

Comparative phenotypic and genomic analysis of the methanogen *Methanomethylovorans thermophila* L2FAW and its phylogenomic placement within the Genome Taxonomy Database

Mathias Wunderer*, Andja Mullaumeri, Andreas O. Wagner and Eva Maria Prem

Abstract

The genome of the methylotrophic methanogen *Methanomethylovorans thermophila* L2FAW is not included in the Genome Taxonomy Database (GTDB) so far, even though the strain was first described in 2005. To evaluate its genomic characteristics and placement in the GTDB, we sequenced the genome of *M. thermophila* L2FAW via *Illumina* shotgun and Oxford Nanopore sequencing and subsequently did hybrid assembly. The assembled genome consists of 2.25 Mbp (contigs ≥ 500 bp) with a G+C content of 40 mol%. The quality of the genome is good, which is already apparent from the low L50 (=1) and L90 (=2) metrics. Our assembled genome was highly similar to the metagenome-assembled genome *Methanomethylovorans* sp014361205 (GCA_014361205.1_ASM1436120v1_genomic) with an average nucleotide identity of 99.9%. Even though KEGG Mapper Reconstruction results revealed that *M. thermophila* L2FAW harbours all the enzymes necessary for acetoclastic and hydrogenotrophic methanogenesis and *gapseq* predicted formate as a potential substrate for *M. thermophila* L2FAW, no metabolic activity could be observed on acetate, H₂-CO₂ (80:20 vol/vol, 2,000 mbar) and on a mixture of H₂-CO₂ and formate in lab tests; thus, the obligate methylotrophic lifestyle of the phenotype was confirmed.

DATA SUMMARY

This whole-genome shotgun project has been deposited at the National Center for Biotechnology Information (NCBI) under the BioProject ID: PRJNA1271748. The BioSample accession no. is SAMN48886148, and the two Sequence Read Archive (SRA) accession nos. are SRX29517605 and SRX29517606.

INTRODUCTION

The methanogenic archaeon *Methanomethylovorans thermophila* L2FAW was isolated from a laboratory-scale thermophilic upflow anaerobic sludge blanket reactor and described in 2005 [1]. It is phylogenetically closely related to *Methanomethylovorans hollandica* DMS1 (98% similarity for the 16S rRNA gene) and represented a novel species in the genus *Methanomethylovorans* [1]. The genus name was proposed in 1999 after the isolation of the type strain *M. hollandica* DMS1 from the sediment of an eutrophic freshwater pond in the Netherlands [2]. *Methanomethylovorans* species are obligate methylotrophs, utilizing methanol, methylated amines, methylated sulphides and tetramethylammonium hydroxide [1, 3]. As elaborated with the *SingleM* software suite implemented in *Sandpiper* [4], *Methanomethylovorans* sp. was also found in other habitats, such as wetlands, biofilms, wastewater, anaerobic digestion plants, soils or mine drainage with relative abundances of up to 30% in, e.g. wetlands (https://sandpiper.qut.edu.au/taxonomy/g__Methanomethylovorans). Coastal wetlands play an integral role in the global carbon cycle, potentially serving as methane sources [5]. In those habitats, methylotrophic methanogenesis is a key metabolic process [6]. In river sediments, *Methanomethylovorans* sp. dominate dimethyl sulphide degradation [7], thereby further contributing to global climate dynamics.

Access Microbiology is an open research platform. Pre-prints, peer review reports, and editorial decisions can be found with the online version of this article. Received 30 September 2025; Accepted 24 April 2026; Published 21 May 2026

Author affiliations: *Department of Microbiology, Universität Innsbruck, Technikerstrasse 25d, 6020 Innsbruck, Austria.

***Correspondence:** Mathias Wunderer, mathias.wunderer@uibk.ac.at

Keywords: hybrid genome assembly; *Illumina* shotgun sequencing; methanogenic archaea; Oxford Nanopore sequencing.

Abbreviations: AF, aligned fraction; ANI, average nucleotide identity; GTDB, Genome Taxonomy Database; MAG, metagenome-assembled genome; NCBI, National Center for Biotechnology Information.

One supplementary figure and five supplementary tables are available with the online version of this article.

001123.v4 © 2026 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License. This article was made open access via a Publish and Read agreement between the Microbiology Society and the corresponding author's institution.

