Ancient origin and evolution of microbial mercury methylation

Heyu Lin[1], Edmund R. R. Moody[2], Tom A. Williams[3] and John W. Moreau[1, 4, *]

[1]School of Geographical, Atmospheric and Earth Sciences, The University of Melbourne, Parkville, Victoria 3010, Australia

[2]School of Earth Sciences, University of Bristol, Bristol, BS8 1RJ, United Kingdom

[3]School of Biological Sciences, University of Bristol, Bristol, BS8 1TQ, United Kingdom

[4]School of Geographical and Earth Sciences, University of Glasgow, Glasgow, G12 8RZ, United Kingdom

KEYWORDS: mercury, methylmercury, evolution, antimicrobial, *hgc* gene, gene loss, horizontal gene transfer, LUCA

*Author for Correspondence: John W. Moreau, School of Geographical and Earth Sciences, University of Glasgow, Glasgow, United Kingdom, +44 141 330 5461, john.moreau@glasgow.ac.uk

Abstract

The origin and evolution of microbial mercury methylation have long remained a mystery. Here we employed genome-resolved phylogenetic analyses to decipher the evolution of the mercury methylating genes *hgcAB*, to constrain the ancestral origin of the *hgc* operon, and to explain its current distribution across the bacterial and archaeal domains. We infer the extent to which vertical descent and horizontal gene transfer have influenced the evolution of this operon and hypothesize that the original function of methylmercury was as an antimicrobial compound on the early Earth. We speculate that subsequent evolution of the detoxifying agent alkylmercury lyase (*merB)* reduced the selective advantage of the mercury methylation activity, resulting in widespread loss of the *hgcAB* genes among archaea and bacteria.

Significance

Neurotoxic methylmercury ($MeHg^{+}$) is synthesized from Hg^{II} in the environment by microorganisms, involving the gene pair *hgcA* and *hgcB*. However, the origin and evolution of these two genes are poorly understood. Our phylogenetic analyses revealed the evolutionary history of the *hgc* gene pair and uncovered a link between the evolution of Hg methylation and demethylation. We show that the *hgc* gene has undergone generally vertical evolution with extensive parallel gene loss, and we propose that mercury methylation evolved as an antimicrobial production mechanism during competition for limited energy resources on the early Earth.

Introduction

Microorganisms play key roles in the global mercury cycle. In soils, sediments and natural waters, microbes mediate the enzymatic reductive volatilization of Hg^{II} to Hg^{0}, the transformation of aqueous Hg^{II} to $MeHg^{+}$, the complexation of aqueous Hg^{II} by bacteriogenic bisulfide, and the adsorption/desorption of Hg^{II} on/from biogenic metal ox(yhydrox)ides (e.g., Barkay et al. 2003; Barkay & Gu 2022; Barkay & Wagner□Döbler 2005; Gilmour et al. 2013; Hsu-Kim et al. 2013; Tebo et al. 2004; Wang et al. 2022). Essentially, microbes control or strongly influence the geochemical speciation of mercury in most natural environments today.

Both industrial and geological sources contribute mercury to the environment (Liu et al. 2011). Industrial mercury sources include fossil fuel combustion, mining, cement production and waste incineration (AMAP Assessment 2011); natural sources primarily involve geothermal activity. In fact, volcanos and hot springs have emitted mercury throughout geologic time since Earth's earliest geologic history (Grasby et al. 2019; Zerkle et al. 2020, 2021). Following the widespread oxygenation of the early ~2.4 billion years ago, Hg^{II} would therefore be commonly found as a trace metal in hydrothermal environments. Thermophiles, as the Earth's earliest microorganisms, would presumably have needed some defense mechanism against the toxicity of Hg^{II}, i.e., to cope with its affinity for protein sulfhydryl groups and the resulting destabilization of enzymes. Previous studies present phylogenetic support for a thermophilic bacterial origin of the *mer* operon, acquired later in Archaea through multiple independent transfer events reflecting the environmental ubiquity of aqueous mercury (Barkay et al. 2010; Boyd & Barkay 2012).

Microbial mercury methylation has received much attention because $MeHg^{+}$ is a potent neurotoxin that bioaccumulates across both marine and terrestrial food webs (Kidd et al. 2012). However, the origin of this process has remained a mystery, largely because $MeHg^{+}$

serves no known biological function. Mercury methylation is encoded by the *hgc* operon, most closely related to the ancient carbon monoxide dehydrogenase/acetyl-CoA synthase gene, *cdh*, which has previously been suggested as being present in the Last Universal Common Ancestor (LUCA) (Adam et al. 2018; Parks et al. 2013; Sousa & Martin 2014; Weiss et al. 2018). CdhD and CdhE (hereafter "CdhDE") are cobalamin-dependent proteins 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). While HgcA is also predicted to be a cobalamin-dependent methyltransferase, it only functions in Hg^{II} methylation and does not take part in the WL pathway, as other components of this pathway are incomplete in characterised *hgcA*-carrying genomes (Date et al. 2019). Similar to *cdh*, the *hgc* gene seems to be broadly distributed across the bacterial and archaeal domains, representing a range of environmental habitats across redox gradients (McDaniel et al. 2020; Capo et al. 2022). Some of these bacterial and archaeal phyla have been experimentally confirmed to methylate $Hg^{II}_{(aq)}$, while others are still being considered as putative methylators encoding homologous *hgcAB* sequences that have not yet been experimentally validated. In lieu of cultivated isolates, computational modelling has been employed to test for consistency of putative HgcAB sequences with the HgcAB structure and functionality of confirmed methylators (Gionfriddo et al. 2016; Lin et al. 2021).

Many studies of mercury 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 co-A formation (Choi et al. 1994) or methionine synthesis (Siciliano & 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 mercury methylating genes, employing genome-resolved

phylogenetic analyses to constrain the ancestral origin of the *hgc* operon and explain its currently known distribution across the bacterial and archaeal domains. We assess the extent to which vertical descent and horizontal gene transfer have shaped the evolution of the *hgc* operon and offer a novel hypothesis for the functional origin of microbial mercury methylation.

