Title:

Comparative genomics analyses indicate differential methylated amine utilisation trait within the member of the genus *Gemmobacter*

Running title:

Methylated amine utilisation trait in *Gemmobacter*

Eileen Kröber [1], Lewis Spurgin[2], Daniela Wischer[3] and Deepak Kumaresan [3#]

[1] Microbial Biogeochemistry, RA Landscape Functioning, ZALF Leibniz Centre for Landscape Research, Müncheberg, Germany

[2] School of Biological Sciences, University of East Anglia, Norwich, UK

[3] School of Biological Sciences & Institute for Global Food Security, Queen's University Belfast, Belfast, Northern Ireland, UK

[#]Correspondence to

Dr D Kumaresan, School of Biological Sciences, Queen's University Belfast, Belfast, Northern Ireland, United Kingdom

Email: d.kumaresan@qub.ac.uk

Abstract

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 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 utilisation as a carbon source. Here, using a comparative genomics approach, we investigate how widespread methylated amine utilisation trait is within the members of the bacterial genus *Gemmobacter*. Five 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 public repositories and used for the analysis. Our results indicate that some members of the genus *Gemmobacter*, namely *G. aquatilis*, *G. caeni* and *G.* sp. LW-1 have the genetic potential of methylated amine utilisation while others (*G. megaterium* and *G. nectariphilus*) have not. Ancestral state reconstruction analysis also suggested that methylated amine utilisation trait might not be ancestral to members of the genus *Gemmobacter* and has been gained. Based on our analysis, we suggest that the trait of methylated amine utilisation within the members of the genus *Gemmobacter* might be independent of their habitat and more randomly distributed.

Importance

Methylated amines are an important carbon and/or nitrogen for microorganisms in the environment. Recent advances in DNA sequencing and availability of bacterial genome sequences have allowed us to perform *in silico* experiments to understand and compare genetic potential. Using a comparative genomic approach, we show differential methylated amine utilisation trait within a bacterial genus i.e. *Gemmobacter* and that the trait is independent of the habitat. Ancestral state reconstruction analysis suggests that members

within the genus *Gemmobacter* that can use methylated amines might have gained the trait. Our study also emphasizes the need to use a holistic approach to understand the role of microorganisms in C and N cycling in the environment, rather than methods based on phylogenetic marker genes.

Introduction

Methylated amines (MAs) are ubiquitous in the environment with a range of natural and anthropogenic sources including the oceans, vegetation, sediments and organic-rich soils, animal husbandry, food industry, pesticides, combustion, sewage, and automobiles, to mention only a few (1-3). Methylated amines are also known to influence earth's climate, via a series of complex biological and chemical interactions (4). Some of the most abundant methylated amines found in the atmosphere are trimethylamine (TMA), dimethylamine (DMA) and monomethylamine (MMA) (1). Microbial metabolism of methylated amines involves aerobic and anaerobic microorganisms, e.g. some methanogenic archaea such as *Methanosarcina* and *Methanomicrobium* can use MAs to produce methane (5-7) while Gram-positive and Gram-negative methylotrophic bacteria can use MAs as carbon and nitrogen source (8). 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 (8). However, recent studies have reported the detection and activity of aerobic methylotrophic bacteria that utilize MAs in a variety of natural and anthropogenic environments (1, 9-12) 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 (9, 14, 15). Methylotrophs can degrade TMA to DMA by using the enzymes TMA dehydrogenase, TMA monooxygenase or TMA methyltransferase, encoded by the genes *tdm*, *tmm* and *mtt*, respectively (16-18). The enzymes DMA dehydrogenase (*dmd*) or DMA monooxygenase (*dmmDABC*) modulate the conversion of DMA to MMA (15, 16). Two distinct pathways have been characterised for the oxidation of MMA (10). The direct MMA-oxidation pathway mediated by a single enzyme (MMA dehydrogenase (MMADH) in gram negative bacteria and MMA oxidase in gram positive bacteria) converts MMA to formaldehyde and releases ammonium (19, 20). 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 ($CH_2 = H_4F$) (3, 10). A stepwise conversion of MMA in the NMG pathway is modulated by the enzymes GMA synthetase (*gmaS*), NMG synthase (*mgsABC*) and NMG dehydrogenase (*mgdABCD*) (3, 8). 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 but not as a carbon source (16, 21, 22).

In a recent study, we isolated an alphaproteobacterial facultative methylotrophic bacterium, *Gemmobacter* sp. LW-1 (recently renamed from *Catellibacterium* (23)) from the Movile Cave ecosystem (24) (Mangalia, Romania) that can use methylated amines as both carbon and nitrogen source (12) and subsequently obtained its genome sequence (25). 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 utiliser in microbial mats from this environment (12). This

was the first report of methylated amine utilisation in a member of the bacterial genus *Gemmobacter*. However, growth on C_1 compounds (methanol and formate) has been previously reported for the genus *Gemmobacter*, i.e. for *G. caeni* (26).

The genus *Gemmobacter* (family *Rhodobacteraceae*) currently comprises ten validated species: *Gemmobacter megaterium* (27), *G. nectariphilum* (23, 28), *G. aquatilis* (29), *G. caeni* (23, 26), *G. aquaticus* (23, 30), *G. nanjingense* (23, 31), *G. intermedius* (32), *G. lanyuensis* (33), *G. tilapiae* (34) and *G. fontiphilus* (23). These species were isolated from a wide range of environments including fresh water environments (freshwater pond (29, 34), freshwater spring (23, 33)), coastal planktonic seaweed (27), white stork nestling (32), waste water and activated sludge (26, 28, 31), suggesting that members of the genus *Gemmobacter* are widely distributed in anthropogenic and natural environments.

Here, using a comparative genomics approach we study how widespread methylated amine utilisation trait is within the members of the genus *Gemmobacter*. We used five isolate genomes for members within the genus *Gemmobacter* (*G*. sp. LW-1, *G. caeni*, *G. aquatilis*, *G. nectariphilus* and *G. megaterium)* available from public repositories (Accessed March 2017 and June 2018). These genomes were used for comparative genomics and ancestral state reconstruction analysis to understand patterns of gain/loss of methylated amine utilisation across the *Gemmobacter* phylogeny.

Materials and Methods

Genome data acquisition

Five *Gemmobacter* genomes (*G. caeni*, *G. aquatilis*, *G. nectariphilus*, *G. megaterium*, *Gemmobacter* sp. LW-1) available through the Integrated Microbial Genomes (IMG) database

(https://img.jgi.doe.gov/) were used for comparative genome analysis (35). Accession numbers and genome characteristics are listed in Supplementary Table S1.

Phylogenetic analysis

Phylogenetic relatedness between the different members of the genus *Gemmobacter* was determined using phylogenetic trees constructed from 16S rRNA gene sequences (nucleotide) and metabolic gene sequences (*gmaS* and *mauA*; amino acids) involved in MMA utilisation. RNAmmer (36) was used to retrieve 16S rRNA gene sequences from the genome sequences. Multiple sequence alignment of 16S rRNA gene sequences was performed using the SINA alignment service via SILVA (37, 38) and subsequently imported into MEGA7 (39) to construct a neighbour-joining nucleotide-based phylogenetic tree (40). Bootstrap analysis was performed with 1000 replicates to provide confidence estimates for phylogenetic tree topologies (41).

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) (42) and curated *gmaS* sequences used for primer design in our previous study (12). Amino acid sequences were aligned in MEGA7 (39) using ClustalW (43) and the alignment was subsequently used to construct maximum likelihood phylogenetic trees based on the JTT matrix-based model (44). Bootstrap analysis was performed with 1000 replicates to provide confidence estimates for phylogenetic tree topologies (41).

Comparative genomic analyses

CGView Comparison Tool (CCT) was used to visually compare the genomes within the genus *Gemmobacter* (45). CCT utilizes BLAST to compare the genomes and the BLAST

results are presented in a graphical map (45). Average Nucleotide Identity (ANI) (46) between different genomes was estimated using one-way ANI (best hit) and two-way ANI (reciprocal best hit) based on Goris *et al*. (47). 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 (48) 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,

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

AF represents alignment fraction.

Pan-genome analysis including average amino acid identity (AAI) analysis, pan-genome tree construction and determination of core and dispensable genes and singletons (unique genes) was carried out using the Efficient Database framework for comparative Genome Analyses using BLAST score Ratios (EDGAR) platform (49).

In order to compare the genetic potential for methylated amine utilisation within the available *Gemmobacter* genomes, known protein sequences involved in methylated amine utilisation pathways (3, 16) were used as query sequences through the BLAST (blastp) program (42) available within the Rapid Annotation using Subsystem Technology (RAST) server (50). The list of protein queries used is given in Supplementary Table S2.

We examined patterns of gain/loss of methylated amine utilisation along the above 16S rRNA gene phylogeny, by performing ancestral state reconstruction analysis using the phytools package in R (51). We used stochastic character mapping (52) to map presence/absence of methylated amine utilisation, assigning a prior probability of one to species known to utilise

methylated amines, zero to those known not to do so, and 0.5 to those where the trait value was unknown. We used MCMC (with the function *make.simmap)* to simulate 1000 stochastic maps, and then we used the function *densityMap* (53) to visualise the aggregate result from the stochastic mapping analysis.

Results and discussion

Phylogenetic relatedness based on the 16S rRNA and metabolic gene

The phylogenetic relatedness of the five members within the genus *Gemmobacter* (*G*. sp. LW-1, *G. caeni*, *G. aquatilis*, *G. nectariphilus* and *G. megaterium*) was resolved based on 16S rRNA gene sequences (Figure 1). Three members of the genus *Gemmobacter* (*G*. sp. LW-1, *G. caeni*, 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 (Figure 1). *G. nectariphilus* and *G. megaterium* sequences clustered together with *Paracoccus kawasakiensis* and other related *Gemmobacter* sequences from marine, fresh water and activated sludge environments (Figure 1). Based on the 16S rRNA gene sequences retrieved from public database, we observed that the members of the genus *Gemmobacter* are widely distributed in anthropogenic (such as activated sludge and clinical environments) and natural environments i.e. fresh water, soil and sediment, and marine environments (Figure 1).

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* sequences were used to construct an amino acid-based phylogenetic tree (Figure 2). Multiple sequence alignment and construction of a phylogenetic tree based on amino-acid sequences clearly separated *gmaS* from *glnA* gene sequences, a homolog of the *gmaS* gene encoding for the enzyme glutamine synthetase functionally unrelated (Figure 2). *gmaS* gene sequences

retrieved from genomes of *G*. sp. LW-1, *G. caeni* 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* as described in Wischer et al. (12) and were closely related to *Paracoccus yeei*, *P*. sp. 1W-5 and *Rhodobacter* sp. 1W-5 (Figure 2).

