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

Short Title: Methanogen methyltransferases homologs

Abstract: Analyses of microbial genomes have revealed unexpectedly wide distributions of enzymes from specialized metabolism, including methanogenesis, 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 medically, environmentally, and industrially important organisms, making them likely to be physiologically relevant. Enzyme annotation, analysis of genomic context, and biochemical experiments suggests these gene clusters play a role in methyl-sulfur and/or methyl-selenide metabolism in numerous anoxic environments, including the human gut microbiome, potentially impacting sulfur and selenium cycling in diverse, anoxic environments.

Keywords: Corrinoid methyltransferases, methanogenesis, selenium metabolism, sulfur metabolism, bacteria

Introduction

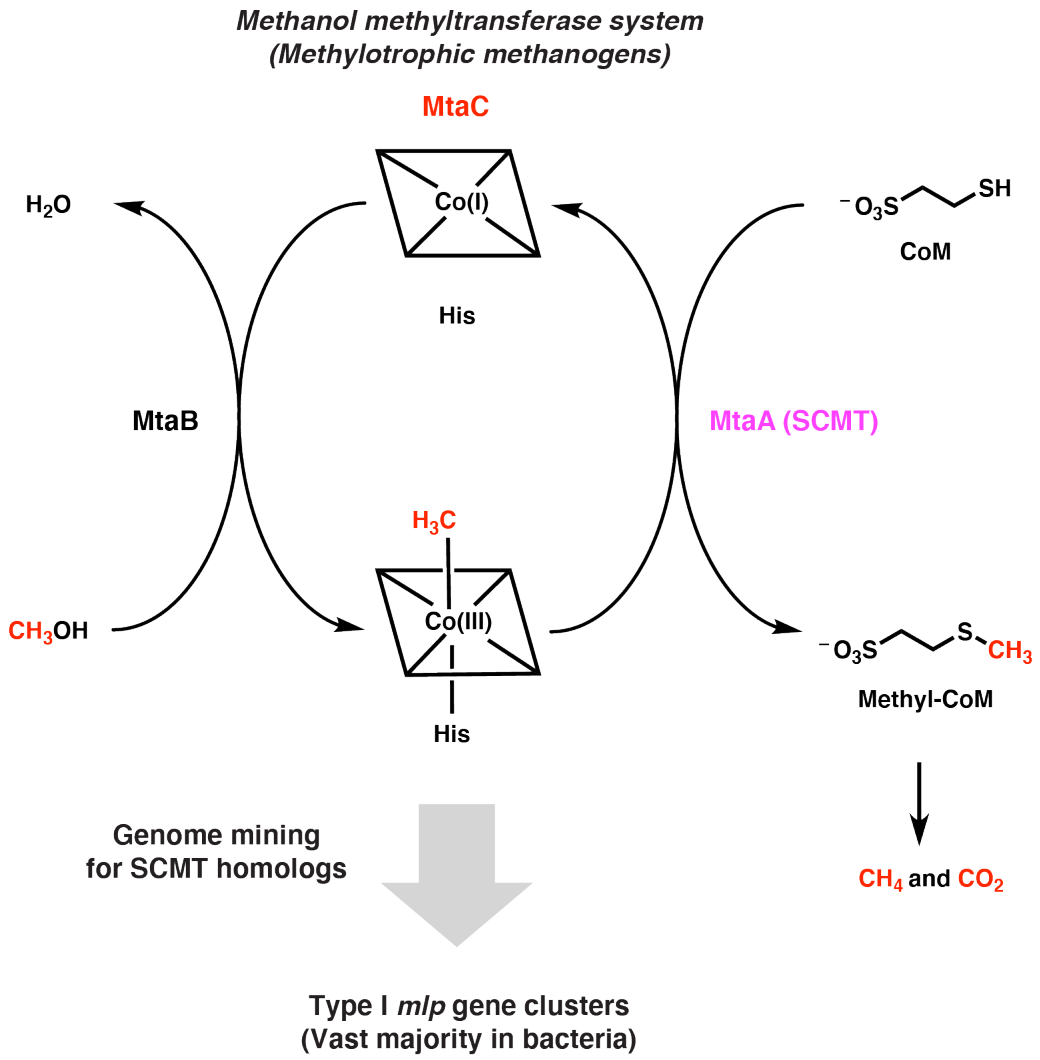
Other the past 50 years, biochemists and microbiologists have made much progress in elucidating the enzymology of methanogenesis (1,2). This reservoir of biochemical knowledge presents the opportunity to identify homologs of methanogen specific enzymes in organisms that are categorically not methanogens. In doing so, we may discover previously unappreciated metabolic pathways with global or ecological

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

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

59 We wondered if this logic extends to the soluble coenzyme M methyltransferases
60 (SCMTs) from methylotrophic methanogenesis. SCMTs are corrinoid-dependent
61 methyltransferases that transfer a methyl group from a methylated corrinoid-binding
62 protein (corrinoid protein) to coenzyme M to form methyl-coenzyme M, the immediate
63 precursor of methane (Fig. 1) (9). SCMTs belong to the very large and sequence-diverse
64 uroporphyrinogen decarboxylase (UroD) superfamily, which includes uroporphyrinogen
65 decarboxylase, cobalamin-independent methionine synthase, epoxide coenzyme M
66 transferase, and a variety of corrinoid methyltransferases including *O*-, *S*-, and halogen-

67 demethylases (10–16). However, we focused on proteins having >25% amino acid
68 identity to a characterized SCMT. We call these enzymes MtaA-like proteins, or Mlps,
69 after MtaA, the methanol-specific SCMT in *Methanosarcina* species (16). To our
70 knowledge, these Mlps have not been assigned a role outside of methanogens.



71

72 **Fig. 1 Searching genomes for the soluble coenzyme M methyltransferase (SCMT) MtaA involved**
73 **in methylotrophic methanogenesis identifies related enzymes in non-methanogens.** Overview of
74 genome mining strategy used in this work. The methanol-specific methanol:corrinoid methyltransferase
75 MtaB transfers the methyl group of methanol to the cognate corrinoid protein MtaC (the corrinoid cofactor
76 is represented by a square containing a cobalt ion). The SCMT MtaA then transfers to the methyl group

77 from MtaC to coenzyme M (CoM). The methyl group incorporated onto CoM is then disproportionated into
78 methane and carbon dioxide or solely reduced to methane, depending on the organism and environmental
79 conditions.
80
81 Here, we propose a prospective role for Mlps in a wide range of microorganisms by
82 identifying a diverse, widespread, and homologous set of gene clusters that encode
83 multiple Mlps, a corrinoid protein and a corrinoid reductive activase (Ram/RACE enzyme).
84 Gene cluster annotation, bioinformatic analysis, and biochemical logic suggests that the
85 encoded enzymes have two functions—methyl-selenide recycling and/or methionine
86 salvage—depending on the organism. These results illustrate the utility of using known
87 enzymes involved in metabolisms of interest to discover uncharacterized microbial gene
88 clusters that likely have intriguing and ecologically important functions.
89
90 **Materials and methods**
91
92 General methods and procedures
93 Manipulations and visualizations of genomic data were performed using Geneious
94 Prime 2020.0.4 software (17). BLASTP searches were conducted through the Geneious
95 interface using either National Center for Biotechnology Information (NCBI) nr/rt database
96 or custom databases generated by the user. In all cases we used the BLOSUM62 matrix
97 and a gap opening penalty of 11 and a gap extension penalty of 1. Protein alignments we
98 performed using MUSCLE (18).
99

100 Identification of *mlp* gene clusters

101 To identify homologs of SCMTs in *C. ljungdahlii*, we generated a protein database of
102 all the predicted proteins encoded in the *C. ljungdahlii* DSM 13528 genome, and
103 performed at BLASTP search of this database using the *Methanosarcina barkeri* MtbA
104 protein sequence (Swiss-Prot ID: O30640) as a query. We then inspected the genomic
105 context of the all the hits with E-values less than 1.0e-10.

106 To identify additional MGCs, we performed BLASTP searches against NCBI's nr/nt
107 database using *C. ljungdahlii* Mlp1, Mlp2, and Mlp3 (locus tags in Table S1) as queries.
108 We collected the top 1000 hits from each BLAST search. For the Mlp1 search, this yielded
109 hits with as low as 19.8% amino acid identity to the query. For the Mlp2 and Mlp3
110 searches, the minimum percent identities of the hits were 28.0% and 24.9%, respectively.
111 We then manually inspected the genomic context of each hit to determine whether they
112 were found in true MGCs. A "true" MGC met the following criteria: 1) it encodes at least
113 two Mlps, a corrinoid protein, and a Ram protein, 2) each of these components are
114 transcribed in the same direction with no intervening divergently-transcribed genes, 3)
115 there is no more than 2 kb of DNA in between the genes noted in 1, and 4) the gene
116 cluster encodes no homologs of tetrahydrofolate methylases within 20 kb of any of the
117 genes noted in 1. Using these methods and definitions, we identified 367 MGCs.

