

1 **Gallionellaceae pangenomic analysis reveals insight into**
2 **phylogeny, metabolic flexibility, and iron oxidation**
3 **mechanisms**

4 Rene L. Hoover^{a,b}, Jessica L. Keffer^b, Shawn W. Polson^{d,e}, Clara S. Chan^{a,b,c}

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

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

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

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

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

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

13 **Abstract**

14 The iron-oxidizing Gallionellaceae drive a wide variety of biogeochemical cycles through
15 their metabolisms and biominerals. To better understand the environmental impacts of
16 Gallionellaceae, we need to improve our knowledge of their diversity and metabolisms,
17 especially any novel iron oxidation mechanisms. Here, we used a pangenomic analysis of 103
18 genomes to resolve Gallionellaceae phylogeny and explore the range of genomic potential. Using
19 a concatenated ribosomal protein tree and key gene patterns, we determined Gallionellaceae has
20 four genera, divided into two groups—iron-oxidizing bacteria (FeOB) *Gallionella*, *Sideroxydans*,
21 and *Ferriphaselus* with known iron oxidases (Cyc2, MtoA) and nitrite-oxidizing bacteria (NOB)
22 *Candidatus Nitrotoga* with nitrite oxidase (Nxr). The FeOB and NOB have similar electron
23 transport chains, including genes for reverse electron transport and carbon fixation. Auxiliary
24 energy metabolisms including S oxidation, denitrification, and organotrophy were scattered
25 throughout the Gallionellaceae FeOB. Within FeOB, we found genes that may represent
26 adaptations for iron oxidation, including a variety of extracellular electron uptake (EEU)
27 mechanisms. FeOB genomes encoded more predicted *c*-type cytochromes overall, notably more
28 multiheme *c*-type cytochromes (MHCs) with >10 CXXCH motifs. These include homologs of
29 several predicted outer membrane porin-MHC complexes, including MtoAB and Uet. MHCs are
30 known to efficiently conduct electrons across longer distances and function across a wide range
31 of redox potentials that overlap with mineral redox potentials, which can help expand the range
32 of usable iron substrates. Overall, the results of pangenome analyses suggest that the
33 Gallionellaceae genera *Gallionella*, *Sideroxydans*, and *Ferriphaselus* are primarily iron
34 oxidizers, capable of oxidizing dissolved Fe²⁺ as well as a range of solid iron or other mineral
35 substrates.

36 **Importance**

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

48 **Introduction**

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

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

62 The Gallionellaceae are named after *Gallionella ferruginea*, first described by Ehrenberg
63 in 1838 (20), and recognizable by its distinctive, twisted iron oxyhydroxide stalk (21). While the
64 type strain, *G. ferruginea* Johan (22) no longer exists, there are seven iron-oxidizing
65 Gallionellaceae isolates (7, 11, 23–26). Some isolates, such as *Ferriphaseelus* spp., appear to be
66 obligate iron oxidizers, while others also grow on non-iron substrates. In addition to iron, *S.*
67 *lithotrophicus* ES-1 grows by thiosulfate oxidation (24, 27) while *Sideroxydans* sp. CL21 shows
68 mixotrophic growth with either lactate or hydrogen (28). Some *Ferrigenium* can also reduce
69 nitrate (29, 30). It is unknown how common it is for Gallionellaceae to use electron
70 donors/acceptors besides Fe(II)/O₂, though these alternate metabolisms may help their success
71 across different environments and fluctuating conditions typical of many oxic-anoxic interfaces.
72 Even so, since all seven Gallionellaceae isolates are neutrophilic aerobic chemolithoautotrophic
73 iron oxidizers, this could be the dominant metabolic niche of Gallionellaceae.

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

81 common than *mtoAB* and is found in nearly all well-characterized neutrophilic FeOB like the
82 Gallionellaceae (32) and Zetaproteobacteria (33), making it a suitable genetic marker for many
83 FeOB. Cyc2 has been demonstrated to oxidize aqueous Fe²⁺ (32), while Mto gene/protein
84 expression has been associated with the oxidation of solid iron minerals (37). However, Cyc2
85 and MtoA may not be the only mechanisms for neutrophilic iron oxidation. There are a number
86 of additional uncharacterized cytochromes and electron transport genes (27, 38) within
87 Gallionellaceae genomes such as isolate *S. lithotrophicus* ES-1 (27, 38), suggesting the existence
88 of novel iron oxidation genes and mechanisms within the family.

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

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

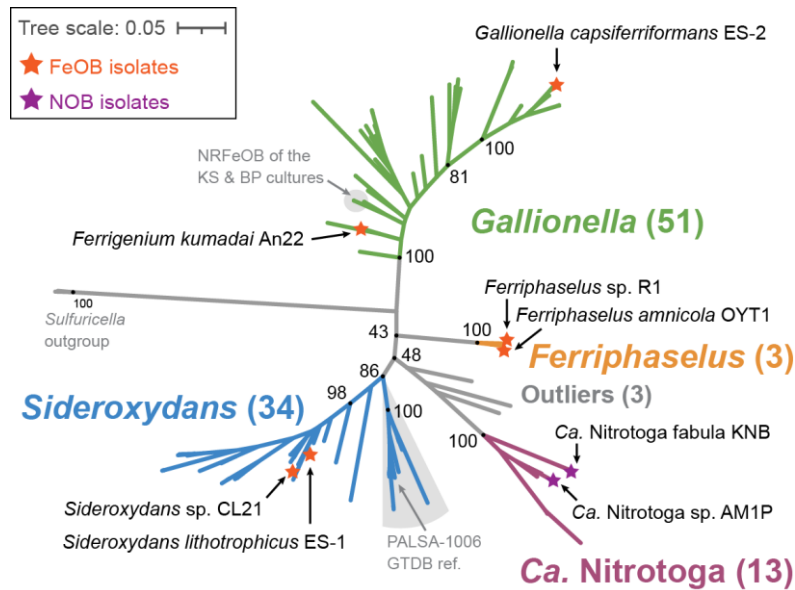
104 assembles genomes (MAGs). We used those sequences to resolve the Gallionellaceae phylogeny
105 and delineate groups of iron and nitrite oxidizers. We searched for known and novel iron
106 oxidation genes, other energy and nutrient metabolisms, and genes found exclusively in FeOB
107 that may represent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

111 **Results**

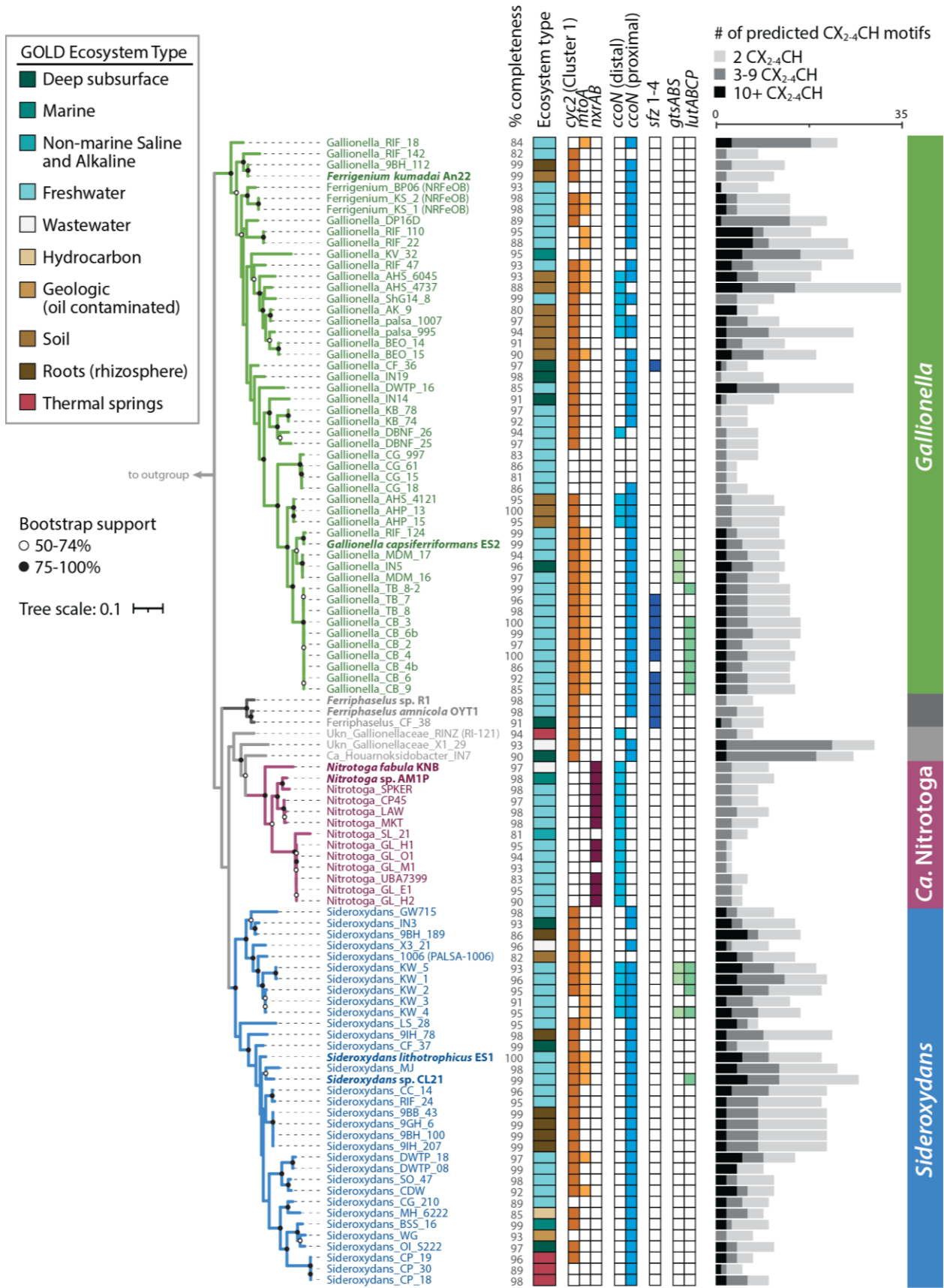
112 **Phylogeny**

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

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



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



136 **FIGURE 2** Maximum likelihood tree of concatenated ribosomal proteins from the
137 Gallionellaceae annotated with source ecosystem and genes for iron oxidases (*cyc2*, *mtaA*),
138 nitrite oxidase (*nxrAB*), terminal oxidase (*ccoN*), stalk formation (*sfz*), and organic utilization
139 (*gtsABS*, *lutABCP*). The bar graph to the right shows the number multiheme cytochromes
140 CXXCH, CX₃CH, and CX₄CH heme-binding motifs. Phylogeny does not correlate to
141 environments, and key genes, including those for multiheme cytochromes, show distinct
142 distributions between iron and nitrite oxidizer clades. Isolates are shown in bold. % completeness
143 = genome completeness calculated with CheckM. Outgroup omitted for space.

