Metagenomic Approaches Unearth Methanotroph Phylogenetic and Metabolic Diversity

Garrett J. Smith\textsuperscript{1}\textsuperscript{*} and Kelly C. Wrighton\textsuperscript{1,2}\textsuperscript{*}

\textsuperscript{1}Department of Microbiology, The Ohio State University, Columbus, OH, USA.
\textsuperscript{2}Department of Soil and Crop Science, Colorado State University, Fort Collins, CO, USA.

\textsuperscript{*}Correspondence: smith.10284@osu.edu and kelly.wrighton@colostate.edu

Abstract

Methanotrophic microorganisms utilize methane as an electron donor and a carbon source. To date, the capacity to oxidize methane is restricted to microorganisms from three bacterial and one archaeal phyla. Most of our knowledge of methanotrophic metabolism has been obtained using highly enriched or pure cultures grown in the laboratory. However, many methanotrophs currently evade cultivation, thus metagenomics provides a complementary approach for gaining insight into currently unisolated microorganisms. Here we synthesize the studies using metagenomics to glean information about methanotrophs. We complement this summary with an analysis of methanotroph marker genes from 235 publicly available metagenomic datasets. We analyse the phylogenetic and environmental distribution of methanotrophs sampled by metagenomics. We also highlight metabolic insights that methanotroph genomes assembled from metagenomes are illuminating. In summary, metagenomics has increased methanotrophic foliage within the tree of life, as well as provided new insights into methanotroph metabolism, which collectively can guide new cultivation efforts. Lastly, given the importance of methanotrophs for biotechnological applications and their capacity to filter greenhouse gases from a variety of ecosystems, metagenomics will continue to be an important component in the arsenal of tools needed for understanding methanotroph diversity and metabolism in both engineered and natural systems.

Introduction

Ferdinand Cohn described a filamentous microorganism that quickly earned infamy for flourishing in European waterworks and obstructing the flow of water in the late 1800s (Cohn, 1870). More than 100 years after the initial discovery of \textit{Crenothrix polyspora}, this organism still evades cultivation. In the past 40 years, ultrastructure evidence (Völker \textit{et al.}, 1977) and the presence of key functional genes (Stoecker \textit{et al.}, 2006) hinted that these pervasive filamentous cells grew using methane as a carbon and energy source. In 2017, metagenomic sequencing, or the untargeted sequencing of DNA directly from a microbial community, led to the first insight into \textit{Crenothrix} genomic contents (Oswald \textit{et al.}, 2017). In combination with other chemical and imaging technologies, here metagenomics confirmed that these mysterious, filamentous cells were methanotrophs. This is only one example of how metagenomics, especially when combined

https://doi.org/10.21775/cimb.033.057
with other methodologies, has demystified the physiology of uncultivated microorganisms (Brown et al., 2015; Solden et al., 2016).

Methanotrophic microorganisms play immense roles in mediating biogeochemical cycles. These organisms can filter 50–97% of methane from ecosystems (Segarra et al., 2015), and therefore have a large impact on the release of this potent greenhouse gas to the atmosphere. In addition to the carbon cycle, it has also been shown that methanotrophs can aid in the removal of 90–99% of the nitrogen during wastewater treatment (Jewell et al., 1992; He et al., 2015). More recently, methanotrophic strains have been harnessed to degrade organic pollutants, produce a variety of industrial precursors, and scavenge metals such as copper (Sullivan et al., 1998; Semrau et al., 2010; Puri et al., 2015; Strong et al., 2015, 2016). Given their important biogeochemical and biotechnological applications, metagenomics is a critical research tool for improved understanding of methanotroph diversity and physiology in natural and manmade settings.

Today, with the power of recovering near-complete to closed genomes directly from microbial communities, axenic cultivation is no longer a requirement for investigating a microorganism’s metabolic capabilities (Solden et al., 2016). Metagenomics, which provides an inventory of the metabolic potential of a community, is especially powerful when utilized in parallel with metatranscriptomics (community RNA sequencing) or metaproteomics (community protein analyses). These latter technologies provide information on gene and protein expression in a microbial community, indicating metabolic processes and organisms that are responding to specific environmental conditions. Here our objective is to describe how metagenomics and enabled ‘omic’ technologies (multi-omics) have altered perspectives on methanotroph environmental distribution, phylogenetic diversity, and metabolism. Leveraging recent metagenomic studies, we (i) summarize key findings from a variety of ecosystems, (ii) describe the methanotroph genomic foliage added to the tree of life, and (iii) discuss how metagenomic analyses of methanotrophs has impacted other scientific disciplines.

**Overview of methanotroph phylogenetic and metabolic diversity**

Methanotrophs are currently assigned to three bacterial and one archaeal phyla (Fig. 3.1). Today all known Bacterial methanotrophs encode methane monoxygenase (MMO), the enzyme complex that oxidizes a C–H bond in methane using oxygen. There are two forms of this enzyme, particulate (pMMO) and soluble (sMMO), and at least one of these is required for aerobic oxidation of methane (Dalton, 1983). Functional marker genes for the particulate (pmaA) and soluble (mmaX) forms are commonly used to probe for bacterial methanotrophs (McDonald et al., 2008; Knief, 2015).

The most well studied bacterial methanotrophs belong to the phylum Proteobacteria specifically within the Gammaproteobacteria and Alphaproteobacteria classes. The Gammaproteobacteria methanotrophs belong to the order Methylocoales, and are often referred to as ‘Type I’ and ‘Type X’ methanotrophs. Within the Alphaproteobacteria, methanotrophs are associated with the order Rhizobiales, referred to as ‘Type II’ methanotrophs (Dalton, 1983). These two Proteobacteria methanotroph classes are distinguished not only by phylogenetic assignment, but also have differences in carbon assimilation pathways and intra-cytoplasmic membrane ultrastructure (Dalton, 1983; Knief, 2015).

The second bacterial phylum containing methanotrophs is the Verrucomicrobia. Here members of the family Methylocycliphiles, have core methane oxidation pathways resembling Proteobacteria methanotrophs. However, unlike the Proteobacteria methanotrophs, these taxa may also utilize hydrogen as an alternative electron donor and require lanthanide metals for growth (Pol et al., 2007, 2014; Mohammadi et al., 2017). The third bacterial phylum of methanotrophs is the Methylocirribilita, which contains the genus ‘Candidatus Methylomirabilis’ and were previously classified as members of the candidate NC10 phylum (Ettwig et al., 2010; Glöckner et al., 2017; Parks et al., 2018). These methanotrophs are proposed to generate molecular oxygen for methane oxidation by reducing nitrite, followed by dismutation of nitric oxide into dinitrogen and dioxygen (Ettwig et al., 2008, 2010; Wu et al., 2015). Thus, these bacteria are proposed to couple aerobic methane oxidation with
denitrification through an intra-aerobic pathway. Across all the bacterial methanotrophs sampled to date, while much of the enzymatic machinery for aerobic methane oxidation is strongly conserved (e.g. MMO), energy generation and other pathways differ between the phylum (Fig. 3.1) or even strains (Hoefman et al., 2014; Heylen et al., 2016).

Unlike the bacterial methanotrophs, the archaeal methanotrophs do not employ MMO. Archaeal ANAerobic MEthanotrophs (ANME) instead reverse the pathway used by the methanogens to produce methane, thereby oxidizing methane in the absence of molecular oxygen. Archaeal methane oxidation can be coupled to numerous electron acceptors including sulfate, nitrate, nitrite, hemic, iron or manganese. These organisms thrive in anoxic environments, and, like methanogens, are subjects to oxygen toxicity (Haroon et al., 2013; Scheller et al., 2016; Cai et al., 2018). The variation in electron acceptors, methane inputs, or biomass differences may partially explain the highly variable rates of anaerobic methane oxidation for these taxa across ecosystems (Knittel and Boetius, 2009).

The genes for the enzyme responsible for anaerobic oxidation of methane by ANME, methyl co-enzyme reductase (MCR), can be used in conjunction with 16S rRNA genes to phylogenetically identify these methanotrophs, and can differentiate them from related methanogens (Hallam et al., 2003). Using these marker genes,
all currently known ANME lineages are members within the Methanomicrobia class of the Euryarchaeota phylum. Within the Methanomicrobia, the ANME-1 appear to be a distinct order, several lineages of ANME-2 form a monophyletic clade within the Methanosarcinales order, while the ANME-3 likely represent a specific genus within the Methanosarcinaceae family (Knittel and Boetius, 2009; Glöckner et al., 2017; Timmers et al., 2017).

