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 pathways that could have global or ecological relevance. Moreover, studying

homologs of methanogen-specific enzymes in non-methanogenic organisms could shed light on the evolution of methanogenic pathways and organisms as well as the potential adaptation of methanogen genes to perform novel functions.

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

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

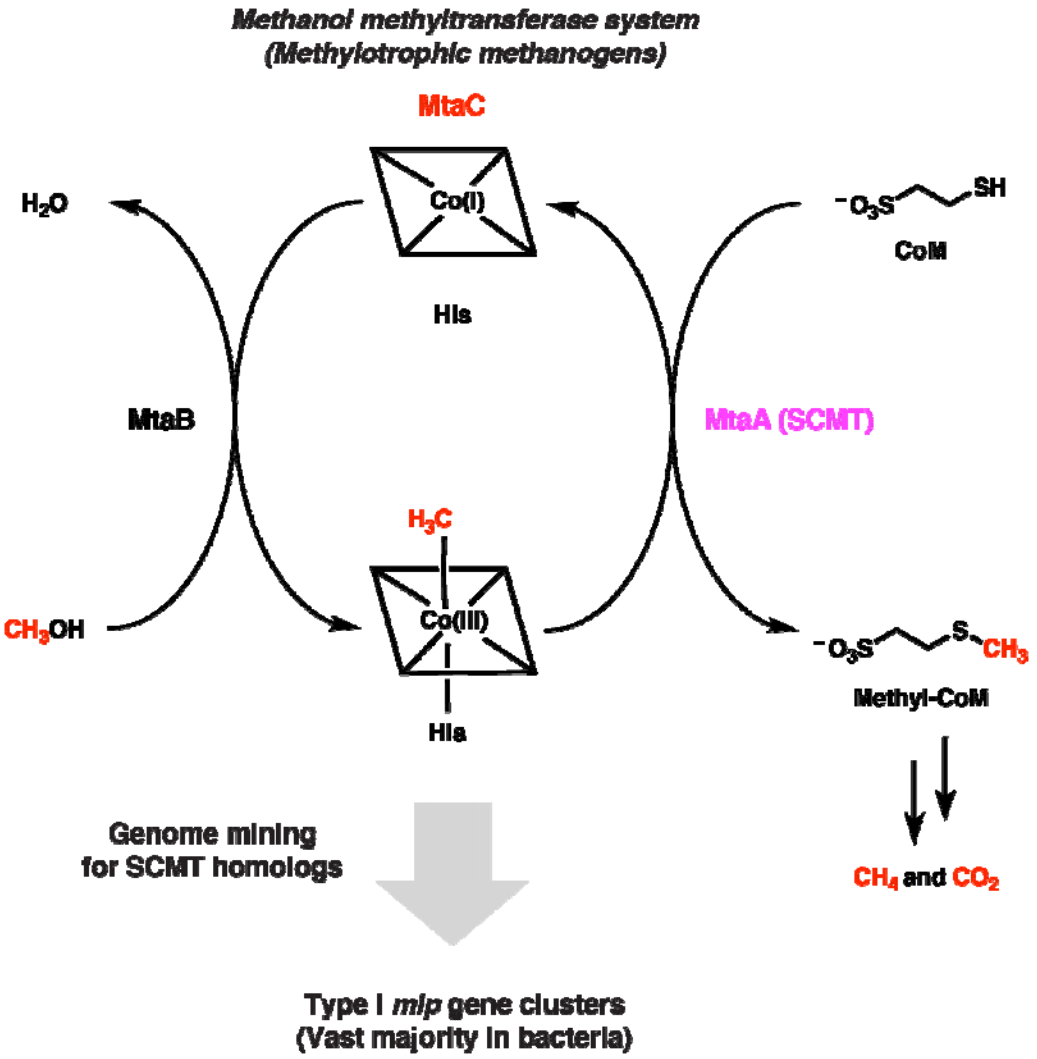


Fig. 1 Searching genomes for the soluble coenzyme M methyltransferase (SCMT) MtaA involved in methylotrophic methanogenesis identifies related enzymes in non-methanogens. Overview of genome mining strategy used in this work. The methanol-specific methanol:corrinoid methyltransferase MtaB transfers the methyl group of methanol to the cognate corrinoid protein MtaC (the corrinoid cofactor

is represented by a square containing a cobalt ion). The SCMT MtaA then transfers to the methyl group from MtaC to coenzyme M (CoM). The methyl group incorporated onto CoM is then disproportionated into methane and carbon dioxide or solely reduced to methane, depending on the organism and environmental conditions.

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

Materials and methods

General methods and procedures

Manipulations and visualizations of genomic data were performed using Geneious Prime 2020.0.4 software (17). BLASTP searches were conducted through the Geneious interface using either National Center for Biotechnology Information (NCBI) nr/rt database or custom databases generated by the user. In all cases we used the BLOSUM62 matrix and a gap opening penalty of 11 and a gap extension penalty of 1. Protein alignments we performed using MUSCLE (18).

Identification of *mlp* gene clusters

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

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

SSN analysis of MGC proteins

We generated all SSNs using the Enzyme Function Initiative Enzyme Similarity Tool (EFI-EST) webtool (19). For each SSN, we first compiled the sequences of characterized superfamily members and superfamily members that are encoded in the 367 MGCs identified above, generating a superfamily-specific “preliminary input list”. We then pruned each list to ensure that only quality sequences (those lacking frameshifts or internal stops) within a certain length window were included. For the Mlp SSN, length window was 200 to 470 amino acids. The resulting sequence list was termed the “input list”. We next input both the appropriate superfamily and input list into the EFI-EST web server. For the Mlp SSN, the superfamily was IPR027980, for the Ram/RACE SSN, it was IPR027980, and for the corrinoid protein SSN it was IPR006158. We selected alignment score thresholds such that most MGC proteins fell into common network clusters, but were still as separated as possible from non-MGC proteins. We selected node representation percent identities to the highest that consistent with efficient use of the Cytoscape GUI on an ordinary laptop MacIntosh computer. For each SSN, we used the UniRef90 option and otherwise default parameters. Once generated, the SSNs were visualized with Cytoscape software (version 3.8.0) using the “preferred” layout (20).

Phylogenetic analysis of *mlp* gene cluster proteins

We compiled proteins encoded by type I MGCs and sequences of related characterized proteins and then aligned the resulting protein sequences using MUSCLE with default parameters, and trimmed the alignments manually. All trees were generated

using FastTree (approximately-maximum likelihood method) with default parameters (21).

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

Phylogenetic membership of type I MGC-encoding organisms

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

Construction of homology models

Clostridium ljungdahlii DSM 13528 Mlps 1, 2, and 3 were submitted to the SWISS-MODEL webserver (23) using *Methanosarcina mazei* MtaA (PDB ID: 4AY7) (24).