The Genome Taxonomy Database (GTDB) follows a phylogenomic approach based on a set of conserved proteins (120 for bacteria and 53 for archaea), primarily related to ribosomes (<https://gtdb.ecogenomic.org/>). It defines higher-rank taxa and species via relative evolutionary divergence and average nucleotide identity (ANI), respectively. Moreover, the National Center for Biotechnology Information (NCBI) Assembly Database functions as a genome repository [8]. The genome of *M. thermophila* L2FAW has not been fully sequenced so far, and it is, therefore, not included in the GTDB. Therefore, this study aimed to sequence the whole genome of *M. thermophila* L2FAW, to compare the genome to the phenotype regarding substrate spectrum and to phylogenomically place the strain within the GTDB to possibly provide information on assigned metagenome-assembled genomes of *Methanomethylovorans* with placeholder names, as well as to gain insights into the geographic distribution of *M. thermophila* L2FAW.

METHODS

Cultivation

The methanogenic archaeon *M. thermophila* L2FAW (DSM 17232) was obtained from DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen (Braunschweig, Germany). The cultivation occurred in a 120-ml serum flask, filled with 50 ml of DSMZ 684 medium with varying carbon sources: 2 ml l⁻¹ methanol, 2.5 g l⁻¹ sodium acetate, 2,000 (±25) mbar H₂-CO₂ (80:20 vol/vol) and 2 g l⁻¹ sodium formate with 2,000 (±25) mbar H₂-CO₂ (80:20 vol/vol). The incubation temperature was 55°C, and 10% (vol/vol) cell suspension was used for inoculation. Media preparation under anoxic conditions occurred according to Wagner *et al.* [9].

Analysis of gas production

Overpressure in the serum flask was measured with a G1113-UT-GE manometer (Greisinger electronic, Germany) using the current atmospheric pressure as reference [Bundesanstalt für Geologie, Geophysik, Klimatologie und Meteorologie (GeoSphere, Austria)]. The biogas composition (H₂, CH₄, CO₂) in the headspace of the serum flask was determined using a GC 2010 gas chromatograph (Shimadzu, Japan) according to Wagner *et al.* [10]. The gas chromatograph was calibrated before each measurement with a reference gas from Messer (Austria) containing 60% CH₄, 30% CO₂ and 5% H₂. CH₄ production was calculated according to Wunderer *et al.* [11].

DNA purification

DNA purification was carried out with the Quick-DNA[™] HMW MagBead Kit (Zymo Research, USA) according to the manufacturer's recommendations. The DNA was quantified with Quant-iT[™] PicoGreen[™] dsDNA Assay Kit (Invitrogen, Germany), resulting in a concentration of 14.1 ng µl⁻¹. Additionally, the DNA quality was checked with a NanoDrop[™] 2000 c (Thermo Scientific) spectrophotometer (260/280=1.82 and 260/230=2.00).

Next- and third-generation sequencing

Illumina shotgun next-generation sequencing was done on a NovaSeq device by Novogene (Cambridge, UK). Third-generation sequencing was done with a MinION Mk1B device (Oxford Nanopore Technologies, Great Britain) in-house using the MinKnow software v24.06.5 (Oxford Nanopore Technologies, Great Britain), an R10.4.1 MinION Flow Cell (FLO-MIN114, Oxford Nanopore Technologies, Great Britain), the Ligation Sequencing Kit V14 (SQK-LSK114, Oxford Nanopore Technologies, Great Britain) and companion products (NEB, USA) according to the manufacturer's recommendation.

Processing of the raw reads and genome assembly

Base calling of the *Nanopore* reads occurred with *Dorado* basecaller (0.7.3-linux-x64), using the automatically downloaded HAC model (<https://github.com/nanoporetech/dorado>). The base called reads were single-read error corrected with the HERRO algorithm (<https://github.com/nanoporetech/dorado>). We used *Filtlong* (<https://github.com/rrwick/Filtlong>) without an external reference for quality filtering of the *Nanopore* reads and removed sequences shorter than 1 kbp, as well as the worst 10% (measured by bp, not by read count). Quality filtering and adapter removal for *Illumina* shotgun reads were done with *Trimmomatic* (<https://github.com/usadellab/Trimmomatic>). Hybrid assembly using *Nanopore* and *Illumina* reads took place with *SPAdes* 4.0.0 (<https://github.com/ablab/spades>) [12–15]. The completeness of the assembled genome was verified with *Bowtie2* (<https://github.com/BenLangmead/bowtie2>). *SAMtools* (<https://github.com/samtools/samtools?tab=readme-ov-file>) was used to sort and index the assembled genome and remove unpaired reads [16]. Polishing of the genome occurred with *Pilon* (<https://github.com/broadinstitute/pilon?tab=readme-ov-file>) and the quality of the assembled genome was evaluated with *QUAST* 5.3.0 (<https://github.com/ablab/quast>) [17].

