

# Lawrence Berkeley National Laboratory

## LBL Publications

### Title

Expanded Phylogenetic Diversity and Metabolic Flexibility of Mercury-Methylating Microorganisms.

### Permalink

<https://escholarship.org/uc/item/4c51571z>

### Journal

mSystems, 5(4)

### ISSN

2379-5077

### Authors

McDaniel, Elizabeth  
Peterson, Ben  
Stevens, Sarah  
et al.

### Publication Date

2020-08-18



### DOI

10.1128/mSystems.00299-20

Peer reviewed



# Expanded Phylogenetic Diversity and Metabolic Flexibility of Mercury-Methylating Microorganisms

 Elizabeth A. McDaniel,<sup>a</sup> Benjamin D. Peterson,<sup>a,f</sup> Sarah L. R. Stevens,<sup>c,d</sup> Patricia Q. Tran,<sup>a,b</sup> Karthik Anantharaman,<sup>a</sup>  
 Katherine D. McMahon<sup>a,e</sup>

<sup>a</sup>Department of Bacteriology, University of Wisconsin—Madison, Madison, Wisconsin, USA

<sup>b</sup>Department of Integrative Biology, University of Wisconsin—Madison, Madison, Wisconsin, USA

<sup>c</sup>Wisconsin Institute for Discovery, University of Wisconsin—Madison, Madison, Wisconsin, USA

<sup>d</sup>American Family Insurance Data Science Institute, University of Wisconsin—Madison, Madison, Wisconsin, USA

<sup>e</sup>Department of Civil and Environmental Engineering, University of Wisconsin—Madison, Madison, Wisconsin, USA

<sup>f</sup>Environmental Chemistry and Technology Program, University of Wisconsin—Madison, Madison, Wisconsin, USA

**ABSTRACT** Methylmercury is a potent bioaccumulating neurotoxin that is produced by specific microorganisms that methylate inorganic mercury. Methylmercury production in diverse anaerobic bacteria and archaea was recently linked to the *hgcAB* genes. However, the full phylogenetic and metabolic diversity of mercury-methylating microorganisms has not been fully unraveled due to the limited number of cultured experimentally verified methylators and the limitations of primer-based molecular methods. Here, we describe the phylogenetic diversity and metabolic flexibility of putative mercury-methylating microorganisms by *hgcAB* identification in publicly available isolate genomes and metagenome-assembled genomes (MAGs) as well as novel freshwater MAGs. We demonstrate that putative mercury methylators are much more phylogenetically diverse than previously known and that *hgcAB* distribution among genomes is most likely due to several independent horizontal gene transfer events. The microorganisms we identified possess diverse metabolic capabilities spanning carbon fixation, sulfate reduction, nitrogen fixation, and metal resistance pathways. We identified 111 putative mercury methylators in a set of previously published permafrost metatranscriptomes and demonstrated that different methylating taxa may contribute to *hgcA* expression at different depths. Overall, we provide a framework for illuminating the microbial basis of mercury methylation using genome-resolved metagenomics and metatranscriptomics to identify putative methylators based upon *hgcAB* presence and describe their putative functions in the environment.

**IMPORTANCE** Accurately assessing the production of bioaccumulative neurotoxic methylmercury by characterizing the phylogenetic diversity, metabolic functions, and activity of methylators in the environment is crucial for understanding constraints on the mercury cycle. Much of our understanding of methylmercury production is based on cultured anaerobic microorganisms within the *Deltaproteobacteria*, *Firmicutes*, and *Euryarchaeota*. Advances in next-generation sequencing technologies have enabled large-scale cultivation-independent surveys of diverse and poorly characterized microorganisms from numerous ecosystems. We used genome-resolved metagenomics and metatranscriptomics to highlight the vast phylogenetic and metabolic diversity of putative mercury methylators and their depth-discrete activities in thawing permafrost. This work underscores the importance of using genome-resolved metagenomics to survey specific putative methylating populations of a given mercury-impacted ecosystem.


**KEYWORDS** genomics, mercury, metagenomics, methylmercury

**Citation** McDaniel EA, Peterson BD, Stevens SLR, Tran PQ, Anantharaman K, McMahon KD. 2020. Expanded phylogenetic diversity and metabolic flexibility of mercury-methylating microorganisms. *mSystems* 5:e00299-20. <https://doi.org/10.1128/mSystems.00299-20>.

**Editor** Angela D. Kent, University of Illinois at Urbana—Champaign

**Copyright** © 2020 McDaniel et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Elizabeth A. McDaniel, [elizabethmcd93@gmail.com](mailto:elizabethmcd93@gmail.com).

 Expanded Phylogenetic Diversity and Metabolic Flexibility of Mercury-Methylating Microorganisms

**Received** 6 April 2020

**Accepted** 29 July 2020

**Published** 18 August 2020

Methylmercury (MeHg) is a potent neurotoxin that biomagnifies upward through food webs. Inorganic mercury deposited in sediments, freshwater, and the global oceans from both natural and anthropogenic sources is converted to bioavailable and biomagnifying organic MeHg (1, 2). The majority of MeHg exposure stems from contaminated marine seafood consumption (3), which is the most problematic for childbearing women and results in severe neuropsychological deficits when developing fetuses are exposed (4, 5). MeHg production in the water column and sediments of freshwater lakes is one of the leading environmental sources that causes fish consumption advisories (6, 7). Biotransformation from inorganic mercury to organic methylmercury has been thought to be carried out by specific anaerobic microorganisms, namely, sulfur-reducing (SRB) and iron-reducing (FeRB) bacteria and methanogenic archaea (8–11). Early identification of mercury-methylating microorganisms relied on testing isolates cultured from anaerobic sediments for MeHg production after spiking with Hg (11, 12). Associations of microbial community structure with Hg speciation and biogeochemical characteristics in the environment were inferred through profiling of the 16S rRNA gene (13, 14). However, mercury methylation potential cannot be predicted based on 16S rRNA phylogenetic signal and is, rather, a species- or strain-specific trait (10, 15–18).

Recently, the *hgcAB* genes were discovered to be required for mercury methylation in several model organisms (19). The *hgcA* gene encodes a corrinoid-dependent enzyme that is predicted to act as a methyltransferase, and *hgcB* encodes a 2[4Fe-4S] ferredoxin that reduces the corrinoid cofactor (19). Importantly, both genes are thus far known to only occur in microorganisms capable of methylating mercury and are both required for methylation activity. This discovery has allowed for high-throughput identification of microorganisms with methylation potential in cultures or diverse environmental data sets based upon *hgcAB* sequence presence (17, 20). Clone and amplicon sequencing has been used as a low-cost method to retrieve *hgcAB* sequences with sufficient sequencing depth to assess the abundance and diversity of the *hgcAB* gene pair in the environment (18, 21–26). Broad-range and clade-specific quantitative PCR (qPCR) primers have been developed to screen for and quantify specific methylating groups in a given site (18, 21, 22). Identification of *hgcAB* sequences on metagenomic contigs has expanded the phylogenetic diversity of putative methylators beyond canonical SRBs, FRBs, and methanogens to include members of the *Aminicenantes*, *Chloroflexi*, *Elusimicrobia*, *Nitrospina*, *Nitrospirae*, *Spirochaetes*, and other groups (18, 20, 27, 28). However, *hgcAB* gene sequencing approaches have been shown to fail to capture the true phylogenetic diversity of putative methylating microorganisms (29), and the metabolic capabilities of identified microorganisms cannot be reliably inferred using these techniques.

Genome-resolved metagenomics is a powerful approach to connect the phylogenetic identity of microorganisms with metabolic potential through reconstructed population genomes from the environment (30). These genomes can then be used in conjunction with metatranscriptomics and metaproteomics to further explore the functional potential, activity, and putative interactions of the microbial community (31). Jones et al. reconstructed metagenome-assembled genomes from two sulfate-impacted lakes in which the dominant putative methylators consisted of members of the *Spirochaetes*, *Aminicenantes*, and PVC superphylum (29). Gionfriddo et al. applied assembly-based and genome-resolved metagenomics to understand microbial mercury resistance and mercury methylation processes in geothermal springs (32). We recently investigated methylation potential across the anoxic food web of Lake Mendota and recovered population genomes of putative methylators among the *Bacteroidetes* and PVC superphylum with fermentation capabilities (33).

In this study, we identified nearly 1,000 genomes containing the *hgcA* marker from publicly available isolate genomes, metagenome-assembled genomes (MAGs), and novel bins assembled from three freshwater lakes. Using this collection of MAGs and reference genomes, we identified putative methylators spanning 30 phyla from diverse ecosystems, including phyla that, to our knowledge, have never been characterized as

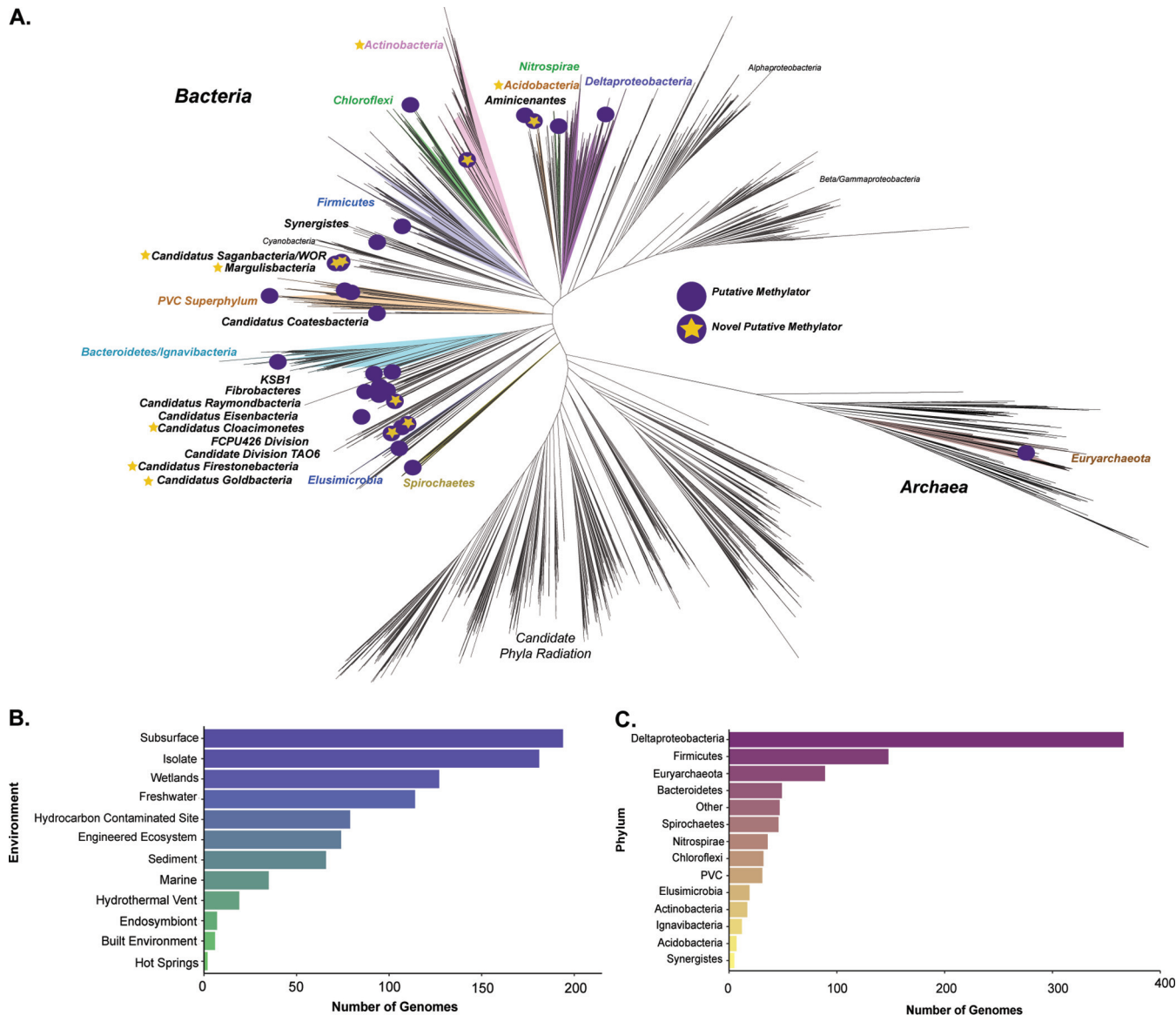
methylating groups. The *hgcAB* phylogenetic signal further confirms that this diverse locus originated through extensive horizontal gene transfer events, and provides insights into the difficulty of predicting the identity of putative methylators using ribosomal sequences or broad-range *hgcA* primers (16, 20, 29). Putative methylators span metabolic guilds beyond canonical SRBs, FRBs, and methanogenic archaea, carrying genes for traits such as nitrogen fixation and metal resistance pathways. To understand *hgcA* expression in the environment, we analyzed depth-discrete metatranscriptomes from a permafrost thawing gradient from which we identified 111 putative methylators. This work demonstrates the significance of reconstructing population genomes from the environment to accurately capture the phylogenetic distribution and metabolic capabilities of putative methylating groups in a given system.

