16S rRNA gene sequences of *Candidatus* Methylumidiphilus (*Methylococcales*), a putative methanotrophic genus in lakes and ponds

Running page head: 16S rRNA genes of *Candidatus* Methylumidiphilus

Antti J Rissanen[1*], Moritz Buck[2] & Sari Peura[3,4]

[1]Faculty of Engineering and Natural Sciences, Tampere University, Korkeakoulunkatu 6, FI-33720, Tampere, Finland

[2]Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, box 7050, SE-75007, Uppsala, Sweden

[3]Department of Forest Mycology and Plant Pathology, Science for Life Laboratory, Swedish University of Agricultural Science, Almas allé 5, SE-75651, Uppsala, Sweden

[4)]Current address: Swedish Nuclear Fuel and Waste Management Co., (SKB), P.O. Box 250, SE-10124 Stockholm, Sweden

***Corresponding author**: Antti J Rissanen, Faculty of Engineering and Natural Sciences, Tampere University, Korkeakoulunkatu 6, FI-33720, Tampere, Finland. Tel: +358 40 1981145; Fax: +358 3 3641392; e-mail: antti.rissanen@tuni.fi

Abstract

A putative novel methanotrophic genus, *Candidatus* Methylumidiphilus (*Methylococcales*), was recently shown to be ubiquitous and one of the most abundant methanotrophic genera in water columns of oxygen-stratified lakes and ponds of boreal and subarctic area. However, it has probably escaped detection in many previous studies that used 16S rRNA gene amplicon sequencing due to insufficient database coverage, which is because previously analysed metagenome assembled genomes (MAGs) affiliated with *Ca*. Methylumidiphilus do not contain 16S rRNA genes. Therefore, we screened MAGs affiliated with the genus for their 16S rRNA gene sequences in a recently published lake and pond MAG dataset. Among 66 MAGs classified as *Ca*. Methylumidiphilus (with completeness over 40% and contamination less than 5%) originating from lakes in Finland, Sweden and Switzerland as well as from ponds in Canada, we could find 5 MAGs each containing one 1532 bp long sequence spanning the V1-V9 regions of the 16S rRNA gene. After removal of sequence redundancy, this resulted in two unique 16S rRNA gene sequences. These sequences represented two different putative species, i.e. *Ca.* Methylumidiphilus alinenensis (Genbank accession: OK236221) as well as another so far unnamed species of *Ca*. Methylumidiphilus (Genbank accession: OK236220). We suggest that including these two sequences in reference databases will enhance 16S rRNA gene–based detection of members of this genus from environmental samples.

Keywords: *Candidatus* Methylumidiphilus, methanotroph, 16S rRNA gene, metagenome, lake, pond

1. INTRODUCTION

Methanotrophic bacteria are widely distributed and play a crucial role in consuming the greenhouse gas methane in natural (wetlands, lakes, oceans, soils) and anthropogenic (wastewater treatment plants, landfills) methane-producing ecosystems (Hanson & Hanson 1996, Kallistova et al. 2005). Currently, their identity, diversity and community structure are commonly studied using polymerase chain reaction (PCR)–based techniques, i.e. high-throughput amplicon sequencing and quantitative PCR, targeting the 16S rRNA gene or the *pmoA* gene encoding the beta subunit of particulate methane monooxygenase (Rissanen et al. 2018, Mayr et al. 2020a, b). The advantage of these PCR-based methods is their cost-effectiveness and speed in the analyses of multiple samples. Yet, recently, more expensive, shotgun metagenomic study methods, which overcome the problem of primer bias/mismatch inherent in PCR–based methods and which allow also insights into the genetic potential of the *in situ* bacterial community, have been employed in studies of methanotrophic communities (e.g. Oswald et al. 2017, Rissanen et al. 2018, Smith & Wrighton 2019, van Grinsven et al. 2020).

The results of DNA sequencing–based taxonomic analyses are dependent on the quality and taxonomic coverage of reference database(s), as the analyses are done by comparing the DNA sequences of samples for similarity with the DNA sequences deposited in databases. Using a PCR–free, 16S rRNA gene–independent shotgun metagenomic sequencing approach, we recently showed that a putative novel genus of methanotrophs, *Candidatus* Methylumidiphilus (order *Methylococcales*), was ubiquitous and one of the most abundant methanotrophic genera in water columns of oxygen–stratified lakes and ponds of boreal and subarctic area (Rissanen et al. 2018, 2020, Martin et al. 2021). The first putative species of this genus was named as *Candidatus* Methyloumidiphilus alinensis [the name later proposed to be changed to *Ca*.

