nitrite assimilation and nitrogen fixation, for instance) was detected in the *Methylobacter* genome.

Uncorrected Proof

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

Keywords

Meromictic lake, metagenomics, methane cycle, methanogens, methanotrophs, metabolic reconstruction.

## Introduction

Methane ($CH_4$) is the second most important greenhouse gas after carbon dioxide ($CO_2$) but with a global warming potential that is 21-fold higher. About one-third of total $CH_4$ emissions in the atmosphere come from natural sources, dominated by wetland emissions (watercourses, swamps, ponds, oceans and lakes). Lakes are estimated to contribute from 6 to 16% of total natural $CH_4$ emissions (Bastviken *et al.* 2004; Kirschke *et al.* 2013). The release from lakes is therefore considerable (Bastviken *et al.* 2011), exceeding that from oceans despite the comparatively small portion of inland waters on the earth's surface (<1%) (Downing *et al.* 2006). In lacustrine systems, most of the $CH_4$ originates from the activity of methanogenic archaea through decomposition of organic matter under anoxic conditions (Conrad, Klose and Noll 2009). After production, as much as 80-90% of the $CH_4$ is captured by aerobic and anaerobic methanotrophs that use this gas as a sole carbon (C) and energy source for growth before it escapes to the atmosphere (Knittel and Boetius 2009; Ettwig *et al.* 2010; Chistoserdova 2015). Most of the $CH_4$ oxidation takes place in the top layer of the sediment, in which $CH_4$ and dioxygen ($O_2$) form steep counter gradients (Auman *et al.* 2000). Hence, by mitigating $CH_4$ emissions, microorganisms act as an efficient biological $CH_4$ filter (Cole *et al.* 1994; Bastviken *et al.* 2004).

Methane has been the focus of intensive research over the last few decades because of its ecological and applied interests. This work has resulted in a detailed understanding of the principal processes linked to its global cycle and the apparition of hundreds of genome sequences for both methanotrophic and methanogenic microorganisms in public databases. Nevertheless, genomic data on methanotrophs are far less numerous in comparison to that of methanogens, mainly due to the very recent discovery of novel biogeochemical processes and microorganisms still having poorly characterized metabolisms (Wu *et al.* 2011; Hernandez *et al.* 2015). For instance, it has recently been demonstrated that aerobic methanotrophs belonging to *Proteobacteria* were capable to couple $CH_4$ oxidation to nitrate ($NO_3^-$) reduction under $O_2$ limitation (Kits, Klotz and Stein 2015), thereby highlighting their unexpected metabolic flexibility and assigning these bacteria a new role at the metabolic intersection of C and nitrogen (N) cycles. Furthermore, data available for lakes are still scarce compared to

Uncorrected Proof

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

other systems, even though these environments are characterized by dynamic cycling of $CH_4$, serving both as its major sources and major sinks. The principal studies have focused on a few models and their sedimentary compartments (Rahalkar *et al.* 2009; Chistoserdova and Lidstrom 2013); but few studies have addressed pelagic freshwater environments (Pernthaler, 2017). However, the importance of this latter compartment is certainly largely underestimated for the overall understanding of $CH_4$ cycling as evidenced by the active $CH_4$ production in the oxygenated waters of many lakes (Schulz *et al.* 2001; Grossart *et al.* 2011; Tang *et al.* 2016) as well as the intensive $CH_4$ oxidation in anoxic waters in stratified lakes (Karr *et al.* 2006). Moreover, permanently stratified lakes (e.g. meromictic) present another interesting characteristic, since only negligible amount of $CH_4$ are released into the atmosphere compared to the more classical shallow lakes where most of the $CH_4$ is consumed at the oxic-anoxic boundary in the water column (Borges *et al.* 2011).

On account of their particular properties, meromictic lakes represent excellent field laboratories to follow biologically mediated redox processes, especially those linked to $CH_4$ cycling. In Lake Pavin, metabarcoding surveys were conducted in order to highlight which microorganisms play a critical role in the most significant $CH_4$ pathways. These studies revealed that most of the methanogens were closely related to the genus *Methanoregula* in the *Methanomicrobiales* class and the methanotrophs to the genus *Methylobacter* (Biderre-Petit *et al.* 2011b; Borrel *et al.* 2011). However, the methods used, based on single marker amplification, only gave insight into the specific richness within these relevant uncultivated species but none into their genomic structure. To overcome these limitations, metagenomics has recently become a powerful tool for collecting information on microbial communities. It provides an all-inclusive picture of their specific functional pathways and allow us to investigate their distribution in response to their different environmental conditions in natural habitats (Kalyuzhnaya *et al.* 2008; Quince *et al.* 2017). In our study, we used a strategy which enabled us to decipher the metabolic potential of the microorganisms involved in methanotrophy and methanogenesis along the Lake Pavin water column through a two-step based method of metagenomic sequencing analysis. The first step consisted in a read-centric approach in order to identify the entire genetic repertoire of planktonic methanotrophs and methanogens and thus, perform a large-scale, strain-level analysis. The second step was an assembly-based approach to resolve the genomic organization for the dominant related-populations thriving in this ecosystem.

Uncorrected Proof

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

## Materials and methods

### Lake Pavin water samples

Lake Pavin is a small (~0.44 $km^2$), deep (92 m), almost circular (~750 m diameter) lake of volcanic origin formed 6,900 years ago and located in the Massif Central (France) (45°29.740N, 2°53.280E), at an altitude of 1,197 m above sea level. Four sampling depths with contrasted $O_2$ tensions were selected: the surface (4 m) with high $O_2$ tension; the oxycline with low $O_2$ tension (53 m) and two last depths in the anoxic zone, one close to the oxic-anoxic interface (65 m; chemocline) and one deeper (80 m), previously shown to contain a high $CH_4$ concentration (Lehours *et al.* 2005; Biderre-Petit *et al.* 2011b; Lopes *et al.* 2011). Water samples were collected on July 2013 from a platform positioned above the deepest point of the lake (92 m deep). A volume of 20 L was collected using a Van Dorn bottle as described previously (Biderre-Petit *et al.* 2011a). Immediately on collection, samples were transferred into sterile bottles and transported to the laboratory on ice within 0.5 to 3h. Serial filtrations (25, 15 and 1.2 µm pore size membrane filters) were performed on each sample to remove large particles and eukaryotic cells. After filtration, tangential ultrafiltration (Amicon pump) was used to concentrate the filtrate containing microorganisms to a final volume of 1 L. As a last step, the microbial biomass was collected on 0.2 µm pore size (pressure <100 mbar) polycarbonate filters (47 mm TSPT Millipore filters, Billerica, Massachusetts, USA) and stored at -80°C to await nucleic acid extraction.

### Physico-chemical parameters

Temperature and dissolved $O_2$ profiles were obtained using a multiparameter probe ProOdO™ (Ysi, Germany). For determination of sulfate ($SO_4^{2-}$), $NO_3^-$ and ammonium ($NH_4^+$) concentration, 3 replicates × 50 ml of sampled waters were filtered through a 0.2 µm syringe filter and stored frozen at -20°C to await analysis using Spectroquant reagent standard kits (Merck, Germany) (Fig. S1, Supporting Information).

### DNA extraction and sequencing

Genomic DNA (gDNA) was extracted using the standard phenol-chloroform method as previously described (Biderre-Petit *et al.* 2011b). DNA concentrations were estimated using a spectrophotometer (NanoDrop ND-1000, Nanodrop products, Wilmington, DE, USA). The metagenomic library preparation and sequencing were performed by GATC Biotech (Konstanz, Germany) from ~6 µg of gDNA for each sample. Sequencing was performed

Uncorrected Proof

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

using the Illumina HiSeq-2000 paired-end technology (2 x 100 bases) which generated from ~95 to 124 x $10^6$ reads depending on the sample, comprising a total of ~41 gigabases (Gb) of sequence data (Table 1).

Read-centric analysis

To facilitate analysis, especially in terms of computing time, a sub-sample of each raw data set containing 2 x $10^6$ randomly picked Illumina reads was generated and then processed through the MG-RAST server pipeline (V3). More than 95% of the reads were retained after quality filtering based on length, ambiguous bases and quality scores. They were then annotated using MG-RAST default parameters (E-value cutoff $10^{-5}$, minimum identity cutoff 60% and minimum alignment length cutoff 15 amino acids for proteic features) (Meyer *et al.* 2008). Sequencing statistics are shown in Table 1. The sequences of genes encoding key enzymes of the metabolisms of interest were extracted from sub-sampled metagenomic datasets through keyword research. For this, specific searches for a selection of 268 enzymes (206 covering central C and N metabolisms for methanotrophs (Table S1, Fig. S2 and S3, Supporting Information) and 62 covering the main methanogenic steps (Table S2, Fig. S4, Supporting Information)) were performed inside the MG-RAST web browser through keyword filter use. The workbench was used to download the FASTA files with reads corresponding to all selected proteins. Given the non-specificity of most of the selected enzymes for a unique metabolic group, the affiliation of all reads downloaded from MG-RAST was manually curated by BLASTX against the NR protein database *via* the web interface and executed with default parameters (https://blast.ncbi.nlm.nih.gov/blast, Jonhson *et al.* 2008). Only reads with the best hits for methanogens and methanotrophs as well as a drastic decrease of similarity with other taxa were retained for further analysis.