## RESULTS AND DISCUSSION

**Identification of diverse, novel, putative mercury-methylating microorganisms.** We identified nearly 1,000 putative bacterial and archaeal mercury methylators spanning 30 phyla among publicly available isolate genomes and MAGs recovered from numerous environments (Fig. 1). Well-known methylators among the *Deltaproteobacteria*, *Firmicutes*, and *Euryarchaeota* represent well more than one-half of all identified putative methylators, with the majority among the *Deltaproteobacteria* (Fig. 1C). We also expanded upon groups of methylators that, until recently, have not been considered to be major contributors to mercury methylation. Putative methylators belonging to the *Spirochaetes* and the PVC superphylum (consisting of *Planctomycetes*, *Verrucomicrobia*, *Chlamydiae*, and *Lentisphaerae* phyla and several candidate divisions) were previously identified from metagenomic contigs (18) and reconstructed MAGs from Jones et al. (29). Additionally, *hgcA* sequences have been detected on metagenomic contigs belonging to the *Nitrospirae*, *Chloroflexi*, and *Elusimicrobia* (18, 27), but overall, it is unknown how these groups contribute to MeHg production. We also detected *hgcA* in phyla that, to our knowledge, have not been shown to carry the gene. We identified 7 *hgcA*<sup>+</sup> *Acidobacteria*, 5 of which come from the permafrost gradient system, where they are considered to be the main plant biomass degraders (31). We also identified 17 *hgcA*<sup>+</sup> *Actinobacteria*, which all belong to poorly characterized lineages within the *Coriobacteriia* and *Thermoleophilina* classes, as described below. A few putative methylators belong to recently described candidate phyla, such as “*Candidatus* Aminicenantes,” “*Candidatus* Firestonebacteria,” and WOR groups (see Fig. S1D in the supplemental material).

We recently reconstructed new MAGs from three freshwater lakes with diverse biogeochemical characteristics. Trout Bog Lake is a small, dimictic humic lake in a rural area near Minocqua, WI, that is surrounded by sphagnum moss, which leaches large amounts of organic carbon. Lake Tanganyika is a meromictic lake with an anoxic monimolimnion in the East African Rift Valley and is the second largest lake in the world both by volume and depth, from which we recently reconstructed nearly 4,000 MAGs (34). Lake Mendota is a large, dimictic eutrophic lake in an urban setting in Madison, WI, with a sulfidic anoxic hypolimnion. From these reconstructed freshwater MAGs, we identified 55 putative mercury methylators. Notably, several understudied PVC clades were recovered from the Lake Mendota hypolimnion. Members of this superphylum are ubiquitous in freshwater and marine environments, display a cosmopolitan distribution, and can make up approximately 7% of the total microbial community in these ecosystems (35–38). In Lake Mendota, members of the PVC superphylum account for approximately 40% of the microbial community in the hypolimnion, the majority of which do not contain *hgcA* (33).

A majority of the identified methylators were from a subsurface aquifer system and thawing permafrost gradient assembled by Anantharaman et al. (39) and Woodcroft et al. (31), respectively (Fig. 1B). Additionally, we identified methylators from hydrocarbon-contaminated sites such as oil tills and sand ponds, sediments and wetlands, engineered systems, hydrothermal vents, and the built environment. Only a few *hgcA*<sup>+</sup> organisms were recovered from marine systems, as the Tara Ocean project mostly



**FIG 1** Phylogenetic distribution and diversity of putative methylators. (A) Putative methylators among the prokaryotic tree of life. The highest quality putative methylator from each identified phylum was selected as a representative among select references from Anantharaman et al. (39) from across the tree of life. The tree was constructed from an alignment of 16 concatenated ribosomal proteins. Purple dots represent putative methylating phyla, and those with stars represent phyla containing putative novel methylating lineages that, to our knowledge, have not been identified as potential mercury methylators before this study. Groups in bold and/or colored are putative methylating groups, whereas a few uncolored groups in nonboldface font are noted only for orientation. (B) Numbers of putative methylating genomes found in various environments or genomes sequenced from isolates. (C) Numbers of putative methylating genomes belonging to each phylum. Phylum names are given for all groups except the *Deltaproteobacteria* class of *Proteobacteria*, as no putative *Proteobacteria* methylators were identified outside the *Deltaproteobacteria*. The group “other” contains several candidate phyla and genomes within groups with few representatives, described in Fig. S1 in the supplemental material.

includes samples from surface waters (40) and therefore would not include traditionally anaerobic methylating microorganisms. Although not as numerous as MAGs from various metagenomic surveys, we identified *hgcA* in more than 150 isolate genomes, with some among the *Bacteroidetes*, *Spirochaetes*, *Nitrospirae*, and *Chloroflexi* (Fig. S1A). Using an extensive set of genomes mostly recovered from anoxic environments, we were able to expand upon the known phylogenetic diversity of putative mercury-methylating microorganisms.

**Novel *Actinobacteria* lineages with putative methylating potential.** To our knowledge, members of the *Actinobacteria* have never been characterized as putative methylators or found to contain the *hgcA* marker. We identified 17 *hgcA*<sup>+</sup> *Actinobac-*



teria (16 of which also contained *hgcB*) among diverse environments, such as our Trout Bog and Lake Mendota freshwater data sets, as well as within the permafrost gradient and MAGs recovered from hydrocarbon contaminated sites (31, 41). In freshwater systems, *Actinobacteria* are ubiquitous and present a cosmopolitan distribution, with specific lineages comprising up to 50% of the total microbial community (42, 43). Freshwater *Actinobacteria* traditionally fall within the class *Actinobacteria* but have a lower abundance distribution in the hypolimnion due to decreasing oxygen concentrations (44, 45). However, none of the *hgcA*<sup>+</sup> *Actinobacteria* belong to ubiquitous freshwater lineages.

All of the 17 identified *hgcA*<sup>+</sup> *Actinobacteria* belong to poorly described classes or ill-defined lineages that may branch outside the established six classes of *Actinobacteria* (46). Nine fall within the *Coriobacteriia* class, four fall within the *Thermoleophilia* class, and the other four diverge from named *Actinobacteria* lineages. Three of the ill-defined *Actinobacteria* MAGs have been designated within the proposed class UBA1414, with two assembled from the permafrost system and one from the Lake Mendota hypolimnion. The other ill-defined *Actinobacteria* MAG was designated within the proposed class RBG-13-55-18 and was also recovered from the permafrost system. While some *Coriobacteriia* have been identified as clinically significant members of the human gut, all identified putative methylators belong to the OPB41 order (47), which refers to the 16S rRNA sequence identified by Hugenholtz et al. from the Obsidian Pool hot spring in Yellowstone National Park (48). Members of the OPB41 order have also been found to subsist along the seafloor of the Baltic Sea, an extreme and nutrient-poor environment (49).

Members of the *Thermoleophilia* are known as heat- and oil-loving microbes due to their growth restriction to only substrate *n*-alkanes (50). Within this deep-branching lineage, the two orders *Solirubrobacterales* and *Thermoleophilales* are currently recognized based on sequenced isolates and the most updated version of Bergey's taxonomy (46, 51). However, our *hgcA*<sup>+</sup> *Thermoleophilia* MAGs from the hypolimnia of Lake Mendota and Trout Bog as well as two other *hgcA*<sup>+</sup> *Thermoleophilia* from the permafrost thawing gradient do not cluster within the recognized orders (Fig. 2). According to the genome taxonomy database (GTDB), these MAGs cluster within a novel order preliminarily named UBA2241, as the only MAGs recovered from this novel order to date have been from the authors' corresponding permafrost system (31, 52).

Based on collected isolates, members of the *Thermoleophilia* are assumed to be obligately aerobic (51). However, the functional potential of *Thermoleophilia* members has been poorly described due to few sequenced isolates and high-quality MAGs (53). We reconstructed the main metabolic pathways of a high-quality *Thermoleophilia* MAG, referred to as MENDH-Thermo (see Table S2 available at <https://figshare.com/articles/dataset/MENDH-Thermo-annotations/11620104>). All components of the glycolytic pathway are present for using a variety of carbohydrates for growth. A partial tricarboxylic acid (TCA) cycle for providing reducing power is present; only missing the step for converting citrate to isocitrate, which can be transported from outside sources. Interestingly, the MENDH-Thermo genome encodes the tetrahydromethanopterin-dependent enzymes encoded by *mch* and *fwdAB*, pointing to either formate utilization or detoxification (54, 55). We could not detect the presence of a putative acetate kinase; therefore, autotrophic growth through the formation of acetate is unlikely. Interestingly, the MENDH-Thermo genome encodes a full pathway for the formation of 4-hydroxybutyrate, a storage polymer (56). As the MENDH-Thermo genome can also synthesize the amino acids glycine and serine, combined with the ability to use reduced carbon compounds and form storage polymers, the MENDH-Thermo genome could exhibit a methylotrophic lifestyle but is missing key steps for tetrahydromethanopterin cofactor synthesis and formate oxidation (57). The identification of putative methylators within the *Thermoleophilia* highlights novel metabolic features not only of putative methylators but also of a poorly described class within *Actinobacteria*.



**FIG 2** Novel putative methylators within the *Thermoleophilia* class of *Actinobacteria*. Phylogenetic tree of publicly available and assembled *Thermoleophilia* MAGs and *Actinobacteria* references. Representative or reference genomes belonging to the other 5 classes of *Actinobacteria* were used, with *Bacillus subtilis* used as an outgroup. Orders are clustered and colored by GTDB-proposed designation. The phylogeny was constructed with RAXML using 100 rapid bootstraps, and nodes with bootstrap support greater than 60 are shown as black circles. Stars denote putative methylators identified in this study.

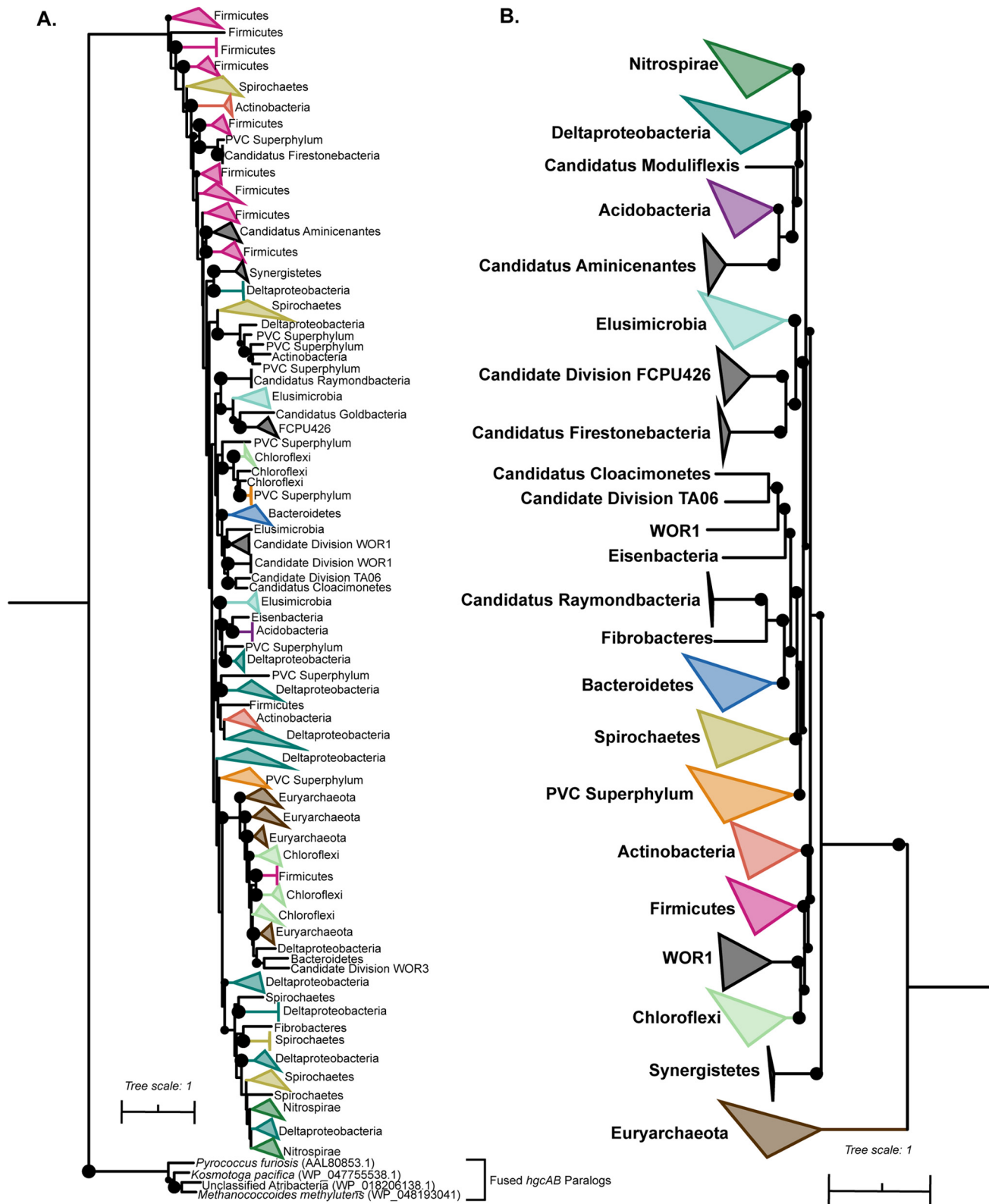
**Implications for horizontal gene transfer of *hgcAB*.** The HgcAB protein phylogeny does not follow a logical species tree phylogeny, as demonstrated by a comparison to a concatenated phylogeny of ribosomal proteins (Fig. 3). For example, *Deltaproteobacteria* HgcAB sequences are some of the most disparate sequences surveyed, clustering with HgcAB sequences of *Actinobacteria*, *Nitrospirae*, *Spirochaetes*, and members of the PVC superphylum. Although *Firmicutes* HgcAB sequences are not as disparate, they also cluster with HgcAB sequences of other phyla such as *Deltaproteobacteria* and *Spirochaetes*. The only HgcAB sequences that clustered monophyletically among our data set are the *Bacteroidetes/Ignavibacteria* groups and *Nitrospirae*. These observations have implications for the underlying evolutionary mechanisms of the methylation pathway and modern methods for identifying methylating populations in the environment.