Expression and purification of *Clostridium ljungdahlii* MtaA-like protein 2 (*Cl*Mlp2)

Clmlp2 gene was PCR-amplified from *C. ljungdahlii* DSM 13528 genomic DNA and cloned using the PIPE method into pSpeedET, which was then transformed into *E. coli* Rosetta 2(DE3)pLysS. The resulting strain was grown aerobically at 37 °C with 180 RPM shaking in four 4 L baffled flasks, each containing 2 L of LB broth (Research Products International) supplemented with 22 mM potassium phosphate (salts from Sigma Aldrich) buffer pH 7.2, 20 mM D-glucose (Sigma Aldrich), 20 mM $MgCl_2$ (Sigma-Aldrich), 50 μg/mL kanamycin sulfate (VWR International), and 35 μg/mL chloramphenicol (Sigma Aldrich). At a OD_{600nm} of 0.6, IPTG (Teknova) was added to a final concentration of 0.5 mM, and the growth temperature was shifted to 16 °C. After 12 hours of incubation at 16 °C, the cultures were harvested by centrifugation and the resulting pellets washed once with 50 mM MOPS (Sigma Aldrich) pH 7.2 before resuspension in buffer A. Buffer A contained 50 mM imidazole, 22 mM potassium phosphate, 500 mM NaCl, and 10% glycerol (by volume). The resuspended pellet was lysed by sonication using a Branson Digital Sonifier in 120 cycles composed of 5 seconds of sonication followed by 15 seconds of rest. The resulting lysate clarified by centrifugation at 50,000 x g for 45 minutes. The extract was loaded onto a 5 mL Hitrap IMAC FF Ni^{2+}-affinity column (GE Life Sciences), which was washed with buffer A. The elution was conducted with a linear gradient from 100% buffer A to 100% buffer B. Buffer B contained 500 mM imidazole, 22 mM potassium phosphate, 500 mM NaCl, and 10% glycerol. The eluted protein was collected, concentrated with a 5 kDa-cutoff Spin-X UF concentrator (Corning), and diluted at least 10-fold in buffer C. Buffer C consisted of 50 mM MOPS, 10% glycerol, pH 7.2. The resulting protein sample was loaded onto a 5 ml Hitrap Q column, which was washed with buffer C, followed with elution with a linear

gradient from 100% buffer C to 100% buffer D. Buffer consisted of 50 mM MOPS, 500 mM NaCl, 10% glycerol, pH 7.2. The resulting fractions were screened for a polypeptide of the expected molecular mass (39.1 kDa) for *Cl*Mlp2. The identified fractions were concentrated and buffer exchanged into buffer E (50 mM MOPS, 200 mM KCl, 10% glycerol, pH 7.2). The resulting protein preparation was made anoxic under an N_2 atmosphere by the flush-evacuation technique using a Schlenk line. It was then flash frozen at in liquid N_2 and stored at –70 °C until use.

Assay of *Cl*Mlp2

The biochemical assay procedure used was described previously (25). The assays were conducted in a Powerwave XS (BioTek) plate reader heated to 37 °C in a quartz 96-well plate (Molecular Devices). All phases of the assay took place under red light and otherwise in the dark in an MBraun chamber under a 100% N_2 atmosphere. 150 µL reaction mixtures (assembled in the plate and without methyl acceptor) were incubated at 37 °C for 2 minutes and were then initiated with addition of methyl acceptor substrate. Following addition of the methyl acceptor, the reaction composition was 50 mM MOPS pH 7.2, 0.4 mM methylcobalamin (Sigma Aldrich), 4.7 µM *Cl*Mlp2, and methyl acceptor at indicated concentrations. Following initiation, the absorbances at 525, 485, and 800 nm were followed. Rates of reaction were calculated from the maximum rate of A_{525nm} decrease sustained over at least 5 minutes during the 1 hour period for which the reaction was followed. The absorbance change rate was then converted using the $\Delta\varepsilon_{525nm}$ = 8.6 mM^{-1} reported in (25). Manipulations and analyses of the data were performed in GraphPad Prism 9.3.1.

Results

While manually inspecting the genome of *Clostridium ljungdahlii* DSM 13528, a bacterium that has received attention as a potential producer of biofuels, we noticed that this organism encodes multiple members of the UroD superfamily. To see if these genes are closely related to SCMTs, we performed BLASTP searches of the SCMT MtbA (*Methanosarcina barkeri* Fusaro, *Mb*MtbA) against the predicted *C. lungdahlii* DSM 13528 proteome. We identified seven hits with an E value lower than 1e-09 (Table S1). The highest hit was CLJU_RS09375, which showed 32.5% amino acid (aa) identity to *Mb*MtbA. It is worth noting that the SCMTs MtbA, MtaA, and MtsA (all from *M. barkeri* Fusaro) show only about 30% aa identity to each other, suggesting that CLJU_RS09375 could be a close relative of the SCMTs. CLJU_RS09375's amino acid sequence identity to the UroD superfamily phenolic *O*-demethylases OdmB and VdmB was only 17.6% and 17.0%, respectively.

To obtain clues as to the physiological role of CLJU_RS09375, we next examined this gene's genomic context. CLJU_RS09375 is encoded in a gene cluster that includes genes encoding a corrinoid protein (CLJU_RS09370), a "bacterial-type" corrinoid protein reductive activase (Ram/RACE protein) (CLJU_RS09380), and two additional Mlps (CLJU_RS09360 and CLJU_RS09365) (Fig. 2a). This gene cluster has been independently predicted to form an operon by OperonDB, an automatic operon prediction database (26). Because this gene cluster encodes multiple Mlps, we named it the *mlp* gene cluster (MGC). We named the three Mlps in the gene cluster Mlp1, Mlp2,

and Mlp3, (specifically, *Cl*Mlp1, 2, and 3) in order of their appearance in the gene cluster from the presumed transcriptional start site (see Fig. 2a).