Results and Discussion

hgc genes originated prior to LUCA

A dataset 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 Table S1 & S2**). Phylogenetic reconstruction of the protein family PF03599 defined three deep-branching clusters, comprising HgcA, CdhD, and CdhE (**Figure 1A**, for details see **Figure S1**). On the hypothesis that CdhD- and CdhE-encoding genes were already present in LUCA (Adam et al. 2018; Sousa & Martin 2014; Weiss et al. 2018), then HgcA also branched prior to LUCA (that is, the divergence between these three protein subfamilies must pre-date LUCA; **Figure 1A**). While robustly rooting single gene trees is challenging, midpoint, Minvar (Mai et al. 2017) and MAD (Tria et al. 2017) rooting approaches all suggest that the root of this protein family lies between CdhD and CdhE+HgcA. In comparison to CdhD and CdhE, which have a relatively broad distribution across modern prokaryotes and broadly congruent evolutionary histories, *hgcA* genes have a more restricted distribution in extant taxa. They are mainly found in Bacteria, although with a more restricted distribution than in Archaea (**Figure 1A**), mostly affiliated with Deltaproteobacteria, Firmicutes, and FCB group (**Figure 1A**). Based on the rooted tree (**Figure 1B**), HgcA has a longer stem than CdhD and CdhE, which might be the

result of accelerated evolution (for example, associated with a change in function) or as a result of gene loss (or lineage extinction) in former HgcA-encoding clades.

hgcA evolved primarily vertically with extensive loss

To investigate the evolutionary history of HgcA, we further enlarged the sample size by retrieving HgcA homologs in UniProt Reference Proteomes database v2022_03 at 75% cutoff (RP75). Two other datasets, 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 dataset. Totally 169 HgcA sequences were collected after removing redundancies (**Supplementary Table S1**). Our analysis suggests that *hgcA* genes were separated into two well-supported clades with high bootstrap support, as shown in **Figure 1A** and **Figure 2** (for details see **Figure S2**). These two clades correspond to a difference in gene structure: the smaller clade comprises genes in which HgcA and HgcB became fused to form a HgcAB gene, while the larger contains single domain HgcA genes. Several Asgard archaea, including Thorarchaeota and Lokiarchaeota, form a clade in the HgcAB subtree, suggesting that the gene fusion occurred prior to the radiation of Asgard phyla. The fused-HgcAB from *Pyrococcus furiosus* has been experimentally tested and shown not to methylate Hg (Podar et al. 2015), but the functionality of other fused-HgcAB proteins is yet to be confirmed. Except for those fused-HgcAB carriers, the functionality of other members in this clade is also unknown. 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; Villar et al. 2020; Tada et al. 2021). 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 unexplored.

By contrast, the majority of (non-fused) HgcA sequences form a clade which broadly follows the universal tree (**Figure 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 is consistent with a pre-LUCA origin for the HgcA gene. However, the taxonomic distribution of HgcA in modern 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, we infer that it must have been lost subsequently in the other bacterial lineages.

Gene loss is a major force driving microbial genome evolution (Bolotin & Hershberg 2016), reflecting environmental change and/or adaptation (Koskiniemi et al. 2012). So far, *hgcA*+ Hg methylators have only been reported for anoxic or suboxic environments (Capo et al. 2022; Lin et al. 2021; Parks et al. 2013). During the evolution of Hg methylators, changes in reduction-oxidation potential might have inhibited Hg methylation and constituted an environmental selective pressure that facilitated the loss of *hgcA* genes. In addition, the long branches of the HgcA subtree in comparison to CdhD/E also suggest that they were subject to a weaker natural selection for gene maintenance, and were more likely to be lost in genetic drift, a common scenario observed for many microorganisms (Bolotin & Hershberg 2016).

A few putative HGT events could be inferred from the larger clade of the HgcA tree e.g., Marinimicrobia-HgcA clustered with Euryarchaeota-HgcA in the archaeal cluster, suggesting possible acquisition of *hgcA* via HGT. However, HGT does not appear to be the main driver for *hgcA* evolution, since only a few transfers were inferred.

hgcA and *hgcB* co-evolved

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 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 gene sequence length (95 amino acids on average), the overall topologies of the *hgcB* tree and the *hgcA* tree were congruent, supporting this inference (**Figure 3**, for details see **Figure S3**). Nevertheless, several *hgcA*+ genomes did not carry neighbouring *hgcB* genes, including all *Nitrospina* and a few Deltaproteobactiera and Firmicutes, potentially because of gene loss during evolution or incomplete transfer events (i.e., only *hgcA* genes were acquired during the HGT events). Another exception is that a complete *hgcB* gene is 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' in the fused-HgcAB from the same host, suggesting the possible acquisition of this gene from closely related microorganisms through HGT.

Phylogenetic distribution of *merB* and its relationship to *hgc*

The *merB* gene encoding organomercury lyase catalyzes the protonolysis of the C-Hg bond in $MeHg^{+}$ and releases Hg^{II} (Boyd & Barkay 2012), an opposite process to the reaction catalyzed by HgcAB proteins. Totally 225 MerB sequences were retrieved from the RP35 dataset (**Supplementary Table S2**). MerB homologs were distributed across both the bacterial and archaeal domains (**Figure 4**, for details see **Figure S4**), 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 Euryarchaeota and one thaumarchaeon, were found to encode MerB. These archaeal MerB homologs fall into bacterial clades, and are not monophyletic. This suggests a bacterial origin of MerB homologs, with transfers into the archaeal domain occurring subsequently. This interpretation aligns with the previous hypothesis that merB was recruited from a mesophilic ancestor recently instead of from LUCA (Christakis et al. 2021).

We mapped the presence/absence of *merB* and *hgc* genes onto the Tree of Life (**Figure 5**, for details see **Figure S5**) to illustrate their phylogenetic relationship. Intriguingly, although the distribution of MerB homologs overlapped with HgcAB homologs, few genomes encode both MerB and HgcAB. Taken together with their opposing activities, this complementary distribution pattern, in which genomes tend to encode either HgcAB or MerB, but not both, suggests a functional conflict between these two genes (i.e., there is utility in encoding for one or the other activity, but not both). Nevertheless, a few genomes encoding both MerB and HgcAB were observed, in which the capability of both Hg methylation and demethylation by *Citrifermentans bemidjiense* Bem was addressed before (Lu et al. 2016), suggesting a potential for both Hg methylation and demethylation among some anaerobic bacteria. The phylogeny and phylogenetic distribution of these enzymes (**Figure 4 & Figure 5**) suggests that HgcAB was the earlier of the two activities to evolve, with MerB originating more recently and spreading by horizontal gene transfer.

Mercury methylation as antimicrobial synthesis by early Earth microbes?