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

155 **Metabolic potential and diversity**

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

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

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

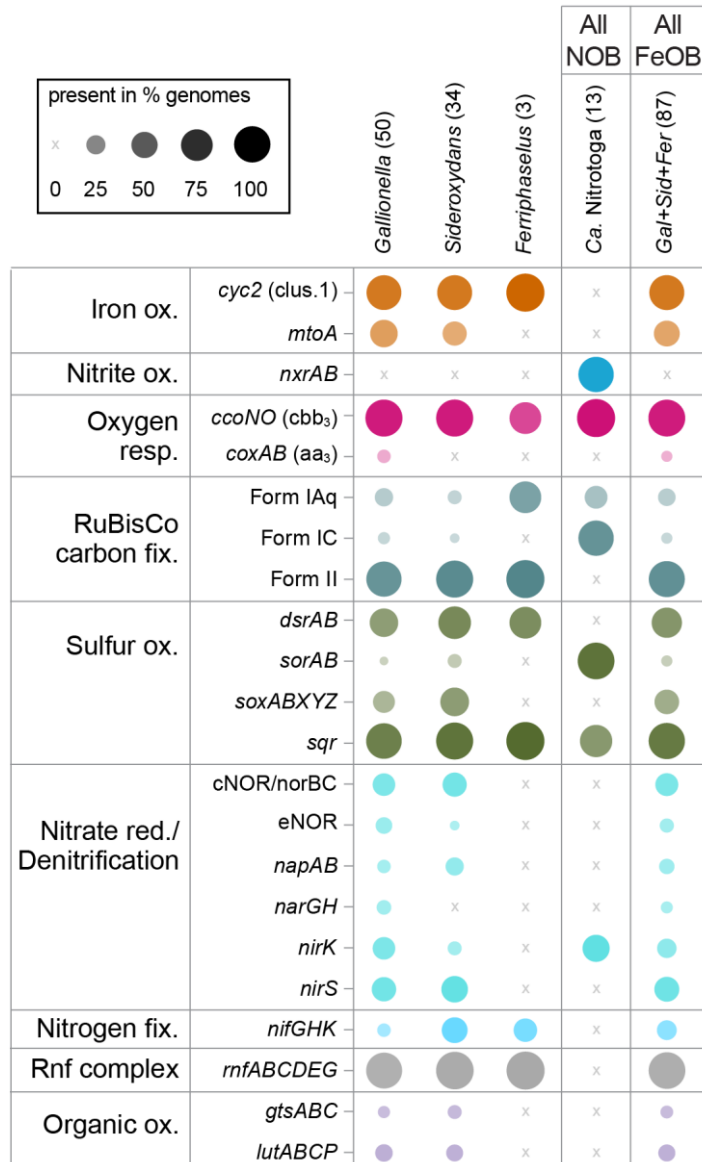
167 Known metabolisms for the few Gallionellaceae isolates suggest *Ca. Nitrotoga* are nitrite
168 oxidizers, while *Sideroxydans*, *Ferriphaselus*, and *Gallionella* are iron oxidizers. We examined
169 the pangenome for the presence of *cyc2* and *mtoA* iron oxidase genes and *nxrAB* nitrite oxidase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

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

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

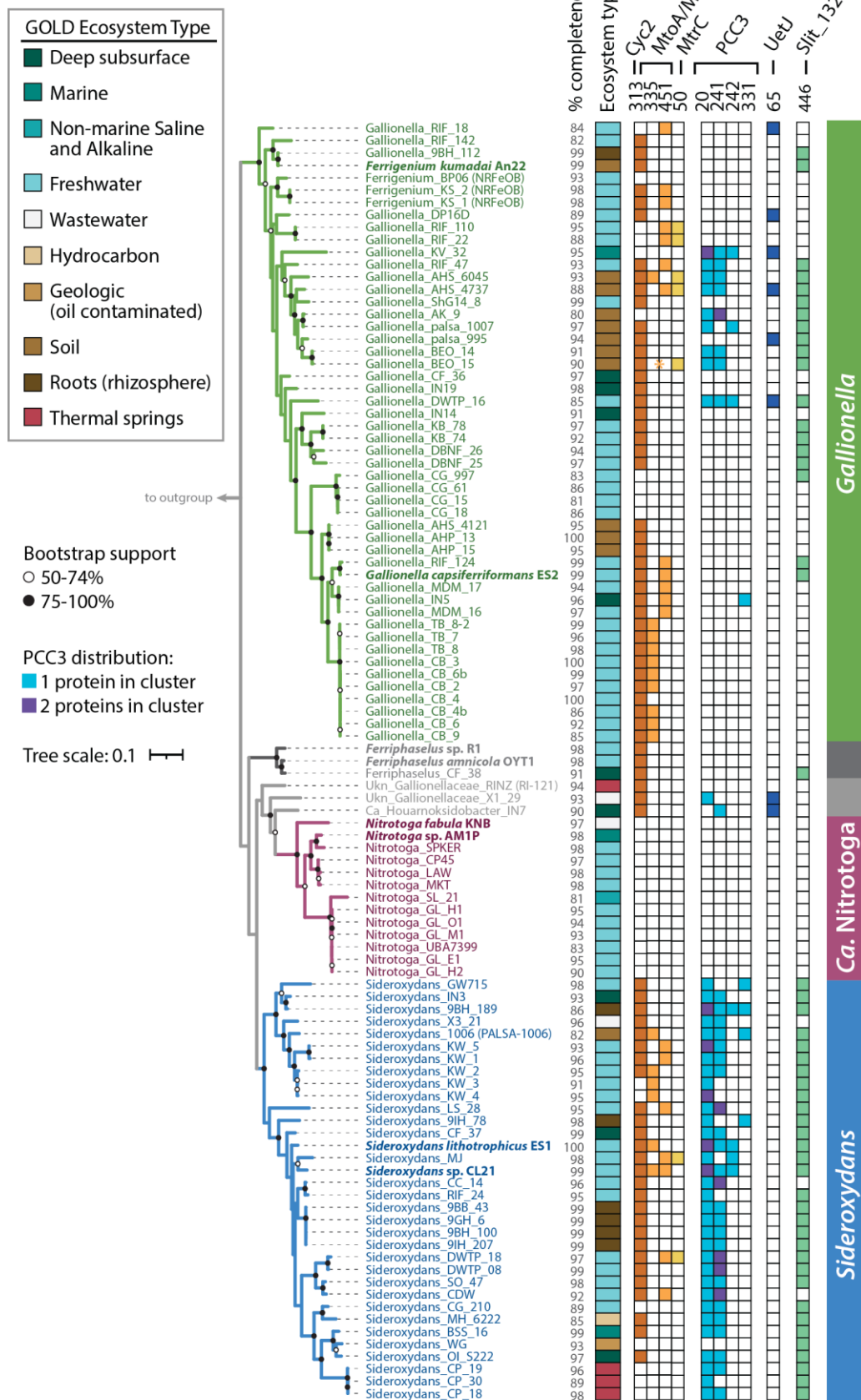
183 The *Ca. Nitrotoga* SL_21 MAG contains only a predicted Cluster 2 Cyc2 homolog.
184 Confidently assigning iron oxidation function to Cluster 2 Cyc2s depends on supporting context,
185 which is lacking in this case. *Ca. Nitrotoga* SL_21 is not from a typical iron-oxidizing
186 environment (permanently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

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



194 **FIGURE 3** Plot showing the percent of genomes in each genus/group with genes for key
 195 metabolic pathways. The plot indicates the Gallionellaceae are aerobic lithoautotrophs with two
 196 main energy metabolisms, iron or nitrite oxidation. Some members also have metabolic potential
 197 for S oxidation and/or denitrification. Numbers in parentheses indicate the total number of
 198 genomes in each group. Color is used to distinguish groups, while dot size and opacity indicate
 199 % presence in the genome groups.