118

119 SSN analysis of MGC proteins

120 We generated all SSNs using the Enzyme Function Initiative Enzyme Similarity Tool
121 (EFI-EST) webtool (19). For each SSN, we first compiled the sequences of characterized

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

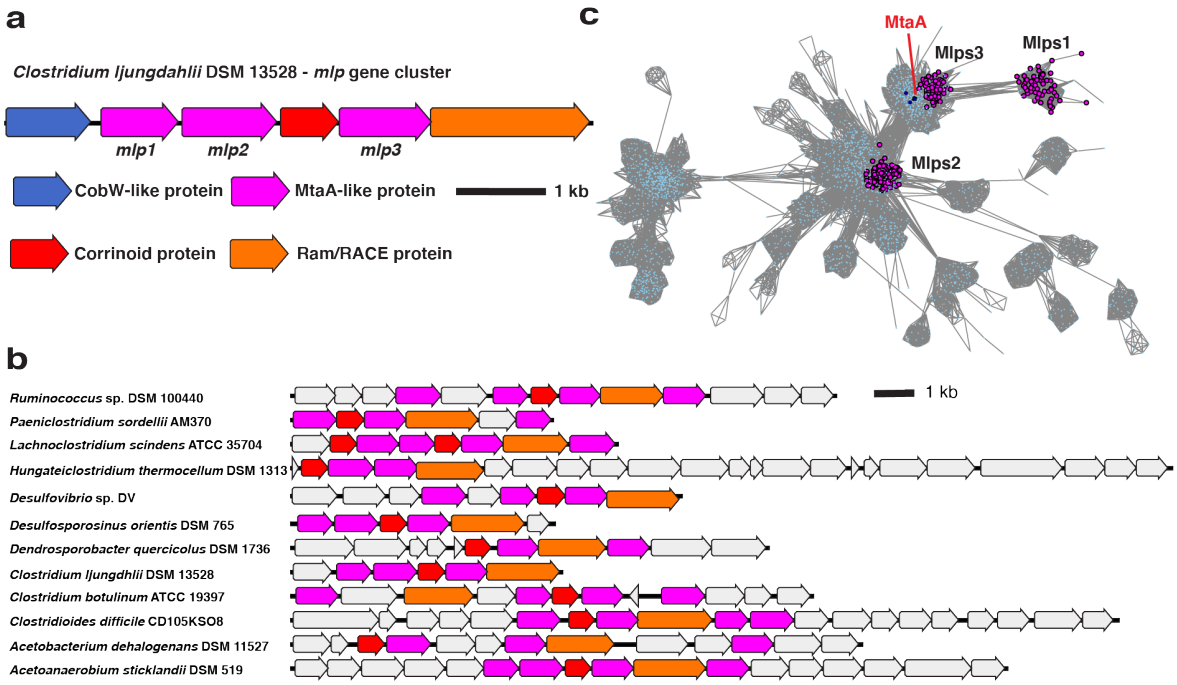
137 Phylogenetic analysis of *mlp* gene cluster proteins

138 We compiled proteins encoded by type I MGCs and sequences of related
139 characterized proteins and then aligned the resulting protein sequences using MUSCLE
140 with default parameters, and trimmed the alignments manually. All trees were generated
141 using FastTree (approximately-maximum likelihood method) with default parameters
142 (21).

143 We generated the rooted phylogenetic tree of the Mlps as follows. We first compiled
144 the sequences of Mlps1, Mlps2, and Mlps3 from type I MGCs included in the above SSN.
145 We added the SCMTs identified in the unrooted tree to this list, as well as the known
146 phenolic *O*-demethylases OdmB and VdmB (UniProt IDs: O87604 and Q6W001,
147 respectively). The SCMTs were identified as the smallest set of branches containing the
148 characterized SCMTs MtaA, MtbA, and MtsA (all from *Methanosarcina barkeri* Fusaro),
149 but excluding the Mlps3. The SCMTs were identified as the smallest clade in the unrooted
150 tree which contained the known SCMTs MtaA, MtbA, and MtsA. The tree was rooted on
151 the branch connecting the *O*-demethylases to the rest of the sequences, based on the
152 hypothesis that they serve as an appropriate outgroup.
153
154 **Results**
155
156 While manually inspecting the genome of *Clostridium ljungdahlii* DSM 13528, a
157 potential producer of biofuels, we noticed that this organism encodes multiple members
158 of the UroD superfamily. To see if these genes are closely related to SCMTs, we
159 performed BLASTP searches of the SCMT MtbA (*Methanosarcina barkeri* Fusaro,
160 *Mb*MtbA) against the predicted *C. lungdahlii* DSM 13528 proteome. We identified seven
161 hits with an E value < 1e-09 (Table S1). The highest hit was CLJU_RS09375, which
162 showed 32.5% amino acid (aa) identity to *Mb*MtbA. It is worth noting that the SCMTs
163 MtbA, MtaA, and MtsA (all from *M. barkeri* Fusaro) show only about 30% aa identity to
164 each other, suggesting that CLJU_RS09375 could be a close relative of the SCMTs.

165 CLJU_RS09375 had only 17.6% and 17.0% aa identity to the UroD superfamily phenolic
166 *O*-demethylases OdmB and VdmB, respectively.

167 To explore the physiological role of CLJU_RS09375, we next examined its
168 genomic context. CLJU_RS09375 is encoded in a gene cluster that includes genes
169 encoding a corrinoid protein (CLJU_RS09370), a "bacterial-type" corrinoid protein
170 reductive activase (Ram/RACE protein) (CLJU_RS09380), and two additional Mlps
171 (CLJU_RS09360 and CLJU_RS09365) (Fig. 2a). This gene cluster has been
172 independently predicted to form an operon by OperonDB, an automatic operon prediction
173 database (26). Because this gene cluster encodes multiple Mlps, we named it the *mlp*
174 gene cluster (MGC). We named the three Mlps in the gene cluster Mlp1, Mlp2, and Mlp3,
175 (specifically, *Cl*Mlp1, 2, and 3) in order of their appearance in the gene cluster from the
176 presumed transcriptional start site (see Fig. 2a).



177

178 **Fig. 2 Mlp gene clusters (MGCs) are similar in genetic composition and protein seqence. a** The
179 MGC from *C. ljungdahlii*. **b** Representative type I MGC. Coloring of the core MGC genes is the same as in
180 a, but non-core genes are colored grey. **c** A protein SSN of the UroD superfamily (IPR000257) showing the
181 clustering of Mlps1, 2, and 3. Nodes representing MtsA, MtbA, and MtaA (all from *M. barkeri* Fusaro) are
182 shown in dark blue. The alignment score cutoff was 40 and the node representation percent cut was 50 %
183 aa identity. The SSN shown is a trimmed version of the full SSN, with all nodes removed that are neither
184 directly or indirectly connected to the Mlps1, 2, or 3.
185

