Title: A diversified, widespread microbial gene cluster encodes homologs of methyltransferases involved in methanogenesis.

Duncan J. Kountz[1] and Emily P. Balskus[1,2 *]

Author information

Duncan J. Kountz djkountz@g.harvard.edu

Emily P. Balskus balskus@chemistry.harvard.edu, orcid.org/0000-0001-5985-5714

[1]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, United States.

[2]Howard Hughes Medical Institute, Harvard University, Cambridge, Massachusetts 02138, United States.

* Corresponding author

Competing interests: None

Abstract: Analyses of microbial genome sequencing data have revealed unexpectedly wide distributions of enzymes from specialized metabolic pathways, including enzymes from methanogens, providing exciting opportunities for discovery. Here, we identify a family of gene clusters (the type 1 *mlp* gene clusters (MGCs)) that encodes homologs of the soluble coenzyme M methyltransferases (SCMTs) involved in methylotrophic methanogenesis and is widespread in bacteria and archaea. Type 1 MGCs are expressed and regulated in a number of medically, environmentally, and industrially important organisms, making them likely to be physiologically relevant. Enzyme annotation and analysis of genomic context suggests these gene clusters are likely to play a role in methyl-sulfur and/or methyl-selenide metabolism in numerous anoxic environments, including the human gut microbiome. Notably, we propose that type 1 MGCs could participate in selenium and methionine salvage pathways that could impact sulfur and selenium cycling in diverse, anoxic environments.

Introduction

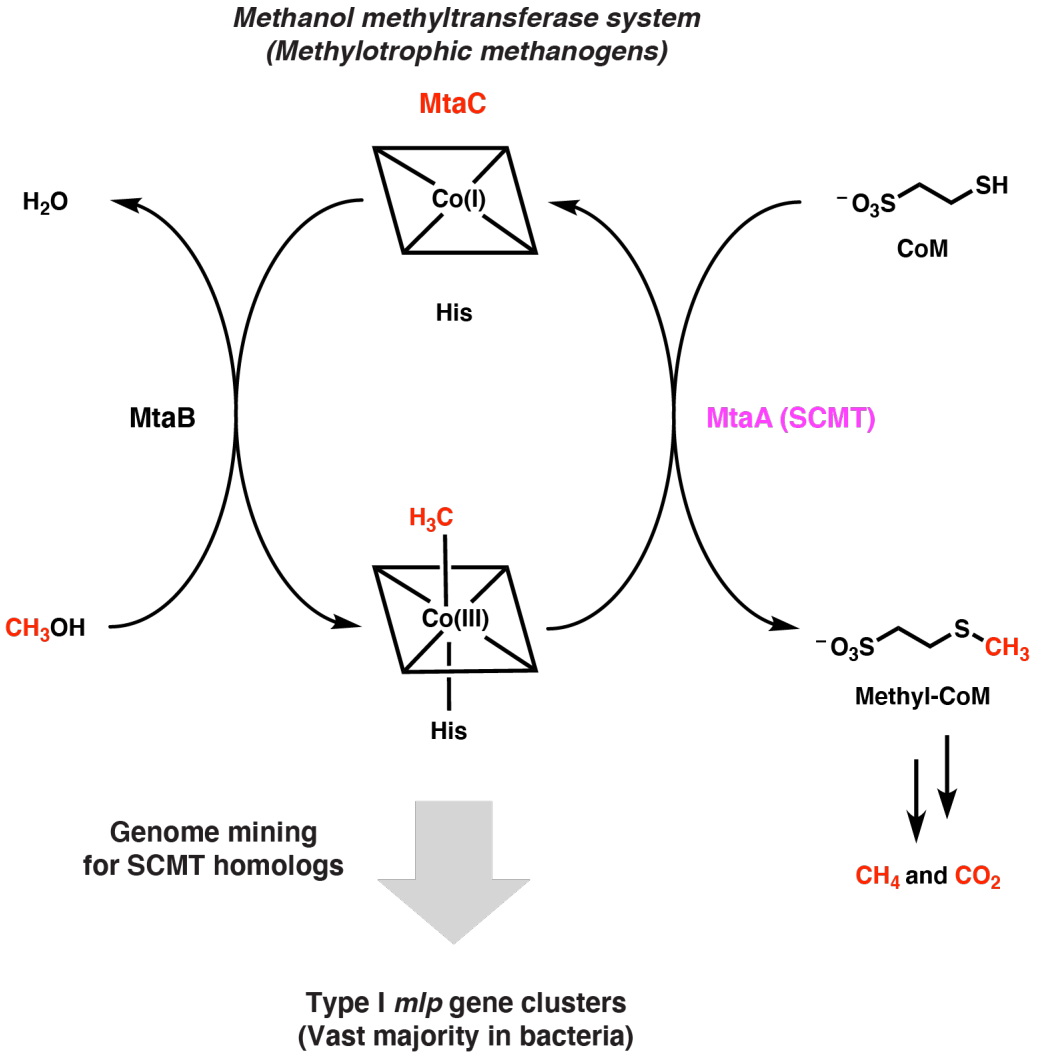
Other the past 50 years, biochemists and microbiologists have made much progress in elucidating the enzymology of methanogenesis. The canonical methanogenic pathways are now well understood (1,2), and the vast majority of the central catabolic enzymes have been purified and at least partially characterized, and their genes identified and sequenced. This reservoir of biochemical knowledge presents the opportunity to identify homologs of methanogen specific enzymes in organisms that are categorically not methanogens. In doing so, we may discover previously unappreciated metabolic

45 pathways that could have global or ecological relevance. Moreover, studying homologs
46 of methanogen-specific enzymes in non-methanogenic organisms could shed light on the
47 evolution of methanogenic pathways and organisms as well as the potential adaptation
48 of methanogen genes to perform novel functions.

49 Methanogenesis can be regarded as a specialized metabolism—its pathways are
50 biochemically quite distinct from those of most organisms and not expected to interface
51 well with other metabolic processes when horizontally transferred. It is therefore
52 interesting to note that methanogen cofactors and homologs of methanogen enzymes are
53 indeed found in non-methanogens. Examples include the use of methanofuran- and
54 methanopterin-dependent enzymes in aerobic methylotrophy (3), the use of deazaflavin
55 redox cofactor F_{420} in bacterial metabolism (4), the use of coenzyme M in alkene
56 catabolism (5), and the use of heterodisulfide reductase-like enzymes in various
57 metabolisms, including sulfate reduction (6), acetogenesis (7), and the degradation of
58 benzoate (8). In each case the "methanogen" enzymes are performing an interesting,
59 novel, and important function.

60 We wondered if this occurrence of methanogen enzymes in non-methanogenic
61 organisms extends to the soluble coenzyme M methyltransferases (SCMTs) from
62 methylotrophic methanogenesis. SCMTs are corrinoid-dependent methyltransferases
63 that are responsible for transferring a methyl group from a methylated corrinoid-binding
64 protein (corrinoid protein) to coenzyme M to form methyl-coenzyme M, the immediate
65 precursor of methane (Fig. 1) (9). SCMTs belong to the very large and sequence-diverse
66 uroporphyrinogen decarboxylase (UroD) superfamily, which includes uroporphyrinogen

decarboxylase, cobalamin-independent methionine synthase, epoxide coenzyme M transferase, and a variety of corrinoid methyltransferases including *O*-, *S*-, and halogen-demethylases (10–16). However, we focused our attention on proteins having greater than 25% amino acid identity to a characterized SCMT. We will refer to such enzymes as MtaA-like proteins, or Mlps, after MtaA, the methanol-specific SCMT in *Methanosarcina* species (16). To our knowledge, these Mlps have not been assigned a role outside of methanogens.



75 **Fig. 1 Searching genomes for the soluble coenzyme M methyltransferase (SCMT) MtaA involved**
76 **in methylotrophic methanogenesis identifies related enzymes in non-methanogens.** Overview of
77 genome mining strategy used in this work. The methanol-specific methanol:corrinoid methyltransferase
78 MtaB transfers the methyl group of methanol to the cognate corrinoid protein MtaC (the corrinoid cofactor
79 is represented by a square containing a cobalt ion). The SCMT MtaA then transfers to the methyl group
80 from MtaC to coenzyme M (CoM). The methyl group incorporated onto CoM is then disproportionated into
81 methane and carbon dioxide or solely reduced to methane, depending on the organism and environmental
82 conditions.

83

84 Here, we propose a prospective role for Mlps in a wide range of microorganisms by
85 identifying a diverse, widespread, and homologous set of microbial gene clusters that
86 encode multiple Mlps, as well as a corrinoid protein and a corrinoid reductive activase
87 (Ram/RACE enzyme). Gene cluster annotation, bioinformatic analysis, and biochemical
88 logic suggests that the encoded enzymes have two functions—methyl-selenide recycling
89 and/or methionine salvage—depending on the organism. These results illustrate the utility
90 of using known enzymes involved in metabolisms of interest to discover uncharacterized
91 microbial gene clusters that likely have intriguing and ecologically important functions.

92

93 **Materials and methods**

94

95 General methods and procedures

96 Manipulations and visualizations of genomic data were performed using Geneious
97 Prime 2020.0.4 software (17). BLASTP searches were conducted through the Geneious
98 interface using either National Center for Biotechnology Information (NCBI) nr/rt database

99 or custom databases generated by the user. In all cases we used the BLOSUM62 matrix
100 and a gap opening penalty of 11 and a gap extension penalty of 1. Protein alignments we
101 performed using MUSCLE (18).
102
103 Identification of *mlp* gene clusters
104 To identify homologs of soluble coenzyme M methyltransferases (SCMTs) in *C.*
105 *ljungdahlii*, we generated a protein database of all the predicted proteins encoded in the
106 *C. ljungdahlii* DSM 13528 genome, and performed at BLASTP search of this database
107 using the *Methanosarcina barkeri* MtbA protein sequence (Swiss-Prot ID: O30640) as a
108 query. We then inspected the genomic context of the all the hits with E-values less than
109 1.0e-10.
110 To identify additional *mlp* gene clusters (MGCs), we performed BLASTP searches
111 against NCBI's nr/nt database using *C. ljungdahlii* Mlp1, Mlp2, and Mlp3 (locus tags in
112 Table S1) as queries. We collected the top 1000 hits from each BLAST search. For the
113 Mlp1 search, this yielded hits with as low as 19.8% amino acid identity to the query. For
114 the Mlp2 and Mlp3 searches, the minimum percent identities of the hits were 28.0% and
115 24.9%, respectively. We then manually inspected the genomic context of each hit to
116 determine whether they were found in true MGCs. A "true" MGC met the following criteria:
117 1) it encodes at least two Mlps, a corrinoid protein, and a Ram protein, 2) each of these
118 components are predicted to be transcribed in the same direction with no intervening
119 divergently-transcribed genes, 3) there is no more than 2 kb of DNA in between the genes
120 noted in 1, and 4) the gene cluster encodes no homologs of tetrahydrofolate methylases