Single marker gene sequencing of the 16S rRNA genes along with bacterial MMO (pmoA, mmoX) and archaeal MCR (mcrA) genes have sampled methanotrophs from a wide range of environments. From these studies it is inferred that a minor fraction of methanotrophic phylogenetic diversity is represented by axenic cultures. In fact, entire clades of methanotrophs still lack a single isolated representative, including ANME and Ca. Methylophilum limosum. By comparing small subunit ribosomal protein S3 (rpS3) genes recovered in metagenomes with known reference isolate genomic sequences, we highlight that metagenomics continues to recover potentially novel species, i.e. 251 out of 641 recovered genes shared < 95% amino acid identity to methanotroph isolate genomes (see Web resources). Furthermore, this includes the identification of numerous potentially novel species, using the same criteria, even among the well-sampled phylogenetic groups such as the Gammaproteobacteria, e.g. 170 out of 539 rpS3 genes recovered from metagenomes (see Web resources). Recent metatranscriptomic studies continue to uncover and define new lineages of methanotrophs, as well as provide necessary metabolic and ecological context, thereby playing a critical role in advancing our understanding of methanotrophy from single genes, to microbial communities, to entire ecosystems.

Metagenomic approaches used to study methanotrophs in situ and in vitro

A ‘metagenome’ is the theoretical collection of genomes or genomic information from members of a given microbial community. While metagenomic methods and sequencing technologies have changed over time, the core principle remains: the untargeted sequencing of DNA from a microbial community. The earliest metagenomic methanotroph studies relied on sequencing libraries containing large inserts of DNA cloned into plasmids (Hallam et al., 2004; Meyerdierks et al., 2005; Ricke et al., 2005; Dumont et al., 2006; Chen and Murrell, 2010). Contemporary metagenomics leverage next-generation sequencing advancements, typically resulting in fragments (reads) with increased sequencing depth per sample (Escobar-Zepeda et al., 2015). However, as sequencing technology continues to improve, metagenomics show increasing promise in complete genome recovery from complex environmental samples.

Metagenomic data can be analysed using two main approaches: gene-centric or genome-centric (Fig. 3.2). In one gene-centric approach, sequencing reads are aligned (mapped) to a database of microbial genomes or reference genes to assign reads to closely related genes, pathways or genomes contained in a database. While both time- and cost-effective, this approach may miss novel genomic information, owing to insufficient representation in databases. In an alternative gene-centric approach, sequencing reads can be de novo assembled into larger genome fragments, which are then compared to nearest neighbour genomes. Here the advantage is that novel functional and phylogenetic markers can be reconstructed from an environmental sample. The limitation of both gene-centric approaches is the lack of a holistic genome content and organization, hindering the precise understanding of the metabolic pathways or phylogenetic signals essential in description of novel taxa and pathways.

In contrast, in the genome-centric approach, the assembled genomic fragments (contiguous sequences, or contigs) are clustered together to form distinct, population-level genome bins. In the past 15 years, these Metagenome-Assembled Genome bins (MAGs) have yielded near-complete or essentially closed genomes from a myriad of environmental samples (Tyson et al., 2004; Sharon and Banfield, 2013; Parks et al., 2017). This approach allows for genome wide comparisons, which provide expanded phylogenetic resolution through the use of conserved housekeeping genes, residing at various locations in the genome (Rinke et al., 2013; Hug et al., 2016; Yeoh et al., 2016; Parks et al., 2017), as well as for the extended analysis of the metabolic pathway content (Anantharaman et al., 2016), virus–host linkages via CRISPR-Cas loci, etc. (Edwards et al., 2016). A combination
of the established bioinformatics platforms and the affordable sequencing technologies now allow for prevalent use of the genome-centric approach. Given sufficient depth of sequencing, this method can recover genomic bins of the quality comparable to the quality of draft genomes from isolate cultures (Bowers et al., 2017), and thus this method represents an important avenue for discovering novel methanotrophs. Here we catalogue the methanotroph MAGs from multiple databases, to provide facile access to these genomic data, and summarize their diversity and ecosystem and high-level genomic information (see Web resources).

Given that all methanotrophs, by definition, are able to consume methane, metagenomic methods have been successfully combined with stable isotope probing (SIP), or incubation of an environmental sample with $^{13}$C-methane, followed by the analysis of $^{13}$C-DNA (Friedrich, 2006). Sequencing of the labelled DNA, referred to as functional metagenomics, followed by gene- or genome-centric analysis, directly links methane consumption to specific taxa (McDonald et al., 2005; Kalyuzhnaya et al., 2008; Chen and Murrell, 2010). Functional metagenomics, with or without SIP, have been proven to be exceptionally useful for examining slow-growing or consortia-dependent methanotroph taxa, providing a majority of genomic information to date for lineages such as ANME and Ca. Methylomirabilis that lack pure culture representatives (discussed below). The difficulties facing both methanotroph isolation and metagenomic sequencing of complex samples can be simplified or overcome by the use of enrichment-based metagenomics.

---

**Figure 3.2** Schematic overview of metagenomics approaches. There are two primary methods of metagenomic analysis: gene-centric that maps reads, genes, or contigs (represented by coloured dashes) to a reference database (boxed genes), and genome-centric that reconstructs genomes (broken circles) independent of reference databases. Functional metagenomics precedes these other methods using an initial enrichment, with or without a label, to more directly target methanotrophs, e.g. $^{13}$C-methane. The utility of these methods and example ecosystems are outlined for each approach. Acronyms are alphabetically as follows: mmo, methane monooxygenase genes; mdh, methanol dehydrogenase genes; mcr, methyl coenzyme M reductase genes; pmo, particulate methane monooxygenase genes; smo, soluble methane monooxygenase genes.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Data and Analysis</th>
<th>Purpose</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene-centric</td>
<td>Gene specific analyses of reads or assembled fragments</td>
<td>Estimate known diversity and abundance</td>
<td>Arctic permafrost, Deepwater Horizon, Rice paddies</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reconstruct known pathways</td>
<td></td>
</tr>
<tr>
<td>Genome-centric</td>
<td>Genomes of strains or closely related populations</td>
<td>Capture new phylogenetic diversity</td>
<td>Lake waters, Marine OMZ, Wetland soils</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Identify novel metabolisms</td>
<td></td>
</tr>
<tr>
<td>Functional</td>
<td>Isotopically labeled DNA (or RNA)</td>
<td>Enrich for low abundance members</td>
<td>Lake Washington, Wastewater treatment, Natural methane seeps</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Environmental Sample</th>
<th>Approach</th>
<th>Purpose</th>
<th>Examples</th>
<th>Purpose</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gene-centric</td>
<td></td>
<td></td>
<td>Genome-centric</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Functional</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Metagenomic sampling of methanotroph phylogeny and physiology across diverse habitats

We queried 741 metagenomes and metatranscriptomes associated with methane cycling from the Integrated Microbial Genomes database (IMG, 4th June 2018), provided by the Joint Genome Institute. We used ribosomal protein S3 (rpS3) genes from known methanotroph genomes to mine a total of 566 methanotroph rpS3 proteins, from 235 metagenomic samples, spanning a variety of taxa and ecosystems (see Web resources). In addition to this gene-specific search, we also recovered the rpS3 genes from 75 methanotroph MAGs (37 inferred methanotroph MAGs lacked this gene) (see Web resources). These two metagenome-derived rpS3 data sets, along with reference rpS3 proteins from the known methanotroph isolate lineages, were used to construct a phylogenetic tree visualizing the methanotroph phylogenetic diversity captured by both isolation and metagenomic sampling (Fig. 3.3). Auxiliary rings on the phylogenetic tree provide additional information for class level assignment, the type of sequencing performed, and the ecosystem from which the sequence was recovered. The dissimilarity between reference isolate genomes and MAGs’ rpS3 genes shows that many MAGs are not well represented by isolates – in only 22 out of 112 MAGs did the rpS3 gene peptides share > 95% amino acid identity with reference genomes. MAGs continue to play an important role in expanding the methanotroph foliage on the tree of life. Using our meta-analysis as a guide, below we highlight some of the prominent metagenomic studies performed across a range of different habitats, and indicate the insights gleaned about methanotrophs from these studies.

