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

Kröber, E.[1], Spurgin L.[2], Wischer, D., Kumaresan, D.[3#]

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

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

[4] 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, 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 thought to be a major sink. Recently we isolated a facultative methylotroph, *Gemmobacter* sp. LW-1, an isolate from the unique environment Movile Cave, Romania, which is capable of methylated amine utilisation as a carbon source. Here, using a comparative genomics approach, we investigate how widespread methylated amine utilisation trait is within the member 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 the 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 *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.

Introduction

Methylated amines (MAs) are ubiquitous in the environment with a variety of natural and anthropogenic sources including the oceans, vegetation, sediments and organic-rich soils, animal husbandry, food industry, pesticides, sewage, and automobiles, to mention only a few (Schade and Crutzen, 1995;Latypova et al., 2010;Ge et al., 2011).

61 Methylated amines are also known to influence earth's climate, via a series of
62 complex biological and chemical interactions (Carpenter et al., 2012). Some of the
63 most abundant methylated amines found in the atmosphere are trimethylamine
64 (TMA), dimethylamine (DMA) and monomethylamine (MMA) (Ge et al., 2011).
65 Microbial metabolism of methylated amines involves both aerobic and anaerobic
66 microorganisms, e.g. some methanogenic archaea such as *Methanosarcina* and
67 *Methanomicrobium* can use MAs to produce methane (Burke et al., 1998;Liu and
68 Whitman, 2008;Lyimo et al., 2009) while Gram-positive and Gram-negative
69 methylotrophic bacteria can use MAs as carbon and nitrogen source (Chen et al.,
70 2010b). Previously, MAs were typically associated with marine ecosystems as they
71 are by-products of degradation of osmolytic chemicals such as glycine betaine,
72 carnitine, choline and trimethylamine N-oxide (Chen et al., 2010b). However, recent
73 studies have reported the detection and activity of aerobic methylotrophic bacteria
74 that utilise MAs in a variety of natural and engineered environments (Chen et al.,
75 2009;Chistoserdova et al., 2009;Chistoserdova, 2011;Ge et al., 2011;Wischer et al.,
76 2015) and could play a major role in global C and N budgets.
77
78 Aerobic methylotrophs are a polyphyletic group of microorganisms capable of
79 utilising one-carbon (C_1) compounds such as methane, methanol or methylated
80 amines as their sole source of carbon and energy (Anthony, 1982;Lidstrom,
81 2006;Chistoserdova et al., 2009). Methylotrophs can degrade TMA to DMA by using
82 the enzymes TMA dehydrogenase, TMA monooxygenase or TMA methyltransferase,
83 encoded by the genes *tdm*, *tmm* and *mtt*, respectively (Paul et al., 2000;Chen,
84 2012;Lidbury et al., 2014). The enzymes DMA dehydrogenase (*dmd*) or DMA
85 monooxygenase (*dmmDABC*) modulate the conversion of DMA to MMA (Lidstrom,

86 2006;Chen, 2012). Two distinct pathways have been characterised for the oxidation of
87 MMA (Chistoserdova, 2011). The direct MMA-oxidation pathway mediated by a
88 single enzyme (MMA dehydrogenase in Gram-negative bacteria and MMA oxidase in
89 Gram-positive bacteria) converts MMA to formaldehyde and releases ammonium
90 (McIntire et al., 1991;Chistoserdova et al., 1994). The alternate pathway, referred to
91 as the *N*-methylglutamate (NMG) pathway or indirect MMA-oxidation pathway, is
92 mediated by three individual enzymes via the oxidation of MMA to gamma-
93 glutamylmethylamide (GMA) and its further degradation to *N*-methylglutamate
94 (NMG) and 5,10-methylenetetrahydrofolate (CH_2 = H_4F) (Latypova et al.,
95 2010;Chistoserdova, 2011). A stepwise conversion of MMA in the NMG pathway is
96 modulated by the enzymes GMA synthetase (*gmaS*), NMG synthase (*mgsABC*) and
97 NMG dehydrogenase (*mgdABCD*) (Chen et al., 2010b;Latypova et al., 2010). The
98 capability to use MMA not only as a source for carbon but also for nitrogen is
99 widespread in bacteria. Notably, the NMG pathway is not only restricted to
100 methylotrophs but also present in non-methylotrophic bacteria that use MMA as a
101 nitrogen but not as a carbon source (Chen et al., 2010a;Chen, 2012;Taubert et al.,
102 2017).
103
104 In a recent study, we isolated an alphaproteobacterial facultative methylotrophic
105 bacterium, *Gemmobacter* sp. LW-1 (recently renamed from *Catellibacterium* (Chen et
106 al., 2013)) from the Movile Cave ecosystem (Kumaresan et al., 2014) (Mangalia,
107 Romania) that can use methylated amines as both carbon and nitrogen source
108 (Wischer et al., 2015) and subsequently obtained its genome sequence (Kumaresan et
109 al., 2015). Using a ^{13}C-MMA DNA based stable-isotope probing (SIP) experiment we
110 also showed that *Gemmobacter* sp. LW-1 was indeed an active MMA utiliser in

111 microbial mats from this environment (Wischer et al., 2015). This was the first report
112 of methylated amine utilisation in a member of the bacterial genus *Gemmobacter*.
113 However, growth on C_1 compounds (methanol and formate) has been reported for the
114 genus *Gemmobacter*, e.g. in *G. caeni* (Zheng et al., 2011).
115
116 The genus *Gemmobacter* (family *Rhodobacteraceae*) currently comprises ten
117 validated species: *Gemmobacter megaterium* (Liu et al., 2014), *G. nectariphilum*
118 (Tanaka et al., 2004;Chen et al., 2013), *G. aquatilis* (Rothe et al., 1987), *G. caeni*
119 (Zheng et al., 2011;Chen et al., 2013), *G. aquaticus* (Liu et al., 2010;Chen et al.,
120 2013), *G. nanjingense* (Zhang et al., 2012;Chen et al., 2013), *G. intermedius*
121 (Kämpfer et al., 2015), *G. lanyuensis* (Sheu et al., 2013b), *G. tilapiae* (Sheu et al.,
122 2013a) and *G. fontiphilus* (Chen et al., 2013). These species were isolated from a
123 wide range of environments including fresh water environments (freshwater pond
124 (Rothe et al., 1987;Sheu et al., 2013a), freshwater spring (Chen et al., 2013;Sheu et
125 al., 2013b)), coastal planktonic seaweed (Liu et al., 2014), white stork nestling
126 (Kämpfer et al., 2015), waste water and activated sludge (Tanaka et al., 2004;Zheng
127 et al., 2011;Zhang et al., 2012), suggesting that members of the genus *Gemmobacter*
128 are widely distributed in engineered and natural environments.
129
130 Here, using a comparative genomics approach we study how widespread methylated
131 amine utilisation trait is within the members of the genus *Gemmobacter*. We used five
132 isolate genomes for members within the genus *Gemmobacter* (*G*. sp. LW-1, *G. caeni*,
133 *G. aquatilis*, *G. nectariphilus* and *G. megaterium)* available from public repositories
134 (Accessed June 2018). These genomes were used for comparative genomics and

135 ancestral state reconstruction analyses to understand patterns of gain/loss of
136 methylated amine utilisation across the *Gemmobacter* phylogeny.
137
138 **Materials and Methods**
139 *Genome data acquisition*
140 Five *Gemmobacter* genomes (*G. caeni*, *G. aquatilis*, *G. nectariphilus*, *G. megaterium*,
141 *Gemmobacter* sp. LW-1) available through the Integrated Microbial Genomes (IMG)
142 database (https://img.jgi.doe.gov/) were used for comparative genome analysis
143 (Markowitz et al., 2013). Accession numbers and genome characteristics are listed in
144 Supplementary Table S1.
145
146 *Phylogenetic analysis*
147 Phylogenetic relatedness between the different members of the genus *Gemmobacter*
148 was determined using phylogenetic trees constructed from 16S rRNA gene sequences
149 (nucleotide) and metabolic gene sequences (*gmaS* and *mauA*; amino acids) involved
150 in MMA utilisation. RNAmmer (Lagesen et al., 2007) was used to retrieve 16S rRNA
151 gene sequences from the genome sequences. Multiple sequence alignment of 16S
152 rRNA gene sequences was performed using the SINA alignment service via SILVA
153 (Pruesse et al., 2007;Pruesse et al., 2012) and subsequently imported into MEGA7
154 (Kumar et al., 2016) to construct a maximum-likelihood nucleotide-based
155 phylogenetic tree (Saitou and Nei, 1987). Bootstrap analysis was performed with
156 1000 replicates to provide confidence estimates for phylogenetic tree topologies
157 (Felsenstein, 1985).
158

