1 **Title:**
2 **Comparative genomics analyses indicate differential methylated amine**
3 **utilisation trait within members of the genus *Gemmobacter***

4
5 Running title:
6 Methylated amine utilisation trait in *Gemmobacter*

7
8 Kröber, E.¹, Peixoto, J.², Cunningham, MR.⁴, Spurgin L.³, Wischer, D.³, Kruger, R.²,
9 Kumaresan, D.^{4#}

10
11
12
13 ¹ Microbial Biogeochemistry, RA Landscape Functioning, ZALF Leibniz Centre for
14 Landscape Research, Müncheberg, Germany

15 ²Department of Cellular Biology, University of Brasília, Brasília, Brazil

16 ³ School of Environmental Sciences, University of East Anglia, Norwich, UK

17 ⁴ School of Biological Sciences & Institute for Global Food Security, Queen's
18 University Belfast, Belfast, Northern Ireland, UK

19
20
21
22 [#]Correspondence to

23 Dr D Kumaresan, School of Biological Sciences, Queen's University Belfast, Belfast,
24 Northern Ireland, United Kingdom, d.kumaresan@qub.ac.uk

25

26

27

28

29

30

31

32

33

34

35

36

37

38 **Abstract**

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

53

54 **Introduction**

55 Methylated amines (MAs) are ubiquitous in the environment with a variety of natural
56 and anthropogenic sources including the oceans, vegetation, sediments and organic-
57 rich soils, animal husbandry, food industry, pesticides, sewage, and automobiles, to
58 mention only a few [1-3]. Methylated amines are also known to influence earth's
59 climate, via a series of complex biological and chemical interactions [4]. Some of the
60 most abundant methylated amines found in the atmosphere are trimethylamine
61 (TMA), dimethylamine (DMA) and monomethylamine (MMA) [1]. Microbial

62 metabolism of methylated amines involves both aerobic and anaerobic
63 microorganisms, e.g. some methanogenic archaea such as *Methanosarcina* and
64 *Methanomicrobium* can use MAs to produce methane [5-7] while Gram-positive and
65 Gram-negative methylotrophic bacteria can use MAs as carbon and nitrogen source
66 [8]. Previously, MAs were typically associated with marine ecosystems as they are
67 by-products of degradation of osmolytic chemicals such as glycine betaine, carnitine,
68 choline and trimethylamine N-oxide [8]. However, recent studies have reported the
69 detection and activity of aerobic methylotrophic bacteria that utilise MAs in a variety
70 of natural and engineered environments [1, 9-12] and could play a major role in
71 global C and N budgets.

72

73 Aerobic methylotrophs are a polyphyletic group of microorganisms capable of
74 utilising one-carbon (C₁) compounds such as methane, methanol or methylated
75 amines as their sole source of carbon and energy [9, 13, 14]. Methylotrophs can
76 degrade TMA to DMA by using the enzymes TMA dehydrogenase, TMA
77 monooxygenase or TMA methyltransferase (under anaerobic conditions by
78 methylotrophic methanogens), encoded by the genes *tdm*, *tmm* and *mtt*, respectively
79 [15-18]. The enzymes DMA dehydrogenase (*dmd*) or DMA monooxygenase
80 (*dmmDABC*) modulate the conversion of DMA to MMA [14, 15]. Two distinct
81 pathways have been characterised for the oxidation of MMA [10]. The direct MMA-
82 oxidation pathway mediated by a single enzyme (MMA dehydrogenase in Gram-
83 negative bacteria and MMA oxidase in Gram-positive bacteria) converts MMA to
84 formaldehyde and releases ammonium [19, 20]. The alternate pathway, referred to as
85 the *N*-methylglutamate (NMG) pathway or indirect MMA-oxidation pathway, is
86 mediated by three individual enzymes via the oxidation of MMA to gamma-

87 glutamylmethanamide (GMA) and its further degradation to *N*-methylglutamate
88 (NMG) and 5,10-methylenetetrahydrofolate ($\text{CH}_2 = \text{H}_4\text{F}$) [3, 10]. A stepwise
89 conversion of MMA in the NMG pathway is modulated by the enzymes GMA
90 synthetase (*gmaS*), NMG synthase (*mgsABC*) and NMG dehydrogenase (*mgdABCD*)
91 [3, 8]. The capability to use MMA not only as a source for carbon but also for
92 nitrogen is widespread in bacteria. Notably, the NMG pathway is not only restricted
93 to methylotrophs but also present in non-methylotrophic bacteria that use MMA as a
94 nitrogen but not as a carbon source [15, 21, 22].

95
96 In a recent study, we isolated an alphaproteobacterial facultative methylotrophic
97 bacterium, *Gemmobacter* sp. LW-1 (recently renamed from *Catellibacterium* [23])
98 from the Movile Cave ecosystem (Mangalia, Romania) [24] that can use methylated
99 amines as both carbon and nitrogen source [12] and subsequently obtained its genome
100 sequence [25]. Using a ^{13}C -MMA DNA based stable-isotope probing (SIP)
101 experiment we also showed that *Gemmobacter* sp. LW-1 was indeed an active MMA
102 utiliser in microbial mats from this environment [12]. This was the first report of
103 methylated amine utilisation in a member of the bacterial genus *Gemmobacter*.
104 However, growth on C_1 compounds (methanol and formate) has been reported for the
105 genus *Gemmobacter*, e.g. in *G. caeni* [26]. The genus *Gemmobacter* (family
106 *Rhodobacteraceae*) currently comprises ten validated species: *Gemmobacter*
107 *megaterium* [27], *G. nectarophilum* [23, 28], *G. aquatilis* [29], *G. caeni* [23, 26], *G.*
108 *aquaticus* [23, 30], *G. nanjingense* [23, 31], *G. intermedius* [32], *G. lanyuensis* [33],
109 *G. tilapiae* [34] and *G. fontiphilus* [23]. These species were isolated from a wide
110 range of environments including fresh water environments (freshwater pond [29, 34],
111 freshwater spring [23, 33]), coastal planktonic seaweed [27], white stork nestling [32],

112 waste water and activated sludge [26, 28, 31], suggesting that members of the genus
113 *Gemmobacter* are widely distributed in engineered and natural environments.