Marine and terrestrial hydrocarbon and methane seeps

Gene-centric metagenomics have provided genomic insights into the anaerobic methane-oxidizing taxa from natural hydrocarbon seeps at Coal Oil Point (California) (Håvelsrud et al., 2011), Troll petroleum reservoir (Norway) (Håvelsrud et al., 2012), Lei-Gong-Hou mud volcano (Taiwan) (Tu et al., 2017), and the Mississippi Canyon (Gulf of Mexico) (Vigneron et al., 2013). In these ecosystems, reaching up to 1100 metres in depth, the ANME types were the most abundant methanotroph taxa. Here, it was inferred that the electron acceptor for methane oxidation was ferric hydroxides or sulfate, the latter reduced through a partnership with sulfate-reducing bacteria (SRB). These results are consistent with much of our understanding of extremely diverse ANME metabolism derived primarily from naturally and artificially enriched vent samples (Knittel and Boetius, 2009), discussed in more detail in sections below. Beyond anaerobic methane oxidation, some additional evidence was provided that aerobic methanotrophs may also be important in marine methane seep systems, as reads mapping to aerobic methanotroph genomes from the Proteobacteria (Methylloccoccales and Rhizobiales), Verrucomicrobia (Methylacidiphilales), and Methylomirabilota (Ca. Methylomirabilis) were also present in many of the metagenomes (Håvelsrud et al., 2011, 2012; Tu et al., 2017). Additionally, gene-centric metagenomics performed on a hydrocarbon seep in the Red Sea reconstructed functional genes for methane oxidation not detected in clone libraries, and was one of the first studies showing that sMMO genes were more abundant than pMMO genes in a natural ecosystem (Abdallah et al., 2014).

In addition to the naturally occurring methane seeps, anthropogenic sources of hydrocarbons also occur in marine and terrestrial systems, and represent another focus for the methanotrophy research. Following the Deepwater Horizon spill in 2010, several metagenomic studies revealed that aerobic Gammaproteobacteria methanotrophs of the order Methylloccoccales must have consumed nearly all the methane in the water column, mitigating methane release to the atmosphere (Mason et al., 2014; King et al., 2015; Yergeau et al., 2015). The methane consumption activity also caused an oxygen anomaly as well as depleted certain metals in the water column (Kessler et al., 2011; King et al., 2015; Shiller et al., 2017). Taken together, these studies demonstrate both the positive and potentially the negative impacts methanotroph activity may have on an ecosystem scale.

An additional hydrocarbon source is oil sand tailing ponds, which are small natural freshwater bodies contaminated with hydrocarbon waste after industrial petroleum extraction from bituminous sands. SIP-enabled gene-centric metagenomics of Canadian tailings pond sediments recovered
pMMO genes most similar to the Gammaproteobacteria Methylococcales member of the genus *Methylocaldum*, inferred to cross-feed with non-methanotrophic methylotrophic taxa via methanol (Saidi-Mehrabad *et al.*, 2013). Methanol-mediated communal metabolism of methane has been demonstrated in other ecosystems (Beck *et al.*, 2013; Krause *et al.*, 2017), suggesting that, while methanotroph and methylotroph genera may differ between ecosystems, the metabolic exchanges may be conserved. Although generally in low quantities in the most marine and terrestrial environments, naturally and artificially hydrocarbon- or methane-enriched environments support diverse communities of methanotrophs that mitigate methane release into the atmosphere.

**Atmospheric methane sinks in dry soils**

Forest soils have long been of interest for the study of methanotrophs, due to their ability to act as net methane sinks by consuming trace (<2 parts per million) quantities of methane present in the atmosphere. Several studies have inferred that
this phenomenon, named high-affinity methane oxidation (HAMO), likely occurred through the activity of uncultivated Alphaproteobacteria methanotrophs (Dunfield et al., 1999; Holmes et al., 1999; Henckel et al., 2000). One of the early SIP-enabled metagenomics projects generated sequences of Bacterial Artificial Chromosomes (BACs) from forest soil, recovering fragments of genomes most closely related to members of the Alphaproteobacteria genus Methylocystis (Dumont et al., 2006). This demonstrated the application of this technique for recovering the genomic contents of methanotrophs that are generally present at low abundance in complex soil communities. In a separate study, sequencing BACs from forest soils provided the first genomic information for the Alphaproteobacterial Upland Soil Cluster (USCa) methanotrophs, a prevalent lineage in forest soils implicated in oxidizing methane at atmospheric pressure levels (Ricke et al., 2005). The gene-centric approach revealed that, based on the pMMO and neighbouring genes, the proposed USCa methanotroph was most similar to the known Alphaproteobacteria methanotrophs (Ricke et al., 2005).

Recently, Pratscher et al. (2018) provided the first morphological information for the USCa, and recovered the first draft genome of USCa methanotroph, by combining targeted cell sorting and metagenomic sequencing. The resulting MAG lacked sMMO genes, but contained several pMMO gene variants (Pratscher et al., 2011, 2018). It was proposed that these different pMMO variants may function in a manner similar to variants previously identified in Methylocystis isolates, showing different substrate affinities towards methane (Baani and Liesack, 2008; Dam et al., 2013). Additionally, the USCa genome encoded pathways for acetate utilization, signifying the capability to utilize alternative carbon substrates as a means of adaptation for HAMO (Pratscher et al., 2011, 2018). Mining of 16S rRNA gene-databases posited that USCa species are not restricted to forest soils, but are also present in lava cave biofilms and arctic permafrost (Pratscher et al., 2018). Similar to these findings, Singleton et al. (2018) recovered a partial USCa methanotroph genome from partially thawed permafrost soils in Sweden. The much wider distribution of USCa methanotrophs, including non-arid environments such as permafrosts, suggests the importance of HAMO on a global scale.

In addition to the USCa methanotrophs, the first nearly complete genome of the divergent Gammaproteobacterial Upland Soil Cluster (USCy) methanotroph has recently been recovered from cold, arid soils. This genome was shown to be significantly divergent from the genomes of known Methylococcales (Edwards et al., 2017) (Fig. 3.3; see Web resources), while encoding typical methylation metabolic pathways. Since their initial discovery in forest soils, the recovery of USCy from Taylor Dry Valley in Antarctica demonstrates their distribution beyond forested environments (Edwards et al., 2017). Importantly, for both USCa and USCy lineages, these MAGs alone cannot explain the specific metabolic requirements or the enzymatic differences that enable HAMO. Given the broad environmental distribution of these organisms (northern latitude forests, Antarctic deserts, and permafrost ecosystems), there is a need for cultivation and physiological studies targeting these lineages.

Freshwater lake sediments
One of the best-characterized ecosystems for microbial C1 cycling is freshwater lake sediments. These systems have anoxic conditions in the deeper sediments, resulting in methanogenesis, with methane diffusing to oxic sediment–water interface, where aerobic methanotrophy occurs. One of the earliest uses of SIP in combination with metagenomics investigated the microbial taxa involved in methylothrophy in Lake Washington sediments (Kalyuzhnaya et al., 2008). Here sediments incubated with 13C-methane largely recapitulated the data from the cultivation efforts, showing that Methylobacter, of the Gammaproteobacteria Methylococcales clade, were the dominant methane-oxidizing taxa (Auman et al., 2000). Additionally, these methane-fed communities were enriched for Betaproteobacteria species Methylothera, which were likely labelled through consuming methanol, a by-product of methane oxidation. This metabolic cross-feeding spurred a series of experiments showing that at lower oxygen: methane mixing ratios, Methylobacter, consistently outcompeted other methanotrophs, and also produced methanol to cross-feed Methylothera (Beck et al., 2011; Oshkin et al., 2015; Krause et al., 2017).
Alternatively, under high oxygen:methane mixing ratios, members of the same family but different genus (*Methylosarcina*) cross-fed different Betaproteobacteria microorganisms (*Methylophilus*) (Hernandez et al., 2015). These findings highlight the role of cooperative metabolic exchanges fuelled by methanotrophy occurring across ecosystems, and the value of both laboratory and field based studies.

**Thawing arctic permafrost**

There is growing interest in the metabolic activity of methanotrophs in arctic peat and permafrost soils. These environments store approximately 1300–1600 petagrams of carbon that is susceptible to release under predicted climate warming scenarios (Dean et al., 2015). Warming temperature trends threaten to thaw permafrost, making the previously inaccessible carbon newly available. The newly accessible organic carbon can be degraded to yield methane, which may be consumed by the methanotrophs or be emitted into the atmosphere, the latter with future climatic feedbacks (MacKelprang et al., 2011).