Gene gain/loss mediated through several horizontal gene transfer (HGT) events was previously suggested to underpin the sparse and divergent phylogenetic distribution of *hgcAB* (17, 19–21, 29). Metabolic functions can be horizontally transferred across diverse phyla through numerous mechanisms, such as phage transduction, transposon-mediated insertion of genomic islands or plasmid exchanges, direct conjugation, and/or gene gain/loss events. We were unable to identify any *hgcA*-like sequences on any viral contigs in the NCBI database or within close proximity to any known insertion sequences or transposases, and all known *hgcA* sequences have been found on the chromosome. Structural and mechanistic similarities between the cobalamin-binding domains of HgcA and the carbon monoxide dehydrogenase/acetyl coenzyme A (acetyl-CoA) synthase (CODH/ACS) of the reductive acetyl-CoA pathway (also known as the Wood-Ljungdahl pathway) suggests associations between the two pathways (19, 20). Interestingly, *hgcA* sequences of *Euryarchaeota* and *Chloroflexi* cluster together, which was recently found to also be true for the CODH of these groups (58). It is plausible that a gene duplication event followed by rampant gene gain/loss and independent transfers of both the CODH and HgcA proteins, respectively, could explain the disparate phylogeny, as has been described for both pathways (19, 20, 58).

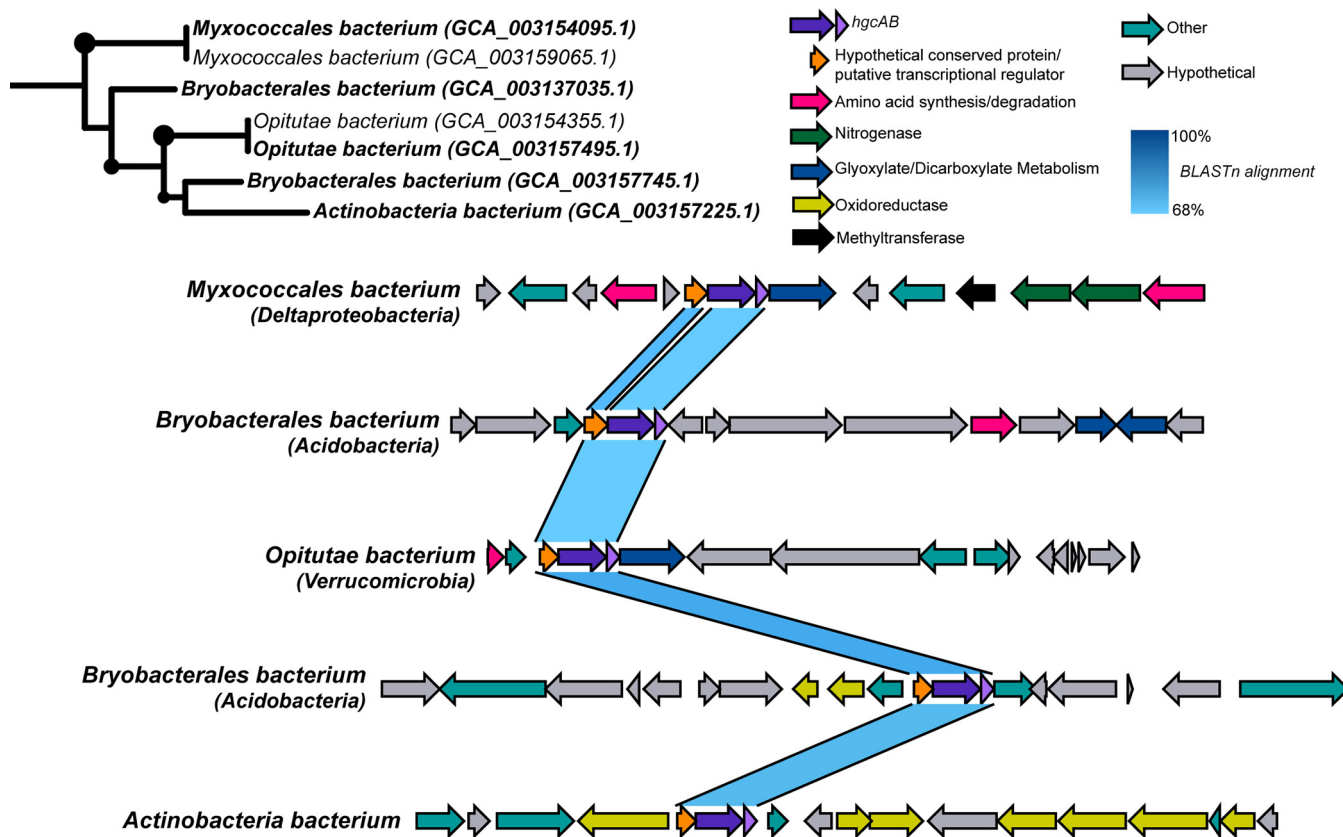
To illustrate an example of potential interphylum HGT of *hgcAB*, we selected methylators assembled from a permafrost system containing *hgcAB* regions with approximately 70% DNA sequence similarity (Fig. 4). Five putative methylators encompassing four different phyla (one *Deltaproteobacteria*, two *Acidobacteria*, one *Verrucomicrobia*, and an *Actinobacteria*) contain highly similar *hgcAB* regions, and their *hgcAB* phylogeny does not match the species tree. All *hgcAB* sequences of these methylators are preceded by a similar hypothetical protein that is annotated as a putative transcriptional regulator in the *Opitutae* genome. We screened for this putative transcriptional regulator among all methylators in our data set and identified it immediately upstream of *hgcAB* in 112 genomes (see Table S3 available at [https://figshare.com/articles/dataset/hgcAB-regulator-resultsUntitled\\_Item/11620107](https://figshare.com/articles/dataset/hgcAB-regulator-resultsUntitled_Item/11620107)). Overall, the gene neighborhoods of these similar permafrost *hgcAB* sequences do not share any other obvious sequence or functional similarities. The *hgcAB* genes are flanked by numerous hypothetical proteins, various amino acid synthesis and degradation genes, and oxidoreductases for core metabolic pathways. More detailed phylogenomic analyses will need to be applied in future studies to specifically pinpoint individual *hgcAB* HGT events and mechanisms.

Broad-range and clade-specific qPCR primers targeting *hgcAB* have been used to screen for putative methylators in the environment (18, 22). However, these primers were constructed based on a few experimentally verified methylators, which surely does not capture the complete or true diversity of methylating microorganisms. We tested the accuracy of these primers *in silico* on the coding regions of *hgcA* in all our identified putative methylators by using the taxonomical assignment of each putative methylator from the full genome classification. Depending on the number of *in silico* mismatches allowed, only experimentally verified methylators could be identified, or *hgcA* sequences belonging to other phyla would be hit (see Table S4 at <https://figshare.com/articles/dataset/hgcA-in-silico-primer-stats/10093535>). For example, *Deltaproteobacteria*-specific *hgcA* qPCR primers





**FIG 3** Discordance of HgcAB phylogeny relative to a concatenated corresponding ribosomal protein phylogeny of select methylators. (A) Phylogeny of HgcAB of select methylators was constructed as described in Materials and Methods. The tree was constructed using RAxML with 100 rapid bootstraps, and nodes with bootstrap support greater than 50 are depicted as black circles. (B) Corresponding concatenated ribosomal protein phylogeny of the select methylators. The tree was constructed using a set of 16 ribosomal proteins (2,470 positions following masking), concatenated, and built using RAxML with 100 rapid bootstraps. Nodes with bootstrap support of greater than 60 are denoted as black circles. A fully annotated HgcAB tree is provided in Fig. S4 at [https://figshare.com/articles/figure/hgcAB\\_annotated\\_tree\\_full/12592085](https://figshare.com/articles/figure/hgcAB_annotated_tree_full/12592085).

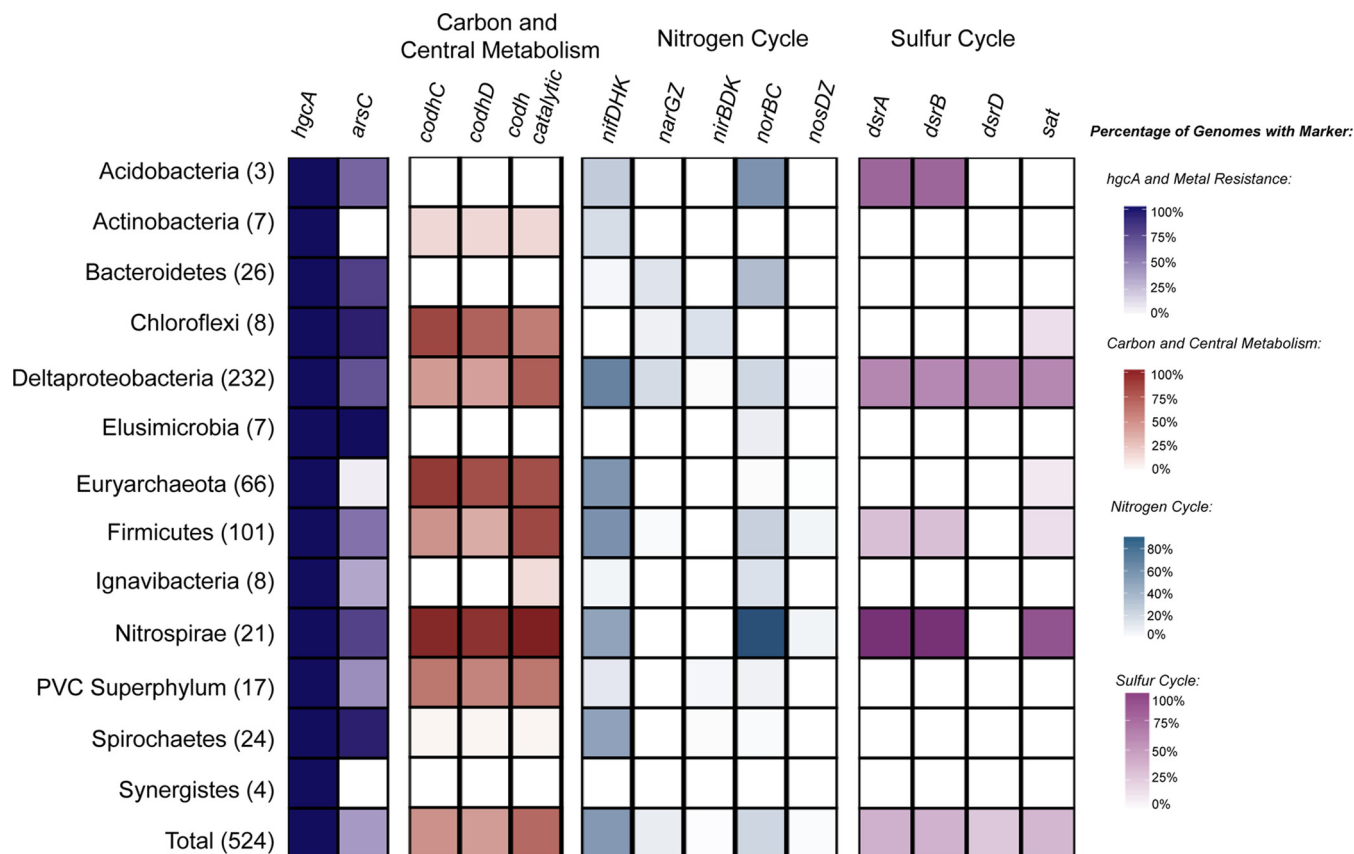


**FIG 4** Gene neighborhoods of diverse permafrost methylators containing similar *hgcAB* regions. Contigs containing similar *hgcAB* regions from putative methylators identified in a permafrost thawing system (31) were compared. Pairwise nucleotide BLAST between the genes for each contig was performed and visualized in Easyfig (106). Colors correspond to predicted functions of each gene. *hgcAB* genes were identified based on curated HMMs and reference sequences, while all other gene assignments were made based on Prokka predictions.

picked up related *Actinobacteria*, *Spirochaetes*, and *Nitrospirae* sequences. Whereas the *Firmicutes*-specific primers detected all experimentally verified *Firmicutes* methylators and a few environmental *hgcA* sequences, overall, they missed the diversity of *Firmicutes* *hgcA* sequences.

Discordance between the HgcAB and ribosomal protein phylogenies suggests that nonspecific amplicon approaches and broad-range qPCR assays will not yield easily interpretable results. Although new broad-range *hgcAB* primers have been developed based on an updated reference database of *hgcAB* sequences (21), the apparent HGT dynamics of this gene pair makes assigning the taxonomy of resulting hits challenging. Instead, finer-scale primers of specific methylating groups could be constructed based on *hgcAB* sequences from assembled population genomes from specific environments. However, this approach relies on sufficient depth of sequencing coverage to adequately assemble *hgcAB*-containing MAGs, which may consist of low-abundance members of the community. For example, we were able to assemble genomes of 35 putative methylators out of 228 total nonredundant genomes from Lake Mendota. Anantharaman et al. (39) and Woodcroft et al. (31) generated genome-resolved metagenomic data sets at terabase scales, from which we were able to identify 111 putative methylators from a total of 2,540 genomes and 111 putative methylators from a total of 1,529 total genomes, respectively. Alternatively, methods such as multiplexed digital droplet PCR (ddPCR) approaches could be applied to simultaneously amplify the 16S rRNA and *hgcAB* genes (59, 60). This approach would allow for higher sequencing depth at a lower cost than metagenomics while also ensuring that methylation status is linked to phylogenetic identity with fidelity.