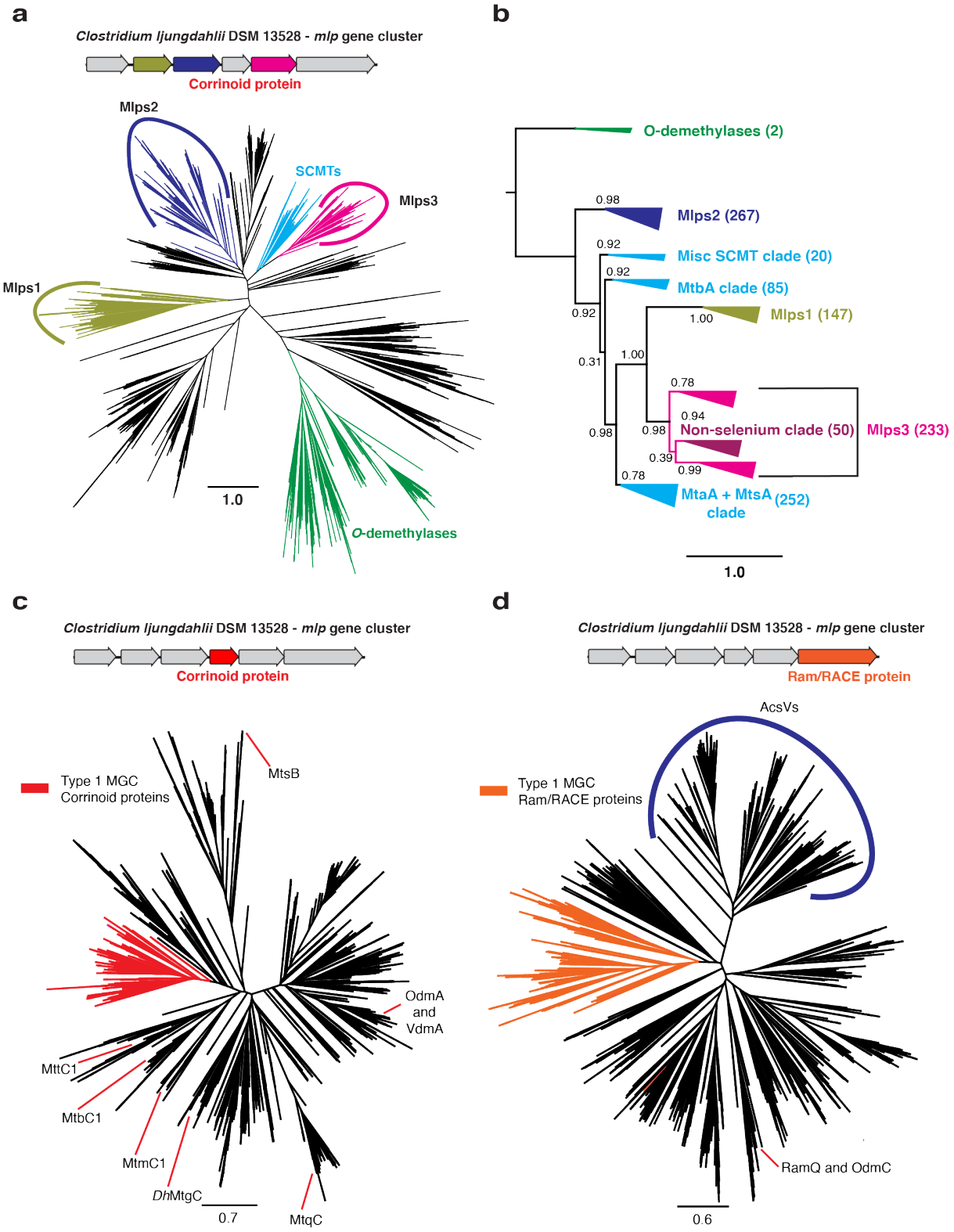
186 Given the relatively high sequence similarity of the MGC Mlps to SCMTs, it seems
187 reasonable to hypothesize that these enzymes are corrinoid methyltransferases. This is
188 consistent with the presence of genes encoding a corrinoid protein and a Ram/RACE
189 protein in the MGC. The former is a methyl group carrier (27), while the latter is an ATP-
190 dependent corrinoid protein reductase that reductively activates the corrinoid protein to
191 its catalytically-active Co(I) state (28,29). Corrinoid methyltransferases, corrinoid
192 proteins, and Ram/RACE proteins are frequently colocalized (30–32). In bacteria, such
193 gene clusters are involved in methylotrophic metabolism and participate in transferring a
194 methyl group from a donor substrate to tetrahydrofolate (H_4folate) via a methyl-corrinoid
195 protein intermediate, a reaction catalyzed by a methyl-corrinoid protein:H_4folate
196 methyltransferase (15,30). In most of the bacterial methyltransferase pathways examined
197 to date, the H_4folate methylase is encoded in the same gene cluster as the corrinoid
198 protein and methyl donor methyltransferase (30,31). It is therefore surprising that H_4folate
199 methyltransferase is missing from the *C. ljungdahlii* MGC. While it is possible that a
200 H_4folate methyltransferase encoded elsewhere in the *C. ljungdahlii* genome could

201 cooperate with the MGC methyltransferase system, it is also possible that H_4folate is not
202 involved or that one of the MGC Mlps is an undiscovered H_4folate methyltransferase.
203 In parallel, we located similar gene clusters in *C. sporogenes* DSM 795 and
204 *Hungateiclostridium* (*Clostridium*) *thermocellum* DSM 1313. Like the MGC from *C.*
205 *ljungdahlii*, the *C. sporogenes* and *H. thermocellum* gene clusters encode multiple Mlps
206 (each with about 30% aa identity to the SCMTs), a corrinoid protein, and a Ram/RACE
207 protein (Fig. 2b). These gene clusters also lack a H_4folate methylase. However, unlike *C.*
208 *ljungdahlii*, which encodes H_4folate methylases elsewhere in its genome, homologs of
209 H_4folate methylases are absent from the *C. sporogenes* and *H. thermocellum* genomes,
210 suggesting that H_4folate is not involved in the pathways encoded by these gene clusters.
211 Because of the similarity of the *C. sporogenes* and *H. thermocellum* gene clusters to the
212 *C. ljungdahlii* MGC, we refer to then as the *Cs*MGC, *Ht*MGC, and *Cl*MGC, respectively.
213 We next wondered if related MGCs are found in additional organisms. To address this,
214 we BLASTed *Cl*Mlp1, 2, and 3 against the NCBI non-redundant protein database. Many
215 of the organisms that had high alignment score hits for *Cl*Mlp1 also encoded high
216 alignment score hits for *Cl*Mlp2 and 3. When we examined the genomic contexts of these
217 hits, we identified gene clusters that resemble the *C. ljungdahlii* MGC (Fig. 2b).
218 We next asked if these MGCs are homologous or analogous. Because components
219 of a corrinoid methyltransferase pathway are commonly encoded in gene clusters
220 (14,15,30–32), evolutionary pressure to colocalize corrinoid methyltransferase genes
221 could have resulted in superficially similar MGCs that share no evolutionary relationship.
222 Alternatively, these gene clusters could be descended from a common ancestor (i.e. are

223 homologous), which would also suggested a shared function in diverse organisms. To
224 address this question, we identified 367 organisms encoding likely MGCs (see materials
225 and methods). We next constructed several protein sequence similarity networks (SSNs)
226 (19), reasoning that if MGCs are homologous, each conserved component should form
227 its own cluster in an SSN of the protein family to which that component belongs. Because
228 there are multiple *mlp* genes in each MGC, there are two SSN analysis outcomes that
229 are consistent with the homology hypothesis. First, all Mlps from the MGCs could be
230 grouped within a single cluster in the SSN, which would be expected if the multiple *mlp*
231 genes in each MGC arose from gene duplication. Second, the MGC Mlps could each
232 represent different lineages and would therefore likely form distinct clusters within an
233 SSN.

234 We first constructed an SSN (Fig. 2c) of the IPR000257 superfamily, which includes
235 the Mlps. We found that 80.9% of the MGC Mlps localized into one of three sub-clusters
236 within a larger cluster (Fig. 2c). Interestingly, these three sub-clusters correspond to the
237 three Mlps from the *C. ljungdahlii* MGC, with *Cl*Mlp1, 2, and 3 being grouped into distinct
238 sub-clusters. Inspection of other MGCs reveals that almost all encode Mlps from clusters
239 2 and 3, while a gene encoding an Mlp from cluster 1 was present a majority of the time,
240 but is often absent. We denote the proteins in the sub-cluster containing *Cl*Mlp1 as Mlps1,
241 and those from sub-clusters containing *Cl*Mlp2 and 3 as Mlps2, and Mlps3, respectively.
242 Altogether, this analysis shows that a large number of MGCs encode at least an Mlp2
243 and an Mlp3. We designate such gene clusters as “type 1 MGCs”.

244 Because >95% of MGCs identified appear to be type 1 MGCs, we selected these
245 MGCs for further study. The fact that the type 1 MGC Mlps fall into three separate clusters
246 in the SSN suggests they might represent three distinct Mlp lineages, rather than a single
247 lineage that arose from recent gene duplication events. To solidify this inference, we
248 constructed a phylogenetic tree of the Mlps and other members of the UroD superfamily
249 that are close relatives of known corrinoid methyltransferases (Figure 3a). The Mlps 1, 2,
250 and 3 are situated in a large branch that includes the SCMTs, to the exclusion of other
251 known UroD superfamily corrinoid methyltransferases such as the aromatic *O*-
252 demethylases (13). When rooted (Fig. 3b), we found that the Mlps 1, 2 and 3 formed
253 distinct, well-supported clades. Interestingly, the Mlps2 form a sister clade to the clade
254 containing the SCMTs and the Mlps1 and Mlps3. Within this clade, the Mlps1 and Mlps3
255 form sister clades emerging from the SCMTs, which therefore appear to be paraphyletic.