Annotation, taxonomic classification and metabolic reconstruction

We downloaded GTDB species representatives of *Methanomethylovorans* sp. (*n*=10) from the NCBI [18] in June 2025 (GTDB release r226). GTDB representatives are non-redundant, curated biological genomes of high quality. Together with our assembled genome of *M. thermophila* L2FAW (*n*=11), ANI values were calculated in a multifactorial manner via *ANIClustmap* v2.0.1 using *skani* as calculation tool [19] (<https://github.com/moshi4/ANIClustmap>) and visualized with the options `--annotation`, `--fig_width 18`, `--cmap_gamma 0.8`, `--cmap_colors`, `--mode skani` (Fig. 1). For information on the aligned fraction

(AF), we used *skani* v0.2.2 with following options: --slow, --min-af 10 and --robust [20]. Furthermore, we downloaded all GTDB species representatives of the family *Methanosarcinaceae* ($n=139$) from the NCBI [21] in February 2026 (GTDB release r226). The 53 marker proteins for archaea were identified and aligned with *GTDB-Tk* (v2.4.1; *GTDB-Tk* reference data version r226) (<https://github.com/GenomeEngine/GTDBTk>) [22]. The alignment of these 140 genomes (139 GTDB species representatives and our assembled genome) was used for the phylogenomic reconstruction with *iqtree* v3.0.1 [23] with the options: -m MFP [24], -bb 1000 [25] and alrt 1000 (<https://github.com/iqtree/iqtree3>). The model finder selected Q.PLANT+F+R7 as the best-fit model according to the Bayesian information criteria. The phylogenomic tree was visualized with the online display tool *iTOL* (<https://itol.embl.de>) [26].

The completeness and the contamination of the genome were determined using *CheckM2* 1.1.0 (<https://github.com/chklovski/CheckM2>) [27]. *Prokka* v1.14.5 (<https://github.com/tseemann/prokka>) [28] was used for the annotation of the genome. The resulting amino acid sequences were processed with *BlastKOALA* (with BlastKEGG Orthology Anks Links Annotation) (<https://www.kegg.jp/blastkoala/>) [29] and visualized with *KEGG Mapper Reconstruction Result* [30]. Additionally, prediction of metabolic pathways, transporter inference, metabolic model construction and multi-step gap filling took place with *gapseq* (<https://github.com/jotech/gapseq>) [31, 32]. The received metabolic model was further processed in *RStudio* 2024.12.1+563 with the R version R-4.4.2 (Posit PBC, USA), using the packages *data.table* [33] and *sybil* [34] (<https://gapseq.readthedocs.io/en/latest/tutorials/archaea.html>) [35]. *SingleM* software suite (<https://github.com/wwood/singlem>) implemented in *Sandpiper* (<https://github.com/wwood/sandpiper>) [4] was used to determine the distribution of the metagenome-assembled genome (MAG) *Methanomethylovorans* sp014361205 and *Methanomethylovorans* sp. in general.

RESULTS AND DISCUSSION

Genome quality assessment

The quality of the assembled genome and its closest match (MAG GCA 014361205.1) is shown in Table S1, available in the online Supplementary Material. In total, 53 contigs were created, of which 3 were longer than 500 bp. The genome comprises 2.27 Mbp (2.25 Mbp contigs ≥ 500 bp) with a G+C content of 40 mol%. According to *CheckM2*, the genome has a completeness of 98.97% and contamination of 0.57%. This is comparable with other in the GTDB listed *Methanomethylovorans* genomes, i.e. 2.15 Mbp for the MAG GCA_014361205.1 (completeness 99.67%) (https://gtdb.ecogenomic.org/genome?gid=GCA_014361205.1) or 2.71 Mbp for the type species of the genus *Methanomethylovorans*, *M. hollandica* GCF_0000328665.1 (completeness 99.84%) (https://gtdb.ecogenomic.org/genome?gid=GCF_0000328665.1). The G+C content is slightly higher

