Genome-resolved metagenomics identifies novel active microbes in biogeochemical cycling within methanol-enriched soil

Michael C. Macey

Abstract
Metagenome assembled genomes (MAGs), generated from sequenced 13C-labelled DNA from 13C-methanol enriched soils, were binned using an ensemble approach. This method produced a significantly larger number of higher-quality MAGs compared to direct binning approaches. These MAGs represent both the primary methanol utilizers and the secondary utilizers labelled via cross-feeding and predation on the labelled methylotrophs, including numerous uncultivated taxa. Analysis of these MAGs enabled the identification of multiple metabolic pathways within these active taxa that have climatic relevance relating to nitrogen, sulfur and trace gas metabolism. This includes denitrification, dissimilatory nitrate reduction to ammonium, ammonia oxidation and metabolism of organic sulfur species. The binning of viral sequence data also yielded extensive viral MAGs, identifying active viral replication by both lytic and lysogenic phages within the methanol-enriched soils. These MAGs represent a valuable resource for characterizing biogeochemical cycling within terrestrial environments.

INTRODUCTION
Soils play a major role in global biogeochemical cycling, including the synthesis and metabolism of gases relevant to climate change (e.g., methane, nitrous oxide, methanol, carbon monoxide). Amongst these volatile, methanol represents the third most emitted volatile organic gas (120–350 Tg yr⁻¹), after methane (535 Tg yr⁻¹) and isoprene (350–520 Tg yr⁻¹) (Opacka et al., 2021). Methanol and its products from atmospheric interaction will act as either a net source or a net sink for radicals, which can in turn influence the lifetime of methane in the atmosphere (Galbally & Kirstine, 2002; Heikes et al., 2002). Methanol can be converted to formic acid by photochemical reactions that can enhance the formation of acid rain (Jacob, 1986).

Methanol-utilising methylotrophs can also play a major role in the cycling of further climate-impacting compounds with relevance to global biogeochemical cycling of the major elements (e.g., nitrous oxide production and consumption via denitrification [Abed et al., 2013; Prosper et al., 2020; Wrage et al., 2004], production and oxidation of dimethyl-sulfide [DMS] [Carrión et al., 2017; Zhang et al., 2019]). Characterisation of the role of this functional guild of microbes in biogeochemical cycling in soils requires the identification of their genomic potential within the soil microbiome. Stable isotope probing (SIP) is a powerful technique that can identify active substrate utilizing in an environment whilst reducing the issues with sequencing depth associated with complex microbial communities and identification of the specific microbes involved in a defined metabolic process (Chen & Murrell, 2010; Radajewski et al., 2002; Thomas et al., 2019). Incubation length is a key component of the design of SIP experiments—identification of environmentally relevant microbes requires incubations with environmentally
relevant substrate concentrations in the absence of further nutritional amendments, as these can influence the microbial community (Cébron et al., 2007; Chen et al., 2008; McDonald et al., 2005); however, this can result in long-term incubations, which in turn enhances the ability of non-target microbes to incorporate the 13C label into their biomass via cross-feeding (Lueders et al., 2006; Macey et al., 2020). However, in addition to the generation of a focussed metagenome enabling the identification of the taxonomic and functional potential of the microbial community, genome-resolved metagenomics can inform the identification of primary and secondary 13C utilizers.

In a previous study, we performed a DNA-SIP experiment with 13C-labelled methanol to identify methylotrophs in pea and wheat rhizosphere soil relative to the unplanted soil to identify the impact of plant growth on the soil community (Macey et al., 2020). This study identified that *Methylobacterium* was enriched in the rhizosphere soils relative to the unplanted soils, potentially due to exudates and plant-generated methanol (Galbally & Kirstine, 2002). This potential for enrichment of soil methylotrophs in association with plants has also been observed via other studies that performed SIP experiments with *Zea mays*, *Arabidopsis thaliana*, grassland and deciduous forest soils (Butterfield et al., 2016; Chaignaud et al., 2018; Li et al., 2019; Morawe et al., 2017). The metagenomes generated from the 13C labelled DNA from the previous Macey et al., 2020 study were analysed with a direct binning approach. This enabled the identification of 10 metagenome-assembled genomes (MAGs), which included the most abundant methanol utilizers across four MAGs and six additional MAGs that represented potential cross-feeders. However, via an ensemble approach, it is possible to identify a greater number of MAGs (Yue et al., 2020) and detect MAGs representative of a greater diversity of taxa in the active soil community. Across multiple studies, direct binning programs are shown to vary in their relative performance across different environments and sample types (Bak et al., 2023; Wang et al., 2023) (e.g., variable numbers of bins, completion, and contamination). Ensemble binning methods allow the integration and dereplication of binning results of multiple direct binning programs—an ensemble binning approach can therefore generate a higher number of good quality MAGs without the high levels of redundancy from directly utilizing the outputs of direct binning programs.