200



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

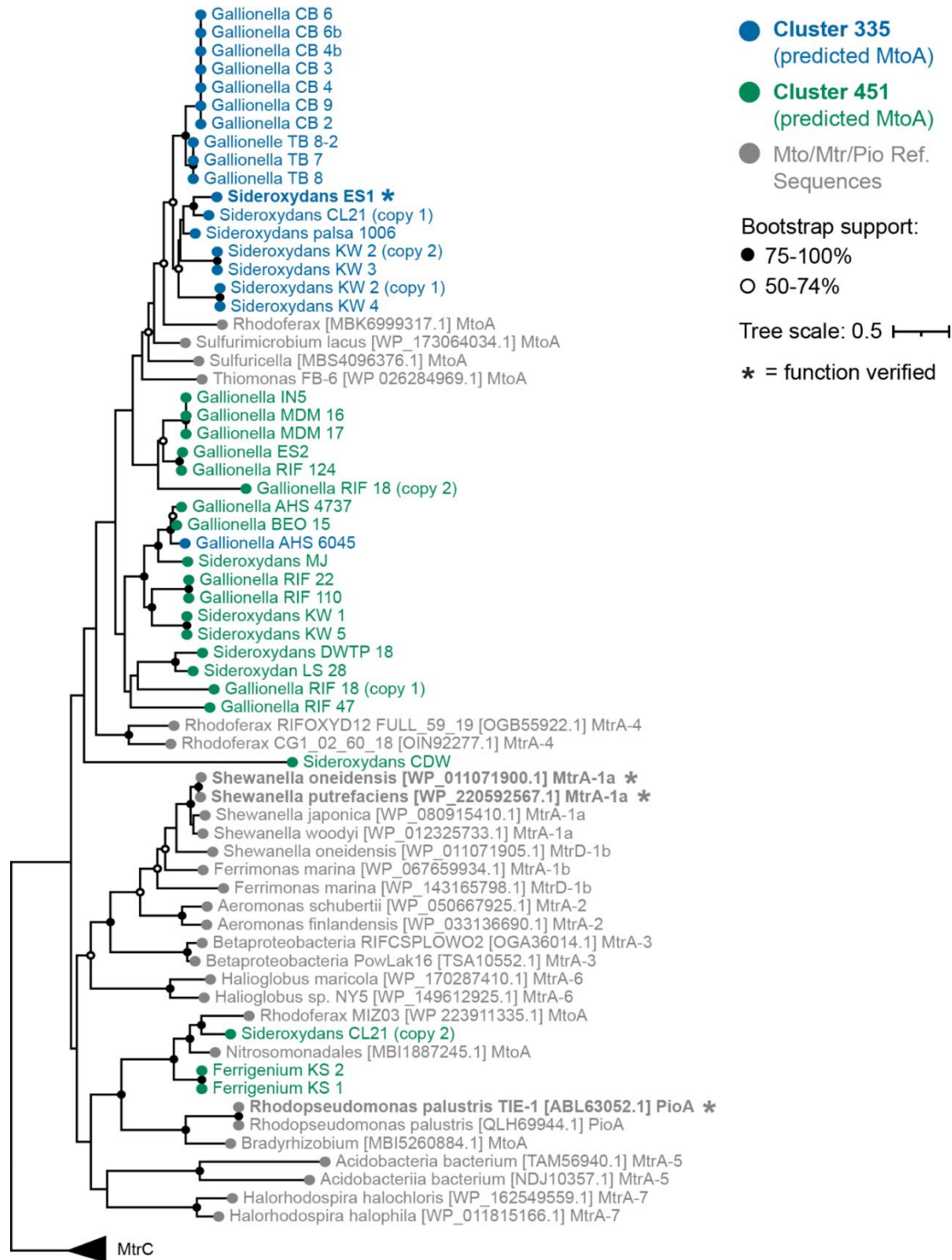
208 ***c*-type cytochromes**

209 Both known iron oxidases (Cyc2 and MtoA) in Gallionellaceae are *c*-type cytochromes
210 that transport electrons across the outer membrane. FeOB use additional *c*-type cytochromes to
211 transport electrons through the periplasm to the rest of the electron transport chain. We reasoned
212 that novel iron oxidation mechanisms may also utilize *c*-type cytochromes, so we searched the
213 Gallionellaceae genomes for proteins containing the CXXCH, CX₃CH, and CX₄CH heme-
214 binding motifs (abbreviated hereafter as CXXCH). There is a stark difference between FeOB and
215 NOB in the distribution of predicted *c*-type cytochromes. FeOB genomes have an average of
216 1.5x more CXXCH-containing proteins than NOB, and only non-NOB genomes encoded
217 proteins with ten or more CXXCH motifs (Fig. 2). The abundance of genes for potential *c*-type
218 cytochromes, in particular multiheme cytochromes (MHCs), suggest the presence of additional
219 iron oxidation mechanisms in the Gallionellaceae FeOB.

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

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

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



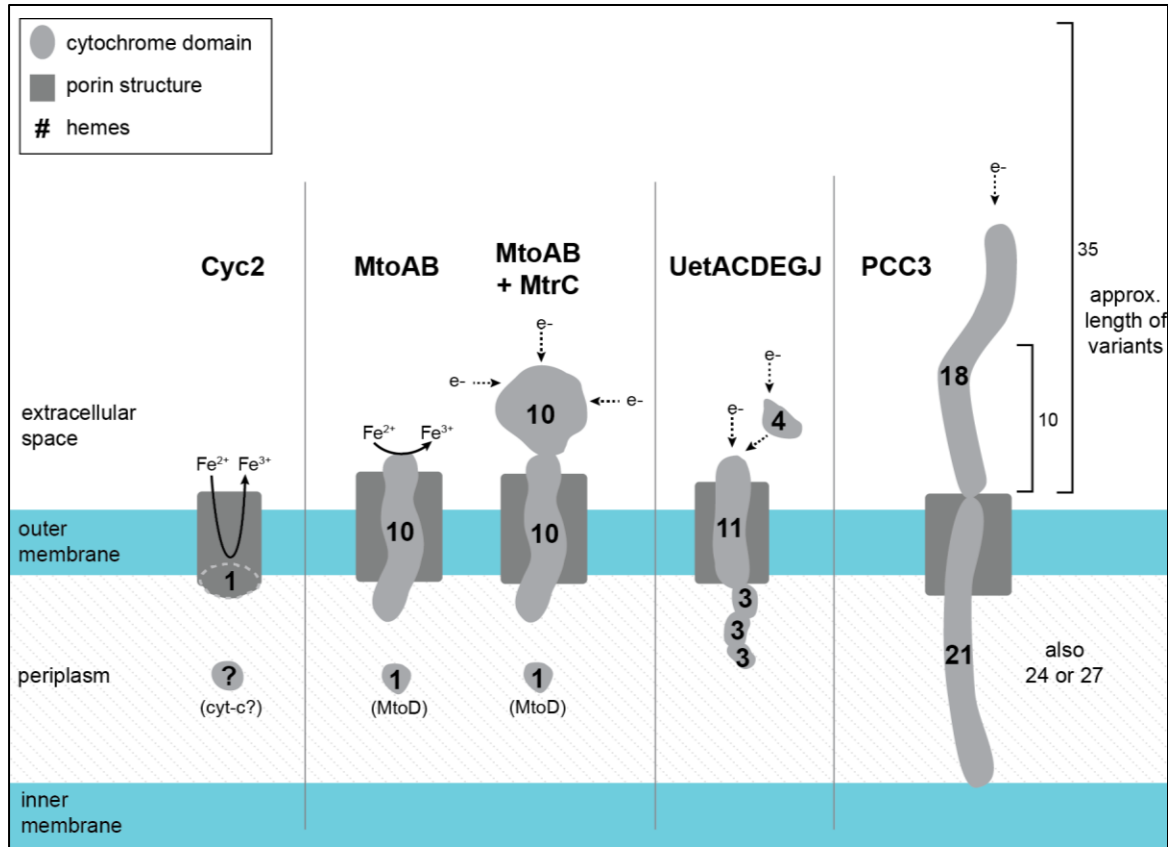
243 **FIGURE 5** Maximum likelihood tree of the predicted MtoA sequences identified in MMSeqs2
 244 Cluster 335 and Cluster 451 along with MtoA reference sequences from NCBI and MtrA
 245 reference sequences from Baker, et al. 2022. Numbers (1a, 1b, 2, 3, 4, 5, 6, and 7) appended after

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

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

Cluster	Functional prediction	# CXXCH, CX ₃ CH, or CX ₄ CH motifs per protein	# FeOB (of 87)
Iron oxidation/reduction proteins			
313	Iron oxidase Cyc2‡	1	70
451	Decaheme <i>c</i> -type cytochrome, DmsE family, MtoA [‡]	10	19
335	Decaheme <i>c</i> -type cytochrome, DmsE family, MtoA [‡]	10	17
50	Decaheme <i>c</i> -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 <i>c</i> 3 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 <i>c</i> [‡] - 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 <i>c</i> ; Octaheme tetrathionate reductase [†]	8	25
253	Sulfite reductase, dissimilatory-type, subunit DsrJ [†]	3	17

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



267 **FIGURE 6** Models of potential Gallionellaceae extracellular electron transfer mechanisms. All
 268 sizes are approximated. Dimensions of Cyc2 with its fused cytochrome-porin and the porin-
 269 cytochrome complexes MtoAB, MtoAB+MtrC, MtoD and Uet drawn from models and
 270 measurements in previous literature (32, 38, 61–63). Illustration of PCC3 is based on
 271 AlphaFold2 predictions (Fig. S2). The number of hemes and size of PCC3 can vary. The 21/18
 272 heme complex of *S. lithotrophicus* ES-1 is depicted along with the estimated length of the 10 and
 273 35 heme variants of the extracellular cytochrome.