The reason(s) why microorganisms evolved to methylate Hg^{II} to $MeHg^{+}$ has long been a mystery. Previous studies have proposed that the process conferred resistance to Hg^{II} toxicity (Trevors 1986). In fact, the ability to produce $MeHg^{+}$ has been shown *not* to confer

214 Hg^{II} resistance (Gilmour et al. 2011). Considering $MeHg^{+}$ is the more toxic and stable form of
215 Hg^{II} compared to inorganic Hg^{II} (Jonas et al. 1984; Gilmour et al. 2011; IARC Working
216 Group on the Evaluation of Carcinogenic Risks to Humans 1993), we hypothesise here that
217 $MeHg^{+}$ could have been employed by early Hg^{II} methylators as an antimicrobial against
218 microorganisms without the ability to metabolize/detoxify this organometallic compound. We
219 note also that the greater lipophilicity of $MeHg^{+}$ over Hg^{2+} would have allowed the former to
220 penetrate cell walls more readily. As the *hgcAB* gene pair likely evolved in LUCA (**Figure**
221 **1A),** Hg^{II} methylation could have stabilised to persist as an early form of antimicrobial
222 production. Microorganisms have evolved a wide range of attack mechanisms to compete
223 against other microbes for limited resources (Granato et al. 2019), and our hypothesis for
224 mercury methylation as antimicrobial production mirrors a similar hypothesis recently put
225 forward for *arsM*, a gene that encodes for arsenic methylation (Li et al. 2021).

226 As a response, other microorganisms may have evolved alkylmercury lyase encoded
227 by *merB* as a defense against $MeHg^{+}$, while Hg^{2+} toxicity could be mitigated via other
228 detoxification systems, e.g., MerA (Boyd & Barkay 2012), iron-coupled redox reactions (Liu
229 & Wiatrowski 2018), abiotic reduction (Gu et al. 2011), or other mechanisms (Christakis et al.
230 2021). Once *merB* evolved, however, *hgcAB* may no longer have offered a selective
231 advantage, and therefore underwent extensive loss during vertical evolution, as inferred by
232 our study. Furthermore, we speculate that bacteria carrying other $MeHg^{+}$ detoxification
233 mechanisms might not need to co-carry *merB* genes; this would explain why MerB homologs
234 are rarely found in sulfate-reducing bacteria (SRB), for example, as SRB produce aqueous
235 bisulfide (HS^{-}) that can transform $MeHg^{+}$ to volatile dimethylmercury (DMHg) which is then
236 removed quickly from the cell (Jonsson et al. 2016). Other studies have observed that $MeHg^{+}$
237 can be rapidly exported from methylating cells after production, and therefore might not
238 accumulate within methylators (Graham et al. 2012; Lin et al. 2015). However, the

mechanisms underpinning $MeHg^+$ export are still poorly understood. Interestingly, extracellular thiol compounds, such as cysteine, were found to facilitate export and desorption of $MeHg^+$ (Lin et al. 2015), possibly due to competitive binding of thiols to receptors on $MeHg^+$ transporters. To the best of 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 $MeHg^+$ transporters, but neither of them is carried by Hg methylators. Therefore, potential cellular transporters for $MeHg^+$ in Hg methylators should be investigated in future studies.

Our study reveals an ancient origin for microbial mercury 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 hypothesise that early mercury methylating microorganisms may have innovatively transformed a ubiquitous aqueous trace metal, Hg^{II} sourced originally from geothermal activity, into a highly effective antimicrobial compound with the ability to enter cells and deactivate enzymes. Prior to the evolution of genomic countermeasures such as *merB*, $MeHg^+$ concentrations in aqueous environments could have increased to toxic levels beyond those typically found in natural environments today, where they are typically constrained by demethylation. We speculate therefore that microbial Hg methylation today represents a vestige of early geosphere-biosphere and microbe-microbe interactions with profound consequences for Earth's biogeochemical mercury cycle.

Materials and Methods

Dataset construction

UniProt Reference Proteomes v2022_03 (Chen et al. 2011) at 35% cutoff (RP35) and 75% cut-off (RP75) were used to establish two datasets. The RP35 dataset contained 5,535 proteomes and was used to study the phylogeny of protein family PF03599 (CdhD, CdhE and

HgcA) and PF03243 (MerB). The RP75 dataset 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 datasets, 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 dataset, followed by removal of any redundancies. HgcAB proteins selected for this study including their metadata are listed in Supplementary Table S1; Other proteins including CdhD, CdhE, and MerB used in this study are listed in Supplementary Table S2.

Determination of genes of interest

Protein sequences belonging to protein families PF03599 and PF03243 from both datasets were determined according to annotation by UniProt and also confirmed with hmmsearch v3.1 (Eddy 1998) against curated HMM 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. 2022) 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 dataset were aligned using MAFFT-linsi v7.453 (Katoh & Standley 2013), resulting in an alignment containing 478 sequences with 2922 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 1000 ultrafast bootstrap (Hoang et al. 2018) replicates (**Figure 1**) The unrooted tree was rooted using three methods: Midpoint-rooting (Swofford et al. 1996), Minimum Variance (MinVar, Mai et al. 2017), and Minimal Ancestor Deviation (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 1000 ultrafast bootstrap (Hoang et al. 2018) replications. Protein sequences were aligned using MAFFT (L-INS-i) v7.453 (Katoh & 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 (**Figure 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 (**Figure 3**).

MerB sequences from the RP35 database were aligned and used to build a ML tree under the best-fit protein model LG+C40+F+G chosen according to BIC. Two MerA proteins (P08332.1 from *Shigella flexneri* and P16171.1 from *Bacillus cereus*) were used as an outgroup to root the tree **(Figure 4)**.

A species tree of *hgc*+ and *merB*+ proteomes (251 proteomes in total) from the RP35 database was reconstructed based on the 27 marker genes proposed by Moody et al. (2022). 27 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 (8024 sites) was used to infer a ML tree under the model of LG+F+R10 (**Figure 5**). All phylogenetic trees described above were visualized using iTOL v6 (Letunic & Bork 2021) and refined with Adobe Illustrator (Adobe Systems Inc., San Jose, CA, USA).

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

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. A special thank goes to the members of Mersorcium network (https://mersorcium.github.io/) for extensive discussions. This publication was made possible in part through support from 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 (URF\R\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.

References

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. doi: 10.1073/pnas.1716667115.

AMAP Assessment. 2011. *Mercury in the Arctic. Arctic Monitoring and Assessment Programme (AMAP)*. Oslo, Norway.

Barkay T, Gu B. 2022. Demethylation─the other side of the mercury methylation coin: A critical review. ACS Environ. Au. 2:77–97. doi: 10.1021/acsenvironau.1c00022.

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. doi: 10.1111/j.1462-2920.2010.02260.x.

Barkay T, Miller SM, Summers AO. 2003. Bacterial mercury resistance from atoms to ecosystems. FEMS Microbiol Rev. 27:355–384. doi: 10.1016/s0168-6445(03)00046-9.

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.Advances in Applied Microbiology pp. 1–52.

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. doi: 10.1038/srep35168.

362 Boyd E, Barkay T. 2012. The mercury resistance operon: from an origin in a geothermal
363 environment to an efficient detoxification machine. Front. Microbiol. 3. doi:
364 10.3389/fmicb.2012.00349.