Several studies by Tveit et al. (2013, 2014) clearly articulated the importance of aerobic Gammaproteobacteria Methylococcales in warming permafrost. In the first two studies, metatranscriptomic data were collected from permafrost soils in Svalbard, Norway. Here 14 nearly full-length assembled 16S rRNA genes (>1330 nucleotides) in the metatranscriptome were assigned to the genus *Methylobacter*. Metatranscriptomic reads were mapped to pMMO, methanol dehydrogenase, and formaldehyde oxidation pathways from reference genomes, further demonstrating that members of *Methylobacter* were primarily responsible for methane oxidation in these permafrost soils. Notably, genes from other methanotrophs were absent, or present at negligible counts (Tveit et al., 2013, 2014), demonstrating the metabolic prevalence of this one genus in these soils. Laboratory mesocosm incubations designed to examine the effects of thaw gradient (temperatures ranging from 1 to 30°C) on these soils also demonstrated that *Methylobacter* remained the dominant methanotroph (Tveit et al., 2015). Taking these studies together, it is possible that in these permafrost soils, active methanotrophy today and under future warming scenarios may be mediated by a single genus of methanotrophic bacteria (Tveit et al., 2015).

Other studies have implicated Alphaproteobacteria as key methanotrophs to respond to permafrost thaw. Mackelprang et al. (2011) used changes in the relative abundance of 16S rRNA and pMMO genes during *in vitro* thaw simulation experiment, to infer members of the genus *Methylocystis* as responsive species. A large-scale study at Stordalen Mire in Sweden found both classes of Proteobacteria methanotrophs to be present, but these were differentially distributed across the thaw gradient (Singleton et al., 2018). Metagenomes from partially thawed and intact permafrost contained mostly reads mapping to Alphaproteobacteria methanotrophs, particularly *Methylocystis* and USCa MAGs, while fully thawed permafrost metagenomes harboured reads that mapped mostly to Gammaproteobacteria Methylococcales MAGs. This trend was matched by the distribution and diversity of pMMO and sMMO genes identified in the metagenomes, as well as metatranscriptomic reads mapping to these genes (Singleton et al., 2018).

Two other studies in the Canadian High Arctic also used metagenomics technologies to investigate methanotroph diversity and activity in permafrost soils. One gene-centric metagenomics study concluded that Gammaproteobacterial Methylococcales were the dominant methanotrophs in the active layer and up to 2 m depth into the arctic permafrost (Yergeau et al., 2010). In contrast, a paired metagenomic, metatranscriptomic, and metaproteomic investigation of another permafrost site noted that Alphaproteobacterial USCa methanotrophs recruited most reads and peptide fragments in all datasets (Lau et al., 2015). It was inferred that the relative aridity and atmospheric methane consumption in this site, compared to other permafrost field locations, may have contributed to the prevalence of USCa. Similarly, Siberian permafrosts were also dominated by Alphaproteobacteria members of the genus *Methylocystis*, but Gammaproteobacteria methanotrophs could also be detected (Rivkina et al., 2016).

In summary, while the active methane-oxidizing communities across several different permafrost habitats were dominated by Proteobacteria methanotrophs, it appeared that the thaw stage was a major controlling factor for the distribution and activity
of different clades of methanotrophs in these permafrost ecosystems. Therefore, understanding the physiological constraints, e.g. temperature, hydrology, methane oxidation rates, etc. of Proteobacteria methanotrophs will be critically important for improving predictions of methane release from Northern thawing permafrost ecosystems. Understanding these dynamics is important for predicting the efficacy of methanotrophs in reducing methane release from ecosystems with large amounts of stored carbon.

**Temperate hydric soils and sediments (peats or wetlands)**

Temperate wetlands can be defined as flooded or seasonally flooded soils located in regions with relatively mild climates, and these include freshwater, coastal, and peatland soils. Despite their small land coverage, on a global scale, these systems account for a disproportionate fraction of the total methane budget (Nazaries et al., 2013; Dean et al., 2018). Below we summarize some of the recent metagenomic investigations of these habitats.

In a boreal peat bog in Minnesota, SIP-enabled metagenomes revealed enrichment for Alphaproteobacteria *Methylcystis* and Gammaproteobacterial *Methylomonas* pMMO genes, at a ratio of about 3:1 (Esson et al., 2016). At increasing depths, the *Methylomonas* signal was eventually below detection, consistent with qPCR and microarray data from the same study site (Esson et al., 2016). Other studies also found that pMMO genes affiliated with *Methylcystis* and other Alphaproteobacteria methanotrophs tended to be more abundant, possibly owing to the acidic pH or to copper limitation (Chen et al., 2008; Kolb and Horn, 2012; Lin et al., 2014; Graham et al., 2017). Corroborating decades of interest in peatland systems, gene-centric analyses have reaffirmed the importance of Alphaproteobacteria methanotrophs in many of these acidic environments as natural filters of methane release.

In a restored peat ecosystem in California, metagenomes from samples collected up to 25 cm into the soil column revealed methanotroph relative abundances ranging from ≈1% to 3% of total microbial community (S. He et al., 2015). Both the number of pMMO genes and the reads mapped to components of methane oxidation pathways were enriched in plant-associated samples, dominated primarily by Alphaproteobacteria *Methylcystis* and *Methylosinus*. Data collected from grasslands, forests, mangroves, and systems replete in plants corroborated the finding that *Methylcystis* types were the most prevalent methanotrophs in soils within the rhizosphere (Xu et al., 2014). Therefore, further understanding of the physiology and metabolism of *Methylcystis* is important for improving models for methane cycling in vegetated environments.

The Great Lakes region on the border between the United States and Canada stores large amounts of carbon (Nahlik and Fennessy, 2016), and wetlands adjacent to Lake Erie are known to be net methane emitters, with seasonal variations in methane output (Angle et al., 2017; Rey-Sanchez et al., 2018). Gene-centric approaches, 16S rRNA gene abundances and pMMO gene transcript inventories, have shown that *Methylobacter* species dominated the methanotroph community (Smith et al., 2018). Furthermore, pMMO transcription was season-dependent, and found to be increased in autumn relative to summer. In contrast, methanogen gene expression was relatively stable over the sampling period, suggesting that the greater methane emissions observed in the summer months might be due to the reduced methanotroph activity and not to increased methanogenesis activity (Smith et al., 2018). A biogeographical survey conducted by Smith et al. (Smith et al., 2018) detected pMMO genes similar to the those of *Methylobacter* lineages found in many methane-emitting Northern latitude environments that include Lake Washington (Chistoserdova, 2011a), the restored Twitchell Island peatlands (He, S. et al., 2015), and the Prairie Potholes region of North Dakota (Dalcin Martins et al., 2017). Despite major differences in the latitude, the prevalence of *Methylobacter* genotypes and their genomic conservation across geographically distinct wetland and permafrost habitats suggests that further studies of the representatives of this genus are relevant for global climate change research (Tveit et al., 2013).

**Rice paddies and agricultural soils**

Rice paddies represent inundated agricultural soils that contribute 4–20% to the atmospheric methane budget (Nazaries et al., 2013; Dean et al., 2018), despite most of the methane generated (up to 90%) being consumed by the methanotrophs (Bao et al.,
methanotrophs were detected in the anoxic zone at surface layers, and no transcripts affiliated with was restricted to the very narrow 2 mm thick oxic surface layers, and no transcripts affiliated with methanotrophs were detected in the anoxic zone at 6–8 cm (Kim and Liesack, 2015).

Seasonal dynamics likely play a role in constraining methanotroph abundance and activity in rice paddies. During the drought season, rice paddies typically serve as net sinks of methane (Harriss et al., 1982; Kelley et al., 1995; Melling et al., 2007; Kolb and Horn, 2012). Incubation of rice paddy soils at high versus near-atmospheric methane mixing ratios found a strong enrichment for Gammaproteobacteria genus Methylosarcina and mild enrichment of Alphaproteobacteria Methylocystis (Cai et al., 2016). From a combination of metatranscriptomics and SIP experiments, it was thought that Methylosarcina responded to the initial spikes of methane (≈9000 ppm methane), consuming methane to produce metabolites that are necessary to support HAMO (≈2 ppm methane) by the Alphaproteobacteria methanotrophs. Interestingly, this study noted that neither of the USC clades were detected in the metagenomes or metatranscriptomes, therefore HAMO observed in these samples was dependent on the known methanotrophs and not on the expected USC-type methanotrophs.