Metagenome assembly

For each full read datasets, raw reads were trimmed using FASTX-Toolkit (v 0.0.13.1; Pearson *et al* 1997) with a quality cut-off of 30 and 50 bp as a minimal sequence length. Trimmed reads were then assembled using IDBA-UD version 1.0.9, which was run in its default iterative mode (k-min=20, k-max=100, k-step=20) (Peng *et al.* 2012). From 287,042 to 512,204 contigs were produced depending on the sample (Table 1, Fig. S5, Supporting Information). In the following step, the reads identified by read-centric analysis were blasted against contigs using BLASTN (version 2.2.26) executed with default parameters. Only the

Uncorrected Proof

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

contigs with a read alignment of E-value $<10^{-4}$ were selected. However, given the possibility of cross-reactions and chimeric contigs, the outputs from these automated analyses were manually curated by BLASTN and BLASTX against the NR databases *via* the NCBI site (https://blast.ncbi.nlm.nih.gov/blast) and executed with default parameters. Of the contigs initially identified by BLASTN, up to 19% of those selected for the methanotrophs (312 contigs in total) and 81% for the methanogens (386 contigs) were finally retained. Gene prediction for the remaining contigs was performed using a MetaGeneAnnotator (Noguchi, Taniguchi and Itoh 2008). All predicted translated genes were compared to Refseq non-redundant (prokaryotes + viruses (O'Leary *et al.* 2016)) and KEGG (Kanehisa *et al.* 2016) databases using BLASTP (Altschul *et al.* 1990) with an E-value threshold of $10^{-5}$ and a percent identity of 60%. The rRNA genes in contigs were identified with BLASTN (E-value $< 10^{-5}$, identity $\geq$ 97%) against the SILVA rRNA database v. 123 (Quast *et al.* 2013). The location of the contigs mapped onto the reference genomes was then visualized using the CGView Server (http://stothard.afns.ualberta.ca/cgview_server/; Grant and Stothard 2008). The BLASTN parameters, used in the CGViewer Server to compare the reference genome sequences to the contig sets, had an E-value cutoff of $10^{-3}$, an alignment length cutoff of 100 bp and a percent identity cutoff of 60%.

Nucleotide accession number

Sequence data are available in GenBank under the accession numbers KY06104-KY06107 and MF076238-MF076545 for the methanotrophic contigs and KY994152-KY994537 for the methanogenic contigs. The raw sequences were archived in MG-RAST under the accession numbers mgm4580845.3 to mgm4580848.3 for the metagenomic reads and the accession numbers mgm4557752.3.100, mgm4557753.3.100, mgm4581104.3.100 and mgm4581106.3.100 for the assembled contigs.

## Results

### An overview of the microbial community composition in Lake Pavin water column

According to the read-centric analysis, the domain *Bacteria* numerically dominated the composition of the microbial communities at all four of the depths sampled. More than 93% of the taxonomically assigned reads matched with bacteria, with a few representatives of archaea (range 0.1-3.7%), eukaryotes (0.7-3.8%) and viruses (0.6-3.9%) (Table 1). Archaeal

Uncorrected Proof

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

reads were more abundant in the anoxic waters (range 3.1-3.7% against 0.1% in oxic waters). The rRNA gene affiliation agreed with the broad taxonomic picture provided by the functional genes with 91-99% of rRNA genes affiliated with *Bacteria*, whereas *Archaea* and *Eukarya* were minor components (up to 6.4% and 4%, respectively) (Table 1). Overall, the most abundantly recovered bacterial rRNA reads were consistent with those usually found in freshwaters, with a domination of *Proteobacteria* (59.8%; range 36.6-73.8%), *Actinobacteria* (12.9%; range 3.3-23%), *Bacteroidetes* (8.4%; range 2.6-21.3%), and, to a lesser extent, *Verrucomicrobia* (4.1%; range 2.2-7.6%), *Firmicutes* (1.9%; range 0-4.8%) and *Cyanobacteria* (1.9%; range 0.4-6.4%) (Fig. 1A and 1B). The bacterial community was also composed of many less-abundant phyla (<5% of the community) characterized by a more limited spatial distribution. The largest diversity was observed at the depth 65 m and the lowest at 4 m (Fig. 1B). The former depth is part of the chemocline, a layer known to offer a variety of ecological niches and thereby generally harboring a highly diverse microbial community. Concerning *Archaea*, corresponding rRNA reads were only detected in the monimolimnion and were all affiliated to the hydrogenotrophic *Methanomicrobiales* (Fig. 1A), mainly to the genus *Methanoregula*.

As *Proteobacteria* account for most of the rRNA reads, we focused on this phylum for a comparative study of its potential involvement in the N, sulfur (S) and $CH_4$ geochemistry throughout the water column. We used the relative abundance of the related-genes and literature knowledge as proxies of the potential relevance of each population *in situ*. For the S cycle, reads affiliated with known sulfur-oxidizing bacteria (SOB) were detected from 53 m to 80 m, with the genera *Sulfuricella* and *Sulfuritalea* (β-*Proteobacteria)* in the oxycline and the genera *Sulfuricurvum, Sulfurimonas* and *Sulfurovum* (*ε-Protetobacteria)* in the anoxic layers. As expected, sulfate-reducing bacteria (SRB), with *Desulfobacca* (*δ-Proteobacteria)* as the main genus, were exclusively detected in the anoxic waters (Fig. 2A). For the N cycle, the potential for nitrification was mostly observed at the oxycline with the *Nitrosomonadales* (β-*Proteobacteria*) as the prevailing ammonium oxidizing bacteria (AOB) (Fig. 2B) and the phylum *Nitrospirae* as the dominant nitrite oxidizing bacteria (NOB) (Fig. 1B). As denitrifying bacteria (NBR) are phylogenetically diverse, and distributed over at least 60 genera, they were therefore likely to occur in all samples, with the β-*Proteobacteria* as the primary candidates; principally the genera *Burkholderia*, *Cupriavidus* and *Ralstonia* (Fig. 2B). As for the $CH_4$ cycle, two proteobacterial types, both recognized as $CH_4$-oxidizing bacteria (MOB) (type I MOBs belonging to γ-*Proteobacteria* (*Methylococcaceae*) and type II MOBs belonging to α-*Proteobacteria* (*Methylocystaceae* and *Beijerinckiaceae*)), were found

Uncorrected Proof

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

in the rRNA pool. The former was the most abundant with *Methylococcaceae* reads detected both at 53 m and 65 m, whereas only a few sequences were affiliated with the latter at 4 m. These belonged to the genus *Methylocystis* within the family *Methylocystaceae* (Fig. 2C). The co-occurrence of the non-methanotrophic methylotrophic *Methylotenera* (*β-Proteobacteria),* in the same areas as *Methylococcaceae,* could suggest a potential relationship between these two groups (Fig. 2C).

**Methanogenesis pathway reconstruction**

From the metagenomic data gathered in Lake Pavin, we reconstructed a draft of the core metabolic pathways used by planktonic methanogens for $CH_4$ production. To do so, we firstly established a list of 62 genes coding for proteins and involved in the pathways of methanogenesis (Fig. S4, Table S2, Supporting Information) and then used all possible appellations known in scientific literature for each protein as queries against MG-RAST read annotations. The current study showed a strong correlation between our functional profile, methanogenic genes being found only in anoxic metagenomes (at depths of 65 and 80 m), and the $CH_4$ concentrations measured along the water column of Lake Pavin as seen in previous studies (Lehours *et al.* 2005; Biderre-Petit *et al.* 2011b; Lopes *et al.* 2011). Up to 0.03% of the reads (1,183) of these two metagenomes matched with 52 out of the 62 genes present in the established list. Of note, 55.1% of these were retrieved at a depth of 65 m and the other 44.9% at a depth of 80 m (Fig. 3A). Most of the reads (99.8%) were assigned to methanogenic archaea, mainly the order *Methanomicrobiales* (96.3%) followed by *Methanosarcinales* (1.1%) (Fig. 3B, Fig. S6, Supporting Information) which suggests a dominance of hydrogenotrophic methanogenesis over acetoclastic methanogenesis. The remaining 0.2% were affiliated with non-methanogenic archaea belonging to the order *Thermoplasmatales*. At the genus level, most of methanogenic reads were identified as the hydrogenotrophic *Methanoregula* (66.3%) (Fig. S6, Supporting Information).

Using the methanogenic reads as queries, a total of 386 contigs with a size ranging from 229 to 6,143 bp were isolated from the assembled metagenomes (Table 2). Of these, 49.7% were retrieved at 65 m and 50.3% at 80 m (Fig. 4A). They also had similar size distribution whatever the depth (Fig. 4B). A total of 526 coding DNA sequences (CDSs) were predicted from the contigs, with an average length of 491 bp (1,851 bp for the largest), corresponding to a protein-coding content of 97% (Table S3, Supporting Information), as well as six rRNA genes and three tRNA genes (Table S4, Supporting Information). Compared to

Uncorrected Proof

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

the overall G+C content (54.1%), the intergenic regions had a lower G+C percentage (44%) and the gene sequences had a higher G+C percentage (54.7%) (Table 2). The rRNA operon was organized in a bacterial-like 5'-16S-23S-5S-3' transcriptional unit and displayed the highest identity with *Methanoregula* strains (Table S4, Supporting Information). Like the reads, the CDSs had the highest frequency of high scoring BLAST hits with the genus *Methanoregula* (88.4%) with an average amino acid identity of 82.6% (range 54.7-100%). The frequencies of top BLASTP hits with other groups were, most importantly, *Methanolinea* (4.2%), *Methanosphaerula* (3.4%) and *Methanoculleus* (1.5%) with the two former being also affiliated with the family *Methanoregulaceae* (Table S3, Supporting Information). Most of the CDSs (90.5%) were associated to functions related to the $CH_4$ metabolism initially targeted; only a low proportion (<5%) of the identified KEGG Orthologous (KO) genes was uncategorized (Fig. 5A). All of the CDSs encoded 77 distinct enzymes. The mapping of all contigs onto the *Methanoregula formicicum* SMSPT (CP003167) reference genome using CGView comparison tool software (Grant and Stothard, 2008) showed that they covered up to 85 kbp of this genome, scattered over 24 genomic regions on the chromosome (Fig. 6A). Only a few rearrangements were observed in comparison to the reference genome (Fig. 6B).

