UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

CULTURE-INDEPENDENT CHARACTERIZATION OF METHANE-OXIDIZING MICROBIAL COMMUNITIES IN UNEXPLORED EXTREME ENVIRONMENTS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

CHUANG LI
Norman, Oklahoma
2021

CULTURE-INDEPENDENT CHARACTERIZATION OF METHANE-OXIDIZING MICROBIAL COMMUNITIES IN UNEXPLORED EXTREME ENVIRONMENTS

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF MICROBIOLOGY AND PLANT BIOLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Lee R. Krumholz, Chair

Dr. Bradley S. Stevenson

Dr. Krithi. Sankaranarayanan

Dr. Ingo B. Schlupp

Dr. Boris. Wawrik

ACKNOWLEDGEMENTS

Coming to a foreign country to pursue a Ph.D degree has been one of my hardest decision, but the process has turned out to be the most meaningful experience in my life. It has been a long journey of learning how to be an eligible scientific researcher, with full of challenges but rewarding my life and future career. Here, I would like to thank the people who have supported and helped me walk throughout my Ph.D program.

First, I would like to thank the University of Oklahoma and the Department of Microbiology and Plant Biology for offering me the opportunity to pursue my Ph.D degree. Specially thanks to the department for offering me the TA jobs for four year, that allows me to support myself and continue on research.

I give greatest thanks to my primary advisor, Lee Krumholz, where my strongest support comes from. I will be forever grateful that he adopted me into his lab and gave me the opportunity to participate in these wonderful research projects. He is always being around to provide me guidance and encouragement, without which I could not have completed my dissertation.

I greatly appreciate the input and feedback from my other current committee members: Bradley S. Stevenson, Krithi. Sankaranarayanan, Ingo B. Schlupp, Boris. Wawrik, and previous member Laura E. Bartley. Their support and professional suggestions have guided me toward the destination of the doctoral degree. I want to specially thank Dr. Sankaranarayanan for the time he spent mentoring me on data analysis.

Thanks to the members of the Krumholz lab, who have made the past few years a very special time in my life. I had the pleasure to mentor three talented and dedicated

undergraduate student MaJoi A. Moore, Hannah G. Bowen and Kyrah F. Kotary, who have helped with so much laboratory work and is greatly appreciated.

Lastly, I must give thanks to my wife, Tian Yuan, gave me the greatest support in life. Many thanks to my parents who are always there cheering me up in the background, and my son who made this journey enjoyable. Especially during the pandemic period, their support allows me to work smoothly and more efficiently from home.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 2

CHAPTER 3

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

ABSTRACT

Methane is the second most abundant greenhouse gas in atmosphere. With increasing emissions and more effective in trapping heat than CO_2, it contributes up to ~20% of total global warming. Methanotrophs playing a critical role in mitigating the amount of methane released to atmosphere, have received attention from the scientific community. However, the distributions of methanotrophs in extreme environments are still poorly understood, but where their roles are likely significant: a) They can be an overlooked sink for methane; b) They may influence the in-situ carbon flow and community-wide functions by serving as the primary producer; c) Novel extremophilic methanotrophs have the potential to be used as bioremediation and biomanufacturing agents with novel metabolic potentials. In an effort to understand better the methanotrophic ecology in unexplored extreme environments, we investigated the composition, diversity, abundance and interacting relationships of methane-oxidizing microbial communities using culture-independent methods in two ecosystems encountered with combination of multiple environmental stresses: including cyanobacterial bloom aggregates (an eutrophic ecosystem with unusual limnological extreme conditions in surface freshwaters including fast-fluctuating oxygen and pH gradients, strong light intensity input, cyanotoxins-tainted poor water quality) and Sanford Underground Research Facility (SURF) in the former Homestake gold mine (an oligotrophic ecosystem with unusual geo-physicochemical extreme conditions in terrestrial subsurface including higher pressure, deficient oxygen, toxic metals, high temperature and no light input).

Fueled by excessive nutrient input and global warming effect, cyanobacterial harmful algal blooms (cyanoHABs) are occurring with increasing frequency around the world, which can severely impact the health of humans and threaten the sustainability of the global freshwaters through releasing toxins and causing anoxia. Although eutrophication and the occurrence of cyanobacteria in oxic surface waters have been positively associated with methane emissions in lakes, the linkage to cyanobacterial blooms has gained little scientific attention. Here, *Microcystis*-blooms aggregate samples were collected from ten lakes across four continents and eight countries. Given the roles of methanotrophs in reducing methane emissions and modulating carbon flow, methanotrophic communities associated with floating *Microcystis* aggregates were surveyed through sequencing of the marker genes. Our results indicate methanotrophs are present in all samples and relatively enriched in three of the ten lakes, with the majority associated with gammaproteobacterial lineages. Methanogenic archaea as the potential methane source in aggregates are also detected in half of the ten lakes. We further quantified the abundances of the three co-occurring microbial groups (cyanobacteria, methanogens and methanotrophs) to infer their relationships at a functional level. Abundance of methanotrophs is strongly positively correlated with those of methanogens, suggesting coupled activities of methanogenesis/methanotrophy. We are using the commonality and correlation across disparate geographies as a first step in assuming the functional connections, but direct measures of activities will be required to address in detail the carbon flow throughout the entire microbial ecosystem. To our knowledge, this is the first study specifically looking into cyanobacteria-associated methanotrophs, which could act as an efficient methane sink for methane produced by the "methane paradox"

phenomenon in oxic surface waters and lead to diversified carbon flow within the aggregate.

Chemolithoautotrophy is generally thought of as a primary carbon and energy source for maintenance of the subsurface microbial ecosystems. As methane is commonly encountered in the deep subsurface biosphere, methanotrophy may play a fundamental role that is comparable to chemolithoautotrophy. However, methanotrophic ecology is particularly poorly characterized in subsurface environments. To comprehensively evaluate the distribution of aerobic methanotrophic population, we extracted the DNA from deeply circulating groundwaters and various types of mine shaft materials in SURF. High-throughput amplicon sequencing of marker genes reveal that methanotrophic populations (mainly classified as *Methyloocccus*) constitute up to ~27% of bacterial community in biofilms collected from mine shaft walls, and a high diversity of methanotrophic lineages are present in subsurface habitats. Meanwhile, several uncultivated methanotrophic lineages unreported in previous subsurface studies are observed here. We further conducted in-depth phylogenetic analysis of these two biomarkers, and the results reveal that several mine OTUs have closest relatives from other subsurface habitats, of which several groups form into a subsurface-colonizing monophylotypes, implying these SURF methanotrophs are indigenous subsurface inhabitants. The composition of bacterial communities is moderately similar between groundwaters and biofilms collected from the same depth level. Type Ib methanotrophs are likely to serve as the primary producer in biofilms as supported by the correlation patterns observed by network analysis. Finally, we successfully reconstructed ten

methanotrophic genomes from metagenomic sequences, uncovering various degree of novel methanotrophic strains were indeed present in subsurface biosphere through AAI analysis and some unusual features (e.g. with possession of the *pmoA2* and *mmoX* genes) are potentially involved in adapting to extreme conditions. To our knowledge, this is the first study specifically looking into the distribution of methanotrophs in non-cave subsurface ecosystem, where diverse, abundant and novel subsurface-adapted indigenous aerobic methane oxidizing bacteria are discovered.

CHAPTER 1: INVESTIGATING METHANOTROPHIC ECOLOGY IN EXTREME ENVIRONMENTS

Chuang Li

BACKGROUND

Methane, known as the second most abundant greenhouse gas, contributes about 20% of the total global warming (Knittel and Boetius, 2009). Although total emission amount is less than carbon dioxide, methane is about 25-fold more effective in trapping infrared radiation than carbon dioxide on a per mole basis (IPCC, 2013). The amount of methane emitted to the atmosphere is determined by the balance of source and sinks. The global methane emission budget is about ~600 Tg per year, with the majority of methane derived from decomposition of organic matter in anoxic environments by microbial methanogenesis. It has been recognized that microbially mediated methane oxidation (methanotrophy) acts as an essential biofilter for consumption of more than half of the methane produced each year before it reaches the atmosphere (Reeburgh, 2003), while methane escaping into the atmosphere is further photochemically oxidized by hydroxyl radicals. After maintaining a relative stable level for about ten years in the 1990s, the methane emissions began to surge (Hartmann et al., 2013), fueled by anthropogenic associated sources including ruminant animals, rice paddy field, fossil fuel burning, landfills and urban waste (IPCC, 2013). The current elevation of the atmospheric methane concentration has invigorated interest of researchers to explore the methane oxidizing microbes (methanotrophs) inhabiting in these man-made environments, as well as other natural ecosystems (including lakes, marine, wetlands, soils), where methanotrophs might play critical roles in mitigating the amount of methane released to atmosphere. Besides as an important methane sink, methanotrophs can be used as appealing bioremediations agents, and are highly suitable for breakdown of organic pollutants, such as trichloroethylene and aromatic compounds, in methane-containing

environments, as their key enzymes involved in methane oxidation have a broad substrate range (Jiang et al., 2010). Furthermore, a broad range of organic molecules, including single-cell protein, polyhydroxybutyrate (PHB), organic acids, pigments, vitamins, and lipids, can be generated from methane by methanotrophs, which provides a low-cost solution for sustainable manufacture of high-value bioproducts (Strong et al., 2015). Due to their importance for climate, bioremediation and biomanufacturing, methanotrophs have been extensively studied with regard to their ecological distribution, diversity, physiology and genomics. However, there are many unexplained features of the atmospheric methane dynamics and limited applications of methanotrophy-based biomanufacturing, as our understandings of the metabolism of methanotrophs are still superficial and mainly rely on a handful of cultivated strains. The cultivation of methanotrophs remains difficult, and one of the main reasons is that methanotrophs intend to establish mutualistic and synergistic interactions with associated community members for survival and proliferation, which relationships are hard to replicate in the laboratory. Their tightly interactions are widespread in complex environmental settings (Ho et al., 2016), with various modes of promotive influences discovered in synthetic cultivation studies, including exchanges of nutrients and volatile organic matters (Iguchi et al., 2011; Ho et al., 2014; Veraart et al., 2018).

Microbial methane oxidation

Aerobic and anerobic methane oxidation processes can occur in natural environments depending on the availability of oxygen. These two processes are carried out by methane-oxidizing bacteria (MOB) and anaerobic methanotrophic archaea

(ANME) under aerobic and anerobic conditions respectively, and utilize completely different pathways. For the aerobic methane oxidation pathway, although methane is one of least reactive organic molecules due to the strong C-H bond (Blanksby and Ellison, 2003), MOB can overcome the high activation energy by incorporation of oxygen to produce methanol, catalyzed by methane monooxygenases (MMO) (Semrau et al., 2010). Two forms of MMO are known: sMMO (cytoplasmic-soluble MMO), only found in small group of MOBs, and pMMO (particulate MMO), widely distributed in almost all MOBs and bounded in specialized internal membrane structures, called ICMs (intracytoplasmic membranes) (Semrau et al., 2010). The possession of unique methane activation enzyme distinguishes MOB from methylotrophic bacteria, which can grow on a number of C1 compounds. Methanol is then converted to formaldehyde by periplasmic PQQ-dependent methanol dehydrogenase (MDH). Formaldehyde is further oxidized to formate via the methanopterin (H4MPT) route and folate (H4F) route. Formate is finally oxidized to carbon dioxide by formate dehydrogenase (FDH). For the majority of MOBs, the methane-derived carbon can be assimilated into cellular biomass via ribulose monophosphate (RuMP) cycle or serine cycle at the level of formaldehyde or formate (transformed to methylene tetrahydrofolate), respectively (Hanson and Hanson, 1996).

In contrast, ANME classified within the archaeal phylum Euryarchaeota conduct anaerobic methane oxidation through a complete reverse methanogenesis pathway with using non-oxygen elements (e.g., nitrate, sulfate and metals) as terminal electron acceptor (Cui et al., 2015). After about 20 years for the discovery of aerobic methanotrophy, members of clades ANME-1, ANME-2a/b/c, and ANME-3, in syntrophic cooperation

with delta-proteobacterial sulfate-reducers, were shown to couple anaerobic oxidation with sulfate reduction in anoxic marine sediments (Boetius et al., 2000). Without associated bacterial consortia, representatives belonging to ANME-2d have been detected to independently couple anaerobic oxidation of methane to the reduction of nitrate to nitrite, with about 10% of which further reduced to ammonium (Haroon et al., 2013; Ettwig et al., 2016). As revealed recently in 2016, ANME-2d could be a versatile methanotrophic lineage and is able to perform metal-dependent anaerobic oxidation of methane using ferric iron or manganese as alternative electron acceptors (Ettwig et al., 2016; Leu et al., 2020). To date, there is no pure culture strains for ANME methanotrophs. Although most ANME lineage members are restrictedly to be present in mesophilic or psychrophilic marine environments (Cui et al., 2015), ANME-2d consortia and their activity are now more frequently detected in terrestrial ecosystem, particularly in low-sulfate, nitrate-rich or metal-rich freshwater sediments (Haroon et al., 2013; Weber et al., 2017; Leu et al., 2020). The methyl-coenzyme M reductase (MCR) is involved in catalyzing the last step of methyl group reduction to methane within methanogenic pathways, while all ANME clades utilizes a homologue of MCR of methanogens to activate methane for formation of methyl-CoM, which is further metabolized into carbon dioxides (Evans et al., 2019). The associated *mcrA* gene is typically used as functional biomarker for investigating methanogenic archaea and ANME communities. Besides Euryarchaeota, recent studies discovered MCR-like enzymes were present in several other archaeal phyla (e.g. Bathyarchaeota and Verstraetearchaeota) indicating anerobic methane metabolisms might be phylogenetically widespread in domain Archaea (Evans et al., 2019).

Taxonomic and phylogenic classifications of methanotrophs

To date, the described MOBs are confined to three bacterial phyla including Proteobacteria (classes alphaproteobacteria and gammaproteobacteria), Verrucomicrobia and candidate phylum NC10. The proteobacterial MOBs displaying the largest taxonomic diversity (> 25 defined genera) have been extensively studied and isolated from circumneutral pH environments. The known diversity of MOBs was expanded by isolation of obligately acidophilic Verrucomicrobial genera, *Methylacidiphilum* (Op den Camp et al., 2009) and *Methylacidimicrobium* (van Teeseling et al., 2014), which can assimilate carbon through fixation of carbon dioxide. The enrichment culture of Candidatus *Methylomirabilis oxyfera* (*M. oxyfera*), belonging to NC10 phylum, was obtained from anoxic canal sediments, that showed the anaerobic oxidation of methane coupled with nitrite reduction (Ettwig et al., 2008). *M. oxyfera* responsible for this process encodes, transcribes, and expresses all the necessary genes for aerobic oxidation of methane and generates its own oxygen intracellularly by reductively disproportionating nitrite via nitric oxide into dinitrogen gas and oxygen (Ettwig et al., 2010). The key enzyme pMMO for the methane activation is confined to be arranged in ICMs in most MOBs, whereas it is surprisingly located at the cytoplasmic membrane in *M. oxyfera* cells which lacks ICMs (Wu et al., 2012). Regardless of the diversity, all MOB members from these three phyla employ the same typical aerobic methane oxidation pathway by scavenging oxygen either from the atmosphere (Proteobacterial and Verrucomicrobial MOBs) or through intracellular generation (NC10 MOBs), while fixing the carbon into biomass through different assimilation pathways. Most proteobacterial

MOBs incorporate C1 compounds via RuMP or serine pathways, while MOBs within phyla Verrucomicrobia and NC10 utilize the Calvin-Benson-Bassham (CBB) cycle.

Taking advantage of the successful cultivation and extensive characterization of MOBs through the 1970s and 1980s, MOBs are initially divided into three main groups (type I/II/X) based on physiological, morphological, ultrastructural, and chemotaxonomic properties. The most distinct properties between type I and type II MOBs are arrangement of ICMs as vesicular discs (type I) or alignment of ICMs to the cell periphery as paired membranes (type II), assimilation of carbon through the RuMP pathway (type I) or serine cycle (type II), and the predominance of C16 (type I) or C18 (type II) signature fatty acids as a major membrane lipid, as well as the capability of nitrogen fixation and the formation of resting stages (Hanson and Hanson, 1996). The third group of MOBs type X is further resolved from type I MOBs with having the following unique features: higher genomic G+C content, the presence of ribulose-1,5-bisphosphate carboxylase, better capability of fixing nitrogen or higher optimum growth temperatures (Hanson and Hanson, 1996). Nonetheless, as many MOBs genera and species are discovered and characterized in recent years, the criteria of this classification system do not guarantee us to clearly and accurately differentiate these three groups. Despite the fact that the carbon assimilation pathways remains the main distinct feature, some other characteristics are no longer exclusively harbored by one or the other group (e.g. membrane fatty acid specific for type I MOB are also discovered in type II (Dedysh et al., 2007)). The group terms are still widely used in the literature, but adapted to the increasing diversity of MOBs (see next paragraph).

Although the traditional classification system is an integrated reflection of morphological, biochemical, and physiological properties of MOBs, the phylogenetic analyses of their marker genes reflect this classification as well from an evolutionary perspective. Type I and type X MOBs belong to the Gammaproteobacteria, while type II to the Alphaproteobacteria. In addition, the recently proposed phylogeny-based classification system includes more pure cultures, as well as enriched cultures and uncultivated lineages represented by environmental marker gene sequences, thus it provides better classification resolution of MOBs (Knief, 2015). Phylogenetic analysis reveals traditionally classified type I and type X gamma-MOBs fall within the family *Methylococcacea*, but form into two distinct subclusters. After reconsidering the phylogenetic relatedness and biochemical characteristics, type X MOBs within *Methylococcacea* should be reclassified under type I group, and now this clade is referred as type Ib. The remaining type I MOBs within *Methylococcacea* are adapted as belonging to type Ia clade. Type II alpha-MOBs are further resolved into two families *Methylocystaceae* (type IIa) and *Beijerinckiaceae* (type IIb). Several recently delineated isolates within family *Methylothermaceae*, harbor biochemical features of type I and II MOBs and are placed between this two groups in the phylogenetic tree, thus this family corresponds to the newly proposed type Ic clade. To be updated in parallel with the phylogeny-based typing, Verrucomicrobial MOBs are referred as type III clade. It is important to note that each subgroup (type Ia/Ib/Ic or type IIa/IIb) contains either cultivated lineages with having representatives isolates or uncultivated lineages purely consisting of environmental sequences. Some deeply-branching lineages are proposed to

be clustered into type Id clade when discussing classification based on functional marker gene phylogeny (Knief, 2015). To avoid the inconsistencies encountered in single gene based phylogeny, the updated genome-based taxonomic classification proposes gamma-MOBs are divided into three families: *Methylomonadaceae* (type Ia), *Methylococcaceae* (type Ib) and *Methylothermaceae* (type Ic), while the alpha-MOB can be merged into one family *Beijerinckiaceae* (type IIa/IIb) (Parks et al., 2018).

Identifying molecular markers

The knowledge of phylogenetic diversity of MOBs is largely derived from culture-independent studies through recovery of their marker genes from environmental samples. The most frequently targeted marker gene is the 16S rRNA gene, and taking advantage of its large database, a variety of taxon-specific primers and probes have been designed and applied in order to detect different groups of MOBs (McDonald et al., 2008). A complementary choice for detection of MOBs is targeting their functional marker genes, which can narrow down the investigation to the methanotrophic populations and thus enable a higher sensitivity for detecting rare MOBs in complex environmental settings (e.g., bloom aggregates). Specifically, methane monooxygenases (pMMO and sMMO) involved in activation of methane are unique to MOBs, thus their respective encoding genes *pmoA* and *mmoX* are particularly used for molecular investigation of MOBs (Hanson and Hanson, 1996). The whole pMMO enzyme is encoded by a three-gene operon (*pmoCAB*) responsible for synthesis of three integral membrane subunits, while sMMO is encoded by a six-gene operon involved in encoding three subunits of the hydroxylase (*mmoXYZ*), the regulatory protein (*mmoB*), the

reductase (*mmoC*), and the putative auxiliary protein (*mmoD*) (Hanson and Hanson, 1996). The pMMO is present in all MOBs with exceptions among type IIb genera *Methylocella* and *Methyloferula*, whereas sMMO only occurs in certain MOBs, thus the *mmoX* sequences are much less frequently detected than *pmoA* sequences. A couple of other functional marker genes that are not unique to MOBs but involved in methylotrophic metabolism (e.g., *mxaF* gene, encoding the large subunit of the methanol dehydrogenase) have also been used to detect MOBs in environmental samples (Kolb and Stacheter, 2013).

The phylogenies derived from *pmoA* gene are largely congruent with those of the 16S rRNA gene, which allows to draw consistent conclusion about the phylogenetic placement of putative MOBs (Knief, 2015). Nevertheless, there are some exceptions about this congruency with the presence of paralogous *pmoA* gene copies. Many MOBs contain multiple copies of the *pmoCAB* operon and they initially appeared to be nearly identical, whereas two very different copies of the *pmoCAB* were recently identified in several type IIa strains (Dunfield et al., 2002; Tchawa Yimga et al., 2003). For example, the *pmoA2* gene located in the second operon is shown to share low identity (~73%) with the closet conventional *pmoA1* gene (Dunfield et al., 2002). Subsequently, up to four distinct *pmoA* gene copies are found in Verrucomicrobial MOBs (Op den Camp et al., 2009). Similar to the occurrence of *pmoA2*, the *pxmA* copy, further distinctly related to conventional *pmoA* sequences, has been detected in several type Ia strains (Tavormina et al., 2011). The function of these paralogous *pmoA* copies remain largely unknown. Another exception is that the atypical *pmoA* sequences of the filamentous gamma-MOB

Crenothrix polyspora is highly divergent from those of all other gamma-MOBs (Knief, 2015).

Although the pMMOs of Verrucomicrobial and NC10 MOBs share a similar architecture and conserved function with canonical pMMOs in proteobacterial MOBs, but their *pmoA* genes are highly divergent from each other (Knief, 2015), and there is lack of common *pmoA* primer targeting all MOBs groups. Three primer sets (A189/A682, A189/mb661, A189/A650) with various degrees of targeting bias are most frequently used (only one is typically used in a study) for detecting proteobacterial MOBs, while specific primers and nested PCR strategy are needed and suggested to amplify *pmoA* genes of Verrucomicrobia and NC10 (Knief, 2015). The primers and probes designed for targeting single marker genes has been incorporated into different molecular techniques (including clone sequencing, DGGE, FISH, real-time PCR, stable isotope probing, microarrays, high-throughput sequencing) to uncover the distribution, diversity, abundance and activity of MOBs in natural ecosystems, including those with extreme conditions (McDonald et al., 2008).

Evidence for methanotrophs in explored extreme environments

Although most aerobic methanotrophs are mesophilic (optimal 10-35°C) and neutrophilic microbes, several isolates were retrieved from various extreme habitats and were described to be tolerant of a broad range of pH (1-11), temperature (0-72°C), and salinities (up to 30%) (Dunfield, 2009) and including adaptation to oxygen-depleted condition. Thermophilic methanotrophs stains are reported to be within three families:

Methylothermaceae (type Ic) and *Methylacidiphilaceae* (type III) and *Methylococcacea* (type Ib). The first thermophilic methanotroph, *Methylothermus* strain HB within *Methylothermaceae*, was isolated from underground hot springs in Hungary and was reported to grow at temperature up to 72℃ with optimally at 62-65℃ (Bodrossy et al., 1999). Until now, this is the only methanotrophic strain having a growth optimum higher than 60℃, but unfortunately, this strain was unable to be maintained and does no longer exist. Subsequently, two related strains within the same genus, *Methylothermus thermalis* (37-67℃) (Tsubota et al., 2005) and *Methylothermus subterraneus* (37-65℃) (Hirayama et al., 2011), respectively isolated from a Japanese hot spring and a gold mine geothermal water, can grow optimally above 55℃. Three thermophilic strains belonging to *Methylacidiphilaceae* with optimum temperature 55-60℃ were obtained from geothermally influenced soils and springs (Op den Camp et al., 2009). The type Ib thermophilic strain *Methylocaldum szegediense* (Bodrossy et al., 1997) and thermotolerant strain *Methylococcus capsulatus* (Whittenbury et al., 1970) can respectively grow at temperature up to 62℃ and 50℃. Molecular surveys of marker genes have been done in high-temperature environments, and most of the retrieved gene sequences are tightly related to *Methylacidiphilum*, *Methylocaldum* and *Methylothermus* lineages. The rates of methane oxidation have been recorded by gas chromatography in geothermally heated soils and hot springs across different regions, and are correlated to the presence of described methanotrophs (Kizilova et al., 2014; Sharp et al., 2014). Isotope-labeled methane has also been observed to be assimilated into biomass and carbon dioxide, that confirms the activity of methanotrophs in the thermophilic ecosystem (Kizilova et al., 2014; Sharp et al., 2014). Thermophilic methanotrophs are

thought to have thermal adaptions for surviving under elevated temperatures, including accumulation of thermolytes (e.g., polyamines and sucrose) to cope with heat stress, and possession of long-chain fatty acids with higher melting point to promote membrane stability (Medvedkova et al., 2007). In particular, the type Ib thermophilic and thermotolerant strains can increase the activity of cytochrome c peroxidase to relieve the impact of increased reactive oxygen species produced under high temperature (Medvedkova et al., 2009).

Relatively fewer psychrophilic methanotrophs were isolated from permanently cold terrestrial environments. The type Ia strain *Methylobacter psychrophilus* is the most psychrophilic methanotroph (3.5-20℃) with an optimum growth temperature at 5-10℃ and was isolated from a tundra soil in Siberia (Omel'chenko et al., 1996). Another psychrophilic type Ia member *Methylosphaera hansonii* has a higher optimum growth temperature (10-13℃), but it is able to grow well at 0℃, which is the lowest growth temperature ever reported in methanotrophs (Bowman et al., 1997). The psychrotolerant type IIb species *Methylocella palustris*, isolated from Siberian peat, prefers the temperatures of 15-25℃ and is the common methanotrophic member in northern wetlands (Dedysh et al., 1998). Although no isolates have been retrieved from cold marine environments, biomarkers of potential psychrophilic type Ia methanotrophs are detected in deep-sea mussels around seeps and hydrothermal vents (Duperron et al., 2007).

Acidophilic methanotrophs are not found in Gammaproteobacteria yet, but more common in Verrucomicrobia and Alphaproteobacteria. Several species within family *Methylacidiphilaceae* in phylum Verrucomicrobia are thermophilic (see above), and

meanwhile these are also extremely acidophilic, with optimum pH growth range among 2.0 and 4.3 (Op den Camp et al., 2009). Members in this family appeared to be exclusively present in geothermal habitats especially associated with acidic conditions. Our poor knowledge about their geographic distribution could be due to the atypical *pmoA* genes they possess are hardly detected by the regular primers. Two mildly acidophilic methanotrophs type IIb species *Methylocella palustris* and *Methylocapsa acidophila* were both isolated from acidic peatlands and have similar pH range (4.2-7.2) for growth, but they differ in other aspects (Dedysh et al., 1998; Dedysh et al., 2002). Methanotrophs that are adapted to high pH are mainly found within type Ia genus *Methylomicrobium*. Several strains of this genera were isolated from soda lakes and are able to grow at pH values up to 11 with the help of their special phenotypic properties (e.g. glycoprotein S-layers contributing to shape maintenance) (Trotsenko and Khmelenina, 2002a, b). These alkaliphilic *Methylomicrobium* strains are also halophiles, and their maximum salt tolerances are less than 8% NaCl. Synthesis of osmoregulatory compounds (e.g., ectoine) is thought to be one of their main adaption mechanisms under halophilic conditions (Trotsenko and Khmelenina, 2002a). The most salt-tolerant halophile *Methylohalobius crimeensis* (type Ic member) was isolated from hypersaline lakes and can grow at NaCl levels as high as 15% (Heyer et al., 2005). Methanotrophy indeed occurs in hypersaline lakes as revealed by a sensitive isotope labelling technique (Trotsenko and Khmelenina, 2002a, b).

Although MOBs typically flourish at aerobic-anerobic interface, many lineages have been detected in various anoxic environments. There are many mechanisms

proposed for their thrive and grow under anoxic conditions, including: methane oxidation coupled to denitrification or mineral reduction (Kits et al., 2015), methane-derived fermentation (Kalyuzhnaya et al., 2013), interaction with oxygenated phototrophs (Milucka et al., 2015). Recently, several type Ia isolates (*Methylosoma difficile* and *Methyloglobulus morosus*) were shown to grow preferably at reduced oxygen concentrations (Knief, 2015). The wide occurrence of MOBs in various extreme conditions, along with a potential of metabolic flexibility, provide a clue that there are many more MOBs present in unexplored extreme environments that await detection and isolation.

RATIONALE OF EXPLORING PLANKTONIC METHANOTROPHS ASSOCIATED WITH BLOOM

Cyanobacterial bloom aggregates

In recent years, the outbreaks of cyanobacterial harmful algal blooms (cyanoHABs), particularly those of toxic species, have been frequently recorded in global freshwater lake systems (Huisman et al., 2018). The bloom events are fueled by high nutrient input and global warming (Paerl et al., 2016). CyanoHABs can alter the trophic structures, functions, sustainability of freshwater systems through depletion of dissolved oxygen, and are threatening the health of humans, livestock, and wildlife by releasing the toxic metabolites (e.g. hepatotoxins and neurotoxins) (Huisman et al., 2018). Basically, most cyanoHABs are composed of planktonic aggregates. Within the aggregate, bloom-causing cyanobacterial species are accompanied by a diverse suite of bacterial groups (Fig. 1.1). Cyanobacteria and associated bacteria tend to exchange necessary and

complementary metabolic products from each other to form a remarkably stable interactome community (Cook et al., 2020). To date, the nutrients (including carbon) and energy flow within the interactome is still poorly understood, which hinders the development of effective solutions for mitigating cyanoHABs.

Unusual extreme conditions associated with bloom aggregates

Traditionally, we think of methane being produced in the anoxic sediments of lake, most of which would be oxidized by methanotrophs in sediment-water interface, with limited amount can escape into the surface water layer. As such, most of previous studies on planktonic methanotrophs mainly focused on the oxic-anoxic transitional water layers within the metalimnion and/or hypolimnion, with much fewer reports on the surface waters. Also, a higher light intensity encountered in surface water layers plays an inhibitory role on oxidation activity of methanotrophs (Murase and Sugimoto, 2005). Distribution of planktonic methanotrophs in undisturbed mid to bottom water layers may imply they prefer to live within environmental niches with lower turnover. However, steep physiochemical gradients occur within bloom aggregate on a small spatial scale, and such gradients can change diurnally or abruptly over a short time. For example, during daytime, the oxygen is saturated in the surrounding water of aggregate, and its concentration can decrease gradually to zero from the outside to the center of aggregate; during darkness, as photosynthesis ceases and microbial respiration increases, internal oxygen can be consumed and anoxic conditions may spread the whole aggregate (Ploug, 2008; Klawonn et al., 2015). Furthermore, the pH within aggregate has been recorded to drop from 9.0 to 7.4 in a few minutes when a light-dark shift event happens (e.g., sun

shaded by the cloud) (Ploug, 2008). Therefore, the unusual limnological extreme conditions (including low availability of methane, light inhibition, fast-fluctuating oxygen and pH) encountered in bloom aggregates seem unfavorable for growth of methanotrophs, which could be the reason there is still lack of report about bloom aggregate associated methanotrophs.

Aerobic methane oxidation coupled to oxygenic photosynthesis in nature

In theory, O_2 released by photoautotroph can be consumed by aerobic methanotrophs, which in turn can produce CO_2 needed for photosynthetic growth. Methane oxidation coupled to oxygenic photosynthesis has been verified and documented in global peatland pools between methanotrophs and waterlogged Sphagnum moss (Raghoebarsing et al., 2005). In lake anoxic waters, methanotrophy is able to be fueled by in situ oxygen generation by associated photosynthetic algae (Milucka et al., 2015). Otherwise, symbiotic interactions between methanotroph and microalgae (van der Ha et al., 2011) or Cyanobacteria (Hill et al., 2017) have also been verified and achieved in artificial laboratory settings. Such coupling activities observed in other natural habitats and lab allows us to speculate that methanotrophs-Cyanobacteria symbiotic system can occur in bloom aggregates. It is likely that methanotrophs profited from O_2 produced by Cyanobacteria. On the other hand, along with intensive photosynthetic activity during daytime, Cyanobacteria might suffer from severely undersaturated CO_2 conditions at specific time scale (Balmer and Downing, 2011), thus the possibility of benefiting from methanotrophy-derived carbon source seems likely. In addition, active methanotrophs have been broadly found on macrophytic algae from eutrophic freshwater habitats

(Yoshida et al., 2014). Even if two processes are uncoupled, such ubiquitously plant/aggregate-associated habitats could allow methanotrophs to tolerate stress, access nutrients and avoid from predation in fast-changing surface freshwater ecosystem.

Potential available methane resources for bloom aggregates

In macroscale distance, the CH_4 consumed by methanotrophs in aggregate can be up-diffused from benthic waters and sediment (Bastviken et al., 2011). During the bloom season, the enhanced eutrophication status of freshwater lakes can promote the growth of photoautotrophs, which are normally terminated by sedimentation, thus supplying plenty of organic matters to the bottom of lakes. Methanogenesis are fueled by organic matter inputs in the lake sediments, which will lead to more methane transported to lake surface waters (Zhang et al., 2021). Compared with diffusion from outside of aggregates, methane generated from inside by associated microbes just needs to be transported over only a much shorter microscale distances to deliver to methanotrophs. Particularly, methanogenic archaea are known to be present in cyanobacterial aggregate (Batista et al., 2019). There could be several other mechanisms (known as "methane paradox") responsible for microbial methane production inside of cyanobacterial aggregate, including photosynthesis-related pathway, enzymatic cleavage of the C-P bond, N_2-fixation related pathway, hypothetical demethylation pathways, which could be carried out by Cyanobacteria and/or heterotrophic bacteria (Tang et al., 2016; Bizic et al., 2020). However, these "unclassical" methanogenic pathways appeared to be much less efficient than archaeal methanogenesis with much lower methane production rate (Bizic et al., 2020).

GOALS OF STUDY ON BLOOM AGGREGATE

Based on the potential symbiotic relationship and methane resources mentioned above, we hypothesize that methanotrophs can occur within cyanobacterial bloom aggregates and, once certain environmental factors/cues are present, they will establish mutualistic and synergistic interactions with other members of the microbial community that will lead to diversified carbon flow within the aggregate. Further, we will try to determine: 1) Is the co-occurrence of methanotroph and cyanobacteria found at the global scale? 2) Is there any potential methane resource within aggregate (e.g., methanogenic archaea)? 3) If methanogens co-occur with methanotrophs, can methanotrophy be coupled to methanogenesis? The marker genes sequencing and qPCR quantification will be combined to characterize the distribution and abundance of methane-metabolic microbes within bloom aggregates collected from ten global lakes. The findings are detailed in the 2rd chapter.

RATIONALE OF INVESTIGATING EXTREMOPHILIC METHANOTROPHS IN SUBSURFACE

The significance of investigating methanotrophs in subsurface biosphere

Although atmospheric methane is mainly attributed to emissions from surface ecosystems (wetlands: 177-284; fossil fuels: 85-105; ruminants: 87-94; landfills: 67-90 Tg per year), geological sources derived from subsurface still contribute up to 75 Tg methane per year and are considered as the fifth most abundant sources of atmospheric methane (Etiope et al., 2008; IPCC, 2013; Houghton et al., 2019). Methanotrophs can act

as the major sink for methane in natural environments, if their regulator roles are not considered, these reported emission budget balances would be substantially overestimated. A broad diversity of methanotrophic communities and their oxidation rate have been well documented in surface ecosystems, that information helps with developing better model to accurately estimate the amount of methane emissions by taking account with the mitigation effect of methanotrophs. However, when estimation of global subsurface methane flux, the contribution of methane oxidation is not included, because studies about the distribution and activity of methanotrophs in subsurface ecosystem is still very limited (Kotelnikova, 2002).

On the other hand, microbes inhabiting oligotrophic habitats, normally encountered in subsurface, tend to cooperate with each other for removing thermodynamic barriers and exchanging metabolites to promote the growth and energy yield (Lau et al., 2016). As being capable utilizing methane as obligate substrate, methanotroph has better growth advantage than other microbes, and may provide methane-derived carbon to support a variety of other types of microorganisms through excretion and lysis. Along with hydrogenotrophic microbes, the traditional primary producer in lithotrophic microbial ecosystems, methanotrophs can further modulate carbon flow and influence biogeochemical processes in subsurface habitats, as they perform in surface ecosystems (Ho et al., 2016).

In parallel with its role as the food base for other organisms, methanotrophs are particularly promising platforms for converting low-cost methane into a series of value-added compounds in biochemicals and biofuels (Nguyen and Lee, 2021). The potential of methane-based biomanufacturing largely depends on the detailed understanding of methane metabolism, but such good knowledge is confined to the isolates obtained from surface environments (Nguyen and Lee, 2021). The subsurface geochemical conditions with shortage of energy and carbon sources have been suggested for favoring the growth of chemolithoautrotrophs with novel metabolic capabilities (Gregory et al., 2019). To survive in the extreme conditions, the extremophilic methanotrophs are expected to adapt to evolve and possibly acquire new and efficient pathways for carbon assimilation, which would potentially advance metabolic engineering in methanotrophs.

Therefore, in-depth investigation of distribution, diversity, abundance, physiological properties and metabolic capabilities of subsurface methanotrophic communities can provide valuable insight to their roles involved in mitigation of subsurface methane flux, and their impact on community-wide functions in lithotrophic microbial ecosystems through methanotrophy-fueled biotic interaction, as well as will pave the way for successfully isolation of industrially useful extremophilic methanotrophs.

Methanotroph can survive and be active in subsurface environments and their distribution is likely shaped by oxygen

Although methanotrophs may play a significant role in subsurface ecosystems, the next fundamental question raises that whether methanotrophs can survive and grown in such unusual environment, associated with extreme temperature, high pressure, low oxygen concentration, toxic metals, and no light input (Rastogi et al., 2009). The availability and concentrations of substrates (methane and oxygen) are considered as the basically determined factors for the occurrence of methanotrophs in natural ecosystem. For instance, the distribution of methanotrophs in lakes vary widely, but they are mostly enriched and active in niches where oxygen-methane counter gradient meets (Bastviken et al., 2008). Methane is a common and dominant gas in subsurface ecosystem with several geological formation pathways (Etiope and Sherwood Lollar, 2013). The deep crystalline rock aquifers, mines and caves are the three relatively well-studied terrestrial deep subsurface environments that can be reached through drillholes for groundwater or tunnel for attached materials (Bomberg and Ahonen, 2017). Metabolically diverse methanogens, including chemolithoautotrophic (mainly found in deeper depth) and organotrophic (mainly found in shallower depths), were prevalent in borehole fluids throughout deep crystalline rock environments and mines (Kotelnikova, 2002; Kietavainen and Purkamo, 2015). This implies geological and biogenic methane is available in subsurface environments, thus methane is unlikely to be the limited factor for the occurrence of methanotrophs. More importantly, isotope signature further indicates biogenic methane is incorporated into the biomass for *in situ* microbial community, likely derived from the activity of *in-situ* methanotrophs (Simkus et al., 2016).

Unfortunately, the marker genes of methanotrophs are only occasionally detected in oxygen-deficient borehole fluid at certain sites (Kietavainen and Purkamo, 2015). In contrast to the poor distribution in groundwaters from crystalline rock and mines, a high abundance and diversity of methanotrophs (mainly within type Ia clade) are well-described in cave samples collected at "air pockets" where has an atmosphere containing methane and oxygen (Hutchens et al., 2004; Karwautz et al., 2018). Furthermore, methanotrophic bacteria has been shown to act as the primary producer to sustain the isolated cave microbial ecosystem (Hutchens et al., 2004; Karwautz et al., 2018), and as the significant subterranean sink for cave atmospheric methane (Fernandez-Cortes et al., 2015). It is likely that the noticeably distinct oxygen concentrations could well explain the varied distribution of aerobic methanotrophs between oxygen-deficient borehole fluid (in crystalline rock and mine) and attached materials (in cave air pockets). Thus, we hypothesize that aerobic methanotrophs can grow and be active in extreme subsurface conditions with development of adaptive metabolisms as long as methane and oxygen are available, but their distributions are largely shaped by in situ oxygen gradients, with more diverse and abundant aerobic methanotrophs inhabiting in niches with higher oxygen concentrations. To test this hypothesis, we planned to target the samples in SURF mine with variable oxygen concentrations, including borehole fluid and tunnel materials. we suspected methanotrophic communities living in well-ventilated tunnel would be more abundant and diverse than those in groundwater, and their ecological impact in mines might be as significant as in reported cave air pockets (Fernandez-Cortes et al., 2015; Karwautz et al., 2018).

SURF as a new portal into subsurface biosphere is ideal for studying methanotrophs

Reports about the methanotrophic communities are much less common for mines than caves and crystalline rock aquifers, and related microbial information are mainly derived from the mines located in Japan and South Africa (Kietavainen and Purkamo, 2015). In the current study, the Sanford Underground Research Facility (SURF) in the former Homestake gold mine (South Dakota, USA) will be used as a new testbed for characterization of extreme methane oxidizing microbial communities living in deep subsurface biosphere. The Homestake gold mine is the deepest mine in the North America, and produced ~1,101 tons of gold through ~125 years of operation (Caddey et al., 1991). Since mining activities ceased in 2001, this mine was transitioned into a dedicated research facility for scientific exploration. SURF maintains 370 miles of underground tunnels that reach up to 8,100 ft below the surface ground (Fig. 1.2). The deep groundwater, circulating in the rock fractures and seeping through mine walls into the tunnel, are characterized by a wide range of geochemical components that sustain subsurface biosphere. Although the methanotrophs are not the major focus in previous studies conducted in SURF, those results imply that SURF is a suitable and ideal site for exploring methanotrophic populations. Geochemical measurements of borehole fluids and pools within the mine tunnels show that methane and oxygen are broadly detected in various SURF samples with a broad range of gradients, and thermodynamic modeling suggests aerobic methane oxidation reaction is feasible and promising in some samples (Osburn et al., 2014). Microbiological studies at SURF have revealed 16S rRNA marker gene sequences of aerobic methanotrophs are present in borehole fluids and tunnel soils (Rastogi et al., 2009; Rastogi et al., 2010; Osburn et al., 2014).

GOALS OF STUDY ON SURF

Overall, the methanotrophic ecology of the terrestrial subsurface is particularly poorly characterized in mine ecosystems. In an effort to better understand the ecological roles of methanotrophs in oligotrophic subsurface habitats, a deeply comprehensive evaluation of aerobic methanotrophic populations in SURF has been conducted as detailed in the 3rd chapter. Borehole groundwaters and different types of mine shaft materials potentially supplied with methane and oxygen were collected for this study (Fig. 3.2). The 16S rRNA gene sequencing and functional marker gene *pmoA* sequencing is combined to describe the composition of methanotrophic population in SURF samples. In-depth phylogenetic analysis of these two marker genes is used to define the potential novel uncultivated methanotrophic lineages. To understand subsurface distribution pattern from a global perspective, diversity of methanotrophs uncovered in SURF samples is compared and discussed with those in deep crystalline rock aquifers, caves and other mines. We construct the relationship between phylogenetic diversity and habitat preferences to determine 1) if specific methanotrophic members uncovered in SURF prefer to inhabit specific locations; 2) if specific methanotrophic members uncovered in SURF exclusively inhabit subsurface habitats.

Moreover, we hypothesize that the methanotrophs found in SURF are generally indigenous to subsurface environments, and those living in tunnels likely originated from the borehole fluids, which will be respectively evaluated according to the subsurface habitat origins of closet relatives and community relatedness among these samples. The biofilms attaching on the tunnel walls appear similar to biofilms described from cave air

pocket, making us to speculate that methanotrophs might be the food base in mine shaft biofilm similarly to those reported in cave biofilm. This suspected role of methanotrophs will be estimated through network analysis on biofilms, while using groundwater network as a control. In addition, we will attempt to reconstruct the genomes of methanotrophs from shot-gun genomic sequencing to confirm the presence of novel methanotrophic strains in subsurface biosphere and infer their potential subsurface-adapted traits.

REFERENCES

Balmer, M.B., and Downing, J.A. (2011) Carbon dioxide concentrations in eutrophic lakes: undersaturation implies atmospheric uptake. *Inland Waters* **1**: 125-132.

Bastviken, D., Cole, J.J., Pace, M.L., and Van de Bogert, M.C. (2008) Fates of methane from different lake habitats: Connecting whole-lake budgets and CH_4 emissions. *J Geophys Res* **113**: G02024.

Bastviken, D., Tranvik, L.J., Downing, J.A., Crill, P.M., and Enrich-Prast, A. (2011) Freshwater methane emissions offset the continental carbon sink. *Science* **331**: 50.

Batista, A.M.M., Woodhouse, J.N., Grossart, H.P., and Giani, A. (2019) Methanogenic archaea associated to Microcystis sp. in field samples and in culture. *Hydrobiologia* **831**: 163-172.

Bizic, M., Klintzsch, T., Ionescu, D., Hindiyeh, M.Y., Gunthel, M., Muro-Pastor, A.M. et al. (2020) Aquatic and terrestrial cyanobacteria produce methane. *Sci Adv* **6**: eaax5343.

Blanksby, S.J., and Ellison, G.B. (2003) Bond dissociation energies of organic molecules. *Acc Chem Res* **36**: 255-263.