Whilst *gmaS* gene sequences were detected in three of the five investigated *Gemmobacter* genomes, *mauA* gene sequences were identified only in the genomes of *G. caeni* and *G*. sp. LW-1 (Supplementary Figure S1). It has been suggested that the NMG pathway for MMA utilisation is more universally distributed and more abundant across proteobacterial methylotrophs than the direct MMA oxidation pathway (13) although the NMG pathway is energetically more expensive compared to the energetically favourable direct MMA oxidation pathway (54). In addition, 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, but not as a carbon source (12, 16).

A comparative genome analysis of members within the genus Gemmobacter

At the time of the analysis, five *Gemmobacter* genomes obtained from isolates from different environments were available (Figure 1 and Table 2). *Gemmobacter* genome sizes range from ~3.96 Mb to ~5.14 Mb with GC contents between 64.71% to 66.19%. Analysis of sequence annotations revealed that on average 91.13% of the genomes consist of coding sequences.

The genomes were compared using the CGView comparison tool (45) was used (Figure 3). *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 (Figure 3). 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 (Figure 3) revealed low identity levels (mostly <88%) between *Gemmobacter* sp. LW-1 and *G. aquatilis*, *G. nectariphilus* and *G. megaterium* across the genomes. 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 2200-2300 kbp (Figure 3) 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 not an artefact arising from a lack of sensitivity in the BLAST analysis. Instead it could be an insertion or deletion of a large DNA fragment such as a prophage or a genomic island. Even though the genomes of *G.* sp. LW-1 and *G. caeni* are closely related, the 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) (46) (Supplementary Table S3 and Supplementary Figure S2A-D) and the average amino acid identity, short AAI (Supplementary Figure S2E). It is generally accepted that an ANI value of >95-96% can be used for species delineation (55-57). Our analysis revealed that *Gemmobacter* sp. LW-1 and *G. caeni* share an ANI value of 98.62 (Supplementary Table S3) 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 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 the EDGAR platform (49), identified metabolic genes present in all *Gemmobacter* species (core genes), two or more *Gemmobacter* species (accessory or dispensable genes), and unique *Gemmobacter* species (singleton genes). A pan-genome tree was constructed (Figure 4A) based on the pan-genome dataset and neighbor-joining method (40). As with the 16S-based phylogenetic tree (Figure 1), the five *Gemmobacter* species formed two main clusters in the pan-genome tree analysis (Figure 4A). The pan-genome tree also confirmed the evolutionary closeness between *Gemmobacter caeni* and *Gemmobacter* sp. LW-1 (Figure 4A). According to pan-genome analysis of the five *Gemmobacter* genomes, a total of 9,286 genes were identified, consisting of 1,806 core genes, 3085 dispensable genes and 305, 1,072, 896, 1,165 and 957 singletons for *G.* sp. LW-1, *G. caeni*, *G. aquatilis*, *G. nectariphilus* and *G. megaterium*, respectively (Figure 4B). On average 53.3% of singletons were identified as having hypothetical functions (Figure 4C). In addition, the number of singletons did not correlate with the size of the genome, which is in contrast with the correlation between the number of genes and the size of the genome.

Methylated amine utilisation, N assimilation and C_1 oxidation

Investigation of the methylated amine utilisation pathways in five *Gemmobacter* species revealed the presence of the genes encoding the enzymes TMA dehydrogenase (*tmd*), TMA monooxygenase (*tmm*), TMA methyltransferase (*mttB*), TMAO demethylase (*tdm*) and DMA monooxygenase in genomes of *G.* sp. LW-1, *G. caeni* and *G. aquatilis* while none of these genes were detected in *G. nectariphilus* or *G. megaterium* (Figure 5). 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 (12). *G*. sp. LW-1, *G. caeni* and *G. aquatilis* could potentially use the TMA oxidation or demethylation pathway to convert TMA to DMA. Based on the genome sequences, it can be suggested that these three species 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 (Figure 5).

In order to compare the distribution of the direct MMA-oxidation pathway and the NMG pathway, the genomes of five *Gemmobacter* species were tested for the presence of *mauA* and *gmaS* gene sequences (Figure 5). 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 (12, 13, 21, 54). Analysis of the genome sequences revealed that both *G*. sp. LW-1 and *G. caeni* possess genes for both MMA oxidation pathways (Figure 5). We have previously shown that *Gemmobacter* sp. LW-1 can use MMA and TMA as both a carbon and nitrogen source (12). Genome sequence of *G. aquatilis* indicated the presence of genes involved only in the NMG pathway. In the facultative methylotroph *Methylobacterium extorquens* AM1 it has been shown that the NMG pathway is advantageous compared to the direct MMA-oxidation pathway (54). Particularly under limiting MMA concentrations (<1 mM) the energetically expensive NMG pathway plays a key role during growth with MMA as a sole nitrogen source. 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 (54). This could suggest that *G. aquatilis* might use the NMG pathway for utilizing MMA as both nitrogen and carbon source. However, growth assays are required to confirm whether *G. aquatilis* can use MMA

as a carbon source. We did not detect genes for either MMA oxidation pathways (*mauA*- and *gmaS*-dependent) 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.

Gemmobacter sp. LW-1 was isolated from the Movile Cave ecosystem (12). Microbial mats and lake water within the cave have been shown to harbor a wide diversity of methylated amine-utilizing bacteria (12, 58). Whilst 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 (12). Similarly, *G. caeni* isolated from activated sludge (26) could possibly use the MAs generated from organic matter degradation. Interestingly, whilst *G. megaterium* was isolated from a marine environment (seaweed (27)) 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 utilisation. Our analysis suggest that the trait for methylated amine utilisation could be independent of the habitat.

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 five *Gemmobacter* species possess the genetic capability for C_1 oxidation and also indicate that tetrahydrofolate (H_4F) is the C_1 carrier (Figure 5). The bifunctional enzyme 5,10-methylene-tetrahydrofolate dehydrogenase/ cyclohydrolase, encoded by the gene *folD*, was detected in all the *Gemmobacter* genomes (Figure 5/ Table 1). Genes encoding key enzymes in the C_1 oxidation pathway via tetrahydromethanopterin (H_4MPT) were not detected. (10). The formate-tetrahydrofolate ligase, encoded by the gene *fhs* (Figure 5), provides C_1 units for biosynthetic pathways (16). 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 five *Gemmobacter* genomes.

The *xoxF* gene, encoding the lanthanide-dependent methanol dehydrogenase (59-61) and the *mxaF* gene, encoding the alpha subunit of the calcium-dependent methanol dehydrogenase (62), are responsible for the oxidation of methanol to formaldehyde (63, 64) and were both detected in genomes of *G*. sp. LW-1, *G. caeni* and *G. aquatilis* but not in *G. nectariphilus* and *G. megaterium* (Figure 5/ Table 1). The lanthanide-dependent methanol dehydrogenase is thought to be more widespread in bacterial genomes (65). The *fae* gene, encoding the formaldehyde-activating enzyme that catalyses the reduction of formaldehyde with H_4MPT was not detected in any of the five *Gemmobacter* genomes confirming that these members of the genus *Gemmobacter* lack the H_4MPT pathway for formaldehyde oxidation (Figure 5/ Table 1). However, in all the *Gemmobacter* genomes we detected genes encoding the glutathione (GSH)-dependent formaldehyde oxidation pathway, including the genes *adhI* / *flhA* and *fghA* coding for the GSH-dependent formaldehyde dehydrogenase and the S-formyl-GSH hydrolase, respectively (Figure 5 and Table 1). Reduction of formaldehyde with GSH can be mediated by the GSH-dependent formaldehyde-activating enzyme, encoded by the gene *gfa* (66). A Gfa-like protein was detected in the genomes of *G*. sp. LW-1, *G. caeni*, *G. necariphilus* and *G. megaterium* but not in *G. aquatilis*. This could perhaps indicate a non-essential role of Gfa in formaldehyde oxidation as suggested by Chen (16) and Wilson *et al*. (67). Investigation of the nitrogen assimilation pathway revealed the presence of the genes encoding glutamine synthetase (GS; *gluL*) and glutamine synthase (GOGAT; *glxB*) in all five *Gemmobacter* genomes. In bacteria this pathway is essential for glutamate synthesis at low ammonium concentrations (16).

Using comparative genome analysis we provide genome-based evidence that the two *Gemmobacter* isolates *G.* sp. LW-1 and *G. caeni* are capable of generating energy from complete oxidation of methylated amines via the H_4F-dependent pathway using either the energetically expensive NMG pathway or the energetically favourable direct MMA oxidation pathway. *Gemmobacter aquatilis* is genetically capable of methylated amine degradation to yield formaldehyde and only encodes the genes for the energetically expensive NMG pathway, which indicates that *G. aquatilis* could use this pathway to use MMA as a nitrogen source. Both *G. nectariphilus* and *G. megaterium* genomes indicate the lack of potential to use methylated amines (Figure 5/Table 1).

Evolution of methylated amine utilisation in Gemmobacter

Three of the five investigated *Gemmobacter* species are genetically capable of methylated amine utilisation, of which two (*G.* sp. LW-1 and *G. caeni*) possess the genes for both MMA utilisation pathways, whereas in the genome of *G. aquatilis* we detected only genes involved in the indirect MMA oxidation pathway (NMG pathway). The two other *Gemmobacter* species (*G. megaterium* and *G. nectariphilus*) do not possess the genes for methylated amine utilisation. Therefore, the question arises if the genes for methylated amine utilisation have been acquired or lost during evolution.

Stochastic character mapping along the 16S rRNA gene phylogeny suggested that the ability to use methylated amines has either been gained or lost multiple times (Supplementary Figure S3). Most likely is that methylated amine utilisation is not ancestral in *Gemmobacter*, and has evolved three times, once in the clade containing *G.* sp. LW-1, *G. caeni* and *G. aquatilis,* once in *Haematobacter* and once in *Paracoccus* (Supplementary Figure S3). An alternative, but less plausible, scenario is that methylated amine utilisation is an ancestral trait in *Gemmobacter,* and has been lost and regained multiple times across the phylogeny.

