


On the Origin and Evolution of Microbial Mercury Methylation

Heyu Lin ^{1,†}, Edmund R.R. Moody², Tom A. Williams ³, and John W. Moreau^{1,4,*},[†]

¹School of Geographical, Atmospheric and Earth Sciences, The University of Melbourne, Parkville, Victoria, Australia

²School of Earth Sciences, University of Bristol, United Kingdom

³School of Biological Sciences, University of Bristol, United Kingdom

⁴School of Geographical and Earth Sciences, University of Glasgow, United Kingdom

[†]These authors contributed equally to this work.

*Corresponding author: E-mail: john.moreau@glasgow.ac.uk.

Accepted: 14 March 2023

Abstract

The origin of microbial mercury methylation has long been a mystery. Here, we employed genome-resolved phylogenetic analyses to decipher the evolution of the mercury-methylating gene, *hgcAB*, constrain the ancestral origin of the *hgc* operon, and explain the distribution of *hgc* in Bacteria and Archaea. We infer the extent to which vertical inheritance and horizontal gene transfer have influenced the evolution of mercury methylators and hypothesize that evolution of this trait bestowed the ability to produce an antimicrobial compound (MeHg⁺) on a potentially resource-limited early Earth. We speculate that, in response, the evolution of MeHg⁺-detoxifying alkylmercury lyase (encoded by *merB*) reduced a selective advantage for mercury methylators and resulted in widespread loss of *hgc* in Bacteria and Archaea.

Key words: mercury, methylmercury, evolution, antimicrobial, *hgc* gene, gene loss, horizontal gene transfer, LUCA.

Significance

Neurotoxic methylmercury (MeHg⁺_(aq)) is synthesized from Hg^{II}_(aq) in the environment by microorganisms possessing the gene pair *hgcAB*. Our phylogenetic analyses elucidate the origin and evolution of the *hgc* operon and support a hypothesis that mercury methylation evolved as an antimicrobial production mechanism, possibly from competition for limited resources on the early Earth. We infer from our analyses that *hgc* has been primarily vertically inherited in Bacteria and Archaea, with extensive parallel loss, and note that few taxa possessing *hgc* also possess the gene encoding for MeHg⁺ demethylation, *merB*. Our findings support the interpretation that *merB* evolved as a defense mechanism against the evolution of microbial Hg^{II}_(aq) methylation.

Introduction

Microorganisms play key roles in the global mercury (Hg) cycle. In soils, sediments, and natural waters, microbes mediate the enzymatic reductive volatilization of Hg^{II}_(aq) to Hg⁰_(g), the transformation of Hg^{II}_(aq) to MeHg⁺_(aq), the complexation of aqueous Hg^{II}_(aq) by bacteriogenic bisulfide, and the adsorption/desorption of Hg^{II}_(aq) on/from biogenic

metal ox(yhydrox)ides (e.g., Barkay et al. 2003; Tebo et al. 2004; Barkay and Wagner-Döbler 2005; Gilmour et al. 2013; Hsu-Kim et al. 2013; Barkay and Gu 2022; Wang et al. 2022). Essentially, microbes control or strongly influence the geochemical speciation of Hg in most natural environments today. Microbial Hg methylation in particular has received much attention because MeHg⁺ is a potent

© The Author(s) 2023. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

neurotoxin that bioaccumulates across both marine and terrestrial food webs (Kidd et al. 2012). The origin of Hg methylation has remained a mystery, largely because MeHg⁺ serves no known biological function.

Microbial Hg methylation requires Hg^{II}_(aq), and on the early Earth, volcanoes would have emitted abundant Hg⁰_(g) and some Hg^{II} (Grasby et al. 2019). However, studies have shown that Hg⁰_(g) can be efficiently oxidized to Hg^{II}_(aq) by reactive Cl⁻_(aq) and Br⁻_(aq) (e.g., Amyot et al. 2005; Seigneur and Lohman 2008; Obrist et al. 2011; Wang et al. 2014). Volcanoes that emitted Hg⁰_(g) on the early Earth could also have emitted reactive HCl and BrO (Bobrowski et al. 2003; Guo and Korenaga 2021), providing oxidants to convert Hg⁰_(g) to Hg^{II}_(aq). Recent work supports a dual role for Archean volcanism in Hg⁰_(g) and halogen emissions and subsequent oxidation of Hg⁰_(g) to Hg^{II}_(aq) (Zerkle et al. 2021), as well as the possibility of ephemeral periods of atmospheric oxygenation prior to the Great Oxidation Event at ~2.4 billion years ago (Meixnerova et al. 2021). Following this event, Hg^{II}_(aq) would have been commonly found as a trace metal in aqueous environments. We assume therefore that Hg^{II}_(aq) would have been available on the early Earth for methylation upon evolution of the ancestral *hgc* (Hg-methylating) operon. Intriguingly, however, the obligately anaerobic and “canonical” Hg methylator, *Pseudodesulfovibrio mercurii*, can apparently oxidize dissolved Hg⁰ directly and methylate the resulting Hg^{II}_(aq) (Colombo et al. 2013).

Microbial Hg methylation is currently thought to require the *hgc* operon, most closely related to the ancient carbon monoxide dehydrogenase/acetyl-CoA synthase gene, *cdh*, suggested to have been present in the last universal common ancestor (“LUCA”; Parks et al. 2013; Sousa and Martin 2014; Weiss et al. 2016, 2018; Adam et al. 2018). The cobalamin-dependent proteins CdhD and CdhE (hereafter “CdhDE”) are present in a highly conserved operon in genomes enabling methyl transfer in the Wood–Ljungdahl (WL) pathway for autotrophic carbon assimilation (Svetlitchnaia et al. 2006). Although HgcA is also predicted to be a cobalamin-dependent methyltransferase, it only functions in Hg^{II}_(aq) methylation and does not seem to take part in the WL pathway, as other components of this pathway are incomplete in characterized *hgcA*-carrying genomes (Date et al. 2019). Similar to *cdh*, *hgc* seems to be broadly distributed across Bacteria and Archaea representing a range of environmental habitats and redox potentials (McDaniel et al. 2020; Capo et al. 2023). Some of these bacteria and archaea have been experimentally confirmed to methylate Hg^{II}_(aq), while others are still considered as “putative methylators” encoding homologous *hgcAB* sequences that have not yet been experimentally validated. In lieu of cultivated isolates, computational modeling has been employed to test for consistency between putative HgcAB sequences and HgcAB structure and functionality

in confirmed methylators (Gionfriddo et al. 2016; Lin et al. 2021).

Many studies of Hg methylation have focused on environmental factors influencing Hg^{II}_(aq) bioavailability for uptake by methylating cells. In addition, previous efforts to identify the biochemical mechanism of Hg^{II}_(aq) methylation have attributed the process to 1) accidental enzymatic catalysis of a methyl group transfer to Hg^{II}_(aq) during either acetyl-CoA formation (Choi et al. 1994) or methionine synthesis (Siciliano and Lean 2002) or 2) an as-yet unidentified pathway (Ekstrom et al. 2003). Here, we focus on understanding the more elusive origin and evolution of Hg-methylating genes, employing genome-resolved phylogenetic analyses to constrain the ancestry of the *hgc* operon and explain its currently known distribution in Bacteria and Archaea. We assess the extent to which vertical descent and horizontal gene transfer (HGT) have shaped *hgc* evolution and propose a hypothesis for the functional origin of microbial Hg methylation as an antimicrobial production mechanism. We also evaluate a potential evolutionary link between *hgc* and *merB*, the latter gene encoding for organomercury lyase (MerB), which lyses the C–Hg bond in MeHg⁺ to release Hg^{II}_(aq) for reductive volatilization (i.e., detoxification) by MerA (Tezuka and Tonomura 1976).