From the read and contig analysis, all of the genes required for growth on $H_2$ and $CO_2$ were uncovered, whether it is those encoding enzymes found in the cytoplasmic cell fraction (steps 1-5 and 7; Fig. S4, Supporting Information) or those encoding the membrane-bound methyltransferase (MtrA-H) (step 6; Fig. S4, Supporting Information). Although the Mtr complex is used both by the hydrogenotrophic and the acetoclastic methanogens, corresponding sequences invariably showed the best matches with the former. From contig overlaps, all operons encoding the key enzymatic complexes involved in the different steps of the hydrogenotrophic pathway were reconstructed, i.e., that of the coenzyme M reductase (*mcrBDCGA)*, the methyltransferase (*mtrEDCBAFGH*) and the formylmethanofuran dehydrogenase (*fmd/fwdFDBAC*), the latter being present in three scattered copies (Fig. 6B). Two separate copies of the component A2 (*atw2*), predicted to encode an ATP-binding protein required in the activation of the Mcr complex, were also detected. The electron-bifurcating complex (a [Ni-Fe] hydrogenase specific to hydrogenotrophic methanogens), which is composed by the cytoplasmic F420 non-reducing hydrogenase (MvhADG) and the heterodisulfide reductase (HdrABC), was also present, constituted by the *mvhD* subunit located directly downstream of the Hdr operon (Fig. 6B). In contrast, no sequence (read or contig) for the second gene class of Hdr complex (*hdrED),* found in acetoclastic methanogens, was detected. The metagenomes also contained genes encoding three other [Ni-

Uncorrected Proof

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

Fe] hydrogenases which are required for the reduction of ferredoxin, whether it is the Eha and Ehb hydrogenases found in hydrogenotrophic methanogens (Gao and Gupta 2007) or the homologous Ech enzyme common to all methanogens. Only the read datasets revealed the sequences for Eha and Ehb subunits while the contigs allowed the reconstruction of a partial Ech complex, organized in three adjacent genes (subunits B, C and D). Finally, a complete gene set for a cytoplasmic F420-dependent hydrogenase (FrhABDG), a key enzyme catalyzing the reversible reduction of coenzyme F420 with $H_2$, and a second separate copy for the *frhB* gene were reconstructed (Fig. 6B).

Regarding C metabolism, in autotrophic $H_2$-utilizing methanogens, carbon monoxide (CO) is reported to work as a C source within the anabolic reductive acetyl-CoA (AcCoA) pathway, also known as the Wood-Ljungdahl (WL) pathway. In this study, two copies of the operon encoding the carbon monoxide complex (Cdh) responsible for $CO_2$ fixation and AcCoA formation (one complete (*cdhABCDE*) and one partial (*cdhCDE*)) were detected in the metanogenomic datasets (Fig. 6B). The enzymes responsible for AcCoA synthesis from acetate, i.e. the acetyl-CoA synthetase (Acs) and phosphotransacetylase (Pta), an alternative to the autotrophic process, were also present in the metagenomes. However, the acetate kinase (Ack) was not detected in any of the read or contig datasets (Fig. S4, Supporting Information). We also found the pyruvate ferredoxin oxidoreductase (Por) enzyme (*porGDAB*) responsible for the conversion of AcCoA to pyruvate (Fig. 6B).

**Methanotrophy pathway reconstruction**

Applying the same method as used for methanogenesis, we reconstructed a draft of the core metabolic pathways of $CH_4$ oxidation and C assimilation for the planktonic MOBs present in Lake Pavin. We also included, in our survey, the inventory of the genes involved in the N cycle on account of the current growing interest in the impact of this cycle on MOB growth and their $CH_4$ oxidation capabilities. To do so, we established a list of 206 proteins (107 for C metabolism and 99 for N cycle) (Table S1, Fig. S2 and S3, Supporting Information) and then processed as previously described. Up to 0.009% metagenomic reads (735 reads), at the four sampling depths, matched with 76 out of the 107 genes involved in central C metabolism. Most of them were recovered from the 53 m metagenome (81.5%), which agreed with the sharp decline in the $CH_4$ concentration observed close to this depth (Lopes *et al.* 2011), followed by 65 m (13.5%), 80 m (4.1%) and 4 m (0.9%) (Fig. 3A). Concerning the N cycle, methanotrophic genes were detected in 53 m and 65 m datasets (Table S5, Supporting

Uncorrected Proof

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

Information) with up to 0.003% of the reads (152 reads) matching with 14 out of 99 targeted genes; 96.7% of them were from 53 m and 3.3% from 65 m. The taxonomic affiliation showed that type I MOBs dominated, representing more than 98.3% of the methanotrophic reads (Fig. 3B), with *Methylobacter*-like sequences representing the most significant proportion of this group (39.4% of total type I MOB sequences; Fig. S7A and B, Supporting Information).

Using MOB reads as queries, a total of 312 contigs ranging from 236 to 12,298 bp and representing a total of 593.27 kb of DNA sequence were isolated from the assembled metagenomes (Table 2); 77.9% of them were retrieved from 53 m, 20.8% from 65 m and 1.3% from 80 m, none was isolated from 4 m (Fig. 4A, Table S6, Supporting Information). The longest contigs were identified at 65 m with a mean size of 2,800 bp against 1,688 bp for 53 m and 646 bp for 80 m (Fig. 4B). A total of 622 CDSs were predicted (Table S7, Supporting Information), with an average length of 875 bp (3,399 bp for the largest), corresponding to a protein-coding content of 91.7%, as well as seven rRNA genes and one tRNA gene. As seen with the methanogens, the intergenic regions had a lower G+C percentage (33.9%) and the gene sequences, a higher G+C percentage (41.9%), compared to the overall G+C content (41.3%) (Table 2). No rRNA operon could be reconstructed from assembled data; 16S rRNA genes had 99 to 100% identity to *Methylobacter* strains and 23S rRNA genes, 92 to 98% identity to *Methylomonas* strains (Table S6). Like the reads, CDSs had the highest frequency of high scoring BLAST hits with the species *Methylobacter tundripaludum* 21/22 (79.1%) with an average amino acid identity of 81.3% (range 46.8-100%). The frequencies of top BLASTP hits with other groups were as follows, *Methylosarcina* (9.2%), *Methylomicrobium* (6.9%), *Methylomonas* (2.9%) and *Methylomarinum* (1.9%) (Table S7, Supporting Information). Even though, as expected, the majority of the CDSs (40.1%) were associated with the targeted metabolisms (mainly carbohydrates with glycolysis, pentose phosphate pathway and TCA cycle (19.3%) and energy (specifically to $CH_4$ (19%) and N (1.1%) metabolisms), this approach highlighted genes involved in many other important processes (Fig. 5B). Compared to methanogenesis, a large proportion (>27%) of the identified KO genes remained uncategorized (Fig. 5, Table S7, Supporting Information). All of the CDSs encoded 361 distinct enzymes. The mapping of all of the contigs onto *M. tundripaludum* 21/22 reference genome showed that they covered up to 381 kbp, scattered over 111 unique regions (Fig. 7A). Gene organization conservation in both order and orientation was observed for most of the genomic regions, except in seven locations where rearrangements were detected, mostly deletions (Fig. 7B).

Uncorrected Proof

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

From the read and CDS analysis, most of gene encoding enzymes responsible for $CH_4$ oxidation to $CO_2$ were uncovered: i) both the particulate and soluble forms of the methane monooxygenases (pMMO and sMMO respectively), the key enzymes responsible for the $CH_4$ oxidation to methanol, ii) the $Ca^{2+}$-dependent (Mxa) and the $Ln^{3+}$-dependent (XoxF) forms of methanol dehydrogenases (MDHs) another hallmark of the methanotrophy responsible for oxidation of methanol to formaldehyde and iii) enzymes required for the pathways involving the two pterin cofactors, i.e. the tetrahydrofolate ($H_4F$) and the tetrahydromethanopterin ($H_4MPT$), both responsible for formaldehyde dissimilation to $CO_2$ (Fig. S2, Table S6, Supporting Information). Furthermore, the expression and the activity of pMMO are known to be tightly regulated by copper (Cu) (for a review see Semrau, DiSpirito and Yoon 2010). In our study, CopA, a Cu-ATPase involved in Cu efflux and CopZ, a Cu chaperone, both known to be organized within a Cu-regulated operon (Sitthisak *et al.* 2007), were found. As for the MDHs, the Mxa cluster showed the same highly conserved gene organization as other organisms with the *mxaF* gene being linked to genes *mxaJ* and *mxaG* (Table S6, Supporting Information). However, the small catalytic subunit *mxaI* was not detected as the assembled contig carrying the complex was not sufficiently long.

In the MOBs, C assimilation takes place either at the level of formaldehyde *via* the ribulose monophosphate (RuMP) pathway or at the level of methylene-$H_4F$ and $CO_2$ *via* the serine cycle. Regarding the RuMP cycle, most of the genes encoding the enzymes involved both in the Embden-Meyerhof-Parnas (EMP) variant and the collateral Enthner-Doudoroff (EDD) variant were present in the metagenomes. The enzymes responsible for the inter-conversions of the C3 intermediate metabolites (pyruvate and phosphoenolpyruvate) resulting from formaldehyde conversion through RuMP were also detected as well as those for a complete tricarboxylic (TCA) cycle. In contrast, none of the sequences detected for Ack, Pta and Acs enzymes, which are required for acetate conversion to AcCoA, were classified as MOBs. The components of the dissimilatory RuMP cycle for formaldehyde oxidation and RuMP regeneration was also present (Fig. S2, Supporting Information). As for the serine cycle, almost all of the related-genes were uncovered from the contigs (Fig. S2, Supporting Information). They clustered on the same genomic fragment (Table S6, Supporting Information). Only the malyl-CoA-lyase (Mcl), the truly specific enzyme of the serine cycle, and the hydroxypyruvate reductase (HPR) were missing from our metagenomic data. Though this cycle is considered as the hallmark of type II MOBs, corresponding sequences invariably showed better matches with type I MOBs. Finally, regarding the inventory of essential genes encoding N metabolisms, the methanotrophic genes involved in $NO_3^-$ transport (*narK*),

Uncorrected Proof

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

denitrification (*nirJ* and *norB*), nitrite assimilation (*nirBD*), N fixation (*nifDK*), $NH_4^+$ transport (*amt*) and assimilation (*glnBDE* and *ntrBC*), and, potentially, nitrification (*hao* homologous sequences), were represented (Fig. S3, Table S6, Supporting Information).