114

115 Here, using a comparative genomics approach we study how widespread methylated
116 amine utilisation trait (i.e. metabolic potential) is within the members of the genus
117 *Gemmobacter*. We used five isolate genomes for members within the genus
118 *Gemmobacter* (*G. sp.* LW-1, *G. caeni*, *G. aquatilis*, *G. nectariphilus* and *G.*
119 *megaterium*) alongside genomes of four closely related organisms within the family
120 *Rhodobacteraceae* to show that the methylated amine utilisation trait is not common
121 within the members of the genus *Gemmobacter*.

122

123 **Materials and Methods**

124 *Genome data acquisition*

125 Five *Gemmobacter* genomes (*G. caeni*, *G. aquatilis*, *G. nectariphilus*, *G. megaterium*,
126 *Gemmobacter sp.* LW-1) available through the Integrated Microbial Genomes (IMG)
127 database (<https://img.jgi.doe.gov/>) were used for comparative genome analysis [35].
128 Accession numbers and genome characteristics are listed in Supplementary Table S1.

129

130 *Phylogenetic analysis*

131 Phylogenetic relatedness between the different members of the genus *Gemmobacter*
132 was determined using phylogenetic trees constructed from 16S rRNA gene sequences
133 (nucleotide) and metabolic gene sequences (*gmaS* and *mauA*; amino acids) involved
134 in MMA utilisation. RNAmmer [36] was used to retrieve 16S rRNA gene sequences
135 from the genome sequences. Multiple sequence alignment of 16S rRNA gene
136 sequences was performed using the SINA alignment service via SILVA [37, 38] and

137 subsequently imported into MEGA7 [39] to construct a maximum-likelihood
138 nucleotide-based phylogenetic tree [40]. Bootstrap analysis was performed with 1000
139 replicates to provide confidence estimates for phylogenetic tree topologies [41].

140

141 To determine phylogenetic affiliations for the protein encoding genes *gmaS* and
142 *mauA*, gene sequences retrieved from the genome sequences were aligned to
143 homologous sequences retrieved from the NCBI Genbank database using Basic Local
144 Alignment Search Tool (BLAST, blastx) [42] and curated *gmaS* sequences used for
145 primer design in our previous study [12]. Amino acid sequences were aligned in
146 MEGA7 [39] using ClustalW [43] and the alignment was subsequently used to
147 construct maximum likelihood phylogenetic trees based on the JTT matrix-based
148 model [44]. Bootstrap analysis was performed with 1000 replicates to provide
149 confidence estimates for phylogenetic tree topologies [41].

150

151 We investigated the phylogeny of nine bacterial species – *Gemmobacter aquatilis*,
152 *Gemmobacter caeni*, *Gemmobacter* sp. LW-1, *Gemmobacter megabacterium* DSM-
153 26375, *Gemmobacter nectarophilus* DSM-15620, *Paracoccus denitrificans* PD1222,
154 *Citricella* sp. SE45, *Roseovarius* sp. TM1035 and *Rhodobacter sphaeroides* 241 –
155 using the protein sequences of 831 orthologs present in single copy in all species
156 identified using OrthoFinder [45] with default settings. For each set of orthologous
157 proteins, multiple alignments were produced using MAFFT [46] with the --auto
158 settings. Subsequently, conserved alignment blocks were identified using trimAl [47]
159 with the option -automated1. The phylogenetic reconstruction analysis was based in a
160 final concatenated alignment consisting of 265,536 amino acid positions, which was
161 the input for RAxML [48] with the option PROTCATLGF. The selection of best

162 amino acid substitution model was performed after implementing the script
163 ProteinModelSelection.pl provided by the RAxML package. In order to assess the
164 environmental distribution of the genus *Gemmobacter*, we used MAPseq [49]
165 (<https://beta.microbeatlas.org>) to survey the relative abundance based on 16S rRNA
166 gene sequences (query sequence: *Gemmobacter aquatilis* DSM3857 (NR_104740.1 at
167 97% cut-off).