365 Capo E et al. 2022. A consensus protocol for the recovery of mercury methylation genes from
366 metagenomes. Mol. Ecol. Resour. 1–15. doi: 10.1111/1755-0998.13687.

367 Chen C et al. 2011. Representative proteomes: a stable, scalable and unbiased proteome set
368 for sequence analysis and functional annotation. PLOS ONE. 6:e18910. doi:
369 10.1371/journal.pone.0018910.

370 Choi S-C, Chase T, Bartha R. 1994. Metabolic pathways leading to mercury methylation in
371 *Desulfovibrio desulfuricans* LS. Appl. Environ. Microbiol. 60:4072–4077. doi:
372 10.1128/aem.60.11.4072-4077.1994.

373 Christakis CA, Barkay T, Boyd ES. 2021. Expanded diversity and phylogeny of mer genes
374 broadens mercury resistance paradigms and reveals an origin for mera among thermophilic
375 archaea. Front. Microbiol. 12. doi: 10.3389/fmicb.2021.682605.

376 Date SS et al. 2019. Kinetics of enzymatic mercury methylation at nanomolar concentrations
377 catalyzed by HgcAB. Appl Env. Microbiol. doi: 10.1128/AEM.00438-19.

378 Eddy SR. 1998. Profile hidden Markov models. Bioinformatics. 14:755–763. doi:
379 10.1093/bioinformatics/14.9.755.

380 Ekstrom EB, Morel FMM, Benoit JM. 2003. Mercury methylation independent of the acetyl-
381 coenzyme A pathway in sulfate-reducing bacteria. Appl Env. Microbiol. 69:5414–5422. doi:
382 10.1128/aem.69.9.5414-5422.2003.

383 Gilmour CC et al. 2013. Mercury methylation by novel microorganisms from new
384 environments. Env. Sci Technol. 47:11810–20. doi: 10.1021/es403075t.

385 Gilmour CC et al. 2011. The sulfate-reducing bacterium *Desulfovibrio desulfuricans* ND132
386 as a model for understanding bacterial mercury methylation. Appl Env. Microbiol. 77:3938–
387 51. doi: 10.1128/AEM.02993-10.

388 Gilmour CC, Bullock AL, McBurney A, Podar M, Elias DA. 2018. Robust mercury
389 methylation across diverse methanogenic *Archaea*. MBio. 9. doi: 10.1128/mBio.02403-17.

390 Gionfriddo CM et al. 2016. Microbial mercury methylation in Antarctic sea ice. Nat
391 Microbiol. 1:16127. doi: 10.1038/nmicrobiol.2016.127.

392 Graham AM et al. 2012. Detailed Assessment of the Kinetics of Hg-Cell Association, Hg
393 Methylation, and Methylmercury Degradation in Several Desulfovibrio Species. Appl.
394 Environ. Microbiol. 78:7337–7346. doi: 10.1128/AEM.01792-12.

395 Granato ET, Meiller-Legrand TA, Foster KR. 2019. The evolution and ecology of bacterial
396 warfare. Curr. Biol. 29:R521–R537. doi: 10.1016/j.cub.2019.04.024.

397 Grasby SE, Them TR, Chen Z, Yin R, Ardakani OH. 2019. Mercury as a proxy for volcanic
398 emissions in the geologic record. Earth-Sci. Rev. 196:102880. doi:
399 10.1016/j.earscirev.2019.102880.

Gu B et al. 2011. Mercury reduction and complexation by natural organic matter in anoxic environments. Proc. Natl. Acad. Sci. 108:1479–1483. doi: 10.1073/pnas.1008747108.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35:518–522. doi: 10.1093/molbev/msx281.

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. 47:2441–56. doi: 10.1021/es304370g.

IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. 1993. Mercury and Mercury Compounds. In: Beryllium, Cadmium, Mercury, and Exposures in the Glass Manufacturing Industry.Vol. 58 International Agency for Research on Cancer: Lyon (FR).

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. doi: 10.1038/srep27958.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30:772–780. doi: 10.1093/molbev/mst010.

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. doi: 10.1016/j.scitotenv.2012.08.057.

Koskiniemi S, Sun S, Berg OG, Andersson DI. 2012. Selection-driven gene loss in bacteria. PLOS Genet. 8:e1002787. doi: 10.1371/journal.pgen.1002787.

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. doi: 10.1093/nar/gkab301.

Li YP et al. 2021. Antimicrobial Activity of Metals and Metalloids. Annu. Rev. Microbiol. 75:175–197. doi: 10.1146/annurev-micro-032921-123231.

Lin H et al. 2021. Mercury methylation by metabolically versatile and cosmopolitan marine bacteria. ISME J. 15:1810–1825. doi: 10.1038/s41396-020-00889-4.

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. doi: 10.1021/acs.estlett.5b00209.

Liu G, Cai Y, O'Driscoll N, Feng X, Jiang G. 2011. Overview of mercury in the environment. In: Environmental chemistry and toxicology of mercury. John Wiley & Sons, Ltd pp. 1–12. doi: 10.1002/9781118146644.ch1.

Liu S, Wiatrowski HA. 2018. Reduction of Hg(II) to Hg(0) by biogenic magnetite from two magnetotactic bacteria. Geomicrobiol. J. 35:198–208. doi: 10.1080/01490451.2017.1362076.

438 Lu X et al. 2016. Anaerobic mercury methylation and demethylation by *Geobacter*
439 *bemidjiensis* Bem. Env. Sci Technol. 50:4366–73. doi: 10.1021/acs.est.6b00401.

440 Mai U, Sayyari E, Mirarab S. 2017. Minimum variance rooting of phylogenetic trees and
441 implications for species tree reconstruction. PLOS ONE. 12:e0182238. doi:
442 10.1371/journal.pone.0182238.

443 McDaniel EA et al. 2020. Expanded phylogenetic diversity and metabolic flexibility of
444 mercury-methylating microorganisms. mSystems. 5. doi: 10.1128/mSystems.00299-20.

445 Moody ER et al. 2022. An estimate of the deepest branches of the tree of life from ancient
446 vertically evolving genes Perry, GH & Koonin, EV, editors. eLife. 11:e66695. doi:
447 10.7554/eLife.66695.

448 Parks JM et al. 2013. The genetic basis for bacterial mercury methylation. Science. 339:1332–
449 5. doi: 10.1126/science.1230667.

450 Podar M et al. 2015. Global prevalence and distribution of genes and microorganisms
451 involved in mercury methylation. Sci Adv. 1:e1500675. doi: 10.1126/sciadv.1500675.

452 Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other
453 things). Methods Ecol. Evol. 3:217–223. doi: 10.1111/j.2041-210X.2011.00169.x.

454 Schmidt HA, Minh BQ, von Haeseler A, Nguyen L-T. 2014. IQ-TREE: a fast and effective
455 stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 32:268–
456 274. doi: 10.1093/molbev/msu300.

