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*

1Department of Symbiosis, Max-Planck Institute for Marine Microbiology, Bremen, Germany.
2School of Biological Sciences, Institute for Global Food Security, Queen’s University Belfast, Belfast, UK.
3Department of Cellular Biology, University of Brasilia, Brasilia, Brazil.
4School 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 an 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.

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 (C1) 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 \textit{tdm}, \textit{tmm} and \textit{mtt}, respectively (Paul et al., 2000; Chen, 2012; Lidbury et al., 2014). The enzymes DMA dehydrogenase (\textit{dmd}) or DMA monooxygenase (\textit{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 (DMA 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-methylene tetrahydrofolate (CH$_2$ = H,F) (Latypova et al., 2010; Chistoserdova, 2011). A stepwise conversion of MMA in the NMG pathway is modulated by the enzymes GMA synthetase (\textit{gmaS}), ‘NMG synthase’ (\textit{mgsABC}) and NMG dehydrogenase (\textit{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 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, \textit{Gemmobacter} sp. LW-1 (recently renamed from \textit{Catellibacterium} (Chen et al., 2013)) from the Movie 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 \textit{Gemmobacter} sp. LW-1 was indeed an active MMA user 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 \textit{Gemmobacter}. However, growth on C$_1$ compounds (methanol and formate) has been reported for the genus \textit{Gemmobacter}, for example in \textit{G. caeni} (Zheng et al., 2011). The genus \textit{Gemmobacter} (family \textit{Rhodobacteraceae}) currently comprises 17 species (summarized in Table S1): \textit{Gemmobacter megaterium} (Liu et al., 2014), \textit{G. nectariphilum} (Tanaka et al., 2004; Chen et al., 2013), \textit{G. aquatilis} (Rothe et al., 1987), \textit{G. caeni} (Zheng et al., 2011; Chen et al., 2013), \textit{G. aquaticus} (Liu et al., 2010; Chen et al., 2013), \textit{G. nanjingense} (Zhang et al., 2012; Chen et al., 2013), \textit{G. intermedius} (Kämpfer et al., 2015), \textit{G. lanyuensis} (Sheu et al., 2013a), \textit{G. tilapia} (Sheu et al., 2013b), \textit{G. fontiphilus} (Chen et al., 2013), \textit{G. straminiformis} (Kang et al., 2017), \textit{G. serpentinus} (Lim et al., 2020), \textit{G. aquarius} (Baek et al., 2020), \textit{G. caeruleus} (Qu et al., 2020), \textit{G. lutiurus} (Yoo et al., 2019), \textit{G. aestuarii} (Hameed et al., 2020) and \textit{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 \textit{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 \textit{Gemmobacter}. We used seven isolate genomes (available in public repositories at the time of this study – March 2020) for members within the genus \textit{Gemmobacter} (G. sp. LW-1, \textit{G. caeni}, \textit{G. aquatilis}, \textit{G. nectariphilum}, \textit{G. megaterium}, \textit{G. lutiurus} YJ-T1-11) alongside genomes of closely related organisms within the family \textit{Rhodobacteraceae} to show that the methylated amine utilization trait is a distinctive feature within selected members of the genus \textit{Gemmobacter}.

Materials and methods

Genome data acquisition

Seven \textit{Gemmobacter} genomes (\textit{G. caeni}, \textit{G. aquatilis}, \textit{G. nectariphilum}, \textit{G. megaterium}, \textit{G. lutiurus} sp. LW-1, \textit{G. sp. HYNO069} and \textit{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 \textit{Gemmobacter} was determined using phylogenetic trees constructed from 16S rRNA gene (nucleotide) and metabolic gene sequences (\textit{gmaS} and \textit{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 phylogenetics 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-Ti-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 computed as follows,

\[ p_r^{\text{intra-species}} \cdot |A| = p_r^{\text{intra-species}} \cdot |A| = \frac{AF = a}{AF = a} \]

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 from four distinct ecosystems: (i) Reactor facilities for treating municipal wastewater (SRR870266), (ii) epiphytic bacterial communities in *Hydrilla verticillata* (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/htm/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,

\[ p_r^{\text{intra-species}} \cdot |A| = p_r^{\text{intra-species}} \cdot |A| = \frac{AF = a}{AF = a} \]
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. HYNO069 sequences clustered together with *Paracoccus kawasakiiensis* 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 HYNO069 were closely related to Rhodobacter *sphaeroides* 2.4.1 whereas *G. megabacterium* and *G. nectariphilus* to *Paracoccus denitrificans* (Fig. 1B). 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 *verticillata* (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. HYNO069, 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. HYNO069 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.

