nature microbiology

Article

https://doi.org/10.1038/s41564-022-01252-3

Methanotrophy by a *Mycobacterium* species that dominates a cave microbial ecosystem

Received: 23 June 2021

Accepted: 14 September 2022

Published online: 3 November 2022

Check for updates

Rob J. M. van Spanning[1] ✉, Qingtian Guan[2], Chrats Melkonian[1,13,14], James Gallant[3], Lubos Polerecky[4], Jean-François Flot[5], Bernd W. Brandt[6], Martin Braster[1], Paul Iturbe Espinoza[1], Joost W. Aerts[1], Marion M. Meima-Franke[7], Sander R. Piersma[8], Catalin M. Bunduc[1], Roy Ummels[3], Arnab Pain[2], Emily J. Fleming[9], Nicole N. van der Wel[10], Vasile D. Gherman[11], Serban M. Sarbu[9,12], Paul L. E. Bodelier[7] and Wilbert Bitter[1,3] ✉

So far, only members of the bacterial phyla Proteobacteria and Verrucomicrobia are known to grow methanotrophically under aerobic conditions. Here we report that this metabolic trait is also observed within the Actinobacteria. We enriched and cultivated a methanotrophic *Mycobacterium* from an extremely acidic biofilm growing on a cave wall at a gaseous chemocline interface between volcanic gases and the Earth's atmosphere. This *Mycobacterium*, for which we propose the name *Candidatus* Mycobacterium methanotrophicum, is closely related to well-known obligate pathogens such as *M. tuberculosis* and *M. leprae*. Genomic and proteomic analyses revealed that *Candidatus* M. methanotrophicum expresses a full suite of enzymes required for aerobic growth on methane, including a soluble methane monooxygenase that catalyses the hydroxylation of methane to methanol and enzymes involved in formaldehyde fixation via the ribulose monophosphate pathway. Growth experiments combined with stable isotope probing using ^{13}C-labelled methane confirmed that *Candidatus* M. methanotrophicum can grow on methane as a sole carbon and energy source. A broader survey based on 16S metabarcoding suggests that species closely related to *Candidatus* M. methanotrophicum may be abundant in low-pH, high-methane environments.

The limits of life and its potential on other planets are often studied using microbial communities from extreme environments[1]. One such community can be found in Sulfur Cave on Puturosu Mountain (Stinky Mountain) in Romania, an area that emits volcanic gases composed predominantly of CO_2 and traces of H_2, H_2S and CH_4 (refs. [2–4]). In this cave, volcanic gases only slowly mix with atmospheric air and form a relatively stable gaseous chemocline. At the interface of this chemocline, the cave walls are covered by a 5–10-cm-wide biofilm (Extended Data Fig. 1). The biofilm is minimally exposed to light, has no contact with external sources of organic carbon or fixed nitrogen, and its only source of liquid water is the condensation of atmospheric water vapour[5]. A special feature of the biofilm is its very low pH (close to 1) most likely resulting from sulphuric acid production during aerobic oxidation of H_2S (ref. [6]). An earlier study based on 16S rRNA gene amplicon sequencing showed that the microbial community of the cave biofilm comprises members of bacterial and archaeal genera known to thrive in highly acidic environments, including *Acidithiobacillus*, *Ferroplasma* and *Thermoplasma*[5]. However, the community is dominated by members of the *Mycobacterium* genus, which was unexpected and prompted us to explore the metabolic traits facilitating their success in such an unusual environment.

A full list of affiliations appears at the end of the paper. ✉e-mail: rob.van.spanning@vu.nl; w.bitter@amsterdamumc.nl

https://doi.org/10.1038/s41564-022-01252-3

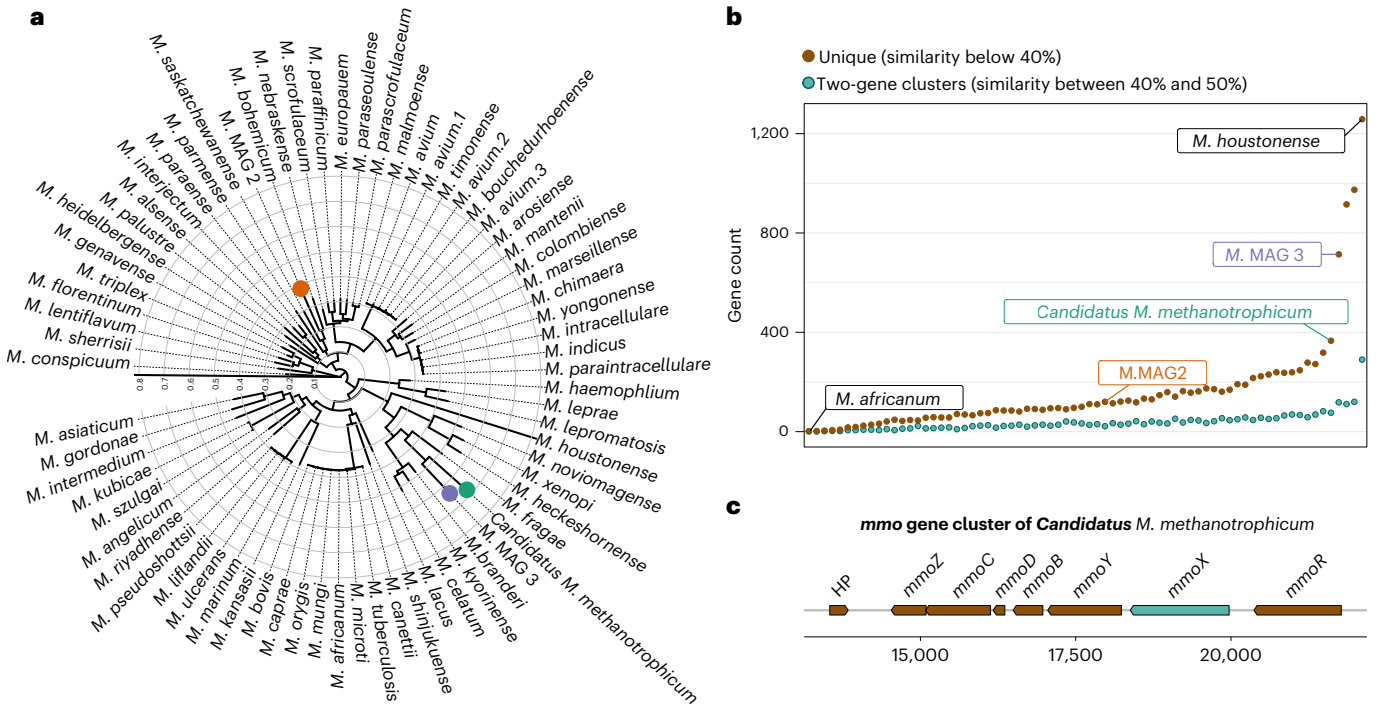


Fig. 1 | Phylogenetic position and unique genetic make-up of the Sulfur Cave mycobacteria. a, Phylogenetic tree of the genus *Mycobacterium*. The tree excludes the four relatively fast-growing genera of the Mycobacteriaceae (*Mycolicibacter*, *Mycolicibacillus*, *Mycolicibacterium* and *Mycobacteroides*). The tree was constructed using 316 monocore gene markers, which are genes common in each of the species within the genus (Methods). Bootstrap values are given in the tree shown in Supplementary Fig. 1. The internal tree scale is based on amino acid sequence similarity. **b**, Count of monocore gene markers unique within the genomes of *Mycobacterium* species shown in **a**. Brown points represent monocore gene markers with a similarity below 40%, while green points represent two-gene orthologous clusters with a similarity between 40% and 50%. **c**, The *mmo* gene cluster encoding sMMO of Candidatus M. methanotrophicum. Genes coloured in brown are unique to *Candidatus* M. methanotrophicum within the genus *Mycobacterium* (similarity below 40%), whereas *mmoX*, shown in green, additionally occurs in *Mycobacterium pseudoshottsii* (similarity 41%).

Some members of the Mycobacteriaceae, such as *Mycobacterium tuberculosis* and *Mycobacterium leprae*[7], are obligate and persistent pathogens, whereas others grow in diverse habitats and display great metabolic flexibility[8–10]. Members of this family were shown to use a variety of carbon and energy sources such as lipids, fatty acids, aliphatics or aromatic hydrocarbons. Some of them grow on petroleum compounds, which is consistent with their genetic potential to synthesize different types of mono- and dioxygenase[11]. Most of the Mycobacteriaceae prefer molecular oxygen as the electron acceptor for respiration, although some species have the potential to use nitrate as an alternative electron acceptor[12].

In this Article, we show that the most abundant *Mycobacterium* in the cave biofilm is methanotrophic. Using metagenomics and proteomics, we identify key enzymes involved in its energy metabolism and biomass production based on CH_4. Furthermore, by combining growth experiments with stable isotope probing, we establish that this species can grow on CH_4 as a sole source of carbon and energy. On the basis of these results, we propose that the most abundant *Mycobacterium* in the cave biofilm be called *Candidatus* Mycobacterium methanotrophicum. With this report of aerobic methanotrophy within the Actinobacteria, we add this phylum to the other two bacterial phyla, Proteobacteria and Verrucomicrobia, where this metabolic trait has been observed[13].

Results

Genomic analysis of the biofilm community

Previous taxonomic analyses using 16S rRNA gene amplicon sequencing revealed that three *Mycobacterium* species are highly abundant in the cave biofilm[5]. To follow up on these observations, we obtained new biofilm samples, isolated DNA and performed metagenomic analyses. We were able to assemble metagenome-assembled genomes (MAGs) of the three previously identified *Mycobacterium* species and denoted them as *M.* MAG 1, 2 and 3 in the order from highest to lowest abundance in the community (Supplementary Table 1). On the basis of the ability to grow methanotrophically (see below), *M.* MAG 1 was named *Candidatus* M. methanotrophicum.

A phylogenetic tree constructed on the basis of 316 monocore gene markers (Fig. 1a and Supplementary Fig. 1) revealed that the three *Mycobacterium* species belong to the slow-growing mycobacteria[14]. The tree positions *Candidatus* M. methanotrophicum close to *M.* MAG 3 in a branch with *M. fragae*, *M. noviomagense* and *M. celatum* (Fig. 1a). The branch is also relatively close to the one containing *M. tuberculosis*. *M.* MAG 2 is positioned closer to *M. bohemicum*. Similar results were obtained when the phylogenetic tree was constructed on the basis of 27 monocore gene markers (Supplementary Fig. 2). Although *M.* MAG 3 is closely related to *Candidatus* M. methanotrophicum, the average nucleotide identity value of 84.47% indicates that they are different species (Supplementary Fig. 3). Bioinformatic analyses also revealed that the 16S rRNA gene of *Candidatus* M. methanotrophicum has a unique indel at positions 180–199, which is absent in those of other Mycobacteriaceae (Supplementary Table 2).

Methane metabolism of *Candidatus* M. methanotrophicum

To explore how the genomic potential of the most abundant *Mycobacterium* was expressed under the special growth conditions in the cave, we used the same biofilm samples to extract proteins and analysed them using the MAGs (Table 1). This analysis suggested that *Candidatus*

https://doi.org/10.1038/s41564-022-01252-3

Table 1 | List of 26 most abundant proteins in the cave biofilm as determined by semi-quantitative proteomics. Only proteins expressed in *Candidatus* M. methanotrophicum are shown. Abbreviations: sMMO[1], soluble methane monooxygenase; AdhD, Zn-dependent alcohol dehydrogenase D; FDH, formate dehydrogenase; 6PGDH, 6-phosphogluconate dehydrogenase; H6PS, hexulose-6-phosphate synthetase; PyDH, pyruvate dehydrogenase; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; MftD, ADH-D associated protein; EF-Tu, elongation factor thermo unstable. FyDH is 52 on the list of abundant proteins and not included in this table. LFQ stands for the relative label-free quantification

Protein	Short ID	Peptide count (unique)	Sequence coverage (%)	Molecular weight (kDa)	Average normalized count per pellet	Average LFQ intensity per pellet (log_{10})
sMMO subunit α (MmoX)	Mmeth_04015	53 (53)	71.1	60.43	140.5	9.88
AdhD	Mmeth_02892	26 (26)	61.4	39.97	135.9	9.55
sMMO subunit γ (MmoZ)	Mmeth_04010	28 (28)	82.3	21.35	101.9	9.85
sMMO subunit β (MmoY)	Mmeth_04014	25 (25)	57.5	44.79	88.2	9.79
GroEL	Mmeth_02899	44 (23)	55.5	56.96	86.0	9.46
Transketolase	Mmeth_00187	35 (26)	56.5	75.58	48.2	9.26
FDH	Mmeth_01089	19 (19)	48.5	42.09	40.0	9.04
EF-Tu	Mmeth_03490	20 (4)	57.1	43.68	37.8	8.94
α-Crystallin	Mmeth_03520	16 (16)	79.0	15.93	34.1	10.0
Sulfite reductase	Mmeth_00631	31 (31)	48.4	62.40	27.3	8.60
Transaldolase	Mmeth_00186	20 (20)	56.5	40.01	26.0	8.69
Pyruvate kinase	Mmeth_00880	18 (12)	44.1	50.53	23.8	8.65
H6PS	Mmeth_00838	9 (9)	60.4	20.89	22.6	8.29
GroES	Mmeth_00273	19 (14)	30.1	56.13	22.0	8.66
DoxX family member	Mmeth_04503	16 (16)	45.2	31.21	21.4	8.65
MftD	Mmeth_02893	14 (14)	39.9	41.11	21.3	8.63
ATP synthase subunit α	Mmeth_01783	21 (15)	35.6	59.21	21.0	8.78
Peroxidase	Mmeth_03095	20 (20)	37.4	80.70	20.2	8.39
Fumarate reductase	Mmeth_02646	23 (23)	40.0	69.98	20.0	8.64
MmoC	Mmeth_04011	14 (14)	42.9	38.40	19.7	8.55
MmoB	Mmeth_04013	9 (9)	48.4	17.50	19.6	8.84
Encapsulin	Mmeth_01555	18 (18)	52.5	28.92	17.7	8.59
ATP synthase subunit β	Mmeth_01785	14 (5)	36.2	52.93	16.9	8.54
PyDH E3 subunit	Mmeth_04124	18 (15)	36.3	59.57	16.5	8.42
GAPDH	Mmeth_00172	13 (9)	47.6	35.86	16.0	8.58
6PGDH	Mmeth_02221	12 (12)	27.9	52.54	15.1	8.50

[1]Enzymes in bold are key to CH_4 metabolism.

