bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

# Gallionellaceae pangenomic analysis reveals insight into phylogeny, metabolic flexibility, and iron oxidation mechanisms

Rene L. Hoover[a,b], Jessica L. Keffer[b], Shawn W. Polson[d,e], Clara S. Chan[a,b,c]

a Microbiology Graduate Program, University of Delaware, Newark, Delaware, USA

b Department of Earth Sciences, University of Delaware, Newark, Delaware, USA

c School of Marine Science and Policy, University of Delaware, Newark, Delaware, USA

d Department of Computer and Information Sciences, University of Delaware, Newark, Delaware, USA

e Center for Bioinformatics and Computational Biology, University of Delaware, Newark, Delaware, USA

Corresponding author: Clara S. Chan, cschan@udel.edu

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

# Abstract

The iron-oxidizing Gallionellaceae drive a wide variety of biogeochemical cycles through their metabolisms and biominerals. To better understand the environmental impacts of Gallionellaceae, we need to improve our knowledge of their diversity and metabolisms, especially any novel iron oxidation mechanisms. Here, we used a pangenomic analysis of 103 genomes to resolve Gallionellaceae phylogeny and explore the range of genomic potential. Using a concatenated ribosomal protein tree and key gene patterns, we determined Gallionellaceae has four genera, divided into two groups–iron-oxidizing bacteria (FeOB) *Gallionella*, *Sideroxydans*, and *Ferriphaselus* with known iron oxidases (Cyc2, MtoA) and nitrite-oxidizing bacteria (NOB) *Candidatus* Nitrotoga with nitrite oxidase (Nxr). The FeOB and NOB have similar electron transport chains, including genes for reverse electron transport and carbon fixation. Auxiliary energy metabolisms including S oxidation, denitrification, and organotrophy were scattered throughout the Gallionellaceae FeOB. Within FeOB, we found genes that may represent adaptations for iron oxidation, including a variety of extracellular electron uptake (EEU) mechanisms. FeOB genomes encoded more predicted *c*-type cytochromes overall, notably more multiheme *c*-type cytochromes (MHCs) with >10 CXXCH motifs. These include homologs of several predicted outer membrane porin-MHC complexes, including MtoAB and Uet. MHCs are known to efficiently conduct electrons across longer distances and function across a wide range of redox potentials that overlap with mineral redox potentials, which can help expand the range of usable iron substrates. Overall, the results of pangenome analyses suggest that the Gallionellaceae genera *Gallionella*, *Sideroxydans*, and *Ferriphaselus* are primarily iron oxidizers, capable of oxidizing dissolved $Fe^{2+}$ as well as a range of solid iron or other mineral substrates.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

**Importance**

Neutrophilic iron-oxidizing bacteria (FeOB) produce copious iron (oxyhydr)oxides that can profoundly influence biogeochemical cycles, notably the fate of carbon and many metals. To fully understand environmental microbial iron oxidation, we need a thorough accounting of iron oxidation mechanisms. In this study we show the Gallionellaceae FeOB have both known iron oxidases as well as uncharacterized multiheme cytochromes (MHCs). MHCs are predicted to transfer electrons from extracellular substrates and likely confer metabolic capabilities that help Gallionellaceae occupy a range of different iron- and mineral-rich niches. Gallionellaceae appear to specialize in iron oxidation, so it makes sense that they would have multiple mechanisms to oxidize various forms of iron, given the many iron minerals on Earth, as well as the physiological and kinetic challenges faced by FeOB. The multiple iron/mineral oxidation mechanisms may help drive the widespread ecological success of Gallionellaceae.

# Introduction

*Gallionella* are one of the oldest known and most well studied iron-oxidizing bacteria (FeOB), yet we are still learning how they oxidize iron and adapt to iron-rich niches. *Gallionella* is the type genus of the family Gallionellaceae, which also includes *Sideroxydans*, *Ferriphaselus*, and *Ferrigenium*. These Gallionellaceae FeOB are found in a wide range of environments, including freshwater creeks, sediment, root rhizospheres, peat, permafrost, deep subsurface aquifers, and municipal waterworks (1–18). FeOB potentially drive the fate of many metals and nutrients via both metabolic reactions and forming iron oxyhydroxides that adsorb and react with many solutes (19). To better understand the biogeochemical effects of Gallionellaceae, we need to improve our knowledge of their phylogeny and metabolic mechanisms, especially for iron

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

oxidation. Recently, the rapid increase in metagenomes from iron-rich environments has significantly expanded the number of available Gallionellaceae genomes, which makes it possible to investigate diversity and mechanisms using genomic analyses of both cultured and uncultured Gallionellaceae.

The Gallionellaceae are named after *Gallionella ferruginea*, first described by Ehrenberg in 1838 (20), and recognizable by its distinctive, twisted iron oxyhydroxide stalk (21). While the type strain, *G. ferruginea* Johan (22) no longer exists, there are seven iron-oxidizing Gallionellaceae isolates (7, 11, 23–26). Some isolates, such as *Ferriphaselus* spp., appear to be obligate iron oxidizers, while others also grow on non-iron substrates. In addition to iron, *S. lithotrophicus* ES-1 grows by thiosulfate oxidation (24, 27) while *Sideroxydans* sp. CL21 shows mixotrophic growth with either lactate or hydrogen (28). Some *Ferrigenium* can also reduce nitrate (29, 30). It is unknown how common it is for Gallionellaceae to use electron donors/acceptors besides Fe(II)/$O_2$, though these alternate metabolisms may help their success across different environments and fluctuating conditions typical of many oxic-anoxic interfaces. Even so, since all seven Gallionellaceae isolates are neutrophilic aerobic chemolithoautotrophic iron oxidizers, this could be the dominant metabolic niche of Gallionellaceae.

In Gallionellaceae and other neutrophilic chemolithotrophic FeOB, there are two known iron oxidases: Cyc2, a fused monoheme cytochrome-porin and MtoAB, a decaheme porin-cytochrome complex (31–33). The *mtoA* (metal oxidation) gene was first identified and characterized in FeOB *S. lithotrophicus* ES-1 (31). The *mtoA* gene is a homolog of both *pioA* (phototrophic iron oxidation), which encodes the PioA iron oxidase in the photoferrotroph *Rhodopseudomonas palustris* TIE-1 (34, 35), and *mtrA* (metal reduction), which encodes the MtrA iron reductase in iron-reducing bacteria (FeRB) *Shewanella* (36). The *cyc2* gene is more

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

common than *mtoAB* and is found in nearly all well-characterized neutrophilic FeOB like the Gallionellaceae (32) and Zetaproteobacteria (33), making it a suitable genetic marker for many FeOB. Cyc2 has been demonstrated to oxidize aqueous $Fe^{2+}$ (32), while Mto gene/protein expression has been associated with the oxidation of solid iron minerals (37). However, Cyc2 and MtoA may not be the only mechanisms for neutrophilic iron oxidation. There are a number of additional uncharacterized cytochromes and electron transport genes (27, 38) within Gallionellaceae genomes such as isolate *S. lithotrophicus* ES-1 (27, 38), suggesting the existence of novel iron oxidation genes and mechanisms within the family.

The Gallionellaceae also includes a recently identified genus, *Candidatus* Nitrotoga, which are chemolithotrophic nitrite-oxidizing bacteria (NOB). Like the iron-oxidizing Gallionellaceae, they are widespread in freshwater and engineered environments, including permafrost (39), coastal sediments (40), freshwater (41), freshwater sediments (42), and the activated sludge of wastewater treatment facilities (43, 44). There are only two isolates, *Ca.* Nitrotoga fabula (43) and *Ca.* Nitrotoga sp. AM1P (45), along with four genomes from enrichment cultures (42). *Ca.* Nitrotoga are adapted to niches with low nitrite, and oxidize it using a distinct high-affinity Nxr nitrite oxidoreductase (39, 42, 46). Extensive iron uptake mechanisms have been detected in *Ca.* Nitrotoga genomes, indicating the importance of iron for growth, likely due to the FeS cluster of Nxr (42). However, neither the isolates nor enrichments are known to oxidize Fe(II). If *Ca.* Nitrotoga lack the capacity to oxidize iron, then we can investigate the iron-oxidizing mechanisms and adaptations of Gallionellaceae through a comparative genomic analysis of iron- versus nitrite-oxidizing members.

Toward this goal, we took advantage of the growing number of environmental metagenomes and collected 103 high quality Gallionellaceae genomes and metagenome

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

assembles genomes (MAGs). We used those sequences to resolve the Gallionellaceae phylogeny and delineate groups of iron and nitrite oxidizers. We searched for known and novel iron oxidation genes, other energy and nutrient metabolisms, and genes found exclusively in FeOB that may represent adaptations for an iron-oxidizing lifestyle. This work increases our understanding of Gallionellaceae family phylogeny and the metabolic traits of its genera. It also highlights some of the key multiheme cytochromes in Gallionellaceae FeOB, which may facilitate extracellular electron uptake (EEU) and the oxidation of different iron substrates.

# Results

## Phylogeny

We collected 103 high quality Gallionellaceae isolate genomes and metagenome assembled genomes (MAGs) from various databases and collections (Table S1). Many of these MAGs were only classified at the family level. To resolve the phylogeny, verify existing classifications, and classify the unknown Gallionellaceae, we constructed a concatenated protein tree (Figure 1) from 13 ribosomal protein sequences. Organisms in the tree formed distinct, well-supported clades that corresponded to the major genera: *Gallionella*, *Sideroxydans*, *Ferriphaselus*, and *Ca.* Nitrotoga (Figure 1). Most of the MAGs previously classified as Gallionellaceae and Gallionellales were found to be either *Gallionella* or *Sideroxydans*, with the exception of one that clustered with the *Ca*. Nitrotoga (*Ca*. Nitrotoga SL_21). Although some genomes formed sub-clades, many were organized along a continuum. Near the base of the *Gallionella* are *Ferrigenium kumadai* An22 and the nitrate-reducing iron-oxidizing bacteria (NRFeOB) of the Straub (KS) and Bremen Pond (BP) enrichments (Figure 1). There is not a clear boundary between the *Gallionella* and the relatively new *Ferrigenium* genus, so we

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

included the *Ferrigenium* and NRFeOB with the *Gallionella* grouping for our analyses. We also constructed a 16S rRNA gene tree containing 24 sequences in our dataset along with 941 high-quality, full-length Gallionellaceae sequences from the SILVA database (Fig. S1), but bootstrap support was weaker and clades were less clearly resolved. Therefore, concatenated ribosomal proteins are a more reliable determinant of Gallionellaceae phylogeny than 16S rRNA genes.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

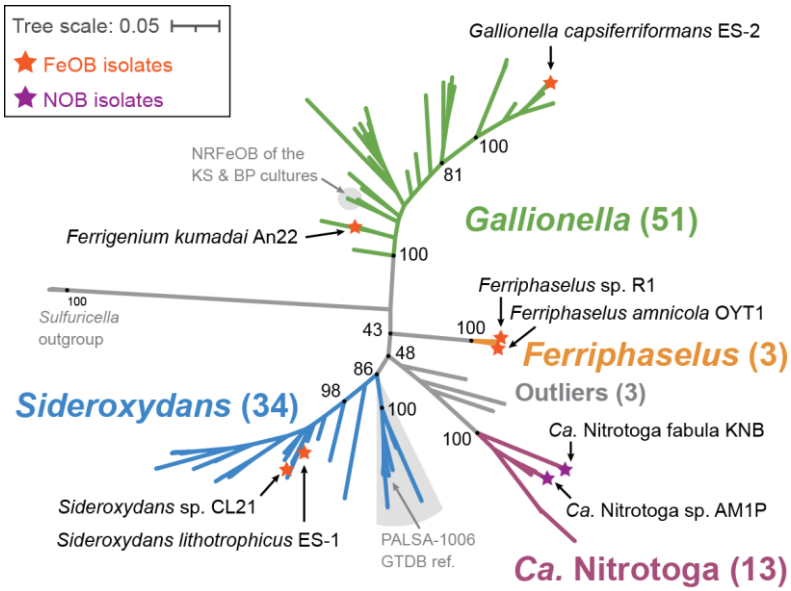
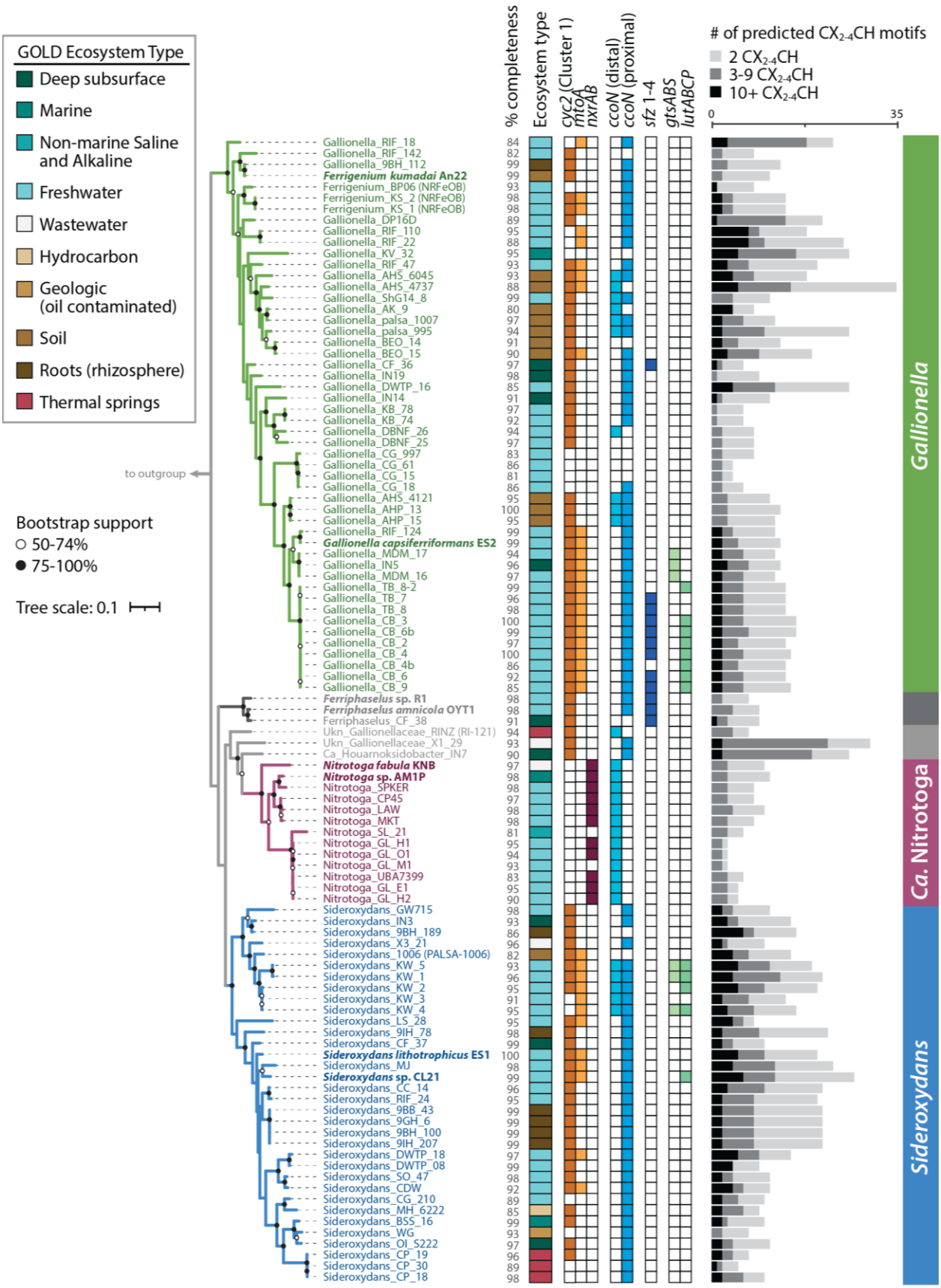