© 2021 The Authors. Environmental Microbiology Reports published by Society for Applied Microbiology and John Wiley & Sons Ltd., Environmental Microbiology Reports
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) (Rodríguez-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 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.

© 2021 The Authors. *Environmental Microbiology Reports* published by Society for Applied Microbiology and John Wiley & Sons Ltd., *Environmental Microbiology Reports*
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 megariterium and Gemmobacter nectariphilus (578 uniquely shared CDS) and between Gemmobacter caeni and Gemmobacter sp. LW1 (360 uniquely shared CDS).

Methylated amine utilization trait in Gemmobacter

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).
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.

© 2021 The Authors. *Environmental Microbiology Reports* published by Society for Applied Microbiology and John Wiley & Sons Ltd., *Environmental Microbiology Reports*
use the enzyme DMA monoxygenase (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 *Methylophilus 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 C1 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 C1 oxidation and also indicate that tetrahydrofolate (H4F) is the C1 carrier (Fig. 5). The bifunctional enzyme 5,10-methylene-tetrahydrofolate dehydrogenase/cyclohydrolase, encoded by the gene *foiD*, was detected in all the *Gemmobacter* genomes (Fig. 5, Table 1). Genes encoding key enzymes in the C1 oxidation pathway via tetrahydrod看着polythioprotein (H4MPT) were not detected (Chistoserdova, 2011). The formate-tetrahydrofolate ligase, encoded by the gene *fhs* (Fig. 5), provides C1 units for biosynthetic pathways (Lidbury et al., 2015a). However, the oxidation of formyl-H4F (CHO-H4F) can also be facilitated by *purU*, the gene encoding for the formyl-H4F dehydrogenase. The formate dehydrogenase (*tdh*) mediates the last step of the C1 oxidation pathway, the oxidation of formate to CO2. The genes for the C1 oxidation pathway via H4F were detected in all *Gemmobacter* genomes.

The *fae* gene, encoding the formaldehyde-activating enzyme that catalyses the reduction of formaldehyde with H4F, was not detected in any of the seven *Gemmobacter* genomes confirming that these members of the genus *Gemmobacter* lack the H4MPT 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 H4F-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.

© 2021 The Authors. *Environmental Microbiology Reports* published by Society for Applied Microbiology and John Wiley & Sons Ltd., *Environmental Microbiology Reports*
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 in comparison to closely related genera within the family Gemmobiacters Rhodobacteraceae.

<table>
<thead>
<tr>
<th>Organism</th>
<th>mtaA</th>
<th>mtdA</th>
<th>mtdB</th>
<th>fhs (ftfL)</th>
<th>adhI (flhA)</th>
<th>gfa</th>
<th>gmaS</th>
<th>mauA</th>
<th>cox</th>
<th>tmm</th>
<th>purU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citrobacter sp. LW-1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Rhodobacter sp. LW-1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Gemmobiacter sp. LW-1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Gemmobiacter sp. LW-2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Gemmobiacter aquatilis DSM-15647</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Gemmobiacter aequarius sp. nov.</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Gemmobiacter nectariphilus DSM-26375</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Gemmobiacter lutimaris YJ-T1-1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Gemmobiacter HYN0069</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Note: + denotes the presence of a gene encoding for a specific protein or enzyme. - denotes the absence of a gene encoding for a specific protein or enzyme. a denotes gfa-like protein-encoding gene. adhI (flhA), glutathione-dependent formaldehyde dehydrogenase; gfa, GSH-dependent formaldehyde-activating enzyme; gma, GMA synthetase; mauA, methylamine dehydrogenase small subunit; tmm, TMA monooxygenase; purU, TMA monooxygenase.

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 Gemmobiacter 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


Methylated amine utilization trait in Gemmobacter 13


**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.