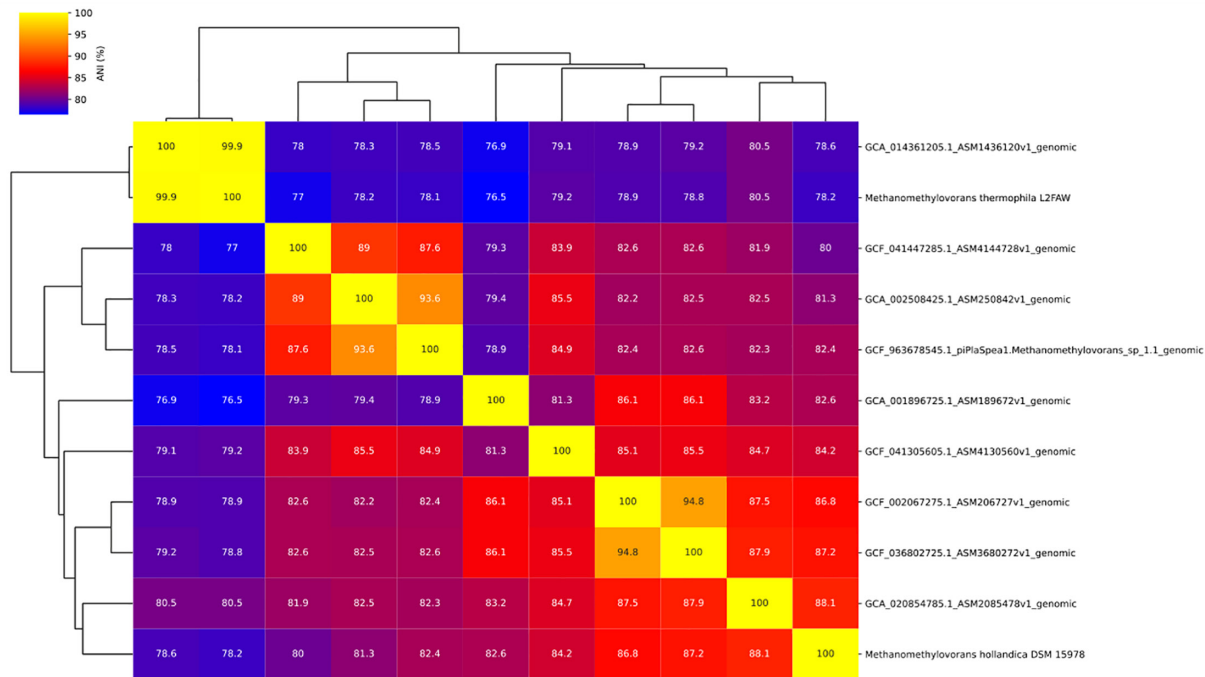


Fig. 1. Comparison of ANI values of the GTDB species representatives of *Methanomethylovorans* sp. with our assembled genome (*M. thermophila* L2FAW). The completeness, align fraction reference and align fraction query of the GTDB species representatives are shown in Table S5.