**Metabolic capabilities of identified putative methylators.** We next characterized the broad metabolic capabilities among 524 highly complete putative methylators.



**FIG 5** Summary of broad metabolic characteristics present among high-quality putative methylators. Genomes with  $>90\%$  completeness and  $<10\%$  redundancy and belonging to phyla with more than 5 genomes in that group were screened. HMM profiles spanning the sulfur, nitrogen, and carbon cycles and metal resistance markers were searched for. The intensity of shading within a cell equates to the corresponding colored legend of percentage of genomes within that group that contain that marker. *hgcA*, gene encoding a mercury methylation corrinoid protein; *arsC*, gene encoding arsenite reductase; *codhC*/*codhD*/*codh catalytic*, genes encoding carbon monoxide dehydrogenase C and D and catalytic subunits; *nifDHK*, genes encoding nitrogenase subunits; *narGZ*, genes encoding nitrate reductase subunits; *nirBDK*, genes encoding nitrite reductase subunits; *norBD*, genes encoding nitric oxide reductase subunits; *nosDZ*, genes encoding nitrous oxide reductase subunits; *dsrABD*, genes encoding dissimilatory sulfite reductase; *sat*, sulfate adenylyltransferase. Individual *nif*, *nar*, *nir*, *nor*, and *nos* subunit presence/absence results were averaged together.

From our analysis of metabolic profiles spanning several biogeochemical cycles (39), the *hgcAB* gene pair is the only genetic marker linking high-quality methylating genomes other than universally conserved markers (Fig. 5). We detected putative methylators with methanogenesis and sulfate reduction pathways, as is expected for archaeal and bacterial methylating guilds, respectively (11, 16, 29). Although sulfate availability has been linked to methylating bacteria and active methylation, very few guilds we examined exhibited dissimilatory sulfate reduction pathways. We detected machinery for dissimilatory sulfate reduction within the *Acidobacteria*, *Deltaproteobacteria*, *Firmicutes*, and *Nitrospirae*. However, for example, only one-half of the *Deltaproteobacteria* methylators screened contained the *dsrABD* subunits for sulfate reduction. Similarly, we detected few guilds with the sulfate adenylyltransferase (*sat*) marker for assimilatory sulfate reduction for incorporating sulfur into amino acids. This suggests that mercury-methylating groups may be composed of guilds other than canonical SRBs and methanogenic archaea.

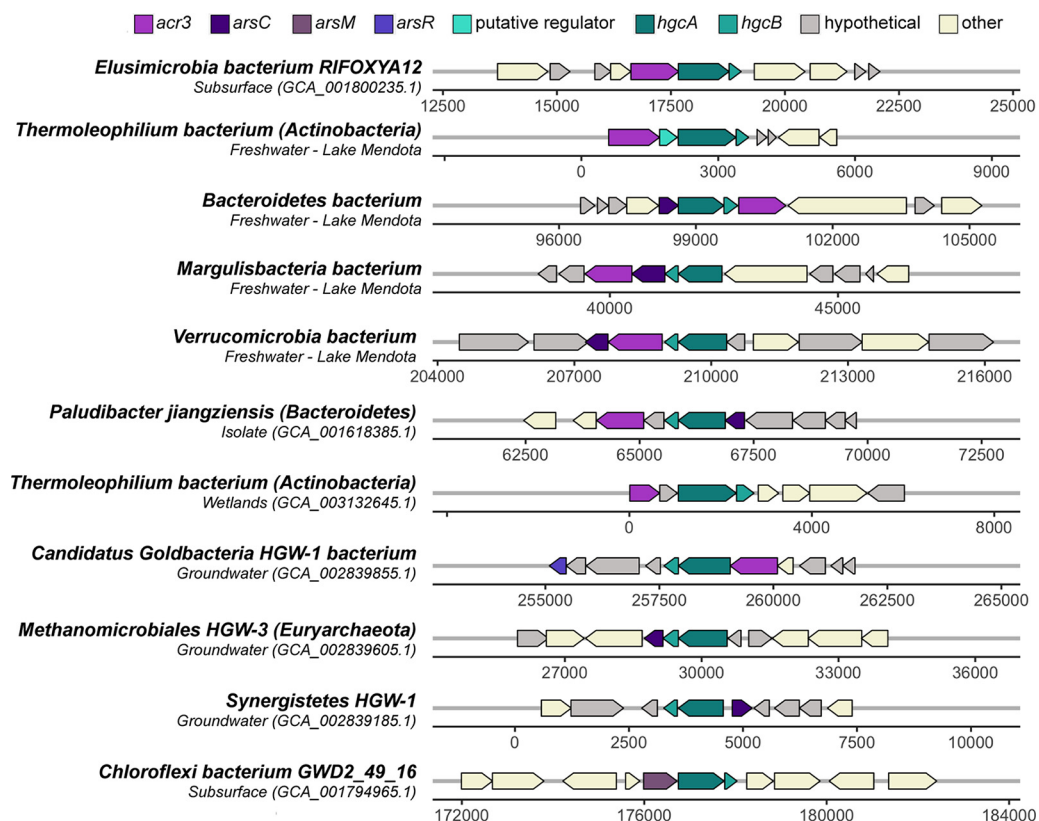
We found that several methylating guilds contain the essential subunits that form the nitrogenase complex for nitrogen fixation. The three main components of the nitrogenase complex are *nifH* and *nifD-nifK*, which encode the essential iron and molybdenum-iron protein subunits, respectively. We detected machinery for the nitrogenase complex within the *Acidobacteria*, *Actinobacteria*, *Deltaproteobacteria*, *Euryarchaeota*, *Firmicutes*, *Nitrospirae*, and *Spirochaetes* methylators. Notably, more than one-half of all *Deltaproteobacteria* methylators and approximately one-half of all *Firmi-*

*cutes* methylators contain all three subunits for the nitrogenase complex. Recently, the presence of nitrate has been suggested to regulate MeHg concentrations (61), leading to nitrate addition in an attempt to control MeHg accumulation in freshwater lakes (62). Although we detected parts of the denitrification pathway in a few lineages, none of the putative methylators contained a full denitrification pathway for fully reducing nitrate to nitrogen gas. Previous studies have identified putative methylators among *Nitrospina* nitrite oxidizers in marine systems (27, 28). Although we did not detect putative *Nitrospina* methylators (possibly due to these *hgcA* sequences being unbinned, within low-quality MAGs, or low-confidence hits), links between the nitrogen cycle and mercury methylation warrant further exploration.

As mentioned above, structural and functional similarities between HgcA and the carbon monoxide dehydrogenase/acetyl-CoA synthase (CODH/ACS) suggest associations between the two pathways. This also raises the question if methylators use the reductive acetyl-CoA pathway for autotrophic growth. We screened for the presence of the *codhC* and *codhD* subunits of the carbon monoxide dehydrogenase, which have been suggested as potential paralogous origins of HgcA (19, 20). Among high-quality putative methylators, we detected both subunits among *Actinobacteria*, *Chloroflexi*, *Deltaproteobacteria*, *Euryarchaeota*, *Firmicutes*, *Nitrospirae*, and members of the PVC superphylum. We also screened for the presence of the CODH catalytic subunit, which would infer acetyl-CoA synthesis. Although we detected the CODH catalytic subunit, it was not detected in all methylating lineages, such as the *Elusimicrobia*, *Synergistes*, and *Spirochaetes*. This suggests other growth strategies than autotrophic carbon fixation among a large proportion of methylators, such as fermentation pathways as suggested previously by Jones et al. (29).

Interestingly, >50% of the putative methylators contain a thioredoxin-dependent arsenate reductase, which detoxifies arsenic through the reduction of arsenate to arsenite (63, 64). As(V) reduction to the even more toxic As(III) is tightly coupled to export from the cell, which has interesting similarities to the mercury methylation system. Hg(II) uptake has been shown to be energy dependent and therefore is potentially imported through active transport mechanisms (65). Hg(II) uptake is highly coupled to MeHg export, suggesting the methylation system may act as a detoxification mechanism against environmental Hg(II) (65). Intriguingly, numerous diverse putative methylators contain *hgcAB* regions that are immediately flanked by different arsenic-related genes or are within a few open reading frames (Fig. 6). We found several examples where the *hgcAB* region is flanked by *acr3*, encoding an arsenite efflux permease, and *arsC*, encoding a thioredoxin- or glutaredoxin-dependent arsenite reductase. Both the arsenite efflux and reductase are part of the *ars* operon for arsenic resistance and detoxification and have somewhat redundant functions in that they both achieve extrusion of toxic arsenic but through different mechanisms (66).

Recently, Goñi-Urriza et al. identified an ArsR-like regulator upstream of *hgcAB* among *Desulfovibrio* and *Pseudodesulfovibrio* methylating strains that is cotranscribed with *hgcA* (67). This family of metalloregulatory transcriptional repressors is responsive to a variety of metals such as As, Zn, and Ni, and canonical ArsR repressors mediate arsenic detoxification. (68, 69). However, some ArsR-like sequences have been identified that share homology to canonical ArsR repressors but do not contain hallmark metal binding sequence signatures (68). Recently, Gionfriddo et al. also identified a putative ArsR-like repressor downstream from *hgcAB*, in addition to ArsB, an arsenical pump membrane protein (32). We identified ArsR-like sequences in close proximity to *hgcAB* as well as *arsM*, an arsenite S-adenosylmethionine methyltransferase. Growing evidence for connections between metal resistance pathways and mercury methylation provide intriguing hypotheses to test concerning the constraints and controls of methylation in the environment. Overall, putative methylators are composed of metabolic guilds other than the classical SRBs and methanogenic archaea and employ diverse growth strategies other than autotrophic carbon fixation.

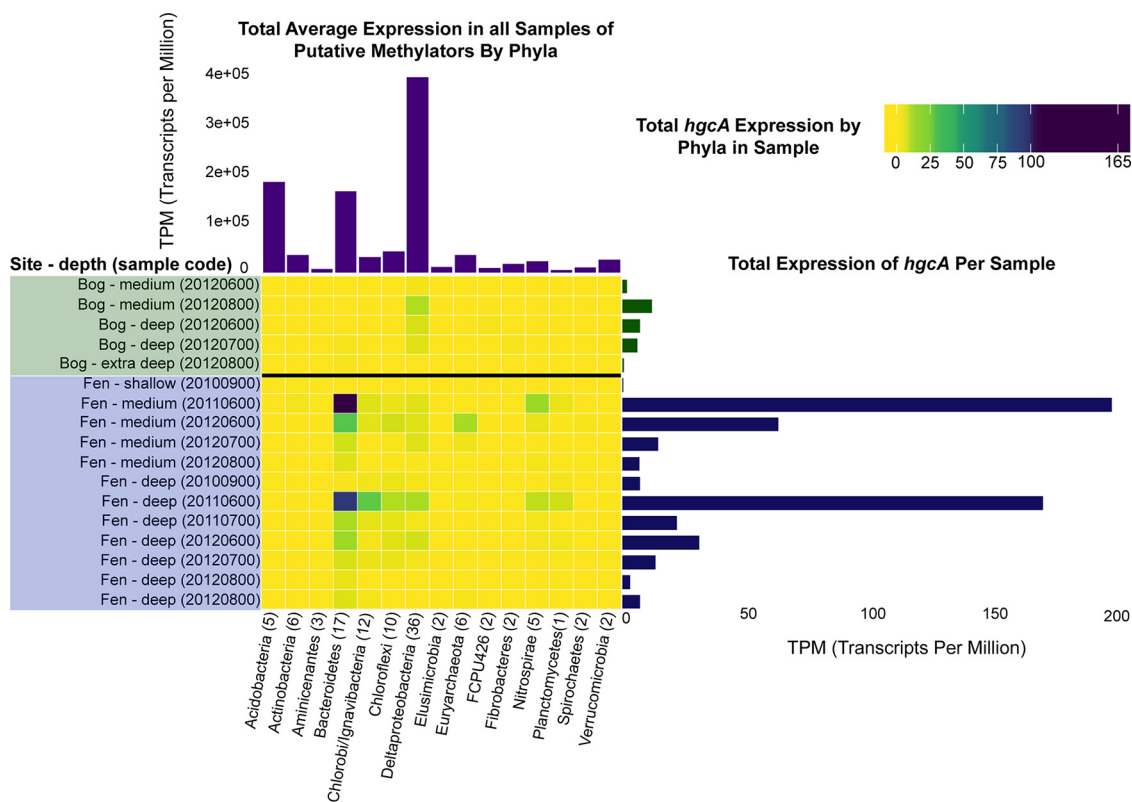


**FIG 6** Gene neighborhoods of select putative methylators demonstrating the association of arsenic metabolism with *hgcAB* genes. Arsenic metabolism-related genes are colored shades of purple, and *hgcAB* genes and the putative regulator are colored shades of blue. *acr3*, gene encoding arsenite efflux; *arsC*, gene encoding glutaredoxin/thioredoxin-dependent arsenite reductase; *arsM*, gene encoding arsenic methyltransferase; *arsR*, gene encoding Ars-R type transcriptional regulator.