Subsurface ecosystems sustained by methane
Microbial life in caves and subsurface ecosystems, which typically lack organic carbon input by plant primary productivity, is dependent on chemolithotrophic metabolisms. Methane is a relatively rich source of energy and carbon in a variety of geographically unrelated subsurface ecosystems (Hutchens et al., 2004; Brankovits et al., 2017; Karwautz et al., 2018). A model cave system in Romania, Movile Cave, undergoes seasonal extremes of methane concentrations due to consumption by up to four distinct bacterial clades of methanotrophs – USCα, USCγ, and two potentially novel lineages (Waring et al., 2017). Metagenomic analysis revealed that aerobic methanotrophic marker genes were present in both the oxic mat and the anoxic sediment, and also showed that the predominant MMO types varied between niches (Kumaresan et al., 2018). In another subsurface environment, an artificial mine in a granitic system in Japan, methane oxidation appeared to entirely rely on the ANME type Archaea, and a nearly complete genome of an ANME-2d organism has been reconstructed that differed from the previously reconstructed ANME-2d genomes by lacking any genes for

2014; Lee et al., 2015). Given the large error in current predictions for methane emissions from these habitats, there is a desire to better understand how agricultural land management impacts the activity of methane-cycling microorganisms. In particular, rice paddies are characterized by drastic oxyclines that occur along the water depth profile and along the distance from the plant roots, conducive of both aerobic and anaerobic methanotrophy (Reim et al., 2012; Lee et al., 2015). A study on Indian rice paddies with different fertilizers by Bhattacharyya et al. taxonomically classified reads, finding that Gammaproteobacteria Methylococcales, particularly Methylobacter and Methylococcus, were the most abundant methanotroph taxa, regardless of fertilizer treatment (Philippot et al., 2010; Bhattacharyya et al., 2017). Alphaproteobacteria and Verrucomicrobia methanotrophs were also detected, but were less abundant. The reconstruction of the formaldehyde assimilation pathways showed predominance of the serine cycle genes, which are present in both Alphaproteobacteria and Gammaproteobacteria methanotrophs (Ward et al., 2004; Takeuchi et al., 2014).

Profiling active methane metabolism along rice paddy soil compartments, including water, soil, phyllo-, rhizo-, and endo-spheres, is a current research area. One study that applied metagenomics in combination with metaproteomics to both the phyllosphere and the rhizosphere has found that proteins for both methane oxidation and nitrogen fixation are represented methanotrophic Alphaproteobacteria and found that rice root metaproteomes mainly recover any peptides mapping to Methylococcales and both sMMO and pMMO were expressed at approximately equal peptide spectra (Knief et al., 2012). In contrast, Bao et al. (2014) failed to recover any peptides mapping to Methylococcales and found that rice root metaproteomes mainly represented methanotrophic Alphaproteobacteria of the Methylocystaceae family, which carry out both methanotrophy and dinitrogen fixation in the vascular bundles and epidermal cells of rice roots. A separate metatranscriptomic analysis of rice paddy laboratory incubations found Methylococcales to be the predominant methane oxidizer, but was restricted to the very narrow 2 mm thick oxic surface layers, and no transcripts affiliated with methanotrophs were detected in the anoxic zone at 6–8 cm (Kim and Liesack, 2015).
denitrification or metal reduction (Ino et al., 2018). Instead, the genome encoded functions for a putative respiratory H₂ oxidation as well as a possible pathway for utilizing zero-valent sulfur to exchange electrons with syntrophic partners (Milucka et al., 2012; Ino et al., 2018). Overall, caves and other subsurface ecosystems are relatively understudied environments, but the few examples in the literature suggest that these ecosystems are home to interesting methanotrophs, and that methane oxidation occurs differentially along ecological niches.

Metagenome-derived genomes reveal methanotroph metabolic variability

In this review, we compiled the genomic information for all 80 isolates and 112 MAGs available via the IMG or the National Center for Biotechnology Information (NCBI), assigned to methanotroph taxa or taxa described as a putative methanotrophs in publications (Fig. 3.4; see Web resources). Reconstruction of MAGs provides genomic information for major clades entirely lacking a single pure culture representative, e.g. ANME, USCα, USCγ or Ca. Methylomirabilis. These MAGs provide essential information about these microorganisms that are phylogenetically and physiologically divergent from the known, isolated methanotrophs. Genome-centric metagenomics approaches have more than doubled methanotroph genomic database sampling in as little time as a decade. Below, we summarize the major discoveries that analyses of MAGs have added to the phylogenetic and metabolic understanding of methanotrophs in natural as well as artificial settings.

MAGs represent the only genomic information from some uncultivated methanotroph lineages

To date there is not a single ANME representative in axenic culture. Genomic information for ANME-1 was first obtained in 2005, linking methanogenesis and 16S rRNA genes on the same contig (Meyerdierks et al., 2005). This finding was followed by the first MAG described for ANME-1 lineage, recovered in 2010 by Meyerdierks et al. (2010) by fosmid sequencing of microbial mats in the Black Sea. From this MAG, the authors reconstructed the first near-complete pathway for Archaeal methane oxidation, inferring the methanogenesis pathway operating in reverse (Hallam et al., 2003; Meyerdierks et al., 2010). Additionally, they recovered the first [Fe–Fe] hydrogenases in Archaea, as well as the multiple multi-haem c type Cytochromes (MHC) that were proposed to act as the electron shuttling intermediates involved in interactions with the sulfate-reducing bacteria (SRB) or in metal reduction (Meyerdierks et al., 2010). These genomes were critical, providing the first evidence for the types of anaerobic energy generation mechanisms encoded by the archaeal methanotrophs. Shortly after, combined metagenomic and metaproteomic/metatranscriptomic studies corroborated the genomic potential postulated from the first ANME-1 genome (Stokke et al., 2012; Krukenberg et al., 2018).

Shortly after the recovery of the first ANME-1 genome, a MAG representative of the ANME-2d lineage was reconstructed from a methane consuming and denitrifying bioreactor inoculated with a mixture of waste sludge and freshwater lake sediments (Haroon et al., 2013). This near-complete genome is referred to as Candidatus Methanoperedens nitroreducens, named after the inferred metabolic capacity to couple methane oxidation, via reverse methanogenesis, to dissimilatory nitrate reduction (Haroon et al., 2013). In contrast to ANME-1, the Ca. Methanoperedens genome contained a full reverse methanogenesis pathway. An additional unique feature of this genome was the presence of only the first gene for denitrification, nar, which would only allow Ca. Methanoperedens nitroreducens to reduce nitrate to nitrite (Haroon et al., 2013). Metatranscriptomic coupled to 15N-nitrate SIP analysis applied to the bioreactor microbial community suggested that the nitrite generated was likely consumed by the nitrite-utilizing microorganisms, including Ca. Methylomirabilis, or by anaerobic ammonia-oxidizing microorganisms (Raghoebarsing et al., 2006; Haroon et al., 2013). Subsequent ANME-2 MAGs from enriched freshwater canal sediments in the Netherlands found that other Ca. Methanoperedens nitroreducens MAGs encoded the capacity to perform dissimilatory nitrate reduction to Ammonia (DNRA, nrfH) (Arshad et al., 2015). Taken together, these studies are important for highlighting the coupling of anaerobic methane oxidation to nitrate reduction.
Since their initial discovery, approximately 20 genomes have been reconstructed for the ANME lineages, including 4 ANME-1 and 15 ANME-2 representatives (Fig. 3.4, see Web resources). MAGs of both ANME-1 and ANME-2 lineages contain a surprising number of MHC genes, suggesting that both groups may be capable of coupling methane oxidation to the reduction of metals (Meyerdierks et al., 2010; Wang et al., 2014; Cai et al., 2018). Furthermore, these analyses also suggested that MHC was the prevailing mode of interaction between the ANME and the SRB among both the thermophilic and the psychrotolerant communities of the, respectively, Guaymas Basin and the Crimean Peninsula, and the mid-Norwegian margin (Meyerdierks et al., 2010; Stokke et al., 2012; Krukenberg et al., 2018), suggesting that this function is important across temperature and geographic ranges. Genomic inferences suggested that ANME-2 have the potential for nitrogen fixation, a finding confirmed by the cellular incorporation of labelled dinitrogen gas (Pernthaler et al., 2008), and in contrast to the other ANME-2 MAGs reconstructed from subsurface ecosystems (Ino et al., 2018). Taken together, metagenomics, often in combination with other techniques, have provided
a wealth of information on the metabolic diversity maintained among ANME lineages.

These reconstructed ANME MAGs have also enabled improved detection of the ANME-type methanotrophs from a wide range of ecosystems. For instance, sequences have been recruited to ANME genomes from deep marine ecosystems and wastewater treatment reactors (Raghoebarsing et al., 2006; Hävelsrud et al., 2011, 2012; Lee et al., 2016; Welte et al., 2016; Cai et al., 2018), but also from rice paddies, wetlands, and peatlands soils, as well as from subsurface locations (Takeuchi et al., 2011; Vaksmaa et al., 2017; Ino et al., 2018). The impact of ANME in ecosystems exposed to dynamic oxygen fluctuations is poorly understood relative to marine and wastewater treatment environments, but more genomic representatives in databases may result in promising new areas for future research.