121 within 20 kb of any of the genes noted in 1. Using these methods and definitions, we
122 identified 367 MGCs.
123
124 SSN analysis of MGC proteins
125 We generated all SSNs using the Enzyme Function Initiative Enzyme Similarity Tool
126 (EFI-EST) webtool (19). For each SSN, we first compiled the sequences of characterized
127 superfamily members and superfamily members that are encoded in the 367 MGCs
128 identified above, generating a superfamily-specific "preliminary input list". We then pruned
129 each list to ensure that only quality sequences (those lacking frameshifts or internal stops)
130 within a certain length window were included. For the Mlp SSN, length window was 200
131 to 470 amino acids. The resulting sequence list was termed the "input list". We next input
132 both the appropriate superfamily and input list into the EFI-EST web server. For the Mlp
133 SSN, the superfamily was IPR027980, for the Ram/RACE SSN, it was IPR027980, and
134 for the corrinoid protein SSN it was IPR006158. We selected alignment score thresholds
135 such that most MGC proteins fell into common network clusters, but were still as
136 separated as possible from non-MGC proteins. We selected node representation percent
137 identities to the highest that consistent with efficient use of the Cytoscape GUI on an
138 ordinary laptop MacIntosh computer. For each SSN, we used the UniRef90 option and
139 otherwise default parameters. Once generated, the SSNs were visualized with Cytoscape
140 software (version 3.8.0) using the "preferred" layout (20).
141
142 Phylogenetic analysis of *mlp* gene cluster proteins

143 We compiled proteins encoded by type I MGCs and sequences of related
144 characterized proteins and then aligned the resulting protein sequences using MUSCLE
145 with default parameters, and trimmed the alignments manually. All trees were generated
146 using FastTree (approximately-maximum likelihood method) with default parameters
147 (21).

148 We generated the rooted phylogenetic tree of the Mlps as follows. We first compiled
149 the sequences of Mlps1, Mlps2, and Mlps3 from type I MGCs included in the above SSN.
150 We added the SCMTs identified in the unrooted tree to this list, as well as the known
151 phenolic *O*-demethylases OdmB and VdmB (UniProt IDs: O87604 and Q6W001,
152 respectively). The SCMTs were identified as the smallest set of branches containing the
153 characterized SCMTs MtaA, MtbA, and MtsA (all from *Methanosarcina barkeri* Fusaro),
154 but excluding the Mlps3. The SCMTs were identified as the smallest clade in the unrooted
155 tree which contained the known SCMTs MtaA, MtbA, and MtsA. The tree was rooted on
156 the branch connecting the *O*-demethylases to the rest of the sequences, based on the
157 hypothesis that they serve as an appropriate outgroup.

158

159 <u>Phylogenetic membership of type I MGC-encoding organisms</u>

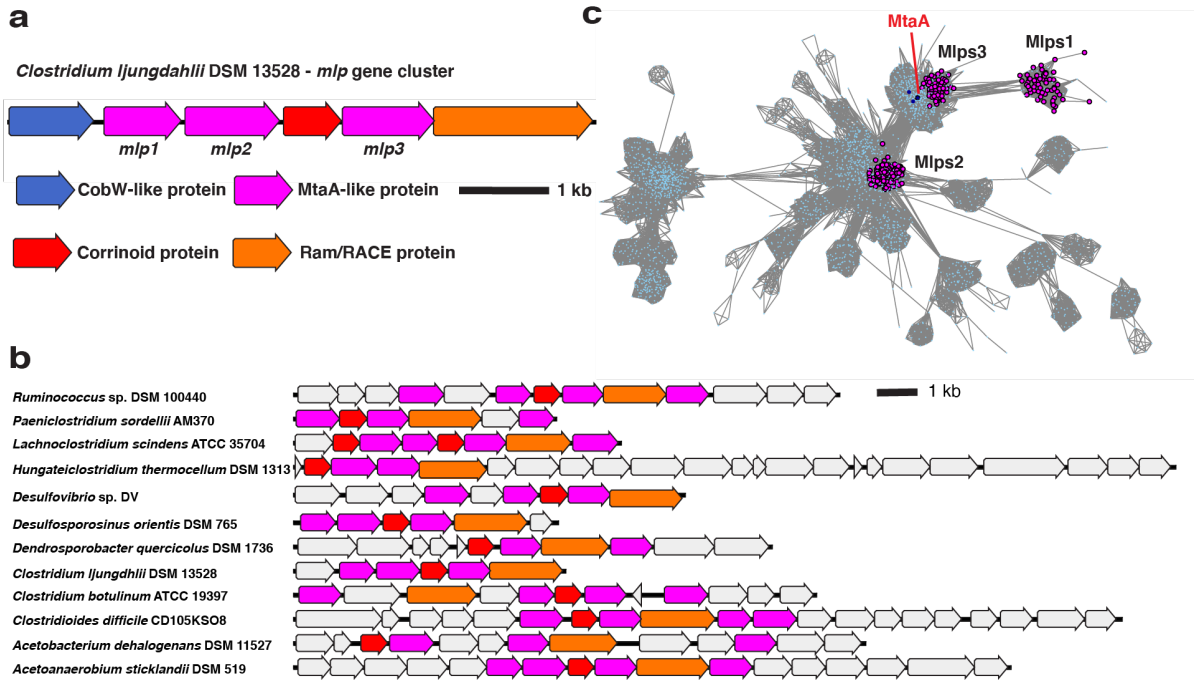
160 We analyzed the phylogenic composition of type I MGC-encoding bacteria and
161 archaea according to the Genomes Taxonomy Database (GTDB) taxonomic system (22).

162

163 <u>Construction of homology models</u>

164 Clostridium ljungdahlii DSM 13528 Mlps 1, 2, and 3 were submitted to the SWISS-
165 MODEL webserver (23) using *Methanosarcina mazei* MtaA (PDB ID: 4AY7) (24).
166
167 **Results**
168
169 While manually inspecting the genome of *Clostridium ljungdahlii* DSM 13528, a
170 bacterium that has received attention as a potential producer of biofuels, we noticed that
171 this organism encodes multiple members of the UroD superfamily. To see if these genes
172 are closely related to SCMTs, we performed BLASTP searches of the SCMT MtbA
173 (*Methanosarcina barkeri* Fusaro, *Mb*MtbA) against the predicted *C. lungdahlii* DSM 13528
174 proteome. We identified seven hits with an E value lower than 1e-09 (Table S1). The
175 highest hit was CLJU_RS09375, which showed 32.5% amino acid (aa) identity to
176 *Mb*MtbA. It is worth noting that the SCMTs MtbA, MtaA, and MtsA (all from *M. barkeri*
177 Fusaro) show only about 30% aa identity to each other, suggesting that CLJU_RS09375
178 could be a close relative of the SCMTs. CLJU_RS09375's amino acid sequence identity
179 to the UroD superfamily phenolic *O*-demethylases OdmB and VdmB was only 17.6% and
180 17.0%, respectively.
181 To obtain clues as to the physiological role of CLJU_RS09375, we next examined
182 this gene's genomic context. CLJU_RS09375 is encoded in a gene cluster that includes
183 genes encoding a corrinoid protein (CLJU_RS09370), a "bacterial-type" corrinoid protein
184 reductive activase (Ram/RACE protein) (CLJU_RS09380), and two additional Mlps
185 (CLJU_RS09360 and CLJU_RS09365) (Fig. 2a). This gene cluster has been

186 independently predicted to form an operon by OperonDB, an automatic operon prediction
187 database (25). Because this gene cluster encodes multiple Mlps, we named it the *mlp*
188 gene cluster (MGC). We named the three Mlps in the gene cluster Mlp1, Mlp2, and Mlp3,
189 (specifically, *Cl*Mlp1, 2, and 3) in order of their appearance in the gene cluster from the
190 presumed transcriptional start site (see Fig. 2a).



191

192 **Fig. 2 Type I MGCs encode orthologous proteins. a** The MGC from *C. ljungdahlii*. **b** Representative
193 type I MGC. Coloring of the core MGC genes is the same as in a, but non-core genes are colored grey. **c**
194 A protein SSN of the UroD superfamily (IPR000257) showing the clustering of Mlps1, 2, and 3. Nodes
195 representing MtsA, MtbA, and MtaA (all from *M. barkeri* Fusaro) are shown in dark blue. The alignment
196 score cutoff was 40 and the node representation percent cut was 50 % aa identity. The SSN shown is a
197 trimmed version of the full SSN, with all nodes removed that are neither directly or indirectly connected to
198 the Mlps1, 2, or 3.

199

200 Given the relatively high sequence similarity of the MGC Mlps to SCMTs, it seems
201 reasonable to hypothesize that these enzymes are corrinoid methyltransferases. This is
202 consistent with the presence of genes encoding a corrinoid protein and a Ram/RACE
203 protein in the MGC. The former is a methyl group carrier (26), while the latter is an ATP-
204 dependent corrinoid protein reductase that reductively activates the corrinoid protein to
205 its catalytically-active Co(I) state (27,28). Corrinoid methyltransferases, corrinoid
206 proteins, and Ram/RACE proteins are frequently found colocalized in gene clusters (29–
207 31). In bacteria, gene clusters encoding components of corrinoid methyltransferase
208 pathways are involved in methylotrophic metabolism and participate in transferring a
209 methyl group from a donor substrate to tetrahydrofolate (H_4folate) via a methyl-corrinoid
210 protein intermediate, a reaction catalyzed by a methyl-corrinoid protein:H_4folate
211 methyltransferase (15,29). In most of the bacterial methyltransferase pathways examined
212 to date, the H_4folate methylase is encoded in the same gene cluster as the corrinoid
213 protein and methyl donor methyltransferase (29,30). It is therefore surprising that H_4folate
214 methyltransferase is missing from the *C. ljungdahlii* MGC. While it is possible that a
215 H_4folate methyltransferase encoded elsewhere in the *C. ljungdahlii* genome could
216 cooperate with the MGC methyltransferase system, it is also possible that H_4folate is not
217 involved in the MGC methyltransferase pathway or that one of the MGC Mlps is an
218 undiscovered H_4folate methyltransferase.