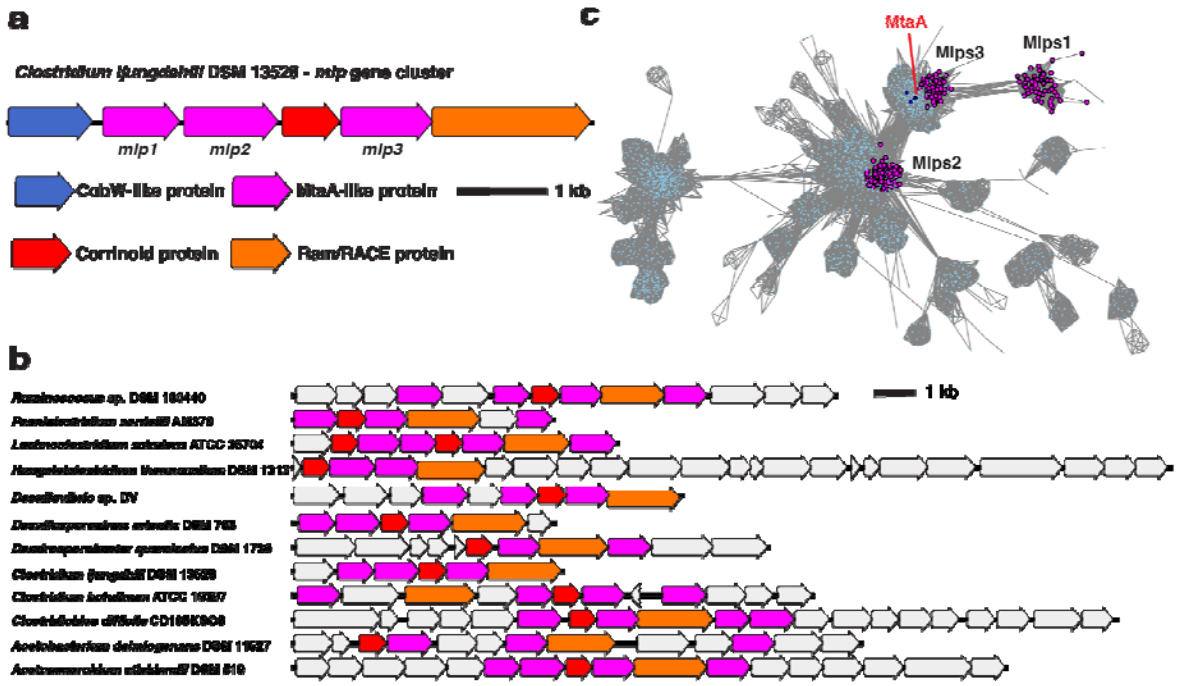


Fig. 2 Mlp gene clusters (MGCs) are similar in genetic composition and protein seqence. a The MGC from *C. ljungdahlii*. **b** Representative type I MGC. Coloring of the core MGC genes is the same as in a, but non-core genes are colored grey. **c** A protein SSN of the UroD superfamily (IPR000257) showing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

Given the relatively high sequence similarity of the MGC Mlps to SCMTs, it seems reasonable to hypothesize that these enzymes are corrinoid methyltransferases. This is consistent with the presence of genes encoding a corrinoid protein and a Ram/RACE protein in the MGC. The former is a methyl group carrier (27), while the latter is an ATP-dependent corrinoid protein reductase that reductively activates the corrinoid protein to

its catalytically-active Co(I) state (28,29). Corrinoid methyltransferases, corrinoid proteins, and Ram/RACE proteins are frequently found colocalized in gene clusters (30–32). In bacteria, gene clusters encoding components of corrinoid methyltransferase pathways are involved in methylotrophic metabolism and participate in transferring a methyl group from a donor substrate to tetrahydrofolate (H_4folate) via a methyl-corrinoid protein intermediate, a reaction catalyzed by a methyl-corrinoid protein:H_4folate methyltransferase (15,30). In most of the bacterial methyltransferase pathways examined to date, the H_4folate methylase is encoded in the same gene cluster as the corrinoid protein and methyl donor methyltransferase (30,31). It is therefore surprising that H_4folate methyltransferase is missing from the *C. ljungdahlii* MGC. While it is possible that a H_4folate methyltransferase encoded elsewhere in the *C. ljungdahlii* genome could cooperate with the MGC methyltransferase system, it is also possible that H_4folate is not involved in the MGC methyltransferase pathway or that one of the MGC Mlps is an undiscovered H_4folate methyltransferase.

In parallel with the identification of the *C. ljungdahlii* MGC, we located similar gene clusters in *C. sporogenes* DSM 795 and *Hungateiclostridium* (*Clostridium*) *thermocellum* DSM 1313. Like the MGC from *C. ljungdahlii*, the gene clusters in *C. sporogenes* and *H. thermocellum* encode multiple Mlps (each with about 30% identity to the SCMTs), a corrinoid protein and a Ram/RACE protein (Fig. 2b). The gene clusters from *C. sporogenes* and *H. thermocellum* also lack a H_4folate methylase. However, unlike *C. ljungdahlii*, which encodes H_4folate methylases elsewhere in its genome, homologs of H_4folate methylases are absent from sequenced *C. sporogenes* and *H. thermocellum* genomes, suggesting that H_4folate might not be involved in the presumptive

methyltransferase pathways encoded by these gene clusters. Because of the similarity of the *C. sporogenes* and *H. thermocellum* gene clusters to the *C. ljungdahlii* MGC, we refer to these gene clusters as the *Cs*MGC, *Ht*MGC, and *Cl*MGC, respectively.

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

We next asked if these MGCs are homologous or analogous. Because components of a corrinoid methyltransferase pathway are commonly encoded in gene clusters (14,15,30–32), it would not be surprising to find that evolutionary pressure to colocalize corrinoid methyltransferase genes resulted in MGCs that appear only superficially similar, but share no evolutionary relationship. Alternatively, the common appearance and genetic content of the MGCs could suggest that these gene clusters are descended from a common ancestor (i.e. are homologous), which would also suggest they likely have a shared function in diverse organisms. To address this question, we first compiled a list of 367 organisms encoding likely MGCs (see materials and methods). To preliminarily assess the similarity of the proteins encoded by these MGCs, we constructed several protein sequence similarity networks (SSNs) (19). We reasoned that if MGCs are homologous, each conserved component of the identified MGCs should preferentially form its own cluster in an SSN of the protein family to which that

component belongs. For instance, in an SSN of the corrinoid protein superfamily, the MGC corrinoid proteins should form their own cluster within the SSN. The same should be true for the Ram/RACE proteins in an SSN of their superfamily. Because there are multiple *mlp* genes in each MGC, there are two SSN analysis outcomes that are consistent with the homology hypothesis. First, all of the Mlps from the MGCs could be grouped within a single cluster in the SSN. This would be expected to be the case if the multiple *mlp* genes in each MGC arose from gene duplication. Second, the MGC Mlps could each represent different lineages, and could therefore likely form distinct clusters within an SSN.

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

Because the overwhelming majority (>95%) of MGCs examined here appear to be type 1 MGCs, we selected the type I MGCs for further study. The fact that the type 1 MGC Mlps fall into three separate clusters in the SSN hints that they might represent

three distinct Mlp lineages, rather than a single lineage within the UroD superfamily that arose from recent gene duplication events. To solidify this inference, we constructed a phylogenetic tree of the Mlps and other members of the UroD superfamily that are close relatives of known corrinoid methyltransferases (Figure 3a). The phylogenetic tree showed that the Mlps 1, 2, and 3 are situated in a large branch that includes the SCMTs, to the exclusion of other known UroD superfamily corrinoid methyltransferases such as the aromatic *O*-demthylases (13). When rooted (Figure 3b), we found that the Mlps 1, 2 and 3 formed distinct, well-supported clades. Interestingly, the Mlps2 form a sister clade to the clade containing the SCMTs and the Mlps1 and Mlps3. Within this clade, the Mlps1 and Mlps3 form sister clades emerging from the SCMTs, which therefore appear to be paraphyletic.