Results and Discussion

hgc Present in LUCA

A data set containing 478 protein sequences belonging to the protein family PF03599 was retrieved from UniProt Reference Proteomes database v2022_03 (Chen et al. 2011) at 35% cutoff (RP35) to study the phylogeny of the protein family (supplementary tables S1 and S2, Supplementary Material online). Phylogenetic reconstruction of the protein family PF03599 defined three deep-branching clusters, comprising HgcA, CdhD, and CdhE (fig. 1A; for details, see supplementary fig. S1, Supplementary Material online). On the assumption that CdhD- and CdhE-encoding genes were present in LUCA (Sousa and Martin 2014; Adam et al. 2018; Weiss et al. 2018), we deduce that *hgcA* was also likely to be present (i.e., these three protein subfamilies diverged in, or before, LUCA; fig. 1A). While robustly rooting single gene trees is challenging, midpoint, Minvar (Mai et al. 2017), and minimal ancestor deviation (MAD) (Tria et al. 2017) rooting approaches all supported a root between CdhD and CdhE + HgcA. In comparison with genes encoding for CdhD and CdhE, which exhibit relatively wide distributions across extant prokaryotes and broadly congruent evolutionary histories, *hgcA* exhibits a more restricted distribution across extant taxa. The latter gene is mainly found in Bacteria, mostly affiliated with Deltaproteobacteria, Firmicutes, and the FCB group, although with a more restricted distribution

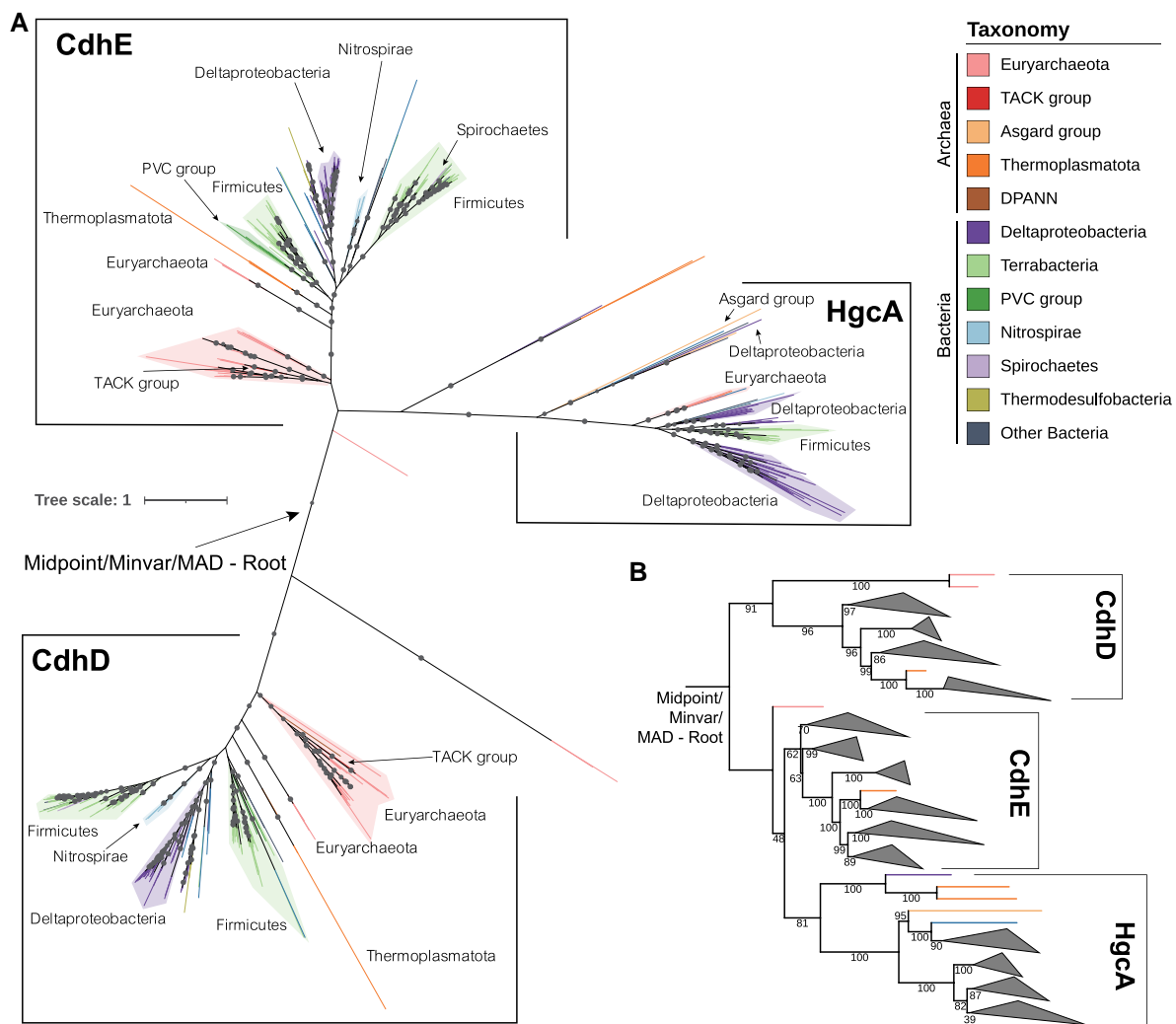


FIG. 1.—Phylogenetic tree of the protein family PF03599. (A) Unrooted tree of the protein family PF03599. The tree was inferred by using the ML method under LG + C50 + F + R model. This analysis involved 478 amino acid sequences with a total of 2,922 positions in the alignments. Different taxonomies are represented by different colors. Ultrafast bootstrap support values were calculated with 1,000 replications, and ultrafast bootstrap values > 90% are shown by black dots at the nodes. (B) PF03599 tree rooted by the midpoint. Clades whose average branch length distance to their leaves are below 1.5 are collapsed with iTOL for better visualization.

than in Archaea (fig. 1A). Based on the rooted protein tree (fig. 1B), HgcA has a longer stem than Cdhd and Cdhe, which might have resulted from accelerated evolution (e.g., associated with a change in function) or gene loss (or lineage extinction) in former HgcA-encoding clades.

hgcA Evolved Primarily Vertically but with Extensive Loss

In our phylogenetic analysis, to increase the number and taxonomic coverage of targets, we further enlarged the sample size by retrieving HgcA homologs in UniProt Reference Proteomes database v2022_03 at 75% cutoff (RP75). Two other data sets, including one containing 700 representative prokaryotic proteomes constructed by Moody et al. (2022) and another containing several novel

hgc carriers published by Lin et al. (2021), were retrieved and incorporated into the RP75 data set. After removing redundancies and including sequences from new phyla, the number of HgcA sequences in the data set was increased from 76 to 169 (supplementary table S1, Supplementary Material online). The resulting HgcA tree was rooted according to the topology of the HgcA subtree in figure 1.

Our analysis suggests that *hgcA* genes are separated into two well-supported clades with high bootstrap support, as shown in figures 1A and 2 (for details, see supplementary fig. S2, Supplementary Material online). These two clades correspond to a difference in gene structure: the smaller clade comprises genes in which *hgcA* and *hgcB* became fused to form *hgcAB*, whereas the larger contains single subunit (*hgcA*) genes. Several Asgard archaea, including

Downloaded from https://academic.oup.com/gbe/article/15/4/evad051/7084592 by University of Glasgow user on 12 April 2023