Conclusions

Three of the five investigated *Gemmobacter* genomes (*G.* sp. LW-1, *G. caeni* and *G. aquatilis*) indicated metabolic potential to utilize methylated amines, of which only two (*G.* sp. LW-1 and *G. caeni*) possess the genes for both the MMA oxidation pathways, the energetically expensive NMG pathway and the energetically favourable direct MMA oxidation pathway. *G.* sp. LW-1 and *G. caeni* are facultative methylotrophs which could potentially use these pathways to utilize MMA as both a carbon and nitrogen source, while potentially *G. aquatilis* could only use the NMG pathway as a nitrogen source. Furthermore, the genomes of *G.* sp. LW-1 and *G. caeni* showed a high similarity to each other (>98%) suggesting that both belong to the same species. *G. megaterium* and *G. nectariphilus* genomes indicated no metabolic potential to utilize MAs. Phylogenetic analysis (16S tree, Figure 1), pan-genome and ANI analysis revealed that *G.* sp. LW-1 and *G. caeni* are closely related, although they were isolated from different environments. Whilst *G. caeni* and *G. nectariphilus* were isolated from a similar environment (activated sludge) it revealed a high amount of evolutionary change from the common ancestor. Overall, these results suggest that the trait for methylated amine utilisation could be independent from the habitat and localised factors could influence the ability of these organisms to use methylated amines. Access to *Gemmobacter* isolates with or without the genetic potential for methylated amine utilisation trait will allow us to perform physiological experiments in future to test how this trait can affect fitness of closely related organisms. Ancestral state reconstruction analysis confirms that across *Gemmobacter* and related genera, methylated amine utilisation has either evolved or been lost multiple times over the evolutionary history of this group. The adaptive or non-adaptive processes behind this pattern remain to be investigated.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

We 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 EDGAR platform is financially supported by the BMBF grant FKZ 031A533 within the de.NBI network.

REFERENCES

1. Ge X, Wexler AS, Clegg SL. 2011. Atmospheric amines – Part I. A review. Atmospheric Environment 45:524-546.
2. Schade GW, Crutzen PJ. 1995. Emission of aliphatic-amines from animal husbandry and their reactions - Potential source of N^2O and Hcn. Journal of Atmospheric Chemistry 22:319-346.
3. Latypova E, Yang S, Wang YS, Wang T, Chavkin TA, Hackett M, Schäfer H, Kalyuzhnaya MG. 2010. Genetics of the glutamate-mediated methylamine utilisation pathway in the facultative methylotrophic beta-proteobacterium *Methyloversatilis universalis* FAM5. Molecular microbiology 75:426-439.
4. Carpenter LJ, Archer SD, Beale R. 2012. Ocean-atmosphere trace gas exchange. Chemical Society Reviews 41:6473-6506.
5. Lyimo TJ, Pol A, Jetten MSM, Op den Camp HJM. 2009. Diversity of methanogenic archaea in a mangrove sediment and isolation of a new *Methanococcoides* strain. Fems Microbiology Letters 291:247-253.
6. Burke SA, Lo SL, Krzycki JA. 1998. Clustered genes encoding the methyltransferases of methanogenesis from monomethylamine. Journal of Bacteriology 180:3432-3440.
7. Liu Y, Whitman WB. 2008. Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea. Ann N Y Acad Sci 1125:171-89.
8. Chen Y, Scanlan J, Song L, Crombie A, Rahman MT, Schäfer H, Murrell JC. 2010. γ-Glutamylmethylamide is an essential intermediate in the metabolism of methylamine by *Methylocella silvestris*. Applied and Environmental Microbiology 76:4530-4537.
9. Chistoserdova L, Kalyuzhnaya MG, Lidstrom ME. 2009. The expanding world of methylotrophic metabolism. Annual Review of Microbiology 63:477-499.
10. Chistoserdova L. 2011. Modularity of methylotrophy, revisited. Environmental Microbiology 13:2603-2622.
11. Chen Y, Wu L, Boden R, Hillebrand A, Kumaresan D, Moussard H, Baciu M, Lu Y, Colin Murrell J. 2009. Life without light: microbial diversity and evidence of sulfur- and ammonium-based chemolithotrophy in Movile Cave. ISME J 3:1093-1104.
12. Wischer D, Kumaresan D, Johnston A, El Khawand M, Stephenson J, Hillebrand-Voiculescu AM, Chen Y, Colin Murrell J. 2015. Bacterial metabolism of methylated amines and identification of novel methylotrophs in Movile Cave. ISME J 9:195-206.
13. Nayak DD, Marx CJ. 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.
14. Anthony C. 1982. The biochemistry of methylotrophs. Academic Press, London.
15. Lidstrom ME. 2006. Aerobic Methylotrophic Prokaryotes, p 618-634. *In* Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E (ed), The Prokaryotes: Volume 2: Ecophysiology and Biochemistry doi:10.1007/0-387-30742-7_20. Springer New York, New York, NY.
16. Chen Y. 2012. Comparative genomics of methylated amine utilisation by marine *Roseobacter* clade bacteria and development of functional gene markers (*tmm*, *gmaS*). Environmental Microbiology 14:2308-2322.
17. Lidbury I, Murrell JC, Chen Y. 2014. Trimethylamine *N*-oxide metabolism by abundant marine heterotrophic bacteria. Proceedings of the National Academy of Sciences 111:2710-2715.
18. Paul L, Ferguson DJ, Krzycki JA. 2000. The trimethylamine methyltransferase gene and multiple dimethylamine methyltransferase genes of *Methanosarcina barkeri*

contain in-frame and read-through amber codons. Journal of Bacteriology 182:2520-2529.

19. McIntire WS, Wemmer DE, Chistoserdova A, Lidstrom ME. 1991. A new cofactor in a prokaryotic enzyme: tryptophan tryptophylquinone as the redox prosthetic group in methylamine dehydrogenase. Science 252:817-824.
20. Chistoserdova AY, Christoserdova LV, McIntire WS, Lidstrom ME. 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.
21. Chen Y, McAleer KL, Murrell JC. 2010. Monomethylamine as a nitrogen source for a nonmethylotrophic bacterium, *Agrobacterium tumefaciens*. Applied and Environmental Microbiology 76:4102-4104.
22. Taubert M, Grob C, Howat AM, Burns OJ, Pratscher J, Jehmlich N, von Bergen M, Richnow HH, Chen Y, Murrell JC. 2017. Methylamine as a nitrogen source for microorganisms from a coastal marine environment. Environmental Microbiology 19:2246-2257.
23. Chen WM, Cho NT, Huang WC, Young CC, Sheu SY. 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*. International Journal of Systematic and Evolutionary Microbiology 63:470-478.
24. Kumaresan D, Wischer D, Stephenson J, Hillebrand-Voiculescu A, Murrell JC. 2014. Microbiology of Movile Cave - A chemolithoautotrophic ecosystem. Geomicrobiology Journal 31:186-193.
25. Kumaresan D, Wischer D, Hillebrand-Voiculescu AM, Murrell JC. 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.
26. Zheng J-W, Chen Y-G, Zhang J, Ni Y-Y, Li W-J, He J, 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*. International Journal of Systematic and Evolutionary Microbiology 61:1921-1926.
27. Liu J-J, Zhang X-Q, Chi F-T, Pan J, Sun C, Wu M. 2014. *Gemmobacter megaterium* sp. nov., isolated from coastal planktonic seaweeds. International Journal of Systematic and Evolutionary Microbiology 64:66-71.
28. Tanaka Y, Hanada S, Manome A, Tsuchida T, Kurane R, Nakamura K, 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. International Journal of Systematic and Evolutionary Microbiology 54:955-959.
29. Rothe B, Fischer A, Hirsch P, Sittig M, Stackebrandt E. 1987. The phylogenetic position of the budding bacteria *Blastobacter aggregatus* and *Gemmobacter aquatilis* gen., nov. sp. nov. Archives of Microbiology 147:92-99.
30. Liu Y, Xu C-J, Jiang J-T, Liu Y-H, Song X-F, Li H, Liu Z-P. 2010. *Catellibacterium aquatile* sp. nov., isolated from fresh water, and emended description of the genus *Catellibacterium* Tanaka et al. 2004. International Journal of Systematic and Evolutionary Microbiology 60:2027-2031.

31. Zhang J, Chen S-A, Zheng J-W, Cai S, Hang B-J, He J, Li S-P. 2012. *Catellibacterium nanjingense* sp. nov., a propanil-degrading bacterium isolated from activated sludge, and emended description of the genus *Catellibacterium*. International Journal of Systematic and Evolutionary Microbiology 62:495-499.
32. Kämpfer P, Jerzak L, Wilharm G, Golke J, Busse H-J, Glaeser SP. 2015. *Gemmobacter intermedius* sp. nov., isolated from a white stork (*Ciconia ciconia*). International Journal of Systematic and Evolutionary Microbiology 65:778-783.
33. Sheu S-Y, Shiau Y-W, Wei Y-T, Chen W-M. 2013. *Gemmobacter lanyuensis* sp. nov., isolated from a freshwater spring. International Journal of Systematic and Evolutionary Microbiology 63:4039-4045.
34. Sheu S-Y, Sheu D-S, Sheu F-S, Chen W-M. 2013. *Gemmobacter tilapiae* sp. nov., a poly-β-hydroxybutyrate-accumulating bacterium isolated from a freshwater pond. International Journal of Systematic and Evolutionary Microbiology 63:1550-1556.
35. Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Pillay M, Ratner A, Huang J, Woyke T, Huntemann M. 2013. IMG 4 version of the integrated microbial genomes comparative analysis system. Nucleic acids research 42:D560-D567.
36. Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic acids research 35:3100-3108.
37. Pruesse E, Peplies J, Glöckner FO. 2012. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics 28:1823-1829.
38. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic acids research 35:7188-7196.
39. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular biology and evolution 33:1870-1874.
40. Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular biology and evolution 4:406-425.
41. Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783-791.
42. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. Journal of molecular biology 215:403-410.
43. Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic acids research 22:4673-4680.
44. Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. Bioinformatics 8:275-282.
45. Grant JR, Arantes AS, Stothard P. 2012. Comparing thousands of circular genomes using the CGView Comparison Tool. BMC genomics 13:202.
46. Rodriguez-R LM, Konstantinidis KT. 2016. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Preprints,
47. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. 2007. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. International journal of systematic and evolutionary microbiology 57:81-91.
48. Konstantinidis KT, Tiedje JM. 2005. Genomic insights that advance the species definition for prokaryotes. Proceedings of the National Academy of Sciences of the United States of America 102:2567-2572.