159 To determine phylogenetic affiliations for the protein encoding genes *gmaS* and
160 *mauA*, gene sequences retrieved from the genome sequences were aligned to
161 homologous sequences retrieved from the NCBI Genbank database using Basic Local
162 Alignment Search Tool (BLAST, blastx) (Altschul et al., 1990) and curated *gmaS*
163 sequences used for primer design in our previous study (Wischer et al., 2015). Amino
164 acid sequences were aligned in MEGA7 (Kumar et al., 2016) using ClustalW
165 (Thompson et al., 1994) and the alignment was subsequently used to construct
166 maximum likelihood phylogenetic trees based on the JTT matrix-based model (Jones
167 et al., 1992). Bootstrap analysis was performed with 1000 replicates to provide
168 confidence estimates for phylogenetic tree topologies (Felsenstein, 1985).
169
170 *Comparative genomic analyses*
171 CGView Comparison Tool (CCT) was used to visually compare the genomes within
172 the genus *Gemmobacter* (Grant et al., 2012). CCT utilises BLAST to compare the
173 genomes and the BLAST results are presented in a DNA-based graphical map (Grant
174 et al., 2012). Average Nucleotide Identity (ANI) (Rodriguez-R and Konstantinidis,
175 2016) between different genomes was estimated using one-way ANI (best hit) and
176 two-way ANI (reciprocal best hit) based on Goris *et al*. (Goris et al., 2007). In
177 addition the whole-genome based average nucleotide identity (gANI) and the p_r^{intra-}
178 ^{species} value were determined for *G*. sp. LW-1 and *G. caeni* (these two genomes
179 revealed the closest ANI) based on Konstantinidis and Tiedje (Konstantinidis and
180 Tiedje, 2005) via the Joint Genome Institute (JGI) platform (https://ani.jgi-
181 psf.org/html/home.php; Version 0.3, April 2014). In order to determine if two
182 genomes belong to the same species, the computation of empirical probabilities (p_r^{intra-}
183 ^{species}) can be calculated as follows,

184 $$p_r^{intra-species}[AF = a, ANI = b] = p_r^{intra-species}[AF = a] * p_r^{intra-species}[ANI = b|AF = a]$$

185 AF represents alignment fraction.

186

187 Pan-genome analysis including average amino acid identity (AAI) analysis, pan-
188 genome tree construction and determination of core and dispensable genes and
189 singletons (unique genes) was carried out using the Efficient Database framework for
190 comparative Genome Analyses using BLAST score Ratios (EDGAR) platform (Blom
191 et al., 2016).

192

193 In order to compare the genetic potential for methylated amine utilisation within the
194 available *Gemmobacter* genomes, known protein sequences involved in methylated
195 amine utilisation pathways (Latypova et al., 2010;Chen, 2012) were used as query
196 sequences through the BLAST (blastp) program (Altschul et al., 1990) available
197 within the Rapid Annotation using Subsystem Technology (RAST) server (Aziz et al.,
198 2008). The list of protein queries used is given in Supplementary Table S2.

199

200 We examined patterns of gain/loss of methylated amine utilisation along the above
201 16S rRNA gene phylogeny, by performing ancestral state reconstruction analysis
202 using the phytools package in R (Revell, 2012). We used stochastic character
203 mapping (Nielsen, 2002) to map presence/absence of methylated amine utilisation,
204 assigning a prior probability of one to species known to utilise methylated amines,
205 zero to those known not to do so, and 0.5 to those where the trait value was unknown.
206 We used MCMC (with the function *make.simmap)* to simulate 1000 stochastic maps,
207 and then we used the function *densityMap* (Revell, 2013) to visualise the aggregate
208 result from the stochastic mapping analysis.

209

210 **Results and discussion**

211 ***Phylogenetic relatedness based on the 16S rRNA and metabolic gene***

212 The phylogenetic relatedness of the five members within the genus *Gemmobacter* (*G.*
213 sp. LW-1, *G. caeni*, *G. aquatilis*, *G. nectariphilus* and *G. megaterium*) was resolved
214 based on 16S rRNA gene sequences (Figure 1). Three members of the genus
215 *Gemmobacter* (*G.* sp. LW-1, *G. caeni*, and *G. aquatilis)* clustered together with
216 several other related *Gemmobacter* and *Rhodobacter* 16S rRNA gene sequences
217 retrieved from fresh water, soil and sediment and activated sludge environments
218 (Figure 1). *G. nectariphilus* and *G. megaterium* sequences clustered together with
219 *Paracoccus kawasakiensis* and other related *Gemmobacter* sequences from marine,
220 fresh water and activated sludge environments (Figure 1). Based on the 16S rRNA
221 gene sequences retrieved from public database, we observed that the members of the
222 genus *Gemmobacter* are widely distributed in engineered (such as activated sludge
223 and clinical environments) and natural environments i.e. fresh water, soil and
224 sediment, and marine environments (Figure 1). GMA synthetase, a key enzyme in the
225 NMG pathway, is encoded by the gene *gmaS*. *gmaS* sequences retrieved from the
226 isolate genomes along with other ratified *gmaS* sequences were used to construct an
227 amino acid-based phylogenetic tree (Figure 2). *gmaS* gene sequences retrieved from
228 genomes of *G.* sp. LW-1, *G. caeni* and *G. aquatilis* clustered within Group I of
229 alphaproteobacterial *gmaS* sequences containing sequences from marine and non-
230 marine bacteria within the orders *Rhodobacterales* and *Rhizobiales* as described in
231 Wischer et al. (Wischer et al., 2015) and were closely related to *Paracoccus yeei*, *P.*
232 sp. 1W-5 and *Rhodobacter* sp. 1W-5 (Figure 2).

233

234 Whilst *gmaS* gene sequences were detected in three of the five investigated
235 *Gemmobacter* genomes, *mauA* gene sequences were identified only in the genomes of
236 *G. caeni* and *G.* sp. LW-1 (Supplementary Figure S1). It has been suggested that the
237 NMG pathway for MMA utilisation is more universally distributed and more
238 abundant across proteobacterial methylotrophs than the direct MMA oxidation
239 pathway (Nayak and Marx, 2015). However, it should be noted that genes encoding
240 for the enzymes within the NMG pathway (*gmaS*) can not only be detected in
241 methylotrophs but also in non-methylotrophic bacteria that use MMA as a nitrogen
242 source, but not as a carbon source (Chen, 2012;Wischer et al., 2015).
243
244 ***A comparative genome analysis of members within the genus Gemmobacter***
245 At the time of the analysis, five *Gemmobacter* genomes obtained from isolates from
246 different environments were available (Figure 1 and Table 2). *Gemmobacter* genome
247 sizes range from ~3.96 Mb to ~5.14 Mb with GC contents between 64.71% to 66.19%
248 (Table 2). Analysis of sequence annotations revealed that on average 91.13% of the
249 genomes consist of coding sequences.
250
251 The genomes were compared using the CGView comparison tool (Grant et al., 2012)
252 (Figure 3). *Gemmobacter* sp. LW-1, isolated from the Movile Cave ecosystem was
253 used as the reference genome and the results of the BLAST comparison with other
254 *Gemmobacter* genomes are represented as a BLAST ring for each genome (Figure 3).
255 Similarities between segments of the reference genome sequence and the other
256 genome sequences are shown by a coloured arc beneath the region of similarity
257 indicating the percentage of similarity as a colour code. Our analysis (Figure 3)
258 revealed low identity levels (mostly <88%) between *Gemmobacter* sp. LW-1 and *G.*

259 *aquatilis*, *G. nectariphilus* and *G. megaterium* across the genomes. Moreover, the
260 analysis suggested several sites of potential insertion/deletion events in the genome of
261 *Gemmobacter* sp. LW-1. Possible insertion/deletion regions can be identified as those
262 gaps in the map where no homology is detected. For example the region between
263 2200-2300 kbp (Figure 3) where a gap can be found in the otherwise contiguous
264 homologous regions between the reference genome *G.* sp. LW-1 and the first of the
265 query genomes (*G. caeni*). This might likely be due to a lack of hits or hits with low
266 identity that can be spurious matches. Since it covers a large region we could possibly
267 rule out that it is not an artefact arising from a lack of sensitivity in the BLAST
268 analysis. Even though the genomes of *G.* sp. LW-1 and *G. caeni* are closely related,
269 our analysis demonstrates that their genomes are not completely identical. Despite the
270 fact that the majority of their genomes indicate very high identity levels (mostly >96-
271 98% as shown by the dominance of dark red colours of the circle representing the
272 BLAST hit identity between *G.* sp. LW-1 and *G. caeni*), many segments appear to be
273 exclusive to *G.* sp. LW-1.
274
275 In order to further resolve the similarity between these genomes we calculated the
276 average nucleotide identity (ANI) (Rodriguez-R and Konstantinidis, 2016)
277 (Supplementary Table S3 and Supplementary Figure S2A-D) and the average amino
278 acid identity,(AAI; Supplementary Figure S2E). It is generally accepted that an ANI
279 value of >95-96% can be used for species delineation (Richter and Rossello-Mora,
280 2009;Kim et al., 2014). Our analysis revealed that *Gemmobacter* sp. LW-1 and
281 *Gemmobacter caeni* share an ANI value of 98.62 (Supplementary Table S3) implying
282 that both are in fact the same species. The genome-based average nucleotide identity
283 (gANI) between *G.* sp. LW-1 and *G. caeni* was calculated as 98.70. The AF was