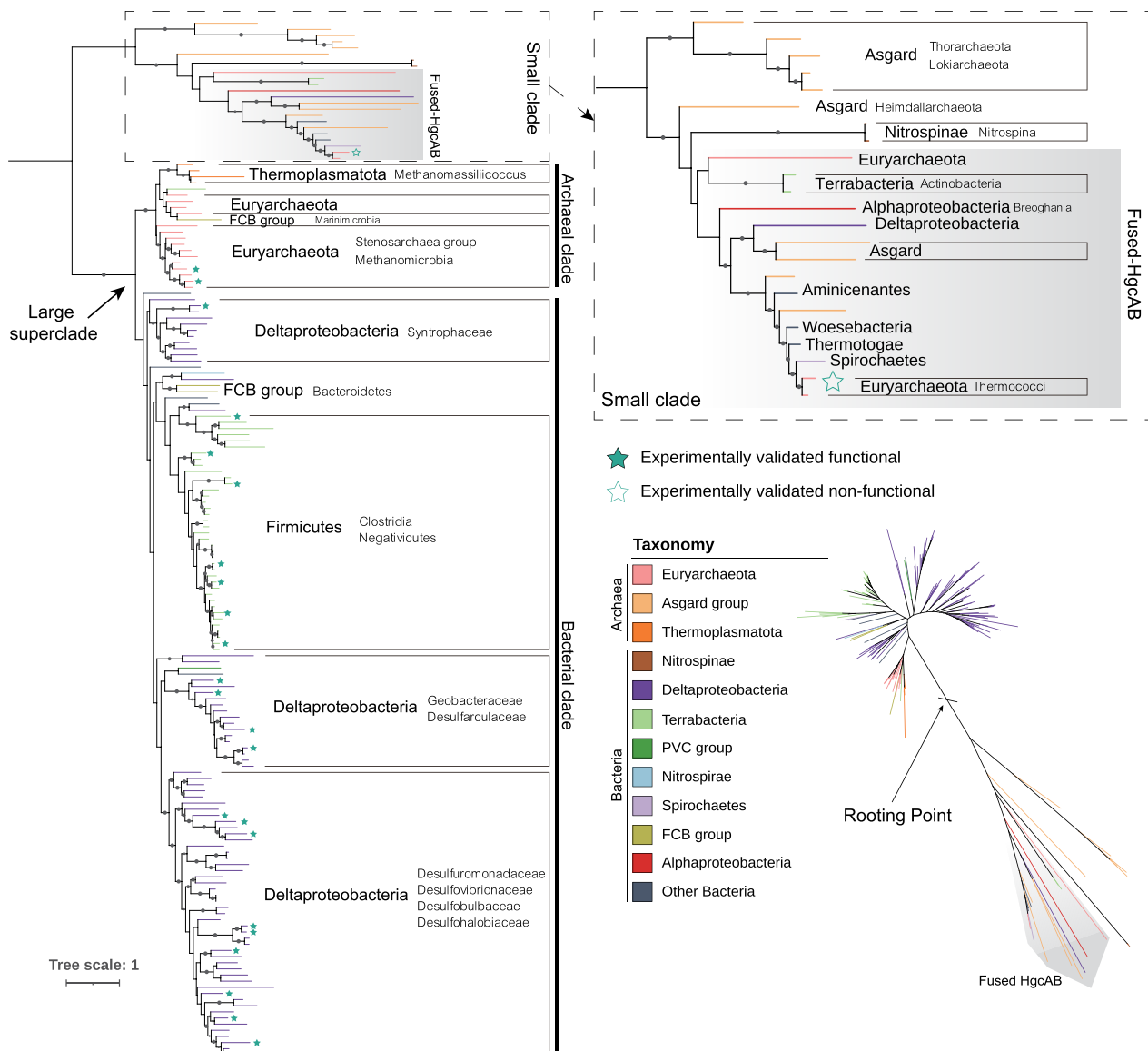


Fig. 2.—Phylogenetic tree of HgcA proteins. The tree was inferred by using the ML method under LG + C60 + F + G model. This analysis involved alignment of 169 amino acid sequences with a total of 493 positions. Different taxonomies are represented by different colors. Ultrabootstrap support values were calculated with 1,000 replications, and ultrabootstrap values > 90% are shown by black dots at the nodes. Experimentally validated functional and non-functional HgcA sequences are labeled by solid and hollow stars. Fused-HgcAB sequences are indicated by a gray background.

some Thorarchaeota and Lokiarchaeota, form a clade in the HgcAB subtree, suggesting that gene fusion occurred prior to the radiation of Asgard phyla. The fused-*hgcAB* of *Pyrococcus furiosus* has been experimentally shown not to methylate Hg^{II}_(aq) (Podar et al. 2015), but the functionality of other fused-*hgcAB* genes, as well as other members of this clade, is yet to be confirmed. For example, *Nitrospina*-like *hgcA* sequences have been reported as the dominant *hgcA* homologs in many marine environments (Gionfriddo et al. 2016; Tada et al. 2020, 2021; Villar et al. 2020). However, the functionality of these sequences has not yet been experimentally confirmed, as no known

hgc+ members of this genus are currently held in isolation. Also, the alphaproteobacterium *Breoghania* sp. L-A4 is the only aerobic *hgcA+* (fused-*hgcAB+*) microorganism found in this analysis, and its methylation capacity under aerobic conditions also remains untested.

By contrast, the majority of (nonfused) HgcA sequences form a clade which broadly follows the universal tree (fig. 2). In particular, the deepest split lies between archaeal and bacterial HgcA homologs. Taken together with the sister relationship of HgcA to CdhD/E, this finding supports an origin of HgcA in or before LUCA. However, the taxonomic distribution of HgcA in extant Bacteria is patchy (being

found mainly in Firmicutes, Deltaproteobacteria, and the FCB group). Therefore, if HgcA was indeed present in LUCA and its gene tree traces the divergence of Archaea and Bacteria, it must have been lost subsequently in the other bacterial lineages. While it is undoubtedly difficult to resolve the evolutionary history of ancient single genes with certainty, it is interesting that an ancient origin for HgcA cannot be excluded based on the inferred tree.

Gene loss is a major force driving microbial genome evolution (Bolotin and Hershberg 2016), reflecting environmental change and/or adaptation (Koskiniemi et al. 2012). So far, *hgcA*+ methylators have primarily been reported for anoxic or suboxic environments (Parks et al. 2013; Lin et al. 2021; Capo et al. 2023). Changes in redox potential that could inhibit Hg^{II}_(aq) methylation may have presented a selective environmental pressure to lose *hgcA*. In addition, long branches of the HgcA subtree, in comparison with CdhD/E, also suggest a weaker natural selection for gene maintenance and greater likelihood of loss through genetic drift, a common scenario observed for many microorganisms (Bolotin and Hershberg 2016).

A few putative HGT events were inferred for the larger clade of the HgcA tree, for example Marinimicrobia-HgcA clustered with Euryarchaeota-HgcA in the archaeal cluster, with 98% coverage and 56% identity by BlastP against HgcA from Theionarchaea archaeon DG-70-1, suggesting a possible lateral acquisition of *hgcA*. This inference is consistent with that of McDaniel et al. (2020), who also suggested HGT played an important role in HgcA evolution. Our study differs from that of McDaniel et al. in suggesting that *hgcA* may date back to the LUCA; this inference is based on the nesting of the HgcA subtree within the broader diversity of the Cdh family (fig. 1B), an aspect of the family's evolutionary history not considered in that analysis, as well as the recovery in our analyses of a deep split between archaeal and bacterial sequences in the *hgcA* subtree, notwithstanding some HGT.

hgcA and *hgcB* Coevolved

The gene *hgcB* is nearly always located immediately downstream of *hgcA*, with rare cases of being one gene apart (Gilmour et al. 2018). A lack of either *hgcA* or *hgcB* in a genome is thought to render the microbial host incapable of Hg^{II}_(aq) methylation (Parks et al. 2013). Therefore, the gene pair *hgcAB* likely evolved together as a conserved operon. Although the HgcB tree was poorly resolved due to short sequence length (95 amino acids on average), the overall topologies of the HgcB tree and the HgcA tree were congruent, supporting this inference (fig. 3; for details, see [supplementary fig. S3, Supplementary Material](#) online). Nevertheless, several *hgcA*+ genomes did not carry neighboring *hgcB* genes, including all *Nitrospina* and a few

Deltaproteobacteria and Firmicutes, potentially because of gene loss during evolution or incomplete transfer events (i.e., only *hgcA* genes were acquired during HGT events). Another exception involves a complete *hgcB* gene present downstream of the fused-*hgcAB* gene in the genome of the deltaproteobacterial endosymbiont, Delta1. This complete HgcB protein had the longest branch length of the HgcB tree and was located in a sister cluster of the “HgcB tail” associated with the fused-*hgcAB* from the same host, suggesting a possible lateral acquisition of *hgcB* from closely related taxa.