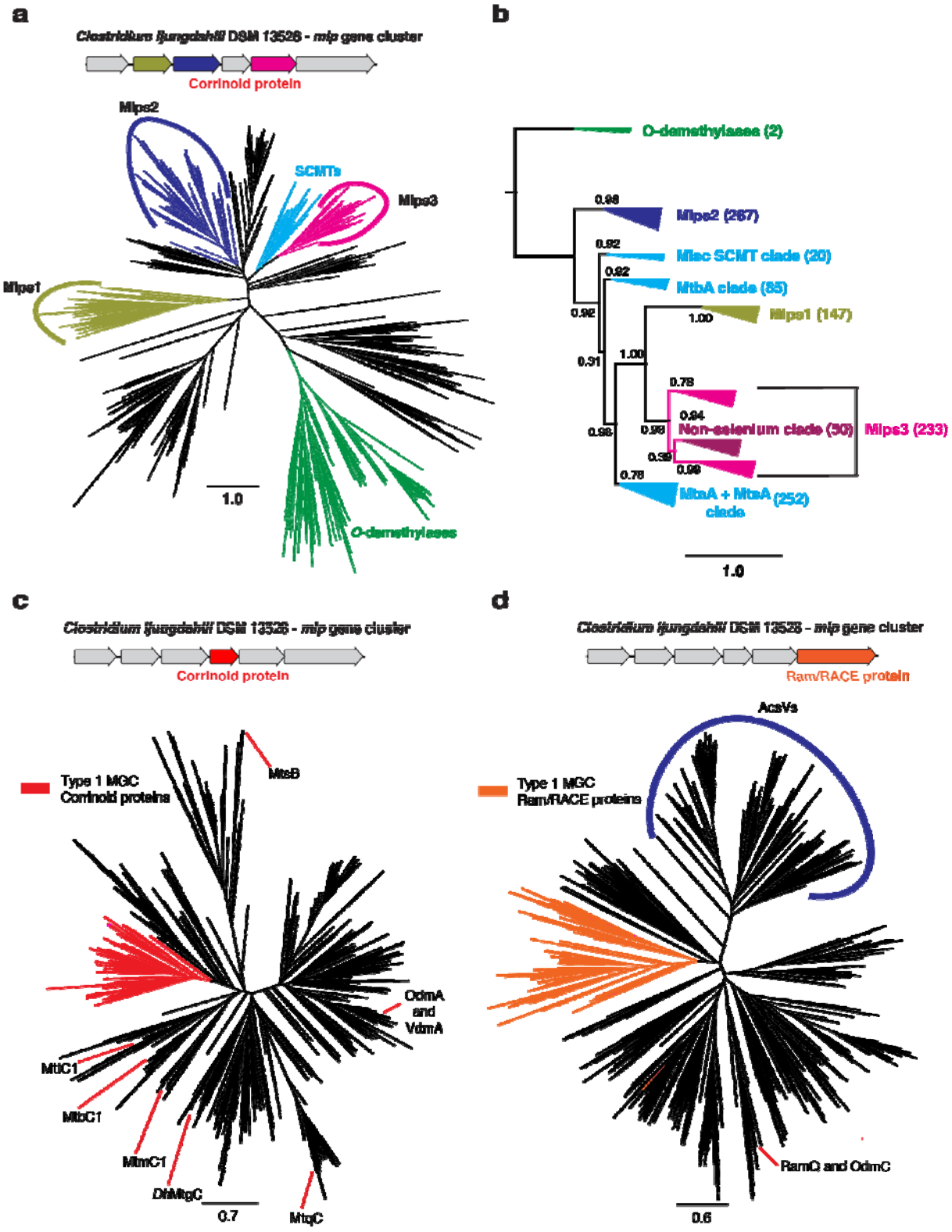


Fig. 3 Type I MGCs encode orthologous proteins. a An approximately-maximum likelihood (FastTree) phylogenetic tree of members of the UroD superfamily methyltransferases. The FastTree

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

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

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

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

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

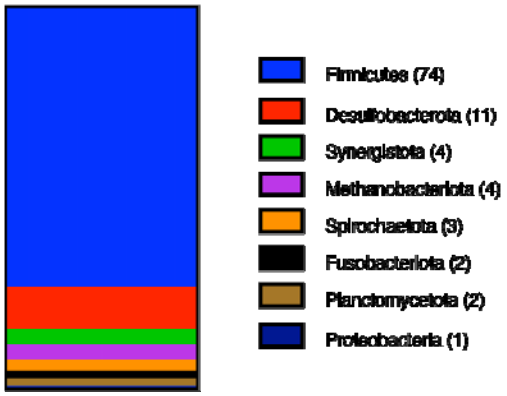


Fig. 4 Type I MGCs are found in several prokaryotic Phyla. Phylum composition of genera encoding type I MGCs. 101 genera are represented. The number of type I MGC-encoding genera within each phylum is enclosed in parentheses.

Many of the organisms encoding type I MGCs are either known acetogens or encode the proteins necessary for acetogenesis. These include the acetogenic *Clostridium* spp., *Sporomusa* spp., certain *Clostridioides difficile* strains, and acetogenic *Treponema* strains. It should be emphasized, however, that many known acetogens do not encode MGCs and many of the organisms that encode MGCs lack the critical *acsA* and *acsB* (genes encoding bacterial anaerobic carbon monoxide dehydrogenase and bacterial acetyl-CoA synthase, respectively) that are required for acetogenesis in all known instances.

We next turned our attention to analyzing additional components of the type 1 MGCs and proposing a biochemical function for these genes. The UroD superfamily includes corrinoid methyltransferases that (de)methylate several heteroatoms including halogens, sulfur atoms, and aryl oxygens (10,12,14,15,33). Several lines of evidence suggest that the type 1 MGCs encode a corrinoid *S*- or *Se*-methyltransferase system.

First, the closest relationship between known UroD superfamily proteins and the type 1 *mlp* cluster Mlps is with the SCMTs (Fig. a and b), which are *S*-methyltransferases. Additional evidence comes from examination of the enzyme's Zn^{2+}-binding site. This Zn^{2+} cofactor coordinates the heteroatom of a substrate, activating it for (de)methylation (24,33,34). The primary coordination sphere of the Zn^{2+} cofactor is His-Cys-Cys in *S*-methyltransferases from the UroD superfamily and beyond, while in *O*-methyltransferases, one or both of the Cys ligands are substituted by carboxylates (Glu or Asp) or are substituted for non-coordinating residues (15,24,33,34). The His-Cys-Cys coordination environment favors the binding of a soft ligand such as a selenium or sulfur atom to the Zn^{2+} center, while substitution of Cys for a carboxylate favors binding of an oxygen atom. The coordination sphere of the type 1 MGC Mlps is (with the exception of a few of the Mlps1) invariably His-Cys-Cys, suggesting that these proteins are *S*- or *Se*-methyltransferases. We constructed homology models of each of the *C. ljungdahlii* Mlps in order to confirm that the His-Cys-Cys triads are likely appropriately positioned to coordinate Zn^{2+} (Fig. S3). Finally, examination of the non-core components of type I MGCs (genes that are found in type I MGCs, but do not encode Mlps, corrinoid proteins, or Ram/RACE proteins) consistently reveals genes involved in sulfur metabolism, including methionine aminotransferases, sulfurtransferases, methionine transporters, and close homologs of the recently-described MarHDK enzyme that catalyzes C–S bond cleavage in the production of ethylene from methylthioethanol (35). These observations suggest a potential role for type I MGCs in methyl-sulfur metabolism.