In this study, we applied an ensemble approach to enhance the identification of the taxonomic and functional potential of the active carbon cycling component of the soil microbiome following enrichment with methanol. Viral MAGs were also assembled and characterised, as these have previously been shown to have a major impact on both enriched diversity and carbon cycling (Barnett & Buckley, 2023; Starr et al., 2021; Trubl et al., 2021).

### EXPERIMENTAL PROCEDURES

#### Sample description and sequencing

In the originating study (Macey et al., 2020), the soil was collected in April 2015 from a naturally grassed and unfertilised part of John Innes Centre Church Farm (Norfolk, UK) (52.6276 N, 1.1786 E). The top 10 cm was removed from a 1 m² section and then soil to 20 cm depth was removed, air dried and sieved through 10 mm and 5 mm sieves. Sterilised *Pisum sativa* and *Triticum aestivum* seeds were then planted and grown in 10 cm × 10 cm pots in bulk soil and growing at 22°C under long day regimes (16:8 h) in plant growth rooms for 4 weeks. Unplanted controls were run alongside. Rhizosphere soil was collected by shaking the plant roots three times to separate non-rhizosphere soil (Bulgarelli et al., 2012). Roots were then vortexed in autoclaved 1× phosphate buffered saline (PBS) (In 1 L of water: NaCl 8 g, KCl 0.2 g, Na2HPO4 1.44 g and KH2PO4 0.24 g) to separate the adhering soil that is defined as rhizosphere soil. 2 g of rhizosphere soil and unplanted soil was dispensed into 120-ml serum vials in triplicate and combined with 40 mL of sterile H2O (soil slurry). Vials were then supplemented with 10 μmol of 13C methanol or 12C methanol. The serum vials were sealed air-tight with rubber seals and then incubated at 30°C without light in a shaking incubator (120 rpm). These vials were incubated alongside unincubated vials with 40 mL of sterile H2O incubated with 100 μM–1 mM methanol to act as standards for the concentration of methanol. The concentration of methanol in the headspace of the serum vials was measured using gas chromatography as detailed in Macey et al., 2020. After depletion of methanol, samples were resupplied with methanol to the same concentration. After 17 days, a total of 200 μmol of C had been consumed in incubations and soil was collected for DNA extraction. DNA was extracted from all soil samples via the Griffiths technique (Griffiths et al., 2000). DNA was processed via caesium chloride density gradient centrifugation to separate the 13C- and 12C-labelled DNA from 1 to 3 μg of DNA from each test group according to established protocols and DNA was recovered by precipitation (Neufeld et al., 2007). Confirmation of 13C labelling was achieved via denaturing gradient gel electrophoresis and 16S rRNA gene amplicon sequencing of the heavy and light fractions of the DNA extracted from the 12C and 13C enriched test groups to confirm the occurrence of enriched taxa, as detailed in Taubert et al., 2015, Taubert et al., 2019 and Macey et al., 2020. The DNA from the 13C-heavy fractions of the methanol SIP experiment was then pooled, quantified, and sequenced by the Centre for Genomic Research at the University of Liverpool. Sequencing was performed using paired-end sequencing (2 × 150 bp) on an Illumina HiSeq 4000.
Metagenome analysis and assembly of metagenome assembled genomes

Short sequences and sequences of poor quality were excluded from the files using the program Trimmomatic (0.36) (Bolger et al., 2014) and the quality of the metagenomes was assessed using QUAST, including the MetaQUAST expansion (5.0.0) (Gurevich et al., 2013). The reads from the pea, wheat and unplanted soil methanol enrichments were pooled to enhance the representation of less abundant taxa and these were processed alongside the individual metagenomes. Reads were assembled using Megahit (1.1.2) (Li et al., 2015). The Trimmomatic and Megahit steps were performed using Kbase (Arkin et al., 2018). The taxonomic profile of the metagenome was analysed using Metaphlan3 (3.0) (Bezhini et al., 2021; Segata et al., 2013).