Phylogenetic Distribution of *merB* and Its Relationship to *hgc*

Microbial detoxification mechanisms for Hg^{II}_(aq) and MeHg⁺_(aq), including the enzymes MerA and MerB, respectively, have been well described (c.f., Robinson and Tuovinen 1984; Barkay et al. 2010; Barkay and Gu 2022). Recent phylogenetic analyses have revealed a thermophilic and archaeal origin for the *mer* operon, acquired later by bacteria through multiple independent transfer events (Christakis et al. 2021) reflective of the environmental advantage of Hg resistance genes (Boyd and Barkay 2012). In total, 225 MerB sequences were retrieved from the RP35 data set ([supplementary table S2, Supplementary Material](#) online). MerB homologs were distributed across Archaea and Bacteria (fig. 4; for details, see [supplementary fig. S4, Supplementary Material](#) online), with most representatives found in Terrabacteria (including Actinobacteria, Firmicutes, and Chloroflexi) and Proteobacteria (including Alphaproteobacteria, Betaproteobacteria, and Gammaproteobacteria). In contrast, only a small number of archaea, including a few euryarchaeotes and one Thaumarchaeon, were found to encode MerB. These archaeal MerB homologs were placed within bacterial MerB clades and were not monophyletic, consistent with the interpretation of recruitment of *merB* to the *mer* operon from a mesophilic ancestor (Christakis et al. 2021), which we infer involved HGT from Bacteria to Archaea.

We mapped the presence/absence of *merB* and *hgc* genes onto the species tree (fig. 5; for details, see [supplementary fig. S5, Supplementary Material](#) online) to investigate their phylogenetic relationships. Intriguingly, although the distribution of MerB homologs overlapped with HgcAB homologs, few genomes encode *both* MerB and HgcAB, signifying that Hg methylators generally cannot also demethylate MeHg⁺. This apparent polarization in gene distribution suggests a functional conflict between these two genes (i.e., there is utility in encoding for one or the other function, but not both). Only a few genomes were found to encode for both MerB and HgcAB; this dual functionality in *Citri fermentans bemidjiense* Bem was previously recognized and discussed (Lu et al. 2016). The phylogenetic distribution of these opposing enzymes

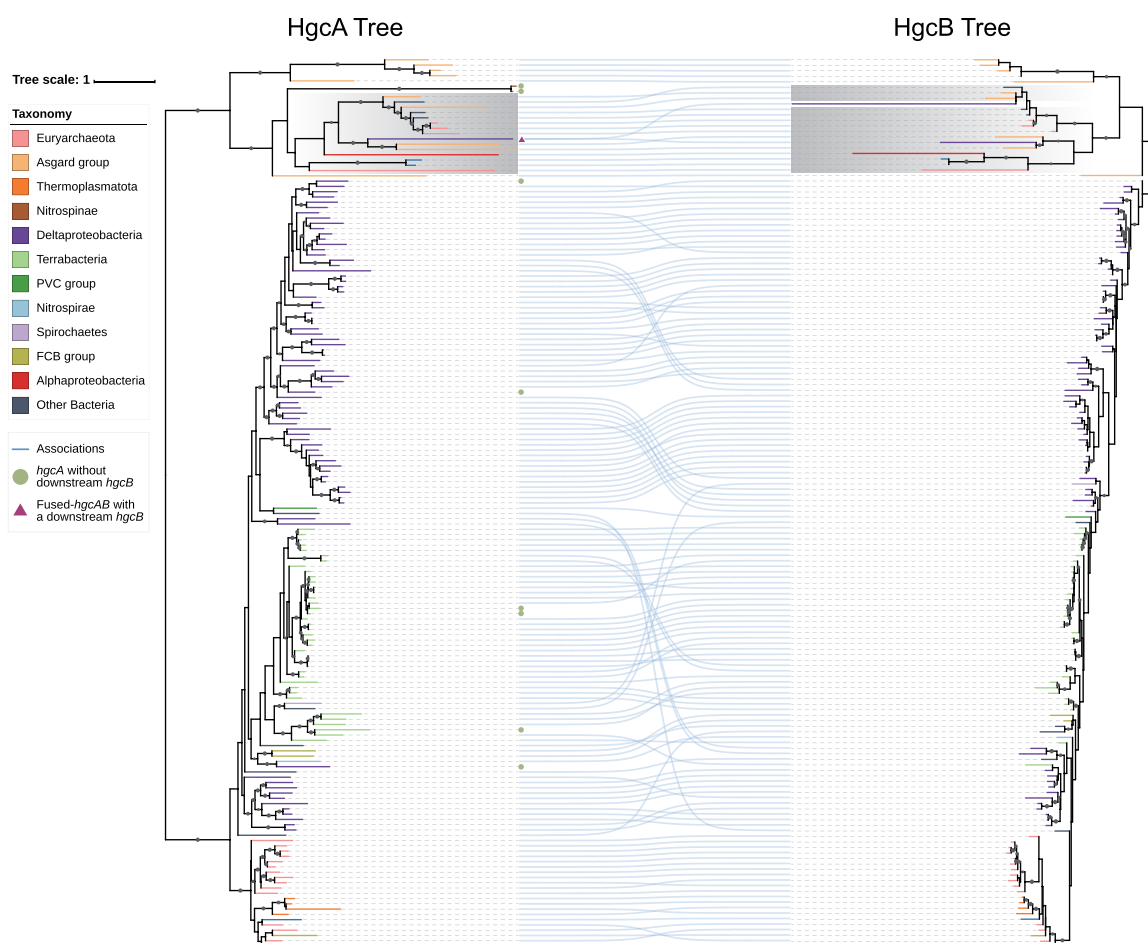


Fig. 3.—Cophylogenetic tree of HgcA and HgcB. The phylogeny of HgcA shown on the left is the same as figure 2. The phylogeny of HgcB shown on the right was inferred by using the ML method under LG + C60 + F + G model. Lines between the two trees connect HgcA (left) and HgcB (right) proteins from the same proteome, illustrating the degree of congruency in their respective phylogenetic associations. This analysis involved alignment of 169 amino acid sequences with a total of 322 positions. Taxonomies of the HgcB sequences are represented in the same colors as shown in the HgcA phylogeny. Ultrabootstrap support values were calculated with 1,000 replications, and ultrabootstrap values > 90% are shown by black dots at the nodes. Gray-shaded blocks describe fused-HgcAB and “HgcB tail” genes in the two trees, respectively. The green dots at the tip of the HgcA tree represent the corresponding *hgca* without downstream *hgcB*, and the red triangle represents the corresponding fused-*hgcAB* with another *hgcB* downstream.

(figs. 4 and 5) suggests that HgcAB was the earlier of the two to evolve, with MerB originating more recently and radiating via HGT.

Mercury Methylation as Antimicrobial Synthesis?

The reason why microorganisms evolved the ability to methylate Hg^{II} to MeHg^+ has long been a mystery. Previous studies proposed that this process might reduce susceptibility to $\text{Hg}^{\text{II}}_{(\text{aq})}$ toxicity (Trevors 1986). But in fact, the ability to produce $\text{MeHg}^+_{(\text{aq})}$ has been shown *not* to confer $\text{Hg}^{\text{II}}_{(\text{aq})}$ resistance in methylating microbes (Gilmour et al. 2011). Previous studies have suggested that $\text{MeHg}^+_{(\text{aq})}$ is more toxic than $\text{Hg}^{\text{II}}_{(\text{aq})}$ to some microbes (Jonas et al. 1984; Gilmour et al. 2011), and we hypothesize here that $\text{Hg}^{\text{II}}_{(\text{aq})}$ methylation may have evolved effectively

as an early antimicrobial against microorganisms without the ability to metabolize/detoxify this organometallic compound. We note that, in seawater, the higher proportion of lipophilic $\text{MeHg}^+_{(\text{aq})}$ (speciated as $\text{MeHgCl}_{(\text{aq})}$) over $\text{Hg}^{\text{II}}_{(\text{aq})}$ (as $\text{HgCl}_{2(\text{aq})}$) renders the former a more effective toxicant to microbial cells (Mason et al. 1995). Microorganisms have evolved a wide range of strategies to compete against other microbes for limited resources (Granato et al. 2019), and our hypothesis that $\text{MeHg}^+_{(\text{aq})}$ effectively originated as an antimicrobial mirrors a similar interpretation for *arsM* encoding for arsenic methylation (Li et al. 2021). As the *hgcAB* gene pair likely evolved in LUCA (fig. 1A), $\text{Hg}^{\text{II}}_{(\text{aq})}$ methylation could have stabilized to persist as an early form of antimicrobial production.