457 Siciliano SD, Lean DRS. 2002. Methyltransferase: An enzyme assay for microbial
458 methylmercury formation in acidic soils and sediments. Environ. Toxicol. Chem. 21:1184–
459 1190. doi: 10.1002/etc.5620210610.

460 Silla-Martínez JM, Capella-Gutiérrez S, Gabaldón T. 2009. trimAl: a tool for automated
461 alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 25:1972–1973. doi:
462 10.1093/bioinformatics/btp348.

463 Sone Y et al. 2017. A Novel Role of MerC in Methylmercury Transport and
464 Phytoremediation of Methylmercury Contamination. Biol. Pharm. Bull. 40:1125–1128. doi:
465 10.1248/bpb.b17-00213.

466 Sone Y et al. 2013. Increase methylmercury accumulation in Arabidopsis thaliana expressing
467 bacterial broad-spectrum mercury transporter MerE. AMB Express. 3:52. doi: 10.1186/2191-
468 0855-3-52.

469 Sousa FL, Martin WF. 2014. Biochemical fossils of the ancient transition from geoenergetics
470 to bioenergetics in prokaryotic one carbon compound metabolism. Biochim. Biophys. Acta
471 BBA - Bioenerg. 1837:964–981. doi: 10.1016/j.bbabio.2014.02.001.

472 Svetlitchnaia T, Svetlitchnyi V, Meyer O, Dobbek H. 2006. Structural insights into
473 methyltransfer reactions of a corrinoid iron–sulfur protein involved in acetyl-CoA synthesis.
474 Proc. Natl. Acad. Sci. 103:14331–14336. doi: 10.1073/pnas.0601420103.

475 Swofford DL, Olsen GJ, Waddell PJ, Hillis DM. 1996. Phylogenetic inference. In: Molecular

systematics.Vol. 15 Sinauer Associates, Inc.: Sunderland, Massachusetts pp. 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. doi: 10.3389/fmicb.2020.01369.

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. doi: doi:10.1128/Spectrum.00833-21.

Tebo BM et al. 2004. Biogenic manganese oxides: properties and mechanisms of formation. Annu. Rev. Earth Planet. Sci. 32:287–328. doi: 10.1146/annurev.earth.32.101802.120213.

Trevors JT. 1986. Mercury methylation by bacteria. J. Basic Microbiol. 26:499–504. doi: 10.1002/jobm.3620260811.

Tria FDK, Landan G, Dagan T. 2017. Phylogenetic rooting using minimal ancestor deviation. Nat. Ecol. Evol. 1:1–7. doi: 10.1038/s41559-017-0193.

Villar E, Cabrol L, Heimburger-Boavida LE. 2020. Widespread microbial mercury methylation genes in the global ocean. Env. Microbiol Rep. doi: 10.1111/1758-2229.12829.

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. doi: https://doi.org/10.1016/j.jhazmat.2021.126964.

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. doi: 10.1371/journal.pgen.1007518.

Zerkle AL et al. 2020. Anomalous fractionation of mercury isotopes in the Late Archean atmosphere. Nat. Commun. 11:1709. doi: 10.1038/s41467-020-15495-3.

Zerkle AL et al. 2021. Sulfur and mercury MIF suggest volcanic contributions to Earth's atmosphere at 2.7 Ga. doi: 10.7185/geochemlet.2124.