In addition to ANME, methanotrophic representatives of the Methylomirabilota, formerly of the candidate NC10 phylum, the Ca. Methylomirabilis genus also lack any pure culture isolates. However, genomic insights into the metabolism of this group are enabled by at least eight MAGs. These genomes were reconstructed from denitrifying wastewater treatment reactors (Raghoebarsing et al., 2006). These anoxic reactor fluids, when supplemented with nitrate and methane, became enriched with two anaerobic methanotroph populations: ANME-2, described above, and Ca. Methylomirabilis (Ettwig et al., 2010). The exact mechanism of nitrogen transformation in this system was largely unknown until the recovery of MAGs from the bioreactor. Analysis of the Ca. Methylomirabilis MAGs identified a paradoxical lack of the nitrous oxide reductase (nos) genes that were expected to be responsible for the stoichiometric conversions of nitrite to dinitrogen gas, observed in previous studies (Ettwig et al., 2010, 2012). To explain this phenomenon, Ettwig et al. (2010) hypothesized that an unknown enzyme, a nitric oxide dismutase (NOD), was responsible for disproportionation of two nitric oxide molecules into one molecular nitrogen and one molecular oxygen. While the biochemistry of this hypothesized enzyme remains to be fully elucidated (Reimann et al., 2016), intracellularly generated oxygen could explain the production dioxygen in cultures with no other exogenous sources, and enable the traditional aerobic methane oxidation using oxygen-dependent pMMO in anoxic conditions (Ettwig et al., 2010).

**MAGs demonstrate the breadth of phylogenetic diversity within the Proteobacteria methanotrophs**

Beyond the lineages comprised exclusively of uncultivated methanotrophs, MAGs have also sampled lineages represented by or closely related to the cultivated methanotrophs (Fig. 3.4). Multiple MAGs have been recovered related to the members of *Methyllobacter*, *Methylomonas*, and *Methylocystis* (Fig. 3.4; see Web resources). Additionally, MAGs reconstructed from marine oxygen minimum zone and marine endosymbiont samples have sampled novel lineages within the Gammaproteobacteria. In this section, we highlight the metabolic information derived MAG representatives of uncultivated Proteobacteria methanotroph lineages.

In marine systems, endosymbiotic methanotrophs have been identified, forming thus far a monophyletic lineage within Gammaproteobacteria *Methylococcales* (Petersen and Dubilier, 2009). One of the most well-known marine animals to host the methanotrophs in its gill tissues is *Bathymodiolus*, a mussel found at the bottom of the ocean (Fisher et al., 1987; Nix et al., 1995). Because endosymbiotic bacteria are difficult to isolate, knowledge on their metabolism relies on these culture-independent methods. Phylogenetic analyses using 16S rRNA and pMMO genes and ultrastructure affiliated the endosymbiotic methanotrophs with the *Methylococcales* genus *Methyloprofundus* (Spiridonova et al., 2006; Tavormina et al., 2015; Sun et al., 2017).

Two recent publications independently reconstructed MAGs from distinct species of *Bathymodiolus* (Ponnudurai et al., 2017; Takishita et al., 2017). The rpS3 genes recovered from these genomes were most closely related to each other and to *Methyloprofundus* (Fig. 3.3; see Web resources). These MAGs were not only enlightening for resolving the identity of the endosymbionts, but also for their function(s) within the host. Ponnudurai et al. (2017) concluded that the methanotroph supplied carbon dioxide for assimilation by other non-methanotrophic endosymbionts, as well as synthesizing amino acids and necessary cofactors for the host.
Takahita et al. (2017) found that the methanotrophic endosymbiont synthesized unique cholesterol intermediates that the mussel could not produce on its own. Further investigations into these relationships using metagenomics will likely uncover additional critical functions these endosymbiotic methanotrophs perform for the host, as well as further insights into the evolution of methanotroph-animal endosymbioses.

Recovery of MAGs from low-oxygen marine ecosystems has enabled the identification of multiple uncultivated, novel Gammaproteobacteria clades. From deep-sea marine vents at the Lau Basin, Methylothermaceae B42 is the only genomic representative of a likely novel genus in the Gammaproteobacteria Methylococcales (Skennerton et al., 2015). This genome shares only 94% 16S rRNA sequence identity with its closest isolated relative in the genus Methylohalobius. Similarly, the first genome from the uncultivated Operational PmoA Unit 3 (OPU3) lineage was recovered from Oxygen Minimum Zone (OMZ) of Gulfo Dulce, sharing <93% 16S rRNA sequence identity with any Methylococcales isolates (Hayashi et al., 2007; Padilla et al., 2017). OPU3 appear to be cosmopolitan in marine systems and are enriched in some OMZ ecosystems and during the Deepwater Horizon oil spill (Tavormina et al., 2010; Lesniewski et al., 2012; Li et al., 2014). Both Methylothermaceae B42 and OPU3 MAGs were sampled from oxygen-limited environments, and other 16S rRNA gene surveys indicate that the OPU3 are found in deep waters (100–2,000 m depth), suggesting tolerance to low dissolved oxygen concentrations (Tavormina et al., 2010), as discussed below. These two studies have both recovered genomic information for uncultivated lineages in marine systems, which likely signify the beginning of future discoveries of novel methanotrophic Gammaproteobacteria taxa in marine environments through metagenomics.

Although most information regarding the distribution and ecology of USC methanotrophs comes from forest soils, MAGs representing both USCα and USCγ have been reconstructed from polar latitudes (Edwards et al., 2017; Pratscher et al., 2018; Singleton et al., 2018). There is increasingly strong evidence that the USCα are a divergent lineage closely related to Alphaproteobacteria genus Methylocapsa (Ricke et al., 2005; Pratscher et al., 2018; Singleton et al., 2018), but evolutionary history of the USCγ is less clear (Edwards et al., 2017). While the pMMO marker genes are similar to Methylococcales genus Methylocaldum, the 16S rRNA gene recovered in the MAG was most similar to microorganisms within the Gammaproteobacteria class Chromatiales (Edwards et al., 2017), similar to its placement upon the fringe of the Methylococcales clade in the rpS3 gene phylogeny (Fig. 3.3). To our knowledge, there are no other close isolated relatives to the USCγ MAG, therefore more cultivation-based sampling is needed to make inferences about metabolic or phylogenetic relationships for this genome (Edwards et al., 2017).

Interestingly, the recovery of USCα MAGs in thawing permafrost by Singleton et al. (2018) uncovered more phylogenetic novelty within the Alphaproteobacteria. This includes the presence of pMMO and sMMO genes in a novel genus of the non-methanotrophic but methylotrophic Alphaproteobacteria family Hyphomicrobiaceae (Singleton et al., 2018). This is not the first occurrence of MMO detected in non-methanotrophic methylotrophic clade; a species of non-methanotrophic isolate genus Methyllocceanibacter was isolated that encoded pMMO genes and could utilize methane (Vekeman et al., 2016a). These genome-centric studies in polar regions have not only improved the phylogenetic resolution of the USC methanotrophs, but have additionally posted a novel taxa capable of oxidizing methane.

Despite being studied for decades, novel lineages of methanotroph within the Proteobacteria are constantly being discovered. This is true of ecosystems ranging from the bottom and deep waters of the ocean to terrestrial ecosystems across the globe. Interestingly, much of the new diversity is found within the existing methanotroph families, e.g. Beijerinckaceae, Methylocystaceae, and Methylococccaceae. Further analyses of these genomes may provide information on the strain-level variations enabling adaptation to different environmental conditions across the globe. It is apparent that natural ecosystems host many novel methanotrophs, and the employment of metagenomics will continue to resolve the vast diversity of methanotrophs by reconstructing MAG representatives of uncultivated lineages.
Methyllococcales MAGs widely encode the denitrification potential, potentially enabling methane oxidation under low oxygen conditions

Using pure cultures, Stein and colleagues provided evidence that Gammaproteobacteria methanotrophs may conserve oxygen for methane oxidation and use nitrate, rather than oxygen, to generate a proton motive force for energy generation (Kits et al., 2015). This metabolic potential has been inferred from multiple, diverse MAGs, located across a range of ecosystems. It is thought that this capacity may enable methane oxidation in environments with low or below detectable oxygen concentrations.