As a defense against $\text{MeHg}^+_{(\text{aq})}$, other microorganisms evolved alkylmercury lyase (MerB) encoded by *merB*, with

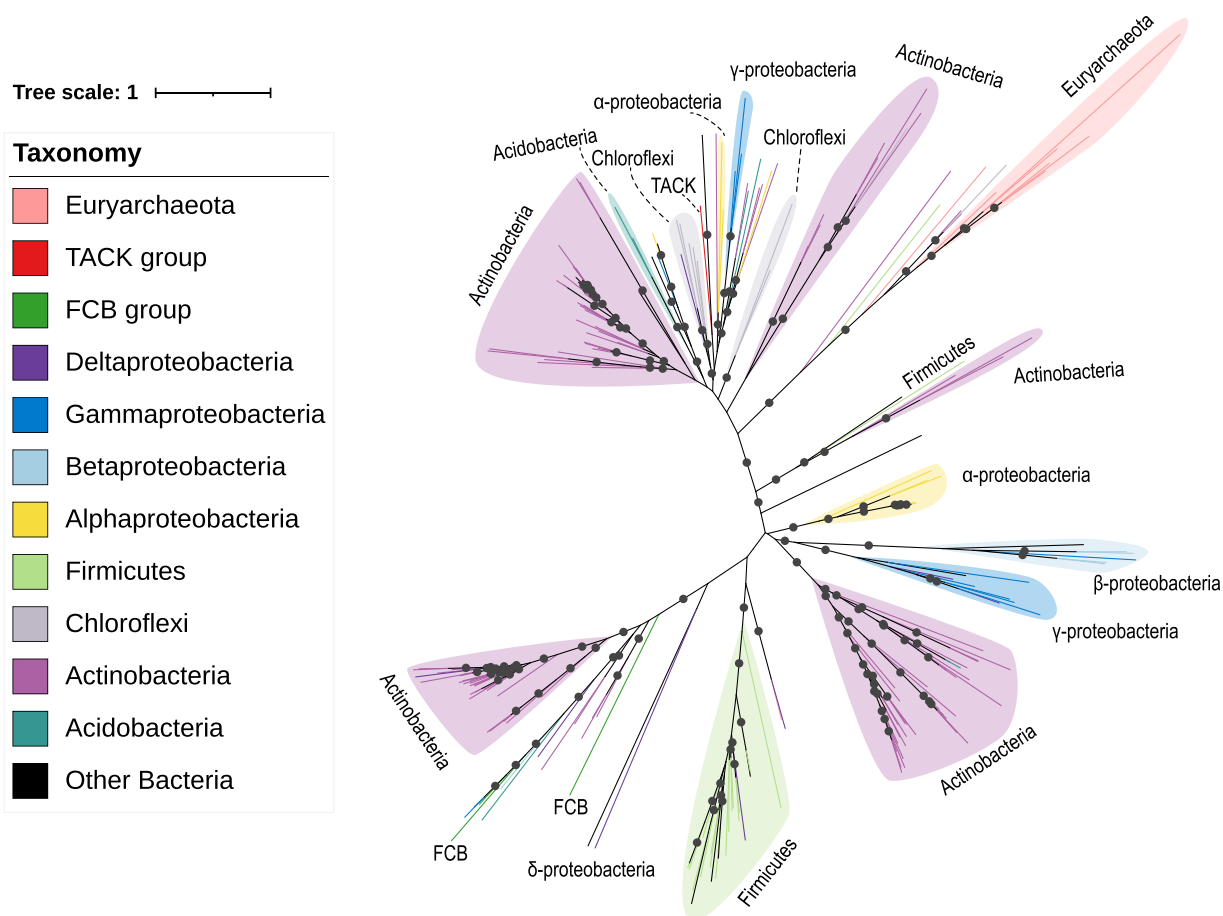


FIG. 4.—Phylogenetic tree of MerB proteins. The tree was inferred by using the ML method under LG + C40 + F + G model. This analysis involved alignment of 223 amino acid sequences with a total of 1,448 positions. Different taxonomies are represented by different colors. Ultrabootstrap support values were calculated with 1,000 replications, and ultrabootstrap values > 90% are shown by black dots at the nodes.

$\text{Hg}^{\text{II}}_{(\text{aq})}$ toxicity mitigated via other detoxification systems, for example, MerA (Boyd and Barkay 2012), iron-coupled redox reactions (Liu and Wiatrowski 2018), abiotic reduction (Gu et al. 2011), or other mechanisms (Christakis et al. 2021). Once *merB* evolved, however, *hgcAB* may no longer have offered a selective advantage and therefore underwent extensive loss during vertical evolution. Furthermore, we suggest that bacteria carrying other $\text{MeHg}^+_{(\text{aq})}$ detoxification mechanisms might not need to also carry *merB*. For example, MerB homologs are rarely found in sulfate-reducing bacteria (SRB), as SRB produce H_2S as a metabolite that reacts with $\text{MeHg}^+_{(\text{aq})}$ to form dimethylmercury (DMHg) which volatilizes efficiently from the cell (Baldi et al. 1993). Jonsson et al. (2016) also found DMHg production resulting from reaction of $\text{MeHg}^+_{(\text{aq})}$ with poorly crystalline iron sulfides; these phases are typically precipitated indirectly by the metabolism of SRB. Other studies have observed that $\text{MeHg}^+_{(\text{aq})}$ is somehow efficiently exported from methylating cells, thus reducing toxicity to the methylator (Graham et al. 2012; Lin et al. 2015). We

recognize, however, that the mechanisms underpinning $\text{MeHg}^+_{(\text{aq})}$ export have not yet been elucidated. Interestingly, extracellular thiol compounds, such as cysteine, were found to facilitate export and desorption of $\text{MeHg}^+_{(\text{aq})}$ (Schaefer and Morel 2009; Lin et al. 2015), possibly due to competitive binding of thiols to receptors on $\text{MeHg}^+_{(\text{aq})}$ transporters. To our knowledge, only two proteins encoded by *mer* operons, MerC (Sone et al. 2017) and MerE (Sone et al. 2013), have been reported as potential $\text{MeHg}^+_{(\text{aq})}$ transporters, but neither of them is carried by $\text{Hg}^{\text{II}}_{(\text{aq})}$ methylators.

Our study has revealed an ancient origin for microbial Hg methylation, evolving from LUCA to radiate extensively throughout the tree of life both vertically, albeit with extensive loss, and to a lesser extent horizontally. We hypothesize that early Hg-methylating microorganisms may have innovatively transformed a ubiquitous aqueous trace metal, $\text{Hg}^{\text{II}}_{(\text{aq})}$, into a highly effective antimicrobial compound. The dual ability of SRB to methylate Hg and excrete sulfide provides serendipitous protection against $\text{MeHg}^+_{(\text{aq})}$