However, consideration of additional evidence suggests that most type I MGCs function in methyl-selenium metabolism. For instance, type 1 MGCs commonly contain

genes provisionally linked to selenium metabolism, such as the recently characterized "dicubane cluster protein" (36–38). Several type 1 MGCs (Fig. 5) contain genes encoding authentic machinery for selenocysteine synthesis, encoding, and decoding (SelABCD), as well as the enzymes of the glycine reductase system, two of which have selenocysteine residues (39,40). A strikingly high percentage (87%) of sequenced species that encode type I MGCs also encode selenocysteine biosynthesis machinery within the same genome. For comparison, a recent study found that only about 21.5% of bacteria contain the requisite genes for selenocysteine biosynthesis, and even in the selenocysteine-rich Clostridia, only 42% of 184 genomes surveyed encoded selenocysteine biosynthesis machinery (36). Not all organisms that make use of selenium encode selenocysteine biosynthetic machinery—some organisms that use the selenium cofactor only encode selenide-water dikinase (SelD) (36). We therefore manually inspected genomes that contain a type I MGC, but lack SelAB or C, finding that several genomes lacked *selABC*, but had *selD* as well as selenium cofactor-dependent enzymes. As noted below, type 1 MGCs from *C. ljungdahlii* and *C. drakei* are upregulated during autotrophic growth on H_2 + CO_2, a condition where there is increased demand for selenoprotein biosynthesis (41). The Cys-Cys-His coordination of the active site Zn^{2+}, in addition to being suggestive of binding a sulfur nucleophile/methyl-sulfur methyl donor, could also be compatible with a selenium nucleophile/methyl-selenide methyl donor. Most importantly, during the preparation of this manuscript, we became aware of a 2004 publication that used proteomic and genetic methods to implicate three type I MGC-encoded proteins (the corrinoid protein and two Mlps2) in the utilization of dimethyl-selenide as a selenium source in the

hydrogenotrophic methanogen *Methanococcus voltae* (42). The authors suggested that the Mlps might demethylate (di)methyl-selenide and transfer the methyl group onto the MGC corrinoid protein, liberating selenide for incorporation into biological macromolecules. While the authors did not biochemically characterize the Mlps in this study, their data strongly support a role for the type I MGCs in methyl-selenol utilization.

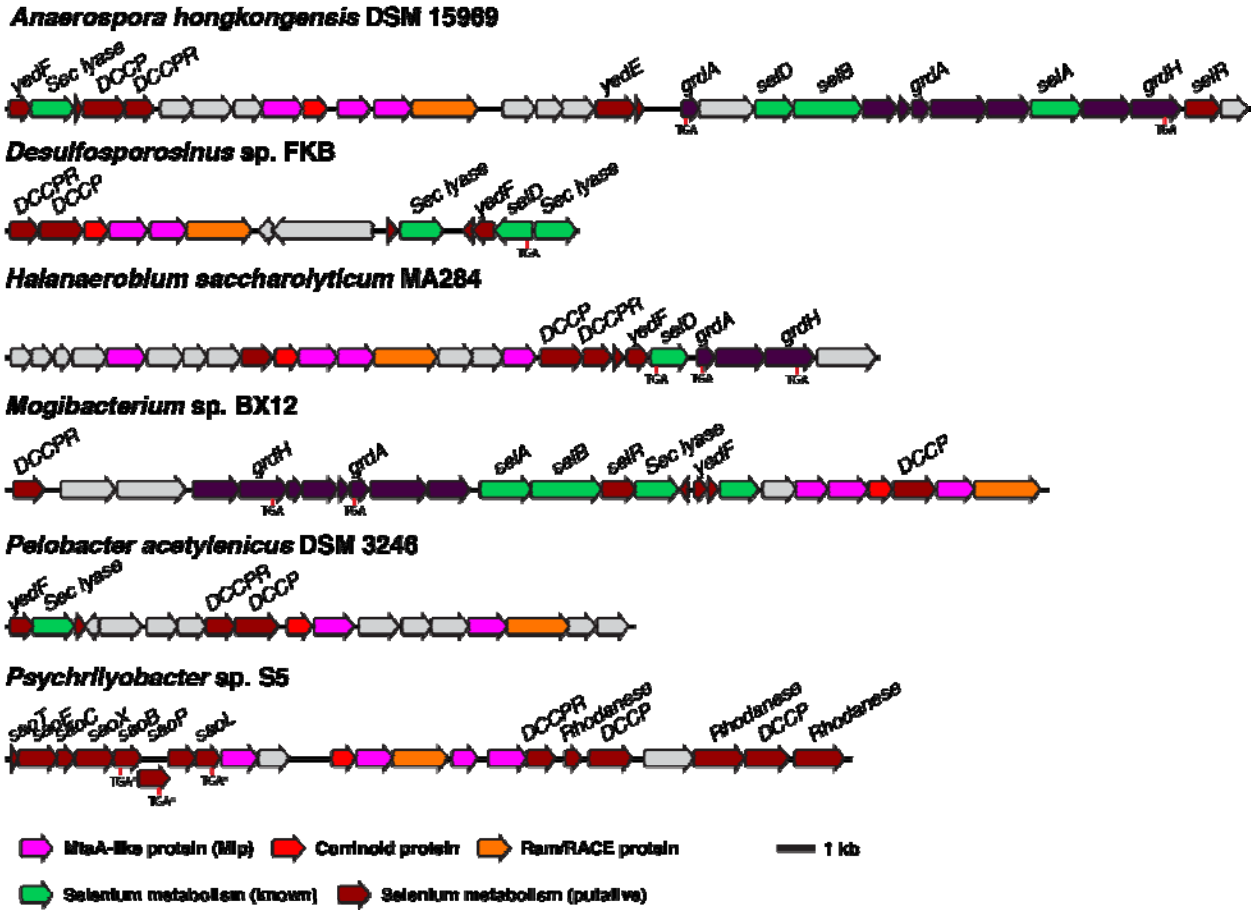


Fig. 5 Type I MGCs often encode genes involved in selenium metabolism. Selenium metabolism gene show are either close homologs of genes with experimentally-confirmed roles in selenium metabolism ("Selenium metabolism (known)") or have been associated with selenium metabolism by bioinformatic methods ("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.

Because type 1 MGCs are present in several model organisms that have been the subjects of proteomic and transcriptomic studies, we examined the expression of these genes in existing datasets for *C. botulinum*, *C. ljungdahlii*, and *H. thermocellum* to gain additional information regarding their functions. In all three organisms, the MGCs are expressed at a detectable level under all conditions examined. First, we found that the *C. botulinum* MGC is upregulated during growth at the organism's optimal growth temperature of 37 °C compared to growth at 15 °C (transcripts 5.27-68.6 fold upregulated at 37 °C) (43). In *C. ljungdahlii*, we found that the type 1 MGC is transcriptionally upregulated roughly 3-5 fold during autotrophic growth on H_2 + CO_2 or CO versus heterotrophic growth on fructose (44–46). The same pattern is found in *C. drakei* (6 to13 fold transcriptional upregulation on H_2 + CO_2 compared to fructose) (41). In *H. thermocellum*, the type 1 MGC is upregulated during growth with supplemental acetate in a strain from which the FeFe hydrogenase maturation enzyme *ΔhydG* had been deleted (47). These findings imply that type 1 MGCs are expressed, indicating their physiological relevance, and respond to environmental and genetic perturbations in industrially and medically significant organisms.