219 In parallel with the identification of the *C. ljungdahlii* MGC, we located similar gene
220 clusters in *C. sporogenes* DSM 795 and *Hungateiclostridium* (*Clostridium*) *thermocellum*
221 DSM 1313. Like the MGC from *C. ljungdahlii*, the gene clusters in *C. sporogenes* and *H.*

222 *thermocellum* encode multiple Mlps (each with about 30% identity to the SCMTs), a
223 corrinoid protein and a Ram/RACE protein (Fig. 2b). The gene clusters from *C.*
224 *sporogenes* and *H. thermocellum* also lack a H_4folate methylase. However, unlike *C.*
225 *ljungdahlii*, which encodes H_4folate methylases elsewhere in its genome, homologs of
226 H_4folate methylases are absent from sequenced *C. sporogenes* and *H. thermocellum*
227 genomes, suggesting that H_4folate might not be involved in the presumptive
228 methyltransferase pathways encoded by these gene clusters. Because of the similarity of
229 the *C. sporogenes* and *H. thermocellum* gene clusters to the *C. ljungdahlii* MGC, we refer
230 to these gene clusters as the *Cs*MGC, *Ht*MGC, and *Cl*MGC, respectively.

231 We next wondered if related MGCs are found in organisms other than *C. ljungdahlii*,
232 *C. sporogenes*, and *H. thermocellum*. To address this, we BLASTed *Cl*Mlp1, 2, and 3
233 against the NCBI non-redundant protein database. We notice that many of the organisms
234 that had high alignment score hits for *Cl*Mlp1 also encoded high alignment score hits for
235 *Cl*Mlp2 and 3. When we examined the genomic contexts of these hits, we found that they
236 are indeed colocalized in gene clusters that resemble the *C. ljungdahlii* MGC (see Fig. 2b
237 for examples).

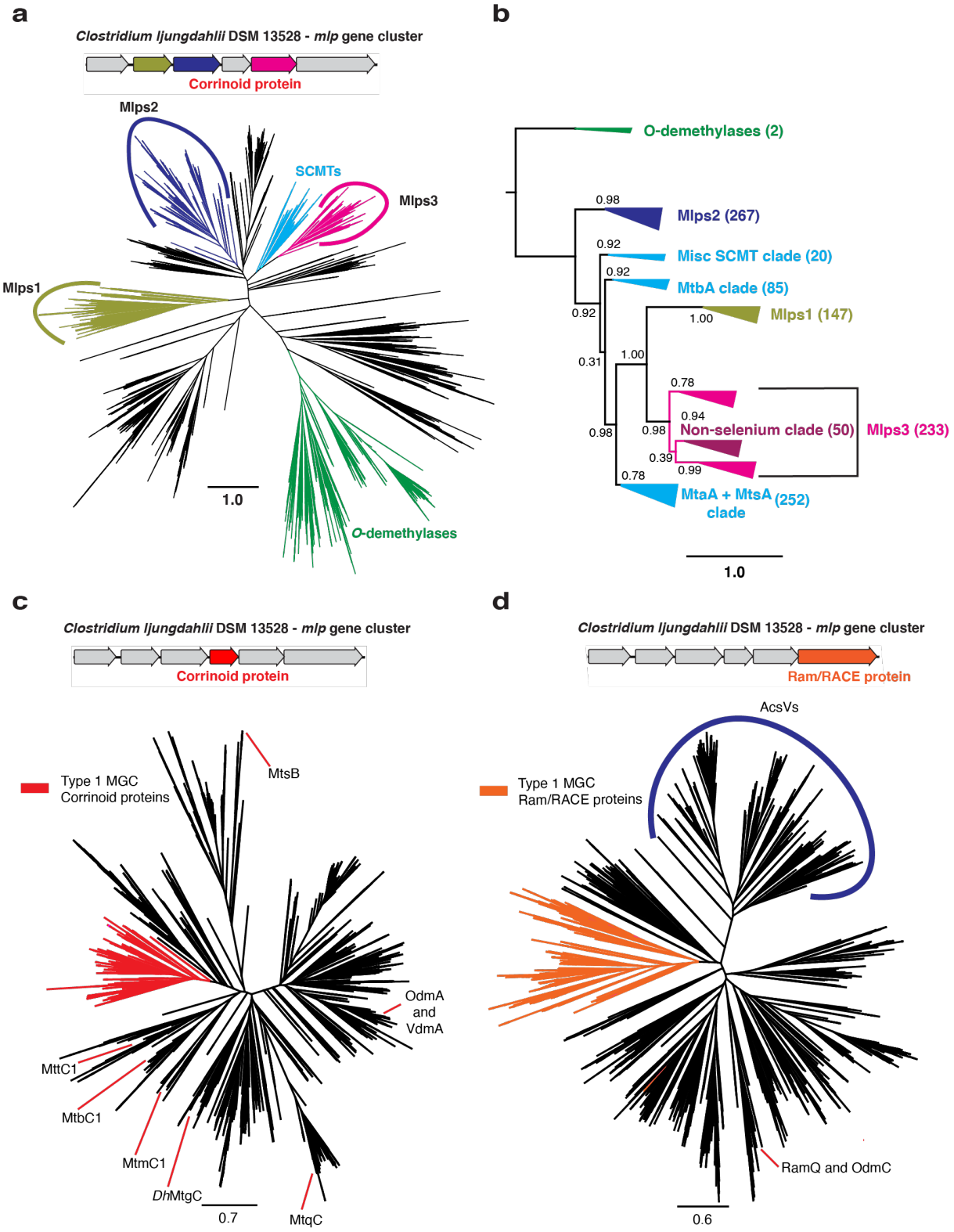
238 We next asked if these MGCs are homologous or analogous. Because components
239 of a corrinoid methyltransferase pathway are commonly encoded in gene clusters
240 (14,15,29–31), it would not be surprising to find that evolutionary pressure to colocalize
241 corrinoid methyltransferase genes resulted in MGCs that appear only superficially similar,
242 but share no evolutionary relationship. Alternatively, the common appearance and genetic
243 content of the MGCs could suggest that these gene clusters are descended from a

244 common ancestor (i.e. are homologous), which would also suggest they likely have a
245 shared function in diverse organisms. To address this question, we first compiled a list of
246 367 organisms encoding likely MGCs (see materials and methods). To preliminarily
247 assess the similarity of the proteins encoded by these MGCs, we constructed several
248 protein sequence similarity networks (SSNs) (19). We reasoned that if MGCs are
249 homologous, each conserved component of the identified MGCs should preferentially
250 form its own cluster in an SSN of the protein family to which that component belongs. For
251 instance, in an SSN of the corrinoid protein superfamily, the MGC corrinoid proteins
252 should form their own cluster within the SSN. The same should be true for the Ram/RACE
253 proteins in an SSN of their superfamily. Because there are multiple *mlp* genes in each
254 MGC, there are two SSN analysis outcomes that are consistent with the homology
255 hypothesis. First, all of the Mlps from the MGCs could be grouped within a single cluster
256 in the SSN. This would be expected to be the case if the multiple *mlp* genes in each MGC
257 arose from gene duplication. Second, the MGC Mlps could each represent different
258 lineages, and could therefore likely form distinct clusters within an SSN.

259 We first constructed an SSN (Fig. 2c) of the IPR000257 superfamily, which includes
260 the Mlps. We found that 80.9% of the MGC Mlps localized into one of three sub-clusters
261 within a larger cluster (Fig. 2c, three clusters of magenta-colored nodes). Interestingly,
262 these three sub-clusters correspond to the three Mlps from the *C. ljungdahlii* MGC, with
263 *Cl*Mlp1, 2, and 3 being grouped into one of each of the sub-clusters. Inspection of other
264 MGCs reveal that almost of these clusters encode Mlps from clusters 2 and 3, while a
265 gene encoding an Mlp from cluster 1 was present a majority of the time, but often absent.

266 We denote the proteins in the sub-cluster containing *Cl*Mlp1 as Mlps1, and those from
267 network sub-clusters containing *Cl*Mlp2 and 3 as Mlps2, and Mlps3, respectively.
268 Altogether, this analysis shows that a large number of MGCs encode at least an Mlp2
269 and an Mlp3. We designate such gene clusters as "type 1 MGCs".

270 Because the overwhelming majority (>95%) of MGCs examined here appear to be
271 type 1 MGCs, we selected the type I MGCs for further study. The fact that the type 1 MGC
272 Mlps fall into three separate clusters in the SSN hints that they might represent three
273 distinct Mlp lineages, rather than a single lineage within the UroD superfamily that arose
274 from recent gene duplication events. To solidify this inference, we constructed a
275 phylogenetic tree of the Mlps and other members of the UroD superfamily that are close
276 relatives of known corrinoid methyltransferases (Figure 3a). The phylogenetic tree
277 showed that the Mlps 1, 2, and 3 are situated in a large branch that includes the SCMTs,
278 to the exclusion of other known UroD superfamily corrinoid methyltransferases such as
279 the aromatic *O*-demthylases (13). When rooted (Figure 3b), we found that the Mlps 1, 2
280 and 3 formed distinct, well-supported clades. Interestingly, the Mlps2 form a sister clade
281 to the clade containing the SCMTs and the Mlps1 and Mlps3. Within this clade, the Mlps1
282 and Mlps3 form sister clades emerging from the SCMTs, which therefore appear to be
283 paraphyletic.