## Discussion

In this study, we investigated the microbial taxa involved in the $CH_4$ cycle which are inhabiting in the waters of Lake Pavin, through the combination of assembly-free and assembly-based analyses of metagenomic data. This two-tiered strategy was adopted because of the high microbial diversity found in the lake waters whereas methanogens and methanotrophs comprise a minor fraction of the natural community. Indeed, methanogens were estimated to contribute to less than 2% of the total Lake Pavin anoxic water community (Lehours *et al.* 2005) while planktonic MOBs, contribute up to 5% of the total cell numbers in most studied lakes (Carini *et al.* 2005; Jones and Lennon 2009; Oswald *et al.* 2015; Samad and Bertilsson 2017). In these conditions, an approach only based on metagenomic assembly analysis would have required much more computational resources and been much more time-consuming without the possibility of assembling whole genomes for these guilds owing to an insufficient coverage (Quince *et al.* 2017). Furthermore, despite its democratization, the use of traditional metagenomics to specifically investigate key metabolic functions in highly diversified microbial communities is still scarce. Most of the metagenomic approaches developed for these issues rely on substrate-specific labeling methods such as stable isotope probing (SIP) before sequencing. However, though used to study benthic methylotrophic populations (Kalyuzhnaya *et al.* 2008; Chistoserdova 2011; Beck *et al.* 2013), this method has never been applied on pelagic populations.

### *Methanoregula* as dominant $CH_4$ producers in anoxic waters

In the present work, the anoxic metagenomic datasets harbored a low proportion of archaea (with a high of 3.7% of taxonomy classified reads), with 99.7% of archaeal selected reads assigned to methanogenic taxa, mostly *Methanomicrobiales* (Fig. 3). This order, widely reported from planktonic environments (Crevecoeur, Vincent and Lovejoy 2016), was designated as an indicator group for freshwater (Auguet, Barberan and Casamayor 2010) and, consequently, considered among the main players in the pivotal ecological functions within this kind of habitat. In our study, no depth-related change in terms of methanogenic read abundance or diversity was observed. This implies that these $CH_4$ producers are likely to

Uncorrected Proof

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

occur at relatively constant proportions across the anoxic waters and would contribute, at least in part, to the $CH_4$ measured in waters of Lake Pavin, as previously demonstrated in other lakes (Iversen, Oremland and Klug 1987; Crowe *et al.* 2011). The genus *Methanoregula* was the dominant taxon, which corroborates our previous findings based on the use of single markers (Biderre-Petit *et al.* 2011b; Denonfoux *et al.* 2013). This genus, for which 16S rRNA sequences were recovered in most freshwater lake clone libraries (Borrel *et al.* 2011), is suspected to have poor salt tolerance (Tong *et al.* 2017).

On the basis of the CDSs information, a draft of methanogenesis pathways was successfully reconstructed for the methanogens in Lake Pavin with the identification of 55 genes, all with encoding proteins directly or indirectly involved in hydrogenotrophic processes. As for the other *Methanoregula* sequenced, the phylotypes living in Lake Pavin are expected to utilize $H_2/CO_2$ for growth, highlighting the importance of hydrogenotrophic methanogenesis in this ecosystem in comparison to the other methanogenic pathways. Concerning central carbon metabolism, we note the presence of the genes for a Cdh complex, known to allow $CO_2$ fixation by the Wood-Ljungdahl pathway. Beyond autotrophic metabolism, the methanogens in Lake Pavin also contained an acetyl-CoA synthetase (Acs) and a phosphotransacetylase (Pta), known to be involved in acetate activation, which has been demonstrated previously in other hydrogenotrophic methanogens (Kouzuma *et al.* 2017). This could be an advantage under limiting growth conditions. Moreover, a large set of hydrogenases was found (Ech, Mvh, Eha, Ehb and Frh). As with most *Methanomicrobiales*, the complementary [NiFe] hydrogenase (MvhAG) was not identified, but the MvhD subunit was present, tightly bound to HdrABC operon. The best explanation was that MvhAG could be replaced by FrhAG (identified as well) to form a functional complex to catalyze methanogenesis through electron bifurcation (Anderson *et al.* 2009; Thauer *et al.* 2010; Kaster *et al.* 2011; Browne *et al.* 2017; Gilmore *et al.* 2017). However, that association has still not been experimentally verified. Finally, the reconstructed Mcr operon, the key enzyme of methanogenesis, displayed the same genomic organization as that described using the capture approach, with the *dtxR* gene (encoding a metal-dependent repressor) located directly upstream (Denonfoux *et al.* 2013). Considering that this organization differs from what is generally observed for sequenced *Methanoregula*, it could suggest that it reflects the adaptation of the methanogens to their environment in Lake Pavin.

The assembly-based analyses of the metagenomic data revealed several variants with few percent nucleotide differences for most of the functional genes characterized for methanogens, thus resulting in an important marker redundancy and a short size for most of

Uncorrected Proof

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

the reconstructed contigs. Such a functional diversity suggests the co-existence of many closely related populations for the hydrogenotrophic *Methanoregula* group living in Lake Pavin. To sum up, in this study, the functional genes, with the potential to shed light onto strain-level heterogeneity, appeared to be much better-performing markers yielding to a taxonomic resolution not achievable by ribosomal RNA genes. Indeed, all recovered rRNA gene sequences (16S, 23S and 5S rRNA genes) were identical while rRNA operons only differed by a few nucleotides in spacer regions. Hence, some of these strains could form separate groups with distinct ecological relevance. However, such conclusions must be considered with care because characterizing strains from assembled metagenomes alone remains challenging on account of the many possible biases such as sequencing errors, fragmentary contigs and even the quality of the assembly (Kelley and Salzberg 2010; Wang, Ye and Tang 2012; Shapiro and Polz 2014).

### *Methylobacter* as dominant MOBs under hypoxic and suboxic conditions

Planktonic MOBs, a functional guild found in all studied lakes regardless of status, depth or water chemistry (Ross *et al.* 1997; Eller *et al.* 2005; Sundh, Bastviken and Tranvik 2005; Kojima, Fukuhara and Fukui 2009; Samad and Bertilsson 2017), constitute an important part of freshwater ecosystems and are considered a key link between sediment and pelagic C flow. However, despite the intensive research carried out over the last few years, major gaps still exists in our fundamental knowledge of this key microbial group. In stratified lakes, it consumes most of the $CH_4$ at the oxic-anoxic interface, but also in anoxic waters, sometimes well-below the chemocline (Biderre-Petit *et al.* 2011b; Peura *et al.* 2012; Blees *et al.* 2014). In this work, the analysis of the metagenomes in Lake Pavin supports previous metagenomic findings for stratified lakes (Peura *et al.* 2015), i.e., a high frequency of methanotrophic genes in the hypoxic and upper anoxic zones while practically absent near the surface. *Methylobacter* was the dominant taxon and thus could be the main active methanotroph species, even under anoxic conditions. This postulation is in accordance with the growing number of studies that point to its dominant role in lakes, whether they are stratified (Taipale, Jones and Tiirola 2009; Blees *et al.* 2014) or not (Grossart *et al.* 2011; Tsutsumi *et al.* 2011; Ullrich *et al.* 2016), as well as in sediments (Kalyuzhnaya *et al.* 2008; Dumont, Pommerenke and Casper 2013; Oshkin *et al.* 2015; Martinez-Cruz *et al.* 2017). Low temperatures were proposed to be a significant driver of its predominance (Tsutsumi *et al.* 2011). Furthermore, comparisons of sample-specific datasets uncovered the presence of distinct populations and richness closely related to *Methylobacter tundripaludum* depending on the depth. This

Uncorrected Proof

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

suggests a vertical shift in its assemblage along the water column with potentially different ecophysiological characteristics for *Methylobacter* strains, leading to niche differentiation, as hypothesized in other studies (Tsutsumi *et al.* 2011; Oshkin *et al.* 2015).

In this work, 361 gene encoding proteins involved in the C, N and energy pathways were identified thus allowing a partial reconstruction of the *Methylobacter* metabolism. This included all three subunits of the pMMO, the key enzyme of $CH_4$ oxidation that was organized in an operon *pmoCAB* as in the other type I MOBs (Trotsenko and Murrell 2008). However, the reads detected for sMMO have better matches with *Methylomonas* and *Crenothrix* than with *Methylobacter*. This also includes the two distinct MDH systems, i.e. the classical $Ca^{2+}$-dependent MxaF type and the $Ln^{3+}$-containing XoxF type. Moreover, it was recently demonstrated that the presence of non-methanotrophic methylotrophs, particularly *Methylotenera*, could induce a change in the expression of MOB MDHs, with a downregulation of the dominant XoxF-type and an upregulation of the MxaF-type (Krause *et al.* 2017). This switch, resulting in a higher methanol release in medium, could benefit methylotroph growth through a cross-feeding mechanism (He *et al.* 2012; Beck *et al.* 2013). The presence of the pelagic *Methylotenera* in this study positively correlated with that of *Methylobacter*, hence suggesting a potential similar cross-feeding mechanism. Furthermore, their co-location with the peak of $NO_3^-$ (Fig. S1, Supporting Information) and low $O_2$ concentrations (Fig. 3) could be explained by a niche adaptation conferred by a putative denitrification capability as previously shown in other ecosystems (Kalyuzhnaya *et al.* 2008; Stein and Klotz 2011; Chistoserdova 2015; Oshkin *et al.* 2015; Oswald *et al.* 2016; Martinez-Cruz *et al.* 2017). Indeed, the coupling of $NO_3^-$ respiration to $CH_4$ oxidation has previously been reported for *Methylobacter* and other MOBs, alone (Kits, Klotz and Stein 2015) or in cooperation (Beck *et al.* 2013; Dumont, Pommerenke and Casper 2013; Zhu *et al.* 2016). However, further investigation is still needed to demonstrate that these species engage in cooperative behavior and the importance of the N cycle in this cooperation. It is interesting to note that the methanotrophs that thrive in Lake Pavin have a rich inventory for nitrogen oxide metabolism.