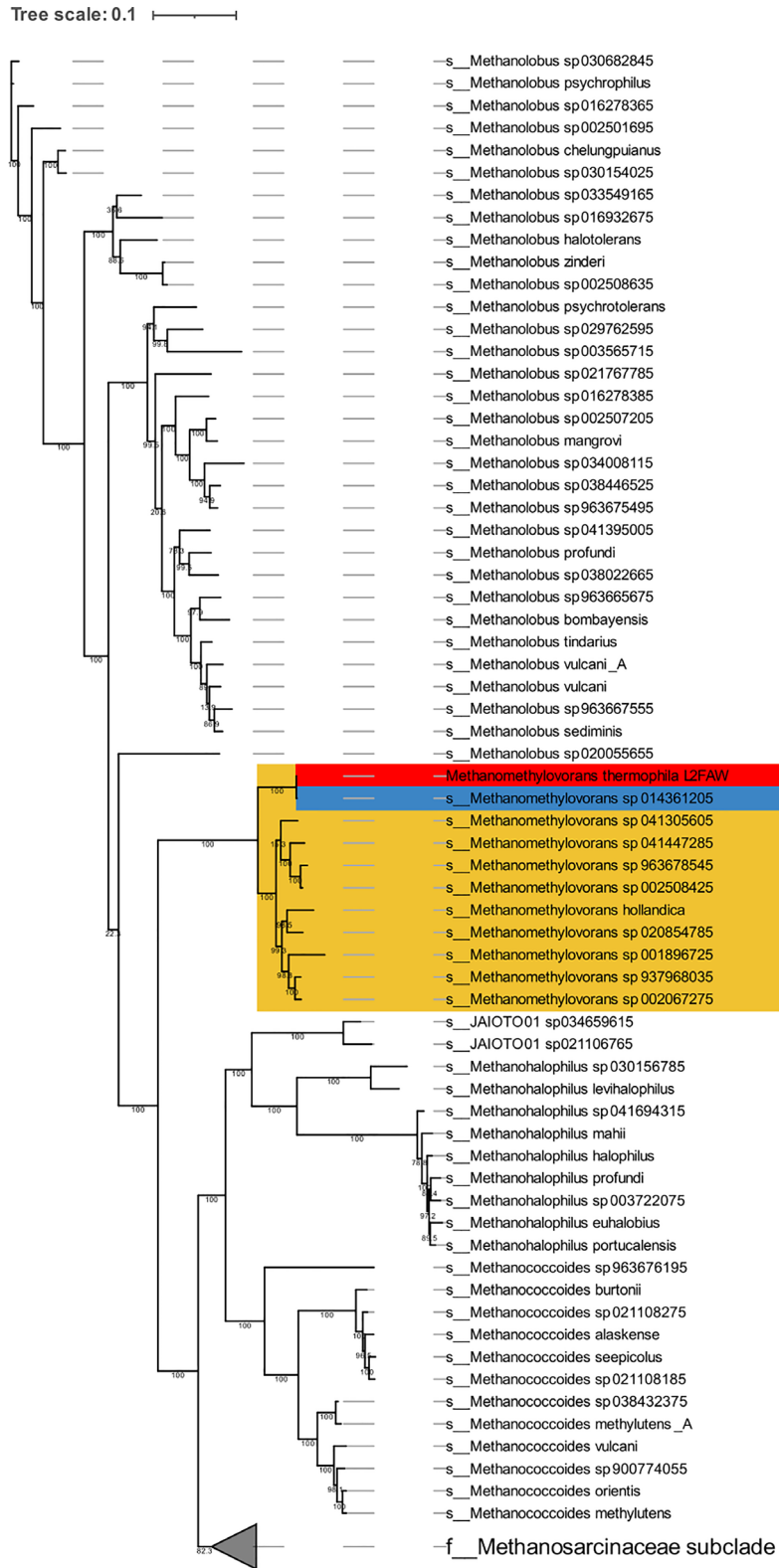


Fig. 2. Phylogenomic tree based on the 53 archaeal marker proteins of all available GTDB species representatives ($n=139$) and our assembled genome *M. thermophila* L2FAW. In red, our assembled genome; in blue, the MAG with the highest similarity to our genome (ANI=99.9%); and in yellow, all members of the genus *Methanomethylovorans*. The given numbers at the branches reflect bootstrap values.

than determined by Jiang et al. [1] (37.6 mol%) and is comparable with that of other *Methanomethylovorans* sp. (34.4 mol% of *M. hollandica* DSM1 and 39.2 mol% of *Methanomethylovorans uponensis*) [1, 36]. N50 and N90 lengths were 1.56 Mb and 685 kb, respectively. The minimum number of contigs needed to cover 50 and 90% of the assembled genome was 1 and 2, respectively (L50=1, L90=2). 99.96% of the reads are mapped with a coverage of 100% (1×), an average coverage depth of 414 and zero uncalled bases per 100 kb. These metrics indicate that hybrid assembly worked well, and the genome is of good quality.

Taxonomic evaluation

The investigated genome of *M. thermophila* L2FAW had the highest similarity with the MAG *Methanomethylovorans* sp014361205 (GCA_014361205.1_ASM1436120v1_genomic), with a genome ANI of 99.9% (Fig. 1). Genomes with an ANI $\geq 95\%$ (or $\geq 97\%$) are usually clustered as the same species [37]. Moreover, 96.5% of the reference and 92.2% of the query were covered by the alignment (AF value) (Table S5). So far, MAG *Methanomethylovorans* sp014361205 was found globally in bioreactor sludge, soil, wastewater or oil production facilities with relative abundances below 5% (https://sandpiper.qut.edu.au/taxonomy/s__Methanomethylovorans%20sp014361205). This is in accordance with the isolation source of *M. thermophila* L2FAW.

All other *Methanomethylovorans* sp. genomes showed ANI and AF values ≤ 80 and $< 40\%$, respectively, when compared with *M. thermophila* L2FAW (Table S5). This indicates a high genomic distance between *M. thermophila* L2FAW and those *Methanomethylovorans* sp. genomes.

According to Parks et al. [37], intragenus species within genera with multiple species show a distribution of ANI values between 78 and 95%. Since all ANI values except for the MAG *Methanomethylovorans* sp014361205 are between 76.52 and 80.38% (Fig. 1), when compared with our assembled genome, we calculated a phylogenomic tree with all available GTDB species representatives of the family *Methanosarcinaceae* (Fig. 2). Our assembled genome (in red) clusters together with the MAG *Methanomethylovorans* sp014361205 (in blue) in the same clade as the sister group of all the other *Methanomethylovorans* sp. (in yellow). This shows us that it belongs to the genus *Methanomethylovorans*, even when the ANI values in the pairwise comparison are mostly $< 80\%$ (for the entire tree, please see Fig. S1).

