environmental microbiology reports

Environmental Microbiology Reports (2021) **00**(00), 00–00 doi:10.1111/1758-2229.12927

Brief Report

Comparative genomics analyses indicate differential methylated amine utilization trait within members of the genus *Gemmobacter*

Eileen Kröber,[1†*] Mark R. Cunningham,[2†] Julianna Peixoto,[3] Lewis Spurgin,[4] Daniela Wischer,[4] Ricardo Kruger [3] and Deepak Kumaresan [2*]

[1]*Department of Symbiosis, Max-Planck Institute for Marine Microbiology, Bremen, Germany.*

[2]*School of Biological Sciences, Institute for Global Food Security, Queen's University Belfast, Belfast, UK.*

[3]*Department of Cellular Biology, University of Brasília, Brasília, Brazil.*

[4]*School of Environmental Sciences, University of East Anglia, Norwich, UK.*

Summary

Methylated amines are ubiquitous in the environment and play a role in regulating the earth's climate via a set of complex biological and chemical reactions. Microbial degradation of these compounds is thought to be a major sink. Recently we isolated a facultative methylotroph, *Gemmobacter* sp. LW-1, an isolate from the unique environment Movile Cave, Romania, which is capable of methylated amine utilization as a carbon source. Here, using a comparative genomics approach, we investigate how widespread methylated amine utilization is within members of the bacterial genus *Gemmobacter*. Seven genomes of different *Gemmobacter* species isolated from diverse environments, such as activated sludge, fresh water, sulphuric cave waters (Movile Cave) and the marine environment were available from the public repositories and used for the analysis. Our results indicate that methylamine utilization is a distinctive feature of selected members of the genus *Gemmobacter*, namely *G. aquatilis*, *G. lutimaris*, *G.* sp. HYN0069, *G. caeni* and *G.* sp. LW-1 have the genetic potential while others (*G. megaterium* and *G. nectariphilus*) have not.

Received 8 September, 2020; accepted 11 January, 2021. *For correspondence. E-mail ekroeber@mpi-bremen.de; Tel. +4942120288230. E-mail d.kumaresan@qub.ac.uk; Tel. +442890972181. †These authors contributed equally to this study.

Introduction

Methylated amines (MAs) are ubiquitous in the environment with a variety of natural and anthropogenic sources including the oceans, vegetation, sediments and organic-rich soils, animal husbandry, food industry, pesticides, sewage and automobiles, to mention only a few (Schade and Crutzen, 1995; Latypova *et al*., 2010; Ge *et al*., 2011). Methylated amines are also known to influence Earth's climate, via a series of complex biological and chemical interactions (Carpenter *et al*., 2012). Some of the most abundant methylated amines found in the atmosphere are trimethylamine (TMA), dimethylamine (DMA) and monomethylamine (MMA) (Ge *et al*., 2011). Microbial metabolism of methylated amines involves both aerobic and anaerobic microorganisms, for example some methanogenic archaea such as *Methanosarcina* and *Methanomicrobium* can use MAs to produce methane (Burke *et al*., 1998; Liu and Whitman, 2008; Lyimo *et al*., 2009) while Gram-positive and Gram-negative methylotrophic bacteria can use MAs as carbon and nitrogen source (Chen *et al*., 2010a). Previously, MAs were typically associated with marine ecosystems as they are by-products of degradation of osmolytic chemicals such as glycine betaine, carnitine, choline and trimethylamine N-oxide (Chen *et al*., 2010b). However, recent studies have reported the detection and activity of aerobic methylotrophic bacteria that utilize MAs in a variety of natural and engineered environments (Chen *et al*., 2009; Chistoserdova *et al*., 2009; Chistoserdova, 2011; Ge *et al*., 2011; Wischer *et al*., 2015) and could play a major role in global C and N budgets.

Aerobic methylotrophs are a polyphyletic group of microorganisms capable of utilizing one-carbon (C_1) compounds such as methane, methanol or methylated amines as their sole source of carbon and energy (Anthony, 1982; Lidstrom, 2006; Chistoserdova *et al*., 2009). Methylotrophs can degrade TMA to DMA by using the enzymes TMA

dehydrogenase, TMA monooxygenase or TMA methyltransferase (under anaerobic conditions by methylotrophic methanogens), encoded by the genes *tdm*, *tmm* and *mtt*, respectively (Paul *et al*., 2000; Chen, 2012; Lidbury *et al*., 2014). The enzymes DMA dehydrogenase (*dmd*) or DMA monooxygenase (*dmmDABC*) modulate the conversion of DMA to MMA (Lidstrom, 2006; Chen, 2012; Lidbury *et al*., 2017). Two distinct pathways have been characterized for the oxidation of MMA (Chistoserdova, 2011). The direct MMA-oxidation pathway mediated by a single enzyme (MMA dehydrogenase in Gram-negative bacteria and MMA oxidase in Gram-positive bacteria) converts MMA to formaldehyde and releases ammonium (McIntire *et al*., 1991; Chistoserdova *et al*., 1994). The alternate pathway, referred to as the *N*-methylglutamate (NMG) pathway or indirect MMA-oxidation pathway, is mediated by three individual enzymes via the oxidation of MMA to gamma-glutamylmethylamide (GMA) and its further degradation to *N*-methylglutamate (NMG) and 5,10-methylenetetrahydrofolate ($CH_2 = H_4F$) (Latypova *et al*., 2010; Chistoserdova, 2011). A stepwise conversion of MMA in the NMG pathway is modulated by the enzymes GMA synthetase (*gmaS*), 'NMG synthase' (*mgsABC*) and NMG dehydrogenase (*mgdABCD*) (Latypova *et al*., 2010; Chen *et al*., 2010a). The capability to use MMA not only as a source for carbon but also for nitrogen is widespread in bacteria. Notably, the NMG pathway is not only restricted to methylotrophs but also present in non-methylotrophic bacteria that use MMA as a nitrogen as an energy source but not as a carbon source (Chen *et al*., 2010b; Chen, 2012; Lidbury *et al*., 2015a; Taubert *et al*., 2017).

In a recent study, we isolated an alphaproteobacterial facultative methylotrophic bacterium, *Gemmobacter* sp. LW-1 (recently renamed from *Catellibacterium* (Chen *et al*., 2013)) from the Movile Cave ecosystem (Mangalia, Romania) (Kumaresan *et al*., 2014) that can use methylated amines as both carbon and nitrogen source (Wischer *et al*., 2015) and subsequently obtained its genome sequence (Kumaresan *et al*., 2015). Using a ^{13}C-MMA DNA-based stable-isotope probing (SIP) experiment, we also showed that *Gemmobacter* sp. LW-1 was indeed an active MMA utilizer in microbial mats from this environment (Wischer *et al*., 2015). This was the first report of methylated amine utilization in a member of the bacterial genus *Gemmobacter*. However, growth on C_1 compounds (methanol and formate) has been reported for the genus *Gemmobacter*, for example in *G. caeni* (Zheng *et al*., 2011). The genus *Gemmobacter* (family *Rhodobacteraceae*) currently comprises 17 species (summarized in Table S1): *Gemmobacter megaterium* (Liu *et al*., 2014), *G. nectariphilum* (Tanaka *et al*., 2004; Chen *et al*., 2013), *G. aquatilis* (Rothe *et al*., 1987), *G. caeni* (Zheng *et al*., 2011; Chen *et al*., 2013), *G. aquaticus* (Liu *et al*., 2010; Chen *et al*., 2013), *G. nanjingense* (Zhang *et al*., 2012; Chen *et al*., 2013), *G. intermedius* (Kämpfer *et al*., 2015), *G. lanyuensis* (Sheu *et al*., 2013a), *G. tilapiae* (Sheu *et al*., 2013b), *G. fontiphilus* (Chen *et al*., 2013), *G. straminiformis* (Kang *et al*., 2017), *G. serpentinus* (Lim *et al*., 2020), *G. aquarius* (Baek *et al*., 2020), *G. caeruleus* (Qu *et al*., 2020), *G. lutimaris* (Yoo *et al*., 2019), *G. aestuarii* (Hameed *et al*., 2020) and *G. changlensis* (Hameed *et al*., 2020). These species were isolated from a wide range of environments including fresh water environments (freshwater pond (Rothe *et al*., 1987, Sheu *et al*., 2013a), freshwater spring (Chen *et al*., 2013; Sheu *et al*., 2013b)), coastal planktonic seaweed (Liu *et al*., 2014), white stork nestling (Kämpfer *et al*., 2015), waste water and activated sludge (Tanaka *et al*., 2004; Zheng *et al*., 2011; Zhang *et al*., 2012), suggesting that members of the genus *Gemmobacter* are widely distributed in engineered and natural environments.

Here, using a comparative genomics approach we study how widespread methylated amine utilization trait (i.e. metabolic potential) is within the members of the genus *Gemmobacter*. We used seven isolate genomes (available in public repositories at the time of this study – March 2020) for members within the genus *Gemmobacter* (*G*. sp. LW-1, *G. caeni*, *G. aquatilis*, *G. nectariphilus*, *G. megaterium*, *G*. sp. HYN0069 *and G. lutimaris* YJ-T1-11) alongside genomes of closely related organisms within the family *Rhodobacteraceae* to show that the methylated amine utilization trait is a distinctive feature within selected members of the genus *Gemmobacter*.