Both the EDD- and EMP-variants of the RuMP cycle, required by type I MOBs to assimilate formaldehyde into biomass, were predicted in our assembled metagenomes. Which of these variants is primarily employed is still today a debated question, though a recent analysis pointed toward the EMP variant as the major route for single C assimilation, contrary to existing assumptions (Kalyuzhnaya *et al.* 2013). We also showed that the *Methylobacter* in Lake Pavin is endowed with the majority of the serine cycle genes, thus agreeing with recent

Uncorrected Proof

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

genomic and proteomic data (de la Torre *et al.* 2015; Ullrich *et al.* 2016; Padilla *et al.* 2017). On the other hand, the interconnection of a partial serine cycle and the EMP-variant might represent an alternative strategy to increase the conversion efficiency of AcCoA production from $CH_4$ (Kalyuzhnaya, Puri and Lidstrom 2015) which, once generated, could enter the TCA cycle for which all essential genes were detected. The presence of putative mixed-acid fermentation and $H_2$ production genes in the *Methylobacter* genome also opens up a possibility for fermentation under $O_2$-limiting conditions as recently proposed for a *Methylomicrobium* species (Kalyuzhanaya et al, 2013). Indeed, the switching capability from a respiratory mode to a fermentation mode understandingly represents an advantage for organisms living in stratified environments such as Lake Pavin, where availability of $O_2$ and other electron acceptors varies. Hence, the very broad metabolic flexibility of MOBs in Lake Pavin might be the key to their presence in the hypoxic and suboxic zones, as this allows them to adjust to shifting environmental settings (Oswald *et al.* 2016).

## Conclusion

Overall, this study expands the current genomic knowledge of pelagic $CH_4$ producers and utilizers in Lake Pavin. In the absence of cultivated species from Lake Pavin water column, metagenomics, with the partial reconstruction of the core parts required for methanogenesis and methanotrophy, revealed to be much more informative than the PCR-dependent approaches based on the use of single marker genes. Indeed, though the latter methods have significantly contributed to the advancement of our understanding of the uncultivated methanogens and MOBs in Lake Pavin (Lehours *et al.* 2007; Biderre-Petit *et al.* 2011b), they didn't provide any insight into their genomic structure. This outcome is of particular importance when considering the microbial communities thriving along the steep redox gradients of $O_2$ minimum zones as these habitats are predicted to expand in response to climate change. Furthermore, the overlapping patterns between the methanogens and the MOBs in the suboxic area of Lake Pavin is consistent with the simultaneous production and consumption of $CH_4$. This knowledge is critical for predicting links between greenhouse gas, C and nutrient fluxes in such ecosystems. However, as all of our results are based on metagenomics (DNA level), further approaches based on metatranscriptomics or metaproteomics are required in future studies to explore the principal functions of the living microorganisms involved in the $CH_4$ production in the water masses of Lake Pavin.

Uncorrected Proof

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

## Funding

The work of C.H. was supported by the Agence Nationale de la Recherche (ANR) through the projects EUREKA (ANR-14-CE02-0004-01). H.G. is supported by a grant from the French Ministry for higher Education and Research.

## Acknowledgements

We thank Agnès Vellet for its efficient technical assistance and Cécile Lepère for reviewing the English version of the manuscript.

## References

Altschul SF, Gish W, Miller W *et al.* Basic local alignment search tool. *J Mol Biol* 1990;**215**:403–10.

Anderson I, Ulrich LE, Lupa B *et al.* Genomic Characterization of Methanomicrobiales Reveals Three Classes of Methanogens. Ahmed N (ed.). *PLoS ONE* 2009;**4**:e5797.

Auguet J-C, Barberan A, Casamayor EO. Global ecological patterns in uncultured Archaea. *ISME J* 2010;**4**:182.

Auman AJ, Stolyar S, Costello AM *et al.* Molecular characterization of methanotrophic isolates from freshwater lake sediment. *Appl Environ Microbiol* 2000;**66**:5259–66.

Bastviken D, Cole J, Pace M *et al.* Methane emissions from lakes: Dependence of lake characteristics, two regional assessments, and a global estimate: LAKE METHANE EMISSIONS. *Glob Biogeochem Cycles* 2004;**18**:n/a – n/a.

Bastviken D, Tranvik LJ, Downing JA *et al.* Freshwater Methane Emissions Offset the Continental Carbon Sink. *Science* 2011;**331**:50–50.

Beck DAC, Kalyuzhnaya MG, Malfatti S *et al.* A metagenomic insight into freshwater methane-utilizing communities and evidence for cooperation between the *Methylococcaceae* and the *Methylophilaceae*. *PeerJ* 2013;**1**:e23.

Biderre-Petit C, Boucher D, Kuever J *et al.* Identification of sulfur-cycle prokaryotes in a low-sulfate lake (Lake Pavin) using *aprA* and 16S rRNA gene markers. *Microb Ecol* 2011a;**61**:313–27.

Biderre-Petit C, Jézéquel D, Dugat-Bony E *et al.* Identification of microbial communities involved in the methane cycle of a freshwater meromictic lake: Methane cycle in a stratified freshwater ecosystem. *FEMS Microbiol Ecol* 2011b;**77**:533–45.

Blees J, Niemann H, Wenk CB *et al.* Micro-aerobic bacterial methane oxidation in the chemocline and anoxic water column of deep south-Alpine Lake Lugano (Switzerland). *Limnol Oceanogr* 2014;**59**:311–24.

Uncorrected Proof

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

Borges AV, Abril G, Delille B *et al.* Diffusive methane emissions to the atmosphere from Lake Kivu (Eastern Africa). *J Geophys Res* 2011;**116**, DOI: 10.1029/2011JG001673.

Borrel G, Jézéquel D, Biderre-Petit C *et al.* Production and consumption of methane in freshwater lake ecosystems. *Res Microbiol* 2011;**162**:832–47.

Browne P, Tamaki H, Kyrpides N *et al.* Genomic composition and dynamics among Methanomicrobiales predict adaptation to contrasting environments. *ISME J* 2017;**11**:87–99.

Busigny V, Planavsky NJ, Jézéquel D *et al.* Iron isotopes in an Archean ocean analogue. *Geochim Cosmochim Acta* 2014;**133**:443–62.

Carini S, Bano N, LeCleir G *et al.* Aerobic methane oxidation and methanotroph community composition during seasonal stratification in Mono Lake, California (USA). *Environ Microbiol* 2005;**7**:1127–38.

Chistoserdova L. Methylotrophy in a Lake: from Metagenomics to Single-Organism Physiology. *Appl Environ Microbiol* 2011;**77**:4705–11.

Chistoserdova L. Methylotrophs in natural habitats: current insights through metagenomics. *Appl Microbiol Biotechnol* 2015;**99**:5763–79.

Chistoserdova L, Lidstrom ME. Aerobic Methylotrophic Prokaryotes. *The Prokaryotes*. Springer, Berlin, Heidelberg, 2013, 267–85.

Cole JJ, Caraco NF, Kling GW *et al.* Carbon Dioxide Supersaturation in the Surface Waters of Lakes. *Science* 1994;**265**:1568–70.

Conrad R, Klose M, Noll M. Functional and structural response of the methanogenic microbial community in rice field soil to temperature change. *Environ Microbiol* 2009;**11**:1844–53.

Crevecoeur S, Vincent WF, Lovejoy C. Environmental selection of planktonic methanogens in permafrost thaw ponds. *Sci Rep* 2016;**6**, DOI: 10.1038/srep31312.

Crowe SA, Katsev S, Leslie K *et al.* The methane cycle in ferruginous Lake Matano: Methane cycle in ferruginous Lake Matano. *Geobiology* 2011;**9**:61–78.

Denonfoux J, Parisot N, Dugat-Bony E *et al.* Gene capture coupled to high-throughput sequencing as a strategy for targeted metagenome exploration. *DNA Res* 2013;**20**:185–96.

Downing JA, Prairie YT, Cole JJ *et al.* The global abundance and size distribution of lakes, ponds, and impoundments. *Limnol Oceanogr* 2006;**51**:2388–97.

Dumont MG, Pommerenke B, Casper P. Using stable isotope probing to obtain a targeted metatranscriptome of aerobic methanotrophs in lake sediment: SIP-metatranscriptomics of methanotrophs. *Environ Microbiol Rep* 2013:n/a – n/a.

Eller G, Deines P, Grey J *et al.* Methane cycling in lake sediments and its influence on chironomid larval Î´13C. *FEMS Microbiol Ecol* 2005;**54**:339–50.

Uncorrected Proof

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

Ettwig KF, Butler MK, Le Paslier D *et al.* Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. *Nature* 2010;**464**:543–8.

Gao B, Gupta RS. Phylogenomic analysis of proteins that are distinctive of Archaea and its main subgroups and the origin of methanogenesis. *BMC Genomics* 2007;**8**:86.

Gilmore SP, Henske JK, Sexton JA *et al.* Genomic analysis of methanogenic archaea reveals a shift towards energy conservation. *BMC Genomics* 2017;**18**, DOI: 10.1186/s12864-017-4036-4.

Grant JR, Stothard P. The CGView Server: a comparative genomics tool for circular genomes. *Nucleic Acids Res* 2008;**36**:W181–4.

Grossart H-P, Frindte K, Dziallas C *et al.* Microbial methane production in oxygenated water column of an oligotrophic lake. *Proc Natl Acad Sci* 2011;**108**:19657–61.

Hernandez ME, Beck DAC, Lidstrom ME *et al.* Oxygen availability is a major factor in determining the composition of microbial communities involved in methane oxidation. *PeerJ* 2015;**3**:e801.

He R, Wooller MJ, Pohlman JW *et al.* Diversity of active aerobic methanotrophs along depth profiles of arctic and subarctic lake water column and sediments. *ISME J* 2012;**6**:1937–48.

Iversen N, Oremland RS, Klug MJ. Big Soda Lake (Nevada). 3. Pelagic methanogenesis and anaerobic methane oxidation. *Limnol Oceanogr* 1987;**32**:804–14.

Jones S, Lennon J. Evidence for limited microbial transfer of methane in a planktonic food web. *Aquat Microb Ecol* 2009;**58**:45–53.