**FIGURE 1** Concatenated ribosomal protein maximum likelihood tree of the Gallionellaceae family showing the four distinct genera: *Gallionella*, *Sideroxydans*, *Ferriphaselus*, and *Ca.* Nitrotoga. Isolates are labeled and annotated with stars. Support values from 1000 bootstraps shown for major branching nodes (black dots). Detailed tree shown in Fig. 2.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.




135

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

**FIGURE 2** Maximum likelihood tree of concatenated ribosomal proteins from the Gallionellaceae annotated with source ecosystem and genes for iron oxidases (*cyc2*, *mtoA*), nitrite oxidase (*nxrAB*), terminal oxidase (*ccoN*), stalk formation (*sfz*), and organic utilization (*gtsABS*, *lutABCP*). The bar graph to the right shows the number multiheme cytochromes CXXCH, $CX_3CH$, and $CX_4CH$ heme-binding motifs. Phylogeny does not correlate to environments, and key genes, including those for multiheme cytochromes, show distinct distributions between iron and nitrite oxidizer clades. Isolates are shown in bold. % completeness = genome completeness calculated with CheckM. Outgroup omitted for space.

We assessed whether there was a relationship between phylogeny and environment. Each genome and MAG was classified with the GOLD classification schema (47) based on pre-existing GOLD classifications, available metadata and publications (Fig. 2, Table S2). The majority of aquatic genomes were from freshwater and groundwater environments while terrestrial genomes were mostly found in soil, peat, and rhizosphere environments. However, some genomes were sequenced from more extreme environments such as thermal hot springs (ENVO:00002181) and acid mine drainage (ENVO:00001997) (Table S2). Gallionellaceae are widespread and can inhabit many different environments, but there was no clear pattern between GOLD Ecosystem Type and broad phylogenetic groupings (Fig. 2). Different Gallionellaceae appear to co-exist in some environments, suggesting niches not captured in the ecosystem classification are controlling Gallionellaceae diversity and environmental distribution.

## Metabolic potential and diversity

The Gallionellaceae family has few isolates, so to uncover the shared metabolic traits of its FeOB members, we compared and contrasted *Gallionella*, *Sideroxydans*, and *Ferriphaselus*

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

genomes to those of the nitrite-oxidizing *Ca.* Nitrotoga. We identified key genes within the pangenome for iron oxidation (including predicted *c*-type cytochromes), carbon fixation, and respiration using a combination of DRAM (48), FeGenie (49), MagicLamp (50), a heme motif counter script (51), and BLAST (52, 53). To further uncover genes and pathways specifically enriched in the iron oxidizers, we used Anvi'o (54–56) to analyze a filtered dataset of only *Gallionella*, *Sideroxydans*, and *Ca*. Nitrotoga genomes that were >97% complete. This approach enabled us to create a comprehensive picture of Gallionellaceae metabolic diversity and pinpoint promising gene clusters that may be adaptations for an iron-oxidizing lifestyle.

**Primary energy metabolisms — iron and nitrite oxidation**

Known metabolisms for the few Gallionellaceae isolates suggest *Ca*. Nitrotoga are nitrite oxidizers, while *Sideroxydans*, *Ferriphaselus*, and *Gallionella* are iron oxidizers. We examined the pangenome for the presence of *cyc2* and *mtoA* iron oxidase genes and *nxrAB* nitrite oxidase genes to determine if that pattern also holds throughout the uncultured Gallionellaceae. As with the isolates, there is a clear delineation between organisms with marker genes for iron versus nitrite oxidation, which corresponds to the phylogenetic groups (Fig. 2, Fig. 3).

The *cyc2* gene is widespread among clades of iron oxidizers, with at least one copy detected in 83% of the FeOB genomes (Fig. 3, Table S3). The *mtoA* gene is found in 41% of the FeOB genomes, and 37% of genomes have both *mtoA* and *cyc2*. In total, 89% have at least one iron oxidase gene, either *cyc2* or *mtoA* (Table S3). Since the dataset includes multiple MAGs with a mean completeness score of 95%, it appears that almost all Gallionellaceae FeOB contain one of these two known mechanisms for iron oxidation. Overall, *cyc2* homologs are more common than *mtoA* (Fig. 2, Fig. 3) and some genomes encode multiple copies of *cyc2* (Table S3). All of the FeOB Gallionellaceae with *cyc2* encode at least one copy that is closely related to

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Cluster 1 Cyc2 (classified as in McAllister, et al. (33)), which has been functionally verified as an iron oxidase (32).

The *Ca.* Nitrotoga SL_21 MAG contains only a predicted Cluster 2 Cyc2 homolog. Confidently assigning iron oxidation function to Cluster 2 Cyc2s depends on supporting context, which is lacking in this case. *Ca.* Nitrotoga SL_21 is not from a typical iron-oxidizing environment (permanently stratified, non-marine, saline lake) and it is not closely related to the functionally verified Cluster 2 Cyc2 representative, *Acidithiobacillus*. Currently, there is no evidence that this sole *Ca.* Nitrotoga Cyc2 is an iron oxidase.

In contrast, *nxrAB* genes are exclusive to the *Ca.* Nitrotoga and copies are present in 85% of the genomes (Fig. 2, Fig. 3). Given that many of the genomes are MAGs with a mean completeness of 94%, distribution of *nxrAB* appears to indicate nitrate oxidation is the main metabolism of *Ca.* Nitrotoga. Thus, our pangenome analysis confirms Gallionellaceae can be divided into two main groups based on primary energy metabolism – FeOB and NOB.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

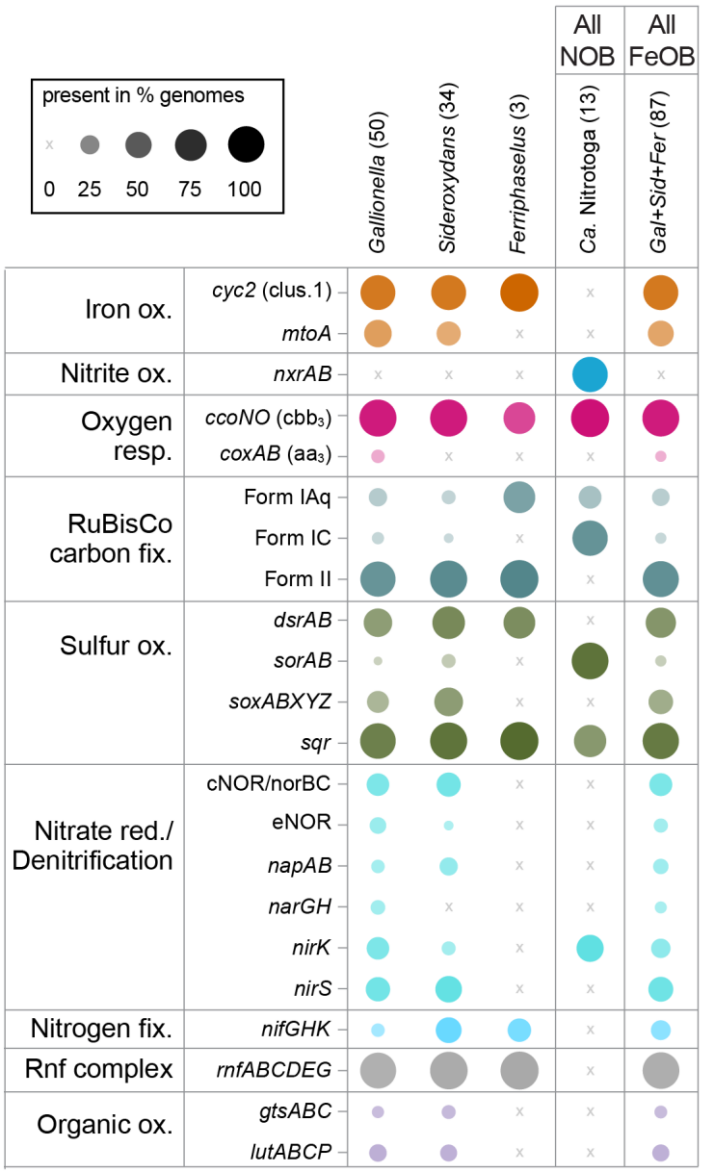
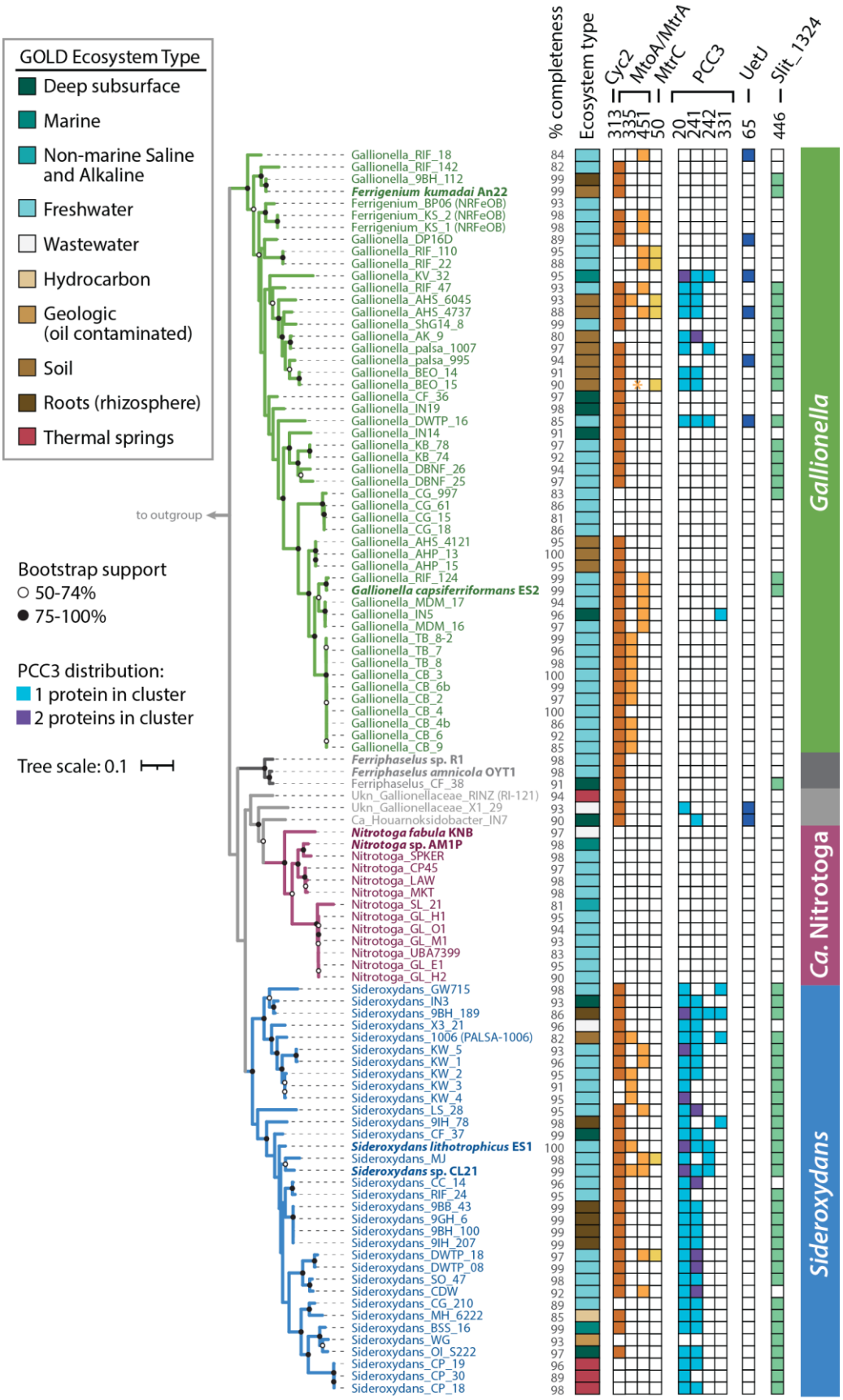



**FIGURE 3** Plot showing the percent of genomes in each genus/group with genes for key metabolic pathways. The plot indicates the Gallionellaceae are aerobic lithoautotrophs with two main energy metabolisms, iron or nitrite oxidation. Some members also have metabolic potential for S oxidation and/or denitrification. Numbers in parentheses indicate the total number of genomes in each group. Color is used to distinguish groups, while dot size and opacity indicate % presence in the genome groups.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

**FIGURE 4** Maximum likelihood tree of concatenated ribosomal proteins from the Gallionellaceae that shows the distribution of MMSeqs2 Clusters that represent predicted cytochromes Cyc2, MtoA, MtoC, PCC3, Uet, and Slit_1324. Asterisk (*) for Gallionella_BEO_15 indicates a partial MtoA sequence was detected using HMMs and verified with BLAST, but was too short to bin into the MMseqs2 MtoA clusters. Isolates are shown in bold. % completeness = genome completeness calculated with CheckM. Outgroup omitted for space.