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

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

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

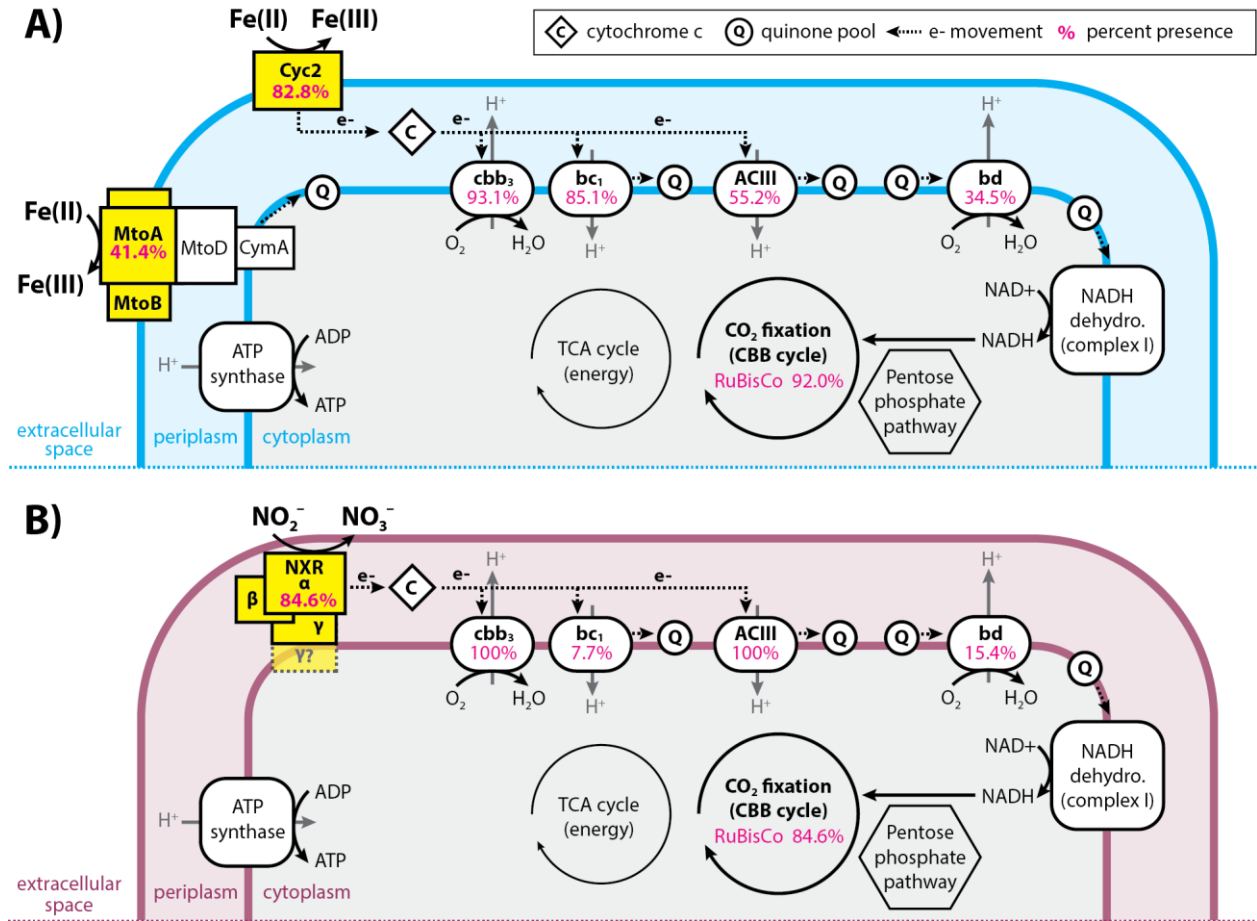
292 **Electron transport chains**

293 We compared electron transport chain component genes of the iron and nitrite oxidizer
294 groups and found them to be largely similar (Fig. 7). High-affinity *cbb*₃-type oxidases are
295 common (Fig. 3), with most genomes containing either the proximal or distal form of *ccoN* (Fig.
296 2) (64). Even the four NRFeOB genomes contain *ccoNO* genes, indicating a potential for both
297 oxygen and nitrate respiration. In contrast, few Gallionellaceae genomes contain *narGH* or
298 *napAB* (6 and 10 genomes, respectively, with no overlap), indicating nitrate respiration is
299 relatively rare overall (Fig. 3, Table S3).

300 In addition to the *cbb₃*-type oxidase genes, 34.5% of iron oxidizers and 15.4% of nitrite
301 oxidizers possess genes for cytochrome *bd*-type oxidases (*cydAB*) (Fig. 7). The presence of *bd*-
302 type oxidase genes often overlaps with *cbb₃*-type oxidase genes (Table S3). Like *cbb₃*-type
303 oxidases, cytochrome *bd*-type oxidases have a high affinity for oxygen and recent studies show
304 they can be more highly expressed than *cbb₃*-type oxidases under low-oxygen, organic-rich
305 conditions (65). Both FeOB and NOB have genes for cytochrome *bc₁* and Alternative complex
306 III (ACIII) quinol oxidase complexes. However, *bc₁* is more common in FeOB (85.1%)
307 compared to NOB (7.7%), while ACIII is more common in NOB (100%) than FeOB (55.2%)
308 (Fig. 7, Table S3). Like the *bd*- and *cbb₃*-type oxidases, presence of *bc₁* and ACIII often overlaps
309 in a single organism, especially in FeOB (Table S3). Possessing both *bd*- and *cbb₃*-type oxidases,
310 and/or having both *bc₁* and ACIII contributes to flexibility within the electron transport chains of
311 Gallionellaceae. The presence of various terminal oxidases implies adaptation to niches where
312 oxygen and organic carbon availability differ or fluctuate.

313 **Carbon fixation**

314 Gallionellaceae isolates grow autotrophically. To determine if the capacity for
315 autotrophic growth is widespread, we analyzed the pangenome for RuBisCo genes (*cbbLS*,
316 *cbbMQ*). Most genomes in the dataset (>91%; 94 of 103 genomes) contain genes for either Form
317 I or Form II RuBisCo (Fig. 3, Table S3). FeOB more commonly have Form II, while NOB only
318 have Form I. Form II enzymes are adapted for medium to high CO₂ and low O₂ concentrations
319 (66) and their predominance in FeOB may correspond to different oxygen niches of FeOB and
320 NOB. The prevalence of RuBisCo genes indicates both iron- and nitrite-oxidizing
321 Gallionellaceae have the capacity to grow autotrophically.



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

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

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

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

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

354 We searched the pangenome for the candidate genes for stalk formation (*sfz/sfb*)
355 identified in the stalk-forming *Ferriphaselus* and Zetaproteobacteria isolates (71, 72). Twelve
356 genomes, restricted to *Gallionella* (9) and *Ferriphaselus* (3) contain the four *sfz/sfb* genes (Fig.
357 2, Table S3). Thus far, all cultured Gallionellaceae stalk formers belong to these two genera,
358 suggesting stalk formation may be limited and not a trait of *Sideroxydans*.

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

376 **TABLE 2** Gene clusters of interest from the Anvi'o pangenome subset that were present in iron-
 377 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 (LoID)	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 (Murl)	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

378 Discussion

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

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

387 **Phylogeny**

388 The Gallionellaceae is composed of four genera: *Gallionella*, *Sideroxydans*,
389 *Ferriphaselus*, and *Ca. Nitrotoga*, based on a concatenated ribosomal protein tree. In
390 comparison, 16S rRNA phylogeny did a poorer job of resolving these genera, so 16S-based
391 identification should be considered tentative, pending availability of genomes. To facilitate
392 consistent classification, the protein sequences and alignments used here (Fig. 1) are available at
393 ([https://figshare.com/projects/Gallionellaceae_Ribosomal_Proteins_for_Concatenated_Tree/157](https://figshare.com/projects/Gallionellaceae_Ribosomal_Proteins_for_Concatenated_Tree/157347)
394 347).

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

405 *Gallionella* and *Ferrigenium*. Moreover, the tree topology suggests continued diversification
406 within both *Gallionella* and *Sideroxydans* largely without the formation of subclades that
407 represent distinct niches. There is one subclade of *Sideroxydans* that corresponds to the GTDB
408 genus level designation PALSA-1006 (Fig. 1). However, ANI/AAI results (Table S5) indicate
409 there is not enough diversity within the Gallionellaceae to justify further splitting the four major
410 genera any further. Additionally, we did not detect any obvious functional difference in PALSA-
411 1006. Given our phylogenetic analysis, ANI/AAI, and similar functional profiles, we recommend
412 keeping them within *Sideroxydans*. Based on the above classification scheme, most of the
413 genomes (84 of 103) fall into either *Gallionella* or *Sideroxydans*.

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

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

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

432 **Iron oxidation and extracellular electron uptake mechanisms**

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

439 Cyc2 has been shown to oxidize dissolved Fe(II) (27, 32, 37). The monoheme Cyc2 is a
440 small fused cytochrome-porin and since aqueous Fe²⁺ is common to many redox transition zones,
441 it makes sense that most FeOB would retain and use the simplest tool. But in Earth's various
442 environments, iron is largely available as minerals (clays, oxides, sulfides) and also bound to
443 organics (e.g. humic substances). The decaheme MtoA has been shown to play roles in the
444 oxidation of mineral-bound Fe(II), specifically Fe(II) smectite clay (37). As a MHC, MtoA may
445 have multiple benefits that help in oxidizing minerals. MtoA has a large redox potential window
446 (-350 to +30 mV; (31, 60)), which could help with oxidation of solids, like smectite, that also
447 have a range of redox potentials (e.g., -600 to +0 mV for SWa-1 vs. -400 to +400 mV for SWy-
448 2; (79)), which change as mineral-bound iron is oxidized or reduced. Assuming the MtoA
449 structure is similar to MtrA, the ten hemes span the membrane, making a wire that conducts
450 from extracellular substrates to periplasmic proteins (62, 80). The multiple hemes allow for

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

457 Like MtrCAB, the predicted PCC3 complex includes both periplasmic and extracellular
458 MHCs and a porin. A key difference is that the PCC3 cytochromes often have more hemes than
459 MtoA/MtrA and MtrC. The greater number of hemes may serve to store electrons, as in a
460 capacitor. They may also conduct across a greater distance; the PCC3 periplasmic MHC, with
461 21-27 hemes, is potentially long enough to span the entire periplasm (as noted by Edwards et al.,
462 (84)). Intraprotein electron transfer between hemes is rapid (85–87); therefore the periplasm-
463 spanning MHC of PCC3 may allow for faster electron transfer compared to complexes
464 containing smaller periplasmic cytochromes like the monoheme MtoD. The extracellular PCC3
465 MHC contains between 10 and 35 hemes, which could extend further from the outer membrane
466 compared to MtrC. Not only would this extend the range of electron transfer, but may also be
467 faster than a “wire” of smaller cytochromes (e.g. *Geobacter* hexaheme OmcS (88)). Increasing
468 oxidation rates via larger MHCs would allow FeOB to oxidize substrates faster. Given that Fe(II)
469 is subject to abiotic oxidation under certain conditions and other organisms may compete for
470 EEU, such kinetic advantages would give FeOB a competitive edge.

471 **Conclusions**

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

474 89, 90). Although this may contribute to an impression that the niche was relatively restricted,
475 16S rRNA sequencing of cultures and environmental samples has revealed both the diversity of
476 Gallionellaceae as well as its prevalence across practically any freshwater and some brackish
477 environments where Fe(II) and O₂ meet. The pangenome shows that Gallionellaceae possess
478 metabolic flexibility to use non-iron substrates, notably sulfur, and the MHCs likely also confer
479 further metabolic capabilities that may help them occupy a range of different iron- and mineral-
480 rich niches. Gallionellaceae thrive in aquifers, soil, and wetlands, all of which have substantial
481 mineral content. Thus, the widespread ecological success of Gallionellaceae may well
482 correspond to genomes that encode a range of iron oxidation mechanisms as well as adaptations
483 for varied environments.