In marine oxygen minimum zones, both the Methylothermaceae B42 and OPU3 MAGs encode genes for denitrification. This includes the capacity to reduce nitrate to nitrous oxide (\textit{nar}, \textit{nap}, \textit{ nir}, \textit{ nor}). Both Methylothermaceae B42 and OPU3 MAGs contained \textit{nar} genes that are divergent from the genes in other Methyllococcales methanotrophs, suggesting a recent horizontal transfer from non-methanotrophs (Skennerton et al., 2015; Padilla et al., 2017). It was theorized that denitrification can support methane oxidation in marine waters where dissolved oxygen is commonly below detection limits. Interestingly, transcripts for one of the \textit{ nar} genes, encoded by the OPU3 MAG, were detected in anoxic marine waters (Padilla et al., 2017). This is so far the only environmental evidence that these denitrification pathways, encoded by aerobic Gammaproteobacteria methanotrophs, are transcribed or expressed \textit{ in situ} (Padilla et al., 2017).

Beyond marine systems, the genomic potential for denitrification has been widely reported in methanotroph MAGs from terrestrial environments. For example, Dumont et al. (2011) observed expression of \textit{Methylobacter}-like denitrification genes in oxic incubations of lake sediments. Additionally, the first genome representative of the mysterious \textit{Crenothrix} clade encoded the metabolic capacity for denitrification, possibly explaining the methane consumption activity in anoxic incubations of freshwater lake waters (Oswald et al., 2017). Lastly, \textit{Methylobacter} MAGs from freshwater wetland soils also possessed the potential for denitrification, but the transcripts for this pathway was absent in metatranscriptomes from these soils (Smith et al., 2018).

The denitrification metabolic potential of Methyllococcales isolates and MAGs has been recently inventoried several times (Padilla et al., 2017; Smith et al., 2018). The most recent analysis by Smith et al. (2018) recovered at least partial denitrification (\textit{ nar}, \textit{ nap}, or \textit{ nir}) encoded in over 45 out of 57 combined MAG and isolate genomes. Moreover, 77% of the 18 reconstructed MAGs contained the capacity for reducing nitrate (Smith et al., 2018). Notably, both analyses failed to identify genes for the final step in denitrification, the reduction of nitrous oxide to dinitrogen gas by nitrous oxide reductase (\textit{ nos}), in any Methyllococcales genomes (Padilla et al., 2017; Smith et al., 2018). Therefore, it remains unclear whether the greenhouse gas mitigation benefits from methane consumption activity across wider oxygen gradients or whether it might be offset by the production of a more potent greenhouse gas nitrous oxide. The prevalence and conservation of this metabolism among Methyllococcales suggests it is likely a valuable metabolic strategy, however more environmental context and expression data are required to truly understand the impacts of methanotroph-mediated denitrification.

Beyond denitrification, methanotrophs may also be able to tolerate much lower oxygen concentrations than previously thought. Genome analysis of MAGs and isolate genomes has identified the presence of high-affinity cytochromes, e.g. cytochrome \textit{bd} ubiquinol–oxidoreductase, in addition to the traditional low-affinity cytochrome \textit{ c} oxidase. The use of these high-affinity oxidases would enable methane oxidation coupled to aerobic respiration in nanomolar range of oxygen (Skennerton et al., 2015; Oswald et al., 2017; Smith et al., 2018). Thus, it is possible that methane oxidation by aerobic Proteobacteria may be occurring in environments with lower dissolved oxygen than currently assumed in global models (Riley et al., 2011; Xu et al., 2016).

Another possible adaptation for maintaining methane consumption in low oxygen conditions is ‘micro-aerobic fermentation’, which has been reported in pure cultures of \textit{Methylomicrobium buryatense} (Kalyuzhnaya et al., 2013; Gilman et al., 2015). Oswald et al. (2017) and Smith et al. (2018) both reconstructed pathways for fermentation in, respectively, \textit{Crenothrix} and \textit{Methylobacter} MAGs. These fermentation pathways would enable the
re-oxidation of NAD(P), and would ultimately lead to the secretion of mixed fermentation products including acetate, lactate, succinate, ethanol, and/or hydrogen. However, discerning the likelihood of this fermentative metabolism from metagenomic or metatranscriptomic data alone is difficult, and therefore future physiological investigations in these and other lineages are warranted to better understand micro-aerobic fermentation by methanotrophs.

A final possible mechanism for consuming methane when oxygen is relatively depleted is by scavenging oxygen via a bacteriohemerythrin. Hemerythrins are able to bind molecular oxygen, and they have been shown to interact with the pMMO enzyme, suggesting a role in directly providing oxygen to the methane oxidation machinery (Karlsen et al., 2005; Kao et al., 2008; Chen et al., 2012). Bacteriohemerythrin genes have been identified in 80% of the methanotroph genomes, including at least four Methyllococcales MAGs, 18 Methyllococcales, six Alphaproteobacteria, and one Verrucomicrobia methanotroph genomes (Rahalkar and Bahulikar, 2018). Their wide distribution strongly suggests that they provide an important and conserved mechanism for sustaining methane consumption during oxygen limitation across the aerobic methanotrophic phyla. In fact, up-regulated transcription of this gene has been reported under oxygen-limited conditions in Methylocystis (Gilman et al., 2017). From these studies, it is clear that MAGs play an important role in elucidating the repertoire of mechanisms supporting methane oxidation under a variety of redox conditions, across a range of environments.

**Methanotroph MAGs and isolate genomes indicate the importance of novel metal biochemistry**

Methanol DeHydrogenases (MDHs) catalyse the oxidation of methanol to formaldehyde, an essential step in aerobic methane oxidation. There are two known MDH enzymes, the canonical and long-studied MxaF type and the novel type, XoxF (Chistoserdova et al., 2009). Despite similar functions, these two different MDH require different metal cofactors: the MxaFI complex requiring calcium, while the XoxF enzyme requiring a light lanthanide series metal (known as Rare Earth Elements, REEs) (Keltjens et al., 2014; Pol et al., 2014; Chistoserdova, 2015, 2016). Low concentrations of REEs have been reported to strongly repress expression of the mxa genes (Chu and Lidstrom, 2016; Chu et al., 2016), suggesting that XoxF may be the ‘preferred’ MDH. Microorganisms using XoxF-type enzymes are now of current interest not only for the ability to transform multi-carbon compounds (Good et al., 2016; Wehrmann et al., 2017), or alter trophic interactions (Krause et al., 2017), but also of for their potential in extracting REEs from electronics (Martinez-Gomez et al., 2016).

Verrucomicrobia methanotrophs rely on the XoxF MDH as all genomes thus far lack MxaF genes and isolate grow poorly in the absence of lanthanide metals. So far, Verrucomicrobia methanotrophs appear to be nearly absent in metagenomes assembled from mesophilic grassland and freshwater environments (Kielak et al., 2010; Cabello-Yeyes et al., 2017; He et al., 2017), and are primarily detected in thermo- and acidophilic geothermal ecosystems (Op den Camp et al., 2009; Sharp et al., 2014; Knief, 2015). One MAG putatively assigned to a Methylocacidiphilaceae (Methylocacidiphilaceae UBA1321) was reconstructed by Parks et al. (2017) from the Twitchell Island restored wetland. However, both the MAG and the metagenome lack the MMO genes (S. He et al., 2015), suggesting this might not be a bona fide Verrucomicrobia methanotroph. This clade is not well understood given the few number of isolates and limited genome-derived genomic information, however their reliance on XoxF type MDH, hydrogen metabolism (Carere et al., 2017; Mohammadi et al., 2017), and heat and acid tolerances make them physiologically unique and interesting.

Early investigations of xoxF distribution found the gene present in all the Proteobacteria methanotrophs, often in addition to the mxaFI-encoded MDH (Chistoserdova, 2011b). Environmental surveys suggested that xoxF might be more prevalent and widespread among uncultivated lineages in natural ecosystems than mxaF (Ramachandran and Walsh, 2015; Taubert et al., 2015). Metagenomic analysis of enrichment cultures of two closely related Methyllococcales and native samples from marine ecosystems found MAGs containing only xoxF MDH (Vekeman et al., 2016b; Padilla et al., 2017). However, as is the case when making
inferences from non-closed genomes, caution must be used when inferring absence of genes from MAGs. Interestingly, metagenomes from freshwater and terrestrial systems have failed to detect \textit{mxaF} in the metagenomes using both gene- and genome-centric approaches (Oswald \textit{et al}., 2017; Pratscher \textit{et al}., 2018; Singleton \textit{et al}., 2018; Smith \textit{et al}., 2018). This evidence, composed of both isolate and metagenomic information, suggests that the \textit{xoxF} is an environmentally relevant MDH, and therefore cultivation strategies for novel methanotrophs should be modified to include lanthanides in enrichment medium.