Materials and methods

Genome data acquisition

Seven *Gemmobacter* genomes (*G. caeni*, *G. aquatilis*, *G. nectariphilus*, *G. megaterium*, *Gemmobacter* sp. LW-1, *G*. sp. HYN0069 *and G. lutimaris* YJ-T1-11) available (accessed in March 2020) through the Integrated Microbial Genomes (IMG) database (https://img.jgi.doe.gov/) were used for comparative genome analysis (Markowitz *et al*., 2013). Accession numbers and genome characteristics are listed in Table S2.

Phylogenetic and phylogenomic analysis

Phylogenetic relatedness between the different members of the genus *Gemmobacter* was determined using phylogenetic trees constructed from 16S rRNA gene (nucleotide) and metabolic gene sequences (*gmaS* and *mauA*; amino acids) involved in MMA utilization. RNAmmer (Lagesen *et al*., 2007) was used to retrieve 16S rRNA gene sequences from the genome sequences. Multiple

sequence alignment of 16S rRNA gene sequences from *Gemmobacter* genomes along with related sequences (retrieved from NCBI) was performed using the SINA (v1.2.11) alignment service via ARB-SILVA (https://www.arb-silva.de/aligner/) (Pruesse *et al*., 2007; Pruesse *et al*., 2012) and subsequently imported into MEGA7 (Kumar *et al*., 2016) to construct a maximum-likelihood nucleotide-based phylogenetic tree (Saitou and Nei, 1987) using the Tamura-Nei model for computing evolutionary distances and bootstrapping with 1000 replicates. To determine phylogenetic affiliations for the protein encoding genes *gmaS* and *mauA*, gene sequences retrieved from the genome sequences were aligned to homologous sequences retrieved from the NCBI Genbank database using Basic Local Alignment Search Tool (BLAST, blastx) (Altschul *et al*., 1990) and curated *gmaS* sequences used for primer design in our previous study (Wischer *et al*., 2015). Amino acid sequences were aligned in MEGA7 (Kumar *et al*., 2016) using ClustalW (Thompson *et al*., 1994) and the alignment was subsequently used to construct maximum likelihood phylogenetic trees based on the JTT matrix-based model (Jones *et al*., 1992). Bootstrap analysis was performed with 1000 replicates to provide confidence estimates for phylogenetic tree topologies (Felsenstein, 1985).

We inferred the phylogenomics of 11 bacterial species – *G. aquatilis*, *G. caeni*, *G*. sp. *LW-1*, *G. megabacterium* DSM-26375, *G. nectariphilus* DSM-15620, *G*. sp HYN00069, *G. lutimaris* YJ-T1-11, *Paracoccus denitrificans* PD1222, *Roseovarius* sp. TM103, *Rhodobacter sphaeroides* 241 and *Methylobacterium nodulans* ORS2060 using 117 single copy phylogenetic marker genes specific to organisms within the class *Alphaproteobacteria* (Emms and Kelly, 2015) via the GToTree (v1.5.22) pipeline (Lee, 2019). For each set of protein sequences retrieved using the HMMER3 tool (Eddy, 2011), multiple alignments were produced using MUSCLE (v3.8.31, default settings) (Edgar, 2004). Subsequently, conserved alignment blocks were identified using trimal (v1.4) (Capella-Gutiérrez *et al*., 2009) with the option -automated1. The phylogenetic reconstruction analysis using the final concatenated alignment was constructed using FastTree2 (v2.1.10) (Price *et al*., 2010) using default settings and 1000 bootstraps. Genome taxonomy was also confirmed using the toolkit GTDB-TK (v1.3.0) (Chaumeil *et al*., 2019).

In order to assess the environmental distribution of the genus *Gemmobacter*, we used MAPseq (Matias Rodrigues *et al*., 2017) (https://beta.microbeatlas.org) to survey the relative abundance in different amplicon and metagenome datasets based on 16S rRNA gene sequences (query sequence: *Gemmobacter aquatilis* DSM3857 (NR_104740.1) at 97% cut-off; accessed in March 2020). We also determined the relative abundance of *Gemmobacter* in 16S rRNA gene sequence datasets from four distinct ecosystems: (i) Reactor facilities for treating municipal wastewater (SRR870266), (ii) epiphytic bacterial communities in *Hydrilla verticillate* (SRR2033800), (iii) human skin microbiome (interdigital web space; SRR1704943) and (iv) dry valley lakes in high altitude (SRR953422).

Comparative genomic analyses

CGView Comparison Tool (CCT) was used to visually compare the genomes within the genus *Gemmobacter* (Grant *et al*., 2012). CCT utilizes BLAST to compare the genomes and the BLAST results are presented in a DNA-based graphical map (Grant *et al*., 2012). Average Nucleotide/Amino Acid Identity (ANI/AAI) (Rodriguez-R and Konstantinidis, 2016) between different genomes was estimated using one-way ANI (best hit) and two-way ANI (reciprocal best hit) based on Goris *et al*. (Goris *et al*., 2007). In addition the whole-genome based average nucleotide identity (gANI) and the $p_r^{\text{intra-species}}$ value were determined for *G*. sp. LW-1 and *G. caeni* (these two genomes revealed the closest ANI) based on Konstantinidis and Tiedje (Konstantinidis and Tiedje, 2005) via the Joint Genome Institute (JGI) platform (https://ani.jgi-psf.org/html/home.php; Version 0.3, April 2014). In order to determine if two genomes belong to the same species, the computation of empirical probabilities ($p_r^{\text{intra-species}}$) can be calculated as follows,

$$p_r^{\text{intra-species}}[AF = a, ANI = b] = p_r^{\text{intra-species}}[AF = a] * p_r^{\text{intra-species}}[ANI = b | AF = a]$$

AF represents alignment fraction. Pan-genome analysis for determination of core and dispensable genes and singletons (unique genes) was carried out using EDGAR v2.0 (Blom *et al*., 2009) using default settings. Estimation of genomic completeness and contamination was carried out using the CheckM (v 1.3.0) program (Parks *et al*., 2015).

In order to compare the genetic potential for methylated amine utilization within the available *Gemmobacter* genomes, known protein sequences involved in methylated amine utilization pathways (Latypova *et al*., 2010; Chen, 2012) were used as query sequences through the BLAST (blastp) program (Altschul *et al*., 1990) available within the Rapid Annotation using Subsystem Technology (RAST) server (Aziz *et al*., 2008). The list of protein queries used is given in Table S3.

Results and discussion

Analysis of phylogenetic relatedness and environmental distribution

The phylogenetic relatedness of members within the genus *Gemmobacter* was resolved based on 16S rRNA

gene sequences (Fig. 1A). Four members of the genus *Gemmobacter* (*G*. sp. LW-1, *G. caeni*, *G*. sp. Lutimaris YJ-T1-11 and *G. aquatilis*) clustered together with several other related *Gemmobacter* and *Rhodobacter* 16S rRNA gene sequences retrieved from fresh water, soil and sediment and activated sludge environments with *G. megaterium* along with sequences from the marine environment. *G. nectariphilus*, *G. megaterium* and *G*. sp. HYN0069 sequences clustered together with *Paracoccus kawasakiensis* and other related *Gemmobacter* sequences from fresh water and activated sludge environments (Fig. 1A). Phylogenomic analysis based on single copy marker genes specific to members within *Alphaproteobacteria* revealed that *G*. sp LW1, *G. caeni* and *G. lutimaris* clustered together and along with *G. aquatilis* and *G*. sp HYN0069 were closely related to *Rhodobacter sphaeroides* 2.4.1 whereas *G. megabacterium and G. nectariphilus* to *Paracoccus denitrificans* (Fig. 1B & Table S5 for genome taxonomy classification).

The environmental distribution of the genus *Gemmobacter* in sequence datasets was determined using the MAPseq tool (v1.22; accessed via www.beta.microbeatlas.org), a reference-based rRNA gene sequence analysis in both amplicon and shotgun metagenome sequences (Matias Rodrigues *et al*., 2017). MAPseq analysis detected *Gemmobacter*-related sequences in 4810 aquatic, 1540 soil, 2040 plant and 1870 animal/human samples (Fig. S1A). Members of the genus *Gemmobacter* are widely distributed in engineered (such as activated sludge and clinical environments) and natural environments, that is fresh water, soil and sediment and marine environments. In order to determine the relative abundance of *Gemmobacter* in specific environments, sequence datasets from four distinct ecosystems were used: (i) reactor facilities for treating municipal wastewater (2.56%), (ii) epiphytic bacterial communities in *Hydrilla verticillate* (6.6%), (iii) human skin microbiome (interdigital web space; 4.19%) and (iv) dry valley lakes in high altitude (3.71%; Fig. 1B).