M. methanotrophicum grows methanotrophically and allowed us to reconstruct the most likely pathways involved in such growth (Fig. 2), as described in the following.

The first step, the conversion of methane to CH_3OH, is most likely catalysed by a soluble methane monooxygenase (sMMO), which is a three-component enzyme capable of breaking the C–H bond in methane and other alkanes or alkenes[15]. This conclusion is supported by the high abundance of the α, β and γ subunits of a di-iron-type sMMO (named MmoX, MmoY and MmoZ), as well as of the accessory components MmoB, MmoC and MmoD, in the biofilm (Fig. 2a–b and Table 1). Moreover, the genes encoding these proteins are exclusively present in the genome of *Candidatus* M. methanotrophicum and not in other known mycobacterial genomes (Fig. 1b,c). Bioinformatic analysis further revealed that the clustering of the *mmo* genes in the genome of *Candidatus* M. methanotrophicum is largely similar to the clustering found in the genomes of *Methylococcus capsulatus* Bath and *Methylosinus trichosporium* OB3b (Extended Data Fig. 2), both of which are known to oxidize methane[16,17]. Also, although the MmoX from *Candidatus* M. methanotrophicum is most closely related to MmoX from *Mycolicibacterium chubuense* NBB4 (Supplementary Figs. 4 and 6), for which methane oxidation has not been demonstrated, it shares several positionally conserved amino acids located within the central region of bacterial multicomponent monooxygenases active on methane (Supplementary Fig. 5). Finally, a notable feature is the presence of a small gene in the sMMO cluster of *Candidatus* M. methanotrophicum. AlphaFold predictions indicate with high confidence that the corresponding protein is folding similarly to MmoD (Supplementary Fig. 7). Possibly, this subunit is required for efficient methane utilization.

Many methanotrophic bacteria use a MxaF-type methanol dehydrogenase to catalyse the conversion of CH_3OH to formaldehyde[18,19]. However, *Candidatus* M. methanotrophicum does not have the so-called *mxa* genes, which are required to make such a canonical MxaF-type methanol dehydrogenase. Instead, our data suggest that this conversion is catalysed by alcohol dehydrogenase D (AdhD; Extended Data Fig. 3), which is the second most abundant protein in the biofilm (Table 1). The closest homologue is a Zn-dependent alcohol dehydrogenase that belongs to the zinc-binding class III-type alcohol dehydrogenases[20–22]. Enzymes of this family act on primary or secondary alcohols with very broad specificity, although they oxidize CH_3OH not as efficiently as ethanol[21]. But since Candidatus M.

https://doi.org/10.1038/s41564-022-01252-3

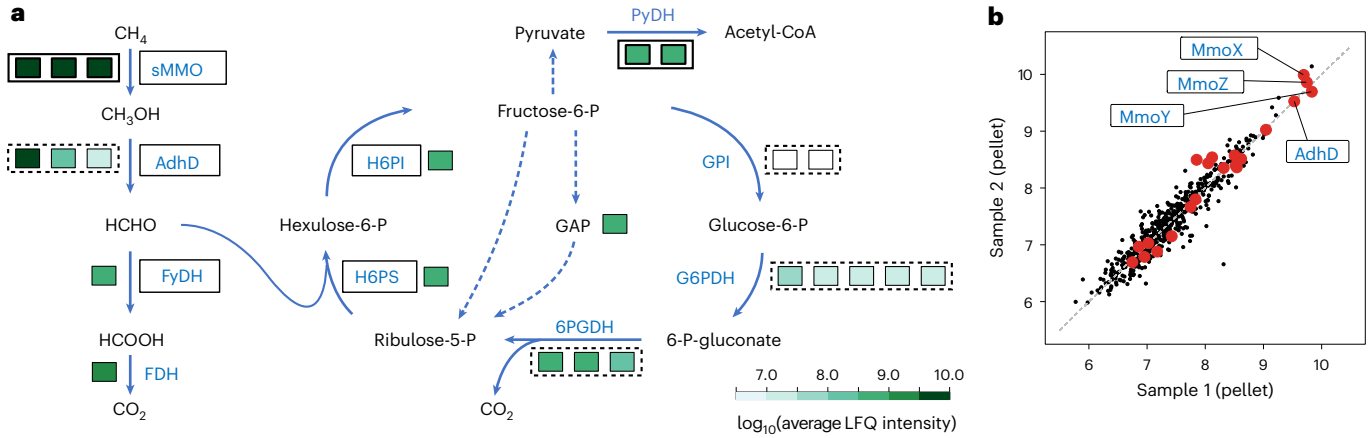


Fig. 2 | Metabolic reconstruction of methane utilization in *Candidatus* M. methanotrophicum. a, Simplified molecular pathways of CH_4 oxidation and the RuMP cycle in *Candidatus* M. methanotrophicum as reconstructed from the metagenomic and proteomic analyses. The metabolic chart is inspired by ref. [73]. In the RuMP cycle, three molecules of ribulose-5-phosphate combine with three molecules of formaldehyde through an aldol condensation, producing three molecules of 3-hexulose-6-phosphate. These are then converted to three molecules of fructose-6-phosphate, one of which is split into GAP and pyruvate. The latter is used for biomass formation, whereas GAP and the other two fructose-6-phosphate molecules are used to regenerate ribulose-5-phosphate[74,75]. Key enzymes that were abundant in the cave biofilm are written in bold. Boxed enzymes are uniquely present in *Candidatus* M. methanotrophicum and not in the other two *Mycobacterium* species found in the biofilm. Dashed arrows represent lumped reactions for biosynthetic purposes and regeneration of ribulose-5-phosphate. Abbreviations (numbers in parentheses indicate the positions on the list of abundant proteins in the biofilm, Table 1): sMMO, soluble methane monooxygenase (1, 3, 4, 20, 21); AdhD, Zn-dependent alcohol dehydrogenase (2); FyDH, formaldehyde dehydrogenase (52); FDH, formate dehydrogenase (7); H6PI, hexulose-6-phosphate isomerase; GPI, glucose phosphate isomerase; G6PDH, glucose-6-phosphate dehydrogenase; 6PGDH, 6-phosphogluconate dehydrogenase (26); H6PS, hexulose-6-phosphate synthetase (13); PyDH, pyruvate dehydrogenase (24); GAP, glyceraldehyde-3-phosphate. The coloured boxes alongside the enzymes show the $\log_{10}$-transformed average of the relative label-free quantification (LFQ) of the proteomics intensities. The coloured boxes are surrounded by a solid-line box when the peptides are subunits of the same enzyme, and by a dashed-line box if the peptides are homologues. White boxes indicate intensities lower than the minimum in the colour scale. **b**, Scatter plot of the $\log_{10}$-transformed LFQ intensities of the proteins derived from two samples of the same biofilm (Pearson correlation coefficient $r = 0.95$, $N = 579$, $P < 2.2 \times 10^{-16}$, two-sided test). Red data points correspond to the enzymes displayed with coloured boxes in **a**.

methanotrophicum is slow-growing (see below), it seems unlikely that CH_3OH oxidation limits the growth rate. The AdhD from *Candidatus* M. methanotrophicum clusters with homologues from Gram-positive bacteria in the phylogenetic tree (Supplementary Fig. 8). Some of these proteins have been purified and shown to function as methanol dehydrogenases[23–25]. Homologous sequences are also present in other members of the Mycobacteriaceae known to be able to grow methylotrophically (Supplementary Figs. 8 and 9).

The final steps in the methane oxidation pathway are most likely catalysed by a formaldehyde and formate dehydrogenase (FyDH and FDH, respectively), which sequentially oxidize formaldehyde via formate to CO_2 (refs. [26,27]). This conclusion is supported by the high abundance of a putative glutathione-independent FyDH and an NAD^+-dependent FDH in the biofilm (Table 1). The gene encoding FyDH is exclusively found in *Candidatus* M. methanotrophicum and not in other known genomes within the Mycobacteriaceae, and its closest homologues are from members of *Aeromicrobium* and *Janibacter* (Supplementary Fig. 10). Furthermore, *Candidatus* M. methanotrophicum seems to use this protein to fully control the partitioning of formaldehyde between assimilation and oxidation, as its genome contains only one gene encoding FyDH and no genes encoding key enzymes of the alternative formaldehyde oxidation pathway involving tetrahydromethanopterin.

Carbon uptake by *Candidatus* M. methanotrophicum is most likely facilitated via the ribulose monophosphate (RuMP) cycle of formaldehyde fixation[28], as indicated by the high abundance of the key enzymes of this pathway, hexulose-6-phosphate synthetase (H6PS) and 6-phosphogluconate dehydrogenase (6PGDH), in the biofilm (Table 1). The RuMP cycle is a widespread pathway used by prokaryotes involved in formaldehyde fixation and detoxification[29]. The genes encoding H6PS and 6PGDH are present only in *Candidatus* M. methanotrophicum and not in the other two mycobacteria found in the biofilm. The enzymes transaldolase and transketolase, which are central in the pentose phosphate pathway[29], are also abundant in the biofilm (Table 1 and Extended Data Fig. 4), indicating that this pathway is highly active in *Candidatus* M. methanotrophicum to generate ribulose-5-phosphate required for the RuMP pathway.

Culturing

In an attempt to culture *Candidatus* M. methanotrophicum, we seeded samples of the cave biofilm in defined medium with a pH of 1.5. The head-space contained a mixture of CO_2 (50%), CH_4 (25%) and air (25%) to mimic the conditions in the cave. After a few months of incubation, we observed some growth in the cultures, in some cases with a small pellicle on top of the medium. Further replenishment with fresh medium ultimately resulted in highly enriched cultures of *Mycobacterium* with a doubling time of around 150–200 days (Fig. 3a). Subsequent sequencing of the almost entire 16S rRNA gene of the cultured cells revealed that its internal V3–V4 region was identical to the amplicon-based sequence obtained from the biofilm. Furthermore, a complete genome sequence of the cultured *Candidatus* M. methanotrophicum was more than 99% identical to *M.* MAG 1 assembled from the metagenome of the biofilm. The difference was most likely due to the lower quality of the DNA extracted from the biofilm.

Proteomic analysis revealed that the cultures were dominated by *Candidatus* M. methanotrophicum (92%) and additionally contained members of the genus *Ferroplasma* (7%) and trace amounts of *Acidothiobacillus* and close relatives of *Thermoplasma* (values correspond to the relative contributions of the proteins detected in the proteome; Extended Data Fig. 5). Detailed inspection of the most abundant proteins from the latter three groups indicated no relationship to methanotrophic or methylotrophic growth. Further analysis of the

https://doi.org/10.1038/s41564-022-01252-3

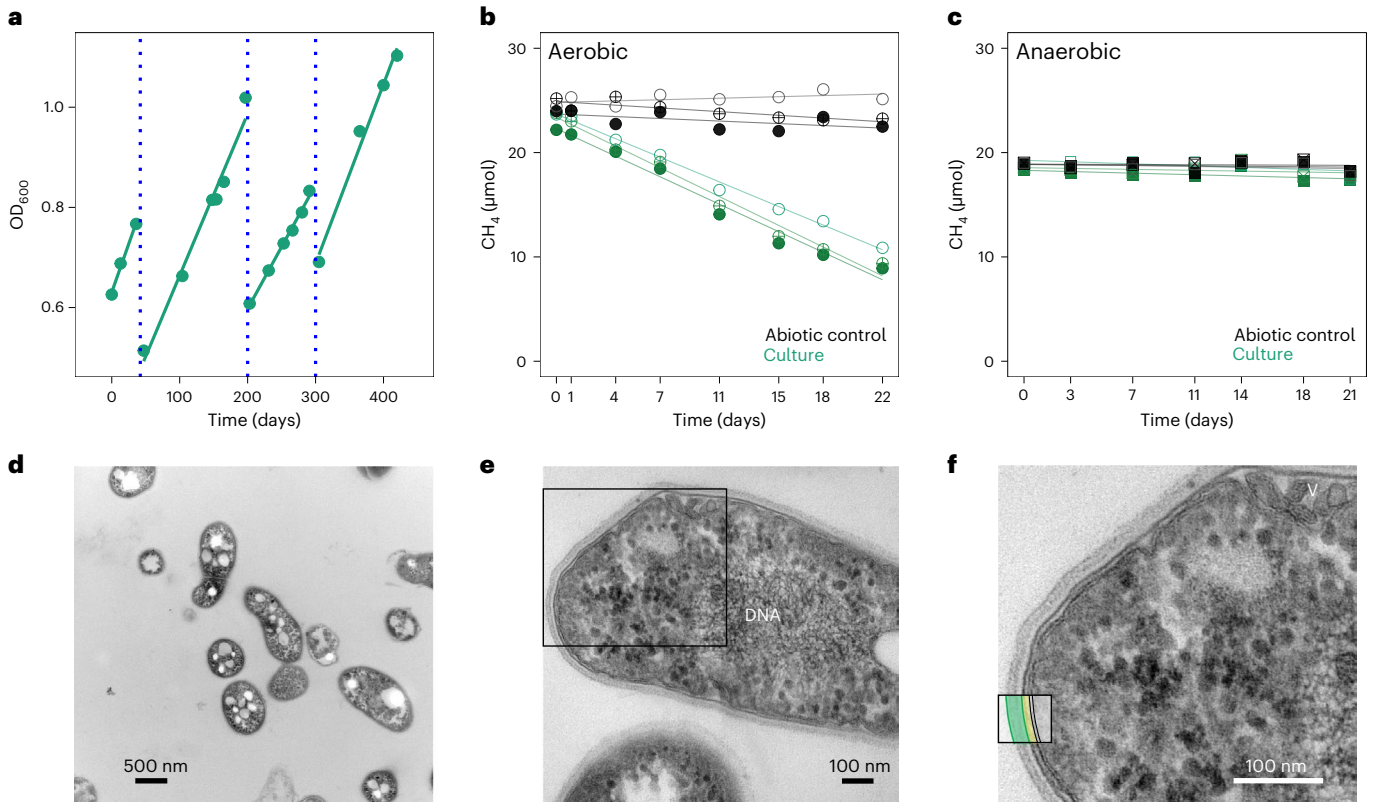


Fig. 3 | Growth curves, methane assimilation and morphology of *Candidatus* M. methanotrophicum. a, Growth curves of *Candidatus* M. methanotrophicum in a highly enriched culture grown on CH_4. Blue dashed lines represent the timepoints where half of the culture was withdrawn and replenished with fresh medium to the original volume. **b**,**c**, Evolution of CH_4 during incubations of *Candidatus* M. methanotrophicum enrichment cultures with CH_4. Incubations were conducted under aerobic (**b**) or anaerobic (**c**) conditions and using live cultures (green symbols) or a sterile NMS medium (black symbols). Different symbols correspond to replicate incubations, and the values correspond to the total amount of CH_4 per incubation bottle. The corresponding CO_2 data are shown in Extended Data Fig. 6c,d. **d**–**f**, Transmission electron micrographs (TEM) of cultured *Candidatus* M. methanotrophicum cells. Shown are representative examples of images obtained from around 100 cells obtained at different levels of resolution and magnification (68,000–150,000×). The boxed area in **e** is shown in **f**. The colour-coded boxed area in **f** shows the capsular layer (green), periplasmic space (yellow) and the plasma membrane (grey). Note the vesicles (V) detected in the periplasmic space. More detailed annotation of the different intracellular structures is provided in Extended Data Fig. 6a,b.

proteome showed that all proteins involved in methane utilization that were produced in the cave biofilm were also highly abundant in the cultures. The only major difference was the high abundance of all ESX-3 secretion system membrane components and two potential ESX-3 substrates in the cultured cells (Supplementary Table 3). As the ESX-3 system is involved in metal transport[30], this difference may indicate a shortage of some metal(s) in our culture medium. In addition to the above genera, the cultures also contained small amounts of fungus from the genus *Acidomyces*, as revealed by microscopy combined with the analysis of the 18S rRNA gene[5].