## *c*-type cytochromes

Both known iron oxidases (Cyc2 and MtoA) in Gallionellaceae are *c*-type cytochromes that transport electrons across the outer membrane. FeOB use additional *c*-type cytochromes to transport electrons through the periplasm to the rest of the electron transport chain. We reasoned that novel iron oxidation mechanisms may also utilize *c*-type cytochromes, so we searched the Gallionellaceae genomes for proteins containing the CXXCH, $CX_3CH$, and $CX_4CH$ heme-binding motifs (abbreviated hereafter as CXXCH). There is a stark difference between FeOB and NOB in the distribution of predicted *c*-type cytochromes. FeOB genomes have an average of 1.5x more CXXCH-containing proteins than NOB, and only non-NOB genomes encoded proteins with ten or more CXXCH motifs (Fig. 2). The abundance of genes for potential *c*-type cytochromes, in particular multiheme cytochromes (MHCs), suggest the presence of additional iron oxidation mechanisms in the Gallionellaceae FeOB.

To find *c*-type cytochromes of interest, all CXXCH-containing proteins were clustered using MMSeqs2 with bidirectional coverage and an 80% alignment cutoff. Clusters of sequences were then classified with representative sequences from isolates using BLAST to query the Uniprot database. If the cluster did not contain a sequence from an isolate, a consensus

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

classification was used. A cluster of monoheme proteins (Cluster 313) was classified as Cyc2 and three clusters of decaheme proteins were classified as MtoA (Clusters 335 and 451) and MtrC (Cluster 50) (Fig. 2, Fig. 4, Table 1). These Cyc2, MtoA, and MtrC clusters largely agree with FeGenie's HMM-based predicted distributions. Since MMSeqs2 generated two clusters of MtoA sequences, we sought to further verify the classifications. We constructed a tree of all Gallionellaceae MtoA sequences along with reference sequences of MtrA from iron-reducing bacteria (Fig. 5) (57). Although there is some separation of Cluster 335 and Cluster 451 MtoA sequences, many clades are not well defined or supported. In fact, backbone support throughout the tree is poor and the tree does not indicate a clear separation of the MtoA and MtrA sequences (Fig. 5).There is some evidence that the direction of electron flow through Mto/Mtr can be reversible (31, 58). So, it may be that the functions of MtoA and MtrA are interchangeable, and in fact they may be indistinguishable proteins that can conduct electrons across the outer membrane in either direction.

The decaheme cytochrome MtrC is the extracellular partner of the iron-reducing MtrAB complex of *Shewanella*. The MtrAB complex is a homolog of the MtoAB complex of FeOB, but MtrC was thought to be exclusive to iron-reducers because there is no MtrC homolog in *S. lithotrophicus* ES-1. However, we found seven MAGs within both *Gallionella* and *Sideroxydans* that encode MtrC (Table S3), suggesting it may also have a role in the energy metabolism of iron oxidizers.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

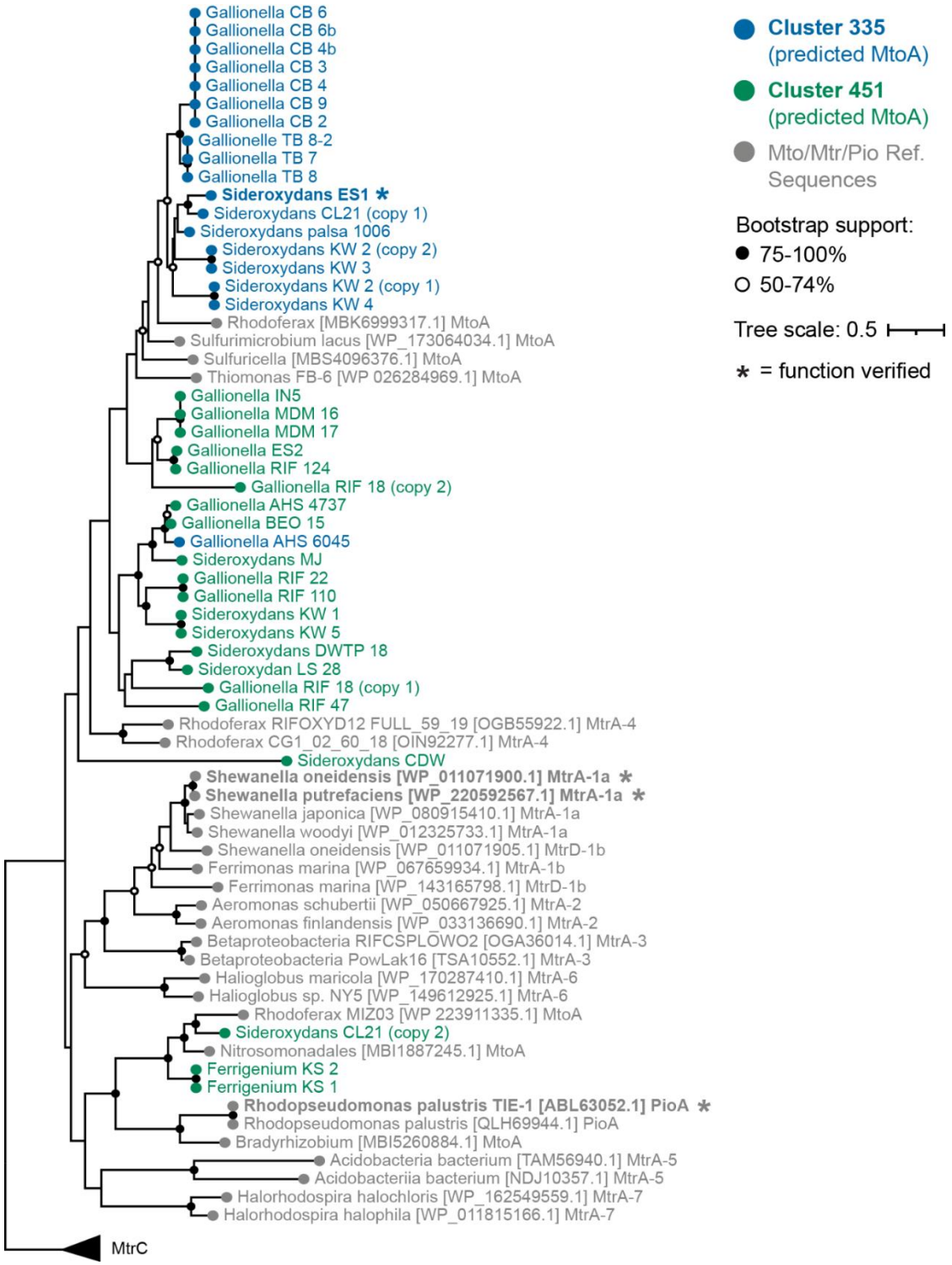



**FIGURE 5** Maximum likelihood tree of the predicted MtoA sequences identified in MMSeqs2 Cluster 335 and Cluster 451 along with MtoA reference sequences from NCBI and MtrA reference sequences from Baker, et al. 2022. Numbers (1a, 1b, 2, 3, 4, 5, 6, and 7) appended after

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Mtr denote reference sequences from the seven MtrA groups defined by Baker, et al., 2022. MtrA-4 indicates the Group 4 Betaproteobacteria. Tree is rooted using MtrC. Support is the result of 500 bootstrap replicates.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

**TABLE 1** Clusters of predicted *c*-type cytochromes and other heme-containing proteins of interest from MMSeqs2. Functional predictions are based on ‡ isolate annotations and NCBI BLAST or † BLAST of sequences from metagenomes in Uniprot.

| Cluster | Functional prediction | # CXXCH, $CX_3CH$, or $CX_4CH$ motifs per protein | # FeOB (of 87) |
|---|---|---|---|
| **Iron oxidation/reduction proteins** | | | |
| 313 | Iron oxidase Cyc2‡ | 1 | 70 |
| 451 | Decaheme c-type cytochrome, DmsE family, MtoA‡ | 10 | 19 |
| 335 | Decaheme c-type cytochrome, DmsE family, MtoA‡ | 10 | 17 |
| 50 | Decaheme c-type cytochrome, OmcA/MtrC family† | 10 | 7 |
| **Potential EET/EEU pathway proteins** | | | |
| 20 | Cytochrome C family protein; potential periplasmic PCC3 subunit‡ | 21, 24, 27 | 42 |
| 241 | Cytochrome C family protein; potential extracellular PCC3 subunit‡ | 10, 11, 12, 13, 14, 15, 16, 18 | 34 |
| 242 | Cytochrome C family protein; potential extracellular PCC3 subunit‡ | 26, 28, 29, 33, 35 | 7 |
| 331 | Cytochrome C family protein; potential extracellular PCC3 subunit‡ | 15, 17 | 5 |
| 65 | Doubled CXXCH motif-containing protein; Cytochrome c3 family protein†, potential UetJ subunit | 11, 12 | 6 |
| 479 | Tetraheme cytochrome - potential UetA subunit | 4 | 6 |
| 330 | Cytochrome C7 domain-containing protein; Triheme cytochrome - potential UetDEG subunit | 3 | 5 |
| 94 | Cytochrome C7 domain-containing protein; Triheme cytochrome - potential UetDEG subunit | 3 | 5 |
| 446 | Diheme cytochrome c‡ - potential Slit_1324 | 2 | 51 |
| **Sensory proteins** | | | |
| 152 | Methyl-accepting chemotaxis sensory transducer; YoaH ‡ | 1 | 54 |
| 40 | Methyl-accepting chemotaxis sensory transducer with Pas/Pac sensor; Aerotaxis receptor ‡ | 1 | 43 |
| 400 | Diguanylate cyclase with PAS/PAC sensor; Cyclic di-GMP phosphodiesterase Gmr ‡ | 1 | 36 |
| **Other** | | | |
| 433 | 2Fe-2S ferredoxin‡ | 1, 2 | 72 |
| 360 | 4Fe-4S ferredoxin iron-sulfur binding domain protein | 1 | 41 |
| 403 | Forkhead-associated protein‡ | 10 | 27 |
| 146 | Cytochrome c; Octaheme tetrathionate reductase† | 8 | 25 |
| 253 | Sulfite reductase, dissimilatory-type, subunit DsrJ† | 3 | 17 |

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Multiheme porin-cytochrome *c* complexes have been proposed to play roles in extracellular electron transport and/or metal oxidation because they provide a conduit for electrons to cross the outer membrane and participate in cellular metabolism (59, 60). One example is the PCC3 complex, identified through bioinformatic analyses of genomes of several FeOB including *S. lithotrophicus* ES-1, which contains a periplasmic MHC, an extracellular MHC, an outer membrane porin, and a conserved inner membrane protein (38). We identified 26 Gallionellaceae FeOB genomes with a complete predicted PCC3 complex, an additional 11 genomes with a partial complex, and four instances where a genome's PCC3 gene cluster encodes two predicted periplasmic cytochromes instead of one (Fig. 4, Table S3). The predicted periplasmic MHCs grouped in MMSeqs2 Cluster 20, while predicted extracellular MHCs grouped in Clusters 241, 242, and 331. The extracellular MHCs exhibited variability in the number of CXXCH heme motifs (10-35; Table 1), which suggests a range of functions in the extracellular PCC3 MHCs. Based on *in silico* protein structure models, PCC3 MHCs appear long and mostly linear (Fig. 6, Fig. S2), suggesting an extended conduction range both intra- and extracellularly.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