Contigs were binned into metagenome-assembled genomes (MAGs) using the binning programs MaxBin2 (2.2.4) (Wu et al., 2016) (version), MetaBAT2 (1.7) (Kang et al., 2019) and Concoct (1.1) (Alneberg et al., 2014). The MAGs were then processed into a non-redundant set of MAGs using the program DasTool (1.1.2) (Sieber et al., 2018) and dRep (3.1) (Olm et al., 2017) with default settings. The completeness, contamination and heterogeneity of these MAGs were assessed using the program CheckM (1.0.13) (Parks et al., 2015). MAGs within 10% of a contamination quality score threshold (5% for high quality and 10% for medium quality) were refined using VizBin (Laczný et al., 2015). The taxonomy of the MAGs was then assessed using GTDBk and MAGs with taxonomic identification above the family level were further assessed using TYGS and Protologger (Chaumeil et al., 2020, 2022; Hitch et al., 2021; Meier-Kolthoff & Göker, 2019). The presence of specific functional genes, metabolic pathways, rRNAs and tRNAs was also screened for using BlastKOALA (Kanehisa et al., 2016) and DRAM (Shaffer et al., 2020). Genes of interest were also screened for by performing local BLAST searches in BioEdit with tblastn against local databases generated from the MAGs (Hall, 1999).