Bodrossy, L., Kovács, K.L., McDonald, I.R., and Murrell, J.C. (1999) A novel thermophilic methane-oxidising γ-Proteobacterium. *FEMS Microbiology Letters* **170**: 335-341.

Bodrossy, L., Holmes, E.M., Holmes, A.J., Kovacs, K.L., and Murrell, J.C. (1997) Analysis of 16S rRNA and methane monooxygenase gene sequences reveals a novel group of thermotolerant and thermophilic methanotrophs, Methylocaldum gen. nov. *Arch Microbiol* **168**: 493-503.

Boetius, A., Ravenschlag, K., Schubert, C.J., Rickert, D., Widdel, F., Gieseke, A. et al. (2000) A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature* **407**: 623-626.

Bomberg, M., and Ahonen, L. (2017) Editorial: Geomicrobes: Life in Terrestrial Deep Subsurface. *Front Microbiol* **8**: 103.

Bowman, J.P., McCammon, S.A., and Skerrat, J.H. (1997) Methylosphaera hansonii gen. nov., sp. nov., a psychrophilic, group I methanotroph from Antarctic marine-salinity, meromictic lakes. *Microbiology* **143 (Pt 4)**: 1451-1459.

Caddey, S.W., Bachman, R.L., Campbell, T.J., Reid, R.R., and Otto, R.P. (1991) The Homestake gold mine, an early Proterozoic iron-formation-hosted gold deposit, Lawrence County, South Dakota. In *Bulletin*.

Cook, K.V., Li, C., Cai, H., Krumholz, L.R., Hambright, K.D., Paerl, H.W. et al. (2020) The global *Microcystis* interactome. *Limnol Oceanogr* **65**: S194-S207.

Cui, M., Ma, A., Qi, H., Zhuang, X., and Zhuang, G. (2015) Anaerobic oxidation of methane: an "active" microbial process. *Microbiologyopen* **4**: 1-11.

Dedysh, S.N., Panikov, N.S., Liesack, W., Grosskopf, R., Zhou, J., and Tiedje, J.M. (1998) Isolation of acidophilic methane-oxidizing bacteria from northern peat wetlands. *Science* **282**: 281-284.

Dedysh, S.N., Khmelenina, V.N., Suzina, N.E., Trotsenko, Y.A., Semrau, J.D., Liesack, W., and Tiedje, J.M. (2002) Methylocapsa acidiphila gen. nov., sp. nov., a novel methane-oxidizing and dinitrogen-fixing acidophilic bacterium from Sphagnum bog. *Int J Syst Evol Microbiol* **52**: 251-261.

Dedysh, S.N., Belova, S.E., Bodelier, P.L.E., Smirnova, K.V., Khmelenina, V.N., Chidthaisong, A. et al. (2007) Methylocystis heyeri sp. nov., a novel type II methanotrophic bacterium possessing 'signature' fatty acids of type I methanotrophs. *Int J Syst Evol Microbiol* **57**: 472-479.

Dunfield, P.F. (2009) Methanotrophy in Extreme Environments. In *eLS*.

Dunfield, P.F., Yimga, M.T., Dedysh, S.N., Berger, U., Liesack, W., and Heyer, J. (2002) Isolation of a Methylocystis strain containing a novel pmoA-like gene. *FEMS Microbiol Ecol* **41**: 17-26.

Duperron, S., Sibuet, M., MacGregor, B.J., Kuypers, M.M., Fisher, C.R., and Dubilier, N. (2007) Diversity, relative abundance and metabolic potential of bacterial endosymbionts in three Bathymodiolus mussel species from cold seeps in the Gulf of Mexico. *Environ Microbiol* **9**: 1423-1438.

Etiope, G., and Sherwood Lollar, B. (2013) ABIOTIC METHANE ON EARTH. *Reviews of Geophysics* **51**: 276-299.

Etiope, G., Lassey, K.R., Klusman, R.W., and Boschi, E. (2008) Reappraisal of the fossil methane budget and related emission from geologic sources. *Geophysical Research Letters* **35**.

Ettwig, K.F., Zhu, B., Speth, D., Keltjens, J.T., Jetten, M.S.M., and Kartal, B. (2016) Archaea catalyze iron-dependent anaerobic oxidation of methane. *Proc Natl Acad Sci U S A* **113**: 12792-12796.

Ettwig, K.F., Shima, S., van de Pas-Schoonen, K.T., Kahnt, J., Medema, M.H., Op den Camp, H.J. et al. (2008) Denitrifying bacteria anaerobically oxidize methane in the absence of Archaea. *Environ Microbiol* **10**: 3164-3173.

Ettwig, K.F., Butler, M.K., Le Paslier, D., Pelletier, E., Mangenot, S., Kuypers, M.M. et al. (2010) Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. *Nature* **464**: 543-548.

Evans, P.N., Boyd, J.A., Leu, A.O., Woodcroft, B.J., Parks, D.H., Hugenholtz, P., and Tyson, G.W. (2019) An evolving view of methane metabolism in the Archaea. *Nat Rev Microbiol* **17**: 219-232.

Fernandez-Cortes, A., Cuezva, S., Alvarez-Gallego, M., Garcia-Anton, E., Pla, C., Benavente, D. et al. (2015) Subterranean atmospheres may act as daily methane sinks. *Nat Commun* **6**: 7003.

Gregory, S.P., Barnett, M.J., Field, L.P., and Milodowski, A.E. (2019) Subsurface Microbial Hydrogen Cycling: Natural Occurrence and Implications for Industry. *Microorganisms* **7**.

Hanson, R.S., and Hanson, T.E. (1996) Methanotrophic bacteria. *Microbiol Rev* **60**: 439-471.

Haroon, M.F., Hu, S., Shi, Y., Imelfort, M., Keller, J., Hugenholtz, P. et al. (2013) Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage. *Nature* **500**: 567-570.

Hartmann, D.L., Klein, T.A.M.G., Rusticucc, M., Alexander, L.V., and Charabi, Y. (2013) Observations: atmosphere and surface. Pp. 1535 in T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J.: 1-25.

Heyer, J., Berger, U., Hardt, M., and Dunfield, P.F. (2005) Methylohalobius crimeensis gen. nov., sp. nov., a moderately halophilic, methanotrophic bacterium isolated from hypersaline lakes of Crimea. *Int J Syst Evol Microbiol* **55**: 1817-1826.

Hill, E.A., Chrisler, W.B., Beliaev, A.S., and Bernstein, H.C. (2017) A flexible microbial co-culture platform for simultaneous utilization of methane and carbon dioxide from gas feedstocks. *Bioresour Technol* **228**: 250-256.

Hirayama, H., Suzuki, Y., Abe, M., Miyazaki, M., Makita, H., Inagaki, F. et al. (2011) Methylothermus subterraneus sp. nov., a moderately thermophilic methanotroph isolated from a terrestrial subsurface hot aquifer. *Int J Syst Evol Microbiol* **61**: 2646-2653.

Ho, A., de Roy, K., Thas, O., De Neve, J., Hoefman, S., Vandamme, P. et al. (2014) The more, the merrier: heterotroph richness stimulates methanotrophic activity. *ISME J* **8**: 1945-1948.

Ho, A., Angel, R., Veraart, A.J., Daebeler, A., Jia, Z., Kim, S.Y. et al. (2016) Biotic Interactions in Microbial Communities as Modulators of Biogeochemical Processes: Methanotrophy as a Model System. *Front Microbiol* **7**: 1285.

Houghton, K.M., Carere, C.R., Stott, M.B., and McDonald, I.R. (2019) Thermophilic methanotrophs: in hot pursuit. *FEMS Microbiol Ecol* **95**.

Huisman, J., Codd, G.A., Paerl, H.W., Ibelings, B.W., Verspagen, J.M.H., and Visser, P.M. (2018) Cyanobacterial blooms. *Nat Rev Microbiol* **16**: 471-483.

Hutchens, E., Radajewski, S., Dumont, M.G., McDonald, I.R., and Murrell, J.C. (2004) Analysis of methanotrophic bacteria in Movile Cave by stable isotope probing. *Environ Microbiol* **6**: 111-120.

Iguchi, H., Yurimoto, H., and Sakai, Y. (2011) Stimulation of methanotrophic growth in cocultures by cobalamin excreted by rhizobia. *Appl Environ Microbiol* **77**: 8509-8515.

IPCC (2013) Climate Change: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge, UK: Cambridge University Press.

Jiang, H., Chen, Y., Jiang, P.-x., Zhang, C., Smith, T., Murrell, J., and Xing, X. (2010) Methanotrophs: Multifunctional bacteria with promising applications in environmental bioengineering. *Biochemical Engineering Journal* **49**: 277-288.

Kalyuzhnaya, M.G., Yang, S., Rozova, O.N., Smalley, N.E., Clubb, J., Lamb, A. et al. (2013) Highly efficient methane biocatalysis revealed in a methanotrophic bacterium. *Nat Commun* **4**: 2785.

Karwautz, C., Kus, G., Stockl, M., Neu, T.R., and Lueders, T. (2018) Microbial megacities fueled by methane oxidation in a mineral spring cave. *ISME J* **12**: 87-100.

Kietavainen, R., and Purkamo, L. (2015) The origin, source, and cycling of methane in deep crystalline rock biosphere. *Front Microbiol* **6**: 725.

Kits, K.D., Klotz, M.G., and Stein, L.Y. (2015) Methane oxidation coupled to nitrate reduction under hypoxia by the Gammaproteobacterium Methylomonas denitrificans, sp. nov. type strain FJG1. *Environ Microbiol* **17**: 3219-3232.

Kizilova, A.K., Sukhacheva, M.V., Pimenov, N.V., Yurkov, A.M., and Kravchenko, I.K. (2014) Methane oxidation activity and diversity of aerobic methanotrophs in pH-neutral and semi-neutral thermal springs of the Kunashir Island, Russian Far East. *Extremophiles* **18**: 207-218.

Klawonn, I., Bonaglia, S., Bruchert, V., and Ploug, H. (2015) Aerobic and anaerobic nitrogen transformation processes in N_2-fixing cyanobacterial aggregates. *ISME J* **9**: 1456-1466.

Knief, C. (2015) Diversity and habitat preferences of cultivated and uncultivated aerobic methanotrophic bacteria evaluated based on *pmoA* as molecular marker. *Front Microbiol* **6**: 1-38.

Knittel, K., and Boetius, A. (2009) Anaerobic oxidation of methane: progress with an unknown process. *Annu Rev Microbiol* **63**: 311-334.

Kolb, S., and Stacheter, A. (2013) Prerequisites for amplicon pyrosequencing of microbial methanol utilizers in the environment. *Front Microbiol* **4**: 268.

Kotelnikova, S. (2002) Microbial production and oxidation of methane in deep subsurface. *Earth-Science Reviews* **58**: 367-395.

Lau, M.C., Kieft, T.L., Kuloyo, O., Linage-Alvarez, B., van Heerden, E., Lindsay, M.R. et al. (2016) An oligotrophic deep-subsurface community dependent on syntrophy is dominated by sulfur-driven autotrophic denitrifiers. *Proc Natl Acad Sci U S A* **113**: E7927-E7936.

Leu, A.O., Cai, C., McIlroy, S.J., Southam, G., Orphan, V.J., Yuan, Z. et al. (2020) Anaerobic methane oxidation coupled to manganese reduction by members of the Methanoperedenaceae. *ISME J* **14**: 1030-1041.

McDonald, I.R., Bodrossy, L., Chen, Y., and Murrell, J.C. (2008) Molecular ecology techniques for the study of aerobic methanotrophs. *Appl Environ Microbiol* **74**: 1305-1315.

Medvedkova, K.A., Khmelenina, V.N., and Trotsenko, Y.A. (2007) Sucrose as a factor of thermal adaptation of the thermophilic methanotroph Methylocaldum szegediense O-12. *Microbiology* **76**: 500-502.

Medvedkova, K.A., Khmelenina, V.N., Suzina, N.E., and Trotsenko, Y.A. (2009) Antioxidant systems of moderately thermophilic methanotrophs Methylocaldum szegediense and Methylococcus capsulatus. *Microbiology* **78**: 670.

Milucka, J., Kirf, M., Lu, L., Krupke, A., Lam, P., Littmann, S. et al. (2015) Methane oxidation coupled to oxygenic photosynthesis in anoxic waters. *ISME J* **9**: 1991-2002.

Murase, J., and Sugimoto, A. (2005) Inhibitory effect of light on methane oxidation in the pelagic water column of a mesotrophic lake (Lake Biwa, Japan). *Limnol Oceanogr* **50**: 1339-1343.

Nguyen, A.D., and Lee, E.Y. (2021) Engineered Methanotrophy: A Sustainable Solution for Methane-Based Industrial Biomanufacturing. *Trends Biotechnol* **39**: 381-396.

Omel'chenko, M.V., Vasil'eva, L.V., Zavarzin, G.A., Savel'eva, N.D., Lysenko, A.M., Mityushina, L.L. et al. (1996) A Novel Psychrophilic Methanotroph of the genus Methylobacter. *Mikrobiologiya* **65**: 384-389.

Op den Camp, H.J., Islam, T., Stott, M.B., Harhangi, H.R., Hynes, A., Schouten, S. et al. (2009) Environmental, genomic and taxonomic perspectives on methanotrophic Verrucomicrobia. *Environ Microbiol Rep* **1**: 293-306.

Osburn, M.R., LaRowe, D.E., Momper, L.M., and Amend, J.P. (2014) Chemolithotrophy in the continental deep subsurface: Sanford Underground Research Facility (SURF), USA. *Front Microbiol* **5**: 610.

Paerl, H.W., Gardner, W.S., Havens, K.E., Joyner, A.R., McCarthy, M.J., Newell, S.E. et al. (2016) Mitigating cyanobacterial harmful algal blooms in aquatic ecosystems impacted by climate change and anthropogenic nutrients. *Harmful Algae* **54**: 213-222.

Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.A., and Hugenholtz, P. (2018) A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* **36**: 996-1004.

Ploug, H. (2008) Cyanobacterial surface blooms formed by Aphanizomenon sp. and Nodularia spumigena in the Baltic Sea: Small-scale fluxes, pH, and oxygen microenvironments. *Limnology and Oceanography* **53**: 914-921.

Raghoebarsing, A.A., Smolders, A.J., Schmid, M.C., Rijpstra, W.I., Wolters-Arts, M., Derksen, J. et al. (2005) Methanotrophic symbionts provide carbon for photosynthesis in peat bogs. *Nature* **436**: 1153-1156.

Rastogi, G., Stetler, L.D., Peyton, B.M., and Sani, R.K. (2009) Molecular analysis of prokaryotic diversity in the deep subsurface of the former Homestake gold mine, South Dakota, USA. *The Journal of Microbiology* **47**: 371-384.

Rastogi, G., Osman, S., Kukkadapu, R., Engelhard, M., Vaishampayan, P.A., Andersen, G.L., and Sani, R.K. (2010) Microbial and mineralogical characterizations of soils collected from the deep biosphere of the former Homestake gold mine, South Dakota. *Microb Ecol* **60**: 539-550.

Reeburgh, W.S. (2003) Global methane biogeochemistry. In: Holland HD, Turekian KK (eds). Treatise on Geochemistry, Vol. 4: The Atmosphere. Amsterdam: Elsevier. 1-25.

Semrau, J.D., DiSpirito, A.A., and Yoon, S. (2010) Methanotrophs and copper. *FEMS Microbiol Rev* **34**: 496-531.

Sharp, C.E., Martinez-Lorenzo, A., Brady, A.L., Grasby, S.E., and Dunfield, P.F. (2014) Methanotrophic bacteria in warm geothermal spring sediments identified using stable-isotope probing. *FEMS Microbiol Ecol* **90**: 92-102.

Simkus, D.N., Slater, G.F., Lollar, B.S., Wilkie, K., Kieft, T.L., Magnabosco, C. et al. (2016) Variations in microbial carbon sources and cycling in the deep continental subsurface. *Geochimica et Cosmochimica Acta* **173**: 264-283.

Strong, P.J., Xie, S., and Clarke, W.P. (2015) Methane as a resource: can the methanotrophs add value? *Environ Sci Technol* **49**: 4001-4018.

Tang, K.W., McGinnis, D.F., Ionescu, D., and Grossart, H.-P. (2016) Methane production in oxic lake waters potentially increases aquatic methane flux to air. *Environ Sci Technol Lett* **3**: 227-233.

Tavormina, P.L., Orphan, V.J., Kalyuzhnaya, M.G., Jetten, M.S., and Klotz, M.G. (2011) A novel family of functional operons encoding methane/ammonia monooxygenase-related proteins in gammaproteobacterial methanotrophs. *Environ Microbiol Rep* **3**: 91-100.

Tchawa Yimga, M., Dunfield, P.F., Ricke, P., Heyer, J., and Liesack, W. (2003) Wide distribution of a novel pmoA-like gene copy among type II methanotrophs, and its expression in Methylocystis strain SC2. *Appl Environ Microbiol* **69**: 5593-5602.

Trotsenko, Y.A., and Khmelenina, V.N. (2002a) The biology and osmoadaptation of haloalkaliphilic methanotrophs. *Mikrobiologiia* **71**: 149-159.

Trotsenko, Y.A., and Khmelenina, V.N. (2002b) Biology of extremophilic and extremotolerant methanotrophs. *Arch Microbiol* **177**: 123-131.

Tsubota, J., Eshinimaev, B.T., Khmelenina, V.N., and Trotsenko, Y.A. (2005) Methylothermus thermalis gen. nov., sp. nov., a novel moderately thermophilic obligate methanotroph from a hot spring in Japan. *Int J Syst Evol Microbiol* **55**: 1877-1884.

van der Ha, D., Bundervoet, B., Verstraete, W., and Boon, N. (2011) A sustainable, carbon neutral methane oxidation by a partnership of methane oxidizing communities and microalgae. *Water Res* **45**: 2845-2854.

van Teeseling, M.C., Pol, A., Harhangi, H.R., van der Zwart, S., Jetten, M.S., Op den Camp, H.J., and van Niftrik, L. (2014) Expanding the verrucomicrobial methanotrophic world: description of three novel species of Methylacidimicrobium gen. nov. *Appl Environ Microbiol* **80**: 6782-6791.

Veraart, A.J., Garbeva, P., van Beersum, F., Ho, A., Hordijk, C.A., Meima-Franke, M. et al. (2018) Living apart together-bacterial volatiles influence methanotrophic growth and activity. *ISME J* **12**: 1163-1166.

Weber, H.S., Habicht, K.S., and Thamdrup, B. (2017) Anaerobic Methanotrophic Archaea of the ANME-2d Cluster Are Active in a Low-sulfate, Iron-rich Freshwater Sediment. *Front Microbiol* **8**: 619.

Whittenbury, R., Phillips, K.C., and Wilkinson, J.F. (1970) Enrichment, isolation and some properties of methane-utilizing bacteria. *Journal of general microbiology* **61**: 205-218.

Wu, M.L., van Alen, T.A., van Donselaar, E.G., Strous, M., Jetten, M.S., and van Niftrik, L. (2012) Co-localization of particulate methane monooxygenase and cd1 nitrite

reductase in the denitrifying methanotroph 'Candidatus Methylomirabilis oxyfera'. *FEMS Microbiol Lett* **334**: 49-56.

Yoshida, N., Iguchi, H., Yurimoto, H., Murakami, A., and Sakai, Y. (2014) Aquatic plant surface as a niche for methanotrophs. *Front Microbiol* **5**: 30.

Zhang, L., Liao, Q., Gao, R., Luo, R., Liu, C., Zhong, J., and Wang, Z. (2021) Spatial variations in diffusive methane fluxes and the role of eutrophication in a subtropical shallow lake. *Sci Total Environ* **759**: 143495.

FIGURE LEGENDS

Figure 1.1. a. *Microcystis*-dominated cyanoHABs outbroken in lake Taihu, China in June of 2019 (credited by Hans W. Paerl). **b.** Microscope photo of bloom aggregate under 100x magnification (credited by Rosie Moon-Escamilla) showing *Microcystis* cells are surrounded by different bacterial strains.

Figure 1.2. Sanford Underground Research Facility (SURF) in former Homestake gold mine located in South Dakota, USA, serving as the platform for experiments in different scientific fields. (This figure was adapted from the National Science Foundation Website: https://www.nsf.gov/news/news_images.jsp?cntn_id=109694).

FIGURES

a.

b.

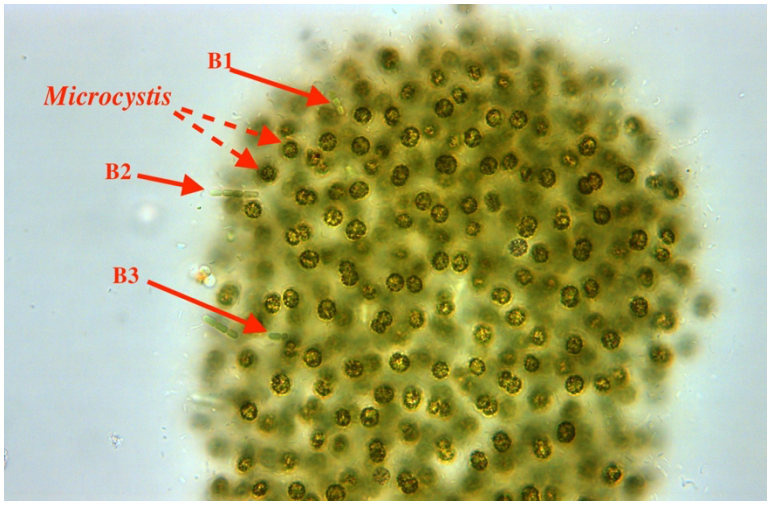


Figure 1.1. Photos of *Microcystis* bloom and aggregates.

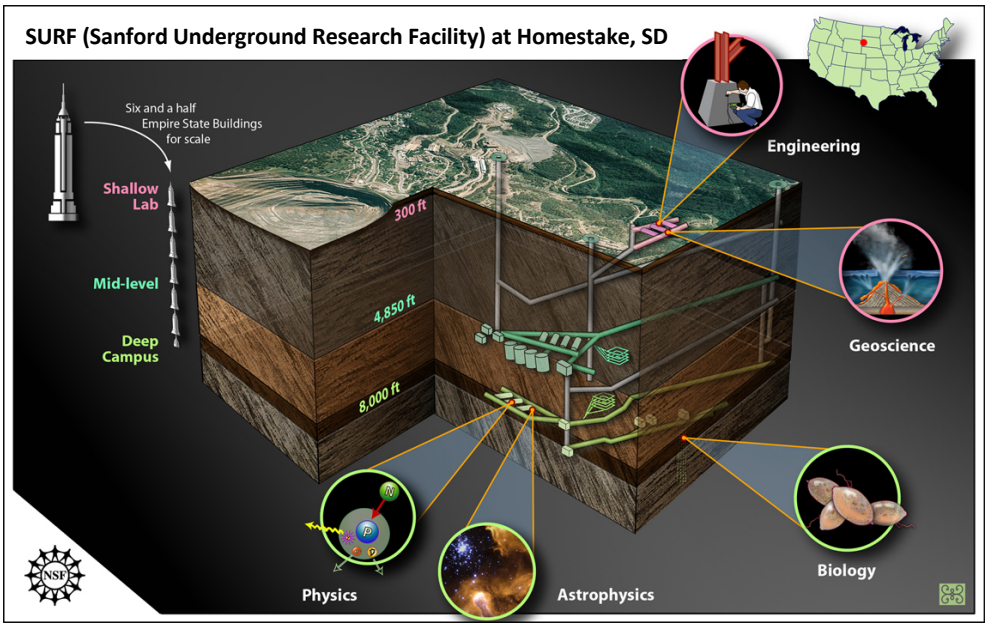


Figure 1.2. Sanford Underground Research Facility (SURF) in former Homestake gold mine located in South Dakota, USA.

CHAPTER 2: GLOBAL CO-OCCURRENCE OF METHANOGENIC ARCHAEA AND METHANOTROPHIC BACTERIA IN MICROCYSTIS AGGREGATES

Chuang Li[1], K. David Hambright[2], Hannah G. Bowen[3], Majoi A. Trammell[4], Hans-Peter Grossart[5], Michele A. Burford[6], David P. Hamilton[7], Helong Jiang[8], Delphine Latour[9], Elisabeth I. Meyer[10], Judit Padisak[11], Richard M Zamor[12] and Lee R. Krumholz[1,*]

[1]Department of Microbiology and Plant Biology; Institute for Energy and the Environment, The University of Oklahoma, Norman, Oklahoma

[2]Plankton Ecology and Limnology Laboratory, Program in Ecology and Evolutionary Biology, and the Geographical Ecology Group, Department of Biology, The University of Oklahoma, Norman, Oklahoma

[3]Department of Microbiology and Immunology, The University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma

[4]Biomedical Research Center, The University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma

[5]Department of Experimental Limnology, Leibniz Institute for Freshwater Ecology and Inland Fisheries, Stechlin, and Institute for Biochemistry and Biology, Potsdam University, Potsdam, Germany

[6]Australian Rivers Institute and School of Environment and Science, Griffith University, Nathan, Queensland, Australia

[7]Australian Rivers Institute, Griffith University, Nathan, Queensland, Australia

[8]State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing, China

[9]Université Clermont Auvergne CNRS, LMGE Aubière Cedex France.

[10]Institute for Evolution and Biodiversity, University of Münster, Münster, Germany

[11]Research Group of Limnology, Centre of Natural Sciences, University of Pannonia, Veszprém, Hungary

[12]Grand River Dam Authority, Vinita Oklahoma

*Correspondence: Lee R. Krumholz, Department of Microbiology and Plant Biology, 770 Van Vleet Oval, University of Oklahoma, Norman, OK 73019, Phone: 405-325-0437, E-mail: krumholz@ou.edu

Keywords:

Cyanobacterial bloom; Methanotrophs; Methanogens; Methane cycle; Freshwater

Formatted for publication in *Environmental Microbiology*

ABSTRACT

Global warming and eutrophication contribute to the worldwide increase in cyanobacterial blooms, and the level of cyanobacterial biomass is strongly associated with rises in methane emissions from surface lake waters. Hence, methane metabolizing microorganisms may be important for modulating carbon flow in cyanobacterial blooms. Here, we surveyed methanogenic and methanotrophic communities associated with floating *Microcystis* aggregates in ten lakes spanning four continents, through sequencing of 16S rRNA and functional marker genes. Methanogenic archaea (mainly *Methanoregula* and *Methanosaeta*) were detectable in five of the ten lakes and constituted the majority (~50 – 90%) of the archaeal community in these lakes. Three of the ten lakes contained relatively more abundant methanotrophs than the other seven lakes, with the methanotrophic genera *Methyloparacoccus*, *Crenothrix*, and an uncultured species related to *Methylobacter* dominating and nearly exclusively found in each of those three lakes. These three are among the five lakes in which methanogens were observed. OTU richness and abundance of methanotrophs were strongly positively correlated with those of methanogens, suggesting that their activities may be coupled. These *Microcystis*-aggregate-associated methanotrophs may be responsible for a hitherto overlooked sink for methane in surface freshwaters, and their co-occurrence with methanogens sheds light on the methane cycle in cyanobacterial aggregates.

INTRODUCTION

Cyanobacterial harmful algal blooms (cyanoHABs) in freshwater lakes have reached unprecedented levels in recent years. This increase is a global ecological issue, mainly

due to depletion of oxygen from surface waters and release of toxins, which can severely impact ecosystem function and human health (Huisman *et al*., 2018). The increased frequency of cyanoHABs are in general linked to excessive nutrient inputs and rising global temperatures (Paerl and Otten, 2013; Mantzouki *et al*., 2018). Most cyanoHABs are composed of planktonic aggregates, usually containing a dominant cyanobacterial species as a primary producer accompanied by diverse heterotrophic bacteria, which may contribute to the proliferation and persistence of cyanobacterial blooms (e.g., through recycling of nutrients, generation of CO_2, and detoxification of oxygen radicals) (Dziallas and Grossart, 2011; Dziallas and Grossart, 2012; Paerl and Otten, 2013; Cook *et al*., 2020). The community composition of the associated heterotrophic bacteria can vary with size of aggregates (Cai *et al*., 2014), while anoxic microniches important for anaerobes are more likely to occur in larger aggregates (Klawonn *et al*., 2015) and scums.

Although we are beginning to get a clearer picture of the nitrogen cycle within cyanobacterial aggregates (Klawonn *et al.*, 2015), the role of methane cycling microbes (methanogens and methanotrophs) has been addressed to a lesser degree, but is clearly very important, as enhanced eutrophication of freshwater lakes promotes growth of photoautotrophs and methane emissions from surface waters of lakes have been positively linked to methane production by Cyanobacteria and other algae (Grossart *et al*., 2011; Bogard *et al*., 2014; Bizic *et al*., 2020; Günthel *et al*., 2020). Increased eutrophication has been predicted to result in a substantial increase (30 – 90%) in methane emissions to the atmosphere over the next century (Beaulieu *et al*., 2019), but the linkage to cyanobacterial blooms has gained little scientific attention.

Microbial methane production in lakes was thought to primarily occur in anoxic sediments, from where up-diffused methane would be depleted by aerobic methane oxidation in the surface layers of sediment or in the water column (Bastviken *et al*., 2008). Yet, there have been several reports showing the presence and activity of methanogenic archaea associated with photoautotrophs in oxic lake waters (Grossart *et al*., 2011; Bogard *et al*., 2014), as well as the occurrence of methanogens in isolated *Microcystis* aggregates (Batista *et al*., 2019). Their presence is likely due to temporary anoxic microniches within planktonic aggregates (Paerl and Pinckney, 1996; Dziallas and Grossart, 2012; Klawonn *et al*., 2015).

Methane oxidation can be rapid in the presence of oxygen, e.g. during seasonal lake turnover (Mayr *et al*., 2020). In addition, methane oxidation coupled to oxygenic photosynthesis, allowing methane oxidation and subsequent symbiotic carbon transfer to the phototroph, has been documented in the anoxic hypolimnion of lakes and in peatlands (Raghoebarsing *et al*., 2005; Milucka *et al*., 2015). Until now, there are no studies on the occurrence and microscale distribution of methanotrophs within cyanobacterial aggregates. Given previous observations of co-occurring aerobic and anerobic microbial nitrogen cycling processes across steep oxygen gradients within cyanobacterial aggregates (Klawonn *et al*., 2015), it is likely that methanotrophy is indeed coupled with archaeal methanogenesis within *Microcystis* aggregates.

At circumneutral pH, aerobic oxidation of methane is carried out by methane-oxidizing bacteria (MOB) belonging to Alphaproteobacteria and Gammaproteobacteria (Hanson and Hanson, 1996). The *pmoA* gene encoding a subunit of the particulate methane monooxygenase is commonly used as a biomarker to detect MOB, as phylogenies derived from it are largely congruent with those of the 16S rRNA gene (Knief, 2015). In addition, methanogenic members within the archaeal phylum Euryarchaeota universally use methyl-coenzyme M reductase, and the associated *mcrA* gene is typically used as functional biomarker for investigating methanogenic archaea communities (Thauer *et al*., 2008).

The goals of this study were (i) to investigate the distribution of methane cycling microbes associated with cyanobacterial blooms and (ii) to investigate their relationships based on abundances and richness of marker genes to predict whether activities of methanogens and methanotrophs are linked to each other. We focused primarily on *Microcystis* blooms in freshwater lakes and collected aggregate samples from ten lakes across four continents (Fig. S2.1).

RESULTS

Cyanobacterial reads were predominant in each aggregate sample (Fig. S2.2), most of which (>82% of Cyanobacteria) were further classified to the *Microcystis* genus, confirming that *Microcystis* was the major bloom-forming component in all aggregate samples.

Molecular characterization of methanogenic archaea in cyanobacterial bloom aggregates

To explore the potential for biotic methane production within the cyanobacterial aggregates, the presence of methanogenic archaea in aggregate DNA extracts was examined by generating archaeal 16S rRNA gene libraries. Although archaeal reads were nearly absent from five of the ten sampled lakes, a total of 301 OTUs from 48,135 high-quality archaeal reads were successfully retrieved from the remaining five lakes (Fig. 2.1a). These latter five included the group of three lakes with a high proportion of methanotrophs (see below, i.e. high MOB group). Methanogens belonging to six orders of the phylum Euryarchaeota, together comprised 50-90% of the total archaeal community across these lakes (Fig. 2.1a). In addition, the other major archaeal lineages were represented by Woesearchaeia and Bathyarchaeia, along with ~0.63% - 3.6% of sequences affiliated with five other phyla (Fig. 2.1a). Within the six methanogenic orders, members of the two genera *Methanosaeta* and *Methanoregula* dominated the methanogenic communities (together accounting for 39.1 to 72.4% of reads) across the five lakes (Fig. 2.1b).

Additionally, we targeted the *mcrA* gene for further proof of their presence and for phylogenetic analysis of methanogens. The *mcrA* genes could not be amplified via PCR in the lakes that did not yield any archaeal 16S gene products. With identical PCR conditions and sequencing runs, the other five lakes yielded the correct sized amplicons with the number of identified *mcrA* reads ranging from 9,936 to 31,851 per lake. Given the consistent distribution pattern observed with the two types of genes, we confirmed

that methanogens occurred in five of ten lakes and were most likely to be either absent or present in negligible numbers in the others. Phylogenetic analysis of representative sequences of the 41 *mcrA* OTUs showed that all fell within the same orders of Euryarchaeota as determined by 16S rRNA gene analysis. The majority of the OTUs determined by *mcrA* analysis were similarly closely related to the methanogenic genera identified by 16S gene analysis (Fig. 2.1&2.2). The composition of methanogenic communities among the lakes uncovered by *mcrA* analysis (Fig. 2.1c) also confirms the dominance of *Methanosaeta* and *Methanoregula* among the detected methanogens. As well, six OTUs within the order Methanomicrobiales robustly form monophyletic clades with environmental clones, supported by high bootstrap values (Fig. 2.2; labelled OTU17, 27,41,56,62,78). OTU17 (the fourth most abundant *mcrA* OTU representing 10.6% of total *mcrA* reads) was broadly distributed among lake samples and clustered with a clone (97.16% similarity) from Lake Dianchi which is known to have persistent blooms (Yang *et al.*, 2016). Using a summary of the range of substrates utilized by each specific methanogenic lineage (Table S2.1) (Evans *et al*., 2019), we observed an abundance of methanogens with the potential to obligately perform hydrogenotrophic or acetoclastic methanogenesis, constituting 87.2 to 96.2% of total methanogens across the five lakes (Fig. S2.3), suggesting that acetate and H_2 constitute their main energy sources, obtained through a direct interaction with fermentative bacteria, rather than being active in degradation of methylated substrates.

As the same lineages of dominant methanogens were observed in previous studies (Grossart *et al.*, 2011, Batista *et al.*, 2019), this raises the question of whether

cyanobacterial aggregates share a similar methanogenic community. The dissimilarity test (Adonis) based on weighted UniFrac distance of *mcrA* genes, showed significant differences in the phylogenetic structure of methanogenic communities (at the OTU level) among five lakes ($R^2 = 0.85$, $p = 0.001$) (Fig. S2.5c). Correspondingly, a principal coordinates analysis (PCoA) plot demonstrated that methanogenic communities (determined with *mcrA)* are clearly divided into three groups (1-Lakes Clarendon and Rotoehu; 2-Wentowsee; 3-Lac de Villerest and Aasee), based on their clear separation along the first axis of the plot (Fig. S2.5c). This pattern was consistent with data from procrustes analysis using the archaeal 16S gene dataset ($m^2 = 0.286$, $p < 0.001$) (Fig. S2.4).

Molecular characterization of MOB community in cyanobacterial bloom aggregates

After removal of 16S cyanobacterial sequences, the remainder were used to describe the MOB community with 5,831 MOB sequences clustered into 35 OTUs (Fig. S2.6). Among the ten lakes, type II MOB belonging to *Methylocystis* were ubiquitous and present at similar levels across all lakes, whereas some specific type I MOB genera were nearly exclusively found in one lake and not in the others (Fig. 2.3a). Specifically, *Methyloparacoccus* and one of the eight OTUs associated with unclassified *Methylomonaceae* co-dominated the MOB community in Lake Clarendon, constituting 1% and 0.74% of the non-cyanobacterial sequences respectively (Fig. 2.3a). The latter unclassified OTU was found to be most similar (99.3%) to a strain of the very recently defined novel genus *Methylicorpusculum* from an oil sands tailings pond (Saidi-Mehrabad *et al*., 2020). In Lac de Villerest, the genus *Crenothrix*, with remarkably high

relative abundance, up to 2.38%, outnumbered the other genera (Fig. 2.3a). The lineage pLW-20 (Nercessian *et al*., 2005) accounted for over 1.45% of non-cyanobacterial sequences in Wentowsee (Fig. 2.3a).

Despite our ability to detect MOB through 16S rRNA gene analysis, most of the sequencing results are related to the predominant Cyanobacteria, and the 16S rRNA gene of rare methanotrophs is not fully recovered. Therefore, the derived diversity estimations are less statistically valid. Thus, we also sequenced the *pmoA* functional marker genes from all samples to more effectively characterize the diversity of the MOB community, as *pmoA* primers being highly specific, can more sensitively detect rare methanotrophs. The amplicon yield of *pmoA* genes ranged from 976 sequences in Belső-tó to 132,557 in Lac de Villerest, with a total of 303,339 sequences clustered into 125 OTUs. The uncultivated lineages are identified in the context of a phylogenetic tree and named according to the habitat where they are predominantly found or first reported (Dumont *et al*., 2014). By comparing the *pmoA* sequences from all ten lakes to reference sequences of each lineage provided by a recent study (Dumont *et al*., 2014), we obtained 7 uncultivated lineages at the genus level. Notably, three different lineages appeared to be the most abundant methanotrophs in each of the three high methanotroph lakes ("lake-cluster-2" constituted 34.4% of pmoA sequences in Lake Clarendon; "LP20" made up 83.9% in Lac de Villerest; and "deep-sea-1" made up 51.9% in Wentowsee) (Fig. 2.3b). These three lineages were present at low levels or not at all in the other lakes. "LP20" and "lake-cluster-2" are respectively related to two recently defined genera *Methyloglobus* and *Methyloparacoccus* (as the dominant *pmoA* OTUs from each of the

uncultivated lineages was 96% similar to the related type strains) and supported by updated classifications in a recent study (Knief, 2015). Among the sequences related to cultivated bacteria, common freshwater genera including *Methylocystis*, *Methylomonas* and *Methylobacter* were commonly observed (Fig. 2.3b). Unfortunately, we were unable to detect the *pmoA* gene of *Crenothrix* sp., uncovered by 16S rRNA gene analysis in Lac de Villerest, due to its atypical gamma-proteobacterial *pmoA* which mismatched with the primers used here (Knief, 2015).

As guided by the strengths of each marker gene, we estimated the diversity indices and the abundance of MOB communities using *pmoA* and 16S rRNA gene sequences, respectively. The MOB members together accounted for over 2% of the non-cyanobacterial communities in each of the three lakes (Lake Clarendon, Lac de Villerest, and Wentowsee), levels of which were approximately 10- to 100-fold higher than in the other seven lakes. Based on these order of magnitude differences, the ten lakes were correspondingly grouped into "high" and "low" MOB categories (Table 2.1). Our inability to detect *Crenothrix* sp. in Lac de Villerest by targeting *pmoA*, likely contributed to the unexpectedly low Shannon diversity in this "high" group lake (Table 2.1).

Quantitative and qualitative relationships of methane metabolizing microbes

To quantify the relationship between methane producers and consumers in *Microcystis* aggregates, qPCR was used in all five lakes where both groups were present. The abundance of each of the two groups, as well as *Microcystis*, was first estimated as cell numbers per μg of DNA (Fig. S2.7). To facilitate comparisons of methanogens and

methanotrophs in a more meaningful way, their original abundances were further normalized by dividing their cell numbers by those of *Microcystis* (Fig. 2.4). In Lake Rotoehu, with the lowest levels, there were 15.4 ± 0.8 methanogen cells and 2.1 ± 0.4 methanotroph cells per million *Microcystis* cells (Fig. 2.4), while in the other four lakes, methanogen numbers ranged from 140 ± 20 (Aasee) to 1390 ± 790 (Lake Clarendon), and methanotroph numbers from 50 ± 20 (Lac de Villerest) to 150 ± 10 (Aasee) cells per million *Microcystis* cells (Fig. 2.4). The abundance of methanotrophs is likely underestimated in Lac de Villerest due to our failure to detect the *Crenothrix* sp. *pmoA*. Linear regression was further conducted to estimate the abundance relatedness between methanotrophs and methanogens or *Microcystis* and demonstrated that cell numbers of methanotrophs were positively correlated to cell numbers of methanogens (R^2=0.4, P<0.05) (Fig. 2.5a). In contrast, cell numbers of *Microcystis* showed no correlation (R^2=0.1, P=0.34) (Fig. 2.5a). Likewise, there was a strong positive correlation between the number of *pmoA* OTUs and the number of *mcrA* OTUs across all ten lakes (R^2=0.56, P<0.0001), but only a weak correlation with the number of cyanobacterial 16S rRNA OTUs (R^2=0.26, P<0.05) (Fig. 2.5b).

Heterotrophic bacteria associated with methanotrophs within the aggregate

Interactions between methanotrophs and heterotrophs are widespread in natural systems (Ho *et al*., 2016). It is therefore reasonable to speculate that methanotrophs could have indirect effects on aggregate community function through interacting with the associated heterotrophs. The potential interacting heterotrophs are likely to be co-enriched with methanotrophs in the three high MOB lakes and abundances should be strongly

correlated with methanotroph abundances. Following this assumption, we first attempted to verify whether heterotrophic communities in high MOB lakes were dissimilar from those in low MOB lakes. Significant differences (but with less strength) in the non-cyanobacterial heterotrophic communities was identified between the high and low MOB lakes using three nonparametric multivariate analysis methods (Table S2.2). As reflected on PCoA analysis, high MOB lakes clustered together and were marginally separated from low MOB lakes along the third axis (Fig. S2.8). In agreement with the observed weak dissimilarity, 30 major family-level lineages were found to differ significantly (p<0.05) in abundances between the two groups of lakes (Fig. 2.6a). Among them, 15 heterotrophic families were significantly more abundant in the high MOB lakes, and also positively correlated with the abundance of methanotrophic lineages (Fig. 2.6b).

DISCUSSION

Presence of methanogenic archaea in *Microcystis* aggregates

Previous studies have shown that *Methanoregula*, *Methanosaeta* and uncultured *Methanomicrobiales* were the only methanogenic lineages exclusively associated with the particle fractions of various *Microcystis* strains in lab cultures (Batista *et al*., 2019) or planktonic photoautotrophs in an oligo-mesotrophic lake (Grossart *et al*., 2011). Six OTUs belonging to uncultured Methanomicrobiales were similarly observed in the *mcrA* gene results **(**Fig. 2.2**)**, indicating that this order likely contains additional novel lineages adapted to cyanobacterial aggregates. The results presented here further revealed the diversity of methanogenic communities by identifying the less common lineages: Methanobacteriales, Methanocellales, Methanomassiliicoccales and

Methanofastidiosales. Aggregates may provide protective habitats and favorable (temporally) anoxic niches, providing support for the notion that methanogenic archaea rarely persist as free-living organisms in surface waters (Grossart *et al.*, 2011, Batista *et al.*, 2019). Production of acetate, H_2 and CO_2 by *Microcystis* strains under dark conditions through intracellular fermentation of stored photosynthates may provide energy for acetoclastic and hydrogenotrophic methanogens in aggregates. (Yasuo and Kawamura, 1984; Moezelaar and Stal, 1994; Dziallas and Grossart, 2012). Alternatively, N_2-fixation could be an important source of H_2 in planktonic aggregates (Paerl and Pinckney, 1996), and N_2-fixing bacteria have been shown to be important components of *Microcystis* aggregates (see below). Methanogens may also be important in N_2 fixation within the aggregates as methanogenic Euryarchaeota are the only known archaea that contain *nif* genes (Leigh, 2000). In the Florida Everglades, methanogenic Euryarchaeota are major contributors to N_2 fixation in soils (Bae *et al*., 2018). The absence of detectable methanogens or other Archaea in five of the lakes suggests that other factors may influence the ability of methanogens to survive within the cyanobacterial aggregate, as well as affect their community composition (Fig. S2.5c). These factors might include the presence of alternative respiratory electron acceptors (sulfate, nitrate) in lakes which could allow the methanogens to be out-competed by other anaerobic respiring bacteria (Thauer *et al*., 2008). This competition is becoming important as many lakes now have increasing sulfate (Tao *et al.*, 2013) and nitrate concentrations. An increase in levels of sulfate reducing bacteria within Desulfovibrionales (i.e., Chaohu) (Fig. S2.9) and high concentrations of nitrate and nitrite (i.e., Belső-tó and FP23) (Table S2.3) observed in lakes lacking methanogens, support this hypothesis. Alternatively, another factor that

could control methanogenic viability within aggregates is the presence/absence of enzymes capable of oxygen radical degradation in other microorganisms which might protect methanogens against oxidative stress under high light conditions (Kim *et al*., 2019).

Co-occurrence of methanogen-associated Archaea

Along with methanogenic Euryarchaeota, the other abundant archaeal classes Woesearchaeia and Bathyarchaeia likely interact with methanogens. Woesearchaeia-methanogen consortia are globally distributed in inland anoxic environments with a syntrophic relationship proposed based on their genomic features (Liu *et al*., 2018). Members of Bathyarchaeia are also widespread in anoxic environments most likely due to a range of metabolic capabilities, including a potential for methane metabolism (Evans *et al*., 2015) and acetogenesis (He *et al*., 2016), likely feeding reducing equivalents to methanogens during anoxic periods. Together, Euryarchaeota, Woesearchaeia and Bathyarchaeia made up 96.4% - 99.3% of the archaeal communities across the five lakes, highlighting the potential importance of Archaea in methane-related carbon flow in cyanobacterial aggregates.