284 calculated to be 0.91, which would result in a computed probability of 0.98
285 suggesting that both genomes might belong to the same species. However, it should
286 be noted that these are draft genomes and a more in depth characterization of their
287 physiology and phenotype is required to delineate these organisms at the level of
288 strain.
289
290 Pan-genome analysis, carried out using the EDGAR platform (Blom et al., 2016),
291 identified metabolic genes present in all *Gemmobacter* species (core genes), two or
292 more *Gemmobacter* species (accessory or dispensable genes), and unique
293 *Gemmobacter* species (singleton genes). A pan-genome tree was constructed (Figure
294 4A) based on the pan-genome dataset and neighbor-joining method (Saitou and Nei,
295 1987). As with the 16S-rRNA gene based phylogenetic tree (Figure 1), the five
296 *Gemmobacter* species formed two main clusters in the pan-genome tree analysis
297 (Figure 4A). The pan-genome tree also confirmed the phylogenetic closeness between
298 *Gemmobacter caeni* and *Gemmobacter* sp. LW-1 (Figure 4A). According to pan-
299 genome analysis of the five *Gemmobacter* genomes, a total of 9,286 genes were
300 identified, consisting of 1,806 core genes, 3085 dispensable genes and 305, 1,072,
301 896, 1,165 and 957 singletons for *G.* sp. LW-1, *G. caeni*, *G. aquatilis*, *G.*
302 *nectariphilus* and *G. megaterium*, respectively (Figure 4B).
303
304 ***Methylated amine utilisation, N assimilation and C_1 oxidation***
305 Investigation of the methylated amine utilisation pathways in five *Gemmobacter*
306 species revealed the presence of the genes encoding enzymes TMA dehydrogenase
307 (*tmd*), TMA monooxygenase (*tmm*), TMAO demethylase (*tdm*) and DMA
308 monooxygenase in genomes of *G.* sp. LW-1, *G. caeni* and *G. aquatilis* while none of

309 these genes were detected in *G. nectariphilus* or *G. megaterium* (Figure 5). These
310 findings are supported by results from a previous study which showed growth of *G.*
311 sp. LW-1 on TMA as a carbon and nitrogen source (Wischer et al., 2015). *G.* sp. LW-
312 1, *G. caeni* and *G. aquatilis* could potentially use the TMA oxidation pathway to
313 convert TMA to DMA. Based on the genome sequences, it can be suggested that
314 these three species could use the enzyme DMA monooxygenase (*dmmDABC*) to
315 oxidize DMA to MMA but not the DMA dehydrogenase since the corresponding
316 protein encoding gene (*dmd*) was not found (Figure 5).
317
318 We also compared the distribution of the direct MMA-oxidation and the NMG
319 pathways in the genomes of five *Gemmobacter* species (Figure 5). The direct MMA-
320 oxidation pathway (*mauA*-dependent) is so far only known to be present in
321 methylotrophic bacteria that can use MMA as a carbon source. Whereas the NMG
322 pathway (*gmaS*-dependent) has been shown to be present in non-methylotrophic
323 bacteria that can use MMA as a nitrogen source (Chen et al., 2010a;Nayak and Marx,
324 2015;Wischer et al., 2015;Nayak et al., 2016). Analysis of the genome sequences
325 revealed that both *G.* sp. LW-1 and *G. caeni* possess genes for both MMA oxidation
326 pathways (Figure 5). We have previously shown that *Gemmobacter* sp. LW-1 can use
327 MMA and TMA as both a carbon and nitrogen source (Wischer et al., 2015). Genome
328 sequence of *G. aquatilis* indicated the presence of genes involved only in the NMG
329 pathway. In the facultative methylotroph *Methylobacterium extorquens* AM1 it has
330 been shown that the NMG pathway is advantageous compared to the direct MMA-
331 oxidation pathway (Nayak et al., 2016). NMG pathway enables facultative
332 methylotrophic bacteria to switch between using MMA as a nitrogen source or as a
333 carbon and energy source whereas the direct MMA oxidation pathway allows for

334 rapid growth on MMA only as the primary energy and carbon source (Nayak et al.,
335 2016). This could suggest that *G. aquatilis* might use the NMG pathway for utilising
336 MMA as both nitrogen and carbon source. However, growth assays are required to
337 confirm whether *G. aquatilis* can use MMA as a carbon source. We did not detect
338 genes for either MMA oxidation pathways in the genome sequences of *G.*
339 *nectariphilus* and *G. megaterium* suggesting the lack of genetic potential of these
340 organisms to use MMA as either C or N source.
341
342 The C_1 units derived from methylated amines need to be further oxidized when the
343 nitrogen is sequestered without assimilation of the carbon from the methylated
344 amines. Genome analysis confirmed that all five *Gemmobacter* species possess the
345 genetic capability for C_1 oxidation and also indicate that tetrahydrofolate (H_4F) is the
346 C_1 carrier (Figure 5). The bifunctional enzyme 5,10-methylene-tetrahydrofolate
347 dehydrogenase/ cyclohydrolase, encoded by the gene *folD*, was detected in all the
348 *Gemmobacter* genomes (Figure 5/ Table 1). Genes encoding key enzymes in the C_1
349 oxidation pathway via tetrahydromethanopterin (H_4MPT) were not detected.
350 (Chistoserdova, 2011). The formate-tetrahydrofolate ligase, encoded by the gene *fhs*
351 (Figure 5), provides C_1 units for biosynthetic pathways (Chen, 2012). However, the
352 oxidation of formyl-H_4F (CHO-H_4F) can also be facilitated by *purU*, the gene
353 encoding for the formyl-H_4F deformylase. The formate dehydrogenase (*fdh*) mediates
354 the last step of the C_1 oxidation pathway, the oxidation of formate to CO_2. The genes
355 for the C_1 oxidation pathway via H_4F were detected in all five *Gemmobacter*
356 genomes.
357

358 The *fae* gene, encoding the formaldehyde-activating enzyme that catalyses the
359 reduction of formaldehyde with H_4MPT was not detected in any of the five
360 *Gemmobacter* genomes confirming that these members of the genus *Gemmobacter*
361 lack the H_4MPT pathway for formaldehyde oxidation (Figure 5/ Table 1).
362 Investigation of the nitrogen assimilation pathway revealed the presence of the genes
363 encoding glutamine synthetase (GS; *gluL*) and glutamine synthase (GOGAT; *glxB*) in
364 all five *Gemmobacter* genomes. In bacteria this pathway is essential for glutamate
365 synthesis at low ammonium concentrations (Chen, 2012).
366
367 Using comparative genome analysis we provide genome-based evidence that the two
368 *Gemmobacter* isolates *G.* sp. LW-1 and *G. caeni* are capable of generating energy
369 from complete oxidation of methylated amines via the H_4F-dependent pathway using
370 either the NMG pathway or the direct MMA oxidation pathway. *Gemmobacter*
371 *aquatilis* is genetically capable of methylated amine degradation to yield
372 formaldehyde and only encodes the genes for the NMG pathway, which indicates that
373 *G. aquatilis* could use this pathway to use MMA as a nitrogen source. Both *G.*
374 *nectariphilus* and *G. megaterium* genomes indicate the lack of potential to use
375 methylated amines (Figure 5/Table 1). Therefore, the question arises if the genes for
376 methylated amine utilisation have been acquired or lost.
377
378 Stochastic character mapping along the 16S rRNA gene phylogeny suggested that the
379 ability to use methylated amines has either been gained or lost multiple times
380 (Supplementary Figure S3). Most likely is that methylated amine utilisation is not
381 ancestral in *Gemmobacter,* and has evolved three times, once in the clade containing
382 *G.* sp. LW-1, *G. caeni* and *G. aquatilis,* once in *Haematobacter* and once in

383 *Paracoccus* (Supplementary Figure S3). An alternative, but less plausible, scenario is
384 that methylated amine utilisation is an ancestral trait in *Gemmobacter,* and has been
385 lost and regained multiple times across the phylogeny.
386
387 *Gemmobacter* sp. LW-1 was isolated from the Movile Cave ecosystem (Wischer et
388 al., 2015). Microbial mats and lake water within the cave have been shown to harbor a
389 wide diversity of methylated amine-utilising bacteria (Wischer et al.,
390 2015;Kumaresan et al., 2018). Whilst the mechanism of MAs production within the
391 system has to be elucidated, it can be speculated that degradation of floating microbial
392 mats (i.e. organic matter) could result in MAs (Wischer et al., 2015). Similarly, *G.*
393 *caeni* isolated from activated sludge (Zheng et al., 2011) could possibly use the MAs
394 generated from organic matter degradation. Interestingly, whilst *G. megaterium* was
395 isolated from a marine environment (seaweed (Liu et al., 2014)) possibly
396 encountering MAs from the degradation of osmolytes such as glycine betaine (*N,N,N*-
397 trimethylglycine) we did not detect metabolic genes involved in methylated amine
398 utilisation. Our analyses suggest that the trait for methylated amine utilisation could
399 be independent of the habitat.
400