Ensemble binning of MAGs from the combined methanol-enriched soil metagenomes generated 48 microbial MAGs of medium quality (>50% completeness and <10% contamination) (Table 1). For MAGs of this quality threshold, Concoct generated 25, MetaBAT 22 and MaxBin 21. DRep did not identify any MAGs as redundant, supporting that the binning programs generated partially distinct sets of MAGs. Prior analysis of the metagenomes (Macey et al., 2020) generated 18 MAGs of at least medium quality, 11 Pseudomonadae data MAGs, with additional detection of five Verrucomicrobiota and one Bdellovibrio and Thermoproteota MAGs. Pseudomonadae and Verrucomicrobiota were also the most abundant phyla in the MAGs generated in this study, with 18 and five MAGs respectively. One Thermoproteota and one Bdellovibrio MAG were also generated. This study also generated MAGs from eight further phyla: Actinobacteria (3), Planctomyceto-ta (2), Patesicibacteria (2), Myxococcota (1), Gemmatimonadota (1), Cyanobacteria (1), Chloroflexota (1), Armatimonadota (1). Bins 5, 12, 19, 30 and 43 all had contamination scores above 10%, identifying them as low-quality MAGs. However, via manual selection of the MAG contigs with VizBin, it was possible to reduce their contamination levels to below 10%, increasing the quality assignment of these MAGs. MAG completeness remained above 50% in all instances. The MAGs included those that could not be assigned to high levels of taxonomic resolution, being classified to the levels of Family (4 MAGs – Hyphomicrobiaceae, Methylophilaceae, Bdellovibrioaceae, Nitrosophaeraeaceae), Order (2 MAGs – Gemmatimonadales, Micavirionales),
<table>
<thead>
<tr>
<th>Classification by GTDBk</th>
<th>Contigs</th>
<th>Completeness %</th>
<th>Contamination %</th>
<th>5S rRNA</th>
<th>16S rRNA</th>
<th>23S rRNA</th>
<th>tRNA count</th>
<th>Abundance in reads*</th>
<th>Relative abundance in community profile*</th>
</tr>
</thead>
<tbody>
<tr>
<td>bin.001.fa Methylophilus</td>
<td>259</td>
<td>66.22</td>
<td>7.51</td>
<td>21</td>
<td>1.2</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin.002.fa Methylibium</td>
<td>100</td>
<td>91.84</td>
<td>9.84</td>
<td>41</td>
<td>0.16</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin.003.fa Methylobacterium</td>
<td>137</td>
<td>86.29</td>
<td>5.22</td>
<td>37</td>
<td>0.28</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin.004.fa Hyphomicrobiaceae</td>
<td>463</td>
<td>94.49</td>
<td>5.11</td>
<td>57</td>
<td>0.24</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin.005.fa Senicyctochromatia UBA7694</td>
<td>1726</td>
<td>77.26</td>
<td>6.41</td>
<td>74</td>
<td>0.26</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin.006.fa Roseimicrobium</td>
<td>833</td>
<td>56.08</td>
<td>3.43</td>
<td>18</td>
<td>0.14</td>
<td>0.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin.007.fa Methylophilaceae MM2</td>
<td>118</td>
<td>93.76</td>
<td>9.93</td>
<td>42</td>
<td>1.8</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin.008.fa Fimbriimonas</td>
<td>268</td>
<td>91.51</td>
<td>1.47</td>
<td>47</td>
<td>0.09</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin.009.fa Myxococca XYA12-FULL-58-9</td>
<td>217</td>
<td>93.31</td>
<td>2.26</td>
<td>46</td>
<td>0.17</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin.010.fa Methylophilaceae</td>
<td>403</td>
<td>82.89</td>
<td>7.95</td>
<td>40</td>
<td>0.94</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin.011.fa Methylophilaceae MM1</td>
<td>137</td>
<td>94.02</td>
<td>6.92</td>
<td>36</td>
<td>0.39</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin.012.fa Bdellovibrionaceae 21-14-0-10-47-8</td>
<td>249</td>
<td>86.74</td>
<td>1.8</td>
<td>28</td>
<td>0.06</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin.013.fa Solunubacteraceae</td>
<td>580</td>
<td>66.87</td>
<td>2.14</td>
<td>35</td>
<td>0.14</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin.014.fa Thermoleophilia Ga0077560</td>
<td>509</td>
<td>70.52</td>
<td>1.03</td>
<td>34</td>
<td>0.08</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin.015.fa Gemmatimonadales GWC2-71-9</td>
<td>932</td>
<td>58</td>
<td>9.6</td>
<td>33</td>
<td>0.09</td>
<td>0.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin.016.fa Verrucimicrobium</td>
<td>1154</td>
<td>80.04</td>
<td>1.39</td>
<td>25</td>
<td>0.1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin.017.fa Patescibacteria UBA1384</td>
<td>69</td>
<td>62.74</td>
<td>1.03</td>
<td>38</td>
<td>0.03</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin.018.fa Methyloceanibacter</td>
<td>510</td>
<td>90.78</td>
<td>7.8</td>
<td>47</td>
<td>0.8</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin.019.fa Phyllophora UBA1161</td>
<td>813</td>
<td>86.33</td>
<td>5.75</td>
<td>49</td>
<td>0.21</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin.020.fa Methylophilaceae MM2</td>
<td>278</td>
<td>86.26</td>
<td>8.3</td>
<td>48</td>
<td>0.73</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin.021.fa Micavirionales UBA2020</td>
<td>136</td>
<td>94.66</td>
<td>6.45</td>
<td>38</td>
<td>0.09</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin.022.fa Methylophilaceae MM2</td>
<td>302</td>
<td>78.39</td>
<td>5.28</td>
<td>31</td>
<td>1.38</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin.023.fa Nitrososphaeraceae TA-21</td>
<td>160</td>
<td>96.15</td>
<td>0.97</td>
<td>37</td>
<td>0.05</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin.024.fa Methylophilaceae MM1</td>
<td>30</td>
<td>86.7</td>
<td>0.06</td>
<td>29</td>
<td>0.14</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin.025.fa Chthoniobacter</td>
<td>553</td>
<td>87.05</td>
<td>3.04</td>
<td>46</td>
<td>0.13</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin.026.fa Ramlibacter</td>
<td>527</td>
<td>65</td>
<td>9.85</td>
<td>28</td>
<td>1.88</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin.027.fa Planctomyces J058</td>
<td>993</td>
<td>57.94</td>
<td>0.72</td>
<td>41</td>
<td>0.07</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin.028.fa Sphingomonas</td>
<td>599</td>
<td>68.13</td>
<td>3.73</td>
<td>29</td>
<td>0.08</td>
<td>0.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin.029.fa Micavirionales UBA2020</td>
<td>256</td>
<td>87.6</td>
<td>8.97</td>
<td>45</td>
<td>0.2</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bin.030.fa Methylacidiphilales UBA1321</td>
<td>571</td>
<td>96.96</td>
<td>5.41</td>
<td>50</td>
<td>0.17</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Class (3 MAGs Thermoleophilia, Physicphaerae, Sericytochromatia) and Phylum (6 MAGs – Planctomycetota, Myxococcota, Actinobacteriota, Chloroflexota and Patescibacteria). These MAGs are of value for characterising members of these uncultivated taxa, including attempting to elucidate their potential role in biogeochemical cycling in the soil environment. Collectively, the MAGs represent 25.77% of the mapped reads, with 11.18% of the mapped reads represented by bin 46, a Methylophilus MAG. Of the five MAGs that represented >1% of the mapped reads, three were also identified as Methylophilaceae (bins 1, 7 and 26). The fourth bin was identified as bin 30, which contained xoxF and was identified as Ramlibacter. The fifth bin representing >1% of the mapped reads were identified as Rariglobus (previously identified as Opitutus in Macey et al., 2020), which was identified as a 1% relative abundance of the enriched microbiome.