Morphology

High-resolution electron microscopy revealed that, although the overall morphology of the cultured *Candidatus* M. methanotrophicum cells resembles that of other mycobacteria, there are several differences (Fig. 3d–f). The width of *Candidatus* M. methanotrophicum cells (628 ± 54 nm, based on 47 cells) is almost two-fold greater than the width of *M. tuberculosis* cells (327 ± 55 nm, based on 45 cells). Also, *Candidatus* M. methanotrophicum cells contain a clear capsular layer (29 ± 3 nm thick) (Fig. 3f), which is typically difficult to discern in other mycobacteria (such as *M. tuberculosis*) without using cryo-EM techniques[31]. Finally, the periplasmic space of *Candidatus* M. methanotrophicum contains small vesicles between the outer membrane and the plasma membrane (Fig. 3f and Extended Data Fig. 6a,b).

Oxidation and assimilation of methane

We performed two assays to confirm that *Candidatus* M. methanotrophicum can oxidize CH_4 and use it as a carbon and energy source for growth. The first assay involved incubation of three replicate cultures for 22 days under either aerobic or anaerobic conditions and with 1% CH_4 in the head-space. The results clearly demonstrate that the consortium enriched with *Candidatus* M. methanotrophicum can oxidize CH_4 to CO_2 using O_2 as the terminal electron acceptor but has no such ability under anaerobic conditions (Fig. 3b,c). Under aerobic conditions, the rate of CO_2 production was initially greater than expected on the basis of CH_4 respiration alone (Extended Data Fig. 6c), possibly because small amounts of *Mycobacterium*-derived organics became available and subsequently mineralized by the other community members following the change in the CH_4 fraction in the head-space from 25% during culturing to 1% during the incubation. After about 4 days of incubation, however, the ratio between the amounts of CO_2 produced and CH_4 consumed stabilized at around 0.66 (one replicate culture) and 1 (two replicate cultures) (Fig. 3b and Extended Data Fig. 6c). The variation in the CO_2:CH_4 ratio possibly reflects differences in the efficiency of CH_4 assimilation among the replicate cultures tested in these incubations. Under anaerobic conditions, small amount of CO_2 was still produced despite the absence of measurable CH_4 oxidation (Fig. 3c and Extended Data Fig. 6d). This production was probably due to anaerobic mineralization (for example, via denitrification using nitrate in the medium)

https://doi.org/10.1038/s41564-022-01252-3

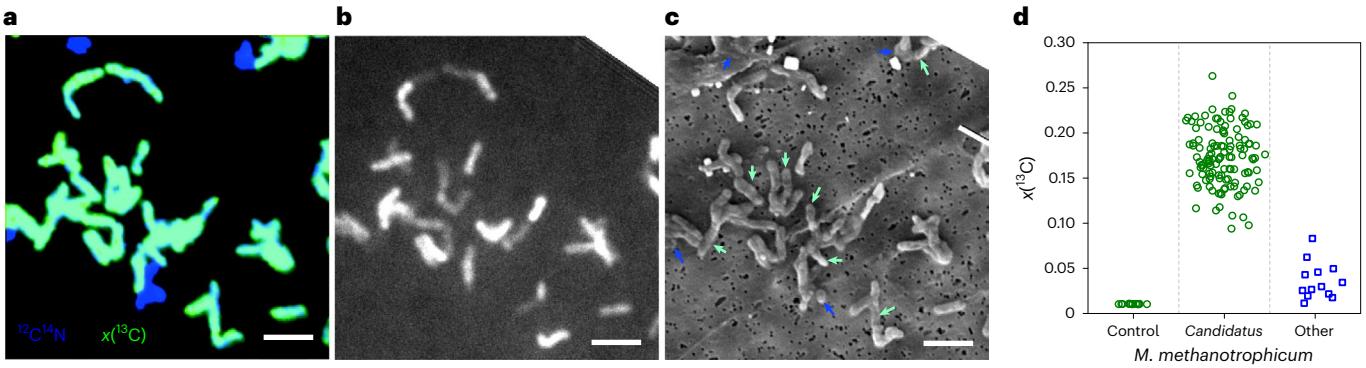


Fig. 4 | Correlative imaging analysis of cells from a *Candidatus* M. methanotrophicum enrichment culture grown on ^{13}C-labelled methane. **a**, The nanoSIMS image shows an overlay between the accumulated counts of $^{12}C^{14}N^-$ (blue), which is a proxy for biomass, and ^{13}C atom fraction (green), which is a measure of methane-derived ^{13}C assimilation. More images are shown in Extended Data Fig. 7a–h. **b**, The fluorescence image shows the intensity of staining with auramine, which is a fluorescent dye that binds specifically to mycolic acids. **c**, The scanning electron microscopy image shows the surface topography. Scale bar in all images is 2 µm. Data in the upper-right corner of images **b** and **c** are missing owing to image rotation. Note the absence of fluorescence from some *Candidatus* M. methanotrophicum cells (marked by cyan arrows in **c**), which is probably due to a high cell-to-cell variability in the efficiency of staining by auramine[76]. Cells marked by blue arrows in **c** probably belong to the other members of the culture community, as suggested by their markedly different morphology and lack of auramine fluorescence. **d**, Cell-specific ^{13}C atom fractions derived from nanoSIMS images. Control cells were incubated with unlabelled methane.

of *Mycobacterium*-derived organics by the other community members during the incubation.

In the second assay, we incubated a culture with ^{13}C-labelled CH_4 (^{13}C atom fraction of 50%) for 110 days and subsequently analysed the ^{13}C labelling of individual cells by nanoscale secondary ion mass spectrometry (nanoSIMS). When this analysis was combined with fluorescence imaging of cells stained with auramine, which is a fluorescent dye that binds specifically to mycolic acids present in the cell wall of mycobacteria[32], the results showed that CH_4-derived carbon was assimilated by *Candidatus* M. methanotrophicum (Fig. 4a,b). Additionally, this analysis revealed the presence of auramine-negative cells with a different morphology compared with that of *Candidatus* M. methanotrophicum (examples marked with blue arrows in Fig. 4c). These cells probably belong to the other members of the culture community. The ^{13}C labelling of these cells was, however, significantly lower than that of *Candidatus* M. methanotrophicum (Fig. 4d), confirming that the ^{13}C labelling of *Candidatus* M. methanotrophicum cells was due to direct assimilation of CH_4-derived carbon rather than transfer from the other community members. When the nanoSIMS analysis was performed on *Candidatus* M. methanotrophicum cells without any post-incubation chemical treatment that could dilute their ^{13}C labelling, the average ^{13}C enrichment of cells (atom fraction of 0.26 ± 0.04) matched well the expected value of 0.247 derived by assuming that the cells grew solely on the provided ^{13}C-labelled CH_4 (Extended Data Fig. 7i,j).

Environmental occurrence of *Candidatus* M. methanotrophicum

To explore whether Sulfur Cave is a unique environment for *Candidatus* M. methanotrophicum, we collected soil and water samples at different locations on Puturosu Mountain (Supplementary Fig. 11), checked for emissions of volcanic gases and the availability of water, measured the pH of the samples and isolated the DNA to determine the community profiles. The results show that *Candidatus* M. methanotrophicum is highly abundant in all types of sample from this area where volcanic gas emissions occur and the pH is around or below 4 (Fig. 5 and Supplementary Table 4). In contrast, *Candidatus* M. methanotrophicum is much less abundant at sites with low or no gas emissions or at sites with a pH above 4. The latter sites also have a higher diversity index (Fig. 5b). Typically, the sampled microbial communities had a highly similar composition if *Candidatus* M. methanotrophicum was abundant (Fig. 5a–c).

To check for other environments where *Candidatus* M. methanotrophicum may be present, we searched the National Center for Biotechnology Information (NCBI) database for the characteristic unique indel in the 16S rRNA gene at positions 180–199. We obtained 52 similar sequences from members of the family Mycobacteriaceae present in diverse environmental samples, including basalts, lava beds, corroded concrete sewer systems and soils with elevated CO_2 (Supplementary Table 2). These data suggest that species closely related to *Candidatus* M. methanotrophicum are widespread all around the world, possibly in low-pH, high-methane environments.

Discussion

Methanotrophy is a metabolic novelty in mycobacteria. Although it has been suggested previously[33,34], the evidence remained inconclusive as it was based solely on the measurement of methane oxidation activity in mycobacterial enrichments and thus lacked specificity. Here we describe the cultivation and identification of a *Mycobacterium* and demonstrate conclusively that it can grow on methane as a sole energy and carbon source. The species originates from a highly acidic biofilm growing on the wall of a cave in Romania, and we propose to call it *Candidatus* M. methanotrophicum. Using methane oxidation assays, we show that a culture enriched in *Candidatus* M. methanotrophicum (92% of the protein content) oxidizes methane to CO_2 under aerobic conditions but not under anaerobic conditions. By combining cultivation on ^{13}C-labelled methane with auramine staining, fluorescence microscopy and quantification of ^{13}C labelling of individual cells by nanoSIMS, we show that *Candidatus* M. methanotrophicum utilizes methane as a sole carbon source for biomass synthesis. Genomic analysis revealed that *Candidatus* M. methanotrophicum is the only species in the family Mycobacteriaceae sequenced so far that possesses a unique combination of genes encoding sMMO, AdhD, FyDH, H6PS and H6PI. This genomic make-up allows *Candidatus* M. methanotrophicum to use CH_4 as an energy source and as a precursor for formaldehyde, which is incorporated into biomass via the RuMP pathway. Finally, proteomic analysis showed that the key enzymes of methane oxidation and of the RuMP pathway are highly abundant in the enrichment cultures as well as in the cave biofilm, indicating that *Candidatus* M. methanotrophicum grows methanotrophically under both laboratory and natural conditions.

An important unresolved question is why *Candidatus* M. methanotrophicum is abundant in the cave biofilm instead of being

https://doi.org/10.1038/s41564-022-01252-3

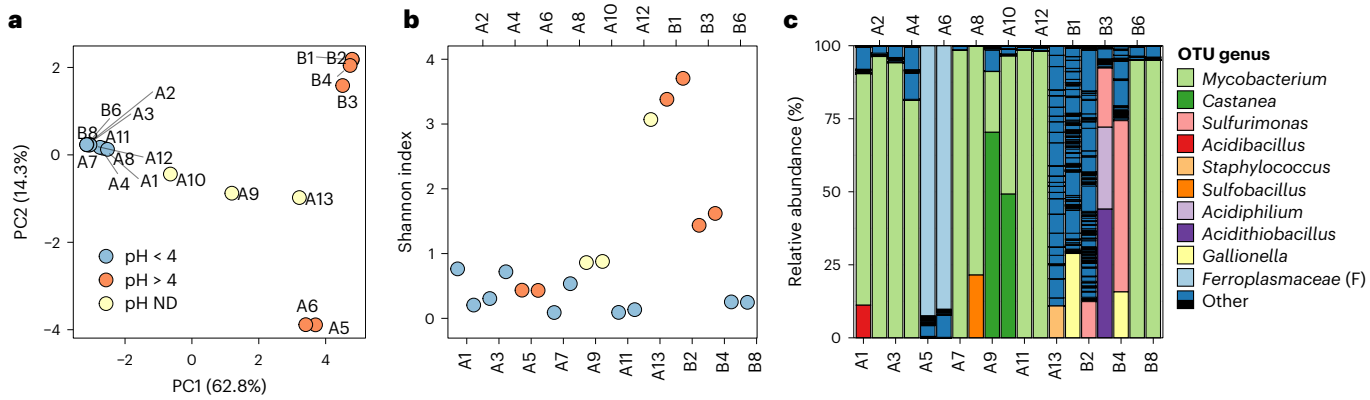


Fig. 5 | Microbial community characteristics based on amplicon-based sequencing of 16S rRNA genes in samples from Puturosu Mountain. **a**, Principal component (PC) analysis plot of Bray–Curtis dissimilarity. **b**, Shannon index of operational taxonomic unit (OTU) abundances. **c**, Relative abundances of genera in the samples. Sample numbers and their locations are provided in Supplementary Table 4. The genera with a relative abundance of more than 10% are shown in **c**, and the rest is merged as 'others' and coloured dark blue or black. ND, not determined. F, genus from the family Ferroplasmaceae.

outcompeted by other types of methanotroph. We hypothesize that this is due to its ability to withstand the extremely low pH. Members of the Mycobacteriaceae have a special membrane with mycolic acids, which makes the cells more robust and protects them against environmental stresses[35]. Our electron microscopy images of *Candidatus* M. methanotrophicum revealed a cell wall with a relatively thick capsular layer, which could play a role in the protection against low pH. Additionally, members of the abundant archaeal genera identified in the biofilm (*Ferroplasma* and *Thermoplasma*) are also known to possess membrane fortification systems (for example, tetraether lipids that form a membrane monolayer) that help them cope with harsh acidic conditions[35]. Finally, our hypothesis is supported by the distribution patterns of mycobacteria in the natural environment. Although we have found only one publication describing mycobacteria from an extremely acidic environment (a *M. parascrofulaceum* strain from a hot spring with pH 3 in Yellowstone National Park[36]), our sampling in the proximity of Sulfur Cave and around Puturosu Mountain consistently revealed very high *Mycobacterium* abundances at strongly acidic sites and much lower abundances at sites with a higher pH. Also in soils, where mycobacteria are common, albeit at relatively low levels (between 0.03% and 3.0% of all 16S rRNA gene reads), their abundance tends to be higher under more acidic conditions[9].