49. Blom J, Kreis J, Spanig S, Juhre T, Bertelli C, Ernst C, Goesmann A. 2016. EDGAR 2.0: an enhanced software platform for comparative gene content analyses. Nucleic Acids Research 44:W22-W28.
50. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M. 2008. The RAST Server: rapid annotations using subsystems technology. BMC genomics 9:75.
51. Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3:217-223.
52. Nielsen R. 2002. Mapping mutations on phylogenies. Systematic biology 51:729-739.
53. Revell LJ. 2013. Two new graphical methods for mapping trait evolution on phylogenies. Methods in Ecology and Evolution 4:754-759.
54. Nayak DD, Agashe D, Lee M-C, Marx CJ. 2016. Selection maintains apparently degenerate metabolic pathways due to tradeoffs in using methylamine for carbon versus nitrogen. Current Biology 26:1-11.
55. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol 57:81-91.
56. Richter M, Rossello-Mora R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A 106:19126-19131.
57. Kim M, Oh H-S, Park S-C, Chun J. 2014. Towards a taxonomic coherence between average nucleotide identity and 16S gene sequence similarity for species demarcation of prokaryotes. International Journal of Systematic and Evolutionary Microbiology 64:346-351.
58. Kumaresan D, Stephenson J, Doxey AC, Bandukwala H, Brooks E, Hillebrand-Voiculescu A, Whiteley AS, Murrell JC. 2018. Aerobic proteobacterial methylotrophs in Movile Cave: genomic and metagenomic analyses. Microbiome 6.
59. Hibi Y, Asai K, Arafuka H, Hamajima M, Iwama T, Kawai K. 2011. Molecular structure of La3+-induced methanol dehydrogenase-like protein in *Methylobacterium radiotolerans*. Journal of Bioscience and Bioengineering 111:547-549.
60. Nakagawa T, Mitsui R, Tani A, Sasa K, Tashiro S, Iwama T, Hayakawa T, Kawai K. 2012. A catalytic role of XoxF1 as La3+-dependent methanol dehydrogenase in *Methylobacterium extorquens* strain AM1. Plos One 7.
61. Pol A, Barends TRM, Dietl A, Khadem AF, Eygensteyn J, Jetten MSM, Op den Camp HJM. 2014. Rare earth metals are essential for methanotrophic life in volcanic mudpots. Environmental Microbiology 16:255-264.
62. Williams PA, Coates L, Mohammed F, Gill R, Erskine PT, Coker A, Wood SP, Anthony C, Cooper JB. 2005. The atomic resolution structure of methanol dehydrogenase from *Methylobacterium extorquens*. Acta Crystallographica Section D-Biological Crystallography 61:75-79.
63. Harms N, Ras J, Koning S, Reijnders WNM, Stouthamer AH, Spanning RGM. 1996. Genetics of C1 metabolism regulation in *Paracoccus denitrificans*, p 126-132. *In* Lidstrom ME, Tabita FR (ed), Microbial growth on C(1) compounds. Kluwer Academic Publishers, Dordrecht, the Netherlands.
64. Christoserdova L, Lidstrom ME. 1997. Molecular and mutational analysis of a DNA region separating two methylotrophy gene clusters in *Methylobacterium extorquens* AM1. Microbiology 143:1729-1736.
65. Lv H, Sahin N, Tani A. 2018. Draft genome and description of *Novimethylophilus kurashikiensis* gen. nov. sp. nov., a new lanthanide-dependent methylotrophic species of *Methylophilaceae*. Environmental Microbiology doi:10.1111/1462-2920.14062.
66. Goenrich M, Bartoschek S, Hagemeier CH, Griesinger C, Vorholt JA. 2002. A glutathione-dependent formaldehyde-activating enzyme (Gfa) from *Paracoccus*

denitrificans detected and purified via two-dimensional proton exchange NMR spectroscopy. J Biol Chem 277:3069-3072.

67. Wilson SM, Gleisten MP, Donohue TJ. 2008. Identification of proteins involved in formaldehyde metabolism by *Rhodobacter sphaeroides*. Microbiology 154:296-305.

Figures and Tables

Figure 1. Phylogenetic tree based on 16S rRNA gene sequences. Star represents the *Gemmobacter* species used for comparative genome analysis. The tree was constructed using the neighbour-joining method for clustering and the maximum composition likelihood method for computing evolutionary distances. Numbers at branches are bootstrap percentages >50% of 1000 replicates. Coloured boxes 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 recently reclassified to *Gemmobacter* (23). Scale bar: 0.02 substitutions per nucleotide position.

Figure 2. Maximum-likelihood phylogenetic tree based on *gmaS* gene sequences. Members of the genus *Gemmobacter* used for genome comparison are represented with a star. Amino acid sequences (GmaS) were aligned using the ClustalW algorithm. The tree was constructed using the maximum-likelihood method based on the JTT matrix-based model. Numbers at branches are bootstrap percentages >50% of 1000 replicates. Amino acid sequences of the glutamine synthetase type III (GlnA) were used as out-group. Scale bar: 0.1 substitutions per amino acid position. MRC, marine *Roseobacter* clade.

Figure 3. DNA BLAST map of *Gemmobacter* genomes. *Gemmobacter* sp. LW-1 was used as a reference genome against *Gemmobacter megaterium* (inner ring), *Gemmobacter aquatilis* (second inner ring), *Gemmobacter nectariphilus* (third ring), and *Gemmobacter caeni* (fourth ring). The fifth and sixth ring (outer rings) represent the CDS (blue), tRNA (maroon), and rRNA (purple) on the reverse and forward strand, respectively. The color scale (inset) shows the level of 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.

Figure 4. Pan-genome analysis of Gemmobacters. (A) Pan-genome tree consisting of five *Gemmobacter* species was constructed using the neighbour-joining method within the EDGAR platform. **(B)** Number of core, dispensable, and specific genes (singletons) of each *Gemmobacter* species. **(C)** Proportion of hypothetical and uncharacterized proteins in the core, dispensable and singleton genome of five *Gemmobacter* species.

Figure 5. Metabolic pathways involved in methylated amine utilisation and one-carbon utilisation 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 nectariphilus* (N) and *Gemmobacter megaterium* (M). The color-coded boxes next to the genes indicate the presence (green) or absence (orange) of a gene in each genome. Yellow, gfa-like

Table 1. Comparative genomic analysis of methylated amine-utilizing genes in genomes-sequenced *Gemmobacters* in comparison to selected marine *Roseobacter* clade bacteria. Shown is the presence (+) or absence (-) of specific genes in the genome sequences.

Supplementary Figure S1. 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 >50% of 1000 replicates. Scale bar: 0.1 substitutions per amino acid. Coloured boxes indicate *Alphaproteobacteria (*yellow), *Gammaproteobacteria (*red) and *Betaproteobacteria (*black).

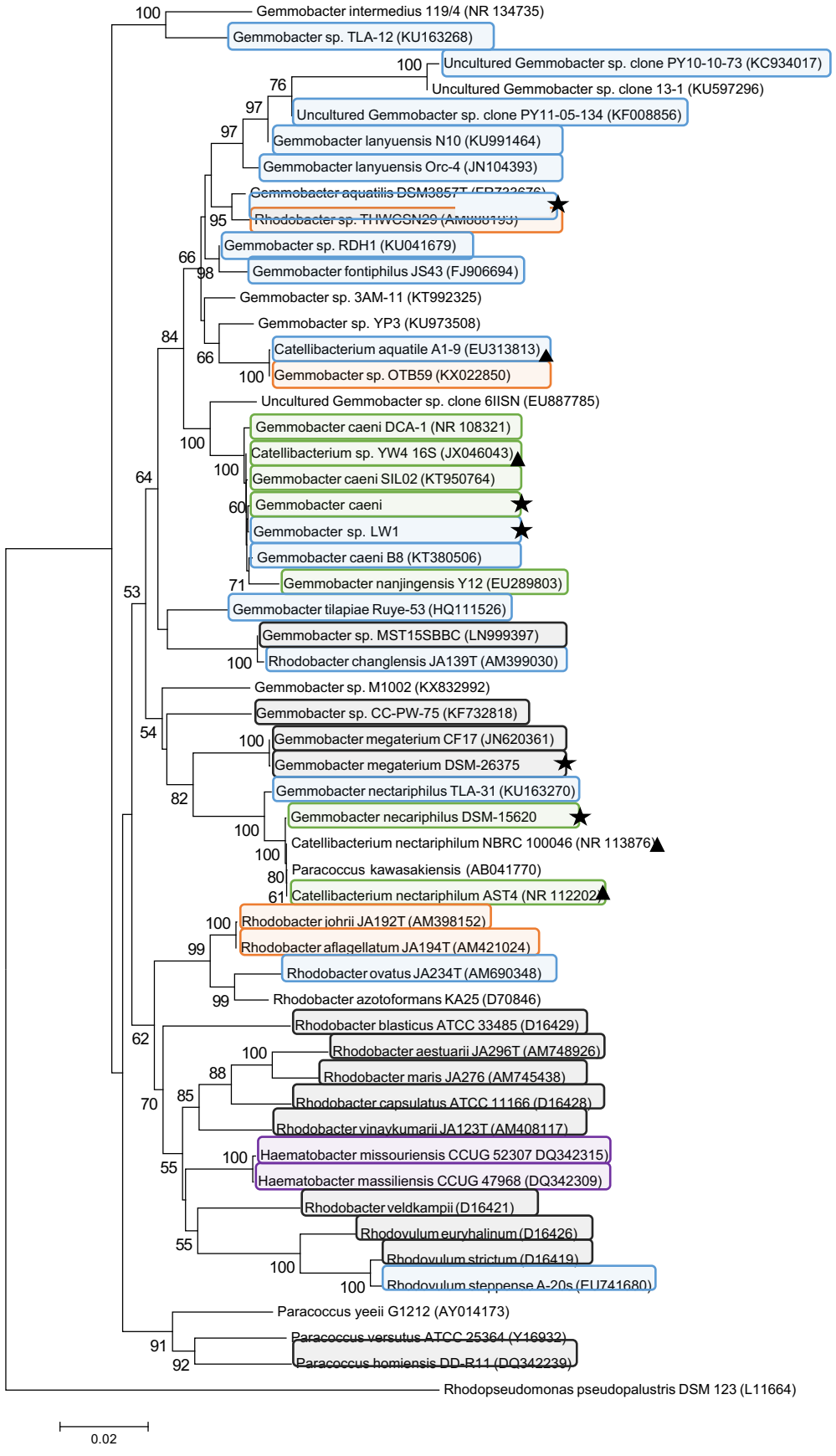
Supplementary Figure S2. (A-D) Average nucleotide identity (ANI) analysis of *Gemmobacter* sp. LW-1 and *Gemmobacter caeni*, *Gemmobacter aquatilis*, *Gemmobacter nectariphilus* and *Gemmobacter megaterium* and (E) AAI analysis between those species