401 **Conclusions**

402 In summary, three of the five investigated *Gemmobacter* genomes (*G.* sp. LW-1, *G.*
403 *caeni* and *G. aquatilis*) indicated metabolic potential to utilise methylated amines, of
404 which only two (*G.* sp. LW-1 and *G. caeni*) possess the genes for both MMA
405 oxidation pathways, the NMG pathway and the direct MMA oxidation pathway. *G.*
406 sp. LW-1 and *G. caeni* are facultative methylotrophs which could potentially use
407 these pathways to utilise MMA as both a carbon and nitrogen source, while

408 potentially *G. aquatilis* could only use the NMG pathway as a nitrogen source.
409 Furthermore, the genomes of *G.* sp. LW-1 and *G. caeni* showed a high similarity to
410 each other (>98%) suggesting that both belong to the same species. *G. megaterium*
411 and *G. nectariphilus* genomes indicated no metabolic potential to utilise MAs.
412 Phylogenetic, pan-genome and ANI analyses revealed that *G.* sp. LW-1 and *G. caeni*
413 are closely related, although they were isolated from different environments. Whilst
414 *G. caeni* and *G. nectariphilus* were isolated from a similar environment (activated
415 sludge) it revealed a high amount of evolutionary change from the common ancestor.
416 Overall, these results suggest that the trait for methylated amine utilisation could be
417 independent from the habitat and localised factors or selection pressures could
418 influence the ability of these organisms to use methylated amines. Access to
419 *Gemmobacter* isolates with or without the genetic potential for methylated amine
420 utilisation trait will allow us to perform physiological experiments in future to test
421 how this trait can affect fitness of closely related organisms. Ancestral state
422 reconstruction analysis confirms that across *Gemmobacter* and related genera,
423 methylated amine utilisation has either evolved or been lost multiple times over the
424 evolutionary history of this group. The adaptive or non-adaptive processes behind this
425 pattern remain to be investigated.
426

427 **Conflict of Interest**

428 The authors declare no conflict of interest.
429

430 **Acknowledgements**

431 We thank Dr Felipe Hernandes Coutinho (Federal University of Rio de Janeiro) for a
432 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

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* 215**,** 403-410.

Anthony, C. (1982). *The biochemistry of methylotrophs.* London: Academic Press.

Aziz, R.K., Bartels, D., Best, A.A., Dejongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., and Kubal, M. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9**,** 75.

Blom, J., Kreis, J., Spanig, S., Juhre, T., Bertelli, C., Ernst, C., and Goesmann, A. (2016). EDGAR 2.0: an enhanced software platform for comparative gene content analyses. *Nucleic Acids Research* 44**,** W22-W28.

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

Carpenter, L.J., Archer, S.D., and Beale, R. (2012). Ocean-atmosphere trace gas exchange. *Chemical Society Reviews* 41**,** 6473-6506.

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

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

Chen, Y., Mcaleer, K.L., and Murrell, J.C. (2010a). Monomethylamine as a nitrogen source for a nonmethylotrophic bacterium, *Agrobacterium tumefaciens*. *Applied and Environmental Microbiology* 76**,** 4102-4104.

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

Chen, Y., Wu, L., Boden, R., Hillebrand, A., Kumaresan, D., Moussard, H., Baciu, M., Lu, Y., and Colin Murrell, J. (2009). Life without light: microbial diversity and evidence of sulfur- and ammonium-based chemolithotrophy in Movile Cave. *ISME Journal* 3**,** 1093-1104.

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

Chistoserdova, L. (2011). Modularity of methylotrophy, revisited. *Environmental Microbiology* 13**,** 2603-2622.

Chistoserdova, L., Kalyuzhnaya, M.G., and Lidstrom, M.E. (2009). The expanding world of methylotrophic metabolism. *Annual Review of Microbiology* 63**,** 477-499.

507 Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the
508 bootstrap. *Evolution* 39**,** 783-791.
509 Ge, X., Wexler, A.S., and Clegg, S.L. (2011). Atmospheric amines – Part I. A review.
510 *Atmospheric Environment* 45**,** 524-546.
511 Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., and
512 Tiedje, J.M. (2007). DNA–DNA hybridization values and their relationship
513 to whole-genome sequence similarities. *International Journal of*
514 *Systematic and Evolutionary Microbiology* 57**,** 81-91.
515 Grant, J.R., Arantes, A.S., and Stothard, P. (2012). Comparing thousands of
516 circular genomes using the CGView Comparison Tool. *BMC genomics* 13**,**
517 202.
518 Jones, D.T., Taylor, W.R., and Thornton, J.M. (1992). The rapid generation of
519 mutation data matrices from protein sequences. *Bioinformatics* 8**,** 275-
520 282.
521 Kämpfer, P., Jerzak, L., Wilharm, G., Golke, J., Busse, H.-J., and Glaeser, S.P. (2015).
522 *Gemmobacter intermedius* sp. nov., isolated from a white stork (Ciconia
523 ciconia). *International Journal of Systematic and Evolutionary Microbiology*
524 65**,** 778-783.
525 Kim, M., Oh, H.-S., Park, S.-C., and Chun, J. (2014). Towards a taxonomic
526 coherence between average nucleotide identity and 16S gene sequence
527 similarity for species demarcation of prokaryotes. *International Journal of*
528 *Systematic and Evolutionary Microbiology* 64**,** 346-351.
529 Konstantinidis, K.T., and Tiedje, J.M. (2005). Genomic insights that advance the
530 species definition for prokaryotes. *Proceedings of the National Academy of*
531 *Sciences of the United States of America* 102**,** 2567-2572.
532 Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary
533 Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and*
534 *Evolution* 33**,** 1870-1874.
535 Kumaresan, D., Stephenson, J., Doxey, A.C., Bandukwala, H., Brooks, E.,
536 Hillebrand-Voiculescu, A., Whiteley, A.S., and Murrell, J.C. (2018). Aerobic
537 proteobacterial methylotrophs in Movile Cave: genomic and metagenomic
538 analyses. *Microbiome* 6.
539 Kumaresan, D., Wischer, D., Hillebrand-Voiculescu, A.M., and Murrell, J.C. (2015).
540 Draft genome sequences of facultative methylotrophs, *Gemmobacter* sp.
541 strain LW1 and *Mesorhizobium* sp. strain 1M-11, isolated from Movile
542 Cave, Romania. *Genome Announcements* 3.
543 Kumaresan, D., Wischer, D., Stephenson, J., Hillebrand-Voiculescu, A., and
544 Murrell, J.C. (2014). Microbiology of Movile Cave - A
545 chemolithoautotrophic ecosystem. *Geomicrobiology Journal* 31**,** 186-193.
546 Lagesen, K., Hallin, P., Rødland, E.A., Stærfeldt, H.-H., Rognes, T., and Ussery, D.W.
547 (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA
548 genes. *Nucleic acids research* 35**,** 3100-3108.
549 Latypova, E., Yang, S., Wang, Y.S., Wang, T., Chavkin, T.A., Hackett, M., Schäfer, H.,
550 and Kalyuzhnaya, M.G. (2010). Genetics of the glutamate‐mediated
551 methylamine utilization pathway in the facultative methylotrophic beta‐
552 proteobacterium *Methyloversatilis universalis* FAM5. *Molecular*
553 *Microbiology* 75**,** 426-439.