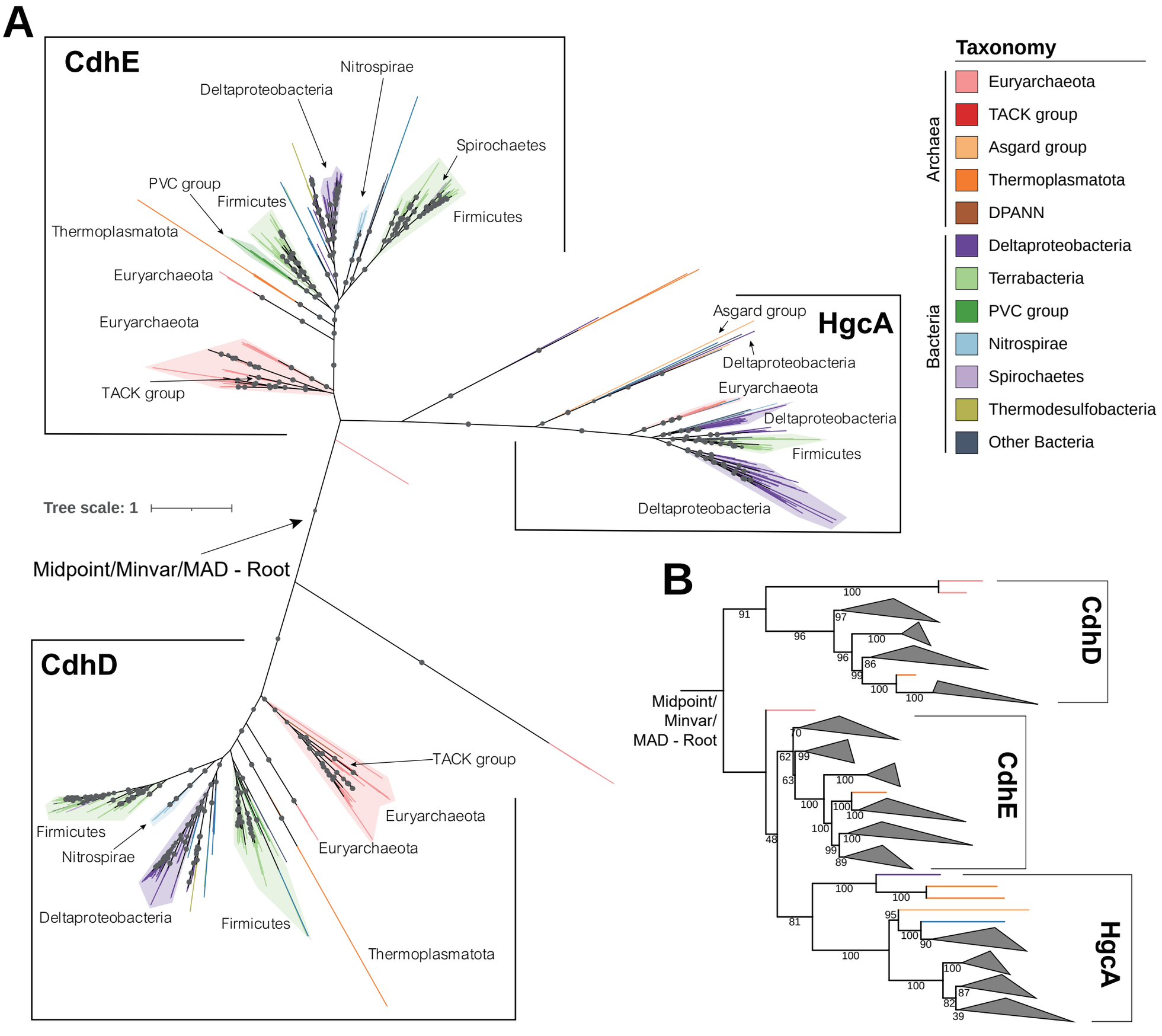
Figure Legends

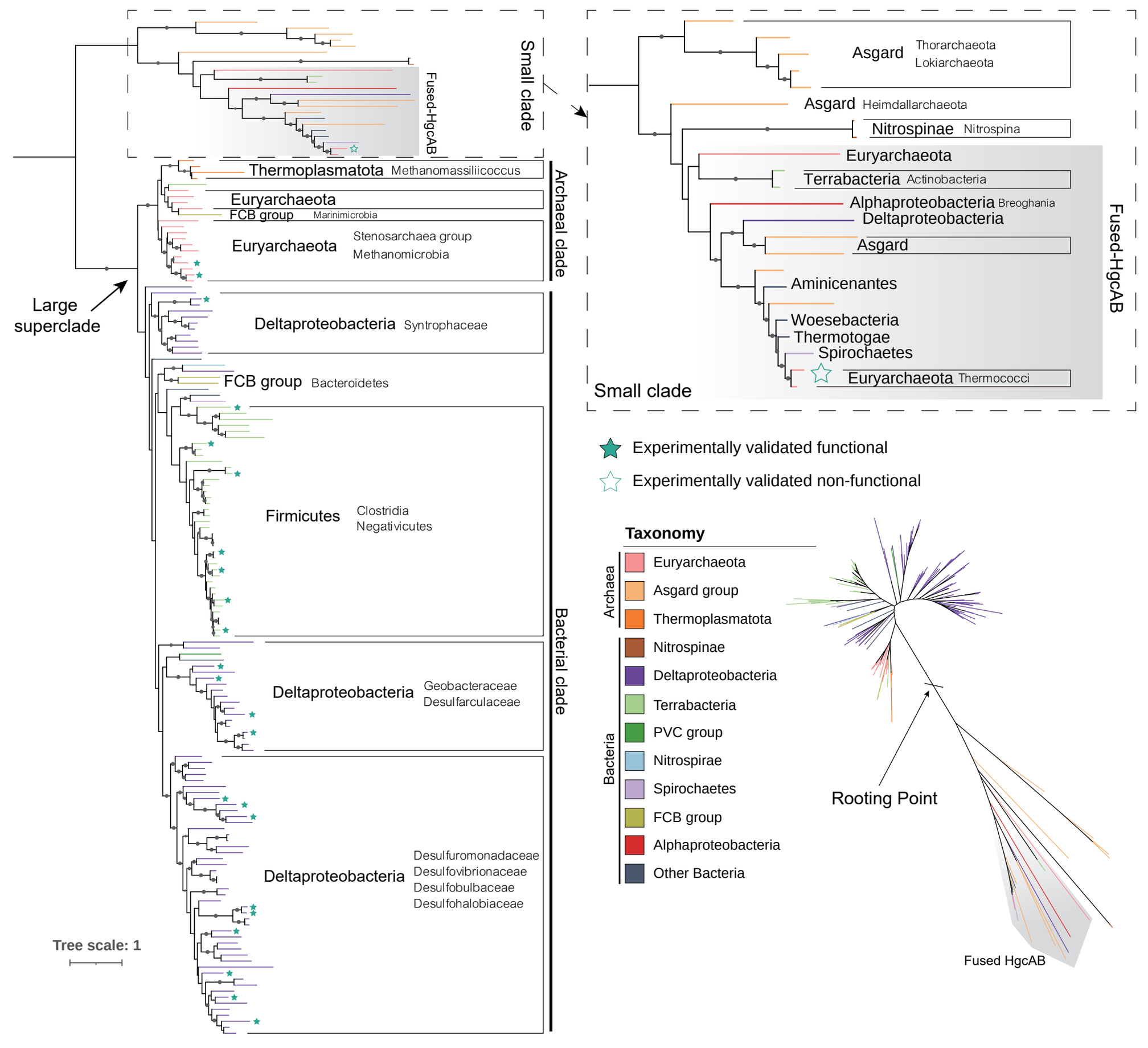
Figure 1. Phylogenetic tree of the protein family PF03599. (A) Unrooted tree of the protein family PF03599. The tree was inferred by using the Maximum Likelihood method under LG+C50+F+R model. This analysis involved 478 amino acid sequences with a total of 2922 positions in the alignments. Different taxonomies are represented by different colors. Ultrafast bootstrap support values were calculated with 1000 replications, and ultrafast

508 bootstrap values > 90% are shown by black dots at the nodes. (B) PF03599 tree rooted by the
509 midpoint. Clades whose average branch length distance to their leaves are below 1.5 are
510 collapsed with iTOL for better visualization.
511
512 **Figure 2. Phylogenetic tree of HgcA proteins.** The tree was inferred by using the Maximum
513 Likelihood method under LG+C60+F+G model. This analysis involved alignment of 169
514 amino acid sequences with a total of 493 positions. Different taxonomies are represented by
515 different colors. Ultrabootstrap support values were calculated with 1000 replications, and
516 ultrabootstrap values > 90% are shown by black dots at the nodes. Experimentally validated
517 functional and non-functional HgcA sequences are labelled by solid and hollow stars. Fused-
518 HgcAB sequences are indicated by a grey background.
519
520 **Figure 3. Co-phylogenetic tree of HgcA and HgcB.** The phylogeny of HgcA shown on the
521 left is the same as Figure 2. The phylogeny of HgcB shown on the right was inferred by using
522 the Maximum Likelihood method under LG+C60+F+G model. This analysis involved
523 alignment of 169 amino acid sequences with a total of 322 positions. Taxonomies of the
524 HgcB sequences are represented in the same colors as shown in the HgcA phylogeny.
525 Ultrabootstrap support values were calculated with 1000 replications, and ultrabootstrap
526 values > 90% are shown by black dots at the nodes. Gray shaded blocks describe fused-
527 HgcAB and 'HgcB tail' genes in the two trees, respectively. Lines connecting the co-
528 phylogenies demonstrate associations between the two trees. The green dots at the tip of the
529 HgcA tree represent the corresponding *hgcA* without downstream *hgcB*, and the red triangle
530 represents the corresponding fused-*hgcAB* with another *hgcB* downstream.
531

Figure 4. Phylogenetic tree of MerB proteins. The tree was inferred by using the Maximum Likelihood method under LG+C40+F+G model. This analysis involved alignment of 225 amino acid sequences with a total of 1547 positions. Different taxonomies are represented by different colors. Ultrabootstrap support values were calculated with 1000 replications, and ultrabootstrap values > 90% are shown by black dots at the nodes. The tree was rooted using two MerA proteins.