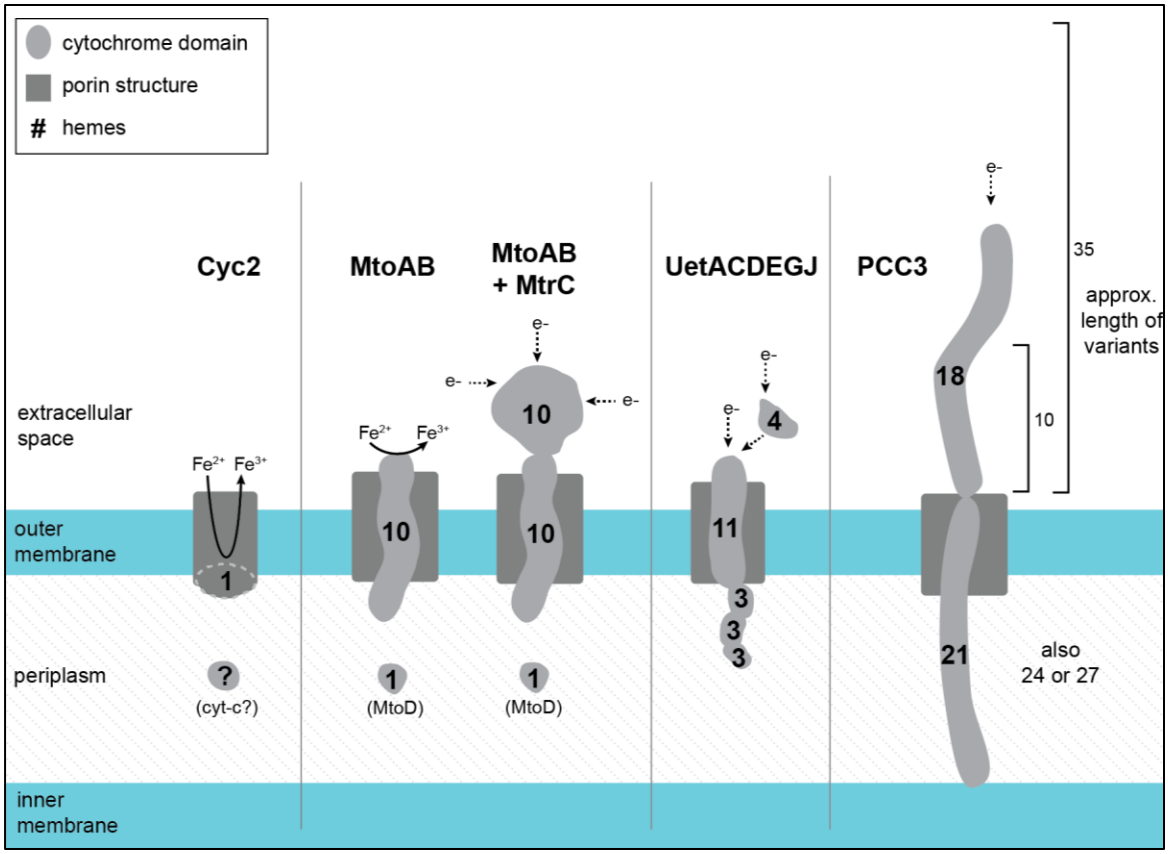



**FIGURE 6** Models of potential Gallionellaceae extracellular electron transfer mechanisms. All sizes are approximated. Dimensions of Cyc2 with its fused cytochrome-porin and the porin-cytochrome complexes MtoAB, MtoAB+MtrC, MtoD and Uet drawn from models and measurements in previous literature (32, 38, 61–63). Illustration of PCC3 is based on AlphaFold2 predictions (Fig. S2). The number of hemes and size of PCC3 can vary. The 21/18 heme complex of *S. lithotrophicus* ES-1 is depicted along with the estimated length of the 10 and 35 heme variants of the extracellular cytochrome.

Another recently described multiheme porin-cytochrome *c* complex is the undecaheme electron transfer (Uet) complex, found in the cathode-oxidizing Tenderiales (61) (Fig. 6). We used a combination of MMSeqs2 and BLAST to identify Uet genes in the Gallionellaceae. While PCC3 is more common to *Sideroxydans* (59%) than *Gallionella* (12%), the Uet pathway appears

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

exclusive to *Gallionella* and two unclassified outliers (Fig. 4). Six *Gallionella* have predicted undecaheme cytochrome (UetJ), extracellular tetraheme cytochrome (UetA), three predicted periplasmic triheme cytochromes (UetDEG), peptidylprolyl isomerase (UetB), and NHL repeat units (UetHI) (Fig. 4, Table S3). We checked for genes encoding the β-barrel porin UetC and found BLAST hits in four of the six genomes (Table S3).

*S. lithotrophicus* ES-1 has a set of periplasmic cytochrome genes without a predicted porin that were highly upregulated during growth on iron, and therefore thought to be involved in iron oxidation (27). The genes encode a cytochrome b (Slit_1321), a hypothetical extracellular protein (Slit_1322), a monoheme cytochrome class I (Slit_1323), a periplasmic diheme cytochrome (Slit_1324; Cluster 446 in Table 1), and a molecular chaperone Hsp33 (Slit_1325). We found homologs of the Slit_1321-1324 gene cluster are common and well-conserved among Gallionellaceae FeOB, present in 50 genomes (Fig. 4, Table S3). These genes may represent a mechanism of periplasmic electron transport, perhaps as part of an iron oxidation/extracellular electron uptake pathway.

**Electron transport chains**

We compared electron transport chain component genes of the iron and nitrite oxidizer groups and found them to be largely similar (Fig. 7). High-affinity *cbb*$_3$-type oxidases are common (Fig. 3), with most genomes containing either the proximal or distal form of *ccoN* (Fig. 2) (64). Even the four NRFeOB genomes contain *ccoNO* genes, indicating a potential for both oxygen and nitrate respiration. In contrast, few Gallionellaceae genomes contain *narGH* or *napAB* (6 and 10 genomes, respectively, with no overlap), indicating nitrate respiration is relatively rare overall (Fig. 3, Table S3).

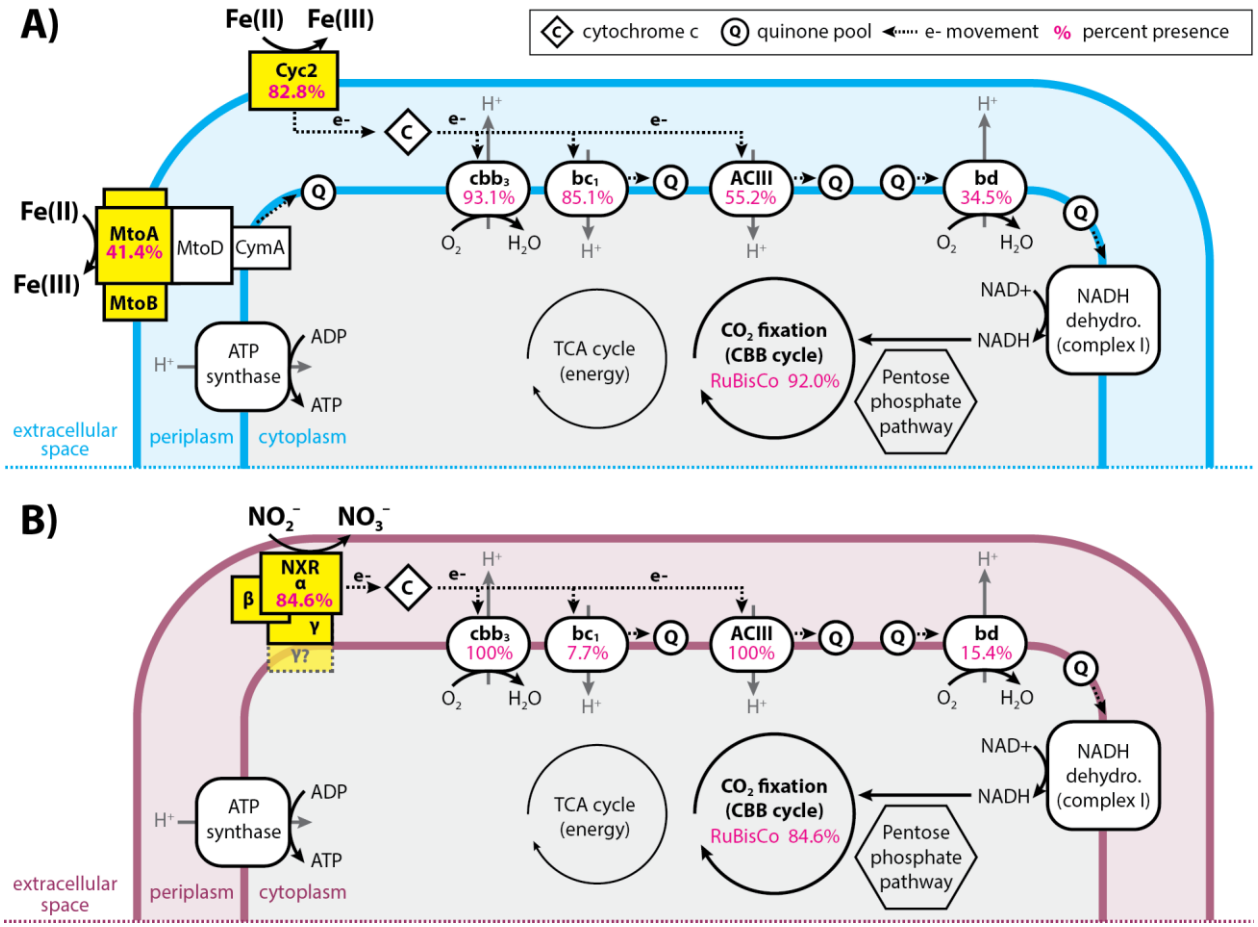
bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In addition to the *cbb*$_3$-type oxidase genes, 34.5% of iron oxidizers and 15.4% of nitrite oxidizers possess genes for cytochrome *bd*-type oxidases (*cydAB*) (Fig. 7). The presence of *bd*-type oxidase genes often overlaps with *cbb*$_3$-type oxidase genes (Table S3). Like *cbb*$_3$-type oxidases, cytochrome *bd*-type oxidases have a high affinity for oxygen and recent studies show they can be more highly expressed than *cbb*$_3$-type oxidases under low-oxygen, organic-rich conditions (65). Both FeOB and NOB have genes for cytochrome *bc*$_1$ and Alternative complex III (ACIII) quinol oxidase complexes. However, *bc*$_1$ is more common in FeOB (85.1%) compared to NOB (7.7%), while ACIII is more common in NOB (100%) than FeOB (55.2%) (Fig. 7, Table S3). Like the *bd*- and *cbb*$_3$-type oxidases, presence of *bc*$_1$ and ACIII often overlaps in a single organism, especially in FeOB (Table S3). Possessing both *bd*- and *cbb*$_3$-type oxidases, and/or having both *bc*$_1$ and ACIII contributes to flexibility within the electron transport chains of Gallionellaceae. The presence of various terminal oxidases implies adaptation to niches where oxygen and organic carbon availability differ or fluctuate.

**Carbon fixation**

Gallionellaceae isolates grow autotrophically. To determine if the capacity for autotrophic growth is widespread, we analyzed the pangenome for RuBisCo genes (*cbbLS*, *cbbMQ*). Most genomes in the dataset (>91%; 94 of 103 genomes) contain genes for either Form I or Form II RuBisCo (Fig. 3, Table S3). FeOB more commonly have Form II, while NOB only have Form I. Form II enzymes are adapted for medium to high $CO_2$ and low $O_2$ concentrations (66) and their predominance in FeOB may correspond to different oxygen niches of FeOB and NOB. The prevalence of RuBisCo genes indicates both iron- and nitrite-oxidizing Gallionellaceae have the capacity to grow autotrophically.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.




322 **FIGURE 7** Diagram showing the similarities and differences between the electron transport
323 chains of (A) iron- versus (B) nitrite-oxidizing Gallionellaceae. Pink numbers indicate the
324 percent of FeOB (A) or NOB (B) genomes that encoded each part of the electron transport chain
325 or RuBisCo.

326 **Auxiliary energy metabolisms**

327 Previous studies showed some Gallionellaceae FeOB possess alternate energy
328 metabolisms such as thiosulfate and lactate oxidation (27, 28). We searched the pangenome for
329 key genes of sulfur, manganese, and organic substrate oxidation pathways to determine how
330 common alternate metabolisms are among Gallionellaceae FeOB. Sulfide:quinone reductase

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(*sqr*) is common to both FeOB and NOB (Fig. 3, Table S3). Sqr can oxidize sulfide, transporting electrons to the quinone pool, although it may be a means of detoxification rather than energy conservation (67, 68). In contrast, both *soxABXYZ* and *dsrAB* are detected exclusively in the iron-oxidizing Gallionellaceae genomes (Fig. 3, Table S3). To predict the oxidative vs. reductive function of *dsrAB*, we constructed a tree using reference sequences from Loy, et al. (69, 70). Gallionellaceae sequences form a discrete clade within the sulfur-oxidizing group (Fig. S3), indicating the DsrAB of Gallionellaceae is likely a reverse dissimilatory sulfite reductase (rDSR). In contrast, the *Ca.* Nitrotoga genomes do not contain *dsr* or *sox* genes. Instead, *Ca.* Nitrotoga have *sorAB*, which may enable oxidation of sulfite to sulfate (Fig. 3). Together these results indicate that although sulfur oxidation is an accessory trait of both iron- and nitrite-oxidizing Gallionellaceae, only certain FeOB appear capable of oxidizing S(0) or thiosulfate.

We analyzed the pangenome for signs of organic utilization. Although not widely distributed, the most common genes were for lactate utilization (*lutABCP*) and sugar transport (*msmX*, *gtsABC*). Only eight *Gallionella* and five *Sideroxydans* genomes, including *Sideroxydans* sp. CL21, have *lutABC* along with the *lutP* lactate permease gene (Fig. 2, Fig. 3, Table S3). Likewise, only six genomes contain *gtsABC* genes for glucose/mannose uptake. None of the NOB contain the *lut* or *gts* genes for organic utilization.

We used BLAST to evaluate the Gallionellaceae genomes for manganese oxidase genes *mcoA*, *moxA*, *mofA*, and *mnxG*. There are a few hits for *mcoA*, *moxA*, and *mofA* genes, but none for *mnxG* (Table S3). Since manganese oxidation activity has not been shown in any of the Gallionellaceae isolates, additional verification is needed to determine whether the genes identified by BLAST are truly Mn oxidases.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

**Other genes distinct to FeOB, potentially related to iron oxidation**

We searched the pangenome for the candidate genes for stalk formation (*sfz/sfb*) identified in the stalk-forming *Ferriphaselus* and Zetaproteobacteria isolates (71, 72). Twelve genomes, restricted to *Gallionella* (9) and *Ferriphaselus* (3) contain the four *sfz/sfb* genes (Fig. 2, Table S3). Thus far, all cultured Gallionellaceae stalk formers belong to these two genera, suggesting stalk formation may be limited and not a trait of *Sideroxydans*.