554 Lidbury, I., Murrell, J.C., and Chen, Y. (2014). Trimethylamine N-oxide
555 metabolism by abundant marine heterotrophic bacteria. *Proceedings of*
556 *the National Academy of Sciences* 111**,** 2710-2715.
557 Lidstrom, M.E. (2006). "Aerobic Methylotrophic Prokaryotes," in *The*
558 *Prokaryotes: Volume 2: Ecophysiology and Biochemistry,* eds. M. Dworkin,
559 S. Falkow, E. Rosenberg, K.-H. Schleifer & E. Stackebrandt. (New York, NY:
560 Springer New York), 618-634.
561 Liu, J.-J., Zhang, X.-Q., Chi, F.-T., Pan, J., Sun, C., and Wu, M. (2014). *Gemmobacter*
562 *megaterium* sp. nov., isolated from coastal planktonic seaweeds.
563 *International Journal of Systematic and Evolutionary Microbiology* 64**,** 66-
564 71.
565 Liu, Y., and Whitman, W.B. (2008). Metabolic, phylogenetic, and ecological
566 diversity of the methanogenic archaea. *Annals of the New York Academy of*
567 *Sciences* 1125**,** 171-189.
568 Liu, Y., Xu, C.-J., Jiang, J.-T., Liu, Y.-H., Song, X.-F., Li, H., and Liu, Z.-P. (2010).
569 *Catellibacterium aquatile* sp. nov., isolated from fresh water, and emended
570 description of the genus *Catellibacterium* Tanaka et al. 2004. *International*
571 *Journal of Systematic and Evolutionary Microbiology* 60**,** 2027-2031.
572 Lyimo, T.J., Pol, A., Jetten, M.S.M., and Op Den Camp, H.J.M. (2009). Diversity of
573 methanogenic archaea in a mangrove sediment and isolation of a new
574 *Methanococcoides* strain. *FEMS Microbiology Letters* 291**,** 247-253.
575 Markowitz, V.M., Chen, I.-M.A., Palaniappan, K., Chu, K., Szeto, E., Pillay, M.,
576 Ratner, A., Huang, J., Woyke, T., and Huntemann, M. (2013). IMG 4 version
577 of the integrated microbial genomes comparative analysis system. *Nucleic*
578 *acids research* 42**,** D560-D567.
579 Mcintire, W.S., Wemmer, D.E., Chistoserdova, A., and Lidstrom, M.E. (1991). A
580 new cofactor in a prokaryotic enzyme: tryptophan tryptophylquinone as
581 the redox prosthetic group in methylamine dehydrogenase. *Science* 252**,**
582 817-824.
583 Nayak, D.D., Agashe, D., Lee, M.-C., and Marx, C.J. (2016). Selection maintains
584 apparentlydegenerate metabolic pathwaysdueto tradeoffsin using
585 methylamine for carbon versus nitrogen. *Current Biology* 26**,** 1-11.
586 Nayak, D.D., and Marx, C.J. (2015). Experimental horizontal gene transfer of
587 methylamine dehydrogenase mimics prevalent exchange in nature and
588 overcomes the methylamine growth constraints posed by the sub-optimal
589 N-Methylglutamate pathway. *Microorganisms* 3**,** 60-79.
590 Nielsen, R. (2002). Mapping mutations on phylogenies. *Systematic Biology* 51**,**
591 729-739.
592 Paul, L., Ferguson, D.J., and Krzycki, J.A. (2000). The trimethylamine
593 methyltransferase gene and multiple dimethylamine methyltransferase
594 genes of *Methanosarcina barkeri* contain in-frame and read-through
595 amber codons. *Journal of Bacteriology* 182**,** 2520-2529.
596 Pruesse, E., Peplies, J., and Glöckner, F.O. (2012). SINA: accurate high-throughput
597 multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 28**,**
598 1823-1829.
599 Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glöckner,
600 F.O. (2007). SILVA: a comprehensive online resource for quality checked
601 and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic*
602 *Acids Research* 35**,** 7188-7196.

603 Revell, L.J. (2012). phytools: an R package for phylogenetic comparative biology
604 (and other things). *Methods in Ecology and Evolution* 3**,** 217-223.
605 Revell, L.J. (2013). Two new graphical methods for mapping trait evolution on
606 phylogenies. *Methods in Ecology and Evolution* 4**,** 754-759.
607 Richter, M., and Rossello-Mora, R. (2009). Shifting the genomic gold standard for
608 the prokaryotic species definition. *Proceedings of the National Academy of*
609 *Sciences* 106**,** 19126-19131.
610 Rodriguez-R, L.M., and Konstantinidis, K.T. (2016). The enveomics collection: a
611 toolbox for specialized analyses of microbial genomes and metagenomes.
612 *PeerJ Preprints*.
613 Rothe, B., Fischer, A., Hirsch, P., Sittig, M., and Stackebrandt, E. (1987). The
614 phylogenetic position of the budding bacteria *Blastobacter aggregatus*
615 and *Gemmobacter aquatilis* gen., nov. sp. nov. *Archives of Microbiology*
616 147**,** 92-99.
617 Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for
618 reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4**,** 406-
619 425.
620 Schade, G.W., and Crutzen, P.J. (1995). Emission of aliphatic-amines from animal
621 husbandry and their reactions - potential source of N_2O and HC_n. *Journal*
622 *of Atmospheric Chemistry* 22**,** 319-346.
623 Sheu, S.-Y., Sheu, D.-S., Sheu, F.-S., and Chen, W.-M. (2013a). *Gemmobacter tilapiae*
624 sp. nov., a poly-β-hydroxybutyrate-accumulating bacterium isolated from
625 a freshwater pond. *International Journal of Systematic and Evolutionary*
626 *Microbiology* 63**,** 1550-1556.
627 Sheu, S.-Y., Shiau, Y.-W., Wei, Y.-T., and Chen, W.-M. (2013b). *Gemmobacter*
628 *lanyuensis* sp. nov., isolated from a freshwater spring. *International*
629 *Journal of Systematic and Evolutionary Microbiology* 63**,** 4039-4045.
630 Tanaka, Y., Hanada, S., Manome, A., Tsuchida, T., Kurane, R., Nakamura, K., and
631 Kamagata, Y. (2004). *Catellibacterium nectariphilum* gen. nov., sp. nov.,
632 which requires a diffusible compound from a strain related to the genus
633 *Sphingomonas* for vigorous growth. *International Journal of Systematic*
634 *and Evolutionary Microbiology* 54**,** 955-959.
635 Taubert, M., Grob, C., Howat, A.M., Burns, O.J., Pratscher, J., Jehmlich, N., Von
636 Bergen, M., Richnow, H.H., Chen, Y., and Murrell, J.C. (2017). Methylamine
637 as a nitrogen source for microorganisms from a coastal marine
638 environment. *Environmental Microbiology* 19**,** 2246-2257.
639 Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the
640 sensitivity of progressive multiple sequence alignment through sequence
641 weighting, position-specific gap penalties and weight matrix choice.
642 *Nucleic Acids Research* 22**,** 4673-4680.
643 Wischer, D., Kumaresan, D., Johnston, A., El Khawand, M., Stephenson, J.,
644 Hillebrand-Voiculescu, A.M., Chen, Y., and Colin Murrell, J. (2015).
645 Bacterial metabolism of methylated amines and identification of novel
646 methylotrophs in Movile Cave. *ISME Journal* 9**,** 195-206.
647 Zhang, J., Chen, S.-A., Zheng, J.-W., Cai, S., Hang, B.-J., He, J., and Li, S.-P. (2012).
648 *Catellibacterium nanjingense* sp. nov., a propanil-degrading bacterium
649 isolated from activated sludge, and emended description of the genus
650 *Catellibacterium*. *International Journal of Systematic and Evolutionary*
651 *Microbiology* 62**,** 495-499.

652 Zheng, J.-W., Chen, Y.-G., Zhang, J., Ni, Y.-Y., Li, W.-J., He, J., and Li, S.-P. (2011).
653 Description of *Catellibacterium caeni* sp. now., reclassification of
654 *Rhodobacter changlensis* Anil Kumar et al. 2007 as *Catellibacterium*
655 *changlense* comb. nov. and emended description of the genus
656 *Catellibacterium*. *International Journal of Systematic and Evolutionary*
657 *Microbiology* 61**,** 1921-1926.
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692

693 **Figures and Tables**

694

695 **Figure 1. Phylogenetic tree based on 16S rRNA gene sequences.** The tree was
696 constructed using the maximum likelihood method for clustering and the Tamura-Nei
697 model for computing evolutionary distances. Numbers at branches are bootstrap
698 percentages >50% of 1000 replicates. Star represents the *Gemmobacter* species used
699 for comparative genome analysis. Coloured boxes represent the habitat where the
700 sequence was retrieved: blue (fresh water), orange (soil and sediment), green
701 (activated sludge), grey (marine), purple (clinical source). Triangles represent
702 sequences that are listed as *Catellibacterium* in the NCBI database, which have been
703 recently reclassified to *Gemmobacter* (Chen et al., 2013). Scale bar: 0.02 substitutions
704 per nucleotide position.