Identification of methylotrophic and non-methylotrophic diversity

The prior analysis identified ten Pseudomonodata MAGs that contained genes encoding the xoxF methanol dehydrogenases and associated accessory genes (e.g., genes encoding for associated cytochromes or the mxaJ, a Periplasmic solute binding protein [Schmidt et al., 2010; Choi et al., 2013]) in the Methylophilales (8), Methylobacterium (1), and Ramlibacter (1) (Macey et al., 2020). Ensemble binning approaches applied in this study identified an additional five MAGS that either contained the xoxF methanol dehydrogenase gene and were assigned to taxa associated with methanol oxidation (Xanthobacteraceae [Tikhonova et al., 2021], Ideonella [Hachisuka et al., 2022] and Methylibium [Nakatsu et al., 2006; Song & Cho, 2007]) or were only assigned to taxa associated with methanol oxidation (Methylocalcuphilum [Hou et al., 2008; Versantvoort et al., 2019] and Methyloceainibacter [Takeuchi et al., 2014, 2019]); whilst none of the 13 species of Ramlibacter is capable of methanol oxidation, members of other genera within the Comamonadaceae contain species confirmed to grow on methanol (Agafonova et al., 2017; Satola et al., 2013) and xoxF genes are present in members of 28 of the 34 genera of the family (Macey et al., 2020). As methanol oxidation is consistently tested in the absence of lanthanides, and these taxa only possess a lanthanide-dependant methanol dehydrogenase (Fitriyanto et al., 2011; Hibi et al., 2011; Nakagawa et al., 2012; Wang et al., 2019), this metabolic trait may be undetected. Surprisingly, mxaFL, the canonical, non-lanthanide dependant methanol dehydrogenase encoding genes were found in only three of the MAGs assembled in this study—bins 3 (Methylobacterium), 28 (Methylocalcuphilum) and 42 (Xanthomonadaceae).