Finally, we sought characterize the activity of an Mlp to obtain insight into the function of type I MGCs. Niess et al.(42) observed that deletion of the gene encoding an Mlp2 (*sdmC*) resulted in loss of the ability of *M. voltae* to use dimethyl-selenide as a selenium source. As this suggests the function of Mlps2 may be to demethylate

dimethyl-selenide and/or methyl-selenol, we therefore selected *Cl*Mlp2 for cloning, expression, and biochemical assays. Recombinant ClMlp2 was assayed with sodium selenide and methylcob(III)lalamin under a strictly anoxic nitrogen atmosphere as described in the *Materials and Methods*. In brief, this UV-visible spectroscopy assay assesses methylation of a methyl acceptor (methylcob(III)alamin:acceptor methyltransferase activity) by measuring the disappearance or shift of a large methylcob(III)alamin absorbance band (λ_{max} ~ 525 nm) when the oxidation and/or ligation state of the cobalt ion is changed. For convenience, we assayed for methylcob(III)alamin:selenide methyl transfer (Fig 6a). This is the reverse reaction of the hypothesized physiological direction of the reaction, which is substantially more difficult to assay. Addition of sodium selenide (666 μM) to the reaction mixture resulted in a slow but significant decrease in A_{525nm} (Fig. 6b). Control reactions lacking either *Cl*Mlp2 or sodium selenide exhibited rates of A_{525nm} decrease that were 13.7- and 8.3-fold slower, respectively, and are attributable to artefactual photolysis of methylcob(III)alamin inherent to measurement of absorbance(48). Addition of sodium sulfide (666 μM) instead of sodium selenide resulted in a 4.3-fold slower rate, which could suggest the enzyme also possesses low sulfide methylating activity. Starting and ending spectra of the reactions only showed significant spectral changes in the selenide and sulfide reactions. In both cases, the spectra showed a shift in λ_{max} from 520 to 473 nm indicative of formation of base-on cob(II)alamin and therefore demethylation (Fig. 6c). The net steady state formation of cob(II)alamin with minimal accumulation of other intermediates (indicated by a steady absorbance trace at the methylcob(III)alamin/cob(II)alamin isosbestic point of 485 nm (Fig S4)) is likely the

result of rapid oxidation of cob(I)alamin, the expected reaction product. Steady-state kinetic analysis revealed an apparent k_{cat} of 0.778 min^{-1} and an apparent K_M for selenide of 79.8 μM with a selenide substate inhibition constant (K_I) of 3830 μM (Fig. 6d). The low k_{cat} is likely due the use of methylcobalamin (rather than the physiological methyl-corrinoid protein) as a methyl donor. The low selenide K_M, which is comparable to the 46 μM selenide K_M measured for *E. coli* selenophosphate synthase (49) suggests that methyl-selenol and/or dimethylselenide are physiological substrates of *Cl*Mlp2. Together, these biochemical studies of *Cl*Mlp2 support the hypothesis that type I MGCs are involved in methyl-selenide metabolism.

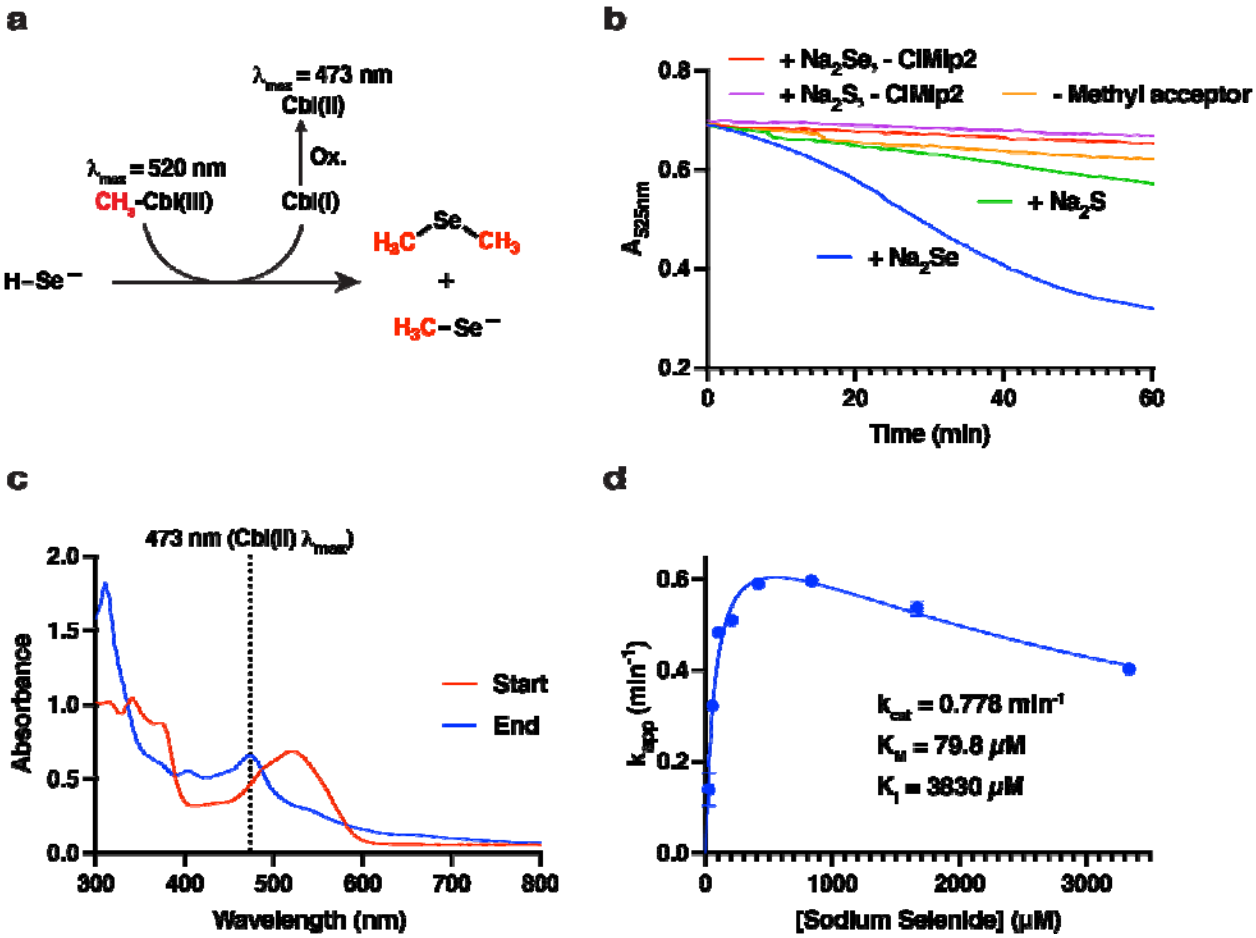


Fig. 6 *C. ljungdahlii* (Cl)Mlp2 catalyzes methylcobalamin:selenide methyl transfer a Schematic of the reaction. Cbl: cobalamin. "Ox." signifies the observed formation of cob(II)alamin, presumably from nonenzymatic reaction of small amounts of diselenides with highly labile cob(I)alamin. **b** 525 nm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