168

169 *Comparative genomic analyses*

170 CGView Comparison Tool (CCT) was used to visually compare the genomes within
171 the genus *Gemmobacter* [50]. CCT utilises BLAST to compare the genomes and the
172 BLAST results are presented in a DNA-based graphical map [50]. Average
173 Nucleotide Identity (ANI) [51] between different genomes was estimated using one-
174 way ANI (best hit) and two-way ANI (reciprocal best hit) based on Goris *et al.* [52].
175 In addition the whole-genome based average nucleotide identity (gANI) and the $p_r^{\text{intra-}}$
176 species value were determined for *G. sp.* LW-1 and *G. caeni* (these two genomes
177 revealed the closest ANI) based on Konstantinidis and Tiedje [53] via the Joint
178 Genome Institute (JGI) platform (<https://ani.jgi-psf.org/html/home.php>; Version 0.3,
179 April 2014). In order to determine if two genomes belong to the same species, the
180 computation of empirical probabilities ($p_r^{\text{intra-species}}$) can be calculated as follows,

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

181 AF represents alignment fraction.

182

183 Pan-genome analysis including average amino acid identity (AAI) analysis, pan-
184 genome tree construction and determination of core and dispensable genes and
185 singletons (unique genes) was carried out using the Efficient Database framework for

186 comparative Genome Analyses using BLAST score Ratios (EDGAR) platform [54].
187 Estimation of genomic completeness and contamination was carried out using the
188 CheckM program [55].

189

190 In order to compare the genetic potential for methylated amine utilisation within the
191 available *Gemmobacter* genomes, known protein sequences involved in methylated
192 amine utilisation pathways [3, 15] were used as query sequences through the BLAST
193 (blastp) program [42] available within the Rapid Annotation using Subsystem
194 Technology (RAST) server [56]. The list of protein queries used is given in
195 Supplementary Table S2.

196

197 **Results and discussion**

198 *Analysis of phylogenetic relatedness*

199 The phylogenetic relatedness of the five members within the genus *Gemmobacter* (*G.*
200 sp. LW-1, *G. caeni*, *G. aquatilis*, *G. nectariphilus* and *G. megaterium*) was resolved
201 based on 16S rRNA gene sequences (Figure 1). Three members of the genus
202 *Gemmobacter* (*G. sp. LW-1*, *G. caeni*, and *G. aquatilis*) clustered together with
203 several other related *Gemmobacter* and *Rhodobacter* 16S rRNA gene sequences
204 retrieved from fresh water, soil and sediment and activated sludge environments
205 (Figure 1). *G. nectariphilus* and *G. megaterium* sequences clustered together with
206 *Paracoccus kawasakiensis* and other related *Gemmobacter* sequences from marine,
207 fresh water and activated sludge environments (Figure 1). Based on the 16S rRNA
208 gene sequences retrieved from public database and MAPseq analyses, we observed
209 that the members of the genus *Gemmobacter* are widely distributed in engineered
210 (such as activated sludge and clinical environments) and natural environments i.e.

211 fresh water, soil and sediment, and marine environments (Figure 1 and Supplementary
212 Figure S3). Phylogenomic analysis based on single copy marker genes
213 (Supplementary Figure S4) revealed that *G. sp* LW1, *G. caeni* and *G. aquatilis* were
214 closely related to *Rhodobacter sphaeroides* 241 whereas *G. megabacterium* and *G.*
215 *nectariphilus* to *Paracoccus denitrificans*.

216

217 GMA synthetase, a key enzyme in the NMG pathway, is encoded by the gene *gmaS*.
218 *gmaS* sequences retrieved from the isolate genomes along with other ratified *gmaS*
219 sequences were used to construct an amino acid-based phylogenetic tree (Figure 2).
220 *gmaS* gene sequences retrieved from genomes of *G. sp.* LW-1, *G. caeni* and *G.*
221 *aquatilis* clustered within Group I of alphaproteobacterial *gmaS* sequences containing
222 sequences from marine and non-marine bacteria within the orders *Rhodobacterales*
223 and *Rhizobiales* as described in Wischer et al. [12] and were closely related to
224 *Paracoccus yeei*, *P. sp.* 1W-5 and *Rhodobacter sp.* 1W-5 (Figure 2). Whilst *gmaS*
225 gene sequences were detected in three of the five investigated *Gemmobacter*
226 genomes, *mauA* gene sequences were identified only in the genomes of *G. caeni* and
227 *G. sp.* LW-1 (Supplementary Figure S1). It has been suggested that the NMG
228 pathway for MMA utilisation is more universally distributed and more abundant
229 across proteobacterial methylotrophs than the direct MMA oxidation pathway [57].
230 However, it should be noted that genes encoding for the enzymes within the NMG
231 pathway (*gmaS*) can not only be detected in methylotrophs but also in non-
232 methylotrophic bacteria that use MMA as a nitrogen source, but not as a carbon
233 source [12, 15].

234

235 ***A comparative genome analysis of members within the genus Gemmobacter***

236 At the time of the analysis, five *Gemmobacter* genomes obtained from isolates from
237 different environments were available (Figure 1 and Table 2). *Gemmobacter* genome
238 sizes range from ~3.96 Mb to ~5.14 Mb with GC contents between 64.71% to 66.19%
239 and genome completeness between 98.61% to 99.69% (Table 2; Supplementary Table
240 1). Analysis of sequence annotations revealed that on average 91.13% of the genomes
241 consist of coding sequences.