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

496 **Methods**

497 **Data collection and curation**

498 Gallionellaceae genomes were collected from the National Center for Biotechnology
499 Information (NCBI) Entrez database (93), the Joint Genome Institute Integrated Microbial
500 Genomes (JGI IMG) database (94), and the European Nucleotide Archive (ENA) at EMBL-EBI
501 database (*Sideroxydans* sp. CL21, *Ca. Nitrotoga fabula* KNB, and the “IN” MAGs (17, 43, 95))
502 (Table S6). We also received non-public genomes from the Ménez Lab at the Université de Paris
503 (3 genomes reconstructed by Aurélien Lecoivre from the Carbfix study in Hengill, Iceland (2);
504 metagenomes available at Sequence Read Archive SRR3731039, SRR3731040, SRR4188484,
505 and SRR4188643), and the Banfield Lab at the University of California, Berkeley (3 genomes
506 reconstructed by Alex Probst from Crystal Geysir in Utah, USA (96)) (Table S6). This initial
507 230-genome dataset included isolate genomes, metagenome assembled genomes (MAGs), and
508 single-cell amplified genomes (SAGs) that were taxonomically classified as members of the
509 Gallionellales order; Gallionellaceae family; or the *Gallionella*, *Sideroxydans*, *Ferriphaseelus*,
510 *Ferrigenium*, or *Ca. Nitrotoga* genera in their respective databases. Duplicate genomes were
511 identified and removed if they had identical accession numbers or their average nucleotide
512 identities (ANI) were 100%. CheckM v1.1.2 (97) was used to assess genome quality. Genomes
513 with lower than 80% completeness and greater than 7% contamination were removed from the
514 dataset. The final filtered dataset, referred to as “the Gallionellaceae” or “the dataset” contained
515 103 genomes, including six of the Gallionellaceae FeOB isolates (Table S1). The seventh isolate,
516 *Sideroxyarcus emersonii* (26), was not published at the time of our main analysis, but a

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

519 **Naming conventions**

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

527 **Ecosystem classifications**

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

536 **Calculation of average amino acid and nucleotide identities**

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

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

541 **Tree construction**

542 **Concatenated ribosomal protein tree**

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

555 **16S rRNA gene tree**

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

561 genes were aligned in SINA (109) along with Gallionellaceae sequences from the Silva database
562 (110) that had >1475 bp and >85-90 sequence quality score. The outgroup is composed of
563 *Thiobacillus*, *Ferritrophicum*, *Sulfuricella*, *Sulfuriferula*, and *Nitrosomonas* sequences acquired
564 from the Silva database. The final alignment contained 965 non-redundant sequences and
565 alignment length was 1500 positions after trimming and masking all sequence gaps greater than
566 70%. A maximum likelihood tree was constructed using RAxML-NG v1.0.3 (107) with the
567 GTR+G model and 300 bootstraps. Family and genus level classifications from the SILVA
568 database were used to annotate the tree in Iroki (111).

569 **Individual protein trees**

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

578 **Pangenome analysis**

579 **Metabolic gene analysis**

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

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

587 **Multiheme cytochrome analysis**

588 To identify potential *c*-type cytochromes we used a modified heme counter script (54) to
589 search for CXXCH, CXXXCH and CXXXXCH motifs within the protein sequences of each
590 genome. The search identified 5,929 protein sequences with one or more CX₂₋₄CH-motifs. To
591 determine which protein sequences were shared between genomes, sequences were clustered
592 using MMSeqs2 (112) with coverage mode 0 for bidirectional coverage of at least 80% of the
593 query and target sequences. Several clusters of interest were identified based on either the
594 number of CX₂₋₄CH-motifs in each sequence or the relative abundance of FeOB sequences in the
595 cluster. Querying with BLASTp (52) against the Uniprot (113) database was used to classify
596 sequences from clusters of interest thereby identifying clusters of predicted *c*-type cytochromes.
597 Isolate sequences were used as representative sequences for cluster classification. If a cluster did
598 not contain an isolate sequence, a consensus classification was used. The subcellular localization
599 of proteins was predicted using a combination of PSORTb v3.0.3 (114) and LocTree3 (115).

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

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

611 **PCC3 modeling**

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

617 **Anvi’o subset analysis**

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

629 **Acknowledgements**

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

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

642 We declare that we have no conflicts of interest.

643 **References**

- 644 1. Emerson D, Fleming EJ, McBeth JM. 2010. Iron-Oxidizing Bacteria: An Environmental
645 and Genomic Perspective. *Annu Rev Microbiol* 64:561–583.
- 646 2. Trias R, Ménez B, le Campion P, Zivanovic Y, Lecourt L, Lecoeuvre A, Schmitt-Kopplin
647 P, Uhl J, Gislason SR, Alfreðsson HA, Mesfin KG, Snæbjörnsdóttir SÓ, Aradóttir ES,
648 Gunnarsson I, Matter JM, Stute M, Oelkers EH, Gérard E. 2017. High reactivity of deep
649 biota under anthropogenic CO₂ injection into basalt. *Nat Commun* 8:1063.
- 650 3. Jewell TNM, Karaoz U, Brodie EL, Williams KH, Beller HR. 2016. Metatranscriptomic
651 evidence of pervasive and diverse chemolithoautotrophy relevant to C, S, N and Fe cycling
652 in a shallow alluvial aquifer. *9. The ISME Journal* 10:2106–2117.
- 653 4. Emerson D, Moyer CL. 2002. Neutrophilic Fe-Oxidizing Bacteria Are Abundant at the
654 Loihi Seamount Hydrothermal Vents and Play a Major Role in Fe Oxide Deposition.
655 *Applied and Environmental Microbiology* 68:3085–3093.
- 656 5. Emerson D, Vet WD. 2015. The Role of FeOB in Engineered Water Ecosystems: A
657 Review. *Journal AWWA* 107:E47–E57.
- 658 6. Wang J, Sickinger M, Ciobota V, Herrmann M, Rasch H, Rösch P, Popp J, Küsel K. 2014.
659 Revealing the microbial community structure of clogging materials in dewatering wells
660 differing in physico-chemical parameters in an open-cast mining area. *Water Research*
661 63:222–233.

- 662 7. Kato S, Krepski S, Chan C, Itoh T, Ohkuma M. 2014. *Ferriphaselus amnicola* gen. nov., sp.
663 nov., a neutrophilic, stalk-forming, iron-oxidizing bacterium isolated from an iron-rich
664 groundwater seep. *International Journal of Systematic and Evolutionary Microbiology*
665 64:921–925.
- 666 8. Fabisch M, Beulig F, Akob DM, Küsel K. 2013. Surprising abundance of *Gallionella*-
667 related iron oxidizers in creek sediments at pH 4.4 or at high heavy metal concentrations.
668 *Front Microbiol* 4.
- 669 9. Fabisch M, Freyer G, Johnson CA, Büchel G, Akob DM, Neu TR, Küsel K. 2016.
670 Dominance of ‘*Gallionella capsiferiformans*’ and heavy metal association with
671 *Gallionella*-like stalks in metal-rich pH 6 mine water discharge. *Geobiology* 14:68–90.
- 672 10. Weiss JV, Rentz JA, Plaia T, Neubauer SC, Merrill-Floyd M, Lilburn T, Bradburne C,
673 Megonigal JP, Emerson D. 2007. Characterization of Neutrophilic Fe(II)-Oxidizing
674 Bacteria Isolated from the Rhizosphere of Wetland Plants and Description of
675 *Ferritrophicum radicolica* gen. nov. sp. nov., and *Sideroxydans paludicola* sp. nov.
676 *Geomicrobiology Journal* 24:559–570.
- 677 11. Lüdecke C, Reiche M, Eusterhues K, Nietzsche S, Küsel K. 2010. Acid-tolerant
678 microaerophilic Fe(II)-oxidizing bacteria promote Fe(III)-accumulation in a fen: Acid-
679 tolerant Fe(II)-oxidizers in a fen. *Environmental Microbiology* 12:2814–2825.
- 680 12. Buongiorno J, Herbert LC, Wehrmann LM, Michaud AB, Laufer K, Røy H, Jørgensen BB,
681 Szykiewicz A, Faiia A, Yeager KM, Schindler K, Lloyd KG. 2019. Complex Microbial

- 682 Communities Drive Iron and Sulfur Cycling in Arctic Fjord Sediments. *Appl Environ*
683 *Microbiol* 85:e00949-19, /aem/85/14/AEM.00949-19.atom.
- 684 13. Patzner MS, Logan M, McKenna AM, Young RB, Zhou Z, Joss H, Mueller CW, Hoeschen
685 C, Scholten T, Straub D, Kleindienst S, Borch T, Kappler A, Bryce C. 2022. Microbial iron
686 cycling during palsa hillslope collapse promotes greenhouse gas emissions before complete
687 permafrost thaw. *Commun Earth Environ* 3:1–14.
- 688 14. Neubauer SC, Emerson D, Megonigal JP. 2002. Life at the Energetic Edge: Kinetics of
689 Circumneutral Iron Oxidation by Lithotrophic Iron-Oxidizing Bacteria Isolated from the
690 Wetland-Plant Rhizosphere. *Applied and Environmental Microbiology* 68:3988–3995.
- 691 15. de Vet W w. j. m., Dinkla I j. t., Abbas B a., Rietveld L c., van Loosdrecht M c. m. 2012.
692 *Gallionella* spp. in trickling filtration of subsurface aerated and natural groundwater.
693 *Biotechnology and Bioengineering* 109:904–912.
- 694 16. Eichinger S, Boch R, Leis A, Koraimann G, Grengg C, Domberger G, Nachtnebel M,
695 Schwab C, Dietzel M. 2020. Scale deposits in tunnel drainage systems – A study on fabrics
696 and formation mechanisms. *Science of The Total Environment* 718:137140.
- 697 17. Bethencourt L, Bochet O, Farasin J, Aquilina L, Borgne TL, Quaiser A, Biget M, Michon-
698 Coudouel S, Labasque T, Dufresne A. 2020. Genome reconstruction reveals distinct
699 assemblages of *Gallionellaceae* in surface and subsurface redox transition zones. *FEMS*
700 *Microbiol Ecol* 96.