705

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

713

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

719 strand, respectively. The colour scale (inset) shows the level of sequence identity with
720 the respective sequences from *G. megaterium, G. aquatilis, G. nectariphilus* and *G.*
721 *caeni*. The locations of genes involved in methylotrophy are indicated at the outside
722 of the map.
723
724 **Figure 4. Pan-genome analysis. (A)** Pan-genome tree consisting of five
725 *Gemmobacter* species was constructed using the neighbour-joining method within the
726 EDGAR platform. **(B)** Number of core, dispensable, and specific genes (singletons)
727 of each *Gemmobacter* species.
728
729 **Figure 5. Metabolic pathways involved in methylated amine utilisation and one-**
730 **carbon utilisation annotated with presence/absence of specific genes in the**
731 **genomes of *Gemmobacter*.** The analysis was based on a five-way comparison among
732 *Gemmobacter* sp. LW-1 (L), *Gemmobacter caeni* (C), *Gemmobacter aquatilis* (A),
733 *Gemmobacter nectariphilus* (N) and *Gemmobacter megaterium* (M). The color-coded
734 boxes next to the genes indicate the presence (green) or absence (orange) of a gene in
735 each genome.
736
737 **Table 1. Comparative genomic analysis of methylated amine-utilising genes in**
738 **genomes-sequenced *Gemmobacters* in comparison to selected marine *Roseobacter***
739 **clade bacteria.** Shown is the presence (+) or absence (-) of specific genes in the
740 genome sequences.
741
742 **Supplementary Figure S1. Maximum-likelihood phylogenetic tree (JTT matrix-**
743 **based) of *mauA* sequences**. Sequences from the genus *Gemmobacter* are marked

744 with a star. Amino acid sequences (MauA) were aligned using the ClustalW
745 algorithm. Numbers at branches are bootstrap percentages >50% of 1000 replicates.
746 Scale bar: 0.1 substitutions per amino acid. Coloured boxes indicate
747 *Alphaproteobacteria (*yellow), *Gammaproteobacteria (*red) and *Betaproteobacteria*
748 *(*black).
749
750 **Supplementary Figure S2. (A-D) Average nucleotide identity (ANI) analysis of**
751 ***Gemmobacter* sp. LW-1 and *Gemmobacter caeni*, *Gemmobacter aquatilis*,**
752 ***Gemmobacter nectariphilus* and *Gemmobacter megaterium* and (E) AAI analysis**
753 **between those species**
754
755 **Supplementary Figure S3. Ancestral state reconstruction of methylated amine**
756 **utilisation along the 16S rRNA gene phylogeny, using stochastic mapping.** Branch
757 colour represents the posterior probability (computed as the relative frequency across
758 stochastic maps) of methylated amine utilisation through the phylogeny. Red indicates
759 a high posterior probability of methylated amine utilisation.
760
761 **Supplementary Table S1. Genome characteristics of the five *Gemmobacter* isolate**
762 **genomes used in this study.**
763
764 **Supplementary Table S2. List of protein queries used for the genome**
765 **comparison with their accession number.**
766
767 **Supplementary Table S3. Average nucleotide identity (ANI)** values between
768 *Gemmobacter* sp. LW-1 and *Gemmobacter caeni*, *Gemmobacter aquatilis*,

769 *Gemmobacter nectariphilus*, *Gemmobacter megaterium*, *Rhodobacter sphaeroides*

770 and *Paracoccus denitrificans*

771

772

Figure 1

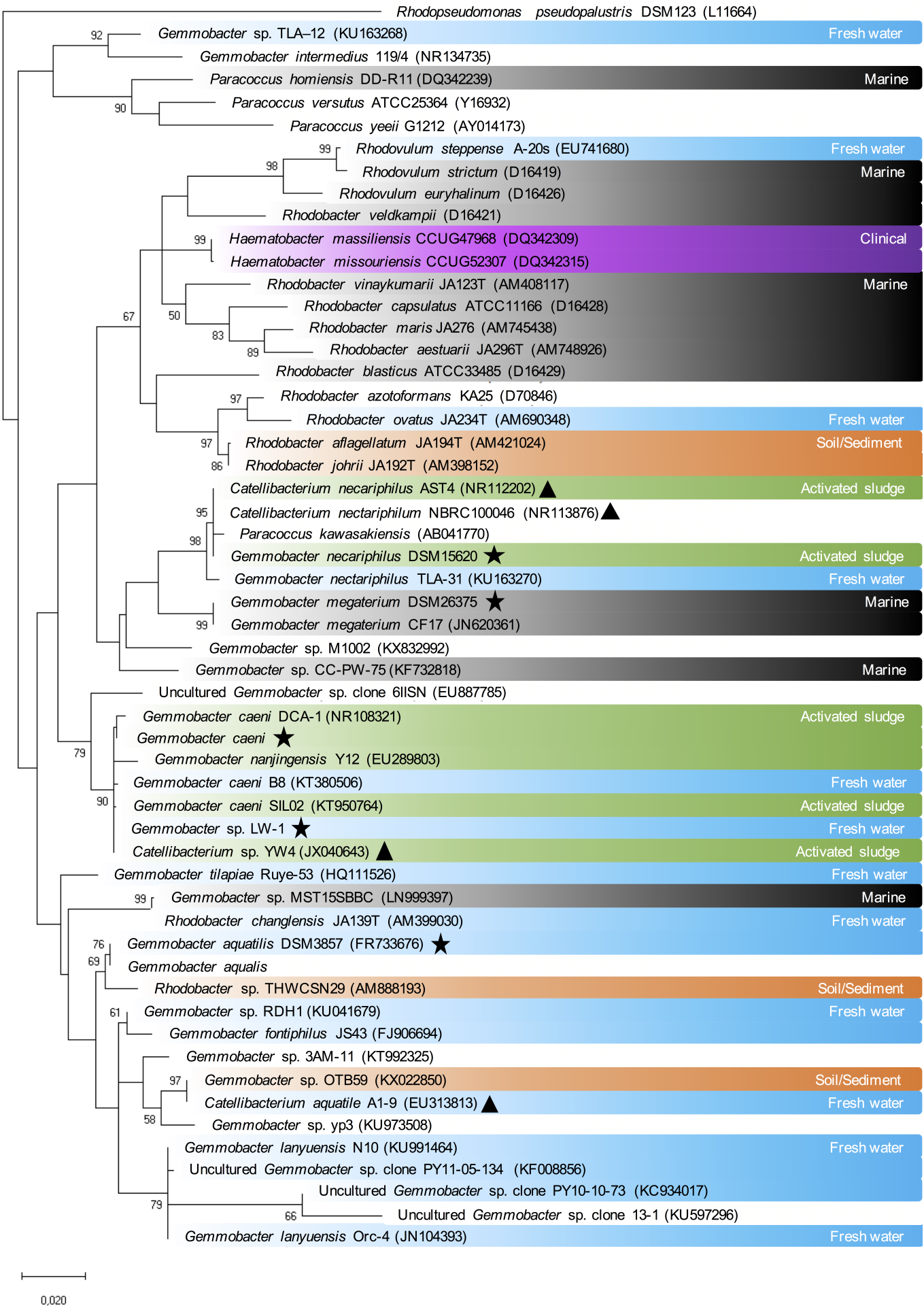

Rhodopseudomonas pseudopalustris DSM123 (L11664)
Gemmobacter sp. TLA–12 (KU163268)
Fresh water
Gemmobacter intermedius 119/4 (NR134735)
Paracoccus homiensis DD-R11 (DQ342239)
Marine
Paracoccus versutus ATCC25364 (Y16932)
Paracoccus yeeii G1212 (AY014173)
Rhodovulum steppense A-20s (EU741680)
Fresh water
Rhodovulum strictum (D16419)
Marine
Rhodovulum euryhalinum (D16426)
Rhodobacter veldkampii (D16421)
Haematobacter massiliensis CCUG47968 (DQ342309)
Clinical
Haematobacter missouriensis CCUG52307 (DQ342315)
Rhodobacter vinaykumarii JA123T (AM408117)
Marine
Rhodobacter capsulatus ATCC11166 (D16428)
Rhodobacter maris JA276 (AM745438)
Rhodobacter aestuarii JA296T (AM748926)
Rhodobacter blasticus ATCC33485 (D16429)
Rhodobacter azotoformans KA25 (D70846)
Rhodobacter ovatus JA234T (AM690348)
Fresh water
Rhodobacter aflagellatum JA194T (AM421024)
Soil/Sediment
Rhodobacter johrii JA192T (AM398152)
Catellibacterium necariphilus AST4 (NR112202)
Activated sludge
Catellibacterium nectariphilum NBRC100046 (NR113876)
Paracoccus kawasakiensis (AB041770)
Gemmobacter necariphilus DSM15620
Activated sludge
Gemmobacter nectariphilus TLA-31 (KU163270)
Fresh water
Gemmobacter megaterium DSM26375
Marine
Gemmobacter megaterium CF17 (JN620361)
Gemmobacter sp. M1002 (KX832992)
Gemmobacter sp. CC-PW-75 (KF732818)
Marine
Uncultured Gemmobacter sp. clone 6IISN (EU887785)
Gemmobacter caeni DCA-1 (NR108321)
Activated sludge
Gemmobacter caeni
Gemmobacter nanjingensis Y12 (EU289803)
Gemmobacter caeni B8 (KT380506)
Fresh water
Gemmobacter caeni SIL02 (KT950764)
Activated sludge
Gemmobacter sp. LW-1
Fresh water
Catellibacterium sp. YW4 (JX040643)
Activated sludge
Gemmobacter tilapiae Ruye-53 (HQ111526)
Fresh water
Gemmobacter sp. MST15SBBC (LN999397)
Marine
Rhodobacter changlensis JA139T (AM399030)
Fresh water
Gemmobacter aquatilis DSM3857 (FR733676)
Gemmobacter aqualis
Rhodobacter sp. THWCSN29 (AM888193)
Soil/Sediment
Gemmobacter sp. RDH1 (KU041679)
Fresh water
Gemmobacter fontiphilus JS43 (FJ906694)
Gemmobacter sp. 3AM-11 (KT992325)
Gemmobacter sp. OTB59 (KX022850)
Soil/Sediment
Catellibacterium aquatile A1-9 (EU313813)
Fresh water
Gemmobacter sp. yp3 (KU973508)
Gemmobacter lanyuensis N10 (KU991464)
Fresh water
Uncultured Gemmobacter sp. clone PY11-05-134 (KF008856)
Uncultured Gemmobacter sp. clone PY10-10-73 (KC934017)
Uncultured Gemmobacter sp. clone 13-1 (KU597296)
Gemmobacter lanyuensis Orc-4 (JN104393)
Fresh water
0,020