256

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

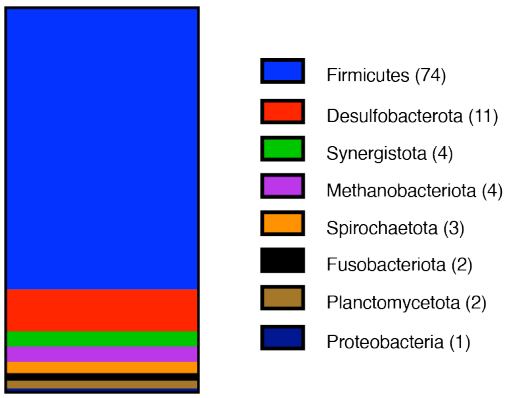
258 phylogenetic tree of members of the UroD superfamily methyltransferases. The FastTree support values

259 for the Mlps1, 2, and 3 clades are 0.95, 0.94, and 1.00, respectively. Scale bar: branch length corresponding
260 to an average of one change per position. **b** A rooted approximately-maximum likelihood (FastTree)
261 phylogenetic tree of selected protein sequences. Shown are the soluble coenzyme M methyltransferases
262 (SCMTs), the Mlps1, 2 and 3, and the phenolic *O*-demethylases OdmB and VdmB (outgroup). Each branch
263 is labeled with the branch's FastTree support value. The number of sequences in each clade are as follows.
264 *O*-demethylases: 2; Mlps2: 267; Misc. SCMT clade: 20; MtbA clade: 85; Mlps1: 147; Mlps3: 233; MtaA +
265 MtsA clade: 252. **c** An approximately-maximum likelihood (FastTree) phylogenetic tree of members of the
266 methyltransferase corrinoid protein family. The type I MGC proteins are shown in red. The FastTree support
267 value for the type I MGC proteins is 0.92. Selected characterized family members are shown: MtbC1,
268 MtmC1, and MttC1 (all from *M. acetivorans* C2A); *Dh*MtgC from *Desulfitobacterium hafniense* Y51; MtqC
269 from *Eubacterium limosum* ATCC 8486; MtsB from *M. barkeri* Fusaro; OdmA and VdmA from
270 *Acetobacterium dehalogenans* DSM 11527. **d** An approximately-maximum likelihood (FastTree)
271 phylogenetic tree of members of the Ram/RACE protein family. The type I MGC proteins are shown in
272 orange. The FastTree support value for the type I MGC proteins is 0.98. A clade containing the corrinoid
273 iron-sulfur protein reductive activases is labeled "AcsVs". The locations of the activases RamQ (*E. limosum*
274 ATCC 8486) and OdmC (*A. dehalogenans* DSM 11527) are shown. Scale bars for all tree represent
275 respective branch length corresponding to an average of one change per position.
276

277 To further test this hypothesis, we generated two additional phylogenetic trees:
278 one containing the type 1 MGC corrinoid proteins and other monomeric methyltransferase
279 corrinoid proteins (Fig. 3c and 3d, respectively), and the other containing the type 1 MGC
280 Ram/RACE proteins and other Ram/RACE proteins (Fig. S4). In each case, the type 1
281 MGC proteins were found in the same branch, with a handful of exceptions. These
282 exceptions could be the result of gene conversion or the replacement of a "canonical"
283 type 1 MGC gene with an isofunctional gene from outside of the MGC.

284 Together, data from the SSN and phylogenetic analyses suggests that type 1 MGCs
285 are homologous in that they encode mutually orthologous proteins. The simplest
286 explanation for the homology of the type 1 MGCs is that they are horizontally transferred
287 as a unit. However, we cannot rule out more complicated scenarios including the transfer
288 of type 1 MGC genes individually into different regions of the genome followed by
289 selective pressure for colocalization. This latter possibility gains some credence when we
290 note the lack of conserved synteny in type 1 MGCs.

291 We next analyzed the phylogenetic distribution of type 1 MGCs (Fig. 4). To date, we
292 have identified over one hundred prokaryotic genera containing at least one type 1 MGC-
293 encoding strain. Of these, only three (*Methanococcus*, *Methanothermococcus*, and
294 *Methanotorris*) are methanogenic archaea. This is noteworthy because members of the
295 Methanococcales are not known to be methylotrophic and therefore would not be
296 expected to encode SCMTs. All type I MGC-encoding genera only obligate anaerobes,
297 with the majority (73%) belonging to the Firmicutes phylum. Although type 1 MGCs are
298 found a limited number of phyla, they are widely distributed across at least 35 orders of
299 bacteria and archaea. From this analysis, we conclude that type 1 MGCs are
300 phylogenetically widespread and typically associated with microorganisms from anoxic
301 environments. Notable species encoding type 1 MGCs include toxigenic clostridia
302 (*Clostridium botulinum*, *Paeniclostridium sordelii*, certain strains of *Clostridioides difficile*)
303 and industrially-relevant clostridia (*Hungateiclostridium* (*Clostridium*) *thermocellum*,
304 *Clostridium ljungdahlii*).



305

306 **Fig. 4 Type I MGCs are found in several prokaryotic Phyla.** Phylum composition of genera encoding
307 type I MGCs. 101 genera are represented. The number of type I MGC-encoding genera within each phylum
308 is enclosed in parentheses.

309

310 We next analyzed additional components of the type 1 MGCs and proposed a
311 biochemical function for these genes. The UroD superfamily includes corrinoid
312 methyltransferases that (de)methylate heteroatom-containing substrates
313 (10,12,14,15,33). Several lines of evidence suggest that the type 1 MGCs encode a
314 corrinoid *S*- or *Se*-methyltransferase system. First, the closest relationship between
315 known UroD superfamily proteins and the type 1 *mlp* cluster Mlps is with the SCMTs (Fig.
316 a and b), which are *S*-methyltransferases. Additional evidence comes from examination
317 of the enzyme's Zn^{2+}-binding site. This Zn^{2+} cofactor coordinates the heteroatom of a
318 substrate, activating it for (de)methylation (24,33,34). The primary coordination sphere of
319 the Zn^{2+} cofactor is His-Cys-Cys in *S*-methyltransferases from the UroD superfamily,
320 while in *O*-methyltransferases, one or both of the Cys ligands are substituted by
321 carboxylates (Glu or Asp) or are substituted for non-coordinating residues (15,24,33,34).
322 The His-Cys-Cys coordination environment favors the binding of a soft ligand such as a
323 selenium or sulfur atom to the Zn^{2+} center, while substitution of Cys for a carboxylate

324 favors binding of an oxygen atom. The coordination sphere of the type 1 MGC Mlps is
325 (with the exception of several Mlps1) invariably His-Cys-Cys, suggesting that these
326 proteins are *S*- or *Se*-methyltransferases. Homology models of the *C. ljungdahlii* Mlps in
327 further confirmed that these residues are likely appropriately positioned to coordinate Zn^{2+}
328 (Fig. S3). Finally, examination of the non-core components of type I MGCs (genes in type
329 I MGCs that do not encode Mlps, corrinoid proteins, or Ram/RACE proteins) consistently
330 reveals genes involved in sulfur metabolism, including methionine aminotransferases,
331 sulfurtransferases, methionine transporters, and close homologs of the recently-
332 described MarHDK enzyme that catalyzes C–S bond cleavage to produce ethylene from
333 methylthioethanol (35). These observations suggest a potential role for many type I MGCs
334 in methyl-sulfur metabolism.

335 However, additional evidence suggests that most type I MGCs function in methyl-
336 selenium metabolism. Many type 1 MGCs contain genes provisionally linked to selenium
337 metabolism, such as the recently characterized "dicubane cluster protein" (36–38).
338 Several type 1 MGCs (Fig. 5) contain genes encoding authentic machinery for
339 selenocysteine synthesis, encoding, and decoding (SelABCD), as well as the enzymes
340 of the glycine reductase system, two of which have selenocysteine residues (39,40).
341 Notably, 87% of sequenced species that encode type I MGCs also encode selenocysteine
342 biosynthesis machinery within the same genome. For comparison, only 21.5% of bacteria
343 contain the requisite genes for selenocysteine biosynthesis, and even in the
344 selenocysteine-rich Clostridia, only 42% of 184 genomes encoded selenocysteine
345 biosynthesis machinery (36). Not all organisms that make use of selenium encode

346 selenocysteine biosynthetic machinery—some organisms that use the selenium cofactor
347 only encode selenide-water dikinase (SelD) (36). We therefore manually inspected
348 genomes that contain a type I MGC, but lack SelAB or C, finding that several genomes
349 lacked *selABC*, but had *selD* as well as selenium cofactor-dependent enzymes. As noted
350 below, type 1 MGCs from *C. ljungdahlii* and *C. drakei* are upregulated during autotrophic
351 growth on H_2 + CO_2, a condition where there is increased demand for selenoprotein
352 biosynthesis (41). Most importantly, during the preparation of this manuscript, we became
353 aware of prior study that used proteomic and genetic methods to implicate three type I
354 MGC-encoded proteins in the utilization of dimethyl-selenide as a selenium source in the
355 hydrogenotrophic methanogen *Methanococcus voltae* (42). The authors suggested that
356 the Mlps might demethylate (di)methyl-selenide and transfer the methyl group onto the
357 MGC corrinoid protein, liberating selenide for incorporation into biological
358 macromolecules. While the authors did not biochemically characterize the Mlps, their data
359 strongly support a role for the type I MGCs in methyl-selenol utilization.