Identification of the greater diversity of MAGs that were assigned to taxa not associated with methanol
oxidation and that do not possess methanol dehydrogenases indicates that these additional MAGs could represent cross-feeding microbes. The presence of these MAGs in the metagenome indicates that they were $^{13}$C labelled via utilisation of substrates generated by the methylotrophs either as a direct consequence of methanol oxidation by-products from methanol oxidation (e.g., formaldehyde, formate, carbon dioxide), alternate substrates (sugars or exudates [Beck et al., 2013; Hernandez et al., 2015; Krause et al., 2017]) and/or labelled cellular material via necrophagy or predation (Barnett et al., 2016; Barnett & Buckley, 2023; Hanajima et al., 2019; López-Mondéjar et al., 2020; Zumsteg et al., 2013). With regards to the utilisation of formaldehyde and formate, all the MAGs proposed as potential methylotrophs contained genes encoding for formaldehyde activating enzymes and formate—tetrahydrofolate ligases. These enzymes perform the condensation of formaldehyde and tetrahydrodismethanopterin to methylene tetrahydrodismethanopterin (Kim et al., 2020; Marx et al., 2003), which is then integrated into further formaldehyde and C1 metabolism pathways (Anthony, 1982). A further 12 MAGs outside of the proposed methylotrophs contained formaldehyde-activating enzymes and no further methylotrophy-related genes, indicating a possible role in resolving issues associated with toxicity as opposed to metabolism (Vorholt et al., 2000; Yurimoto et al., 2005). Further to these genes, 20 complete formate dehydrogenase encoding genes were found across 10 MAGs; Bins 3, 4, 7 and 24 all contained two, and bins 12, 13, 26, 27, 28 and 38 contained one. Of these 10 MAGs, only bin 27, identified as Nitrososphaeraceae, was not a proposed methylotroph.

Prior SIP experiments have identified non-methanotrophic methylotrophs and non-methylotrophs labelled in $^{13}$C-methane SIP experiments via the proposed metabolism of leaked labelled compounds (Beck et al., 2013; Taubert et al., 2019). Based on the shared metabolic pathway for complete methanol oxidation, these represent potential substrates in this experiment. The metabolic capabilities of the different groups indicate that they could have used many $^{13}$C compounds potentially produced by the methylotrophs. This includes carbon dioxide, compounds exuded by the methylotrophs or the cellular components of lysed methylotrophs (Pankratov et al., 2008; Noar & Buckley, 2009; Dumont et al., 2011). Necromass can represent a major organic carbon source in oligotrophic environments (Chatzigiannidou et al., 2018); therefore, it is also possible that some of the heterotrophic taxa incorporated the $^{13}$C label into their DNA via consumption of cellular material from dead labelled cells, as has previously been shown to occur in SIP experiments performed with $^{13}$C enriched dead Escherichia coli cells (Hanajima et al., 2019). These means of $^{13}$C labelling are well-characterised issues with long-term SIP experiments, indicating that the timescale of the methanol enrichment series was sufficient to allow this cross-feeding to occur; this labelling also reinforces the balance between supplying an environmentally relevant concentration of a substrate, incorporating sufficient $^{13}$C to enable the collection of sufficient DNA for metagenome approaches and avoiding supplementation with additional nutrients to expedite the labelling of the cells. Furthermore, whilst not the primary utilizers of the labelled substrate, labelling that arises because of cross-feeding does permit the identification and characterisation of active members of the soil microbial community. Bins 9 and 14 were assigned to bacterial taxa associated with predation of microbial cells (Myxococccota and Bdellovibrionacea respectively). This labelling of predatory bacteria was previously observed in SIP experiments performed with $^{13}$C methane or $^{13}$C enriched prey cells and various terrestrial innocula (Barnett et al., 2016; Hutchens et al., 2004; Lueders et al., 2004). The methylotrophs enriched in these experiments were exclusively Gram-negative, effectively generating an abundant source of labelled carbon for this specific metabolic guild of microbe.