Methylumidiphilus alinenensis (Oren et al. 2020)], which was represented by an abundant metagenome–assembled genome (MAG) in the water samples of boreal Lake Alinen Mustajärvi (Rissanen et al. 2018). Furthermore, in the same study, an abundant operational taxonomic unit (OTU), which was detected in simultaneous high–throughput 16S rRNA gene amplicon sequencing analysis, was affiliated with the genus based on its identical position in the phylogenetic tree with the position of MAG of *Ca*. Methylumidiphilus alinenensis in the phylogenomic tree (Rissanen et al. 2018). Interestingly, analyses by Rissanen et al. (2018) suggested that the genus had probably not been classified as a methanotroph (*Methylococcales*) at all (i.e. it was classified as unclassified *Gammaproteobacteria*) in previous 16S rRNA gene–based analyses using older Silva 119 (released 24 July 2014) and 123 (23 July 2015) databases, while starting with Silva 128 database (29 Sep 2016) it was classified correctly as *Methylococcales*. In our recent study, where we compared the results of taxonomic classification of shotgun metagenomic reads of subarctic and boreal lakes and ponds between a 16S rRNA gene–independent and a 16S rRNA gene–dependent approach, which used Silva 132 database (13 Dec 2017), the results suggested that 16S rRNA gene sequences of *Ca*. Methylumidiphilus were classified as unknown *Methylococcales* (Martin et al. 2021). To aid in correctly classifying the 16S rRNA genes of this genus, a previously published clone library sequence from Lake Alinen Mustajärvi (Genbank, HE616416, 830bp) was determined to represent *Ca*. Methylumidiphilus alinenensis based on its identical position in the phylogenetic tree with the position of MAG of *Ca*. Methylumidiphilus alinenensis in the phylogenomic tree as well as on its high identity (99.7 %) with the representative sequence (288 bp) of the afore-mentioned 16S rRNA gene–based OTU affiliated with the species (Rissanen et al. 2018). HE616416 was then used as a database sequence in some subsequent 16S rRNA gene analyses (Thamdrup et al.

2019, Rissanen et al. 2020). However, the 16S rRNA gene–based phylogenetic position of *Ca*. Methylumidiphilus remains to be confirmed (Knief 2019), as 16S rRNA gene sequences are not available from the previously reconstructed MAGs representing the genus (Rissanen et al. 2018, 2020). In addition, HE616416 covers only V1-V5 regions of the 16S rRNA gene making it impossible to use it as a reference sequence in studies focusing on V6-V9 regions. Modern PCR–based amplicon sequencing analyses using long-read sequencing technologies (PacBio or Oxford Nanopore) covering the whole V1-V9 regions of 16S rRNA gene as well as PCR-free shotgun metagenomic–based 16S rRNA gene analyses would also require full-length or almost full-length 16S rRNA gene sequences, thus as references.

Metagenomic assembly and binning approaches typically reconstruct 16S rRNA genes of only part of the MAGs of the target organisms, for example of lake methanotrophs (Oswald et al. 2017, Rissanen et al. 2020, van Grinsven et al. 2020). Therefore, screening of multiple MAGs representing the organism(s) of interest is needed to find MAGs containing 16S rRNA genes. The recently published shotgun metagenomic dataset from water columns of lakes and ponds by Buck et al. (2021), on which the aforementioned results by Martin et al. (2021) on the ubiquitousness and abundance of *Ca*. Methylumidiphilus were based on, provides a great source of MAGs taxonomically affiliated with *Ca*. Methylumidiphilus. Therefore, with the aim to provide 16S rRNA gene sequences representing *Ca*. Methylumidiphilus to be included in reference databases, we screened these MAGs for their 16S rRNA genes.