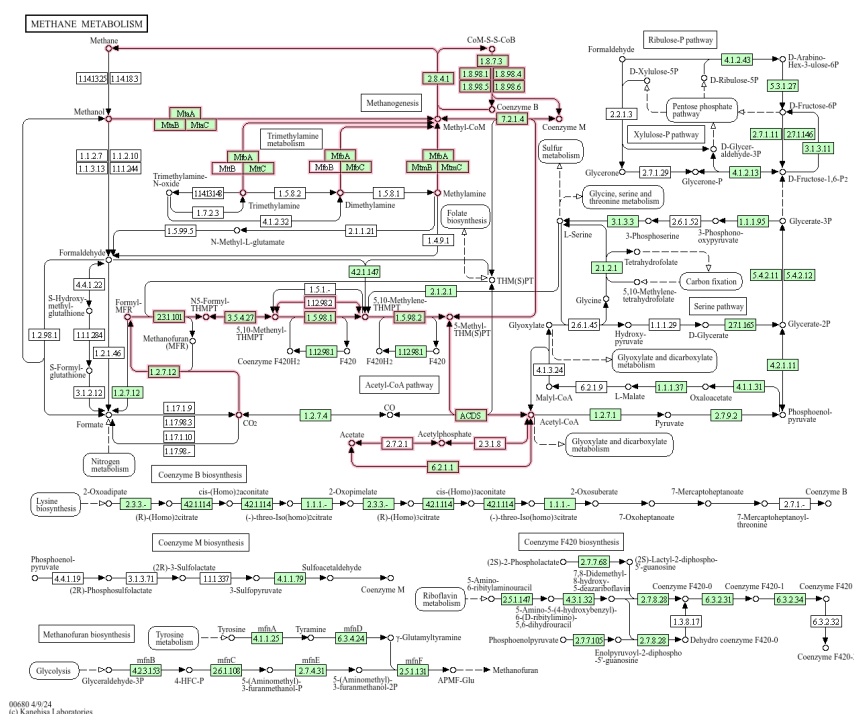


Fig. 3. Methane metabolism map created with BlastKOALA and KEGG Mapper Reconstruction Result. In green are the enzymes that are present in the genome of *M. thermophila* L2FAW. Red lines show the methanogenic pathways: M00567 (CO₂ → methane, complete 8/8), M00357 (acetate → methane, complete 5/5), M00356 (methanol → methane, complete 3/3) and M00563 (mono-, di- and trimethylamine → methane, complete 4/4).

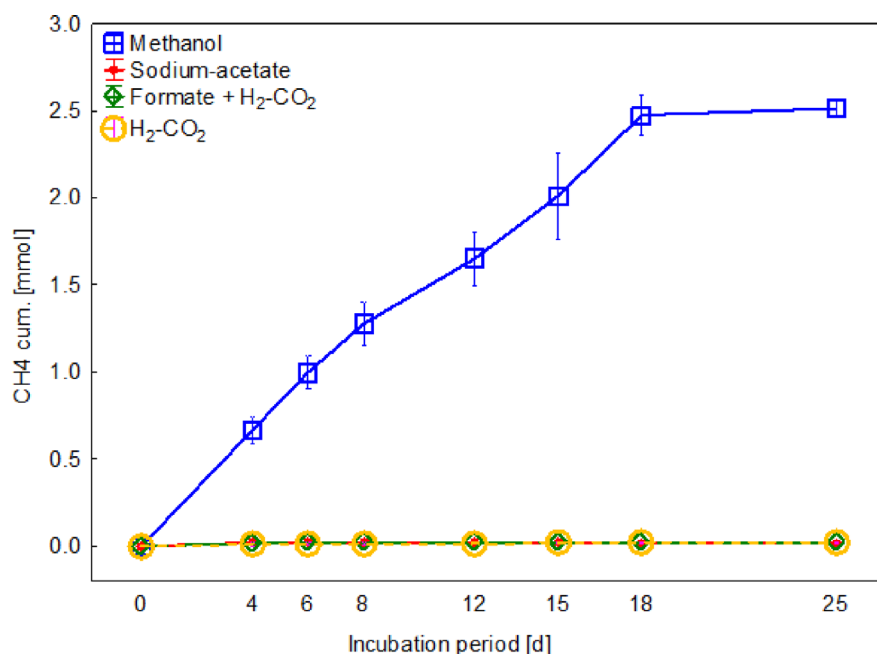


Fig. 4. Cumulative methane production of *M. thermophila* L2FAW on various carbon sources: methanol, sodium acetate, H₂-CO₂ and a mixture of H₂-CO₂ and sodium formate.