Using the Anvi'o subset of only genomes >97% complete, we identified several gene clusters that were present and abundant only in *Gallionella* and *Sideroxydans*, but lacked prior connection to an iron-oxidizing lifestyle. These included distinct gene clusters with COG functional annotations for: Cell Wall/Membrane/Envelope Biogenesis, Cytoskeleton formation, Signal Transduction Mechanisms, and Energy Production and Conversion (Table 2, Table S4). Clusters for Cell Wall/Membrane/Envelope Biogenesis may indicate FeOB have specific adaptations for housing *c*-type cytochromes and EET mechanisms in the outer membrane, or to avoid encrustation by iron oxides. Clusters for Energy Production and Conversion included ferredoxin (Fdx) and subunits of the RnfABCDEG complex. The Rnf complex was originally discovered for its role in N fixation, in which it oxidizes NADH and generates reduced ferredoxin that donates electrons to nitrogenase (73). More recent studies have shown Rnf complexes can conserve energy under anaerobic conditions (74–76) and, as a low potential electron donor, ferredoxin can transfer electrons to many metabolic pathways including some that produce secondary metabolites (77). Not all Gallionellaceae with Rnf complex genes have *nifDHK* nitrogenase genes, implying Gallionellaceae Rnf and ferredoxin have functions beyond N fixation. Although their specific function in Gallionellaceae FeOB are unknown, their ubiquity implies utility for FeOB and an area for additional research.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

**TABLE 2** Gene clusters of interest from the Anvi'o pangenome subset that were present in iron-oxidizing *Gallionella* and *Sideroxydans*, but absent in nitrite-oxidizing *Ca.* Nitrotoga.

| COG20 Category | COG20 Function | Gene Cluster ID |
|---|---|---|
| Cell wall/ membrane/ envelope biogenesis | Lipid carrier protein ElyC involved in cell wall biogenesis, DUF218 family (ElyC) | GC_00001120 |
| | ABC-type lipoprotein export system, ATPase component (LolD) | GC_00000969 |
| | ADP-heptose synthase, bifunctional sugar kinase/ adenylyltransferase (RfaE) | GC_00001059, GC_00001084 |
| | ADP-heptose:LPS heptosyltransferase (RfaF) | GC_00001100 |
| | Glycosyltransferase involved in cell wall biosynthesis (RfaB) | GC_00001179 |
| | Outer membrane protein TolC | GC_00000022, GC_00000920 |
| | Glutamate racemase (MurI) | GC_00001047 |
| | Murein L,D-transpeptidase YafK | GC_00001108 |
| Cytoskeleton | Cytoskeletal protein CcmA, bactofilin family | GC_00000987 |
| Energy production and conversion | Na+ translocating ferredoxin: NAD+ oxidoreductase RNF, RnfA | GC_00000042 |
| | Na+ translocating ferredoxin: NAD+ oxidoreductase RNF, RnfB | GC_00001082 |
| | Na+ translocating ferredoxin: NAD+ oxidoreductase RNF, RnfC | GC_00001069 |
| | Na+ translocating ferredoxin: NAD+ oxidoreductase RNF, RnfD | GC_00001055 |
| | Na+ translocating ferredoxin: NAD+ oxidoreductase RNF, RnfE | GC_00001071 |
| | Na+ translocating ferredoxin: NAD+ oxidoreductase RNF, RnfG | GC_00001096 |
| | Ferredoxin (Fdx) | GC_00001052 |
| | Cytochrome c-type biogenesis protein CcmH/NrfF | GC_00001058 |
| | Cytochrome c-type biogenesis protein CcmH/NrfG | GC_00001078 |
| Signal transduction mechanisms | PAS domain \| GAF domain \| HAMP domain \| Cyclic di-GMP metabolism protein | GC_00000006 |
| | cAMP-binding domain of CRP or a regulatory subunit of cAMP-dependent protein kinases \| Small-conductance mechanosensitive channel MscK | GC_00001152 |

# Discussion

The Gallionellaceae family is historically known for its iron-oxidizing members, but recently a new candidate genus of nitrite oxidizers, *Ca.* Nitrotoga, was identified (39). Comparing their genomes to those of FeOB genera has helped identify genes and pathways

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

related to iron oxidation since *Ca.* Nitrotoga isolates have no documented capacity for that metabolism (39, 40, 42, 43, 45). We resolved the phylogeny of the Gallionellaceae and verified *Ca.* Nitrotoga lacked iron oxidation marker genes. Given separate groups of FeOB and NOB, we used a pangenomic approach to identify shared features of the Gallionellaceae, as well as FeOB-specific genes that may represent novel iron oxidation pathways.

## Phylogeny

The Gallionellaceae is composed of four genera: *Gallionella*, *Sideroxydans*, *Ferriphaselus*, and *Ca.* Nitrotoga, based on a concatenated ribosomal protein tree. In comparison, 16S rRNA phylogeny did a poorer job of resolving these genera, so 16S-based identification should be considered tentative, pending availability of genomes. To facilitate consistent classification, the protein sequences and alignments used here (Fig. 1) are available at (https://figshare.com/projects/Gallionellaceae_Ribosomal_Proteins_for_Concatenated_Tree/157347).

The new phylogeny provides a framework for understanding the diversity and major metabolisms of the Gallionellaceae. They are members of Nitrosomonadales, which contain many chemolithotrophic S and N oxidizers. Like their closest relatives, the Sulfuricellaceae (78), many Gallionellaceae retain the ability to oxidize sulfur (Fig. 3, Fig. S3). The Gallionellaceae tree (Fig. 1) shows a deeply branching split between genera, with each of the two major genera, *Gallionella* and *Sideroxydans*, containing a continuum of diversity. Within the *Gallionella*, the isolates *G. capsiferriformans* ES-2 and *Ferrigenium kumadai* An22 bracket the *Gallionella*, with An22 deeply branching and ES-2 at the crown. *F. kumadai* An22 was originally classified as *Ferrigenium* based on 16S rRNA distance (25). However, our analyses do not show any clear phylogenetic clustering or functional distinction, with which we could draw a line between

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

*Gallionella* and *Ferrigenium*. Moreover, the tree topology suggests continued diversification within both *Gallionella* and *Sideroxydans* largely without the formation of subclades that represent distinct niches. There is one subclade of *Sideroxydans* that corresponds to the GTDB genus level designation PALSA-1006 (Fig. 1). However, ANI/AAI results (Table S5) indicate there is not enough diversity within the Gallionellaceae to justify further splitting the four major genera any further. Additionally, we did not detect any obvious functional difference in PALSA-1006. Given our phylogenetic analysis, ANI/AAI, and similar functional profiles, we recommend keeping them within *Sideroxydans*. Based on the above classification scheme, most of the genomes (84 of 103) fall into either *Gallionella* or *Sideroxydans*.

Phylogenetic diversity corresponds to functional diversity that can drive Gallionellaceae success in a variety of environments. Many *Gallionella* and *Sideroxydans* do not appear to be obligate iron oxidizers, and some may not be obligate aerobes. Auxiliary metabolisms for S, N, and C are present to varying degrees throughout the iron-oxidizing genera and are not associated with specific sub-groups. Some FeOB from organic-rich environments, such as *Sideroxydans* sp. CL21, have genes for organoheterotrophy. Other FeOB show metabolic flexibility in additional lithotrophic metabolisms, such as oxidation of S or potentially Mn, elements that often co-occur with Fe in the environment. Some Gallionellaceae may also thrive in oxygen-poor environments by reducing nitrate, although this capability appears rare. Such traits contribute to diversity in the Gallionellaceae FeOB genera, which appear to acquire and/or retain additional energy and nutrient metabolisms to adapt to a range of environments.

*Ca.* Nitrotoga stands out as an exception within the Gallionellaceae. The pangenome analysis shows that *Ca.* Nitrotoga have distinctive genomic content (Fig. S4). They do not appear to have the capacity for iron oxidation based on available physiological evidence and the

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

genomic analyses presented here. The similarities in Gallionellaceae FeOB and *Ca*. Nitrotoga electron transport chains enable them to meet the shared challenge of conserving energy from high-potential electron donors. However, *Ca*. Nitrotoga are a distinct clade that appear to have evolved from the FeOB to occupy a nitrite oxidation niche.

## Iron oxidation and extracellular electron uptake mechanisms

The Gallionellaceae FeOB genomes encode a wide variety of predicted *c*-type cytochromes. Of these cytochromes, many appear to be associated with the outer membrane, implying a role in extracellular electron transport. Cyc2 is present in the majority of Gallionellaceae FeOB genomes, while multiheme cytochromes (MHC) Mto/Mtr, Uet, and PCC3 are less common, each with different distribution patterns (Fig. 4), suggesting the different cytochromes play distinct roles.

Cyc2 has been shown to oxidize dissolved Fe(II) (27, 32, 37). The monoheme Cyc2 is a small fused cytochrome-porin and since aqueous $Fe^{2+}$ is common to many redox transition zones, it makes sense that most FeOB would retain and use the simplest tool. But in Earth's various environments, iron is largely available as minerals (clays, oxides, sulfides) and also bound to organics (e.g. humic substances). The decaheme MtoA has been shown to play roles in the oxidation of mineral-bound Fe(II), specifically Fe(II) smectite clay (37). As a MHC, MtoA may have multiple benefits that help in oxidizing minerals. MtoA has a large redox potential window (-350 to +30 mV; (31, 60)), which could help with oxidation of solids, like smectite, that also have a range of redox potentials (e.g., −600 to +0 mV for SWa-1 vs. −400 to +400 mV for SWy-2; (79)), which change as mineral-bound iron is oxidized or reduced. Assuming the MtoA structure is similar to MtrA, the ten hemes span the membrane, making a wire that conducts from extracellular substrates to periplasmic proteins (62, 80). The multiple hemes allow for

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

transfer of multiple electrons at a time (59). Some MAGs with *mtoAB* also encode the extracellular decaheme cytochrome MtrC. In *Shewanella,* the MtrCAB complex requires MtrC to reduce solid minerals (ferrihydrite), while MtrAB alone can only reduce dissolved Fe(III) and electrodes (81–83). Likewise, Gallionellaceae MtrC may help increase interactions with different minerals. Some Gallionellaceae FeOB may retain genes for both Cyc2 and MtoAB (with or without MtrC) to oxidize different Fe(II) substrates in their environments.

Like MtrCAB, the predicted PCC3 complex includes both periplasmic and extracellular MHCs and a porin. A key difference is that the PCC3 cytochromes often have more hemes than MtoA/MtrA and MtrC. The greater number of hemes may serve to store electrons, as in a capacitor. They may also conduct across a greater distance; the PCC3 periplasmic MHC, with 21-27 hemes, is potentially long enough to span the entire periplasm (as noted by Edwards et al., (84)). Intraprotein electron transfer between hemes is rapid (85–87); therefore the periplasm-spanning MHC of PCC3 may allow for faster electron transfer compared to complexes containing smaller periplasmic cytochromes like the monoheme MtoD. The extracellular PCC3 MHC contains between 10 and 35 hemes, which could extend further from the outer membrane compared to MtrC. Not only would this extend the range of electron transfer, but may also be faster than a “wire” of smaller cytochromes (e.g. *Geobacter* hexaheme OmcS (88)). Increasing oxidation rates via larger MHCs would allow FeOB to oxidize substrates faster. Given that Fe(II) is subject to abiotic oxidation under certain conditions and other organisms may compete for EEU, such kinetic advantages would give FeOB a competitive edge.

## Conclusions

Gallionellaceae, specifically *Gallionella*, is best known for lithoautotrophically oxidizing iron to make mineral stalks that come together to form microbial mats at groundwater seeps (18,

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

89, 90). Although this may contribute to an impression that the niche was relatively restricted, 16S rRNA sequencing of cultures and environmental samples has revealed both the diversity of Gallionellaceae as well as its prevalence across practically any freshwater and some brackish environments where Fe(II) and $O_2$ meet. The pangenome shows that Gallionellaceae possess metabolic flexibility to use non-iron substrates, notably sulfur, and the MHCs likely also confer further metabolic capabilities that may help them occupy a range of different iron- and mineral-rich niches. Gallionellaceae thrive in aquifers, soil, and wetlands, all of which have substantial mineral content. Thus, the widespread ecological success of Gallionellaceae may well correspond to genomes that encode a range of iron oxidation mechanisms as well as adaptations for varied environments.

It is becoming clear that there are multiple ways to oxidize iron, though we have varying levels of evidence for gene/protein function (60, 91, 92). Validating iron oxidation genes/proteins is painstaking work due to challenging cultures, low yield, few genetic systems, and the fact that iron interferes with many molecular extractions and assays. And yet, there are likely even more iron oxidation mechanisms, so we need to be strategic about choosing genes/proteins for deeper characterization. Our pangenome analysis gives a wider view of the distribution and frequency of potentially novel iron oxidation genes, which will help us to prioritize investigations. Furthermore, the varied outer membrane-associated cytochromes inspire us to investigate relationships between structure and function. Why are there so many different multiheme cytochromes? Is there substrate specificity, kinetic advantages, battery-like functions, or some utility we have yet to consider? Addressing these questions will help us understand how these proteins and pathways shape microbial transformations of varied Earth materials.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

# Methods

## Data collection and curation

Gallionellaceae genomes were collected from the National Center for Biotechnology Information (NCBI) Entrez database (93), the Joint Genome Institute Integrated Microbial Genomes (JGI IMG) database (94), and the European Nucleotide Archive (ENA) at EMBL-EBI database (*Sideroxydans* sp. CL21, *Ca.* Nitrotoga fabula KNB, and the "IN" MAGs (17, 43, 95)) (Table S6). We also received non-public genomes from the Ménez Lab at the Université de Paris (3 genomes reconstructed by Aurélien Lecoeuvre from the Carbfix study in Hengill, Iceland (2); metagenomes available at Sequence Read Archive SRR3731039, SRR3731040, SRR4188484, and SRR4188643), and the Banfield Lab at the University of California, Berkeley (3 genomes reconstructed by Alex Probst from Crystal Geyser in Utah, USA (96)) (Table S6). This initial 230-genome dataset included isolate genomes, metagenome assembled genomes (MAGs), and single-cell amplified genomes (SAGs) that were taxonomically classified as members of the Gallionellales order; Gallionellaceae family; or the *Gallionella*, *Sideroxydans*, *Ferriphaselus*, *Ferrigenium*, or *Ca.* Nitrotoga genera in their respective databases. Duplicate genomes were identified and removed if they had identical accession numbers or their average nucleotide identities (ANI) were 100%. CheckM v1.1.2 (97) was used to assess genome quality. Genomes with lower than 80% completeness and greater than 7% contamination were removed from the dataset. The final filtered dataset, referred to as "the Gallionellaceae" or "the dataset" contained 103 genomes, including six of the Gallionellaceae FeOB isolates (Table S1). The seventh isolate, *Sideroxyarcus emersonii* (26), was not published at the time of our main analysis, but a

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

supplemental of its key metabolic genes and MHCs (Table S7) shows it has similar patterns to *Sideroxydans*.