Johnson M, Zaretskaya I, Raytselis Y, *et al.* NCBI BLAST: a better web interface. *Nucleic Acids Res* 2008;**36**:W5–W9.

Kalyuzhnaya MG, Lapidus A, Ivanova N *et al.* High-resolution metagenomics targets specific functional types in complex microbial communities. *Nat Biotechnol* 2008;**26**:1029–34.

Kalyuzhnaya MG, Puri AW, Lidstrom ME. Metabolic engineering in methanotrophic bacteria. *Metab Eng* 2015;**29**:142–52.

Kalyuzhnaya MG, Yang S, Rozova ON *et al.* Highly efficient methane biocatalysis revealed in a methanotrophic bacterium. *Nat Commun* 2013;**4**, DOI: 10.1038/ncomms3785.

Kanehisa M, Sato Y, Kawashima M *et al.* KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 2016;**44**:D457–62.

Karr EA, Ng JM, Belchik SM *et al.* Biodiversity of Methanogenic and Other Archaea in the Permanently Frozen Lake Fryxell, Antarctica. *Appl Environ Microbiol* 2006;**72**:1663–6.

Kaster A-K, Moll J, Parey K *et al.* Coupling of ferredoxin and heterodisulfide reduction via electron bifurcation in hydrogenotrophic methanogenic archaea. *Proc Natl Acad Sci* 2011;**108**:2981–6.

Uncorrected Proof

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

Kelley DR, Salzberg SL. Clustering metagenomic sequences with interpolated Markov models. *BMC Bioinformatics* 2010;**11**:544.

Keltjens JT, Pol A, Reimann J *et al.* PQQ-dependent methanol dehydrogenases: rare-earth elements make a difference. *Appl Microbiol Biotechnol* 2014;**98**:6163–83.

Kirschke S, Bousquet P, Ciais P *et al.* Three decades of global methane sources and sinks. *Nat Geosci* 2013;**6**:813–23.

Kits KD, Klotz MG, Stein LY. Methane oxidation coupled to nitrate reduction under hypoxia by the Gammaproteobacterium *M ethylomonas denitrificans* , sp. nov. type strain FJG1: Denitrifying metabolism in *M. denitrificans* FJG1. *Environ Microbiol* 2015;**17**:3219–32.

Knittel K, Boetius A. Anaerobic Oxidation of Methane: Progress with an Unknown Process. *Annu Rev Microbiol* 2009;**63**:311–34.

Kojima H, Fukuhara H, Fukui M. Community structure of microorganisms associated with reddish-brown iron-rich snow. *Syst Appl Microbiol* 2009;**32**:429–37.

Kouzuma A, Tsutsumi M, Ishii S *et al.* Non-autotrophic methanogens dominate in anaerobic digesters. *Sci Rep* 2017;**7**, DOI: 10.1038/s41598-017-01752-x.

Krause SMB, Johnson T, Samadhi Karunaratne Y *et al.* Lanthanide-dependent cross-feeding of methane-derived carbon is linked by microbial community interactions. *Proc Natl Acad Sci* 2017;**114**:358–63.

Lehours A-C, Bardot C, Thenot A *et al.* Anaerobic Microbial Communities in Lake Pavin, a Unique Meromictic Lake in France. *Appl Environ Microbiol* 2005;**71**:7389–400.

Lehours A-C, Evans P, Bardot C *et al.* Phylogenetic Diversity of Archaea and Bacteria in the Anoxic Zone of a Meromictic Lake (Lake Pavin, France). *Appl Environ Microbiol* 2007;**73**:2016–9.

Lopes F, Viollier E, Thiam A *et al.* Biogeochemical modelling of anaerobic vs. aerobic methane oxidation in a meromictic crater lake (Lake Pavin, France). *Appl Geochem* 2011;**26**:1919–32.

Martinez-Cruz K, Leewis M-C, Herriott IC *et al.* Anaerobic oxidation of methane by aerobic methanotrophs in sub-Arctic lake sediments. *Sci Total Environ* 2017;**607-608**:23–31.

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

Noguchi H, Taniguchi T, Itoh T. MetaGeneAnnotator: Detecting Species-Specific Patterns of Ribosomal Binding Site for Precise Gene Prediction in Anonymous Prokaryotic and Phage Genomes. *DNA Res Int J Rapid Publ Rep Genes Genomes* 2008;**15**:387–96.

O'Leary NA, Wright MW, Brister JR *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* 2016;**44**:D733–45.

Uncorrected Proof

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

Oshkin IY, Beck DA, Lamb AE *et al.* Methane-fed microbial microcosms show differential community dynamics and pinpoint taxa involved in communal response. *ISME J* 2015;**9**:1119.

Oswald K, Milucka J, Brand A *et al.* Light-dependent aerobic methane oxidation reduces methane emissions from seasonally stratified lakes. *PLoS One* 2015;**10**:e0132574.

Oswald K, Milucka J, Brand A *et al.* Aerobic gammaproteobacterial methanotrophs mitigate methane emissions from oxic and anoxic lake waters: Methane oxidation in Lake Zug. *Limnol Oceanogr* 2016;**61**:S101–18.

Padilla CC, Bertagnolli AD, Bristow LA *et al.* Metagenomic Binning Recovers a Transcriptionally Active Gammaproteobacterium Linking Methanotrophy to Partial Denitrification in an Anoxic Oxygen Minimum Zone. *Front Mar Sci* 2017;**4**, DOI: 10.3389/fmars.2017.00023.

Peng Y, Leung HCM, Yiu SM *et al.* IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 2012;**28**:1420–8.

Pernthaler J. Competition and niche separation of pelagic bacteria in freshwater habitats. *Environ Microbiol* 2017;**19**:2133–50.

Peura S, Eiler A, Bertilsson S *et al.* Distinct and diverse anaerobic bacterial communities in boreal lakes dominated by candidate division OD1. *ISME J* 2012;**6**:1640.

Peura S, Sinclair L, Bertilsson S *et al.* Metagenomic insights into strategies of aerobic and anaerobic carbon and nitrogen transformation in boreal lakes. *Sci Rep* 2015;**5**:12102.

Quast C, Pruesse E, Yilmaz P *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2013;**41**:D590–6.

Quince C, Walker AW, Simpson JT *et al.* Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol* 2017;**35**:833–44.

Rahalkar M, Deutzmann J, Schink B *et al.* Abundance and Activity of Methanotrophic Bacteria in Littoral and Profundal Sediments of Lake Constance (Germany). *Appl Environ Microbiol* 2009;**75**:119–26.

Ross JL, Boon PI, Ford P *et al.* Detection and quantification with 16S rRNA probes of planktonic methylotrophic bacteria in a floodplain lake. *Microb Ecol* 1997;**34**:97–108.

Samad MS, Bertilsson S. Seasonal Variation in Abundance and Diversity of Bacterial Methanotrophs in Five Temperate Lakes. *Front Microbiol* 2017;**8**, DOI: 10.3389/fmicb.2017.00142.

Schulz M, Faber E, Hollerbach A *et al.* The methane cycle in the epilimnion of Lake Constance. *Fundam Appl Limnol* 2001;**151**:157–76.

Semrau JD, DiSpirito AA, Yoon S. Methanotrophs and copper. *FEMS Microbiol Rev* 2010;**34**:496–531.

Uncorrected Proof

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

Shapiro BJ, Polz MF. Ordering microbial diversity into ecologically and genetically cohesive units. *Trends Microbiol* 2014;**22**:235–47.

Sitthisak S, Knutsson L, Webb JW *et al.* Molecular characterization of the copper transport system in Staphylococcus aureus. *Microbiology* 2007;**153**:4274–83.

Stein LY, Klotz MG. Nitrifying and denitrifying pathways of methanotrophic bacteria. *Biochem Soc Trans* 2011;**39**:1826–31.

Sundh I, Bastviken D, Tranvik LJ. Abundance, Activity, and Community Structure of Pelagic Methane-Oxidizing Bacteria in Temperate Lakes. *Appl Environ Microbiol* 2005;**71**:6746–52.

Taipale S, Jones R, Tiirola M. Vertical diversity of bacteria in an oxygen-stratified humic lake, evaluated using DNA and phospholipid analyses. *Aquat Microb Ecol* 2009;**55**:1–16.

Tang KW, McGinnis DF, Ionescu D *et al.* Methane Production in Oxic Lake Waters Potentially Increases Aquatic Methane Flux to Air. *Environ Sci Technol Lett* 2016;**3**:227–33.

Thauer RK, Kaster A-K, Goenrich M *et al.* Hydrogenases from Methanogenic Archaea, Nickel, a Novel Cofactor, and H $_2$ Storage. *Annu Rev Biochem* 2010;**79**:507–36.

Tong C, Cadillo-Quiroz H, Zeng ZH *et al.* Changes of community structure and abundance of methanogens in soils along a freshwater–brackish water gradient in subtropical estuarine marshes. *Geoderma* 2017;**299**:101–10.

De la Torre A, Metivier A, Chu F *et al.* Genome-scale metabolic reconstructions and theoretical investigation of methane conversion in Methylomicrobium buryatense strain 5G(B1). *Microb Cell Factories* 2015;**14**, DOI: 10.1186/s12934-015-0377-3.

Trotsenko YA, Murrell JC. Metabolic Aspects of Aerobic Obligate Methanotrophy⋆. *Advances in Applied Microbiology*. Vol 63. Elsevier, 2008, 183–229.

Tsutsumi M, Iwata T, Kojima H *et al.* Spatiotemporal variations in an assemblage of closely related planktonic aerobic methanotrophs: Spatiotemporal variations in planktonic methanotrophic assemblage. *Freshw Biol* 2011;**56**:342–51.

Ullrich N, Casper P, Otto A *et al.* Proteomic evidence of methanotrophy in methane-enriched hypolimnetic lake water: Proteomics of Lake Methanotrophs. *Limnol Oceanogr* 2016;**61**:S91–100.