Global distribution of MOB communities in cyanobacterial bloom aggregates

In order to fully understand the role of MOB in cyanobacterial bloom aggregates, it is critical to understand their distribution on a global scale. This worldwide survey of bacterial 16S rRNA genes showed that proteobacterial MOB were ubiquitously present in *Microcystis* aggregates, but most abundant in three of the ten lakes (i.e., the high MOB

lakes). One of a variety of environmental parameters or the availability of methanotroph substrates (i.e., CH_4, O_2 and C_1-compounds) could directly or indirectly affect the distribution of methanotrophic genera in the lakes. Although C_1 methylated compounds are potential substrates for MOB, they are commonly found at low concentrations in the oxic epilimnion of lakes (Watson and Juttner, 2017). Diurnal O_2 fluctuation (high during the day and low at night) and its relative importance at specific times of the bloom season (Chen *et al.*, 2016), likely influence the taxonomic composition and allow for alpha-MOB (see below) to dominate as these are thought to be better adapted to fluctuating O_2 concentrations in surface waters (Reis *et al.*, 2020). Although limited lake methane data is available, Chaohu recorded a relatively high CH_4 concentration (0.36 μM CH_4), compared to the lower CH_4 levels (< 0.1 μM) recorded in two of the five lakes where methanogens occurred (Table S2.3). This provides some support for the argument that in the five lakes where methanogenic archaea are present (see below), methane is cycled internally with little spillover to the surface water. On the other hand, MOB in the other five lakes may be fed by CH_4 diffusing from outside of aggregates. However, more information is needed to conclusively determine this.

Environmental effects were also considered. The higher DOC concentrations observed in Lake Clarendon and Wentowsee (Table S2.3) may have increased the activity and abundance of MOB by relieving light inhibition in surface waters (Thottathil *et al.*, 2018) or by providing carbon for anaerobic bacteria living within the aggregates to produce methane. Previous results from a high MOB lake (Lac de Villerest) along with two low MOB lakes (Grand Lake and Lake Kinneret), showed MOB present in their mid water

column where diffusive CH_4/O_2 counter-gradients meet during the stratification period (Junier *et al.*, 2010; Morrison *et al.*, 2017; Pradeep Ram *et al.*, 2019). Considering that Lac de Villerest was sampled during the early period of turnover (October), while the other two lakes were sampled at the beginning of the stratification period (Table S2.3), we speculate that abundant MOB may have been circulated to the surface during the fall turnover in Lac de Villerest and subsequently recruited into bloom aggregates.

Although there have been few studies specifically focusing on MOB associated with cyanobacterial blooms, MOB-related lineages have been previously reported in *Microcystis*-bloom containing lakes (Yang *et al*., 2017; Akins *et al*., 2018; Guedes *et al*., 2018; Yang *et al*., 2019). For example, the eutrophic Dianchi had a higher abundance of MOB in sediments in comparison to a nearby mesotrophic lake (Yang *et al*., 2019). In another study, members of the MOB lineage Methylococcales were found to be present in the water column and temporally associated with *Microcystis* in Funil Reservoir (Brazil) (Guedes *et al*., 2018).

Analysis of both 16S rRNA and *pmoA* genes showed that type I gamma-MOB represent abundant methane oxidizers in *Microcystis* aggregates. As type I MOB are often more abundant than type II MOB in lake waters (Biderre-Petit *et al*., 2011; Blees *et al*., 2014; Kojima *et al*., 2014; Samad and Bertilsson, 2017), it is possible that MOB within the aggregates were recruited from the water column. For example, *Methyloparacoccus* (referred to as "lake-cluster-2" in Fig. 2.3b) was the most abundant MOB in Lake Clarendon and was recently isolated from pond water in South Africa and Japan

(Hoefman *et al*., 2014b). Related clones have also been observed in several lakes where *Microcystis*-blooms are present, including Dianchi (Yang *et al*., 2019), Taihu (Wu *et al*., 2007) and Kastoria Lake (Kormas *et al*., 2010). In Wentowsee, pLW-20 and "deep-sea-1" were the most abundant MOB lineages as determined by 16S and *pmoA*, respectively. The predominant OTU in each of these lineages was most closely related to the 16S rRNA sequence (94% nucleotide sequence identity) and *pmoA* sequence (93% amino acid sequence identity) of *Methylobacter marinus A45* (Lidstrom, 1988; Bowman *et al*., 1993) which requires NaCl for growth and may be capable of N_2 fixation (Flynn *et al*., 2016). *Methylocystis* was the only MOB genus, verified by both primer sets, to be present in all lakes. Both *pmoA* (98.7% amino acid sequence identity) and 16S rRNA gene (99.8% nucleotide sequence identity) derived OTUs were most closely related to those of *Methylocystis parvus*, with the ability to form lipid cysts as a resting stage (Whittenbury *et al*., 1970; Bowman *et al*., 1993), and to intracellularly ferment hydroxybutyrate under anoxic conditions (Vecherskaya *et al*., 2009). These and other potential adaptive traits may explain their ubiquitous distribution pattern. The detection of the 16S rRNA gene of *Crenothrix* sp. in Lac de Villerest was consistent with previous findings that this filamentous MOB was more abundant in *Microcystis*-associated aggregates than free living in the water column (Yang *et al*., 2017). Our *pmoA* results indicate that the *Methyloglobulus*-like lineage "LP20" represents another important MOB group in Lac de Villerest. *Methyloglobulus* has been described to grow optimally at reduced oxygen concentrations (Deutzmann et al., 2014a). This suggests that these MOB communities are well adapted to life in *Microcystis* aggregates, where high nutrients (Yang *et al*., 2019)

and reduced oxygen concentrations (Blees et al., 2014; Deutzmann et al., 2014a; Oswald et al., 2017) as well as particle-associations (Yang *et al*., 2017) are common.

It is important to note that, under anoxic conditions, methane-dependent nitrate and nitrite reduction can be respectively catalyzed by ANME-2d associated archaeal methanotrophs (Haroon *et al.*, 2013) and the bacterial genus *Methylomirabilis* in phylum NC10 (Ettwig *et al.*, 2008). Not surprisingly, neither *Methylomirabilis*-like 16S rRNA gene sequences (Fig. 2.3a) nor ANME-2d associated *mcrA* (Fig. 2.2) or related 16S rRNA gene sequences (Fig. 2.1) were retrieved from any aggregate samples, whereas positive control sediments and filters effectively amplified these sequences (data not shown). These observations concur with the lack of evidence for the occurrence of anaerobic methanotrophs in surface waters, presumably due to the slow growth rate (Doubling time: T_d > 1 month (Kampman *et al.*, 2012; Zhu *et al.*, 2012; Vaksmaa *et al.*, 2017)) restricting them to habitats with lower turnover, such as the lake and groundwater sediments (Deutzmann *et al.*, 2014b). Acidophilic methanotrophs from the phylum Verrucomicrobia (Dunfield *et al.*, 2007) were also not detected in aggregates (Fig. 2.3a), perhaps due to the unfavorable neutral to alkaline pH that occurs throughout the bloom season.

MOB capable of directly influencing the N-cycle in bloom aggregates

Microcystis, as a non-diazotrophic cyanobacterium, is solely reliant on fixed N sources derived from lake water or from N-cycling reactions conducted by its associated bacteria (Cook *et al*., 2020). The role of N_2-fixing bacteria becomes critical during the midsummer period of maximal bloom growth, during which dissolved inorganic N decreases dramatically (Xu *et al*., 2010; Jankowiak and Gobler, 2020). As such,

Rhizobiales, Azospirillales (Fig. S2.9) (Yang *et al*., 2017; Kim *et al*., 2019; Jankowiak and Gobler, 2020) and many MOB lineages having N_2-fixation potential (Auman *et al*., 2001), could increase N availability to *Microcystis*. Possible methanotrophic taxa could include members from the abundant genera *Crenothrix* (Oswald *et al*., 2017), *Methyloglobus* (Deutzmann et al., 2014a), *Methylicorpusculum* (Saidi-Mehrabad *et al*., 2020), and also ubiquitous genera such as *Methylobacter* (Kalyuzhnaya *et al*., 2015; Flynn *et al*., 2016), *Methylomonas* (Hoefman *et al*., 2014a; Kalyuzhnaya *et al*., 2015), and *Methylocystis* (Whittenbury *et al*., 1970; Bowman *et al*., 1993). O_2-depleted microzones are usually required for facilitating N_2-fixation in cyanobacterial aggregates (Paerl and Bebout, 1988; Paerl and Pinckney, 1996). Daylight promotes photosynthesis resulting in high O_2 levels, whereas dark respiration of organic compounds contributes to O_2 depletion (Dziallas and Grossart, 2012; Klawonn *et al*., 2015). As such, N_2-fixation in cyanobacterial aggregates has been shown to be far more active during darkness (Klawonn *et al*., 2015).

Likewise, the presence of anoxic or suboxic microzones could help us understand the occurrence of previously reported aggregate-associated denitrifying bacteria (Yang *et al*., 2017; Jankowiak and Gobler, 2020). Many of the MOB genera found here are also known to contain genes encoding enzymes involved in denitrification (Hoefman *et al*., 2014b; Smith *et al*., 2018; Saidi-Mehrabad *et al*., 2020) and methane oxidation coupled to denitrification occurs in type I MOB under hypoxia (Kits *et al*., 2015b; Kits *et al*., 2015a). *Crenothrix* and *Methylobacter* probably take advantage of denitrification to grow within the oxygen-deficient water column (Oswald *et al*., 2017; van Grinsven *et al*.,

2020). However, denitrification reactions in cyanobacterial aggregates require N oxides to diffuse from oxic zones where nitrification occurs (Klawonn *et al*., 2015). Nitrifying bacteria, (*Nitrosomonadaceae*) were present in all *Microcystis* aggregates in all sampled lakes and were observed in other studies (Louati *et al*., 2015; Jankowiak and Gobler, 2020) suggesting a likely source of N-oxides.

Potential coupling of methane production and consumption between methanogenic archaea and MOB

Given the co-occurrence of MOB and methanogenic archaea within *Microcystis* aggregates, the question remains as to whether their metabolisms are coupled. Previous work has demonstrated a positive relationship between biodiversity and ecosystem function on highly specialized metabolic processes carried out by individual groups of microbes (Wohl *et al*., 2004; Peter *et al*., 2011). For example, diversity indices and absolute abundance of MOB have been demonstrated to be positively correlated with methane consumption rates in a wetland (Bodelier *et al*., 2013), upland soil (Levine *et al*., 2011) and in surface lake waters (Reis *et al.*, 2020). In lab studies with cyanobacteria-associated soil aggregates, the abundance of methanogenic archaea was positively correlated with the methane production rate (Angel *et al*., 2012). Our results also reveal a positive correlation between methanogens and MOB total numbers as well as between methanogen and MOB richness, suggesting that the activities of both groups are indeed coupled.

In order for methanogenesis and methanotrophy to be coupled, both would likely occur in overlapping time scales. Most methanogenic archaea are known to be sensitive to O_2. Several groups, however, can tolerate O_2 exposure (Huser *et al*., 1982; Jarrell, 1985), but are most likely only active under highly reduced conditions. A number of studies have suggested that *Crenothrix* (Oswald *et al*., 2017), *Methylobacter* (Biderre-Petit *et al*., 2011; Blees *et al*., 2014; van Grinsven *et al*., 2020) and unclassified lineages of gamma-MOB (Oswald *et al*., 2016; Cabrol *et al*., 2020) are abundant and active in O_2-depleted lake water columns. The most abundant *pmoA* OTU in Wentowsee was most closely related to clones (HF674404: 98.7% and AB930855: 96.8% similarity) previously obtained from a deep anoxic lake water layer (Blees *et al*., 2014; Kojima *et al*., 2014) . Presumably, these MOB are also growing in low O_2 environments using a combination of aerobic and anaerobic (nitrate/nitrite respiration or fermentation) metabolisms (Vecherskaya *et al*., 2009; Kalyuzhnaya *et al*., 2013; Oswald *et al*., 2017; van Grinsven *et al*., 2020). For temporal alignment of these two groups, both methanogens and MOB would likely cooperate within microzones exhibiting fluctuating oxygen conditions with low concentrations during darkness and high concentrations when light is present. Previous work also suggests that daytime light results in a lower methane oxidation rate, due to photoinhibition of MOB (Murase and Sugimoto, 2005; Tang *et al.*, 2014; Tang *et al.*, 2016; Thottathil *et al.*, 2018). It is therefore likely that darkness is the niche for temporal coupling of MOB and methanogenic archaea activities to co-occur within *Microcystis* aggregates. While our analysis is based entirely on relative abundances, direct measures of coupled activities will be required to address in detail the potential for methanogenesis and methanotrophy to occur in the cyanobacterial aggregates, even

though such work will be a real challenge to perform in such a dynamic microscale system.

Coupling of methanogenic archaea activities could potentially provide MOB access to much higher methane concentrations than might be available solely by diffusion from anoxic benthic waters and sediments (Bastviken *et al*., 2011). Several possible alternative mechanisms for microbial methane production could exist within cyanobacterial aggregates (Tang *et al*., 2014; Tang *et al*., 2016; Bizic *et al*., 2020). However, none of the alternatives could produce methane at levels that compare to those of active archaeal methanogens (Bizic *et al*., 2020), indicating that within anoxic/suboxic cyanobacterial aggregates where methanogenic archaea are present, they represent an important source of methane, in particular when oxygen levels are low. In their absence and at increasing oxygen concentrations, alternative sources of methane such as from cyanobacterial oxic methane production will be more important.

MOB: a potential hidden driver of carbon flow within cyanobacterial aggregates

The co-occurrence and co-variation of methanotrophic and non-methanotrophic bacteria in cyanobacterial aggregates can be used to predict the existence of ecological interactions among various microorganisms. For example, abundance-based correlations have predicted that redox partnerships exist between MOB and their associated microbiome in anoxic waters (Cabrol *et al*., 2020). Here, the abundance (based on relative read levels) of the three methanotroph-containing families (*Methylomonaceae*, *Methylococcaceae* and *Beijerinckiaceae*) (Fig. 2.6) were significantly higher in the high

MOB lakes, supporting our results on the separation of the ten lakes into two groups. Several families identified here to be enriched with methanotrophs (e.g., *Microscillaceae*, *Aeromonadaceae*, *Caldilineaceae* and *Solibacteraceae*) are chemo-organotrophs and usually present in organic compound enriched habitats, where their functional potentials are not well understood. Previous stable isotope probing studies have suggested that members of the *Burkholderiaceae* and *Bdellovibrionaceae* may feed either on metabolites released during methane oxidation or directly on cells of MOB in natural consortia (Morris *et al*., 2002; Ho *et al*., 2016), explaining their associations observed in this study. Along with Bacteria and Archaea, methane-derived carbon can also be assimilated into the biomass of grazing Eukarya (Murase and Frenzel, 2007), which could be further utilized by obligate intracellular parasitic bacteria from *Paracaedibacteraceae* (Hess *et al*., 2016) and *Rickettsiaceae* (Gortz and Brigge, 1998), two other groups enriched in *Microcystis* aggregates of high MOB lakes. The heterotrophic family *Ilumatobacteraceae* also enriched here was found to be enriched with *Methyloglobulus* in bathypelagic water layers of Lake Baikal (Cabello-Yeves *et al*., 2020), suggesting a functional or evolutionary relationship. Such cross-feeding interactions allow us to hypothesize that a heterotrophic food web dependent on the additional carbon contributed by MOB could be present.

The co-occurrence of methanogens, nitrifiers and specific heterotrophs with MOB at the scale of hundreds of micrometers within cyanobacterial aggregates suggests to us that their functions interact to directly link the carbon and nitrogen cycles as illustrated in Fig.

2.7. These interactions, could allow a stable and co-evolving population of organisms to thrive within the aggregates (Garcia *et al*., 2015).

In conclusion, our results provide compelling molecular evidence for the co-occurrence of methanotrophs and methanogenic archaea in *Microcystis* aggregates from a global perspective. The absolute abundances and richness of MOB communities and methanogenic communities were positively correlated across global lake samples, suggesting that methane cycling is important within these cyanobacterial aggregates. Yet, their *in-situ* activities and interactions require further exploration.

MATERIALS AND METHODS

Bloom samples and collection

Cyanobacterial bloom samples dominated by *Microcystis* were collected from 12 lakes (Cook *et al*., 2020). Here, only ten of those lakes were used for studying methane metabolizing organisms within *Microcystis* aggregates because the other two lakes were missing replicates. The ten lakes spanned a 90° latitudinal and 270° longitudinal gradient (Fig. S2.1) and their general limnological information was included as well (Table S2.3). It is important to note that this information is a summary, based on recently recorded data and in most cases, was not directly collected during sampling trips. The size of lakes varies greatly, from 0.001 (FP23) to 770 km^2 (Chaohu). The five lakes where methanogens were present appear to be restricted to high-latitude regions. Methane, DOC and stratification data were used to help explain the differential distribution of methanotrophs (see above). Surface water samples were collected and left undisturbed for

10 min. The concentrated floating cyanobacterial biomass on the surface was filtered through triplicate sterile Nitex screens (100-µm pore size) to collect large *Microcystis* aggregates. The Nitex screens were then submerged in DNA preservative (Zymo Research, Irvine, CA, USA), shipped at ambient temperature, and stored at -20°C prior to DNA extraction (for details, see (Cook *et al*., 2020)).

DNA extraction, Amplification and Illumina sequencing

DNA was extracted from Nitex screens using Quick-DNA Fecal/Soil Microbe Miniprep Kits (Zymo Research). The amplicons of bacterial 16S rRNA genes sequenced in our previous study (Cook *et al*., 2020) were used for exploring the methanotrophic community, since *in-silico* evaluation results indicated that the primers (Table S2.4a) used, had broad coverage of bacterial methanotrophic lineages (Table S2.4b). In this study, the archaeal 16S rRNA, *mcrA* (targeting methanogenic Archaea) and *pmoA* (targeting methanotrophic Bacteria) genes were amplified from the same DNA template using the corresponding primers and thermocycler conditions as summarized in (Table S2.4a). For each of the purified amplicons, 5 µl were added to a second PCR mixture containing the appropriate index primers using the above PCR conditions for another eight cycles, to allow each sample to receive a unique "barcodeF+barcodeR" combination sequence to allow demultiplexing post-pooling. The mixed sample was sequenced on an Illumina MiSeq platform at the Oklahoma Medical Research Foundation (300 bp paired-end).

Quantitative real-time PCR

Quantification of methanogens, methanotrophs and *Microcystis* abundance was performed via qPCR, targeting *mcrA*, *pmoA* and 16S rRNA genes *of Microcystis*. All qPCR assays were run on a CFX96 Touch RT-PCR System (Bio-Rad Laboratories, Hercules, CA, USA) with the procedures provided in Table S2.4a. Standard curves were created by serially diluting known amounts of PCR products of 16S rRNA, *mcrA* and *pmoA* genes from *Microcystis aeruginosa*, *Methanospirillum hungatei* and *Methylosinus trichosporium,* respectively, cloned into the pCR4-TOPO vector (Invitrogen, Carlsbad, CA, USA) and amplified with vector-specific primers. Efficiency (E) and R^2 were calculated from the standard curve using the Bio-Rad CFX Maestro software and summarized for each target gene: E = 80.6%, R^2=0.999 (16S rRNA); E = 90.7%, R^2=0.993 (*mcrA*); E = 97.5%, R^2=0.941 (*pmoA*).

Sequence processing and analysis

Paired-end reads were merged with USEARCH (Edgar, 2010) with the length restricted to 400-550 bp and a maximum error set to 1. The joined fragments with undetermined nucleotides were removed, and fragments were truncated where there were three consecutive low-quality nucleotides (quality score < 20), followed by discarding sequences that had less than 75% of their original length.

For archaeal 16S rRNA gene sequences, the remaining high-quality fragments were de-replicated and clustered into OTUs at the 97% similarity level with UPARSE (Edgar, 2013). Taxonomic annotations were assigned to each OTU sequence with the RDP Classifier (Wang *et al*., 2007) against the SILVA-SSUv132 Database (Quast *et al*., 2013),

and the OTUs annotated as Bacteria were removed. The samples with low sequencing depth and OTUs with an abundance below 0.005% were discarded. The potential methanogenic OTUs were identified based on the recent review on classification of methanogens (Evans *et al*., 2019).

For *mcrA* and *pmoA* sequences, the remaining high-quality sequences that did not match (using BLAST) a reference collection of *mcrA* (Euryarchaeota and Verstraetearchaeota) or bacterial *pmoA* sequences at a cutoff e value of 1e-10 were discarded. Functional gene OTUs were picked with UCLUST (Edgar, 2010) in QIIME (Caporaso *et al*., 2010) at a minimum sequence identity of 84% (*mcrA*) and 87% (*pmoA*) (Degelmann *et al*., 2010; Yang *et al*., 2014). The OTU nucleotide sequences were aligned with MAFFT (Katoh and Standley, 2013) and phylogenetic distances were calculated with QIIME. For the *pmoA* gene, taxonomic assignments of OTUs were made with UCLUST against Dumont's *pmoA* database (Dumont *et al*., 2014). To obtain detailed taxonomic assignments and phylogenetic relationships from *mcrA* OTUs, representative amino-acid sequences (translated with FrameBot (Wang *et al*., 2013)) from each OTU were extracted for performing a web-based BLAST search to retrieve closer relatives, which were combined into the reference collection mentioned above. After aligning all the sequences with MAFFT, the maximum likelihood phylogenetic tree of *mcrA* amino-acid sequences was constructed with RAxML 8.2.12 (Stamatakis, 2014) and organized in ARB v6.0.4 (Ludwig *et al*., 2004).

To determine the composition of methanotrophic communities among lakes using bacterial 16S rRNA genes, we used the non-cyanobacterial OTU table generated previously (Cook *et al*., 2020). The gamma-MOB were affiliated with three families (*Methylococcaceae*, *Methylomonaceae* and *Methylothermaceae*), while the alpha-MOB have recently been merged into *Beijerinckiaceae* (Parks *et al*., 2018). The OTU sequences of these families were extracted from the table (Cook *et al*., 2020).

Statistical analyses

Phylogenetic structure dissimilarities were compared for methanogenic communities in the five lakes using the weighted UniFrac distance and displayed in principal coordinates analysis (PCoA) plots. Procrustes analysis (Peres-Neto and Jackson, 2001) was conducted to determine the similarity of the PCoA between pairs of samples from 16S rRNA and functional gene datasets. The analysis provided a m^2 value to demonstrate similarity between the two datasets and associated *p*-value based on 999 Monte Carlo simulations. Non-parametric tests and PCoA plot were used to assess and display the phylogenetic dissimilarity of non-cyanobacterial communities between two groups of lakes. STAMP was used to identify the bacterial families which differed significantly between the two groups of lakes (Parks *et al*., 2014). Correlations between the relative abundance of each of the different families and that of combined methanotrophic lineages were further performed in the R statistical environment.

Data availability

Raw reads in this study have been deposited into the NCBI Sequence Read Archive (SRA) with the Project number PRJNA663707 (under the accession numbers SRR12649448-SRR12649525), and PRJNA575023 (under the accession numbers SRR10903090-SRR10903098). PRJNA575023 is available to public. PRJNA663707 can be accessed by the reviewers and editors via the following link: https://dataview.ncbi.nlm.nih.gov/object/PRJNA663707?reviewer=r81bq16sgeo6usaosq21c5ev9a. Once the manuscript is accepted and published, such data will be released to the public.

ACKNOWLEDGEMENTS

We thank Katherine V. Cook, Christopher Abin, Rosa Moon-Escamilla and Christopher Garner for helpful discussion, Arianto Santoso for measuring methane in Lake Rotoehu, Alan E. Wilson, Assaf Sukenik and Ann Chuang for providing samples. This work was supported by the National Science Foundation grant No. 1736255 under a subcontract from South Dakota School of Mines & Technology through the EPSCoR RII Track-2 FEC program, by DEB-1831061 through the Dimensions of Biodiversity program, and by a grant from the German Science Foundation (DFG-GR1540/21-2). Following were the author contributions. CL, LRK and KDH conceived of the study design; CL, LRK, HGB and MAT performed analysis; CL and LRK drafted the manuscript; CL, LRK, KDH, HPG critically revised the paper; HPG, MAB, DPH, HJ, DL, EIM, JP and RMZ provided samples and lake chemistry data. All authors gave final approval for publication and agree to be held accountable for the work performed therein.

Conflict of interest

There are no competing financial interests.

REFERENCES

Akins, L.N., Ayayee, P., and Leff, L.G. (2018) Composition and diversity of cyanobacteria-associated and free-living bacterial communities during cyanobacterial blooms. *Ann Microbiol* **68**: 493-503.

Angel, R., Claus, P., and Conrad, R. (2012) Methanogenic archaea are globally ubiquitous in aerated soils and become active under wet anoxic conditions. *ISME J* **6**: 847-862.

Auman, A.J., Speake, C.C., and Lidstrom, M.E. (2001) *nifH* sequences and nitrogen fixation in type I and type II methanotrophs. *Appl Environ Microbiol* **67**: 4009-4016.

Bae, H.S., Morrison, E., Chanton, J.P., and Ogram, A. (2018) Methanogens are major contributors to nitrogen fixation in soils of the Florida Everglades. *Appl Environ Microbiol* **84**: e02222-02217.

Bastviken, D., Cole, J.J., Pace, M.L., and Van de Bogert, M.C. (2008) Fates of methane from different lake habitats: Connecting whole-lake budgets and CH_4 emissions. *J Geophys Res* **113**: G02024.

Bastviken, D., Tranvik, L.J., Downing, J.A., Crill, P.M., and Enrich-Prast, A. (2011) Freshwater methane emissions offset the continental carbon sink. *Science* **331**: 50.

Batista, A.M.M., Woodhouse, J.N., Grossart, H.P., and Giani, A. (2019) Methanogenic archaea associated to Microcystis sp. in field samples and in culture. *Hydrobiologia* **831**: 163-172.

Beaulieu, J.J., DelSontro, T., and Downing, J.A. (2019) Eutrophication will increase methane emissions from lakes and impoundments during the 21st century. *Nat Commun* **10**: 1375.

Biderre-Petit, C., Jezequel, D., Dugat-Bony, E., Lopes, F., Kuever, J., Borrel, G. et al. (2011) Identification of microbial communities involved in the methane cycle of a freshwater meromictic lake. *FEMS Microbiol Ecol* **77**: 533-545.

Bizic, M., Klintzsch, T., Ionescu, D., Hindiyeh, M.Y., Gunthel, M., Muro-Pastor, A.M. et al. (2020) Aquatic and terrestrial cyanobacteria produce methane. *Sci Adv* **6**: eaax5343.

Blees, J., Niemann, H., Wenk, C.B., Zopfi, J., Schubert, C.J., Kirf, M.K. et al. (2014) Micro-aerobic bacterial methane oxidation in the chemocline and anoxic water column of deep south-Alpine Lake Lugano (Switzerland). *Limnol Oceanogr* **59**: 311-324.

Bodelier, P.L., Meima-Franke, M., Hordijk, C.A., Steenbergh, A.K., Hefting, M.M., Bodrossy, L. et al. (2013) Microbial minorities modulate methane consumption through niche partitioning. *ISME J* **7**: 2214-2228.

Bogard, M.J., del Giorgio, P.A., Boutet, L., Chaves, M.C., Prairie, Y.T., Merante, A., and Derry, A.M. (2014) Oxic water column methanogenesis as a major component of aquatic CH_4 fluxes. *Nat Commun* **5**: 5350.

Bowman, J.P., Sly, L.I., Nichols, P.D., and Hayward, A.C. (1993) Revised taxonomy of the methanotrophs: Description of *Methylobacter* gen. nov., emendation of *Methylococcus*, validation of *Methylosinus* and *Methylocystis* species, and a proposal that the family *Methylococcaceae* includes only the group I methanotrophs. *Int J Syst Evol Microbiol* **43**: 735-753.

Burford, M.A., and O'donohue, M.J. (2006) A comparison of phytoplankton community assemblages in artificially and naturally mixed subtropical water reservoirs. *Freshwater Biology* **51**: 973-982.

Cabello-Yeves, P.J., Zemskaya, T.I., Zakharenko, A.S., Sakirko, M.V., Ivanov, V.G., Ghai, R., and Rodriguez-Valera, F. (2020) Microbiome of the deep Lake Baikal, a unique oxic bathypelagic habitat. *Limnol Oceanogr* **65**: 1471-1488.

Cabrol, L., Thalasso, F., Gandois, L., Sepulveda-Jauregui, A., Martinez-Cruz, K., Teisserenc, R. et al. (2020) Anaerobic oxidation of methane and associated microbiome in anoxic water of Northwestern Siberian lakes. *Sci Total Environ* **736**: 139588.

Cai, H., Jiang, H., Krumholz, L.R., and Yang, Z. (2014) Bacterial community composition of size-fractioned aggregates within the phycosphere of cyanobacterial blooms in a eutrophic freshwater lake. *PLoS One* **9**: e102879.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K. et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335-336.

Chen, X., Jiang, H., Sun, X., Zhu, Y., and Yang, L. (2016) Nitrification and denitrification by algae-attached and free-living microorganisms during a cyanobacterial bloom in Lake Taihu, a shallow Eutrophic Lake in China. *Biogeochemistry* **131**: 135-146.

Cook, K.V., Li, C., Cai, H., Krumholz, L.R., Hambright, K.D., Paerl, H.W. et al. (2020) The global *Microcystis* interactome. *Limnol Oceanogr* **65**: S194-S207.

Degelmann, D.M., Borken, W., Drake, H.L., and Kolb, S. (2010) Different atmospheric methane-oxidizing communities in European beech and Norway spruce soils. *Appl Environ Microbiol* **76**: 3228-3235.

Deutzmann, J.S., Hoppert, M., and Schink, B. (2014a) Characterization and phylogeny of a novel methanotroph, *Methyloglobulus morosus* gen. nov., spec. nov. *Syst Appl Microbiol* **37**: 165-169.

Deutzmann, J.S., Stief, P., Brandes, J., and Schink, B. (2014b) Anaerobic methane oxidation coupled to denitrification is the dominant methane sink in a deep lake. *Proc Natl Acad Sci U S A* **111**: 18273.

Dumont, M., Lüke, C., Deng, Y., and Frenzel, P. (2014) Classification of *pmoA* amplicon pyrosequences using BLAST and the lowest common ancestor method in MEGAN. *Front Microbiol* **5**: 1-11.

Dunfield, P.F., Yuryev, A., Senin, P., Smirnova, A.V., Stott, M.B., Hou, S. et al. (2007) Methane oxidation by an extremely acidophilic bacterium of the phylum Verrucomicrobia. *Nature* **450**: 879-882.

Dziallas, C., and Grossart, H.P. (2011) Temperature and biotic factors influence bacterial communities associated with the cyanobacterium *Microcystis* sp. *Environ Microbiol* **13**: 1632-1641.

Dziallas, C., and Grossart, H.-P. (2012) Microbial interactions with the cyanobacterium *Microcystis aeruginosa* and their dependence on temperature. *Mar Biol* **159**: 2389-2398.

Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460-2461.

Edgar, R.C. (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* **10**: 996-998.

Ettwig, K.F., Shima, S., van de Pas-Schoonen, K.T., Kahnt, J., Medema, M.H., Op den Camp, H.J. et al. (2008) Denitrifying bacteria anaerobically oxidize methane in the absence of Archaea. *Environ Microbiol* **10**: 3164-3173.

Evans, P.N., Parks, D.H., Chadwick, G.L., Robbins, S.J., Orphan, V.J., Golding, S.D., and Tyson, G.W. (2015) Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science* **350**: 434-438.

Evans, P.N., Boyd, J.A., Leu, A.O., Woodcroft, B.J., Parks, D.H., Hugenholtz, P., and Tyson, G.W. (2019) An evolving view of methane metabolism in the Archaea. *Nat Rev Microbiol* **17**: 219-232.

Flynn, J.D., Hirayama, H., Sakai, Y., Dunfield, P.F., Klotz, M.G., Knief, C. et al. (2016) Draft genome sequences of gammaproteobacterial methanotrophs isolated from marine ecosystems. *Genome Announc* **4**: e01629-01615.

Garcia, S.L., Buck, M., McMahon, K.D., Grossart, H.P., Eiler, A., and Warnecke, F. (2015) Auxotrophy and intrapopulation complementary in the 'interactome' of a cultivated freshwater model community. *Mol Ecol* **24**: 4449-4459.

Gortz, H.D., and Brigge, T. (1998) Intracellular bacteria in protozoa. *Naturwissenschaften* **85**: 359-368.

Grossart, H.P., Frindte, K., Dziallas, C., Eckert, W., and Tang, K.W. (2011) Microbial methane production in oxygenated water column of an oligotrophic lake. *Proc Natl Acad Sci U S A* **108**: 19657-19661.

Guedes, I.A., Rachid, C., Rangel, L.M., Silva, L.H.S., Bisch, P.M., Azevedo, S., and Pacheco, A.B.F. (2018) Close link between harmful cyanobacterial dominance and associated bacterioplankton in a tropical eutrophic reservoir. *Front Microbiol* **9**: 1-13.

Günthel, M., Klawonn, I., Woodhouse, J., Bižić, M., Ionescu, D., Ganzert, L. et al. (2020) Photosynthesis-driven methane production in oxic lake water as an important contributor to methane emission. *Limnol Oceanogr* **9999**: 1-13.

Hanson, R.S., and Hanson, T.E. (1996) Methanotrophic bacteria. *Microbiol Rev* **60**: 439-471.

Haroon, M.F., Hu, S., Shi, Y., Imelfort, M., Keller, J., Hugenholtz, P. et al. (2013) Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage. *Nature* **500**: 567-570.

He, K., Wang, H.W., Hu, X.K., An, Y.F., Zhong, J.C., and Zhang, L. (2021) Emission fluxes and pathways of methane in different eutrophic areas of Lake Chaohu. *China Environmental Science* 1-13.

He, Y., Li, M., Perumal, V., Feng, X., Fang, J., Xie, J. et al. (2016) Genomic and enzymatic evidence for acetogenesis among multiple lineages of the archaeal phylum Bathyarchaeota widespread in marine sediments. *Nat Microbiol* **1**: 16035.

Hess, S., Suthaus, A., and Melkonian, M. (2016) "*Candidatus* Finniella" (*Rickettsiales*, *Alphaproteobacteria*), novel endosymbionts of viridiraptorid amoeboflagellates (Cercozoa, Rhizaria). *Appl Environ Microbiol* **82**: 659-670.

Ho, A., Angel, R., Veraart, A.J., Daebeler, A., Jia, Z., Kim, S.Y. et al. (2016) Biotic interactions in microbial communities as modulators of biogeochemical processes: methanotrophy as a model system. *Front Microbiol* **7**: 1-11.

Hoefman, S., van der Ha, D., Boon, N., Vandamme, P., De Vos, P., and Heylen, K. (2014a) Niche differentiation in nitrogen metabolism among methanotrophs within an operational taxonomic unit. *BMC Microbiology* **14**: 83.

Hoefman, S., van der Ha, D., Iguchi, H., Yurimoto, H., Sakai, Y., Boon, N. et al. (2014b) *Methyloparacoccus murrellii* gen. nov., sp. nov., a methanotroph isolated from pond water. *Int J Syst Evol Microbiol* **64**: 2100-2107.

Huisman, J., Codd, G.A., Paerl, H.W., Ibelings, B.W., Verspagen, J.M.H., and Visser, P.M. (2018) Cyanobacterial blooms. *Nat Rev Microbiol* **16**: 471-483.

Huser, B.A., Wuhrmann, K., and Zehnder, A.J.B. (1982) *Methanothrix soehngenii* gen. nov. sp. nov., a new acetotrophic non-hydrogen-oxidizing methane bacterium. *Arch Microbiol* **132**: 1-9.

Jankowiak, J.G., and Gobler, C.J. (2020) The composition and function of microbiomes within *Microcystis* colonies are significantly different than native bacterial assemblages in two North American lakes. *Front Microbiol* **11**: 1-26.

Jarrell, K.F. (1985) Extreme oxygen sensitivity in methanogenic archaebacteria. *BioScience* **35**: 298-302.

Junier, P., Kim, O.-S., Eckert, W., Casper, P., Imhoff, J.F., Witzel, K.-P., and Hadas, O. (2010) Methane- and ammonia-oxidizing bacteria at the chemocline of Lake Kinneret (Israel). *Aquatic Microbial Ecology* **58** 241-248.

Kalyuzhnaya, M.G., Lamb, A.E., McTaggart, T.L., Oshkin, I.Y., Shapiro, N., Woyke, T., and Chistoserdova, L. (2015) Draft genome sequences of gammaproteobacterial methanotrophs isolated from lake washington sediment. *Genome Announc* **3**: e00103-00115.

Kalyuzhnaya, M.G., Yang, S., Rozova, O.N., Smalley, N.E., Clubb, J., Lamb, A. et al. (2013) Highly efficient methane biocatalysis revealed in a methanotrophic bacterium. *Nat Commun* **4**: 2785.

Kampman, C., Hendrickx, T.L., Luesken, F.A., van Alen, T.A., Op den Camp, H.J., Jetten, M.S. et al. (2012) Enrichment of denitrifying methanotrophic bacteria for application after direct low-temperature anaerobic sewage treatment. *J Hazard Mater* **227-228**: 164-171.

Katoh, K., and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**: 772-780.

Kim, M., Shin, B., Lee, J., Park, H.Y., and Park, W. (2019) Culture-independent and culture-dependent analyses of the bacterial community in the phycosphere of cyanobloom-forming *Microcystis aeruginosa*. *Sci Rep* **9**: 20416.

Kits, K.D., Klotz, M.G., and Stein, L.Y. (2015a) Methane oxidation coupled to nitrate reduction under hypoxia by the Gammaproteobacterium *Methylomonas denitrificans*, sp. nov. type strain FJG1. *Environ Microbiol* **17**: 3219-3232.

Kits, K.D., Campbell, D.J., Rosana, A.R., and Stein, L.Y. (2015b) Diverse electron sources support denitrification under hypoxia in the obligate methanotroph *Methylomicrobium album* strain BG8. *Front Microbiol* **6**: 1-11.

Klawonn, I., Bonaglia, S., Bruchert, V., and Ploug, H. (2015) Aerobic and anaerobic nitrogen transformation processes in N_2-fixing cyanobacterial aggregates. *ISME J* **9**: 1456-1466.

Knief, C. (2015) Diversity and habitat preferences of cultivated and uncultivated aerobic methanotrophic bacteria evaluated based on *pmoA* as molecular marker. *Front Microbiol* **6**: 1-38.

Kojima, H., Tokizawa, R., Kogure, K., Kobayashi, Y., Itoh, M., Shiah, F.K. et al. (2014) Community structure of planktonic methane-oxidizing bacteria in a subtropical reservoir characterized by dominance of phylotype closely related to nitrite reducer. *Sci Rep* **4**: 5728.

Kormas, K.A., Vardaka, E., Moustaka-Gouni, M., Kontoyanni, V., Petridou, E., Gkelis, S., and Neofitou, C. (2010) Molecular detection of potentially toxic cyanobacteria and their associated bacteria in lake water column and sediment. *World J Microbiol Biotechnol* **26**: 1473-1482.

Leigh, J.A. (2000) Nitrogen fixation in methanogens: the archaeal perspective. *Curr Issues Mol Biol* **2**: 125-131.

Levine, U.Y., Teal, T.K., Robertson, G.P., and Schmidt, T.M. (2011) Agriculture's impact on microbial diversity and associated fluxes of carbon dioxide and methane. *ISME J* **5**: 1683-1691.

Lidstrom, M.E. (1988) Isolation and characterization of marine methanotrophs. *Antonie Van Leeuwenhoek* **54**: 189-199.

Liu, X., Li, M., Castelle, C.J., Probst, A.J., Zhou, Z., Pan, J. et al. (2018) Insights into the ecology, evolution, and metabolism of the widespread Woesearchaeotal lineages. *Microbiome* **6**: 102.

Louati, I., Pascault, N., Debroas, D., Bernard, C., Humbert, J.F., and Leloup, J. (2015) Structural diversity of bacterial communities associated with bloom-forming

freshwater Cyanobacteria differs according to the cyanobacterial genus. *PLoS One* **10**: e0140614.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar et al. (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363-1371.

Mantzouki, E., Lurling, M., Fastner, J., de Senerpont Domis, L., Wilk-Wozniak, E., Koreiviene, J. et al. (2018) Temperature effects explain continental scale distribution of cyanobacterial toxins. *Toxins (Basel)* **10**: 156.

Mayr, M.J., Zimmermann, M., Dey, J., Brand, A., Wehrli, B., and Burgmann, H. (2020) Growth and rapid succession of methanotrophs effectively limit methane release during lake overturn. *Commun Biol* **3**: 108.

Milucka, J., Kirf, M., Lu, L., Krupke, A., Lam, P., Littmann, S. et al. (2015) Methane oxidation coupled to oxygenic photosynthesis in anoxic waters. *ISME J* **9**: 1991-2002.

Moezelaar, R., and Stal, L.J. (1994) Fermentation in the unicellular cyanobacterium *Microcystis* PCC7806. *Arch Microbiol* **162**: 63-69.

Morris, S.A., Radajewski, S., Willison, T.W., and Murrell, J.C. (2002) Identification of the functionally active methanotroph population in a peat soil microcosm by stable-isotope probing. *Appl Environ Microbiol* **68**: 1446-1453.

Morrison, J.M., Baker, K.D., Zamor, R.M., Nikolai, S., Elshahed, M.S., and Youssef, N.H. (2017) Spatiotemporal analysis of microbial community dynamics during seasonal stratification events in a freshwater lake (Grand Lake, OK, USA). *PLoS One* **12**: e0177488.

Murase, J., and Sugimoto, A. (2005) Inhibitory effect of light on methane oxidation in the pelagic water column of a mesotrophic lake (Lake Biwa, Japan). *Limnol Oceanogr* **50**: 1339-1343.

Murase, J., and Frenzel, P. (2007) A methane-driven microbial food web in a wetland rice soil. *Environ Microbiol* **9**: 3025-3034.

Nercessian, O., Noyes, E., Kalyuzhnaya, M.G., Lidstrom, M.E., and Chistoserdova, L. (2005) Bacterial populations active in metabolism of C_1 compounds in the sediment of Lake Washington, a freshwater lake. *Appl Environ Microbiol* **71**: 6885-6899.

Oswald, K., Milucka, J., Brand, A., Hach, P., Littmann, S., Wehrli, B. et al. (2016) Aerobic gammaproteobacterial methanotrophs mitigate methane emissions from oxic and anoxic lake waters. *Limnol Oceanogr* **61**: S101-S118.

Oswald, K., Graf, J.S., Littmann, S., Tienken, D., Brand, A., Wehrli, B. et al. (2017) *Crenothrix* are major methane consumers in stratified lakes. *ISME J* **11**: 2124-2140.

Paerl, H.W., and Bebout, B.M. (1988) Direct measurement of O_2-depleted microzones in marine *Oscillatoria*: Relation to N_2 fixation. *Science* **241**: 442-445.

Paerl, H.W., and Pinckney, J.L. (1996) A mini-review of microbial consortia: Their roles in aquatic production and biogeochemical cycling. *Microb Ecol* **31**: 225-247.

Paerl, H.W., and Otten, T.G. (2013) Harmful cyanobacterial blooms: causes, consequences, and controls. *Microb Ecol* **65**: 995-1010.

Parks, D.H., Tyson, G.W., Hugenholtz, P., and Beiko, R.G. (2014) STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* **30**: 3123-3124.

Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.A., and Hugenholtz, P. (2018) A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* **36**: 996-1004.

Peres-Neto, P.R., and Jackson, D.A. (2001) How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia* **129**: 169-178.

Peter, H., Beier, S., Bertilsson, S., Lindstrom, E.S., Langenheder, S., and Tranvik, L.J. (2011) Function-specific response to depletion of microbial diversity. *ISME J* **5**: 351-361.

Pradeep Ram, A.S., Keshri, J., and Sime-Ngando, T. (2019) Distribution patterns of bacterial communities and their potential link to variable viral lysis in temperate freshwater reservoirs. *Aquatic Sciences* **81**: 72.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P. et al. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**: D590-596.

Raghoebarsing, A.A., Smolders, A.J., Schmid, M.C., Rijpstra, W.I., Wolters-Arts, M., Derksen, J. et al. (2005) Methanotrophic symbionts provide carbon for photosynthesis in peat bogs. *Nature* **436**: 1153-1156.

Reis, P.C.J., Thottathil, S.D., Ruiz-González, C., and Prairie, Y.T. (2020) Niche separation within aerobic methanotrophic bacteria across lakes and its link to methane oxidation rates. *Environmental Microbiology* **22**: 738-751.

Saidi-Mehrabad, A., Kits, D.K., Kim, J.J., Tamas, I., Schumann, P., Khadka, R. et al. (2020) *Methylicorpusculum oleiharenae* gen. nov., sp. nov., an aerobic methanotroph isolated from an oil sands tailings pond. *Int J Syst Evol Microbiol* **70**: 2499-2508.

Samad, M.S., and Bertilsson, S. (2017) Seasonal variation in abundance and diversity of bacterial methanotrophs in five temperate lakes. *Front Microbiol* **8**: 1-12.

Siakwa, A.N.K. (2018) Statistical techniques in analyzing the relationship between phytoplankton and environmental variables in Belső-tó, Tihany (Hungary). *Master Thesis (University of Pannonia, Hungary)*.