GMA synthetase, a key enzyme in the NMG pathway, is encoded by the gene *gmaS*. *gmaS* sequences retrieved from the isolate genomes along with other ratified *gmaS* and glutamine synthetase type III (GlnA; as outgroup) sequences were used to construct an amino acid-based phylogenetic tree (Fig. 2). *gmaS* gene sequences retrieved from genomes of *G*. sp. LW-1, *G. caeni*, *G*. sp. HYN0069, *G. lutimaris* YJ-T1-11 and *G. aquatilis* clustered within Group I of alphaproteobacterial *gmaS* sequences containing sequences from marine and non-marine bacteria within the orders *Rhodobacterales* and *Rhizobiales* (Wischer *et al*., 2015) and were closely related to *Paracoccus yeei*, *P*. sp. 1 W-5 and *Rhodobacter* sp. 1 W-5 (Fig. 2). While *gmaS* gene sequences were detected in five of the seven investigated *Gemmobacter* genomes, *mauA* gene sequences were identified only in the genomes of *G. caeni*, *G. lutimaris* YJ-T1-11 and *G*. sp. LW-1 (Fig. 3) that clustered together in phylogenomic analysis (Fig. 1B). It has been suggested that the NMG pathway for MMA utilization is more universally distributed and more abundant across proteobacterial methylotrophs than the direct MMA oxidation pathway (Nayak and Marx, 2015). However, it should be noted that genes encoding for the enzymes within the NMG pathway (*gmaS*) can not only be detected in methylotrophs but also in non-methylotrophic bacteria that use MMA as a nitrogen source for energy, but not as a carbon source (Chen, 2012; Wischer *et al*., 2015; Lidbury *et al*., 2015b).

A comparative genome analysis of members within the genus Gemmobacter

At the time of the analysis, seven *Gemmobacter* genomes obtained from isolates from different environments were available (Fig. 1B). *Gemmobacter* genome sizes range from ~3.96 Mb to ~5.14 Mb with GC contents between 63.95% and 66.19% and genome completeness between 98.31% and 99.70% (Table S2). Analysis of sequence annotations revealed that on average 98.44% of the genomes consist of coding sequences (CDS).

The genomes were compared using the CGView comparison tool (Grant *et al*., 2012) (Fig. 4). *Gemmobacter* sp. LW-1, isolated from the Movile Cave ecosystem was used as the reference genome and the results of the BLAST comparison with other *Gemmobacter* genomes are represented as a BLAST ring for each genome (Fig. 4). Similarities between segments of the reference genome sequence and the other genome sequences are shown by a coloured arc beneath the region of similarity indicating the percentage of similarity as a colour code. Our analysis (Fig. 4) revealed low amino acid sequence identity levels (mostly <88%) between *Gemmobacter* sp. LW-1 and *G. aquatilis*, *G. nectariphilus*, *G. megaterium* and *G*. sp. HYN0069 across the genomes. Higher identity levels (>90%) were detected between *Gemmobacter* sp. LW-1 and *G. caeni* and *G. lutimaris*. Moreover, the analysis suggested several sites of potential insertion/deletion events in the genome of *Gemmobacter* sp. LW-1. Possible insertion/deletion regions can be identified as those gaps in the map where no homology is detected. For example, the region between 2200 and 2300 kbp (Fig. 4) where a gap can be found in the otherwise contiguous homologous regions between the reference genome *G*. sp. LW-1 and the first of the query genomes (*G. caeni*). This might likely be due to a lack of hits or hits with low identity that can be spurious matches. Since it covers a large region, we could possibly rule out that it is an artefact arising from a lack of sensitivity in the BLAST analysis. Even though the genomes of *G*. sp. LW-1 and *G. caeni* are closely related,

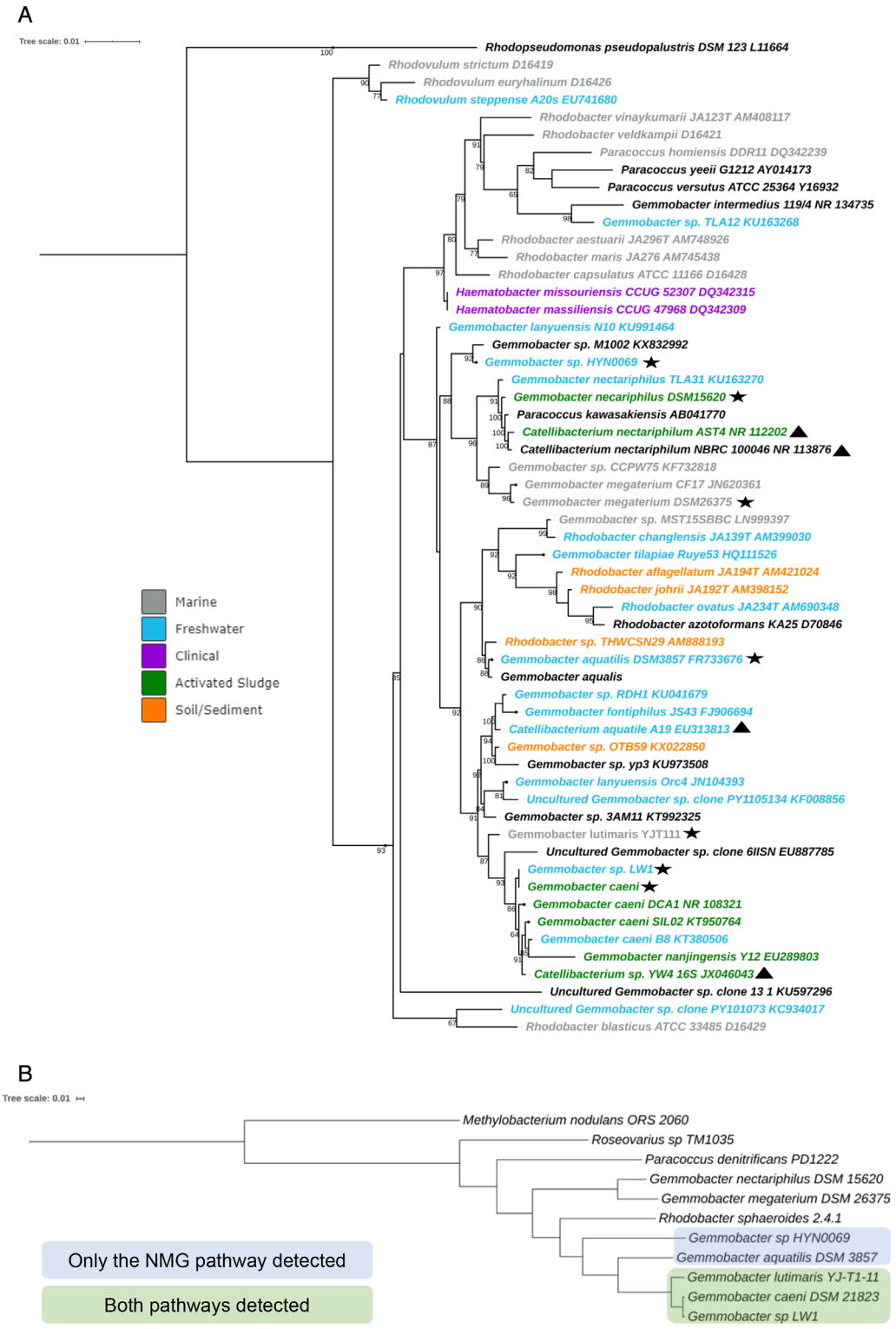


Fig 1. A. Phylogenetic tree based on 16S rRNA gene sequences. The tree was constructed using the maximum likelihood method for clustering and the Tamura-Nei model for computing evolutionary distances. Numbers at branches are bootstrap percentages >50% of 1000 replicates. Star represents the *Gemmobacter* species used for comparative genome analysis. Coloured fonts represent the habitat where the sequence was retrieved: blue (fresh water), orange (soil and sediment), green (activated sludge), grey (marine), purple (clinical source). Triangles represent sequences that are listed as *Catellibacterium* in the NCBI database, which have been reclassified to *Gemmobacter* (Chen *et al*., 2013). Scale bar: 0.01 substitutions per nucleotide position. B. Phylogenomics tree of genomes within the genus *Gemmobacter* and closely related organisms within *Alphaproteobacteria*. Scale bar: 0.01 substitutions per nucleotide position.