**Transcriptional activity of putative methylators in a permafrost system.** While the connection between the presence of the *hgcAB* genes and methylation status has been well established in cultured isolates, there have been few studies investigating *hgcAB* transcriptional activity under laboratory conditions (70) or in a specific environment (71). Additionally, the mere presence of a gene does not predict its functional dynamics or activity in the environment. Assessing microbial populations actively contributing to MeHg production is important for identifying constraints on this process, and genome-resolved metatranscriptomics may be a useful approach to do so in the environment. Woodcroft et al. performed extensive spatiotemporal sampling across three sites of the Stordalen Mire peatland in northern Sweden paired with genome-resolved metagenomics, metatranscriptomics, and metaproteomics to understand microbial contributions to carbon cycling (31). This site includes well-drained but mostly intact palsas, intermediately thawed bogs with *Sphagnum* moss, and completely thawed fens. We identified 111 *hgcA*<sup>+</sup> genomes among MAGs assembled from this permafrost system spanning well-known methylators belonging to the *Deltaproteobacteria*, *Firmicutes*, and *Euryarchaeota* but also less-studied putative methylators within the *Acidobacteria*, *Actinobacteria*, *Elusimicrobia*, and *Verrucomicrobia*. We chose this extensive data set to explore depth-discrete activity of *hgcA*, since it is largely an anaerobic environment and methylation activity has been shown to occur in this type of system (20, 72, 73). Furthermore, there are growing concerns about MeHg production in polar regions due to thawing permafrost from global warming effects (74).

We specifically investigated the site and depth-discrete expression patterns of *hgcA* within phylum-level groups (Fig. 7). We were unable to detect *hgcA* transcripts at any time point or depth among the palsa sites or the shallow bog depths; therefore, these





**FIG 7** Transcriptional activity of putative methylators in a permafrost thawing gradient. Counts of mapped metatranscriptomic reads to 111 putative methylators identified in a permafrost thawing gradient in bog and fen samples (31). (Center) Total reads mapping to the *hgcA* gene within a phylum per sample normalized by transcripts per million (TPM). Heat map color scale represents a continuous gradient between 0 and 100 TPM, and a break between 100 and 165 TPM to better visualize results in the low range. (Right) Total *hgcA* counts per sample in TPM. (Top) Average overall expression of genomes within a phylum across all bog and fen samples in TPM.

sites and depths are not included. Lack of *hgcA* transcript detection in the frozen palsa sites may be connected to the relationship between increased methylation activity with increased amounts of oxidized organic matter (75, 76), which would more likely be present in the thawing bog and fen sites. The highest total *hgcA* transcript abundance was detected in the fen sites, which are the most thawed along the permafrost gradient. The highest *hgcA* expression was contributed by the *Bacteroidetes* in the medium and deep depths of the fen sites across multiple time points, followed by those belonging to *Deltaproteobacteria* in the medium and deep depths of the bog sites. Interestingly, *hgcA* expression in the bog samples was almost exclusively dominated by *Deltaproteobacteria*, whereas the fen samples contained a more diverse collection of phyla expressing *hgcA*, albeit *Bacteroidetes* was the most dominant. In the deep fen samples, groups such as the *Ignavibacteria*, *Chloroflexi*, and *Nitrospirae* also contributed to *hgcA* expression. Overall, methylators belonging to *Bacteroidetes* and *Deltaproteobacteria* also constituted the most highly transcriptionally active members across all samples.

All of the putative *Bacteroidetes* methylators transcribing *hgcA* belong to the vadinHA17 class. The name vadinHA17 was given to the 16S rRNA gene clone recovered from an anaerobic digester treating winery wastewater, referring to the vinasse anaerobic digester of Narbonne (VADIN) (77). Members of the vadinHA17 group remain uncultured and are thought to be involved in hydrolyzing and degrading complex organic matter, cooccurring with methanogenic archaea (78, 79). Currently, the three publicly available *Bacteroidetes* strains containing *hgcA* belong to the *Paludibacteraceae* and *Marinilabiliaceae* and are not closely related to members of the vadinHA17 class. However, the dominant *Deltaproteobacteria* member expressing *hgcA* in the bog

samples is classified within the *Syntrophobacterales* order, specifically, a member of *Smithella* spp. Among the publicly available sequenced *hgcA*<sup>+</sup> *Deltaproteobacteria* isolates, 3 are among the *Syntrophobacterales* order, including the *Smithella* sp. F21 strain. This methylating strain may be an appealing experimental system to further explore *hgcA* expression patterns under different conditions.

For genomes in which we identified confident *hgcB* hits, *hgcB* expression patterns mostly mirrored that of *hgcA* (see Fig. S2A). Additionally, we identified a conserved putative regulator immediately upstream of *hgcAB* as described above that was also present in some of the permafrost MAGs. Among genomes that contained the putative regulator, we detected transcripts associated with *Chloroflexi* and *Deltaproteobacteria*, corresponding to sites and depths at which these groups also demonstrate similar *hgcAB* expression patterns (Fig. S2B). We also compared *hgcA* expression to that of the housekeeping gene *rpoB*, which encodes the beta subunit of RNA polymerase (see Fig. S3). There are some *hgcA*<sup>+</sup> groups in which we detected appreciable *rpoB* transcripts but could not detect *hgcA* transcripts in any samples, such as the *Acidobacteria*, *Aminicenantes*, and *Fibrobacteres*. This suggests that some *hgcA*<sup>+</sup> organisms are transcriptionally active but may not contribute to methylmercury production in this ecosystem. Additionally, for groups such as the *Bacteroidetes*, *Deltaproteobacteria*, and *Nitrospirae* in which we detected *hgcA* transcripts, we detected *rpoB* transcripts for these groups in samples in which *hgcA* was not expressed.

Our findings have implications for understanding MeHg production in the environment, as currently, the presence of *hgcAB* is used to connect specific microorganisms to biogeochemical characteristics. However, it may be possible that only a few groups exhibit *hgcAB* activity and actually contribute to MeHg production. Although this study does not contain environmental data on inorganic mercury or MeHg concentrations, these results provide insights into *hgcA* expression for further follow-up. Previous studies of *hgcA* expression in the lab and the environment have inferred that *hgcA* is constitutively expressed (70, 71). Additionally, *hgcAB* expression by *Desulfovibrio dechloracetivorans* strain BerOc1 was found to have no quantifiable relationship to methylmercury production and depended upon growth conditions rather than induction of the gene pair (70). Therefore, further work is required to understand the regulation and environmental conditions leading to *hgcAB* expression and, ultimately, methylation activity. Although previous studies have inferred that *hgcAB* expression may not be connected to actual methylmercury production, genome-resolved metatranscriptomics can begin to assess these questions in the natural environment, since laboratory experiments of pure cultures may not reflect environmental conditions or dynamics. Future experiments could integrate genome-resolved metagenomics and metatranscriptomics with Hg/MeHg assays in sites with MeHg accumulation to understand if the expression of specific *hgcA*<sup>+</sup> populations contributes to environmentally relevant MeHg levels. Experiments performed with concurrent collections of biogeochemical characteristics could begin to parse which environmental conditions contribute to community activity and, specifically, *hgcAB* activity. Careful replication of conditions and necessary controls could also illuminate other genes and pathways associated with the activity of the *hgcAB* genes (80). Although it is entirely possible that *hgcAB* expression may indeed not be connected to actual methylmercury production, future research addressing these questions will overall expand our currently limited understanding of *hgcAB* expression.

**Conclusions.** In this study, we expanded upon the known diversity of microorganisms that are likely to perform mercury methylation. We demonstrated that putative methylators encompass 30 phylum-level lineages and diverse metabolic guilds using a set of publicly available isolate genomes, MAGs, and novel freshwater MAGs. Apparent extensive HGT of the diverse *hgcAB* region poses unique challenges for detecting methylating populations in the environment, in which “universal” amplicon or even group-specific approaches may not accurately reflect the true phylogenetic origin of specific methylators. Furthermore, genome-resolved

metatranscriptomics of putative methylators in a thawing permafrost system revealed that specific methylating populations are transcriptionally active at different sites and depths. These results highlight the importance of moving beyond identifying the mere presence of a gene to infer a specific function and investigating transcriptionally active populations. Overall, genome-resolved omics techniques are an appealing approach for accurately assessing the controls and constraints of microbial MeHg production in the environment.

## MATERIALS AND METHODS

**Accessed data sets, sampled sites, and metagenomic assembly.** We used a combination of publicly available MAGs, sequenced isolates, and newly assembled MAGs from three freshwater lakes. We sampled Lake Tanganyika in the East African Rift Valley, Lake Mendota in Madison, WI, and Trout Bog Lake near Minocqua, WI. We collected depth-discrete samples along the northern basin of Lake Tanganyika at two stations (Kigoma and Mahale). DNA extraction, metagenomic sequencing of 24 samples from Lake Tanganyika, and subsequent genome assembly and refinement were performed as described by Tran et al. (34). Briefly, 24 samples were passed through a 0.2- $\mu\text{m}$ -pore-size fraction filter and used for shotgun metagenomic sequencing on the Illumina HiSeq 2500 platform at the Department of Energy Joint Genome Institute. Each of the 24 metagenomes was individually assembled with metaSPAdes and binned into population genomes using a combination of MetaBat, MetaBat2, and MaxBin (81–83). Across these binning approaches, approximately 4,000 individual bins were assembled. To dereplicate identical bins assembled through different platforms and from different samples, we applied DasTool and dRep to keep the highest-quality bin for a given cluster, resulting in 803 total MAGs (84, 85). Using completeness and contamination cutoffs of  $>70\%$  and  $<10\%$ , respectively, this resulted in 431 MAGs. From this data set, we used 16 MAGs that we identified as confident *hgcA* hits from criteria described below.

Five depth-discrete samples were collected from the Lake Mendota hypolimnion using a peristaltic pump onto a 0.2- $\mu\text{m}$  Whatman filter and were used for Illumina shotgun sequencing on the HiSeq 4000 at the QB3 sequencing center in Berkeley, CA, as described by Peterson et al. (33). Samples were individually assembled using metaSPAdes, and population genomes were constructed using a combination of the MetaBat2, Maxbin, and CONCOCT binning algorithms (81–83, 86). Resulting MAGs were aggregated with DasTool and dereplicated into representative sets by pairwise average nucleotide identity comparisons (84). This resulted in a total of 228 bins that were each manually refined and inspected for uniform differential coverage using *anvi'o* (87), 35 of which were used in this study through identifying confident *hgcA* hits from the criteria described below.

Population genomes from samples collected from Trout Bog Lake were assembled and manually curated as previously described (35, 88, 89), from which we used four genomes that contained confident *hgcA* hits from the criteria described below. All publicly available isolate genomes and MAGs were accessed from GenBank in August of 2019, resulting in  $>200,000$  genomes in which to search for *hgcA*. Genomes listed by Jones et al. (29) were accessed from JGI/IMG in March of 2019 at GOLD study identification (ID) Gs0130353. Only genomes of medium quality according to MiMAG standards with  $>50\%$  completeness and  $<10\%$  redundancy were kept for downstream analyses, as calculated with CheckM (90, 91).

**Identification of putative methylators.** We built a hidden Markov model (HMM) profile of the HgcA protein using a collection of full-length HgcA protein sequences from experimentally verified methylating organisms (19). The constructed HMM profile was then used to identify putative mercury-methylating bacteria and archaea. For all genome sequences, open reading frames and protein-coding genes were predicted using Prodigal (92). We used *hmmsearch* in the *hmmer* program to search all predicted protein sequences with the HMM profile with an E value cutoff of  $1\text{e}-50$  (93). Sequence hits with an E value cutoff score  $<1\text{e}-50$  and/or a score of 300 were removed from the results to ensure high confidence in all hits. All archaeal and bacterial HgcA hits were concatenated and aligned with MAFFT (94). Alignments were manually visualized using AliView (95), and sequences without the conserved cap-helix domain [G(I/V)NVWCAAGK] reported by Parks et al. (19) were removed. Although these cutoff criteria may be stringent, we found that an E value cutoff of  $1\text{e}-50$  and a threshold cutoff score of 300 returned hits that nearly all contained the conserved cap-helix domain and manual inspection and curation of the resulting alignment were further reduced.

From this set of putative methylators, the taxonomy of each genome was assigned and/or confirmed using both automatic and manual classification approaches. Each genome was automatically classified using the genome taxonomy database toolkit (GTDB-tk) with default parameters (52). Additionally, each genome was manually classified using a set of 16 ribosomal protein markers (96). If a genome's taxonomy could not be resolved between the GTDB-tk classification and the manual ribosomal protein classification method, the genome was removed from the data set (approximately 30 genomes). We kept specific genomes designated "unclassified," as a few genomes with this designation belong to newly assigned phyla that have not been characterized in the literature, such as the proposed phyla BMS3A and Moduliflexota in the GTDB. This resulted in a total of 904 putative methylators used for downstream analyses as described. All genome information for each putative methylator, including taxonomic assignment by each method, quality, and genome statistics, is provided in Table S1 available at <https://figshare.com/articles/dataset/mehg-metadata/10062413> and summarized in Fig. S1 in the supplemental material.

**hgcAB and ribosomal phylogenies.** We selected the highest-quality methylator genome from each phylum (total of 30 individual phyla containing a putative methylator) to place in a prokaryotic tree of life. Bacterial and archaeal references (the majority from Anantharaman et al. [39]) were screened for the 16 ribosomal protein HMMs along with the select 30 methylators. Individual ribosomal protein hits were aligned with MAFFT (94) and concatenated. The tree was constructed using FastTree and further visualized and edited using the interactive tree of life (iTOL) online tool (97, 98) to highlight methylating phyla and novel methylating groups.