Figure 5. Species tree of *hgcA*+ and *merB*+ genomes in the RP35 database. The tree was inferred by using the Maximum Likelihood method under LG+F+R10 model. This analysis involved alignment of 251 amino acid sequences with a total of 8024 positions. Different taxonomies are represented by different colors. The presence of genes *hgcA*/fused *hgcAB*/*hgcB*/*merB* in the genomes were represented by different symbols and colors in the outer circle. Ultrabootstrap support values were calculated with 1000 replications, and ultrabootstrap values > 90% are shown by black dots at the nodes.


A
CdhE
Nitrospirae
Deltaproteobacteria
Spirochaetes
PVC group
Firmicutes
Firmicutes
Thermoplasmatota
Euryarchaeota
Euryarchaeota
TACK group
HgcA
Asgard group
Deltaproteobacteria
Euryarchaeota
Deltaproteobacteria
Firmicutes
Deltaproteobacteria
Tree scale: 1
Midpoint/Minvar/MAD - Root
CdhD
TACK group
Euryarchaeota
Euryarchaeota
Firmicutes
Nitrospirae
Deltaproteobacteria
Firmicutes
Thermoplasmatota
Taxonomy
Archaea
Euryarchaeota
TACK group
Asgard group
Thermoplasmatota
DPANN
Bacteria
Deltaproteobacteria
Terrabacteria
PVC group
Nitrospirae
Spirochaetes
Thermodesulfobacteria
Other Bacteria
B
Midpoint/
Minvar/
MAD - Root
CdhD
CdhE
HgcA


Small clade
Fused-HgcAB
Large superclade
Thermoplasmatota Methanomassiliicoccus
Euryarchaeota
FCB group Marinimicrobia
Euryarchaeota Stenosarchaea group
Methanomicrobia
Archaeal clade
Deltaproteobacteria Syntrophaceae
FCB group Bacteroidetes
Firmicutes Clostridia
Negativicutes
Deltaproteobacteria Geobacteraceae
Desulfarculaceae
Deltaproteobacteria Desulfuromonadaceae
Desulfovibrionaceae
Desulfobulbaceae
Desulfohalobiaceae
Bacterial clade
Tree scale: 1
Asgard Thorarchaeota
Lokiarchaeota
Asgard Heimdallarchaeota
Nitrospinae Nitrospina
Euryarchaeota
Terrabacteria Actinobacteria
Alphaproteobacteria Breoghania
Deltaproteobacteria
Asgard
Aminicenantes
Woesebacteria
Thermotogae
Spirochaetes
Euryarchaeota Thermococci
Small clade
Fused-HgcAB
Experimentally validated functional
Experimentally validated non-functional
Taxonomy
Archaea
Euryarchaeota
Asgard group
Thermoplasmatota
Bacteria
Nitrospinae
Deltaproteobacteria
Terrabacteria
PVC group
Nitrospirae
Spirochaetes
FCB group
Alphaproteobacteria
Other Bacteria
Rooting Point
Fused HgcAB