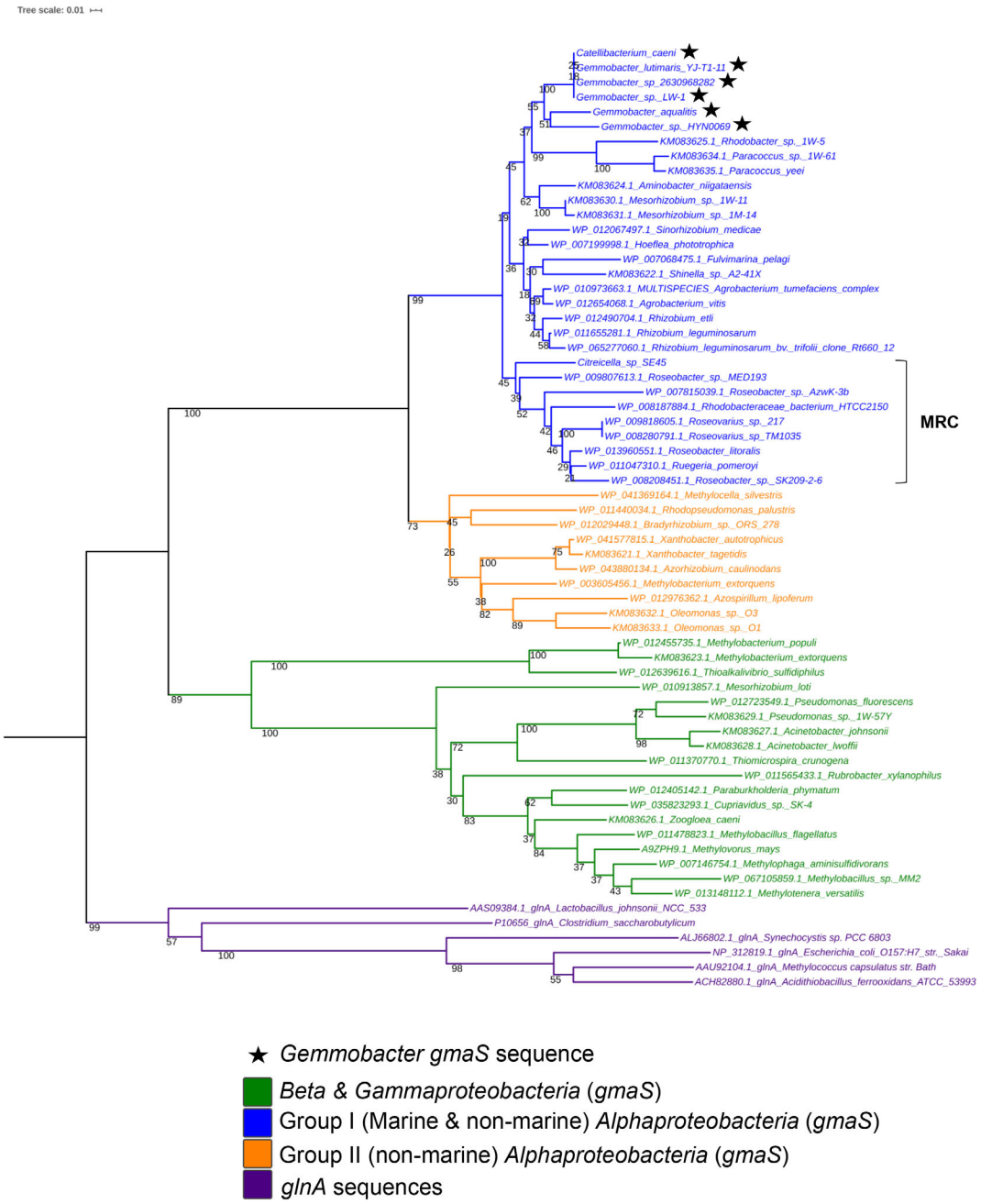


Fig 2. Maximum-likelihood phylogenetic tree based on GmaS sequences. The tree was constructed using amino acid sequences (GmaS) using the maximum-likelihood method based on the JTT matrix-based model. Members of the genus *Gemmobacter* used for genome comparison are represented with a star. Numbers at branches are bootstrap percentages of 1000 replicates. Amino acid sequences of the glutamine synthetase type III (GlnA) were used as out-group. MRC, marine *Roseobacter* clade. Scale bar: 0.1 substitutions per amino acid position.

our analysis demonstrates that their genomes are not completely identical. Despite the fact that the majority of their genomes indicate very high identity levels (mostly >96%–98% as shown by the dominance of dark red colours of the circle representing the BLAST hit identity between *G*. sp. LW-1 and *G. caeni*, many segments appear to be exclusive to *G*. sp. LW-1.

In order to further resolve the similarity between these genomes we calculated the average nucleotide identity (ANI) (Rodriguez-R and Konstantinidis, 2016) (Table S4 and Fig. S2A–F). It is generally accepted that an ANI value of >95%–96% can be used for species delineation (Richter and Rossello-Mora, 2009; Kim *et al*., 2014). Our analysis revealed that *Gemmobacter* sp. LW-1 and

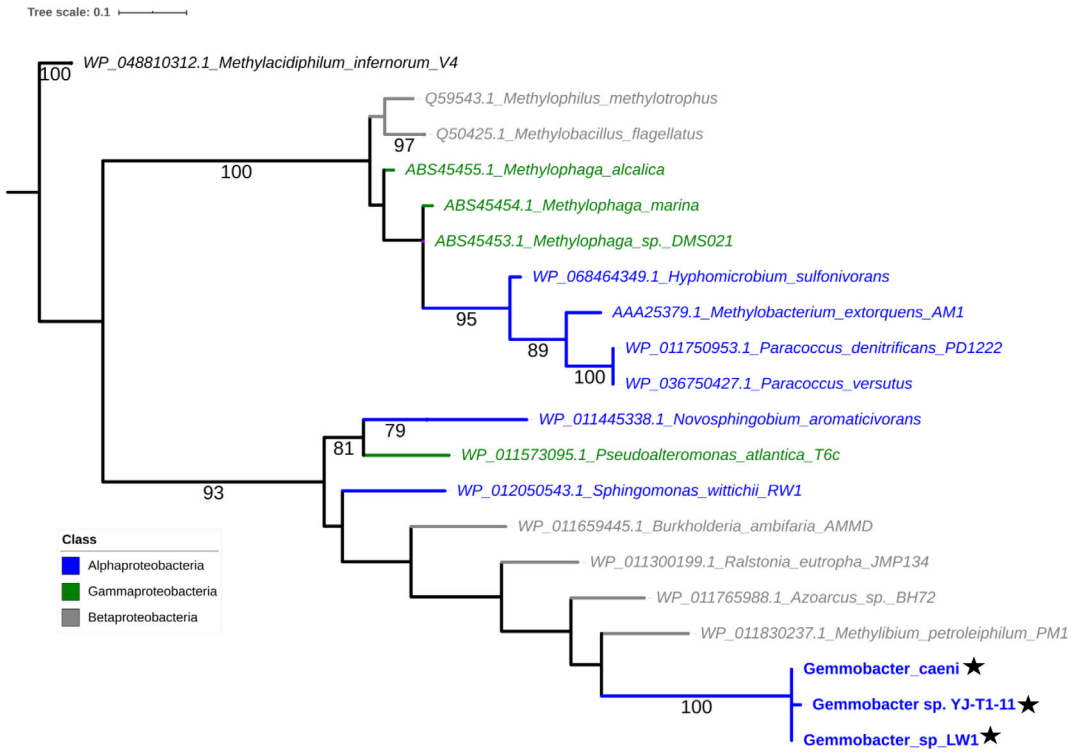


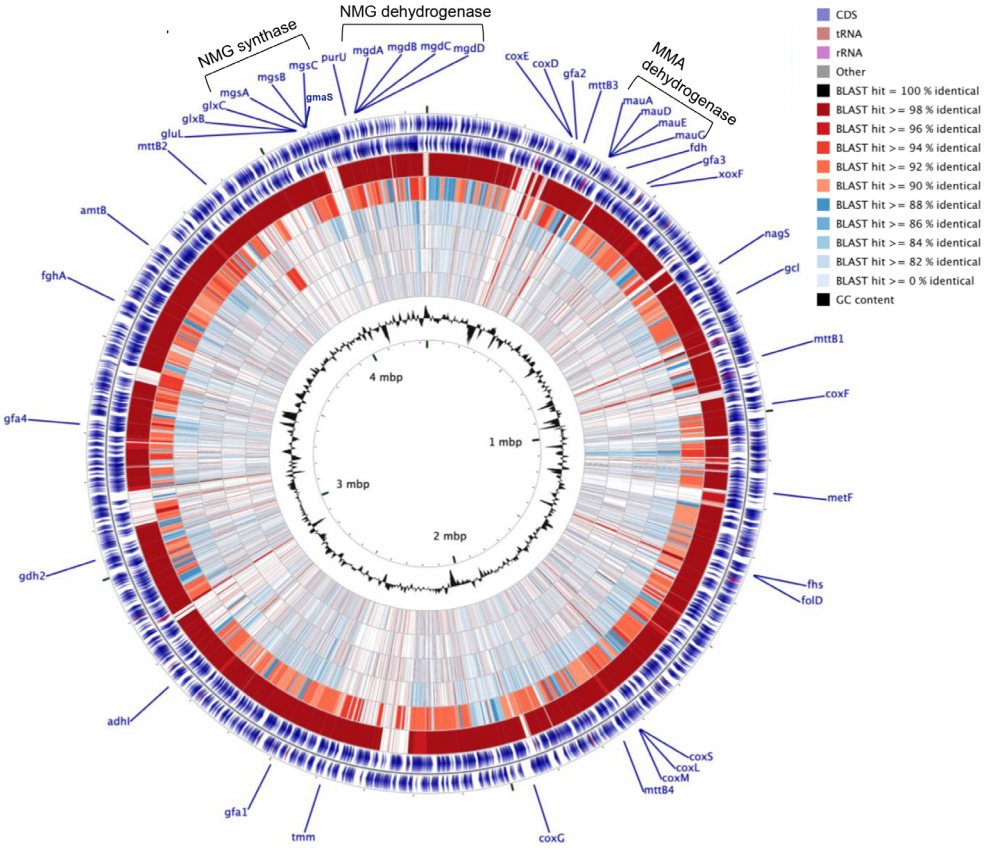
Fig 3. Maximum-likelihood phylogenetic tree (JTT matrix-based) of *mauA* sequences. Sequences from the genus *Gemmobacter* are marked with a star. Amino acid sequences (MauA) were aligned using the ClustalW algorithm. Numbers at branches are bootstrap percentages of 1000 replicates. Scale bar: 0.1 substitutions per amino acid. Coloured boxes indicate *Alphaproteobacteria* (blue), *Gammaproteobacteria* (green) and *Betaproteobacteria* (grey).