Discussion

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

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

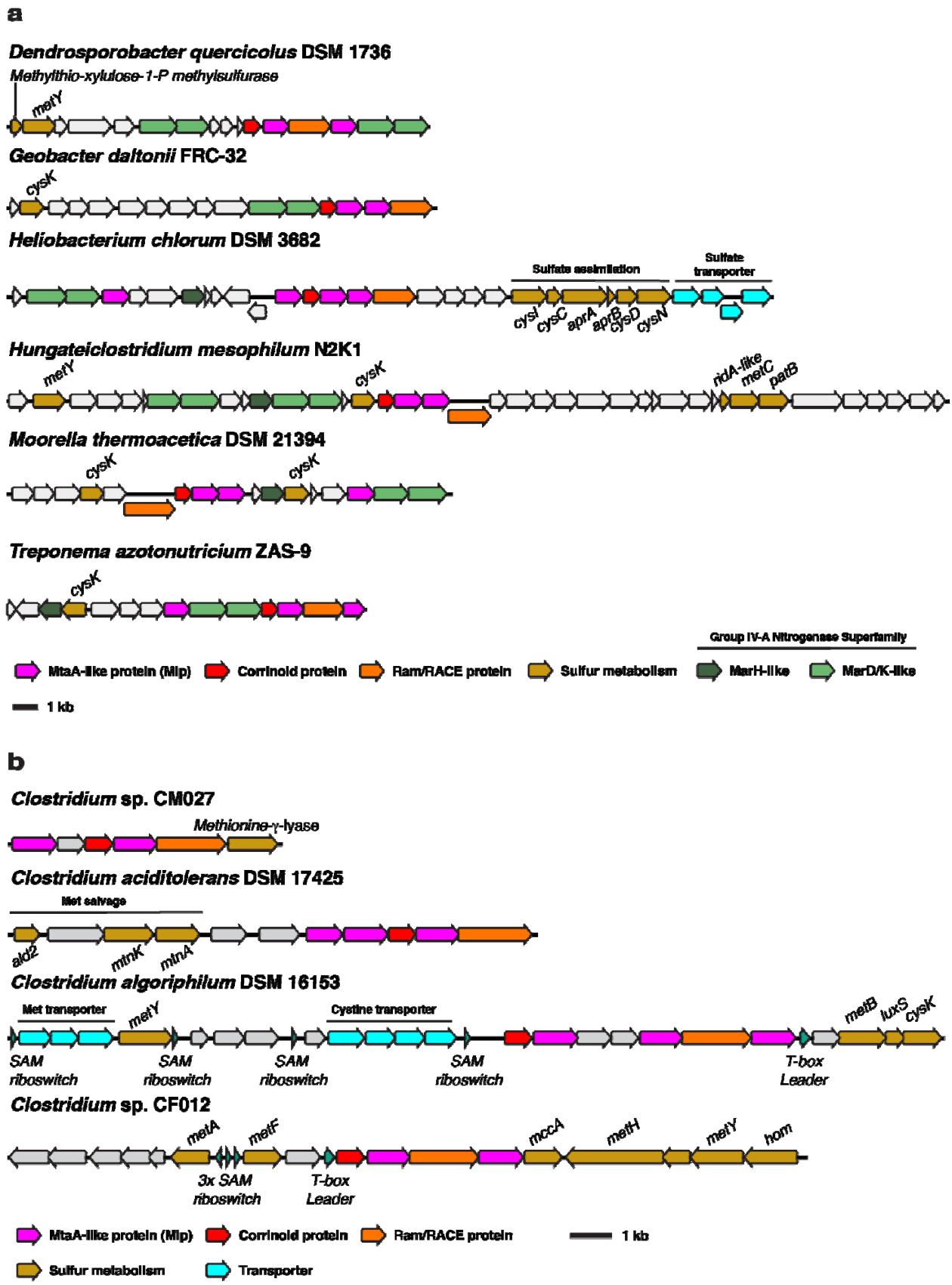
The potential physiological roles of such reactivity may be explored through considering features of the organisms that possess type I MGCs. It seems clear from earlier studies that type I MGCs can function in methyl-selenol valorization (39). However, those authors implicated only the corrinoid protein and two closely-related Mlps2 in dimethyl-selenide utilization—the more distantly related *M. voltae* Mlp1 and Mlp3 (which are also present in the *M. voltae* type I MGC) appear to have gone unnoticed. It seems likely that the full type I MGC contains methyltransferases for demethylation of several methyl-selenium compounds. While the genetic data in Ref. 39 implicates dimethyl-selenide and/or methyl-selenol as substates of the type I MGC Mlps2 (an inference that our assay of *Cl*Mlp2 supports), we propose that the Mlps1 and Mlps3 may be responsible for demethylation of other methyl-selenium compounds such as methyl-selenocysteine and selenomethionine and concomitant methylation of the gene clusters' encoded corrinoid proteins. The methylated corrinoid protein must be demethylated and the methyl group transferred to a methyl-accepting metabolite if the pathway is to continue to operate. Accordingly, we propose that the methyl-selenide-derived methyl group is subsequently transferred from the methylated corrinoid protein to a thiol containing compound such as sulfide. The resulting methyl-sulfur metabolite could then be excreted or recycled into methionine (or, in the type I MGC encoders from the Methanococcales, Mlp3 could methylate coenzyme M, making a presumably small contribution to methane production). Methyl-selenium of various forms are present in various environments, including (based on dietary considerations) the human gut (50). It may be that most type I MGCs function to valorize these methyl-selenium compounds. Nonetheless, we believe an additional explanation is required, as many organisms

(such as *C. ljungdahlii* and *M. voltae*) upregulate their type I MGCs even when methyl-selenium compounds are not added to growth media. This observation suggests that there may be an endogenous source of methyl-selenium compounds. Indeed, it has been suggested and evidentially supported that the selenide generated as an intermediate in selenoprotein biosynthesis can react non-enzymatically with *S*-adenosyl-methionine (SAM) (51). This presents a problem for a cell that has a high demand for selenium metabolism: methyl selenide formation from SAM both wastes valuable selenium and leads to a potentially toxic dead-end metabolite. The enzymes encoded by type I MGCs could hypothetically solve this quandary by recycling the methyl selenide into methionine.