284

285 **Fig. 3 Type I MGCs encode orthologous proteins. a** An approximately-maximum likelihood (FastTree)

286 phylogenetic tree of members of the UroD superfamily methyltransferases. The FastTree support values

287 for the Mlps1, 2, and 3 clades are 0.95, 0.94, and 1.00, respectively. Scale bar: branch length corresponding
288 to an average of one change per position. **b** A rooted approximately-maximum likelihood (FastTree)
289 phylogenetic tree of selected protein sequences. Shown are the soluble coenzyme M methyltransferases
290 (SCMTs), the Mlps1, 2 and 3, and the phenolic *O*-demethylases OdmB and VdmB (outgroup). Each branch
291 is labeled with the branch's FastTree support value. The number of sequences in each clade are as follows.
292 *O*-demethylases: 2; Mlps2: 267; Misc. SCMT clade: 20; MtbA clade: 85; Mlps1: 147; Mlps3: 233; MtaA +
293 MtsA clade: 252. **c** An approximately-maximum likelihood (FastTree) phylogenetic tree of members of the
294 methyltransferase corrinoid protein family. The type I MGC proteins are shown in red. The FastTree support
295 value for the type I MGC proteins is 0.92. Selected characterized family members are shown: MtbC1,
296 MtmC1, and MttC1 (all from *M. acetivorans* C2A); *Dh*MtgC from *Desulfitobacterium hafniense* Y51; MtqC
297 from *Eubacterium limosum* ATCC 8486; MtsB from *M. barkeri* Fusaro; OdmA and VdmA from
298 *Acetobacterium dehalogenans* DSM 11527. **d** An approximately-maximum likelihood (FastTree)
299 phylogenetic tree of members of the Ram/RACE protein family. The type I MGC proteins are shown in
300 orange. The FastTree support value for the type I MGC proteins is 0.98. A clade containing the corrinoid
301 iron-sulfur protein reductive activases is labeled "AcsVs". The locations of the activases RamQ (*E. limosum*
302 ATCC 8486) and OdmC (*A. dehalogenans* DSM 11527) are shown. Scale bars for all tree represent
303 respective branch length corresponding to an average of one change per position.
304

305 To further explore the sequence similarity relationships between MGCs, we first
306 generated SSNs for the corrinoid protein (IPR006158) and Ram/RACE protein
307 (IPR027980) superfamilies. These SSNs were similar: each contained a cluster
308 consisting of proteins from MGCs (Fig. S2). The only exceptions amongst corrinoid
309 proteins were found in cases where two corrinoid proteins were present in the MGC, one
310 of which was located outside of the type I MGC network cluster. We refer to this second
311 corrinoid protein as an "auxiliary." A comparatively small number (13 in total) of type I

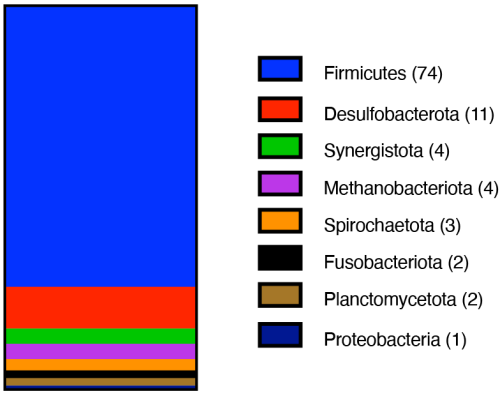
312 MGC Ram/RACE proteins were found outside the main network cluster. Interestingly, all
313 but two of these non-canonical Ram/RACE proteins were found in MGC that also
314 encoded a non-pyrrolysine homolog of the methylamine methyltranasferase, MtmB. Ten
315 of these MGCs encoded an auxiliary corrinoid protein. These data are consistent with
316 the hypothesis that type 1 MGCs are homologous to each other, but not confirmatory, as
317 they do not use phylogenetic algorithms. To further test this hypothesis, we generated
318 two additional phylogenetic trees: one containing the type 1 MGC corrinoid proteins and
319 other monomeric methyltransferase corrinoid proteins (Fig. 3c and 3d, respectively), and
320 the other containing the type 1 MGC Ram/RACE proteins and other Ram/RACE proteins
321 (Fig. S4). In each case, the type 1 MGC proteins were found in the same branch, with a
322 handful of exceptions. These exceptions could be the result of gene conversion or the
323 replacement of a "canonical" type 1 MGC gene with an isofunctional gene from outside
324 of the MGC.

325 Together, the data from the SSN and phylogenetic analyses suggests that type 1
326 MGCs are homologous in that they encode mutually orthologous proteins (i.e. the
327 conserved type 1 MGC Mlps are all orthologs, the type 1 MGC corrinoid proteins are all
328 orthologs, and the type 1 MGC Ram/RACE proteins are all orthologs.) The simplest
329 explanation for the homology of the type 1 MGCs is that they are preserved and
330 horizontally transferred as a unit. However, we cannot rule out more complicated
331 scenarios such as one in which the type 1 MGC genes are transferred individually into
332 different regions of the genome and then brought together by selective pressure. This

333 latter possibility gains some credence when we note the lack of conserved synteny in type
334 1 MGCs.

335 Having established that type 1 MGCs are homologous, we sought to analyze the
336 phylogenetic distribution of these gene clusters (Fig. 4). To date, we have identified over
337 one hundred genera (all of them prokaryotes) containing at least one strain encoding a
338 type 1 MGC. Of these genera, only three (*Methanococcus*, *Methanothermococcus*, and
339 *Methanotorris*) are archaeal. These three archaeal genera are of methanogens belonging
340 to the Methanococcales order. This is noteworthy because members of the
341 Methanococcales are not known to be methylotrophic and therefore would not be
342 expected to encode SCMTs. Of genera that encode type I MGCs, all but two
343 (*Rhodoblastus* and *Rhodovastum*) contain only obligate anaerobes, with the majority
344 (73%) belonging to the Firmicutes phylum and a minor contribution from the
345 *Desulfobacterota* phylum (composing 11% of the type I MGC-encoding genera). Although
346 type 1 MGCs are found largely in a limited number of phyla, they are widely distributed in
347 those phyla at the order level, as type 1 MGCs are found in at least 35 orders of bacteria
348 and archaea. The two most well represented orders are the *Peptostreptococcales* and
349 *Clostridiales*, which contain 14 and 13 identified type 1 MGC-encoding genera,
350 respectively. From this analysis, we conclude that type 1 MGC are phylogenetically
351 widespread and are associated with microorganisms that are restricted to anoxic
352 environments. Individual species encoding type 1 MGCs include a number of interesting
353 organisms, such as toxigenic clostridia (*Clostridium botulinum*, *Paeniclostridium sordelii*,
354 certain strains of *Clostridioides difficile*) and industrially-relevant clostridia

355 (*Hungateiclostridium* (*Clostridium*) *thermocellum*, *Clostridium ljungdahlii* and its close
356 relatives).
357



358
359
360 **Fig. 4 Type I MGCs are found in several prokaryotic Phyla.** Phylum composition of genera encoding
361 type I MGCs. 101 genera are represented. The number of type I MGC-encoding genera within each phylum
362 is enclosed in parentheses.
363
364 Many of the organisms encoding type I MGCs are either known acetogens or encode
365 the proteins necessary for acetogenesis. These include the acetogenic *Clostridium* spp.,
366 *Sporomusa* spp., certain *Clostridioides difficile* strains, and acetogenic *Treponema*
367 strains. It should be emphasized, however, that many known acetogens do not encode
368 MGCs and many of the organisms that encode MGCs lack the critical *acsA* and *acsB*
369 (genes encoding bacterial anaerobic carbon monoxide dehydrogenase and bacterial
370 acetyl-CoA synthase, respectively) that are required for acetogenesis in all known
371 instances.

372 We next turned our attention to analyzing additional components of the type 1 MGCs
373 and proposing a biochemical function for these genes. The UroD superfamily includes
374 corrinoid methyltransferases that (de)methylate several heteroatoms including halogens,
375 sulfur atoms, and aryl oxygens (10,12,14,15,32). Several lines of evidence suggest that
376 the type 1 MGCs encode a corrinoid *S*- or *Se*-methyltransferase system. First, the closest
377 relationship between known UroD superfamily proteins and the type 1 *mlp* cluster Mlps is
378 with the SCMTs (Fig. a and b), which are *S*-methyltransferases. Additional evidence
379 comes from examination of the enzyme's Zn^{2+}-binding site. This Zn^{2+} cofactor coordinates
380 the heteroatom of a substrate, activating it for (de)methylation (24,32,33). The primary
381 coordination sphere of the Zn^{2+} cofactor is His-Cys-Cys in *S*-methyltransferases from the
382 UroD superfamily and beyond, while in *O*-methyltransferases, one or both of the Cys
383 ligands are substituted by carboxylates (Glu or Asp) or are substituted for non-
384 coordinating residues (15,24,32,33). The His-Cys-Cys coordination environment favors
385 the binding of a soft ligand such as a selenium or sulfur atom to the Zn^{2+} center, while
386 substitution of Cys for a carboxylate favors binding of an oxygen atom. The coordination
387 sphere of the type 1 MGC Mlps is (with the exception of a few of the Mlps1) invariably
388 His-Cys-Cys, suggesting that these proteins are *S*- or *Se*-methyltransferases. We
389 constructed homology models of each of the *C. ljungdahlii* Mlps in order to confirm that
390 the His-Cys-Cys triads are likely appropriately positioned to coordinate Zn^{2+} (Fig. S3).
391 Finally, examination of the non-core components of type I MGCs (genes that are found in
392 type I MGCs, but do not encode Mlps, corrinoid proteins, or Ram/RACE proteins)
393 consistently reveals genes involved in sulfur metabolism, including methionine

394 aminotransferases, sulfurtransferases, methionine transporters, and close homologs of
395 the recently-described MarHDK enzyme that catalyzes C–S bond cleavage in the
396 production of ethylene from methylthioethanol (34). These observations suggest a
397 potential role for type I MGCs in methyl-sulfur metabolism.

398 However, consideration of additional evidence suggests that most type I MGCs
399 function in methyl-selenium metabolism. For instance, type 1 MGCs commonly contain
400 genes provisionally linked to selenium metabolism, such as the recently characterized
401 "dicubane cluster protein" (35–37). Several type 1 MGCs (Fig. 5) contain genes encoding
402 authentic machinery for selenocysteine synthesis, encoding, and decoding (SelABCD),
403 as well as the enzymes of the glycine reductase system, two of which have
404 selenocysteine residues (38,39). A strikingly high percentage (87%) of sequenced
405 species that encode type I MGCs also encode selenocysteine biosynthesis machinery
406 within the same genome. For comparison, a recent study found that only about 21.5% of
407 bacteria contain the requisite genes for selenocysteine biosynthesis, and even in the
408 selenocysteine-rich Clostridia, only 42% of 184 genomes surveyed encoded
409 selenocysteine biosynthesis machinery (35). Not all organisms that make use of selenium
410 encode selenocysteine biosynthetic machinery—some organisms that use the selenium
411 cofactor only encode selenide-water dikinase (SelD) (35). We therefore manually
412 inspected genomes that contain a type I MGC, but lack SelAB or C, finding that several
413 genomes lacked *selABC*, but had *selD* as well as selenium cofactor-dependent enzymes.
414 As noted below, type 1 MGCs from *C. ljungdahlii* and *C. drakei* are upregulated during
415 autotrophic growth on H_2 + CO_2, a condition where there is increased demand for

416 selenoprotein biosynthesis (40). The Cys-Cys-His coordination of the active site Zn^{2+}, in
417 addition to being suggestive of binding a sulfur nucleophile/methyl-sulfur methyl donor,
418 could also be compatible with a selenium nucleophile/methyl-selenide methyl donor. Most
419 importantly, during the preparation of this manuscript, we became aware of a 2004
420 publication that used proteomic and genetic methods to implicate three type I MGC-
421 encoded proteins (the corrinoid protein and two Mlps2) in the utilization of dimethyl-
422 selenide as a selenium source in the hydrogenotrophic methanogen *Methanococcus*
423 *voltae* (41). The authors suggested that the Mlps might demethylate (di)methyl-selenide
424 and transfer the methyl group onto the MGC corrinoid protein, liberating selenide for
425 incorporation into biological macromolecules. While the authors did not biochemically
426 characterize the Mlps in this study, their data strongly support a role for the type I MGCs
427 in methyl-selenol utilization.