Gemmobacter caeni share an ANI value of 98.62 (Table S4) implying that both are in fact the same species. The genome-based average nucleotide identity (gANI) between *G*. sp. LW-1 and *G. caeni* was calculated as 98.70. The AF was calculated to be 0.91, which would result in a computed probability of 0.98 suggesting that both genomes might belong to the same species. However, it should be noted that these are draft genomes and a more in-depth characterization of their physiology and phenotype is required to delineate these organisms at the level of strain.

Pan-genome analysis, carried out using EDGAR v2.0 (Blom *et al*., 2009), identified genes present in all *Gemmobacter* species (core genes), two or more *Gemmobacter* species (accessory or dispensable genes), and unique *Gemmobacter* species (singleton genes). According to pan-genome analysis of the seven *Gemmobacter* genomes, a total of 10 976 genes were identified, of which 50% were singletons (5.492 CDS), 35.7% were dispensable (3921 CDS) and 14.2% were shared by all seven *Gemmobacter* genomes (core genome, 1563 CDS; Fig. 3A). The UpSet plot (Lex *et al*., 2014) shows the number of CDS in the core genome, the singletons but also the number of CDS shared by the different *Gemmobacter* genomes (Fig. 3C). It also confirms the phylogeny of the phylogenetic tree based on the core genome between all seven *Gemmobacter* genomes (Fig. 3B) showing a high similarity between *Gemmobacter megaterium* and *Gemmobacter nectariphilus* (578 uniquely shared CDS) and between *Gemmobacter caeni* and *Gemmobacter* sp. LW1 (360 uniquely shared CDS).

Methylated amine utilization, N assimilation and C_1 oxidation

Investigation of the methylated amine utilization pathways in seven *Gemmobacter* genomes revealed the presence of the genes encoding enzymes TMA dehydrogenase (*tmd*), TMA monooxygenase (*tmm*) and TMAO demethylase (*tdm*) in genomes of *G*. sp. LW-1, *G. caeni*, *G*. sp. HYN0069, *G. lutimaris* and *G. aquatilis* while none of these genes were detected in *G. nectariphilus* or *G. megaterium* (Fig. 5) indicating the metabolic potential of these organisms to use the TMA oxidation pathway to convert TMA to DMA. These findings are supported by results from a previous study which showed growth of *G*. sp. LW-1 on TMA as a carbon and nitrogen source (Wischer *et al*., 2015). Based on the genome sequences, it can be suggested that these five *Gemmobacter* could



Protein	Corresponding gene(s)
TMA monooxygenase	*tmm*
TMA methyltransferase	*mttB1, mttB2, mttB3, mttB4*
MMA dehydrogenase	*mauA, mauD, mauE, mauG*
GMA synthetase	*gmaS*
NMG synthase	*mgsA, mgsB, mgsC*
NMG dehydrogenase	*mgdA, mgdB, mgdC, mgdD*
HCHO-activating enzyme	*gfa1, gfa2, gfa3, gfa4*
GSH-dependent HCHO dehydrogenase	*adhI*
S-formyl-GSH hydrolase	*fghA*
Glutamine synthetase (GS)	*gluL*
Glutamate synthase (GOGAT)	*glxB, glxC*
Methylene-H_4F-dehydro-genase/cyclohydrolase	*folD*
Formate-H_4F ligase	*fhs*
Formyl-H_4F deformylase	*purU*
Formate dehydrogenase	*fdh*
Carbon monoxide dehydrogenase	*coxD, coxE, coxF,coxG, coxL, coxM, coxS*
Methanol dehydrogenase	*xoxF*
N-acetylglutamate synthase	*nagS*
Gamma-glutamylcysteine synthetase	*gcl*
Methylenetetrahydrofolate reductase	*metF*
Glutamate dehydrogenase	*gdh2*
Ammonium transporter	*amtB*

Fig 4. DNA BLAST map of *Gemmobacter* genomes. *Gemmobacter* sp. LW-1 was used as a reference genome against *Gemmobacter megaterium* (inner ring), *Gemmobacter* sp. HYN0069, (second inner ring), *Gemmobacter nectariphilus* (third ring), *Gemmobacter aquatilis* (fourth ring), *Gemmobacter lutimaris* (fifth ring) and *Gemmobacter caeni* (sixth ring). The seventh and eight ring (outer rings) represent the CDS (blue), tRNA (maroon), and rRNA (purple) on the reverse and forward strand, respectively. The colour scale (inset) shows the level of amino acid sequence identity with the respective sequences from *G. megaterium*, *G. aquatilis*, *G. nectariphilus* and *G. caeni*. The locations of genes involved in methylotrophy are indicated at the outside of the map.

use the enzyme DMA monooxygenase (*dmmDABC*) to oxidize DMA to MMA but not the DMA dehydrogenase since the corresponding protein encoding gene (*dmd*) was not found (Fig. 5).

We also compared the distribution of the direct MMA-oxidation and the NMG pathways in the genomes of seven *Gemmobacter* species (Fig. 5) and gene arrangement (Fig. S4). The direct MMA-oxidation pathway (*mauA*-dependent) is so far only known to be present in methylotrophic bacteria that can use MMA as a carbon source. Whereas the NMG pathway (*gmaS*-dependent) has been shown to be present in non-methylotrophic bacteria that can use MMA as a nitrogen source for energy (Chen *et al*., 2010a; Nayak and Marx, 2015; Wischer *et al*., 2015; Nayak *et al*., 2016). Analysis of the genome sequences revealed that *G*. sp. LW-1, *G. lutimaris* and *G. caeni* possess genes for both MMA oxidation pathways (Fig. 5). We have previously shown that *Gemmobacter* sp. LW-1 can use MMA and TMA as both a carbon and nitrogen source (Wischer *et al*., 2015). Genome sequences of *G. aquatilis* and *G*. sp. HYN0069 indicated the presence of genes involved only in the NMG pathway. In the facultative methylotroph *Methylorubrum extorquens* AM1, it has been shown that the NMG pathway is advantageous compared to the direct MMA-oxidation pathway (Nayak *et al*., 2016). NMG pathway enables facultative methylotrophic bacteria to switch between using MMA as a nitrogen source or as a carbon and energy source whereas the direct MMA oxidation pathway allows for rapid growth on MMA only as the primary energy and carbon source (Nayak *et al*., 2016). This could suggest that *G. aquatilis* and *G*. sp. HYN0069 might use the NMG pathway for utilizing MMA as both nitrogen and carbon source. However, growth assays are required to confirm whether both organisms can use MMA as a carbon source. We did not detect genes for either MMA oxidation pathways in the genome sequences of *G. nectariphilus* and *G. megaterium* suggesting the lack of genetic potential of these organisms to use MMA as either C or N source.

The C_1 units derived from methylated amines need to be further oxidized when the nitrogen is sequestered without assimilation of the carbon from the methylated amines. Genome analysis confirmed that all seven *Gemmobacter* species possess the genetic capability for C_1 oxidation and also indicate that tetrahydrofolate (H_4F) is the C_1 carrier (Fig. 5). The bifunctional enzyme 5,10-methylene-tetrahydrofolate dehydrogenase/cyclohydrolase, encoded by the gene *folD*, was detected in all the *Gemmobacter* genomes (Fig. 5, Table 1). Genes encoding key enzymes in the C_1 oxidation pathway via tetrahydromethanopterin (H_4MPT) were not detected (Chistoserdova, 2011). The formate-tetrahydrofolate ligase, encoded by the gene *fhs* (Fig. 5), provides C_1 units for biosynthetic pathways (Lidbury *et al*., 2015a). However, the oxidation of formyl-H_4F (CHO-H_4F) can also be facilitated by *purU*, the gene encoding for the formyl-H_4F deformylase. The formate dehydrogenase (*fdh*) mediates the last step of the C_1 oxidation pathway, the oxidation of formate to CO_2. The genes for the C_1 oxidation pathway via H_4F were detected in all *Gemmobacter* genomes.

The *fae* gene, encoding the formaldehyde-activating enzyme that catalyses the reduction of formaldehyde with H_4MPT, was not detected in any of the seven *Gemmobacter* genomes confirming that these members of the genus *Gemmobacter* lack the H_4MPT pathway for formaldehyde oxidation (Fig. 5 and Table 1). Investigation of the nitrogen assimilation pathway revealed the presence of the genes encoding glutamine synthetase (GS; *glnA*) and glutamate synthase (GOGAT; *gltB*) in all seven *Gemmobacter* genomes. In bacteria this pathway is essential for glutamate synthesis at low ammonium concentrations (Chen, 2012).

Conclusion

Using comparative genome analysis, we provide genome-based evidence that the three *Gemmobacter* isolates *G*. sp. LW-1, *G. lutimaris* and *G. caeni* are capable of generating energy from complete oxidation of methylated amines via the H_4F-dependent pathway using either the NMG pathway or the direct MMA oxidation pathway. Both *Gemmobacter aquatilis* and *G*. sp. HYN0069 are genetically capable of methylated amine degradation to yield formaldehyde and only encode the genes for the NMG pathway, which indicates that these organisms could use this pathway to use MMA as a nitrogen source for energy. Both *G. nectariphilus* and *G. megaterium* genomes indicate the lack of potential to use methylated amines.