2. MATERIALS AND METHODS

We used previously published MAG dataset from 41 stratified lakes and ponds mainly located in the boreal and subarctic regions, but also from one tropical reservoir and one temperate lake

(Buck et al. 2021). See Buck et al. (2021) on detailed report of the sample collection, DNA extraction, library preparation, sequencing and bioinformatic analyses (trimming/filtering, assembly, metagenomic binning). Furthermore, Buck et al. (2021) used checkM (v. 1.0.13) for assessing the prokaryotic completeness and redundancy of the MAGs (Parks et al. 2015), while GTDB-Tk (version 102 with database release 89) (Parks et al. 2018) as well as SourMASH's lca classifier (Brown & Irber 2016) were used for their taxonomic classification. Finally, Buck et al. (2021) clustered the MAGs, starting with 40% complete genomes with less than 5% contamination, into metagenomic operational taxonomic units (mOTUs) at 95 % level of average nucleotide identity (ANI) calculated using fastANI (v. 1.3) (Jain et al. 2018).

For our analyses, we chose MAGs with genus level taxonomic classification of "d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Methylococcales;f__Methylococcaceae;g__AMB10-2013", which had completeness over 40% and contamination less than 5%. The genus level name "g__AMB10-2013" denotes the MAG of *Candidatus* Methylumidiphilus alinenensis (GCA_003242955) discovered from water column of boreal lake Alinen Mustajärvi (Rissanen et al. 2018). The chosen MAGs were functionally annotated using Prokka (v. 1.14.6) (Seemann 2014), which included detection of rRNA genes using barrnap (v. 0.9) (Seemann 2018). Phylogenetic trees based on 16S rRNA genes were built from gene alignments generated in Mega X (Kumar et al. 2018) using the maximum–likelihood algorithm (GTRGAMMA–model) with 100 bootstrap replicates in RAxML (v. 8.2.12) (Stamatakis 2014). Furthermore, phylogenomic trees including reference genomes as well as representative MAGs of mOTUs affiliated to *Ca*. Methylumidiphilus (i.e. with genus level taxonomic classification of "g__AMB10-2013") were built from protein alignments generated in PhyloPhlAn (v. 3.0.58; with PhyloPhlAn database incl. 400 universal marker genes) (Asnicar et al. 2020) using the

maximum-likelihood algorithm (PROTCATLG–model) with 100 bootstrap replicates in RAxML (v. 8.2.12) (Stamatakis 2014).

3. RESULTS AND DISCUSSION

In Buck et al. (2021) dataset, there were 66 MAGs, which had completeness over 40% and contamination less than 5% and with taxonomic assignment to *Ca.* Methylumidiphilus (i.e. with genus level taxonomic classification: "g__AMB10-2013"). These MAGs were classified into 12 mOTUs (Fig. 1), whose representative genomes originated from lakes in Finland, Sweden and Switzerland as well as from ponds in Canada (Buck et al. 2021). Hence, besides being present in boreal and subarctic lakes and ponds as already shown by Martin et al. (2021), *Ca*. Methylumidiphilus was also noticed to inhabit a lake in temperate area, i.e Lake Loclat in Switzerland (Buck et al. 2021). Of the mOTUs, mOTU 0341, 2711, 1471, 1599 and 2021 were represented by more than one MAG, i.e. 42, 6, 4, 4 and 3 MAGs, respectively, while each of the other mOTUs included only one MAG (Fig. 1). Our previously studied MAGs of *Ca*. Methylumidiphilus originating from boreal lakes, i.e. *Ca*. Methylumidiphilus alinenensis from Lake Alinen Mustajärvi and bin-0959 from Lake Lovojärvi, were also included in phylogenomic tree analysis, with a result indicating that they belong to mOTUs 0341 and 1599, respectively (Fig. 1) (Rissanen et al. 2018, 2020).

Fragments of 16S rRNA genes were found in 15 out of the 66 studied MAGs. Of these, 6 MAGs included almost full length 16S rRNA gene sequences (1530-1532 bp; the length of full length 16S rRNA gene is about 1550 bp) and were chosen for further analyses, while all other had lengths less than 1200 bp. In the preliminary taxonomic classification analyses using blastn (Altschul et al. 1990), one of the 16S rRNA gene sequences (from bin-1515 GCA_903920655.1)