- 701 18. Emerson D, Revsbech NP. 1994. Investigation of an Iron-Oxidizing Microbial Mat
702 Community Located near Aarhus, Denmark: Field Studies. *Applied and Environmental*
703 *Microbiology* 60:4022–4031.
- 704 19. Kappler A, Bryce C. 2017. Cryptic biogeochemical cycles: unravelling hidden redox
705 reactions. *Environmental Microbiology* 19:842–846.
- 706 20. Ehrenberg C. 1838. *Die infusionsthierchen Als vollkommene organismen*. Leipzig: Leopold
707 Voss.
- 708 21. Ghiorse WC. 1984. Biology of iron- and manganese-depositing bacteria. *Annual Review of*
709 *Microbiology* 38:515–550.
- 710 22. Hallbeck L, Pedersen K. 1991. Autotrophic and mixotrophic growth of *Gallionella*
711 *ferruginea*. *Microbiology*, 137:2657–2661.
- 712 23. Krepski ST, Hanson TE, Chan CS. 2012. Isolation and characterization of a novel
713 biomineral stalk-forming iron-oxidizing bacterium from a circumneutral groundwater seep:
714 A novel stalk-forming Fe-oxidizing bacterium. *Environmental Microbiology* 14:1671–
715 1680.
- 716 24. Emerson D, Moyer C. 1997. Isolation and Characterization of Novel Iron-Oxidizing
717 Bacteria That Grow at Circumneutral pH. *Applied and Environmental Microbiology* 63:9.
- 718 25. Khalifa A, Nakasuji Y, Saka N, Honjo H, Asakawa S, Watanabe T. 2018. *Ferrigenium*
719 *kumadai* gen. nov., sp. nov., a microaerophilic iron-oxidizing bacterium isolated from a

- 720 paddy field soil. *International Journal of Systematic and Evolutionary Microbiology*
721 68:2587–2592.
- 722 26. Kato S, Itoh T, Iino T, Ohkuma M. 2022. *Sideroxyarcus emersonii* gen. nov. sp. nov., a
723 neutrophilic, microaerobic iron- and thiosulfate-oxidizing bacterium isolated from iron-rich
724 wetland sediment. *International Journal of Systematic and Evolutionary Microbiology* 72.
- 725 27. Zhou N, Keffer JL, Polson SW, Chan CS. 2022. Unraveling Fe(II)-Oxidizing Mechanisms
726 in a Facultative Fe(II) Oxidizer, *Sideroxydans lithotrophicus* Strain ES-1, via Culturing,
727 Transcriptomics, and Reverse Transcription-Quantitative PCR. *Applied and Environmental*
728 *Microbiology* 88:e01595-21.
- 729 28. Cooper RE, Finck J, Chan C, Küsel K. 2023. Mixotrophy broadens the ecological niche
730 range of the iron oxidizer *Sideroxydans* sp. CL21 isolated from an iron-rich peatland.
731 *FEMS Microbiology Ecology* fiac156.
- 732 29. Huang Y-M, Jakus N, Straub D, Konstantinidis KT, Blackwell N, Kappler A, Kleindienst
733 S. 2022. ‘*Candidatus ferrigenium straubiae*’ sp. nov., ‘*Candidatus ferrigenium bremense*’
734 sp. nov., ‘*Candidatus ferrigenium altingense*’ sp. nov., are autotrophic Fe(II)-oxidizing
735 bacteria of the family Gallionellaceae. *Systematic and Applied Microbiology* 45:126306.
- 736 30. Straub KL, Benz M, Schink B, Widdel F. 1996. Anaerobic, nitrate-dependent microbial
737 oxidation of ferrous iron. *Applied and Environmental Microbiology* 62:1458–1460.
- 738 31. Liu J, Wang Z, Belchik SM, Edwards MJ, Liu C, Kennedy DW, Merkley ED, Lipton MS,
739 Butt JN, Richardson DJ, Zachara JM, Fredrickson JK, Rosso KM, Shi L. 2012.
740 Identification and Characterization of MtoA: A Decaheme c-Type Cytochrome of the

- 741 Neutrophilic Fe(II)-Oxidizing Bacterium *Sideroxydans lithotrophicus* ES-1. *Front Microbio*
742 3.
- 743 32. Keffer JL, McAllister SM, Garber AI, Hallahan BJ, Sutherland MC, Rozovsky S, Chan CS.
744 2021. Iron Oxidation by a Fused Cytochrome-Porin Common to Diverse Iron-Oxidizing
745 Bacteria. *mBio* 0:e01074-21.
- 746 33. McAllister SM, Polson SW, Butterfield DA, Glazer BT, Sylvan JB, Chan CS. 2020.
747 Validating the *Cyc2* Neutrophilic Iron Oxidation Pathway Using Meta-omics of
748 *Zetaproteobacteria* Iron Mats at Marine Hydrothermal Vents. *mSystems* 5:e00553-19,
749 /msystems/5/1/msys.00553-19.atom.
- 750 34. Jiao Y, Newman DK. 2007. The *pio* Operon Is Essential for Phototrophic Fe(II) Oxidation
751 in *Rhodospseudomonas palustris* TIE-1. *Journal of Bacteriology* 189:1765–1773.
- 752 35. Bose A, Gardel EJ, Vidoudez C, Parra EA, Girguis PR. 2014. Electron uptake by iron-
753 oxidizing phototrophic bacteria. *Nat Commun* 5:3391.
- 754 36. Beliaev AS, Saffarini DA. 1998. *Shewanella putrefaciens* *mtrB* Encodes an Outer
755 Membrane Protein Required for Fe(III) and Mn(IV) Reduction. *Journal of Bacteriology*
756 180:6292–6297.
- 757 37. Zhou N, Kupper RJ, Catalano JG, Thompson A, Chan CS. 2022. Biological Oxidation of
758 Fe(II)-Bearing Smectite by Microaerophilic Iron Oxidizer *Sideroxydans lithotrophicus*
759 Using Dual *Mto* and *Cyc2* Iron Oxidation Pathways. *Environ Sci Technol*
760 <https://doi.org/10.1021/acs.est.2c05142>.

- 761 38. He S, Barco RA, Emerson D, Roden EE. 2017. Comparative Genomic Analysis of
762 Neutrophilic Iron(II) Oxidizer Genomes for Candidate Genes in Extracellular Electron
763 Transfer. *Front Microbiol* 8.
- 764 39. Alawi M, Lipski A, Sanders T, Eva-Maria-Pfeiffer, Spieck E. 2007. Cultivation of a novel
765 cold-adapted nitrite oxidizing betaproteobacterium from the Siberian Arctic. 3. *The ISME*
766 *Journal* 1:256–264.
- 767 40. Ishii K, Fujitani H, Soh K, Nakagawa T, Takahashi R, Tsuneda S. 2017. Enrichment and
768 Physiological Characterization of a Cold-Adapted Nitrite-Oxidizing *Nitrotoga* sp. from an
769 Eelgrass Sediment. *Applied and Environmental Microbiology* 83:e00549-17,
770 /aem/83/14/e00549-17.atom.
- 771 41. Podowski JC, Paver SF, Newton RJ, Coleman ML. 2022. Genome Streamlining,
772 Proteorhodopsin, and Organic Nitrogen Metabolism in Freshwater Nitrifiers. *mBio*
773 13:e02379-21.
- 774 42. Boddicker AM, Mosier AC. 2018. Genomic profiling of four cultivated *Candidatus*
775 *Nitrotoga* spp. predicts broad metabolic potential and environmental distribution. *The ISME*
776 *Journal* 12:2864–2882.
- 777 43. Kitzinger K, Koch H, Lückner S, Sedlacek CJ, Herbold C, Schwarz J, Daebeler A, Mueller
778 AJ, Lukumbuzya M, Romano S, Leisch N, Karst SM, Kirkegaard R, Albertsen M, Nielsen
779 PH, Wagner M, Daims H. 2018. Characterization of the First “*Candidatus Nitrotoga*”
780 Isolate Reveals Metabolic Versatility and Separate Evolution of Widespread Nitrite-
781 Oxidizing Bacteria. *mBio* 9:e01186-18, /mbio/9/4/mBio.01186-18.atom.

- 782 44. Lückner S, Schwarz J, Gruber-Dorninger C, Spieck E, Wagner M, Daims H. 2015. Nitrotoga
783 -like bacteria are previously unrecognized key nitrite oxidizers in full-scale wastewater
784 treatment plants. *3. The ISME Journal* 9:708–720.
- 785 45. Ishii K, Fujitani H, Sekiguchi Y, Tsuneda S. 2020. Physiological and genomic
786 characterization of a new ‘Candidatus Nitrotoga’ isolate. *Environmental Microbiology* 22.
- 787 46. Daims H, Lückner S, Wagner M. 2016. A New Perspective on Microbes Formerly Known as
788 Nitrite-Oxidizing Bacteria. *Trends in Microbiology* 24:699–712.
- 789 47. Ivanova N, Tringe SG, Liolios K, Liu W-T, Morrison N, Hugenholtz P, Kyrpides NC.
790 2010. A call for standardized classification of metagenome projects. *Environmental*
791 *Microbiology* 12:1803–1805.
- 792 48. Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa SL, Solden LM, Liu P,
793 Narrowe AB, Rodríguez-Ramos J, Bolduc B, Gazitúa MC, Daly RA, Smith GJ, Vik DR,
794 Pope PB, Sullivan MB, Roux S, Wrighton KC. 2020. DRAM for distilling microbial
795 metabolism to automate the curation of microbiome function. *Nucleic Acids Res* 48:8883–
796 8900.
- 797 49. Garber AI, Neilson KH, Okamoto A, McAllister SM, Chan CS, Barco RA, Merino N.
798 2020. FeGenie: A Comprehensive Tool for the Identification of Iron Genes and Iron Gene
799 Neighborhoods in Genome and Metagenome Assemblies. *Frontiers in Microbiology* 11.
- 800 50. Garber AI. 2020. MagicLamp: toolkit for annotation of ’omics datasets using curated HMM
801 sets (2020: MagicLamp). GitHub repository. [https://github.com/Arkadiy-](https://github.com/Arkadiy-Garber/MagicLamp)
802 [Garber/MagicLamp](https://github.com/Arkadiy-Garber/MagicLamp).