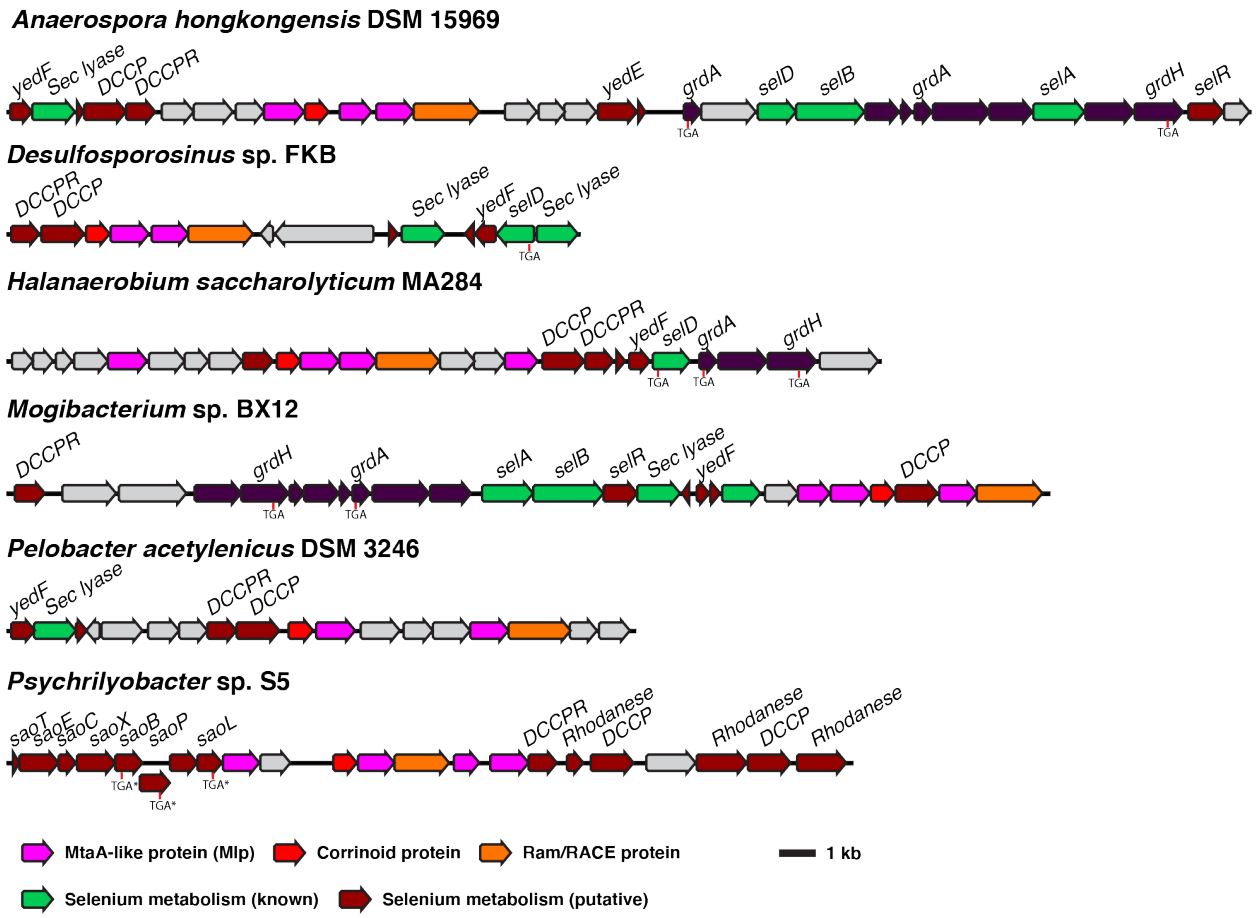


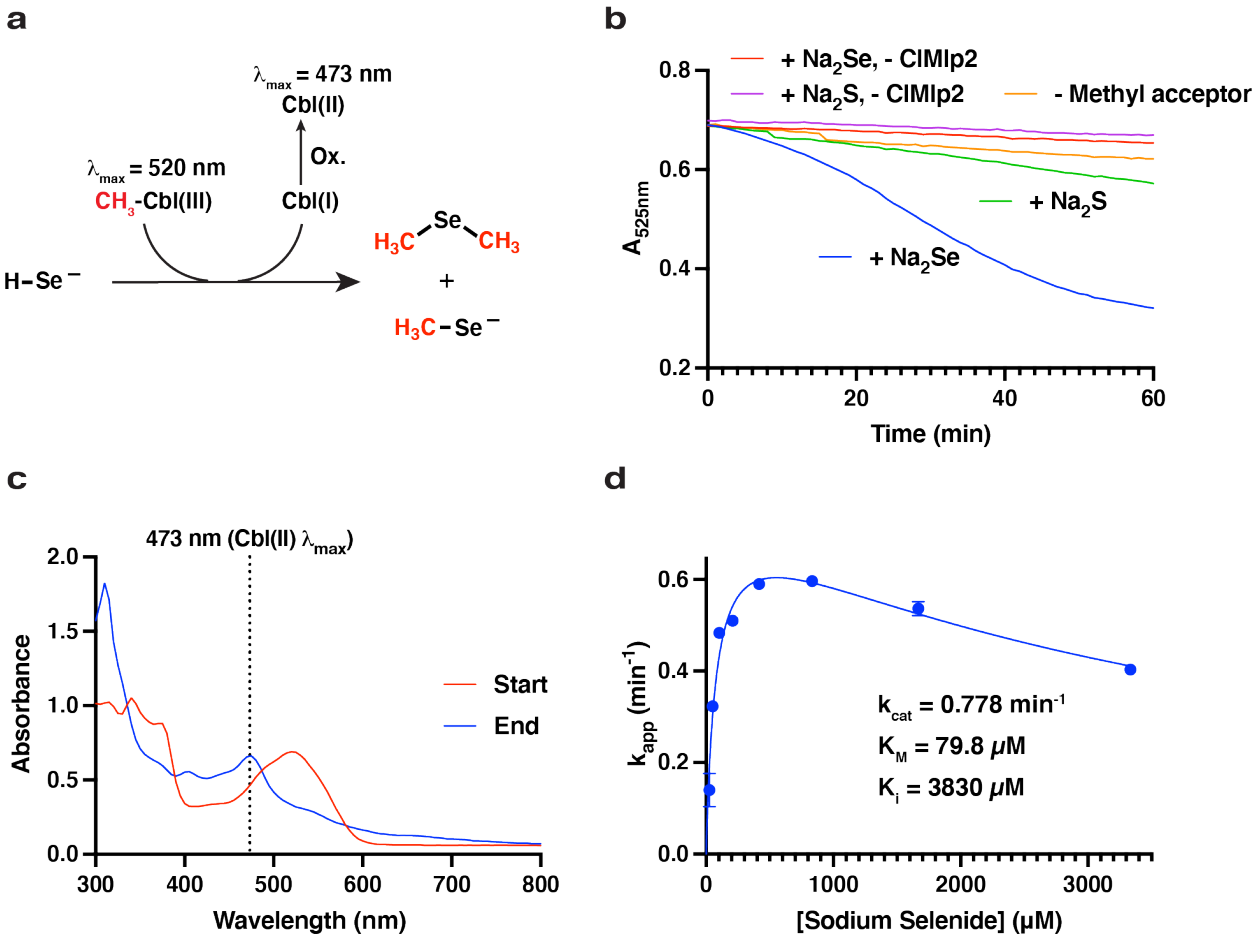
Fig. 5 Type I MGCs often encode genes involved in selenium metabolism. Selenium metabolism gene show are either close homologs of genes with experimentally-confirmed roles in selenium metabolism ("Selenium metabolism (known)") or have been associated with selenium metabolism by bioinformatic methods ("Selenium metabolism (putative)") (36,37). The positions of selenocysteine codons are make with "TGA." "TGA*" indicates a selenocysteine codon that has been bioinformatically predicted, but not experimentally confirmed (37). 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 (36). Sao: putative "selenocysteine-assisted organometallic operon genes.