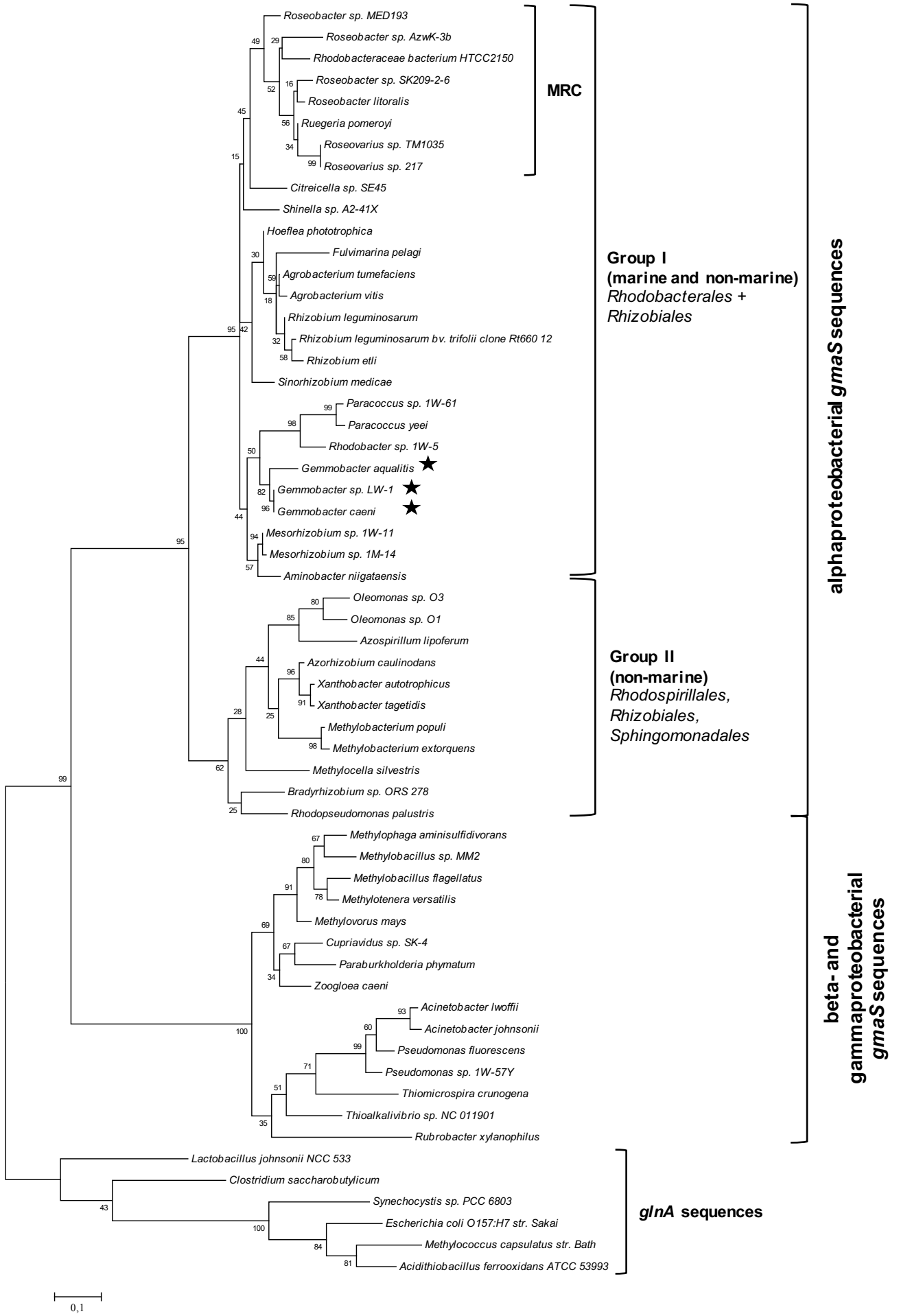
Supplementary Figure S3. Ancestral state reconstruction of methylated amine utilisation along the 16S rRNA gene phylogeny, using stochastic mapping. Branch colour represents the posterior probability (computed as the relative frequency across stochastic maps) of methylated amine utilisation through the phylogeny. Red indicates a high posterior probability of methylated amine utilisation. Star represents the *Gemmobacter* species used for comparative genome analysis.

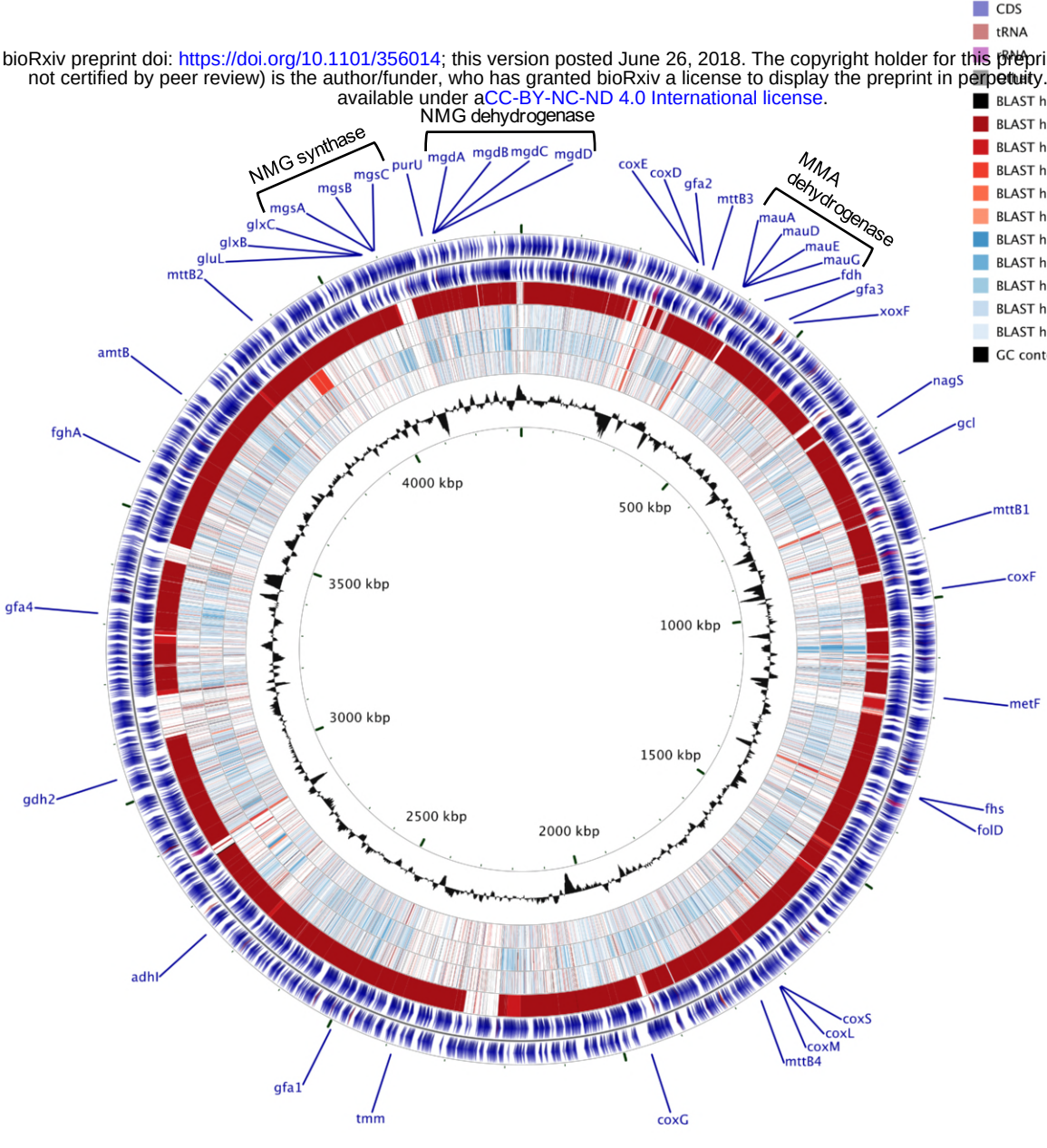
Supplementary Table S1. Genome characteristics of the five *Gemmobacter* isolate genomes used in this study.

Supplementary Table S2. List of protein queries used for the genome comparison with their accession number.

Supplementary Table S3. Average nucleotide identity (ANI) values between *Gemmobacter* sp. LW-1 and *Gemmobacter caeni*, *Gemmobacter aquatilis*, *Gemmobacter nectariphilus*, *Gemmobacter megaterium*, *Rhodobacter sphaeroides* and *Paracoccus denitrificans*