We created an HMM profile of the HgcB protein using full-length HgcB sequences from experimentally verified methylators as described above for HgcA (17, 93). We searched for HgcB in the set of 904 putative methylators identified from HgcA presence as described above. We were able to identify HgcB-like sequences in all 904 putative methylators but further screened these hits as follows. All putative HgcB hits had to have a threshold cutoff score greater than 70 and contain the conserved ferredoxin-binding motif (CXXCXXC) as reported by Parks et al. (19). Additionally, we required the resulting hit to be either directly downstream of *hgcA* or within five open reading frames, as some experimentally verified methylators, such as *Desulfovibrio africanus* Walvis Bay, have been observed to have genes between *hgcAB* (19). The resulting HgcB hit also had to be on the same contig as HgcA to ensure a confident hit and potentially include it in the HgcAB phylogeny. We recognize that this criterion might be overly conservative and stringent, potentially excluding some genomes with both genes (e.g., *hgcB* on a separate contig) or fused *hgcAB* hits, but we preferred to avoid any false positives. This resulted in 844 of 904 identified methylators containing a confident HgcB hit to be used for the HgcAB phylogeny. A given genome had to contain 12 or more ribosomal protein markers to be included in the HgcAB tree used to compare against a concatenated ribosomal protein phylogeny. Additionally, we removed sequences with low bootstrap support using RogueNaRok (99). Overall, this resulted in 650 HgcAB sequences used to create a representative HgcAB tree to compare to the corresponding concatenated ribosomal protein tree.

The HgcAB protein sequences were aligned separately using MUSCLE and concatenated (100). The alignment was uploaded to the Galaxy server for filtering with BMGE1.1 using the BLOSUM30 matrix, a threshold and gap rate cutoff score of 0.5, and a minimum block size of 5 (101, 102). Four fused HgcAB sequences were obtained from GenBank for *Kosmotoga pacifica* (WP\_047755538), *Methanococcoides methylutens* (WP\_048193041), *Pyrococcus furiosus* DSM 3638 (AAL80853), and an unclassified *Atribacteria* (WP\_018206138) as used by Gionfriddo et al. (27) as paralogs to root the tree and aligned using MUSCLE (100). A consensus alignment of the concatenated HgcAB sequences and the fused HgcAB sequences was created with MUSCLE (100). A phylogenetic tree of the hits was constructed using RAxML with 100 rapid bootstraps (103). The tree was rooted with the four fused HgcAB sequences and edited using iTOL (98). For visualization ease, we collapsed clades by the dominant monophyletic phylum when possible. For example, if a monophyletic clade contained a majority (approximately 80% to 90%) of *Deltaproteobacteria* sequences and one *Nitrospirae* sequence, we collapsed and identified the clade as a whole as *Deltaproteobacteria*. This was performed to highlight broad trends in the disparate phylogenetic structure of identified HgcAB sequences. To reduce complexity, identified HgcAB sequences from candidate phyla containing few individuals were not assigned individual colors.

The corresponding ribosomal tree derived from genomes carrying the representative HgcAB sequences was constructed using a collection of 16 ribosomal proteins provided in the metabolisHMM package (96, 104), requiring that each putative methylator contain at least 12 ribosomal proteins. Each marker was searched for in each genome using specific HMM profiles for each ribosomal protein, individually aligned using MAFFT (94), and then concatenated together. A phylogenetic tree was constructed using RAxML with 100 bootstraps and visualized using iTOL (98, 103). A phylogenetic tree of HgcAB with noncollapsed branches, most resolved taxonomical names, and bootstrap scores is provided in Fig. S4 available at [https://figshare.com/articles/figure/hgcAB\\_annotated\\_tree\\_full/12592085](https://figshare.com/articles/figure/hgcAB_annotated_tree_full/12592085).

**Functional annotations, metabolic reconstructions, and *in silico* PCR analysis.** All genomes were functionally annotated using Prokka (105). The predicted gene sequence for each *hgcA* open reading frame was obtained using the locus tag of the predicted *hgcA* protein with the above-described HMM-based approach. *In silico* PCR with the broad-range *hgcA* primers and group-specific qPCR *hgcA* primers were tested using Geneious (Biomatters Ltd., Auckland, NZ), with primer sequences from Christensen et al. (22). The settings allow for both 0 and 2 mismatches, and the primer amplification “hit” results were compared to the full-genome classification of the given *hgcA* gene sequence. For genome synteny analysis of select permafrost methylators, Easyfig was used to generate pairwise BLAST results and view alignments (106). We did not use one of the *Myxococcales* MAGs as it was exactly identical to the other. The *Opitutae* genome GCA\_003154355 from the UBA assembled set was not used, because *hgcA* was found on only a very short contig.

Broad metabolic capabilities were characterized across all methylating organisms based on curated sets of metabolic HMM profile sets provided in the metabolisHMM package (39, 104). An HMM of the putative transcriptional regulator was built using the alignment of the five putative transcriptional regulators with similar *hgcAB* sequences with MUSCLE and hmmbuild (93, 100) and searched for among all methylators with hmsearch (93). A genome was considered to have a hit for the putative transcriptional regulator if the identified protein contained a search hit E value of greater than  $1e-10$  and was immediately upstream of *hgcAB* or 1 to 2 genes upstream. Raw presence/absence results are provided in Table S5 at <https://figshare.com/articles/dataset/raw-metabolic-marker-results-high-quality-genomes/10203530>.

Metabolic reconstruction of the MENDH-*Thermoleophilia* bin was supplemented with annotations using the KofamKOALA distribution of the KEGG database and parsed for significant hits (107) (annotations available in Table S2 at <https://figshare.com/articles/dataset/MENDH-Thermo-annotations/11620104>). The phylogeny of the MENDH-*Thermoleophilia* MAG was confirmed by downloading representative/reference sequences from GenBank and Refseq for the entire *Actinobacteria* phylum, depending on which database/designation was available for the particular class. All *Thermoleophilia* class genomes were downloaded as designated by the Genome Taxonomy Database (52). We included *Bacillus subtilis* strain 168 as an outgroup. The phylogeny was constructed using the alignment of concatenated ribosomal proteins as described above. The tree was constructed using RAxML with 100 rapid bootstraps and visualized with iTOL (98, 103).

**Transcriptional activity of putative methylators in a permafrost system.** To investigate the transcriptional activity of methylators in the environment, we tracked transcriptional activity of the identified methylating organisms in a permafrost thaw gradient. (31). All 26 raw metatranscriptomes from the palsa, bog, and fen sites sequenced on the NextSeq Illumina platform were downloaded from NCBI from BioProject accession number [PRJNA386568](https://bioinformatics.ncbi.nlm.nih.gov/bioproject/PRJNA386568). Paired-end sequences were quality filtered and adapters removed using fastp (108). Each transcriptome was mapped to the indexed open reading frames of the 111 putative methylators identified in this ecosystem using kallisto (109). Annotated open reading frames and size in base pairs for each genome were supplied using Prokka (105). For every sample, counts were normalized by calculating transcripts per million (TPM) (110).

Open reading frames for *hgcA*, *hgcB*, and a putative conserved regulator were identified for each genome from the HMM profile annotations as described above. We did not detect expression of any of the putative methylators in any of the palsa samples and therefore did not include these samples in the results. Expression for *hgcA*, *hgcB*, and the regulator within a phylum was calculated by total TPM expression of that gene within that phylum. For *hgcA*, total *hgcA* expression within a sample was calculated as the total TPM-normalized expression of *hgcA* counts within that sample. The total average expression of methylators within a phylum was calculated by adding all TPM-normalized counts for all genomes within a phylum and averaged by the number of genomes within that phylum to represent the average activity of that group compared to *hgcA* activity. Expression of *hgcA* was compared to that of the housekeeping gene *rpoB*, which encodes the RNA polymerase beta subunit and predicted from Prokka annotations.

**Data availability.** Lake Tanganyika genomes are available at <https://osf.io/pmhae/>. Lake Mendota genomes are available at <https://osf.io/9vwgt/>. The four Trout Bog genomes can be accessed from JGI/IMG under accession IDs [2582580680](https://img.jgi.doe.gov/data/record/2582580680), [2582580684](https://img.jgi.doe.gov/data/record/2582580684), [2582580694](https://img.jgi.doe.gov/data/record/2582580694), and [2593339183](https://img.jgi.doe.gov/data/record/2593339183). A complete workflow including all code and analysis workflows can be found at <https://github.com/elizabethmcd/MEHG>. All supplementary data files and tables are available on Figshare at [https://figshare.com/projects/Expanded\\_Diversity\\_and\\_Metabolic\\_Flexibility\\_of\\_Microbial\\_Mercury\\_Methylation/70361](https://figshare.com/projects/Expanded_Diversity_and_Metabolic_Flexibility_of_Microbial_Mercury_Methylation/70361) under the CC-BY 4.0 license.

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**FIG S1**, TIF file, 0.5 MB.

**FIG S2**, PDF file, 0.7 MB.

**FIG S3**, PDF file, 1.1 MB.

## ACKNOWLEDGMENTS

We thank the many authors and groups for making a vast and valuable amount of genomic, metagenomic, and metatranscriptomic data publicly available, as our work was greatly improved by these data sets. We thank the anonymous reviewers of the manuscript, as their feedback and critiques strengthened our work. We thank members of the McMahan lab for constructive feedback on drafts of the manuscript, specifically, Amber White. We also thank Kristopher Kieft and Adam Breister for providing NCBI viral searches and tree of life ribosomal sequences, respectively. We thank the U.S. Department of Energy Joint Genome Institute for sequencing and assembly (CSPs 394 and 2796). This research was performed in part using the computer resources and assistance of the Center for High-Throughput Computing (CHTC) at UW-Madison in the Department of Computer Sciences. CHTC is supported by UW-Madison, the Advanced Computing Initiative, the Wisconsin Alumni Research Foundation, the Wisconsin Institute for Discovery, and the National Science Foundation. This research was also performed in part using the Wisconsin Energy Institute computing cluster, which is supported by the Great Lakes Bioenergy Research Center as part of the U.S. Department of Energy Office of Science.

The U.S. National Science Foundation North Temperate Lakes Long-Term Ecological Research site (NTL-LTER DEB-1440297) provided funding to the Microbial Observatory for long-term sampling of Lake Mendota and Trout Bog. Funding to E.A.M. was provided by a fellowship through the Department of Bacteriology at the University of



Wisconsin—Madison. Funding was also provided to K.D.M. by the Wisconsin Alumni Research Foundation and University of Wisconsin Sea Grant Institute under grants from the National Sea Grant College Program, National Oceanic and Atmospheric Administration, U.S. department of Commerce, and from the State of Wisconsin (federal grant number NA14OAR4170092, project number R/HCE-22).