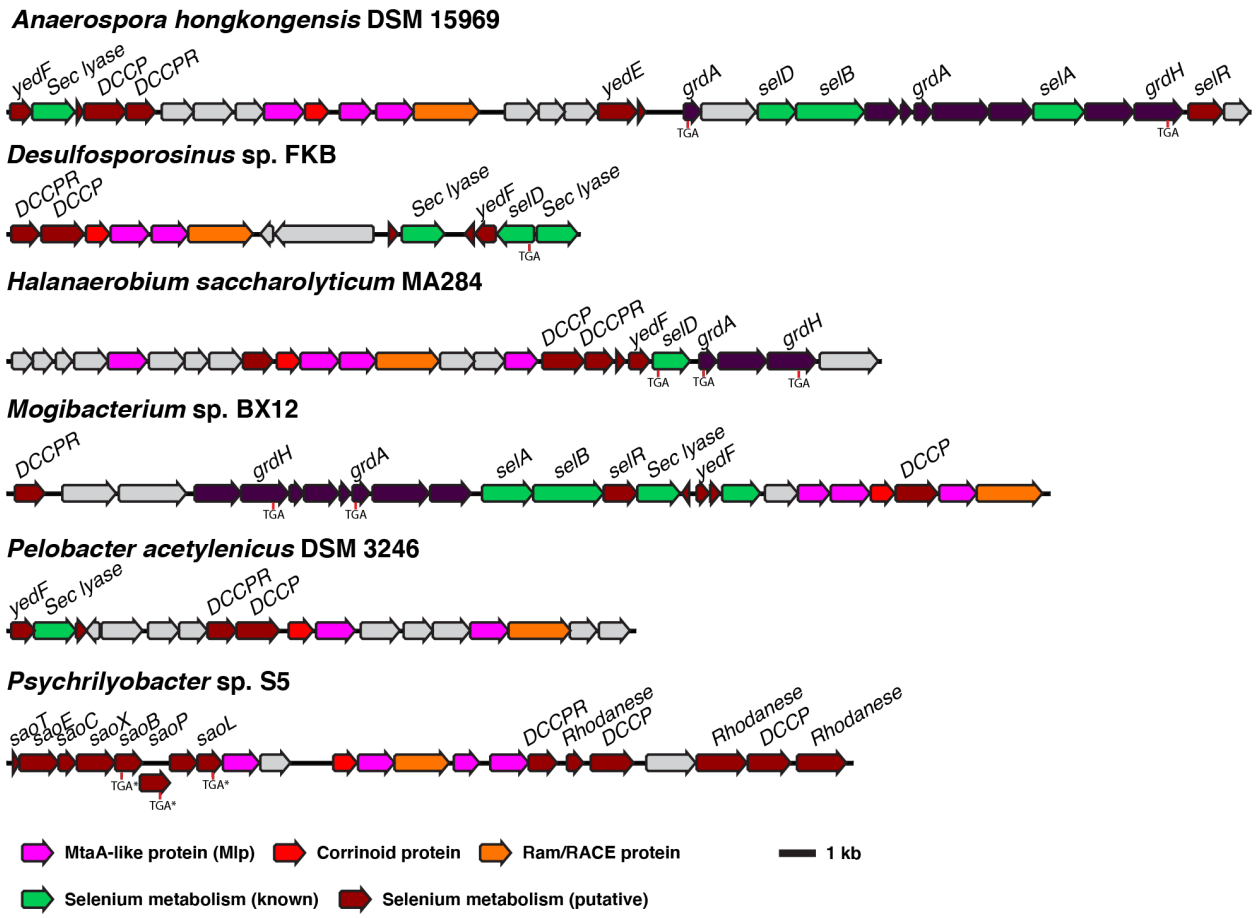


Fig. 5 Type I MGCs often encode genes involved in selenium metabolism. Selenium metabolism gene show are either close homologs of genes with experimentally-confirmed roles in selenium metabolism ("Selenium metabolism (known)") or have been associated with selenium metabolism by bioinformatic methods ("Selenium metabolism (putative)") (35,36). The positions of selenocysteine codons are make with "TGA." "TGA*" indicates a selenocysteine codon that has been bioinformatically predicted, but not experimentally confirmed (36). Sec lyase: selenocysteine lyase. DCCP: dicubane cluster protein. DCCPR: dicubane cluster protein reductase. grdA: glycine reductase subunit A. grdH: glycine betaine reductase subunit H. selA: *L*-seryl-tRNA(Sec) selenium transferase. selB: selenocysteine-specific translation elongation factor. selD: selenide water dikinase. selR: putative selenium transcriptional regulator (35). Sao: putative "selenocysteine-assisted organometallic operon genes.

440 Because type 1 MGCs are present in several model organisms that have been the
441 subjects of proteomic and transcriptomic studies, we examined the expression of these
442 genes in existing datasets for *C. botulinum*, *C. ljungdahlii*, and *H. thermocellum* to gain
443 additional information regarding their functions. In all three organisms, the MGCs are
444 expressed at a detectable level under all conditions examined. First, we found that the *C.*
445 *botulinum* MGC is upregulated during growth at the organism's optimal growth
446 temperature of 37 °C compared to growth at 15 °C (transcripts 5.27-68.6 fold upregulated
447 at 37 °C) (42). In *C. ljungdahlii*, we found that the type 1 MGC is transcriptionally
448 upregulated roughly 3-5 fold during autotrophic growth on H_2 + CO_2 or CO versus
449 heterotrophic growth on fructose (43–45). The same pattern is found in *C. drakei* (6 to13
450 fold transcriptional upregulation on H_2 + CO_2 compared to fructose) (40). In *H.*
451 *thermocellum*, the type 1 MGC is upregulated during growth with supplemental acetate in
452 a strain from which the FeFe hydrogenase maturation enzyme *ΔhydG* had been deleted
453 (46). These findings imply that type 1 MGCs are expressed, indicating their physiological
454 relevance, and respond to environmental and genetic perturbations in industrially and
455 medically significant organisms.
456
457 **Discussion**
458
459 Methylotrophic methanogenesis is a unique mode of metabolism restricted to a subset
460 of the archaea, so it is surprising to find homologs of enzymes from this pathway in non-
461 methanogens. Here, we identify a diversified, widespread set of gene clusters—type 1

462 MGCs—that encode homologs of the SCMTs involved in methylotrophic
463 methanogenesis. Type I MGCs are found exclusively in anaerobes, and are most
464 common in the Firmicutes phylum. In these organisms, there is a strong tendency for type
465 I MGCs to be found in genomes that also contain selenocysteine biosynthesis and
466 decoding machinery.

467 What reactions do the enzymes encoded by type 1 MGCs catalyze and what is the
468 physiological function of those reactions? The presence of a corrinoid protein, a corrinoid
469 reductive activase, and homologs of corrinoid methyltransferases (Mlps) implies that the
470 type 1 MGCs encode components of a corrinoid methyltransferase pathway. Protein
471 sequence comparisons indicate that the Mlps are likely *S*- or *Se*-methyltransferases that
472 (de)methylate thiol and selenol metabolites.

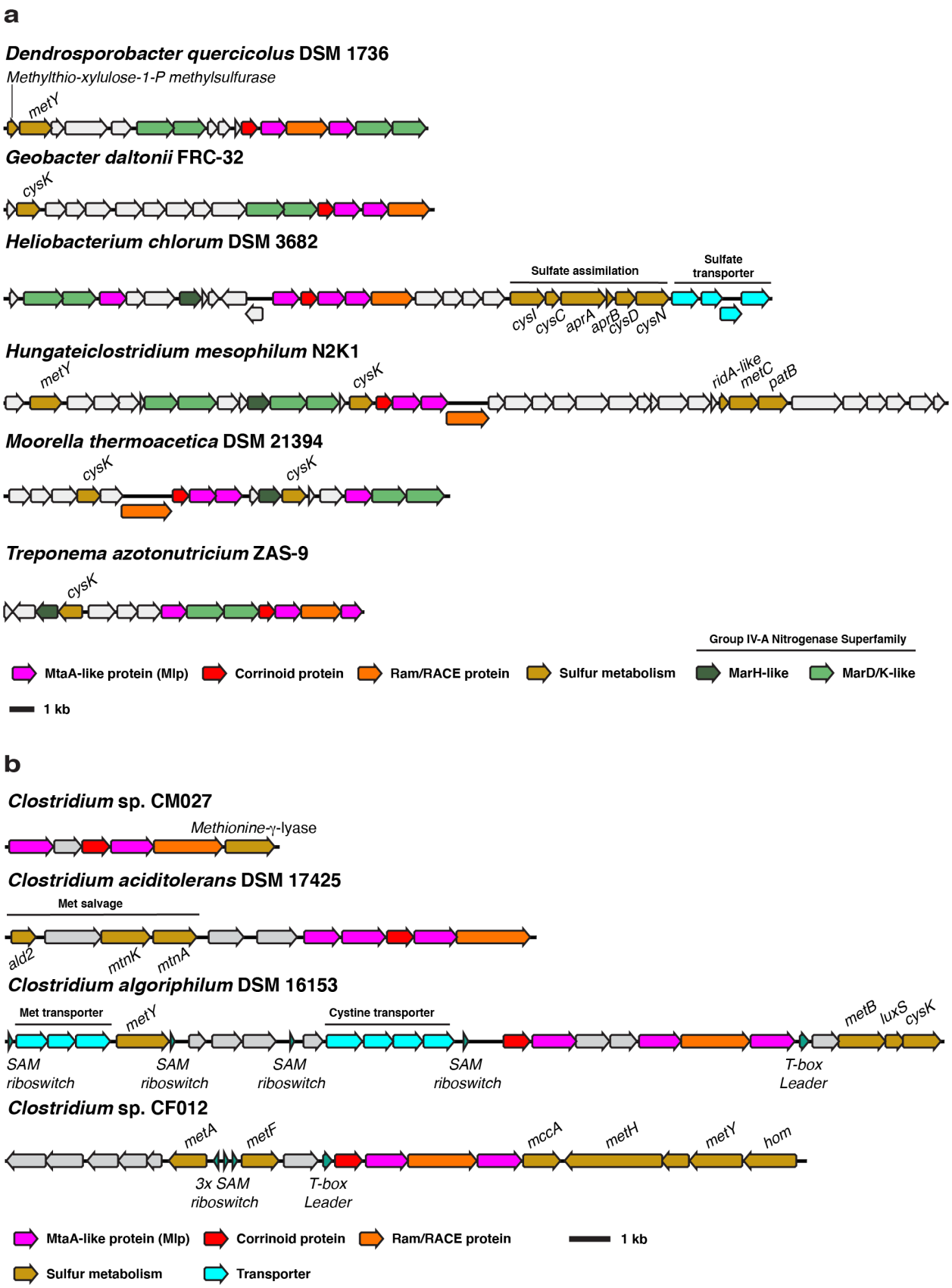
473 The potential physiological roles of such reactivity may be explored through
474 considering features of the organisms that possess type I MGCs. It seems clear from
475 earlier studies that type I MGCs can function in methyl-selenol valorization (39). However,
476 those authors implicated only the corrinoid protein and two closely-related Mlps2 in
477 dimethyl-selenide utilization—the more distantly related *M. voltae* Mlp1 and Mlp3 (which
478 are also present in the *M. voltae* type I MGC) appear to have gone unnoticed. It seems
479 likely that the full type I MGC contains methyltransferases for demethylation of several
480 methyl-selenium compounds. While the genetic data in Ref. 39 implicates dimethyl-
481 selenide and/or methyl-selenol as substates of the type I MGC Mlps2, we propose that
482 the Mlps1 and Mlps3 may be responsible for demethylation of other methyl-selenium
483 compounds such as methyl-selenocysteine and selenomethionine and concomitant