242

243 The genomes were compared using the CGView comparison tool [50] (Figure 3).
244 *Gemmobacter* sp. LW-1, isolated from the Mobile Cave ecosystem was used as the
245 reference genome and the results of the BLAST comparison with other *Gemmobacter*
246 genomes are represented as a BLAST ring for each genome (Figure 3). Similarities
247 between segments of the reference genome sequence and the other genome sequences
248 are shown by a coloured arc beneath the region of similarity indicating the percentage
249 of similarity as a colour code. Our analysis (Figure 3) revealed low identity levels
250 (mostly <88%) between *Gemmobacter* sp. LW-1 and *G. aquatilis*, *G. nectarophilus*
251 and *G. megaterium* across the genomes. Moreover, the analysis suggested several
252 sites of potential insertion/deletion events in the genome of *Gemmobacter* sp. LW-1.
253 Possible insertion/deletion regions can be identified as those gaps in the map where
254 no homology is detected. For example, the region between 2200-2300 kbp (Figure 3)
255 where a gap can be found in the otherwise contiguous homologous regions between
256 the reference genome *G. sp. LW-1* and the first of the query genomes (*G. caeni*).
257 This might likely be due to a lack of hits or hits with low identity that can be spurious
258 matches. Since it covers a large region we could possibly rule out that it is not an
259 artefact arising from a lack of sensitivity in the BLAST analysis. Even though the
260 genomes of *G. sp. LW-1* and *G. caeni* are closely related, our analysis demonstrates

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

266

267 In order to further resolve the similarity between these genomes we calculated the
268 average nucleotide identity (ANI) [51] (Supplementary Table S3 and Supplementary
269 Figure S2A-D) and the average amino acid identity,(AAI; Supplementary Figure
270 S2E). It is generally accepted that an ANI value of >95-96% can be used for species
271 delineation [58, 59]. Our analysis revealed that *Gemmobacter sp. LW-1* and
272 *Gemmobacter caeni* share an ANI value of 98.62 (Supplementary Table S3) implying
273 that both are in fact the same species. The genome-based average nucleotide identity
274 (gANI) between *G. sp. LW-1* and *G. caeni* was calculated as 98.70. The AF was
275 calculated to be 0.91, which would result in a computed probability of 0.98
276 suggesting that both genomes might belong to the same species. However, it should
277 be noted that these are draft genomes and a more in depth characterization of their
278 physiology and phenotype is required to delineate these organisms at the level of
279 strain.

280

281 Pan-genome analysis, carried out using the EDGAR platform [54], identified
282 metabolic genes present in all *Gemmobacter* species (core genes), two or more
283 *Gemmobacter* species (accessory or dispensable genes), and unique *Gemmobacter*
284 species (singleton genes). A pan-genome tree was constructed (Figure 4A) based on
285 the pan-genome dataset and neighbor-joining method [40]. As with the 16S-rRNA

gene 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 phylogenetic 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).

Methylated amine utilisation, N assimilation and C₁ oxidation

Investigation of the methylated amine utilisation pathways in five *Gemmobacter* species revealed the presence of the genes encoding enzymes TMA dehydrogenase (*tmd*), TMA monooxygenase (*tmm*), 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 pathway to convert TMA to DMA. Based on the genome sequences, it can be suggested that these three species could 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).

We also compared the distribution of the direct MMA-oxidation and the NMG pathways in the genomes of five *Gemmobacter* species (Figure 5) and gene arrangement (Supplementary Figure S5). 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, 21, 57, 60]. 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 [60]. 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 [60]. This could suggest that *G. aquatilis* might use the NMG pathway for utilising 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 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.

330

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 C₁ oxidation pathway via tetrahydromethanopterin (H₄MPT) were not detected. [10]. The formate-tetrahydrofolate ligase, encoded by the gene *fhs* (Figure 5), provides C₁ units for biosynthetic pathways [15]. However, the oxidation of formyl-H₄F (CHO-H₄F) can also be facilitated by *purU*, the gene encoding for the formyl-H₄F deformylase. The formate dehydrogenase (*fdh*) 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 five *Gemmobacter* genomes.

The *fae* gene, encoding the formaldehyde-activating enzyme that catalyses the reduction of formaldehyde with H₄MPT was not detected in any of the five *Gemmobacter* genomes confirming that these members of the genus *Gemmobacter* lack the H₄MPT pathway for formaldehyde oxidation (Figure 5/ Table 1). 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 [15].

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 H₄F-dependent pathway using either the NMG pathway or the direct MMA oxidation pathway. *Gemmobacter aquatilis* is genetically capable of methylated amine degradation to yield formaldehyde and only encodes the genes for the NMG pathway, which indicates that

361 *G. aquatilis* could use this pathway to use MMA as a nitrogen source. Both *G.*
362 *nectariphilus* and *G. megaterium* genomes indicate the lack of potential to use
363 methylated amines (Figure 5/Table 1).

364

365 *Gemmobacter* sp. LW-1 was isolated from the Movile Cave ecosystem [12].
366 Microbial mats and lake water within the cave have been shown to harbor a wide
367 diversity of methylated amine-utilising bacteria [12, 61]. Whilst the mechanism of
368 MAs production within the system has to be elucidated, it can be speculated that
369 degradation of floating microbial mats (i.e. organic matter) could result in MAs [12].
370 Similarly, *G. caeni* isolated from activated sludge [26] could possibly use the MAs
371 generated from organic matter degradation. Interestingly, whilst *G. megaterium* was
372 isolated from a marine environment (seaweed [27]) possibly encountering MAs from
373 the degradation of osmolytes such as glycine betaine (*N,N,N*-trimethylglycine) we did
374 not detect metabolic genes involved in methylated amine utilisation. Our genome-
375 based analyses suggest that the potential for methylated amine utilisation is
376 differential within the members of the genus *Gemmobacter* that requires confirmation
377 using growth/activity assays.