## REFERENCES

- Wood JM. 1974. Biological cycles for toxic elements in the environment. *Science* 183:1049–1052. <https://doi.org/10.1126/science.183.4129.1049>.
- Hammerschmidt CR, Fitzgerald WF. 2006. Methylmercury in freshwater fish linked to atmospheric mercury deposition. *Environ Sci Technol* 40:7764–7770. <https://doi.org/10.1021/es061480i>.
- Sunderland EM, Li M, Bullard K. 2018. Decadal changes in the edible supply of seafood and methylmercury exposure in the United States. *Environ Health Perspect* 126:017006. <https://doi.org/10.1289/EHP2644>.
- Mahaffey KR, Clickner RP, Jeffries RA. 2009. Adult women's blood mercury concentrations vary regionally in the United States: association with patterns of fish consumption (NHANES 1999-2004). *Environ Health Perspect* 117:47–53. <https://doi.org/10.1289/ehp.11674>.
- Schober SE, Sinks TH, Jones RL, Bolger PM, McDowell M, Osterloh J, Garrett ES, Canady RA, Dillon CF, Sun Y, Joseph CB, Mahaffey KR. 2003. Blood mercury levels in US children and women of childbearing age, 1999-2000. *JAMA* 289:1667–1674. <https://doi.org/10.1001/jama.289.13.1667>.
- Krabbenhoft DP. 2004. Methylmercury contamination of aquatic ecosystems: a widespread problem with many challenges for the chemical sciences. In Norling P, Wook-Black F, Masciangioli TM (ed), *Water and sustainable development: opportunities for the chemical sciences: a workshop report to the Chemical Sciences Roundtable*. National Academies Press, Washington, DC.
- Hsu-Kim H, Kucharzyk KH, Zhang T, Deshusses MA. 2013. Mechanisms regulating mercury bioavailability for methylating microorganisms in the aquatic environment: a critical review. *Environ Sci Technol* 47:2441–2456. <https://doi.org/10.1021/es304370g>.
- Choi S-C, Chase T, Bartha R. 1994. Metabolic pathways leading to mercury methylation in *Desulfovibrio desulfuricans*. *Appl Environ Microbiol* 60:4072–4077. <https://doi.org/10.1128/AEM.60.11.4072-4077.1994>.
- Kerin EJ, Gilmour CC, Roden E, Suzuki MT, Coates JD, Mason RP. 2006. Mercury methylation by dissimilatory iron-reducing bacteria. *Appl Environ Microbiol* 72:7919–7921. <https://doi.org/10.1128/AEM.01602-06>.
- Yu R-Q, Reinfelder JR, Hines ME, Barkay T. 2013. Mercury methylation by the methanogen *Methanospirillum hungatei*. *Appl Environ Microbiol* 79:6325–6330. <https://doi.org/10.1128/AEM.01556-13>.
- Compeau GC, Bartha R. 1985. Sulfate-reducing bacteria: principal methylators of mercury in anoxic estuarine sediment. *Appl Environ Microbiol* 50:498–502. <https://doi.org/10.1128/AEM.50.2.498-502.1985>.
- Fleming EJ, Mack EE, Green PG, Nelson DC. 2006. Mercury methylation from unexpected sources: molybdate-inhibited freshwater sediments and an iron-reducing bacterium. *Appl Environ Microbiol* 72:457–464. <https://doi.org/10.1128/AEM.72.1.457-464.2006>.
- Devereux R, Winfrey MR, Winfrey J, Stahl DA. 1996. Depth profile of sulfate-reducing bacterial ribosomal RNA and mercury methylation in an estuarine sediment. *FEMS Microbiol Ecol* 20:23–31. <https://doi.org/10.1111/j.1574-6941.1996.tb00301.x>.
- Macalady JL, Mack EE, Nelson DC, Scow KM. 2000. Sediment microbial community structure and mercury methylation in mercury-polluted Clear Lake, California. *Appl Environ Microbiol* 66:1479–1488. <https://doi.org/10.1128/aem.66.4.1479-1488.2000>.
- Ranchou-Peyruse M, Monperrus M, Bridou R, Duran R, Amouroux D, Salvado JC, Guyoneaud R. 2009. Overview of mercury methylation capacities among anaerobic bacteria including representatives of the sulphate-reducers: implications for environmental studies. *Geomicrobiol J* 26:1–8. <https://doi.org/10.1080/01490450802599227>.
- Gilmour CC, Bullock AL, McBurney A, Podar M, Elias DA. 2018. Robust mercury methylation across diverse methanogenic *Archaea*. *mBio* 9:e02403-17. <https://doi.org/10.1128/mBio.02403-17>.
- Gilmour CC, Podar M, Bullock AL, Graham AM, Brown SD, Somenahally AC, Johs A, Hurt RA, Bailey KL, Elias DA. 2013. Mercury methylation by novel microorganisms from new environments. *Environ Sci Technol* 47:11810–11820. <https://doi.org/10.1021/es403075t>.
- Christensen GA, Gionfriddo CM, King AJ, Moberly JG, Miller CL, Somenahally AC, Callister SJ, Brewer H, Podar M, Brown SD, Palumbo AV, Brandt CC, Wymore AM, Brooks SC, Hwang C, Fields MW, Wall JD, Gilmour CC, Elias DA. 2019. Determining the reliability of measuring mercury cycling gene abundance with correlations with mercury and methylmercury concentrations. *Environ Sci Technol* 53:8649–8663. <https://doi.org/10.1021/acs.est.8b06389>.
- Parks JM, Johs A, Podar M, Bridou R, Hurt RA, Smith SD, Tomanicek SJ, Qian Y, Brown SD, Brandt CC, Palumbo AV, Smith JC, Wall JD, Elias DA, Liang L. 2013. The genetic basis for bacterial mercury methylation. *Science* 339:1332–1335. <https://doi.org/10.1126/science.1230667>.
- Podar M, Gilmour CC, Brandt CC, Soren A, Brown SD, Crable BR, Palumbo AV, Somenahally AC, Elias DA. 2015. Global prevalence and distribution of genes and microorganisms involved in mercury methylation. *Sci Adv* 1:e1500675. <https://doi.org/10.1126/sciadv.1500675>.
- Gionfriddo CM, Wymore AM, Jones DS, Wilpiszkeski RL, Lynes MM, Christensen GA, Soren A, Gilmour CC, Podar M, Elias DA. 2020. An improved hgcAB primer set and direct high-throughput sequencing expand Hg-methylator diversity in nature. *bioRxiv* <https://doi.org/10.1101/2020.03.10.983866>.
- Christensen GA, Wymore AM, King AJ, Podar M, Hurt RA, Santillan EU, Soren A, Brandt CC, Brown SD, Palumbo AV, Wall JD, Gilmour CC, Elias DA. 2016. Development and validation of broad-range qualitative and clade-specific quantitative molecular probes for assessing mercury methylation in the environment. *Appl Environ Microbiol* 82:6068–6078. <https://doi.org/10.1128/AEM.01271-16>.
- Liu YR, Yu RQ, Zheng YM, He JZ. 2014. Analysis of the microbial community structure by monitoring an Hg methylation gene (*hgcA*) in paddy soils along an Hg gradient. *Appl Environ Microbiol* 80:2874–2879. <https://doi.org/10.1128/AEM.04225-13>.
- Bravo AG, Peura S, Buck M, Ahmed O, Mateos-Rivera A, Ortega SH, Schaefer JK, Bouchet S, Tolu J, Björn E, Bertilsson S. 2018. Methanogens and iron-reducing bacteria: the overlooked members of mercury-methylating microbial communities in boreal lakes. *Appl Environ Microbiol* 84:e01774-18. <https://doi.org/10.1128/AEM.01774-18>.
- Bravo AG, Zopfi J, Buck M, Xu J, Bertilsson S, Schaefer JK, Poté J, Cosio C. 2018. *Geobacteraceae* are important members of mercury-methylating microbial communities of sediments impacted by waste water releases. *ISME J* 12:802–812. <https://doi.org/10.1038/s41396-017-0007-7>.
- Xu J, Buck M, Eklöf K, Ahmed OO, Schaefer JK, Bishop K, Skjellberg U, Björn E, Bertilsson S, Bravo AG. 2019. Mercury methylating microbial communities of boreal forest soils. *Sci Rep* 9:518. <https://doi.org/10.1038/s41598-018-37383-z>.
- Gionfriddo CM, Tate MT, Wick RR, Schultz MB, Zemla A, Thelen MP, Schofield R, Krabbenhoft DP, Holt KE, Moreau JW. 2016. Microbial mercury methylation in Antarctic sea ice. *Nat Microbiol* 1:16127. <https://doi.org/10.1038/nmicrobiol.2016.127>.
- Villar E, Cabrol L, Heimbürger-Boavida L. 2020. Widespread microbial mercury methylation genes in the global ocean. *Environ Microbiol Rep* 12:277–287. <https://doi.org/10.1111/1758-2229.12829>.
- Jones DS, Walker GM, Johnson NW, Mitchell CPJ, Coleman Wasik JK, Bailey JV. 2019. Molecular evidence for novel mercury methylating microorganisms in sulfate-impacted lakes. *ISME J* 13:1659–1675. <https://doi.org/10.1038/s41396-019-0376-1>.
- Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS, Banfield JF. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428:37–43. <https://doi.org/10.1038/nature02340>.
- Woodcroft BJ, Singleton CM, Boyd JA, Evans PN, Emerson JB, Zayed AAF, Hoelzle RD, Lambertson TO, McCalley CK, Hodgkins SB, Wilson RM, Purvine SO, Nicora CD, Li C, Frolking S, Chanton JP, Crill PM, Saleska SR,

- Rich VI, Tyson GW. 2018. Genome-centric view of carbon processing in thawing permafrost. *Nature* 560:49–54. <https://doi.org/10.1038/s41586-018-0338-1>.
32. Gionfriddo CM, Stott MB, Power JF, Ogorek JM, Krabbenhoft DP, Wick R, Holt K, Chen L-X, Thomas BC, Banfield JF, Moreau JW. 2020. Genome-resolved metagenomics and detailed geochemical speciation analyses yield new insights into microbial mercury cycling in geothermal springs. *Appl Environ Microbiol* 86:e00176-20. <https://doi.org/10.1128/AEM.00176-20>.
  33. Peterson BD, McDaniel EA, Schmidt AG, Lepak RF, Tran PQ, Marick RA, Ogorek JM, DeWild JF, Krabbenhoft DP, McMahon KD. 3 April 2020. Mercury methylation trait dispersed across diverse anaerobic microbial guilds in a eutrophic sulfate-enriched lake. *bioRxiv* <https://doi.org/10.1101/2020.04.01.018762>.
  34. Tran PQ, McIntyre PB, Kraemer BM, Vadeboncoeur Y, Kimirei IA, Tamatamah R, McMahon KD, Anantharaman K. 8 November 2019. Depth-discrete eco-genomics of Lake Tanganyika reveals roles of diverse microbes, including candidate phyla, in tropical freshwater nutrient cycling. *bioRxiv* <https://doi.org/10.1101/834861>.
  35. He S, Stevens SLR, Chan L-K, Bertilsson S, del Rio TG, Tringe SG, Malmstrom RR, McMahon KD. 2017. Ecophysiology of freshwater *Verrucomicrobia* inferred from metagenome-assembled genomes. *mSphere* 2:e00277-17. <https://doi.org/10.1128/mSphere.00277-17>.
  36. Tran P, Ramachandran A, Khawasik O, Beisner BE, Rautio M, Huot Y, Walsh DA. 2018. Microbial life under ice: metagenome diversity and *in situ* activity of *Verrucomicrobia* in seasonally ice-covered lakes. *Environ Microbiol* 20:2568–2584. <https://doi.org/10.1111/1462-2920.14283>.
  37. Chiang E, Schmidt ML, Berry MA, Biddanda BA, Burtner A, Johengen TH, Palladino D, Deneff VJ. 2018. *Verrucomicrobia* are prevalent in north-temperate freshwater lakes and display class-level preferences between lake habitats. *PLoS One* 13:e0195112. <https://doi.org/10.1371/journal.pone.0195112>.
  38. Cabello-Yeves PJ, Ghai R, Mehrshad M, Picazo A, Camacho A, Rodriguez-Valera F. 2017. Reconstruction of diverse verrucomicrobial genomes from metagenome datasets of freshwater reservoirs. *Front Microbiol* 8:2131. <https://doi.org/10.3389/fmicb.2017.02131>.
  39. Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, Thomas BC, Singh A, Wilkins MJ, Karaoz U, Brodie EL, Williams KH, Hubbard SS, Banfield JF. 2016. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat Commun* 7:13219. <https://doi.org/10.1038/ncomms13219>.
  40. Tully BJ, Graham ED, Heidelberg JF. 2018. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. *Sci Data* 5:170203. <https://doi.org/10.1038/sdata.2017.203>.
  41. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW. 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* 2:1533–1542. <https://doi.org/10.1038/s41564-017-0012-7>.
  42. Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S. 2011. A guide to the natural history of freshwater lake bacteria. *Microbiol Mol Biol Rev* 75:14–49. <https://doi.org/10.1128/MMBR.00028-10>.
  43. Newton RJ, Jones SE, Helmus MR, McMahon KD. 2007. Phylogenetic ecology of the freshwater *Actinobacteria* acl lineage. *Appl Environ Microbiol* 73:7169–7176. <https://doi.org/10.1128/AEM.00794-07>.
  44. Crump BC, Armbrust EV, Baross JA. 1999. Phylogenetic analysis of particle-attached and free-living bacterial communities in the Columbia river, its estuary, and the adjacent coastal ocean. *Appl Environ Microbiol* 65:3192–3204. <https://doi.org/10.1128/AEM.65.7.3192-3204.1999>.
  45. Glöckner FO, Zaichikov E, Belkova N, Denissova L, Pernthaler J, Pernthaler A, Amann R. 2000. Comparative 16S rRNA analysis of lake bacterioplankton reveals globally distributed phylogenetic clusters including an abundant group of actinobacteria. *Appl Environ Microbiol* 66:5053–5065. <https://doi.org/10.1128/aem.66.11.5053-5065.2000>.
  46. Ludwig W, Euzéby J, Schumann P, Busse H-J, Trujillo ME, Kämpfer P, Whitman WB. 2012. Road map of the phylum *Actinobacteria*, p 1–28. *Bergey's Manual of Systematic Bacteriology*. Springer, New York, NY.
  47. Bisanz JE, Soto-Perez P, Lam KN, Bess EN, Haiser HJ, Allen-Vercoe E, Rekdal VM, Balskus EP, Turnbaugh PJ. 1 May 2018. Illuminating the microbiome's dark matter: a functional genomic toolkit for the study of human gut *Actinobacteria*. *bioRxiv* <https://doi.org/10.1101/304840>.
  48. Hugenholtz P, Pitulle C, Hershberger KL, Pace NR. 1998. Novel division level bacterial diversity in a Yellowstone hot spring. *J Bacteriol* 180:366–376. <https://doi.org/10.1128/JB.180.2.366-376.1998>.
  49. Bird JT, Tague ED, Zinke L, Schmidt JM, Steen AD, Reese B, Marshall IPG, Webster G, Weightman A, Castro HF, Campagna SR, Lloyd KG. 2019. Uncultured microbial phyla suggest mechanisms for multi-thousand-year subsistence in Baltic Sea sediments. *mBio* 10:e02376-18. <https://doi.org/10.1128/mBio.02376-18>.
  50. Zarilla KA, Perry JJ. 1986. Deoxyribonucleic acid homology and other comparisons among obligately thermophilic hydrocarbonoclastic bacteria, with a proposal for *Thermoleophilum minutum* sp. nov. *Int J Syst Bacteriol* 36:13–16. <https://doi.org/10.1099/00207713-36-1-13>.
  51. Suzuki K, Whitman WB. 2015. *Thermoleophilum* class. nov., p 1–4. *Bergey's manual of systematics of Archaea and Bacteria*. John Wiley & Sons, Ltd., Chichester, UK.
  52. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 36:996–1004. <https://doi.org/10.1038/nbt.4229>.
  53. Hu D, Zang Y, Mao Y, Gao B. 2019. Identification of molecular markers that are specific to the class *Thermoleophilum*. *Front Microbiol* 10:1185. <https://doi.org/10.3389/fmicb.2019.01185>.
  54. Vorholt JA, Chistoserdova L, Stolyar SM, Thauer RK, Lidstrom ME. 1999. Distribution of tetrahydromethanopterin-dependent enzymes in methylotrophic bacteria and phylogeny of methenyl tetrahydromethanopterin cyclohydrolases. *J Bacteriol* 181:5750–5757. <https://doi.org/10.1128/JB.181.18.5750-5757.1999>.
  55. Crowther GJ, Kosály G, Lidstrom ME. 2008. Formate as the main branch point for methylotrophic metabolism in *Methylobacterium extorquens* AM1. *J Bacteriol* 190:5057–5062. <https://doi.org/10.1128/JB.00228-08>.
  56. Anderson AJ, Dawes EA. 1990. Occurrence, metabolism, metabolic role, and industrial uses of bacterial polyhydroxyalkanoates. *Microbiol Rev* 54:450–472. <https://doi.org/10.1128/MMBR.54.4.450-472.1990>.
  57. Chistoserdova L, Chen SW, Lapidus A, Lidstrom ME. 2003. Methylotrophy in *Methylobacterium extorquens* AM1 from a genomic point of view. *J Bacteriol* 185:2980–2987. <https://doi.org/10.1128/jb.185.10.2980-2987.2003>.
  58. Adam PS, Borrel G, Gribaldo S. 2018. Evolutionary history of carbon monoxide dehydrogenase/acetyl-CoA synthase, one of the oldest enzymatic complexes. *Proc Natl Acad Sci U S A* 115:E1166–E1173. <https://doi.org/10.1073/pnas.1716667115>.
  59. Hindson BJ, Ness KD, Masquelier DA, Belgrader P, Heredia NJ, Makarewicz AJ, Bright IJ, Lucero MY, Hiddessen AL, Legler TC, Kitano TK, Hodel MR, Petersen JF, Wyatt PW, Steenblock ER, Shah PH, Bousse LJ, Troup CB, Mellen JC, Wittmann DK, Erndt NG, Cauley TH, Koehler RT, So AP, Dube S, Rose KA, Montesclaros L, Wang S, Stumbo DP, Hodges SP, Romine S, Milanovich FP, White HE, Regan JF, Karlin-Neumann GA, Hindson CM, Saxonov S, Colston BW. 2011. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal Chem* 83:8604–8610. <https://doi.org/10.1021/ac202028g>.
  60. Selvaraj V, Maheshwari Y, Hajeri S, Chen J, McCollum TG, Yokomi R. 2018. Development of a duplex droplet digital PCR assay for absolute quantitative detection of “*Candidatus* liberibacter asiaticus”. *PLoS One* 13:e0197184. <https://doi.org/10.1371/journal.pone.0197184>.
  61. Todorova SG, Driscoll CT, Matthews DA, Effler SW, Hines ME, Henry EA. 2009. Evidence for regulation of monomethyl mercury by nitrate in a seasonally stratified, eutrophic lake. *Environ Sci Technol* 43:6572–6578. <https://doi.org/10.1021/es900887b>.
  62. Matthews DA, Babcock DB, Nolan JG, Prestigiacomo AR, Effler SW, Driscoll CT, Todorova SG, Kuhr KM. 2013. Whole-lake nitrate addition for control of methylmercury in mercury-contaminated Onondaga Lake, NY. *Environ Res* 125:52–60. <https://doi.org/10.1016/j.envres.2013.03.011>.
  63. Achour-Rokbani A, Cordi A, Poupin P, Bauda P, Billard P. 2010. Characterization of the *ars* gene cluster from extremely arsenic-resistant *Methylobacterium* sp. strain A33. *Appl Environ Microbiol* 76:948–955. <https://doi.org/10.1128/AEM.01738-09>.
  64. Mukhopadhyay R, Rosen BP. 2002. Arsenate reductases in prokaryotes and eukaryotes. *Environ Health Perspect* 110 Suppl 5:745–748. <https://doi.org/10.1289/ehp.02110s5745>.
  65. Schaefer JK, Rocks SS, Zheng W, Liang L, Gu B, Morel FMM. 2011. Active transport, substrate specificity, and methylation of Hg(II) in anaerobic bacteria. *Proc Natl Acad Sci U S A* 108:8714–8719. <https://doi.org/10.1073/pnas.1105781108>.