Figure 2

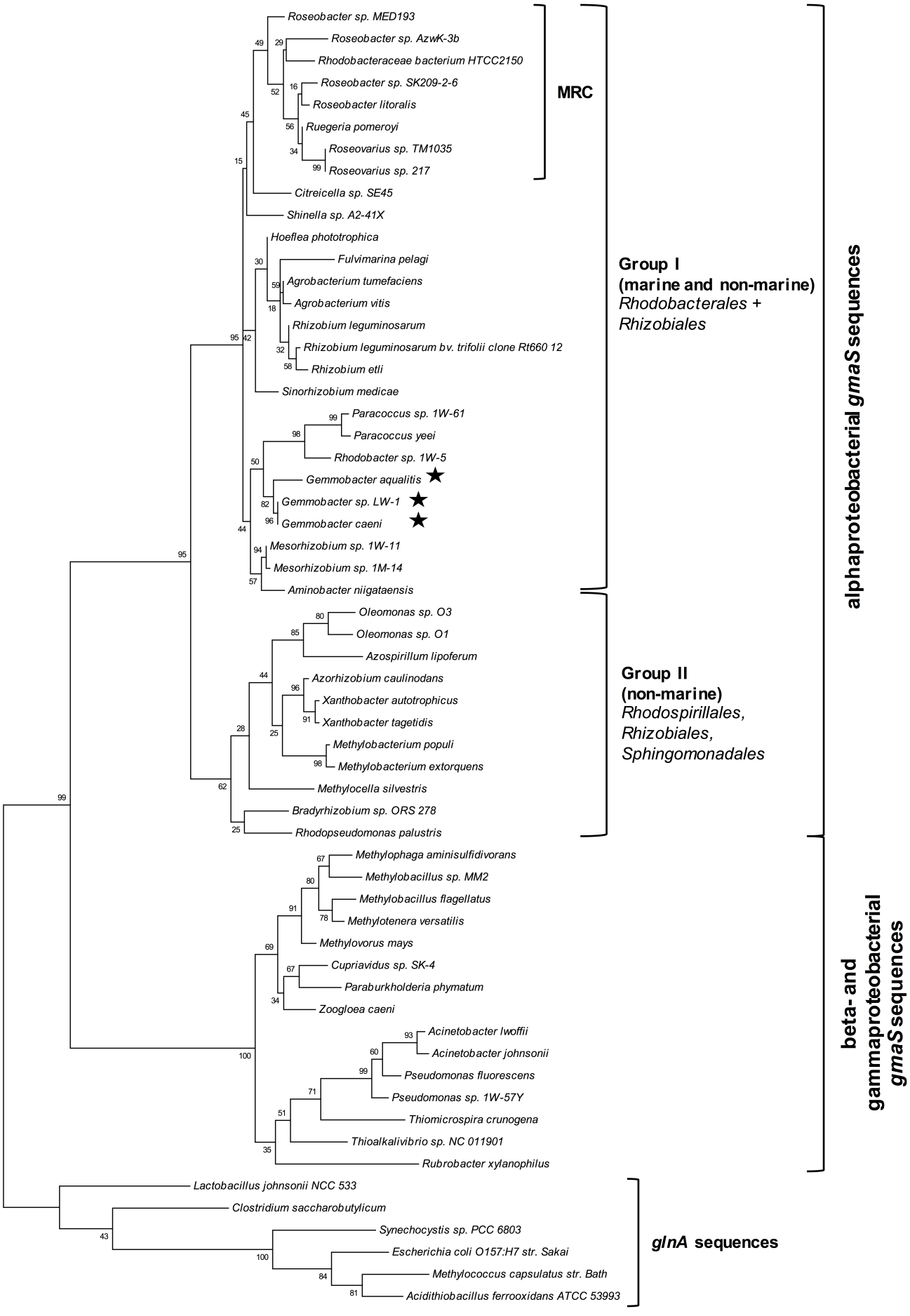
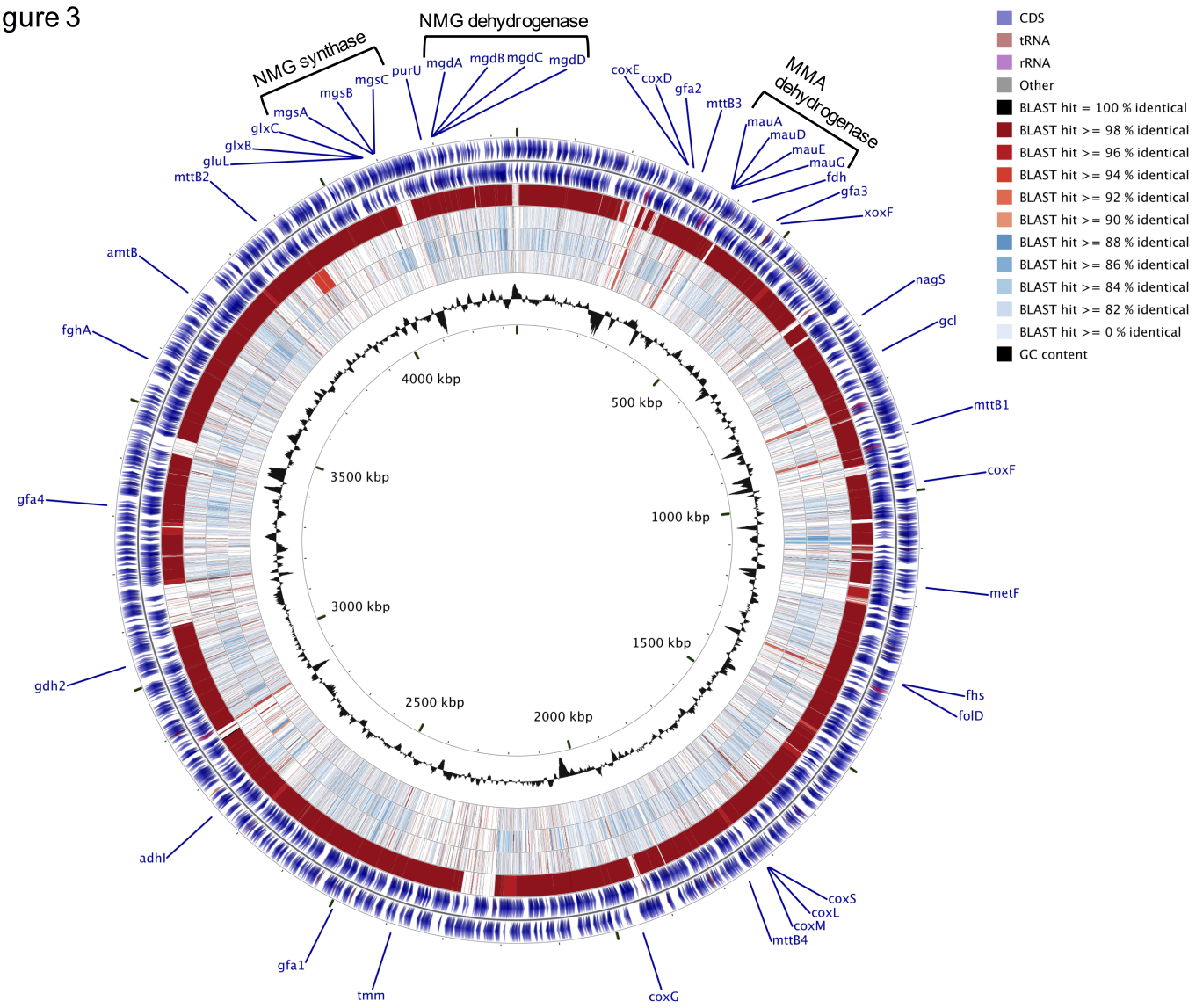