372 Because type 1 MGCs are present in several model organisms, we examined the
373 expression of these genes in existing proteomic and transcriptomic datasets for *C.*
374 *botulinum*, *C. ljungdahlii*, and *H. thermocellum* to gain additional information regarding
375 their functions. In all three organisms, the MGCs are expressed at a detectable level
376 under all conditions examined. First, we found that the *C. botulinum* MGC is upregulated
377 during growth at the organism's optimal growth temperature of 37 °C compared to growth
378 at 15 °C (transcripts 5.27-68.6 fold upregulated at 37 °C) (43). In *C. ljungdahlii*, we found
379 that the type 1 MGC is transcriptionally upregulated roughly 3-5 fold during autotrophic
380 growth on H_2 + CO_2 or CO versus heterotrophic growth on fructose (44–46). The same
381 pattern is found in *C. drakei* (6 to 13 fold transcriptional upregulation on H_2 + CO_2
382 compared to fructose) (41). In *H. thermocellum*, the type 1 MGC is upregulated during
383 growth with supplemental acetate in a strain from which the FeFe hydrogenase
384 maturation enzyme *ΔhydG* had been deleted (47). These findings imply that type 1 MGCs
385 are expressed, indicating their physiological relevance, and respond to environmental and
386 genetic perturbations in industrially and medically significant organisms.

387 Finally, we sought characterize the activity of an Mlp to obtain further insight into the
388 function of type I MGCs. Niess et al.(42) observed that deletion of the gene encoding an
389 Mlp2 (*sdmC*) resulted in loss of the ability of *M. voltae* to use dimethyl-selenide as a
390 selenium source, suggesting the function of Mlps2 may be to demethylate dimethyl-
391 selenide and/or methyl-selenol. We therefore selected *Cl*Mlp2 for biochemical assays.
392 Recombinant ClMlp2 was assayed with sodium selenide and methylcob(III)lalamin under
393 a strictly anoxic nitrogen atmosphere as described in the *Materials and Methods*. In brief,

394 this UV-visible spectroscopy assay assesses methylation of a methyl acceptor by
395 measuring the disappearance or shift of a large methylcob(III)alamin absorbance band
396 (λ_{max} ~ 525 nm) when the oxidation and/or ligation state of the cobalt ion is changed. For
397 convenience, we assayed for methylcob(III)alamin:selenide methyl transfer (Fig 6a), the
398 reverse of the hypothesized physiological reaction. Addition of sodium selenide (666 μM)
399 to the reaction mixture resulted in a slow but significant decrease in A_{525nm} (Fig. 6b).
400 Control reactions lacking either *Cf*Mlp2 or sodium selenide exhibited rates of A_{525nm}
401 decrease that were 13.7- and 8.3-fold slower, respectively, and are attributable to
402 artefactual photolysis of methylcob(III)alamin inherent to measurement of absorbance
403 (48). Addition of sodium sulfide (666 μM) instead of sodium selenide resulted in a 4.3-fold
404 slower rate, which could suggest the enzyme also possesses low sulfide methylating
405 activity. Starting and ending spectra of the reactions only showed significant spectral
406 changes in the selenide and sulfide reactions. In both cases, the spectra showed a shift
407 in λ_{max} from 520 to 473 nm indicative of formation of base-on cob(II)alamin and therefore
408 demethylation (Fig. 6c). The net steady state formation of cob(II)alamin with minimal
409 accumulation of other intermediates (indicated by a steady absorbance trace at the
410 methylcob(III)alamin/cob(II)alamin isosbestic point of 485 nm (Fig S4)) is likely the result
411 of rapid oxidation of cob(I)alamin, the expected reaction product. Steady-state kinetic
412 analysis revealed an apparent k_{cat} of 0.778 min^{-1} and an apparent K_M for selenide of 79.8
413 μM with a selenide substate inhibition constant (K_I) of 3830 μM (Fig. 6d). The low k_{cat} is
414 likely due the use of methylcobalamin (rather than the physiological methyl-corrinoid
415 protein) as a methyl donor. The low selenide K_M, which is comparable to the 46 μM

416 selenide K_M measured for *E. coli* selenophosphate synthase (49), suggests that methyl-
417 selenol and/or dimethylselenide are physiological substrates of *Cl*Mlp2. Together, these
418 biochemical studies of *Cl*Mlp2 support the hypothesis that type I MGCs are involved in
419 methyl-selenide metabolism.



420
421
422 **Fig. 6 *C. ljungdahlii* (Cl)Mlp2 catalyzes methylcobalamin:selenide methyl transfer a** Schematic of the
423 reaction. Cbl: cobalamin. "Ox." signifies the observed formation of cob(II)alamin, presumably from
424 nonenzymatic reaction of small amounts of diselenides with highly labile cob(I)alamin. **b** 525 nm
425 (methylcob(III)alamin peak) traces indicating disappearance of methylcob(III)alamin. **c** Starting (t = 0 min)
426 and ending (t = 60 min) spectra for the "+ Na_2Se" reaction shown in **b**. **d** Steady state kinetic analysis of
427 ClMlp2 showing apparent kinetic parameters. Points and error bars represent the mean and standard
428 deviations from three independent reactions.

429

430 **Discussion**

431

432 Methylotrophic methanogenesis is restricted to a subset of the archaea, so it is
433 surprising to find homologs of enzymes from this pathway in non-methanogens. Here, we
434 identify a diversified, widespread set of gene clusters—type 1 MGCs—that encode
435 homologs of the SCMTs involved in methylotrophic methanogenesis. Type I MGCs are
436 found exclusively in anaerobes and are most common in the Firmicutes phylum. There is
437 a strong tendency for type I MGCs to be found in genomes that also contain
438 selenocysteine biosynthesis and decoding machinery.

439 What reactions do the enzymes encoded by type 1 MGCs catalyze and what is
440 their physiological function? The presence of a corrinoid protein, a corrinoid reductive
441 activase, and homologs of corrinoid methyltransferases (Mlps) implies that the type 1
442 MGCs encode components of a corrinoid methyltransferase pathway. Protein sequence
443 comparisons indicate that the Mlps are likely *S*- or *Se*-methyltransferases that
444 (de)methylate thiol and selenol metabolites.

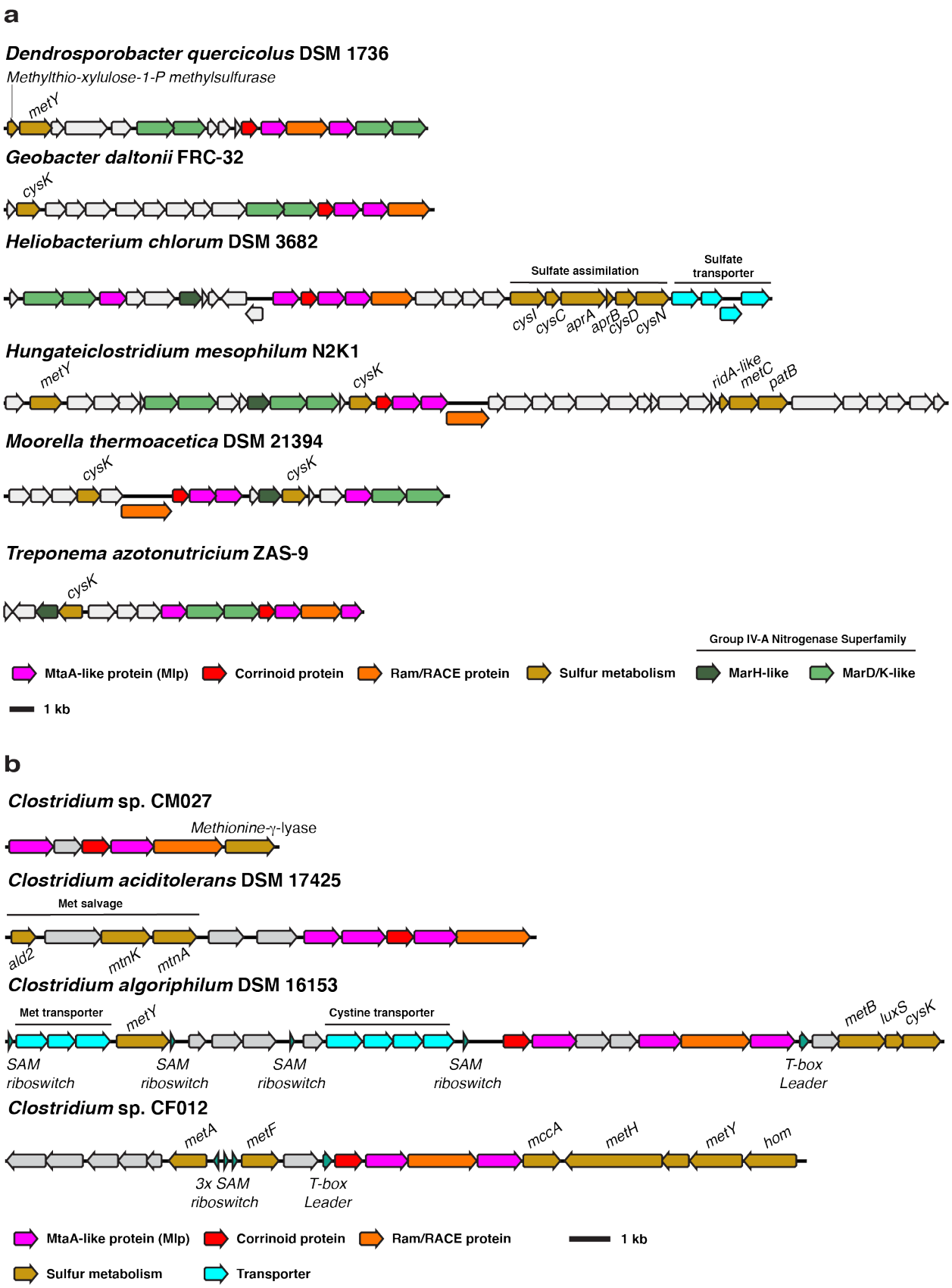
445 The potential physiological roles of such reactivity may be explored through
446 considering features of the organisms that possess type I MGCs. Earlier studies clearly
447 implicate type I MGCs in methyl-selenol valorization (39). However, those authors
448 examined only the corrinoid protein and two closely-related Mlps2 in dimethyl-selenide
449 utilization—the more distantly related *M. voltae* Mlp1 and Mlp3 appear to have gone
450 unnoticed. It seems likely that the full type I MGC contains methyltransferases for