Genome annotation and metabolic prediction

According to *Prokka*, the genome has 2,168 CDS, of which 1,107 could not be assigned and classified as hypothetical proteins (Table S2). The genome encodes two 5S rRNAs, three 16S rRNAs, two 23S rRNAs, 3 repeated regions (CRISPR) and 44 tRNAs. MAG GCA_014361205.1 contains slightly fewer CDS with 2,139 and 47 tRNAs (Table S2) (https://gtadb.ecogenomic.org/genome?gid=GCA_014361205.1). For comparison, the genome *Methanomethylovorans hollandica* GCF_0000328665.1 contains 2,636 CDS and 52 tRNA (https://gtadb.ecogenomic.org/genome?gid=GCF_0000328665.1). The amino acid annotation with *BlastKOALA* and subsequent visualization of the results with *KEGG Mapper Reconstruction Result* revealed that *M. thermophila* harbours all the enzymes that are necessary for acetoclastic and hydrogenotrophic methanogenesis, besides methylotrophic methanogenesis from methanol, mono-, di- and trimethylamine (Fig. 3). The predicted metabolic model (Tables S3 and S4) also showed that formate is a potential substrate for *M. thermophila* and indicates its potential for hydrogenotrophic methanogenesis.

However, even if all enzymes required for acetoclastic and hydrogenotrophic methanogenesis are present, no growth could be observed on H₂-CO₂ (80:20 2,000 mbar), a mixture of formate and H₂-CO₂ (80:20 vol/vol, 2,000 mbar) and acetate in lab tests (Fig. 4). These results confirm the findings of Jiang *et al.* [1], who could demonstrate that *M. thermophila* can only grow on methanol, mono-, di- and trimethylamine but not on H₂-CO₂ (80:20 vol/vol, 1,700 mbar), formate or acetate as sole carbon and energy source.

Funding information

This research was funded by the Austrian Science Fund and in part by the alpha+ Foundation [10.55776/P33838 and 10.55776/ESP7170024]. For open access purposes, the author has applied a CC BY public copyright license to any author accepted manuscript version arising from this submission.

Author contributions

M.W.: Investigation, data curation, conceptualization, visualization, methodology, writing – original draft and writing – review and editing. A.M.: Investigation and data curation. A.O.W.: Validation, resources, project administration, funding acquisition and supervision. E.M.P.: Data curation, conceptualization, validation, writing – review and editing, funding acquisition and supervision.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

- Jiang B, Parshina SN, van Doesburg W, Lomans BP, Stams AJM. *Methanomethylovorans thermophila* sp. nov., a thermophilic, methylotrophic methanogen from an anaerobic reactor fed with methanol. *Int J Syst Evol Microbiol* 2005;55:2465–2470.
- Lomans BP, Maas R, Luderer R, Op den Camp HJ, Pol A, *et al.* Isolation and characterization of *Methanomethylovorans hollandica* gen. nov., sp. nov., isolated from freshwater sediment, a methylotrophic methanogen able to grow on dimethyl sulfide and methanethiol. *Appl Environ Microbiol* 1999;65:3641–3650.