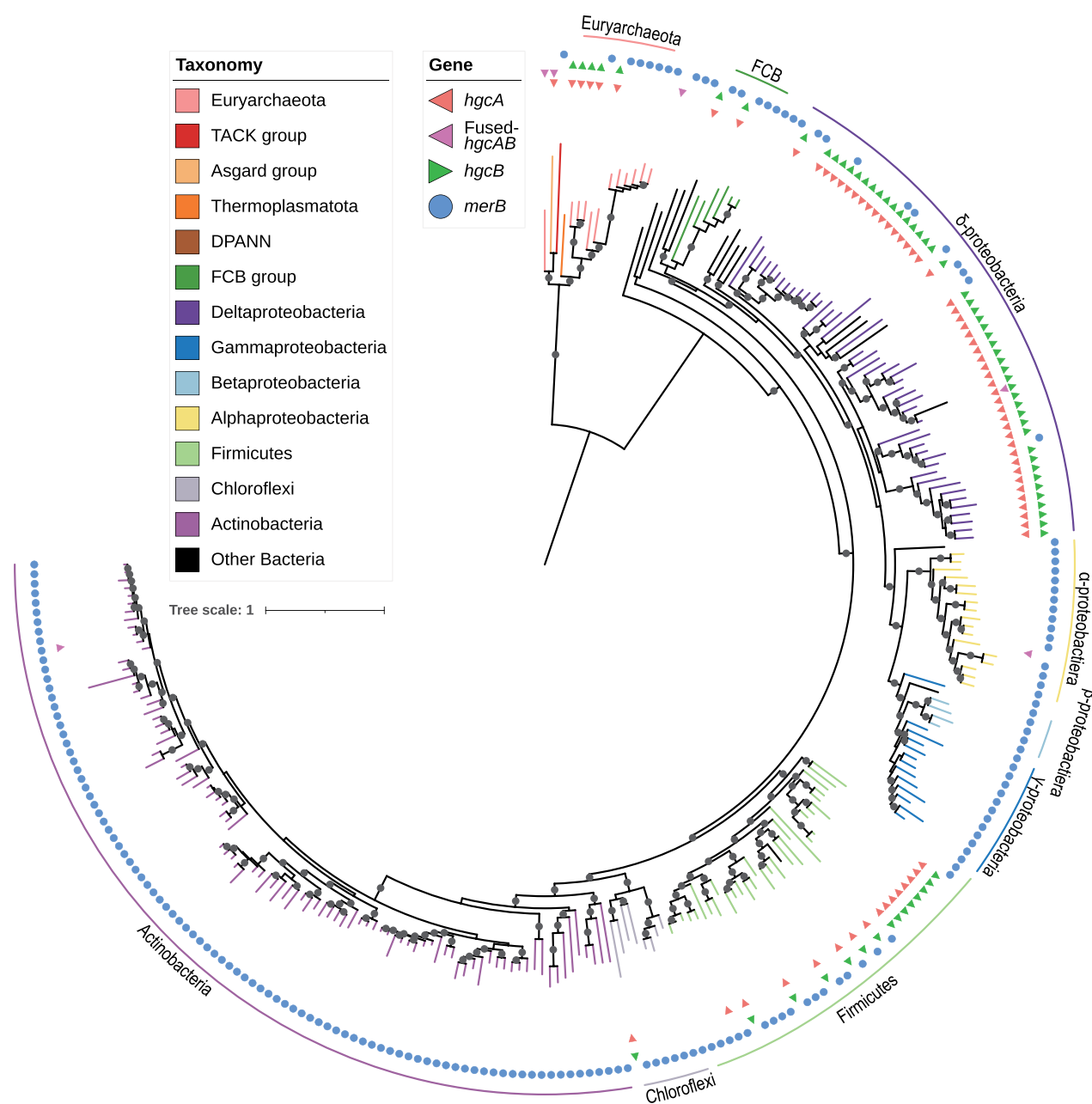


FIG. 5.—Species tree of *hgcA*+ and *merB*+ genomes in the RP35 data set. The tree was inferred from a concatenate of 27 marker genes derived from Moody et al. (2022) by using the ML method under LG + F + R10 model. This analysis involved alignment of 251 amino acid sequences with a total of 8,024 positions. Different taxonomies are represented by different colors. The presence of genes *hgcA*/*fused-hgcAB*/*hgcB*/*merB* in the genomes was represented by different symbols and colors in the outer circle. Ultrabootstrap support values were calculated with 1,000 replications, and ultrabootstrap values > 90% are shown by black dots at the nodes.

toxicity. Methylmercury degradation can also be abiotically photocatalyzed, so it is unlikely that shallow aqueous environments where ultraviolet (UV) light could penetrate would have allowed for toxic buildup of $\text{MeHg}^+_{(aq)}$. We speculate that anoxic sediments where volcanogenic Hg and either geogenic or bacteriogenic sulfide were copresent (such as seafloor sediments near hydrothermal vents) may have

provided an ideal biogeochemical setting for the evolution of microbial Hg methylation. Prior to the evolution of genomic countermeasures such as *merB*, $\text{MeHg}^+_{(aq)}$ concentrations in some natural environments could have increased to toxic levels beyond those typically found today, where they are constrained by both biotic and abiotic demethylation mechanisms. We conclude that microbial $\text{Hg}^{II}_{(aq)}$ methylation in

extant microorganisms represents the genomic vestige of an early geosphere–biosphere interaction with major ramifications for Earth’s mercury cycle.

Materials and Methods

Data Set Construction

UniProt Reference Proteomes v2022_03 (<https://proteininformationresource.org/rps/>) at 35% cutoff (RP35) and 75% cut-off (RP75) were used to establish two data sets. UniProt Reference Proteomes is a database containing representative proteomes selected from a large proteome database with more than 30,000 proteomes from UniProtKB (Chen et al. 2011). Every proteome in the database is comprised of all coding sequences translated from a whole-genome sequence. The RP35 data set contained 5,535 proteomes and was used to study the phylogeny of protein family PF03599 (CdhD, CdhE, and HgcA) and PF03243 (MerB). The RP75 data set contained 17,551 proteomes and was used to study the phylogeny of HgcA individually. In order to enlarge the diversity and sample size of HgcA sequences, two other data sets, including one containing 700 representative prokaryotic proteomes constructed by Moody et al. (2022) and another containing several novel *hgc* carriers published by Lin et al. (2021), were retrieved and incorporated into the RP75 data set, followed by removal of any redundancies. HgcAB proteins selected for this study including their metadata are listed in [supplementary table S1, Supplementary Material](#) online; other proteins including CdhD, CdhE, and MerB used in this study are listed in [supplementary table S2, Supplementary Material](#) online.

Determination of Genes of Interest

Protein sequences belonging to protein families PF03599 and PF03243 from both data sets were determined according to annotation by UniProt and also confirmed with hmmsearch v3.1 (Eddy 1998) against curated hidden Markov models provided by InterPro (<https://www.ebi.ac.uk/interpro/entry/pfam>). HgcA sequences were extracted from the PF03599 family using hmmsearch against the Hg-MATE database (Capo et al. 2023) and further determined by the conserved motif N(V/I)WC(A/S). CdhD and CdhE sequences were extracted from the PF03599 family using BlastP v2.11.0 against the experimentally validated CdhD (Q57577) and CdhE (Q57576) sequences from *Methanocaldococcus jannaschii* JAL-1, respectively. HgcB sequences were determined by searching for 4Fe4S proteins encoded by genes adjacent to *hgcA* genes (two open reading frames on either side of *hgcA*) in the genome. MerB sequences were further confirmed from the PF03243

protein family by hmmsearch against the MerB database constructed by Christakis et al. (2021).

Reconstruction of Phylogenetic Trees

Protein sequences belonging to the PF03599 protein family and included in the RP35 data set were aligned using MAFFT-linsi v7.453 (Katoh and Standley 2013), resulting in an alignment containing 478 sequences with 2,922 columns. The alignment was used to build a maximum likelihood (ML) tree using IQ-TREE v2.0.3 (Schmidt et al. 2014) under the best-fitting model of LG + C50 + F + R. Branch supports were estimated with 1,000 ultrafast bootstrap (Hoang et al. 2018) replicates (fig. 1) The unrooted tree was rooted using three methods: midpoint rooting (Swofford et al. 1996), minimum variance (MinVar, Mai et al. 2017), and MAD (Tria et al. 2017), respectively. Similar to the methods described above, all trees in this study were built using IQ-TREE v2.0.3 (Schmidt et al. 2014) with 1,000 ultrafast bootstrap (Hoang et al. 2018) replications. Protein sequences were aligned using MAFFT (L-INS-i) v7.453 (Katoh and Standley 2013).