Smith, G.J., Angle, J.C., Solden, L.M., Borton, M.A., Morin, T.H., Daly, R.A. et al. (2018) Members of the genus *Methylobacter* are inferred to account for the majority of aerobic methane oxidation in oxic soils from a freshwater wetland. *mBio* **9**: e00815-00818.

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312-1313.

Tang, K.W., McGinnis, D.F., Ionescu, D., and Grossart, H.-P. (2016) Methane production in oxic lake waters potentially increases aquatic methane flux to air. *Environ Sci Technol Lett* **3**: 227-233.

Tang, K.W., McGinnis, D.F., Frindte, K., Brüchert, V., and Grossart, H.-P. (2014) Paradox reconsidered: Methane oversaturation in well-oxygenated lake waters. *Limnol Oceanogr* **59**: 275-284.

Tao, Y., Yuan, Z., Fengchang, W., and Wei, M. (2013) Six-decade change in water chemistry of large freshwater Lake Taihu, China. *Environ Sci Technol* **47**: 9093-9101.

Tavormina, P.L., Ussler, W., and Orphan, V.J. (2008) Planktonic and sediment-associated aerobic methanotrophs in two seep systems along the North American margin. *Appl Environ Microbiol* **74**: 3985-3995.

Thauer, R.K., Kaster, A.K., Seedorf, H., Buckel, W., and Hedderich, R. (2008) Methanogenic archaea: ecologically relevant differences in energy conservation. *Nat Rev Microbiol* **6**: 579-591.

Thottathil, S.D., Reis, P.C.J., del Giorgio, P.A., and Prairie, Y.T. (2018) The Extent and Regulation of Summer Methane Oxidation in Northern Lakes. *Journal of Geophysical Research: Biogeosciences* **123**: 3216-3230.

Vaksmaa, A., Guerrero-Cruz, S., van Alen, T.A., Cremers, G., Ettwig, K.F., Luke, C., and Jetten, M.S.M. (2017) Enrichment of anaerobic nitrate-dependent methanotrophic 'Candidatus Methanoperedens nitroreducens' archaea from an Italian paddy field soil. *Appl Microbiol Biotechnol* **101**: 7075-7084.

van Grinsven, S., Sinninghe Damste, J.S., Abdala Asbun, A., Engelmann, J.C., Harrison, J., and Villanueva, L. (2020) Methane oxidation in anoxic lake water stimulated by nitrate and sulfate addition. *Environ Microbiol* **22**: 766-782.

Vecherskaya, M., Dijkema, C., Saad, H.R., and Stams, A.J. (2009) Microaerobic and anaerobic metabolism of a *Methylocystis parvus* strain isolated from a denitrifying bioreactor. *Environ Microbiol Rep* **1**: 442-449.

Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**: 5261-5267.

Wang, Q., Quensen, J.F., 3rd, Fish, J.A., Lee, T.K., Sun, Y., Tiedje, J.M., and Cole, J.R. (2013) Ecological patterns of *nifH* genes in four terrestrial climatic zones explored with targeted metagenomics using FrameBot, a new informatics tool. *mBio* **4**: e00592-00513.

Watson, S.B., and Juttner, F. (2017) Malodorous volatile organic sulfur compounds: Sources, sinks and significance in inland waters. *Crit Rev Microbiol* **43**: 210-237.

Whittenbury, R., Phillips, K.C., and Wilkinson, J.F. (1970) Enrichment, isolation and some properties of methane-utilizing bacteria. *Microbiology* **61**: 205-218.

Wohl, D.L., Arora, S., and Gladstone, J.R. (2004) Functional redundancy supports biodiversity and ecosystem functioin in a closed and constant environment. *Ecology* **85**: 1534-1540.

Wu, X., Xi, W., Ye, W., and Yang, H. (2007) Bacterial community composition of a shallow hypertrophic freshwater lake in China, revealed by 16S rRNA gene sequences. *FEMS Microbiol Ecol* **61**: 85-96.

Xu, H., Paerl, H.W., Qin, B., Zhu, G., and Gaoa, G. (2010) Nitrogen and phosphorus inputs control phytoplankton growth in eutrophic Lake Taihu, China. *Limnol Oceanogr* **55**: 420-432.

Yang, C., Wang, Q., Simon, P.N., Liu, J., Liu, L., Dai, X. et al. (2017) Distinct network interactions in particle-associated and free-living bacterial communities during a *Microcystis aeruginosa* bloom in a plateau lake. *Front Microbiol* **8**: 1-15.

Yang, S., Liebner, S., Alawi, M., Ebenhoh, O., and Wagner, D. (2014) Taxonomic database and cut-off value for processing *mcrA* gene 454 pyrosequencing data by MOTHUR. *J Microbiol Methods* **103**: 3-5.

Yang, Y., Dai, Y., Wu, Z., Xie, S., and Liu, Y. (2016) Temporal and Spatial Dynamics of Archaeal Communities in Two Freshwater Lakes at Different Trophic Status. *Front Microbiol* **7**: 451.

Yang, Y., Chen, J., Tong, T., Li, B., He, T., Liu, Y., and Xie, S. (2019) Eutrophication influences methanotrophic activity, abundance and community structure in freshwater lakes. *Sci Total Environ* **662**: 863-872.

Yasuo, A., and Kawamura, S. (1984) Hydrogen evolution by *Microcystis aeruginosa* in darkness. *Agric Biol Chem* **48**: 2595-2596.

Zhang, L., Liao, Q., Gao, R., Luo, R., Liu, C., Zhong, J., and Wang, Z. (2021) Spatial variations in diffusive methane fluxes and the role of eutrophication in a subtropical shallow lake. *Sci Total Environ* **759**: 143495.

Zhu, B., van Dijk, G., Fritz, C., Smolders, A.J., Pol, A., Jetten, M.S., and Ettwig, K.F. (2012) Anaerobic oxidization of methane in a minerotrophic peatland: enrichment of nitrite-dependent methane-oxidizing bacteria. *Appl Environ Microbiol* **78**: 8657-8665.

Zohary, T., Sukenik, A., Berman, T., and Nishri, A. (2014) Lake Kinneret: Ecology and Management. *Springer, Heidelberg*.

References for supplementary tables only:

Angel, R., Claus, P., and Conrad, R. (2012) Methanogenic archaea are globally ubiquitous in aerated soils and become active under wet anoxic conditions. ISME J 6: 847-862.

Burford, M.A., and O'donohue, M.J. (2006) A comparison of phytoplankton community assemblages in artificially and naturally mixed subtropical water reservoirs. Freshwater Biology 51: 973-982.

Cook, K.V., Li, C., Cai, H., Krumholz, L.R., Hambright, K.D., Paerl, H.W. et al. (2020) The global Microcystis interactome. Limnol Oceanogr 65: S194-S207.

Evans, P.N., Boyd, J.A., Leu, A.O., Woodcroft, B.J., Parks, D.H., Hugenholtz, P., and Tyson, G.W. (2019) An evolving view of methane metabolism in the Archaea. Nat Rev Microbiol 17: 219-232.

Fischer, M.A., Gullert, S., Neulinger, S.C., Streit, W.R., and Schmitz, R.A. (2016) Evaluation of 16S rRNA gene primer pairs for monitoring microbial community structures showed high reproducibility within and low comparability between datasets generated with multiple archaeal and bacterial primer pairs. Front Microbiol 7: 1297.

He, K., Wang, H.W., Hu, X.K., An, Y.F., Zhong, J.C., and Zhang, L. (2021) Emission fluxes and pathways of methane in different eutrophic areas of Lake Chaohu. China Environmental Science 1-13.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., and Glockner, F.O. (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res 41: e1.

Kolb, S., Knief, C., Stubner, S., and Conrad, R. (2003) Quantitative detection of methanotrophs in soil by novel pmoA-targeted real-time PCR assays. Appl Environ Microbiol 69: 2423-2429.

McKay, L.J., Hatzenpichler, R., Inskeep, W.P., and Fields, M.W. (2017) Occurrence and expression of novel methyl-coenzyme M reductase gene (mcrA) variants in hot spring sediments. Sci Rep 7: 7252.

Neilan, B.A., Jacobs, D., Del Dot, T., Blackall, L.L., Hawkins, P.R., Cox, P.T., and Goodman, A.E. (1997) rRNA sequences and evolutionary relationships among toxic and nontoxic cyanobacteria of the genus Microcystis. Int J Syst Bacteriol 47: 693-697.

Siakwa, A.N.K. (2018) Statistical techniques in analyzing the relationship between phytoplankton and environmental variables in Belső-tó, Tihany (Hungary). Master Thesis (University of Pannonia, Hungary).

Steinberg, L.M., and Regan, J.M. (2008) Phylogenetic comparison of the methanogenic communities from an acidic, oligotrophic fen and an anaerobic digester treating municipal wastewater sludge. Appl Environ Microbiol 74: 6663-6671.

Tavormina, P.L., Ussler, W., and Orphan, V.J. (2008) Planktonic and sediment-associated aerobic methanotrophs in two seep systems along the North American margin. Appl Environ Microbiol 74: 3985-3995.

Winkel, M., Mitzscherling, J., Overduin, P.P., Horn, F., Winterfeld, M., Rijkers, R. et al. (2018) Anaerobic methanotrophic communities thrive in deep submarine permafrost. Sci Rep 8: 1291.

Zhang, L., Liao, Q., Gao, R., Luo, R., Liu, C., Zhong, J., and Wang, Z. (2021) Spatial variations in diffusive methane fluxes and the role of eutrophication in a subtropical shallow lake. Sci Total Environ 759: 143495.

Zohary, T., Sukenik, A., Berman, T., and Nishri, A. (2014) Lake Kinneret: Ecology and Management. Springer, Heidelberg.

TABLES

Groups[a]	Lakes	Diversity indices[b]			Relative abundance[c]
		Observed_OTUs	Shannon	Phylogenetic[d]	
Low	Aasee	45.7 ± 4.2	3.6 ± 0.3	6.8 ± 0.4	0.100% ± 0.037%
	Belső-tó	4.0 ± 1.4	1.3 ± 0.2	1.6 ± 0.4	0.044% ± 0.017%
	Chaohu	10.7 ± 4.6	1.9 ± 0.1	2.8 ± 0.6	0.050% ± 0.004%
	FP23	13.3 ± 2.4	2.8 ± 0.5	3.3 ± 0.3	0.162% ± 0.048%
	Grand Lake	10.0 ± 2.2	2.6 ± 0.2	2.9 ± 0.3	0.038% ± 0.004%
	Lake Kinneret	3.7 ± 1.2	1.4 ± 0.4	1.6 ± 0.4	0.019% ± 0.002%
	Lake Rotoehu	28 ± 7.3	3.6 ± 0.3	4.9 ± 0.9	0.194% ± 0.019%
High	Lake Clarendon	31.7 ± 2.9	3.2 ± 0.2	5.4 ± 0.3	2.296% ± 0.416%
	Lac de Villerest	61.3 ± 15	1.9 ± 0.6	8.6 ± 1.7	2.976% ± 0.287%
	Wentowsee	69.7 ± 23.3	3.7 ± 0.3	9.5 ± 2.6	2.355% ± 0.262%

Table 2.1. The diversity indices of methanotrophic communities and their relative abundance within the non-cyanobacterial community at each lake.

[a] Lakes were grouped into two categories: Low - 7 lakes with a low relative abundance of methanotrophs; High - 3 lakes with a higher relative abundance of methanotrophs.

[b] Estimation was based on *pmoA* sequence analysis.

[c] Estimation was based on 16S rRNA gene sequence analysis.

[d] Phylogenetic diversity determined using Faith's PD whole tree.

Identified lineage of methanogen from phylum Euryarchaeota	Taxonomic order of the lineage	Representative Strain	Metabolism type				Reference
			Hydrogenotrophic	Aceticlastic	Methylotrophic	Methylotrophic (H2-dependent)	
g_*Methanobacterium*	Methanobacteriales	*Methanobacterium formicicium DSM 3637*	√				(Evans *et al*., 2019)
g_*Methanocella*	Methanocellales	*Methanocella paludicola SANAE*	√				
g_Rice Cluster I[a]	Methanocellales		√				
g_*Methanolinea*	Methanomicrobiales	*Methanolinea* sp. *SDB*	√				
g_*Methanoregula*	Methanomicrobiales	*Methanoregula boonei* sp. *6A8*	√				
g_*Methanospirillum*	Methanomicrobiales	*Methanospirillum hungatei JF-1*	√				
o_Methanomicrobiales	Methanomicrobiales	*Methanomicrobium mobile BP*	√				
g_*Methanosaeta*	Methanosarcinales	*Methanosaeta concilii GP6*		√			
g_*Methanosarcina*	Methanosarcinales	*Methanosarcina acetivorans C2A*	√	√	√		
g_*Methanomethylovorans*[b]	Methanosarcinales	*Methanomethylovorans hollandica DSM 15978*	√	√	√		
o_Methanofastidiosales	Methanofastidiosales	*Ca. Methanofastidiosa methylthiophilus Bin006*[c]				√	
f_*Methanomassiliicoccaceae*	Methanomassiliicoccales	*Methanomassiliicoccus luminyensis B10*				√	

Table S2.1. The potential types of metabolism for the methanogenic lineages identified in this study. Each lineage was identified using both 16S gene and *mcrA* sequence analysis except where noted.

[a] Only identified here by archaeal 16S rRNA gene sequencing

[b] Only identified here by *mcrA* gene sequencing

[c] The metabolism was inferred from genomic evidence.

Groups	MRPP	*P*-value	ANOSIM	*P*-value	Adonis	*P*-value
Among ten lakes	0.8033	0.001	0.987	0.001	0.961	0.001
High vs Low	0.0915	0.001	0.399	0.002	0.165	0.001

Table S2.2. Overall non-cyanobacterial community structure comparison using three nonparametric statistical methods.

Abbreviations: MRPP, multiresponse permutation procedure; ANOSIM, analysis of similarity; Adonis, permutational multivariate analysis of variance;

High, lakes with a high abundance of methanotrophs (Clarendon; Villerest; Wentowsee);

Low, lakes with a low abundance of methanotrophs (the other 7 lakes).

Lakes ID	1	2	3	4	5	6	7	8	9	10
	Aasee	Belső-tó	Chaohu	Clarendon	FP23	Grand Lake	Kinneret	Rotoehu	Villerest	Wentowse e
Sampling Time	2017/07	2016/09	2016/05	2017/03	2016/12	2017/06	2017/02	2017/02	2016/10	2016/10
Latitude	51.963655	46.913568	31.551434	-27.509	32.64902	36.661	32.819005	-38.124045	45.981896	53.059841
Longitude	7.607909	17.901674	117.536792	152.348	-85.486988	-94.901	35.580765	176.256412	4.03621	13.22442
Area (km²)	0.42	0.32	770	3.4	0.001	188	168.7	7.9	30	2.79
Max Depth (m)	1.9	2.3	3.7	10	2	50	41.7	13.5	59	4
T (℃)	21.8	14.3 (3.8-28.1)	22.8 (7.5-36.3)[3b]	23.4	20.1 (9.7-27.8)	24	(15.5-28)[7a]	15.7 (9.2-23.5)	16.3	17
pH	9.31	8.9 (8.5-9.2)	8.67 (8.0-10.1)[3b]	8.7	8.4 (6.9-11.6)	8	(8.1-9)[7a]	7.8	7.4[9a]	>9.2
TN (mg/L)	>0.32[1a]	>7.613 (5.74-9.15)[2a]	4.268[3a]	2.1	1.52 (0.82-2.68)	1.28	NA	0.29	1.1[9c]	1.83[10a]
TP (µg/L)	293	245.2 (48.2-1340.9)	184[3a]	130	197 (20.5-308)	128	(15.5-93)[7a]	23.1	75[9c]	89[10a]
Secchi Depth (m)	0.5	0.3	0.34[3a]	1.8	0.23 (0.14-0.31)	0.73	(2.4-3.8)[7c]	3.5	2.2	0.4[10a]
Chl-α (µg/L)	127.5	79.2 (11.8-235.3)	23.9[3a]	14[4a]	77.3 (20.4-235)	6.5	(20-115)[7a]	5.8	23.9	90[10a]
DO (mg/L)	10.7	10.5 (8.5-12.7)	9.27[3a]	9.0	7.2 (0.3-11.6)	5.9	(7.0-12.2)[7a]	7.8 (6.1-9.8)	7.5	11.6
CH_4 (µM)	NA	NA	0.36[3a]	NA	NA	NA	NA	0.07 (0.03-0.12)[8a]	NA	< 0.1
DOC (mg/L)	NA	NA	7.3[3a]	31	NA	NA	3.91[7b]	3.3	7.6 (5.6-9.4)[9b]	30.6
NO_3-N (mg/L)	0.276	7.423 (5.61-8.86)	1.0 (0.417-1.88)[3b]	NA	0.53 (0-1.68)	0.52[6a]	(0-0.4)[7a]	0.0066[8b]	0.59[9c]	0.12
NO_2-N (µg/L)	16	24.8 (1.1-219.3)	NA	NA	62.2 (0-270)	NA	(0-44.8)[7a]	< 2	21[9c]	25
Stratification*	N	N	N	N	N	Y*[1]	Y*[2]	P	Y*[3]	N

Table S2.3. Table containing lake information and chemistry for each of the lakes studied here.

Note:

NA: Not Available.

*: Y: has a long period of distinct stratification each year; P: Polymictic – periods of stratification may occur but are very short; N: No significant stratification; *[1] Stratification from Jun to Sep; *[2] Stratification from Mar to Dec; *[3] Stratification from early Jun to late Sep.

Lake 1: All entries were collected on 2017 July 19th within a week of real sampling date; [1a] DIN value (TN unavailable).

Lake 2: All entries are annual means (range) of Nov 2016 - Oct 2017; [2a] DIN value (TN unavailable).

Lake 3: [3a] Annual means of June 2016 - May 2017; [3b] Seasonal mean (range) of Oct 2019 - Aug 2020.

Lake 4: [4a] One sampling point on Oct 2013; Other entries are annual means of 2011-2019.

Lake 5: All entries are annual means (range) of Nov 2018 - Oct 2019.

Lake 6: All entries are means of 2001 – 2013 summer (May to Aug); [6a] Combined data for (NO_3-N + NO_2-N)

Lake 7: [7a] (Range) of 2001-2011; [7b] Annual mean of 2000-2010; [7c] (Range) of 1990 - 2011

Lake 8: [8a] Mean (range) of 2014 summer (Jan to Mar); Other entries are annual mean of 2011 to 2014; [8b] Combined data for (NO_3-N + NO_2-N).

Lake 9: [9a] Collected on 2015 Oct 26th; [9b] Mean (range) of 2012 (Apr to Nov); [9c] Collected on 2016 Oct 18th; Other entries were on 2016 Oct 16th within a week of real sampling date.

Lake 10: [10a] Seasonal mean of 2014; Other entries were directly collected during sampling.

Data Resouces:

Lake 1:

Monitoring data kindly provided by City of Münster, Germany

Lake 2:

(Siakwa, 2018)

Lake 3:

(Zhang et al., 2021)

(He et al., 2021)

Lake 4:

Unpublished data recorded by local water authority and kindly provided by Dr. Burford lab from Griffith University.

The water authority's water quality data was also used in (Burford and O'donohue, 2006).

Lake 5:

Unpublished data recorded and kindly provided by Dr. Wilson lab from Auburn University and the water quality analyses methods are described and cited in their published papers (http://wilsonlab.com/pubs.html)

Lake 6:

Monitoring data kindly provided by Grand River Dam Authority (GRDA) of Oklahoma, USA

Lake 7:

(Zohary et al., 2014)

Lake 8:

Bay of Plenty Regional Council [BOPRC] 2016. Rotorua Lakes Water Quality Report 2014/2015.

Bay of Plenty Regional Council Environmental Publication 2016/06, Whakatane, New Zealand. (https://cdn.boprc.govt.nz/media/566926/rotorua-lakes-report-2014_2015.pdf).

Methane concentration was recorded and kindly provided by Dr. Arianto Santoso from the University of Waikato. New Zealand.

Lake 9:

From open database: OSUR (http://osur.eau-loire-bretagne.fr/exportosur/Accueil).

Lake 10:

Measured by the "Landesamt fuer Umwelt" of the county of Brandenburg, Germany with monitoring data at: https://mluk.brandenburg.de/w/seen/8000158152799.pdf.

Unpublished data recorded and kindly provided by Leibniz Institute of Freshwater Ecology and Inland Fisheries, Germany.

Primers	Target Gene	Sequence (5'-3')	Size [bp]	$T_{(exp)}$[a] & (no. of cycles)	Reference
Sequencing Illumina Miseq[b]					
mlas-mod-F mcrA-rev-R	Methanogens mcrA gene	GGYGGTGTMGGDTTCACMCARTA CGTTCATBGCGTAGTTVGGRTAGT	469	62-60 touchdown/55 final & (5/30)	(Angel *et al*., 2012; McKay *et al*., 2017)[c]
A189F-mod Mb661R-mod	Methanotrophs pmoA gene	GGNGACTGGGAYTTCTGG CHGGMGCAACGTCYTTACC	491	55 & (28)	This study modified from: (Kolb *et al*., 2003; Tavormina *et al*., 2008)[d]
S-D-Bact-0341-b-S-17 S-D-Bact-0785-a-A-21	Bacteria 16S rRNA gene	CCTACGGGNGGCWGCAG GACTACHVGGGTATCTAATCC	464	55 & (28)	(Klindworth *et al*., 2013)
S-D-Arch-0349-a-S-17 S-D-Arch-0786-a-A-20	Archaea 16S rRNA gene	GYGCASCAGKCGMGAAW GGACTACVSGGGTATCTAAT	457	56 & (28)	(Fischer *et al*., 2016; Winkel *et al*., 2018)
Quantitative PCR (qPCR)					
mlas-F mcrA-rev-R	Methanogens mcrA gene	GGTGGTGTMGGDTTCACMCARTA CGTTCATBGCGTAGTTVGGRTAGT	469	60 & (40)	(Steinberg and Regan, 2008; Winkel *et al*., 2018)
A189F Mb661R	Methanotrophs pmoA gene	GGNGACTGGGACTTCTGG CCGGMGCAACGTCYTTACC	491	55 & (40)	(Kolb *et al*., 2003)
MICR184F MICR431R	*Microcystis* 16S rRNA gene	GCCGCRAGGTGAAAMCTAA AATCCAAARACCTTCCTCCC	230	56 & (35)	(Neilan *et al*., 1997)

Table S2.4a. Oligonucleotides primers used in this study.

[a] The annealing temperature was used in the amplification reaction.

[b] For all of the PCR experiments in which libraries were prepared, the forward primers were modified with an overhanging M13 sequence on the 5' end.

[c] This primer set was developed from a database of 5200 *mcrA* sequences and demonstrated to recover well-known and novel methanogen *mcrA* sequences from the archaeal phyla Euryarchaeota and Verstraetearchaeota respectively.

[d] The *pmoA* primer set (A189F/Mb661R) has been previously used, but with limited coverage of methanotrophic lineages, while another *pmoA* primer set (wcpmoA189f/ wcpmo*A*661r) had higher recovery of *pmoA* genes especially from the water column (Tavormina *et al*., 2008). By taking advantage of the strengths of each primer set, we modified the primer set for this study (forward: A189F-mod, reverse: Mb661R-mod). This was done by replacing one nucleotide position with a degenerate nucleotide in both forward and reverse primers.

[e] DNA from *Methylosinus trichosporium* was used as a positive control for *pmoA*, while *Methanospirillum hungatei* DNA was a positive control for *mcrA* and archaeal 16S rRNA. No template PCR reactions were used as the negative control.

[f] qPCR procedure: Each reaction was performed in a total 20-µl reaction volume containing 10 µl of 2 × PowerUp SYBR Green Master Mix (Applied Biosystems), 1µl of each primer (10 µM) and 2 or 4µl of template DNA. *Microcystis*-specific 16S rRNA gene qPCR reactions were at 95℃ for 5 min, followed by 35 cycles of 95℃ for 30s, 56℃ for 45s, and 72℃ for 45s. The *mcrA* and *pmoA* qPCR reactions were at 95℃ for 5 min, followed by 40 cycles of 95℃ for 30s, 60℃ (*mcrA*) or 55℃ (*pmoA*) for 45s, and 72℃ for 45s. Melt curve analysis was performed from 55 to 95℃ with a 0.5℃ temperature increase per cycle (5s) at the end of each PCR run.

Primers	Target Region	Coverage of domain no Mismatch[a]			Coverage of methane metabolism microorganisms no Mismatch			
		Archaea	Bacteria	Eukarya				
					Methanotrophic lineages (in domain Bacteria)[b]			
S-D-Bact-0341-b-S-17 S-D-Bact-0785-a-A-21	BacV34	0.6%	86.6%	0	o_Methylococcales 93.6%	f_ *Beijerinckiaceae* 93.0%	g_ *Methylacidiphilum* 85.7%	g_ *Methylomirabilis* 96.2%
					Methanogenic lineages (in domain Archaea)[c]			
S-D-Arch-0349-a-S-17 S-D-Arch-0786-a-A-20	ArchV34	75.2%	0	0	o_Methanopyrales 83.3% o_Methanococcales 88.8%	o_Methanocellales 78.4 o_Methanobacteriales 85.2%	o_Methanomicrobiales 85.4% c_Thermococci 73.4%	o_Methanosarcinales 82.3% c_Thermoplasmata 79.7%

Table S2.4b. *In-silico* evaluation results of the primers targeting 16S rRNA gene.

The coverage of the primers was evaluated using the SILVA TestPrimer tool with the SILVA-Ref-NR database (version SSU r132).

[a] indicates the percentage of potentially covered taxa in 3 different domains while allowing no mismatches during primer annealing.

[b] indicates the percentage of potentially covered taxa in 4 different lineages containing aerobic methanotrophs while allowing no mismatches during primer annealing; the following taxonomic levels are used: o - order, f - family, g - genus.

[c] indicates the percentage of potentially covered taxa in 8 different lineages containing methanogens while allowing no mismatches during primer annealing; the following taxonomic levels are used: o - order, c - class.

FIGURE LEGENDS

Figure 2.1. Distribution of methanogenic lineages and other archaeal phyla across five lakes. **a** Composition of the archaeal community based on archaeal 16S rRNA gene sequencing. The blue color lineages in the dashed line represent the six methanogenic orders. The composition of the methanogenic community at the genus level is shown in **b** based on archaeal 16S rRNA gene sequencing and **c** based on *mcrA* gene sequencing. The following taxonomic levels are used: p - phylum, c - class, o - order, f - family, g - genus.

Figure 2.2. Maximum likelihood phylogenetic tree based on aligned partial amino-acid sequences of the *mcrA* gene. The tree was calculated with RAxML 8.2.12 using PROTMIX-JTT evolutionary model and using the rapid bootstrap algorithm with 100 iterations. Bootstrap values above 50% are displayed next to the nodes. Shaded clades contain the OTU sequences that were detected in this study. The clades were clustered into eight methanogenic orders as indicated by background color. The *mcrA* containing lineage from the phylum Bathyarchaeota was used as outgroup.

Figure 2.3. Distribution of methanotrophic lineages across ten lakes. The lakes are arranged with the three lakes with the highest number of methanotrophs on the right. The following taxonomic levels are used: o - order, f - family, g - genus. **a** The relative abundance of methanotrophic lineages in the non-cyanobacterial community as revealed by bacterial 16S rRNA gene sequencing. The scale bar shows the fraction of each lineage. A grey color indicates that the relative abundance was below 0.005%. The right bars are colored by family *Beijerinckiaceae* (red), *Methylomonaceae* (green), *Methylococcaceae*

(blue) and represent the number of OTUs within each group. **b** The relative abundance of methanotrophic lineages in methanotrophic communities as revealed by *pmoA* sequence analysis.

Figure 2.4. The normalized abundance of methanogen and methanotroph cell numbers relative to *Microcystis* cell numbers from five lakes where methanogens were detected. *Microcystis* was enumerated by qPCR of its 16S rRNA gene, methanogens using *mcrA* and methanotrophs using *pmoA* on bloom aggregate samples.

Figure 2.5. The abundance (quantitative) and richness (qualitative) of methanogens were both positively and linearly correlated with that of methanotrophs within bloom aggregate samples. **a** Linear regression of abundance between methanotroph cells and either methanogens or *Microcystis* cells. Calculations were based on the 5 lakes where all three types of organisms co-localized. **b** Linear regression of richness between methanotrophs and either methanogens or Cyanobacteria. Calculations included three individual samples from each of 10 lakes. The methanotrophs and methanogens were enumerated by targeting *pmoA* and *mcrA* respectively, while Cyanobacteria and *Microcystis* were enumerated based on the 16S rRNA gene. Richness data were based on sequence analysis and abundance using qPCR.

Figure 2.6. Bacterial families potentially interacting with methanotrophs. Note that X-axes increase from right to left to allow visualization of groups. **a)** Bacterial families with significant differences in abundance (in the non-cyanobacterial community) between the

high- and low-group lakes. The bacterial families with $P < 0.05$ (tested using nonparametric *t*-test) and average relative abundance > 0.1% in non-cyanobacterial community are listed. The highly abundant families are below the dashed line and use the bottom X-axis with the others using the upper X-axis. **b)** Spearman's rank correlation of the abundance between these families and methanotrophs. Black bars show a positive correlation while the gray bars show a negative correlation. The coefficient rho value from 0 to ±1 in the bottom axis denote the degree of the correlation from weak to strong. Only the families with correlation significances of $P < 0.05$ are shown.

Figure 2.7. Schematic overview demonstrating how methanotrophs potentially act as an interconnecting bridge between the C and N cycle in *Microcystis* aggregates. This schematic depicts an aggregate functioning in the absence of light. Dashed arrows indicate diffusion processes. Solid arrows indicate metabolic pathways which could be catalyzed by methanogens (blue)/methanotrophs (red)/nitrifiers (yellow)/others (black). Pathways are discussed in the text.

Figure S2.1. Distribution of the 10 lakes studied here on the world map (modified from (Cook *et al*., 2020)). The recovered bloom samples represent a 90° latitudinal and 270° longitudinal gradient. AU: Australia; CN: China; DE: Germany; FR: France; HU: Hungary; IL: Israel; NZ: New Zealand; US: United States. Samples were collected by our collaborators during the *Microcystis* blooms between May 2016 and July 2017.

Figure S2.2. High-quality reads of bacterial communities from 10 global lakes separated into non-Cyanobacteria and Cyanobacteria. The percentage label represents the relative abundance of Cyanobacteria in the whole bacterial community. The majority of Cyanobacteria (82.2% to 99.7%) in these lakes are classified as *Microcystis* sp.

Figure S2.3. Composition of methanogenic communities based on substrate metabolism type. **a** Based on archaeal 16S rRNA gene sequence analysis. **b** Based on *mcrA* sequence analysis.

Figure S2.4. Procrustes analysis of the phylogenic structure of methanogenic communities. Both datasets only included the five lakes with methanogens with three bio-replicates per lake (n = 15 for each dataset). The 15 blue balls at the end of the red line represent the structure of methanogenic community based on the archaeal 16S rRNA gene, while the other 15 blue balls at the end of the white line represent the corresponding lake samples determined by *mcrA*. m^2 is the sum of the squared phylogenic distances among 15 pairs of lake samples. The *p*-value is non-parametric and is based on 999 Monte Carlo simulations. As supported by *p* and m^2 values, the phylogenic structure of methanogenic communities determined by these two primer sets were similar.

Figure S2.5. Phylogenetic structure of methanogenic communities from five lakes as determined by *mcrA* sequencing. **a** Rarefaction curve of *mcrA* reads from five lakes with the rarefied depth at 2,650 sequences. Rarefaction curves reached asymptotes

demonstrating that we did capture the full sequence diversity. **b** Procrustes analysis of the phylogenic structure of methanogenic communities estimated by raw and rarefied datasets (cutoff = 2,650 sequences), corresponding to the 15 blue balls at the end of the red line and the white line respectively. The *p*-value is non-parametric and is based on 999 Monte Carlo simulations. As supported by low *p* and m^2 values, the rarified depth was sufficient to represent the native methanogenic community structure. **c** Principal coordinates analysis (PCoA) of the phylogenetic structure of methanogenic communities based on the rarified dataset. Based on the nonparametric statistical testing result (Adonis), the methanogenic communities of five lakes were structurally unique from each other.

Figure S2.6. Combined overview of the most abundant non-Cyanobacteria bacterial orders (outer ring) and MOB (inner bars) from the ten lakes. The outer ring color from darker to lighter represents the abundance of bacterial orders from high to low. The Type I and II methanotrophs are within the order Methylococcales (0.7% abundance of all non-Cyanobacteria) and are a subset of the family *Beijerinckiaceae* (11% of the total Rhizobiales) respectively. The left inner bar indicates the composition of bloom methanotrophs within the order Methylococcales; the right inner bar indicates the composition of bloom bacteria within the family *Beijerinckiaceae* containing Type II methanotrophs.

Figure S2.7. The abundance of methanogens, methanotrophs and *Microcystis* from five lakes. Due to the limited availability of template samples, each lake only had two bio-

replicates, except for lake Aasee having three bio-replicates. Error bars represent the standard deviation of technical triplicate samples. We assumed that each methanogen or methanotroph cell contained a single copy of the *mcrA* or *pmoA* gene respectively, while each *Microcystis* cell contains two copies of the 16S rRNA gene. *Microcystis* was more abundant than the methanogens and methanotrophs in all lakes. The methane cycling microbes were less abundant in lake Rotoehu than the other four lakes. Methanogens were more abundant than methanotrophs in all samples.

Figure S2.8. Principal coordinates analysis (PCoA) of weighted UniFrac beta diversity of the non-cyanobacterial community of the OTU table rarefied to 7105 read depth. Axis labels (PC1, PC2, PC3) indicate the percent variation explained by each axis. Low (n=21) and High (n=9) refers to the "7 lakes with a low abundance of methanotrophs" and the "3 lakes with a high abundance of methanotrophs" respectively, and with each lake having 3 bio-replicates. Dashed line marginally clusters the 3 lakes with a high abundance of methanotrophs. The high abundance lakes were resolved from the others in the PC1 vs PC3 and PC2 vs PC3 plots.

Figure S2.9. Distribution of non-cyanobacterial bacterial orders across the ten lakes. The three lakes enriched with methanotrophs are on the right. The orders within Alpha - and Gamma - proteobacterial classes are clustered together and indicated by red and blue background color, respectively. A $\log_2$ scale bar is used and a white color indicated that the order was present in less than 0.01% of the non-cyanobacterial community in each lake. The two arrows corresponded to the orders containing aerobic methanotrophs, and

one of the orders Methylococcales was enriched in the right three lakes, where their corresponding abundances are listed.

FIGURES

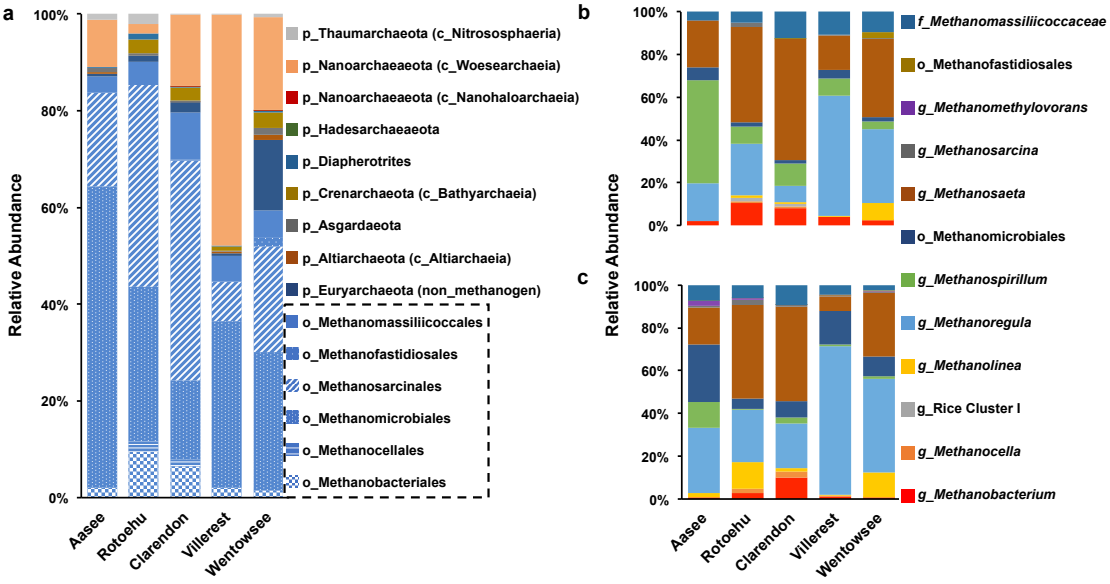


Figure 2.1. Distribution of methanogenic lineages and other archaeal phyla across five lakes.

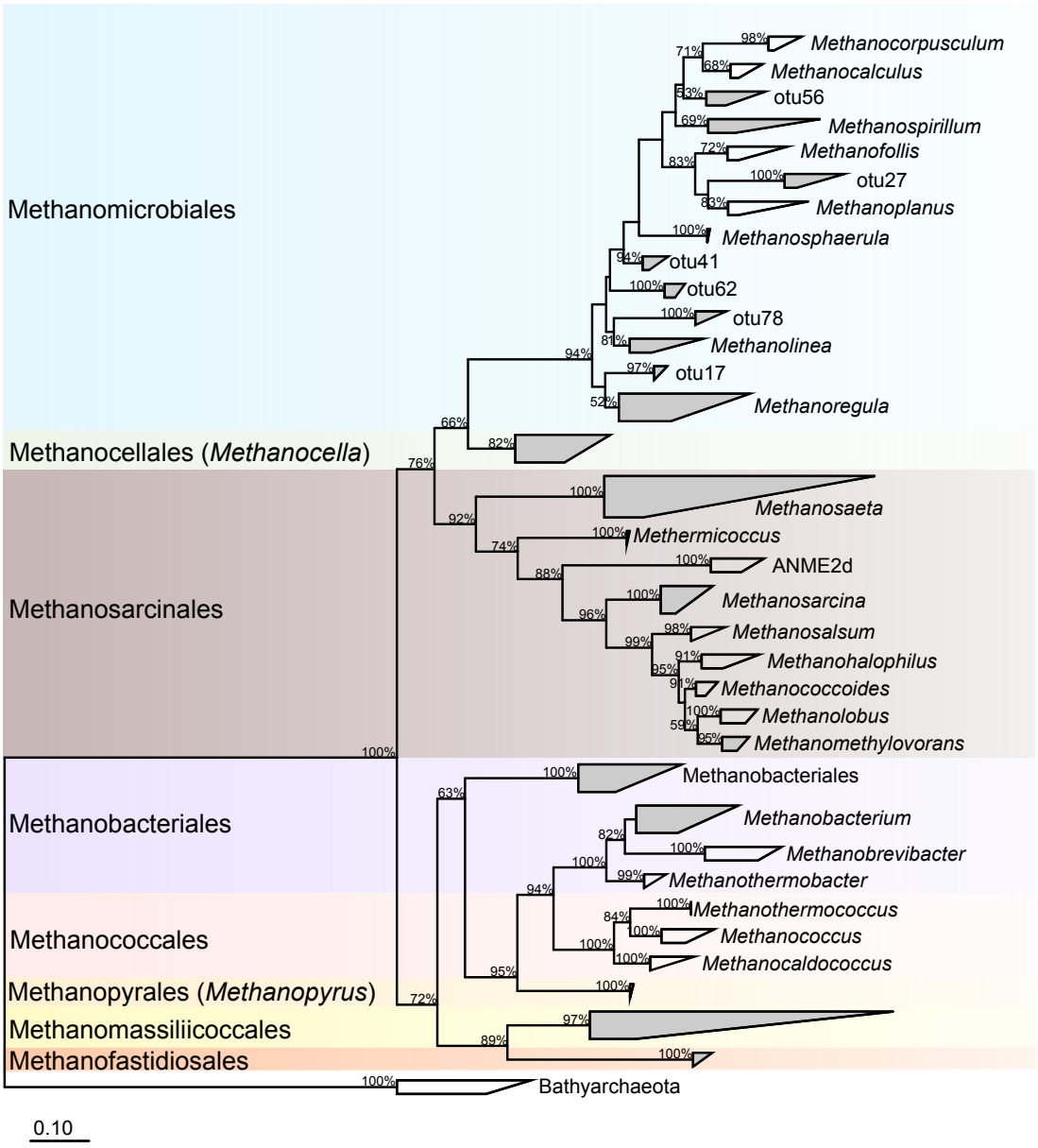


Figure 2.2. Maximum likelihood phylogenetic tree based on aligned partial amino-acid sequences of the *mcrA* gene.

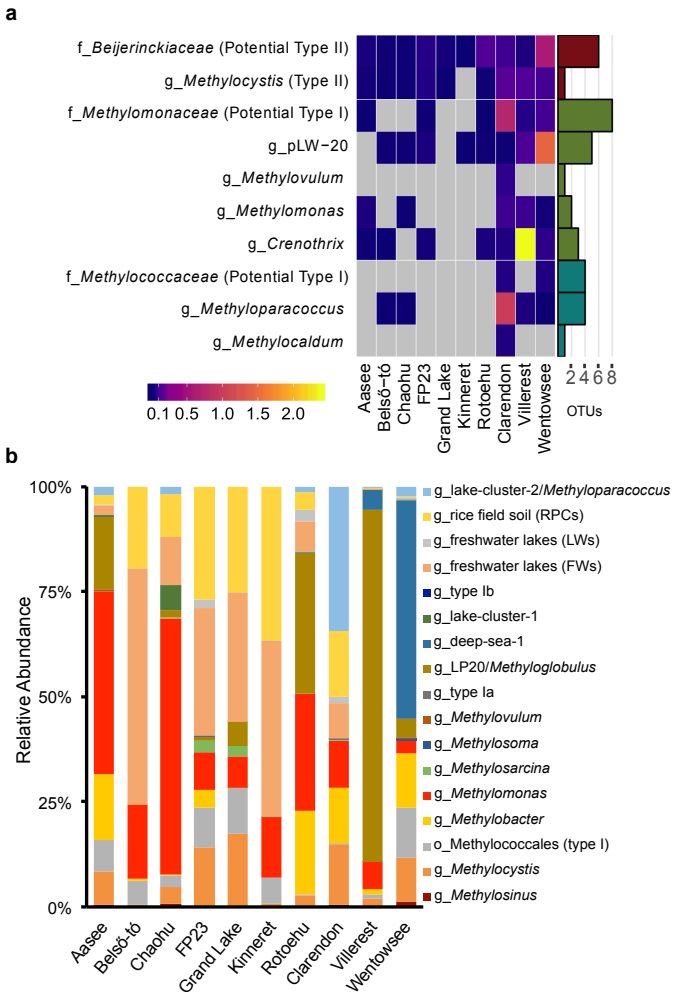


Figure 2.3. Distribution of methanotrophic lineages across ten lakes.

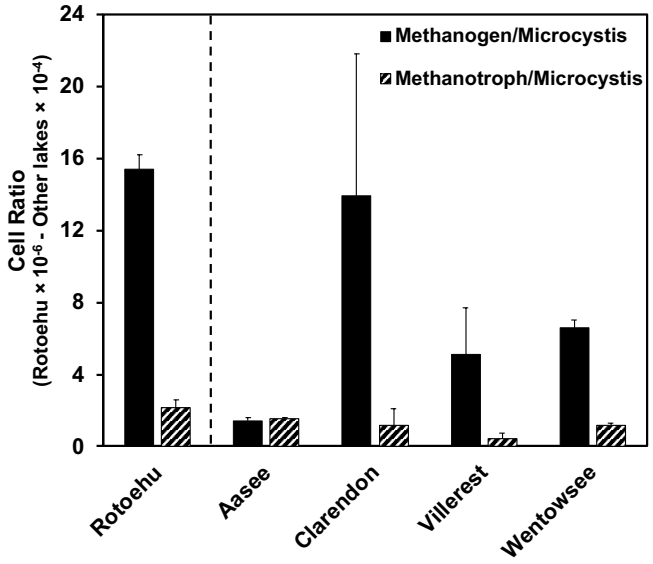


Figure 2.4. The normalized abundance of methanogen and methanotroph cell numbers relative to *Microcystis* cell numbers from five lakes where methanogens were detected.

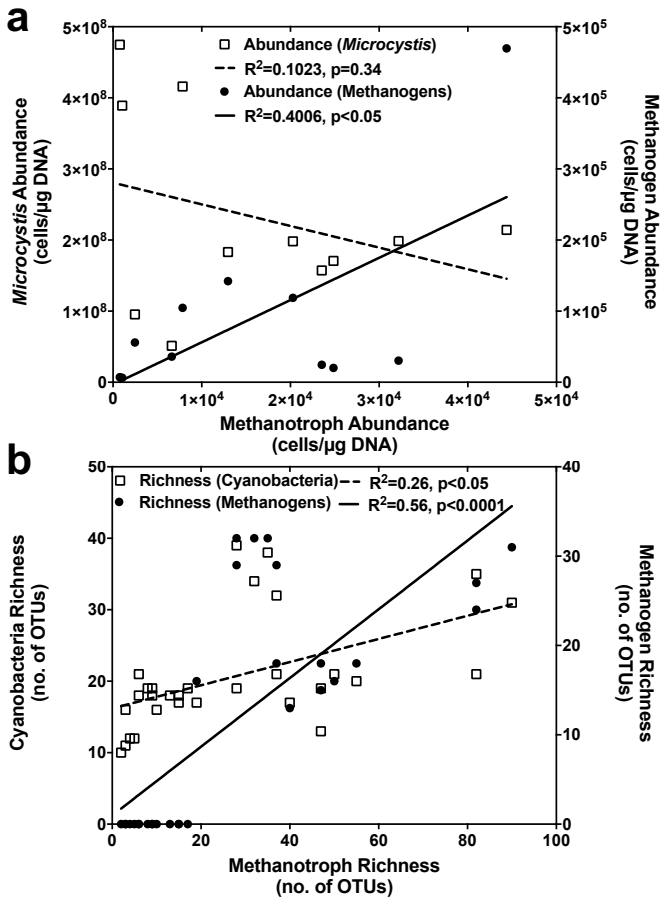


Figure 2.5. The abundance (quantitative) and richness (qualitative) of methanogens were both positively and linearly correlated with that of methanotrophs within bloom aggregate samples.