Gemmobacter sp. LW-1 was isolated from the Movile Cave ecosystem (Wischer *et al*., 2015). Microbial mats and lake water within the cave have been shown to harbour a wide diversity of methylated amine-utilizing bacteria (Wischer *et al*., 2015; Kumaresan *et al*., 2018). While the mechanism of MAs production within the system has to be elucidated, it can be speculated that degradation of floating microbial mats (i.e. organic matter) could result in MAs (Wischer *et al*., 2015). Similarly, *G. caeni* isolated from activated sludge (Zheng *et al*., 2011) could possibly use the MAs generated from organic matter degradation. Interestingly, while *G. megaterium* was isolated from a marine environment (seaweed) (Liu *et al*., 2014) possibly encountering MAs from the degradation of osmolytes such as glycine betaine (*N*,*N*,*N*-trimethylglycine), we did not detect metabolic genes involved in methylated amine utilization.

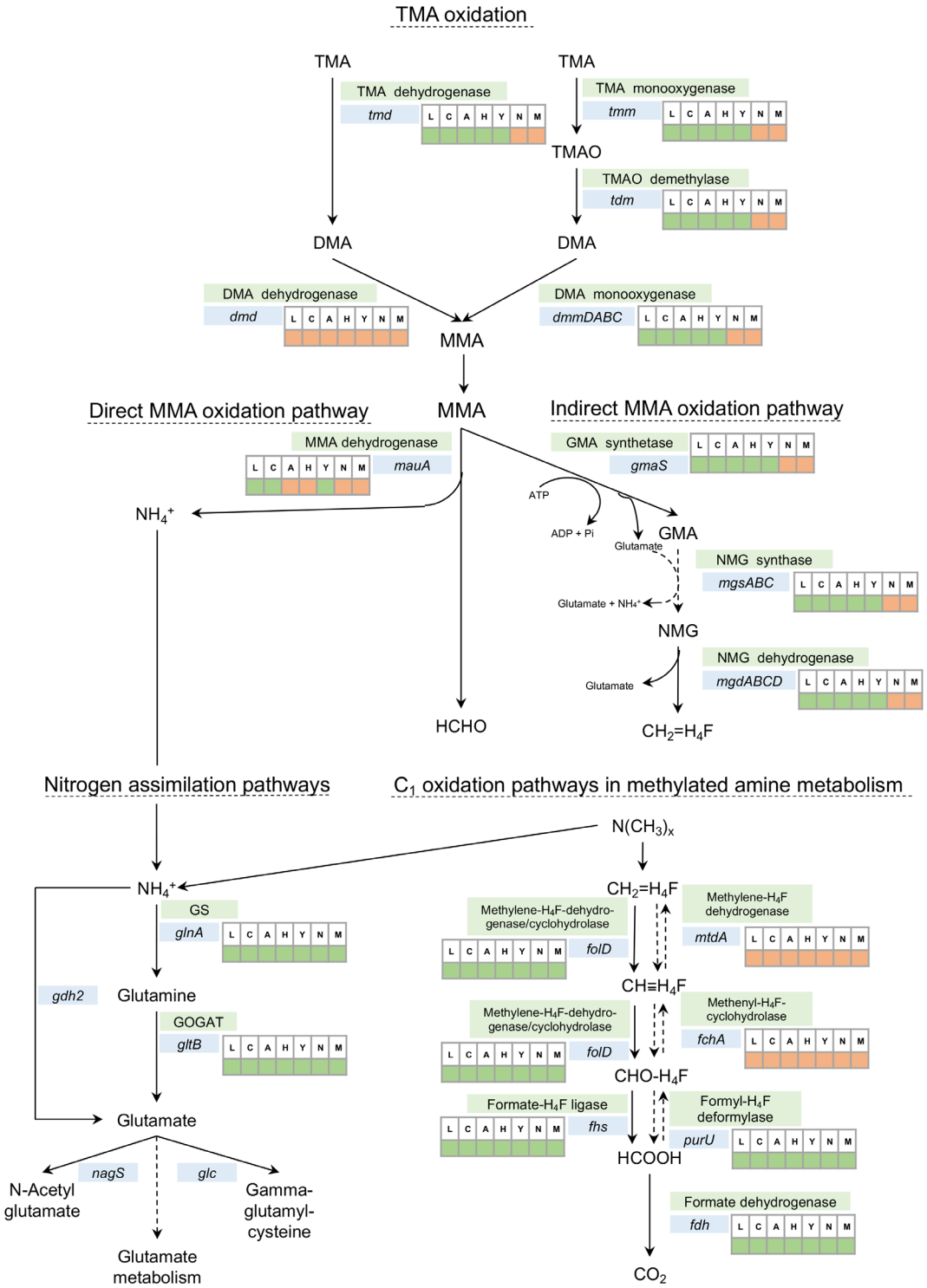


Fig 5. Metabolic pathways involved in methylated amine utilization and one-carbon utilization annotated with presence/absence of specific genes in the genomes of *Gemmobacter*. The analysis was based on a five-way comparison among *Gemmobacter* sp. LW-1 (L), *Gemmobacter caeni* (C), *Gemmobacter aquatilis* (A), *Gemmobacter* sp. HYN0069 (H), *Gemmobacter lutimaris* (Y),*Gemmobacter nectariphilus* (N) and *Gemmobacter megaterium* (M). The colour-coded boxes next to the genes indicate the presence (green) or absence (orange) of a gene in each genome.

Based on the 16S rRNA gene analysis, the genus Gemmobacter appears to be polyphyletic; however, the relatedness of *gmaS* follows established taxonomy. Our study highlights the need for further research into evolutionary implications on methylated amine utilization trait not only in *Gemmobacter* but also across other members

Table 1. Comparative genomic analysis of methylated amine-utilizing genes in genomes-sequenced *Gemmobacters* in comparison to closed related genus within the family *Rhodobacteraceae*.

Organism	*mauA*	*gmaS*	*tmm*	*fae*	*folD*	*purU*	*fhs (fltL)*	*gfa*	*adhI (flhA)*	*fghA*	*fdh*	*amtB*	*mtdA/ mtdB*	*mchA*	*ftr*	*fmdA /B/C*	*cox*
Citreicella sp. *SE45*	–	+	+	–	+	+	+	+	+	+	+	+	–/–	–	–	–	+
Roseovarius sp. *TM1035*	–	+	+	–	+	–	+	–	+	+	+	+	–/–	–	–	–	+
Paracoccus denitrificans PD1222	–	–	–	–	+	+	+	+	+	+	+	+	–/–	–	–	–	–
Rhodobacter sphaeroides 241	–	–	+	–	+	+	+	+	+	+	+	+	–/–	–	–	–	+
Gemmobacter caeni	+	+	+	–	+	+	+	+[a]	+	+	+	+	–/–	–	–	–	+
Gemmobacter sp. *LW-1*	+	+	+	–	+	+	+	+[a]	+	+	+	+	–/–	–	–	–	+
Gemmobacter aquatilis	–	+	+	–	+	+	+	–	+	+	+	+	–/–	–	–	–	+
Gemmobacter nectariphilus DSM-15620	–	–	–	–	+	+	+	+[a]	+	+	+	+	–/–	–	–	–	+
Gemmobacter megaterium DSM-26375	–	–	–	–	+	+	+	+[a]	+	+	+	+	–/–	–	–	–	+
Gemmobacter lutimaris YJ-T1-11	+	+	+	–	+	+	+	+	+	+	+	+	–/–	–	–	+	+
Gemmobacter sp. *HYN0069*	–	+	+	–	+	+	+	–	+	+	+	+	–/–	–	–	+	+

Note: "a" denotes gfa-like protein-encoding gene

adhI (flhA), glutathione-dependent formaldehyde dehydrogenase; *amtB*, ammonium transporter; *cox*, carbon monoxide dehydrogenase; *fae*, formaldehyde-activating enzyme; *fdh*, formate dehydrogenase; *fghA*, S-formyl-glutathione hydrolase; *fhs (fltL)*, formyl-H4F synthetase; *fmdA/B/C*, formylmethanofuran dehydrogenase; *folD*, methylene-H4F dehydrogenase/cyclohydroloase; *ftr*, formylmethanofuran-H_4MPT-N-formyltransferase; *gfa*, GSH-dependent formaldehyde-activating enzyme; *gmaS*, GMA synthetase; mauA, methylamine dehydrogenase small subunit; *mchA*, methenyl-H_4MPT cyclohydrolase; *mtdA/mtdB*, methylene-H_4F dehydrogenase/methylene-H_4MTP dehydrogenase; *purU*, formyl-H4F deformylase; *tmm*, TMA monooxygenase.