To mitigate potential contamination of HgcB tail sequences in the HgcA alignment, the fused-HgcAB proteins were aligned separately and the “HgcB tail” (last 79 positions) was manually removed. The resulting “HgcA head” of the fused-HgcAB proteins was aligned with other HgcA sequences. Finally, aligned “HgcB tail” sequences were subsequently added back to the alignment. The first 327 positions of the alignment were poorly aligned, therefore manually removed. The final alignment contained 169 sequences with 493 columns and was used to infer a ML tree under the LG + C60 + F + G model (fig. 2). The tree was rerooted between the two major groups of HgcA sequences according to the above PF03599 protein family tree.

HgcB sequences and “HgcB tail” sequences from the fused-HgcAB proteins were aligned and used to infer a ML tree using a similar method as the PF03599 protein family tree described above under the LG + C60 + F + G model. Phylogenetic congruence between the HgcA and the HgcB sequences was inferred and visualized using cophylo implemented in the R package phytools v1.0-3 (Revell 2012) based on the topology of the HgcA and the HgcB ML trees described above. Nodes of both HgcA and HgcB trees were allowed to be rotated by the program to optimize vertical matching of tips (fig. 3).

In total, 223 MerB sequences were predicted from proteomes in the RP35 data set. These sequences were aligned and used to build a ML tree under the best-fit protein model LG + C40 + F + G chosen according to Bayesian Information Criterion (BIC) (fig. 4).

A species tree of *hgc+* and *merB+* proteomes (251 proteomes in total) from the RP35 data set was reconstructed

based on the 27 marker genes proposed by Moody et al. (2022). Twenty-seven HMM profiles were created individually based on marker gene alignments and concatenated using HMMER v3.1. The 27 marker homologs in the 251 proteomes were extracted by hmmsearch against the concatenated hmm profile with an *E* value of $1e-10$. The 27 marker homologs were then aligned respectively and concatenated. Poorly aligned regions were trimmed using trimAl v1.2 (Silla-Martínez et al. 2009) with parameters “-resoverlap 0.55 -seqoverlap 60 -automated1.” The resulting supermatrix (8,024 sites) was used to infer a ML tree under the model of LG + F + R10 (fig. 5). All phylogenetic trees described above were visualized using iTOL v6 (Letunic and Bork 2021) and refined with Adobe Illustrator (Adobe Systems Inc., San Jose, CA, USA).

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

Acknowledgments

We thank Yao-ban Chan and Qiuyi Li at The University of Melbourne for valuable discussions and advice in the early stage of this study. We also acknowledge helpful discussions with Caitlin Gionfriddo, Ben Peterson, and Eric Capo. This publication was made possible in part through support from a Strategic Australian Postgraduate Award (PhD scholarship) to H.L. from the Environmental Microbiology Research Initiative (EMRI) at The University of Melbourne (J.W.M.) and laboratory start-up funding from the University of Glasgow to J.W.M., the Moore Foundation (<https://doi.org/10.37807/GBMF9741> to T.A.W.), the Royal Society (URFR\201024) to T.A.W., and the John Templeton Foundation (62220 to T.A.W. and E.R.R.M.). The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of the John Templeton Foundation.

Data Availability

All data generated in this study including amino acid alignments and phylogenetic trees are deposited in Figshare: <https://doi.org/10.6084/m9.figshare.21428523>.

Literature Cited

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*. 115:E1166–E1173.
 Amyot M, Morel FM, Ariya PA. 2005. Dark oxidation of dissolved and liquid elemental mercury in aquatic environments. *Environ Sci Technol*. 39(1):110–114.

Baldi F, Pepi M, Filippelli M. 1993. Methylmercury resistance in *Desulfovibrio desulfuricans* strains in relation to methylmercury degradation. *Appl Environ Microbiol*. 59(8):2479–2485.
 Barkay T, Gu B. 2022. Demethylation—the other side of the mercury methylation coin: a critical review. *ACS Environ Au*. 2:77–97.
 Barkay T, Kritee K, Boyd E, Geesey G. 2010. A thermophilic bacterial origin and subsequent constraints by redox, light and salinity on the evolution of the microbial mercuric reductase. *Environ Microbiol*. 12:2904–2917.
 Barkay T, Miller SM, Summers AO. 2003. Bacterial mercury resistance from atoms to ecosystems. *FEMS Microbiol Rev*. 27:355–384.
 Barkay T, Wagner-Döbler I. 2005. Microbial transformations of mercury: potentials, challenges, and achievements in controlling mercury toxicity in the environment. In: *Advances in applied microbiology*. Volume 57. pp. 1–52.
 Bobrowski N, Hönninger G, Galle B, Platt U. 2003. Detection of bromine monoxide in a volcanic plume. *Nature* 423(6937):273–276.
 Bolotin E, Hershberg R. 2016. Bacterial intra-species gene loss occurs in a largely clocklike manner mostly within a pool of less conserved and constrained genes. *Sci Rep*. 6:35168.
 Boyd ES, Barkay T. 2012. The mercury resistance operon: from an origin in a geothermal environment to an efficient detoxification machine. *Front Microbiol*. 3:349.
 Capo E, et al. 2023. A consensus protocol for the recovery of mercury methylation genes from metagenomes. *Mol Ecol Resour*. 23(1):190–204.
 Chen C, et al. 2011. Representative proteomes: a stable, scalable and unbiased proteome set for sequence analysis and functional annotation. *PLoS One* 6:e18910.
 Choi S-C, Chase T, Bartha R. 1994. Metabolic pathways leading to mercury methylation in *Desulfovibrio desulfuricans* LS. *Appl Environ Microbiol*. 60:4072–4077.
 Christakis CA, Barkay T, Boyd ES. 2021. Expanded diversity and phylogeny of mer genes broadens mercury resistance paradigms and reveals an origin for mera among thermophilic archaea. *Front Microbiol*. 12:682605.
 Colombo MJ, Ha J, Reinfelder JR, Barkay T, Yee N. 2013. Anaerobic oxidation of Hg(0) and methylmercury formation by *Desulfovibrio desulfuricans* ND132. *Geochim Cosmochim Acta*. 112:166–177.
 Date SS, et al. 2019. Kinetics of enzymatic mercury methylation at nanomolar concentrations catalyzed by HgcAB. *Appl Environ Microbiol*. 85(13):e00438-19.
 Eddy SR. 1998. Profile hidden Markov models. *Bioinformatics* 14:755–763.
 Ekstrom EB, Morel FMM, Benoit JM. 2003. Mercury methylation independent of the acetyl-coenzyme A pathway in sulfate-reducing bacteria. *Appl Environ Microbiol*. 69:5414–5422.
 Gilmour CC, et al. 2011. The sulfate-reducing bacterium *Desulfovibrio desulfuricans* ND132 as a model for understanding bacterial mercury methylation. *Appl Environ Microbiol*. 77:3938–3951.
 Gilmour CC, et al. 2013. Mercury methylation by novel microorganisms from new environments. *Env Sci Technol*. 47:11810–11820.
 Gilmour CC, Bullock AL, McBurney A, Podar M, Elias DA. 2018. Robust mercury methylation across diverse methanogenic *Archaea*. *MBio* 9(2):e02403-17.
 Gionfriddo CM, et al. 2016. Microbial mercury methylation in Antarctic sea ice. *Nat Microbiol*. 1:16127.
 Graham AM, et al. 2012. Detailed assessment of the kinetics of Hg-cell association, Hg methylation, and methylmercury degradation in several *Desulfovibrio* species. *Appl Environ Microbiol*. 78:7337–7346.
 Granato ET, Meiller-Legrand TA, Foster KR. 2019. The evolution and ecology of bacterial warfare. *Curr Biol*. 29:R521–R537.