378

379 **Conclusions**

380 In summary, three of the five investigated *Gemmobacter* genomes (*G. sp.* LW-1, *G.*
381 *caeni* and *G. aquatilis*) indicated metabolic potential to utilise methylated amines, of
382 which only two (*G. sp.* LW-1 and *G. caeni*) possess the genes for both MMA
383 oxidation pathways, the NMG pathway and the direct MMA oxidation pathway. *G.*
384 *sp.* LW-1 and *G. caeni* are facultative methylotrophs which could potentially use
385 these pathways to utilise 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. nectarophilus* genomes indicated no metabolic potential to utilise MAs. Phylogenetic, pan-genome and ANI analyses revealed that *G. sp.* LW-1 and *G. caeni* are closely related, although they were isolated from different environments. Overall, our results suggest that the trait for methylated amine utilisation could be independent from the habitat and localised factors or selection pressures 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.

398

399 **Conflict of Interest**

400 The authors declare no conflict of interest.

401

402 **Acknowledgements**

403 We thank Dr Felipe Hernandes Coutinho (Federal University of Rio de Janeiro) for a
404 fruitful discussion on genomic regions of insertion and deletion in *Gemmobacter sp.*
405 LW-1. The EDGAR platform is financially supported by the BMBF grant FKZ
406 031A533 within the de.NBI network. This manuscript has been published as preprint
407 at the preprint server for biology (bioRxiv.org, Kröber et al., 2018).

408

409

410

411

412

413

414

REFERENCES

415

416

417 1.

Ge X, Wexler AS, Clegg SL (2011) Atmospheric amines – Part I. A review.

418

Atmospheric Environment 45: 524-546. doi:

419

<http://dx.doi.org/10.1016/j.atmosenv.2010.10.012>

420

2.

Schade GW, Crutzen PJ (1995) Emission of Aliphatic-Amines from Animal

421

Husbandry and Their Reactions - Potential Source of N₂O and HCN. J

422

Atmos Chem 22: 319-346. doi: Doi 10.1007/Bf00696641

423

3.

Latypova E, Yang S, Wang YS, Wang T, Chavkin TA, Hackett M, Schäfer H,

424

Kalyuzhnaya MG (2010) Genetics of the glutamate - mediated

425

methylamine utilization pathway in the facultative methylotrophic beta -

426

proteobacterium *Methyloversatilis universalis* FAM5. Molecular

427

microbiology 75: 426-439.

428

4.

Carpenter LJ, Archer SD, Beale R (2012) Ocean-atmosphere trace gas

429

exchange. Chemical Society Reviews 41: 6473-6506. doi:

430

10.1039/c2cs35121h

431

5.

Lyimo TJ, Pol A, Jetten MSM, Op den Camp HJM (2009) Diversity of

432

methanogenic archaea in a mangrove sediment and isolation of a new

433

Methanococcoides strain. Fems Microbiol Lett 291: 247-253. doi:

434

10.1111/j.1574-6968.2008.01464.x

435

6.

Burke SA, Lo SL, Krzycki JA (1998) Clustered genes encoding the

436

methyltransferases of methanogenesis from monomethylamine. J

437

Bacteriol 180: 3432-3440.

438

7.

Liu Y, Whitman WB (2008) Metabolic, phylogenetic, and ecological

439

diversity of the methanogenic archaea. Ann N Y Acad Sci 1125: 171-189.

440

doi: 10.1196/annals.1419.019

441

8.

Chen Y, Scanlan J, Song L, Crombie A, Rahman MT, Schäfer H, Murrell JC

442

(2010) γ -Glutamylmethylamide is an essential intermediate in the

443

metabolism of methylamine by *Methylocella silvestris*. Applied and

444

environmental microbiology 76: 4530-4537.

445

9.

Chistoserdova L, Kalyuzhnaya MG, Lidstrom ME (2009) The Expanding

446

World of Methylotrophic Metabolism. Annual Review of Microbiology 63:

447

477-499. doi: 10.1146/annurev.micro.091208.073600

448

10.

Chistoserdova L (2011) Modularity of methylotrophy, revisited.

449

Environmental Microbiology 13: 2603-2622.

450

11.

Chen Y, Wu L, Boden R, Hillebrand A, Kumaresan D, Moussard H, Baci M,

451

Lu Y, Colin Murrell J (2009) Life without light: microbial diversity and

452

evidence of sulfur- and ammonium-based chemolithotrophy in Movile

453

Cave. ISME J 3: 1093-1104.

454

12.

Wischer D, Kumaresan D, Johnston A, El Khawand M, Stephenson J,

455

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. doi:
- 10.1038/ismej.2014.102
13. Anthony C (1982) The biochemistry of methylotrophs. Academic Press,
- London
14. Lidstrom ME (2006) Aerobic Methylotrophic Prokaryotes. In: Dworkin, M,
- Falkow, S, Rosenberg, E, Schleifer, K-H, Stackebrandt, E (eds.) The
- Prokaryotes: Volume 2: Ecophysiology and Biochemistry. Springer New
- York, New York, NY, pp. 618-634
15. Chen Y (2012) Comparative genomics of methylated amine utilization by
- marine *Roseobacter* clade bacteria and development of functional gene
- markers (tmm, gmaS). Environmental Microbiology 14: 2308-2322.
16. 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.
17. 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. J Bacteriol 182: 2520-2529.
18. Lidbury I, Mausz MA, Scanlan DJ, Chen Y (2017) Identification of
- dimethylamine monooxygenase in marine bacteria reveals a metabolic
- bottleneck in the methylated amine degradation pathway. ISME J 11:
- 1592-1601. doi: 10.1038/ismej.2017.31
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, Chistoserdova 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. doi: 10.1111/1462-
- 2920.13709
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*
- changelensis* 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. doi: 10.1099/ijs.0.042051-0