Vekeman B, Speth D, Wille J *et al.* Genome characteristics of two novel type I methanotrophs enriched from North Sea sediments containing exclusively a lanthanide-dependent XoxF5-type methanol dehydrogenase. *Microb Ecol* 2016;**72**:503–9.

Wang M, Ye Y, Tang H. A *de Bruijn* Graph Approach to the Quantification of Closely-Related Genomes in a Microbial Community. *J Comput Biol* 2012;**19**:814–25.

Uncorrected Proof

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

Wu ML, Ettwig KF, Jetten MSM *et al.* A new intra-aerobic metabolism in the nitrite-dependent anaerobic methane-oxidizing bacterium *Candidatus* "Methylomirabilis oxyfera." *Biochem Soc Trans* 2011;**39**:243–8.

Zhu J, Wang Q, Yuan M *et al.* Microbiology and potential applications of aerobic methane oxidation coupled to denitrification (AME-D) process: A review. *Water Res* 2016;**90**:203–15.

Uncorrected Proof

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

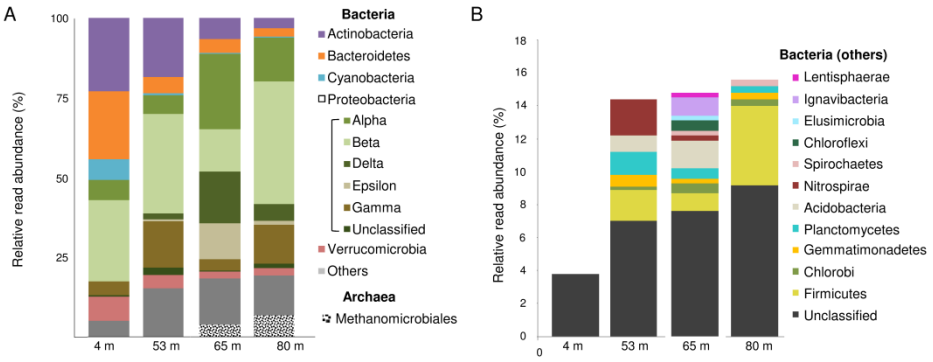



**Figure 1. Taxonomic affiliation of rRNA sequences across depths.** The sequences were assigned at the phylum-level, except the *Proteobacteria* which were represented by class. A) All bacterial and archaeal phyla represented by >5% of the total sequences in any metagenome. 'Others' corresponds to the unclassified bacterial sequences and phyla represented by <5% of the total sequences. B) Bacterial phyla represented by <5% of the total sequences in any metagenome with variable abundance as a function of depth.

Uncorrected

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

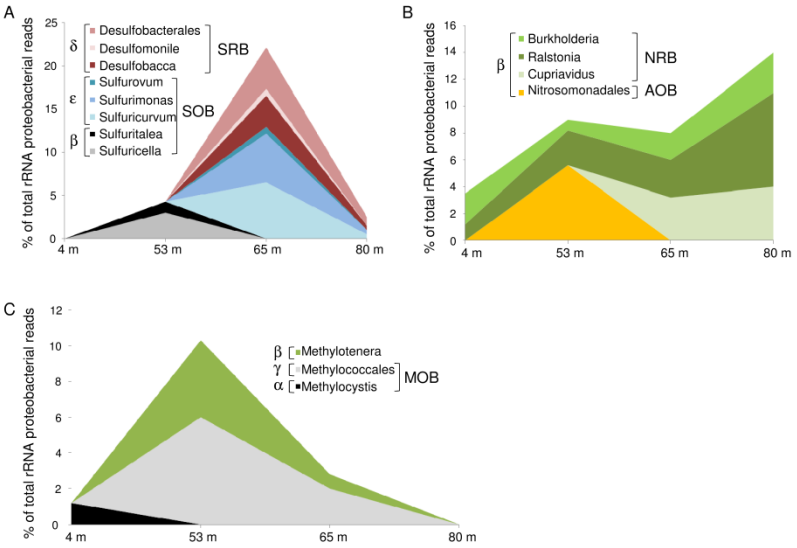



**Figure 2. Functional activities of proteobacterial classes in the water column of Lake Pavin linked with the principal biogeochemical cycles based on rRNA sequence (16S and 23S) affiliation.** A) Taxonomic groups potentially involved in sulfur metabolisms, B) Taxonomic groups potentially involved in nitrogen metabolisms, C) Taxonomic groups potentially involved in methane consumption (*Methylocystis* and *Methylococcaceae*) as well as the abundance of the non methanotrophic, methylotrophic *Methylotenera* genus. AOB, Ammonium Oxidizing Bacteria; MOB, Methane Oxidizing Bacteria; NRB, Nitrate Reducing Bacteria; SOB, Sulfur-Oxidizing Bacteria; SRB, Sulfate Reducing Bacteria.

Uncorrected

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

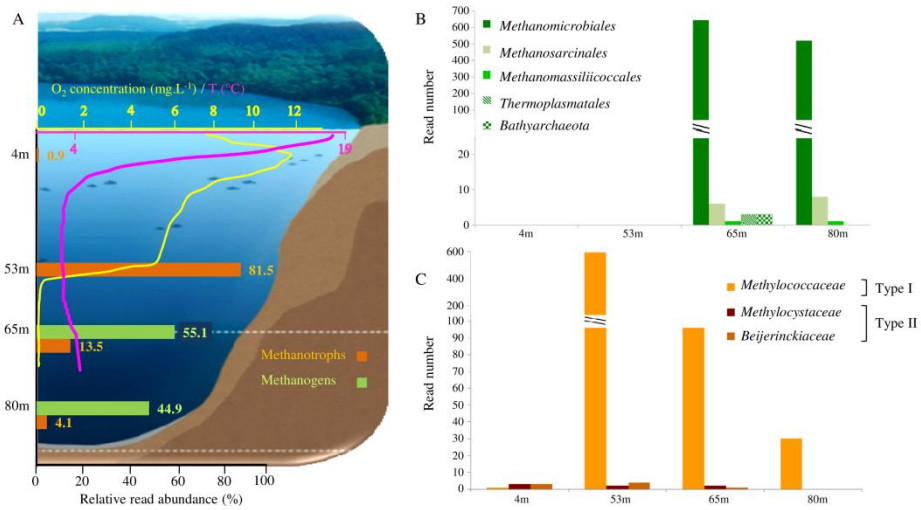



**Figure 3.** A) Vertical cross section of the water column of Lake Pavin showing temperature and dissolved oxygen content as well as read distribution associated to methanogens and methanotrophs. B) Read abundance of archaeal orders depending on the depth. C) Read abundance of type I and II MOBs depending on the depth.

Uncorrected

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

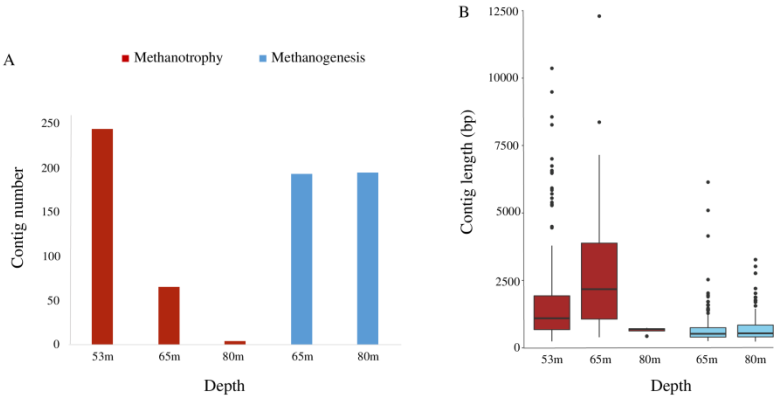



**Figure 4. Contig distribution analysis.** A) Contig number distribution depending on metabolic groups and depths. B) Boxplots of contig size distribution according to metabolic groups and depths.

Uncorrected

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

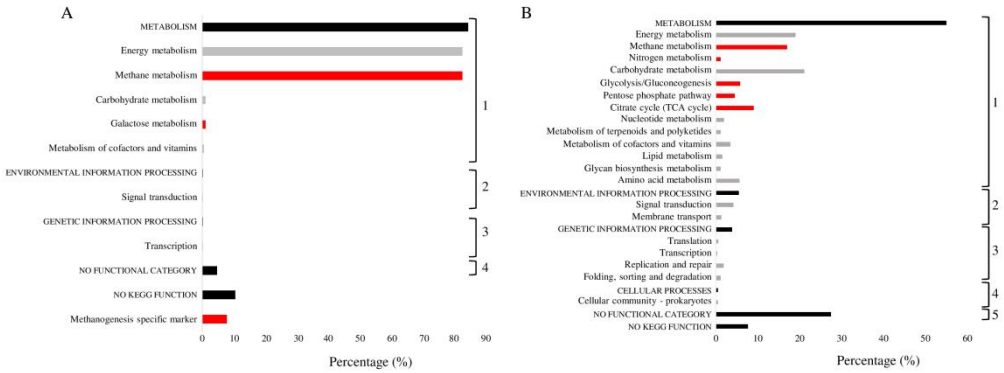



**Figure 5. KEGG pathway assignment of CDSs.** A) Distribution of methanogenic CDSs to different KEGG subcategories belonging to four main categories (hierarchical categories (HC) noted in capital letters). B) Distribution of methanotrophic CDSs to different KEGG subcategories belonging to five HCs. In each HC, the most represented subcategories appear in grey for the functional categories and in red for the pathways. Each CDS was classified into a single category. No functional category: CDSs poorly characterized in the KEGG database. No KEGG function: CDSs without KEGG affiliation.