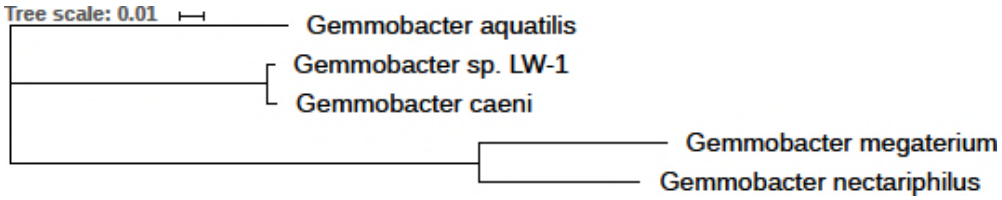

Gemmobacter intermedius 119/4 (NR 134735)
Gemmobacter sp. TLA-12 (KU163268)
Uncultured Gemmobacter sp. clone PY10-10-73 (KC934017)
Uncultured Gemmobacter sp. clone 13-1 (KU597296)
Uncultured Gemmobacter sp. clone PY11-05-134 (KF008856)
Gemmobacter lanyuensis N10 (KU991464)
Gemmobacter lanyuensis Orc-4 (JN104393)
Gemmobacter aquatilis DSM3857T (FR733676)
Rhodobacter sp. THWCSN29 (AM888193)
Gemmobacter sp. RDH1 (KU041679)
Gemmobacter fontiphilus JS43 (FJ906694)
Gemmobacter sp. 3AM-11 (KT992325)
Gemmobacter sp. YP3 (KU973508)
Catellibacterium aquatile A1-9 (EU313813)
Gemmobacter sp. OTB59 (KX022850)
Uncultured Gemmobacter sp. clone 6IISN (EU887785)
Gemmobacter caeni DCA-1 (NR 108321)
Catellibacterium sp. YW4 16S (JX046043)
Gemmobacter caeni SIL02 (KT950764)
Gemmobacter caeni
Gemmobacter sp. LW1
Gemmobacter caeni B8 (KT380506)
Gemmobacter nanjingensis Y12 (EU289803)
Gemmobacter tilapiae Ruye-53 (HQ111526)
Gemmobacter sp. MST15SBBC (LN999397)
Rhodobacter changlensis JA139T (AM399030)
Gemmobacter sp. M1002 (KX832992)
Gemmobacter sp. CC-PW-75 (KF732818)
Gemmobacter megaterium CF17 (JN620361)
Gemmobacter megaterium DSM-26375
Gemmobacter nectariphilus TLA-31 (KU163270)
Gemmobacter necariphilus DSM-15620
Catellibacterium nectariphilum NBRC 100046 (NR 113876)
Paracoccus kawasakiensis (AB041770)
Catellibacterium nectariphilum AST4 (NR 112202)
Rhodobacter johrii JA192T (AM398152)
Rhodobacter aflagellatum JA194T (AM421024)
Rhodobacter ovatus JA234T (AM690348)
Rhodobacter azotoformans KA25 (D70846)
Rhodobacter blasticus ATCC 33485 (D16429)
Rhodobacter aestuarii JA296T (AM748926)
Rhodobacter maris JA276 (AM745438)
Rhodobacter capsulatus ATCC 11166 (D16428)
Rhodobacter vinaykumarii JA123T (AM408117)
Haematobacter missouriensis CCUG 52307 DQ342315)
Haematobacter massiliensis CCUG 47968 (DQ342309)
Rhodobacter veldkampii (D16421)
Rhodovulum euryhalinum (D16426)
Rhodovulum strictum (D16419)
Rhodovulum steppense A-20s (EU741680)
Paracoccus yeeii G1212 (AY014173)
Paracoccus versutus ATCC 25364 (Y16932)
Paracoccus homiensis DD-R11 (DQ342239)
Rhodopseudomonas pseudopalustris DSM 123 (L11664)
0.02


Roseobacter sp. MED193
Roseobacter sp. AzwK-3b
Rhodobacteraceae bacterium HTCC2150
Roseobacter sp. SK209-2-6
Roseobacter litoralis
Ruegeria pomeroyi
Roseovarius sp. TM1035
Roseovarius sp. 217
MRC
Citreicella sp. SE45
Shinella sp. A2-41X
Hoeflea phototrophica
Fulvimarina pelagi
Agrobacterium tumefaciens
Agrobacterium vitis
Rhizobium leguminosarum
Rhizobium leguminosarum bv. trifolii clone Rt660 12
Rhizobium etli
Sinorhizobium medicae
Group I
(marine and non-marine)
Rhodobacterales +
Rhizobiales
Paracoccus sp. 1W-61
Paracoccus yeei
Rhodobacter sp. 1W-5
Gemmobacter aqualitis
Gemmobacter sp. LW-1
Gemmobacter caeni
Mesorhizobium sp. 1W-11
Mesorhizobium sp. 1M-14
Aminobacter niigataensis
Oleomonas sp. O3
Oleomonas sp. O1
Azospirillum lipoferum
Azorhizobium caulinodans
Xanthobacter autotrophicus
Xanthobacter tagetidis
Methylobacterium populi
Methylobacterium extorquens
Methylocella silvestris
Bradyrhizobium sp. ORS 278
Rhodopseudomonas palustris
Group II
(non-marine)
Rhodospirillales,
Rhizobiales,
Sphingomonadales
alphaproteobacterial gmaS sequences
Methylophaga aminisulfidivorans
Methylobacillus sp. MM2
Methylobacillus flagellatus
Methylotenera versatilis
Methylovorus mays
Cupriavidus sp. SK-4
Paraburkholderia phymatum
Zoogloea caeni
Acinetobacter lwoffii
Acinetobacter johnsonii
Pseudomonas fluorescens
Pseudomonas sp. 1W-57Y
Thiomicrospira crunogena
Thioalkalivibrio sp. NC 011901
Rubrobacter xylanophilus
beta- and
gammaproteobacterial
gmaS sequences
Lactobacillus johnsonii NCC 533
Clostridium saccharobutylicum
Synechocystis sp. PCC 6803
Escherichia coli O157:H7 str. Sakai
Methylococcus capsulatus str. Bath
Acidithiobacillus ferrooxidans ATCC 53993
glnA sequences
0,1

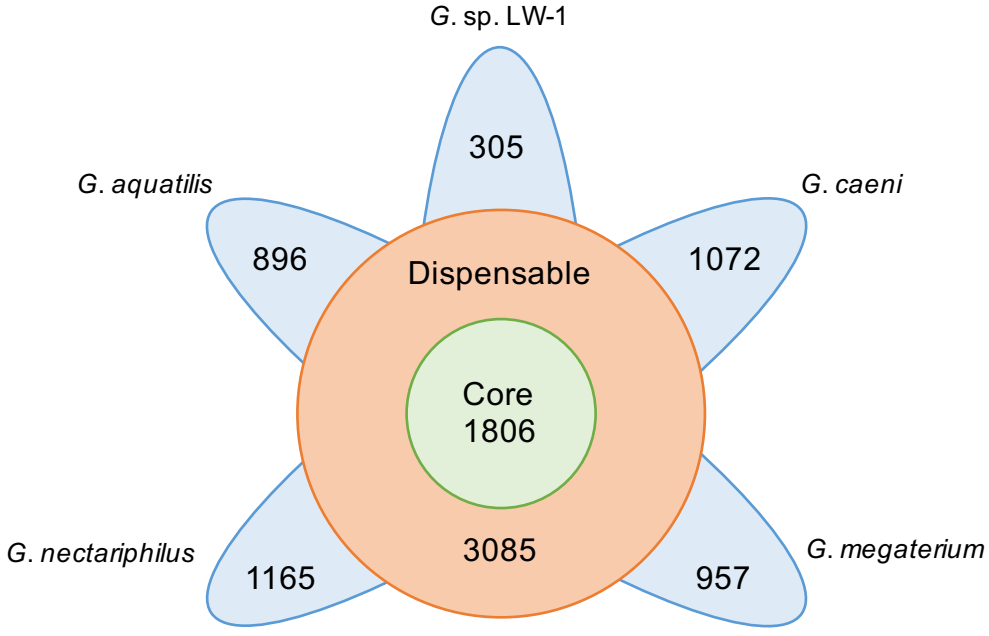


Protein	Corresponding gene(s)
TMA monooxygenase	tmm
TMA methyltransferase	mttB1, mttB2, mttB3, mttB4
MMA dehydrogenase	mauA, mauD, mauE, mauG
NMG synthase	mgsA, mgsB, mgsC
NMG dehydrogenase	mgdA, mgdB, mgdC, mgdD
HCHO-activating enzyme	gfa1, gfa2, gfa3, gfa4
GSH-dependent HCHO dehydrogenase	adhI
S-formyl-GSH 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

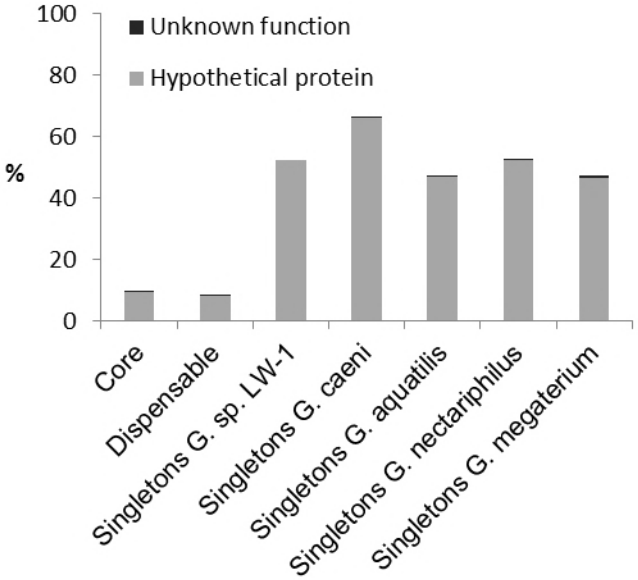
A

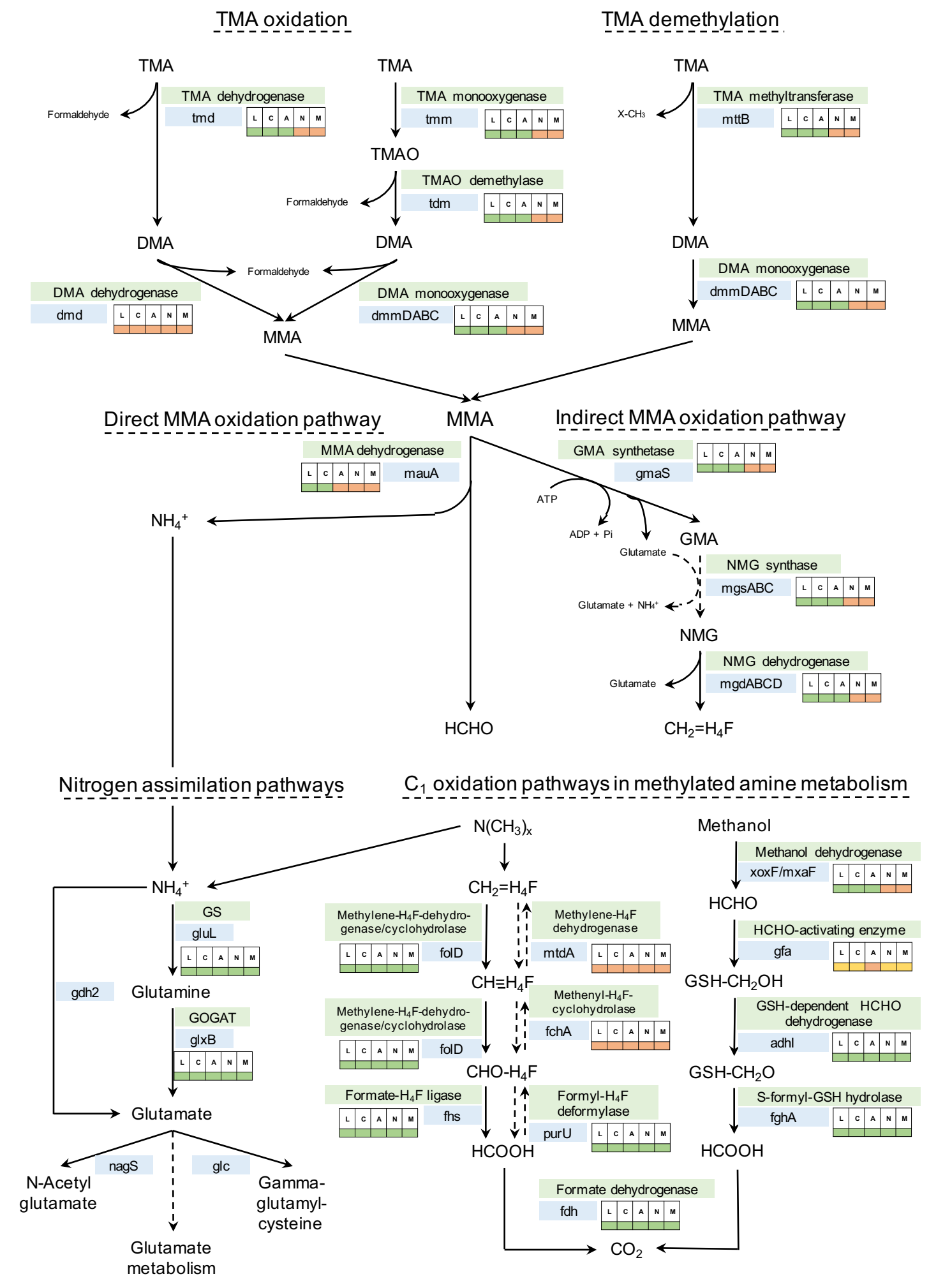

Tree scale: 0.01
Gemmobacter aquatilis
Gemmobacter sp. LW-1
Gemmobacter caeni
Gemmobacter megaterium
Gemmobacter nectariphilus