**Naming conventions**

To assign simple, unique names to the metagenomes, codes were appended to genus-level names based on sample location and bin IDs (Table S1, Table S6, Table S8). Isolates retained their own unique names. Organisms that were taxonomically classified in their original databases at the family Gallionellaceae or order Gallionellales were, if possible, classified at lower taxonomic levels using a combination of AAI, 16S rRNA (if available), classification through the Genome Taxonomy Database Toolkit (GTDB-Tk) (98), and placement in the concatenated ribosomal protein tree (Fig. 1 and Fig. 2).

**Ecosystem classifications**

To assess whether metabolic diversity correlated to ecosystem type, each genome was assigned to an ecosystem based on the Genomes OnLine Database (GOLD) (99) schema which leverages Environmental Ontology (EnvO) classifications (100). A genome's pre-existing classification from IMG was used if available. Genomes without prior classification were categorized based on published descriptions of their sample sites and “habitat” information listed in their database of origin. Based on the GOLD classifications (Table S2), genomes were examined for patterns of correspondence between ecosystems and phylogenetic and/or metabolic diversity.

## Calculation of average amino acid and nucleotide identities

Average amino acid identity (AAI) and average nucleotide identities (ANI) were computed to assess the similarity of genomes in the curated data set (Table S5). AAI was

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

calculated using CompareM (101). ANI was calculated using FastANI in Kbase (102). Final AAI and ANI tables were formatted using Microsoft Excel.

## Tree construction

### Concatenated ribosomal protein tree

A concatenated tree of ribosomal proteins (Fig. 1) was constructed to determine the phylogenetic relationships of genomes in the Gallionellaceae dataset. Two *Sulfuricella* genomes, *Sulfuricella* sp. T08 and *Sulfuricella* 3300027815, where included as an outgroup to root the tree. Use of a *Sulfuricella* outgroup was based on previous literature (103, 104), which identified *Sulfuricella* and other members of the Sulfuricellaceae family as near neighbors of Gallionellaceae. The concatenated sequences were composed of 13 small and large ribosomal proteins (L19, L20, L28, L17, L9_C, S16, L21p, L27, L35p, S11, S20p, S6, S9) present in 94 or more of the 105 genomes including the outgroup. Protein sequences were aligned in Geneious v.10.2.6 (105) using MUSCLE (106). Ends of the alignments were manually trimmed and regions with over 70% gaps were masked, after which sequences were concatenated. The tree was constructed using RAxML-NG v1.0.3 (107) with the maximum likelihood method, LG+G model, and 1000 bootstraps. The final tree was visualized and annotated with iTOL (108).

### 16S rRNA gene tree

We constructed a 16S rRNA gene tree (Fig. S1) composed of sequences from our dataset combined with a selection of sequences from the SILVA database to determine how well 16S rRNA resolves Gallionellaceae phylogeny compared to the concatenated ribosomal protein tree. Full length (~1500 bp) 16S rRNA genes were retrieved from 24 of the Gallionellaceae genomes using Anvio's 'anvi-get-sequences-for-hmm-hits' command for "Ribosomal_RNA_16S." These

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

genes were aligned in SINA (109) along with Gallionellaceae sequences from the Silva database (110) that had >1475 bp and >85-90 sequence quality score. The outgroup is composed of *Thiobacillus*, *Ferritrophicum*, *Sulfuricella*, *Sulfuriferula*, and *Nitrosomonas* sequences acquired from the Silva database. The final alignment contained 965 non-redundant sequences and alignment length was 1500 positions after trimming and masking all sequence gaps greater than 70%. A maximum likelihood tree was constructed using RAxML-NG v1.0.3 (107) with the GTR+G model and 300 bootstraps. Family and genus level classifications from the SILVA database were used to annotate the tree in Iroki (111).

### Individual protein trees

Trees for DsrAB (Fig. S3) and Mto/Mtr (Fig. 5) were constructed from Gallionellaceae protein sequences along with reference sequences from NCBI, Loy, et al. and Baker, et al. (57, 69). Sequences were aligned with MUSCLE (106), ends were manually trimmed, and regions with over 70% sequence gaps were masked in Geneious v.10.2.6 (105). For the Dsr tree, DsrA and DsrB sequences were concatenated. Trees were constructed using RAxML-NG v1.0.3 (107) with the LG+G model. Branch support for Mto/Mtr tree is based on 500 bootstraps and support for the DsrAB tree is based on 300 bootstraps. Final trees were visualized and annotated with Iroki (111).

## Pangenome analysis

### Metabolic gene analysis

We used the Distilled and Refined Annotation of Metabolism (DRAM) v0.0.2 (48) within KBase (102), LithoGenie within MagicLamp (50), and FeGenie (49) to identify key metabolic genes indicative of various oxidation, respiration, and carbon utilization pathways. NCBI

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

BLAST+ (53) was used to identify additional genes for eNOR, cNOR, SorAB, Mn oxidases, LutABCP, and stalk formation. We then analyzed the presence/absence of the metabolic genes and looked for patterns across the concatenated protein tree, between genera, and between FeOB versus NOB.

**Multiheme cytochrome analysis**

To identify potential *c*-type cytochromes we used a modified heme counter script (54) to search for CXXCH, CXXXCH and CXXXXCH motifs within the protein sequences of each genome. The search identified 5,929 protein sequences with one or more $CX_{2\text{-}4}CH$-motifs. To determine which protein sequences were shared between genomes, sequences were clustered using MMSeqs2 (112) with coverage mode 0 for bidirectional coverage of at least 80% of the query and target sequences. Several clusters of interest were identified based on either the number of $CX_{2\text{-}4}CH$-motifs in each sequence or the relative abundance of FeOB sequences in the cluster. Querying with BLASTp (52) against the Uniprot (113) database was used to classify sequences from clusters of interest thereby identifying clusters of predicted *c*-type cytochromes. Isolate sequences were used as representative sequences for cluster classification. If a cluster did not contain an isolate sequence, a consensus classification was used. The subcellular localization of proteins was predicted using a combination of PSORTb v3.0.3 (114) and LocTree3 (115).

Some MHCs were predicted to be part of Mto, PCC3, or Uet porin-cytochrome complexes. Therefore, we wanted to determine if the genes for these MHCs were colocalized in their respective genomes with genes for β-barrel porins, periplasmic proteins, and inner membrane proteins previously identified in the literature (38, 61). We searched for the associated genes using BLASTp and amino acid reference sequences from *S. lithotrophicus* ES-1 (MtoB, MtoD, CymA), *Gallionella* AHS-4737 (MtoC), and *Ca.* Tenderia electrophaga

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(UetBCDEFGHI). The locus tags of BLASTp hits were then compared to locus tags of the MHCs to evaluate synteny and colocalization. The same method was used to determine if diheme *c*-type cytochromes from MMseqs2 cluster 446 which includes Slit_1324 were colocalized with a cytochrome b (Slit_1321), hypothetical extracellular protein (Slit_1322), monoheme cytochrome class I (Slit_1323), and molecular chaperone Hsp33 (Slit_1325).

**PCC3 modeling**

To model predicted PCC3 proteins, we used ColabFold: AlphaFold2 using MMseqs2 (116). Setting included using MSA mode "MMseqs2 (UniRef+environmental)," pair mode "unpaired+paired," protein structure prediction with "AlphaFold2-ptm," and complex prediction with "AlphaFold-multimer-v2" (117, 118). The best scoring model was rendered in PyMol v2.5.4 (119).

**Anvi'o subset analysis**

We used the Anvi'o v7 (54, 56) to build a pangenome database of all *Gallionella* (16), *Sideroxydans* (15), and *Ca.* Nitrotoga (6) genomes that were over 97% complete (Fig. S4) to analyze for additional genes important to FeOB lifestyles. Genes were clustered within the Anvi'o pangenome using a min-bit parameter of 0.5 and an mcl inflation parameter of 2. The Anvi'o pangenome was used to compare gene clusters across the dataset and to bin: 1) near-core (found in >85% of genomes), 2) accessory (found in >1 but <85% of genomes), and 3) strain specific (found in a single genome) sets of gene clusters. Gene annotations were assigned in Anvi'o using Prodigal (120) and functional annotations for Anvi'o gene clusters were assigned using the NCBI's Database of Clusters of Orthologous Genes (COGs) (121, 122). Data tables of the binned Anvi'o gene clusters were analyzed to identify gene clusters found in the near-core genomes of *Gallionella* and *Sideroxydans* but absent in *Ca.* Nitrotoga.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

# Acknowledgements

This research was funded by the National Science Foundation (EAR-1833525 to C.S.C. and S.W.P., MCB-1817651 to C.S.C.) and the Office of Naval Research (N00014-17-1-2640 C.S.C.). R. H. was also supported by fellowships from University of Delaware Graduate College and the Microbiology Program/Unidel Foundation. Support from the University of Delaware Center for Bioinformatics and Computational Biology Core Facility (RRID:SCR_017696) including use of the BIOMIX compute cluster was made possible through funding from Delaware INBRE (NIH P20GM103446), the State of Delaware, and the Delaware Biotechnology Institute.

We thank Aurélien Lecoeuvre, Emmanuelle Gérard, Bénédicte Ménez, Alex Probst, and Jill Banfield for allowing us to use unpublished genomes from their private collections; and all those who granted us permission to use their publically available, unpublished genomes from the NCBI and IMG databases (Table S1, Table S6).

We declare that we have no conflicts of interest.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

# References

1. Emerson D, Fleming EJ, McBeth JM. 2010. Iron-Oxidizing Bacteria: An Environmental and Genomic Perspective. Annu Rev Microbiol 64:561–583.

2. Trias R, Ménez B, le Campion P, Zivanovic Y, Lecourt L, Lecoeuvre A, Schmitt-Kopplin P, Uhl J, Gislason SR, Alfreðsson HA, Mesfin KG, Snæbjörnsdóttir SÓ, Aradóttir ES, Gunnarsson I, Matter JM, Stute M, Oelkers EH, Gérard E. 2017. High reactivity of deep biota under anthropogenic CO2 injection into basalt. Nat Commun 8:1063.

3. Jewell TNM, Karaoz U, Brodie EL, Williams KH, Beller HR. 2016. Metatranscriptomic evidence of pervasive and diverse chemolithoautotrophy relevant to C, S, N and Fe cycling in a shallow alluvial aquifer. 9. The ISME Journal 10:2106–2117.

4. Emerson D, Moyer CL. 2002. Neutrophilic Fe-Oxidizing Bacteria Are Abundant at the Loihi Seamount Hydrothermal Vents and Play a Major Role in Fe Oxide Deposition. Applied and Environmental Microbiology 68:3085–3093.

5. Emerson D, Vet WD. 2015. The Role of FeOB in Engineered Water Ecosystems: A Review. Journal AWWA 107:E47–E57.

6. Wang J, Sickinger M, Ciobota V, Herrmann M, Rasch H, Rösch P, Popp J, Küsel K. 2014. Revealing the microbial community structure of clogging materials in dewatering wells differing in physico-chemical parameters in an open-cast mining area. Water Research 63:222–233.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

7. Kato S, Krepski S, Chan C, Itoh T, Ohkuma M. 2014. Ferriphaselus amnicola gen. nov., sp. nov., a neutrophilic, stalk-forming, iron-oxidizing bacterium isolated from an iron-rich groundwater seep. International Journal of Systematic and Evolutionary Microbiology 64:921–925.

8. Fabisch M, Beulig F, Akob DM, Küsel K. 2013. Surprising abundance of Gallionella-related iron oxidizers in creek sediments at pH 4.4 or at high heavy metal concentrations. Front Microbiol 4.

9. Fabisch M, Freyer G, Johnson CA, Büchel G, Akob DM, Neu TR, Küsel K. 2016. Dominance of ‘Gallionella capsiferriformans’ and heavy metal association with Gallionella-like stalks in metal-rich pH 6 mine water discharge. Geobiology 14:68–90.

10. Weiss JV, Rentz JA, Plaia T, Neubauer SC, Merrill-Floyd M, Lilburn T, Bradburne C, Megonigal JP, Emerson D. 2007. Characterization of Neutrophilic Fe(II)-Oxidizing Bacteria Isolated from the Rhizosphere of Wetland Plants and Description of Ferritrophicum radicicola gen. nov. sp. nov., and Sideroxydans paludicola sp. nov. Geomicrobiology Journal 24:559–570.

11. Lüdecke C, Reiche M, Eusterhues K, Nietzsche S, Küsel K. 2010. Acid-tolerant microaerophilic Fe(II)-oxidizing bacteria promote Fe(III)-accumulation in a fen: Acid-tolerant Fe(II)-oxidizers in a fen. Environmental Microbiology 12:2814–2825.

12. Buongiorno J, Herbert LC, Wehrmann LM, Michaud AB, Laufer K, Røy H, Jørgensen BB, Szynkiewicz A, Faiia A, Yeager KM, Schindler K, Lloyd KG. 2019. Complex Microbial

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Communities Drive Iron and Sulfur Cycling in Arctic Fjord Sediments. Appl Environ Microbiol 85:e00949-19, /aem/85/14/AEM.00949-19.atom.

13. Patzner MS, Logan M, McKenna AM, Young RB, Zhou Z, Joss H, Mueller CW, Hoeschen C, Scholten T, Straub D, Kleindienst S, Borch T, Kappler A, Bryce C. 2022. Microbial iron cycling during palsa hillslope collapse promotes greenhouse gas emissions before complete permafrost thaw. 1. Commun Earth Environ 3:1–14.