was only distantly related (with 85.8 % identity) to the partial 16S rRNA gene sequence HE616416 (length 830 bp) suggested to represent *Ca*. Methylumidiphilus alinenensis (Rissanen et al. 2018), and was actually most closely affiliated with *Methylobacter* (98.5 % identity with *Methylobacter tundripaludum* SV96, NR_042107), and hence probably came from a wrongly binned contig. In contrast, the other 5 of the 16S rRNA gene sequences had high identity with HE616416 (96.0–99.6 % identity) as well as with the shorter representative sequences of the 16S rRNA gene–based OTUs suggested to represent *Ca*. Methylumidiphilus in previously studied Lake Alinen Mustajärvi, i.e. OTU 9 (length 288 bp) (identity 97.2–100 %) (Rissanen et al. 2018), and Lake Lovojärvi, i.e. OTU 229 (length 253 bp) (identity 94.5–94.9 %) (Rissanen et al. 2020), and were, thus, chosen for further analyses. The phylogenetic tree analysis confirmed the phylogenetic position of these 16S rRNA gene sequences as they formed a distinct cluster, with *Methyloterricola* and *Methylospira* as their neighbouring genera (Fig. 2), which agrees with previous phylogenetic analyses with HE616416 (Rissanen et al. 2018, Knief 2019). The 16S rRNA gene sequences formed two clusters, one including three identical 16S rRNA gene sequences representing mOTU 0341 (submitted to Genbank with accession: OK236221), and the other including two identical 16S rRNA gene sequences representing mOTU 2711 (Genbank accession: OK236220) (Fig. 2). The blastn-analysed identities of the 16S rRNA gene sequences of these clusters to those of *Methylospira palustris* (90.9% and 90.8% identity for mOTUs 2711 and 0341, respectively) and *Methyloterricola oryzae* (91.1% and 91.6% identity for mOTUs 2711 and 0341, respectively) were much lower than the suggested 94.5% identity threshold to delineate different genera (Yarza et al. 2014), which further confirms their taxonomic assignment to a different genus than *Methylospira* and *Methyloterricola*. In addition, their identities to each other (97.5% identity between mOTU 2711 and 0341) were much higher than 94.5%, suggesting

194 that they belong to same genus. Phylogenomic analyses as well as the high identity of the 16S
195 rRNA gene sequences of mOTU 0341 to HE616416 (99.6% identity) further suggests that
196 mOTU 0341 represents *Ca*. Methylumidiphilus alinenensis (Fig. 1). In addition, both
197 phylogenomic as well as 16S rRNA gene analyses suggest that mOTU 2711 represents a
198 different, so far unnamed, species of *Ca*. Methylumidiphilus.

199 In this study, we provided for the first time almost full length 16S rRNA gene sequences
200 representing the putative methanotrophic genus, *Ca*. Methylumidiphilus, which is ubiquitous in
201 water columns of lakes and ponds of boreal and subarctic area (Buck et al. 2021, Martin et al.
202 2021), and according to this study, is also present in a temperate lake, Lake Loclat, in
203 Switzerland (Fig. 1). Furthermore, the distribution of *Ca*. Methylumidiphilus very likely extends
204 to also other ecosystems, as suggested by recent *pmoA* gene–based phylogenetic analyses, which
205 show that the *pmoA* gene of MAG of *Ca*. Methylumidiphilus alinenensis belongs to the Lake
206 Washington (LW)–cluster, which includes *pmoA* sequences from wetlands, peatlands and lake
207 sediments (Rissanen et al. 2018, 2020, Knief et al. 2019). Hence, we suggest that including the
208 provided 16S rRNA gene sequences in reference databases will enhance the 16S rRNA gene–
209 based detection of members of *Ca*. Methylumidiphilus in further studies of microbial
210 communities of lakes and other aquatic ecosystems.

211

212 **Acknowledgements**

213 This study was supported by Kone Foundation (Grant No. 201803224) for AJR. The sequencing
214 was funded by a grant from the Science for Life Laboratory biodiversity program and SciLifeLab
215 fellows program. The computations were performed on resources provided by SNIC through

Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Project SNIC snic2020-5-19 as well as on resources provided by CSC – IT Center for Science.