Uncorrected

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

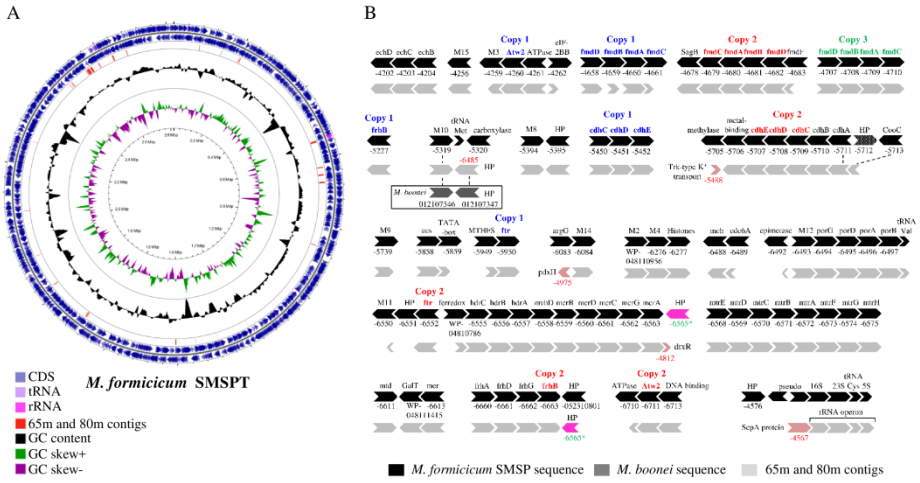



**Figure 6. Graphical representation of contig mapping onto *Methanoregula formicicum* SMSP chromosome used as genome reference (CP003167).** A) Circular map of *M. formicicum* SMSP chromosome performed with a CGview comparison tool (Grant and Stothard, 2008). From outer to inner circle: genes on forward strand; genes on reverse strand; genes carried by contigs; GC content and GC skew. B) Representation of the reconstruction of genomic regions for methanogens from contig overlaps. *M. formicicum* SMSP used as the genome reference is represented in black, *Methanoregula boonei* (CP000780) in black grey. The accession number is indicated under each gene with a dash replacing WP_01528. The name of the genes in multiple copies is represented in color. The genes in light-red in contigs indicate substitution in comparison to reference while the genes in pink indicate a position change. HP: hypothetical protein.

Uncorrected

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

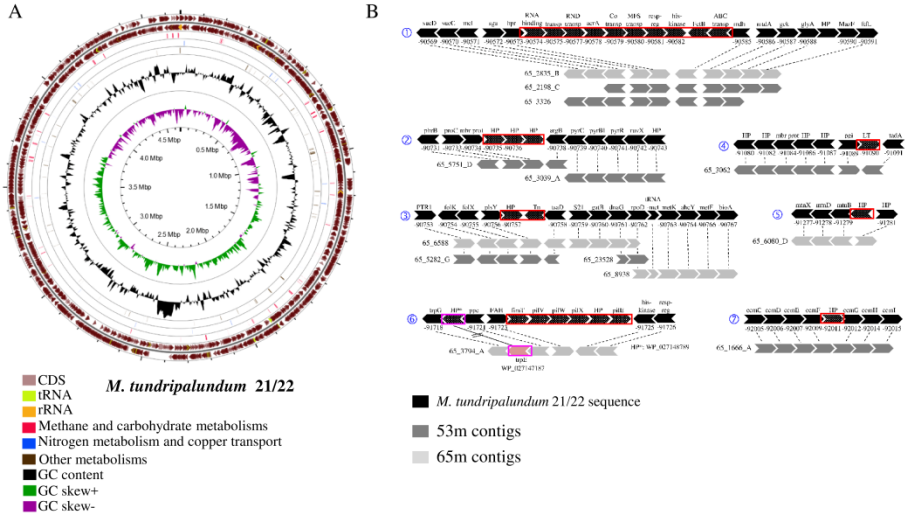



**Figure 7. Graphical representation of contig mapping onto *Methylobacter tundripaludum* 21/22 chromosome used as genome reference (JMLA01000001).** A) Circular map of *M. tundripaludum* 21/22 chromosome performed with a CGview comparison tool (Grant and Stothard, 2008). From outer to inner circle: genes on forward strand; genes on reverse strand; genes carried by contigs and related to methane and carbohydrate metabolisms; CDSs related to nitrogen metabolism and copper transport; all remaining CDSs; GC content and GC skew. B) Representation of contigs with a different gene organization in comparison to reference. Genes are represented by arrows. Red rectangles indicate genes present on reference genome but not in contigs. Pink rectangles indicate a gene substitution at one position. Contigs isolated from 53 m depths are represented in black grey and those isolated from 65 m depth in light grey. Genome reference is represented in black. Accession number is indicated under each gene with a dash replacing WP_0068. HP: hypothetical protein.

**Figure conclusion: Representation of the principal processes related to nitrogen, sulfur and methane metabolisms identified through metagenomics as well as microorganisms potentially involved in each.** The sediment is saturated with methane and ammonium which diffuse toward the top before to be captured by anaerobic and aerobic microbial groups that use them for growth. The blue arrows indicate the methane transport by diffusion and its transformation processes. The ammonium transport and transformations are indicated by brown arrows. The red arrows indicate the different steps of the sulfur cycle identified in this study. Grey arrows indicate oxygen diffusion. **1**, Different phylotypes of methanogenic archaea all related to the genus *Methanoregula.* They could produce methane through hydrogenotrophic pathway. **2**, Methanotrophic bacteria belonging to the genus *Methylobacter*. The distribution of the related-phylotypes could be dependent on the depth and oxygen conditions. **3**, Methylotrophic bacteria related to the genus *Methylotenera*. They could consume methanol potentially released by methanotrophs. **4**, Ammonium oxidizing bacteria belonging to *Nitrosomonadales*. **5**, Nitrite oxidizing bacteria belonging to the phylum

Uncorrected

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

*Nitrospirae*. **6** and **7**, Sulfur-oxidizing bacteria belonging to *Beta-proteobacteria* (6) and *Epsilon-proteobacteria* (**7**). **8**, Sulfate reducing bacteria. Although several genera were identified, the dominant taxon was related to the genus *Desulfobacca*.

Uncorrected Proof

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

**Table 1. Sequencing statistics of free-assembly and assembled metagenomes.**

| Free-assembly metagenomes | | 4 m | 53 m | 65 m | 80 m |
|---|---|---|---|---|---|
| Raw datasets | Number of reads | 108,099,890 | 120,529,997 | 123,680,966 | 95,654,671 |
| High quality reads | Number of reads after QC control | 102,305,570 | 114,673,466 | 119,466,504 | 84,488,684 |
| | Total Gbp | 10.1 | 11.3 | 11.7 | 8.1 |
| Sub-sampled datasets treated with MG-RAST (V3) | MG-RAST ID | mgm4580845.3 | mgm4580848.3 | mgm4580846.3 | mgm4580847.3 |
| | Number of ramdomly selected reads | 2,000,000 | 2,000,000 | 2,000,000 | 2,000,000 |
| | Number of reads after QC control | 1,967,724 | 1,986,645 | 1,990,038 | 1,915,809 |
| | Total Mbp | 194.9 | 196.7 | 194.9 | 187.7 |
| | Mean sequence length (bp) | 99±6 | 99±6 | 98±8 | 98±8 |
| | Mean GC ratio (%) | 51±12 | 51±11 | 57±12 | 57±11 |
| | Reads with predicted feature (%) | 93.3 | 92.8 | 92.3 | 92.3 |
| | Number of predicted proteins | 1,513,681 | 1,750,233 | 1,767,635 | 1,686,082 |
| | Number of identified proteins | 333 | 407 | 436 | 492 |
| | Functionally | 79.9 | 78.8 | 77.8 | 79.1 |

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

| | | | | | |
|---|---|---|---|---|---|
| | assigned proteins (%) | | | | |
| | Taxonomy classified reads (%) | | | | |
| | Bacteria | 93.8 | 98.0 | 94.2 | 95.4 |
| | Archaea | 0.1 | 0.1 | 3.7 | 3.1 |
| | Eukarya | 3.8 | 0.7 | 1.3 | 0.7 |
| | Viruses | 4.9 | 1.0 | 0.6 | 0.6 |
| | Unclassified | 0.1 | 0.1 | 0.2 | 0.1 |
| | Number of rRNA genes (16S/18S and 23S/28S) | 249 | 419 | 381 | 298 |
| | Bacteria (%) | 94.4 | 99.1 | 93.2 | 91.3 |
| | Archaea (%) | 0 | 0 | 3.7 | 6.4 |
| | Eukarya (%) | 4 | 0.9 | 3.1 | 1.6 |
| | Not affiliated | 1.6 | 0 | 0 | 0.7 |
| Assembled metagenomes | | 4 m | 53 m | 65 m | 80 m |
| Contiguous sequences | MG-RAST ID (preprocess passed) | mgm4557753.3.100 | mgm4581106.3.100 | mgm4581104.3.100 | mgm4557752.3.100 |
| | Number of contigs | 287,042 | 474,901 | 512,204 | 413,815 |
| | Total Gbp | 0.16 | 0.47 | 0.54 | 0.17 |
| | Largest contig | 5 | 299,47 | 290,012 | 4,146 |
| | N50 | 1,336 | 1,694 | 1,97 | 1,025 |

Uncorrected Proof

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018

**Table 2. Contig statistics for methanogens and methanotrophs.**

Table 2. Contig statistics for methanogens and methanotrophs.

| | Methanogens | Methanotrophs |
|---|---|---|
| Number of selected contigs | 386 | 312 |
| Assembly size (bp) | 271 065 | 593 276 |
| Mean G+C ratio | 54.1% | 41.3% |
| Minimum contig length (bp) | 229 | 236 |
| Maximum contig length (bp) | 6 143 | 12 298 |
| Mean contig length (bp) | 718 | 1 901 |
| N50 of contigs | 777 | 2 730 |
| N90 of contigs | 384 | 907 |
| Number of rRNA genes | 6 | 7 |
| Number of tRNA genes | 3 | 1 |
| Number of CDS | 526 | 622 |
| Mean G+C ratio of CDS | 54.7% | 41.9% |
| Mean CDS lentgh (bp) | 491 | 875 |
| Number of unique proteins | 77 | 321 |
| Mean G+C ratio of intergenic regions | 44% | 33.9% |

bp: base pair; CDS: coding DNA sequence

Uncorrected Proof

Downloaded from https://academic.oup.com/femsec/advance-article-abstract/doi/10.1093/femsec/fiy183/5092586 by St Francis Xavier University user on 10 September 2018