HgcA Tree

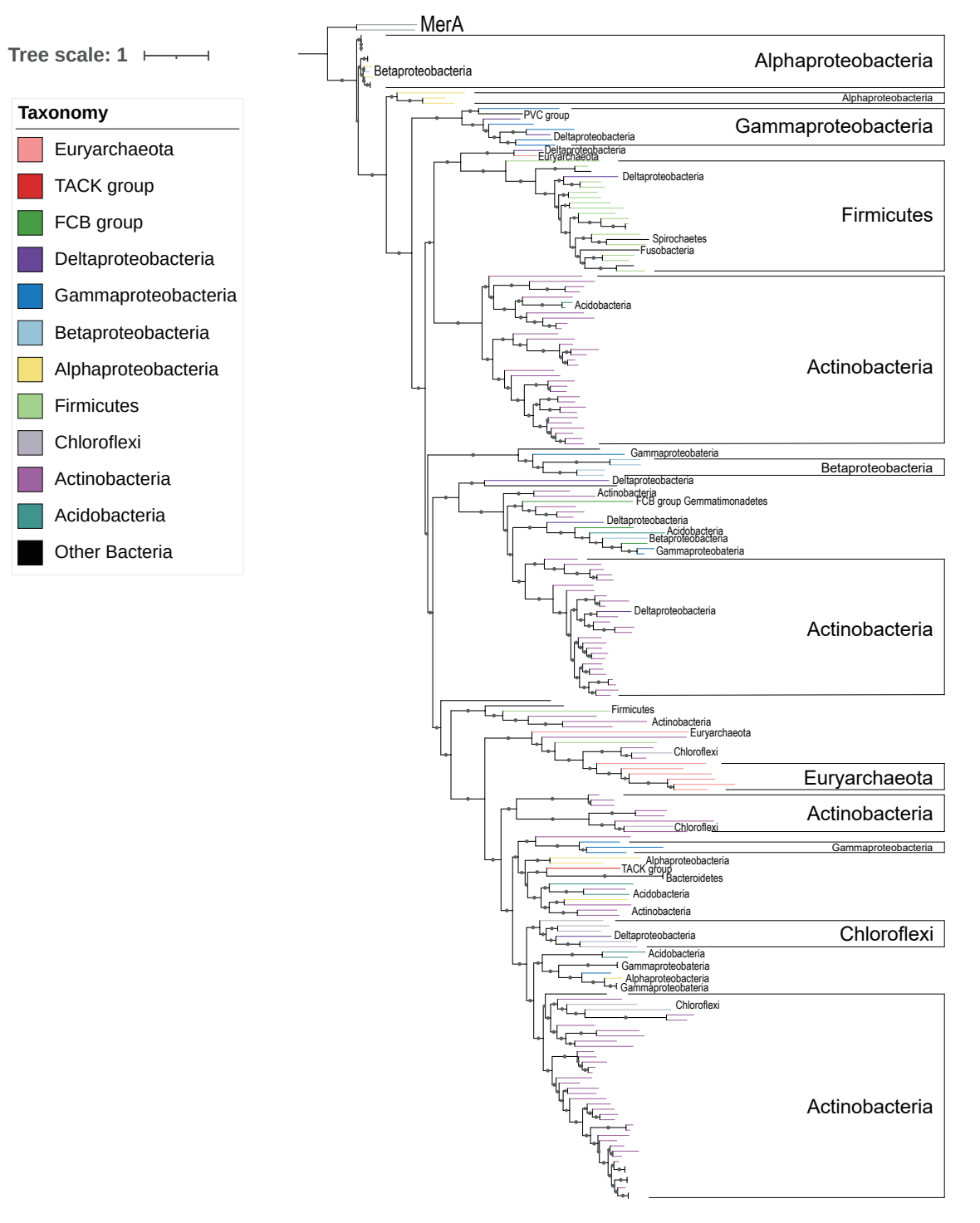
HgcB Tree

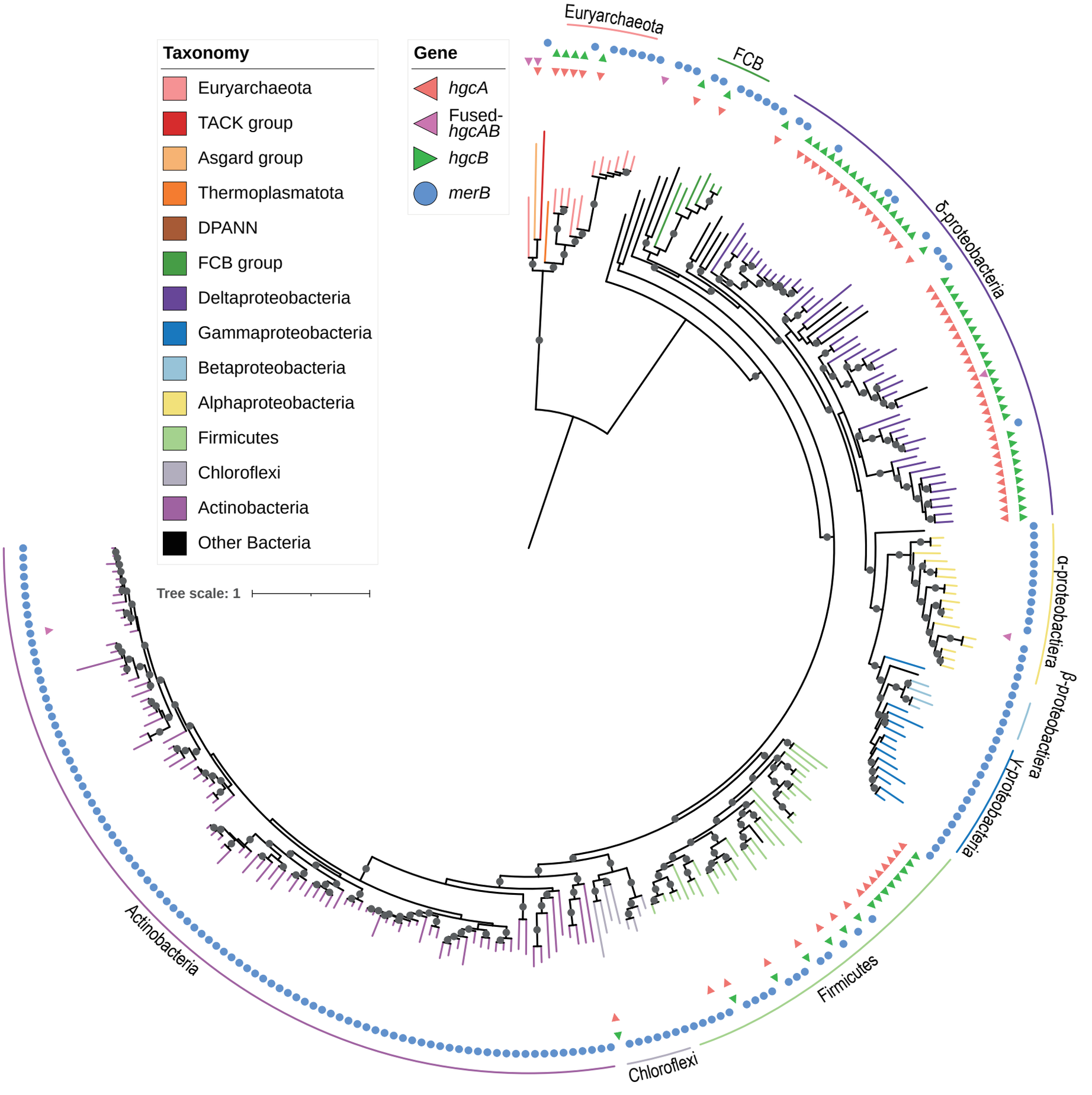
Tree scale: 1

Taxonomy

- Euryarchaeota
- Asgard group
- Thermoplasmatota
- Nitrospinae
- Deltaproteobacteria
- Terrabacteria
- PVC group
- Nitrospirae
- Spirochaetes
- FCB group
- Alphaproteobacteria
- Other Bacteria

- Associations
- *hgcA* without downstream *hgcB*
- Fused-*hgcAB* with a downstream *hgcB*


Tree scale: 1
Taxonomy
Euryarchaeota
TACK group
FCB group
Deltaproteobacteria
Gammaproteobacteria
Betaproteobacteria
Alphaproteobacteria
Firmicutes
Chloroflexi
Actinobacteria
Acidobacteria
Other Bacteria
MerA
Betaproteobacteria
Alphaproteobacteria
Alphaproteobacteria
Gammaproteobacteria
PVC group
Deltaproteobacteria
Deltaproteobacteria
Euryarchaeota
Deltaproteobacteria
Firmicutes
Spirochaetes
Fusobacteria
Acidobacteria
Actinobacteria
Gammaproteobateria
Betaproteobacteria
Deltaproteobacteria
Actinobacteria
FCB group Gemmatimonadetes
Deltaproteobacteria
Acidobacteria
Betaproteobacteria
Gammaproteobateria
Deltaproteobacteria
Actinobacteria
Firmicutes
Actinobacteria
Euryarchaeota
Chloroflexi
Euryarchaeota
Actinobacteria
Chloroflexi
Gammaproteobacteria
Alphaproteobacteria
TACK group
Bacteroidetes
Acidobacteria
Actinobacteria
Chloroflexi
Deltaproteobacteria
Acidobacteria
Gammaproteobateria
Alphaproteobacteria
Gammaproteobateria
Chloroflexi
Actinobacteria


Taxonomy
Euryarchaeota
TACK group
Asgard group
Thermoplasmatota
DPANN
FCB group
Deltaproteobacteria
Gammaproteobacteria
Betaproteobacteria
Alphaproteobacteria
Firmicutes
Chloroflexi
Actinobacteria
Other Bacteria
Gene
hgcA
Fused-hgcAB
hgcB
merB
Tree scale: 1
Euryarchaeota
FCB
δ-proteobacteria
α-proteobactiera
β-proteobactiera
γ-proteobacteria
Firmicutes
Chloroflexi
Actinobacteria