451 demethylation of several methyl-selenium compounds. While the genetic data in Ref. 39
452 implicates dimethyl-selenide and/or methyl-selenol as substates of the type I MGC Mlps2
453 (an inference that our assay of *Cl*Mlp2 supports), we propose that the Mlps1 and Mlps3
454 may be responsible for demethylation of other methyl-selenium compounds such as
455 methyl-selenocysteine and selenomethionine and concomitant methylation of the gene
456 cluster's encoded corrinoid proteins. We propose that the methyl-selenide-derived methyl
457 group is subsequently transferred from the methylated corrinoid protein to a thiol-
458 containing compound such as sulfide. The resulting methyl-sulfur metabolite could then
459 be excreted or recycled into methionine. Forms of methyl-selenium are present in many
460 environments, including the human gut (50). It may be that most type I MGCs function to
461 valorize these methyl-selenium compounds. However, many organisms (such as *C.*
462 *ljungdahlii* and *M. voltae*) upregulate their type I MGCs even when methyl-selenium
463 compounds are not added to growth media. This observation suggests that there may be
464 an endogenous source of methyl-selenium compounds. Indeed, it has been suggested
465 and evidentially supported that the selenide generated as an intermediate in
466 selenoprotein biosynthesis can react non-enzymatically with *S*-adenosylmethionine
467 (SAM) (51). This presents a problem for a cell that has a high demand for selenium
468 metabolism: methyl selenide formation from SAM wastes valuable selenium and
469 generates a potentially toxic dead-end metabolite. The enzymes encoded by type I MGCs
470 may solve this quandary by recycling the methyl selenide into methionine.

471 Some organisms, such as *H. thermocellum*, encode and express type I MGCs but lack
472 genes encoding known selenoenzymes or other indicators of selenium metabolism. We

473 found that all the selenium-trait deficient organisms have Mlps3 that fall into a single clade
474 of the Mlp phylogenetic tree (Fig. 3b). We suggest that the corresponding type I MGCs
475 are involved in methyl-sulfur metabolism. While one group of type I MGCs (>90%) is likely
476 involved in methyl-selenium metabolism, the other, which encodes Mlps3 from the non-
477 selenium clade, probably has a selenium-independent function.

478 Evidence from genomic context suggests that the non-selenium MGCs could be
479 involved in methionine salvage, defined as a metabolic pathway that converts methylthio-
480 adenosine (MTA) into methionine (52). MTA is a toxic by-product of the biosynthesis of
481 key metabolites derived from SAM (53). Methionine salvage pathways therefore support
482 sulfur metabolism both by detoxifying a by-product and preserving the metabolically
483 expensive methylthiol group of MTA (54). The so-called “universal” methionine salvage
484 pathway is O_2-dependent, and methionine salvage pathways that operate under
485 anaerobic conditions have only recently been discovered (52). The anaerobic methionine
486 salvage pathway from *Rhodospirillum rubrum* was recently shown to use the methylthio-
487 alkane reductase MarHDK system, homologs of the nitrogenase subunits NifHDK,
488 respectively (35). Remarkably, *MarHDK*-like genes belonging to the nitrogenase
489 superfamily IV-A group are prevalent in or near type I MGCs from the non-selenium clade
490 and, when not present in the type I MGC itself, are invariably present elsewhere in the
491 genome (Fig. 7a). This suggests that the non-selenium type I MGCs encode enzymatic
492 machinery for either methionine salvage, sulfur valorization from methyl-sulfide
493 compounds, or both.