B


G. sp. LW-1
305
G. aquatilis
896
G. caeni
1072
Dispensable
Core
1806
3085
G. nectariphilus
1165
957
G. megaterium

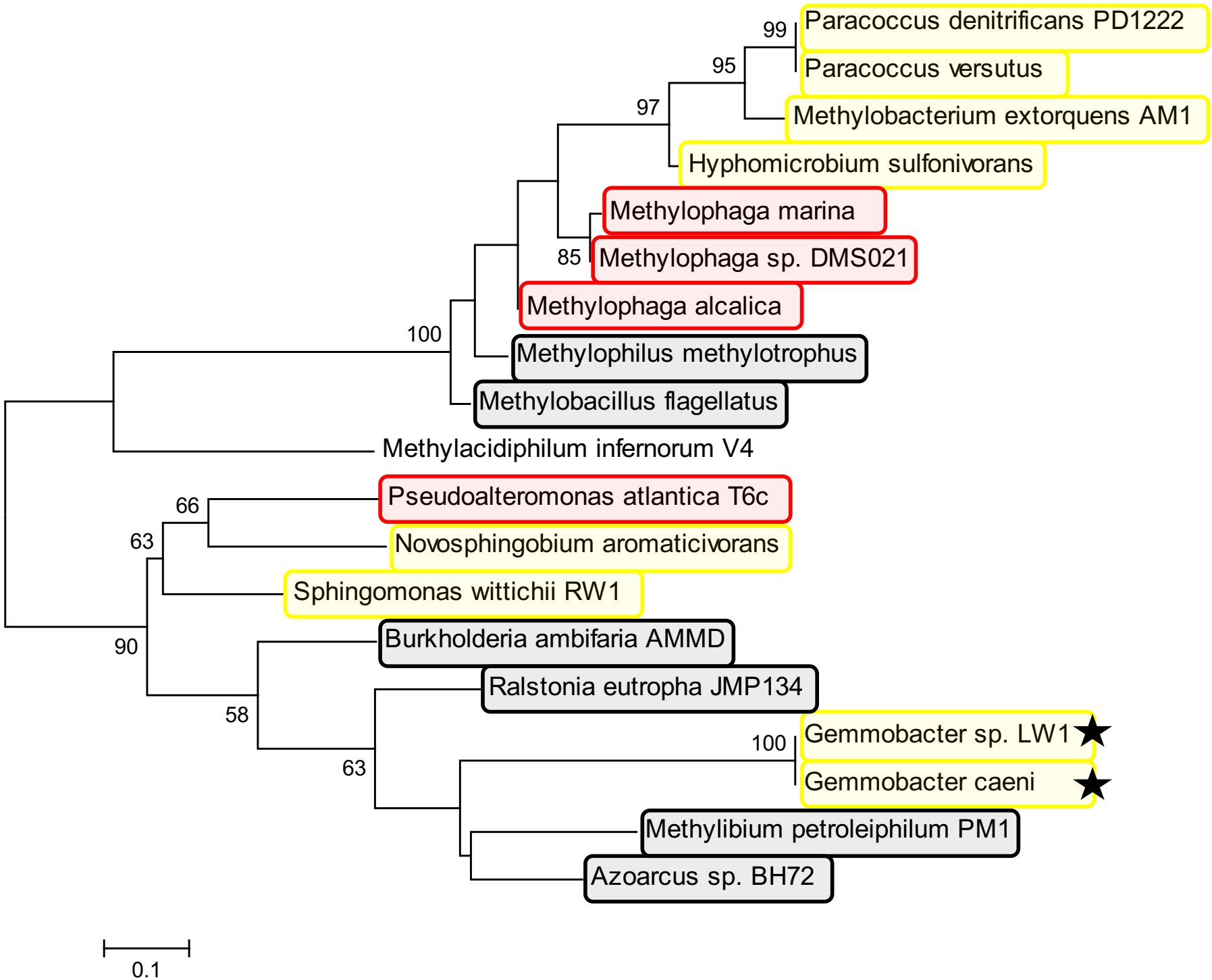
C


100
80
60
40
20
0
%
Unknown function
Hypothetical protein
Core
Dispensable
Singletons G. sp. LW-1
Singletons G. caeni
Singletons G. aquatilis
Singletons G. nectariphilus
Singletons G. megaterium

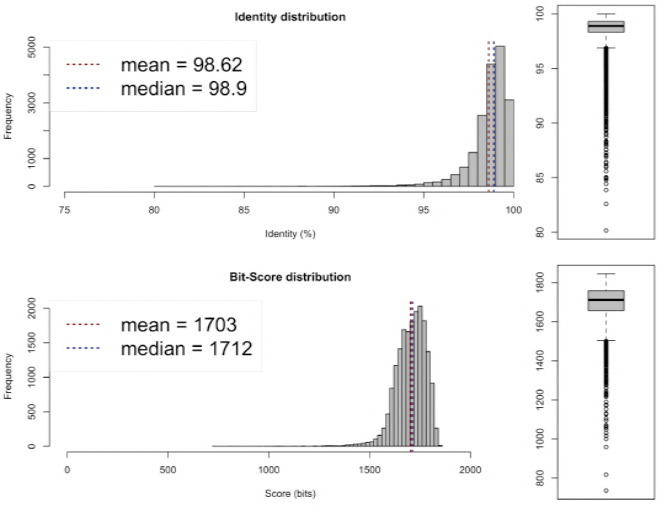

TMA oxidation
TMA demethylation
TMA
Formaldehyde
TMA dehydrogenase
tmd
TMA monooxygenase
tmm
TMAO
TMAO demethylase
tdm
X-CH3
TMA methyltransferase
mttB
DMA
DMA dehydrogenase
dmd
DMA monooxygenase
dmmDABC
MMA
Direct MMA oxidation pathway
Indirect MMA oxidation pathway
MMA dehydrogenase
mauA
GMA synthetase
gmaS
ATP
ADP + Pi
GMA
Glutamate
NMG synthase
mgsABC
Glutamate + NH4+
NMG
NMG dehydrogenase
mgdABCD
NH4+
HCHO
CH2=H4F
Nitrogen assimilation pathways
C1 oxidation pathways in methylated amine metabolism
N(CH3)x
Methanol
Methanol dehydrogenase
xoxF/mxaF
GS
gluL
gdh2
Glutamine
GOGAT
glxB
Glutamate
nagS
glc
N-Acetyl glutamate
Gamma-glutamyl-cysteine
Glutamate metabolism
Methylene-H4F-dehydrogenase/cyclohydrolase
folD
Methylene-H4F dehydrogenase
mtdA
CH≡H4F
Methenyl-H4F-cyclohydrolase
fchA
CHO-H4F
Formate-H4F ligase
fhs
Formyl-H4F deformylase
purU
HCOOH
Formate dehydrogenase
fdh
CO2
HCHO-activating enzyme
gfa
GSH-CH2OH
GSH-dependent HCHO dehydrogenase
adhI
GSH-CH2O
S-formyl-GSH hydrolase
fghA
HCOOH
L C A N M

Organism	*gmaS*	*tmm*	*xoxF*	*fae*	*foldD*	*purU*	*fhs (ftfL)*	*gfa*	*adhI (flhA)*	*fghA*	*fdh*	*amtB*	*mtdA/ mtdB*	*mchA*	*ftr*	*fmdA /B/C*	*mauA*	*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]	+	+	+	+	-/-	-	-	-	-	+

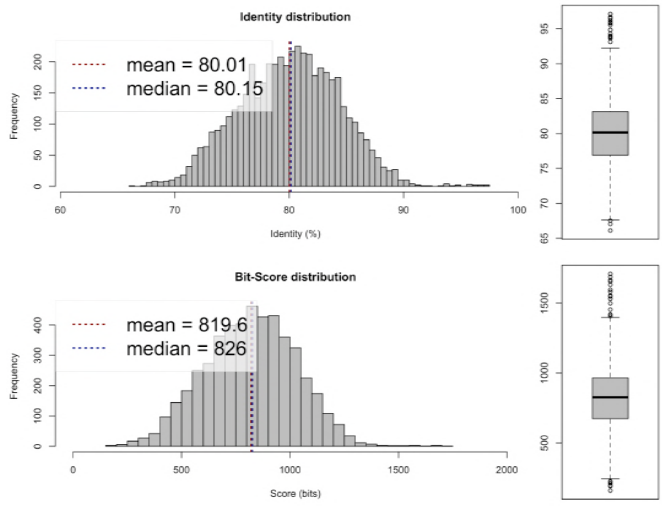
[a] gfa-like protein-encoding gene


Paracoccus denitrificans PD1222
Paracoccus versutus
Methylobacterium extorquens AM1
Hyphomicrobium sulfonivorans
Methylophaga marina
Methylophaga sp. DMS021
Methylophaga alcalica
Methylophilus methylotrophus
Methylobacillus flagellatus
Methylacidiphilum infernorum V4
Pseudoalteromonas atlantica T6c
Novosphingobium aromaticivorans
Sphingomonas wittichii RW1
Burkholderia ambifaria AMMD
Ralstonia eutropha JMP134
Gemmobacter sp. LW1
Gemmobacter caeni
Methylibium petroleiphilum PM1
Azoarcus sp. BH72
99
95
97
85
100
66
63
90
58
100
63
0.1