66. Rosen BP. 2002. Biochemistry of arsenic detoxification. *FEBS Lett* 529: 86–92. [https://doi.org/10.1016/S0014-5793\(02\)03186-1](https://doi.org/10.1016/S0014-5793(02)03186-1).
67. Goñi-Urriza M, Klopp C, Ranchou-Peyruse M, Ranchou-Peyruse A, Monperrus M, Khalfaoui-Hassani B, Guyoneaud R. 2020. Genome insights of mercury methylation among *Desulfovibrio* and *Pseudodesulfovibrio* strains. *Res Microbiol* 171:3–12. <https://doi.org/10.1016/j.resmic.2019.10.003>.
68. Busenlehner LS, Pennella MA, Giedroc DP. 2003. The SmtB/ArsR family of metalloregulatory transcriptional repressors: structural insights into prokaryotic metal resistance. *FEMS Microbiol Rev* 27:131–143. [https://doi.org/10.1016/S0168-6445\(03\)00054-8](https://doi.org/10.1016/S0168-6445(03)00054-8).
69. Murphy JN, Saltikov CW. 2009. The ArsR repressor mediates arsenite-dependent regulation of arsenate respiration and detoxification operons of *Shewanella* sp. strain ANA-3. *J Bacteriol* 191:6722–6731. <https://doi.org/10.1128/JB.00801-09>.
70. Goñi-Urriza M, Corsellis Y, Lanceluor L, Tessier E, Gury J, Monperrus M, Guyoneaud R. 2015. Relationships between bacterial energetic metabolism, mercury methylation potential, and *hgcA/hgcB* gene expression in *Desulfovibrio dechloroacetivorans* BerOc1. *Environ Sci Pollut Res Int* 22:13764–13771. <https://doi.org/10.1007/s11356-015-4273-5>.
71. Bravo AG, Loizeau J-L, Dranguet P, Makri S, Björn E, Ungureau VG, Slaveykova VI, Cosio C. 2016. Persistent Hg contamination and occurrence of Hg-methylating transcript (*hgcA*) downstream of a chlor-alkali plant in the Olt River (Romania). *Environ Sci Pollut Res Int* 23: 10529–10541. <https://doi.org/10.1007/s11356-015-5906-4>.
72. Loseto LL, Lean DRS, Siciliano SD. 2004. Snowmelt sources of methylmercury to high arctic ecosystems. *Environ Sci Technol* 38:3004–3010. <https://doi.org/10.1021/es035146n>.
73. Oiffer L, Siciliano SD. 2009. Methyl mercury production and loss in Arctic soil. *Sci Total Environ* 407:1691–1700. <https://doi.org/10.1016/j.scitotenv.2008.10.025>.
74. Barkay T, Kroer N, Poulain AJ. 2011. Some like it cold: microbial transformations of mercury in polar regions. *Polar Res* 30:15469. <https://doi.org/10.3402/polar.v30i0.15469>.
75. St Louis VL, Rudd JWM, Kelly CA, Bodaly RA, Paterson MJ, Beaty KG, Hesslein RH, Heyes A, Majewski AR. 2004. The rise and fall of mercury methylation in an experimental reservoir. *Environ Sci Technol* 38: 1348–1358. <https://doi.org/10.1021/es034424f>.
76. Kelly CA, Rudd JWM, Bodaly RA, Roulet NP, StLouis VL, Heyes A, Moore TR, Schiff S, Aravena R, Scott KJ, Dyck B, Harris R, Warner B, Edwards G. 1997. Increases in fluxes of greenhouse gases and methyl mercury following flooding of an experimental reservoir. *Environ Sci Technol* 31:1334–1344. <https://doi.org/10.1021/es960493i>.
77. Godon JJ, Zumstein E, Dabert P, Habouzit F, Moletta R. 1997. Molecular microbial diversity of an anaerobic digester as determined by small-subunit rDNA sequence analysis. *Appl Environ Microbiol* 63:2802–2813. <https://doi.org/10.1128/AEM.63.7.2802-2813.1997>.
78. Rojas P, Rodríguez N, de la Fuente V, Sánchez-Mata D, Amils R, Sanz JL. 2018. Microbial diversity associated with the anaerobic sediments of a soda lake (Mono Lake, California, USA). *Can J Microbiol* 64:385–392. <https://doi.org/10.1139/cjm-2017-0657>.
79. Baldwin SA, Khoshnoodi M, Rezaeebashi M, Taupp M, Hallam S, Mattes A, Sanei H. 2015. The microbial community of a passive biochemical reactor treating arsenic, zinc, and sulfate-rich seepage. *Front Bioeng Biotechnol* 3:27. <https://doi.org/10.3389/fbioe.2015.00027>.
80. Crits-Christoph A, Diamond S, Butterfield CN, Thomas BC, Banfield JF. 2018. Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. *Nature* 558:440–444. <https://doi.org/10.1038/s41586-018-0207-y>.
81. Kang DD, Froula J, Egan R, Wang Z. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 3:e1165. <https://doi.org/10.7717/peerj.1165>.
82. Wu Y-W, Tang Y-H, Tringe SG, Simmons BA, Singer SW. 2014. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* 2:26. <https://doi.org/10.1186/2049-2618-2-26>.
83. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 27:824–834. <https://doi.org/10.1101/gr.213959.116>.
84. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol* 3:836–843. <https://doi.org/10.1038/s41564-018-0171-1>.
85. Olm MR, Brown CT, Brooks B, Banfield JF. 2017. DRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J* 11:2864–2868. <https://doi.org/10.1038/ismej.2017.126>.
86. Alneberg J, Bjarnason BS, De Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C. 2014. Binning metagenomic contigs by coverage and composition. *Nat Methods* 11:1144–1146. <https://doi.org/10.1038/nmeth.3103>.
87. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* 3:e1319. <https://doi.org/10.7717/peerj.1319>.
88. Bendall ML, Stevens SL, Chan L-K, Malfatti S, Schwientek P, Tremblay J, Schackwitz W, Martin J, Pati A, Bushnell B, Froula J, Kang D, Tringe SG, Bertilsson S, Moran MA, Shade A, Newton RJ, McMahon KD, Malmstrom RR. 2016. Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. *ISME J* 10:1589–1601. <https://doi.org/10.1038/ismej.2015.241>.
89. Linz AM, He S, Stevens SL, Anantharaman K, Rohwer RR, Malmstrom RR, Bertilsson S, McMahon KD. 2018. Freshwater carbon and nutrient cycles revealed through reconstructed population genomes. *PeerJ* 6:e6075. <https://doi.org/10.7717/peerj.6075>.
90. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>.
91. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloë-Fadrosch EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA, Yooseph S, Sutton G, Glöckner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Ettema TJG, Tighe S, Konstantinidis KT, Liu W-T, Baker BJ, Rattai T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Kyrpides NC, Schirmer M, Garrity GM, Hugenholtz P, Sutton G, et al. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* 35:725–731. <https://doi.org/10.1038/nbt.3893>.
92. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <https://doi.org/10.1186/1471-2105-11-119>.
93. Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol* 7:e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>.
94. Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30:3059–3066. <https://doi.org/10.1093/nar/gkf436>.
95. Larsson A. 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30:3276–3278. <https://doi.org/10.1093/bioinformatics/btu531>.
96. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hershendorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC, Banfield JF. 2016. A new view of the tree of life. *Nat Microbiol* 1:16048. <https://doi.org/10.1038/nmicrobiol.2016.48>.
97. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. <https://doi.org/10.1371/journal.pone.0009490>.
98. Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 44:W242–W245. <https://doi.org/10.1093/nar/gkw290>.
99. Aberer AJ, Krompass D, Stamatakis A. 2013. Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice. *Syst Biol* 62:162–166. <https://doi.org/10.1093/sysbio/sys078>.
100. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <https://doi.org/10.1093/nar/gkh340>.
101. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Grüning BA, Guerler A, Hillman-Jackson J, Hiltmann S, Jalili V, Rasche H, Soranzo N, Goecks J, Taylor J, Nekrutenko A, Blankenberg D. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res* 46:W537–W544. <https://doi.org/10.1093/nar/gky379>.
102. Criscuolo A, Gribaldo S. 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative

- regions from multiple sequence alignments. *BMC Evol Biol* 10:210. <https://doi.org/10.1186/1471-2148-10-210>.
103. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.
  104. McDaniel EA, Anantharaman K, McMahon KD. 21 December 2019. metabolisHMM: phylogenomic analysis for exploration of microbial phylogenies and metabolic pathways. *bioRxiv* <https://doi.org/10.1101/2019.12.20.884627>.
  105. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
  106. Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: a genome comparison visualizer. *Bioinformatics* 27:1009–1010. <https://doi.org/10.1093/bioinformatics/btr039>.
  107. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H. 2019. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* 36: 2251–2252. <https://doi.org/10.1093/bioinformatics/btz859>.
  108. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>.
  109. Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* 34:525–527. <https://doi.org/10.1038/nbt.3519>.
  110. Wagner GP, Kin K, Lynch VJ. 2012. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci* 131:281–285. <https://doi.org/10.1007/s12064-012-0162-3>.