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

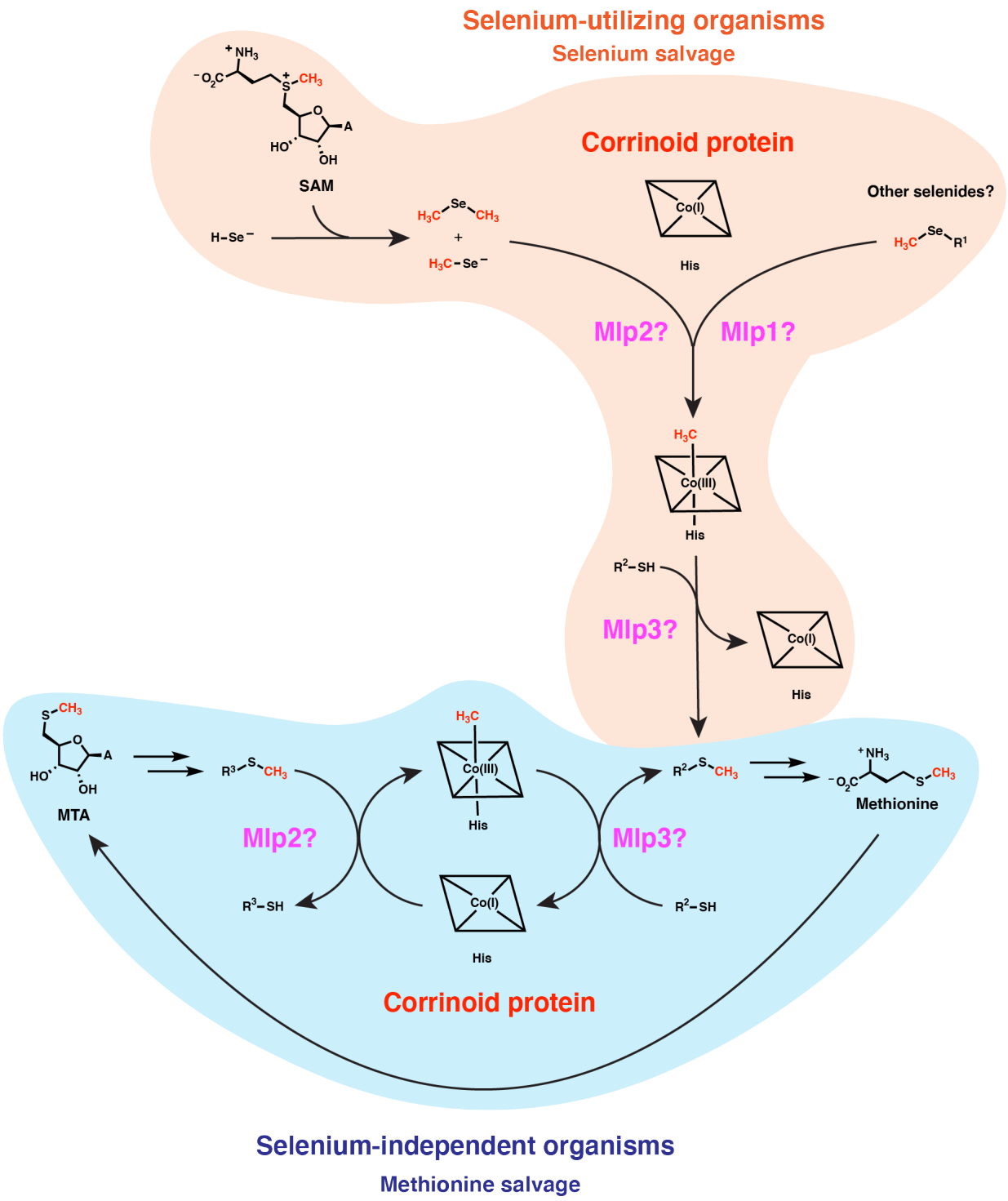
494

495 **Fig. 7 Type I MGCs from encoding an Mlp3 from the non-selenium clade also encode putative sulfur**
496 **and methionine metabolism genes. a** Non-selenium type I MGCs associated with group IV-A nitrogenase
497 superfamily genes and other genes involved in sulfur metabolism. Methylthioribulose-1-P methylsulfurase
498 is a methionine salvage enzyme originally described in *Rhodospirillum rubrum* (55). **b** Other non-selenium
499 type I MGCs associated with methionine metabolism. MetY: *O*-acetyl-homoserine sulfhydrylase. CysK:
500 cysteine synthase. CysI: assimilatory sulfite reductase. CysC: adenylyl-sulfate kinase. AprAB:
501 phosphoadeninosine-phospho-sulfate reductase subunits A and B. CysND: sulfate adenylyltransferase
502 subunits 1 and 2. MetC cystathionine beta-lyase. PatB: alternate cystathionine beta-lyase. Ald2: methylthio-
503 ribulose-1-phosphate aldolase. MtnK: methylthio-ribose kinase. MtnA: methylthio-ribose-1-phosphate
504 isomerase. MetB: cystathionine gamma-synthase. LuxS: *S*-ribosyl-homocysteine lyase. MetA: homoserine-
505 *O*-succinyltransferase. MetF: methylene-tetrahydrofolate reductase. MccA: *O*-acetyl-serine dependent
506 cystathionine beta-synthase. MetH: cobalamin-dependent methionine synthase. Hom: homoserine
507 dehydrogenase. SAM: *S*-adenosyl-methionine.
508

509 Additionally, the sporadic distribution of type I MGCs across phylogenetic trees is
510 potentially consistent with a role in methionine salvage, as such pathways are
511 dispensable under many growth conditions (54) and are not strictly conserved even in
512 closely related strains (56).

513 We propose how a type I MGC-encoded methionine salvage pathway could operate
514 in Fig. 8, considering a type I MGC that encodes only two Mlps. The MarHDK-like
515 enzymes may have a role in such a pathway, generating methyl-sulfide that might serve
516 as the methyl donor for the MGC methyltransferase system (R^1-SCH_3 with R^1 = H).
517 Alternatively, the MarHDK-like proteins might serve as part of a second, parallel
518 methionine salvage pathway, or even as a branch of a topologically bifurcating pathway
519 with two routes to methionine.

520 Intriguingly, this proposed methionine salvage pathway uses a distinct biochemical
521 logic from related pathways. Characterized methionine salvage pathways all incorporate
522 the methylthiol group of MTA intact into methionine (35,52,53,56). In the proposed
523 methionine salvage pathway involving the type 1 MGCs, the methylthio-ether of MTA is
524 dismantled into a methyl group which is carried by the corrinoid protein, leaving the free
525 thiol behind. This pathway architecture theoretically avoids production of volatile methyl-
526 sulfur intermediates that could be more readily lost from the cell. The proposed role of a
527 corrinoid cofactor in methionine salvage is also novel. Ultimately, biochemical
528 characterization of the enzymes encoded by type I MGCs will be required to test this
529 proposal.



530

531 **Fig. 8 Type I MGCs likely encode components of corrinoid-dependent selenium salvage or**
532 **methionine salvage pathways.** The vast majority of type I MGC encoders likely use these
533 methyltransferase systems to recycle methylated selenides formed from the non-enzymatic reaction of SAM

534 with selenol compounds, potentially recycling the methyl group into methionine. A smaller subset of type I
535 MGC encoders in the ref seq database (about 10%) lack known selenium metabolism traits. We propose
536 that these MGCs have evolved to participate in a methionine salvage pathway. MTA: 5'-methylthio-
537 adenosine. His: histidine. A: adenine.
538

539 In summary, we have identified a class of microbial gene clusters encoding
540 components of probable corrinoid *S*- and *Se*-methyltransferase pathways. The type I
541 MGCs are found in phylogenetically and metabolically diverse organisms encompassing
542 >100 microbial genera. Bioinformatic analyses and preliminary biochemical studies
543 suggest type I MGCs participate in methyl-selenide metabolism and/or methionine
544 salvage. Ultimately, this previously unappreciated example of methanogen specific
545 enzymes encoded outside of the archaea could provide insights into the vitally important
546 fields of sulfur and selenium cycling.
547

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

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

558

559 **Data Availability Statement:** All relevant data has been made available in the main text,
560 Supplementary Information, and associated data files.

561

562 **References**

563 1. Kurth JM, Op den Camp HJM, Welte CU. Several ways one goal—methanogenesis
564 from unconventional substrates. Appl Microbiol Biotechnol. 2020 Aug;104(16):6839–
565 54.

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

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

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

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

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

583 7. Mock J, Wang S, Huang H, Kahnt J, Thauer RK. Evidence for a Hexaheteromeric
584 Methylenetetrahydrofolate Reductase in Moorella thermoacetica. J Bacteriol. 2014 Sep
585 15;196(18):3303–14.

586 8. Huwiler SG, Löffler C, Anselmann SEL, Stärk HJ, von Bergen M, Flechsler J, et al.
587 One-megadalton metalloenzyme complex in *Geobacter metallireducens* involved in
588 benzene ring reduction beyond the biological redox window. Proc Natl Acad Sci USA.
589 2019 Feb 5;116(6):2259–64.

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

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

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

599 12. Fu H, Metcalf WW. Genetic Basis for Metabolism of Methylated Sulfur Compounds
600 in Methanosarcina Species. O'Toole GA, editor. J Bacteriol. 2015 Apr 15;197(8):1515–
601 24.

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

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

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

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

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

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

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

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

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

626 22. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify
627 genomes with the Genome Taxonomy Database. Hancock J, editor. Bioinformatics.
628 2020 Mar 1;36(6):1925–7.

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

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

635 25. Roberts DL, Zhao S, Doukov T, Ragsdale SW. The reductive acetyl coenzyme A
636 pathway: sequence and heterologous expression of active
637 methyltetrahydrofolate:corrinoid/iron-sulfur protein methyltransferase from Clostridium
638 thermoaceticum. J Bacteriol. 1994 Oct;176(19):6127–30.

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

641 27. Matthews RG. Cobalamin-Dependent Methyltransferases. Acc Chem Res. 2001
642 Aug 1;34(8):681–9.

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

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

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

654 31. Schilhabel A, Studenik S, Vödisch M, Kreher S, Schlott B, Pierik AY, et al. The
655 Ether-Cleaving Methyltransferase System of the Strict Anaerobe *Acetobacterium*
656 *dehalogenans*: Analysis and Expression of the Encoding Genes. J Bacteriol. 2009 Jan
657 15;191(2):588–99.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

703 48. Hogenkamp HPC. The Photolysis of Methylcobalamin *. Biochemistry. 1966 Feb
704 1;5(2):417–22.

705 49. Veres Z, Tsai L, Scholz TD, Politino M, Balaban RS, Stadtman TC. Synthesis of 5-
706 methylaminomethyl-2-selenouridine in tRNAs: 31P NMR studies show the labile
707 selenium donor synthesized by the selD gene product contains selenium bonded to
708 phosphorus. Proc Natl Acad Sci USA. 1992 Apr;89(7):2975–9.

709 50. Nogueira CW, Zeni G, Rocha JBT. Organoselenium and Organotellurium
710 Compounds: Toxicology and Pharmacology. Chem Rev. 2004 Dec 1;104(12):6255–
711 86.

712 51. Fernandes AP, Wallenberg M, Gandin V, Misra S, Tisato F, Marzano C, et al.
713 Methylselenol Formed by Spontaneous Methylation of Selenide Is a Superior Selenium
714 Substrate to the Thioredoxin and Glutaredoxin Systems. Netto LES, editor. PLoS ONE.
715 2012 Nov 30;7(11):e50727.

716 52. Albers E. Metabolic characteristics and importance of the universal methionine
717 salvage pathway recycling methionine from 5′-methylthioadenosine. :11.

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

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

724 55. Warlick BP, Evans BS, Erb TJ, Ramagopal UA, Sriram J, Imker HJ, et al. 1-
725 Methylthio- D -xylulose 5-Phosphate Methylsulfurylase: A Novel Route to 1-Deoxy- D -
726 xylulose 5-Phosphate in *Rhodospirillum rubrum*. Biochemistry. 2012 Oct
727 23;51(42):8324–6.

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

732

733

734

735