484 methylation of the gene clusters' encoded corrinoid proteins. The methylated corrinoid
485 protein must be demethylated and the methyl group transferred to a methyl-accepting
486 metabolite if the pathway is to continue to operate. Accordingly, we propose that the
487 methyl-selenide-derived methyl group is subsequently transferred from the methylated
488 corrinoid protein to a thiol containing compound such as sulfide. The resulting methyl-
489 sulfur metabolite could then be excreted or recycled into methionine (or, in the type I MGC
490 encoders from the Methanococcales, Mlp3 could methylate coenzyme M, making a
491 presumably small contribution to methane production). Methyl-selenium of various forms
492 are present in various environments, including (based on dietary considerations) the
493 human gut (47). It may be that most type I MGCs function to valorize these methyl-
494 selenium compounds. Nonetheless, we believe an additional explanation is required, as
495 many organisms (such as *C. ljungdahlii* and *M. voltae*) upregulate their type I MGCs even
496 when methyl-selenium compounds are not added to growth media. This observation
497 suggests that there may be an endogenous source of methyl-selenium compounds.
498 Indeed, it has been suggested and evidentially supported that the selenide generated as
499 an intermediate in selenoprotein biosynthesis can react non-enzymatically with *S*-
500 adenosyl-methionine (SAM) (48). This presents a problem for a cell that has a high
501 demand for selenium metabolism: methyl selenide formation from SAM both wastes
502 valuable selenium and leads to a potentially toxic dead-end metabolite. The enzymes
503 encoded by type I MGCs could hypothetically solve this quandary by recycling the methyl
504 selenide into methionine.

505 Some organisms, such as *H. thermocellum* encode and express type I MGCs, but lack
506 genes encoding known selenoenzymes or other indicators of selenium metabolism. On
507 further inspection, we found that all of the selenium-trait deficient organisms have Mlps3
508 that fall into a single clade of the Mlp phylogenetic tree (Fig. 3b). We refer this this clade
509 as the "non-selenium" clade, and suggest that the corresponding type I MGCs are
510 involved in methyl-sulfur metabolism. In fact, interrogation of the Mlp3 sequences from
511 *Sporomusa ovata* indicates that this organism encodes two separate type I MGCs, one
512 with an Mlp3 from the "non-selenium" clade, and one with an Mlp from elsewhere in the
513 Mlp phylogenetic tree. It therefore seems that type I MGCs have at least two presumably
514 distinct functions. One group, which makes up more than 90% of the type I MGCs profiled
515 in this study, is likely involved in methyl-selenium metabolism, while the other, which
516 encodes Mlps3 from the non-selenium clade, probably has a different, selenium-
517 independent function.

518 At present, evidence from genomic context suggests that the non-selenium MGCs
519 could be involved in a methionine salvage pathway, defined as a metabolic pathway that
520 converts methylthio-adenosine (MTA) into methionine (49). MTA is a toxic by-product of
521 the biosynthesis of key metabolites derived from *S*-adenosylmethionine (SAM) including
522 polyamines, *N*-acyl serine homolactones, and secondary metabolites (50). Methionine
523 salvage pathways therefore support sulfur metabolism both by detoxifying a by-product
524 and preserving the metabolically expensive methylthiol group of MTA (51). The so-called
525 "universal" methionine salvage pathway is O_2-dependent, and methionine salvage
526 pathways that operate under anaerobic conditions have only recently begun to be

527 discovered (49). The anaerobic methionine salvage pathway from *Rhodospirillum rubrum*
528 was recently shown to dependent on the methylthio-alkane reductase MarHDK system,
529 homologs of the nitrogenase subunits NifHDK, respectively (34). *MarHDK*-like genes
530 belonging to the nitrogenase superfamily IV-A group are remarkably prevalent in or near
531 type I MGCs from the non-selenium clade and when not present in the type I MGC itself,
532 they are invariably present elsewhere in the genome in question (Fig. 6A). Although the
533 enzymatic activity of the IV-A group nitrogenase superfamily members is not known
534 (genetically characterized MarDK are in the IV-C group), the relatively high sequence
535 identity (~40-50% amino acid identity) of proteins from group IV-A to those in IV-C could
536 suggest a similar function in the reduction of carbon-sulfur bonds. This is consistent with
537 their typical genomic context, which usually includes various genes putatively involved in
538 sulfur metabolism. This suggests that the non-selenium type I MGCs encode enzymatic
539 machinery for either methionine salvage, sulfur valorization from methyl-sulfide
540 compounds, or both. This inference is further supported by the c. The *Clostridium*
541 *algoriphilum* DSM 16153 MGC (which encodes an Mlp3 from the non-selenium clade)
542 includes three predicted SAM riboswitches upstream of the MGC, as well as genes
543 predicted to encoded homocysteine synthase, cysteine synthase, a methionine ABC
544 transporter, and a *S*-ribosylhomocysteine lyase. *Clostridium* sp. CF012 possesses
545 another highly suggestive non-selenium MGC, which contains five SAM riboswitches, a
546 T-box leader, and a near-complete methionine biosynthetic pathway, including a
547 cobalamin-dependent methionine synthase (MetH).


a
Dendrosporobacter quercicolus DSM 1736
Methylthio-xylulose-1-P methylsulfurase
metY
Geobacter daltonii FRC-32
cysK
Heliobacterium chlorum DSM 3682
Sulfate assimilation
Sulfate transporter
cysI
cysC
aprA
aprB
cysD
cysN
Hungateiclostridium mesophilum N2K1
metY
cysK
ridA-like
metC
patB
Moorella thermoacetica DSM 21394
cysK
cysK
Treponema azotonutricium ZAS-9
cysK
Group IV-A Nitrogenase Superfamily
MtaA-like protein (Mlp)
Corrinoid protein
Ram/RACE protein
Sulfur metabolism
MarH-like
MarD/K-like
1 kb
b
Clostridium sp. CM027
Methionine-γ-lyase
Clostridium aciditolerans DSM 17425
Met salvage
ald2
mtnK
mtnA
Clostridium algoriphilum DSM 16153
Met transporter
metY
Cystine transporter
metB
luxS
cysK
SAM riboswitch
SAM riboswitch
SAM riboswitch
SAM riboswitch
T-box Leader
Clostridium sp. CF012
metA
metF
mccA
metH
metY
hom
3x SAM riboswitch
T-box Leader
MtaA-like protein (Mlp)
Corrinoid protein
Ram/RACE protein
1 kb
Sulfur metabolism
Transporter

548

549 **Fig. 6 Type I MGCs from encoding an Mlp3 from the non-selenium clade also encode putative sulfur**
550 **and methionine metabolism genes. a** Non-selenium type I MGCs associated with group IV-A nitrogenase
551 superfamily genes and other genes involved in sulfur metabolism. Methylthioribulose-1-P methylsulfurase
552 is a methionine salvage enzyme originally described in *Rhodospirillum rubrum* (52). **b** Other non-selenium
553 type I MGCs associated with methionine metabolism. MetY: *O*-acetyl-homoserine sulfhydrylase. CysK:
554 cysteine synthase. CysI: assimilatory sulfite reductase. CysC: adenylyl-sulfate kinase. AprAB:
555 phosphoadeninosine-phospho-sulfate reductase subunits A and B. CysND: sulfate adenylyltransferase
556 subunits 1 and 2. MetC cystathionine beta-lyase. PatB: alternate cystathionine beta-lyase. Ald2: methylthio-
557 ribulose-1-phosphate aldolase. MtnK: methylthio-ribose kinase. MtnA: methylthio-ribose-1-phosphate
558 isomerase. MetB: cystathionine gamma-synthase. LuxS: *S*-ribosyl-homocysteine lyase. MetA: homoserine-
559 *O*-succinyltransferase. MetF: methylene-tetrahydrofolate reductase. MccA: *O*-acetyl-serine dependent
560 cystathionine beta-synthase. MetH: cobalamin-dependent methionine synthase. Hom: homoserine
561 dehydrogenase. SAM: *S*-adenosyl-methionine.

562 As noted above, the sequences of the Mlps (in particular their Zn^{2+}-binding sites) are
563 compatible with (de)methylation of both methyl-sulfur compounds and of methyl-selenium
564 compounds. Additionally, the sporadic distribution of type I MGCs across phylogenetic
565 trees is potentially consistent with a role in methionine salvage, as such pathways are
566 dispensable under many growth conditions (51) and are known to be not strictly
567 conserved even in closely related strains. Indeed, many organisms lack a methionine
568 salvage pathway, and in *E. coli*, methionine salvage is a strain-specific feature (53).