Literature cited

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Asnicar F, Thomas AM, Beghini F, Mengoni C, Manara S, Manghi P, Zhu Q, Bolzan M, Cumbo F, May U, Sanders JG, Zolfo M, Kopylova E, Pasolli E, Knight R, Mirarab S, Huttenhower C, Segata N (2020) Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. Nat Commun 11:2500

Brown C., Irber L (2016) Sourmash: a library for MinHash sketching of DNA. J Open Source Softw 1:27

Buck M, Garcia SL, Fernandez L, Martin G, Martinez-Rodriguez GA, Saarenheimo J, Zopfi J, Bertilsson S, Peura S (2021) Comprehensive dataset of shotgun metagenomes from oxygen stratified freshwater lakes and ponds. Sci Data 8:131

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

Hanson RS, Hanson TE (1996) Methanotrophic bacteria. Microbiol Rev 60:439–471

Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S (2018) High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114

Kallistova AY, Kevbrina M V, Nekrasova VK, Glagolev M V, Serebryanaya MI, Nozhevnikova AN (2005) Methane oxidation in landfill cover soil. Microbiology 74:608–614

Knief C (2019) Diversity of methane cycling microorganisms in soils and their relation to oxygen. Curr Issues Mol Biol 33:23–56

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549

Martin G, Rissanen AJ, Garcia SL, Mehrshad M, Buck M, Peura S (2021) *Candidatus* Methylumidiphilus drives peaks in methanotrophic relative abundance in stratified lakes and ponds across Northern landscapes. Front Microbiol 12:669937

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

Mayr MJ, Zimmermann M, Guggenheim C, Brand A, Bürgmann H (2020b) Niche partitioning of methane-oxidizing bacteria along the oxygen–methane counter gradient of stratified

lakes. ISME J 14:274–287

Oren A, Garrity GM, Parker CT, Chuvochina M, Trujillo ME (2020) Lists of names of prokaryotic Candidatus taxa. Int J Syst Evol Microbiol 70:3956–4042

Oswald K, Graf JS, Littmann S, Tienken D, Brand A, Wehrli B, Albertsen M, Daims H, Wagner M, Kuypers MMM, Schubert CJ, Milucka J (2017) *Crenothrix* are major methane consumers in stratified lakes. ISME J 11:2124-2140

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

Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055

Rissanen A, Saarenheimo J, Tiirola M, Peura S, SL A, Karvinen A, Nykänen H (2018) Gammaproteobacterial methanotrophs dominate methanotrophy in aerobic and anaerobic layers of boreal lake waters. Aquat Microb Ecol 81:257–276

Rissanen AJ, Saarela T, Jäntti H, Buck M, Peura S, Aalto SL, Ojala A, Pumpanen J, Tiirola M, Elvert M, Nykänen H (2020) Vertical stratification patterns of methanotrophs and their genetic controllers in water columns of oxygen-stratified boreal lakes. FEMS Microbiol Ecol 97:fiaa252

Seemann T (2018) Barrnap: BAsic Rapid Ribosomal RNA Predictor. https://github.com/tseemann/barrnap

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

Smith GJ, Wrighton KC (2019) Metagenomic approaches unearth methanotroph phylogenetic and metabolic diversity. Curr Issues Mol Biol 33:57–84

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

Thamdrup B, Steinsdóttir HGR, Bertagnolli AD, Padilla CC, Patin N V, Garcia-Robledo E, Bristow LA, Stewart FJ (2019) Anaerobic methane oxidation is an important sink for methane in the ocean's largest oxygen minimum zone. Limnol Oceanogr 64:2569–2585

Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer K-H, Whitman WB, Euzéby J, Amann R, Rosselló-Móra R (2014) Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nat Rev Microbiol 12:635–645