- 505 24. Kumaresan D, Wischer D, Stephenson J, Hillebrand-Voiculescu A, Murrell
506 JC (2014) Microbiology of Movile Cave - A chemolithoautotrophic
507 ecosystem. *Geomicrobiology Journal* 31: 186-193.
- 508 25. Kumaresan D, Wischer D, Hillebrand-Voiculescu AM, Murrell JC (2015)
509 Draft Genome Sequences of Facultative Methylophs, *Gemmobacter* sp.
510 Strain LW1 and *Mesorhizobium* sp. Strain 1M-11, Isolated from Movile
511 Cave, Romania. *Genome Announc* 3. doi: 10.1128/genomeA.01266-15
- 512 26. Zheng J-W, Chen Y-G, Zhang J, Ni Y-Y, Li W-J, He J, Li S-P (2011)
513 Description of *Catellibacterium caeni* sp. nov., reclassification of
514 *Rhodobacter changlensis* Anil Kumar et al. 2007 as *Catellibacterium*
515 *changelense* comb. nov. and emended description of the genus
516 *Catellibacterium*. *International Journal of Systematic and Evolutionary*
517 *Microbiology* 61: 1921-1926.
- 518 27. Liu J-J, Zhang X-Q, Chi F-T, Pan J, Sun C, Wu M (2014) *Gemmobacter*
519 *megaterium* sp. nov., isolated from coastal planktonic seaweeds.
520 *International Journal of Systematic and Evolutionary Microbiology* 64: 66-
521 71. doi: doi:10.1099/ijs.0.050955-0
- 522 28. Tanaka Y, Hanada S, Manome A, Tsuchida T, Kurane R, Nakamura K,
523 Kamagata Y (2004) *Catellibacterium nectarophilum* gen. nov., sp. nov.,
524 which requires a diffusible compound from a strain related to the genus
525 *Sphingomonas* for vigorous growth. *International Journal of Systematic*
526 *and Evolutionary Microbiology* 54: 955-959. doi:
527 doi:10.1099/ijs.0.02750-0
- 528 29. Rothe B, Fischer A, Hirsch P, Sittig M, Stackebrandt E (1987) The
529 phylogenetic position of the budding bacteria *Blastobacter aggregatus*
530 and *Gemmobacter aquatilis* gen., nov. sp. nov. *Archives of Microbiology*
531 147: 92-99. doi: 10.1007/bf00492911
- 532 30. Liu Y, Xu C-J, Jiang J-T, Liu Y-H, Song X-F, Li H, Liu Z-P (2010)
533 *Catellibacterium aquatile* sp. nov., isolated from fresh water, and
534 emended description of the genus *Catellibacterium* Tanaka et al. 2004.
535 *International Journal of Systematic and Evolutionary Microbiology* 60:
536 2027-2031. doi: doi:10.1099/ijs.0.017632-0
- 537 31. Zhang J, Chen S-A, Zheng J-W, Cai S, Hang B-J, He J, Li S-P (2012)
538 *Catellibacterium nanjingense* sp. nov., a propanil-degrading bacterium
539 isolated from activated sludge, and emended description of the genus
540 *Catellibacterium*. *International Journal of Systematic and Evolutionary*
541 *Microbiology* 62: 495-499. doi: doi:10.1099/ijs.0.029819-0
- 542 32. Kämpfer P, Jerzak L, Wilharm G, Golke J, Busse H-J, Glaeser SP (2015)
543 *Gemmobacter intermedius* sp. nov., isolated from a white stork (*Ciconia*
544 *ciconia*). *International Journal of Systematic and Evolutionary*
545 *Microbiology* 65: 778-783. doi: doi:10.1099/ijs.0.000012
- 546 33. Sheu S-Y, Shiau Y-W, Wei Y-T, Chen W-M (2013) *Gemmobacter lanyuensis*
547 sp. nov., isolated from a freshwater spring. *International Journal of*
548 *Systematic and Evolutionary Microbiology* 63: 4039-4045. doi:
549 doi:10.1099/ijs.0.052399-0
- 550 34. Sheu S-Y, Sheu D-S, Sheu F-S, Chen W-M (2013) *Gemmobacter tilapia* sp.
551 nov., a poly- β -hydroxybutyrate-accumulating bacterium isolated from a
552 freshwater pond. *International Journal of Systematic and Evolutionary*
553 *Microbiology* 63: 1550-1556. doi: doi:10.1099/ijs.0.044735-0