Biogeochemical cycling in the enriched soil

Multiple genes involved in the cycling of climatically relevant compounds were identified in the MAGs. Genes encoding for the complete dissimilatory nitrate reduction to ammonium (DNRA) ($\text{narGHI}$ and $\text{nirBD}$) were detected in three MAGs, all non-methylotrophic organisms (two Planctomycetota and one Chthoniobacter). As DNRA is typically associated with heterotrophy, this supports the identification of these labelled taxa as cross-feeders. This metabolism is also associated with the conservation of nitrogen within an ecosystem (Pandey et al., 2020; Zhang et al., 2015)—these microbes may therefore play a role in the cycling of nitrogen within the soil environment as well as conserving the limited nitrogen within the batch culture approach of the enrichment. A MAG that was taxonomically assigned as an ammonia oxidising archaeon (AOA), being assigned to the Nitrososphaeraceae. This MAG contained a gene encoding for an ammonia monoxygenase, which performs the oxidation of ammonia to nitrite. As AOA have been shown to prefer environments with lower ammonia, this supports the low nutrient conditions of the enrichment series and the soil. It is also possible that there is an interchange of nitrogenous species between the AOA and DNRA. The enrichment of these functional groups also indicates that the enrichments may have transitioned to anoxic or microaerophilic conditions despite replenishment of the headspace (Macey et al., 2020). The enrichment of the AOA is also of note because they have been
identified as inhibited by methanol concentrations that overlap with the supplied methanol concentration in this experiment (Macey et al., 2020; Oudova-Rivera et al., 2023).

Three MAGs contained genes encoding the incomplete denitriﬁcation pathway—Solirubrobacteraceae missing the nitrous oxide reductase and Hyphomicrobiaceae and Sphingomonas only possessing the nitrate and nitrite reductase encoding genes, generating NO. Bins 24 and 26, taxonomically identiﬁed as Methylophilaceae (MM2) were the only MAGs that contained genes encoding for the complete denitriﬁcation pathway, as previously identified in the characterisation of the isolate Methylobacillus sp. MM2 (Macey et al., 2018). This is of note because N₂O is the third most emitted greenhouse gas, with 300 times the forcing of CO₂ and a 120-year atmospheric lifespan, it is the most potent. Nitrous oxide is produced by microbes through either incomplete denitriﬁcation, converted from nitric oxide, or through the abiotic oxidation of hydroxylamine by ammonia-oxidising microbes (Prosser et al., 2020). The production of N₂O is also inﬂuenced by the other nitrogen cycling processes that increase the pool of nitrogenous species for denitriﬁcation (e.g., Annamox: Pajares & Ramos, 2019). Active microbes in this soil environment therefore appear to possess the genetic potential to have a major impact on inorganic and disimilatory nitrogen metabolism.

With regards to carbon cycling, genes were identiﬁed that encoded the aerobic carbon monoxide dehydrogenase enzyme that catalyses the oxidation of carbon monoxide to carbon dioxide. Carbon Monoxide is strongly connected to methane and is considered “Kyoto’s Forgotten Gas” (Fry et al., 2013), having been omitted from all current climate accords. CO oxidation-related genes are broadly present in the microbiomes of terrestrial environments (Cordero et al., 2019) but the physiological process is poorly deﬁned with regards to the ecology and physiology of CO-oxidising microbes, with some organisms able to utilise this as either an energy source or a carbon and energy source. These genes were identiﬁed in four of the MAGs (Thermophiles, Solirubrobacteraceae, Ramlibacter, and Hyphomicrobiaceae); The genomes of isolated Ramlibacter species and Solirubrobacteraceae MAGs generated from DNA extracted from desert soils have both been shown to contain carbon monoxide dehydrogenase genes (Ray et al., 2022). Species within the Hyphomicrobiaceae have also been conﬁrmed as capable of carbon monoxide oxidation (Volpiano et al., 2021). This observed genetic potential within taxonomically diverse and active soil microbes indicates a potential relevance of carboxydrotrophy within this soil environment.