14. Neubauer SC, Emerson D, Megonigal JP. 2002. Life at the Energetic Edge: Kinetics of Circumneutral Iron Oxidation by Lithotrophic Iron-Oxidizing Bacteria Isolated from the Wetland-Plant Rhizosphere. Applied and Environmental Microbiology 68:3988–3995.

15. de Vet W w. j. m., Dinkla I j. t., Abbas B a., Rietveld L c., van Loosdrecht M c. m. 2012. Gallionella spp. in trickling filtration of subsurface aerated and natural groundwater. Biotechnology and Bioengineering 109:904–912.

16. Eichinger S, Boch R, Leis A, Koraimann G, Grengg C, Domberger G, Nachtnebel M, Schwab C, Dietzel M. 2020. Scale deposits in tunnel drainage systems – A study on fabrics and formation mechanisms. Science of The Total Environment 718:137140.

17. Bethencourt L, Bochet O, Farasin J, Aquilina L, Borgne TL, Quaiser A, Biget M, Michon-Coudouel S, Labasque T, Dufresne A. 2020. Genome reconstruction reveals distinct assemblages of Gallionellaceae in surface and subsurface redox transition zones. FEMS Microbiol Ecol 96.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

18. Emerson D, Revsbech NP. 1994. Investigation of an Iron-Oxidizing Microbial Mat Community Located near Aarhus, Denmark: Field Studies. Applied and Environmental Microbiology 60:4022–4031.

19. Kappler A, Bryce C. 2017. Cryptic biogeochemical cycles: unravelling hidden redox reactions. Environmental Microbiology 19:842–846.

20. Ehrenberg C. 1838. Die infusionsthierchen Als vollkommene organismen. Leipzig: Leopold Voss.

21. Ghiorse WC. 1984. Biology of iron- and manganese-depositing bacteria. Annual Review of Microbiology 38:515–550.

22. Hallbeck L, Pedersen K. 1991. Autotrophic and mixotrophic growth of Gallionella ferruginea. Microbiology, 137:2657–2661.

23. Krepski ST, Hanson TE, Chan CS. 2012. Isolation and characterization of a novel biomineral stalk-forming iron-oxidizing bacterium from a circumneutral groundwater seep: A novel stalk-forming Fe-oxidizing bacterium. Environmental Microbiology 14:1671–1680.

24. Emerson D, Moyer C. 1997. Isolation and Characterization of Novel Iron-Oxidizing Bacteria That Grow at Circumneutral pH. Applied and Environmental Microbiology 63:9.

25. Khalifa A, Nakasuji Y, Saka N, Honjo H, Asakawa S, Watanabe T. 2018. Ferrigenium kumadai gen. nov., sp. nov., a microaerophilic iron-oxidizing bacterium isolated from a

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

paddy field soil. International Journal of Systematic and Evolutionary Microbiology 68:2587–2592.

26. Kato S, Itoh T, Iino T, Ohkuma M. 2022. Sideroxyarcus emersonii gen. nov. sp. nov., a neutrophilic, microaerobic iron- and thiosulfate-oxidizing bacterium isolated from iron-rich wetland sediment. International Journal of Systematic and Evolutionary Microbiology 72.

27. Zhou N, Keffer JL, Polson SW, Chan CS. 2022. Unraveling Fe(II)-Oxidizing Mechanisms in a Facultative Fe(II) Oxidizer, Sideroxydans lithotrophicus Strain ES-1, via Culturing, Transcriptomics, and Reverse Transcription-Quantitative PCR. Applied and Environmental Microbiology 88:e01595-21.

28. Cooper RE, Finck J, Chan C, Küsel K. 2023. Mixotrophy broadens the ecological niche range of the iron oxidizer Sideroxydans sp. CL21 isolated from an iron-rich peatland. FEMS Microbiology Ecology fiac156.

29. Huang Y-M, Jakus N, Straub D, Konstantinidis KT, Blackwell N, Kappler A, Kleindienst S. 2022. ‘Candidatus ferrigenium straubiae’ sp. nov., ‘Candidatus ferrigenium bremense’ sp. nov., ‘Candidatus ferrigenium altingense’ sp. nov., are autotrophic Fe(II)-oxidizing bacteria of the family Gallionellaceae. Systematic and Applied Microbiology 45:126306.

30. Straub KL, Benz M, Schink B, Widdel F. 1996. Anaerobic, nitrate-dependent microbial oxidation of ferrous iron. Applied and Environmental Microbiology 62:1458–1460.

31. Liu J, Wang Z, Belchik SM, Edwards MJ, Liu C, Kennedy DW, Merkley ED, Lipton MS, Butt JN, Richardson DJ, Zachara JM, Fredrickson JK, Rosso KM, Shi L. 2012. Identification and Characterization of MtoA: A Decaheme c-Type Cytochrome of the

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Neutrophilic Fe(II)-Oxidizing Bacterium Sideroxydans lithotrophicus ES-1. Front Microbio 3.

32. Keffer JL, McAllister SM, Garber AI, Hallahan BJ, Sutherland MC, Rozovsky S, Chan CS. 2021. Iron Oxidation by a Fused Cytochrome-Porin Common to Diverse Iron-Oxidizing Bacteria. mBio 0:e01074-21.

33. McAllister SM, Polson SW, Butterfield DA, Glazer BT, Sylvan JB, Chan CS. 2020. Validating the Cyc2 Neutrophilic Iron Oxidation Pathway Using Meta-omics of *Zetaproteobacteria* Iron Mats at Marine Hydrothermal Vents. mSystems 5:e00553-19, /msystems/5/1/msys.00553-19.atom.

34. Jiao Y, Newman DK. 2007. The pio Operon Is Essential for Phototrophic Fe(II) Oxidation in Rhodopseudomonas palustris TIE-1. Journal of Bacteriology 189:1765–1773.

35. Bose A, Gardel EJ, Vidoudez C, Parra EA, Girguis PR. 2014. Electron uptake by iron-oxidizing phototrophic bacteria. Nat Commun 5:3391.

36. Beliaev AS, Saffarini DA. 1998. Shewanella putrefaciens mtrB Encodes an Outer Membrane Protein Required for Fe(III) and Mn(IV) Reduction. Journal of Bacteriology 180:6292–6297.

37. Zhou N, Kupper RJ, Catalano JG, Thompson A, Chan CS. 2022. Biological Oxidation of Fe(II)-Bearing Smectite by Microaerophilic Iron Oxidizer Sideroxydans lithotrophicus Using Dual Mto and Cyc2 Iron Oxidation Pathways. Environ Sci Technol https://doi.org/10.1021/acs.est.2c05142.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

38. He S, Barco RA, Emerson D, Roden EE. 2017. Comparative Genomic Analysis of Neutrophilic Iron(II) Oxidizer Genomes for Candidate Genes in Extracellular Electron Transfer. Front Microbiol 8.

39. Alawi M, Lipski A, Sanders T, Eva-Maria-Pfeiffer, Spieck E. 2007. Cultivation of a novel cold-adapted nitrite oxidizing betaproteobacterium from the Siberian Arctic. 3. The ISME Journal 1:256–264.

40. Ishii K, Fujitani H, Soh K, Nakagawa T, Takahashi R, Tsuneda S. 2017. Enrichment and Physiological Characterization of a Cold-Adapted Nitrite-Oxidizing Nitrotoga sp. from an Eelgrass Sediment. Applied and Environmental Microbiology 83:e00549-17, /aem/83/14/e00549-17.atom.

41. Podowski JC, Paver SF, Newton RJ, Coleman ML. 2022. Genome Streamlining, Proteorhodopsin, and Organic Nitrogen Metabolism in Freshwater Nitrifiers. mBio 13:e02379-21.

42. Boddicker AM, Mosier AC. 2018. Genomic profiling of four cultivated Candidatus Nitrotoga spp. predicts broad metabolic potential and environmental distribution. The ISME Journal 12:2864–2882.

43. Kitzinger K, Koch H, Lücker S, Sedlacek CJ, Herbold C, Schwarz J, Daebeler A, Mueller AJ, Lukumbuzya M, Romano S, Leisch N, Karst SM, Kirkegaard R, Albertsen M, Nielsen PH, Wagner M, Daims H. 2018. Characterization of the First " *Candidatus* Nitrotoga" Isolate Reveals Metabolic Versatility and Separate Evolution of Widespread Nitrite-Oxidizing Bacteria. mBio 9:e01186-18, /mbio/9/4/mBio.01186-18.atom.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

44. Lücker S, Schwarz J, Gruber-Dorninger C, Spieck E, Wagner M, Daims H. 2015. Nitrotoga -like bacteria are previously unrecognized key nitrite oxidizers in full-scale wastewater treatment plants. 3. The ISME Journal 9:708–720.

45. Ishii K, Fujitani H, Sekiguchi Y, Tsuneda S. 2020. Physiological and genomic characterization of a new 'Candidatus Nitrotoga' isolate. Environmental Microbiology 22.

46. Daims H, Lücker S, Wagner M. 2016. A New Perspective on Microbes Formerly Known as Nitrite-Oxidizing Bacteria. Trends in Microbiology 24:699–712.

47. Ivanova N, Tringe SG, Liolios K, Liu W-T, Morrison N, Hugenholtz P, Kyrpides NC. 2010. A call for standardized classification of metagenome projects. Environmental Microbiology 12:1803–1805.

48. Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa SL, Solden LM, Liu P, Narrowe AB, Rodríguez-Ramos J, Bolduc B, Gazitúa MC, Daly RA, Smith GJ, Vik DR, Pope PB, Sullivan MB, Roux S, Wrighton KC. 2020. DRAM for distilling microbial metabolism to automate the curation of microbiome function. Nucleic Acids Res 48:8883–8900.

49. Garber AI, Nealson KH, Okamoto A, McAllister SM, Chan CS, Barco RA, Merino N. 2020. FeGenie: A Comprehensive Tool for the Identification of Iron Genes and Iron Gene Neighborhoods in Genome and Metagenome Assemblies. Frontiers in Microbiology 11.

50. Garber AI. 2020. MagicLamp: toolkit for annotation of 'omics datasets using curated HMM sets (2020: MagicLamp). GitHub repository. https://github.com/Arkadiy-Garber/MagicLamp.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

51. McAllister SM. 2016. Heme Counter. GitHub repository. https://github.com/seanmcallister/heme_counter.

52. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410.

53. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

54. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ 3:e1319.

55. Shaiber A, Willis AD, Delmont TO, Roux S, Chen L-X, Schmid AC, Yousef M, Watson AR, Lolans K, Esen ÖC, Lee STM, Downey N, Morrison HG, Dewhirst FE, Mark Welch JL, Eren AM. 2020. Functional and genetic markers of niche partitioning among enigmatic members of the human oral microbiome. Genome Biology 21:292.

56. Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, Fink I, Pan JN, Yousef M, Fogarty EC, Trigodet F, Watson AR, Esen ÖC, Moore RM, Clayssen Q, Lee MD, Kivenson V, Graham ED, Merrill BD, Karkman A, Blankenberg D, Eppley JM, Sjödin A, Scott JJ, Vázquez-Campos X, McKay LJ, McDaniel EA, Stevens SLR, Anderson RE, Fuessel J, Fernandez-Guerra A, Maignien L, Delmont TO, Willis AD. 2021. Community-led, integrated, reproducible multi-omics with anvi'o. 1. Nature Microbiology 6:3–6.

57. Baker IR, Conley BE, Gralnick JA, Girguis PR. 2022. Evidence for Horizontal and Vertical Transmission of Mtr-Mediated Extracellular Electron Transfer among the Bacteria. mBio https://doi.org/10.1128/mbio.02904-21.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

58. Ross DE, Flynn JM, Baron DB, Gralnick JA, Bond DR. 2011. Towards Electrosynthesis in Shewanella: Energetics of Reversing the Mtr Pathway for Reductive Metabolism. PLOS ONE 6:e16649.

59. Salgueiro CA, Dantas JM. 2016. Multiheme Cytochromes, 1st ed. Springer Berlin Heidelberg, Berlin, Heidelberg. http://link.springer.com/10.1007/978-3-642-44961-1.

60. Paquete CM, Morgado L, Salgueiro CA, Louro RO. 2022. Molecular Mechanisms of Microbial Extracellular Electron Transfer: The Importance of Multiheme Cytochromes. 6. Frontiers in Bioscience-Landmark 27:174.

61. Eddie BJ, Bird LJ, Pelikan C, Mussmann M, Martínez-Pérez C, Pinamang P, Malanoski AP, Glaven SM. 2022. Conservation of Energetic Pathways for Electroautotrophy in the Uncultivated Candidate Order Tenderiales. mSphere 0:e00223-22.

62. Edwards MJ, White GF, Butt JN, Richardson DJ, Clarke TA. 2020. The Crystal Structure of a Biological Insulated Transmembrane Molecular Wire. Cell 181:665-673.e10.

63. Beckwith CR, Edwards MJ, Lawes M, Shi L, Butt JN, Richardson DJ, Clarke TA. 2015. Characterization of MtoD from Sideroxydans lithotrophicus: a cytochrome c electron shuttle used in lithoautotrophic growth. Frontiers in Microbiology 6.

64. Ducluzeau A-L, Ouchane S, Nitschke W. 2008. The cbb3 Oxidases Are an Ancient Innovation of the Domain Bacteria. Molecular Biology and Evolution 25:1158–1166.

65. Trojan D, Garcia-Robledo E, Meier DV, Hausmann B, Revsbech NP, Eichorst SA, Woebken D. 2021. Microaerobic Lifestyle at Nanomolar O2 Concentrations Mediated by

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Low-Affinity Terminal Oxidases in Abundant Soil Bacteria. mSystems https://doi.org/10.1128/mSystems.00250-21.

66. Badger MR, Bek EJ. 2008. Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO2 acquisition by the CBB cycle. J Exp Bot 59:1525–1541.