288

289

290

291

292

293

294

295

296

297

298

299

Fig.1

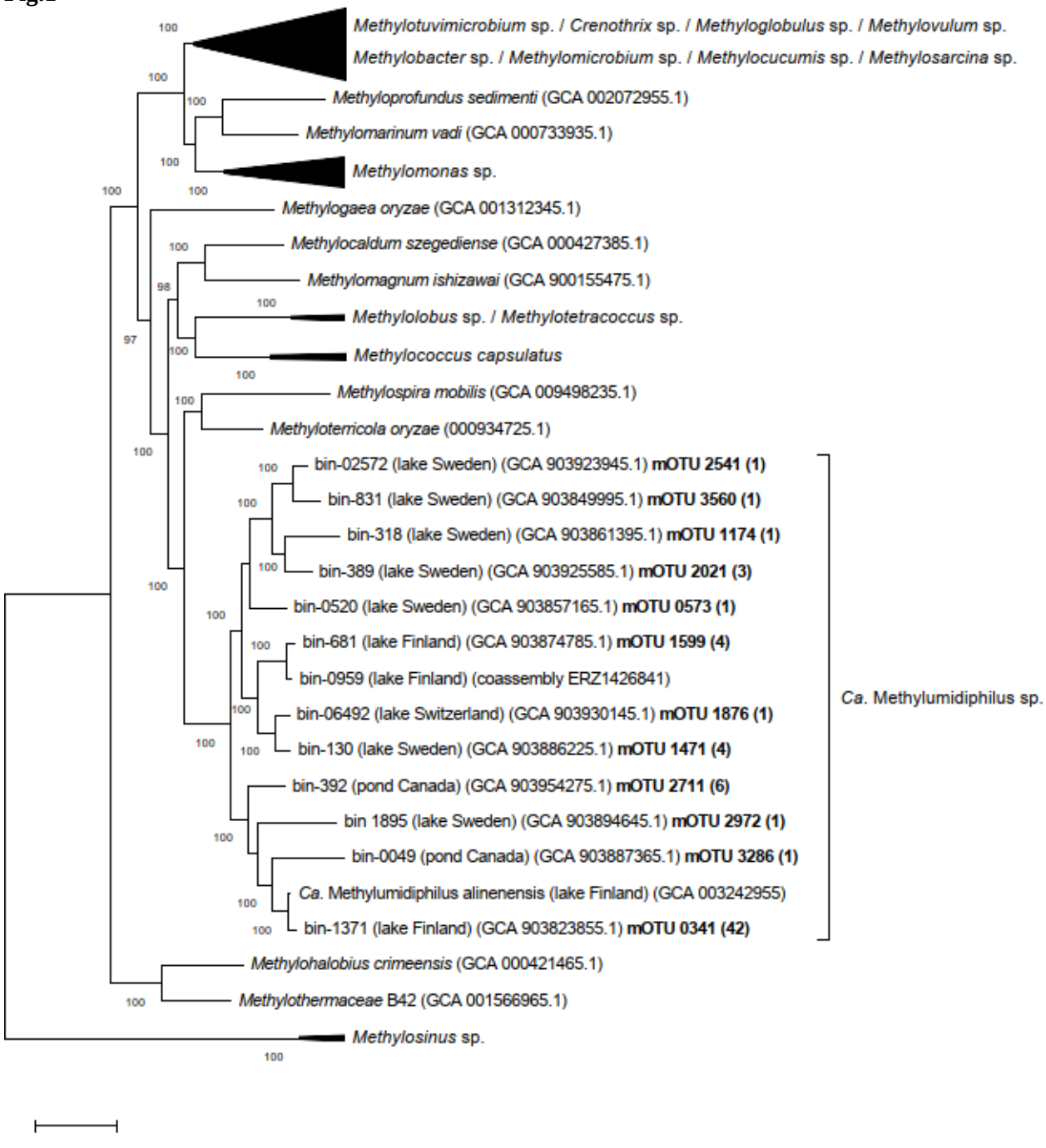

Fig. 1. Phylogenomic tree (PhyloPhlAn) of *Methylococcales* (outgroup = alphaproteobacterial genus *Methylosinus*). The tree shows representative metagenome assembled genomes (MAGs) of metagenomic operational taxonomic units (mOTU) affiliated with *Candidatus* Methylumidiphilus in the Buck et al. (2021) dataset, as well as MAGs, which we analysed previously, i.e. *Ca*. Methylumidiphilus alinenensis and bin-0959 (Rissanen et al. 2018, 2020). mOTU number as well as the number of MAGs belonging to each of the mOTUs (in brackets after mOTU number) are highlighted with bold text. The tree was constructed using the maximum-likelihood algorithm with the PROTCATLG–model in RAxML (v.