Future studies are required to elucidate the role of *Candidatus* M. methanotrophicum in the Sulfur Cave biofilm and other habitats. We hypothesize that this species is the primary producer of methane-derived cellular biomass in the Sulfur Cave biofilm and that other members of the biofilm community profit from it, most likely by feeding on necromass containing carbohydrates, peptides and fatty acids as secondary sources of carbon and energy[37]. A survey of databases revealed that close relatives of *Candidatus* M. methanotrophicum are widespread worldwide at sites of volcanic origin or acidic niches with methane formation and might thus play a globally important role in mitigating CH_4 emissions from such environments[38,39].

Methods

Biofilm sampling in Sulfur Cave

Biofilm samples were collected from the wall of Sulfur Cave in Romania (46.1° N, 26.0° E; Extended Data Fig. 1) with sterile cotton swabs and stored in sterile Nunc cryotubes. The pH of the samples was determined on-site with pH indicator paper 1–14 (90224, Macherey-Nagel). To preserve the biological samples intended for growth experiments, a few drops of sterile Milli-Q water were added before transport. Samples intended for proteomics and metagenomics were kept cold with cooling elements. After transport, saline solution was added to suspend the cells from the biofilm material and remove larger particles such as rock grit. Afterwards, the cell suspension was split into two halves, with one half used for metagenomic and the other half for proteomic analysis.

Proteomic analysis

Suspended biomass of the cave biofilm or enrichment cultures was washed and lysed using a bead beater and used to prepare protein fractions for proteomics[40]. Both soluble and insoluble proteins were analysed. All tandem mass spectra were analysed using MaxQuant version 1.6.10 (ref. [41]) and searched against the concatenated protein database containing entries from both the crude and cave samples. Peak list generation of label-free tandem mass spectra was performed within MaxQuant using default parameters and the built-in Andromeda search engine. Enzyme specificity was set to consider fully tryptic peptides with the allowance of two missed cleavages. Oxidation of methionine and N-terminal acetylation were allowed as variable modifications, whereas carbamidomethylation of cysteine was allowed as a fixed modification. A protein and peptide false discovery rate of less than 1% was employed in MaxQuant with matches between runs enabled. Proteins that contained similar peptides that could not be differentiated on the basis of tandem mass spectrometry analysis alone were grouped to satisfy the principles of parsimony. Reverse database hits, contaminants and proteins identified only by site modifications were omitted in the subsequent data analysis.

DNA isolation, amplification and sequencing of the cave biofilm

Samples were collected from the biofilm in Sulfur Cave (Extended Data Fig. 1) as well as from diverse sites on Puturosu Mountain (Supplementary Table 4 and Supplementary Fig. 11). DNA for metagenomics of the biofilm samples was extracted using the DNeasy PowerSoil Kit (Qiagen, Benelux BV). DNA for Illumina sequencing-based community profiling of the sites on Puturosu Mountain was extracted using the ZymoBIOMICS DNA Miniprep Kit (Zymo Research). Amplification of the 16S rRNA gene was carried out with universal primers 8F (5′-AGAGTTTGATYMTGGCTCAG-3′) and 1512R (5′-ACGGYTACCTTGTTACGACTT-3′) (ref. [42]). Reactions were performed in a Thermocycler (Biometra, Analytik Jena). PCR products were purified and sequenced by Macrogen Europe. These sequences were compared with known 16S rRNA gene sequences deposited in the GenBank database using the BLAST search at the NCBI (http://www.ncbi.nlm.gov/BLAST). For PCR amplification, the DNA of the samples was diluted to 0.01 ng μl^{-1} and distributed in a 96-well plate for PCR amplification (V3–V4 region of the 16S rRNA gene). The reaction in

each well was done using 10 mM dNTP mix, 2 U μl^{-1} Phusion Hot Start II High-Fidelity DNA Polymerase (Thermo Fisher Scientific), 0.1 ng DNA template and 24 primers (8 forward primers; 16 reverse primers) that contain Illumina adapters and an eight-nucleotide index barcode sequence[43]. Thirty-five-cycle amplification reactions were completed with 10 s at 98 °C (denaturation), 30 s at 55 °C (annealing), and 30 s at 72 °C (elongation). The PCR products were pooled calculating an equimolar mix for paired-end sequencing of amplicons on the Illumina MiSeq platform (Illumina)[44].

Metagenomic analyses

Crude genomic DNA was subjected to DNA library preparation using Ovation Ultralow System V2 according to the manufacturer's instructions and sequenced on a HiSeq4000 platform. The raw reads were filtered with BBDuk (http://jgi.doe.gov/data-and-tools/bb-tools/) to remove Illumina adapters, phiX and low-quality bases from both ends to Q30. The reads cleaned in this way were then assembled with MEGAHIT assembler[45], and the resulting assemblies were binned using gbtools[46] and annotated with Prokka[47] and eggNOG[48]. The contamination level and completeness of the assemblies were examined by CheckM[49]. An overview of the statistics of the assemblage of the three mycobacterial MAGs is presented in Supplementary Table 1, and their corresponding KEGG orthologue profiles were compared with those from reference genomes of *M. tuberculosis* and *M. gordonae* (Supplementary Fig. 12). The MAG relative abundance was calculated using CoverM (v0.6.1) (https://github.com/wwood/CoverM).

Complete genome of *Candidatus* M. methanotrophicum

DNA for genomic analysis of the cultured *Candidatus* M. methanotrophicum cells was extracted using the DNeasy PowerSoil Kit (Qiagen, Benelux BV) according to the manufacturer's instructions. An Illumina library was prepared using the Ovation Ultralow System V2 according to the manufacturer's instructions and sequenced in the PE 2 × 150 bp mode on a HiSeq4000 platform. The resulting reads were trimmed using Skewer's PE mode[50], resulting in 1.6 Gbp of trimmed data. In addition, genomic DNA from the enrichment culture was isolated using the High Pure PCR Template preparation kit (Roche) and sequenced with the Oxford Nanopore MinION technology (FLO-MIN106) at Utrecht Sequencing Facility (USEQ), resulting in 6.3 Gbp of sequences with a N50 length of 11.5 kbp and an average length of 5.9 kbp.

The Illumina and Nanopore data were assembled together using the hybrid approach implemented in the program Unicycler (version 0.5.0) (ref. [51]). The genome of *Candidatus* M. methanotrophicum comprises a 4,719,641 bp circular chromosome and a 282,730 bp linear plasmid. Overall, the complete genome of *Candidatus* M. methanotrophicum and *M.* MAG 1 had 5,194 and 4,580 open reading frames, respectively. Using a 98% Cluster Database at High Identity with Tolerance (CD-HIT) identity threshold, 5,324 orthologous clusters were identified, 4,264 of which were made up of one orthologue in each genome. In contrast, 727 single-gene clusters were present in *Candidatus* M. methanotrophicum but not in *M.* MAG 1, whereas 277 single-gene clusters were found in *M.* MAG 1 but not in *Candidatus* M. methanotrophicum. The remaining 56 orthologous clusters had more than 2 genes and comprised a total of 242 genes. The sMMO gene cluster was present in both *M.* MAG 1 and the complete genome of *Candidatus* M. methanotrophicum.

Structural protein predictions

Template-free structural predictions were generated using AlphaFold[52,53]. Protein complex prediction was done using AlphaFold Multimer[54]. Molecular visualization was done in ChimeraX (version 1.255) (ref. [55]).

Phylogenomics

The three cave mycobacterial MAGs together with 69 publicly available *Mycobacterium* genomes sequences[14] were used as input for the phylogenomic analyses. The average nucleotide identity between all the genome pairs was calculated with fastANI[56] (Supplementary Fig. 3). CD-HIT on 70% and 90% amino acid identity thresholds along with a custom-made R script were used to identify two sets of monocore gene marker sets, one with 316 and the other with 27 marker genes, respectively[57,58]. The trees were constructed using PhyloPhlAn (version 3.0) (ref. [59]) using the approach described in a recent phylogenetic study[14]. The PhyloPhlAn parameters '–accurate' and '–diversity low' were applied, which translate to the usage of a pfasum60 substitution matrix. Trimming was performed by removing columns with at least one nucleotide appearing above a threshold of 0.99. The multiple sequence alignments (MSAs) were checked and cleaned from fragmentary entries. The program Diamond was used for sequence alignments[60], MUSCLE to perform MSAs[61], trimAl for further phylogenetic analyses[62], fasttreeMP to infer approximately-maximum-likelihood phylogenetic trees[63], and the final best tree topologies were inferred using RAxML[64]. The comparison of the two trees was performed with the R package dendextend[65].

For the phylogenetic analyses of MmoX, MmoB, AdhD and FdhA, the MSAs were performed using the mafft[66] program and by selecting the parameter '–auto'. The corresponding trees were produced using the RAxML program and by selecting the parameter 'PROTGAMMAAUTO', which automatically searches the best protein model. Also, a 10,000 bootstrap value was selected for all trees[64]. The unique genes of the members in the *Mycobacterium* tree were identified using a custom-made R script on the 40% CD-HIT orthologous groups and visualized using R. The gene clusters encoding sMMO were identified using a custom-made R script on the 40% CD-HIT orthologous groups and the corresponding .gff files[57]. The visualization was performed using R and the R package gggenes[58]. The trees were visualized with iTOL and ggtree[67,68]. For all trees, the internal tree scale was visualized on the basis of amino acid sequence similarity.

Culturing

In April 2018, liquid cultures were started from freshly collected cave biofilm samples. A 15 ml tube was filled with 2 ml of PBS (8 g l^{-1} NaCl; 0.2 g l^{-1} KCl; 1.44 g l^{-1} $Na_2PO_4{\cdot}7H_2O$; 0.24 g l^{-1} KH_2PO_4, pH 7.4) after which a small amount of the biofilm was used as inoculum. This biofilm suspension was shaken overnight at 25 °C in the dark to detach the bacteria as much as possible. Subsequently, 50 µl of the cell suspension was added to 3 ml of nitrogen mineral salts medium (NMS; 1 g l^{-1} KNO_3; 0.54 g l^{-1} KH_2PO_4; 399.36 mg l^{-1} $MgSO_4{\cdot}7H_2O$; 15 mg l^{-1} $CaCl_2{\cdot}2H_2O$; 1 ml l^{-1} trace elements stock solution containing: 5 g l^{-1} EDTA; 2 g l^{-1} $FeSO_4{\cdot}7H_2O$; 0.1 g l^{-1} $ZnSO_4{\cdot}7H_2O$; 0.03 g l^{-1} $MnCl_2{\cdot}4H_2O$; 0.2 g l^{-1} $CoCl_2{\cdot}6H_2O$; 0.1 g l^{-1} $CuCl_2{\cdot}5H_2O$; 0.02 g l^{-1} $NiCl_2{\cdot}6H_2O$; and 0.03 g l^{-1} Na_2MoO_4). The pH was set at 1.5 using 1.0 M H_2SO_4. Petri dishes (Greiner, catalogue number 627161) with 3 ml NMS and 50 µl of a biofilm suspension were incubated at 25 °C in the dark in custom-built chambers[69] with a total volume of 200 ml containing 50% CO_2, 25% CH_4 and 25% air in the head-space. After consistent growth of the primary culture was established in March 2019, a second culture was started form the primary culture. From May 2019 onwards, the optical density at 600 nm (OD_{600}) was measured to assess the growth rate of the cultures. Phase contrast images were taken during the incubations using a ZEISS AXIO imager M1 microscope.

Various attempts to further purify and upscale cultivation were not successful. The only strategy that ensured stable growing cultures was by growing 3 ml cultures as a thin liquid layer in Petri dishes until OD_{600} exceeded 1.0 and subsequently replacing part of the culture volume by the NMS medium such that OD_{600} does not decrease below 0.6. The replaced culture volume was than used to start a subculture, which could subsequently be used for testing or experimentation. This approach could be classified as 'semi fed-batch' cultivation with a partial replacement of the culture by fresh medium. Attempts to remove or inactivate the fungus, either by filtration or by using the

antibiotic natamycin, failed as they always led to cessation of growth of the *Mycobacterium* cells.

Electron microscopy

Cultured cells of *Candidatus* M. methanotrophicum, *M. tuberculosis* Mtb72 ATCC 35801 and *M. smegmatis* were imaged by electron microscopy. The latter two species, used as controls, were cultured until OD_{600} = 0.1 in Middlebrook 7H9 medium, supplemented with oleic albumin dextrose catalase and glycerol. All cultures were fixed in phosphate buffer containing a mixture of 4% paraformaldehyde and 1% glutaraldehyde. Subsequently, they were post-fixed in 1% osmium tetroxide, dehydrated through an ethanol graded series of 70–100%, impregnated with epon–ethanol (1:1) and embedded in an Embed 812 resin (Electron Microscopy Sciences). Finally, ultrathin sections (50–70 nm) were cut with a diamond knife and transferred to copper grids for analysis by electron microscopy (Tecnai 12, FEI with a VELETA and Xarosa camera).

Methane oxidation and assimilation assays

The primary culture started in April 2018 was maintained and propagated in the NMS medium (pH 1.5) as described above (section 'Culturing'). In the period from April 2018 to October 2021, the culture was diluted with a fresh NMS medium six times. On 5 October 2021, a subculture with $OD_{600} \approx 0.9$ was used to start new cultures, which were grown in parallel using either ^{13}C-labelled (^{13}C atom fraction of 50%) or unlabelled CH_4 in the head-space. The ^{13}C-labelled CH_4 was prepared by mixing equal amounts of 99.9% $^{13}CH_4$ (Campro Scientific) with unlabelled CH_4. This cultivation continued until OD_{600} reached 0.95 (110 days). During this time, the head-space in the cultures was replaced three times to maintain relatively stable levels of CH_4 (volume fraction of 25%). Small samples collected at the end of these incubations were analysed by nanoSIMS to quantify the ^{13}C labelling of individual cells (see below). The remaining biomass of the cultures grown on ^{13}C-labelled CH_4 was split into three equal parts (each with a volume of 2.5 ml and $OD_{600} \approx 0.7$) and subsequently used to assay CH_4 consumption and CO_2 production.