569 We propose how a type I MGC-encoded methionine salvage pathway could operate
570 in Fig. 7, considering the simpler case of a type I MGC that encodes only two Mlps, such
571 as the gene cluster from *H. thermocellum*. We suggest that MTA is metabolized to an
572 unknown intermediate (R^1-SCH_3). One of the Mlps then catalyzes methyl transfer from
573 R^1-SCH_3 to the corrinoid protein. The second Mlp then transfers the methyl group from

574 the methylated corrinoid protein to a second metabolite, R^2-SH. The sulfur atom of R^1-SH
575 might be recycled into R^2-SH by an unknown pathway. R^2-SCH_3 could be converted into
576 methionine, or perhaps might be methionine. The MarHDK-like enzymes may have a role
577 in such a pathway, generating methyl-sulfide that might serve as the methyl donor for the
578 MGC methyltransferase system (R^1-SCH_3 with R^1 = H). Alternatively, the MarHDK-like
579 proteins might serve as part of a second, parallel methionine salvage pathway, or even
580 as a branch of a topologically bifurcating pathway with two routes to methionine.

581 Intriguingly, this proposed methionine salvage pathway uses a distinct biochemical
582 logic from related pathways. Characterized methionine salvage pathways all incorporate
583 the methylthiol group of MTA intact into methionine (34,49,50,53). In the proposed
584 methionine salvage pathway involving the type 1 MGCs, the methylthio-ether of MTA is
585 dismantled into a methyl group which is carried by the corrinoid protein, leaving the free
586 thiol behind. This pathway architecture theoretically avoids production of volatile methyl-
587 sulfur intermediates that could be more readily lost from the cell. The proposed role of a
588 corrinoid cofactor in methionine salvage is also novel. Ultimately, biochemical
589 characterization of the enzymes encoded by type I MGCs will be required to test this
590 proposal.

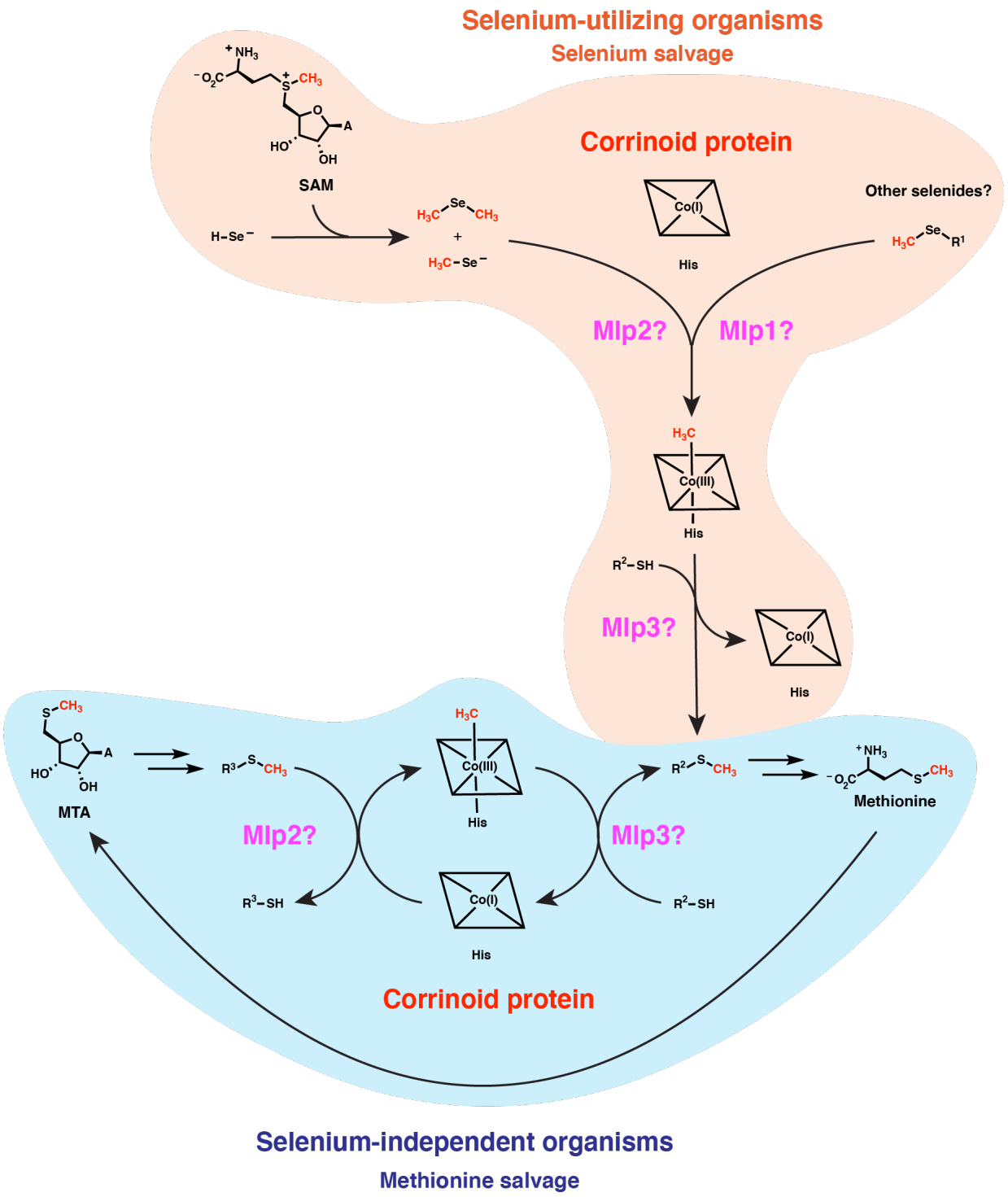


Fig. 7 Type I MGCs likely encode components of corrinoid-dependent selenium salvage or methionine salvage pathways. The vast majority of type I MGC encoders likely use these methyltransferase systems to recycle methylated selenides formed from the non-enzymatic reaction of SAM

595 with selenol compounds, potentially recycling the methyl group into methionine. A smaller subset of type I
596 MGC encoders in the ref seq database (about 10%) lack known selenium metabolism traits. We propose
597 that these MGCs have evolved to participate in a methionine salvage pathway. MTA: 5'-methylthio-
598 adenosine. His: histidine. A: adenine.
599

600 In summary, we have identified a class of widespread gene clusters encoding
601 components of probable corrinoid *S*- and *Se*-methyltransferase pathways. The type I
602 MGCs are found in phylogenetically and metabolically diverse microorganisms
603 encompassing >100 microbial genera. Bioinformatic analyses suggest roles for type I
604 MGCs in methyl-selenide metabolism and/or methionine salvage, setting the stage for
605 experimental characterization. Ultimately, this previously unappreciated example of
606 methanogen specific enzymes encoded outside of the archaea could provide insights into
607 the vitally important fields of sulfur and selenium cycling.
608

609 **Acknowledgements:** D.J.K. acknowledges support from National Institutes of Health
610 Training Grant #5T32GM095450. This work was supported by a Howard Hughes Medical
611 Institute (HHMI)-Gates Faculty Scholar Award (OPP1158186) and a National Science
612 foundation (NSF) Alan T. Waterman Award (CHE-20380529) to E.P.B. E.P.B. is Howard
613 Hughes Medical Institute Investigator. We also note that this article is subject to HHMI's
614 Open Access to Publications policy. HHMI lab heads have previously granted a
615 nonexclusive CC BY 4.0 license to the public and a sublicensable license to HHMI in their
616 research articles. Pursuant to those licenses, the author-accepted manuscript of this

article can be made freely available under a CC BY 4.0 license immediately upon publication.

Data Availability Statement: All relevant data has been made available in the main text and Supplementary Information.

References

1. Kurth JM, Op den Camp HJM, Welte CU. Several ways one goal—methanogenesis from unconventional substrates. Appl Microbiol Biotechnol. 2020 Aug;104(16):6839–54.

2. Thauer RK, Kaster AK, Seedorf H, Buckel W, Hedderich R. Methanogenic archaea: ecologically relevant differences in energy conservation. Nat Rev Microbiol. 2008 Aug;6(8):579–91.

3. Chistoserdova L. Wide Distribution of Genes for Tetrahydromethanopterin/Methanofuran-Linked C1 Transfer Reactions Argues for Their Presence in the Common Ancestor of Bacteria and Archaea. Front Microbiol [Internet]. 2016 Sep 13 [cited 2023 May 29];7. Available from: http://journal.frontiersin.org/Article/10.3389/fmicb.2016.01425/abstract

4. Grinter R, Greening C. Cofactor F420: an expanded view of its distribution, biosynthesis and roles in bacteria and archaea. FEMS Microbiology Reviews. 2021 Sep 8;45(5):fuab021.

638 5. Allen JR, Clark DD, Krum JG, Ensign SA. A role for coenzyme M (2-
639 mercaptoethanesulfonic acid) in a bacterial pathway of aliphatic epoxide carboxylation.
640 Proc Natl Acad Sci USA. 1999 Jul 20;96(15):8432–7.

641 6. Venceslau SS, Stockdreher Y, Dahl C, Pereira IAC. The “bacterial heterodisulfide”
642 DsrC is a key protein in dissimilatory sulfur metabolism. Biochimica et Biophysica Acta
643 (BBA) - Bioenergetics. 2014 Jul;1837(7):1148–64.

644 7. Mock J, Wang S, Huang H, Kahnt J, Thauer RK. Evidence for a Hexaheteromeric
645 Methylenetetrahydrofolate Reductase in Moorella thermoacetica. J Bacteriol. 2014 Sep
646 15;196(18):3303–14.

647 8. Huwiler SG, Löffler C, Anselmann SEL, Stärk HJ, von Bergen M, Flechsler J, et al.
648 One-megadalton metalloenzyme complex in *Geobacter metallireducens* involved in
649 benzene ring reduction beyond the biological redox window. Proc Natl Acad Sci USA.
650 2019 Feb 5;116(6):2259–64.

651 9. Wongnate T, Ragsdale SW. The Reaction Mechanism of Methyl-Coenzyme M
652 Reductase. Journal of Biological Chemistry. 2015 Apr;290(15):9322–34.

653 10. Tallant TC, Paul L, Krzycki JA. The MtsA Subunit of the Methylthiol:Coenzyme M
654 Methyltransferase of Methanosarcina barkeri Catalyses Both Half-reactions of
655 Corrinoid-dependent Dimethylsulfide: Coenzyme M Methyl Transfer. Journal of
656 Biological Chemistry. 2001 Feb;276(6):4485–93.

657 11. Oelgeschläger E, Rother M. *In vivo* role of three fused corrinoid/methyl transfer
658 proteins in *Methanosarcina acetivorans*. Molecular Microbiology. 2009
659 Jun;72(5):1260–72.

660 12. Fu H, Metcalf WW. Genetic Basis for Metabolism of Methylated Sulfur Compounds
661 in Methanosarcina Species. O'Toole GA, editor. J Bacteriol. 2015 Apr 15;197(8):1515–
662 24.

663 13. Kaufmann F, Wohlfarth G, Diekert G. O-Demethylase from Acetobacterium
664 dehalogenans . Substrate specificity and function of the participating proteins. Eur J
665 Biochem. 1998 May;253(3):706–11.