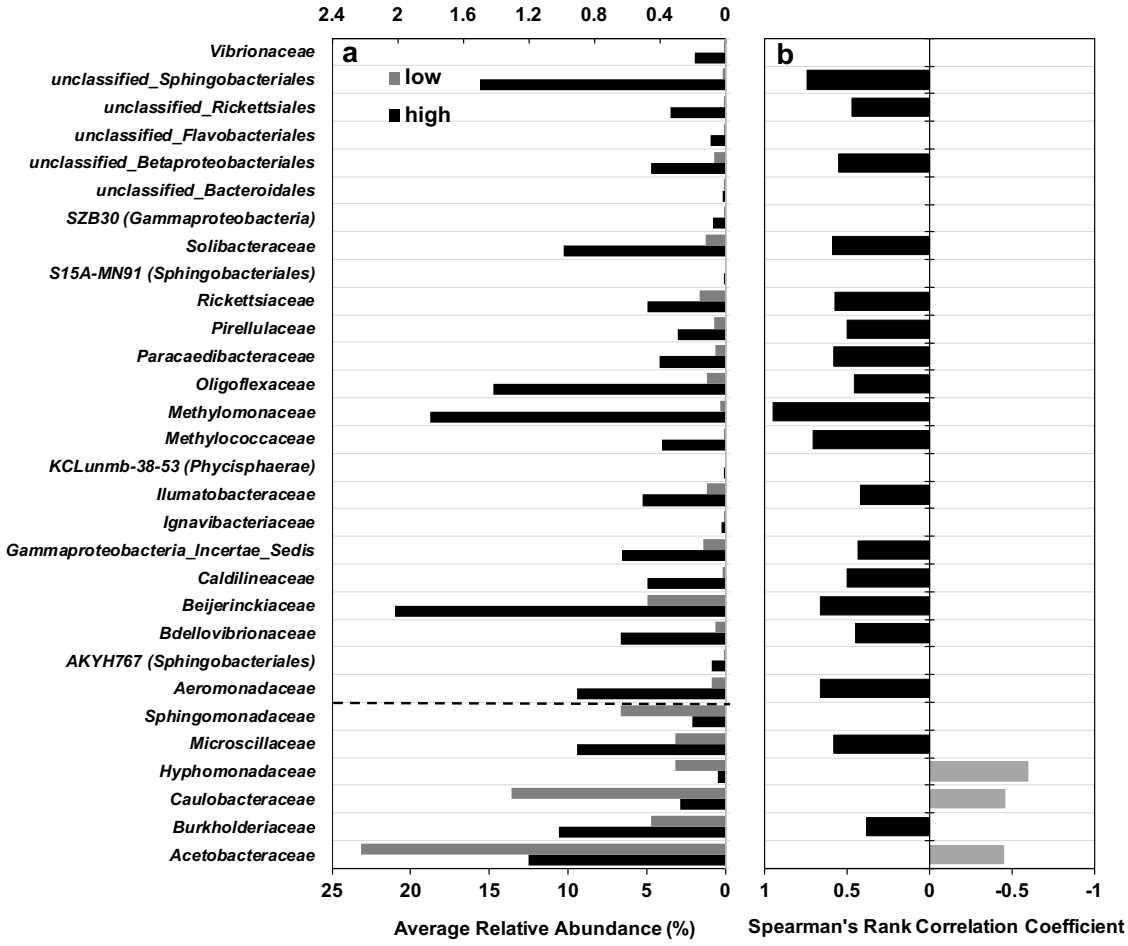


Figure 2.6. Bacterial families potentially interacting with methanotrophs.

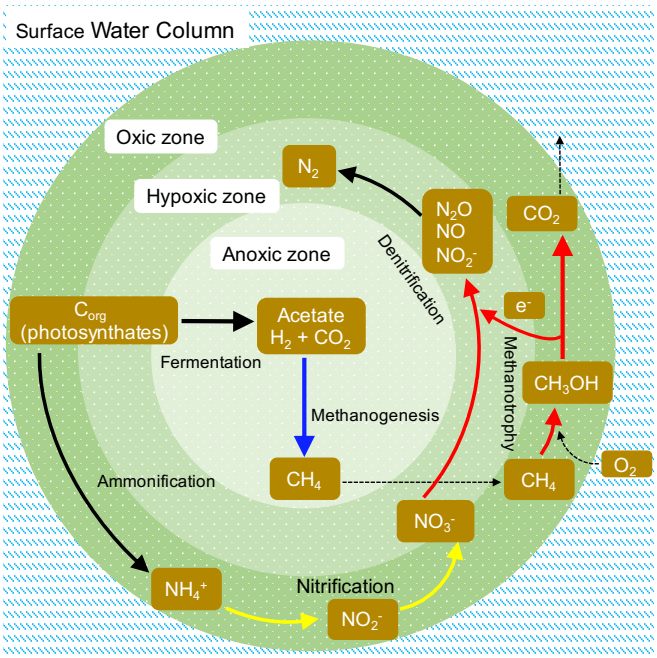


Figure 2.7. Schematic overview demonstrating how methanotrophs potentially act as an interconnecting bridge between the C and N cycle in *Microcystis* aggregates.

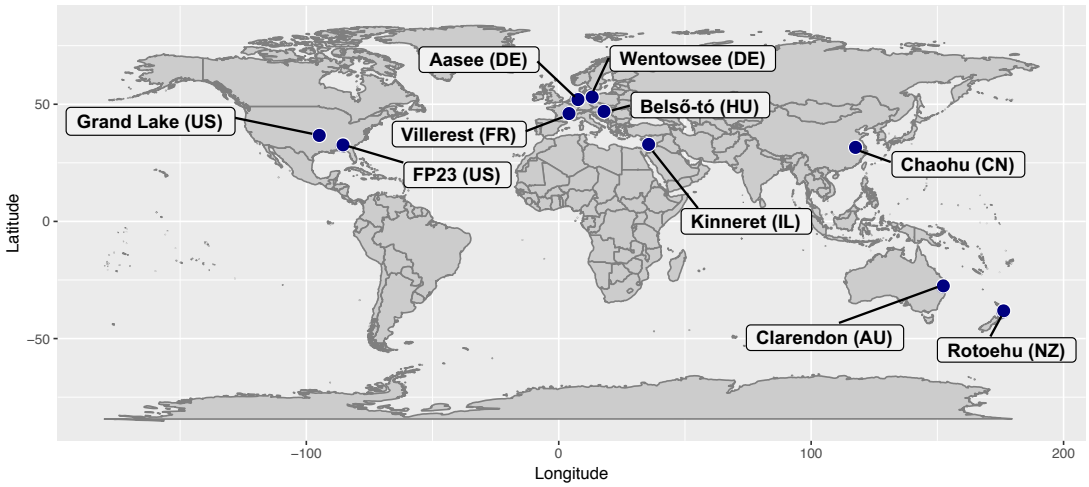


Figure S2.1. Distribution of the 10 lakes studied here on the world map.

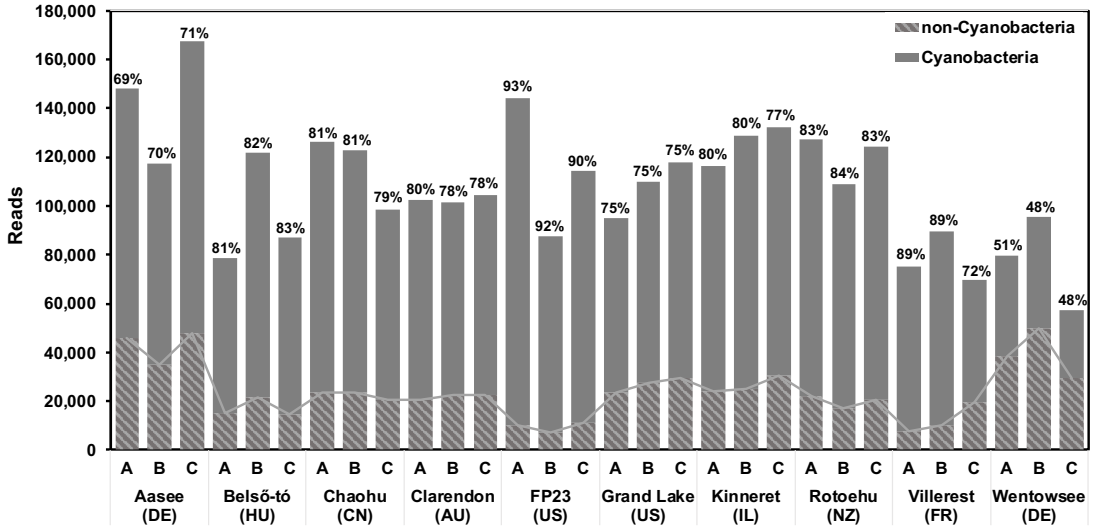


Figure S2.2. High-quality reads of bacterial communities from 10 global lakes separated into non-Cyanobacteria and Cyanobacteria.

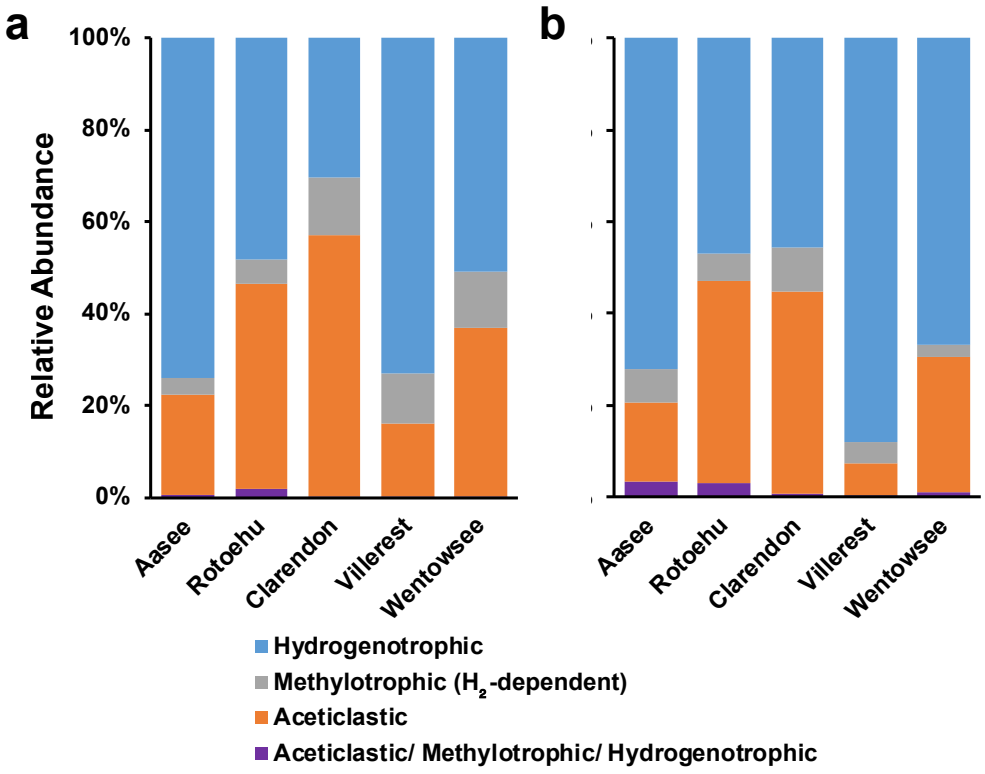


Figure S2.3. Composition of methanogenic communities based on substrate metabolism type.

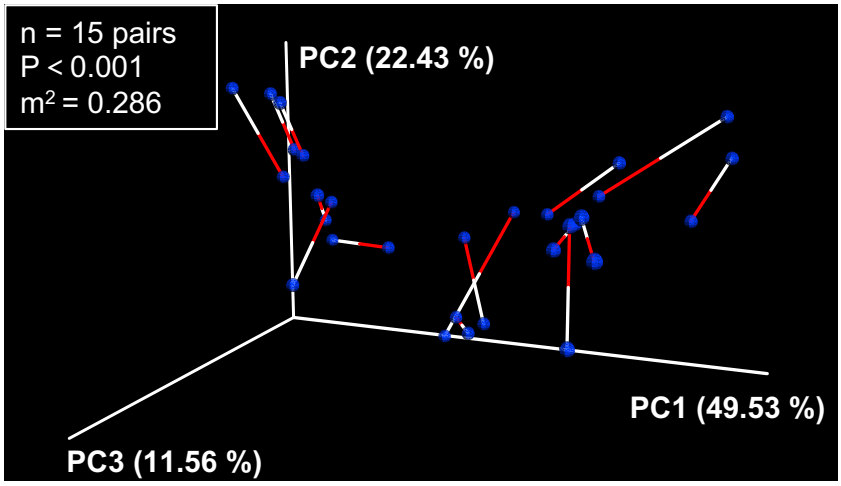


Figure S2.4. Procrustes analysis of the phylogenic structure of methanogenic communities.

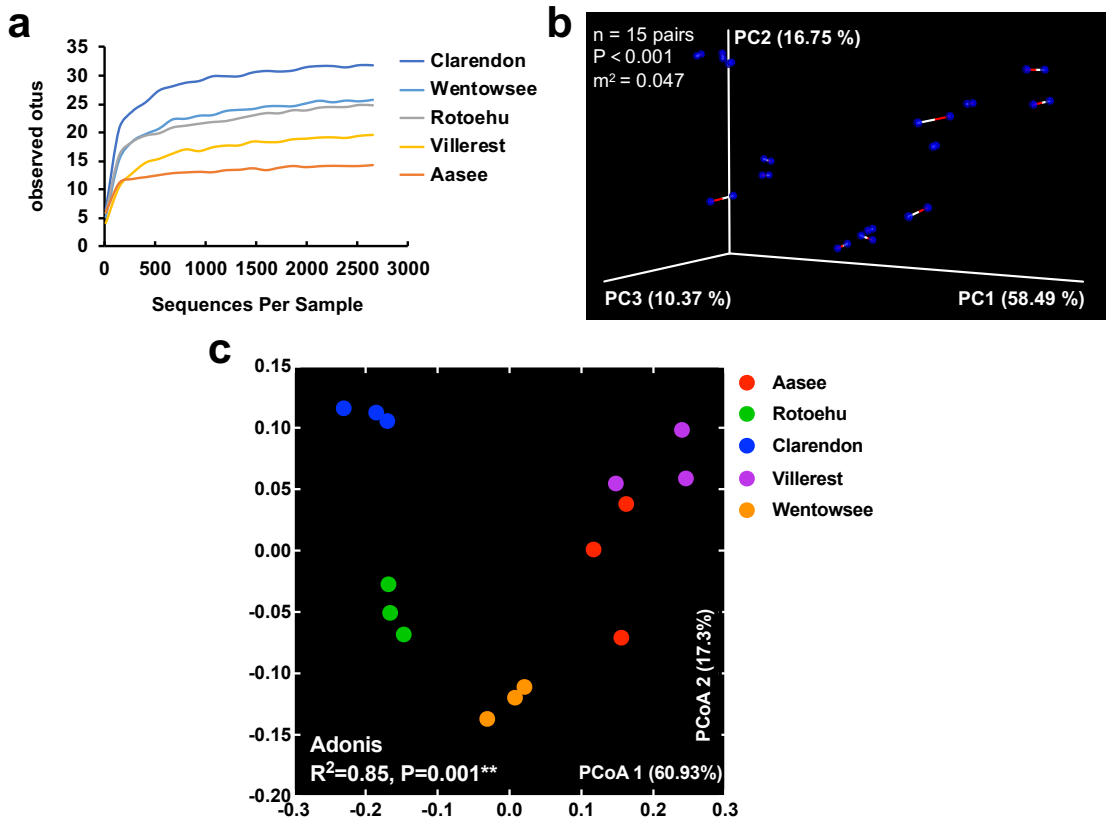


Figure S2.5. Phylogenetic structure of methanogenic communities from five lakes as determined by *mcrA* sequencing.

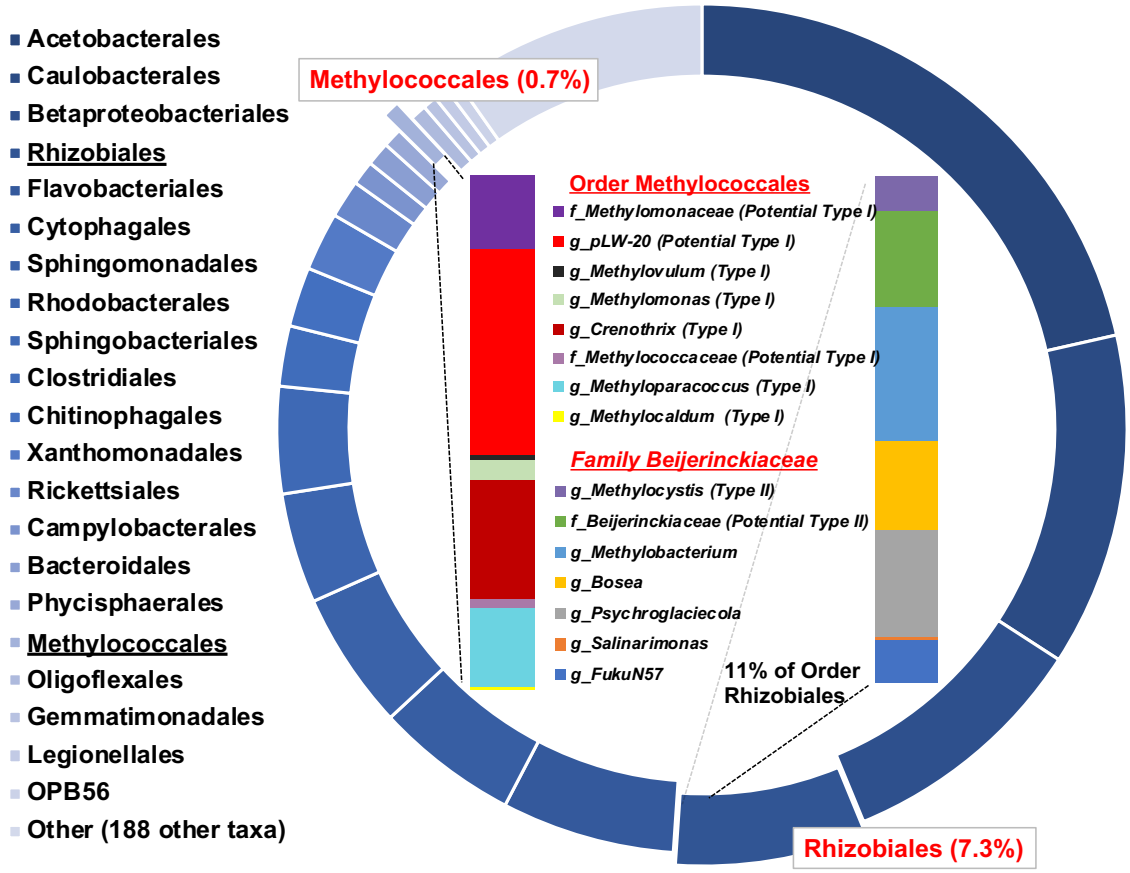


Figure S2.6. Combined overview of the most abundant non-Cyanobacteria bacterial orders (outer ring) and MOB (inner bars) from the ten lakes.

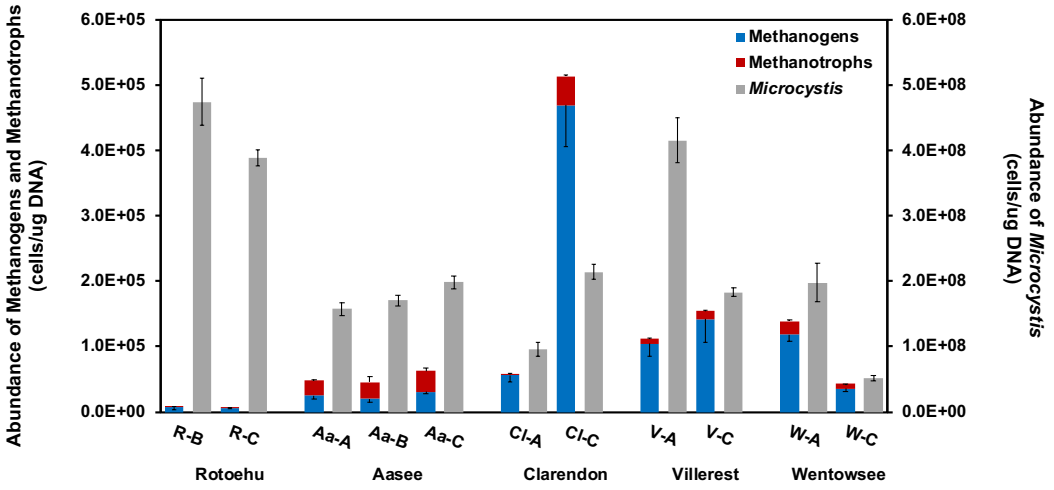


Figure S2.7. The abundance of methanogens, methanotrophs and *Microcystis* from five lakes.

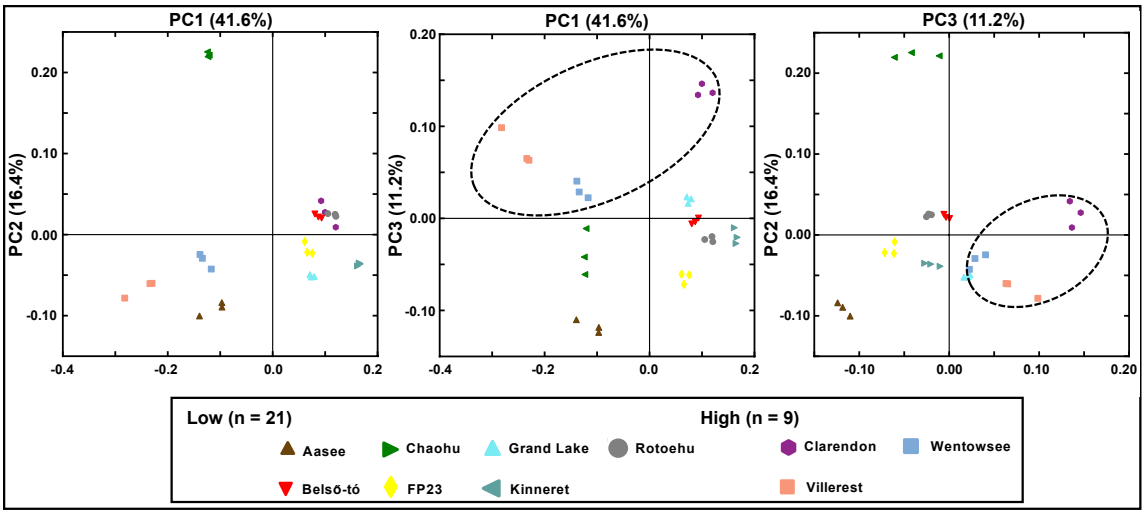


Figure S2.8. Principal coordinates analysis (PCoA) of weighted UniFrac beta diversity of the non-cyanobacterial community of the OTU table rarefied to 7105 read depth.

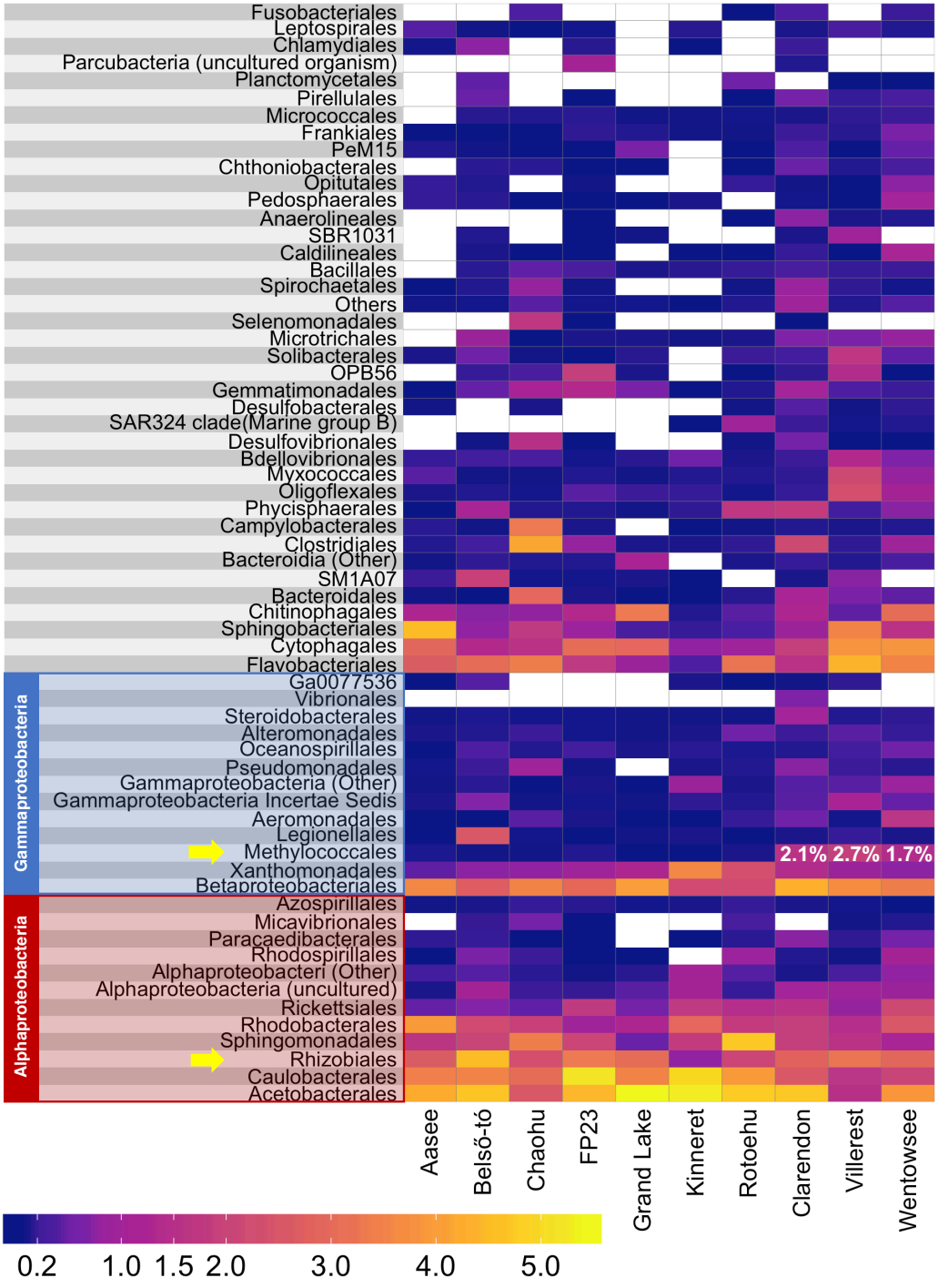


Figure S2.9. Distribution of non-cyanobacterial bacterial orders across the ten lakes.

CHAPTER 3: DIVERSITY AND ECOLOGICAL NICHE PREFERENCES OF SUBSURFACE-ADAPTED METHANOTROPHIC BACTERIA FROM THE DEEP BIOSPHERE OF A GOLD MINE

Chuang Li and Lee R. Krumholz

Keywords:

Subsurface; SURF; Mine; Methanotroph; Oligotrophy

In preparation for *Frontiers in Microbiology*

ABSTRACT

Methane is one of the most common energy and carbon sources in the deep subsurface biosphere, and methanotrophy is hypothesized to sustain the many diverse microbial communities alongside chemolithoautotrophy under oligotrophic conditions. The extent of the contribution of methanotrophy to primary production remains largely unknown, due to our lack of knowledge about the distribution, diversity and abundance of methanotrophs in subsurface ecosystems. Here, we specifically characterize the methanotrophic populations from borehole fluids and tunnel materials in the former Homestake gold mine. Through 16S rRNA and *pmoA* gene sequence analyses, a high diversity of aerobic methanotrophs within a broad range of clades (type Ia, Ib, Id, IIa, IIb, NC10 phylum) along with previously unreported novel lineages were observed at several locations in the mine. As type Ib Methanotrophs (mainly as *Methylococcus*) enriched in biofilm (up to ~27% in bacterial community), type IIa (mainly as *Methylocystis*) dominated in groundwater and several novel lineages exclusively occurred in underwater pink deposit, subsurface methanotrophs tended to occupy different ecological niches with specific geochemical conditions. Closest related sequence to several mine OTUs were exclusively observed in other subsurface habitats, suggesting these obligate subsurface-colonizing methanotrophs were indigenous inhabitants and potentially adapted to subterranean life.

Bacterial populations inhabiting in tunnel biofilms shared moderately similar community compositions with those in the same-depth groundwater, indicating methanotrophs trapped/concentrated in biofilms were likely originated from the feeding

groundwater. Based on the correlation patterns observed by network analysis, type Ib methanotrophs could be the primary producers in tunnel biofilms. We reconstructed ten methanotrophic genomes from metagenomes (MAGs) enabling the identification of the degree of novelty for subsurface methanotrophic strains. Some genomic characteristics of novel methanotrophs (e.g. with possession of the *pmoA2* and *mmoX* genes) are potentially involved in adapting to extreme conditions. This in-depth survey improved our understanding about the distribution of methanotrophs and revealed remarkably diverse, abundant and novel subsurface-adapted indigenous aerobic methane oxidizing bacteria (MOBs) were overlooked but present in deep subsurface biosphere, and their ecological roles serving as subsurface methane sink and food base has been previously underestimated.

INTRODUCTION

The terrestrial subsurface is an isolated oligotrophic ecosystem that often cannot acquire photosynthesis-derived energy and carbon resources. However, the biomass pools of the terrestrial subsurface biosphere are often comparable to those measured in surficial and marine ecosystems (Onstott et al., 1998; Whitman et al., 1998), and up to 30% of the total terrestrial subsurface biomass was estimated to exist at depth deeper than 2,000 ft below the surface ground (Whitman et al., 1998). In such deep subsurface oligotrophic ecosystems, hydrogen seems to be abundant and common, thus typically thought as the primary energy source to maintain the occurrence of microbial ecosystems (Lollar et al., 2014). The availability of hydrogen and inorganic carbon in groundwater apparently favors the growth of chemolithoautotrophic extremophiles, with ability to obtain energy through hydrogenotrophic reactions (including homoacetogenesis, methanogenesis and metal-dependent respiration) with along inorganic carbon fixation (Moser et al., 2005; Lin et al., 2006b; Lever et al., 2013; Magnabosco et al., 2016). Despite of several possible origins (e.g. migrated from the surface and buried detrital organic-rich deposits during sedimentation), dissolved organic carbon is generally scarce or even absent in most types of subsurface environments (Krumholz, 2000). Besides hydrogen and carbon dioxide, methane is another common constituent in deep subsurface environments with several geological and biogenic formation pathways. Regardless of origin, a total of ~3×10^{14} m^3 of gas hydrate (mainly present as CH_4) is estimated to occur in the global subsurface (Boswell and Collett, 2011). Alongside hydrogenotrophy, methane-derived carbons could constitute a large fraction of total terrestrial subsurface biomass, but their extent of the contributions are largely unknown.

Anaerobic methanotrophic archaea (ANME) and aerobic methane-oxidizing bacteria (MOBs) are two specialized groups that can utilize methane as the sole energy and carbon source. MOBs, mainly with representatives from alpha- and gamma-proteobacteria, are capable of converting methane to an array of bioavailable organic compounds, with higher free-energy yield than anaerobic methanotrophy. The roles of MOBs as primary producers in a broad range of surface ecosystems has been well-characterized, and the previous studies based on stable isotope experiment indicated that MOBs could provide carbon for associated microbes in soils (Ho et al., 2016), for Sphagnum mosses in peatlands (Raghoebarsing et al., 2005), for grazing microeukaryotes in rice paddy soils (Murase and Frenzel, 2007) and for zooplankton in lakes (Sanseverino et al., 2012).

Considering the limited availability of substrates in the oligotrophic subsurface, it is reasonable to hypothesize that methanotrophy-derived biomass could also provide energy and carbon fueling the subsurface microbial food chain. Over the past two decades, deep-living microbial communities have been widely studied in several systems including borehole fluids from a deep crystalline rock aquifer, groundwater and attached materials from caves, and borehole fluid and tunnel materials from mines (Bomberg and Ahonen, 2017). Specifically, although methanogenesis and distribution of methanogens (e.g. hydrogenotrophic and aceticlastic methanogens respectively inhibited in deep and shallow levels) were well characterized in deep crystalline rock biosphere, the occurrences of MOBs and methanotrophy were poorly documented. The marker genes

sequences of MOBs were only detected in limited sites with less frequency, implies that MOBs are not the major player in deep rock aquifer, probably due to the unfavorable deficient oxygen level normally encountered in groundwater (Kietavainen and Purkamo, 2015). Otherwise, studies targeting MOBs within a cave ecosystem suggest that assimilation of methane-derived carbon can sustain the whole microbial food web in air pockets where the habitats were in contact with methane and oxygen (Hutchens et al., 2004; Karwautz et al., 2018). Mine studies on the other hand, have largely focused on methanogenic archaea, suggested to produce CH_4 in South African mines, whereas comparably little information is available about the occurrence and ecological roles of MOBs in mine ecosystems (Takai et al., 2001; Ward et al., 2004; Lin et al., 2005; Moser et al., 2005; Gihring et al., 2006; Lin et al., 2006a; Lin et al., 2006b; Davidson et al., 2011). Although methanotrophic bacteria have not been a major focus in studies on mines, isotopic data indicated that incorporation of CH_4 into the biomass through methanotrophy occurred in borehole fluids ranging from 1.1 to 3.3 km below land surface in several South African mines, while fixation of CO_2 through chemolithoautotrophy were present as the dominant pathway at the deeper level (Simkus et al., 2016). The extent of the contribution of methanotrophs to methane sinks and to primary production within the subsurface can only be estimated if we were to develop a more comprehensive understanding of their distribution and diversity among various subsurface habitats.

The Homestake gold mine (South Dakota, USA), active for 125 years, was the deepest mine (tunnels as deep as 8,100 ft) and had the largest gold productivity (~1101 tons) in the North America (Caddey et al., 1991). The mining activities ceased in 2001,

and the site was transitioned into the Sanford Underground Research Facility (SURF) in 2007, providing a window for the geoscientists, physicists, biologists and engineers to explore the subsurface environments (Osburn et al., 2014). Although SURF has been studied to a lesser extent than the South Africa mines, previous microbiological studies at SURF have reported the geochemical conditions (Osburn et al., 2014), a high diversity chemolithotrophic community (Rastogi et al., 2009; Rastogi et al., 2010; Waddell et al., 2010; Osburn et al., 2014; Momper et al., 2017b), and annotated the metabolic potentials of bacteria (Momper et al., 2017a) within different types of samples collected in the mine. Methane and oxygen did not appear to be limiting as they were detected in SURF borehole fluids (Osburn et al., 2014), whereas no in-depth study has been performed on characterization of planktonic and attached methanotrophic communities in SURF mine.

In an effort to understand the distribution of deep-dwelling methanotrophic inhabitants, we first performed 16S rRNA and functional marker genes analyses on borehole fluids and attached tunnel materials each in different geochemical niches. The borehole fluids of mines are expected to share many similar features to groundwater in deep crystalline rock aquifer, while the field CH_4 and O_2 conditions of tunnels are comparable to those in cave air pockets. Thus, a comprehensive sampling effort could allow us to compare the occurrence of MOBs living in the similar geochemical conditions across different subsurface environments, as well as enable us to construct relationships between phylogenetic diversity and habitat preferences. For example, massive biofilms that occur in caves have also been noticed on the SURF tunnel walls and might affect methane oxidation in the deep biosphere through enhanced interactions,

thus the community-wide relationships were explored based on network analysis with the result showing the significant roles of MOBs in biofilm communities. Community composition relatedness were compared across different SURF samples to infer the hydrological connectivity and origins of microbes in biofilms. Metagenome assembled genomes (MAGs) with methanotrophic potential from Proteobacteria were recovered from the same DNA template, and phylogenomic, AAI and marker gene analyses were used to identify the taxonomic classifications and degree of novelty in relation to cultivated MOB strains. These methanotrophic MAGs were interrogated for their subsurface-adaptive capabilities.

MATERIALS AND METHODS

Field Sampling

SURF, also formerly known as the Homestake Gold Mine located in Lead (SD, USA), maintains 370 miles of underground tunnels with deepest level up to 8,100 ft under ground, serving as a well-known repository for exploration of an uncharacterized subsurface microbiome. The SURF samples discussed in this study were obtained during two sampling trips in 2018 and 2019 from deep (4,850 ft) and shallow (1,700 ft) levels (Fig. 3.1). Biofilm samples and fluids from boreholes were collected at the 4,850 ft level. The exact locations, geochemical and physical features for fluids in three boreholes (Borehole 8 - Fig. 3.2a, Borehole B - Fig. 3.2b, Borehole D - Fig. 3.2c) have been profiled in a previous study (Osburn et al., 2014). In particular, the relatively higher dissolved methane concentrations in fluids of the 3 boreholes gained our attentions. They served as the pioneering spots for characterization of planktonic methane-oxidizing communities and were sampled during our earlier sampling effort in 2018. Thick wall-hanging microbial biofilms, feeding on reducing fluids from the fractures in the tunnel wall, are located about 100 m away from borehole 8. Wall biofilms (Fig. 3.2d) were initially collected in 2018 sampling trip, while extra additional orange black biofilms (Fig. 3.2e) and white biofilms (from the same location as the wall biofilms) (Fig. 3.2f) were sampled in 2019. The orange black biofilms can be easily differentiated from other biofilms based on its unique color. In addition to samples originated from walls, other samples were retrieved at 1,700 ft level in 2019 expedition including: pink deposit (Fig. 3.2g) and brown crust (Fig. 3.2h), located under shallow water, white stalactite from the

top of tunnel (Fig. 3.2i), and white fungus (Fig. 3.2j) the only sample without direct contact by groundwater.

Planktonic communities from three boreholes were collected in the field by passing ~2 L of fracture fluids through 47 mm PES filter with 0.2 μm pore size (Sterile Tech, Queensbury, NY). Sampled filters were folded with cell side facing inward and inserted into 2 ml eppendorf tubes. Other samples were directly collected into 50 mL falcon tubes with pre-sterilized spatulas. All samples were immediately frozen on dry ice in the field, and then transported back to the laboratory and stored at -80 ℃ until further analysis.

DNA extraction, PCR amplification and amplicon sequencing

Prior to extraction, filters were cut into pieces to facilitate being submerged in lysis buffer. Genomic DNA was extracted from filter pieces and from other samples using DNeasy PowerSoil Pro Kits (Qiagen, Valencia, CA) with bead beating to homogenously lyse the cells. “No template” extraction and PCR reactions were used as negative controla to check for contamination. ZymoBIOMICS Microbial Community DNA Standard (Zymo Research, Irvine, CA) was used as a positive control for 16S rRNA gene analysis to assess sequencing bias, while DNA from *Methylosinus trichosporium* (OB3B) as a positive control for *pmoA* gene analysis. Primers (341F/785R) (Klindworth et al., 2013) were used to amplify the V3&V4 regions of the 16S rRNA gene in the domain Bacteria, and the *pmoA* primers (A189F/Mb661R) (Kolb et al., 2003) were used to amplify the functional marker gene of methanotrophic lineages.

Both primer sets were used successfully to recover methanotroph DNA sequence from environmental samples in our previous survey (see chapter 2), where the amplification conditions and library preparation processes were described.

Briefly, two rounds of PCR amplifications were conducted with the 1st round to recover the target gene from the sample, and 2nd round to attach the barcode sequences into both ends of amplicons. Each sample received a unique "barcodeF+barcodeR" combination sequence (see chapter 2) to allow demultiplexing post-pooling. Dual-indexed amplicons of all samples were pooled in equal amounts and purified by running on an agarose gel and then with the QIAquick Gel Extraction Kit (Qiagen, Valencia, CA). The purified mixed sample was then prepared using the Miseq Reagent Kit (v3) and sequenced with the Illumina Miseq platform (300 bp paired-end) at the Oklahoma Medical Research Foundation.

Amplicon sequence processing and phylogenetic analysis

The demultiplexed reads were merged using USEARCH v11.0.667 (Edgar, 2010) with the length restricted to 400-550 bp and maximum expected errors set to 1. Joined fragments with either undetermined nucleotides (Ns) or unmatched with primers were removed. Fragments were further truncated at positions where there were three consecutive low-quality nucleotides (Q-score less than 20). Barcode and primer sequence regions were removed and then short fragments were discarded using Cutadapt v1.18 (Martin, 2011).

For the bacterial 16S rRNA gene dataset, the remaining high-quality fragments were de-replicated and clustered into OTUs at the 97% similarity level with UPARSE (Edgar, 2013). Taxonomy was annotated up to the most specific level for each OTU sequence with the RDP's naive Bayesian rRNA Classifier (Wang et al., 2007) against the SILVA SSUv132 database (Quast et al., 2013), and the OTUs annotated as Archaea were removed. OTUs generated from sequenced positive and negative PCR controls were used to correct the noise-based counts, and OTUs with an abundance below 0.001% were discarded. The potential methanotrophic OTUs were determined and extracted based on the classification of methanotrophs summarized in a review (Knief, 2015) and three samples with low numbers of methanotrophic reads were excluded from the methanotrophic community profile. To determine the phylogenetic position of methanotrophic OTUs, reference sequences from relevant environmental clones and methanotrophic isolates, were aligned with query OTUs using the SINA online alignment tool (Pruesse et al., 2012). SINA-based alignments were validated in ARB v6.0.4 (Ludwig et al., 2004), and then a basic phylogenetic tree of nearly full-length reference sequences (> 1200 bp) was generated using ARB's neighbor-joining algorithm with Jukes Cantor correction. Query OTUs and 11 short reference sequences were subsequently added into the base tree with ARB parsimony methods, and the tree was displayed in iTOL (Letunic and Bork, 2019).

For the *pmoA* dataset, high-quality fragments not matching (using local BLAST) a reference collection of bacterial *pmoA* sequences at a cutoff e-value of 1e-10 were discarded. The *pmoA* gene sequences of methanotrophic isolates have been organized

before (Knief, 2015) and were downloaded here as the reference collection, consisting of members of the phyla Proteobacteria, Verrucomicrobia and NC10. OTUs were picked with UCLUST (Edgar, 2010) in QIIME v1.9.1 (Caporaso et al., 2010) at a minimum sequence identity of 87%, which approximately corresponds to the species level of methanotrophs (Degelmann et al., 2010). Preliminary taxonomic assignments of *pmoA* OTUs were made with UCLUST against Dumont's *pmoA* database (Dumont et al., 2014). The classified *pmoA* OTU sequences were translated to amino-acid sequences with the FrameBot program (Wang et al., 2013) in the RDP functional gene pipeline and those with frame shifts or an abundance below 0.01% were removed. To refine taxonomic identities and infer phylogenetic relatedness of *pmoA* OTUs, the closest neighbors of representative amino-acid sequences from each OTU were retrieved through conducting a web-based BLAST search, and were subsequently combined into the reference collection mentioned above. Following alignment of all the sequences with MAFFT v7 (Katoh and Standley, 2013), the maximum likelihood phylogenetic tree of *pmoA* amino-acid sequences was constructed with RAxML 8.2.12 (Stamatakis, 2014) on the CIPRES Science Gateway (Miller et al., 2010), and visualized in iTOL. The uncultivated lineages are identified in the context of a tree and named according to the phylogenetic affiliations of affiliated clones recorded in a recent study (Knief, 2015).

Diversity estimates, community relatedness, structure comparison (beta-diversity estimates) and statistical analysis

Various alpha diversity indices (Chao1, Shannon, PD_whole_tree) were determined on 16S rRNA gene generated libraries in QIIME after being rarified to a

maximum even sequence depth. The number of OTUs shared between all ten samples were visualized with a plot from UpsetR (Conway et al., 2017) to demonstrate the relatedness in community structure across groups of samples. Taxonomic structure dissimilarities were compared for bacterial communities in the three groups of samples (biofilms, fluids of boreholes, and four 1,700 ft samples) using Bray-Curtis distance and displayed in non-metric multidimensional scaling (NMDS) plots generated in the R statistical environment with the vegan package (Oksanen et al., 2020) and ggplot2 (Wickham, 2016). Adonis-tests (an analysis of variance) was performed to assess the significance of taxonomic dissimilarity of bacterial communities among these three groups of samples by running 1000 permutations in vegan.