Shown is the presence (+) or absence (–) of specific genes in the genome sequences.

within the bacterial domain. Furthermore, results from this study suggest that the trait for methylated amine utilization within the genus *Gemmobacter* could be independent from the habitat and localized factors (e.g. substrate availability) or selection pressures could influence the ability of these organisms to use methylated amines. It has been well-established that the direct MMA oxidation pathway allows organisms to achieve faster growth on MMA and genes encoding for MaDH enzyme (*mau* cluster) can be acquired by horizontal gene transfer (Nayak *et al*., 2015). We also show that members of the genus *Gemmobacter* are widespread in the environment and their contribution to carbon and nitrogen cycling via methylotrophic modules require detailed characterization. Access to *Gemmobacter* isolates with or without the genetic potential for methylated amine utilization trait will allow us to perform physiological experiments in future to test how this trait can affect fitness while growing on methylated amines and also understand the eco-evolutionary factors that shape the physiology of *Gemmobacter* from different environments.

Acknowledgements

The authors thank Dr Felipe Hernandes Coutinho (Federal University of Rio de Janeiro) for a fruitful discussion on genomic regions of insertion and deletion in *Gemmobacter* sp. LW-1. The authors thank Paul Stothard for using their in-house pipeline of the CGView comparison tool. The EDGAR platform is financially supported by the BMBF grant FKZ 031A533 within the de.NBI network. This manuscript has been published as preprint at the preprint server for biology (bioRxiv.org, Kröber *et al*., 2018). The authors also thank the anonymous reviewers for their constructive comments.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403–410.

Anthony, C. (1982) *The Biochemistry of Methylotrophs*. London: Academic Press.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., *et al*. (2008) The RAST server: rapid annotations using subsystems technology. *BMC Genomics* **9**: 75.

Baek, M.-G., Shin, S.-K., and Yi, H. (2020) Gemmobacter aquarius sp. nov., Runella rosea sp. nov. and Flavobacterium fluviale sp. nov., isolated from the Namhangang River system. *Int J Syst Evol Microbiol*: ijsem004455 **70**(11), 5640–5647.

Blom, J., Albaum, S.P., Doppmeier, D., Pühler, A., Vorhölter, F.-J., Zakrzewski, M., and Goesmann, A. (2009) EDGAR: a software framework for the comparative analysis of prokaryotic genomes. *BMC Bioinform* **10**: 154.

Burke, S.A., Lo, S.L., and Krzycki, J.A. (1998) Clustered genes encoding the methyltransferases of methanogenesis from monomethylamine. *J Bacteriol* **180**: 3432–3440.

Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973.

Carpenter, L.J., Archer, S.D., and Beale, R. (2012) Ocean-atmosphere trace gas exchange. *Chem Soc Rev* **41**: 6473–6506.

Chaumeil, P.A., Mussig, A.J., Hugenholtz, P., and Parks, D. H. (2019) GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* **36**(6), 1925–1927.

Chen, W.M., Cho, N.T., Huang, W.C., Young, C.C., and Sheu, S.Y. (2013) Description of Gemmobacter fontiphilus sp nov., isolated from a freshwater spring, reclassification of Catellibacterium nectariphilum as Gemmobacter nectariphilus comb. nov., Catellibacterium changlense as Gemmobacter changlensis comb. nov., Catellibacterium aquatile as Gemmobacter aquaticus nom. nov., Catellibacterium caeni as Gemmobacter caeni comb. nov., Catellibacterium nanjingense as Gemmobacter nanjingensis comb. nov., and emended description of the genus Gemmobacter and of Gemmobacter aquatilis. *Int J Syst Evol Microbiol* **63**: 470–478.

Chen, Y. (2012) Comparative genomics of methylated amine utilization by marine *Roseobacter* clade bacteria and development of functional gene markers (tmm, gmaS). *Environ Microbiol* **14**: 2308–2322.

Chen, Y., McAleer, K.L., and Murrell, J.C. (2010a) Monomethylamine as a nitrogen source for a nonmethylotrophic bacterium, Agrobacterium tumefaciens. *Appl Environ Microbiol* **76**: 4102–4104.

Chen, Y., Scanlan, J., Song, L., Crombie, A., Rahman, M.T., Schäfer, H., and Murrell, J.C. (2010b) γ-Glutamylmethylamide is an essential intermediate in the metabolism of methylamine by Methylocella silvestris. *Appl Environ Microbiol* **76**: 4530–4537.

Chen, Y., Wu, L., Boden, R., Hillebrand, A., Kumaresan, D., Moussard, H., *et al*. (2009) Life without light: microbial diversity and evidence of sulfur- and ammonium-based chemolithotrophy in Movile cave. *ISME J* **3**: 1093–1104.

Chistoserdova, A.Y., Christoserdova, L.V., McIntire, W.S., and Lidstrom, M.E. (1994) Genetic organization of the mau gene cluster in Methylobacterium extorquens AM1: complete nucleotide sequence and generation and characteristics of mau mutants. *J Bacteriol* **176**: 4052–4065.

Chistoserdova, L. (2011) Modularity of methylotrophy, revisited. *Environ Microbiol* **13**: 2603–2622.

Chistoserdova, L., Kalyuzhnaya, M.G., and Lidstrom, M.E. (2009) The expanding world of Methylotrophic metabolism. *Annu Rev Microbiol* **63**: 477–499.

Eddy, S.R. (2011) Accelerated profile HMM searches. *PLoS Comput Biol* **7**: e1002195.

Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform* **5**: 113.

Emms, D.M., and Kelly, S. (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* **16**: 157.

Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**: 783–791.

Ge, X., Wexler, A.S., and Clegg, S.L. (2011) Atmospheric amines – part I. a review. *Atmos Environ* **45**: 524–546.

Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., and Tiedje, J.M. (2007) DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* **57**: 81–91.

Grant, J.R., Arantes, A.S., and Stothard, P. (2012) Comparing thousands of circular genomes using the CGView comparison tool. *BMC Genomics* **13**: 202.

Hameed, A., Shahina, M., Lin, S.-Y., Chen, W.-M., Hsu, Y.-H., Lai, W.-A., and Young, C.-C. (2020) Description of Gemmobacter aestuarii sp. nov., isolated from estuarine surface water and reclassification of Cereibacter changlensis as Gemmobacter changlensis. *Arch Microbiol*, **202**(5), 1035–1042.

Jones, D.T., Taylor, W.R., and Thornton, J.M. (1992) The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* **8**: 275–282.

Kämpfer, P., Jerzak, L., Wilharm, G., Golke, J., Busse, H.-J., and Glaeser, S.P. (2015) Gemmobacter intermedius sp. nov., isolated from a white stork (*Ciconia ciconia*). *Int J Syst Evol Microbiol* **65**: 778–783.

Kang, J.Y., Kim, M.-J., Chun, J., Son, K.P., and Jahng, K.Y. (2017) Gemmobacter straminiformis sp. nov., isolated from an artificial fountain. *Int J Syst Evol Microbiol* **67**: 5019–5025.

Kim, M., Oh, H.-S., Park, S.-C., and Chun, J. (2014) Towards a taxonomic coherence between average nucleotide identity and 16S gene sequence similarity for species demarcation of prokaryotes. *Int J Syst Evol Microbiol* **64**: 346–351.

Konstantinidis, K.T., and Tiedje, J.M. (2005) Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci U S A* **102**: 2567–2572.

Kumar, S., Stecher, G., and Tamura, K. (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* **33**: 1870–1874.

Kumaresan, D., Stephenson, J., Doxey, A.C., Bandukwala, H., Brooks, E., Hillebrand-Voiculescu, A., *et al*. (2018) Aerobic proteobacterial methylotrophs in Movile Cave: genomic and metagenomic analyses. *Microbiome* **6**: 1.

Kumaresan, D., Wischer, D., Hillebrand-Voiculescu, A.M., and Murrell, J.C. (2015) Draft genome sequences of facultative Methylotrophs, Gemmobacter sp. strain LW1 and Mesorhizobium sp. strain 1M-11, isolated from Movile cave, Romania. *Genome Announc* **3**(6), e01266–15.

Kumaresan, D., Wischer, D., Stephenson, J., Hillebrand-Voiculescu, A., and Murrell, J.C. (2014) Microbiology of Movile cave - a chemolithoautotrophic ecosystem. *Geomicrobiol J* **31**: 186–193.

Lagesen, K., Hallin, P., Rødland, E.A., Stærfeldt, H.-H., Rognes, T., and Ussery, D.W. (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* **35**: 3100–3108.

Latypova, E., Yang, S., Wang, Y.S., Wang, T., Chavkin, T. A., Hackett, M., *et al*. (2010) Genetics of the glutamate-mediated methylamine utilization pathway in the facultative methylotrophic beta-proteobacterium Methyloversatilis universalis FAM5. *Mol Microbiol* **75**: 426–439.

Lee, M.D. (2019) GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics* **35**: 4162–4164.

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., and Pfister, H. (2014) UpSet: visualization of intersecting sets. *IEEE Trans Vis Comput Graph* **20**: 1983–1992.

Lidbury, I., Kimberley, G., Scanlan, D.J., Murrell, J.C., and Chen, Y. (2015a) Comparative genomics and mutagenesis analyses of choline metabolism in the marine Roseobacter clade. *Environ Microbiol* **17**: 5048–5062.

Lidbury, I., Mausz, M.A., Scanlan, D.J., and Chen, Y. (2017) Identification of dimethylamine monooxygenase in marine bacteria reveals a metabolic bottleneck in the methylated amine degradation pathway. *ISME J* **11**: 1592–1601.