Further to the cox genes, the MAGs identiﬁed as Solirubrobacteraceae and Hyphomicrobiaceae also possessed the genes encoding for a methanethiol oxidase; this enzyme performs the oxidation of methanethiol to formaldehyde and hydrogen sulfide. Additional MAGs that contain this gene include the Gemmatimonadales, Ideonella and Methylophilaceae. Additional methanethiol metabolism-related genes, encoding Methanethiol S Transferases, which catalyses the methylation of methanethiol (MeSH) to yield dimethylsulphide (DMS), were also found in multiple MAGs representative of both methylotrophic (Hyphomicrobiaceae and Methylophilaceae) and non-methylotrophic taxa (Verrucomicrobiium, Planctomycetota and Tardiphaga); members of the species of Hyphomicrobiaceae have been conﬁrmed as capable of DMS oxidation (Pol et al., 1994) and Methylophilaceae have previously been indicated as active in terrestrial DMS degradation (Eyice et al., 2015), but the other identiﬁed taxa identiﬁed in the MAGs remain unconﬁrmed. The metabolism of methanethiol and DMS represents a major component of the global sulfur cycle (Carrión et al., 2017, 2019), indicating an additional role for the active soil microbiome on the metabolism of climatically relevant sulfur species. Screening of the MAGs also identiﬁed two MAGs identiﬁed as members of the Candidatus Patescibacteria, which contains the bacterial candidate phyla radiation. Analysis of these MAGs identiﬁed a minimal presence of metabolic pathways, reinforcing the streamlined nature of the genomes of these taxa (Lemos et al., 2019; Narasingarao et al., 2012; Tian et al., 2020). However, these MAGs did contain rRNA genes, which allowed their genetic identity to be better resolved to Candidatus Nomurabacteria (bin 35) and Candidatus Berkelbacteria (bin 21).

Identification of viral diversity

Vibrant extracted 644 Viral MAGs from the methanol-enriched soil metagenome (Supplementary Figure 2). RNA viruses may also be present and enriched in these samples, but these cannot be detected with the sequencing approach used in this study. Of these, 626 were low quality, 16 were medium quality and two were high quality, with 560 lytic and 84 lysogenic. Four of the low-quality viral MAGs were identiﬁed as circular, with the rest all identiﬁed as linear. The two high-quality MAGs were identiﬁed as lytic and did not cluster with any other phage in the RefSeq or NCBI databases, but screening of individual genes from the genome identiﬁed matches against the genomes of betaproteobacteria; this suggests that this may represent the host diversity of this virus. The most abundant pathways within the viral MAGs are related to amino acid, metabolite and cofactor synthesis (Supplementary Figure 3); these pathways are related to the enhanced replication of viral particles (Zhang et al., 2021). The detection of this high number of viral MAGs in a metagenome generated from the 13C-labelled DNA from an SIP experiment would indicate that these phages were either actively infecting and lysing cells at the termination of the experiment or were lysogenic and contained in the
genomes of actively replicating microbes. The high abundance of lytic phages also supports that this enhanced viral activity was in response to the increased abundance of specific taxa within the soil. Phages have previously been identified in SIP experiments investigating carbon cycling in soil (Barnett & Buckley, 2023; Greenlon et al., 2022; Haig et al., 2015; Lee et al., 2012), with labelling identified across a variety of 13C enrichment regimes (e.g., labelled plant tissue and 13CO2 fed plant rhizo-exudates); this ability to delineate the active phage population from preserved environmental DNA represents a powerful technique to investigate their population dynamics and the resultant impact on carbon cycling.

CONCLUSION

The generation of focussed metagenomes via SIP experiments represents a powerful means to identify the active taxa performing a specific metabolic group within an environment. Whilst cross-feeding consistently occurs across SIP experiments, the ability to identify the genomic contexts of labelled taxa enhances the means through which primary utilizers can be identified. Furthermore, the characterisation of MAGs involved in the secondary utilisation of this carbon represents an approach to further elucidate the biogeochemical cycling and genetic diversity of uncultivated taxa.

AUTHOR CONTRIBUTIONS

Michael C. Macey: Data curation (equal); investigation (equal); methodology (equal); writing — original draft (equal); writing — review and editing (equal).

ACKNOWLEDGEMENTS

This work was supported by the Norwich Research Park BBSRC Doctoral Training Program, the Earth and Life Systems Alliance (ELSA) at the University of East Anglia, and Research England via the “Expanding Excellence in England” grant number 124.18.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The original metagenome data are available in GenBank under BioProject number PRJNA533040.

ORCID

Michael C. Macey https://orcid.org/0000-0002-1800-0469

REFERENCES


**SUPPORTING INFORMATION**

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Macey, M.C. (2024) Genome-resolved metagenomics identifies novel active microbes in biogeochemical cycling within methanol-enriched soil. Environmental Microbiology Reports, 16(2), e13246. Available from: https://doi.org/10.1111/1758-2229.13246