- 803 51. McAllister SM. 2016. Heme Counter. GitHub repository.
804 https://github.com/seanmcallister/heme_counter.
- 805 52. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search
806 tool. *J Mol Biol* 215:403–410.
- 807 53. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009.
808 BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.
- 809 54. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. 2015.
810 Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* 3:e1319.
- 811 55. Shaiber A, Willis AD, Delmont TO, Roux S, Chen L-X, Schmid AC, Yousef M, Watson
812 AR, Lolans K, Esen ÖC, Lee STM, Downey N, Morrison HG, Dewhirst FE, Mark Welch
813 JL, Eren AM. 2020. Functional and genetic markers of niche partitioning among enigmatic
814 members of the human oral microbiome. *Genome Biology* 21:292.
- 815 56. Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, Fink I, Pan JN, Yousef
816 M, Fogarty EC, Trigodet F, Watson AR, Esen ÖC, Moore RM, Clayssen Q, Lee MD,
817 Kivenson V, Graham ED, Merrill BD, Karkman A, Blankenberg D, Eppley JM, Sjödin A,
818 Scott JJ, Vázquez-Campos X, McKay LJ, McDaniel EA, Stevens SLR, Anderson RE,
819 Fuessel J, Fernandez-Guerra A, Maignien L, Delmont TO, Willis AD. 2021. Community-
820 led, integrated, reproducible multi-omics with anvi'o. 1. *Nature Microbiology* 6:3–6.
- 821 57. Baker IR, Conley BE, Gralnick JA, Girguis PR. 2022. Evidence for Horizontal and Vertical
822 Transmission of Mtr-Mediated Extracellular Electron Transfer among the Bacteria. *mBio*
823 <https://doi.org/10.1128/mbio.02904-21>.

- 824 58. Ross DE, Flynn JM, Baron DB, Gralnick JA, Bond DR. 2011. Towards Electrosynthesis in
825 *Shewanella*: Energetics of Reversing the Mtr Pathway for Reductive Metabolism. PLOS
826 ONE 6:e16649.
- 827 59. Salgueiro CA, Dantas JM. 2016. *Multiheme Cytochromes*, 1st ed. Springer Berlin
828 Heidelberg, Berlin, Heidelberg. <http://link.springer.com/10.1007/978-3-642-44961-1>.
- 829 60. Paquete CM, Morgado L, Salgueiro CA, Louro RO. 2022. Molecular Mechanisms of
830 Microbial Extracellular Electron Transfer: The Importance of Multiheme Cytochromes. 6.
831 *Frontiers in Bioscience-Landmark* 27:174.
- 832 61. Eddie BJ, Bird LJ, Pelikan C, Mussmann M, Martínez-Pérez C, Pinamang P, Malanoski
833 AP, Glaven SM. 2022. Conservation of Energetic Pathways for Electroautotrophy in the
834 Uncultivated Candidate Order Tenderiales. *mSphere* 0:e00223-22.
- 835 62. Edwards MJ, White GF, Butt JN, Richardson DJ, Clarke TA. 2020. The Crystal Structure
836 of a Biological Insulated Transmembrane Molecular Wire. *Cell* 181:665-673.e10.
- 837 63. Beckwith CR, Edwards MJ, Lawes M, Shi L, Butt JN, Richardson DJ, Clarke TA. 2015.
838 Characterization of MtoD from *Sideroxydans lithotrophicus*: a cytochrome c electron
839 shuttle used in lithoautotrophic growth. *Frontiers in Microbiology* 6.
- 840 64. Ducluzeau A-L, Ouchane S, Nitschke W. 2008. The *cbb3* Oxidases Are an Ancient
841 Innovation of the Domain Bacteria. *Molecular Biology and Evolution* 25:1158–1166.
- 842 65. Trojan D, Garcia-Robledo E, Meier DV, Hausmann B, Revsbech NP, Eichorst SA,
843 Wobken D. 2021. Microaerobic Lifestyle at Nanomolar O₂ Concentrations Mediated by

- 844 Low-Affinity Terminal Oxidases in Abundant Soil Bacteria. *mSystems*
845 <https://doi.org/10.1128/mSystems.00250-21>.
- 846 66. Badger MR, Bek EJ. 2008. Multiple Rubisco forms in proteobacteria: their functional
847 significance in relation to CO₂ acquisition by the CBB cycle. *J Exp Bot* 59:1525–1541.
- 848 67. Griesbeck C, Schütz M, Schödl T, Bathe S, Nausch L, Mederer N, Vielreicher M, Hauska
849 G. 2002. Mechanism of Sulfide-Quinone Reductase Investigated Using Site-Directed
850 Mutagenesis and Sulfur Analysis. *Biochemistry* 41:11552–11565.
- 851 68. Marcia M, Ermler U, Peng G, Michel H. 2009. The structure of *Aquifex aeolicus*
852 sulfide:quinone oxidoreductase, a basis to understand sulfide detoxification and respiration.
853 *Proceedings of the National Academy of Sciences* 106:9625–9630.
- 854 69. Loy A, Duller S, Baranyi C, Mußmann M, Ott J, Sharon I, Béjà O, Le Paslier D, Dahl C,
855 Wagner M. 2009. Reverse dissimilatory sulfite reductase as phylogenetic marker for a
856 subgroup of sulfur-oxidizing prokaryotes. *Environmental Microbiology* 11:289–299.
- 857 70. Müller AL, Kjeldsen KU, Rattei T, Pester M, Loy A. 2015. Phylogenetic and
858 environmental diversity of DsrAB-type dissimilatory (bi)sulfite reductases. *5. ISME J*
859 9:1152–1165.
- 860 71. Kato S, Ohkuma M, Powell DH, Krepiski ST, Oshima K, Hattori M, Shapiro N, Woyke T,
861 Chan CS. 2015. Comparative Genomic Insights into Ecophysiology of Neutrophilic,
862 Microaerophilic Iron Oxidizing Bacteria. *Front Microbiol* 6.

- 863 72. Koeksoy E, Bezuidt OM, Bayer T, Chan CS, Emerson D. 2021. Zetaproteobacteria Pan-
864 Genome Reveals Candidate Gene Cluster for Twisted Stalk Biosynthesis and Export. *Front*
865 *Microbiol* 12:1374.
- 866 73. Schmehl M, Jahn A, Meyer zu Vilsendorf A, Hennecke S, Masepohl B, Schuppler M,
867 Marxer M, Oelze J, Klipp W. 1993. Identification of a new class of nitrogen fixation genes
868 in *Rhodobacter capsalatus*: a putative membrane complex involved in electron transport to
869 nitrogenase. *Molec Gen Genet* 241:602–615.
- 870 74. Tremblay P-L, Zhang T, Dar SA, Leang C, Lovley DR. 2012. The Rnf Complex of
871 *Clostridium ljungdahlii* Is a Proton-Translocating Ferredoxin:NAD⁺ Oxidoreductase
872 Essential for Autotrophic Growth. *mBio* 4:e00406-12.
- 873 75. Downing BE, Gupta D, Nayak DD. 2022. The dual role of a multi-heme cytochrome in
874 methanogenesis: MmcA is important for energy conservation and carbon metabolism in
875 *Methanosarcina acetivorans*. *bioRxiv* <https://doi.org/10.1101/2022.07.20.500911>.
- 876 76. Liu Y, Chen H, Van Treuren W, Hou B-H, Higginbottom SK, Dodd D. 2022. *Clostridium*
877 *sporogenes* uses reductive Stickland metabolism in the gut to generate ATP and produce
878 circulating metabolites. *Nat Microbiol* 7:695–706.
- 879 77. Child SA, Bradley JM, Pukala TL, Svistunenko DA, Le Brun NE, Bell SG. 2018. Electron
880 transfer ferredoxins with unusual cluster binding motifs support secondary metabolism in
881 many bacteria. *Chem Sci* 9:7948–7957.

- 882 78. Watanabe T, Kojima H. 2019. Sulfuricella, p. 1–6. *In* Whitman, WB, Rainey, F, Kämpfer,
883 P, Trujillo, M, Chun, J, DeVos, P, Hedlund, B, Dedysh, S (eds.), *Bergey’s Manual of*
884 *Systematics of Archaea and Bacteria*, 1st ed. Wiley.
- 885 79. Gorski CA, Klüpfel LE, Voegelin A, Sander M, Hofstetter TB. 2013. Redox Properties of
886 Structural Fe in Clay Minerals: 3. Relationships between Smectite Redox and Structural
887 Properties. *Environ Sci Technol* 47:13477–13485.
- 888 80. Gupta D, Sutherland MC, Rengasamy K, Meacham JM, Kranz RG, Bose A. 2019.
889 Photoferrotrophs Produce a PioAB Electron Conduit for Extracellular Electron Uptake.
890 *mBio* 10:e02668-19.
- 891 81. Coursolle D, Baron DB, Bond DR, Gralnick JA. 2010. The Mtr Respiratory Pathway Is
892 Essential for Reducing Flavins and Electrodes in *Shewanella oneidensis*. *Journal of*
893 *Bacteriology* 192:467–474.
- 894 82. Beliaev AS, Saffarini DA, McLaughlin JL, Hunnicutt D. 2001. MtrC, an outer membrane
895 decahaem c cytochrome required for metal reduction in *Shewanella putrefaciens* MR-1.
896 *Molecular Microbiology* 39:722–730.
- 897 83. Gorby YA, Yanina S, McLean JS, Rosso KM, Moyles D, Dohnalkova A, Beveridge TJ,
898 Chang IS, Kim BH, Kim KS, Culley DE, Reed SB, Romine MF, Saffarini DA, Hill EA, Shi
899 L, Elias DA, Kennedy DW, Pinchuk G, Watanabe K, Ishii S, Logan B, Nealon KH,
900 Fredrickson JK. 2006. Electrically conductive bacterial nanowires produced by *Shewanella*
901 *oneidensis* strain MR-1 and other microorganisms. *Proceedings of the National Academy of*
902 *Sciences* 103:11358–11363.

- 903 84. Edwards MJ, Richardson DJ, Paquete CM, Clarke TA. 2020. Role of multiheme
904 cytochromes involved in extracellular anaerobic respiration in bacteria. *Protein Science*
905 29:830–842.
- 906 85. Richardson DJ, Butt JN, Fredrickson JK, Zachara JM, Shi L, Edwards MJ, White G, Baiden
907 N, Gates AJ, Marritt SJ, Clarke TA. 2012. The ‘porin–cytochrome’ model for microbe-to-
908 mineral electron transfer. *Molecular Microbiology* 85:201–212.
- 909 86. White GF, Shi Z, Shi L, Wang Z, Dohnalkova AC, Marshall MJ, Fredrickson JK, Zachara
910 JM, Butt JN, Richardson DJ, Clarke TA. 2013. Rapid electron exchange between surface-
911 exposed bacterial cytochromes and Fe(III) minerals. *Proceedings of the National Academy*
912 *of Sciences* 110:6346–6351.
- 913 87. van Wonderen JH, Adamczyk K, Wu X, Jiang X, Piper SEH, Hall CR, Edwards MJ, Clarke
914 TA, Zhang H, Jeuken LJC, Sazanovich IV, Towrie M, Blumberger J, Meech SR, Butt JN.
915 2021. Nanosecond heme-to-heme electron transfer rates in a multiheme cytochrome
916 nanowire reported by a spectrally unique His/Met-ligated heme. *Proceedings of the*
917 *National Academy of Sciences* 118:e2107939118.
- 918 88. Wang F, Gu Y, O’Brien JP, Yi SM, Yalcin SE, Srikanth V, Shen C, Vu D, Ing NL,
919 Hochbaum AI, Egelman EH, Malvankar NS. 2019. Structure of Microbial Nanowires
920 Reveals Stacked Hemes that Transport Electrons over Micrometers. *Cell* 177:361-369.e10.
- 921 89. Chan CS, McAllister SM, Leavitt AH, Glazer BT, Krepski ST, Emerson D. 2016. The
922 Architecture of Iron Microbial Mats Reflects the Adaptation of Chemolithotrophic Iron
923 Oxidation in Freshwater and Marine Environments. *Front Microbiol* 7.