666 14. Kurth JM, Nobu MK, Tamaki H, de Jonge N, Berger S, Jetten MSM, et al.
667 Methanogenic archaea use a bacteria-like methyltransferase system to demethoxylate
668 aromatic compounds. ISME J [Internet]. 2021 Jun 18 [cited 2021 Aug 8]; Available
669 from: http://www.nature.com/articles/s41396-021-01025-6

670 15. Naidu D, Ragsdale SW. Characterization of a Three-Component Vanillate *O* -
671 Demethylase from *Moorella thermoacetica*. J Bacteriol. 2001 Jun;183(11):3276–81.

672 16. Sauer K, Thauer RK. Methanol:coenzyme M methyltransferase from
673 *Methanosarcina barkeri* -- substitution of the corrinoid harbouring subunit MtaC by free
674 cob(I)alamin: Substitution of *M. barkeri* Mta MtaC by free cob(I)alamin. European
675 Journal of Biochemistry. 1999 May;261(3):674–81.

676 17. Geneious [Internet]. Available from: https://www.geneious.com

677 18. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high
678 throughput. Nucleic Acids Research. 2004 Mar 8;32(5):1792–7.

679 19. Oberg N, Zallot R, Gerlt JA. EFI-EST, EFI-GNT, and EFI-CGFP: Enzyme Function
680 Initiative (EFI) Web Resource for Genomic Enzymology Tools. Journal of Molecular
681 Biology. 2023 Jul;435(14):168018.

682 20. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape:
683 A Software Environment for Integrated Models of Biomolecular Interaction Networks.
684 Genome Res. 2003 Nov;13(11):2498–504.

685 21. Price MN, Dehal PS, Arkin AP. FastTree 2 – Approximately Maximum-Likelihood
686 Trees for Large Alignments. Poon AFY, editor. PLoS ONE. 2010 Mar 10;5(3):e9490.

687 22. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify
688 genomes with the Genome Taxonomy Database. Hancock J, editor. Bioinformatics.
689 2020 Mar 1;36(6):1925–7.

690 23. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al.
691 SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic
692 Acids Research. 2018 Jul 2;46(W1):W296–303.

693 24. Hoeppner A, Thomas F, Rueppel A, Hensel R, Blankenfeldt W, Bayer P, et al.
694 Structure of the corrinoid:coenzyme M methyltransferase MtaA from *Methanosarcina*
695 *mazei*. Acta Crystallogr D Biol Crystallogr. 2012 Nov 1;68(11):1549–57.

696 25. Okuda S, Yoshizawa AC. ODB: a database for operon organizations, 2011 update.
697 Nucleic Acids Research. 2011 Jan 1;39(Database):D552–5.

698 26. Matthews RG. Cobalamin-Dependent Methyltransferases. Acc Chem Res. 2001
699 Aug 1;34(8):681–9.

700 27. Siebert A, Schubert T, Engelmann T, Studenik S, Diekert G. Veratrol-O-
701 demethylase of Acetobacterium dehalogenans: ATP-dependent reduction of the
702 corrinoid protein. Arch Microbiol. 2005 Sep;183(6):378–84.

703 28. Ferguson T, Soares JA, Lienard T, Gottschalk G, Krzycki JA. RamA, a Protein
704 Required for Reductive Activation of Corrinoid-dependent Methylamine
705 Methyltransferase Reactions in Methanogenic Archaea. Journal of Biological
706 Chemistry. 2009 Jan;284(4):2285–95.

707 29. Ticak T, Kountz DJ, Girosky KE, Krzycki JA, Ferguson DJ. A nonpyrrolysine
708 member of the widely distributed trimethylamine methyltransferase family is a glycine
709 betaine methyltransferase. Proceedings of the National Academy of Sciences. 2014
710 Oct 28;111(43):E4668–76.

711 30. Schilhabel A, Studenik S, Vödisch M, Kreher S, Schlott B, Pierik AY, et al. The
712 Ether-Cleaving Methyltransferase System of the Strict Anaerobe *Acetobacterium*
713 *dehalogenans*: Analysis and Expression of the Encoding Genes. J Bacteriol. 2009 Jan
714 15;191(2):588–99.

715 31. Burke SA, Lo SL, Krzycki JA. Clustered Genes Encoding the Methyltransferases
716 of Methanogenesis from Monomethylamine. J Bacteriol. 1998 Jul;180(13):3432–40.

717 32. Sauer K, Thauer RK. Methyl-coenzyme M formation in methanogenic archaea:
718 Involvement of zinc in coenzyme M activation. European Journal of Biochemistry. 2000
719 May;267(9):2498–504.

720 33. Krüer M, Haumann M, Meyer-Klaucke W, Thauer RK, Dau H. The role of zinc in
721 the methylation of the coenzyme M thiol group in methanol:coenzyme M
722 methyltransferase from *Methanosarcina barkeri*: New insights from X-ray absorption
723 spectroscopy. European Journal of Biochemistry. 2002 Apr;269(8):2117–23.

724 34. North JA, Narrowe AB, Xiong W, Byerly KM, Zhao G, Young SJ, et al. A
725 nitrogenase-like enzyme system catalyzes methionine, ethylene, and methane
726 biogenesis. Science. 2020 Aug 28;369(6507):1094–8.

727 35. Lin J, Peng T, Jiang L, Ni JZ, Liu Q, Chen L, et al. Comparative Genomics Reveals
728 New Candidate Genes Involved in Selenium Metabolism in Prokaryotes. Genome
729 Biology and Evolution. 2015 Mar;7(3):664–76.

730 36. Haft DH, Gwadz M. Eight Unexpected Selenoprotein Families in Organometallic
731 Biochemistry in Clostridium difficile, in ABC Transport, and in Methylmercury
732 Biosynthesis. O'Toole G, editor. J Bacteriol. 2023 Jan 26;205(1):e00259-22.

733 37. Jeoung JH, Dobbek H. ATP-dependent substrate reduction at an [Fe_8S_9] double-
734 cubane cluster. Proc Natl Acad Sci USA. 2018 Mar 20;115(12):2994–9.

735 38. Böck A, Forchhammer K, Heider J, Leinfelder W, Sawers G, Veprek B, et al.
736 Selenocysteine: the 21st amino acid. Mol Microbiol. 1991 Mar;5(3):515–20.

737 39. Wells M, Basu P, Stolz JF. The physiology and evolution of microbial selenium
738 metabolism. Metallomics. 2021 May 31;13(6):mfab024.

739 40. Song Y, Lee JS, Shin J, Lee GM, Jin S, Kang S, et al. Functional cooperation of
740 the glycine synthase-reductase and Wood–Ljungdahl pathways for autotrophic growth
741 of Clostridium drakei. SYSTEMS BIOLOGY.

742 41. Niess UM, Klein A. Dimethylselenide Demethylation Is an Adaptive Response to
743 Selenium Deprivation in the Archaeon *Methanococcus voltae*. J Bacteriol. 2004
744 Jun;186(11):3640–8.

745 42. Dahlsten E, Isokallio M, Somervuo P, Lindström M, Korkeala H. Transcriptomic
746 Analysis of (Group I) Clostridium botulinum ATCC 3502 Cold Shock Response. Popoff
747 MR, editor. PLoS ONE. 2014 Feb 24;9(2):e89958.

748 43. Al-Bassam MM, Kim JN, Zaramela LS, Kellman BP, Zuniga C, Wozniak JM, et al.
749 Optimization of carbon and energy utilization through differential translational
750 efficiency. Nat Commun. 2018 Oct 26;9(1):4474.

751 44. Tan Y, Liu J, Chen X, Zheng H, Li F. RNA-seq-based comparative transcriptome
752 analysis of the syngas-utilizing bacterium Clostridium ljungdahlii DSM 13528 grown
753 autotrophically and heterotrophically. Mol BioSyst. 2013;9(11):2775.

754 45. Aklujkar M, Leang C, Shrestha PM, Shrestha M, Lovley DR. Transcriptomic
755 profiles of Clostridium ljungdahlii during lithotrophic growth with syngas or H2 and CO2
756 compared to organotrophic growth with fructose. Sci Rep. 2017 Oct 13;7(1):13135.

757 46. Biswas R, Wilson CM, Giannone RJ, Klingeman DM, Rydzak T, Shah MB, et al.
758 Improved growth rate in Clostridium thermocellum hydrogenase mutant via perturbed
759 sulfur metabolism. Biotechnol Biofuels. 2017 Dec;10(1):6.

760 47. Nogueira CW, Zeni G, Rocha JBT. Organoselenium and Organotellurium
761 Compounds: Toxicology and Pharmacology. Chem Rev. 2004 Dec 1;104(12):6255–
762 86.

763 48. Fernandes AP, Wallenberg M, Gandin V, Misra S, Tisato F, Marzano C, et al.
764 Methylselenol Formed by Spontaneous Methylation of Selenide Is a Superior Selenium
765 Substrate to the Thioredoxin and Glutaredoxin Systems. Netto LES, editor. PLoS ONE.
766 2012 Nov 30;7(11):e50727.

767 49. Albers E. Metabolic characteristics and importance of the universal methionine
768 salvage pathway recycling methionine from 5′-methylthioadenosine. :11.

769 50. Sauter M, Moffatt B, Saechao MC, Hell R, Wirtz M. Methionine salvage and S-
770 adenosylmethionine: essential links between sulfur, ethylene and polyamine
771 biosynthesis. Biochemical Journal. 2013 Apr 15;451(2):145–54.

772 51. Miller AR, North JA, Wildenthal JA, Tabita FR. Two Distinct Aerobic Methionine
773 Salvage Pathways Generate Volatile Methanethiol in Rhodopseudomonas palustris.
774 Cavanaugh CM, editor. mBio. 2018 May 2;9(2):e00407-18.

775 52. Warlick BP, Evans BS, Erb TJ, Ramagopal UA, Sriram J, Imker HJ, et al. 1-
776 Methylthio- D -xylulose 5-Phosphate Methylsulfurylase: A Novel Route to 1-Deoxy- D -
777 xylulose 5-Phosphate in *Rhodospirillum rubrum*. Biochemistry. 2012 Oct
778 23;51(42):8324–6.

779 53. North JA, Wildenthal JA, Erb TJ, Evans BS, Byerly KM, Gerlt JA, et al. A
780 bifunctional salvage pathway for two distinct S-adenosylmethionine by-products that is
781 widespread in bacteria, including pathogenic *Escherichia coli*. Mol Microbiol. 2020
782 May;113(5):923–37.

783

784

785