The CH_4 consumption and CO_2 production assays were done in 60 ml crimp cap bottles closed using washed and boiled butyl rubber stoppers. The head-space was initially a mixture of air and CH_4. The initial volume fraction of CH_4 in the head-space was decreased from 25% (used during the preceding cultivation) to 1% to avoid technical complications with the measurement of high CH_4 concentrations. Incubations were conducted at 25 °C in the dark and continued for 22 days, during which the head-space was regularly sampled (0.1 ml) for CH_4 and CO_2 quantification (see below). Afterwards, the head-space was flushed with N_2 gas for 2 h to remove any residual oxygen, CH_4 was injected to reach the final volume fraction of 1% and the incubations continued for another 22 days under anaerobic conditions. Incubations of bottles containing only the sterile NMS medium were conducted in parallel and used as abiotic controls.

CH_4 and CO_2 quantification

Methane concentration in the head-space was quantified by injecting 0.1 ml of head-space into a gas chromatograph (Ultra GC, Interscience) equipped with a flame ionization detector and an Rt-Q-BOND capillary column (length 30 m, inner diameter 0.32 mm; Restek, Interscience). Helium was used as a carrier gas, and oven temperature was set at 80 °C. Head-space CO_2 concentration was measured in the same injection by an flame ionization detector after CO_2 was converted into CH_4 by a methanizer. Chromeleon Chromatography Data System 7.1 (CDS, Thermo Fisher Scientific) software was used to analyse the obtained gas chromatograms. The amounts (µmol) of CH_4 and CO_2 in the bottle were calculated from the measured concentrations and known volume of the head-space.

Fluorescence and nanoSIMS imaging

To link the identity of cells with their ability to assimilate CH_4-derived carbon, we first stained the cells grown on ^{13}C-labelled CH_4 with auramine and then analysed them by fluorescence microscopy and nanoSIMS. Auramine was used because it binds specifically to mycolic acids[32] and can be visualized through green fluorescence emitted upon excitation with UV light. The staining was done by incubating the collected cells in a filtered phenolic auramine staining solution (Merck Sigma) for 30 min. Cells were subsequently collected by centrifugation, washed with an acidified methanol solution and PBS, and deposited on a polycarbonate filter (2.5 cm diameter, 0.2 mm pore size, Millipore). Fluorescence images were acquired with a Nikon Eclipse Ti microscope. During this step, each imaged region was marked using a focused UV-light beam, which made it possible to find the same region back during the nanoSIMS analysis. This correlative imaging analysis was performed on eight fields of view yielding values for a total of 18 and 137 cells from the incubation with unlabelled and ^{13}C-labelled methane, respectively (the latter number split between 124 cells of *Candidatus* M. methanotrophicum and 13 cells of the other members in the enriched culture; data shown in Fig. 4 and Extended Data Fig. 7a–h).

In addition to the nanoSIMS analysis of auramine-stained cells, we also analysed cells that were deposited only on polycarbonate filters and air dried for 2 days immediately after the incubations with ^{13}C-labelled or unlabelled methane. This analytical approach was chosen to avoid isotope dilution effects, which can occur owing to post-incubation chemical procedures such as fixation or staining[70]. Cells analysed in this way originated from an independent incubation experiment, and the analysis was performed on nine fields of view yielding data for 40 and 110 cells of *Candidatus* M. methanotrophicum incubated with unlabelled and ^{13}C-labelled methane, respectively (summary of data shown in Extended Data Fig. 7j).

nanoSIMS measurements were performed with the NanoSIMS 50L instrument (Cameca) operated at Utrecht University. Before the nanoSIMS analysis, the filters with cells were sputter coated with a 10 nm gold layer, cut into small circular pieces (5 mm diameter) to make them suitable for the nanoSIMS sample holder, and imaged with a Neoscope II JCM-6000 scanning electron microscope (JEOL) to check the sample quality in terms of cell integrity or cell density. During the analysis, areas of interest were first pre-sputtered with Cs^+ ions until secondary ion yields stabilized. Subsequently, the primary Cs^+-ion beam was scanned over the sample (areas between 5 µm × 5 µm and 40 µm × 40 µm in size, dwell time of 1 ms per pixel) while detecting secondary ions $^{12}C^-$, $^{13}C^-$, $^{16}O^-$, $^{12}C^{14}N^-$, $^{31}P^-$ and $^{32}S^-$. To increase the overall signal, the same area was imaged multiple times, and the resulting ion count images were aligned and accumulated. Data were processed with the Look@NanoSIMS software[71]. Regions of interest corresponding to individual cells or cell clusters were drawn manually using the counts of $^{12}C^{14}N^-$ as a proxy for cell biomass. The ^{13}C atom fractions were calculated as $x(^{13}C) = {}^{13}C^-/({}^{12}C^- + {}^{13}C^-)$ using the total counts of $^{12}C^-$ and $^{13}C^-$ accumulated over the region of interest pixels.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Raw data from the proteomic analysis of the Sulfur Cave biofilm, annotation files and supplementary data can be found in a Zenodo repository[72]. Illumina 16S rRNA gene amplicon data from the Sulfur Cave biofilm are associated with the NCBI BioProject PRJNA675490. Illumina reads from the sequencing of the two metagenomes from the cave and one from the culture as well as the MAGs *M.* MAGs 1, 2 and 3 are deposited at ENA (PRJEB45004) with the accession numbers ERR10036468, ERR10036469, ERR10036470, ERS6581338, ERS6581340

and ERS6581341, respectively and accession CAJUXY010000000. The 16S rRNA gene sequence obtained from the *Candidatus* M. methanotrophicum culture is available at GenBank (MW243585). The Illumina and Nanopore reads from this culture as well as the resulting whole-genome assembly are available at NCBI under BioProject PRJNA837300.

References

1. Martins, Z. et al. Earth as a tool for astrobiology—a European perspective. *Space Sci. Rev.* **209**, 43–81 (2017).
2. Vaselli, O. et al. A geochemical traverse across the Eastern Carpathians (Romania): constraints on the origin and evolution of the mineral water and gas discharges. *Chem. Geol.* **182**, 637–654 (2002).
3. Frunzeti, N., Baciu, C., Etiope, G. & Pfanz, H. Geogenic emission of methane and carbon dioxide at Beciu mud volcano, (Berca-Arbănaşi hydrocarbon-bearing structure, Eastern Carpathians, Romania). *Carpathian J. Earth Environ. Sci.* **7**, 159–166 (2012).
4. Althaus, T., Niedermann, S. & Erzinger, J. Noble gas studies of fluids and gas exhalations in the East Carpathians, Romania. *Chem. Erde Geochem.* **60**, 189–207 (2000).
5. Sarbu, S. M. et al. Sulfur Cave (Romania), an extreme environment with microbial mats in a CO_2–H_2s/O_2 gas chemocline dominated by mycobacteria. *Int. J. Speleol.* **47**, 173–187 (2018).
6. Jones, D. S., Schaperdoth, I. & Macalady, J. L. Metagenomic evidence for sulfide oxidation in extremely acidic cave biofilms. *Geomicrobiol. J.* **31**, 194–204 (2014).
7. Cosma, C. L., Sherman, D. R. & Ramakrishnan, L. The secret lives of the pathogenic mycobacteria. *Annu. Rev. Microbiol.* **57**, 641–676 (2003).
8. Cook, G. M. et al. Physiology of Mycobacteria. *Advances in Microbial Physiology* 81–182. (2009).
9. Walsh, C. M., Gebert, M. J., Delgado-Baquerizo, M., Maestre, F. T. & Fierer, N.A global survey of mycobacterial diversity in soil. *Appl. Environ. Microbiol.* https://doi.org/10.1128/aem.01180-19 (2019).
10. Khan, A. & Sarkar, D. Nitrate reduction pathways in mycobacteria and their implications during latency. *Microbiology* **158**, 301–307 (2012).
11. Brezna, B., Khan, A. A. & Cerniglia, C. E. Molecular characterization of dioxygenases from polycyclic aromatic hydrocarbon-degrading *Mycobacterium* spp. *FEMS Microbiol. Lett.* **223**, 177–183 (2003).
12. Cook, G. M., Hards, K., Vilchèze, C., Hartman, T. & Berney, M. Energetics of respiration and oxidative phosphorylation in mycobacteria. *Microbiol. Spectr.* https://doi.org/10.1128/microbiolspec.mgm2-0015-2013 (2014).
13. Guerrero-Cruz, S. et al. Methanotrophs: discoveries, environmental relevance, and a perspective on current and future applications. *Front. Microbiol.* https://doi.org/10.3389/fmicb.2021.678057 (2021).
14. Gupta, R. S., Lo, B. & Son, J. Phylogenomics and comparative genomic studies robustly support division of the genus *Mycobacterium* into an emended genus *Mycobacterium* and four novel genera. *Front. Microbiol.* https://doi.org/10.3389/fmicb.2018.00067 (2018).
15. Colby, J., Stirling, D. I. & Dalton, H. The soluble methane mono-oxygenase of *Methylococcus capsulatus* (Bath). Its ability to oxygenate *n*-alkanes, *n*-alkenes, ethers, and alicyclic, aromatic and heterocyclic compounds. *Biochem. J.* **165**, 395–402 (1977).
16. Stainthorpe, A., Lees, V., Salmond, G. P., Dalton, H. & Murrell, J. The methane monooxygenase gene cluster of *Methylococcus capsulatus* (Bath). *Gene* **91**, 27–34 (1990).
17. Cardy, D. L. N., Laidler, V., Salmond, G. P. C. & Murrell, J. C. The methane monooxygenase gene cluster of *Methylosinus trichosporium*: cloning and sequencing of the *mmoc* gene. *Arch. Microbiol.* **156**, 477–483 (1991).
18. Chistoserdova, L., Chen, S.-W., Lapidus, A. & Lidstrom, M. E. Methylotrophy in *Methylobacterium extorquens* AM1 from a genomic point of view. *J. Bacteriol.* **185**, 2980–2987 (2003).
19. Harms, N. & van Spanning, R. J. M. C1 metabolism in *Paracoccus denitrificans*: genetics of *Paracoccus denitrificans*. *J. Bioenerg. Biomembr.* **23**, 187–210 (1991).
20. Norin, A., Piersma, S. R., Duine, J. A. & Jörnvall, H. Nicotinoprotein (NAD^+-containing) alcohol dehydrogenase: structural relationships and functional interpretations. *Cell. Mol. Life Sci.* **60**, 999–1006 (2003).
21. Schenkels, P. & Duine, J. A. Nicotinoprotein (NADH-containing) alcohol dehydrogenase from *Rhodococcus erythropolis* DSM 1069: an efficient catalyst for coenzyme-independent oxidation of a broad spectrum of alcohols and the interconversion of alcohols and aldehydes the EMBL accession number for the sequence reported in this paper is p81747. *Microbiology* **146**, 775–785 (2000).
22. Haft, D. H. Bioinformatic evidence for a widely distributed, ribosomally produced electron carrier precursor, its maturation proteins, and its nicotinoprotein redox partners. *BMC Genomics* https://doi.org/10.1186/1471-2164-12-21 (2011).
23. Dubey, A. A., Wani, S. R. & Jain, V. Methylotrophy in mycobacteria: dissection of the methanol metabolism pathway in *Mycobacterium smegmatis*. *J. Bacteriol.* https://doi.org/10.1128/jb.00288-18 (2018).
24. de Vries, G. E., Arfman, N., Terpstra, P. & Dijkhuizen, L. Cloning, expression, and sequence analysis of the *Bacillus methanolicus* C1 methanol dehydrogenase gene. *J. Bacteriol.* **174**, 5346–5353 (1992).
25. Bystrykh, L. V. et al. Electron microscopic analysis and structural characterization of novel NADP(H)-containing methanol: *N*, *N*′-dimethyl-4-nitrosoaniline oxidoreductases from the Gram-positive methylotrophic bacteria *Amycolatopsis methanolica* and *Mycobacterium gastri* MB19. *J. Bacteriol.* **175**, 1814–1822 (1993).
26. Diab, F. Succinate-mediated catabolite repression control on the production of glycine betaine catabolic enzymes in *Pseudomonas aeruginosa* PAO1 under low and elevated salinities. *Microbiology* **152**, 1395–1406 (2006).
27. Quayle, J. in *Carbohydrate Metabolism* (ed. Shukla, A. K.) 360–364 (Elsevier, 1966).
28. Quayle, J. R. Microbial assimilation of C1 compounds. *Biochem. Soc. Trans.* **8**, 1–10 (1980).
29. Kato, N., Yurimoto, H. & Thauer, R. K. The physiological role of the ribulose monophosphate pathway in bacteria and archaea. *Biosci. Biotechnol. Biochem.* **70**, 10–21 (2006).
30. Serafini, A., Pisu, D., Palù, G., Rodriguez, G. M. & Manganelli, R. The ESX-3 secretion system is necessary for iron and zinc homeostasis in mycobacterium tuberculosis. *PLoS ONE* **8**, e78351 (2013).
31. Sani, M. et al. Direct visualization by cryo-EM of the mycobacterial capsular layer: a labile structure containing ESX-1-secreted proteins. *PLoS Pathog.* **6**, e1000794 (2010).
32. Howell, D. G. & McDaniel, H. A. Fluorescent staining of mycobacteria in bovine tissues with auramine O dye—a comparative evaluation of a modified staining procedure. *Proc. Annu. Meet. U. S. Anim. Health Assoc.* **71**, 500–508 (1967).
33. Reed, W. M. & Dugan, P. R. Isolation and characterization of the facultative methylotroph *Mycobacterium* ID-y. *Microbiology* **133**, 1389–1395 (1987).
34. Kambara, H. et al. Environmental factors affecting the community of methane-oxidizing bacteria. *Microbes Environ.* **37**, n/a (2022). 10.1264/jsme2.ME21074