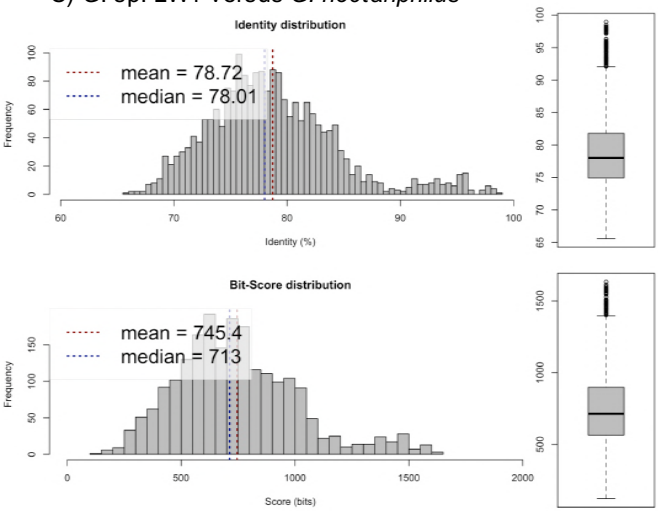
A) *G.* sp. LW1 versus *G. caeni*



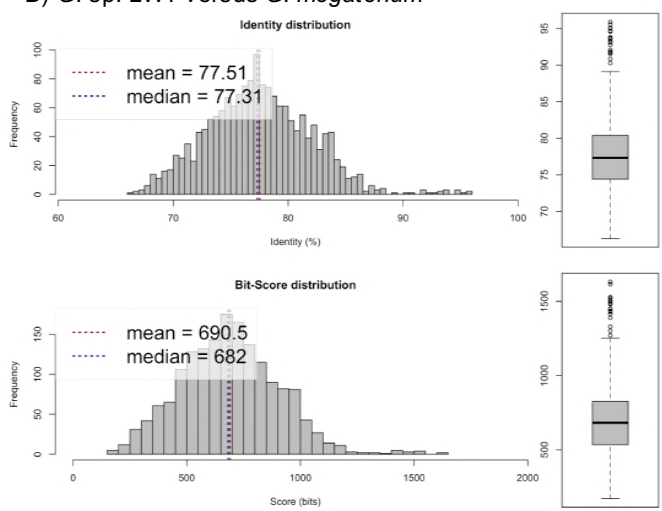
B) *G.* sp. LW1 versus *G. aquatilis*



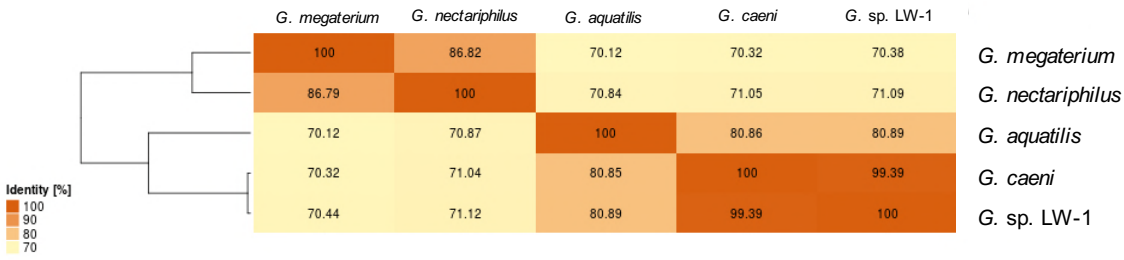
C) *G.* sp. LW1 versus *G. nectariphilus*



D) *G.* sp. LW1 versus *G. megaterium*



E) AAI analysis




Rhodopseudomonas pseudopalustris DSM 123 L11664
Gemmobacter sp. TLA−12 KU163268
Gemmobacter intermedius 119 4 NR 134735
Paracoccus homiensis DD−R11 DQ342239
Paracoccus versutus ATCC 25364 Y16932
Paracoccus yeeii G1212 AY014173
Rhodovulum steppense A−20s EU741680
Rhodovulum strictum D16419
Rhodovulum euryhalinum D16426
Rhodobacter veldkampii D16421
Haematobacter massiliensis CCUG 47968 DQ342309
Haematobacter missouriensis CCUG 52307 DQ342315
Rhodobacter vinaykumarii JA123T AM408117
Rhodobacter capsulatus ATCC 11166 D16428
Rhodobacter maris JA276 AM745438
Rhodobacter aestuarii JA296TAM748926
Rhodobacter blasticus ATCC 33485 D16429
Rhodobacter azotoformans KA25 D70846
Rhodobacter ovatus JA234T AM690348
Rhodobacter aflagellatum JA194T AM421024
Rhodobacter johrii JA192T AM398152
Catellibacterium nectariphilum NBRC 100046 NR 113876
Paracoccus kawasakiensis AB041770
Catellibacterium nectariphilum AST4 NR 112202
Gemmobacter necariphilus DSM−15620 ★
Gemmobacter nectariphilus TLA−31 KU163270
Gemmobacter megaterium DSM−26375 ★
Gemmobacter megaterium CF17 JN620361
Gemmobacter sp. CC−PW−75 KF732818
Gemmobacter sp. M1002 KX832992
Rhodobacter changlensis JA139T AM399030
Gemmobacter sp. MST15SBBC LN999397
Gemmobacter tilapiae Ruye−53 HQ111526
Gemmobacter caeni ★
Gemmobacter sp. LW1 ★
Catellibacterium sp. YW4 16S JX046043
Gemmobacter caeni DCA−1 NR 108321
Gemmobacter caeni SIL02 KT950764
Gemmobacter nanjingensis Y12 EU289803
Gemmobacter caeni B8 KT380506
Uncultured Gemmobacter sp. clone 6IISN EU887785
Gemmobacter fontiphilus JS43 FJ906694
Gemmobacter sp. RDH1 KU041679
Gemmobacter aqualis
Gemmobacter aquatilis DSM3857T FR733676 ★
Rhodobacter sp. THWCSN29 AM888193
Gemmobacter sp. 3AM−11 KT992325
Catellibacterium aquatile A1 9 EU313813
Gemmobacter sp. OTB59 KX022850
Gemmobacter sp. yp3 KU973508
Gemmobacter lanyuensis Orc−4 JN104393
Gemmobacter lanyuensis N10 KU991464
Uncultured Gemmobacter sp. clone PY11 05 134 KF008856
Uncultured Gemmobacter sp. clone 13 1 KU597296
Uncultured Gemmobacter sp. clone PY10 10 73 KC934017
0
PP(state=Present)
1
length=0.054

Organism	Isolation	IMG genome ID	Genome size (Number of total bases)	Gene count (Number of total genes)	CDS count (Number of CDS genes)	rRNA count (Number of rRNA genes)	Protein coding genes with function prediction	Protein coding genes without function prediction
G. sp. LW-1	Movile Cave	2648501906	4,344,434	4,312	4,223	18	3,443	780
G. caeni	Activated Sludge	2739367660	5,135,347	5,228	5,122	3	3,856	1,266
G. aquatilis	Aquatic Pond	2615840721	3,964,038	3,918	3,856	5	3,105	751
G. nectariphilus	Activated Sludge	2524614723	4,523,636	4,462	4,405	5	3,394	1,011
G. megaterium	Seaweed	2681813556	4,168,088	4,075	4,020	3	3,273	747

Protein	Short name	Accession number (NCBI)
Ammonium transporter	AmtB	WP_054302626
Carbon monoxide dehydrogenase	Cox	KRS16390
DMA dehydrogenase	Dmd	S57961
DMA monooxygenase	DmmD, DmmA, DmmB, DMMC	WP_011047316, WP_044028112, WP_011047318, WP_011047319
Formaldehyde-activating enzyme	Fae	KGJ23184
Formate dehydrogenase	Fdh	WP_054303846
Formyl-H_4F deformylase	PurU	WP_028029509
Formyl-H_4F synthetase	Fhs/FtfL	Q83WS0
Formylmethanofuran dehydrogenase	FmdA, FmdB, FmdC	CAF30065, CAF30067, CAF30066
Formylmethanofuran-H_4MPT-N-formyltransferase	Ftr	AKJ40646
Glutamine synthetase	GluL	WP_054301068
Glutamine:2-oxoglutarate aminotransferase (glutamate synthase)	GlxB	EHP89794
Glutathione-dependent formaldehyde dehydrogenase	AdhI/FlhA	YP_352634
GMA synthetase	GmaS	KM083625
GSH-dependent formaldehyde-activating enzyme	Gfa	WP_011746365
Methanol dehydrogenase-like protein	XoxF	ACS40517
Methanol dehydrogenase subunit 1	MxaF	P16027
Methenyl-H_4F cyclohydrolase	FchA	Q49135
Methenyl-H_4MPT cyclohydrolase	MchA	Q8KMJ5
Methylene-H_4F dehydrogenase	MtdA	CAD13312
Methylene-H_4F dehydrogenase/ cyclohydrolase	FolD	KNX40634
Methylene-H_4MTP dehydrogenase	MtdB	CAD13313
MMA dehydrogenase large subunit	MauG	WP_082401055
MMA dehydrogenase small subunit	MauA	AAF03760
N-methyl-glutamate dehydrogenase	MgdA	EPX83758
N-methyl-glutamate synthase	MgsA, MgsB, MgsC	KGM86398, KGM86397, KGM86396
S-formyl-glutathione hydrolase	FghA	YP_001967351
TMA dehydrogenase	Tmd	ABG35545
TMA methyltransferase	MttB	O93658
TMA monooxygenase	Tmm	AAV94838
TMAO demethylase	Tdm	ACK52488

	G. sp. LW-1	*G. caeni*	*G. aquatilis*	*G. nectariphilus*	*G. megaterium*	*Rhodobacter sphaeroides*	*Paracoccus denitrificans*
***G*. sp. LW-1**		98.62	80.01	78.72	77.51	77.91	77.67
G. caeni	98.62		80.04	79.77	77.74	77.81	77.70
G. aquatilis	80.01	80.04		77.97	77.73	77.89	77.11
G. nectariphilus	78.72	79.77	77.97		82.82	77.74	78.37
G. megaterium	77.51	77.74	77.73	82.82		77.36	77.78
Rhodobacter sphaeroides	77.91	77.81	77.89	77.74	77.36		78.42
Paracoccus denitrificans	77.67	77.70	77.11	78.37	77.78	78.42	