3. Iguchi A, Takemura Y, Danshita T, Kurihara T, Aoki M, et al. Isolation and physiological properties of methanogenic archaea that degrade tetramethylammonium hydroxide. *Appl Microbiol Biotechnol* 2023;107:3047–3056.
4. Woodcroft BJ, Aroney STN, Zhao R, Cunningham M, Mitchell JAM, et al. Comprehensive taxonomic identification of microbial species in metagenomic data using SingleM and Sandpiper. *Nat Biotechnol* 2025.
5. Rosentreter JA, Al-Haj AN, Fulweiler RW, Williamson P. Methane and nitrous oxide emissions complicate coastal blue carbon assessments. *Global Biogeochem Cycles* 2021;35:e2020GB006858.
6. Hall N, Wong WW, Lappan R, Ricci F, Jeppe KJ, et al. Coastal methane emissions driven by aerotolerant methanogens using seaweed and seagrass metabolites. *Nat Geosci* 2025;18:854–861.
7. Tsola SL, Prevodnik AA, Sinclair LF, Sanders IA, Economou CK, et al. Methanomethylivorans are the dominant dimethylsulfide-degrading methanogens in gravel and sandy river sediment microcosms. *Environ Microbiome* 2024;19:51.
8. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, et al. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res* 2022;50:D785–D794.
9. Wagner AO, Markt R, Mutschlechner M, Lackner N, Prem EM, et al. Medium preparation for the cultivation of microorganisms under strictly anaerobic/anoxic conditions. *J Vis Exp* 2019;150.
10. Wagner AO, Malin C, Lins P, Illmer P. Effects of various fatty acid amendments on a microbial digester community in batch culture. *Waste Manag* 2011;31:431–437.
11. Wunderer M, Markt R, Lackner N, Wagner AO. The glutamyl tail length of the cofactor F_{420} in the methanogenic Archaea *Methanosarcina thermophila* and *Methanoculleus thermophilus*. *Sci Total Environ* 2022;809:151112.
12. Antipov D, Korobeynikov A, McLean JS, Pevzner PA. hybridSPAdes: an algorithm for hybrid assembly of short and long reads. *Bioinformatics* 2016;32:1009–1015.
13. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes De Novo assembler. *CPin Bioinformatics* 2020;70:e102.
14. Prjibelski AD, Vasilinets I, Bankevich A, Gurevich A, Krivosheeva T, et al. ExSPAnDer: a universal repeat resolver for DNA fragment assembly. *Bioinformatics* 2014;30:i293–301.
15. Vasilinets I, Prjibelski AD, Gurevich A, Korobeynikov A, Pevzner PA. Assembling short reads from jumping libraries with large insert sizes. *Bioinformatics* 2015;31:3262–3268.
16. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, et al. Twelve years of SAMtools and BCFtools. *Gigascience* 2021;10:giab008.
17. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* 2018;34:i142–i150.
18. Sayers EW, Beck J, Bolton EE, Brister JR, Chan J, et al. Database resources of the National Center for Biotechnology Information in 2025. *Nucleic Acids Res* 2025;53:D20–D29.
19. Shimoyama Y. ANIclustermap: a tool for drawing ANI clustermap between all-vs-all microbial genomes. online verfügbar unter: Github. 2022. <https://github.com/moshi4/ANIclustermap>
20. Shaw J, Yu YW. Fast and robust metagenomic sequence comparison through sparse chaining with skani. *Nat Methods* 2023;20:1661–1665.
21. Sayers EW, Bolton EE, Fine AM, Kelly C, Kim S, et al. Database resources of the National Center for Biotechnology Information in 2026. *Nucleic Acids Res* 2026;54:D20–D27.
22. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* 2022;38:5315–5316.
23. Wong T, Ly-Trong N, Ren H, Baños H, Roger A, et al. IQ-TREE 3: Phylogenomic Inference Software using Complex Evolutionary Models. *Life Sciences* 2025. DOI: 10.32942/X2P62N.
24. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017;14:587–589.
25. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* 2018;35:518–522.
26. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 2021;49:W293–W296.
27. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat Methods* 2023;20:1203–1212.
28. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.
29. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* 2016;428:726–731.
30. Kanehisa M, Sato Y, Kawashima M. KEGG mapping tools for uncovering hidden features in biological data. *Protein Sci* 2022;31:47–53.
31. De Bernardini N, Zampieri G, Campanaro S, Zimmermann J, Waschina S, et al. pan-Draft: automated reconstruction of species-representative metabolic models from multiple genomes. *Genome Biol* 2024;25:S.
32. Zimmermann J, Kaleta C, Waschina S. gapseq: informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models. *Genome Biol* 2021;22:S.
33. Barrett T, Dowle M, Srinivasan A, Gorecki J, Chirico M, et al. CRAN: Contributed Packages; 2006
34. Mikhael PG, Wohlwend J, Yala A, Karstens L, Xiang J, et al. Sybil: a validated deep learning model to predict future lung cancer risk from a single low-dose chest computed tomography. *J Clin Oncol* 2023;41:2191–2200.
35. Gelius-Dietrich G, Desouki AA, Fritzscheier CJ, Lercher MJ. Sybil-efficient constraint-based modelling in R. *BMC Syst Biol* 2013;7:S.
36. Cha I-T, Min U-G, Kim S-J, Yim KJ, Roh SW, et al. *Methanomethylivorans uponensis* sp. nov., a methylotrophic methanogen isolated from wetland sediment. *Antonie Van Leeuwenhoek* 2013;104:1005–1012.
37. Parks DH, Chuvochina M, Chaumeil P-A, Rinke C, Mussig AJ, et al. A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat Biotechnol* 2020;38:1079–1086.