- 554 35. Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Pillay M, Ratner
555 A, Huang J, Woyke T, Huntemann M (2013) IMG 4 version of the
556 integrated microbial genomes comparative analysis system. *Nucleic acids*
557 *research* 42: D560-D567.
- 558 36. Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW
559 (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA
560 genes. *Nucleic acids research* 35: 3100-3108.
- 561 37. Pruesse E, Peplies J, Glöckner FO (2012) SINA: accurate high-throughput
562 multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 28:
563 1823-1829.
- 564 38. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO
565 (2007) SILVA: a comprehensive online resource for quality checked and
566 aligned ribosomal RNA sequence data compatible with ARB. *Nucleic acids*
567 *research* 35: 7188-7196.
- 568 39. Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary
569 Genetics Analysis version 7.0 for bigger datasets. *Molecular biology and*
570 *evolution* 33: 1870-1874.
- 571 40. Saitou N, Nei M (1987) The neighbor-joining method: a new method for
572 reconstructing phylogenetic trees. *Molecular biology and evolution* 4:
573 406-425.
- 574 41. Felsenstein J (1985) Confidence limits on phylogenies: an approach using
575 the bootstrap. *Evolution* 39: 783-791.
- 576 42. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local
577 alignment search tool. *Journal of molecular biology* 215: 403-410.
- 578 43. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the
579 sensitivity of progressive multiple sequence alignment through sequence
580 weighting, position-specific gap penalties and weight matrix choice.
581 *Nucleic acids research* 22: 4673-4680.
- 582 44. Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of
583 mutation data matrices from protein sequences. *Bioinformatics* 8: 275-
584 282.
- 585 45. Emms DM, Kelly S (2015) OrthoFinder: solving fundamental biases in
586 whole genome comparisons dramatically improves orthogroup inference
587 accuracy. *Genome biology* 16: 157. doi: 10.1186/s13059-015-0721-2
- 588 46. Nakamura T, Yamada KD, Tomii K, Katoh K (2018) Parallelization of
589 MAFFT for large-scale multiple sequence alignments. *Bioinformatics* 34:
590 2490-2492. doi: 10.1093/bioinformatics/bty121
- 591 47. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool
592 for automated alignment trimming in large-scale phylogenetic analyses.
593 *Bioinformatics* 25: 1972-1973.
- 594 48. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis
595 and post-analysis of large phylogenies. *Bioinformatics* 30: 1312-1313.
- 596 49. Matias Rodrigues JF, Schmidt TSB, Tackmann J, von Mering C (2017)
597 MAPseq: highly efficient k-mer search with confidence estimates, for
598 rRNA sequence analysis. *Bioinformatics* 33: 3808-3810. doi:
599 10.1093/bioinformatics/btx517
- 600 50. Grant JR, Arantes AS, Stothard P (2012) Comparing thousands of circular
601 genomes using the CGView Comparison Tool. *BMC genomics* 13: 202.

- 602 51. Rodriguez-R LM, Konstantinidis KT (2016) The enveomics collection: a
603 toolbox for specialized analyses of microbial genomes and metagenomes.
604 PeerJ Preprints.
- 605 52. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje
606 JM (2007) DNA-DNA hybridization values and their relationship to
607 whole-genome sequence similarities. *International journal of systematic
608 and evolutionary microbiology* 57: 81-91.
- 609 53. Konstantinidis KT, Tiedje JM (2005) Genomic insights that advance the
610 species definition for prokaryotes. *P Natl Acad Sci USA* 102: 2567-2572.
611 doi: 10.1073/pnas.0409727102
- 612 54. Blom J, Kreis J, Spanig S, Juhre T, Bertelli C, Ernst C, Goesmann A (2016)
613 EDGAR 2.0: an enhanced software platform for comparative gene content
614 analyses. *Nucleic Acids Research* 44: W22-W28. doi:
615 10.1093/nar/gkw255
- 616 55. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW (2015)
617 CheckM: assessing the quality of microbial genomes recovered from
618 isolates, single cells, and metagenomes. *Genome research*: gr.
619 186072.186114.
- 620 56. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K,
621 Gerdes S, Glass EM, Kubal M (2008) The RAST Server: rapid annotations
622 using subsystems technology. *BMC genomics* 9: 75.
- 623 57. Nayak DD, Marx CJ (2015) Experimental horizontal gene transfer of
624 methylamine dehydrogenase mimics prevalent exchange in nature and
625 overcomes the methylamine growth constraints posed by the sub-optimal
626 N-Methylglutamate pathway. *Microorganisms* 3: 60-79.
- 627 58. Richter M, Rossello-Mora R (2009) Shifting the genomic gold standard for
628 the prokaryotic species definition. *Proceedings of the National Academy
629 of Sciences of the United States of America* 106: 19126-19131.
- 630 59. Kim M, Oh H-S, Park S-C, Chun J (2014) Towards a taxonomic coherence
631 between average nucleotide identity and 16S gene sequence similarity for
632 species demarcation of prokaryotes. *International Journal of Systematic
633 and Evolutionary Microbiology* 64: 346-351.
- 634 60. Nayak DD, Agashe D, Lee M-C, Marx CJ (2016) Selection maintains
635 apparently degenerate metabolic pathways due to tradeoffs in using
636 methylamine for carbon versus nitrogen. *Current Biology* 26: 1-11.
- 637 61. Kumaresan D, Stephenson J, Doxey AC, Bandukwala H, Brooks E,
638 Hillebrand-Voiculescu A, Whiteley AS, Murrell JC (2018) Aerobic
639 proteobacterial methylotrophs in Movile Cave: genomic and metagenomic
640 analyses. *Microbiome* 6. doi: ARTN 1
641 10.1186/s40168-017-0383-2
642
643

644 **Figures and Tables**

645
646 **Figure 1. Phylogenetic tree based on 16S rRNA gene sequences.** The tree was
647 constructed using the maximum likelihood method for clustering and the Tamura-Nei

648 model for computing evolutionary distances. Numbers at branches are bootstrap
649 percentages >50% of 1000 replicates. Star represents the *Gemmobacter* species used
650 for comparative genome analysis. Coloured boxes represent the habitat where the
651 sequence was retrieved: blue (fresh water), orange (soil and sediment), green
652 (activated sludge), grey (marine), purple (clinical source). Triangles represent
653 sequences that are listed as *Catellibacterium* in the NCBI database, which have been
654 recently reclassified to *Gemmobacter* [23]. Scale bar: 0.02 substitutions per
655 nucleotide position.