Conclusions/discussion

Metagenomics has greatly advanced the field of methane oxidation by providing an unbiased sequencing approach to successfully capture gene- and genomic-level information for novel lineages, improving our understanding of both the phylogenetic and metabolic diversity encompassed by methanotrophs. In the past decade metagenomics has more than doubled methanotroph genomic databases by reconstructing MAGs, which now includes representatives of clades that contain few or no isolated representatives. Specifically, these enigmatic lineages include members of the Proteobacteria (e.g. USC\textalpha, USC\textgamma, and \textit{Crenothrix}), members of the anaerobic methanotrophs (ANME-1, ANME-2), and of the genus \textit{Ca. Methylomirabilis} formerly of the candidate NC10 phylum. Additionally, MAGs have been recovered that are closely related to the well characterized methanotrophs (e.g. \textit{Methylobacter}), which open up the possibility of investigating finely resolved differences in genotypes that allow for specific environmental adaption.

Most importantly, metagenomic surveys have been critical to furthering our understanding of the methanotroph distribution across ecosystems. These studies have highlighted roles for methane oxidation from the depths of the ocean, to the tissues of endosymbionts, to agricultural soils. From MAGs assembled across these ecosystems, it is now possible to extend physiological insights from cultivation studies to the field scale. For instance, recent appreciation for the versatility of methanotroph energy generation pathways, has highlighted the capacity of these organisms to withstand fluctuating redox conditions. Moreover, metagenomic studies from marine, freshwater, and polar systems have found surprisingly low numbers of \textit{mxaF} genes, compared with the number of \textit{xoxF}, potentially further signifying the importance of this enzyme and informing cultivation efforts. It is clear that metagenomics will continue to play an important role in furthering understanding of methanotroph physiology and phylogenetic diversity.

Future trends

While we acknowledge that our analysis was not comprehensive of all current metagenomic sequencing projects (741 available in IMG before 1 June 2016 queried), geographic bias was obvious in our meta-analysis (Fig. 3.5). For example, far more methanotroph genomes have been recovered from metagenomes from the northern hemisphere, mostly in the United States and Europe, than anywhere else. The limitation in knowledge about methanotrophs in more equatorial latitudes may have important global ramifications, as tropical regions account for \approx50–60\% of the emissions from wetland systems (Dean \textit{et al}., 2018). Substantially increased metagenomic sequencing of more equatorial regions is likely to uncover novel methanotrophs involved in mediating the balance between methane source and sinks.

We also note that metagenomic sampling of methanotrophs in Eastern and Southern regions were similarly sparse. These include major rice cultivation areas such as China and India. Rice paddies are important ecosystems that provide food for a large global population, but also represent major sources of methane to the atmosphere. It is imperative that we continue to study the distribution and activity of methanotrophs in these and other agricultural systems, to better understand methane oxidation with different agricultural practices and climatic changes. As metagenomic sequencing becomes more cost efficient and commonplace, it is anticipated that methanotroph genomes will be better sampled across the globe in the near future.

Technological developments that continue to improve ease and efficiency of metagenomic sequencing and analysis will enable better gene- and genome-centric approaches. Sequencing technologies that generate long reads, for example
Nanopore (Feng et al., 2015) and PacBio (Rhoads and Au, 2015), are already showing great promise for improved metagenomic assemblies, yielding greater recovery of genes, pathways, and MAGs among complex, diverse samples. In addition, we are hopeful that pairing metagenomics with other tools (e.g. metatranscriptomics, metaproteomics, SIP, Nano-Sims, fluorescent in situ microscopy, or exometabolomics) will further elucidate the physical and metabolic interactions between methanotrophs and other community members. Coupling metagenomics with measurements of gene or protein expression have already proven invaluable for firmly confirming methane oxidation by the *Crenothrix* filaments (Oswald et al., 2017), nuances of trace metal metabolism in ANME (Glass et al., 2014), and will likely continue to prove to be exceptionally powerful for exploring methanotroph ecology and function.

According to metabolic reconstructions of ANME and *Ca. Methylomirabilis* MAGs to date, members of both lineages should be capable of coupling methane oxidation to respiration independently of consortia, although no isolated representatives currently exist. Increased sampling of these lineages both in the laboratory and the field may offer new physiological variants that better withstand axenic growth. For instance, MAGs from ANME and *Ca. Methylomirabilis* have been sampled from wastewater treatment and natural methane seep ecosystems (Welte et al., 2016); however, these same taxa are also detected and
often prevalent in wetlands, rice paddies, and freshwater lakes (Figs. 3.3 and 3.5) (Haroon et al., 2013; Welte et al., 2016; Narrew et al., 2017; in ‘t Zandt et al., 2018; Graf et al., 2018). These latter habitats represent an untapped potential for examination of the metabolic versatility of these uncultivated methanotroph lineages. Additionally, further genomic investigations of stable consortia may also offer insights into factors that hinder cultivation of these strains. Isolation of these microorganisms may be inhibited by their extremely slow growth rates, but this is unlikely to be the sole cause. Additional physiological-genomic studies may elucidate dependencies on other microorganisms, for example amino acid or vitamin auxotrophies, or the detoxification of toxic metabolites.

MAGs have highlighted the metabolic versatility that may enable methane oxidation along broader redox gradients, however much of this flexibility is currently only theoretical or poorly understood. An example of a currently only hypothetical metabolism is the reverse methanogenesis by Bathyarchaeota, encoding a divergent MCR gene and using pathways similar to the ones in ANME (Evans et al., 2015; Zhou et al., 2018). While there is no direct evidence for this metabolism, it is expected to yield acetate that could be exchanged with SRB. Further understanding of these exchanges may help explain isotopic anomalies already observed in some habitats (Zhou et al., 2018).

The role of dissimilatory nitrate reduction in the aerobic Methylococcales requires further investigation. Leveraging the large number of cultivated representatives within this clade to better understand oxygen responses in the Methylococcales seems warranted, given the broad distribution and activity of these taxa across the globe. For instance, some of these genomes encode up to four different, but possibly overlapping energy generation mechanisms: low-affinity cytochromes for oxic aerobic respiration, high-affinity cytochromes for suboxic aerobic respiration, respiratory nitrate reduction, and micro-aerobic fermentation (Kits et al., 2015; Skennerton et al., 2015; Padilla et al., 2017; Smith et al., 2018). It is currently almost completely unknown how these metabolic pathways are regulated, or what effects they have on methane oxidation rates. New discoveries could drastically alter our perception of the redox gradients across which Methylococcales can consume methane, as well as the metabolites secreted by these methanotrophs.

Additionally, there is increasing evidence that the metabolic versatility of methanotrophs is greater than was historically assumed. For example, the utilization of hydrogen has been observed in Verrucomicrobia, and may be possible in some Proteobacteria methanotrophs (Carere et al., 2017; Mohammadi et al., 2017; Singleton et al., 2018; Smith et al., 2018). While the role of hydrogen metabolisms, evident in Verrucomicrobia methanotrophs, is still unclear, hydrogen production or consumption by methanotrophs may expand the niches occupied by methanotrophs as well as their function in microbial communities. These proposed metabolisms require verification using axenic cultures, however metagenomic sequencing coupled with activity quantification of specific genes or pathways may continue to elucidate the distribution and impact of these genes or pathways under in situ conditions.

Web resources

We would like to acknowledge the Organization of Methanotroph Genome Analysis (OMeGA) for their invaluable contributions to the genomic databases of methane oxidizing microorganisms. Methanotroph genomes sequenced and analysed by OMeGA are available online at IMG/MER (https://img.jgi.doe.gov/cgi-bin/mer/main.cgi) or at the National Center for Biotechnology (https://www.ncbi.nlm.nih.gov/). We would like to thank IMG/MER for hosting the repository of metagenomic information that can be readily mined, and Simon Roux for assistance and guidance for the data mining.

Online supplemental sequences, metadata, tree generation commands, and R scripts used for the analyses presented here are available at the Wrighton laboratory’s GitHub website: https://github.com/TheWrightonLab/Methanotroph_rpS3Analyses_SmithWrighton2018. This repository includes minimally modified rpS3 amino acids sequences for the methanotroph reference and metagenomic sequences, the alignment used for phylogenetic tree generation, the phylogenetic tree displayed in Fig. 3.3, and the metadata tables used as inputs for information displayed in Figs 3.3, 3.4, and 3.5, and finally the R script for recreation of the analyses and figures.
References