35. Draper, P. The outer parts of the mycobacterial envelope as permeability barriers. *Front. Biosci.* **3**, d1253–1261 (1998).
36. Santos, R., Fernandes, J., Fernandes, N., Oliveira, F. & Cadete, M. *Mycobacterium parascrofulaceum* in acidic hot springs in Yellowstone National Park. *Appl. Environ. Microbiol.* **73**, 5071–5073 (2007).
37. Melkonian, C. et al. High biodiversity in a benzene-degrading nitrate-reducing culture is sustained by a few primary consumers. *Commun. Biol.* https://doi.org/10.1038/s42003-021-01948-y (2021).
38. Cavicchioli, R. et al. Scientists' warning to humanity: microorganisms and climate change. *Nat. Rev. Microbiol.* **17**, 569–586 (2019).
39. Pratt, C. & Tate, K. Mitigating methane: emerging technologies to combat climate change's second leading contributor. *Environ. Sci. Technol.* **52**, 6084–6097 (2018).
40. Phan, T. H. et al. EspH is a hypervirulence factor for mycobacterium marinum and essential for the secretion of the ESX-1 substrates EspE and EspF. *PLoS Pathog.* **14**, e1007247 (2018).
41. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.***11**, 2301–2319 (2016).
42. Weisburg, W. G., Barns, S. M., Pelletier, D. A. & Lane, D. J. 16S ribosomal DNA amplification for phylogenetic study. *J. Bacteriol.* **173**, 697–703 (1991).
43. Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq illumina sequencing platform. *Appl. Environ. Microbiol.* **79**, 5112–5120 (2013).
44. Iturbe-Espinoza, P. et al. Effects of DNA preservation solution and DNA extraction methods on microbial community profiling of soil. *Folia Microbiol.* https://doi.org/10.1007/s12223-021-00866-0 (2021).
45. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via Succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
46. Seah, B. K. B. & Gruber-Vodicka, H. R. gbtools: interactive visualization of metagenome bins in R. *Front. Microbiol.* https://doi.org/10.3389/fmicb.2015.01451 (2015).
47. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
48. Jensen, L. J. et al. eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic Acids Res.* **36**, D250–D254 (2007).
49. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
50. Jiang, H., Lei, R., Ding, S.-W. & Zhu, S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* **15**, 182 (2014).
51. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* **13**, e1005595 (2017).
52. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
53. Mirdita, M. et al. ColabFold—making protein folding accessible to all. *Nat Methods.* **19**, 679–682 (2022).
54. Evans, R. et al. Protein complex prediction with AlphaFold-Multimer. Preprint at *bioRxiv* https://doi.org/10.1101/2021.10.04.463034 (2021).
55. Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2020).
56. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90k prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* https://doi.org/10.1038/s41467-018-07641-9 (2018).
57. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
58. R: a language and environment for statistical computing (R Foundation for Statistical Computing, 2020).
59. Asnicar, F. et al. Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. *Nat. Commun.* https://doi.org/10.1038/s41467-020-16366-7 (2020).
60. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2014).
61. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792–1797 (2004).
62. Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
63. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
64. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
65. Galili, T. dendextend: an R package for visualizing, adjusting, and comparing trees of hierarchical clustering. *Bioinformatics* https://doi.org/10.1093/bioinformatics/btv428 (2015).
66. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
67. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).
68. Yu, G. Using ggtree to visualize data on tree-like structures. *Curr. Protoc. Bioinformatics* https://doi.org/10.1002/cpbi.96 (2020).
69. Veraart, A. J. et al. Living apart together—bacterial volatiles influence methanotrophic growth and activity. *ISME J.* **12**, 1163–1166 (2018).
70. Meyer, N. R., Fortney, J. L. & Dekas, A. E. NanoSIMS sample preparation decreases isotope enrichment: magnitude, variability and implications for single-cell rates of microbial activity. *Environ. Microbiol.* **23**, 81–98 (2020).
71. Polerecky, L. et al. Look@NanoSIMS—a tool for the analysis of nanoSIMS data in environmental microbiology. *Environ. Microbiol.* **14**, 1009–1023 (2012).
72. Guan, Q. A methanotrophic *Mycobacterium* dominates a cave microbial ecosystem. *Zenodo* https://zenodo.org/record/4767037 (2021).
73. Fei, Q. et al. Bioconversion of natural gas to liquid fuel: opportunities and challenges. *Biotechnol. Adv.* **32**, 596–614 (2014).
74. Vorholt, J. Cofactor-dependent pathways of formaldehyde oxidation in methylotrophic bacteria. *Arch. Microbiol.* **178**, 239–249 (2002).
75. Hanson, R. S. & Hanson, T. E. Methanotrophic bacteria. *Microbiol. Rev.* **60**, 439–471 (1996).
76. Deb, C. et al. A novel in vitro multiple-stress dormancy model for mycobacterium tuberculosis generates a lipid-loaded, drug-tolerant, dormant pathogen. *PLoS ONE* **4**, e6077 (2009).

Acknowledgements

The authors thank C. Murrell for valuable discussions, C. P. Antony for assistance in metagenomics, A. Grootemaat for TEM imaging,

M. Kienhuis for technical support with nanoSIMS analysis, I. Grigoriev for support during fluorescence microsopy, and Z. Para and B. Hegyeli, the Romanian custodians, for facilitating the experiments in Sulfur Cave. Also, we thank Utrecht Sequencing Facility (useq.nl) for providing sequencing service and data. The nanoSIMS facility at Utrecht University was financed through a large infrastructure grant by the Netherlands Organisation for Scientific Research (NWO, grant no. 175.010.2009.011). C.M. was supported by the Dutch Research Council, as part of the MiCRop Consortium (NWO/OCW grant no. 024.004.014). A.P. and Q.G. are supported by a faculty baseline grant (BAS/1/1020-01-01) from KAUST to A.P. The authors also thank members of the Bioscience Core Laboratory in KAUST for providing assistance with the generation of raw genome sequence datasets. PIE research was funded by FONDECYT-CONCYTEC (216-2015-FONDECYT). This is publication number 7467 of the Netherlands Institute of Ecology (NIOO-KNAW).

Author contributions

R.J.M.v.S., organizing ideas, writing, sampling, pathway reconstructions and physiology. Q.G., R.U., J.G., C.M.B. and A.P., metagenomics and proteomics. C.M., writing, pathway reconstructions, phylogenomics and bioinformatic analyses. L.P., writing, ^{13}C-methane experiments, fluorescence and nanoSIMS imaging, data analysis and interpretation. J.-F.F., whole-genome assembly and annotation. E.J.F., fruitful discussions. B.W.B., Illumina sequencing and data processing. J.W.A., sampling and fruitful discussions. M.B. and R.U., technical support. P.I.E., DNA extractions. M.M.M.-F. and P.L.E.B., culturing and ^{13}C-methane experiments. S.R.P. and R.U., mass spectrometry. C.M.B., mycobacterial physiology. N.N.v.d.W., TEM imaging. V.D.G., sampling and culturing. S.M.S., sampling, culturing and fruitful discussions. W.B., organizing ideas, writing, sampling and mycobacterial genetics.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41564-022-01252-3.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41564-022-01252-3.

Correspondence and requests for materials should be addressed to Rob J. M. van Spanning or Wilbert Bitter.

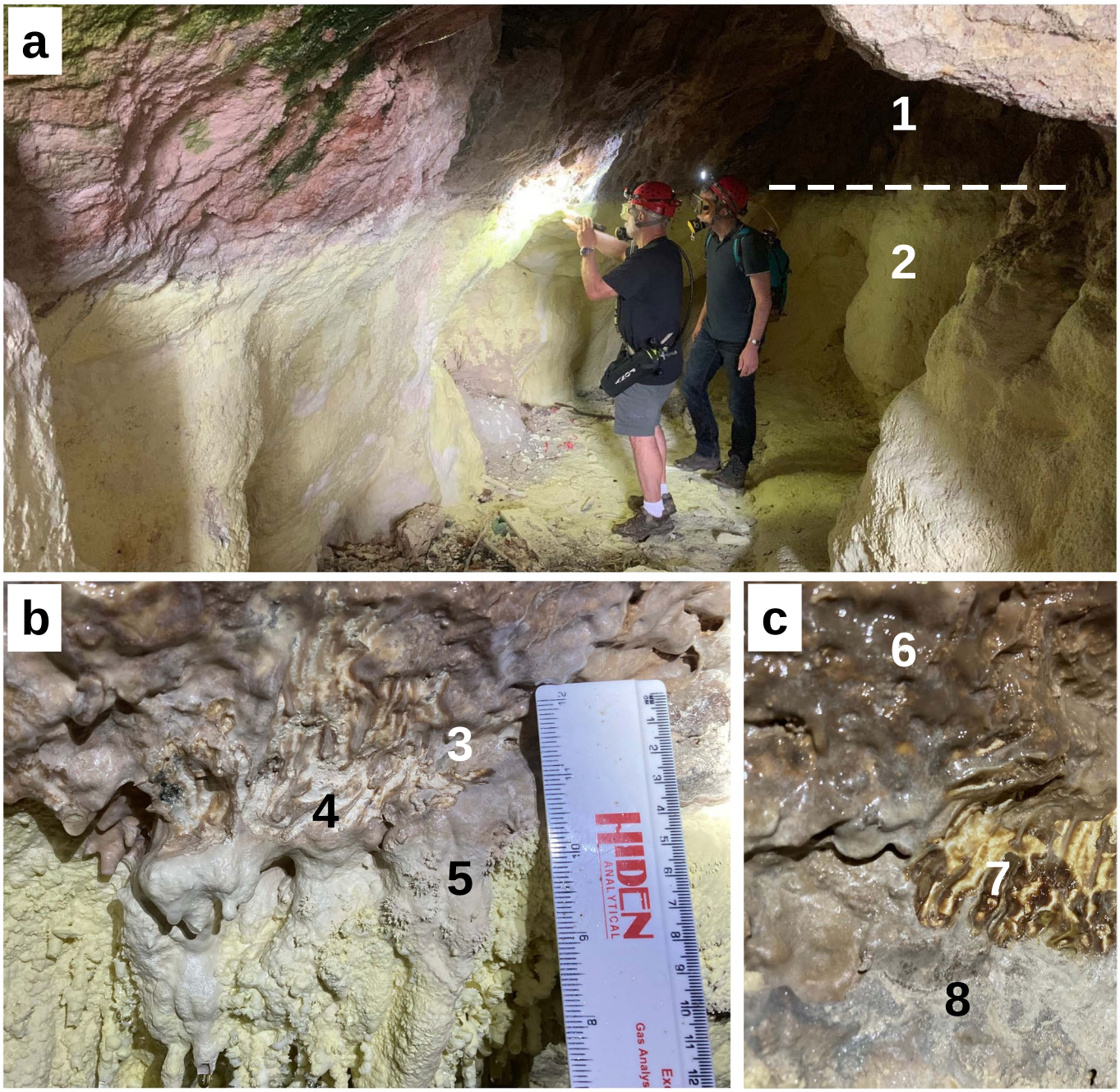
Peer review information Nature Microbiology thanks Marina Khalyuzhnaya, William Inskeep and Florin Musatfor their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

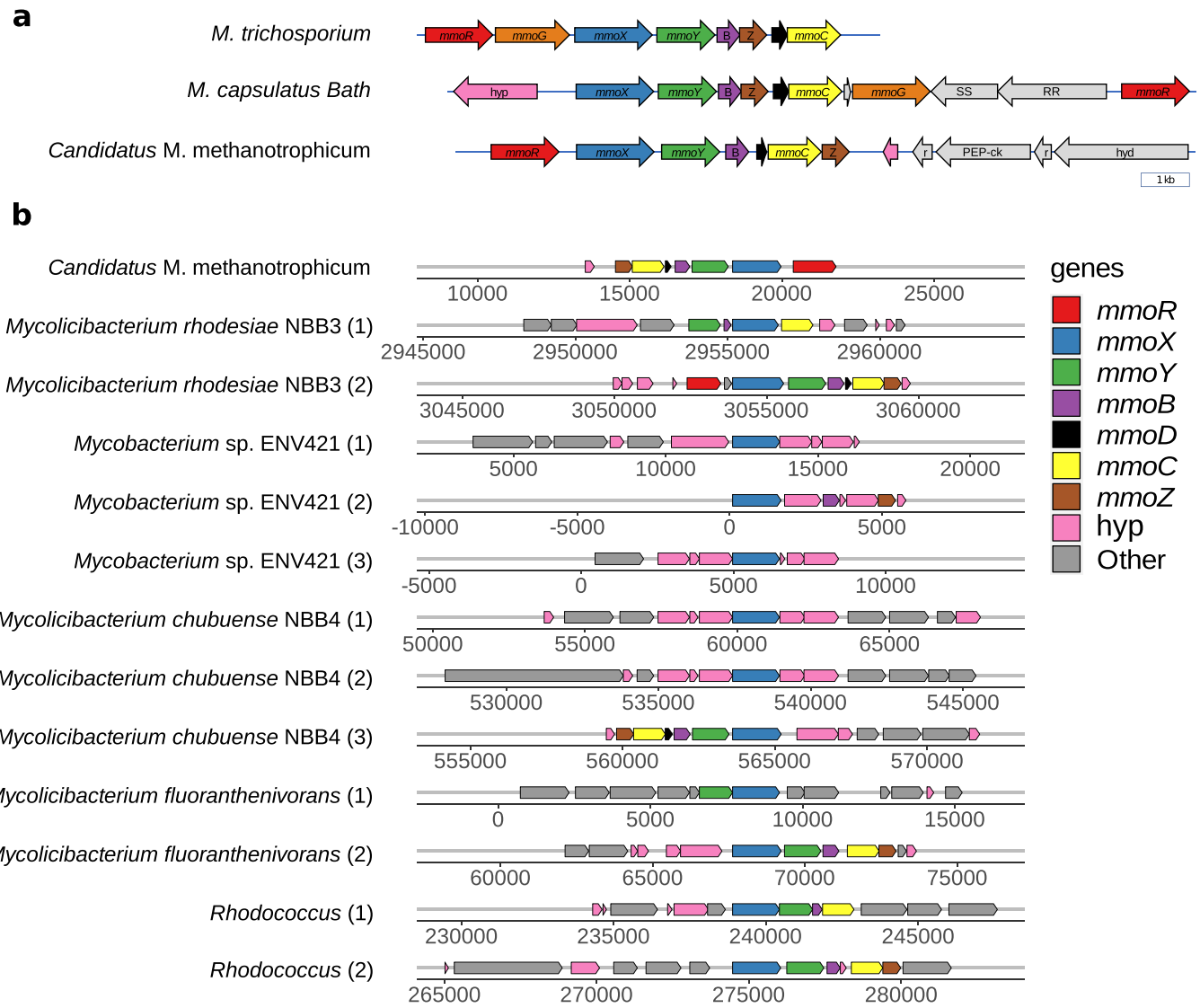
[1]Section Molecular Microbiology, A-LIFE, AIMMS, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. [2]Bioscience program, Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology, Tuwal, Saudi Arabia. [3]Department of Medical Microbiology and Infection Control, Amsterdam University Medical Centers, Amsterdam, the Netherlands. [4]Department of Earth Sciences, Utrecht University, Utrecht, the Netherlands. [5]Interuniversity Institute of Bioinformatics in Brussels—(IB)2 and Department of Organismal Biology, Université libre de Bruxelles, Brussels, Belgium. [6]Department of Preventive Dentistry, Academic Centre for Dentistry Amsterdam (ACTA), University of Amsterdam and Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. [7]Department of Microbial Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, the Netherlands. [8]Department of Medical Oncology, OncoProteomics Laboratory, Amsterdam University Medical Centers, Amsterdam, the Netherlands. [9]Department of Biological Sciences, California State University, Chico, CA, USA. [10]Electron Microscopy Center Amsterdam, Amsterdam University Medical Centers, Amsterdam, the Netherlands. [11]Hydrotechnical Engineering Department, Politehnica University of Timişoara, Timişoara, Romania. [12]Emil Racoviță Institute of Speleology, Bucharest, Romania. [13]Present address: Theoretical Biology and Bioinformatics, Science for Life, Utrecht University, Utrecht, the Netherlands. [14]Present address: Bioinformatics Group, Wageningen University and Research, Wageningen, the Netherlands. ✉e-mail: rob.van.spanning@vu.nl; w.bitter@amsterdamumc.nl