Roseobacter sp. MED193
Roseobacter sp. AzwK-3b
Rhodobacteraceae bacterium HTCC2150
Roseobacter sp. SK209-2-6
Roseobacter litoralis
Ruegeria pomeroyi
Roseovarius sp. TM1035
Roseovarius sp. 217
MRC
Citreicella sp. SE45
Shinella sp. A2-41X
Hoeflea phototrophica
Fulvimarina pelagi
Agrobacterium tumefaciens
Agrobacterium vitis
Rhizobium leguminosarum
Rhizobium leguminosarum bv. trifolii clone Rt660 12
Rhizobium etli
Sinorhizobium medicae
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
Group I
(marine and non-marine)
Rhodobacterales +
Rhizobiales
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

Figure 3



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

Figure 4A

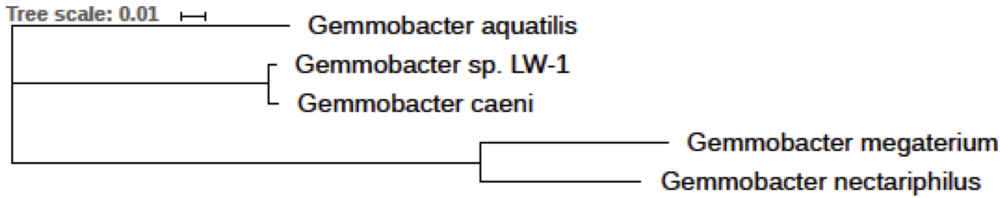


Figure 4B

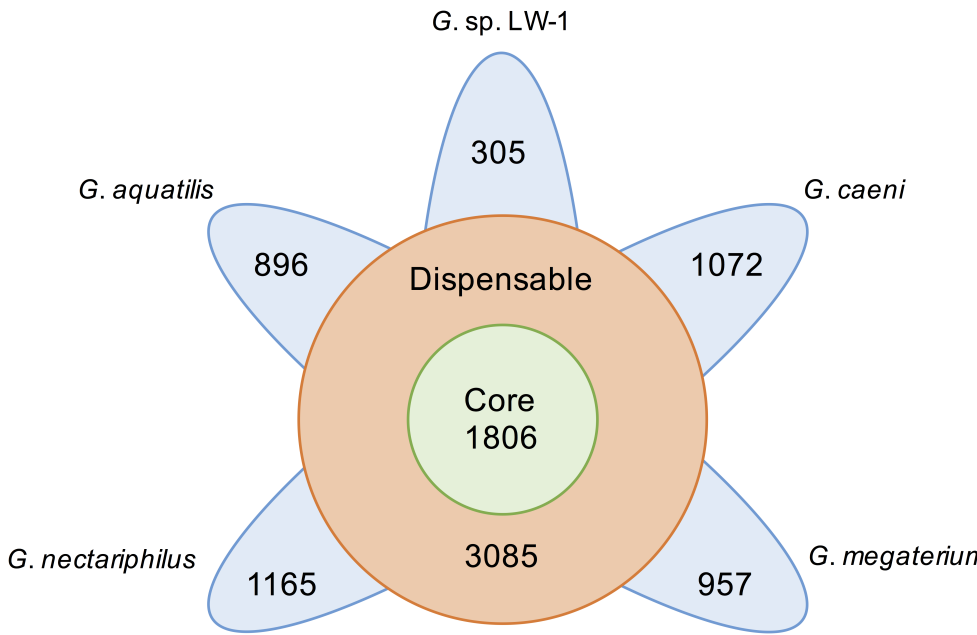


Figure 5

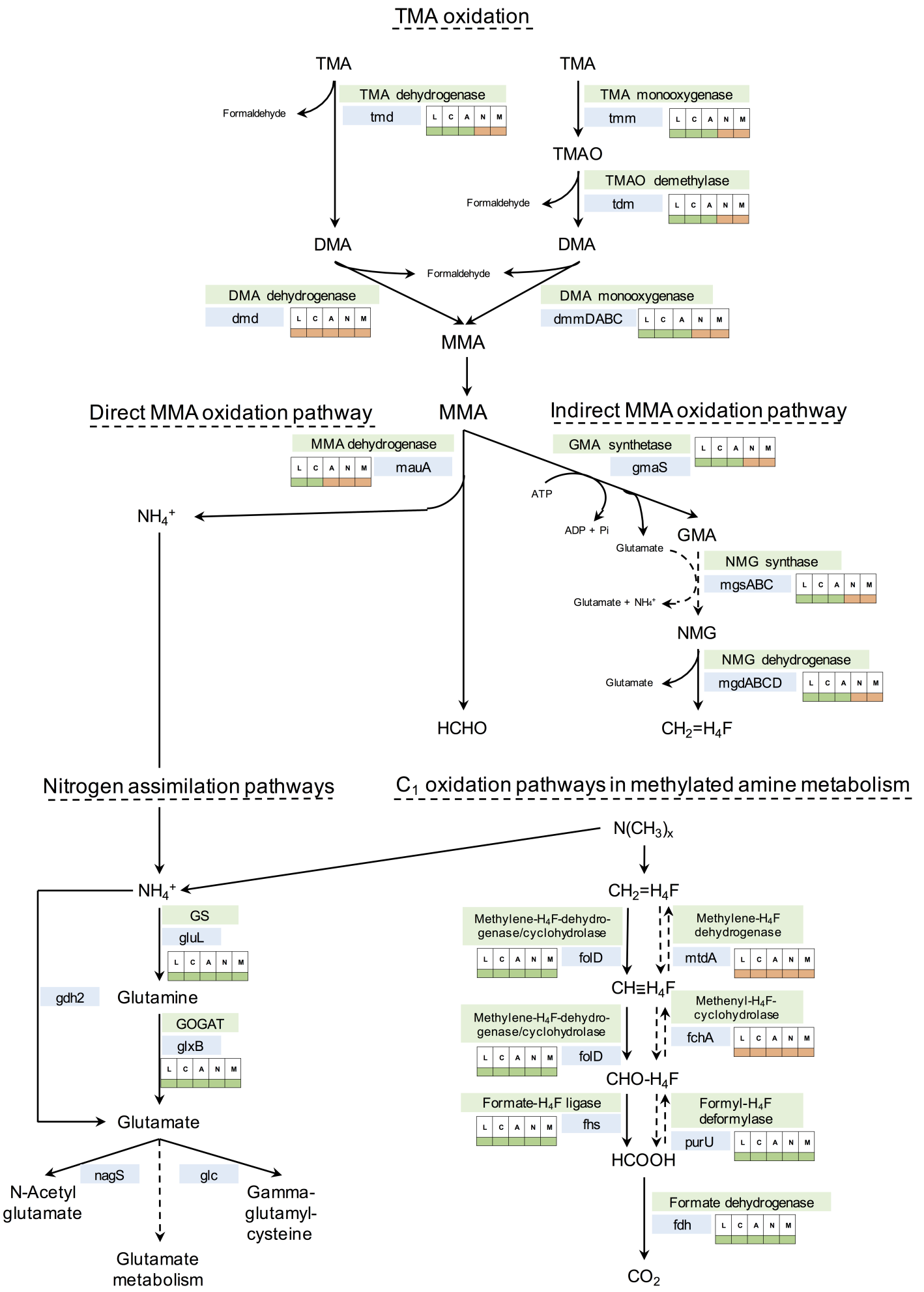

TMA oxidation
TMA
TMA
Formaldehyde
TMA dehydrogenase
tmd
TMA monooxygenase
tmm
TMAO
TMAO demethylase
tdm
Formaldehyde
DMA
DMA
Formaldehyde
DMA dehydrogenase
dmd
DMA monooxygenase
dmmDABC
MMA
MMA
Direct MMA oxidation pathway
Indirect MMA oxidation pathway
MMA dehydrogenase
mauA
GMA synthetase
gmaS
ATP
ADP + Pi
Glutamate
GMA
NH4+
NMG synthase
mgsABC
Glutamate + NH4+
NMG
NMG dehydrogenase
mgdABCD
Glutamate
HCHO
CH2=H4F
Nitrogen assimilation pathways
C1 oxidation pathways in methylated amine metabolism
N(CH3)x
NH4+
CH2=H4F
GS
gluL
Methylene-H4F-dehydro-genase/cyclohydrolase
folD
Methylene-H4F dehydrogenase
mtdA
gdh2
Glutamine
CH≡H4F
GOGAT
glxB
Methylene-H4F-dehydro-genase/cyclohydrolase
folD
Methenyl-H4F-cyclohydrolase
fchA
CHO-H4F
Glutamate
Formate-H4F ligase
fhs
Formyl-H4F deformylase
purU
HCOOH
nagS
glc
N-Acetyl glutamate
Gamma-glutamyl-cysteine
Glutamate metabolism
Formate dehydrogenase
fdh
CO2

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	*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