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

At present, evidence from genomic context suggests that the non-selenium MGCs could be involved in a methionine salvage pathway, defined as a metabolic pathway that converts methylthio-adenosine (MTA) into methionine (52). MTA is a toxic by-product of the biosynthesis of key metabolites derived from *S*-adenosylmethionine (SAM) including polyamines, *N*-acyl serine homolactones, and secondary metabolites (53). Methionine salvage pathways therefore support sulfur metabolism both by detoxifying a by-product and preserving the metabolically expensive methylthiol group of MTA (54). The so-called “universal” methionine salvage pathway is O_2-dependent, and methionine salvage pathways that operate under anaerobic conditions have only recently begun to be discovered (52). The anaerobic methionine salvage pathway from *Rhodospirillum rubrum* was recently shown to dependent on the methylthio-alkane reductase MarHDK system, homologs of the nitrogenase subunits NifHDK, respectively (35). *MarHDK*-like genes belonging to the nitrogenase superfamily IV-A group are remarkably prevalent in or near type I MGCs from the non-selenium clade and when not present in the type I MGC itself, they are invariably present elsewhere in the genome in question (Fig. 7a). Although the enzymatic activity of the IV-A group nitrogenase superfamily members is not known (genetically characterized MarDK are in the IV-C group), the relatively high sequence identity (~40-50% amino acid identity) of proteins from group IV-A to those in IV-C could suggest a similar function in the reduction of carbon-sulfur bonds. This is consistent with their typical genomic context, which usually includes various genes putatively involved in sulfur metabolism. This suggests that the non-selenium type I MGCs encode enzymatic machinery for either methionine salvage, sulfur valorization from methyl-sulfide compounds, or both. This inference is further

supported by the c. The *Clostridium algoriphilum* DSM 16153 MGC (which encodes an Mlp3 from the non-selenium clade) includes three predicted SAM riboswitches upstream of the MGC, as well as genes predicted to encoded homocysteine synthase, cysteine synthase, a methionine ABC transporter, and a *S*-ribosylhomocysteine lyase. *Clostridium* sp. CF012 possesses another highly suggestive non-selenium MGC, which contains five SAM riboswitches, a T-box leader, and a near-complete methionine biosynthetic pathway, including a 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 (Mip)
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
aldi2
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 (Mip)
Corrinoid protein
Ram/RACE protein
1 kb
Sulfur metabolism
Transporter

647

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

As noted above, the sequences of the Mlps (in particular their Zn^{2+}-binding sites) are compatible with (de)methylation of both methyl-sulfur compounds and of methyl-selenium compounds. Additionally, the sporadic distribution of type I MGCs across phylogenetic trees is potentially consistent with a role in methionine salvage, as such pathways are dispensable under many growth conditions (54) and are known to be not strictly conserved even in closely related strains. Indeed, many organisms lack a methionine salvage pathway, and in *E. coli*, methionine salvage is a strain-specific feature (56).

We propose how a type I MGC-encoded methionine salvage pathway could operate in Fig. 8, considering the simpler case of a type I MGC that encodes only two Mlps, such as the gene cluster from *H. thermocellum*. We suggest that MTA is metabolized to an unknown intermediate (R^1-SCH_3). One of the Mlps then catalyzes methyl transfer from R^1-SCH_3 to the corrinoid protein. The second Mlp then transfers the methyl group

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

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

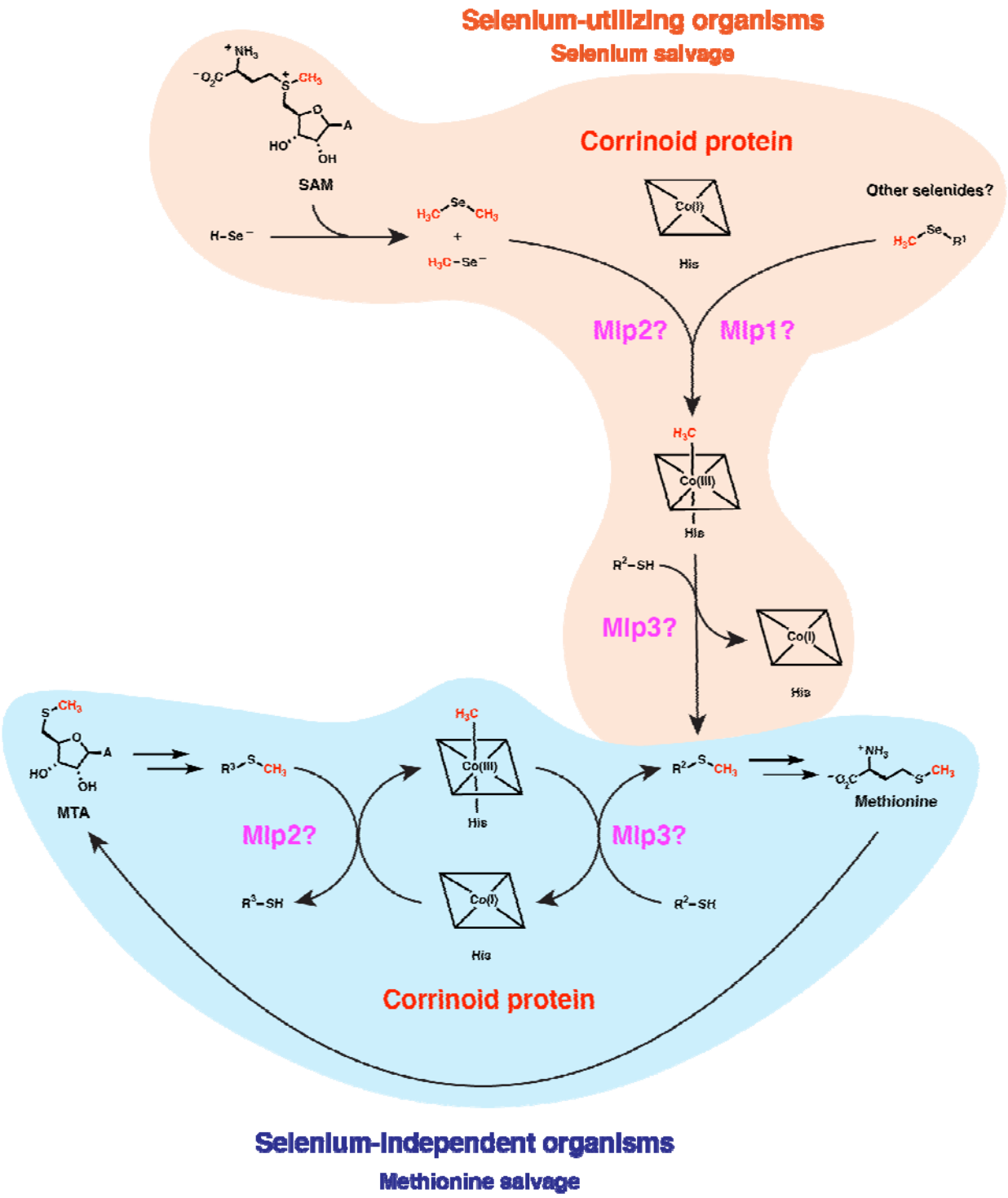


Fig. 8 Type I MGCs likely encode components of corrinoid-dependent selenium salvage or methionine salvage pathways. The vast majority of type I MGC encoders likely use these methyltransferase systems to recycle methylated selenides formed from the non-enzymatic reaction of

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

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

Acknowledgements: D.J.K. acknowledges support from National Institutes of Health Training Grant #5T32GM095450. This work was supported by a Howard Hughes Medical Institute (HHMI)-Gates Faculty Scholar Award (OPP1158186) and a National Science foundation (NSF) Alan T. Waterman Award (CHE-20380529) to E.P.B. E.P.B. is Howard Hughes Medical Institute Investigator. We also note that this article is subject to HHMI's Open Access to Publications policy. HHMI lab heads have previously granted a nonexclusive CC BY 4.0 license to the public and a sublicensable license to HHMI in their 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, Supplementary Information, and associated data files.

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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