- 924 90. Hallbeck L, Pedersen K. 2014. The Family Gallionellaceae, p. 853–858. *In* Rosenberg, E,
925 DeLong, EF, Lory, S, Stackebrandt, E, Thompson, F (eds.), *The Prokaryotes:*
926 *Alphaproteobacteria and Betaproteobacteria*. Springer, Berlin, Heidelberg.
- 927 91. Hedrich S, Schlömann M, Johnson DB. 2011. The iron-oxidizing proteobacteria.
928 *Microbiology* 157:1551–1564.
- 929 92. Ilbert M, Bonnefoy V. 2013. Insight into the evolution of the iron oxidation pathways.
930 *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1827:161–175.
- 931 93. Geer LY, Marchler-Bauer A, Geer RC, Han L, He J, He S, Liu C, Shi W, Bryant SH. 2010.
932 *The NCBI BioSystems database*. *Nucleic Acids Res* 38:D492–D496.
- 933 94. Chen I-MA, Chu K, Palaniappan K, Ratner A, Huang J, Huntemann M, Hajek P, Ritter S,
934 Varghese N, Seshadri R, Roux S, Woyke T, Eloe-Fadrosh EA, Ivanova NN, Kyrpides NC.
935 2021. The IMG/M data management and analysis system v.6.0: new tools and advanced
936 capabilities. *Nucleic Acids Research* 49:D751–D763.
- 937 95. Cooper RE, Wegner C-E, McAllister SM, Shevchenko O, Chan CS, Küsel K. 2020. Draft
938 *Genome Sequence of Sideroxydans* sp. Strain CL21, an Fe(II)-Oxidizing Bacterium.
939 *Microbiol Resour Announc* 9:e01444-19.
- 940 96. Probst AJ, Ladd B, Jarett JK, Geller-McGrath DE, Sieber CMK, Emerson JB,
941 Anantharaman K, Thomas BC, Malmstrom RR, Stieglmeier M, Klingl A, Woyke T, Ryan
942 MC, Banfield JF. 2018. Differential depth distribution of microbial function and putative
943 symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. *Nat*
944 *Microbiol* 3:328–336.

- 945 97. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM:
946 assessing the quality of microbial genomes recovered from isolates, single cells, and
947 metagenomes. *Genome Res* 25:1043–1055.
- 948 98. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2020. GTDB-Tk: a toolkit to classify
949 genomes with the Genome Taxonomy Database. *Bioinformatics* 36:1925–1927.
- 950 99. Mukherjee S, Stamatis D, Bertsch J, Ovchinnikova G, Sundaramurthi JC, Lee J,
951 Kandimalla M, Chen I-MA, Kyrpides NC, Reddy TBK. 2021. Genomes OnLine Database
952 (GOLD) v.8: overview and updates. *Nucleic Acids Research* 49:D723–D733.
- 953 100. Buttigieg PL, Morrison N, Smith B, Mungall CJ, Lewis SE, the ENVO Consortium. 2013.
954 The environment ontology: contextualising biological and biomedical entities. *Journal of*
955 *Biomedical Semantics* 4:43.
- 956 101. Parks D. 2014. CompareM. Github repository. <https://github.com/dparks1134/CompareM>.
- 957 102. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware
958 D, Perez F, Canon S, Sneddon MW, Henderson ML, Riehl WJ, Murphy-Olson D, Chan SY,
959 Kamimura RT, Kumari S, Drake MM, Brettin TS, Glass EM, Chivian D, Gunter D, Weston
960 DJ, Allen BH, Baumohl J, Best AA, Bowen B, Brenner SE, Bun CC, Chandonia J-M, Chia
961 J-M, Colasanti R, Conrad N, Davis JJ, Davison BH, DeJongh M, Devoid S, Dietrich E,
962 Dubchak I, Edirisinghe JN, Fang G, Faria JP, Frybarger PM, Gerlach W, Gerstein M,
963 Greiner A, Gurtowski J, Haun HL, He F, Jain R, Joachimiak MP, Keegan KP, Kondo S,
964 Kumar V, Land ML, Meyer F, Mills M, Novichkov PS, Oh T, Olsen GJ, Olson R, Parrello
965 B, Pasternak S, Pearson E, Poon SS, Price GA, Ramakrishnan S, Ranjan P, Ronald PC,

- 966 Schatz MC, Seaver SMD, Shukla M, Sutormin RA, Syed MH, Thomason J, Tintle NL,
967 Wang D, Xia F, Yoo H, Yoo S, Yu D. 2018. KBase: The United States Department of
968 Energy Systems Biology Knowledgebase. 7. *Nature Biotechnology* 36:566–569.
- 969 103. Watanabe T, Kojima H, Fukui M. 2015. *Sulfuriferula multivorans* gen. nov., sp. nov.,
970 isolated from a freshwater lake, reclassification of ‘*Thiobacillus plumbophilus*’ as
971 *Sulfuriferula plumbophilus* sp. nov., and description of *Sulfuricellaceae* fam. nov. and
972 *Sulfuricellales* ord. nov. *International Journal of Systematic and Evolutionary*
973 *Microbiology*, 65:1504–1508.
- 974 104. Kojima H, Fukui M. 2010. *Sulfuricella denitrificans* gen. nov., sp. nov., a sulfur-oxidizing
975 autotroph isolated from a freshwater lake. *International Journal of Systematic and*
976 *Evolutionary Microbiology*, 60:2862–2866.
- 977 105. Geneious v.10.2.6. <https://www.geneious.com>.
- 978 106. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
979 throughput. *Nucleic Acids Research* 32:1792–1797.
- 980 107. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast,
981 scalable and user-friendly tool for maximum likelihood phylogenetic inference.
982 *Bioinformatics* 35:4453–4455.
- 983 108. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic
984 tree display and annotation. *Nucleic Acids Research* 49:W293–W296.

- 985 109. Pruesse E, Peplies J, Glöckner FO. 2012. SINA: Accurate high-throughput multiple
986 sequence alignment of ribosomal RNA genes. *Bioinformatics* 28:1823–1829.
- 987 110. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO.
988 2013. The SILVA ribosomal RNA gene database project: improved data processing and
989 web-based tools. *Nucleic Acids Res* 41:D590–D596.
- 990 111. Moore RM, Harrison AO, McAllister SM, Polson SW, Wommack EK. 2020. Iroki:
991 automatic customization and visualization of phylogenetic trees. *PeerJ* e8584.
- 992 112. Steinegger M, Söding J. 2017. MMseqs2 enables sensitive protein sequence searching for
993 the analysis of massive data sets. *Nat Biotechnol* 35:1026–1028.
- 994 113. The UniProt Consortium. 2021. UniProt: the universal protein knowledgebase in 2021.
995 *Nucleic Acids Research* 49:D480–D489.
- 996 114. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster
997 LJ, Brinkman FSL. 2010. PSORTb 3.0: improved protein subcellular localization prediction
998 with refined localization subcategories and predictive capabilities for all prokaryotes.
999 *Bioinformatics* 26:1608–1615.
- 1000 115. Goldberg T, Hecht M, Hamp T, Karl T, Yachdav G, Ahmed N, Altermann U, Angerer P,
1001 Ansorge S, Balasz K, Bernhofer M, Betz A, Cizmadija L, Do KT, Gerke J, Greil R,
1002 Joerdens V, Hastreiter M, Hembach K, Herzog M, Kalemanov M, Kluge M, Meier A, Nasir
1003 H, Neumaier U, Prade V, Reeb J, Sorokoumov A, Troshani I, Vorberg S, Waldraff S, Zierer
1004 J, Nielsen H, Rost B. 2014. LocTree3 prediction of localization. *Nucleic Acids Research*
1005 42:W350–W355.

- 1006 116. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. 2022.
1007 ColabFold: making protein folding accessible to all. 6. *Nat Methods* 19:679–682.
- 1008 117. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K,
1009 Bates R, Židek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A,
1010 Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E,
1011 Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals
1012 O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. 2021. Highly accurate protein
1013 structure prediction with AlphaFold. 7873. *Nature* 596:583–589.
- 1014 118. Evans R, O’Neill M, Pritzel A, Antropova N, Senior A, Green T, Židek A, Bates R,
1015 Blackwell S, Yim J, Ronneberger O, Bodenstein S, Zielinski M, Bridgland A, Potapenko A,
1016 Cowie A, Tunyasuvunakool K, Jain R, Clancy E, Kohli P, Jumper J, Hassabis D. 2022.
1017 Protein complex prediction with AlphaFold-Multimer. *bioRxiv*
1018 <https://doi.org/10.1101/2021.10.04.463034>.
- 1019 119. Schrodinger, LLC. 2010. The PyMOL Molecular Graphics System (2.5.4).
- 1020 120. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal:
1021 prokaryotic gene recognition and translation initiation site identification. *BMC*
1022 *Bioinformatics* 11:119.
- 1023 121. Galperin MY, Kristensen DM, Makarova KS, Wolf YI, Koonin EV. 2019. Microbial
1024 genome analysis: the COG approach. *Brief Bioinform* 20:1063–1070.

1025 122. Galperin MY, Wolf YI, Makarova KS, Vera Alvarez R, Landsman D, Koonin EV. 2021.

1026 COG database update: focus on microbial diversity, model organisms, and widespread

1027 pathogens. *Nucleic Acids Research* 49:D274–D281.

1028