- Grasby SE, Them TR, Chen Z, Yin R, Ardakani OH. 2019. Mercury as a proxy for volcanic emissions in the geologic record. *Earth-Sci Rev.* 196:102880.
- Gu B, et al. 2011. Mercury reduction and complexation by natural organic matter in anoxic environments. *Proc Natl Acad Sci.* 108:1479–1483.
- Guo M, Korenaga J. 2021. A halogen budget of the bulk silicate Earth points to a history of early halogen degassing followed by net re-gassing. *Proc Natl Acad Sci U S A.* 118(51):e2116083118.
- Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFB002: improving the ultrafast bootstrap approximation. *Mol Biol Evol.* 35:518–522.
- Hsu-Kim H, Kucharzyk KH, Zhang T, Deshusses MA. 2013. Mechanisms regulating mercury bioavailability for methylating microorganisms in the aquatic environment: a critical review. *Env Sci Technol.* 47:2441–2456.
- Jonas RB, Gilmour CC, Stoner DL, Weir MM, Tuttle JH. 1984. Comparison of methods to measure acute metal and organometal toxicity to natural aquatic microbial communities. *Appl Environ Microbiol.* 47:1005–1011.
- Jonsson S, Mazrui NM, Mason RP. 2016. Dimethylmercury formation mediated by inorganic and organic reduced Sulfur surfaces. *Sci Rep.* 6:27958.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30:772–780.
- Kidd KA, et al. 2012. Biomagnification of mercury through lake trout (*Salvelinus namaycush*) food webs of lakes with different physical, chemical and biological characteristics. *Sci Total Environ.* 438:135–143.
- Koskiniemi S, Sun S, Berg OG, Andersson DI. 2012. Selection-driven gene loss in bacteria. *PLoS Genet.* 8:e1002787.
- Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49:W293–W296.
- Li YP, et al. 2021. Antimicrobial activity of metals and metalloids. *Annu Rev Microbiol.* 75:175–197.
- Lin H, et al. 2021. Mercury methylation by metabolically versatile and cosmopolitan marine bacteria. *ISME J.* 15:1810–1825.
- Lin H, Lu X, Liang L, Gu B. 2015. Thiol-facilitated cell export and desorption of methylmercury by anaerobic bacteria. *Environ Sci Technol Lett.* 2:292–296.
- Liu S, Wiatrowski HA. 2018. Reduction of Hg(II) to Hg(0) by biogenic magnetite from two magnetotactic bacteria. *Geomicrobiol J.* 35:198–208.
- Lu X, et al. 2016. Anaerobic mercury methylation and demethylation by *Geobacter bemidjensis* Bem. *Env Sci Technol.* 50:4366–4373.
- Mai U, Sayyari E, Mirarab S. 2017. Minimum variance rooting of phylogenetic trees and implications for species tree reconstruction. *PLoS One* 12:e0182238.
- Mason RP, Reinfelder JR, Morel FM. 1995. Bioaccumulation of mercury and methylmercury. *Water Air Soil Pollut.* 80:915–921.
- McDaniel EA, et al. 2020. Expanded phylogenetic diversity and metabolic flexibility of mercury-methylating microorganisms. *mSystems* 5(4):e00299–20.
- Meixnerová J, et al. 2021. Mercury abundance and isotopic composition indicate subaerial volcanism prior to the end-Archean “whiff” of oxygen. *Proc Natl Acad Sci U S A.* 118(33):e2107511118.
- Moody ER, et al. 2022. An estimate of the deepest branches of the tree of life from ancient vertically evolving genes. Perry, GH & Koonin, EV, editors. *eLife* 11:e66695.
- Obrist D, et al. 2011. Bromine-induced oxidation of mercury in the mid-latitude atmosphere. *Nat Geosci.* 4(1):22–26.
- Parks JM, et al. 2013. The genetic basis for bacterial mercury methylation. *Science* 339:1332–1335.
- Podar M, et al. 2015. Global prevalence and distribution of genes and microorganisms involved in mercury methylation. *Sci Adv.* 1:e1500675.
- Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol.* 3:217–223.
- Robinson JB, Tuovinen O. 1984. Mechanisms of microbial resistance and detoxification of mercury and organomercury compounds: physiological, biochemical, and genetic analyses. *Microbiol Rev.* 48(2):95–124.
- Schaefer JK, Morel FM. 2009. High methylation rates of mercury bound to cysteine by *Geobacter sulfurreducens*. *Nat Geosci.* 2(2):123–126.
- Schmidt HA, Minh BQ, von Haeseler A, Nguyen L-T. 2014. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32:268–274.
- Seigneur C, Lohman K. 2008. Effect of bromine chemistry on the atmospheric mercury cycle. *J Geophys Res Atmospheres.* 113(D23).
- Siciliano SD, Lean DRS. 2002. Methyltransferase: an enzyme assay for microbial methylmercury formation in acidic soils and sediments. *Environ Toxicol Chem.* 21:1184–1190.
- Silla-Martínez JM, Capella-Gutiérrez S, Gabaldón T. 2009. Trimal: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973.
- Sone Y, et al. 2013. Increase methylmercury accumulation in *Arabidopsis thaliana* expressing bacterial broad-spectrum mercury transporter MerE. *AMB Express* 3:52.
- Sone Y, et al. 2017. A novel role of MerC in methylmercury transport and phytoremediation of methylmercury contamination. *Biol Pharm Bull.* 40:1125–1128.
- Sousa FL, Martin WF. 2014. Biochemical fossils of the ancient transition from geoenergetics to bioenergetics in prokaryotic one carbon compound metabolism. *Biochim Biophys Acta BBA—Bioenerg.* 1837:964–981.
- Svetlitchnaia T, Svetlitchnyi V, Meyer O, Dobbek H. 2006. Structural insights into methyltransfer reactions of a corrinoid iron–sulfur protein involved in acetyl-CoA synthesis. *Proc Natl Acad Sci.* 103:14331–14336.
- Swofford DL, Olsen GJ, Waddell PJ, Hillis DM. 1996. Phylogenetic inference. In: *Molecular systematics*. Sunderland, Massachusetts: Sinauer Associates, Inc.. Vol. 15. p. 407–514.
- Tada Y, Marumoto K, Takeuchi A. 2020. *Nitrospina*-like bacteria are potential mercury methylators in the mesopelagic zone in the East China Sea. *Front Microbiol.* 11:1369.
- Tada Y, Marumoto K, Takeuchi A, Gralnick JA. 2021. *Nitrospina*-like bacteria are dominant potential mercury methylators in both the Oyashio and Kuroshio Regions of the Western North Pacific. *Microbiol Spectr.* 9:e00833–21.
- Tebo BM, et al. 2004. Biogenic manganese oxides: properties and mechanisms of formation. *Annu Rev Earth Planet Sci.* 32:287–328.
- Tezuka T, Tonomura K. 1976. Purification and properties of an enzyme catalyzing the splitting of carbon-mercury linkages from mercury-resistant *Pseudomonas* K-62 strain: 1. Splitting enzyme 1. *The Journal of Biochemistry* 80(1):79–87.
- Trevors JT. 1986. Mercury methylation by bacteria. *J Basic Microbiol.* 26:499–504.
- Tria FDK, Landan G, Dagan T. 2017. Phylogenetic rooting using minimal ancestor deviation. *Nat Ecol Evol.* 1:1–7.
- Villar E, Cabrol L, Heimbürger-Boavida LE. 2020. Widespread microbial mercury methylation genes in the global ocean. *Env Microbiol Rep.* 12(3):277–287.

- Wang P, et al. 2014. Catalytic oxidation of Hg⁰ by MnOx–CeO₂/γ-Al₂O₃ catalyst at low temperatures. *Chemosphere* 101: 49–54.
- Wang J, Dai J, Chen G, Jiang F. 2022. Role of sulfur biogeochemical cycle in mercury methylation in estuarine sediments: a review. *J Hazard Mater.* 423:126964.
- Weiss MC, et al. 2016. The physiology and habitat of the last universal common ancestor. *Nat Microbiol.* 1:1–8.
- Weiss MC, Preiner M, Xavier JC, Zimorski V, Martin WF. 2018. The last universal common ancestor between ancient Earth chemistry and the onset of genetics. *PLoS Genet.* 14:e1007518.
- Zerle AL et al. 2021. Sulfur and mercury MIF suggest volcanic contributions to Earth's atmosphere at 2.7 ga. *Geochem Persp Lett.* 18: 48–52.

Associate editor: Brian Golding