Lidbury, I., Murrell, J.C., and Chen, Y. (2014) Trimethylamine N-oxide metabolism by abundant marine heterotrophic bacteria. *Proc Natl Acad Sci* **111**: 2710–2715.

Lidbury, I.D., Murrell, J.C., and Chen, Y. (2015b) Trimethylamine and trimethylamine N-oxide are supplementary energy sources for a marine heterotrophic bacterium: implications for marine carbon and nitrogen cycling. *ISME J* **9**: 760–769.

Lidstrom, M.E. (2006) Aerobic Methylotrophic prokaryotes. In *The Prokaryotes: Volume 2: Ecophysiology and Biochemistry*. Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., and Stackebrandt, E. (eds). New York, NY: Springer, pp. 618–634.

Lim, K., Kannan, A.D., Shobnam, N., Mahmood, M., Lee, J., and Im, J. (2020) Gemmobacter serpentinus sp. nov., isolated from conserved forages. *Int J Syst Evol Microbiol* **70**: 4224–4232.

Liu, J.-J., Zhang, X.-Q., Chi, F.-T., Pan, J., Sun, C., and Wu, M. (2014) Gemmobacter megaterium sp. nov., isolated from coastal planktonic seaweeds. *Int J Syst Evol Microbiol* **64**: 66–71.

Liu, Y., and Whitman, W.B. (2008) Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea. *Ann N Y Acad Sci* **1125**: 171–189.

Liu, Y., Xu, C.-J., Jiang, J.-T., Liu, Y.-H., Song, X.-F., Li, H., and Liu, Z.-P. (2010) Catellibacterium aquatile sp. nov., isolated from fresh water, and emended description of the genus Catellibacterium Tanaka et al. 2004. *Int J Syst Evol Microbiol* **60**: 2027–2031.

Lyimo, T.J., Pol, A., Jetten, M.S.M., and Op den Camp, H.J. M. (2009) Diversity of methanogenic archaea in a mangrove sediment and isolation of a new Methanococcoides strain. *FEMS Microbiol Lett* **291**: 247–253.

Markowitz, V.M., Chen, I.-M.A., Palaniappan, K., Chu, K., Szeto, E., Pillay, M., *et al*. (2013) IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res* **42**: D560–D567.

Matias Rodrigues, J.F., Schmidt, T.S.B., Tackmann, J., and von Mering, C. (2017) MAPseq: highly efficient k-mer search with confidence estimates, for rRNA sequence analysis. *Bioinformatics* **33**: 3808–3810.

McIntire, W.S., Wemmer, D.E., Chistoserdova, A., and Lidstrom, M.E. (1991) A new cofactor in a prokaryotic enzyme: tryptophan tryptophylquinone as the redox prosthetic group in methylamine dehydrogenase. *Science* **252**: 817–824.

Nayak, D.D., Agashe, D., Lee, M.-C., and Marx, C.J. (2016) Selection maintains apparentlydegenerate metabolic pathwaysdueto tradeoffsin using methylamine for carbon versus nitrogen. *Curr Biol* **26**: 1–11.

Nayak, D.D., and Marx, C.J. (2015) Experimental horizontal gene transfer of methylamine dehydrogenase mimics prevalent exchange in nature and overcomes the methylamine growth constraints posed by the sub-optimal N-Methylglutamate pathway. *Microorganisms* **3**: 60–79.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res*: gr. 186072.186114. **25**(7), 1043–1055.

Paul, L., Ferguson, D.J., and Krzycki, J.A. (2000) The trimethylamine methyltransferase gene and multiple dimethylamine methyltransferase genes of Methanosarcina barkeri contain in-frame and read-through amber codons. *J Bacteriol* **182**: 2520–2529.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010) FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* **5**: e9490.

Pruesse, E., Peplies, J., and Glöckner, F.O. (2012) SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**: 1823–1829.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glöckner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35**: 7188–7196.

Qu, J.-H., Ma, W.-W., Zhou, J., Wang, X.-F., Lu, W.-L., Qu, L.-B., and Wang, L.-F. (2020) Gemmobacter caeruleus sp. nov., a novel species originating from lake sediment. *Int J Syst Evol Microbiol* **70**: 1987–1992.

Richter, M., and Rossello-Mora, R. (2009) Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A* **106**: 19126–19131.

Rodriguez-R, L.M., and Konstantinidis, K.T. (2016) The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Preprints.

Rothe, B., Fischer, A., Hirsch, P., Sittig, M., and Stackebrandt, E. (1987) The phylogenetic position of the budding bacteria Blastobacter aggregatus and Gemmobacter aquatilis gen., nov. sp. nov. *Arch Microbiol* **147**: 92–99.

Saitou, N., and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406–425.

Schade, G.W., and Crutzen, P.J. (1995) Emission of aliphatic-amines from animal husbandry and their reactions – potential source of N2o and Hcn. *J Atmos Chem* **22**: 319–346.

Sheu, S.-Y., Sheu, D.-S., Sheu, F.-S., and Chen, W.-M. (2013a) Gemmobacter tilapiae sp. nov., a poly-β-hydroxybutyrate-accumulating bacterium isolated from a freshwater pond. *Int J Syst Evol Microbiol* **63**: 1550–1556.

Sheu, S.-Y., Shiau, Y.-W., Wei, Y.-T., and Chen, W.-M. (2013b) Gemmobacter lanyuensis sp. nov., isolated from a freshwater spring. *Int J Syst Evol Microbiol* **63**: 4039–4045.

Tanaka, Y., Hanada, S., Manome, A., Tsuchida, T., Kurane, R., Nakamura, K., and Kamagata, Y. (2004) Catellibacterium nectariphilum gen. nov., sp. nov., which

requires a diffusible compound from a strain related to the genus Sphingomonas for vigorous growth. *Int J Syst Evol Microbiol* **54**: 955–959.

Taubert, M., Grob, C., Howat, A.M., Burns, O.J., Pratscher, J., Jehmlich, N., *et al*. (2017) Methylamine as a nitrogen source for microorganisms from a coastal marine environment. *Environ Microbiol* **19**: 2246–2257.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680.

Wischer, D., Kumaresan, D., Johnston, A., El Khawand, M., Stephenson, J., Hillebrand-Voiculescu, A.M., *et al*. (2015) Bacterial metabolism of methylated amines and identification of novel methylotrophs in Movile Cave. *ISME J* **9**: 195–206.

Yoo, Y., Lee, D.W., Lee, H., Kwon, B.-O., Khim, J.S., Yim, U.H., *et al*. (2019) Gemmobacter lutimaris sp. nov., a marine bacterium isolated from a tidal flat. *Int J Syst Evol Microbiol* **69**: 1676–1681.

Zhang, J., Chen, S.-A., Zheng, J.-W., Cai, S., Hang, B.-J., He, J., and Li, S.-P. (2012) Catellibacterium nanjingense sp. nov., a propanil-degrading bacterium isolated from activated sludge, and emended description of the genus Catellibacterium. *Int J Syst Evol Microbiol* **62**: 495–499.

Zheng, J.-W., Chen, Y.-G., Zhang, J., Ni, Y.-Y., Li, W.-J., He, J., and Li, S.-P. (2011) Description of Catellibacterium caeni sp. now., reclassification of Rhodobacter changlensis Anil Kumar et al. 2007 as Catellibacterium changlense comb. nov. and emended description of the genus Catellibacterium. *Int J Syst Evol Microbiol* **61**: 1921–1926.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Figure S1 (A) Distribution of *Gemmobacter* in different ecosystems (output from MAPseq tool) with average abundance in the environmental samples and number of samples available is indicated within parenthesis. (B) Relative abundance of *Gemmobacter* in four distinct ecosystems (top 20 genus).

Fig. S2. (A–F) Average nucleotide identity (ANI) analysis of *Gemmobacter* sp. LW-1 and *Gemmobacter caeni*, *Gemmobacter aquatilis*, *Gemmobacter nectariphilus* and *Gemmobacter megaterium* and (G) AAI analysis between those species and related *Rhodobacter sphaeroides*.

Fig. S3. Pan-genome analysis of seven *Gemmobacter* genomes. (A) Fractional pan genome representation. (B) Phylogenetic tree of seven *Gemmobacter* species based on the core genome calculated between those seven species. (C) UpSet plot (Lex *et al*., 2014) of the pan-genome of the seven *Gemmobacter* genomes. *Gemmobacter* sp. LW1 was chosen as reference genome. (D) Circular representation of the *Gemmobacter* genomes indicating the core genome shared between all seven *Gemmobacter* genomes.

Fig. S4. Arrangement of genes involved in methylated amine utilization.

Table S1. Overview of isolated Gemmobacter species.

Table S2. Genome characteristics of the seven *Gemmobacter* isolate genomes used in this study.

Table S3. List of protein queries used for the genome comparison with their accession number.

Table S4. (A) Average nucleotide identity (ANI) and (B) average amino acid identity (AAI) values between *Gemmobacter* sp. LW-1 and *Gemmobacter caeni*, *Gemmobacter aquatilis*, *Gemmobacter nectariphilus*, *Gemmobacter megaterium*, *Rhodobacter sphaeroides* and *Paracoccus denitrificans*.

Table S5. Genome taxonomy classification of *Gemmobacter* genomes using the toolkit GTDB-TK.