656

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

664

665 **Figure 3. DNA BLAST map of *Gemmobacter* genomes.** *Gemmobacter* sp. LW-1
666 was used as a reference genome against *Gemmobacter megaterium* (inner ring),
667 *Gemmobacter aquatilis* (second inner ring), *Gemmobacter nectarophilus* (third ring),
668 and *Gemmobacter caeni* (fourth ring). The fifth and sixth ring (outer rings) represent
669 the CDS (blue), tRNA (maroon), and rRNA (purple) on the reverse and forward
670 strand, respectively. The colour scale (inset) shows the level of sequence identity with
671 the respective sequences from *G. megaterium*, *G. aquatilis*, *G. nectarophilus* and *G.*

672 *caeni*. The locations of genes involved in methylotrophy are indicated at the outside
673 of the map.

674

675 **Figure 4. Pan-genome analysis. (A)** Pan-genome tree consisting of five
676 *Gemmobacter* species was constructed using the neighbour-joining method within the
677 EDGAR platform. **(B)** Number of core, dispensable, and specific genes (singletons)
678 of each *Gemmobacter* species.

679

680 **Figure 5. Metabolic pathways involved in methylated amine utilisation and one-**
681 **carbon utilisation annotated with presence/absence of specific genes in the**
682 **genomes of *Gemmobacter*.** The analysis was based on a five-way comparison among
683 *Gemmobacter* sp. LW-1 (L), *Gemmobacter caeni* (C), *Gemmobacter aquatilis* (A),
684 *Gemmobacter nectarophilus* (N) and *Gemmobacter megaterium* (M). The color-coded
685 boxes next to the genes indicate the presence (green) or absence (orange) of a gene in
686 each genome.

687

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

692

693 **Supplementary Figure S1.** Maximum-likelihood phylogenetic tree (JTT matrix-
694 based) of *mauA* sequences. Sequences from the genus *Gemmobacter* are marked with
695 a star. Amino acid sequences (MauA) were aligned using the ClustalW algorithm.
696 Numbers at branches are bootstrap percentages >50% of 1000 replicates. Scale bar:

697 0.1 substitutions per amino acid. Coloured boxes indicate *Alphaproteobacteria*
698 (yellow), *Gammaproteobacteria* (red) and *Betaproteobacteria* (black).

699

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

704

705 **Supplementary Figure S3.** Distribution of *Gemmobacter* in different ecosystems

706

707 **Supplementary Figure S4.** Maximum-likelihood phylogenetic tree of concatenated
708 single copy marker genes from genomes of *Gemmobacter* and closely related
709 organisms.

710

711 **Supplementary Figure S5.** Arrangement of genes involved in methylated amine
712 utilisation

713

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

716

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

719

720 **Supplementary Table S3.** Average nucleotide identity (ANI) values between
721 *Gemmobacter* sp. LW-1 and *Gemmobacter caeni*, *Gemmobacter aquatilis*,

722 *Gemmobacter nectarophilus*, *Gemmobacter megaterium*, *Rhodobacter sphaeroides*

723 and *Paracoccus denitrificans*

724

725

Figure 1

0.020

Figure 2

Figure 3

Figure 4A

Figure 4B

Figure 5

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

Organism	<i>gmaS</i>	<i>tmm</i>	<i>xoxF</i>	<i>fae</i>	<i>foldD</i>	<i>purU</i>	<i>fhs</i> (<i>ftfL</i>)	<i>gfa</i>	<i>adhI</i> (<i>flhA</i>)	<i>fghA</i>	<i>fdh</i>	<i>amtB</i>	<i>mtdA</i> / <i>mtdB</i>	<i>mchA</i>	<i>ftt</i>	<i>fmdA</i> /B/C	<i>mauA</i>	<i>cox</i>
<i>Citricella</i> sp. SE45	+	+	+	-	+	+	+	+	+	+	+	+	-/-	-	-	-	-	+
<i>Roseovarius</i> sp. TM1035	+	+	+	-	+	-	+	-	+	+	+	+	-/-	-	-	-	-	+
<i>Paracoccus denitrificans</i> PD1222	-	-	+	-	+	+	+	+	+	+	+	+	-/-	-	-	-	-	-
<i>Rhodobacter sphaeroides</i> 241	-	+	+	-	+	+	+	+	+	+	+	+	-/-	-	-	-	-	+
<i>Gemmobacter caeni</i>	+	+	+	-	+	+	+	+ ^a	+	+	+	+	-/-	-	-	-	+	+
<i>Gemmobacter</i> sp. LW-1	+	+	+	-	+	+	+	+ ^a	+	+	+	+	-/-	-	-	-	+	+
<i>Gemmobacter aquatilis</i>	+	+	+	-	+	+	+	-	+	+	+	+	-/-	-	-	-	-	+
<i>Gemmobacter nectarophilus</i> DSM-15620	-	-	-	-	+	+	+	+ ^a	+	+	+	+	-/-	-	-	-	-	+
<i>Gemmobacter megaterium</i> DSM-26375	-	-	-	-	+	+	+	+ ^a	+	+	+	+	-/-	-	-	-	-	+

^a *gfa*-like protein-encoding gene