https://doi.org/10.1038/s41564-022-01252-3



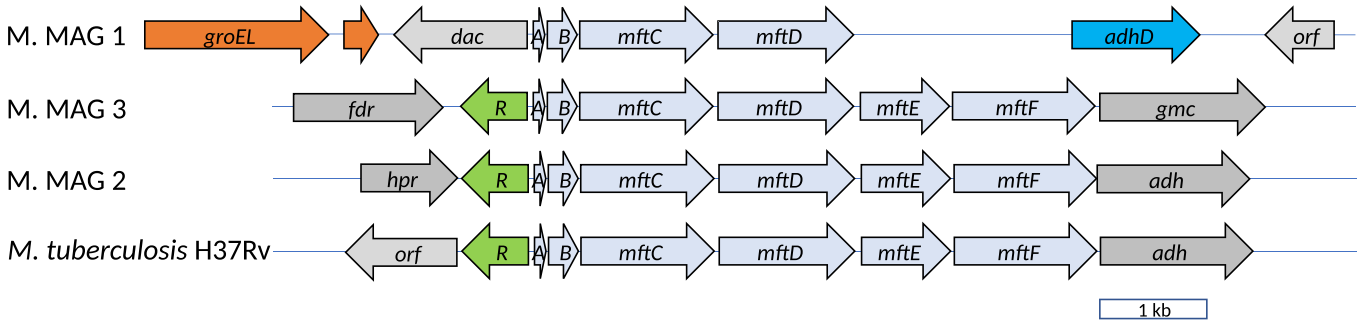
Extended Data Fig. 1 | Sulfur Cave on the Puturosu Mountain (Stinky Mountain), Romania, from which *Candidatus* M. methanotrophicum was enriched. (a) General overview of the cave. **(b)** Detailed images of the cave biofilms. The dashed line in panel **a** marks the stable gaseous chemocline between the volcanic gases (below the chemocline) and atmospheric air (above the chemocline). The cave walls below the chemocline are yellow due to sulphur deposition (marks 2, 5, 8), whereas no such depositions are present above the chemocline (marks 1, 3, 6). Biofilms on the cave walls are only present at the chemocline interface (mark 4). Mark 7 shows the bare cave wall after biofilm sampling.

https://doi.org/10.1038/s41564-022-01252-3



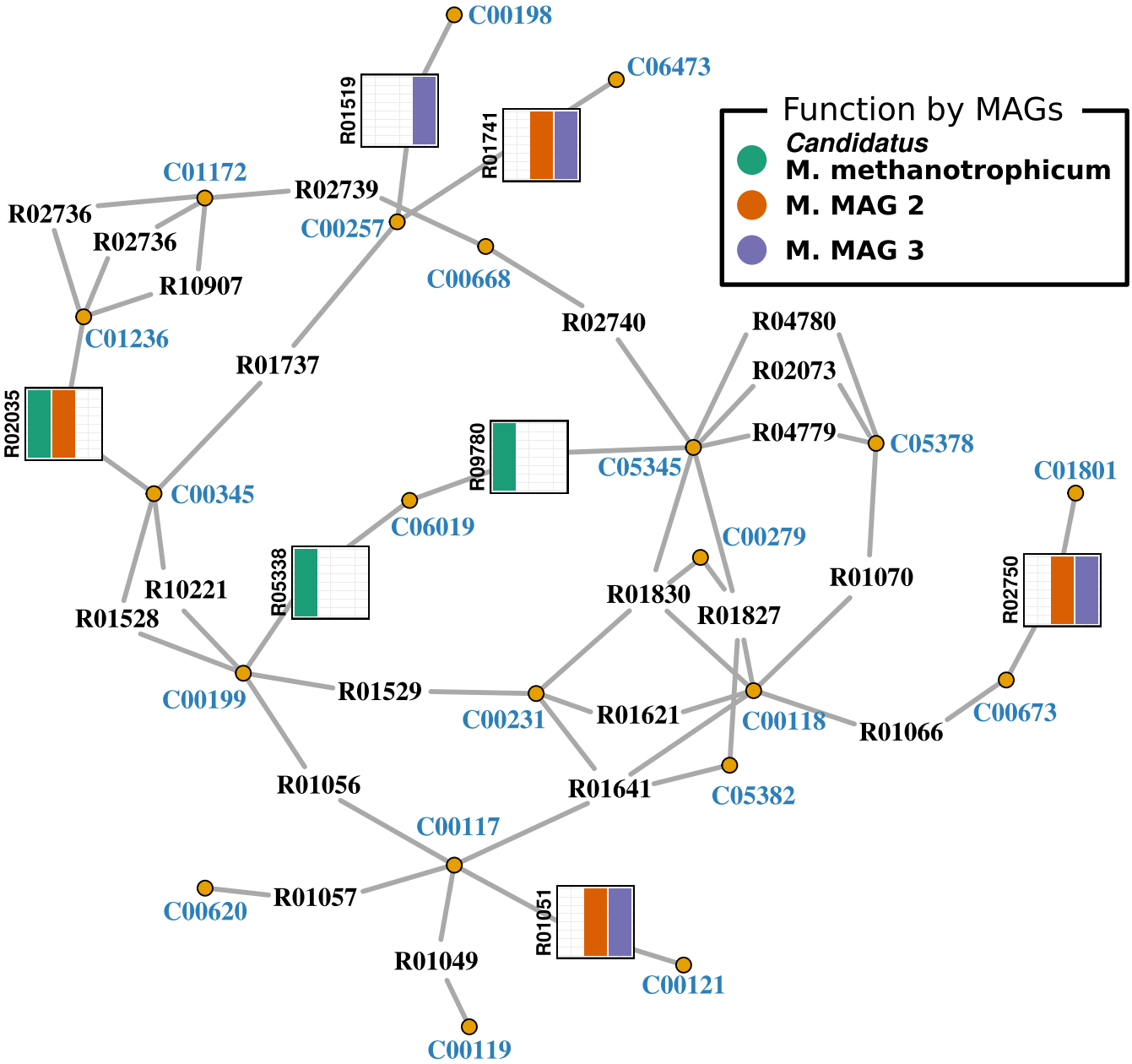
Extended Data Fig. 2 | Maps of sMMO gene clusters in known methanotrophs and selected members of the order Mycobacteriales. (a) Gene clusters encoding sMMO in *Methylosinu strichosporium* OB3b, *Methylococcus capsulatus* Bath, which are known methanotrophs, and *Candidatus* M. methanotrophicum from this study. **Abbreviations:** hyp, gene encoding a hypothetical protein; B, *mmoB*; D, *mmoD*; Z, *mmoZ*; *hyd*, hydrogenase gene cluster; empty arrow, unknown ORF; *r*, hypothetical transcriptional regulator; *PEP-ck*, gene encoding phosphoenolpyruvate carboxykinase; *SS*, signal sensor of a two-component regulatory system; *RR*, response regulator of a two-component regulatory system. Each structural *mmo* gene is shown in distinct color, regulatory *mmoR* and *groEL* homologue *mmoG* genes are in red and orange, respectively. *Candidatus* M. methanotrophicum does not have *mmoG* in close proximity, but does have *groEL* genes elsewhere (see Extended Data Fig. 3). **(b)** Gene maps of genomic regions containing one or more of the *mmoRXYBDCZ* genes. The maps are shown for the species in Cluster 1 of the phylogenetic tree shown in Supplementary Fig. 4. Note that, in this figure, the direction of the arrows does *not* reflect the direction of gene transcription.

https://doi.org/10.1038/s41564-022-01252-3



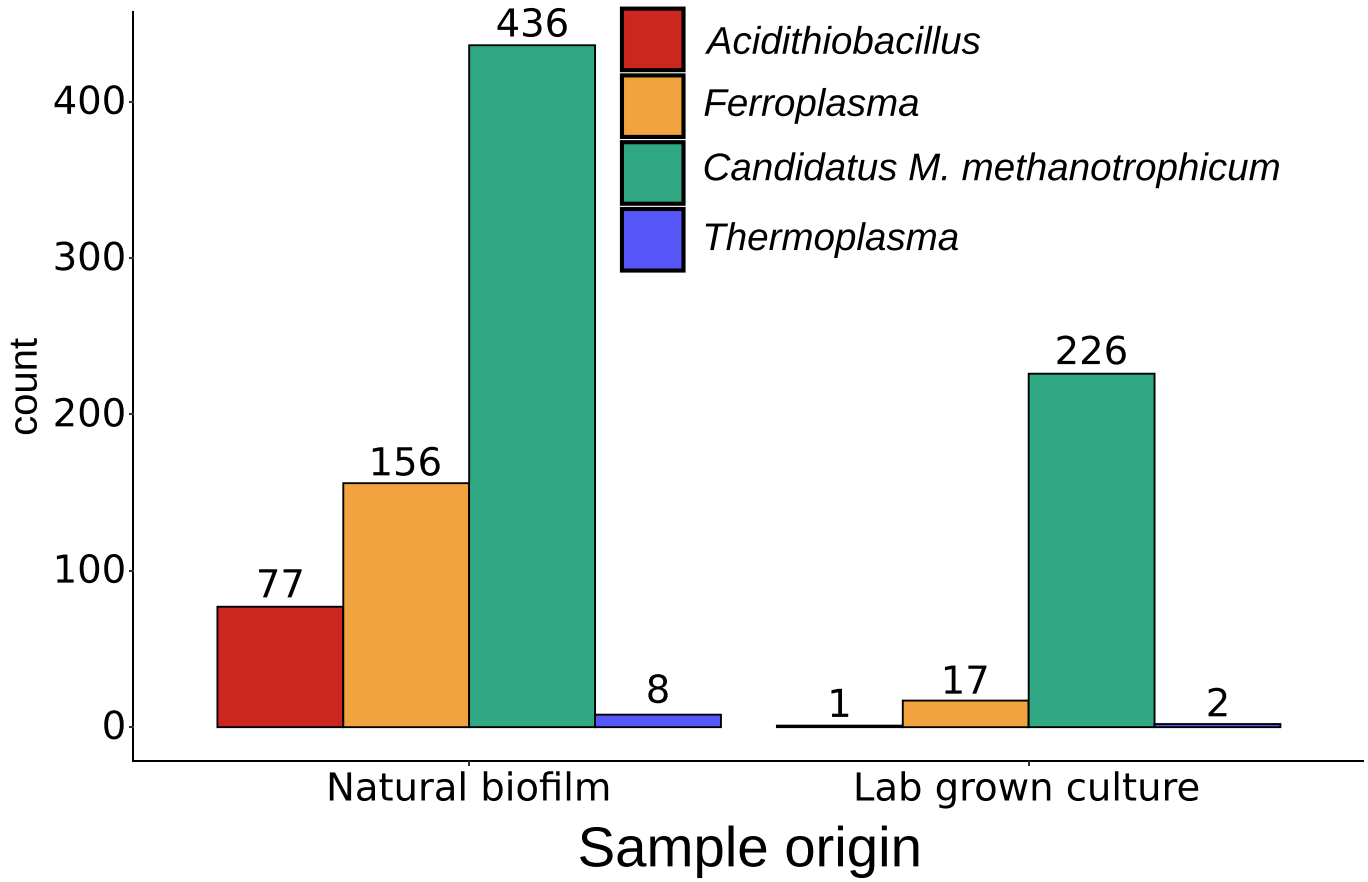
Extended Data Fig. 3 | Maps of *mft* gene clusters in selected *Mycobacterium* species. Abbreviations: M. MAG 1, *Candidatus* M. methanotrophicum; *R*, *mftR*; *A*, *mftA*; *B*, *mftB*; *dac*, D-aminocyclase gene; *adh*, alcohol dehydrogenase gene; *hpr*, hydroxypyruvate reductase gene; *fdr*, Ferredoxin-NAD(P)$^+$ reductase gene; *gmc*, GMC-type oxidoreductase gene. *adh* from M. MAG 2 (ORF 0656) is closely related to the one from *M. tuberculosis* H37Rv (see also Supplementary Fig. 8).