Network construction and analysis

The OTU co-occurrence network was constructed to predict biotic interactions (mutualistic and competitive interactions implied by positive and negative correlations respectively) among bacterial populations in SURF biofilms and groundwaters. To generate reliable correlation patterns, the numbers of sequences in each sample were rarified to the same depth (15,000 and 40,000 reads for biofilm and groundwater bacterial groups respectively), and only OTUs with average abundance $\geq 0.05\%$ and present in at least 8 of 12 samples in each group were retained for downstream network construction. Pairwise Spearman's rank correlations were calculated for OTUs in each group based on their abundances using package Hmisc in R (Harrell, 2008). Strong correlations were validated by determining coefficient values $|r| \geq 0.6$ and p-value < 0.05, and those with strong correlations were selected for generating the network. Modules were determined

by the fast greedy modularity optimization using the igraph package in R (Csardi and Nepusz, 2006), and network graphs were created with Gephi 0.9.2 (Bastian et al., 2009) using a Fruchterman Reingold layout. The connectivity of nodes (OTUs) were decided by two topological indices: Z_i value (representing the connectivity of node i within-module) and P_i value (representing the connectivity of node i among modules). The nodes with higher Z_i value are classified as module hubs, while with higher P_i value are connectors (Guimera and Nunes Amaral, 2005).

Metagenomic sequencing, assembly and binning

Guided by amplicon analysis results, which showed us that some samples had high levels of novel methanotrophs, five samples: wall biofilm, orange black biofilm, white biofilm, borehole D and underwater pink deposit were chosen for shot-gun metagenomic sequencing. DNA concentration was determined with the Qubit dsDNA HS Assay kit on a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA). Library preparation and sequencing with the Illumina NovaSeq platform (SP, 150 bp paired-end) were done at the Oklahoma Medical Research Foundation.

Quality trimming on raw reads was conducted with Trimmomatic (Bolger et al., 2014) with default settings. High-quality reads (> 99% of raw reads retained) from each of the five samples were assembled individually into contigs in the *de novo* mode using megahit v1.1.2 (Li et al., 2015). Coverage depth information (% of reads that are matched to the assembled contigs) was calculated for contig by mapping with the high-quality short reads from the same samples using bowtie2 v2.2.9 (Langmead and Salzberg, 2012),

with the output converted through SAMtools v1.5 (Li et al., 2009) to the binary format required for binning. Metagenome assembled genomes (MAGs) were generated from contigs greater than 1500 bp using MetaBAT v2.12.1 (Kang et al., 2015). The quality of MAGs was validated using CheckM 1.0.18 (Parks et al., 2015) by referring to completeness (percentage of unique single-copy marker genes present within the bin) and level of contamination (multiple presence of a single-copy marker gene).

Taxonomic affiliations and novelty of MAGs

Putative taxonomic identities were assigned for each MAG by comparing sequence with the CheckM database. High-quality MAGs belonging to phylum Proteobacteria with completeness estimates of more than 80% and less than 5% contamination were selected for inferring their phylogenetic relationship with proteobacterial MOB type strains. The quality cut-off chosen here fulfills the minimum standards for characterization of uncultured species (Konstantinidis et al., 2017). Genome sequences of several proteobacterial MOB isolates were downloaded from NCBI GenBank database as phylogenetic references. Through the UBCG pipeline (Na et al., 2018), 92 bacterial core marker genes of query MAGs and reference genomes were identified using prodigal v2.6.3 (Hyatt et al., 2012) and HMMER v3.1b2 (Finn et al., 2011). Based on concatenated marker gene alignments in MAFFT, a maximum-likelihood phylogenomic tree was constructed using FastTree v2.1.10 (Price et al., 2010) with the JTT amino-acid substitution model and gene support indices (GSI) calculated to indicate how many core genes support each branch. Taxonomic identities of MAGs were

determined based on their placement in a phylogenomic tree in relation to closest reference genomes.

The MAGs located within methanotrophic clades on the phylogenomic tree were chosen to assess their degree of novelty based on Average Amino Acid Identity (AAI). Pairwise calculation of AAI with MOB reference genomes were performed with the Enveomics Toolbox (Rodriguez-R and Konstantinidis, 2016) and displayed with the pheatmap package (Kolde). If two query genomes were encountered with high AAI value, Average Nucleotide Identity (ANI) was used to evaluate the dissimilarity between these two related genomes. Taxonomic identities and novelty were further validated with information from conserved marker genes as follows. 16S rRNA genes and subunit genes of pMMO and sMMO (found only in MOB) were extracted from the MAGs based on annotation (see below). Their identities with closest MOB neighbors were determined via BLAST with the NCBI database. Two 16S rRNA gene sequences and three *pmoA* sequences derived from MAGs were further pooled with corresponding amplicon datasets, aligned together, and included in respective phylogenetic trees as described above.

Functional annotation

Protein-coding genes of MAGs were identified and annotated from contigs using Prokka v1.13 (Seemann, 2014). Annotated protein sequences were uploaded to the BlastKOALA (Kanehisa et al., 2016) and MAPLE v2.3.1 (Takami et al., 2016) servers, where KEGG orthology (KO) numbers were assigned with BLAST searches against a

nonredundant set of genes in the KEGG GENES database with the BBH (bi-directional best hit) method to identify orthologs. We imported the KO identifiers into the online KEGG mapper to enable reconstruction of associated KEGG metabolic pathways. In MAPLE, genes were grouped into functional modules defined by KEGG and the metabolic pathway completion was also assessed based on related functional module completion ratios (MCRs). The annotations and KEGG pathways were manually refined and compared to the reference genomes of closest MOB relatives as described above.

RESULTS

Composition of bacterial communities from diverse sample types and MOB enriched in biofilms

A total of 1,896,719 high-quality 16S rRNA gene amplicon reads were obtained from 40 SURF samples, with 4 bio-replicates for ten types of samples. As the composition of bacterial classes are generally consistent among bio-replicates (Fig. 3.3), their reads were collapsed together to represent community structure in each sample type. Planktonic communities within fluids of three boreholes were composed of much less Proteobacteria than attached communities within other types of tunnel samples (Fig. 3.5a). Strikingly, the fluid from borehole 8 harbored extremely diverse as well as rare taxa, along with the highest abundance of unassigned bacteria and members of the Patescibacteria superphylum (also known as Candidate Phyla Radiation, CPR), indicating that microbes in this site probably possess a high metabolic diversity and that their ecological roles are largely unexplored. Fluid from borehole B uniquely contains a large fraction of the novel phylum Acetothermia (former OP1). Proteobacteria along with Bacteroidetes are the major populations in all three biofilm samples (Fig. 3.5a). Particularly, the second most abundant class *Bacteroidia* are the well-known biofilm-forming bacteria, normally occurring at the late phase of biofilm formation (De Tender et al., 2017). At the shallow 1,700 ft level, the white fungus community was composed of primarily 3 bacterial classes (Fig. 3.5a), in parallel to the lowest alpha diversity in this habitat (Fig. 3.6). The composition of communities was nearly consistent across the remaining three samples from shallow level, which were distinguished from other samples with having more Actinobacteria members (Fig. 3.5a). By comparing diversity

measures across samples (white fungus excluded due to its unusual dry sample feature), diversity levels increased from deep to shallow level and the biofilm samples were generally the least diverse (Fig. 3.6). This diversity discrepancy could be attributed to the influence of distinct geochemical patterns associated with depth and different sample types.

A subset of 62,528 MOB-affiliated reads were filtered out from the 16S dataset for characterization of methanotrophic communities from 26 total samples (includes replicates), derived from 7 locations (Fig. 3.4). Otherwise, a total of 359,052 *pmoA* sequences were obtained from 30 individual samples, derived from 8 locations (Fig. 3.4). The effort of sequencing both 16S and pmoA marker genes allows us to successfully capture the majority of MOB lineages in subsurface samples. Two marker gene analysis supported that bacterial methanotrophs were most likely to be absent or occur in negligible numbers in fluid of borehole 8 and the underwater brown crust (Fig. 3.5b&c). Contrastingly, we were only able to detect *pmoA* genes of MOBs (i.e., *Methylocystis* and *Methylomonas*) in ceiling stalactite (Fig. 3.5c) probably due to the higher sensitivity of *pmoA* primer to environmental rare MOB templates. Methanotrophic members were enriched in biofilm samples with a relative abundance surprisingly reaching up to ~27%, and dominated by *Methylococcus* (Fig. 3.5b&c). In contrast to dominance of type I gamma-MOB in other samples, fluid of borehole D were nearly exclusively occupied by type II alpha-MOB (Fig. 3.5b&c). Interestingly, underwater pink deposit appears to possess a high diversity of novel methanotrophic potentials with the occurrence of many deeply branching uncultivated lineages including IheB2-23 and *Methylomirabilis* based

on 16S rRNA gene analysis (Fig. 3.5b) as well as LL_F11 from *pmoA* analysis (Fig. 3.5c), and the OTUs belonging to these lineages were distantly related to traditional well-known proteobacterial-MOB (see below).

Potential subsurface-adaptive MOB subclusters present in (type Ia, Ib, IIa, IIb and NC10) lineages

To better infer the phylogenetic relatedness of methanotrophic OTUs, information about the habitat from which the closest environmental relatives were collected is displayed in node names (Figs. 3.7&3.8). For 16S rRNA gene dataset, 15 OTUs were positioned in the tree with either closest to cultivated MOB types strains (Otu 2, 3698, 1167, 2676, 1711, 1209) or uncultivated clones (the other 9 OTUs), within type Ia, Ib, IIa, IIb and NC10 phyla (Fig. 3.7). Despite the high nucleotide sequence identity (98.8%) to two *Methylococcus* type strains, the most abundant Otu 6 in three biofilm samples was most closely related to the 16S rRNA sequence of subsurface clones retrieved from two mines located in South Africa and Russia with identity at 99.8% and 99.5% respectively, where they clustered together to develop into a distinct subclade within *Methylococcus* genus. Likewise, another 5 OTUs (Otu 2820, 134, 408, 1746, 1717) were phylogenetically related to clone sequences exclusively derived from other subsurface habitats, and OTUs together with associated clones formed into remarkably subsurface-specific clusters within a broad range of MOB lineages (*Methylomirabilis*-like, *Beijerinckiaceae*, *Methylococcaceae* and IheB2-23) (Fig. 3.7). As such, the lifestyle of these six OTUs-associated MOB strains could be especially adaptive for subsurface extreme conditions, but still tend to occupy different ecological niches in SURF mine, as

Otu 2820/1746/1717 appeared to be confined to underwater pink deposit, and Otu 6/408 was mainly present in orange black biofilm, as well as Otu 134 dominantly distributed in wall biofilm (Fig. 3.7).

Phylogenetic placement of SURF *pmoA* sequences

We targeted the *pmoA* genes from the same DNA extracts for further proof of the occurrence and phylogenetic relationships of the SURF MOB community. A total of 12 out of 14 *pmoA* OTUs were affiliated with clades of type Ia, Ib, IIa, IIb (Fig. 3.8) and was consistent with the finding in our 16S rRNA gene analysis, which strengthened that a high degree of MOB phylogenetic diversity was indeed present in subsurface environments. Unfortunately, we were unable to retrieve *pmoA* genes of *Methylomirabilis*-like member in NC10 phyla, detected by 16S rRNA gene analysis, due to its atypical bacterial *pmoA* that mismatched with the primers applied here. However, the *pmoA* primer did recover the type Id lineage named LL_F11 not detected in in the 16S libraries. The representative sequence of *pmoA* OTU *denovo3* classified within LL_F11 lineage showed only 72% amino-acid identity to the *pmoA* sequence from members of the type III clade, but interestingly was nearly identical (98% and 99%) to two subsurface *pmoA* clones. It is important to point out that denovo17 and denovo20, both enriched in fluids of borehole D were respectively identical to the *pmoA1* and *pmoA2* originating from the two operons of *Methylocystis sp. B8* strain (Fig. 3.8), implying these two OTUs were likely derived from the same strain inhabiting borehole D groundwaters. Although both OTUs (denovo14 and denovo29) within *Methylomonas genus* were closely related to cave sequences (Fig. 3.8), their *pmoA* sequences were

apparently resolved from each other due to low nucleotide identity (88%), and both were mainly retrieved from biofilms and white fungus habitats respectively.

Subsurface ecological niches of SURF methanotrophs

In addition, we connected marker gene diversity to habitat origins of the OTU sequences based on associated pie charts (Fig. 3.7&3.8). In general, the distribution pattern of both marker genes as denoted by the size of pie indicated that methanotrophic reads were mostly classified within type IIa and type Ib clades with noticeable larger pie size (Fig. 3.7&3.8). As further shown by the pie chart, type IIa-affiliated reads were mainly retrieved from fluids of borehole D, while type Ib-affiliated reads were predominately distributed in the 3 biofilm samples (Fig. 3.7&3.8), implying that in-site distinct geochemical differences between fluid and biofilm could lead to such skewed occurrence pattern. However, 16S rRNA gene read counts in the white biofilm was low (Fig. 3.7), but type Ib-affiliated *pmoA* read counts were at a comparable level to those in the other two biofilms (Fig. 3.8) probably because rare MOB members of the white biofilm were poorly recovered by 16S primers.

Comparisons of community composition

A total of 2,102 16S-based OTUs were obtained from ten SURF samples, and the numbers of OTUs in each sample varied from 79 from white fungus up to 1,095 for the underwater pink deposit (Fig. 3.9). To compare the community relatedness among samples, the numbers of OTUs shared among samples were explored using an upsetR plot. It is reasonable to envisage that the sampling depth and samples type associated

with various geochemical conditions could determine the occurrence and activity of microbes and further influence community composition. Thus, the shared patterns of the plot were color-coded based on sample types and depth (Fig. 3.9). The first 4 highest vertical black bars and 15 color-coded bars respectively demonstrated that a high number of OTUs were exclusively shared among shallow-level samples and among deep-level samples (Fig. 3.9), implying that the depth was indeed a relevant factor in controlling community composition. Sample type also played a role in community differences for the deep-level samples, as evident by 7 red-coded and 5 blue-coded patterns separately confined to groundwaters and biofilms (Fig. 3.9). However, the effect of sample types was limited as the community structure of groundwaters and biofilms still demonstrated a degree of relatedness, as denoted by the orange-coded pattern (Fig. 3.9). Such relatedness relationship was further reinforced by the ten OTUs shared by all 3 biofilms and groundwaters (Fig. 3.9).

A upsetR plot validates community relatedness from a qualitative perspective by considering the presence/absence of OTUs. By also taking into account the abundance, non-metric multidimensional scaling (NMDS) analysis integrated with Bray-Curtis distance was applied to compare and visualize dissimilarities in community structure between all samples. The NMDS plot revealed that samples from biofilms, fluids of boreholes and shallow level samples were clustered into three groups with distinct nonoverlapping 95% confidence intervals (Fig. 3.10). This observation agrees with the dissimilarity test (Adonis) showing significant differences in the taxonomic structure of

bacterial communities (at the OTU level) between each pair of three groups (*p*-value <0.001).

Network characteristics and association patterns of MOB

OTU-based correlation networks were generated to predict the interactive relationships among bacterial species in biofilms and groundwaters. All biofilms and groundwaters contain a total of 536 and 1,107 16S derived OTUs, which were used for network construction. To facilitate the comparisons, identical thresholds were applied to the two networks. The size of the biofilm network was smaller than groundwater network (85 nodes Vs 126 nodes) (Fig. 3.11a & 3.12), but with only 5 shared nodes, indicating the potential interactions occurring in biofilms and groundwaters likely mediated by distinct bacterial populations. Either positive or negative correlation could occur between nodes in the network, as represented by red and blue edges respectively. The nodes in biofilm network were more positively correlated, as 65% of connections were represented by red edges (Fig. 3.11a), whereas positive (52%) and negative correlation (48%) types had almost equal weights in groundwater network (Fig. 3.12). Such varied ratio of correlation types implied microbes inhabiting in biofilms tend to cooperate each other for enhancing the community-wide productivity, while competitive interactions was important for planktonic microbes to acquire substrates from oligotrophic groundwaters. For the biofilm network, there were 85 nodes linked by 1253 edges, and most of the nodes were affiliated into Gammaproteobacteria (23.5%), Alphaproteobacteria (17.6%) and Bacteroidetes (16.5%). To further explore the correlation patterns, the networks were clustered into modules (color-coded) and the connectivity (within-module and among-

module) of each node were calculated. Four nodes (i.e., OTUs 187-*Thiobacillus*, 13-*Methylophilaceae*, 6-*Methylococcus* 78-*Methylococcaceae*) likely served as hubs in their own modules having larger within-module connectivity than other nodes (Fig. 3.11b). Another node, OTU 408-*Methylococcaceae* connects to 13 nodes and 20 nodes respectively from orange and purple coded modules. It had high among-module connectivity (Fig. 3.11), and it could act as connector between these two modules. For the groundwater network, 126 nodes were linked by 2777 edges and mainly affiliated into Deltaproteobacteria (22.2%), Firmicutes (11.1%) and Actinobacteria (10.3%) (Fig. 3.12). The node OTU 101-Anaerolineales had the highest within-module connectivity, while another node OTU 233-*Desulfobacteraceae* had the highest among-module connectivity (data not shown).

In contrast to MOB nodes absent in the groundwater network, six MOB nodes were present and placed into 3 modules in the biofilm network, which were further utilized to identify MOB-associated bacteria. Five MOB nodes harbored more positive correlations (as from 62% to 76% of connections represented by red edges) with less positive correlations for node OTU138, implying MOB preferred to construct mutualistic interactions with associated bacterial members. Given the gamma-MOB was the dominant population in biofilms (Fig. 3.5b), likely having the potential to sustain the microbial ecosystems, it was more appealing to identify the bacterial members tightly associated with gamma-MOB. All three gamma-MOB nodes (OTUs 6/78/243) were commonly and positively correlated with 12 other nodes, strongly supporting that these

species are tightly associated with MOB and methanotrophy likely have an important role in modulating bio-interactions in biofilms.

Identification of methanotrophic MAGs

Shot-gun sequencing was performed on the three biofilms, fluids from borehole D and the underwater pink deposit as these samples had abundant MOB. The sequencing effort produced tens of millions of paired-end reads for each sample, ranging from 25,046,138 (for white biofilm) to 42,091,175 (for pink deposit). These quality-filtered short reads were *de-novo* assembled into long contigs for individual samples, generating lowest and highest numbers of contigs for the white biofilm (137,688) and pink deposit (1,617,635) respectively. A total of 357 MAGs were recovered from the contigs for all five samples. A subset of 26 high-quality MAGs (with >80% completeness and <5% contamination) preliminarily identified as Gammaproteobacteria and Alphaproteobacteria in CheckM were chosen for identifying their phylogenetic relationships with cultivated MOB type strains.

Up-to-date bacterial core gene sets were extracted from MAGs and translated into amino-acid sequences. Phylogenomic analysis of concatenated marker genes showed that 5 and 2 proteobacterial MAGs respectively derived from underwater pink deposit and white biofilm were distantly related to MOB type strains and not included in the tree. Meanwhile, nine MAGs (5 from orange black biofilm, 3 from wall biofilm and 1 from fluid of borehole D) were phylogenetically related to the MOB type strains (Fig. 3.13 & 3.15), while the remaining ten MAGs recovered from these 3 samples were closer to non-

MOB lineages in Proteobacteria phylum (data not shown). Along with phylogenomic identification, genomic information of these nine MAGs and MOB type strains was also used to calculate pair-wise AAI values. The gamma-MOB phylogenomic tree showed that MAGs BF-4-SURF.Bins.6 and OBB-3-SURF.Bins.7 together with two *Methylococcus sp.* isolates clustered into a monophyletic genus-level lineage with short phylogenetic distances. This was consistent with the high AAI value between these two bins (89.62 and 89.45%) and one type strain (Bath) (Fig. 3.14), while the ANI value (99.7%) between these two query bins suggested they could be derived from the same species inhabiting the two biofilm samples. Likewise, as supported by closest isolate and the high AAI value (94.71%), OBB-3-SURF.Bins.48 can be identified as *Methylotetracoccus sp.* (Fig. 3.13 & 3.14). The remaining 3 MAGs were highly divergent from the closest reference genomes with low AAI values. OBB-3-SURF.Bins.8 had an AAI of 66.36% to *M. oryzae*, and OBB-3-SURF.Bins.25/OBB-3-SURF.Bins.20 is 67.75%/67.41% to *M. ishizawai*, implying they were potentially novel MOB genera within the type Ib clade (Fig. 3.13 & 3.14). The alpha-MOB tree revealed that MAGs MD-6A-SURF.Bins.48 and BF-4-SURF.Bins.20 grouped together with reference genomes of *Methylocystis sp.*, as further suggested by their 86.06% and 84.84% AAI values to *M.rosea* (Fig. 3.15 & 3.16). Deep branching MAG BF-4-SURF.Bins.21 displayed low AAI similarity (68.16%) to the genome from *M.aurea*, and was likely assembled from a novel genus in type IIb clade (Fig. 3.15 & 3.16). In addition, we successfully recovered a groundwater MAG MD-6A-SURF.Bins.52 potentially classified into anaerobic methanotrophic clade ANME-2d based on it shared 73.7% AAI with the genome of *M. nitroreducens* (ANME-2d representative). Basic genome characteristics of

the above ten MAGs identified as having methanotrophic potential were listed (Table 3.1). The genome size and GC content of each MAG were consistent with their closest cultivated relatives, and 9 bacterial MAGs appeared to contain larger genome size with higher GC ratio than ANME-2d archaeal MAG (Table 3.1).

The conserved marker genes of methane monooxygenase (pMMO and sMMO) and 16S rRNA genes were obtained from annotated MAGs and blasted against NCBI database to retrieve the closest relatives. The identities of closest neighbors of single marker genes were consistent with the closest hits of AAI analysis for 8 of 10 MAGs (Table 3.2), providing additional support that they indeed belong to diverse MOB lineages with various degrees of novelty. The MAG OBB-3-SURF.Bins.48 shared high AAI and *pmoC* similarities with *M. oryzae*, but its *mmoX* gene was closest to two other strains (Table 3.2). Since a previous study showed that *M. oryzae* does not contain *mmoX* gene (Ghashghavi et al., 2019), our analysis implied this MAG could be derived from a novel species of *Methylotetracoccus* containing sMMO. Likewise, for MD-6A-SURF.Bins.48, the closest blast hits differed from the result of AAI analysis (Table 3.2). This was because that the query genes (*pmoA/B/C*) were derived from the second operon of pMMO, which were highly similar to the second copy gene set *pmoA2/B2/C2* possessed by two metagenomic assembly, whereas the closest reference genome of *M.rosea* in AAI analysis doesn't contain the second operon of pMMO (Knief, 2015). To link the shot-gun sequencing analysis with the amplicon datasets, we put five marker gene sequences (two16S and three *pmoA*) derived from MAGs into the phylogenetic tree (Fig. 3.7&3.8). All five sequences were almost identical (>99%) to an amplicon OTU

within the same lineage derived from the same sample as the MAG, and their phylogenetic identities (Fig. 3.7&3.8) were identical to the ones of associated MAGs revealed by phylogenomic and AAI analyses.

DISCUSSION

Comparison of SURF MOB and associated bacterial populations with other subsurface studies

Comparison of borehole fluid communities with another SURF study: A previous study (Osburn et al., 2014) surveyed the geochemical data of fluids from a variety of SURF boreholes (including our sampled 3 boreholes), and energy calculations revealed high energy yield from the activities of methane/sulfur/iron/nitrogen/manganese oxidizers. The MOB family *Methylococcaceae* were shown to be present and most common in borehole 8 with high energy density for aerobic methanotrophy, but nearly absent from the other two boreholes (Osburn et al., 2014). In contrast, our two marker gene analyses showed that that MOB were likely absent or present at negligible levels in fluids of borehole 8, but highly enriched in borehole D and with relatively low abundance in borehole B (Fig 3.5). Such inconsistent distribution pattern of MOB could be caused by different library preparation procedures. It should be noted that we extracted the DNA through a DNeasy PowerSoil Pro Kit as it was widely used in previous subsurface community surveys (Inagaki et al., 2003; Hirayama et al., 2005; Gihring et al., 2006; Onstott et al., 2009; Rastogi et al., 2009; Rastogi et al., 2010) and applied bacteria-specific primer (341F/785R) for targeting the V3V4 regions of MOB template, whereas a phenol-chloroform DNA extraction method and prokaryotic-specific primer (515F/806R) for targeting the V4 region was applied in the earlier SURF study (Osburn et al., 2014). However, the dominant bacterial populations were generally comparable between two studies, consistently showing that 3 boreholes contained a high abundance of sulfate reducing class Deltaproteobacteria (i.e., Desulfarculales, Desulfobacterales,

Desulfovibrionales, Desulfuromonadales) (Osburn et al., 2014). Meanwhile, candidate phyla and unassigned sequences indeed comprised a larger proportion of the community in borehole 8 (Osburn et al., 2014). The differences in the observed MOB communities in two studies could also be explained if the planktonic MOB populations were not a core microbiome in groundwater and unstable across two distinct sampling periods in two studies, but sulfate reducers were likely to be the central player and stable in borehole groundwaters.

Comparison of borehole fluid communities with other subsurface environments: As the in-situ oxygen-deficient conditions in deep fractured rocks are unfavorable for the growth and activity of MOBs, they are usually a relatively minor component of the prokaryotic community in deep subsurface groundwater and their distributions were less frequently recorded. However, the *mmo* gene sequence of *Methylococcus* was successfully retrieved from fluid-filled fractures in the Beatrix Gold Mine (Welkom, South Africa) at the depth of 1,339 m (Magnabosco et al., 2018). The *pmoA* genes and transcripts closely from *Methylococcus* and *Methylocaldum* were also detected in fracture fluids from a depth of 500 m in Outokumpu (Finland) (Rajala et al., 2015), while marker genes of *Methylomonas* were retrieved from the 600 m level (Purkamo et al., 2015). Relative to limited data on methane oxidizers in borehole fluid, subterranean methanogens and microbial methanogenesis are frequently observed in anaerobic groundwaters from deep crystalline rock environments (Canada, Finland, South Africa, Sweden, USA) (Kietavainen and Purkamo, 2015). In SURF 4,850 ft groundwaters, the methanogenic Archaea (mainly genus *Methanobacterium*) were

enriched in fluids of borehole B, but almost absent in other two boreholes (data not shown). Such additional biogenic methane resource likely contributed to one order of magnitude higher concentration of dissolved methane in borehole B relative to the other two (Osburn et al., 2014). Aside from aerobic methanotrophs, one MAG derived from anaerobic methanotrophic clade (ANME-2d) was only recovered from fluids of borehole D (Table 3.2). This is in agreement with previous findings that ANME-2d clade associated marker gene was detected in deep fluids from the Pyhasalmi mine (Finland) (Miettinen et al., 2015) and the genome of ANME-2d was assembled from the groundwater collected in Mizumani (Japan) (Ino et al., 2018), while other anaerobic methanotrophic clade member ANME-1a was also detected in deep fracture fluid at specific sulphate-methane transition depth in Olkiluoto (Finland) (Nyyssönen et al., 2012; Bomberg et al., 2015).

Comparison of attached communities derived from various tunnel habitats: Aerobic methane oxidizers dominate and are most active at the redox interface where oxygen-methane counter gradient exists and have been found in a wide variety of environments, including the sediment-water layer of lakes (Bastviken et al., 2008), the water-air interface in cave groundwater (Kumaresan et al., 2014) and the soil-air interface of wetland plant rhizosphere (Conrad, 2007). It is therefore reasonable to speculate that aerobic methanotrophs were abundant and active at tunnel interface niches that were open to the oxygenated tunnel air and fed by methane-containing groundwater either seeping from mine walls or percolating up to the tunnel floor. Attached samples under such condition were collected from the ceiling, floor and walls of the tunnel. Unlike other

samples constantly fed by the groundwater, the white fungus was relatively dry and attached to a rotted log. The microbes inhabiting it would not have much access to nutrients and carbon sources dissolved in groundwater, obtaining nutrients from wood decomposing bacteria and fungi, leading to the lowest diversity in this sample (Fig. 3.6). Even under these conditions, both marker gene analyses showed that *Methylomonas* were present in white fungus (Fig. 3.5b&c), and denovo29 *pmoA* sequence associated strain could be specially adapted to this niche as it was almost exclusively retrieved from white fungus (Fig. 3.8). It is possible that methane needed by the above methanotrophs is obtained as groundwater releases it to the tunnel atmosphere or directly from methanogenesis in the attached wood. Nematode species likely grazing on bacteria were detected in stalactites from three different mines in South Africa based on DGGE analysis of 18S rRNA gene sequences, and were living inside the stalactites with bacteria clearly identified by electron microscopy (Borgonie et al., 2015). MOBs can serve as an important food source for zooplankton in aquatic ecosystems (Sanseverino et al., 2012), and may be particularly important in the subsurface stalactites which is oligotrophic. Materials in the two pools with pink and brown colors were likely quite different and the brown crust was unfavorable for the growth of MOB as negligible methanotrophic reads were retrieved. However, the pink deposit associate geochemical properties presumably promoted the growth of novel MOBs from atypical uncultivated lineages (i.e., NC10, LL_F11, IheB2-23) (Fig. 3.5b&c). Actinobacteria known for their role in organic substance turnover were exclusively enriched in both brown and pink pools, implying that abundant organic substrates were available in both systems and favored the survival of organotrophs. Our lack of geochemical information on the two pools makes a

comparison with other mine systems more difficult. The three biofilms on the tunnel wall had a similar methanotrophic community composition, but the abundances of MOBs varied. To the best of our knowledge, the orange black biofilm was the most MOB enriched sample so far in the deep subsurface biosphere, and the its properties likely supported the abundance of MOB. This assumption was in agreement with that other types of biofilms next to borehole 2 and 5 at the 800 ft level in SURF which lacked MOBs (Osburn et al., 2014).

Phylogenetic diversity of SURF MOB and their habitat preference

Although several cultivated MOB lineages have been uncovered in previous subsurface microbial investigations (see next paragraph), there is still no uncultivated MOB lineages revealed in these surveys, due to lack of in-depth phylogenetic analysis of methanotrophic marker gene sequences. To identify uncultivated lineages, we constructed two comprehensive phylogenetic trees for inferring the relationships between putative MOB sequences retrieved in SURF and their closest relatives (Fig. 3.7&3.8). The uncultivated lineages exclusively containing environmental sequences were identified in the context of our phylogenetic trees and named by referring to a recent study (Knief, 2015). Particularly, the methanotrophic 16S and *pmoA* marker gene sequences recovered in SURF samples, together with their closest environmental sequences, clustered into five uncultivated MOB lineages (USCα, LL_FF11, RPCs, FWs, IheB2-23) (Fig. 3.7&3.8), which lineages haven't been reported in deep terrestrial subsurface biosphere so far. USCα *pmoA* gene sequences were predominantly found in forest soils, and also detected in hydromorphic and grassland soil (Knief, 2015). Stable

Isotope experiments proved that USCα members are capable of assimilating acetate into their biomass (Pratscher et al., 2011). This facultative advantage would enable them to be more competitive in SURF wall biofilm communities. A *pmoA*-specific methanotrophic clade type Id composed of about ten uncultivated lineages were proposed recently based on its intermediate phylogenetic placement between type I and type II or III (Knief, 2015). One of the type Id lineages LL_F11 (Knief, 2015), firstly reported in landfill-cover soil (Henneberger et al., 2012), was also found in underwater pink deposit where likely contains abundant organic matters. Among the other uncultured lineages observed here, RPC has been found as the predominant population in rice field, aquatic ecosystems, bogs and upland soils (Knief, 2015), while recently defined lineage FWs are broadly distributed but most frequently retrieved in aquatic ecosystems (Dumont et al., 2014). With compromised habitat specificity, it was not surprising both type Ib methanotrophic lineages RPCs and FWs occurred as minor groups in groundwater-saturated biofilms. A rDNA clone of IheB2-23 was first reported in deep-sea hydrothermal vent (Suzuki et al., 2004). The two affiliated OTUs (1717&1746) were distantly related (~92% identity) to the known type Ia and Ib isolates (Fig. 3.7). Although placed into IheB2-23 lineage, these two OTUs were closer to subsurface clones and together formed into a subcluster diverging from to traditional deep-sea members (Fig. 3.7). This implied that subsurface-specific cluster members of IheB2-23 lineage may have originated from ancient seas.

For the cultivated MOB lineages, acidophilic methanotrophs (type III) from the phylum Verrucomicrobia (Dunfield et al., 2007) and thermophilic methanotrophs (type

Ic) from the family *Methylothermaceae* were not detected in any of our SURF samples. Although type III clade members have not been reported in subsurface samples either, the type Ic strain *Methylothermus subterraneus* was isolated from subsurface hot aquifer water from a Japanese gold mine and grows between 37-65 ℃ (Hirayama et al., 2011). The neutral to alkaline pH and a narrow range of temperature (10-32.8 ℃) recorded previously throughout the SURF groundwaters (Osburn et al., 2014) would most likely prevent the growth of these two lineages. Our analyses of the two marker genes showed the presence of other cultivated lineages, including type Ia (*Methylomonas*, *Methylosarcina*), type Ib (*Methylococcus*, *Methylocaldum* and *Methylotetrococcus*) and type IIa (*Methylocystis*, *Methylosinus*). These lineages were frequently observed in other subsurface environments. The *Methylococcus* marker gene was detected in borehole fluid from the gold mine in South Africa (Lin et al., 2006a; Magnabosco et al., 2018), and *Methylococcus*, *Methylocaldum, Methylocystis* were found to be living near a subsurface geothermal water stream in a Japanese gold mine (Hirayama et al., 2005). As well, *Methylomonas*, *Methylocystis*, *Methylobacter* were detected in borehole groundwaters in several deep granitic aquifers (Aspo and Forsmark, Sweden; Olkiluoto and Outokumpu, Finland) (Kalyuzhnaya et al., 1999; CHI FRU, 2008; Kutvonen et al., 2015; Purkamo et al., 2015).

Materials on the cave water surface or attached to cave rocks occupy similar niches to the mine tunnel samples, where the methane released from groundwater meets with the oxygenated cave atmosphere. Thus, the MOB communities have been better characterized in caves than the other subsurface ecosystems. Cave studies revealed that

MOBs inhabiting these interface niches were actively feeding on methane (Hutchens et al., 2004) and developed as the dominant populations there, mainly affiliated within the type Ia cultivated lineages (i.e., *Methylovulum*, *Crenothrix* and *Methylobacter*, and *Methylomicrobium*) (Karwautz et al., 2018). The dominance of type Ia lineages in cave differs from the SURF tunnel biofilms (dominated by type Ib lineages) likely due to temperature differences (8-11℃ in cave and 10-32.8 ℃ in SURF groundwaters) (Osburn et al., 2014; Karwautz et al., 2018).

By analyzing the two marker genes, we could successfully retrieved the methanotrophic sequences from type Ia, Ib, Id, IIa, IIb, NC10 clades. Except for cave ecosystem, the broad occurrence of methanotrophic lineages in the other subsurface biosphere has not been systematically reported. Our findings help fill this gap, showing that diverse aerobic methanotrophs were likely to be the dominant players in subsurface habitats other than in borehole groundwater. These MOBs inhabiting SURF tunnels can act as an important sink of methane and their significances on the subsurface carbon cycling remain to be further characterized.

Origins of MOBs in SURF

Despite long terms of intense study on life in the deep subsurface, a key question cannot be avoided is that whether the new identified organism is indigenous or accidently introduced from the surface biosphere by human activities. Based on a recently published perspective, the origin can be inferred from its geographic distribution pattern and the habitats of closest relatives in the same phylotypes (Colman et al., 2017). For example,

strain *Ca. D. audaxviator* were widespread in several groundwaters in South African gold mine (Chivian et al., 2008). Based on the fact that genus-level phylotypes have only been retrieved from deep boreholes (Itavaara et al., 2011; Tiago and Verissimo, 2013), authors concluded that this strain and associated neighbors in genus *Ca. Desulforudis* are generally indigenous to subsurface environments. The Otu6 -*Methylococcus capsulatus* sequences were consistently detected in borehole waters at the 4,850 ft level in SURF mine across 6 years (identical with hit JX434259 recorded on 2012 in NCBI database) (Fig. 3.7). The closest relatives (subspecies level as identity > 99.5%) were distributed in groundwaters of mines in South Africa (Lin et al., 2006a; Magnabosco et al., 2018), Russia (KU517778 in Fig. 3.7) and Japan (Hirayama et al., 2005). Stable occurrence of the Otu6 clone sequence in a longterm scale and subspecies-level phylotypes of Otu6 only found in deep groundwaters across the geographically distinct sites provided the strong evidence that Otu6-associated MOB was an endemic inhabitant. In the same manner, the close relatives of another six OTUs were uniquely associated with subsurface environments (e.g., cave wall, subsurface aquifer, mine soil, mine rock biofilm and deep-well water), and phylogenetically distinct with surface phylotypes. These phylogeny results, and when considered with their capability of methanotrophy in methane-supplied subsurface environments, suggested that these OTUs-associated MOBs were truly restricted to subsurface ecosystem.

The deep subsurface ecosystem is often connected between different sampling sites, enabling exchange of geochemical components and microbial cells through groundwater through a subsurface fracture network. There is a limited number of

diagnostic tools to test the connectivity in deep reservoirs (Ligtenberg, 2005). The in-situ microbial community profile was thought to be ideal bio-signature for identifying natural connectivity. In a system with greater hydrological connectivity, we could find a high degree of the similarity of microbial populations across sites. A recent study showed that microbial community composition was highly similar between two borehole fluids, but substantially distinct with other boreholes distributed throughout the SURF (Zhang et al., 2020). They suggested that these two boreholes are intersected with the same fracture, and their connectivity was later corroborated by log and camera investigation (Zhang et al., 2020). As such, it is reasonable to assume that the microbes (including MOBs) found in biofilms originated from the feeding groundwater, and to test this connectivity based on the community relatedness between biofilms and borehole fluids. Specifically, we hypothesize that the microbial cells in groundwaters were initially trapped and then concentrated in rock biofilms, but due to a difference in geochemical conditions (e.g. O_2 concentration) and biological interactions contained in biofilms, some microbial group members (e.g. MOB) outcompete others and dominate the communities. If this hypothesis holds true, many OTUs will be shared between biofilm and groundwater communities (qualitative perspective) to reflect connectivity, but abundance-based community structure should be dissimilar (quantitative perspective) to reflect the influence of environmental conditions and interactions associated with biofilms and groundwaters. There were few shared OTUs between shallow and deep level samples, however, biofilms and borehole fluids exclusively shared a number of common OTUs at the same depth (Fig. 3.9), suggesting limited degree of connectivity among the same depth samples. The community structures were significantly different between biofilms

and borehole fluids (Fig. 3.10), indicating that sample type associated unique features indeed have a control on community structure from an abundance-based perspective. In parallel with the relatedness of bacterial communities, a high number of methanotrophic OTUs were confined to be present in deep level samples and shared between biofilms and borehole fluids (Fig. 3.7&3.8). Thus, microbial exchange likely occurred from groundwater to biofilms.

MOBs act as the food base in SURF tunnel biofilms

Due to the oligotrophic nature of subsurface environments, it is not hard to envisage that microbes inhabiting the same niche cooperate together to avoid thermodynamic barriers, exchanging diverse metabolites to maximize the community-wide growth and energy yield (Lau et al., 2016). Stable isotope probing experiment have revealed that turnover of methane occurs in Movile cave surface water and mats among MOBs and a diverse range of associated non-methanotrophic bacteria (Hutchens et al., 2004). Especially considering that a high abundance of MOBs inhabiting the SURF biofilms, it is reasonable to hypothesize that MOBs are the primary producers at SURF similarly to the Movile cave to sustain a diverse community of heterotrophs in the mine ecosystems. As natural microbial interactions are poorly ascertained in laboratory studies, network analysis of co-occurrence patterns has been widely used to investigate potential ecological interactions between microbial taxa in freshwater lakes (Yang et al., 2017), soils (Barberan et al., 2012) and rhizosphere (Zhang et al., 2018). Thus, to test the hypothesis, co-occurrence network of the bacterial community in SURF biofilms was performed, and the topological features of MOB nodes uncovered in networks was used

to estimate the roles of MOBs on modulating the interactions. Otherwise, given biomarker gene sequences of MOB were less frequently detected in groundwaters in previous studies, we assume MOBs doesn't play a significant role in interactions in SURF groundwater, which was tested through network analysis as well.

There is only limited network analysis data in subsurface microbial ecosystems. The exception is one subsurface study of microbial network patterns in groundwater collected from Outokumpu bedrock has identified two keystone lineages Burkholderiales and Clostridiales that having broad interactions with associated microbial members, whereas methanotrophic lineages were not observed in this reported network (Purkamo et al., 2016). In SURF groundwater, the most connected nodes within or between modules (e.g. representatives of Anaerolineales, *Desulfobacteraceae*) were involved in carbon fixation (Kadnikov et al., 2018) and sulfate reduction (Kuever, 2014) (Fig. 3.12). Also, there were no MOB nodes observed in the SURF groundwater network. These observations suggested that methanotrophs might not be the primary producers in groundwaters, while chemolithoautotrophs (e.g. sulfate reducer) likely provided organic materials to maintain a variety of other co-occurred microbial types. MOB playing a negligible role in groundwater microbial interaction network was reasonable, when considering the oxygen-deficient condition normally encountered in groundwaters was unfavorable for the growth and activity of planktonic MOBs.

However, MOBs likely play central roles in the biofilms as they have the most connections within and among modules in biofilm network (Fig. 3.11a&b). This

provides further evidence that aerobic methanotrophy fuels the biofilm communities. As being the dominant population in biofilm with occupying more network nodes, type Ib MOB are thought to have more significant roles than type IIa *Methylocystis* (corresponding to one node). Based on the positive correlation with three type Ib nodes, members of 12 nodes (non MOB) were identified to potentially cooperate with MOBs (Fig. 3.11a). Except for the *Burkholderiaceae*, lineages of these 12 nodes were different microbes that interact with MOBs in lakes, grassland soil, rice paddy, and tailing ponds (Ho et al., 2016). This suggests that distinct microbial populations were involved in the flow of carbon after CH_4 oxidation within surface and subsurface ecosystems. The metabolically flexible family *Burkholderiaceae* has been observed to be capable of feeding on metabolites excreted during methane oxidation or directly on biomass of MOB in several surface ecosystems (Ho et al., 2016), and can contribute to the carbon exchange in a variety of deep subsurface aquifers as well (Brazelton et al., 2012; Tiago and Verissimo, 2013; Purkamo et al., 2016), explaining their association with MOB observed in SURF biofilm. The sulphur-oxidizing *Thiobacillus* with the highest within module connectivity here was also found to be enriched with MOBs in cave (Hutchens et al., 2004), that were both likely to be the common contributors to subsurface heterotrophic food web through chemoautotrophy and methanotrophy respectively. Co-occurrence patterns additionally identified a number of putative MOB partners involved in heterotrophic turnover, including *Bryobacter, Dongia, Hyphomonas,* Ignavibacteriales, Kryptoniales, *Microscillaceae, Phycisphaeraceae* and Thermodesulfovibrionia. Under oligotrophic conditions with no input of photosynthetically fixed carbon, it allows us to

hypothesize that these heterotrophs actively interact with the associated primary carbon producer MOBs through cross-feeding on methane-derived carbon.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation grant No. 1736255 under a subcontract from South Dakota School of Mines & Technology through the EPSCoR RII Track-2 FEC program. We thank Christopher Abin, Rosa Moon-Escamilla and Christopher Garner for help in collecting samples. Many thanks to Krithi. Sankaranarayanan for use of their server and help tutoring me in metagenomic analysis. We want to especially thank staff at SURF, including Jaret Heise and Tom Reagan, for providing access to the deep subsurface and assistance during the collection of groundwater samples.

REFERENCES

Barberan, A., Bates, S.T., Casamayor, E.O., and Fierer, N. (2012) Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J* **6**: 343-351.

Bastian, M., Heymann, S., and Jacomy, M. (2009) Gephi: an open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media*.

Bastviken, D., Cole, J.J., Pace, M.L., and Van de Bogert, M.C. (2008) Fates of methane from different lake habitats: Connecting whole-lake budgets and CH_4 emissions. *J Geophys Res* **113**: G02024.

Bolger, A.M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114-2120.

Bomberg, M., and Ahonen, L. (2017) Editorial: Geomicrobes: Life in Terrestrial Deep Subsurface. *Front Microbiol* **8**: 103.

Bomberg, M., Nyyssonen, M., Pitkanen, P., Lehtinen, A., and Itavaara, M. (2015) Active Microbial Communities Inhabit Sulphate-Methane Interphase in Deep Bedrock Fracture Fluids in Olkiluoto, Finland. *Biomed Res Int* **2015**: 979530.

Borgonie, G., Linage-Alvarez, B., Ojo, A., Shivambu, S., Kuloyo, O., Cason, E.D. et al. (2015) Deep subsurface mine stalactites trap endemic fissure fluid Archaea, Bacteria, and Nematoda possibly originating from ancient seas. *Front Microbiol* **6**: 833.

Boswell, R., and Collett, T.S. (2011) Current perspectives on gas hydrate resources. *Energy & Environmental Science* **4**: 1206-1215.

Brazelton, W.J., Nelson, B., and Schrenk, M.O. (2012) Metagenomic evidence for h(2) oxidation and h(2) production by serpentinite-hosted subsurface microbial communities. *Front Microbiol* **2**: 268.

Caddey, S.W., Bachman, R.L., Campbell, T.J., Reid, R.R., and Otto, R.P. (1991) The Homestake gold mine, an early Proterozoic iron-formation-hosted gold deposit, Lawrence County, South Dakota. In *Bulletin*.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K. et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335-336.

CHI FRU, E. (2008) Constraints in the colonization of natural and engineered subterranean igneous rock aquifers by aerobic methane-oxidizing bacteria inferred by culture analysis. *Geobiology* **6**: 365-375.

Chivian, D., Brodie, E.L., Alm, E.J., Culley, D.E., Dehal, P.S., DeSantis, T.Z. et al. (2008) Environmental genomics reveals a single-species ecosystem deep within Earth. *Science* **322**: 275-278.

Colman, D.R., Poudel, S., Stamps, B.W., Boyd, E.S., and Spear, J.R. (2017) The deep, hot biosphere: Twenty-five years of retrospection. *Proc Natl Acad Sci U S A* **114**: 6895-6903.

Conrad, R. (2007) Microbial Ecology of Methanogens and Methanotrophs. In *Advances in Agronomy*: Academic Press, pp. 1-63.

Conway, J.R., Lex, A., and Gehlenborg, N. (2017) UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**: 2938-2940.

Csardi, G., and Nepusz, T. (2006) The igraph software package for complex network research. *Int J Complex Syst* 1-9.

Davidson, M.M., Silver, B.J., Onstott, T.C., Moser, D.P., Gihring, T.M., Pratt, L.M. et al. (2011) Capture of Planktonic Microbial Diversity in Fractures by Long-Term Monitoring of Flowing Boreholes, Evander Basin, South Africa. *Geomicrobiology Journal* **28**: 275-300.

De Tender, C., Devriese, L.I., Haegeman, A., Maes, S., Vangeyte, J., Cattrijsse, A. et al. (2017) Temporal Dynamics of Bacterial and Fungal Colonization on Plastic Debris in the North Sea. *Environ Sci Technol* **51**: 7350-7360.

Degelmann, D.M., Borken, W., Drake, H.L., and Kolb, S. (2010) Different atmospheric methane-oxidizing communities in European beech and Norway spruce soils. *Appl Environ Microbiol* **76**: 3228-3235.

Dumont, M., Lüke, C., Deng, Y., and Frenzel, P. (2014) Classification of *pmoA* amplicon pyrosequences using BLAST and the lowest common ancestor method in MEGAN. *Front Microbiol* **5**: 1-11.

Dunfield, P.F., Yuryev, A., Senin, P., Smirnova, A.V., Stott, M.B., Hou, S. et al. (2007) Methane oxidation by an extremely acidophilic bacterium of the phylum Verrucomicrobia. *Nature* **450**: 879-882.

Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460-2461.

Edgar, R.C. (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* **10**: 996-998.

Finn, R.D., Clements, J., and Eddy, S.R. (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* **39**: W29-37.

Ghashghavi, M., Belova, S.E., Bodelier, P.L.E., Dedysh, S.N., Kox, M.A.R., Speth, D.R. et al. (2019) Methylotetracoccus oryzae Strain C50C1 Is a Novel Type Ib Gammaproteobacterial Methanotroph Adapted to Freshwater Environments. *mSphere* **4**.

Gihring, T.M., Moser, D.P., Lin, L.H., Davidson, M., Onstott, T.C., Morgan, L. et al. (2006) The Distribution of Microbial Taxa in the Subsurface Water of the Kalahari Shield, South Africa. *Geomicrobiology Journal* **23**: 415-430.

Guimera, R., and Nunes Amaral, L.A. (2005) Functional cartography of complex metabolic networks. *Nature* **433**: 895-900.

Harrell, F.E. (2008) Hmisc: harrell miscellaneous R package version: 3.5–2.

Henneberger, R., Luke, C., Mosberger, L., and Schroth, M.H. (2012) Structure and function of methanotrophic communities in a landfill-cover soil. *FEMS Microbiol Ecol* **81**: 52-65.

Hirayama, H., Takai, K., Inagaki, F., Yamato, Y., Suzuki, M., Nealson, K.H., and Horikoshi, K. (2005) Bacterial community shift along a subsurface geothermal water stream in a Japanese gold mine. *Extremophiles* **9**: 169-184.

Hirayama, H., Suzuki, Y., Abe, M., Miyazaki, M., Makita, H., Inagaki, F. et al. (2011) Methylothermus subterraneus sp. nov., a moderately thermophilic methanotroph isolated from a terrestrial subsurface hot aquifer. *Int J Syst Evol Microbiol* **61**: 2646-2653.

Ho, A., Angel, R., Veraart, A.J., Daebeler, A., Jia, Z., Kim, S.Y. et al. (2016) Biotic Interactions in Microbial Communities as Modulators of Biogeochemical Processes: Methanotrophy as a Model System. *Front Microbiol* **7**: 1285.

Hutchens, E., Radajewski, S., Dumont, M.G., McDonald, I.R., and Murrell, J.C. (2004) Analysis of methanotrophic bacteria in Movile Cave by stable isotope probing. *Environ Microbiol* **6**: 111-120.

Hyatt, D., LoCascio, P.F., Hauser, L.J., and Uberbacher, E.C. (2012) Gene and translation initiation site prediction in metagenomic sequences. *Bioinformatics* **28**: 2223-2230.

Inagaki, F., Takai, K., Hirayama, H., Yamato, Y., Nealson, K.H., and Horikoshi, K. (2003) Distribution and phylogenetic diversity of the subsurface microbial community in a Japanese epithermal gold mine. *Extremophiles* **7**: 307-317.

Ino, K., Hernsdorf, A.W., Konno, U., Kouduka, M., Yanagawa, K., Kato, S. et al. (2018) Ecological and genomic profiling of anaerobic methane-oxidizing archaea in a deep granitic environment. *ISME J* **12**: 31-47.

Itavaara, M., Nyyssonen, M., Kapanen, A., Nousiainen, A., Ahonen, L., and Kukkonen, I. (2011) Characterization of bacterial diversity to a depth of 1500 m in the Outokumpu deep borehole, Fennoscandian Shield. *FEMS Microbiol Ecol* **77**: 295-309.

Kadnikov, V.V., Mardanov, A.V., Beletsky, A.V., Banks, D., Pimenov, N.V., Frank, Y.A. et al. (2018) A metagenomic window into the 2-km-deep terrestrial subsurface aquifer revealed multiple pathways of organic matter decomposition. *FEMS Microbiology Ecology* **94**.

Kalyuzhnaya, M.G., Khmelenina, V.N., Kotelnikova, S., Holmquist, L., Pedersen, K., and Trotsenko, Y.A. (1999) Methylomonas scandinavica sp. nov., a new methanotrophic psychrotrophic bacterium isolated from deep igneous rock ground water of Sweden. *Syst Appl Microbiol* **22**: 565-572.

Kanehisa, M., Sato, Y., and Morishima, K. (2016) BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol* **428**: 726-731.

Kang, D.D., Froula, J., Egan, R., and Wang, Z. (2015) MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**: e1165.

Karwautz, C., Kus, G., Stockl, M., Neu, T.R., and Lueders, T. (2018) Microbial megacities fueled by methane oxidation in a mineral spring cave. *ISME J* **12**: 87-100.

Katoh, K., and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**: 772-780.

Kietavainen, R., and Purkamo, L. (2015) The origin, source, and cycling of methane in deep crystalline rock biosphere. *Front Microbiol* **6**: 725.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., and Glockner, F.O. (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* **41**: e1.

Knief, C. (2015) Diversity and habitat preferences of cultivated and uncultivated aerobic methanotrophic bacteria evaluated based on *pmoA* as molecular marker. *Front Microbiol* **6**: 1-38.

Kolb, S., Knief, C., Stubner, S., and Conrad, R. (2003) Quantitative detection of methanotrophs in soil by novel pmoA-targeted real-time PCR assays. *Appl Environ Microbiol* **69**: 2423-2429.

Kolde, R. Pheatmap: pretty heatmaps, R package v. 16 (R Foundation for Statistical Computing, 2012).

Konstantinidis, K.T., Rosselló-Móra, R., and Amann, R. (2017) Uncultivated microbes in need of their own taxonomy. *The ISME Journal* **11**: 2399-2406.

Krumholz, L.R. (2000) Microbial communities in the deep subsurface. *Hydrogeology Journal* **8**: 4-10.

Kuever, J. (2014) The Family Desulfobacteraceae. In *The Prokaryotes: Deltaproteobacteria and Epsilonproteobacteria*. Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., and Thompson, F. (eds). Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 45-73.

Kumaresan, D., Wischer, D., Stephenson, J., Hillebrand-Voiculescu, A., and Murrell, J.C. (2014) Microbiology of Movile Cave—A Chemolithoautotrophic Ecosystem. *Geomicrobiology Journal* **31**: 186-193.

Kutvonen, H., Rajala, P., Carpén, L., and Bomberg, M. (2015) Nitrate and ammonia as nitrogen sources for deep subsurface microorganisms. *Frontiers in Microbiology* **6**.

Langmead, B., and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357-359.

Lau, M.C., Kieft, T.L., Kuloyo, O., Linage-Alvarez, B., van Heerden, E., Lindsay, M.R. et al. (2016) An oligotrophic deep-subsurface community dependent on syntrophy is

dominated by sulfur-driven autotrophic denitrifiers. *Proc Natl Acad Sci U S A* **113**: E7927-E7936.

Letunic, I., and Bork, P. (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research* **47**: W256-W259.

Lever, M.A., Rouxel, O., Alt, J.C., Shimizu, N., Ono, S., Coggon, R.M. et al. (2013) Evidence for microbial carbon and sulfur cycling in deeply buried ridge flank basalt. *Science* **339**: 1305-1308.

Li, D., Liu, C.M., Luo, R., Sadakane, K., and Lam, T.W. (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**: 1674-1676.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N. et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

Ligtenberg, J.H. (2005) Detection of fluid migration pathways in seismic data: implications for fault seal analysis. *Basin Research* **17**: 141-153.

Lin, L.-H., Hall, J., Onstott, T.C., Gihring, T., Lollar, B.S., Boice, E. et al. (2006a) Planktonic Microbial Communities Associated with Fracture-Derived Groundwater in a Deep Gold Mine of South Africa. *Geomicrobiology Journal* **23**: 475-497.

Lin, L.-H., Hall, J., Lippmann-Pipke, J., Ward, J.A., Sherwood Lollar, B., DeFlaun, M. et al. (2005) Radiolytic H2 in continental crust: Nuclear power for deep subsurface microbial communities. *Geochemistry, Geophysics, Geosystems* **6**.

Lin, L.H., Wang, P.L., Rumble, D., Lippmann-Pipke, J., Boice, E., Pratt, L.M. et al. (2006b) Long-term sustainability of a high-energy, low-diversity crustal biome. *Science* **314**: 479-482.

Lollar, B.S., Onstott, T.C., Lacrampe-Couloume, G., and Ballentine, C.J. (2014) The contribution of the Precambrian continental lithosphere to global H2 production. *Nature* **516**: 379-382.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar et al. (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363-1371.

Magnabosco, C., Ryan, K., Lau, M.C., Kuloyo, O., Sherwood Lollar, B., Kieft, T.L. et al. (2016) A metagenomic window into carbon metabolism at 3 km depth in Precambrian continental crust. *ISME J* **10**: 730-741.

Magnabosco, C., Timmers, P.H.A., Lau, M.C.Y., Borgonie, G., Linage-Alvarez, B., Kuloyo, O. et al. (2018) Fluctuations in populations of subsurface methane oxidizers in coordination with changes in electron acceptor availability. *FEMS Microbiol Ecol* **94**.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**: 3.

Miettinen, H., Kietavainen, R., Sohlberg, E., Numminen, M., Ahonen, L., and Itavaara, M. (2015) Microbiome composition and geochemical characteristics of deep subsurface high-pressure environment, Pyhasalmi mine Finland. *Front Microbiol* **6**: 1203.

Miller, M.A., Pfeiffer, W., and Schwartz, T. (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *2010 Gateway Computing Environments Workshop (GCE)*, pp. 1-8.

Momper, L., Jungbluth, S.P., Lee, M.D., and Amend, J.P. (2017a) Energy and carbon metabolisms in a deep terrestrial subsurface fluid microbial community. *ISME J* **11**: 2319-2333.

Momper, L., Kiel Reese, B., Zinke, L., Wanger, G., Osburn, M.R., Moser, D., and Amend, J.P. (2017b) Major phylum-level differences between porefluid and host rock bacterial communities in the terrestrial deep subsurface. *Environmental Microbiology Reports* **9**: 501-511.

Moser, D.P., Gihring, T.M., Brockman, F.J., Fredrickson, J.K., Balkwill, D.L., Dollhopf, M.E. et al. (2005) Desulfotomaculum and Methanobacterium spp. dominate a 4- to 5-kilometer-deep fault. *Appl Environ Microbiol* **71**: 8773-8783.

Murase, J., and Frenzel, P. (2007) A methane-driven microbial food web in a wetland rice soil. *Environ Microbiol* **9**: 3025-3034.

Na, S.I., Kim, Y.O., Yoon, S.H., Ha, S.M., Baek, I., and Chun, J. (2018) UBCG: Up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction. *J Microbiol* **56**: 280-285.

Nyyssönen, M., Bomberg, M., Kapanen, A., Nousiainen, A., Pitkänen, P., and Itävaara, M. (2012) Methanogenic and Sulphate-Reducing Microbial Communities in Deep Groundwater of Crystalline Rock Fractures in Olkiluoto, Finland. *Geomicrobiology Journal* **29**: 863-878.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'hara, R.B., and Simpson, G.L. (2020) Vegan: Community ecology package. R Package version 2.5-7. Available from http://CRAN.R-project.org/package=vegan.

Onstott, T.C., Phelps, T.J., Kieft, T., Colwell, F.S., Balkwill, D.L., Fredrickson, J.K., and Brockman, F.J. (1998) A global perspective on the microbial abundance and activity in the deep subsurface. *In: Seckbach J, editor; Enigmatic Microorganisms and Life in Extreme Environments; Boston: Kluwer*: 489-499.

Onstott, T.C., McGown, D.J., Bakermans, C., Ruskeeniemi, T., Ahonen, L., Telling, J. et al. (2009) Microbial communities in subpermafrost saline fracture water at the Lupin Au mine, Nunavut, Canada. *Microb Ecol* **58**: 786-807.

Osburn, M.R., LaRowe, D.E., Momper, L.M., and Amend, J.P. (2014) Chemolithotrophy in the continental deep subsurface: Sanford Underground Research Facility (SURF), USA. *Front Microbiol* **5**: 610.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**: 1043-1055.

Pratscher, J., Dumont, M.G., and Conrad, R. (2011) Assimilation of acetate by the putative atmospheric methane oxidizers belonging to the USCalpha clade. *Environ Microbiol* **13**: 2692-2701.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010) FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5**: e9490.

Pruesse, E., Peplies, J., and Glockner, F.O. (2012) SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**: 1823-1829.

Purkamo, L., Bomberg, M., Nyyssonen, M., Kukkonen, I., Ahonen, L., and Itavaara, M. (2015) Heterotrophic communities supplied by ancient organic carbon predominate in deep fennoscandian bedrock fluids. *Microb Ecol* **69**: 319-332.

Purkamo, L., Bomberg, M., Kietäväinen, R., Salavirta, H., Nyyssönen, M., Nuppunen-Puputti, M. et al. (2016) Microbial co-occurrence patterns in deep Precambrian bedrock fracture fluids. *Biogeosciences* **13**: 3091-3108.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P. et al. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**: D590-596.

Raghoebarsing, A.A., Smolders, A.J., Schmid, M.C., Rijpstra, W.I., Wolters-Arts, M., Derksen, J. et al. (2005) Methanotrophic symbionts provide carbon for photosynthesis in peat bogs. *Nature* **436**: 1153-1156.

Rajala, P., Bomberg, M., Kietavainen, R., Kukkonen, I., Ahonen, L., Nyyssonen, M., and Itavaara, M. (2015) Rapid Reactivation of Deep Subsurface Microbes in the Presence of C-1 Compounds. *Microorganisms* **3**: 17-33.

Rastogi, G., Stetler, L.D., Peyton, B.M., and Sani, R.K. (2009) Molecular analysis of prokaryotic diversity in the deep subsurface of the former Homestake gold mine, South Dakota, USA. *The Journal of Microbiology* **47**: 371-384.

Rastogi, G., Osman, S., Kukkadapu, R., Engelhard, M., Vaishampayan, P.A., Andersen, G.L., and Sani, R.K. (2010) Microbial and mineralogical characterizations of soils collected from the deep biosphere of the former Homestake gold mine, South Dakota. *Microb Ecol* **60**: 539-550.

Rodriguez-R, L., and Konstantinidis, K. (2016) The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. In: PeerJ Preprints.

Sanseverino, A.M., Bastviken, D., Sundh, I., Pickova, J., and Enrich-Prast, A. (2012) Methane carbon supports aquatic food webs to the fish level. *PLoS One* **7**: e42723.

Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068-2069.

Simkus, D.N., Slater, G.F., Lollar, B.S., Wilkie, K., Kieft, T.L., Magnabosco, C. et al. (2016) Variations in microbial carbon sources and cycling in the deep continental subsurface. *Geochimica et Cosmochimica Acta* **173**: 264-283.

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312-1313.

Suzuki, Y., Inagaki, F., Takai, K., Nealson, K.H., and Horikoshi, K. (2004) Microbial diversity in inactive chimney structures from deep-sea hydrothermal systems. *Microb Ecol* **47**: 186-196.

Takai, K., Moser, D.P., DeFlaun, M., Onstott, T.C., and Fredrickson, J.K. (2001) Archaeal diversity in waters from deep South African gold mines. *Appl Environ Microbiol* **67**: 5750-5760.

Takami, H., Taniguchi, T., Arai, W., Takemoto, K., Moriya, Y., and Goto, S. (2016) An automated system for evaluation of the potential functionome: MAPLE version 2.1.0. *DNA Res* **23**: 467-475.

Tiago, I., and Verissimo, A. (2013) Microbial and functional diversity of a subterrestrial high pH groundwater associated to serpentinization. *Environ Microbiol* **15**: 1687-1706.

Waddell, E.J., Elliott, T.J., Vahrenkamp, J.M., Roggenthen, W.M., Sani, R.K., Anderson, C.M., and Bang, S.S. (2010) Phylogenetic evidence of noteworthy microflora from the subsurface of the former Homestake gold mine, Lead, South Dakota. *Environ Technol* **31**: 979-991.

Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**: 5261-5267.

Wang, Q., Quensen, J.F., 3rd, Fish, J.A., Lee, T.K., Sun, Y., Tiedje, J.M., and Cole, J.R. (2013) Ecological patterns of *nifH* genes in four terrestrial climatic zones explored with targeted metagenomics using FrameBot, a new informatics tool. *mBio* **4**: e00592-00513.

Ward, J.A., Slater, G.F., Moser, D.P., Lin, L.H., Lacrampe-Couloume, G., Bonin, A.S. et al. (2004) Microbial hydrocarbon gases in the Witwatersrand Basin, South Africa: Implications for the deep biosphere1 1Associate editor: R. Summons. *Geochimica et Cosmochimica Acta* **68**: 3239-3250.

Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998) Prokaryotes: the unseen majority. *Proc Natl Acad Sci U S A* **95**: 6578-6583.

Wickham, H. (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Yang, C., Wang, Q., Simon, P.N., Liu, J., Liu, L., Dai, X. et al. (2017) Distinct network interactions in particle-associated and free-living bacterial communities during a *Microcystis aeruginosa* bloom in a plateau lake. *Front Microbiol* **8**: 1-15.

Zhang, B., Zhang, J., Liu, Y., Shi, P., and Wei, G. (2018) Co-occurrence patterns of soybean rhizosphere microbiome at a continental scale. *Soil Biology and Biochemistry* **118**: 178-186.

Zhang, Y., Dekas, A.E., Hawkins, A.J., Parada, A.E., Gorbatenko, O., Li, K., and Horne, R.N. (2020) Microbial Community Composition in Deep-Subsurface Reservoir Fluids Reveals Natural Interwell Connectivity. *Water Resources Research* **56**: e2019WR025916.

TABLES

Table 3.1. General Features of potentially methanotrophic MAGs.

SURF Bin	Genome size (Mbp)	Number contigs	GC Content	N50	Number genes	Completeness	Contamination
BF-4-SURF.Bins.6	3.49	178	63.9	33,282	3,263	99.63	2.19
OBB-3-SURF.Bins.7	3.55	118	63.9	46,976	3,311	99.63	2.08
OBB-3-SURF.Bins.48	5.12	126	63.0	83,564	4,527	99.08	3.62
OBB-3-SURF.Bins.8	4.72	217	62.6	39,827	4,185	99.08	2.96
OBB-3-SURF.Bins.25	5.16	350	64.0	26,623	4,730	97.67	2.81
OBB-3-SURF.Bins.20	5.96	180	61.5	55,701	5,250	98.71	2.90
MD-6A-SURF.Bins.48	3.59	123	61.6	67,609	3,678	99.22	1.43
BF-4-SURF.Bins.20	3.20	413	63.7	11,071	3,210	93.03	1.00
BF-4-SURF.Bins.21	3.18	37	61.4	125,884	2,904	95.51	0
MD-6A-SURF.Bins.52	2.85	396	48.3	9,429	3,161	96.08	3.27

Note: "BF-∗" recovered from wall biofilm; "OBB-∗" recovered from orange black biofilm; "MD-∗" recovered from fluid of borehole manifold D.

Table 3.2. The most closely related species to each MAG based on a comparison of marker genes and AAI values of methanotrophic cultures in the NCBI database.

SURF Bin	Marker genes[a]	Nt/AA Similarity[b]	Closest hits of Blast[c]	AAI Similarity[d]	Closest hits of AAI[e]
BF-4-SURF.Bins.6	*mmoX*	97.53	*Methylococcus capsulatus*	89.47	-
OBB-3-SURF.Bins.7	***pmoA**/B*	98.79/96.14	*Methylococcus capsulatus*	89.34	-
	mmoX	97.72	*Methylococcus capsulatus*		
OBB-3-SURF.Bins.48	*pmoC*	97.61	*Methylotetracoccus oryzae*	94.71	*Methylotetracoccus oryzae*
	mmoX	93.92	*Methylococcus capsulatus*		
		93.92	*Candidatus Methylospira mobilis*		
OBB-3-SURF.Bins.8	*pmoC*	94.21	*Methylotetracoccus oryzae*	66.36	-
OBB-3-SURF.Bins.25	*pmoC*	82.30	*Methylomagnum ishizawai*	67.75	-
OBB-3-SURF.Bins.20	*pmoA/B/C*	91.20/75.43/90.48	*Methylomagnum ishizawai*	67.41	-
	mmoX	97.34	*Methylomagnum ishizawai*		
MD-6A-SURF.Bins.48	***pmoA**/B/C*	100/98.37/99.22	*Methylocystis sp.* B8 from wetland	86.06	*Methylocystis rosea*
		98.05/95.34/96.9	*Methylocystis sp*. from groundwater		
BF-4-SURF.Bins.20	**16S** (738 bp)	98.64	*Methylocystis rosea*	84.84	-
BF-4-SURF.Bins.21	**16S** (740 bp)	96.79	*Methylocapsa aurea*	68.16	-
	***pmoA**/B/C*	73.47/68.8/83.3	*Methylocapsa aurea*		
MD-6A-SURF.Bins.52	16S (1463 bp)	98.36	*Candidatus Methanoperedens nitroreducens*	73.73	-

Note:

a. 16S rRNA gene and/or full-length subunit genes of pMMO (*pmoA/B/C*) and sMMO (*mmoX*) with high-quality annotations were chosen for comparisons. The five genes highlighted with bold were included in the 16S rRNA and *pmoA* gene phylogenetic tree.

b. 16S rRNA genes were searched using blastn, while functional marker genes (*pmoA/B/C* and *mmoX*) were searched using blastp.

c. Lists the closest metagenomic assembly for MD-6A-SURF.Bins.48 and the closest isolates or enrichments for other entries.

d. The AAI values were extracted from corresponding AAI heatmap.

e. "-" indicates the taxon of "Closest hits using AAI" is identical with those shown in the column of "Closest hits using Blast". Or else, the specific taxon was separately listed.

FIGURE LEGENDS

Figure 3.1. Schematic diagram of a cross section of the Sanford Underground Research Facility (SURF) with plan view of main science levels from 300 ft to 4,850 ft underground. Samples for this study were retrieved from 1,700 ft and 4,850 ft levels. Gray lines illustrate the network of tunnels, ramps, shafts and hoists. Diagram is adapted from (https://indico.cern.ch/event/606690/).

Figure 3.2. Photos of three biofilm samples (d - Wall Biofilm, e - Orange Black Biofilm, f - White Biofilm) and three boreholes (a - Borehole 8, b - Borehole B, c - Borehole D) at the 4,850 ft level, as well as the four samples at the 1,700 ft level (g - Underwater Pink Deposit, h - White Fungus, i - Ceiling Stalactite, j - Underwater Brown Crust).

Figure 3.3. Composition of bacterial classes in the ten subsurface samples resolved into 4 bio-replicates per sample type. The eight most abundant classes (average relative abundance more than 2.4% across all samples) were listed along with their associated phylum-level phylogenies.

Figure 3.4. Composition of MOB lineages in subsurface samples uncovered using *pmoA* gene (a) and bacterial 16S rRNA gene (b) sequence analysis respectively. In the white fungus, methanotrophs were only detected in two bio-replicates. In the ceiling stalactite, methanotrophs were only detected by *pmoA* gene sequence analysis.

Figure 3.5. Relative abundance of bacterial classes (a) and MOB lineages (b & c) in the ten subsurface samples based on bacterial 16S rRNA gene sequencing (a & b) and *pmoA* gene sequencing (c). The samples are arranged from deep (4,850 ft, 3 biofilms and 3 groundwaters) to shallow levels (1,700 ft, 4 other types of tunnel samples). a. The 27 abundant classes with at least 2% of any individual sample and their associated phylum-level phylogenies were shown. Rare classes with less than 2% in every sample were grouped together as "Classes <2% abund.", while the sequences that could not be classified with accuracy at the phylum level were denoted as "Unassigned Bacteria". b. The bottom percentage numbers indicate the relative abundance of MOB in bacterial communities. MOB were not detected by 16S rRNA gene primers in three of ten samples. c. MOB were not amplified with *pmoA* primers in two of ten samples. b & c shared a legend of MOB lineages, where () indicate that the corresponding lineages were only detected by 16S rRNA gene or *pmoA* gene sequencing analysis and "/" indicate two separate lineage names were assigned based on the two marker genes (16S/*pmoA*) respectively. Each bar was collapsed from four bio-replicates, except methanotrophic populations of white fungus from two replicates (see Fig. 3.4).

Figure 3.6. Ecological diversity indices (Chao1 richness estimator, Shannon diversity and PD_whole_tree) of bacterial communities in the ten subsurface samples. "PD_whole_tree" is based on Faith's Phylogenetic Diversity, calculated by summing the phylogenetic distances of all microbial members in specific sample.

Figure 3.7. Neighbor joining tree showing the phylogenetic relationships of SURF methanotrophic OTUs with cultivated MOB strains and the closest environmental relatives based on 16S rRNA gene sequences including two 16S sequences extracted from MAGs. A base tree was initially constructed in ARB and bootstrapped at 1000 iterations (>50% indicated by black dots). The short methanotrophic OTU amplicon sequences and 16S rRNA gene sequences of MAGs were inserted into the base tree using parsimony methods in ARB. Colored branches contain the query sequences detected in this study. Blue labeled nodes indicate that *pmoA* gene sequence analysis (Fig. 3.8) retrieved the corresponding query sequences from the MOB lineages, whereas query entries highlighted with red were only retrieved through 16S rRNA gene analysis. MOB lineages are distinguished with outside range colors, and the color patterns were shared with the *pmoA* phylogenetic tree (Fig. 3.8) and its related composition bar (Fig. 3.5b). As shown in the legend, un-bolded lineages were excluded from the *pmoA* tree due to the absence of a *pmoA* gene (e.g., *Methyloferula*) or the presence of an unusual *pmoA* gene (e.g., *Crenothrix)* and lineage names in brackets displayed the associated names in *pmoA* tree. OTUs include a pie chart, with colors demonstrating the seven samples to which the majority of reads were derived. The size of the pies corresponds to the OTU total read counts (log_{10} transformed). The tree was rooted with sequences of two methanogenic strains (CP002565, AM114193). The scale bar represents 0.1 substitution per nucleotide position.

Figure 3.8. *pmoA* gene based Maximum likelihood tree showing the phylogenetic relationship of SURF methanotrophic OTUs and three *pmoA* sequences extracted from

MAGs with related environmental sequences and cultivated MOB strains. This amino acid based tree was derived using by RAxML using the JTT substitution and GAMMA rate heterogeneity models as well as the rapid bootstrap algorithm with 100 iterations (>50% indicated by black dots). Colored text denotes the query *pmoA* sequences detected in this study with blue indicating that 16S rRNA gene analysis (Fig. 3.7) retrieved the corresponding query sequences from the consistent MOB lineages, whereas entries highlighted with red were only retrieved through *pmoA* gene analysis. The MOB lineage legend corresponds to the one shown in 16S rRNA gene phylogenetic tree (Fig. 3.7) and its composition bar (Fig. 3.5c). In the legend, un-bolded lineages were not observed in the 16S rRNA gene tree and lineage names in brackets displayed the associated names in 16S rRNA gene tree. OTUs include a pie chart, with colors demonstrating the eight samples to which the majority of reads were derived The size of the pies corresponds to the OTU total read counts ($\log_{10}$ transformed). The tree was rooted with *amoA* sequences of two ammonia-oxidizing strains (X90822, FOBH01000004). The scale bar represents 0.1 substitution per amino acid position.

Figure 3.9. UpsetR plot comparing numbers of shared OTUs among ten SURF samples with a shared frequency (denoted on top of each bar) ≥ 10. Matrix rows include the ten samples arranged in order from shallow level 1,700 ft (top) to deep level 4,850 ft (bottom), with their total OTU numbers shown in the left horizonal bars. Columns of the matrix show OTU sharing patterns. The vertical bar chart shows OTUs numbers exclusive to those groups, arranged in order from high to low. A single dot denotes non-shared OTUs within that group, with linked dots showing numbers of OTUs shared and

exclusive to those linked samples. Red, blue and orange colors respectively depict shared patterns present within 3 groundwaters (7 patterns), within 3 biofilms (5 patterns) and between groundwaters and biofilms (3 patterns).

Figure 3.10. Non-metric multidimensional scaling (NMDS) ordination of Bray-Curtis distances among SURF samples. Shapes differentiate samples by depth, while colors differentiate the ten samples. 95% confidence interval ellipses are plotted for each of three groups of samples (Biofilm, Borehole, Others) and show separation among groups.

Figure 3.11. a. The OTU co-occurrence network of the bacterial community in SURF biofilms. The size of each node is proportional to the number of connected edges. The color of nodes separates the different module types. Positive and negative correlations are respectively displayed as red and blue lines. The label denotes the most resolved taxonomic affiliation of each node/OTU with MOB taxa highlighted with red text. b. Distribution of nodes based on their topological features. Each dot represents an OTU belonging to MOB (red) and non-MOB (gray). The putative keystone role of MOB within the biofilm network was evaluated based on the scatter plot of their within-module connectivity and among-module connectivity.

Figure 3.12. The OTU co-occurrence network of the bacterial community in SURF borehole groundwaters. The size of each node is proportional to the number of connected edges. The color of nodes separates the different module types. Positive and negative

correlations are respectively displayed as red and blue lines. The label denotes the most resolved taxonomic affiliation of each node.

Figure 3.13. Genome-based phylogenetic tree of gamma-MOB MAGs reconstructed from SURF biofilms. Phylogenomic tree shows the position of query MAGs (red text) in relation to gamma-MOB reference genomes based on the comparative sequence analysis of 92 bacterial core marker genes. The type Ib MOB reference genomes are shown with blue text and branches. Gene support index values are indicated on the branches. Two non-MOB strains within Gamma-proteobacteria were used as the outgroup to root the tree. Bar, 0.1 substitutions per amino acid position. "BF-∗" recovered from wall biofilm; "OBB-∗" recovered from orange black biofilm.

Figure 3.14. Heatmap based on AAI similarity matrix, displaying genome relatedness (AAI value %) between gamma-MOB MAGs and other gamma-MOB type strain genomes. The AAI values between pairs of genomes are depicted using a relative thermal color scale. Hierarchical clustering of the genomes is based on pairwise AAI similarity. "BF-∗" recovered from wall biofilm; "OBB-∗" recovered from orange black biofilm.

Figure 3.15. Genome-based phylogenetic tree of alpha-MOB MAGs recovered from SURF biofilms and groundwater. Phylogenomic tree shows the position of query MAGs (red text) in relation to alpha-MOB type genomes based on the comparative sequence analysis of 92 bacterial core marker genes. The alpha-MOB type reference strains are highlighted with blue text and branches. Gene support index values are indicated on the

branches. Six non-MOB strains within Alpha-proteobacteria were used as the outgroup to root the tree. Bar, 0.1 substitutions per amino acid position. “BF-∗” recovered from wall biofilm; “MD-∗” recovered from fluid of borehole manifold D.

Figure 3.16. Heatmap based on AAI similarity matrix, displaying genome relatedness (AAI value %) between alpha-MOB MAGs and other alpha-MOB type reference genomes. The AAI values between pairs of genomes use the related thermal color scale. Hierarchical clustering of the genomes is based on pairwise AAI similarity. “BF-∗” recovered from wall biofilm; “MD-∗” recovered from fluid of borehole manifold D.

Figure 3.17. Heatmap based on AAI similarity matrix displaying genome relatedness (AAI value %) between a likely methanotrophic MAG in the domain Archaea, reference methanogen genomes and anaerobic methanotrophs (ANME) representative genomes. The AAI values between pairs of genomes use the related thermal color scale. Hierarchical clustering of the genomes is based on pairwise AAI similarity. “MD-∗” recovered from fluid of borehole manifold D.

FIGURES

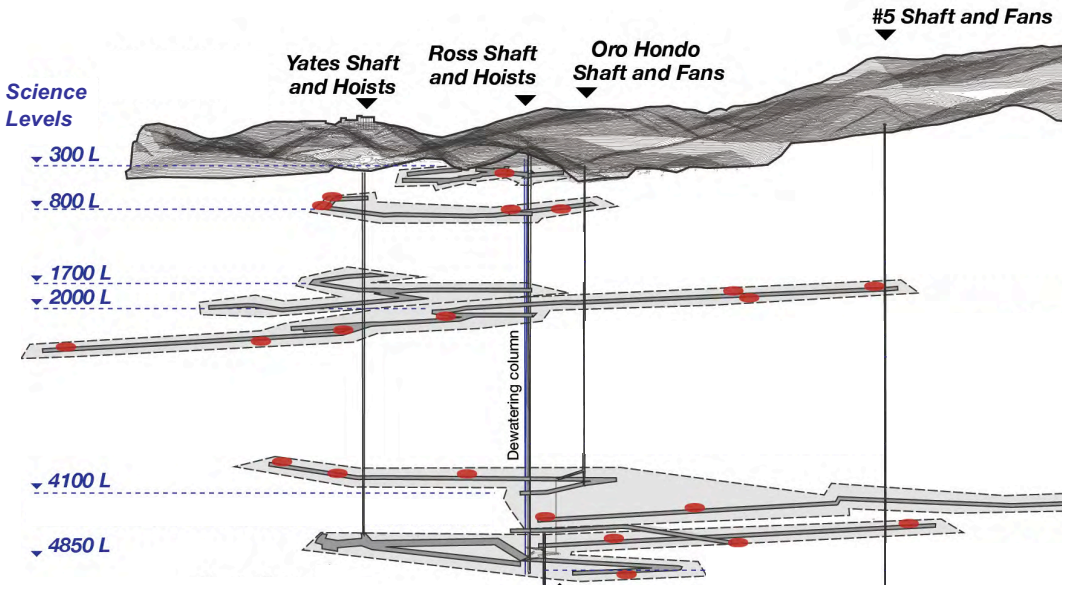
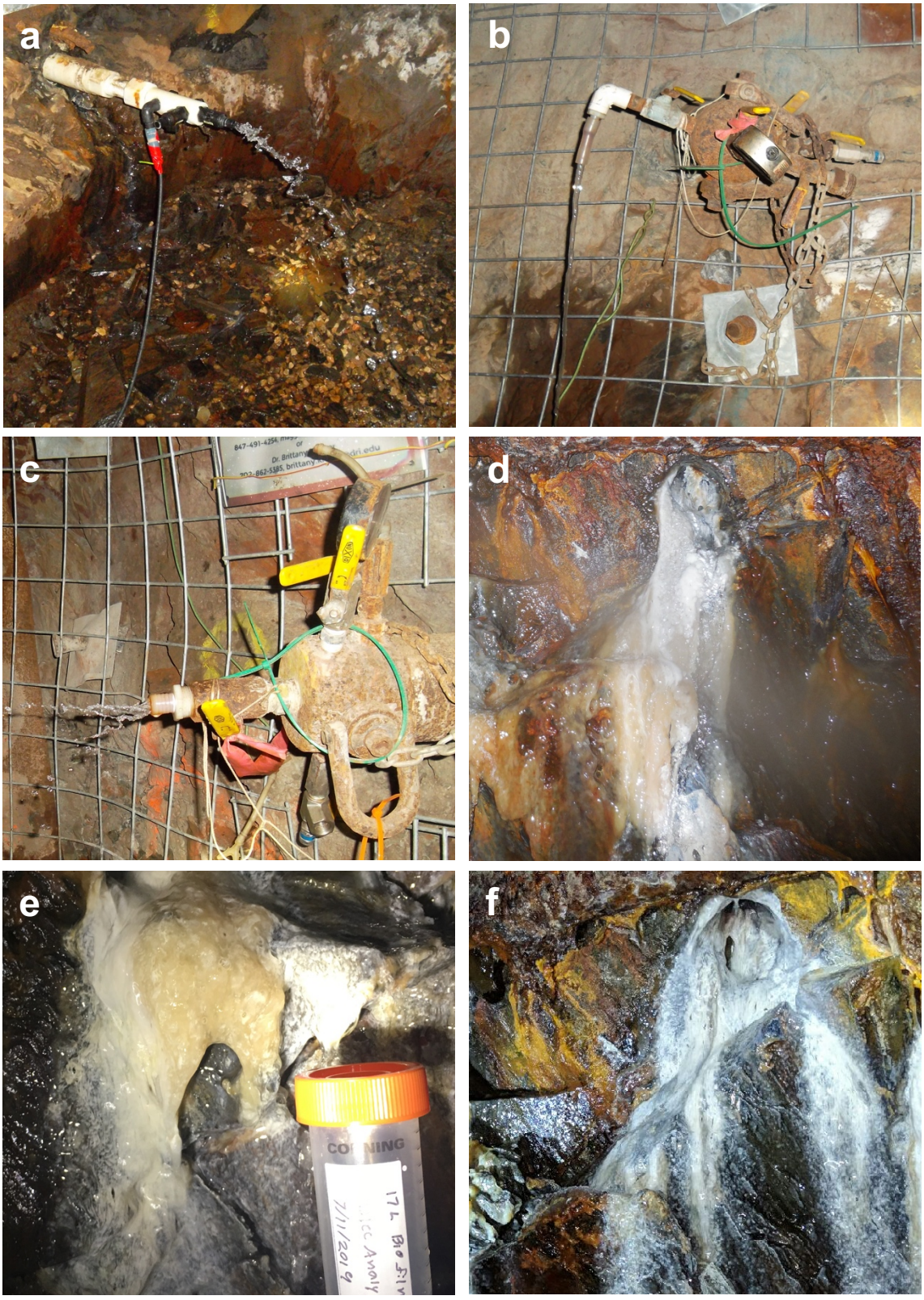


Figure 3.1. Schematic diagram of a cross section of the Sanford Underground Research Facility (SURF)


a
b
c
d
e
f

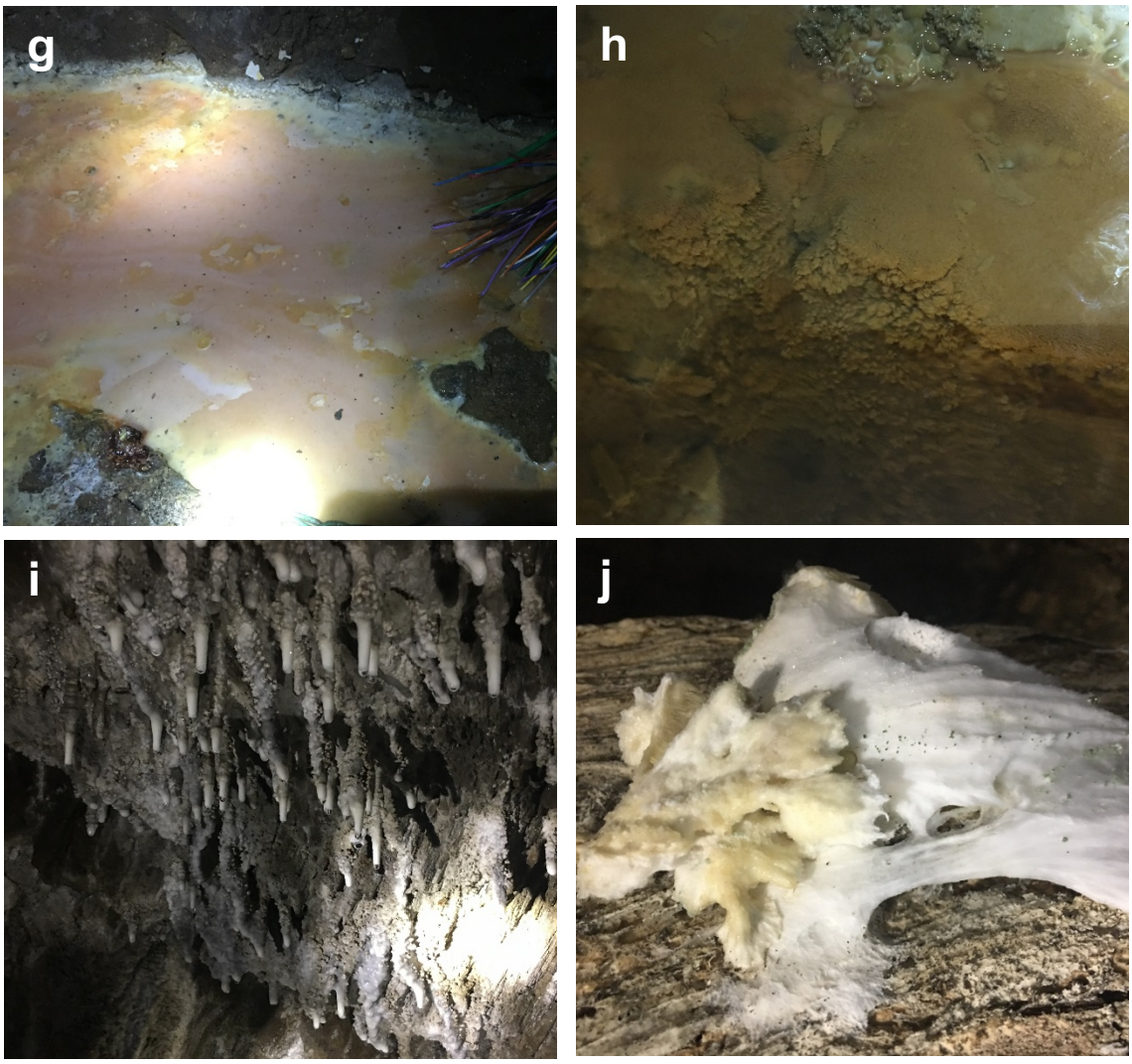


Figure 3.2. Photos of SURF samples.

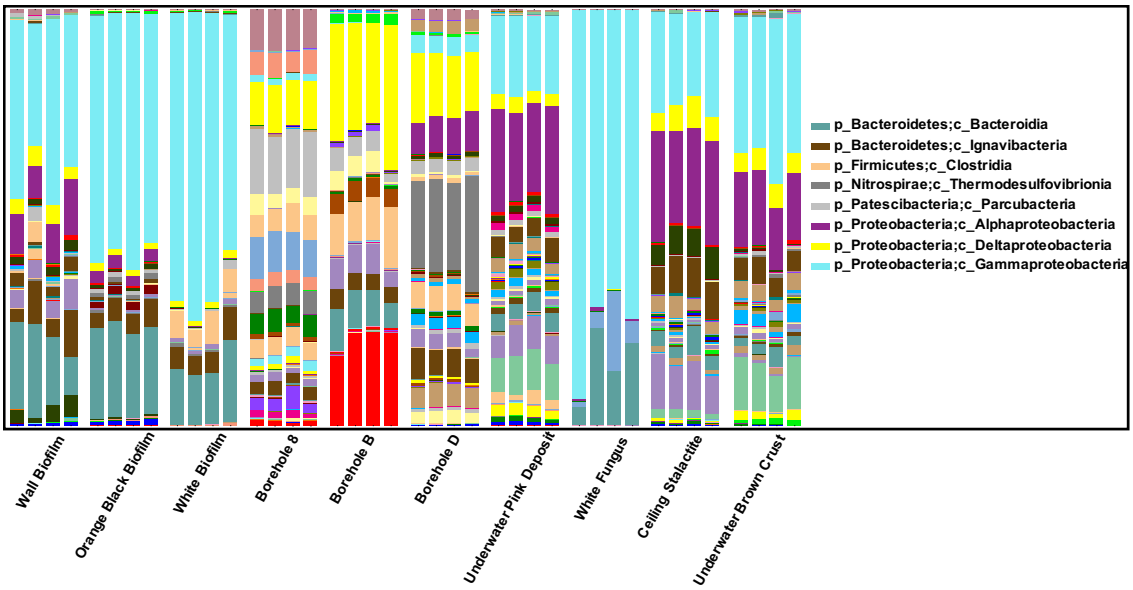


Figure 3.3. Composition of bacterial classes in the ten subsurface samples resolved into 4 bio-replicates per sample type.

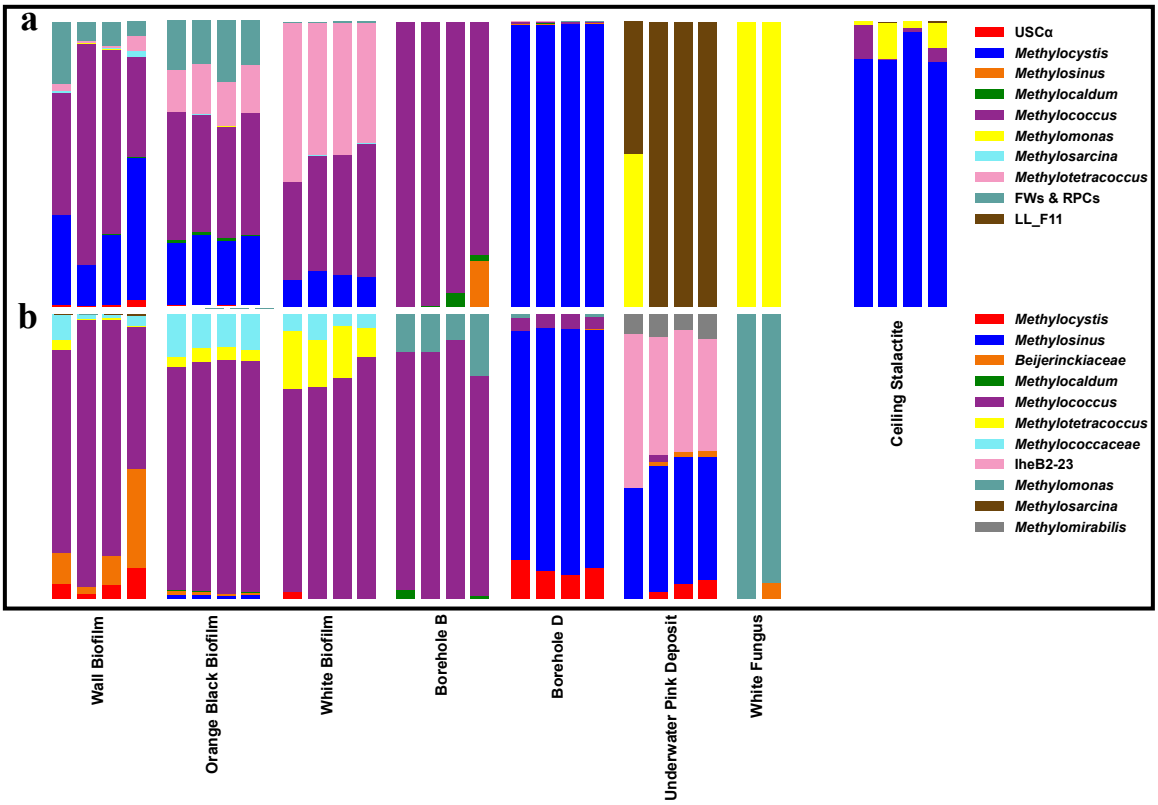


Figure 3.4. Composition of MOB lineages in subsurface samples uncovered using *pmoA* gene (a) and bacterial 16S rRNA gene (b) sequence analysis respectively.

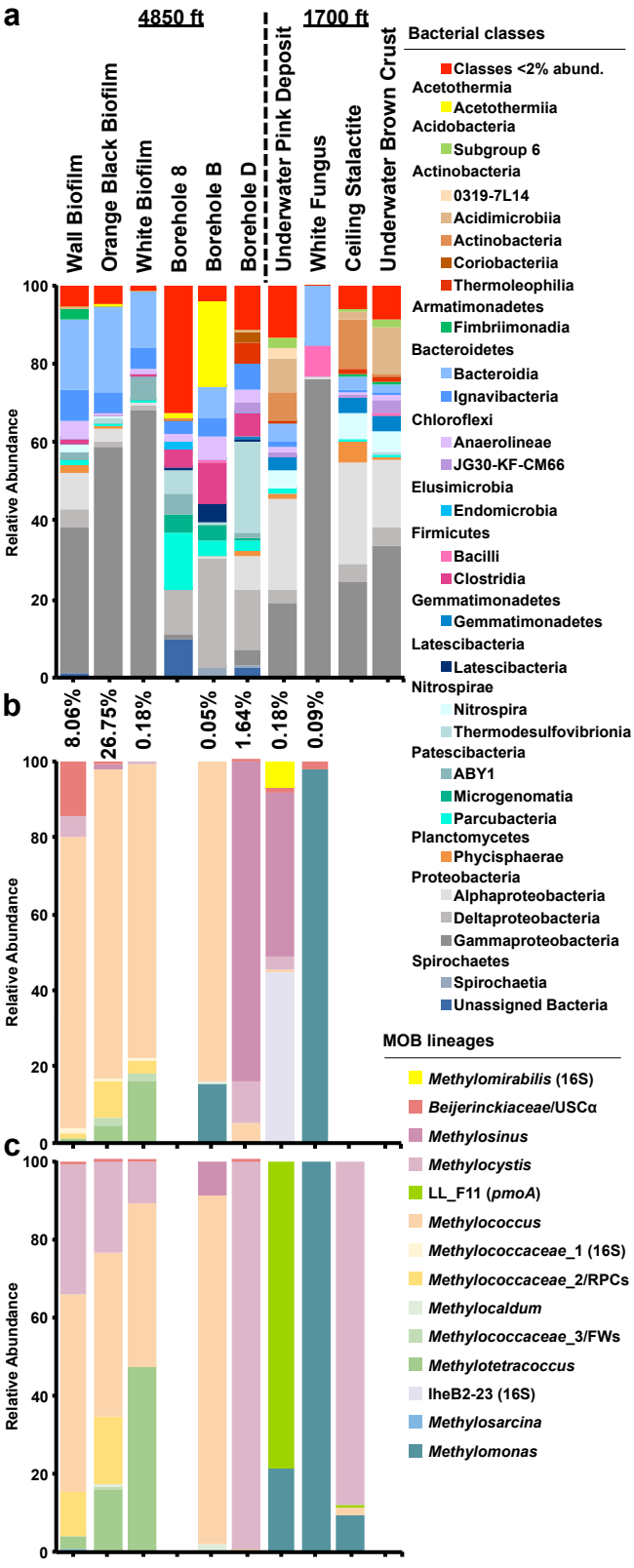


Figure 3.5. Relative abundance of bacterial classes (a) and MOB lineages (b & c).

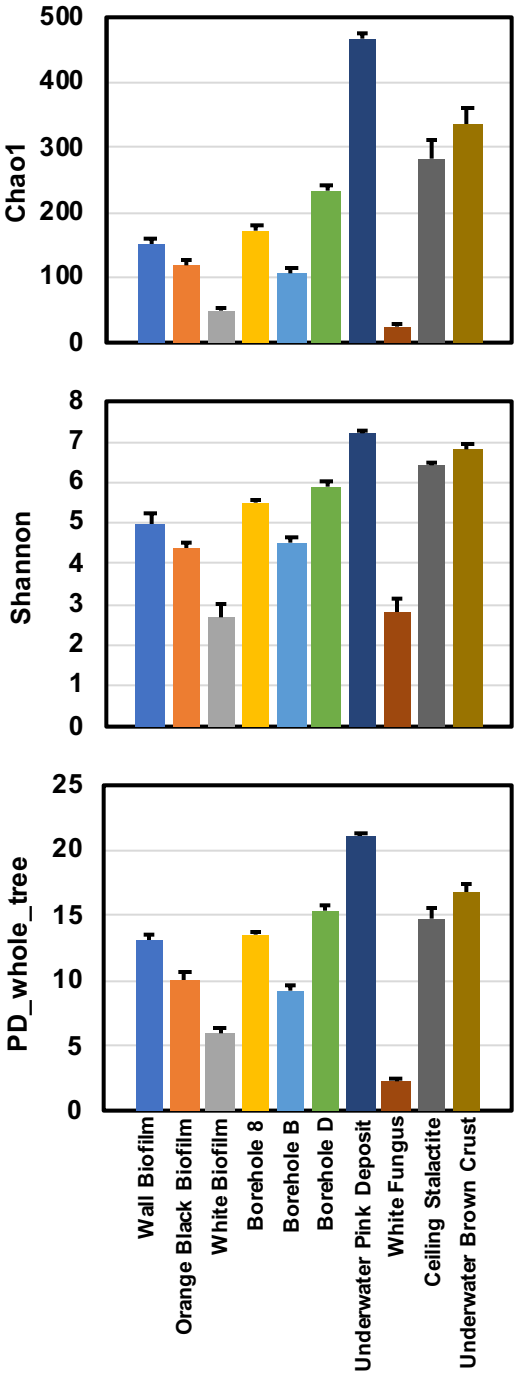


Figure 3.6. Ecological diversity indices of bacterial communities.

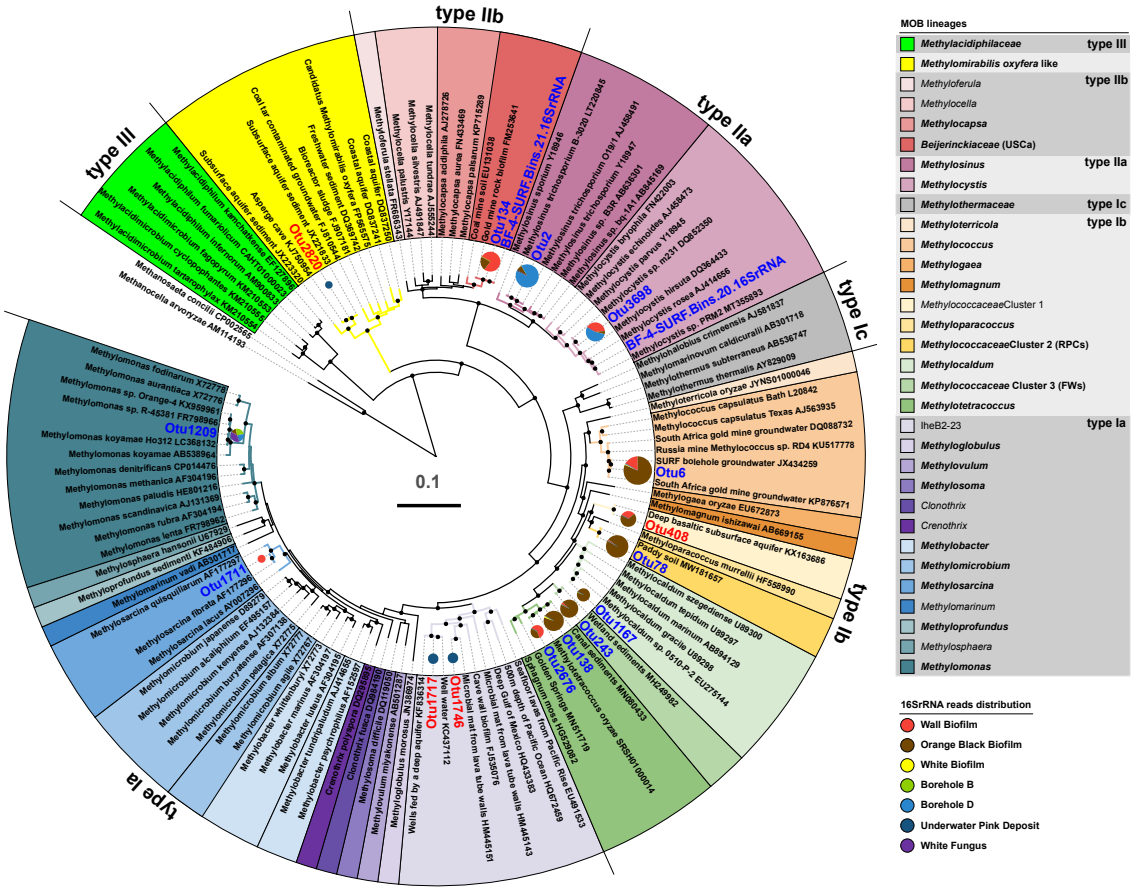


Figure 3.7. Neighbor joining tree showing the phylogenetic relationships of SURF methanotrophic OTUs with cultivated MOB strains and the closest environmental relatives based on 16S rRNA gene sequences.

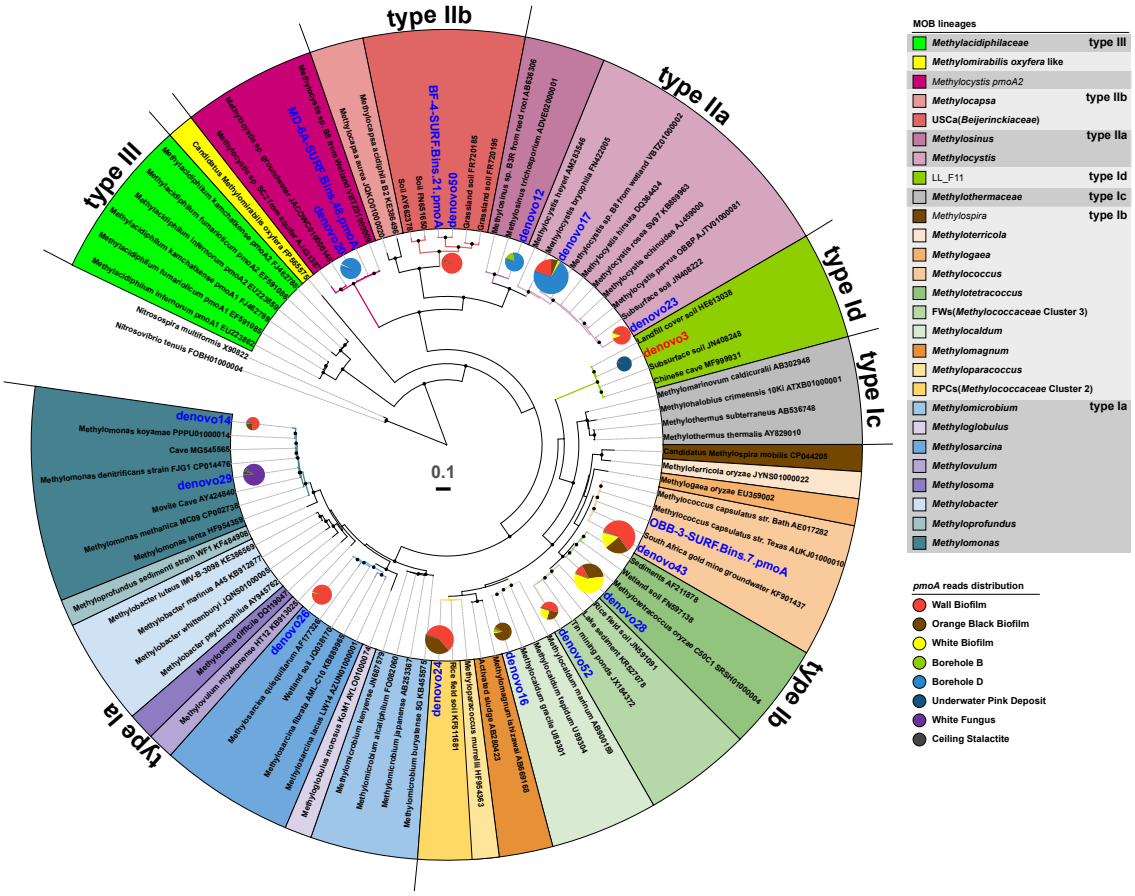


Figure 3.8. ***pmoA*** **gene based Maximum likelihood tree showing the phylogenetic relationship of SURF methanotrophic OTUs and three** ***pmoA*** **sequences extracted from MAGs with related environmental sequences and cultivated MOB strains.**

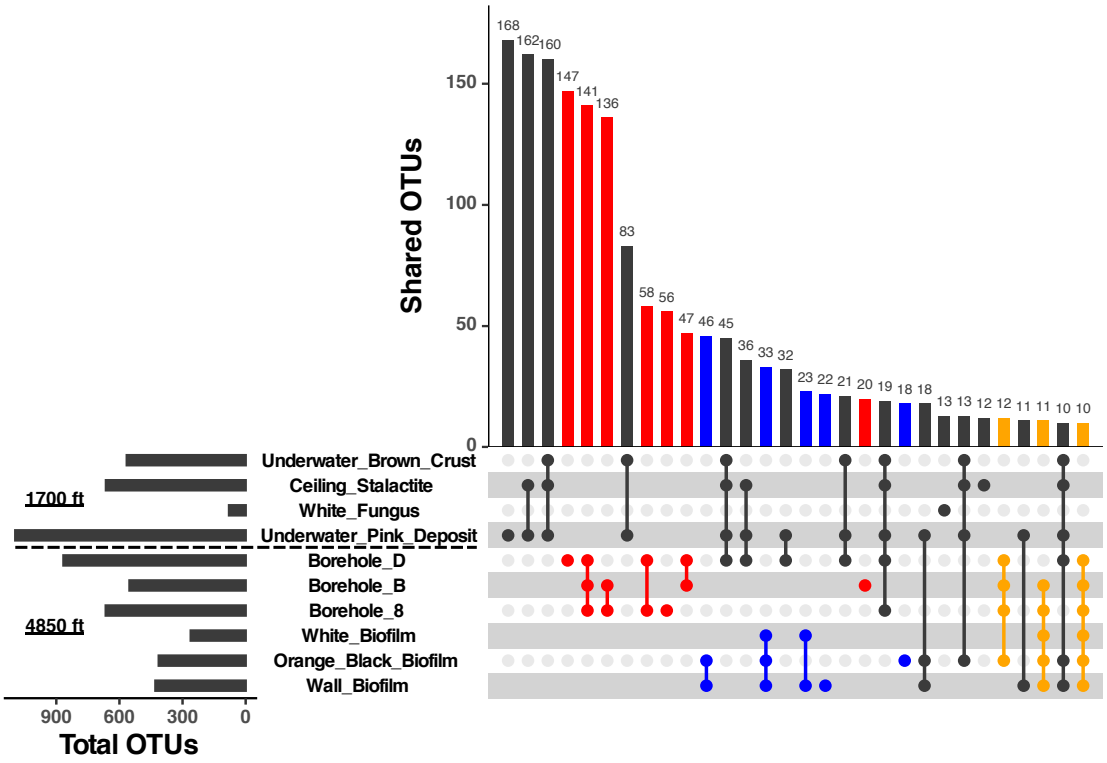


Figure 3.9. UpsetR plot comparing numbers of shared OTUs among ten SURF samples.

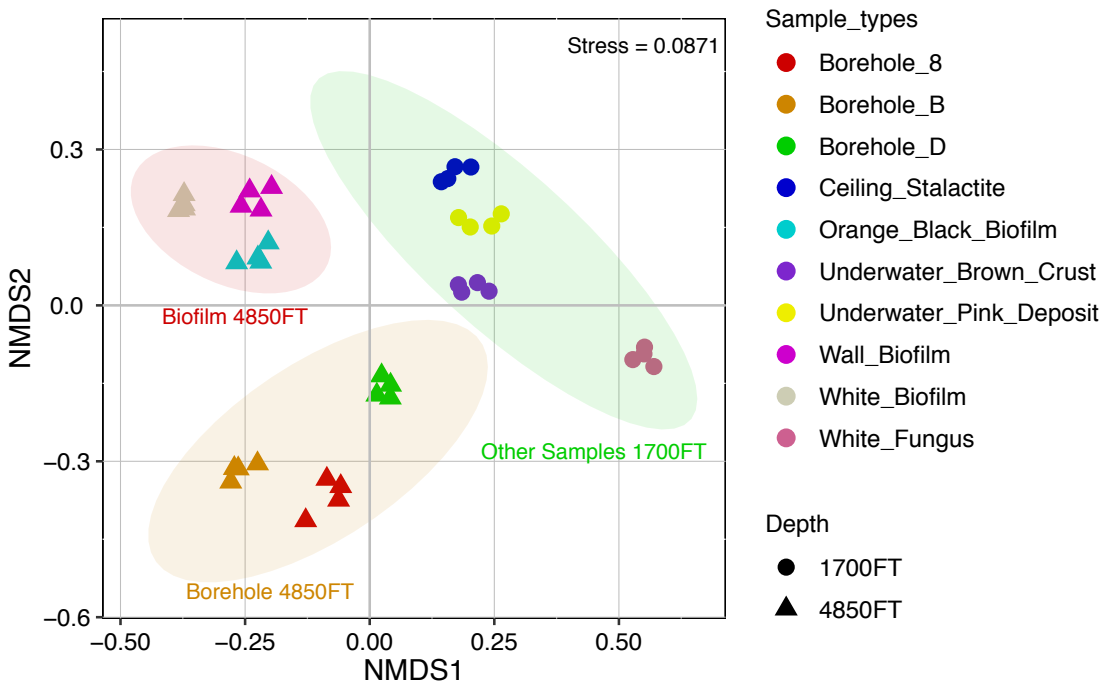


Figure 3.10. Non-metric multidimensional scaling (NMDS) ordination of Bray-Curtis distances among SURF samples.

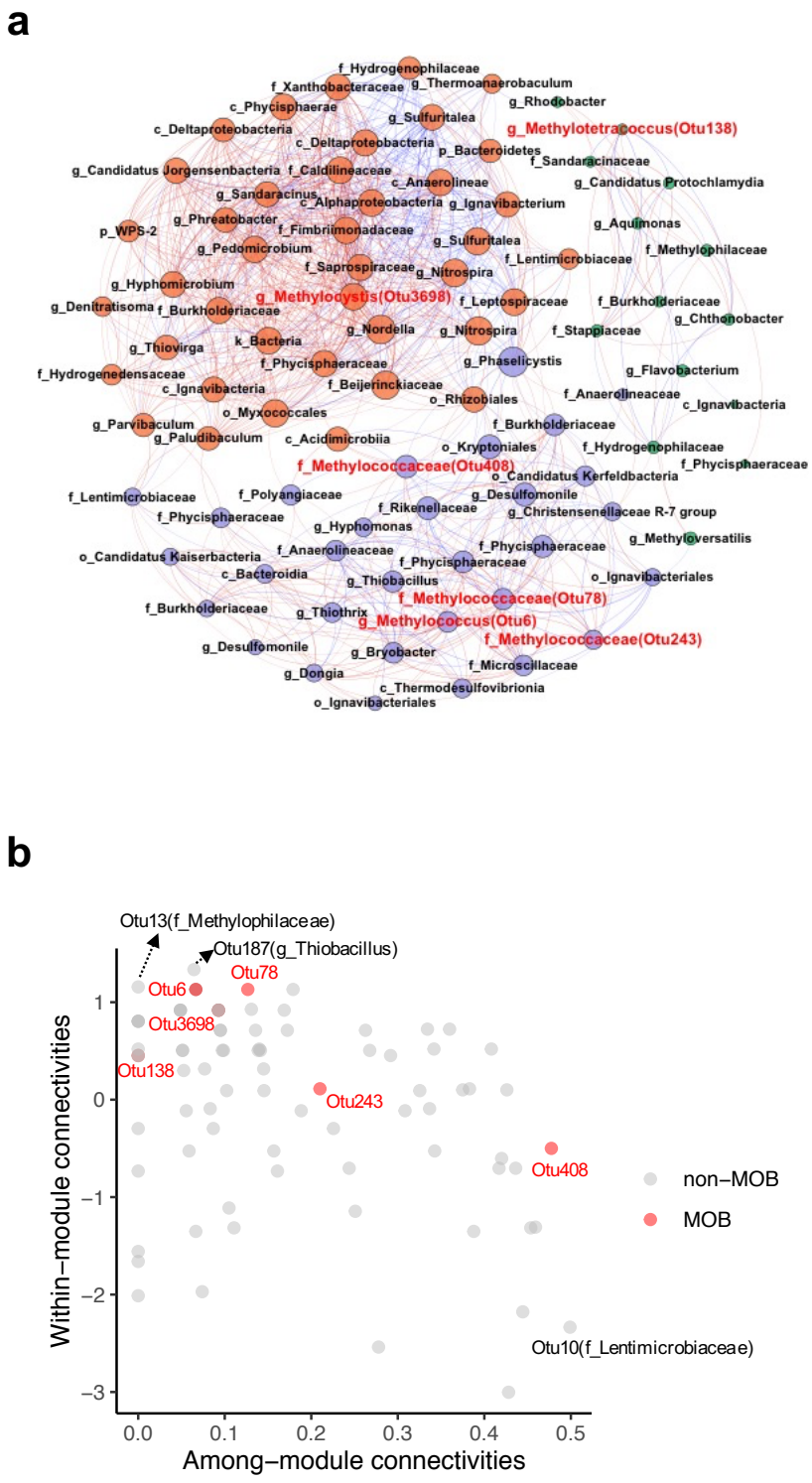


Figure 3.11. a. The OTU co-occurrence network of the bacterial community in SURF biofilms. b. Distribution of nodes based on their topological features.

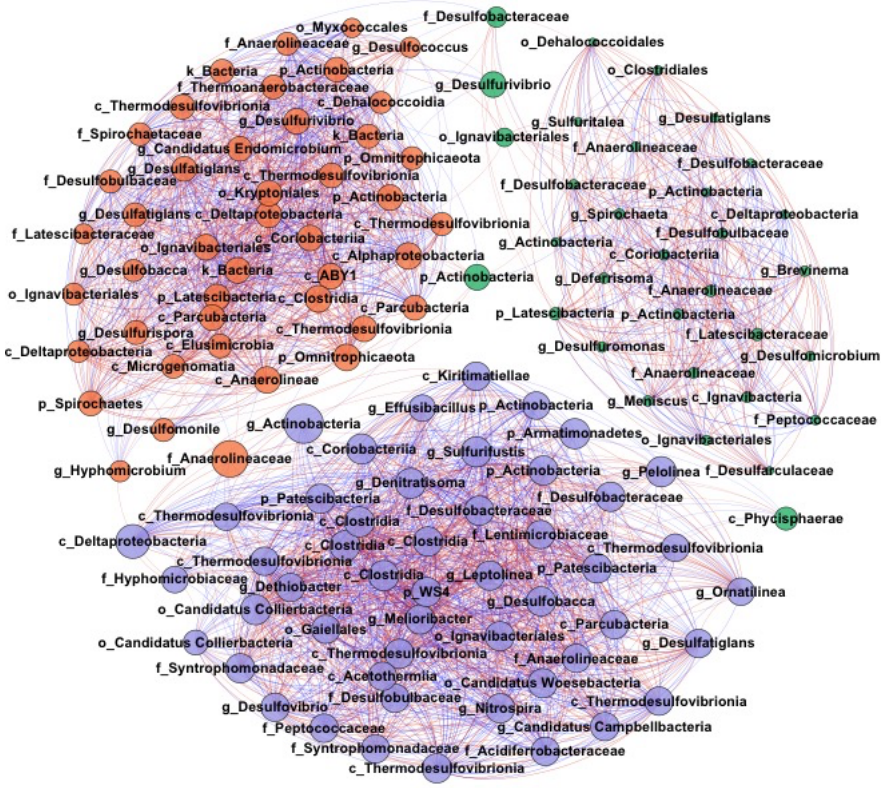


Figure 3.12. The OTU co-occurrence network of the bacterial community in SURF borehole groundwaters.

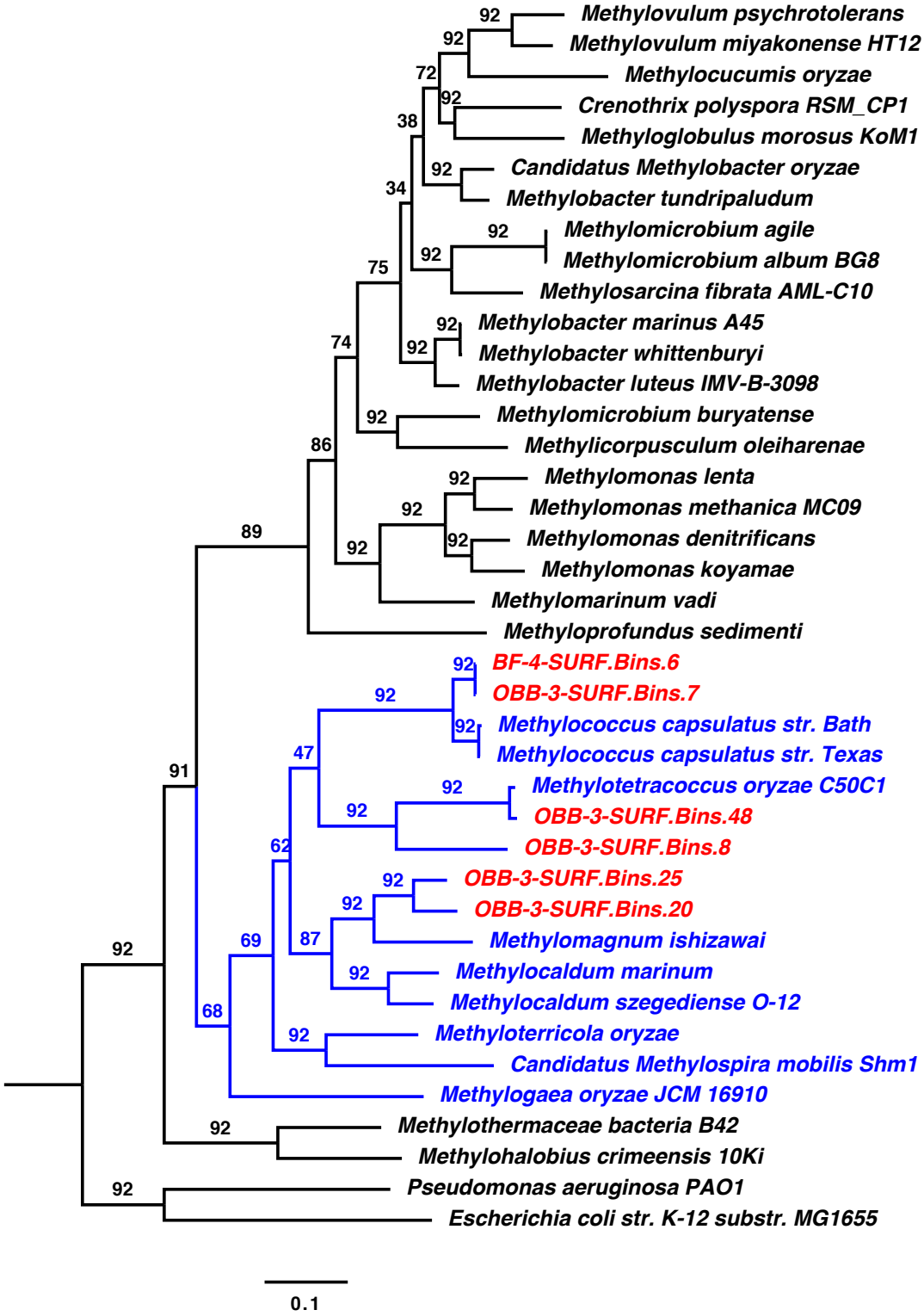


Figure 3.13. Genome-based phylogenetic tree of gamma-MOB MAGs reconstructed from SURF biofilms.

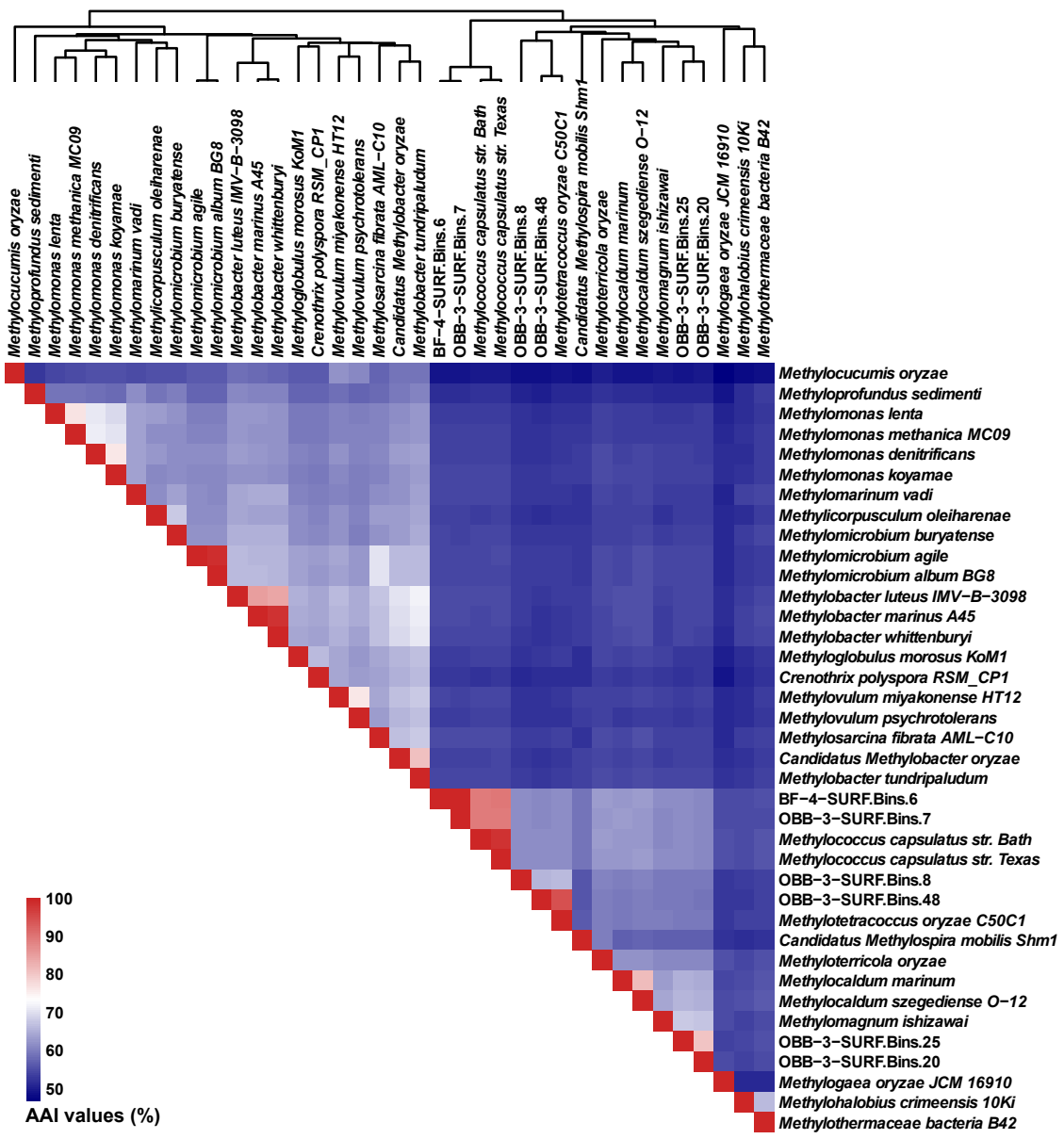


Figure 3.14. Heatmap based on AAI similarity matrix, displaying genome relatedness (AAI value %) between gamma-MOB MAGs and other gamma-MOB type strain genomes.

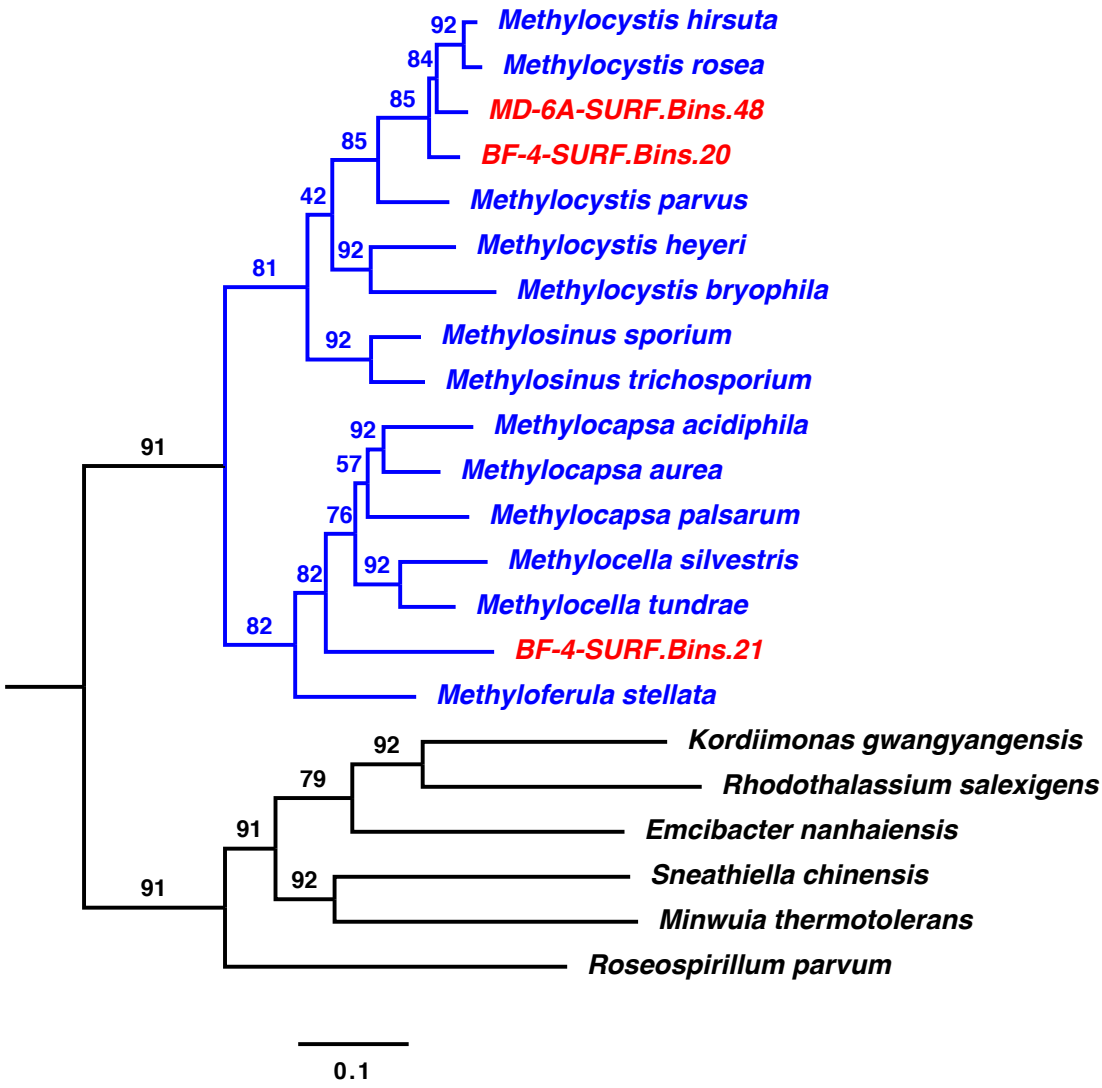


Figure 3.15. Genome-based phylogenetic tree of alpha-MOB MAGs recovered from SURF biofilms and groundwater.

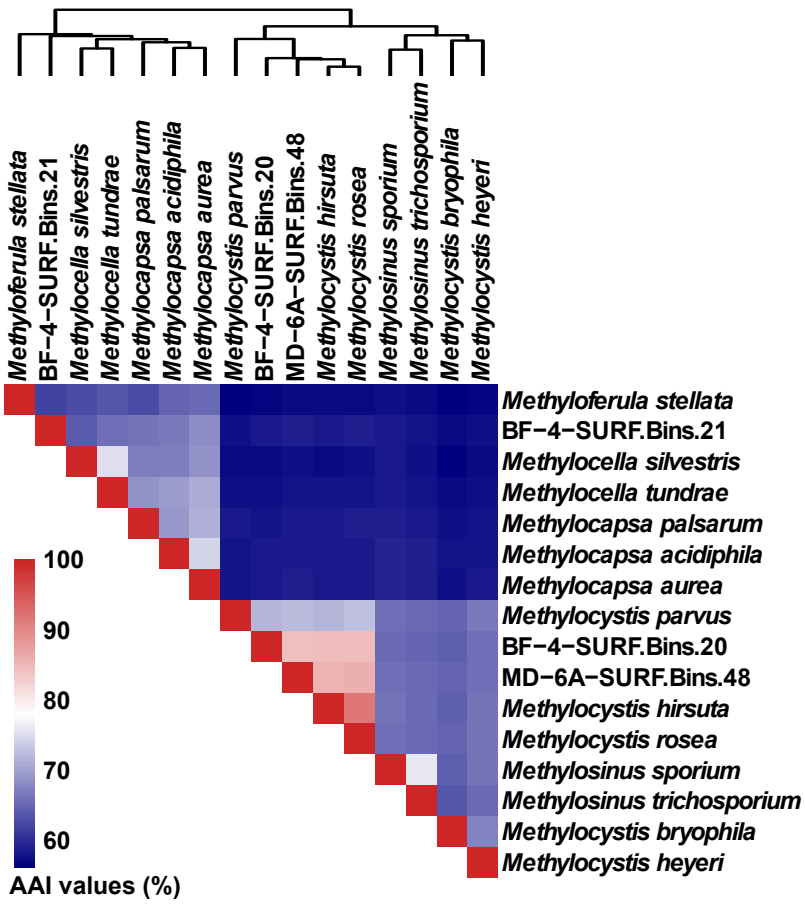


Figure 3.16. Heatmap based on AAI similarity matrix, displaying genome relatedness (AAI value %) between alpha-MOB MAGs and other alpha-MOB type reference genomes.

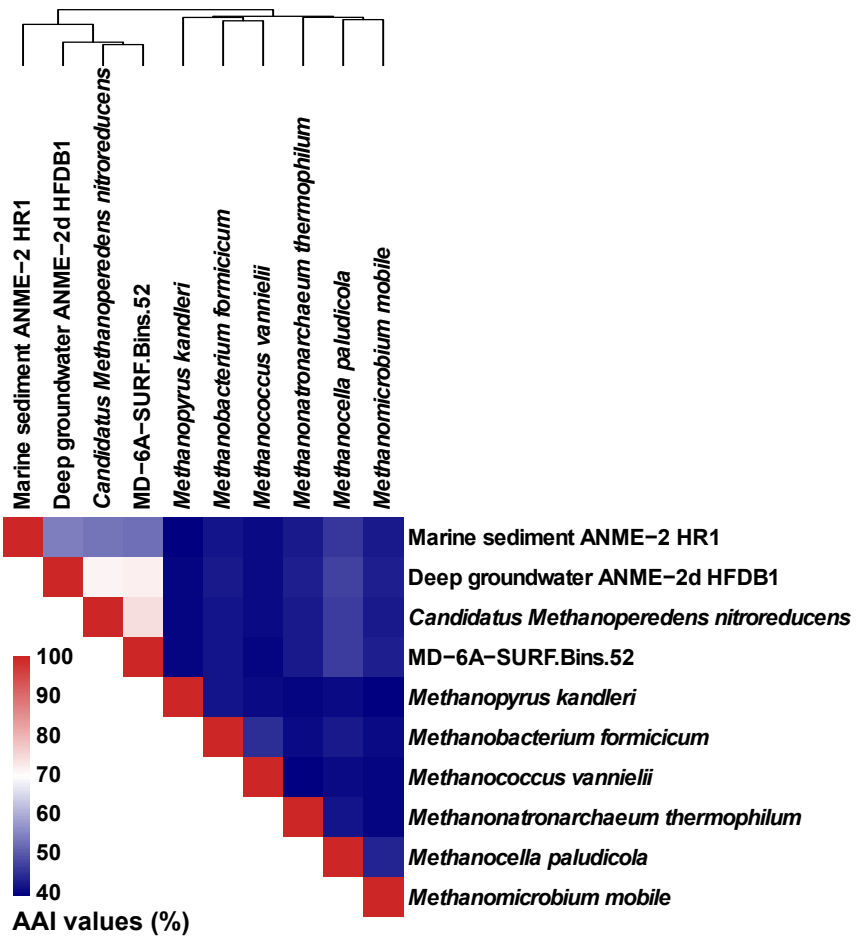


Figure 3.17. Heatmap based on AAI similarity matrix displaying genome relatedness (AAI value %) between a likely methanotrophic MAG in the domain Archaea, reference methanogen genomes and anaerobic methanotrophs (ANME) representative genomes.

APPENDIX

LIST OF ABBREVIATIONS

ANME	Anaerobic methanotrophic archaea
AAI	Average amino acid identity
Adonis	Permutational multivariate analysis of variance
ANI	Average nucleotide identity
BLAST	Basic Local Alignment Search Tool
bp	Base pairs (nucleic acid)
CBB	Calvin-Benson-Bassham
cyanoHABs	cyanobacterial harmful algal blooms
DGGE	Denaturing gradient gel electrophoresis
DNA	Deoxyribonucleic acid
DO	Dissolved oxygen
DOC	Dissolved organic carbon
FISH	Fluorescence in situ hybridization
ft	Feet
ICMs	Intracytoplasmic membranes
KEGG	Kyoto Encyclopedia of Genes and Genomes
km^2	Square kilometer
MAFFT	Multiple alignment using fast Fourier transform
MAG	Metagenome-assembled genome
MCR	Methyl-coenzyme M reductase

MOB	Methane-oxidizing bacteria
NCBI	National Center for Biotechnology Information
NMDS	Non-metric multidimensional scaling
OTU	Operational taxonomic unit
PCoA	Principal coordinates analysis
PCR	Polymerase Chain Reaction
pMMO	Particulate methane monooxygenases
QIIME	Quantitative Insights Into Microbial Ecology
qPCR	Quantitative polymerase chain reaction
RAxML	Randomized Axelerated Maximum Likelihood
RDP	Ribosomal Database Project
rRNA	Ribosomal ribonucleic acid
RuMP	Ribulose monophosphate
sMMO	Cytoplasmic-soluble methane monooxygenases
SURF	Sanford Underground Research Facility
TN	Total Nitrogen
TP	Total Phosphorus