8.2.12) (Stamatakis 2014). The numbers at the nodes indicate the percentage of occurrence in 100 bootstrapped trees (bootstrap values ≥70% are shown). The tree was collapsed from some of the branches to make the phylogenomic position of *Ca*. Methylumidiphilus sp. visually clearer

Fig.2

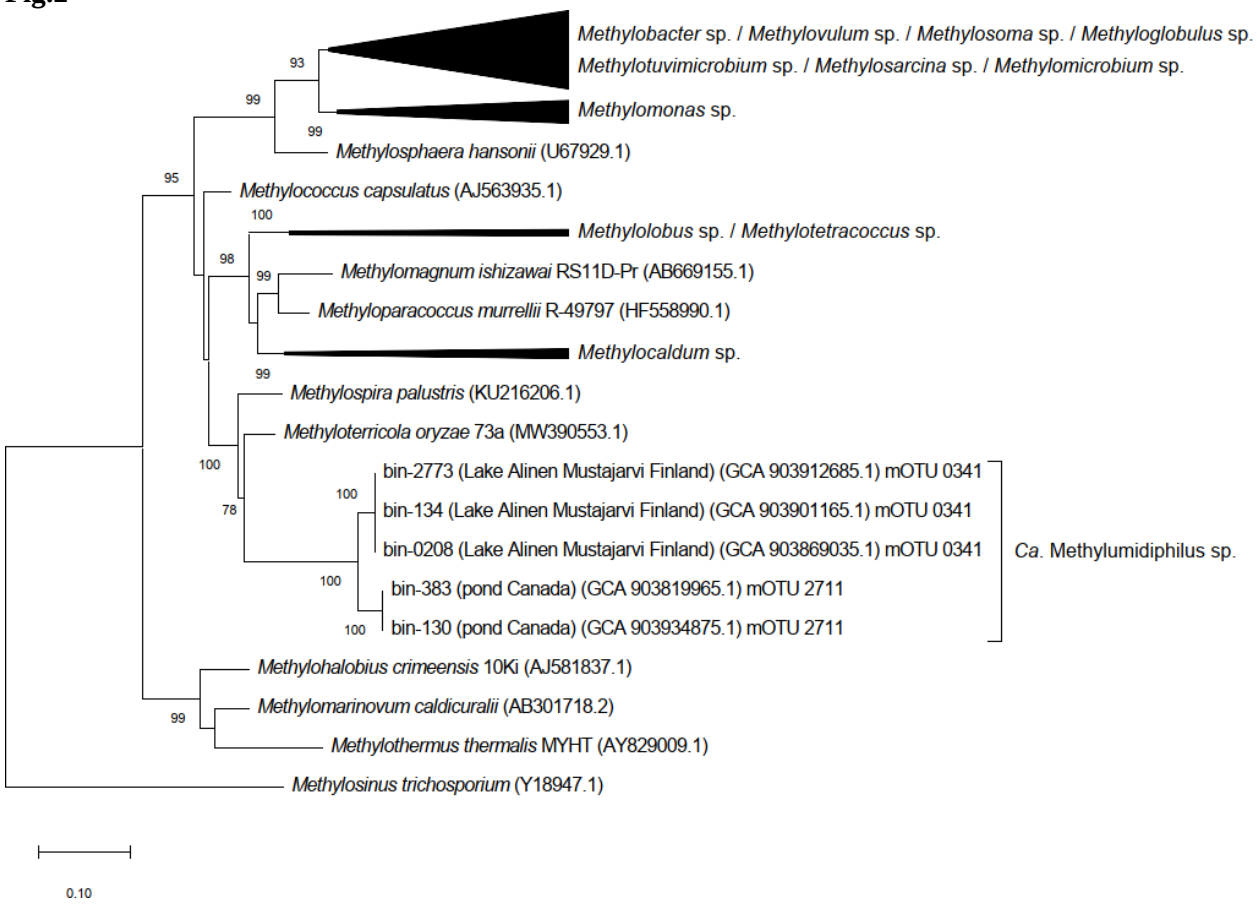

Fig. 2. Phylogenetic tree based on 16S rRNA genes of *Methylococcales* (outgroup = alphaproteobacterial genus *Methylosinus*). The tree shows 16S rRNA gene sequences, spanning V1-V9 regions of the 16S rRNA gene, detected in 5 MAGs affiliated with *Ca*. Methylumidiphilus. The mOTU number of the MAGs is also shown (see Fig. 1). The tree was constructed using the maximum-likelihood algorithm with the GTRGAMMA–model in RAxML (v. 8.2.12) (Stamatakis 2014). The numbers at the nodes indicate the percentage of occurrence in 100 bootstrapped trees (bootstrap values ≥70% are shown). The tree was collapsed from some of the branches to make the phylogenetic position of *Ca*. Methylumidiphilus sp. visually clearer