67. Griesbeck C, Schütz M, Schödl T, Bathe S, Nausch L, Mederer N, Vielreicher M, Hauska G. 2002. Mechanism of Sulfide-Quinone Reductase Investigated Using Site-Directed Mutagenesis and Sulfur Analysis. Biochemistry 41:11552–11565.

68. Marcia M, Ermler U, Peng G, Michel H. 2009. The structure of Aquifex aeolicus sulfide:quinone oxidoreductase, a basis to understand sulfide detoxification and respiration. Proceedings of the National Academy of Sciences 106:9625–9630.

69. Loy A, Duller S, Baranyi C, Mußmann M, Ott J, Sharon I, Béjà O, Le Paslier D, Dahl C, Wagner M. 2009. Reverse dissimilatory sulfite reductase as phylogenetic marker for a subgroup of sulfur-oxidizing prokaryotes. Environmental Microbiology 11:289–299.

70. Müller AL, Kjeldsen KU, Rattei T, Pester M, Loy A. 2015. Phylogenetic and environmental diversity of DsrAB-type dissimilatory (bi)sulfite reductases. 5. ISME J 9:1152–1165.

71. Kato S, Ohkuma M, Powell DH, Krepski ST, Oshima K, Hattori M, Shapiro N, Woyke T, Chan CS. 2015. Comparative Genomic Insights into Ecophysiology of Neutrophilic, Microaerophilic Iron Oxidizing Bacteria. Front Microbiol 6.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

72. Koeksoy E, Bezuidt OM, Bayer T, Chan CS, Emerson D. 2021. Zetaproteobacteria Pan-Genome Reveals Candidate Gene Cluster for Twisted Stalk Biosynthesis and Export. Front Microbiol 12:1374.

73. Schmehl M, Jahn A, Meyer zu Vilsendorf A, Hennecke S, Masepohl B, Schuppler M, Marxer M, Oelze J, Klipp W. 1993. Identification of a new class of nitrogen fixation genes in Rhodobacter capsalatus: a putative membrane complex involved in electron transport to nitrogenase. Molec Gen Genet 241:602–615.

74. Tremblay P-L, Zhang T, Dar SA, Leang C, Lovley DR. 2012. The Rnf Complex of Clostridium ljungdahlii Is a Proton-Translocating Ferredoxin:NAD+ Oxidoreductase Essential for Autotrophic Growth. mBio 4:e00406-12.

75. Downing BE, Gupta D, Nayak DD. 2022. The dual role of a multi-heme cytochrome in methanogenesis: MmcA is important for energy conservation and carbon metabolism in Methanosarcina acetivorans. bioRxiv https://doi.org/10.1101/2022.07.20.500911.

76. Liu Y, Chen H, Van Treuren W, Hou B-H, Higginbottom SK, Dodd D. 2022. Clostridium sporogenes uses reductive Stickland metabolism in the gut to generate ATP and produce circulating metabolites. 5. Nat Microbiol 7:695–706.

77. Child SA, Bradley JM, Pukala TL, Svistunenko DA, Le Brun NE, Bell SG. 2018. Electron transfer ferredoxins with unusual cluster binding motifs support secondary metabolism in many bacteria. Chem Sci 9:7948–7957.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

78. Watanabe T, Kojima H. 2019. Sulfuricella, p. 1–6. *In* Whitman, WB, Rainey, F, Kämpfer, P, Trujillo, M, Chun, J, DeVos, P, Hedlund, B, Dedysh, S (eds.), Bergey's Manual of Systematics of Archaea and Bacteria, 1st ed. Wiley.

79. Gorski CA, Klüpfel LE, Voegelin A, Sander M, Hofstetter TB. 2013. Redox Properties of Structural Fe in Clay Minerals: 3. Relationships between Smectite Redox and Structural Properties. Environ Sci Technol 47:13477–13485.

80. Gupta D, Sutherland MC, Rengasamy K, Meacham JM, Kranz RG, Bose A. 2019. Photoferrotrophs Produce a PioAB Electron Conduit for Extracellular Electron Uptake. mBio 10:e02668-19.

81. Coursolle D, Baron DB, Bond DR, Gralnick JA. 2010. The Mtr Respiratory Pathway Is Essential for Reducing Flavins and Electrodes in Shewanella oneidensis. Journal of Bacteriology 192:467–474.

82. Beliaev AS, Saffarini DA, McLaughlin JL, Hunnicutt D. 2001. MtrC, an outer membrane decahaem c cytochrome required for metal reduction in Shewanella putrefaciens MR-1. Molecular Microbiology 39:722–730.

83. Gorby YA, Yanina S, McLean JS, Rosso KM, Moyles D, Dohnalkova A, Beveridge TJ, Chang IS, Kim BH, Kim KS, Culley DE, Reed SB, Romine MF, Saffarini DA, Hill EA, Shi L, Elias DA, Kennedy DW, Pinchuk G, Watanabe K, Ishii S, Logan B, Nealson KH, Fredrickson JK. 2006. Electrically conductive bacterial nanowires produced by Shewanella oneidensis strain MR-1 and other microorganisms. Proceedings of the National Academy of Sciences 103:11358–11363.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

84. Edwards MJ, Richardson DJ, Paquete CM, Clarke TA. 2020. Role of multiheme cytochromes involved in extracellular anaerobic respiration in bacteria. Protein Science 29:830–842.

85. Richardson DJ, Butt JN, Fredrickson JK, Zachara JM, Shi L, Edwards MJ, White G, Baiden N, Gates AJ, Marritt SJ, Clarke TA. 2012. The 'porin–cytochrome' model for microbe-to-mineral electron transfer. Molecular Microbiology 85:201–212.

86. White GF, Shi Z, Shi L, Wang Z, Dohnalkova AC, Marshall MJ, Fredrickson JK, Zachara JM, Butt JN, Richardson DJ, Clarke TA. 2013. Rapid electron exchange between surface-exposed bacterial cytochromes and Fe(III) minerals. Proceedings of the National Academy of Sciences 110:6346–6351.

87. van Wonderen JH, Adamczyk K, Wu X, Jiang X, Piper SEH, Hall CR, Edwards MJ, Clarke TA, Zhang H, Jeuken LJC, Sazanovich IV, Towrie M, Blumberger J, Meech SR, Butt JN. 2021. Nanosecond heme-to-heme electron transfer rates in a multiheme cytochrome nanowire reported by a spectrally unique His/Met-ligated heme. Proceedings of the National Academy of Sciences 118:e2107939118.

88. Wang F, Gu Y, O'Brien JP, Yi SM, Yalcin SE, Srikanth V, Shen C, Vu D, Ing NL, Hochbaum AI, Egelman EH, Malvankar NS. 2019. Structure of Microbial Nanowires Reveals Stacked Hemes that Transport Electrons over Micrometers. Cell 177:361-369.e10.

89. Chan CS, McAllister SM, Leavitt AH, Glazer BT, Krepski ST, Emerson D. 2016. The Architecture of Iron Microbial Mats Reflects the Adaptation of Chemolithotrophic Iron Oxidation in Freshwater and Marine Environments. Front Microbiol 7.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

90. Hallbeck L, Pedersen K. 2014. The Family Gallionellaceae, p. 853–858. *In* Rosenberg, E, DeLong, EF, Lory, S, Stackebrandt, E, Thompson, F (eds.), The Prokaryotes: Alphaproteobacteria and Betaproteobacteria. Springer, Berlin, Heidelberg.

91. Hedrich S, Schlömann M, Johnson DB. 2011. The iron-oxidizing proteobacteria. Microbiology 157:1551–1564.

92. Ilbert M, Bonnefoy V. 2013. Insight into the evolution of the iron oxidation pathways. Biochimica et Biophysica Acta (BBA) - Bioenergetics 1827:161–175.

93. Geer LY, Marchler-Bauer A, Geer RC, Han L, He J, He S, Liu C, Shi W, Bryant SH. 2010. The NCBI BioSystems database. Nucleic Acids Res 38:D492–D496.

94. Chen I-MA, Chu K, Palaniappan K, Ratner A, Huang J, Huntemann M, Hajek P, Ritter S, Varghese N, Seshadri R, Roux S, Woyke T, Eloe-Fadrosh EA, Ivanova NN, Kyrpides NC. 2021. The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities. Nucleic Acids Research 49:D751–D763.

95. Cooper RE, Wegner C-E, McAllister SM, Shevchenko O, Chan CS, Küsel K. 2020. Draft Genome Sequence of *Sideroxydans* sp. Strain CL21, an Fe(II)-Oxidizing Bacterium. Microbiol Resour Announc 9:e01444-19.

96. Probst AJ, Ladd B, Jarett JK, Geller-McGrath DE, Sieber CMK, Emerson JB, Anantharaman K, Thomas BC, Malmstrom RR, Stieglmeier M, Klingl A, Woyke T, Ryan MC, Banfield JF. 2018. Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. Nat Microbiol 3:328–336.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

97. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055.

98. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36:1925–1927.

99. Mukherjee S, Stamatis D, Bertsch J, Ovchinnikova G, Sundaramurthi JC, Lee J, Kandimalla M, Chen I-MA, Kyrpides NC, Reddy TBK. 2021. Genomes OnLine Database (GOLD) v.8: overview and updates. Nucleic Acids Research 49:D723–D733.

100. Buttigieg PL, Morrison N, Smith B, Mungall CJ, Lewis SE, the ENVO Consortium. 2013. The environment ontology: contextualising biological and biomedical entities. Journal of Biomedical Semantics 4:43.

101. Parks D. 2014. CompareM. Github repository. https://github.com/dparks1134/CompareM.

102. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, Sneddon MW, Henderson ML, Riehl WJ, Murphy-Olson D, Chan SY, Kamimura RT, Kumari S, Drake MM, Brettin TS, Glass EM, Chivian D, Gunter D, Weston DJ, Allen BH, Baumohl J, Best AA, Bowen B, Brenner SE, Bun CC, Chandonia J-M, Chia J-M, Colasanti R, Conrad N, Davis JJ, Davison BH, DeJongh M, Devoid S, Dietrich E, Dubchak I, Edirisinghe JN, Fang G, Faria JP, Frybarger PM, Gerlach W, Gerstein M, Greiner A, Gurtowski J, Haun HL, He F, Jain R, Joachimiak MP, Keegan KP, Kondo S, Kumar V, Land ML, Meyer F, Mills M, Novichkov PS, Oh T, Olsen GJ, Olson R, Parrello B, Pasternak S, Pearson E, Poon SS, Price GA, Ramakrishnan S, Ranjan P, Ronald PC,

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Schatz MC, Seaver SMD, Shukla M, Sutormin RA, Syed MH, Thomason J, Tintle NL, Wang D, Xia F, Yoo H, Yoo S, Yu D. 2018. KBase: The United States Department of Energy Systems Biology Knowledgebase. 7. Nature Biotechnology 36:566–569.

103. Watanabe T, Kojima H, Fukui M. 2015. Sulfuriferula multivorans gen. nov., sp. nov., isolated from a freshwater lake, reclassification of 'Thiobacillus plumbophilus' as Sulfuriferula plumbophilus sp. nov., and description of Sulfuricellaceae fam. nov. and Sulfuricellales ord. nov. International Journal of Systematic and Evolutionary Microbiology, 65:1504–1508.

104. Kojima H, Fukui M. 2010. Sulfuricella denitrificans gen. nov., sp. nov., a sulfur-oxidizing autotroph isolated from a freshwater lake. International Journal of Systematic and Evolutionary Microbiology, 60:2862–2866.

105. Geneious v.10.2.6. https://www.geneious.com.

106. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32:1792–1797.

107. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35:4453–4455.

108. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Research 49:W293–W296.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

109. Pruesse E, Peplies J, Glöckner FO. 2012. SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics 28:1823–1829.

110. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:D590–D596.

111. Moore RM, Harrison AO, McAllister SM, Polson SW, Wommack EK. 2020. Iroki: automatic customization and visualization of phylogenetic trees. PeerJ e8584.

112. Steinegger M, Söding J. 2017. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. 11. Nat Biotechnol 35:1026–1028.

113. The UniProt Consortium. 2021. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Research 49:D480–D489.

114. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, Brinkman FSL. 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26:1608–1615.

115. Goldberg T, Hecht M, Hamp T, Karl T, Yachdav G, Ahmed N, Altermann U, Angerer P, Ansorge S, Balasz K, Bernhofer M, Betz A, Cizmadija L, Do KT, Gerke J, Greil R, Joerdens V, Hastreiter M, Hembach K, Herzog M, Kalemanov M, Kluge M, Meier A, Nasir H, Neumaier U, Prade V, Reeb J, Sorokoumov A, Troshani I, Vorberg S, Waldraff S, Zierer J, Nielsen H, Rost B. 2014. LocTree3 prediction of localization. Nucleic Acids Research 42:W350–W355.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

116. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. 2022. ColabFold: making protein folding accessible to all. 6. Nat Methods 19:679–682.

117. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. 2021. Highly accurate protein structure prediction with AlphaFold. 7873. Nature 596:583–589.

118. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, Žídek A, Bates R, Blackwell S, Yim J, Ronneberger O, Bodenstein S, Zielinski M, Bridgland A, Potapenko A, Cowie A, Tunyasuvunakool K, Jain R, Clancy E, Kohli P, Jumper J, Hassabis D. 2022. Protein complex prediction with AlphaFold-Multimer. bioRxiv https://doi.org/10.1101/2021.10.04.463034.

119. Schrodinger, LLC. 2010. The PyMOL Molecular Graphics System (2.5.4).

120. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119.

121. Galperin MY, Kristensen DM, Makarova KS, Wolf YI, Koonin EV. 2019. Microbial genome analysis: the COG approach. Brief Bioinform 20:1063–1070.

bioRxiv preprint doi: https://doi.org/10.1101/2023.01.26.525709; this version posted February 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

122. Galperin MY, Wolf YI, Makarova KS, Vera Alvarez R, Landsman D, Koonin EV. 2021. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. Nucleic Acids Research 49:D274–D281.