https://doi.org/10.1038/s41564-022-01252-3



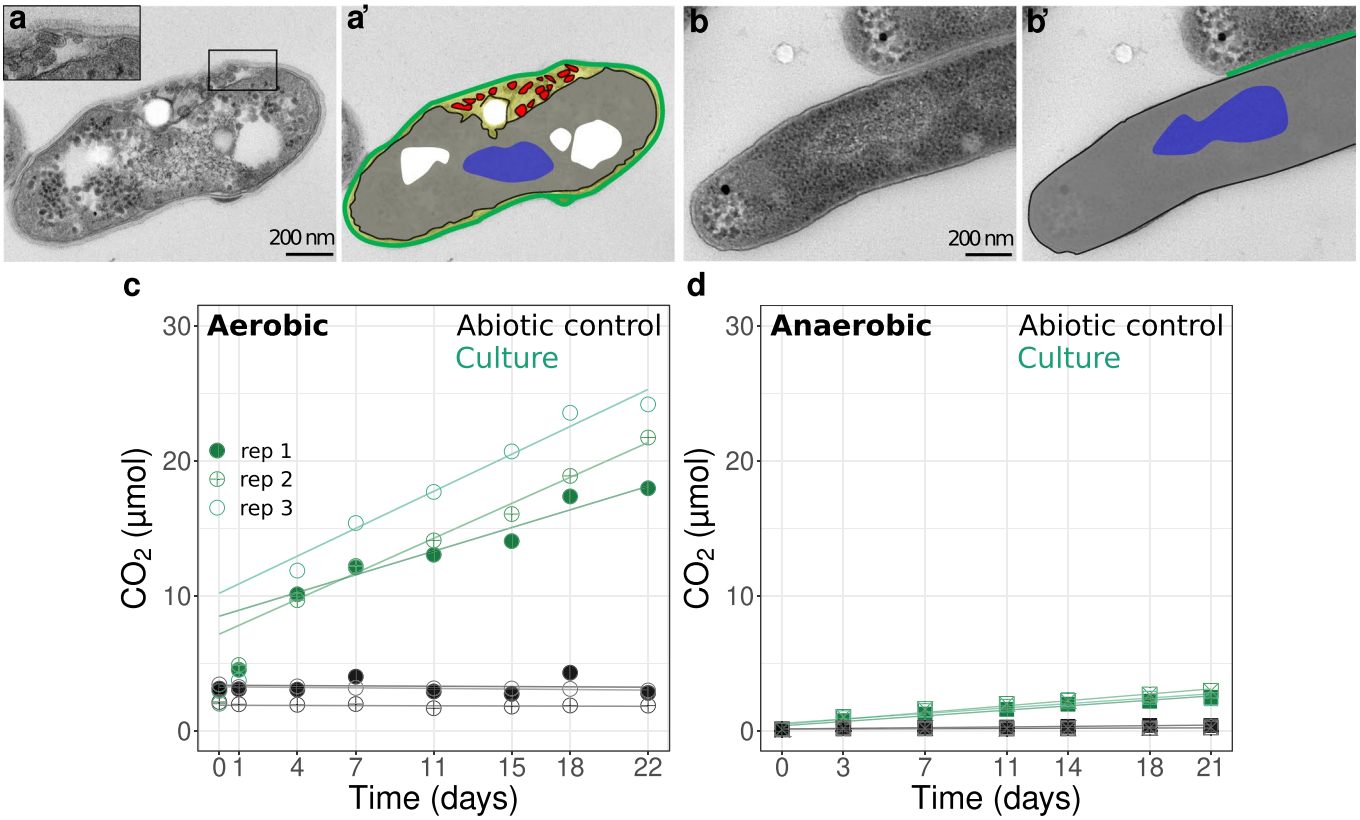
Extended Data Fig. 4 | Metabolic graph of the pentose phosphate pathway based on a corresponding KEGG map (map00030). Shown are reactions of enzymes potentially encoded by the genes of *Candidatus* M. methanotrophicum (M. MAG 1, green), M. MAG 2 (red) and M. MAG 3 (purple) with the corresponding KEGG reaction IDs. Reactions with no color codes are present in all three M. MAGs. Orange circles represent intermediate compounds (corresponding KEGG compound IDs are shown in blue). **KEGG reaction names:** R05338: D-arabino-hex-3-ulose-6-phosphate formaldehyde-lyase (D-ribulose-5-phosphate-forming); R09780: D-arabino-hex-3-ulose-6-phosphate isomerase; R01741: D-Gluconate:(acceptor) 2-oxidoreductase; R02750: ATP:2-deoxy-D-ribose 5-phosphotransferase; R01051: ATP:D-ribose 5-phosphotransferase.

https://doi.org/10.1038/s41564-022-01252-3



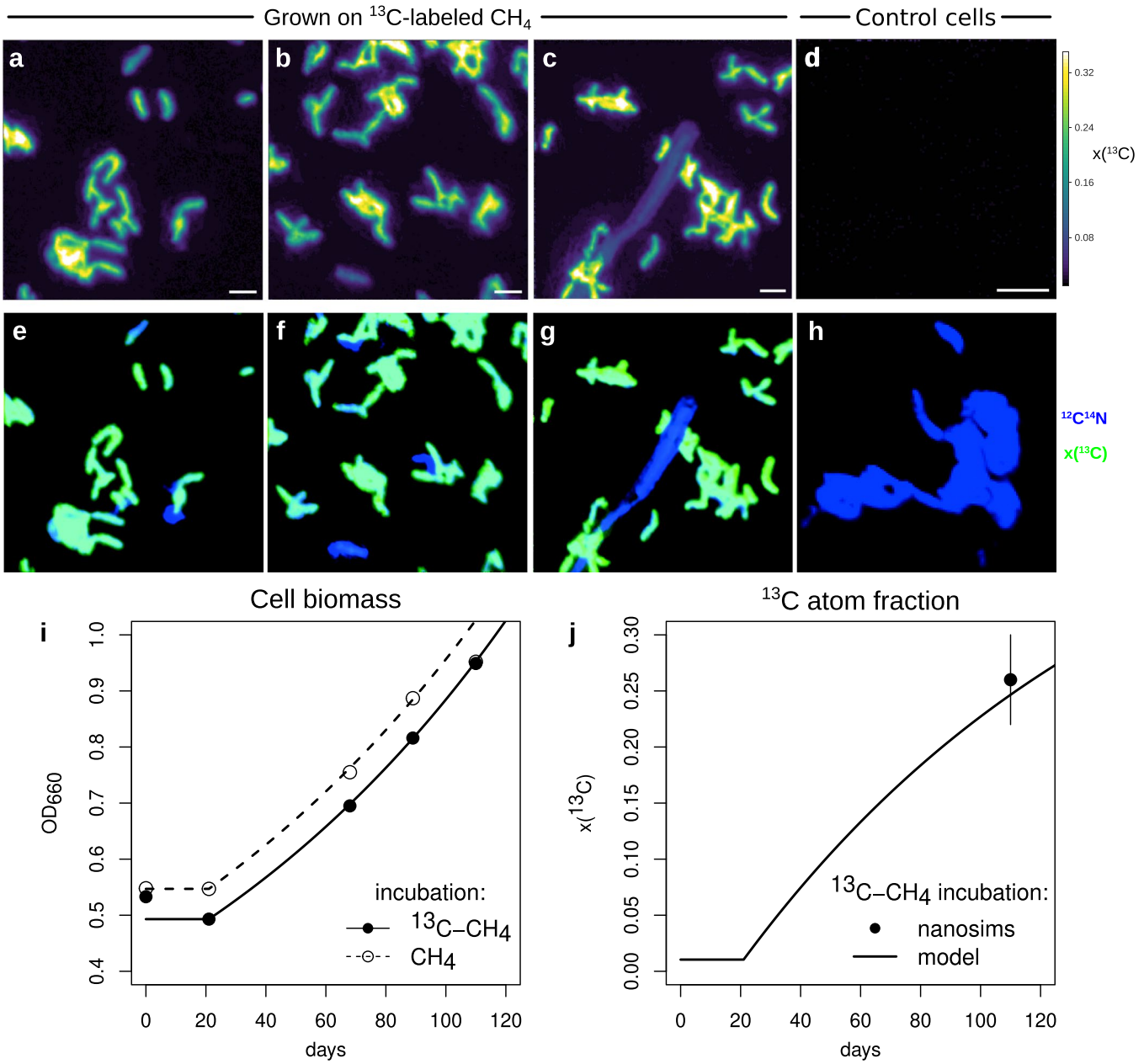
Extended Data Fig. 5 | Number of unique proteins detected by LFQ intensity in the cave biofilm and in the enrichment culture of *Candidatus* M. methanotrophicum. Only proteins from four prokaryotic species that remain in the enrichment culture are shown (see legend). Note the increasing relative contribution of proteins from *Candidatus* M. methanotrophicum in the enrichment culture compared to the cave biofilm.

https://doi.org/10.1038/s41564-022-01252-3



Extended Data Fig. 6 | Morphology and CO_2 production of *Candidatus* M. methanotrophicum. **(a–b)** Electron micrographs of *Candidatus* M. methanotrophicum (**a**) and *M. smegmatis* (**a**). Inset in panel **a** shows vesicles in the periplasm of *Candidatus* M. methanotrophicum at a higher magnification. Schematic representations of the micrographs in panels **a'** and **b'** show the cytosol (grey), the capsular layer (green), the plasma membrane (black line), the periplasm (yellow), the DNA (blue), the elucent areas (white), and the vesicles (red). **(c–d)** Evolution of CO_2 during incubations of *Candidatus* M. methanotrophicum enrichment cultures with CH_4. Incubations were conducted under aerobic **(c)** or anaerobic **(d)** conditions using either live cultures (green symbols) or a sterile NMS medium (black symbols). Different symbols correspond to replicate cultures (rep 1–3), and the values correspond to the total amount of CO_2 per incubation bottle. Green lines show fits of the experimental data within the time interval of 4–22 days **(c)** and 0–22 days **(d)** with a linear model. For the replicate culture 1 incubated under aerobic conditions (rep 1), the slope of this linear model is significantly different from the corresponding negative slope of the linear model characterizing the removal of CH_4 during the incubation (two-sided ANOVA, $F = 6.79$, $p = 0.03$; CH_4 data shown in Fig. 3b). In contrast, the corresponding slopes are not significantly different for replicate cultures 2 ($F = 0.0118$, $p = 0.92$) and 3 ($F = 3.38$, $p = 0.10$). For both the aerobic and anaerobic incubations, the abiotic controls showed no significant variation in the CO_2 amounts over time in comparison to the live cultures.

https://doi.org/10.1038/s41564-022-01252-3



Extended Data Fig. 7 | Additional nanoSIMS images of cells from *Candidatus* M. methanotrophicum enrichment cultures and cell biomass dynamics. **(a–d)** Images of the ^{13}C atom fraction, which is a measure of carbon assimilation from methane provided during the incubation. **(e–h)** Images of an overlay between the $^{12}C^{14}N$ ion counts intensity (blue), which is a proxy for biomass, and the ^{13}C atom fraction (green). Cells shown in panels **a–c** and **e–g** were grown on ^{13}C-labelled methane for 110 days, while cells in panels **d** and **h** were grown on unlabelled methane (control cells). Some cells in panels **e–g** appear blue because their ^{13}C labeling is significantly lower compared to cells that appear cyan, although it was still significant compared to the control cells (see Fig. 4d). Note the filament in panels **c** and **g**, which belongs to a fungus from the genus *Acidomyces*. **(i)** Cell biomass as a function of time in two parallel subcultures of *Candidatus* M. methanotrophicum grown on CH_4 as the sole carbon and energy source. One culture used ^{13}C-labelled CH_4 (^{13}C atom fraction of 0.5), the other one used unlabelled CH_4. Lines show the modelled biomass assuming an exponential growth with a doubling time of 94 days (solid line) and 98 days (dashed line) and a lag phase of 21 days. The decrease in the doubling time from 150–200 in the original culture to 94–98 days in these subcultures was possibly due to small changes in the culturing conditions combined with improved growth properties of *Candidatus* M. methanotrophicum. **(j)** ^{13}C atom fractions in the cells of *Candidatus* M. methanotrophicum grown on ^{13}C-labelled CH_4. Symbol shows the mean value, error bar corresponds to the standard deviation (calculated based on the measurement of $N = 110$ cells, where the cells were not treated by any post-incubation chemical procedure prior to the nanoSIMS analysis). Solid line shows the ^{13}C atom fraction modelled based on the assumption that the ^{13}C-labelled CH_4 was the sole carbon and energy source and the growth characteristics were as shown in panel **i**.

nature portfolio

Corresponding author(s): Rob van Spanning and Wilbert Bitter

Last updated by author(s): Jul 4, 2021

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a | Confirmed

- [x] The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
- [x] A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- [x] The statistical test(s) used AND whether they are one- or two-sided
  Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- [x] A description of all covariates tested
- [x] A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- [x] A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- [x] For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted
  Give P values as exact values whenever suitable.
- [x] For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- [x] For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- [x] Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No custom software, open-source or custom code used to collect data
Data analysis	Used software for analysis: fastANI, PhyloPhlAn v3.0, BBDuk, MEGAHIT assembler, Prokka, EggNOG, CheckM, CoverM, MaxQuant version 1.6.10

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The 16S rRNA gene sequence of cultivated M. methanotrophicum has the NCBI accession number MW243585. Illumina 16S rRNA gene amplicon data are associated with the NCBI BioProject PRJNA675490. The metagenomics and pure culture sequencing samples are deposited into ENA under the study number PRJEB45004. Specifically, the accession numbers for the assemblies used in the present study are ERS6581339, ERS6581338, ERS6581340 and ERS6581341 for M. methanotrophicum VU-WI (enriched culture), M. MAG 1, M. MAG 2 and M. MAG 3, respectively. Corresponding annotation files and extra supplementary data can be found in Zenodo repository (https://zenodo.org/record/4904224#.YOG44R1RUXo).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We studied the composition of a microbial community in a cave biofilm with a major focus on a (new) Mycobacterium species that was later on enriched by growth on methane as carbon and energy source.
Research sample	Samples were obtained from a biofilm at a gas chemocline in Sulfur cave of Puturosu Mountain, Romania .
Sampling strategy	For the pools and springs, the sampling strategy was to number the samples in chronological order without any link to measurable characteristics (i.e. pH, cloudiness, salinity etc).
Data collection	Data was collected from protein and DNA samples obtained directly from the biofilm samples (extraction performed by Catalin Bunduc, Roy Ummels and Martin Braster) or DNA obtained from the enriched methane culture (extraction performed by Roy Ummels). Data collection was performed by Sander Piersma (proteomics) or Qingtian Guan (genomics and metagenomics).
Timing and spatial scale	Collection of field samples from the cave was performed on two separate occasions, one year after the other. Collection of pool and spring samples was done in the second field trip during two consecutive days.
Data exclusions	No data was excluded from the analysis.
Reproducibility	To be able to link the proteomics data with the metagenomics data, the cave biofilm samples were homogenized, split in half and processed separately. For the pools and spring samples, we determined and described all individual samples, no reproducibility samples were included.
Randomization	Samples for the pools and springs were numbered in chronological order of sampling, subsequently processed, after which the data was combined and analyzed. As such, no bias in sample covariates (such as pH) was introduced.
Blinding	As outlined above pool and spring samples were numbered by one researcher (Rob van Spanning) and extracted by another researcher (Martin Braster). The different readouts (PCR data and pH data) were combined after they came available.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	The cave samples conditions are relatively stable in different weather conditions. The pool and spring samples were obtained in May, with a temperature of 20 degrees Celsius, and dry weather.
Location	All samples were obtaining from the cave or pools/springs around Puturosu mountain, Romania, 46:1°N, 26:0°E.
Access & import/export	Local scientists are participating in the study (Serban Sarbu and Vasile Gherman) and were in contact with the caretakers and local authorities. Biological material was directly stored in sterile clean plastic tubes.
Disturbance	The cave biofilm samples were estimated to be less than 0,1% of the total biofilm, for pools/springs this percentage was even much lower.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging