



HAL
open science

A pan-genomic approach reveals novel *Sulfurimonas* clade in the ferruginous meromictic Lake Pavin

Corinne Biderre-Petit, Damien Courtine, Claire Hennequin, Pierre Galand, Stefan Bertilsson, Didier Debroas, Arthur Monjot, Cécile Lepère, Anna-maria Divne, Corentin Hochart

► To cite this version:

Corinne Biderre-Petit, Damien Courtine, Claire Hennequin, Pierre Galand, Stefan Bertilsson, et al.. A pan-genomic approach reveals novel *Sulfurimonas* clade in the ferruginous meromictic Lake Pavin. *Molecular Ecology Resources*, In press, 10.1111/1755-0998.13923 . hal-04381944v2

HAL Id: hal-04381944

<https://hal.science/hal-04381944v2>

Submitted on 10 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Title**

2 A pan-genomic approach reveals novel *Sulfurimonas* clade in the ferruginous meromictic Lake

3 Pavin

4

5 **Running title:** *Sulfurimonas* pan-genomic analysis

6

7 **Author names and affiliation**

8 Corinne Biderre-Petit^{1*}, Damien Courtine¹, Claire Hennequin¹, Pierre E. Galand², Stefan

9 Bertilsson³, Didier Debroas¹, Arthur Monjot¹, Cécile Lepère¹, Anna-Maria Divne⁴, Corentin

10 Hochart²

11 ¹Université Clermont Auvergne, CNRS, Laboratoire Microorganismes : Génome et

12 Environnement, F-63000, Clermont-Ferrand, France

13 ²Sorbonne Universités, CNRS, Laboratoire d'Ecogéochimie des Environnements Benthiques

14 (LECOB), Observatoire Océanologique de Banyuls, Banyuls sur Mer, France

15 ³Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences

16 and Science for Life Laboratory, Sweden

17 ⁴Department of Cell and Molecular Biology, SciLifeLab, Uppsala University, Uppsala, Sweden

18

19 *For correspondence: Corinne Biderre-Petit, 1 Impasse Amélie Murat -Bat BioA 63178 Aubière

20 (France). E-mail: corinne.petit@uca.fr, Tel: +33(0)473405139; Fax+33 (0)473407670.

21 **Abstract**

22 The permanently anoxic waters in meromictic lakes create suitable niches for the growth of
23 bacteria using sulfur metabolisms like sulfur oxidation. In Lake Pavin, the anoxic water mass
24 hosts an active cryptic sulfur cycle that interacts narrowly with iron cycling, however the
25 metabolisms of the microorganisms involved are poorly known. Here we combined
26 metagenomics, single-cell genomics, and pan-genomics to further expand our understanding
27 of the bacteria and the corresponding metabolisms involved in sulfur oxidation in this
28 ferruginous sulfide- and sulfate-poor meromictic lake. We highlighted two new species within
29 the genus *Sulfurimonas* that belong to a novel clade of chemotrophic sulfur oxidizers exclusive
30 to freshwaters. We moreover conclude that this genus holds a key-role not only in limiting
31 sulfide accumulation in the upper part of the anoxic layer but also constraining carbon,
32 phosphate, and iron cycling.

33

34 **Keywords:** Meromictic lake, freshwater, sulfur oxidation, *Sulfurimonas*, pan-genome.

35

36 **Introduction**

37 Sulfur cycling is an important driver of microbial growth and influences element conversions
38 in most environments (Berben et al., 2019). In the absence of oxygen, the sulfide produced as
39 part of this cycle (Booker et al., 2017; Grein et al., 2013; Ravot et al., 2006) is further oxidized
40 mostly in environments with low oxygen levels and/or limited free transition metals. In some
41 aquatic environments, biological rates of sulfide oxidation can be orders of magnitude higher
42 than rates of abiotic sulfide oxidation. Microbial oxidation can thereby be of vital importance,
43 as sulfide is potentially toxic to aerobic organisms and consumption in anoxic waters can

44 effectively prevent its efflux to the aerobic waters. Bacteria involved in sulfur oxidation (SOBs)
45 are phylogenetically and metabolically diverse and spread across various environments.
46 Moreover, although sulfur-based metabolism appeared almost 3.5 billion years ago, many
47 enzymes and enzymatic pathways involved in this process remain to be discovered (Meier et
48 al., 2017; Nosalova et al., 2023; Wacey et al., 2011).

49 In most anaerobic sub-surface systems, the dominant chemolithotrophic SOBs belong to the
50 co-occurring *Sulfurovum*, *Sulfuricurvum* and *Sulfurimonas*-related species (*Campylobacteria*
51 formerly *Epsilonproteobacteria*). *Sulfurimonas* representatives thrive in a broad spectrum of
52 environments (Han & Perner, 2015) including pelagic redox zones (Grote et al., 2012; Hamilton
53 et al., 2014; Henkel et al., 2021; Salmaso, 2019), hydrothermal vents (Hu et al., 2021; Sikorski
54 et al., 2010; Takai et al., 2006; S. Wang, Jiang, et al., 2021; S. Wang, Shao, et al., 2021),
55 petroleum reservoirs (Lahme et al., 2019; Tian et al., 2017), groundwaters (Anantharaman et
56 al., 2016; Probst et al., 2018) and mud volcanoes (Ratnikova et al., 2020). They are particularly
57 abundant in oxygen minimum zones (15 to 30% of all microorganisms (Callbeck et al., 2021;
58 van Vliet et al., 2021)) and deep-sea hydrothermal vents (up to 70% of all microorganisms
59 (Campbell et al., 2006; Han & Perner, 2015; L. Wang et al., 2017)). Their widespread
60 occurrence is largely explained by their versatile metabolisms linked to the use of a wide
61 variety of electron donors (e.g. sulfide, thiosulfate, S^0 , H_2) and acceptors (e.g. NO_3^- , O_2 , MnO_2)
62 and by their adaptative abilities (e.g. aerobic to anoxic growth conditions, osmotic stress, etc)
63 (Götz et al., 2018; Molari et al., 2023; Rogge et al., 2017; Taylor et al., 2018; Zeng et al., 2021).
64 In addition to chemolithoautotrophic oxidation of sulfur compounds, they also perform
65 chemolithoautotrophic disproportionation of thiosulfate and elemental sulfur that may
66 represent a previously unrecognized primary production process (S. Wang et al., 2023).

67 *Sulfurimonas* spp. have mostly been isolated from hydrothermal vents and marine sediments.
68 However, high-scale DNA sequencing showed that they are widespread in fresh waters such
69 as aquifers (Anantharaman et al., 2016; Probst et al., 2018; Tian et al., 2017), springs (Hahn et
70 al., 2022), and lakes (Diao et al., 2018; Fang et al., 2022), especially in the anoxic zone of
71 meromictic (permanently stratified) systems (Berg et al., 2019; Block et al., 2021; Hamilton et
72 al., 2014; Salmaso, 2019). In the meromictic Lake Pavin, the co-occurring *Sulfuricurvum* and
73 *Sulfurimonas*-related species comprise up to 12% of the total bacterial community at the oxic-
74 anoxic boundary (Biderre-Petit, Boucher, et al., 2011). However, compared to most other
75 euxinic meromictic lakes (both anoxic and sulfidic), Lake Pavin contains very low free sulfide
76 concentrations (<15µM) in the upper part of its monimolimnion, thus representing a good
77 model for studying the ecology and metabolic traits of *Sulfurimonas*.

78 The small surface (44 ha) and great depth (92m) of Lake Pavin impede light penetration below
79 the oxycline, resulting in the fixation of carbon predominantly by chemolithoautotrophic
80 microorganisms in the anoxic bottom layer (*i.e.* monimolimnion). In this sulfide and sulfate-
81 poor (<20 µm) water mass, about 80% of total sulfide detected is in the electroactive FeS
82 colloidal form (Bura-Nakić et al., 2012). Berg et al. (2019) showed that the cryptic sulfur cycle
83 in these waters is tightly interacting with the biogeochemical iron cycle, suggesting that the
84 microorganisms involved in these two cycles contribute to mineral formation through
85 biomineralization processes (Miot et al., 2016). The bacteria and metabolisms involved remain
86 however poorly known.

87 The main goal of our study was to test the hypothesis that *Sulfurimonas* strains that thrive in
88 Lake Pavin are specific to freshwater systems, and act as key-players in tightly linked sulfur,
89 carbon, phosphate, and iron cycles. We therefore investigated the distribution of

90 *Sulfurimonas* spp. and their adaptation processes along the anoxic water column. Ribotypes
91 and metabolisms of this genus were studied *via* metabarcoding, genome-resolved
92 metagenomics, and single cell genomics, and compared with publicly available data from all
93 environments shown to host *Sulfurimonas* spp by using a pan-genomic approach.

94

95 **Materials and methods**

96

97 **Sample collection**

98 Water samples for Single Amplified Genomes (SAGs) were collected from Lake Pavin on
99 December 2015 at 65 and 80m depth, while samples for metagenomes were collected on
100 October 2015 at 50, 65, 70 and 80m. All samples were recovered from a platform positioned
101 above the deepest point of the lake (*i.e.* 92m deep) using a Van Dorn bottle. Samples for SAGs
102 were cryopreserved in 1X TE, 5% glycerol (final concentration) for later cell-sorting, and
103 biomass for metagenome sequencings (called MetaG2, N=4) were concentrated by filtration
104 onto 0.22µm pore size polycarbonate (47-mm diameter, TSTP, Millipore, Billerica,
105 Massachusetts, USA) and stored at -80°C for later DNA extraction.

106

107 **Single amplified genomes (SAGs)**

108 Single-cell sorting, whole genome amplification and SAG screening were mainly performed as
109 previously described ((Alneberg et al., 2018), for details see Supplemental material). Thirty-
110 one SAGs that carried 16S rRNA sequences, among which one affiliated to the genus
111 *Sulfurimonas*, were selected for genome sequencing. Their DNA content was quantified with
112 Qubit dsDNA HS assay kit (ThermoFisher Scientific, Waltham MA) and diluted to 0.2ng/µl in

113 nuclease-free water. MiSeq sequencing libraries were prepared with Nextera XT DNA Library
114 Preparation Kit and Nextera XT Index Kit (Illumina, San Diego CA), and run according to the
115 manufacturer's instructions. After Qubit quantification, the libraries were pooled in equimolar
116 concentration and sequenced on a MiSeq with the V3 2x300bp strategy. Subsequently, we
117 used the *bbduck2* and *bbnorm* programs from the BBTools package (release 12.10.2015)
118 (<http://jgi.doe.gov/data-and-tools/bbtools/>) to remove sequencing adapters and discard the
119 low quality reads (option: trimq = 28, maq = 20, maxns = 0, and minlength = 25) as well as to
120 normalize the quality-controlled reads (parameters target = 40, and mindepth = 0). The
121 remaining reads were then assembled with SPAdes v3.5.0 with `--careful` and `--sc` options
122 (Bankevich et al., 2012). Completeness and contamination were estimated by CheckM v1.0.5
123 (Parks et al., 2015). Gene prediction for the SAGs was performed using Prodigal V2.6.3 (Hyatt
124 et al., 2010).

125

126 **Metagenome sequencing, assembly, and binning**

127 For MetaG2 metagenomes (N=4), genomic DNA was extracted using a standard phenol-
128 chloroform method as previously described (Biderre-Petit, Jézéquel, et al., 2011). DNA
129 libraries were prepared and sequenced by GenoScreen (Lille, France), with an Illumina
130 HiSeq2500 sequencer to generate 2x150bp paired-end reads. After sequencing, reads were
131 preprocessed to remove Nextera adapters using the *bbduck* program as described above. Prior
132 to assembly, raw paired-end reads were filtered using Trimmomatic v.0.33 (option:
133 LEADING=28, TRAILING=28, SLIDINGWINDOW=4:15 and MINLEN=30) (Bolger et al., 2014).
134 Then, IDBA-UD v1.1.1 was used to assemble each sample (Peng et al., 2012) with the default
135 parameters and both pair-end reads (`-r` entry) plus single-end reads (`--long` entry).

136 To further assemble additional (low coverage or less abundant) contigs, we performed an
137 additional co-assembly of all Lake Pavin metagenomes samples, *i.e.* MetaG2 obtained in the
138 present study and MetaG1 obtained from a previous study (Biderre-Petit et al., 2019). The
139 MetaG1 (N=4) and MetaG2 (N=4) datasets were co-assembled using MEGAHIT v.1.1.1 with
140 the ‘meta-sensitive’ preset option (D. Li et al., 2015) followed by the filtering of the resulting
141 contigs by length (>2.5kb, *i.e.* the minimum contig length generally used for the initial binning
142 (Kang et al., 2015)).

143 To bin the contigs, the process was as follows: (i) metagenomic reads were individually
144 mapped against contigs using the Burrows-Wheeler-Algorithm (BWA) v.0.7.17 (Li and Durbin
145 2009) and stored as BAM files using Samtools, (ii) each BAM file was processed by *anvi-profile*
146 as implemented in *anvi'o* to get the coverage and abundance information for each contig, (iii)
147 *anvi-merge* was used to combine single profiles into a merged profile database, (iv) contigs
148 were clustered *via* the automatic binning algorithm CONCOCT v.1.0.0 (Alneberg et al., 2014)
149 with a maximal number of clusters set to 100 to minimize ‘fragmentation errors’, finally (v)
150 the resulting bins were manually curated through the *Anvi'o* interactive interface using *anvi-*
151 *refine* v6.1 (Eren et al., 2015) and as detailed by the *Anvi'o* tutorial
152 (<https://merenlab.org/2015/05/11/anvi-refine/>). The manually curated bins will hereafter be
153 referred to as Metagenome-Assembled Genomes (MAGs). MAG abundance across MetaG1
154 and MetaG2 datasets was assessed with the *anvi-summarize* program from *Anvi'o*.

155

156 **Quality evaluation and MAG taxonomic analysis**

157 The SAGs and MAGs genome quality was evaluated using both *Anvi'o* and the ‘lineage
158 workflow’ of CheckM v1.0.12 (Parks et al., 2015). Completeness/completion and
159 contamination/redundancy metrics were then used to compute the mean completeness

160 (mcpl) and the mean contamination (mctn) values to classify genome quality according to
161 established community standards (Bowers et al., 2017). Finally, quality scores were calculated
162 as follows: $Q\text{-score} = mcpl - 5 \times mctn$; and $Q'\text{-score} = mcpl - 5 \times mctn + mctn \times (\text{Checkm strain}$
163 $\text{heterogeneity})/100 + 0.5 \times \log[N50]$.

164 The Genome Taxonomy Database (GTDB, release r202) (Parks et al., 2022) and associated
165 taxonomic classification toolkit (GTDB-Tk, v1.7.0, default parameters) (Chaumeil et al., 2020)
166 were adopted to perform MAGs taxonomic classification on the basis of a set of 120 bacterial
167 and 122 archaeal gene markers.

168 To produce *Sulfurimonas* MAG/SAG hybrid assemblies of superior contiguity and accuracy, we
169 tested two different pipelines allowing genome assembly, reconciliation, and merging, namely
170 CISA v1.3 (Lin & Liao, 2013) and GARM v0.7.5 (Soto-Jimenez et al., 2014). Hybrid genome
171 quality was evaluated as described before for SAGs and MAGs.

172

173 ***Sulfurimonas* MAGs and SAG abundance**

174 Prediction of 16S rDNA genes was performed on *Sulfurimonas* SAG and MetaG2 contigs, using
175 HMM search (Huang et al., 2009) with the SILVA database (Yarza et al., 2014) in rna_hmm
176 tools (https://github.com/meb-team/Public_Tools). Predicted sequences were checked with
177 BLASTN+ v.2.2.31+ (Altschul et al., 1990) searches against the SILVA database (option :
178 perc_identity 97% and evaluate $1e^{-10}$) (Yarza et al., 2014). 16S rDNA sequences were clustered
179 using cd-hit-est (v4.6) (-c 0,95) (W. Li & Godzik, 2006, p.). Quality-controlled metagenomic
180 reads from all samples were individually mapped against the clustered sequences. Mapping
181 was performed using BWA (v.0.7.17) with default parameters and relative abundances were
182 calculated with BAM-Tk (<https://github.com/meb-team/BAM-Tk>).

183

184 **Comparative genomic analysis of *Sulfurimonas* MAGs, SAGs and reference genomes**

185 For performing a comparative genomic analysis between the *Sulfurimonas* MAGs and SAG
186 (identified here) and the reference genomes, we first downloaded the 17 *Sulfurimonas*
187 genomes available from NCBI Genbank (December 2022). These genomes include 15 species
188 among which one is represented by two strains (*S. xiamenensis* 1-1N and *S. xiamenensis* CVO),
189 as well as one strain (*Sulfurimonas* sp. SWIR-19) (Table S1). We onwards refer to these 17
190 datasets as *Sulfurimonas* reference genomes for convenience. *Sulfurimonas* sp. HSL-3221
191 (CP087998.1) was excluded from the study because of its high GC% (>57%) and low 16S rRNA
192 gene sequence similarity with *Sulfurimonas* representatives (GC% ranging from 33.3 to 38.8%;
193 up to 93% 16S rRNA identity with *S. lithotrophica*), suggesting that it belongs to another genus.
194 We also downloaded 55 MAGs and one additional SAG affiliated with *Sulfurimonas*
195 (completeness >70%, redundancy <4%, and GC% <40%) from NCBI genome database and
196 IMG/M (<https://img.jgi.doe.gov>) web portals (Table S2).

197 The average nucleotide identity (ANI) and amino acid identity (AAI) enabled delineation of
198 operational units at the genome level. ANIb was calculated using pyANI v0.2.12 (Pritchard et
199 al., 2015). It was estimated by aligning fragments of 1020nt (Goris et al., 2007) with BLASTN+
200 (Altschul et al., 1990; Camacho et al., 2009) and averaging the sequence identity between
201 pairs of genomes. AAI was calculated using the program CompareM v0.1.2
202 (<https://github.com/dparks1134/comparem>) and the parameters aai_wf --proteins --
203 sensitive. The matrices were then plotted as a heatmap and clustered using clustermap from
204 the Seaborn statistical data visualization package, v0.11.2, using default parameters (Waskom,
205 2021).

206 Analyses of acquired resistance genes were performed by ResFinder v4.1 (Zankari et al., 2012).

207 Insertion sequence (IS) elements were manually annotated according to the ISfinder database

208 (<http://www-is.biotoul.fr/>). Icefinder v2 (M. Liu et al., 2019) was used to identify putative
209 integrative and conjugative elements (ICE) and integrative mobilizable elements (IME). Finally,
210 VFDB (B. Liu et al., 2019) was used, with the genus *Helicobacter* as reference, to highlight
211 putative virulence factors and TAFinder v2 (Xie et al., 2018) with the option “4.2 unannotated
212 complete genome” to identify the presence of toxin-antitoxin (TA) loci. Contigs were also
213 manually analyzed to complete automatic annotations performed by the different tools, using
214 both BLAST searches and predictions performed by Prodigal (Hyatt et al., 2010).

215

216 **Pan-genomics analysis**

217 The pan-genome was computed to identify the clusters of shared genes (GC) according to
218 amino acid sequence homology. Firstly, *anvi'o* v7.1 (Eren et al., 2015) was used to determine
219 the pan-genome for the 17 complete reference genomes using Diamond (*--sensitive*) for
220 sequence alignment (Buchfink et al., 2021) and the Markov cluster algorithm (MCL) for
221 clustering (Van Dongen, 2008). The other parameters were kept as default except the
222 inflation, set to 8. Gene collections were classified as follows: core genome included GCs
223 present in all genomes; accessory genome included two collections – dispensable (present in
224 at least two genomes, *hereafter referred to as* shell GCs) and unique (singletons, present in
225 just one genome, *hereafter referred to as* cloud GCs). Next, these results were used to assign
226 a category to the proteins predicted from the Lake Pavin SAG and MAGs assemblies using a
227 BLASTP v2.9.0+ sequence alignment. Proteins without BLASTP hit were compared in an all-
228 against-all BLASTP, e-value $<10^{-5}$ and query coverage $>80\%$, to extract GCs present only in a
229 unique Lake Pavin strain. GC functional annotation was performed using eggNOG-mapper
230 (version *emapper-2.1.9*) based on eggNOG orthology data (Huerta-Cepas et al., 2019).

231

232 **Phylogeny**

233 A phylogenomic tree was generated with the program GToTree v1.8.2 (Lee, 2019) using a set
234 of 74 markers spanning all bacterial genomes. The representative *Sulfurimonas* genomes were
235 downloaded from NCBI RefSeq (Tables S1 and S2). GToTree identified target genes from
236 genomes and MAGs using HMMER3, aligned each gene set using MUSCLE program v5.1
237 (Edgar, 2022) and performed automated trimming with trimAl v1.4.1 ("-automated1")
238 (Capella-Gutiérrez et al., 2009). Individual gene alignments were concatenated to construct a
239 species tree using FastTree 2.1.11 the default parameters (Price et al., 2010).

240 Protein sequences to the NapA, PiT, SulP, YeeE, ferritin, sulfide: quinone reductases, and DsrE
241 proteins of *Sulfurimonas* genomes were collected using accession numbers found in literature
242 and for those not present in the literature, by searching the NCBI non-redundant database by
243 protein BLAST search. A phylogenetic tree was generated for each protein, from aligned
244 protein sequences (MUSCLE program (Edgar, 2004)) using MEGA software package version 7.0
245 (Kumar et al., 2016). Maximum likelihood (ML) method was used based on the Whelan and
246 Goldman model (JTT model, discrete Gamma distribution (5 categories)) (Whelan & Goldman,
247 2001), with a bootstrap iteration of 1000. All positions with less than 95% site coverage were
248 eliminated.

249

250 **Results**

251

252 **Reconstruction of *Sulfurimonas* genomes from Lake Pavin and species delineation**

253 The metagenome assembly from all depths represented ≈ 1000 bins (Table S3) among which
254 two binned MAGs were affiliated to the genus *Sulfurimonas*, one (Bin58) being >94% complete

255 and high-quality and another (Bin76), >57% complete and medium quality (Table S4). One SAG
256 (733-F2, 27.5% completeness, 0.2% single-marker-gene redundancy) recovered from the
257 water column was also affiliated to this genus (Table S5). Across the genome regions of Bin58
258 and SAG 733-F2 showing homology, the within-cluster sequence identity was close to or 100%
259 (sequence inversions were observed at some contig extremities), suggesting that both
260 represent the same genomic population (hereafter referred to as LP strain 1). However,
261 merging SAG 733-F2 and Bin58 genomic data led to lower estimates of completeness, higher
262 contamination, and lower Q-score (Table S6), these two genomes were therefore kept
263 separate in further analyses. For Bin76, the within-cluster sequence identity dropped below
264 85%, suggesting that this MAG corresponds to a second strain (referred to as LP strain 2).

265 Bin58 and SAG 733-F2 had ANI_b and AAI of 99.8% and 98.7% respectively, and >99% with MAG
266 *LacPavin_0419_WC90* (Jaffe et al., 2022). The corresponding numbers for Bin76 fall below
267 84%, while it shares >97% with MAG *LacPavin_0818_WC45* (Jaffe et al., 2022) (Fig. 1). These
268 results confirm that Lake Pavin hosts two distinct *Sulfurimonas* species (threshold used for
269 species demarcation >95% (Kim et al., 2014; Thompson et al., 2013)). Additionally, these
270 values did not exceed 83% when compared to publicly available data for the genus
271 *Sulfurimonas* (Fig. S1 and S2), suggesting that LP strains 1 and 2 represent novel species. This
272 conclusion was supported by analyses of the 16S rRNA sequence from SAG 733-F2 with the
273 closest database hit being *S. gotlandica* at 96.3% sequence identity. Moreover, the
274 phylogenomic analysis assigned the LP strains to a clade which includes mostly strains from
275 natural freshwater systems (Fig. 2A).

276 Unlike SAG 733-F2, no 16S rRNA gene was found in the recovered MAGs. By contrast, we
277 detected three 16S rRNA sequences in the metagenome assemblies that exhibited 100%
278 homology with the one from SAG 733-F2 (Fig. 2B). Regarding 5S rRNA genes, a copy was

279 present in Bin58. We also isolated contigs with 5S rRNA genes from metagenome assemblies
280 that were 100% identical between each other and with Bin58. These rRNA genes were flanked
281 by various genomic regions (Fig. 2B).

282 Abundance profiles based on mapping reads from the Lake Pavin metagenomes revealed a
283 similar vertical pattern for both SAG 733-F2 and Bin58, with a maximum relative abundance
284 between ~65m and 70m (Fig. 3A and B). A similar pattern was observed for a 16S rRNA OTU
285 with 100% sequence similarity to the SAG 733-F2 (*i.e.* *Sulfu.* 1, Fig. 3C) based on amplicons
286 from both DNA and RNA samples previously obtained for the Lake Pavin water column
287 (Biderre-Petit et al., 2016). This OTU also showed a significant abundance in the upper part of
288 the anoxic zone (~60-63m), suggesting that LP strain 1 (*i.e.* SAG 733-F2/Bin58) expand across
289 a wider zone along the depth profile. In comparison, Bin76 had a marked peak in relative
290 abundance at 65m (Fig. 3A). Its distribution pattern was similar to the one of the second most
291 abundant *Sulfurimonas* OTU (*Sulfu.* 2, Fig. 3C), suggesting that they represent the same
292 phylotype.

293 Among the 17 *Sulfurimonas* reference genomes, *S. sediminis* S2-6 and *S. sp.* strain SWIR-19
294 have ANI_b and AAI \geq 96% (Fig. 1), and 16S rRNA gene sequences that are 99.3-99.6% identical
295 (considering only the most similar copies of SWIR-19, Table S7). This suggests that they belong
296 to the same species. The question of heterotypic synonymy for *S. sediminis* S2-6 and *S.*
297 *hydrogeniphila* might also arise. Indeed, they share similar genome wide identity (94% AAI,
298 92% ANI) and 16S rRNA genes that are 99.3% identical (Table S7) along with many shared
299 biochemical and physiological features (S. Wang, Jiang, et al., 2021; S. Wang, Shao, et al.,
300 2021).

301

302 **Metabolic potential of the LP strains**

303 Lake Pavin SAG and MAGs were assessed for the presence of specific genome-encoded
304 functional traits involved in numerous processes, including mediation of carbon, nitrogen,
305 sulfur, and dihydrogen (H₂) cycles, oxidative stress, motility, vitamin biosynthesis and
306 transport (Table S8, Fig. 4).

307 ***Sulfur metabolism.***

308 Bin58 had the potential to use bisulfide (HS⁻) as electron donor *via* either sulfide:quinone
309 reductase (Sqr) or flavocytochrome c sulfide dehydrogenase (Fcc) pathways. Like in most
310 *Sulfurimonas* representatives, Type II (SqrB), IV (SqrD) and VI (SqrF) (Fig. S3, Table S9) were
311 identified, with two copies of Type II as in *S. marina*, *S. hongkongensis* and *S. gotlandica*.
312 Moreover, the presence of a FccB homolog suggests that this pathway could also be used to
313 catalyze initial sulfide oxidation (Fig. 5A). The less complete Bin76 genome had only one Type
314 II and one Type IV Sqr. Only the *soxCDY₂Z₂H* gene cluster for oxidation of reduced sulfur
315 compounds was found in both LP strains. In *S. xiamenensis* CVO and *S. denitrificans*, that
316 display this phenotype, S⁰ oxidation correlates to high expression of an OprD-like porin
317 (CVO_00715 and Suden_1917, respectively, both sharing 57% amino acid sequence identity),
318 for which a homolog (55% amino acid similarity) was found in Bin58. The elemental sulfur
319 reduction in LP strains could also be mediated by the polysulfide reductase-like
320 molybdoenzymes (psrBCA) (Fig. 5A, Table S8).

321 Among the three major protein families identified for the transport of oxidized sulfur
322 compounds, two (SulP and PiT) clustered with clades involved in the bicarbonate:Na⁺ symport
323 and inorganic phosphate transport, respectively (Fig. S4). The only similar sequence to our PiT
324 protein was identified in a MAG (90% protein identity, accession number OHE15791.1). By

325 **contrast**, the third family of transporters, encoding the YeeE/YedE proteins, clustered with
326 thiosulfate transporters (Tanaka et al., 2020) and is predicted to be involved in thiosulfate ion
327 transfer into the cytoplasm. In Bin76, the *yeeE* genes were tandemly organized in an operon
328 with other genes involved in sulfur metabolism, *i.e.* rhodanese (*rhd*)-like and thioredoxin
329 genes (Fig. S5). Only *Ca. S. marisnigri* contained *yeeE* homologs with a similar tandem
330 organization (HUE87_05335 and HUE87_05340, > 72% amino acid sequence identity). Finally,
331 TauE-like exporters might transport sulfite from cytoplasm to the periplasm (Gwak et al.,
332 2022).

333 Both LP strains possessed a single homolog of the DsrE-like protein (Fig. S6A), involved in
334 cytoplasmic sulfur trafficking during the sulfur oxidation. By comparison, zero to five DsrE-like
335 proteins were detected in the full range of *Sulfurimonas* genomes and MAGs, with the lowest
336 number typically found in freshwater strains (Table S10). Among all proteins identified in
337 *Sulfurimonas* genomes, only one grouped in a cluster previously determined by L.-J. Liu et al.
338 (2014), *i.e.* DsrE2A in a MAG (SAMN18877703), while the others formed eight additional
339 clusters, herein referred to as DsrEa to DsrEh (Fig. S6B). Their distribution may not be random,
340 as some of them are absent in for instance the freshwater strains (Table S10). The LP strain
341 DsrE proteins belonged to the cluster DsrEa and contained three strictly conserved cysteine
342 residues (Fig. S6A).

343 Regarding other cytoplasmic processes, a sulfur metabolizing pathway for assimilatory sulfate
344 reduction was also present in the LP strains with homologs to Cys proteins (CysD, E, K, I) and
345 a sulfate adenylytransferase (Sat, Fig. 5A) (Han & Perner, 2015).

346 ***Nitrogen metabolism.***

347 All components required for the denitrification pathway were detected in LP strains (Fig. 5B,
348 Table S11), with a periplasmic nitrate reductase catalytic subunit (NapA) organized in a well
349 conserved operon. Only *S. pluma* is currently known to lack this operon (Molari et al., 2023).
350 Moreover, as previously shown, NapA from hydrothermal vent strains represent a
351 monophyletic group distinct from genomes obtained from other ecosystems (Fig. S7). The LP
352 strains also contain genes encoding ammonium (*atmB1* and *atmB2* genes) and nitrate (*nark*
353 homolog) uptake. These compounds can subsequently be funneled into amino acid
354 metabolism *via* the ferredoxin-nitrite reductase (*nirA*) and the GS-GOGAT/GDH pathways (Fig.
355 5B). Unlike Bin58 and most other *Sulfurimonas* genomes, Bin76 contains a high-affinity nitrate
356 ABC transporter (NrtABC) for nitrate/nitrite uptake for which homologs have only been found
357 in *Ca. S. marisnigri*, *S. litotrophica*, *S. marina*, *S. pluma* and *S. gotlandica*. Bin76 genes also
358 encode key proteins for N₂ fixation (*nifHD*, *nifA*, *nifT*) with homologs only detected in *Ca. S.*
359 *marisnigri*, *Ca. S. baltica*, *S. litotrophica* and *S. marina*. However, while both *Candidatus* strains
360 can fix N₂ (Henkel et al., 2021), *S. marina* was seemingly unable to grow with N₂ as the sole
361 nitrogen source (Z. Wang et al., 2022).

362 ***H₂*, *O₂*, oxidative stress and central carbon metabolisms.**

363 Most *Sulfurimonas* can grow with H₂ as an energy source. The LP strains contained [NiFe]-
364 hydrogenases of groups 1a and 1b that mediate respiratory H₂ uptake, and group 4e,
365 representing a multisubunit membrane-bound Energy-Conserving Hydrogenase (Ech) that
366 couples ferredoxin oxidation to H₂ evolution (Table S8). Unlike group 1 [NiFe]-hydrogenases,
367 which is ubiquitous in *Sulfurimonas spp.*, group 4e is present only in marine pelagic species (*S.*
368 *gotlandica*, *Candidatus S. marisnigri* and *Ca. S. baltica*) and in both LP strains.

369 Like other *Sulfurimonas* genomes (Molari et al., 2023), LP strains carry various systems for
370 uptake and intracellular availability control of iron and manganese. Among them, a gene

371 encodes a DNA-binding protein belonging to the ferritin-like diiron-carboxylate protein
372 superfamily (known as Dps). This gene is found in almost half of the *Sulfurimonas* genomes
373 (Fig. S8). In contrast, a second member within this gene superfamily (*i.e.* a gene encoding for
374 a 24-subunit non-heme ferritin) was only found in four *Sulfurimonas* reference genomes (Fig.
375 S8). In addition to the ferritin, LP genomes have genes encoding a superoxide dismutase (SOD),
376 producing H₂O₂ and O₂ as byproducts, and a superoxide reductase (SOR), producing only H₂O₂
377 (Table S8, Fig. 4A), both designed to specifically minimize stress and cell damage from reactive
378 oxygen species (ROS) (Johnson & Hug, 2019; Lushchak, 2011).

379 In assessing the potential for aerobic respiration, both LP strains have a gene coding for the
380 cbb3-like high-O₂ affinity cytochrome, as all other *Sulfurimonas* spp. This cytochrome is active
381 in anaerobic processes such as dinitrogen (N₂) fixation and denitrification (Bell et al., 2020). In
382 addition, two hemerythrins, non-heme intracellular O₂-binding proteins, were present in LP
383 strains, one with a single-domain and one with a multi-domain that belongs to the
384 hemerythrin-GGDEF family (Table S8). We also detected an Aer-PAS domain primary energy
385 sensor for motility as well genes coding for a flagellum and chemotaxis (Fig. 4A, Table S8).

386 Finally, both LP strains contain all essential genes for carbon fixation *via* the reverse
387 tricarboxylic acid (rTCA) cycle, a common autotrophic carbon fixation pathway of all
388 *Sulfurimonas* spp. (Fig. 4A).

389

390 **Mobilome and defense mechanism analysis**

391 ***MGE in LP strains***

392 MGE is a generic term for a multitude of genomic sequences (plasmid, prophage, ICE, IME,
393 insertion sequences, etc.) transmitted vertically with cell division or through horizontal

394 transfer (HGT). Most of such genomic features were identified in the LP strain genomes *via*
395 either automatic or manual annotations (Table S12). An IME was found in both LP strains while
396 a prophage was present only in Bin58. The IME encodes a tyrosine integrase and harbors a
397 compact mobilization module with two genes encoding relaxases (one of which is a canonical
398 MobC) and one gene encoding a non-canonical coupling protein (Type IV secretory system
399 conjugative DNA transfer family protein) (Fig. 6). Only one MAG affiliated to *Arcobacter* sp.,
400 reconstructed from cold oxic seafloor (Tully et al., 2018), have been shown to host a similar
401 element with a conserved gene order (Fig. 6). With regards to the prophage, genomic
402 elements with a similar organization and content have only been reported for *Sulfurimonas*
403 MAGs (Fig. S9). Both the IME and prophage carried many **flexible** genes, mostly encoding
404 proteins of unknown function. Therefore, their role in the acquisition of novel adaptive
405 phenotypes remains unknown.

406 ***Defense systems***

407 Among all defense mechanisms identified, Toxin-Antitoxin (TA) systems were prominent,
408 representing small genetic modules with implications in a multitude of bacterial lifestyle
409 adaptation (Fraikin et al., 2020). Among the six TA types (types I-VI) currently classified, type
410 II, that function *via* protein-protein interactions, is the most well characterized and the most
411 prevalent across bacterial genomes (Kamruzzaman et al., 2021; Yeo, 2018). Here, we identified
412 a repertoire of 51 different type II TA systems in the *Sulfurimonas* genomes involving 16 toxin
413 families that are mostly involved in translational arrest (Table 1). Each of the 17 reference
414 genomes had 1 to 28 TA pairs through its chromosome. For those genomes featuring plasmids,
415 one (*S. parvalvinellae*) and five additional loci (*S. aquatica*) were identified. For LP strains, we
416 identified 12 TA pairs in Bin76, nine in Bin58 and eight in SAG 733-F2, as well as the Clp
417 proteases involved in their degradation (Table S12). Overall, the TA pattern was similar for all

418 species of a specific environment even if the TA pair number per species may varied (Table 1).
419 In contrast, both TA number and pattern are kept across strains of a same species, regardless
420 if they are from the same (*S. sediminis* and SWIR-19) or different habitats (*S. xiamenensis*
421 strains). Interestingly, two species (*S. paralvinellae* and *S. marina*) have only one TA pair.
422 However, compared to the species that live in the same environment, these species seem to
423 be restricted to more specific niches. Indeed, *S. paralvinellae*, in hydrothermal vents, lives
424 associated with a host, unlike any of the other species and *S. marina*, in marine sediments,
425 thrives at 2000m depth while the other species were retrieved from less than 250m deep. In
426 addition to type II, a putative type IV TA system (AbiE), widespread in bacteria in which it
427 ensures phage resistance (Dy et al., 2014), was present in seven genomes with an occurrence
428 per genome ranging from one to five (Table 1).

429 **Resistome**

430 The resistome analysis identified resistance to several antibiotic classes within the LP strain 1
431 genomes including beta-lactams, quinolones, macrolides and fusaric acid (FA). We found a
432 gene encoding an AmpC beta-lactamase (Table S12), with the only *Sulfurimonas* homolog
433 being a gene found in a MAG from groundwater (Finland, SAMN12221366) that exhibits the
434 same genomic context, *i.e.* associated with two amino acid ABC transporter substrate-binding
435 proteins and a HAMP domain-containing histidine kinase. Interestingly, LP strain 1 is the only
436 known *Sulfurimonas* representative harboring genes involved in FA resistance. The *fuaABC*
437 operon consists of a transcription factor of the family MarR, an inner membrane efflux pump,
438 a periplasmic efflux transporter and an outer membrane protein. Our strain also has other
439 efflux pump systems belonging to MATE, ABC (MAcAB-TolC) and RND (AcrAB-TolC) families
440 that could contribute to the generation of resistant phenotypes (Table S8).

441 **Virulome**

442 Virulome prediction identified genes involved in adherence, lipid A synthesis and biofilm
443 formation (Table S12). LP strain 1 features all traits required for biofilm formation capacity
444 (chemotaxis, flagellum-assisted motility, EPS production and export). Moreover, we identified
445 a *pgaBCD* locus that is required for a biofilm adhesin synthesis (poly- β -1,6-GlcNAc or PGA)
446 (Wang et al., 2004), but where the *pgaA* subunit encoding the outer membrane protein is
447 lacking. Homologous genes were only detected in *S. gotlandica* and *S. denitrificans*, also
448 devoid of the *pgaA* subunit.

449

450 **Pan-genome analysis of the genus *Sulfurimonas***

451 Comparative analysis of the 17 *Sulfurimonas* reference genomes revealed a pan-genome
452 composed of 9021 gene clusters (GCs) among which 915 core GCs, 3508 shell GCs and 4598
453 cloud GCs (Fig. 7A), which is consistent with previous studies (Molari et al., 2023; S. Wang,
454 Jiang, et al., 2021). With regards to the LP genomes (SAG 733-F2, Bin58 and Bin76), they
455 contained 272 to 842 core GCs, 460 to 1152 shell GCs, and 57 to 192 cloud GCs (Fig. 7B).

456 Among the shell GCs identified in LP strains, 340 to 946 belonged to the shell GCs when
457 analyzing the 17 *Sulfurimonas* reference genomes while 96 to 206 belonged to the cloud GCs
458 (Fig. 7B). This is what is generally observed with the addition of new genomes in pan-genomic
459 analyses: certain GCs firstly identified as cloud are subsequently shared by more than one
460 genome after adding additional genomes resulting in their updated classification as shell GCs.

461 In this study, 316 distinct GCs classified as cloud GCs when analyzing the 17 *Sulfurimonas*
462 reference genomes were reclassified into shell GCs when adding LP genomes (Table S13). The
463 prediction of their function *via* assignment to the Clusters of Orthologous Groups (COGs)
464 classification shows that 39% (123 GCs) could not be categorized due to the absence of
465 common features with other proteins. The 61% remaining clusters fell into 18 COGs with the

466 categories « function unknown » (S, 18%), « signal transduction mechanisms » (T, 27.3%) and
467 « cell wall, membrane, envelope biogenesis » (M, 10.8%) being the biggest three clusters
468 (Table S13).

469 Comparative analysis based on the cloud GCs identified in LP strains revealed 104 GCs found
470 only in SAG 733-F2, with 127 and 38 exclusive to Bin58 and Bin76, respectively (Fig. 7C). The
471 vast majority of these were not categorized (76%, 66.2% and 71.1%, respectively). For the
472 classified ones, nearly a quarter fall into the categories « S » and « R (only general predicted
473 function) » (Table S14). The other most common COGs were related to « mobilome (X) » and
474 « defense mechanisms (V) » (Fig. 8). For LP strain 1, proteins related to « amino-acid transport
475 and metabolism (E) », « replication, recombination, and repair (L) » and « T » were also
476 identified in lower proportions. In addition, a few cloud GCs (22 in total, Fig. 7C) were shared
477 by both strains, most of which (95.4%) had a COG classification. The categories « inorganic ion
478 transport and metabolism » (P), « cell cycle control, cell division, chromosome partitioning
479 (D) », « intracellular trafficking, secretion, and vesicular transport (U) » and M were the main
480 ones (Fig. 8, Table S9). Among these proteins, some could be relevant for the LP strain fitness,
481 as for instance, a chromate-resistance protein, involved in the reduction of chromate
482 accumulation (element with toxic and mutagenic effects), porins of the OprB family, central
483 in carbohydrate transport, and the Min system proteins (MinE, MinC, MinD) (Fig. 4). LP strains
484 differ from all *Sulfurimonas* genomes by the presence of the MinE protein (Fig. S5) that jointly
485 with MinD enables oscillatory behavior of the Min system. In its absence, the system remains
486 static (Taviti & Beuria, 2019). In addition to its involvement in cell division, this system is also
487 crucial for cellular processes such as bacterial motility, RNA decay, colonization, and virulence.

488

489 **Discussion**

490 In this comparative study of Lake Pavin microbiota, we identified two uncharacterized
491 *Sulfurimonas* species inhabiting the bottom anoxic zone using both MAGs and SAGs data. The
492 clear difference we observed in level of completeness between SAG and MAGs affiliated to
493 this genus is congruent with that was previously observed by Alneberg et al. (2018) and
494 attributed this to constraints related to the WGA during single-genome amplification. The
495 MAG quality was ensured by low level of redundancy, rejecting the possibility of contaminant
496 metagenome contigs included in the binning and assembly. Moreover, our abundance profiles
497 suggest that the two LP strains may differ in their ecological preferences related to oxygen
498 (O₂) and electron acceptor/donor availability. Such coexistence of *Sulfurimonas* species within
499 a single ecosystem with overlapping niches has never been described before.

500 The presence of various genes in the vicinity of the ribosomal genes for the LP Lake Pavin
501 strain 1 implies multiple copies per genome, as observed for most *Sulfurimonas* members
502 (Table S1). Furthermore, phylogenomic analysis as well as the diversity and distribution
503 pattern of specific genes (*e.g.* *dsrE* genes) suggest that LP strains form, with other freshwater
504 representatives, a distinct freshwater clade within this genus. Regarding all the other
505 *Sulfurimonas* strains analysed in this study, *S. sediminis* S2-6 and *S. sp.* strain SWIR-19 exhibit
506 a high degree of genome relatedness. This finding allows us to propose renaming the latter to
507 *S. sediminis* strain SWIR-19, similarly to the recent renaming of *S. sp.* strain CVO (renamed *S.*
508 *xiamenensis* strain CVO (Henkel et al., 2021)). The high genetic similarity, also observed with
509 the *S. hydrogeniphila* strain, could be explained by the fact that they all originate from the
510 Indian Ocean (Table S1) with the most closely related *S. sediminis* S2-6 and SWIR-19 isolated
511 from the same area (GPS coordinates 37.78S 49.64E and 37.28S 49.65E, respectively (S. Wang,
512 Shao, et al., 2021)), while the third, *S. hydrogeniphila*, was recovered from a more distant
513 region (6.21N 60.31E (S. Wang, Jiang, et al., 2021)). This is consistent with the identification of

514 ecotypes specific to an environment (as defined by Molari et al., (2023); e.g. brackish water
515 samples, hydrothermal plume samples, etc...) defined on the basis of oligotypes (amplicon
516 sequence variants distinguished on the basis of information entropy determined for each
517 nucleotide position (Eren, Morrison, et al., 2015)).

518 From a functional point of view, like other *Sulfurimonas*-spp. from terrestrial subsurface
519 environments, the LP strains display only the *soxCDY₂Z₂H* genes and not the *soxXY₁Z₁AB* genes
520 (Lahme et al., 2020). The products of this gene cluster are involved in S⁰ oxidation with
521 thiosulfate as metabolic end product. Cell attachment to S⁰ particles are required for their
522 oxidation and different mechanisms were suggested, depending on the *Sulfurimonas* spp. (e.g.
523 flagellar proteins, Type-IV pilin-like proteins). Recently, an outer membrane protein (OprD-like
524 porin) was also implicated in the S⁰ oxidation in strain CVO for which a homolog was found in
525 LP strains. Here, the question whether this protein is directly involved in degradation of S⁰ or
526 facilitates its transport into the cell remains open. Data available for other proteins involved
527 in the sulfur cycle are also poor. This is notably the case for the DsrE proteins for which
528 investigations are needed to identify not only their function but also the proteins they may
529 interact with. For instance, the DsrE2A protein, organized in a *rhd-tusA-dsrE2A* gene cluster,
530 was shown to mediate sulfur transfer and sulfite oxidation in cytoplasm (Dahl, 2015; L.-J. Liu
531 et al., 2014). Since the gene encoding DsrEa in LP strains is in a slightly different cluster (*tusA*
532 gene replaced by the protein-disulfide reductase (*dsbD* gene)), its true function in the cell
533 could be cryptic. We also detected in the LP strain genomes both a PiT transporter homolog,
534 closer to inorganic phosphate transporters than to sulfur compounds (Marietou et al., 2018),
535 and proteins involved in polyphosphate formation and degradation (Ppk). This suggests the LP
536 strains have the capacity to accumulate polyphosphate under favorable conditions as a

537 cellular storage of energy, analogous to the *Sulfurimonas* subgroup GD17/GD1T isolated from
538 Baltic Sea (Möller et al., 2019).

539 The presence of various systems for intracellular iron availability control in LP strain genome
540 suggests its potential role as electron donors. Also, though essential, iron can be a harmful
541 element for the cell by catalyzing the formation of ROS by reacting with H₂O₂ in a Fenton
542 reaction. Almost half of the *Sulfurimonas* genomes, including LP strains, encode a Dps-like
543 protein known to minimize damage from such oxidative stress. Indeed, this protein quenches
544 both Fe²⁺ and H₂O₂ while physically protecting genomic DNA and hence protects the cell
545 against Fenton-reaction (Bradley et al., 2020).

546 Numerous genes implicated in biofilm formation and motility are present in LP strain
547 genomes. Some of these genes were found only in a very limited number of species (*e.g. pga*,
548 hemerythrin genes). With regard to one of the two hemerythrins found, the GGDEF domain
549 located at its C terminus is believed to be involved in the synthesis of a second messenger that
550 contributes to the control of motility, virulence and biofilm formation (French et al., 2008;
551 Römling et al., 2013). This suggests a link between O₂ levels and the cyclic-di-GMP signaling
552 network in LP strains. However, how the interaction of the hemerythrin-like domain with O₂
553 alters the activity of the associated signaling domain is still not known and should be
554 investigated. Furthermore, the importance of O₂ in *Sulfurimonas* strains motility is reinforced
555 by the presence, in the few strains where this hemerythrin where detected (including LP
556 strains), of an Aer-PAS domain primary energy sensor for motility, known to respond to
557 changes in O₂ concentration. Hence, in cooperation with chemotaxis, such traits could confer
558 an ability for the cell to position themselves in a microenvironment appropriate for their
559 metabolic needs (Maschmann et al., 2022).

560 For the PGA adhesin, since it has been hypothesized that the PgaA subunit mediates its
561 translocation and/or docking to the cell surface, it would be interesting to investigate how
562 PGA is released at the surface in these strains in its absence. Moreover, both LP strains and *S.*
563 *gotlandica* all thrive in the pelagic redox zone of the water column, whether in freshwater lake
564 or the Baltic Sea. Therefore, although this has not yet been demonstrated experimentally,
565 capacity to live in biofilm at such redox transition zones might confer ecological advantages,
566 such as efficient nutrient cycling, stress tolerance and resistance to grazing. As *S. gotlandica* is
567 a cultivated species, it would be possible to elucidate some of these issues and determine the
568 importance of *pga* and hemerythrin genes in biofilm formation.

569 We suggest that MGEs might play important roles in the ability of *Sulfurimonas* spp. to
570 disperse. Among them, IMEs have never been reported in *Sulfurimonas* before and,
571 interestingly, the gene associations within IME in LP strains differs from the mandatory ones
572 described by Coluzzi et al. (2017) in *Streptococci*. For TAs, [the wide variety of associated](#)
573 [antitoxins suggests that each TA type II pair may function independently.](#) Moreover, since
574 variations in their patterns are observed both between environments and between species,
575 further work is needed to investigate how such genomic features contribute to *Sulfurimonas*
576 strain acclimation to variable conditions and particular environment. [Indeed, these](#)
577 [characteristics may, in part, explain the low number of TA pairs observed in certain strains.](#)

578 The LP strains also harbor a broad arsenal of virulence and resistance determinants that may
579 demonstrate their capacity for genetic recombination and enable rapid and efficient
580 adaptation to the local environment. Among the resistance determinants, the FA operon
581 shows an organization that is consistent with the finding that Gram-negative bacteria
582 resistance to high FA would be conferred by an efflux pump (Crutcher et al., 2017).

583 Nevertheless, the ability of our strain to grow when exposed to FA remains to be
584 experimentally confirmed. Moreover, the similarity of the MarR sequences with those of
585 *Sulfuricurvum kujiense* (Table S14) suggests an acquisition *via* HGT. Therefore, further studies
586 are needed to shed light on the evolutionary and ecological forces underpinning the early
587 evolution and differentiation of these freshwater strains, as well as their ability to disperse
588 globally.

589 The upper boundary of the Lake Pavin monimolimnion, where *Sulfurimonas* live, is dynamic
590 and uncertain as it responds to weather-induced mixing. Our results suggest that LP strain 1,
591 like *S. gotlandica*, has a strategy based on polyphosphate storage capacity that can be used as
592 an energy source to sustain cell integrity and reach favorable niches when substrate
593 availability changes. This capacity to accumulate polyphosphate has also been demonstrated
594 for magnetotactic bacteria in the monimolimnion of Lake Pavin (Rivas-Lamelo et al., 2017),
595 suggesting that this process might be more widely used than expected in this ecosystem.
596 Furthermore, the LP strain 1 was detected in a zone of the water column where conditions
597 varied from anoxic, ferruginous with elemental sulfur ($\approx 60\text{m}$, S: $3.3\mu\text{M}$) to slightly sulfidic
598 ($\approx 70\text{m}$, H_2S : $15\mu\text{M}$). Together, this suggests *Sulfurimonas* could be key-players in the
599 interconnected cycling of sulfur, carbon, phosphate, and iron cycles in Lake Pavin. Hence this
600 lake may be a promising environment to gain new knowledge about freshwater *Sulfurimonas*
601 strains and to elucidate their interactions with the environment.

602

603 **Funding**

604 The work was supported by the Agence Nationale de la Recherche (ANR) through the project
605 EUREKA (ANR-14-CE02-0004-01).

606

607 Acknowledgements

608 The authors would like to acknowledge support from the Genomics infrastructure services at
609 Science for Life Laboratory (<https://www.scilifelab.se>). Single Cell Sorting, whole genome
610 amplification, screening and library preparation were performed at the Microbial Single Cell
611 Genomics Facility (MSCG). Sequencing was performed by the SNP&SEQ Technology Platform.
612 The facilities are part of the Uppsala node at the National Genomics Infrastructure (NGI)
613 Sweden and Science for Life Laboratory. The SNP&SEQ Platform is also supported by the
614 Swedish Research Council and the Knut and Alice Wallenberg Foundation. We are also grateful
615 to the Mésocentre Clermont Auvergne University (<https://mesocentre.uca.fr/>) for providing
616 help, computing, and storage resources. We thank Maria Hammond (MSCG, Upsala), H  l  ne
617 Gardon (PhD, Universit   Clermont-Auvergne) and Vincent Dardot (M2 student, Universit  
618 Clermont-Auvergne) for efficient technical assistance and expertise.

619

620 References

- 621 Alneberg, J., Bjarnason, B. S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., Lahti, L., Loman, N. J.,
622 Andersson, A. F., & Quince, C. (2014). Binning metagenomic contigs by coverage and
623 composition. *Nature Methods*, *11*(11), 1144-1146. <https://doi.org/10.1038/nmeth.3103>
- 624 Alneberg, J., Karlsson, C. M. G., Divne, A.-M., Bergin, C., Homa, F., Lindh, M. V., Hugerth, L. W., Ettema,
625 T. J. G., Bertilsson, S., Andersson, A. F., & Pinhassi, J. (2018). Genomes from uncultivated
626 prokaryotes: A comparison of metagenome-assembled and single-amplified genomes.
627 *Microbiome*, *6*(1), 173. <https://doi.org/10.1186/s40168-018-0550-0>
- 628 Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search
629 tool. *Journal of Molecular Biology*, *215*(3), 403-410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- 630
- 631 Anantharaman, K., Brown, C. T., Hug, L. A., Sharon, I., Castelle, C. J., Probst, A. J., Thomas, B. C., Singh,
632 A., Wilkins, M. J., Karaoz, U., Brodie, E. L., Williams, K. H., Hubbard, S. S., & Banfield, J. F. (2016).
633 Thousands of microbial genomes shed light on interconnected biogeochemical processes in an
634 aquifer system. *Nature Communications*, *7*(1), 13219. <https://doi.org/10.1038/ncomms13219>
- 635 Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko,
636 S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev,
637 M. A., & Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications
638 to Single-Cell Sequencing. *Journal of Computational Biology*, *19*(5), 455-477.
639 <https://doi.org/10.1089/cmb.2012.0021>
- 640 Bell, E., Lamminm  ki, T., Alneberg, J., Andersson, A. F., Qian, C., Xiong, W., Hettich, R. L., Frutschi, M.,
641 & Bernier-Latmani, R. (2020). Active sulfur cycling in the terrestrial deep subsurface. *The ISME*
642 *Journal*, *14*(5), 1260-1272. <https://doi.org/10.1038/s41396-020-0602-x>
- 643 Berben, T., Overmars, L., Sorokin, D. Y., & Muyzer, G. (2019). Diversity and Distribution of Sulfur
644 Oxidation-Related Genes in Thioalkalivibrio, a Genus of Chemolithoautotrophic and
645 Haloalkaliphilic Sulfur-Oxidizing Bacteria. *Frontiers in Microbiology*, *10*, 160.
646 <https://doi.org/10.3389/fmicb.2019.00160>
- 647 Berg, J. S., J  z  quel, D., Duverger, A., Lamy, D., Laberty-Robert, C., & Miot, J. (2019). Microbial diversity
648 involved in iron and cryptic sulfur cycling in the ferruginous, low-sulfate waters of Lake Pavin.
649 *PLOS ONE*, *14*(2), e0212787. <https://doi.org/10.1371/journal.pone.0212787>
- 650 Biderre-Petit, C., Boucher, D., Kuever, J., Alberic, P., J  z  quel, D., Chebance, B., Borrel, G., Fonty, G., &
651 Peyret, P. (2011). Identification of sulfur-cycle prokaryotes in a low-sulfate lake (Lake Pavin)
652 using *aprA* and 16S rRNA gene markers. *Microbial Ecology*, *61*(2), 313-327.
653 <https://doi.org/10.1007/s00248-010-9769-4>

654 Biderre-Petit, C., Dugat-Bony, E., Mege, M., Parisot, N., Adrian, L., Moné, A., Denonfoux, J.,
655 Peyretailade, E., Debroas, D., Boucher, D., & Peyret, P. (2016). Distribution of Dehalococcoidia
656 in the Anaerobic Deep Water of a Remote Meromictic Crater Lake and Detection of
657 Dehalococcoidia-Derived Reductive Dehalogenase Homologous Genes. *PLOS ONE*, *11*(1),
658 e0145558. <https://doi.org/10.1371/journal.pone.0145558>

659 Biderre-Petit, C., Jézéquel, D., Dugat-Bony, E., Lopes, F., Kuever, J., Borrel, G., Viollier, E., Fonty, G., &
660 Peyret, P. (2011). Identification of microbial communities involved in the methane cycle of a
661 freshwater meromictic lake: Methane cycle in a stratified freshwater ecosystem. *FEMS*
662 *Microbiology Ecology*, *77*(3), 533-545. <https://doi.org/10.1111/j.1574-6941.2011.01134.x>

663 Biderre-Petit, C., Taib, N., Gardon, H., Hochart, C., & Debroas, D. (2019). New insights into the pelagic
664 microorganisms involved in the methane cycle in the meromictic Lake Pavin through
665 metagenomics. *FEMS Microbiology Ecology*, *95*(3). <https://doi.org/10.1093/femsec/fiy183>

666 Block, K. R., O'Brien, J. M., Edwards, W. J., & Marnocha, C. L. (2021). Vertical structure of the bacterial
667 diversity in meromictic Fayetteville Green Lake. *MicrobiologyOpen*, *10*(4).
668 <https://doi.org/10.1002/mbo3.1228>

669 Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence
670 data. *Bioinformatics*, *30*(15), 2114-2120. <https://doi.org/10.1093/bioinformatics/btu170>

671 Booker, A. E., Borton, M. A., Daly, R. A., Welch, S. A., Nicora, C. D., Hoyt, D. W., Wilson, T., Purvine, S.
672 O., Wolfe, R. A., Sharma, S., Mouser, P. J., Cole, D. R., Lipton, M. S., Wrighton, K. C., & Wilkins,
673 M. J. (2017). Sulfide Generation by Dominant *Halanaerobium* Microorganisms in Hydraulically
674 Fractured Shales. *MSphere*, *2*(4), e00257-17. <https://doi.org/10.1128/mSphereDirect.00257-17>

675

676 Bowers, R. M., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T. B. K., Schulz, F., Jarett, J.,
677 Rivers, A. R., Eloie-Fadrosch, E. A., Tringe, S. G., Ivanova, N. N., Copeland, A., Clum, A., Becraft,
678 E. D., Malmstrom, R. R., Birren, B., Podar, M., Bork, P., Weinstock, G. M., ... Woyke, T. (2017).
679 Minimum information about a single amplified genome (MISAG) and a metagenome-
680 assembled genome (MIMAG) of bacteria and archaea. *Nature Biotechnology*, *35*(8), 725-731.
681 <https://doi.org/10.1038/nbt.3893>

682 Bradley, J. M., Svistunenko, D. A., Wilson, M. T., Hemmings, A. M., Moore, G. R., & Le Brun, N. E. (2020).
683 Bacterial iron detoxification at the molecular level. *Journal of Biological Chemistry*, *295*(51),
684 17602-17623. <https://doi.org/10.1074/jbc.REV120.007746>

685 Bura-Nakić, E., Viollier, E., & Ciglenciki, I. (2012). Electrochemical and Colorimetric Measurements
686 Show the Dominant Role of FeS in a Permanently Anoxic Lake *Environmental Science &*
687 *Technology*, *47*(2), 741-749. <https://doi.org/10.1021/es303603j>

688 Callbeck, C. M., Canfield, D. E., Kuypers, M. M. M., Yilmaz, P., Lavik, G., Thamdrup, B., Schubert, C. J.,
689 & Bristow, L. A. (2021). Sulfur cycling in oceanic oxygen minimum zones. *Limnology and*
690 *Oceanography*, *66*(6), 2360-2392. <https://doi.org/10.1002/lno.11759>

691 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009).
692 BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*(1), 421.
693 <https://doi.org/10.1186/1471-2105-10-421>

694 Campbell, B. J., Engel, A. S., Porter, M. L., & Takai, K. (2006). The versatile ϵ -proteobacteria: Key players
695 in sulphidic habitats. *Nature Reviews Microbiology*, *4*(6), 458-468.
696 <https://doi.org/10.1038/nrmicro1414>

697 Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón T. (2009). trimAl: a tool for automated
698 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972-1973.
699 <https://doi.org/10.1093/bioinformatics/btp348>

700 Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2020). GTDB-Tk: A toolkit to classify
701 genomes with the Genome Taxonomy Database. *Bioinformatics*, *36*(6), 1925-1927.
702 <https://doi.org/10.1093/bioinformatics/btz848>

703 Coluzzi, C., Guédon, G., Devignes, M.-D., Ambroset, C., Loux, V., Lacroix, T., Payot, S., & Leblond-
704 Bourget, N. (2017). A Glimpse into the World of Integrative and Mobilizable Elements in

705 Streptococci Reveals an Unexpected Diversity and Novel Families of Mobilization Proteins.
706 *Frontiers in Microbiology*, 8. <https://doi.org/10.3389/fmicb.2017.00443>

707 Crutcher, F. K., Puckhaber, L. S., Stipanovic, R. D., Bell, A. A., Nichols, R. L., Lawrence, K. S., & Liu, J.
708 (2017). Microbial Resistance Mechanisms to the Antibiotic and Phytotoxin Fusaric Acid. *Journal*
709 *of Chemical Ecology*, 43(10), 996-1006. <https://doi.org/10.1007/s10886-017-0889-x>

710 Dahl, C. (2015). Cytoplasmic sulfur trafficking in sulfur-oxidizing prokaryotes. *IUBMB Life*, 67(4),
711 268-274. <https://doi.org/10.1002/iub.1371>

712 Diao, M., Huisman, J., & Muyzer, G. (2018). Spatio-temporal dynamics of sulfur bacteria during oxic—
713 Anoxic regime shifts in a seasonally stratified lake. *FEMS Microbiology Ecology*, 94(4).
714 <https://doi.org/10.1093/femsec/fiy040>

715 Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space
716 complexity. *BMC Bioinformatics*, 5(1), 113. <https://doi.org/10.1186/1471-2105-5-113>

717 Edgar, R. C. (2022). Muscle5: High-accuracy alignment ensembles enable unbiased assessments of
718 sequence homology and phylogeny. *BMC Bioinformatics*, 13(1), 6968. <https://doi.org/10.1038/s41467-022-34630-w>

719

720 Eren, A. M., Esen, Ö. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., & Delmont, T. O. (2015).
721 Anvi'o: An advanced analysis and visualization platform for 'omics data. *PeerJ*, 3, e1319.
722 <https://doi.org/10.7717/peerj.1319>

723 Fang, Y., Liu, J., Yang, J., Wu, G., Hua, Z., Dong, H., Hedlund, B. P., Baker, B. J., & Jiang, H. (2022).
724 Compositional and Metabolic Responses of Autotrophic Microbial Community to Salinity in
725 Lacustrine Environments. *MSystems*, 7(4), e00335-22. <https://doi.org/10.1128/msystems.00335-22>

726

727 Fraikin, N., Goormaghtigh, F., & Van Melderen, L. (2020). Type II Toxin-Antitoxin Systems: Evolution
728 and Revolutions. *Journal of Bacteriology*, 202(7), e00763. [https://doi.org/10.1128/JB.00763-](https://doi.org/10.1128/JB.00763-19)
729 19

730 French, C. E., Bell, J. M. L., & Ward, F. B. (2008). Diversity and distribution of hemerythrin-like proteins
731 in prokaryotes. *FEMS Microbiology Letters*, 279(2), 131-145. [https://doi.org/10.1111/j.1574-](https://doi.org/10.1111/j.1574-6968.2007.01011.x)
732 6968.2007.01011.x

733 Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., & Tiedje, J. M. (2007).
734 DNA–DNA hybridization values and their relationship to whole-genome sequence similarities.
735 *International Journal of Systematic and Evolutionary Microbiology*, 57(1), 81-91.
736 <https://doi.org/10.1099/ijs.0.64483-0>

737 Götz, F., Longnecker, K., Kido Soule, M. C., Becker, K. W., McNichol, J., Kujawinski, E. B., & Sievert, S.
738 M. (2018). Targeted metabolomics reveals proline as a major osmolyte in the
739 chemolithoautotroph *Sulfurimonas denitrificans*. *MicrobiologyOpen*, 7(4), e00586.
740 <https://doi.org/10.1002/mbo3.586>

741 Grein, F., Ramos, A. R., Venceslau, S. S., & Pereira, I. A. C. (2013). Unifying concepts in anaerobic
742 respiration: Insights from dissimilatory sulfur metabolism. *Biochimica et Biophysica Acta (BBA)*
743 *- Bioenergetics*, 1827(2), 145-160. <https://doi.org/10.1016/j.bbabi.2012.09.001>

744 Grote, J., Schott, T., Bruckner, C. G., Glöckner, F. O., Jost, G., Teeling, H., Labrenz, M., & Jürgens, K.
745 (2012). Genome and physiology of a model *Epsilonproteobacterium* responsible for sulfide
746 detoxification in marine oxygen depletion zones. *Proceedings of the National Academy of*
747 *Sciences*, 109(2), 506-510. <https://doi.org/10.1073/pnas.1111262109>

748 Gwak, J.-H., Awala, S. I., Nguyen, N.-L., Yu, W.-J., Yang, H.-Y., von Bergen, M., Jehmlich, N., Kits, K. D.,
749 Loy, A., Dunfield, Peter. F., Dahl, C., Hyun, J.-H., & Rhee, S.-K. (2022). Sulfur and methane
750 oxidation by a single microorganism. *Proceedings of the National Academy of Sciences*,
751 119(32), e2114799119. <https://doi.org/10.1073/pnas.2114799119>

752 Hahn, C. R., Farag, I. F., Murphy, C. L., Podar, M., Elshahed, M. S., & Youssef, N. H. (2022). Microbial
753 Diversity and Sulfur Cycling in an Early Earth Analogue: From Ancient Novelty to Modern
754 Commonality. *MBio*, 13(2), e00016-22. <https://doi.org/10.1128/mbio.00016-22>

755 Hamilton, T. L., Bovee, R. J., Thiel, V., Sattin, S. R., Mohr, W., Schaperdoth, I., Vogl, K., Gilhooly III, W.
756 P., Lyons, T. W., Tomsho, L. P., Schuster, S. C., Overmann, J., Bryant, D. A., Pearson, A., &

757 Macalady, J. L. (2014). Coupled reductive and oxidative sulfur cycling in the phototrophic plate
758 of a meromictic lake. *Geobiology*, 12(5), 451-468. <https://doi.org/10.1111/gbi.12092>

759 Han, Y., & Perner, M. (2015). The globally widespread genus *Sulfurimonas*: Versatile energy
760 metabolisms and adaptations to redox clines. *Frontiers in Microbiology*, 6.
761 <https://doi.org/10.3389/fmicb.2015.00989>

762 Henkel, J. V., Vogts, A., Werner, J., Neu, T. R., Spröer, C., Bunk, B., & Schulz-Vogt, H. N. (2021).
763 Candidatus *Sulfurimonas marisnigri* sp. Nov. And Candidatus *Sulfurimonas baltica* sp. Nov.,
764 thiotrophic manganese oxide reducing chemolithoautotrophs of the class *Campylobacteria*
765 isolated from the pelagic redoxclines of the Black Sea and the Baltic Sea. *Systematic and*
766 *Applied Microbiology*, 44(1), 126155. <https://doi.org/10.1016/j.syapm.2020.126155>

767 Hu, Q., Wang, S., Lai, Q., Shao, Z., & Jiang, L. (2021). *Sulfurimonas indica* sp. Nov., a hydrogen- and
768 sulfur-oxidizing chemolithoautotroph isolated from a hydrothermal sulfide chimney in the
769 Northwest Indian Ocean. *International Journal of Systematic and Evolutionary Microbiology*,
770 71(1). <https://doi.org/10.1099/ijsem.0.004575>

771 Huang, Y., Gilna, P., & Li, W. (2009). Identification of ribosomal RNA genes in metagenomic fragments.
772 *Bioinformatics*, 25(10), 1338-1340. <https://doi.org/10.1093/bioinformatics/btp161>

773 Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R.,
774 Letunic, I., Rattei, T., Jensen, L. J., von Mering, C., & Bork, P. (2019). eggNOG 5.0: A hierarchical,
775 functionally and phylogenetically annotated orthology resource based on 5090 organisms and
776 2502 viruses. *Nucleic Acids Research*, 47(D1), D309-D314.
777 <https://doi.org/10.1093/nar/gky1085>

778 Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal:
779 Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*,
780 11(1), 119. <https://doi.org/10.1186/1471-2105-11-119>

781 Jaffe, A. L., Bardot, C., Lejeune, A; L., Liu, J., Colombet, J., Perrière, F., Billard, H., Castelle, C. J., Lehours,
782 A.-C., & Banfield, J. F. (2023). *Microbiome*, 11(1), 14. <https://doi.org/10.1186/s40168-022-01416-7>

783

784 Johnson, L. A., & Hug, L. A. (2019). Distribution of reactive oxygen species defense mechanisms across
785 domain bacteria. *Free Radical Biology and Medicine*, 140, 93-102.
786 <https://doi.org/10.1016/j.freeradbiomed.2019.03.032>

787 Kamruzzaman, M., Wu, A. Y., & Iredell, J. R. (2021). Biological Functions of Type II Toxin-Antitoxin
788 Systems in Bacteria. *Microorganisms*, 9(6), 1276.
789 <https://doi.org/10.3390/microorganisms9061276>

790 Kang, D. K., Froula, J., Egan, R., & Wang, Z. (2015). MetaBAT, an efficient tool for accurately
791 reconstructing single genomes from complex microbial communities. *PeerJ*, 27(3), e1165.
792 <https://doi.org/10.7717/peerj.1165>

793 Kim, M., Oh, H.-S., Park, S.-C., & Chun, J. (2014). Towards a taxonomic coherence between average
794 nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of
795 prokaryotes. *International Journal of Systematic and Evolutionary Microbiology*, 64(Pt_2),
796 346-351. <https://doi.org/10.1099/ijms.0.059774-0>

797 Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version
798 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, 33(7), 1870-1874.
799 <https://doi.org/10.1093/molbev/msw054>

800 Lahme, S., Callbeck, C. M., Eland, L. E., Wipat, A., Enning, D., Head, I. M., & Hubert, C. R. J. (2020).
801 Comparison of sulfide-oxidizing *Sulfurimonas* strains reveals a new mode of thiosulfate
802 formation in subsurface environments. *Environmental Microbiology*, 22(5), 1784-1800.
803 <https://doi.org/10.1111/1462-2920.14894>

804 Lahme, S., Enning, D., Callbeck, C. M., Menendez Vega, D., Curtis, T. P., Head, I. M., & Hubert, C. R. J.
805 (2019). Metabolites of an Oil Field Sulfide-Oxidizing, Nitrate-Reducing *Sulfurimonas* sp. Cause
806 Severe Corrosion. *Applied and Environmental Microbiology*, 85(3), e01891-18.
807 <https://doi.org/10.1128/AEM.01891-18>

808 Lee, M. D. (2019). GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics*, 35(20), 4162-
809 4164. <https://doi.org/10.1093/bioinformatics/btz188>

810 Li, D., Liu, C.-M., Luo, R., Sadakane, K., & Lam, T.-W. (2015). MEGAHIT: An ultra-fast single-node
811 solution for large and complex metagenomics assembly via succinct *de Bruijn* graph.
812 *Bioinformatics*, 31(10), 1674-1676. <https://doi.org/10.1093/bioinformatics/btv033>

813 Li, W., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or
814 nucleotide sequences. *Bioinformatics*, 22(13), 1658-1659.
815 <https://doi.org/10.1093/bioinformatics/btl158>

816 Lin, S.-H., & Liao, Y.-C. (2013). CISA: Contig Integrator for Sequence Assembly of Bacterial Genomes.
817 *PLoS ONE*, 8(3), e60843. <https://doi.org/10.1371/journal.pone.0060843>

818 Liu, B., Zheng, D., Jin, Q., Chen, L., & Yang, J. (2019). VFDB 2019: A comparative pathogenomic platform
819 with an interactive web interface. *Nucleic Acids Research*, 47(D1), D687-D692.
820 <https://doi.org/10.1093/nar/gky1080>

821 Liu, L.-J., Stockdreher, Y., Koch, T., Sun, S.-T., Fan, Z., Josten, M., Sahl, H.-G., Wang, Q., Luo, Y.-M., Liu,
822 S.-J., Dahl, C., & Jiang, C.-Y. (2014). Thiosulfate Transfer Mediated by DsrE/TusA Homologs
823 from Acidothermophilic Sulfur-oxidizing Archaeon *Metallosphaera cuprina*. *Journal of*
824 *Biological Chemistry*, 289(39), 26949-26959. <https://doi.org/10.1074/jbc.M114.591669>

825 Liu, M., Li, X., Xie, Y., Bi, D., Sun, J., Li, J., Tai, C., Deng, Z., & Ou, H.-Y. (2019). ICEberg 2.0: An updated
826 database of bacterial integrative and conjugative elements. *Nucleic Acids Research*, 47(D1),
827 D660-D665. <https://doi.org/10.1093/nar/gky1123>

828 Louca, S., Doebeli, M., & Parfrey, L. W. (2018). Correcting for 16S rRNA gene copy numbers in
829 microbiome surveys remains an unsolved problem. *Microbiome*, 6(1), 41.
830 <https://doi.org/10.1186/s40168-018-0420-9>

831 Lushchak, V. I. (2011). Adaptive response to oxidative stress: Bacteria, fungi, plants and animals.
832 *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, 153(2),
833 175-190. <https://doi.org/10.1016/j.cbpc.2010.10.004>

834 Marietou, A., Røy, H., Jørgensen, B. B., & Kjeldsen, K. U. (2018). Sulfate Transporters in Dissimilatory
835 Sulfate Reducing Microorganisms: A Comparative Genomics Analysis. *Frontiers in*
836 *Microbiology*, 9, 309. <https://doi.org/10.3389/fmicb.2018.00309>

837 Maschmann, Z. A., Chua, T. K., Chandrasekaran, S., Ibáñez, H., & Crane, B. R. (2022). Redox properties
838 and PAS domain structure of the *Escherichia coli* energy sensor Aer indicate a multistate
839 sensing mechanism. *Journal of Biological Chemistry*, 298(12), 102598.
840 <https://doi.org/10.1016/j.jbc.2022.102598>

841 Meier, D. V., Pjevac, P., Bach, W., Hourdez, S., Girguis, P. R., Vidoudez, C., Amann, R., & Meyerdierks,
842 A. (2017). Niche partitioning of diverse sulfur-oxidizing bacteria at hydrothermal vents. *The*
843 *ISME Journal*, 11(7), 1545-1558. <https://doi.org/10.1038/ismej.2017.37>

844 Miot, J., Jézéquel, D., Benzerara, K., Cordier, L., Rivas-Lamello, S., Skouri-Panet, F., Férard, C., Poinot,
845 M., & Duprat, E. (2016). Mineralogical Diversity in Lake Pavin : Connections with Water Column
846 Chemistry and Biomineralization Processes. *Minerals*, 6(2), 24.
847 <https://doi.org/10.3390/min6020024>

848 Molari, M., Hassenrueck, C., Laso-Pérez, R., Wegener, G., Offre, P., Scilipoti, S., & Boetius, A. (2023). A
849 hydrogenotrophic *Sulfurimonas* is globally abundant in deep-sea oxygen-saturated
850 hydrothermal plumes. *Nature Microbiology*, 8(4), 651-665. <https://doi.org/10.1038/s41564-023-01342-w>

851

852 Möller, L., Laas, P., Rogge, A., Goetz, F., Bahlo, R., Leipe, T., & Labrenz, M. (2019). *Sulfurimonas*
853 subgroup GD17 cells accumulate polyphosphate under fluctuating redox conditions in the
854 Baltic Sea: Possible implications for their ecology. *The ISME Journal*, 13(2), 482-493.
855 <https://doi.org/10.1038/s41396-018-0267-x>

856 Nosalova, L., Píknova, M., Kolesarova, M., & Pristas, P. (2023). Cold Sulfur Springs—Neglected Niche
857 for Autotrophic Sulfur-Oxidizing Bacteria. *Microorganisms*, 11(6), 1436.
858 <https://doi.org/10.3390/microorganisms11061436>

859 Parks, D. H., Chuvochina, M., Rinke, C., Mussig, A. J., Chaumeil, P.-A., & Hugenholtz, P. (2022). GTDB:
860 An ongoing census of bacterial and archaeal diversity through a phylogenetically consistent,
861 rank normalized and complete genome-based taxonomy. *Nucleic Acids Research*, *50*(D1),
862 D785-D794. <https://doi.org/10.1093/nar/gkab776>

863 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM : Assessing
864 the quality of microbial genomes recovered from isolates, single cells, and metagenomes.
865 *Genome Research*, *25*(7), 1043-1055. <https://doi.org/10.1101/gr.186072.114>

866 Peng, Y., Leung, H. C. M., Yiu, S. M., & Chin, F. Y. L. (2012). IDBA-UD: A de novo assembler for single-
867 cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, *28*(11),
868 1420-1428. <https://doi.org/10.1093/bioinformatics/bts174>

869 Perisin, M., Vetter, M., Gilbert, J. A., & Bergelson, J. (2016). 16Stimulator: Statistical estimation of
870 ribosomal gene copy numbers from draft genome assemblies. *The ISME Journal*, *10*(4),
871 1020-1024. <https://doi.org/10.1038/ismej.2015.161>

872 [Price, M.N., Dehal, P.S., & Arkin, A.P. \(2010\). FastTree 2 – Approximately Maximum-Likelihood Trees
873 for Large Alignments. *PLOS ONE* *5*\(3\), e9490. <https://doi.org/10.1371/journal.pone.0009490>](#)

874 Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G., & Toth, I. K. (2015). Genomics and taxonomy
875 in diagnostics for food security: Soft-rotting enterobacterial plant pathogens. *Analytical
876 Methods*, *8*(1), 12-24. <https://doi.org/10.1039/C5AY02550H>

877 Probst, A. J., Ladd, B., Jarett, J. K., Geller-McGrath, D. E., Sieber, C. M. K., Emerson, J. B., Anantharaman,
878 K., Thomas, B. C., Malmstrom, R. R., Stieglmeier, M., Klingl, A., Woyke, T., Ryan, M. C., &
879 Banfield, J. F. (2018). Differential depth distribution of microbial function and putative
880 symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. *Nature
881 Microbiology*, *3*(3), 328-336. <https://doi.org/10.1038/s41564-017-0098-y>

882 Ratnikova, N. M., Slobodkin, A. I., Merkel, A. Y., Kopitsyn, D. S., Kevbrin, V. V., Bonch-Osmolovskaya, E.
883 A., & Slobodkina, G. B. (2020). *Sulfurimonas crateris* sp. Nov., a facultative anaerobic sulfur-
884 oxidizing chemolithoautotrophic bacterium isolated from a terrestrial mud volcano.
885 *International Journal of Systematic and Evolutionary Microbiology*, *70*(1), 487-492.
886 <https://doi.org/10.1099/ijsem.0.003779>

887 Ravot, G., Magot, M., Ollivier, B., Patel, B. K. C., Ageron, E., Grimont, P. A. D., Thomas, P., & Garcia, J.-
888 L. (2006). *Haloanaerobium congolense* sp. Nov., an anaerobic, moderately halophilic,
889 thiosulfate- and sulfur-reducing bacterium from an African oil field. *FEMS Microbiology Letters*,
890 *147*(1), 81-88. <https://doi.org/10.1111/j.1574-6968.1997.tb10224.x>

891 Rivas-Lamelo, S., Benzerara, K., Lefèvre, C. T., Monteil, C. L., Jézéquel, D., Menguy, N., Viollier, E.,
892 Guyot, F., Féraud, C., Poinot, M., Skouri-Panet, F., Trcera, N., Miot, J., & Duprat, E. (2017).
893 Magnetotactic bacteria as a new model for P sequestration in the ferruginous Lake Pavin.
894 *Geochemical Perspectives Letters*, *5*, 35-41. <https://doi.org/10.7185/geochemlet.1743>

895 Rogge, A., Vogts, A., Voss, M., Jürgens, K., Jost, G., & Labrenz, M. (2017). Success of
896 chemolithoautotrophic SUP05 and *Sulfurimonas* GD17 cells in pelagic Baltic Sea redox zones is
897 facilitated by their lifestyles as *K*- and *r*-strategists: SUP05 and *Sulfurimonas* in sulfidic redox
898 zones. *Environmental Microbiology*, *19*(6), 2495-2506. <https://doi.org/10.1111/1462-2920.13783>

900 Römling, U., Galperin, M. Y., & Gomelsky, M. (2013). Cyclic di-GMP: The First 25 Years of a Universal
901 Bacterial Second Messenger. *Microbiology and Molecular Biology Reviews*, *77*(1), 1-52.
902 <https://doi.org/10.1128/MMBR.00043-12>

903 Salmaso, N. (2019). Effects of Habitat Partitioning on the Distribution of Bacterioplankton in Deep
904 Lakes. *Frontiers in Microbiology*, *10*, 2257. <https://doi.org/10.3389/fmicb.2019.02257>

905 Sikorski, J., Munk, C., Lapidus, A., Djao, O. D. N., Tapia, R., Goodwin, L., Pitluck, S., Liolios, K., Ivanova,
906 N., Mavromatis, K., Mikhailova, N., Pati, A., Sims, D., Meincke, L., Brettin, T., Detter, J. C., Chen,
907 A., Palaniappan, K., Land, M., ... Klenk, H.-P. (2010). *Sulfurimonas autotrophica* type strain
908 (OK10). *Standards in Genomic Sciences*, *3*(2), 194-202. <https://doi.org/10.4056/sigs.1173118>

- 909 Soto-Jimenez, L., Estrada, K., & Sanchez-Flores, A. (2014). GARM: Genome Assembly, Reconciliation
910 and Merging Pipeline. *Current Topics in Medicinal Chemistry*, 14(3), 418-424.
911 <https://doi.org/10.2174/1568026613666131204110628>
- 912 Takai, K., Suzuki, M., Nakagawa, S., Miyazaki, M., Suzuki, Y., Inagaki, F., & Horikoshi, K. (2006).
913 *Sulfurimonas paralvinellae* sp. Nov., a novel mesophilic, hydrogen- and sulfur-oxidizing
914 chemolithoautotroph within the *Epsilonproteobacteria* isolated from a deep-sea hydrothermal
915 vent polychaete nest, reclassification of *Thiomicrospira denitrificans* as *Sulfurimonas*
916 *denitrificans* comb. Nov. And emended description of the genus *Sulfurimonas*. *International*
917 *Journal of Systematic and Evolutionary Microbiology*, 56(8), 1725-1733.
918 <https://doi.org/10.1099/ij.s.0.64255-0>
- 919 Tanaka, Y., Yoshikaie, K., Takeuchi, A., Ichikawa, M., Mori, T., Uchino, S., Sugano, Y., Hakoshima, T.,
920 Takagi, H., Nonaka, G., & Tsukazaki, T. (2020). Crystal structure of a YeeE/YedE family protein
921 engaged in thiosulfate uptake. *Science Advances*, 6(35), eaba7637.
922 <https://doi.org/10.1126/sciadv.aba7637>
- 923 Taviti, A. C., & Beuria, T. K. (2019). Bacterial Min proteins beyond the cell division. *Critical Reviews in*
924 *Microbiology*, 45(1), 22-32. <https://doi.org/10.1080/1040841X.2018.1538932>
- 925 Taylor, G. T., Suter, E. A., Pachiadaki, M. G., Astor, Y., Edgcomb, V. P., & Scranton, M. I. (2018). Temporal
926 shifts in dominant sulfur-oxidizing chemoautotrophic populations across the Cariaco Basin's
927 redoxcline. *Deep Sea Research Part II: Topical Studies in Oceanography*, 156, 80-96.
928 <https://doi.org/10.1016/j.dsr2.2017.11.016>
- 929 Thompson, C. C., Chimetto, L., Edwards, R. A., Swings, J., Stackebrandt, E., & Thompson, F. L. (2013).
930 Microbial genomic taxonomy. *BMC Genomics*, 23(14), 913. [https://doi.org/10.1186/1471-](https://doi.org/10.1186/1471-2164-14-913)
931 [2164-14-913](https://doi.org/10.1186/1471-2164-14-913)
- 932 Tian, H., Gao, P., Chen, Z., Li, Y., Li, Y., Wang, Y., Zhou, J., Li, G., & Ma, T. (2017). Compositions and
933 Abundances of Sulfate-Reducing and Sulfur-Oxidizing Microorganisms in Water-Flooded
934 Petroleum Reservoirs with Different Temperatures in China. *Frontiers in Microbiology*, 8, 143.
935 <https://doi.org/10.3389/fmicb.2017.00143>
- 936 Tully, B. J., Wheat, C. G., Glazer, B. T., & Huber, J. A. (2018). A dynamic microbial community with high
937 functional redundancy inhabits the cold, oxic seafloor aquifer. *The ISME Journal*, 12(1),
938 1-16. <https://doi.org/10.1038/ismej.2017.187>
- 939 van Vliet, D. M., von Meijenfildt, F. A. B., Dutilh, B. E., Villanueva, L., Sinninghe Damsté, J. S., Stams, A.
940 J. M., & Sánchez-Andrea, I. (2021). The bacterial sulfur cycle in expanding dysoxic and euxinic
941 marine waters. *Environmental Microbiology*, 23(6), 2834-2857. [https://doi.org/10.1111/1462-](https://doi.org/10.1111/1462-2920.15265)
942 [2920.15265](https://doi.org/10.1111/1462-2920.15265)
- 943 Wacey, D., Kilburn, M. R., Saunders, M., Cliff, J., & Brasier, M. D. (2011). Microfossils of sulphur-
944 metabolizing cells in 3.4-billion-year-old rocks of Western Australia. *Nature Geoscience*, 4(10),
945 698-702. <https://doi.org/10.1038/ngeo1238>
- 946 Wang, L., Cheung, M. K., Liu, R., Wong, C. K., Kwan, H. S., & Hwang, J.-S. (2017). Diversity of Total
947 Bacterial Communities and Chemoautotrophic Populations in Sulfur-Rich Sediments of
948 Shallow-Water Hydrothermal Vents off Kueishan Island, Taiwan. *Microbial Ecology*, 73(3),
949 571-582. <https://doi.org/10.1007/s00248-016-0898-2>
- 950 Wang, S., Jiang, L., Hu, Q., Cui, L., Zhu, B., Fu, X., Lai, Q., Shao, Z., & Yang, S. (2021). Characterization of
951 *Sulfurimonas hydrogeniphila* sp. Nov., a Novel Bacterium Predominant in Deep-Sea
952 Hydrothermal Vents and Comparative Genomic Analyses of the Genus *Sulfurimonas*. *Frontiers*
953 *in Microbiology*, 12, 626705. <https://doi.org/10.3389/fmicb.2021.626705>
- 954 Wang, S., Jiang, L., Xie, S., Alain, K., Wang, Z., Wang, J., Liu, D., & Shao, Z. (2023). Disproportionation of
955 Inorganic Sulfur Compounds by Mesophilic Chemolithoautotrophic *Campylobacterota*.
956 *MSystems*, 8(1), e00954-22. <https://doi.org/10.1128/msystems.00954-22>
- 957 Wang, S., Shao, Z., Lai, Q., Liu, X., Xie, S., Jiang, L., & Yang, S. (2021). *Sulfurimonas sediminis* sp. Nov., a
958 novel hydrogen- and sulfur-oxidizing chemolithoautotroph isolated from a hydrothermal vent
959 at the Longqi system, southwestern Indian ocean. *Antonie van Leeuwenhoek*, 114(6), 813-822.
960 <https://doi.org/10.1007/s10482-021-01560-4>

- 961 Wang, X., Preston, J. F., & Romeo, T. (2004). The *pgaABCD* Locus of *Escherichia coli* Promotes the
962 Synthesis of a Polysaccharide Adhesin Required for Biofilm Formation. *Journal of Bacteriology*,
963 186(9), 2724-2734. <https://doi.org/10.1128/JB.186.9.2724-2734.2004>
- 964 Wang, Z., Wang, S., Lai, Q., Wei, S., Jiang, L., & Shao, Z. (2022). *Sulfurimonas marina* sp. Nov., an
965 obligately chemolithoautotrophic, sulphur-oxidizing bacterium isolated from a deep-sea
966 sediment sample from the South China Sea. *International Journal of Systematic and*
967 *Evolutionary Microbiology*, 72(10), 005582. <https://doi.org/10.1099/ijsem.0.005582>
- 968 Waskom, M. (2021). seaborn: Statistical data visualization. *Journal of Open Source Software*, 6(60),
969 3021. <https://doi.org/10.21105/joss.03021>
- 970 Whelan, S., & Goldman, N. (2001). A General Empirical Model of Protein Evolution Derived from
971 Multiple Protein Families Using a Maximum-Likelihood Approach. *Molecular Biology and*
972 *Evolution*, 18(5), 691-699. <https://doi.org/10.1093/oxfordjournals.molbev.a003851>
- 973 Wiedenheft, B., Mosolf, J., Willits, D., Yeager, M., Dryden, K. A., Young, M., & Douglas, T. (2005). An
974 archaeal antioxidant: Characterization of a Dps-like protein from *Sulfolobus solfataricus*.
975 *Proceedings of the National Academy of Sciences*, 102(30), 10551-10556.
976 <https://doi.org/10.1073/pnas.0501497102>
- 977 Xie, Y., Wei, Y., Shen, Y., Li, X., Zhou, H., Tai, C., Deng, Z., & Ou, H.-Y. (2018). TADB 2.0: An updated
978 database of bacterial type II toxin–antitoxin loci. *Nucleic Acids Research*, 46(D1), D749-D753.
979 <https://doi.org/10.1093/nar/gkx1033>
- 980 Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F. O., Ludwig, W., Schleifer, K.-H., Whitman, W. B., Euzéby,
981 J., Amann, R., & Rosselló-Móra, R. (2014). Uniting the classification of cultured and uncultured
982 bacteria and archaea using 16S rRNA gene sequences. *Nature Reviews Microbiology*, 12(9),
983 635-645. <https://doi.org/10.1038/nrmicro3330>
- 984 Yeo, C. C. (2018). GNAT toxins of bacterial toxin-antitoxin systems: Acetylation of charged tRNAs to
985 inhibit translation: GNAT toxins of bacterial toxin-antitoxin systems. *Molecular Microbiology*,
986 108(4), 331-335. <https://doi.org/10.1111/mmi.13958>
- 987 Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F. M., &
988 Larsen, M. V. (2012). Identification of acquired antimicrobial resistance genes. *Journal of*
989 *Antimicrobial Chemotherapy*, 67(11), 2640-2644. <https://doi.org/10.1093/jac/dks261>
- 990 Zeng, X., Alain, K., & Shao, Z. (2021). Microorganisms from deep-sea hydrothermal vents. *Marine Life*
991 *Science & Technology*, 3(2), 204-230. <https://doi.org/10.1007/s42995-020-00086-4>
- 992

993 **Data accessibility**

994 The raw sequences of the metagenomic datasets MetaG1, previously described in Biderre-
995 Petit et al (2019), are archived in MG-RAST under accession numbers mgm4580845.3 to
996 mgm4580848.3 for the metagenomic reads and the accession numbers mgm4557752.3.100,
997 mgm4557753.3.100, mgm4581104.3.100 and mgm4581106.3.100 for the assembled contigs.
998 Metabarcoding data are available at the NCBI SRA database under the BioProject accession
999 number PRJNA975855 and SRA accession numbers SRX20521566 to SRX20521576. Raw
1000 sequences of the metagenomic datasets MetaG2 as well as the *Sulfurimonas* SAG and MAGs
1001 produced as part of the present study are available at NCBI SRA database under the BioProject
1002 accession PRJNA983208.

1003

1004 **ORCID**

1005 Corinne Biderre-Petit <https://orcid.org/0000-0001-7962-4171>

1006 Arthur Monjot <https://orcid.org/0000-0002-6978-4785>

1007 Cécile Lepère <https://orcid.org/0000-0003-4767-0477>

1008 Didier Debroas <https://orcid.org/0000-0002-9915-1268>

1009 Pierre Galand <https://orcid.org/0000-0002-2238-3247>

1010 Stefan Bertilsson <https://orcid.org/0000-0002-4265-1835>

1011

1012 **Conflict of interests**

1013 The authors declare that they have no known competing interests.

1014

1015 **Author contributions**

1016 **AM** conceived figures, involvement in writing-review and editing; **CBP** was responsible for
1017 conceptualization, supervision, data curation, data analysis and interpretation and writing-
1018 original draft; **Che** was involved in formal analysis and writing - review and editing; **CHo** was
1019 responsible for metagenome assembly and analysis, data curation, methodology and writing
1020 - review and editing; **CL** was involved in writing - original draft , **DC** participate to the pan-
1021 genome analysis, data curation, methodology, writing - review and editing; **PEG** was
1022 responsible for sampling and funding acquisition (ANR EUREKA), writing - review and editing;
1023 **DD** was responsible for sampling and funding acquisition (ANR EUREKA); **SB, AMD** and **MH**
1024 were responsible for SAG preparation and sequencing, writing - review and editing. All authors
1025 contributed to the manuscript and approved the final version.

1026

1027 **Figure titles and legends**

1028 **Figure 1: Pairwise average similarity calculations for the 17 *Sulfurimonas* reference genomes**
1029 **and the Lake Pavin strains.** The MAGs and SAG obtained in the present study are indicated in
1030 orange (Bin58, Bin76 and SAG 733-F2) and MAGs from the Jaffe et al' study (2023), in blue
1031 (MAG 0419_WC90, and MAG 0818_WC45). Upper matrix: pairwise average nucleotide
1032 identity (ANIb); lower matrix: pairwise average amino acid identity (AAI). Lake Pavin strains
1033 from the present study are indicated in orange and those from the Jaffe et al' study, in blue
1034 (Jaffe et al., 2023).

1035 **Figure 2: Phylogenomic analysis of Lake Pavin strains and ribosomal genes detected in Lake**
1036 **Pavin datasets. A)** Phylogenomic tree encompassing *Sulfurimonas* genomes (isolates and
1037 MAGs) made using GToTree (Lee, 2019) and a set of 74 bacterial marker genes. MAGs
1038 obtained in the present are indicated in orange and those from Jaffe et al. (2023), in blue.
1039 Stars indicated *Sulfurimonas* strains found associated to a host. Bootstrap values based on
1040 1000 replicates; only values above 75% are shown at branch nodes. **B)** Schematic
1041 representation of the ribosomal RNA gene genomic context found in the MetaG2
1042 metagenomes as well as in Bin58 and SAG 733-F2 assemblies. Boxed arrows indicate genes
1043 and their direction of transcription. For each gene, the closest relative is indicated in the right
1044 part of the figure, identified either by a letter or by a number.

1045 **Figure 3: Relative abundance of the Lake Pavin *Sulfurimonas* strains along the lake water**
1046 **column. A)** Relative abundance of Bin58 and Bin76 MAGs in metagenomes produced in 2013
1047 (MetaG1) and 2015 (MetaG2). **B).** Relative abundance of SAG 733-F2 in MetaG2 metagenomes
1048 based on mapping the original 16S rRNA metagenomic reads onto the 16S rRNA cluster with
1049 seed-sequence 100% identical to the SAG 733-F2 sequence. **C)** Relative abundance (% of total
1050 community) in the 454 metabarcoding libraries constructed from gDNA (MetaB 16S gDNA)
1051 and cDNA samples (MetaB 16S cDNA) collected along the Lake Pavin water column. Sulfu1:

1052 OTU with seed-sequence 100% identical to the SAG 733-F2 sequence; Sulfu2: the second most
1053 abundant OTU affiliated to the genus *Sulfurimonas* and with less than 97% identity with the
1054 SAG 733-F2 sequence.

1055 **Figure 4: Overview of the metabolism of the Lake Pavin *Sulfurimonas* strains inferred from**
1056 **their genome. A)** General functions. Transporter families (ABC, MFS, P-type ATPase,
1057 symporter, etc) are represented by different colors and forms. **B)** Functions only carried by
1058 Lake Pavin (LP) strains (strain 1, strain 2 or both – black boxes) or shared by one or both LP
1059 strains with only one other *Sulfurimonas* genome (blue box).

1060 **Figure 5: Overview of the putative pathways of oxidative and assimilation metabolisms for**
1061 **sulfur (A) and of reductive and assimilation metabolisms for nitrogen (B) within the Lake**
1062 **Pavin strains** (for details about genes involved in sulfur pathways see Table S9, and for
1063 nitrogen pathways, Table S10).

1064 **Figure 6: Modular structure of integrative and mobilizable element (IME) found in**
1065 ***Sulfurimonas* strains from Lake Pavin.** The gene blocks conserved between SAG 733-F2 and
1066 Bin58, SAG 733-F2 and Bin76 and SAG 733-F2 and *Arcobacter* sp. are shown. Proteins not
1067 found in other *Sulfurimonas* genomes appear in bold; those found only in a few genomes, in
1068 bold italics.

1069 **Figure 7: Pan-genome analysis of the 17 *Sulfurimonas* reference genomes used in the**
1070 **present study and the Lake Pavin strains. A)** Flower plot displaying the numbers of core gene
1071 clusters (GCs) (in the center), flexible (shell) GCs (in the annulus) and unique (cloud) GCs (in
1072 the petals) for the 17 *Sulfurimonas* genomes. **B)** Number of core, flexible (shell), and unique
1073 (cloud) GCs for SAG 733-F2, Bin58 and Bin76 genomes. **C)** Venn diagram displaying the number
1074 of cloud GCs shared by lake Pavin strains or specific to each. The number of genes
1075 corresponding to the GC number is indicated in parentheses in order: SAG 733-F2, Bin58,
1076 Bin76.

1077 **Figure 8: Clusters of Orthologous Groups (COGs) classification of the cloud GCs identified in**
1078 **Lake Pavin strains.**

1079 **Table 1: Classification and distribution of Type II and IV toxin-antitoxin systems in**
1080 ***Sulfurimonas* species.**

1081 **Figure S1: Pairwise average nucleotide identity (ANI_b) heatmap of *Sulfurimonas* genomes,**
1082 **MAGs and Bin58.**

1083 **Figure S2: Pairwise average amino acid identity (AAI) heatmap of *Sulfurimonas* genomes,**
1084 **MAGs and Bin58.**

1085 **Figure S3: Maximum likelihood phylogenetic tree of deduced protein sequences of the**
1086 **sulfide:quinone reductase (Spr).** Number at nodes represent bootstrap values (1 000
1087 replicates of the original dataset), only number above 50% are shown. The scale bar
1088 represents the average number of substitutions per site. The accession number for the
1089 sequences are given at the right side of species name. The protein sequences from Lake Pavin
1090 strains are indicated by a dot for Bin58 and a circle for Bin76, with colors depending on the
1091 *sqr* type: type II in red (*SqrB*), type IV in green (*SqrD*) and type VI in blue (*SqrF*).

1092 **Figure S4: Maximum likelihood phylogenetic trees of deduced protein sequences of two**
1093 **potential sulfate transporters (PiT and SulP). On the left:** PiT protein sequence tree. The two
1094 protein clades are surrounded with different colors and given a letter label (A, B). Reference
1095 sequences: PITA, inorganic phosphate transporter (UniProt P0AFJ7), PITB: inorganic
1096 phosphate transporter (UniProt P0AFJ7), Cys Pit: sulfate transporter (BSUA_01689). **On the**
1097 **right:** SulP protein sequence tree. The three protein clades are surrounded with different
1098 colors and given a letter label (A, B, C) while separate protein groupings are denoted by
1099 subscript number (i, ii). Reference sequences: BICA (Aii), bicarbonate transporter (UniProt
1100 Q14SY0), DAU_Salm (B): dicarboxylic transporter (ACY88617), Riv1739c and SulP_Pseudo
1101 (PA1647) (C): sulfate transporter. Number at nodes represent bootstrap values (1 000
1102 replications of the original dataset), only number above 50% are shown. The scale bar
1103 represents the average number of substitutions per site. The accession number are given for
1104 all sequences. Bin58 and SAG 733-F2 sequences are identified by a red dot.

1105 **Figure S5: Schematic representation of yee gene genomic context in the Bin76 genome.** All
1106 genes potentially involved in sulfur cycling are in ochre-brown color. Each box represents
1107 genes in the contig for which a homolog was present only in a restricted number of
1108 *Sulfurimonas* genomes and MAGs. Their genomic context in the Lake Pavin strain 1 is also
1109 shown (Node for SAG 733-F2 and locus tag, *i.e.* K119_, for Bin58). Ochre arrows: genes
1110 encoding YeeE proteins; green arrows: Min operon; black blue arrow: gene encoding a high
1111 potential iron-sulfur protein.

1112 **Figure S6: Sequence alignment of DsrE homologs (A) and phylogenetic tree (B).** In **A**, the
1113 sequences are identified by their locus tags. The active site cysteine conserved in all sequences
1114 is labeled with an asterisk and highlighted in cyan. For clarity, only those parts of the alignment
1115 are shown to contain conserved cysteine residues (N and C terminus, cysteines highlighted in
1116 yellow). In blue: sequence identifiers of *Sulfurimonas* genomes. Bins are highlighted in red and
1117 their motifs in bold italics. In **B**, the evolutionary history was inferred using the maximum
1118 likelihood method. Number at nodes represent bootstrap values (1 000 replicates of the
1119 original dataset), only number above 50% are shown. The scale bar represents the average
1120 number of substitutions per site. The accession number are given for all sequences. Bin
1121 sequences are identified by a red dot, *Sulfurimonas* genome sequences by a blue diamond.
1122 Subclusters DsrE2A, DsrE3A, DsrE3B, DsrE4 and DsrE5 were those previously described by Liu
1123 et al (2014). DsrE corresponds to the classical DsrE/TusD subunit from the DsrEFH operon that
1124 acts as cytoplasmic acceptor for sulfur. C* corresponds to the active site cysteine conserved
1125 in all sequences.

1126 **Figure S7: Maximum likelihood phylogenetic trees of deduced protein sequences of**
1127 **periplasmic nitrate reductase catalytic subunit (NapA).** Number at nodes represent bootstrap
1128 values (1 000 replicates of the original dataset), only number above 50% are shown. The scale
1129 bar represents the average number of substitutions per site. The accession number are given
1130 for all sequences. Each dot color corresponds to an environmental category for *Sulfurimonas*
1131 genomes: blue, freshwater (Bin58 sequence); yellow: marine sediment; orange: marine water;
1132 light yellow: brackish water, green: oil field water, gray: mud volcano and light brown:
1133 hydrothermal vent.

1134 **Figure S8: Maximum likelihood phylogenetic tree of deduced protein sequences of the**
1135 **ferritin-like diiron-carboxylate protein superfamily.** Number at nodes represent bootstrap
1136 values (1 000 replicates of the original dataset), only number above 50% are shown. The scale
1137 bar represents the average number of substitutions per site. The accession numbers for the
1138 sequences are given at the right side of species name. For the *Sulfurimonas* genomes, the
1139 ferritin belonging to the DNA-binding protein from nutrient starved cells (Dps) family is
1140 identified by a green dot while that belonging to the 24-subunit non-heme ferritin family, by
1141 a red dot.

1142 **Figure S9: Modular structure of integrated prophage found in Lake Pavin strain 1.** The gene
1143 blocks conserved in the two MAGs affiliated to the *Sulfurimonas* genus (UBA and RIFOXYD2)
1144 are shown. Proteins found in no other *Sulfurimonas* genomes appear in red while those found
1145 only in a few genomes, in blue.

1146 **Table S1:** Characteristics of the 17 *Sulfurimonas* reference genomes and strains from the water
1147 column of Lake Pavin.

1148 **Table S2:** List of *Sulfurimonas* MAGs and SAGs downloaded from the public databases and
1149 used in this study.

1150 **Table S3:** List of Bins produced by *de novo* assembly of pooled reads from all metagenomes
1151 (MetaG1 and MetaG2) collected along the Lake Pavin water column.

1152 **Table S4:** Statistics on the assembly of SAG and MAGs of *Sulfurimonas* genus obtained from
1153 Lake Pavin, completeness and contamination estimates.

1154 **Table S5:** Taxonomic affiliation of SAGs obtained at 65 and 80 m depth in Lake Pavin water
1155 column based on their 16S rRNA gene sequence (Sanger sequencing).

1156 **Table S6:** Quality evaluation of hybrid assembly genomes and of Bin58 MAG and SAG_733-F2
1157 used to produce them.

1158 **Table S7:** Multiple sequence alignment of the different 16S rRNA gene copies from
1159 *Sulfurimonas* sp. SWIR19, *S. sediminis* S2-6 and *S. hydrogeniphila* using Clustal Omega tool
1160 (<https://www.ebi.ac.uk/Tools/msa/clustalo/>).

1161 **Table S8:** Specific functions involved in numerous processes, including: carbon, nitrogen,
1162 sulfur, hydrogen (H₂) cycling, oxidative stress, motility, vitamin biosynthesis, transport, etc,
1163 identified in Lake Pavin strains.

1164 **Table S9:** List of genes involved in sulfur cycling identified in Lake Pavin strains.

1165 **Table S10:** DsrE-like protein diversity and distribution in *Sulfurimonas* strains.

1166 **Table S11:** List of genes involved in nitrogen cycling identified in Lake Pavin strains.

1167 **Table S12:** Predicted mobilome, resistome and virulome in MAGs and SAG sequences (* genes
1168 found only in Lake Pavin strains).

1169 **Table S13:** List of cloud gene clusters found both in *Sulfurimonas* reference genomes and Lake
1170 Pavin strains to be reclassified into shell GC and their classification into COG categories.

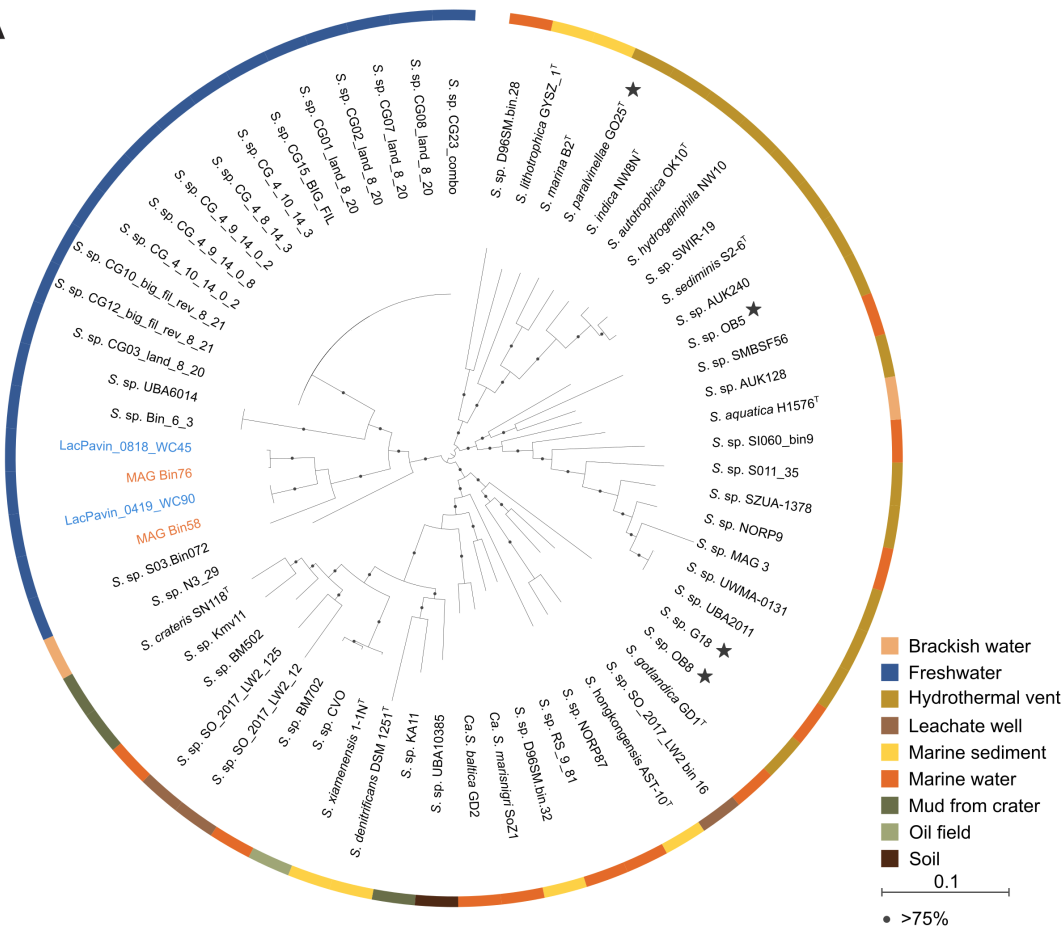
1171 **Table S14:** Cloud gene clusters identified in Lake Pavin strains for which a COG classification
1172 was found.
1173

AAI (lower matrix) vs ANIb (upper matrix)

MAG 0419_WC90	100	100	100	82	81	73	72	73	73	75	75	72	73	73	72	74	74	74	73	74	73
SAG 733-F2	99	100	100	83	82	73	73	73	73	75	75	73	73	73	72	73	74	74	74	74	74
MAG BIN58	100	99	100	82	81	73	73	73	73	75	75	72	73	73	72	72	75	75	75	73	74
MAG BIN76	84	82	84	100	97	74	74	74	74	76	76	73	74	74	74	74	76	76	76	75	75
MAG 0818_WC45	82	82	83	97	100	74	73	73	73	75	75	72	73	73	73	73	75	75	75	74	74
<i>S. autotrophica</i>	69	68	69	71	70	100	82	82	82	74	74	76	76	74	74	73	74	74	74	73	73
<i>S. hydrogeniphila</i>	69	67	69	70	69	84	100	92	92	73	73	76	75	73	74	73	73	73	73	72	72
<i>S. sp SWIR-19</i>	69	68	69	70	69	84	94	100	96	73	73	76	75	73	74	73	73	73	73	72	73
<i>S. sediminis</i>	69	67	69	71	69	85	94	96	100	73	73	76	75	73	74	73	73	73	73	72	72
<i>S. sp CVO</i>	72	70	72	74	72	70	70	69	70	100	99	73	74	74	73	74	76	76	75	75	77
<i>S. xiamenensis</i>	72	70	72	74	73	70	70	70	70	99	100	73	74	73	73	74	76	76	75	75	77
<i>S. paralvinellae</i>	69	68	69	71	69	77	77	77	77	70	70	100	79	73	74	73	73	72	73	72	72
<i>S. indica</i>	69	67	70	71	70	77	76	76	76	70	70	82	100	74	75	73	73	73	73	73	73
<i>S. aquatica</i>	69	68	69	71	70	73	72	71	72	70	70	72	72	100	74	73	74	74	75	74	73
<i>S. marina</i>	68	66	68	70	68	72	72	71	71	69	69	73	73	70	100	74	73	73	74	73	73
<i>S. litotrophica</i>	68	67	69	71	68	69	68	68	68	70	71	68	69	69	69	100	73	73	74	73	73
Candidatus <i>S. baltica</i>	71	70	72	73	72	69	69	69	69	75	76	69	69	70	69	70	100	88	76	74	75
Candidatus <i>S. marisnigri</i>	72	69	72	73	72	70	69	69	69	76	76	69	70	70	69	70	90	100	77	75	75
<i>S. gotlandica</i>	72	71	72	74	72	70	70	70	70	73	74	70	70	71	69	70	75	75	100	77	75
<i>S. hongkongensis</i>	70	69	70	72	71	69	69	69	69	72	72	69	69	69	68	70	72	72	77	100	75
<i>S. denitrificans</i>	70	68	70	72	70	68	68	68	68	78	78	68	68	68	68	68	68	73	73	71	71
<i>S. crateris</i>	71	69	72	73	71	69	69	69	69	81	82	69	69	69	69	69	75	75	72	71	77

Figure 1

A



B

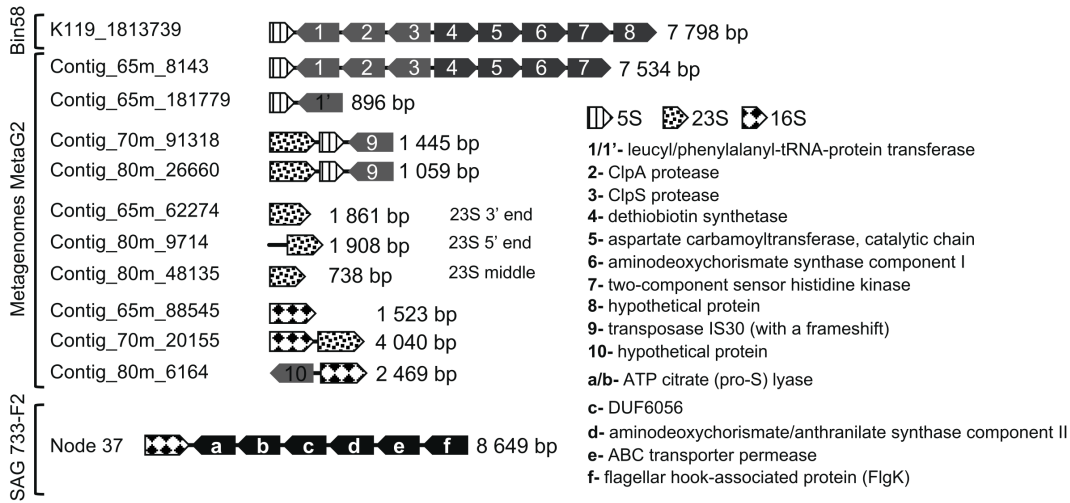


Figure 2

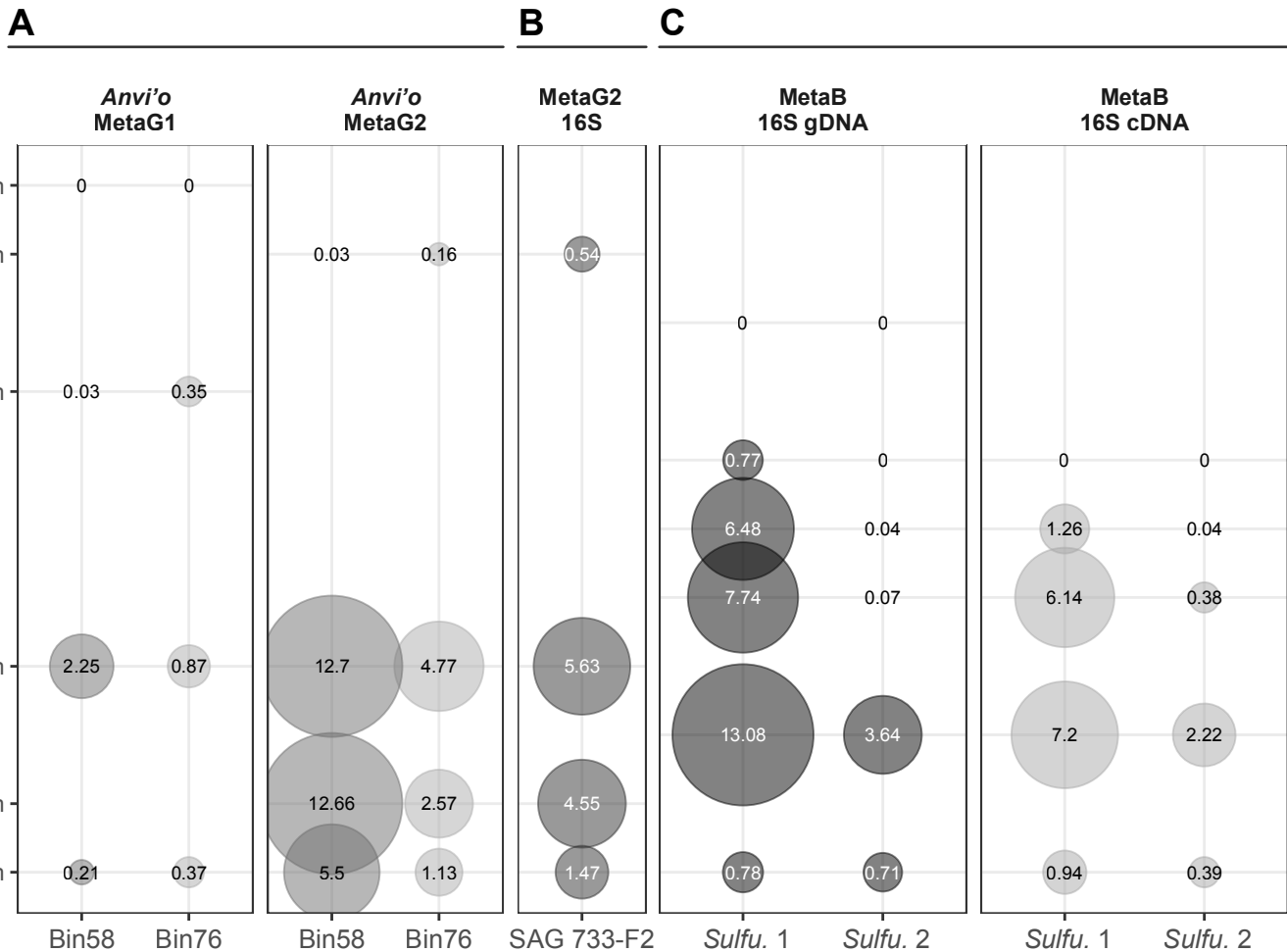


Figure 3

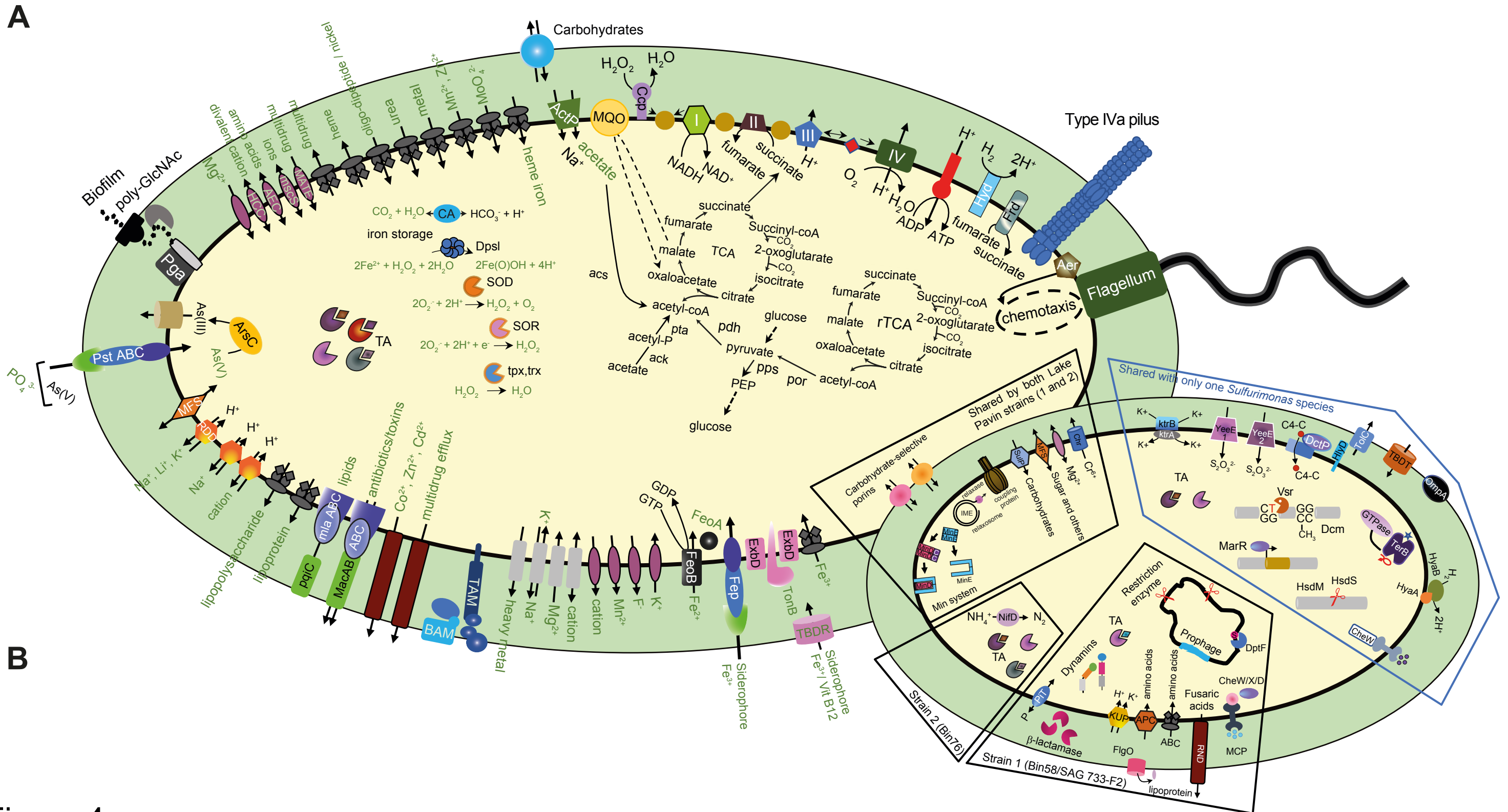


Figure 4

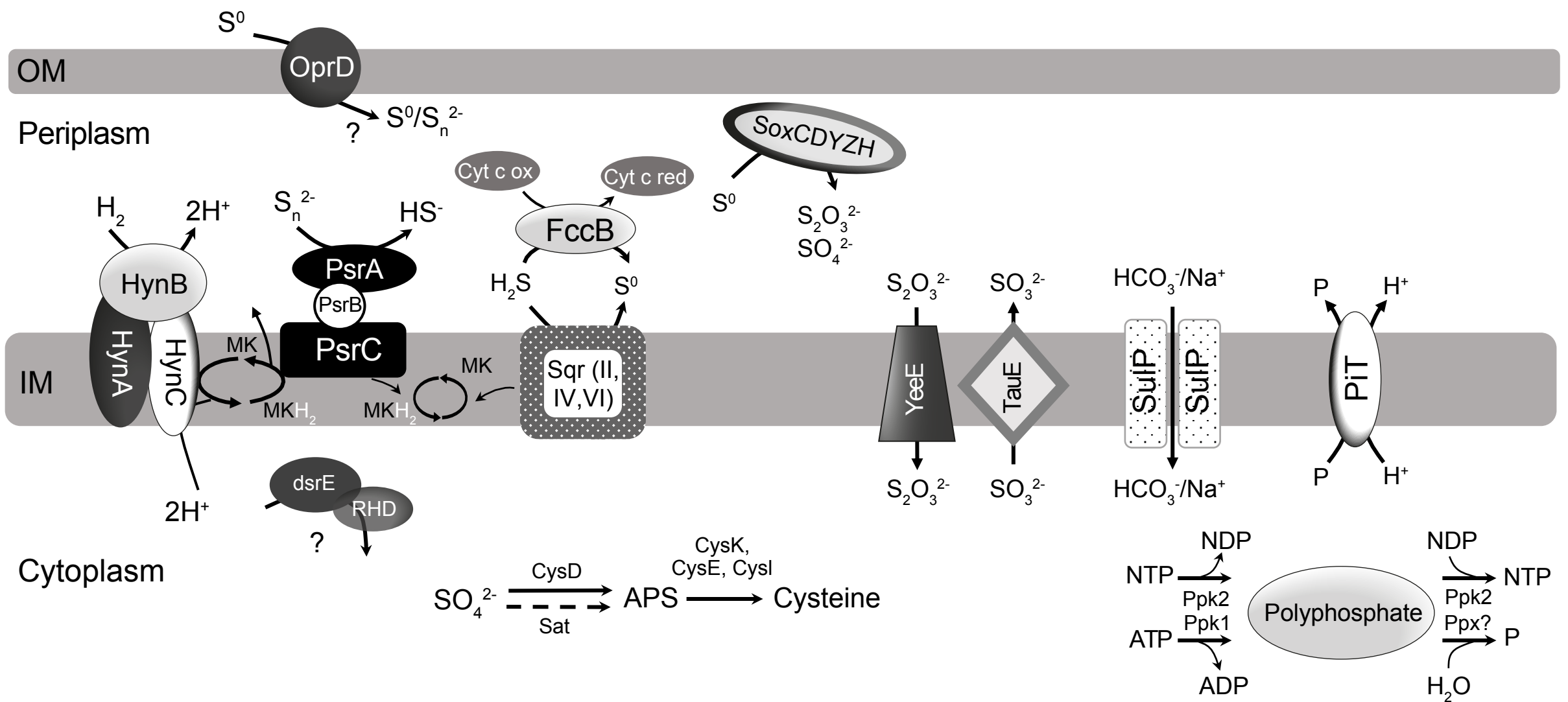
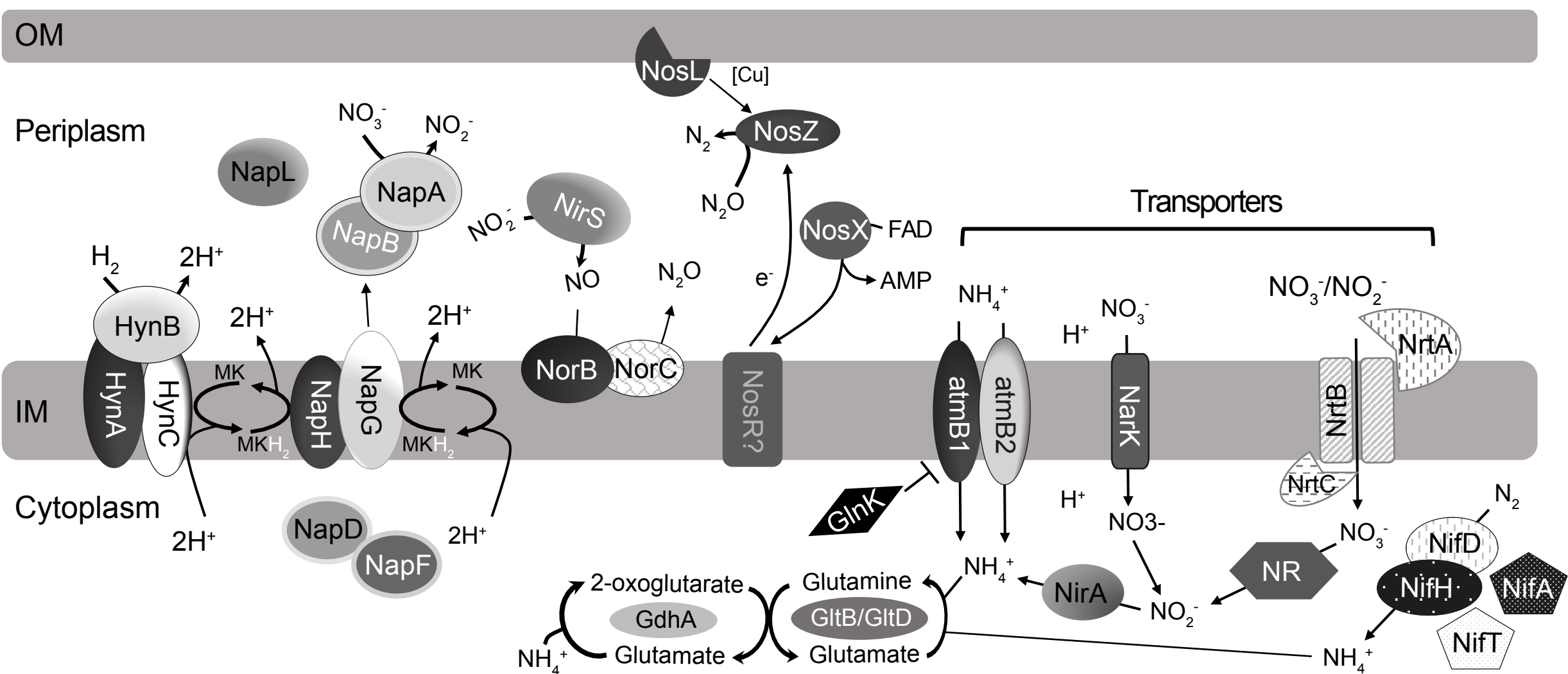
A**B**

Figure 5

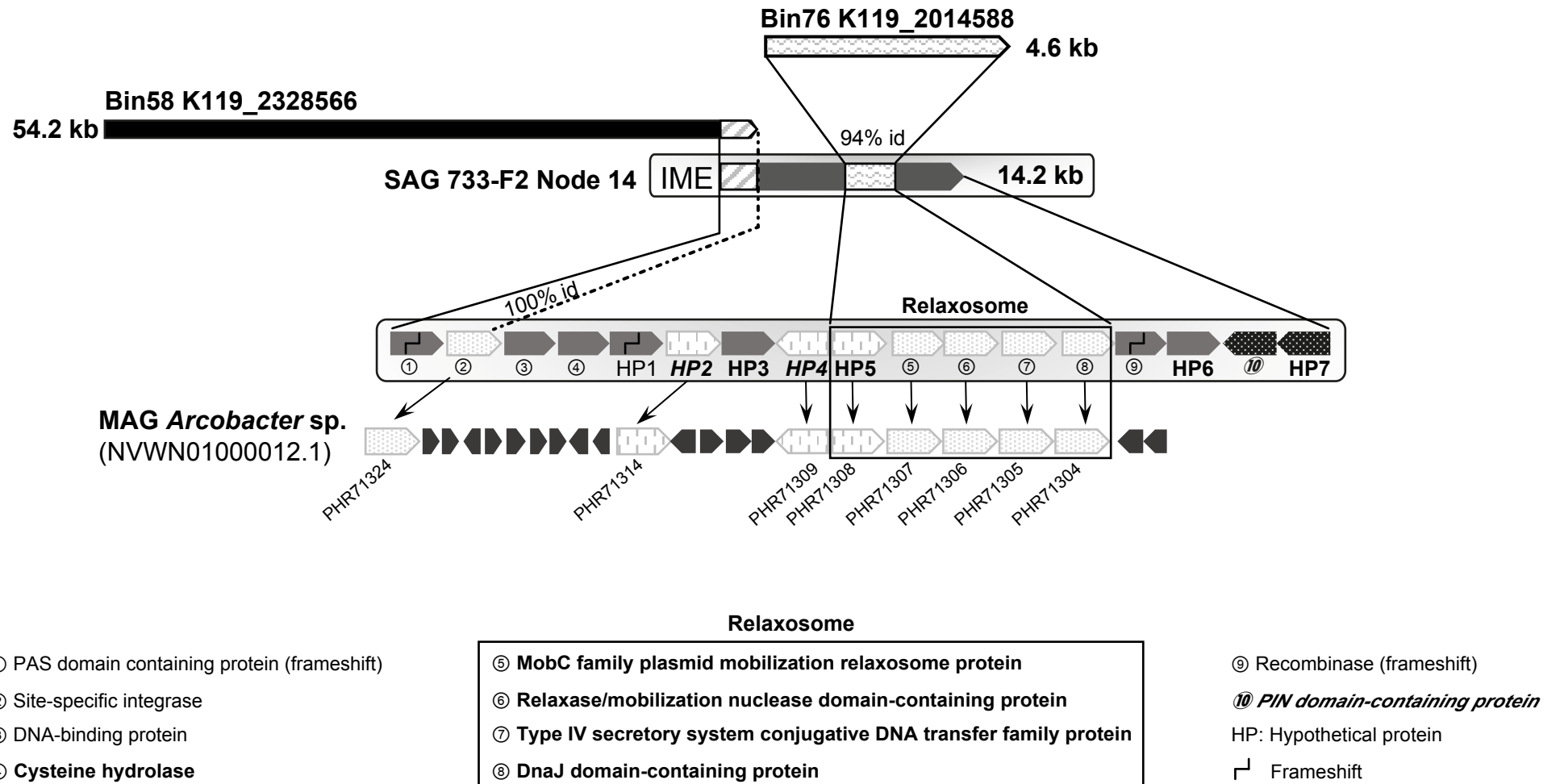


Figure 6

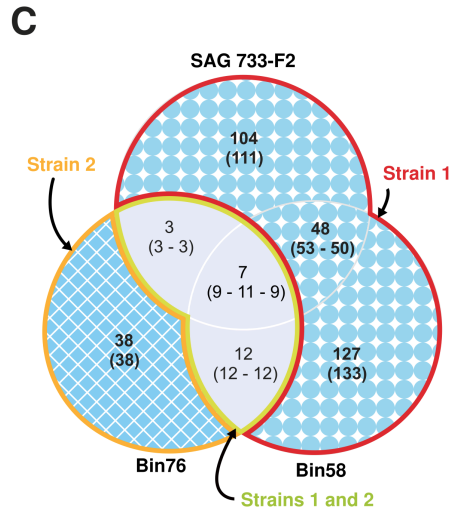
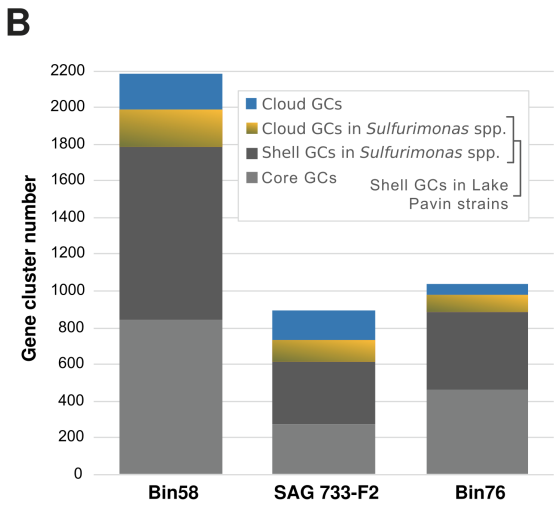
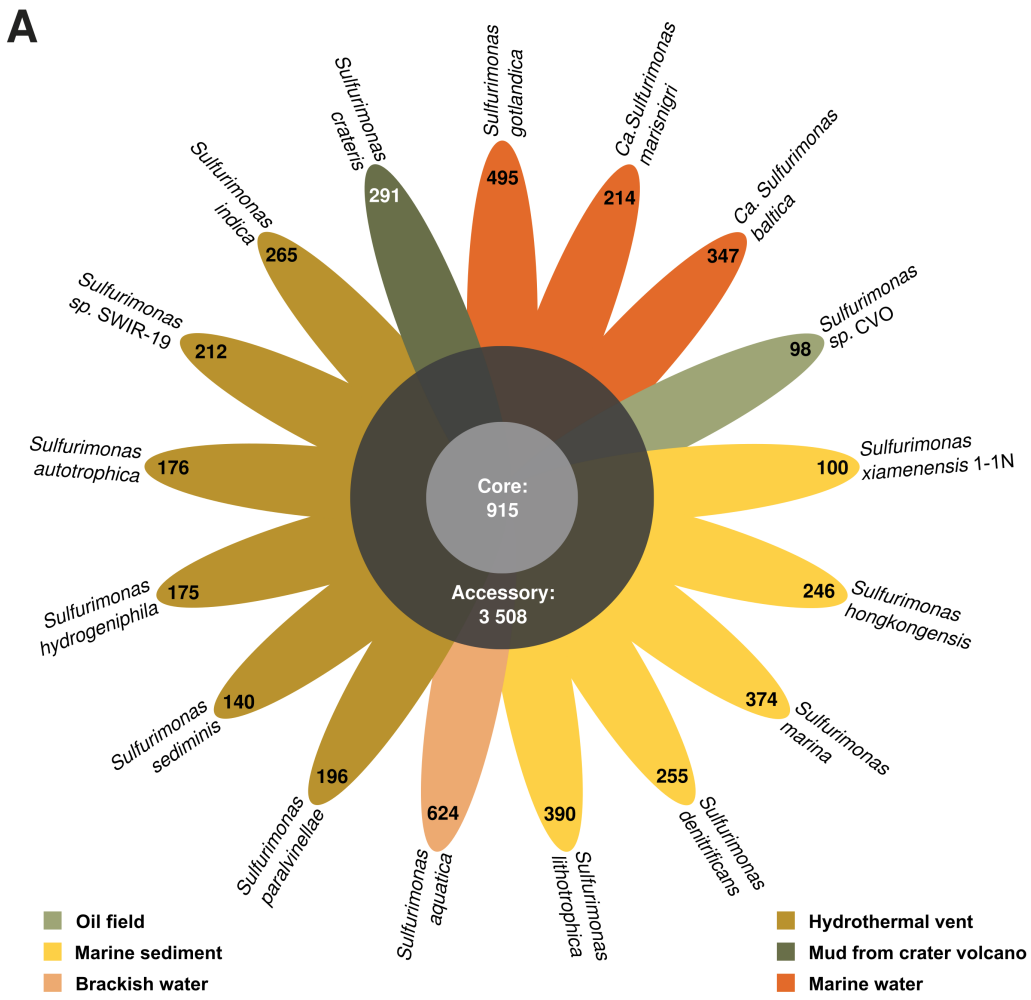


Figure 7

COG hierarchy	Function	COG category	Strain 1 Bin58/SAG 733-F2	Strain 2 Bin76	Strains 1 and 2	
Information storage and processing	Translation, ribosomal structure and biogenesis	J	1			
	Translation, ribosomal structure and biogenesis	JKL	1			
	Transcription	KT	1			
	Replication, recombination and repair	L	5			
Cellular processes and signaling	Cell cycle control, cell division, chromosome partitioning	D	1		3	
	Cell wall/membrane/envelope biogenesis	M			2	
	Cell wall/membrane/envelope biogenesis	MU	1			
	Cell motility	N	3			
	Posttranslational modification, protein turnover, chaperones	O	2			
	Signal transduction mechanisms	T	4			
	Intracellular trafficking, secretion, and vesicular transport	U	1		3	
	Defense mechanisms	V	7	5		
	Mobilome: prophages, transposons	X	17	1	2	
	Metabolism	Energy production and conversion	C		1	
		Energy production and conversion	CO	1		
		Amino acid transport and metabolism	E	3		2
		Nucleotide transport and metabolism	F	1		
Carbohydrate, transport metabolism		G	2		1	
Lipid transport and metabolism		I			1	
Inorganic ion transport and metabolism		P	1		4	
Secondary metabolites biosynthesis		Q	2			
Poorly characterized		Function unknown	S	14	1	2
		General function prediction only	IR	1		
Figure 8	General function prediction only	R	7	1	2	

Table 1

Family	Toxin	Antitoxin	Cell process affected	Stage blocked	Loci numbers																					
					<i>S. paravalvinellae</i>	<i>S. autotrophica</i>	<i>S. indica</i>	<i>S. hydrogeniphila</i>	<i>S. sediminis</i> ‡	<i>S. sp. SWR19</i> ‡	<i>S. aquatica</i>	<i>S. gottlandica</i>	<i>S. Candidatus</i>	<i>S. Candidatus baltica</i>	<i>S. marina</i>	<i>S. denitrificans</i>	<i>S. hongkongensis</i>	<i>S. litotrophica</i>	<i>S. xiamenensis</i>	<i>S. xiamenensis 1-1N</i> *	<i>S. crateris</i>	SAG 733-F2	Bin58	Bin76		
Type II	RelE	HTH	Translation	RNase ribosome-dependent targeting mRNA	1+	10	9	18	22	19	5+5	3	25	11	1	13	13	9	7	7	28	8	9	12		
		RHH																								
		Phd/YefM																								
		CopG																								
		XRE																								
		HigA																								
		PHD																								
		CcdA																								
	AM																									
	DUF																									
	HP																									
	Orphan																									
	YafQ	RelB/DinJ																								
		CopG																								
		PHD																								
		HP																								
	Txe/YoeB	Orphan																								
		PHD																								
		Phd/YefM																								
	HigB	HP																								
		TR																								
		HigA																								
		DNAb																								
		HTH																								
	BrnT	HP																								
		BrnA																								
		CopG																								
		HTH																								
	PemK/MazF	HP																								
		Orphan																								
		CopG																								
		Phd/YefM																								
MazE																										
DNAb																										
VapC	HP																									
	Orphan																									
PIN	CopG																									
	Phd/YefM																									
	CopG																									
HipA	HTH																									
	HP																									
	HicB																									
GNAT	RR																									
	RR																									
	RR																									
NTase	HEPN																									
	DUF86																									
	HP																									
Zeta	HP																									
	HP																									
Yail/YqxD	GIY-YIG																									
	Orphan																									
MMP	HTH																									
	HP																									
Type IV	AbiEii/AbiGii	AbiEi	Anti-phage defense	Viral replication																						
		HTH																								
		DUF																								
		HP																								
		Orphan																								

■: Hydrothermal vent, □: Brackish water, ■: Marine water, ■: Marine sediment, ■: Oil field, ■: Mud from crater, ■: Freshwater

AM: Addition module, DNAb: DNA binding protein, DUF: Protein of unknown function, HP: Hypothetical protein, HTH: Helix turn helix, MMP: Mobile Mystery protein B, NTase: nucleotidyltransferase, PHD: Prevent-host-death family, RHH: Ribbon-helix-helix domain, RR: Response regulator transcription factor, TR: Transcription regulator

nd: not determined

Chromosomal location:

●: 1 copy, ★: 2 copies, ◆: 3 copies, ■: 4 copies, ☆: 5 copies, ✧: 7 copies, ○: orphan toxin (1 copy).

Plasmid location:

●: 1 copy

‡ Strains of the species *S. sediminis*, delineated by double lines

* Strains of the species *S. xiamenensis*, delineated by double lines

Y: black: chromosomal location, pink: plasmid location

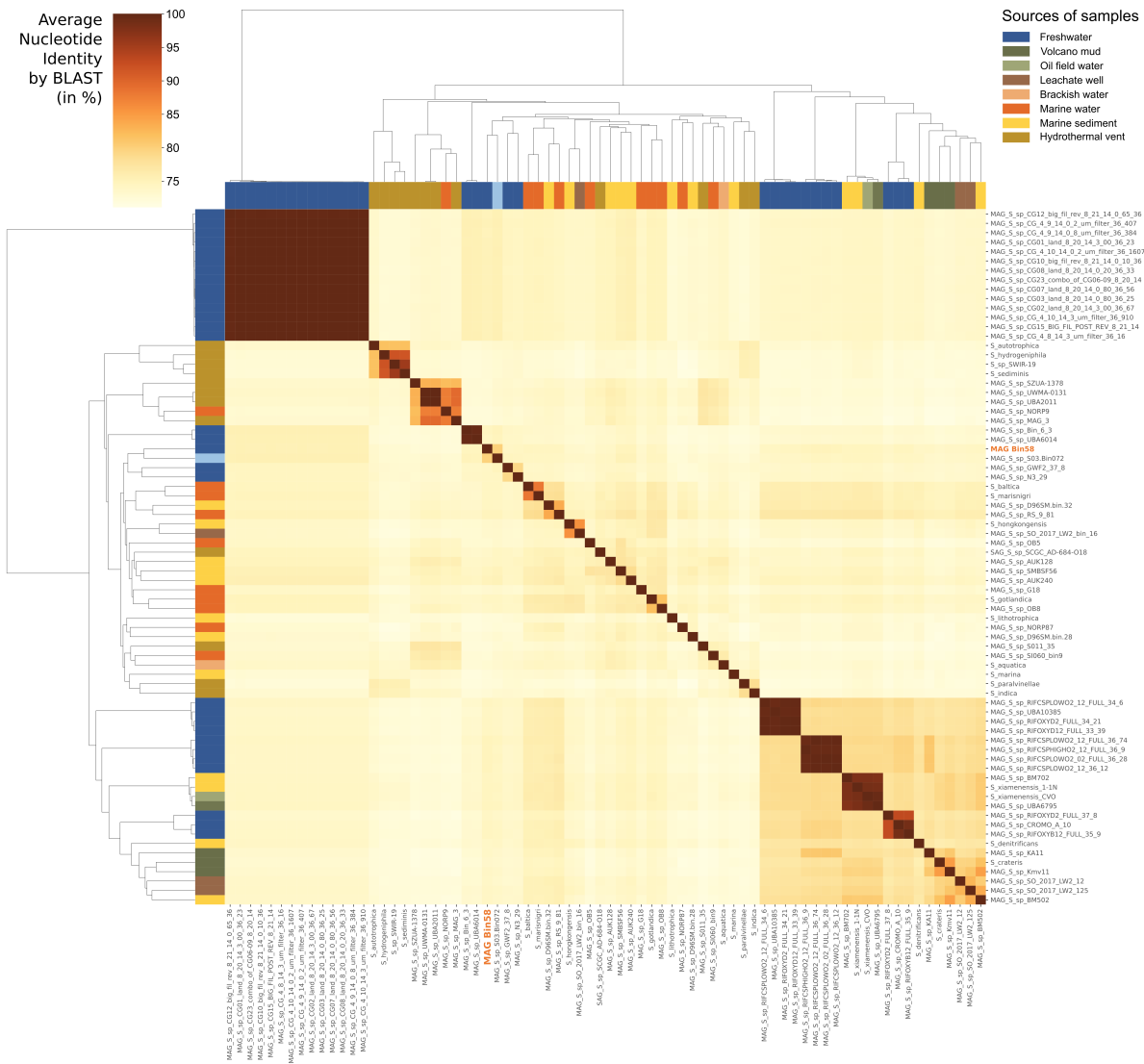


Figure S1

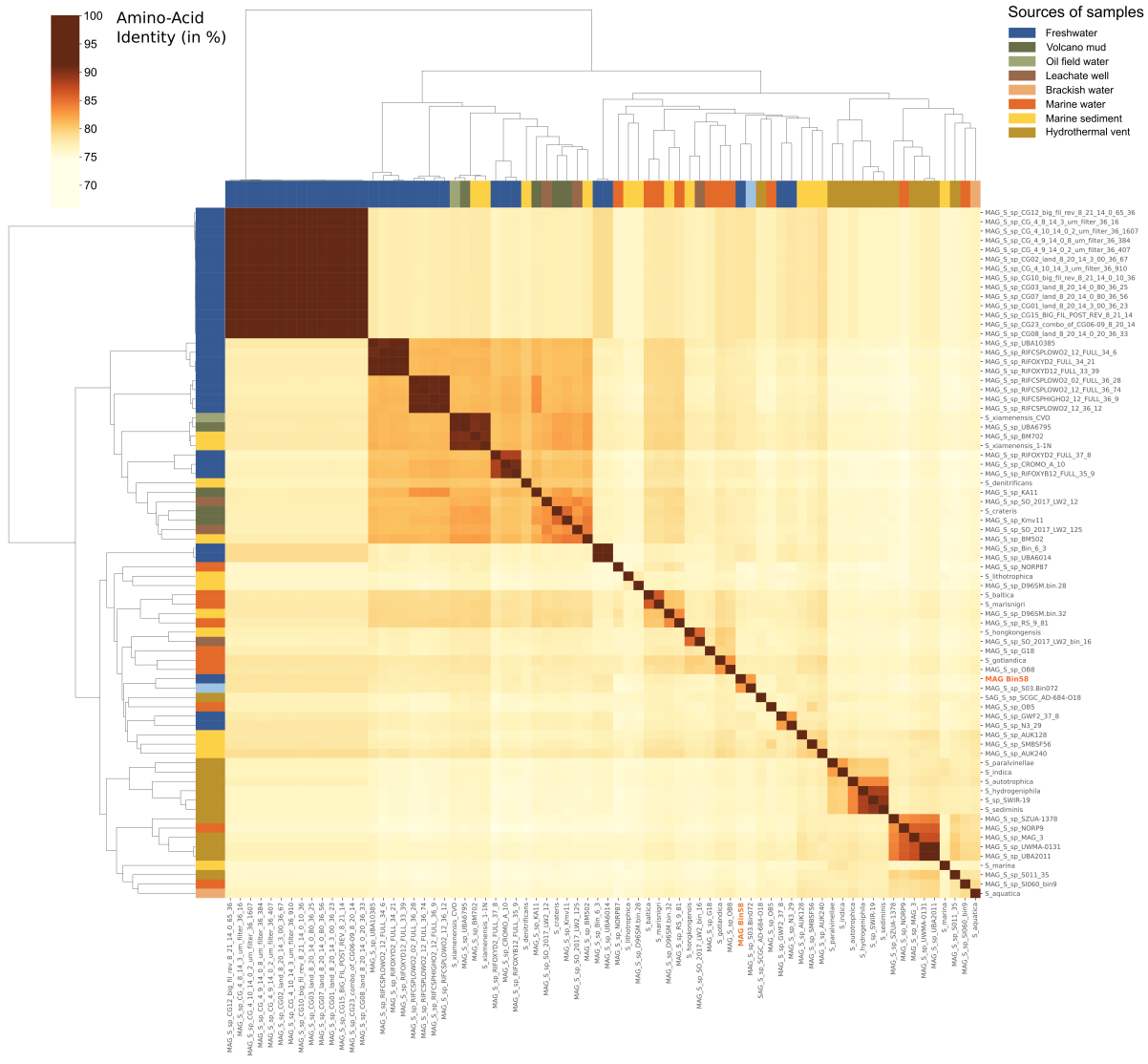


Figure S2

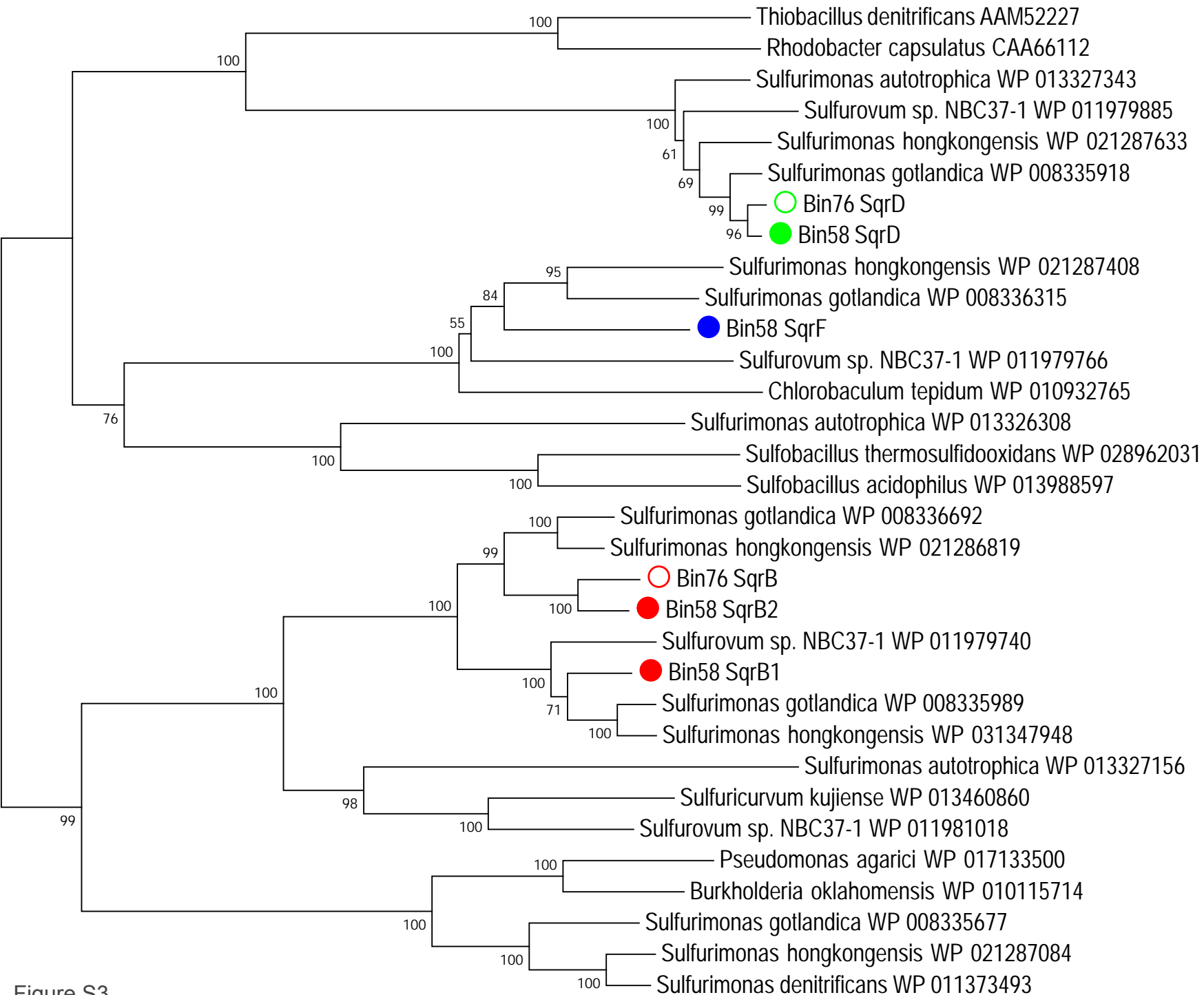


Figure S3

0.2

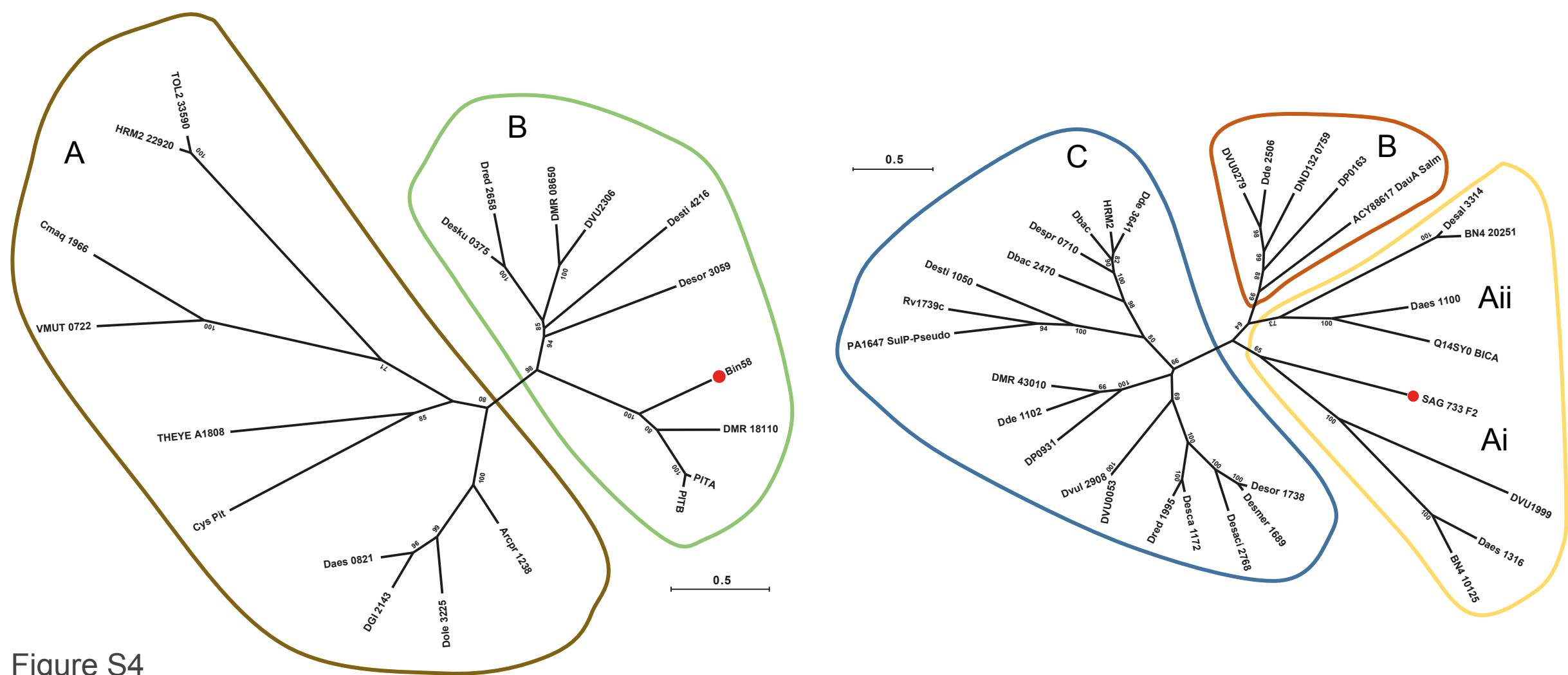
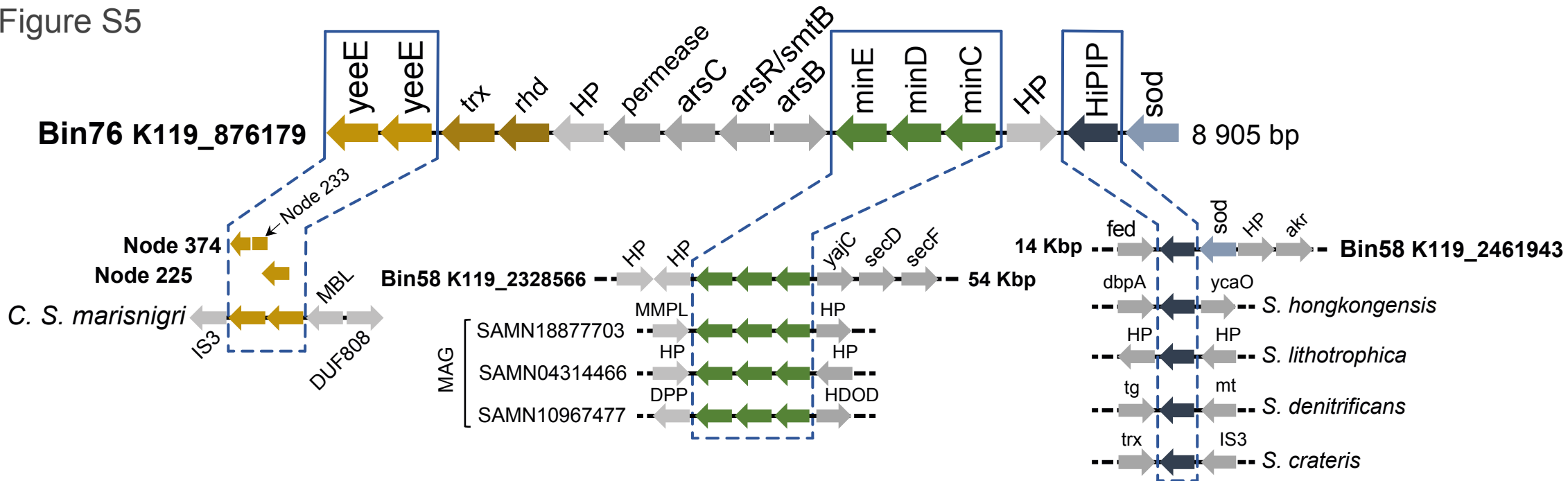


Figure S4

Figure S5



akr: aldo-keto reductase, fed: ferredoxin, ars: arsenate reductase genes, dbpA: decorin-binding protein A, DPP:divergent polysaccharide deacetylase, fed: federroxinHDOD:HDOD domain-containing protein, HiPIP: high potential iron-sulfur proteins, IS3: transposase, MAG: metagenome-assembled genome, MBL: MBL Fold metallo hydrolase, minCDE: Min system, MMPL:MMPL family transporter, sod: superoxide dismutase, rhf: rodhanese-like, tg: transglycosylase, mt:mutarotase, trx: thioredoxin, yajC/secCDF: secretion system, ycaO: ribosomal protein S12 methylthiotransferase accessory factor, HP: hypothetical protein.
 Node: contig from SAG 733-F2.

A Figure S6

N-terminus

TK90	27	TPHRCATPFYLGAVMASMDVDVY	DsrE3B
Atc_2345	24	TAHRCATPFYLGALLSSMDAEV	
LCX93_RS03235	36	IVYQCDFPDIQRIHMLNTINNA	DsrEd
ENK67_03485	36	VVYQCDFADPKRVHMLNTLNNA	
LCX93_03195	21	VVFDSSGNASYIKSRMLVIGKT	DsrEf
SUN_RS08545	20	AVFDCCSSDADYIKSRMLVIGKT	
AS592_RS10230	24	AVFDCCSSDNADYIKSRMLVIGKT	DsrEg
FJR48_RS04845	21	AVFDCAAGDMKFVLTTRMKLLDMT	
JXK04_04865	23	AVFDCAAKDMKFVASRMFLIEES	
LCX_RS03235	71	INVVALGPELQYMMKDFKGTG	DsrEd
ENK67_03485	71	IDIVALGPELQYVMKDFKDTG	
LCX93_03195	57	VALTLHGSSVAMVANDYDMIV	DsrEf
SUN_RS08545	56	FVITLHGGSVSMVSKYKEFV	
AS592_RS10230	60	FAITLHGGSVSMVSKYEMIV	DsrEg
FJR48_RS04845	56	FVLTIHSKCTPIVVKNSDNKD	
JXK04_04865	58	FVLTIHSGTPIAVEEGDDE	

Ahos : *Acidianus hospitalis* W1, **Alvin** : *Allochromonium vinosum* DSM 180, **AS592** : *Sulfurovum rufiae* strain 1812E, **Atc** : *Acidithiobacillus caldus* SM-1, **b3344** : *Escherichia coli* str. K-12, **B649** : *Candidatus Sulfituricum* sp. RIFRC-1, **ENK67** : *Flavobacteriia* bacterium isolate HyVt-598, **FHQ18** : *Deferribacter autotrophicus*, **I6M62** : *Acinetobacter courvalinii*, **IMZ28** : *Sulfurovum indicum* strain ST-419, **JXK04** : *Campylobacteriales* bacterium isolate ZodW, **Lferr** : *Acidithiobacillus ferrooxidans* ATCC 53993, **Mcup** : *Metallospheara cuprina*, **MOP43** : *Arcobacteraceae* bacterium isolate SO_2017_CLC, **Ppha** : *Pelodictyon phaeoclatratiforme* strain DSM 5477, **PSEBR** : *Pseudomonas brassicaearum*, **Saci** : *Sulfolobus acidocaldarius* DSM 639, **SUN** : *Sulfurovum* sp. NBC37-1, **TK90** : *Thioalkalivibrio* sp. K90mix, **YH65** : *Sulfurovum lithotrophicum* strain ATCC BAA-797, **KGZ62** : *Sulfurimonas* sp. CROMO_A_10, **LCX93** : *Sulfurimonas* sp. SWIR-19, **FJR48** : *Sulfurimonas lithotrophica*, **GJV85** : *Sulfurimonas aquatica*.

C-terminus

Bin58	58	NDLVEMLKKMYASGIKLQACGTCNARCGIFK	DsrEa
Bin76	58	YDLVEMLKKMYASGINLQACGTCNARCGIFK	
FHQ18_10630	61	QDLVQMLKELYHSGVELKVCQARCGVVK	DsrEb
LCX93_RS10685	58	YDLAEMLQTMKDGISLQVCGTCNARCGIFK	
YH65_RS07735	58	YDLEMLKKMYTRGINLQVCGTCNARCGLFF	DsrEc
B649_02020	58	HDLVDMKKLYEEGISLQACGTCNARCGLFF	
GJV85_RS05995	56	SDLQERVKVQVATGVKFNACVVDVDDYRVSK	DsrEd
IMZ28_RS04185	52	EALCEVTKVKQKSGVTFSCCIVCSDDYGVTE	
SULKU_RS10025	52	AELQTMVTKVLETVGKASVQVVCSDDYGVTE	DsrEe
LCX93_RS07285	84	KETAVRVDMALQYDVEFVACGNTMRTLHIKE	
BM227_RS06525	83	KKIAYVRVSLQTYDVEFIACGNTMRTLKIKP	DsrEf
LCX93_RS03235	109	VSREFKGLKQLGGDNIHFFACGNTMCKKHVKD	
ENK67_03485	109	ASRIKGLKQLGGDNIKMFACNNTMDKKNVKP	DsrEg
LCX93_RS10690	92	ELQKRIASADTYDVEFLMCKASMPKHNLEA	
IMZ28_RS08640	93	ILAKRIDSLIKTYDVEFLMCGAAMVKNKLAQA	DsrEh
AS592_RS11895	94	DLAKRISAMAKNYDVEFLMCGAAMPKNKLEA	
LCX93_03195	87	QSQQYLKLLAKRKNVTITACAMSLAHNAIEK	DsrE2A
SUN_RS08545	86	KAQEHLKTLATRKKVIVICAMSLDANAIEK	
AS592_RS10230	90	KAQELRVLAEKRGVIVICAMSLASNAIEQ	DsrE2B
FJR48_RS04845	79	AIHQKLTQLKREHNVKIEACKIATNKFYK	
JXK04_04865	82	GISQRLKLLHAEYGVKIEACEIALERFGEIK	DsrE3A
FJR48_RS09400	84	TSIKDMLSSLVKKGGVAVVCGMNAKQVEVKA	
MOP43_03615	85	KSIRDMLISLSKNGGTVMLCGMCASFEEIIT	DsrE3B
KGZ62_05410	112	ATIPELRELQAEVVFIAQMTVELFGFEH	
Alvin_2601	100	ASIEELRELQAEVVKMIACQMTVDLDFDPK	DsrE4
Lferr_2186	97	ASIEELRELQAEVVRMIACQMTVDLFEFDT	
Atc_2353	97	ASLEELRELQAEVVRMIACQMTVDLFEFDT	DsrE5
Ppha_2326	111	ATIEQLRGMQCFEGRFRIACQMTMEVFGEK	
Ahos_1698	71	TWDQLIMQAKEVGEVVKVYACSTTMEFFGLKR	DsrE/TusD
Mcup_0682	80	MWHQLVKEAKDVGEVVKVYACSTTMEFFGIKK	
Saci_0337	74	NPFIRFFDMAKENGVMKYVCVQSLKDMCHMN	
Mcup_0681	74	NPFIHFDMARDNNGVMKYVCVQSLKDMCHMN	
Ahos_1699	68	NPFIHFDMARDNNGVMKYVCVQSLKDMCHMN	
TK90_0639	75	KLKIDFIDAKRAGVRLHVCQPALPGYRIDA	
Atc_2345	72	KKVIEFIRDANAGVEIHVCRPALPGYEIPA	
Saci_0332	67	DFTGLFKKAKKSGKVKIYACSYASKLFFNYSK	
Mcup_1724	66	NVFEHFKAKKSGKVKIYACSYASKIHGLDK	
Mcup_1706	62	KYYLDNLFELAGEDLEITACEFGMRVKDVHE	
I6M62_RS02065	53	DDQRNLKQEWQKLSIRLPVCSAALARGITD	
PSEBR_a3552	58	DLPQQWRTEFVSDNQLDGVVCIAAALRRGVLN	
Alvin_1253	59	HIVNRWAEAEQYELDMVVCVAAALRRGIVD	
b3344	59	DLVRAWQQLNAHQGVALNICVAAALRRGVVD	

* : active cysteine residue common to all DsrE-like proteins

B

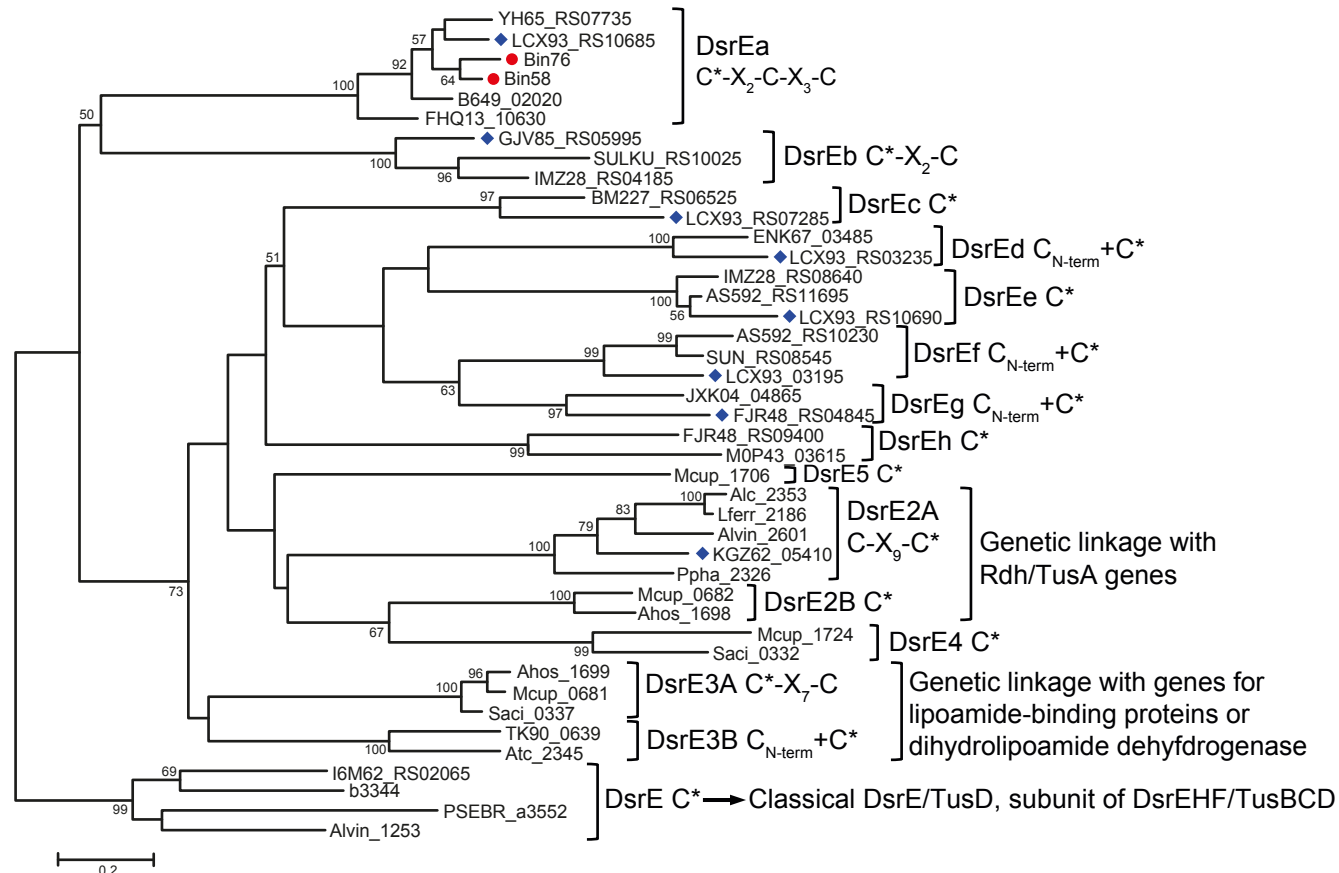
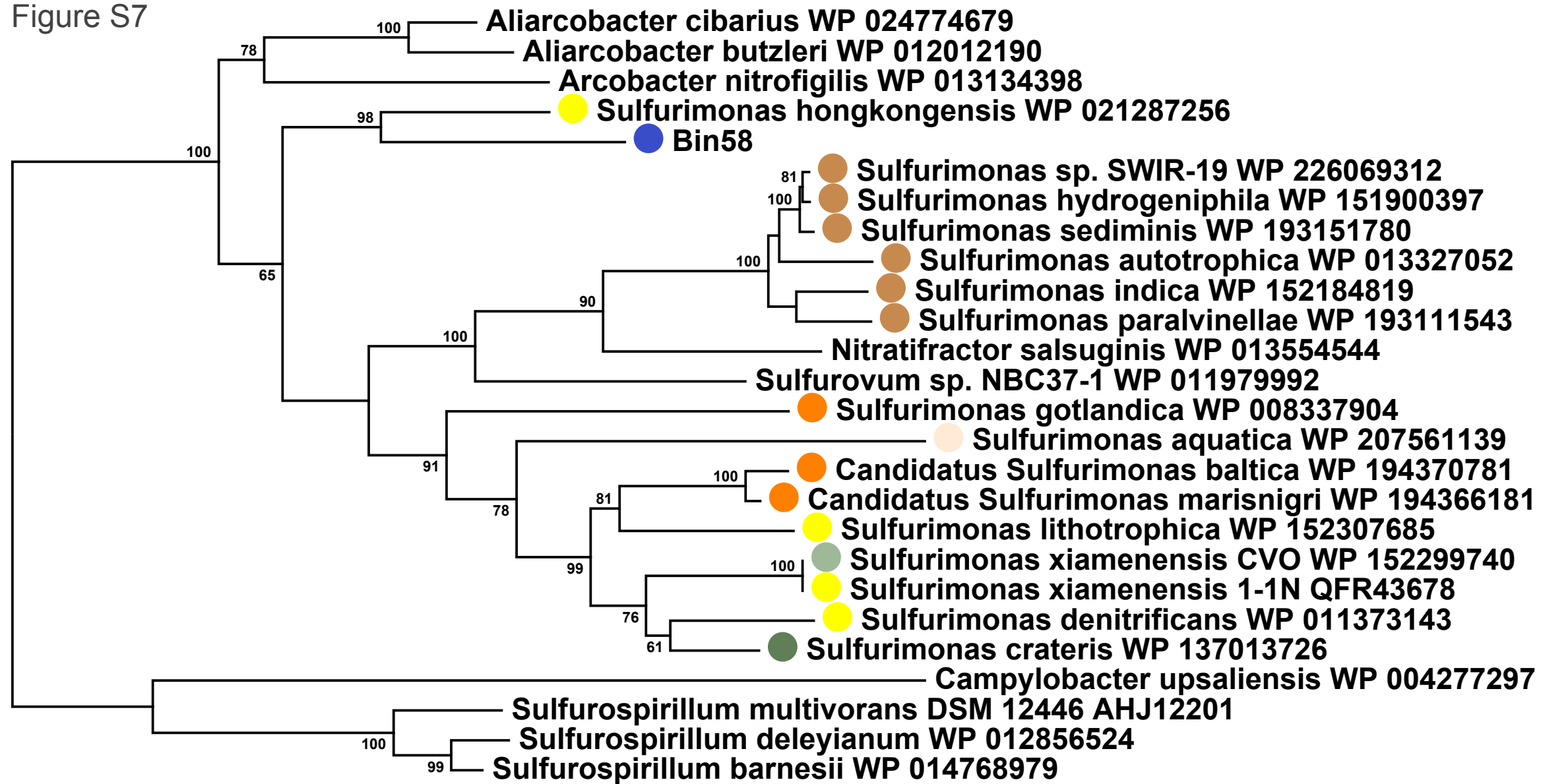
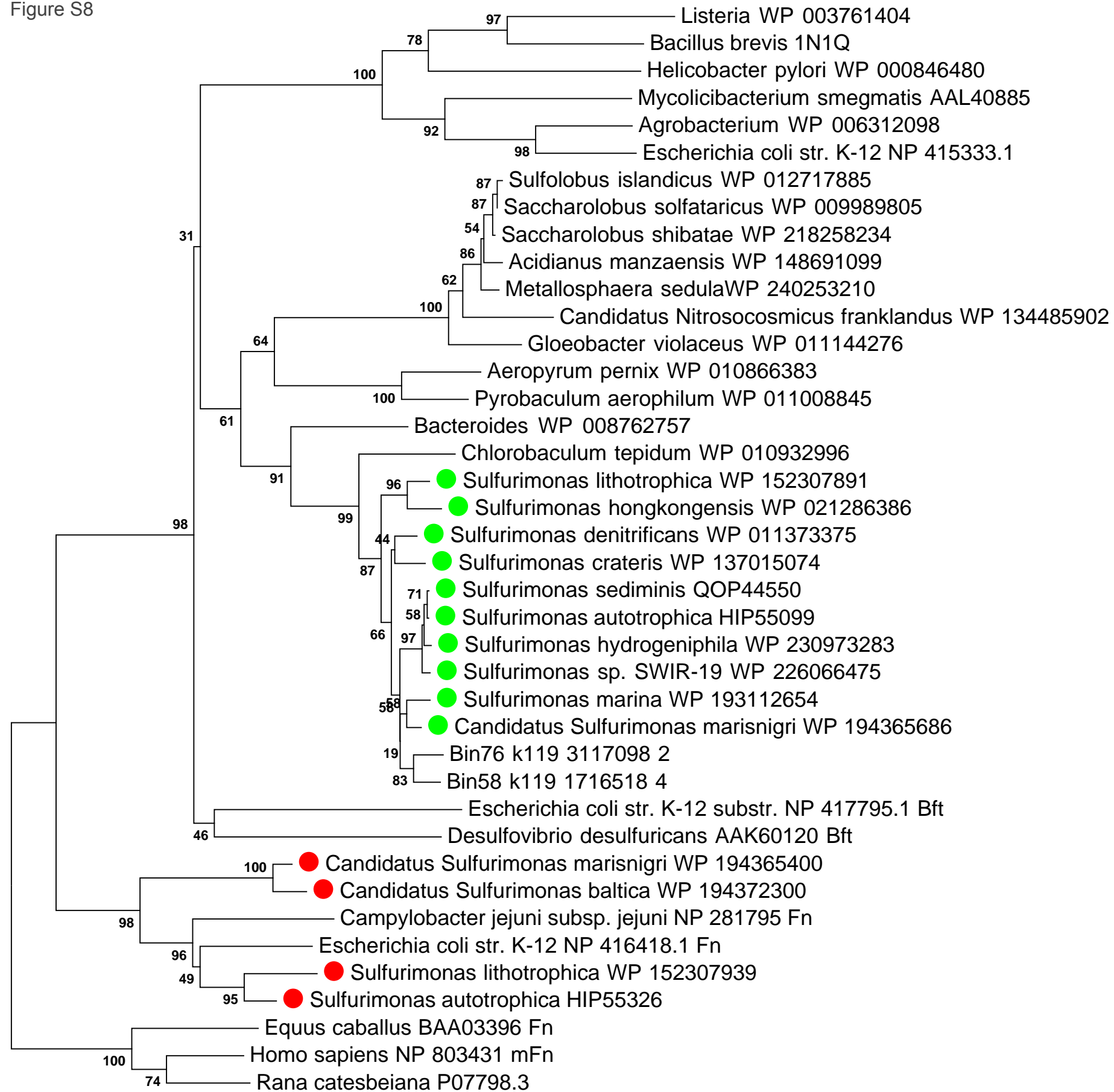


Figure S7



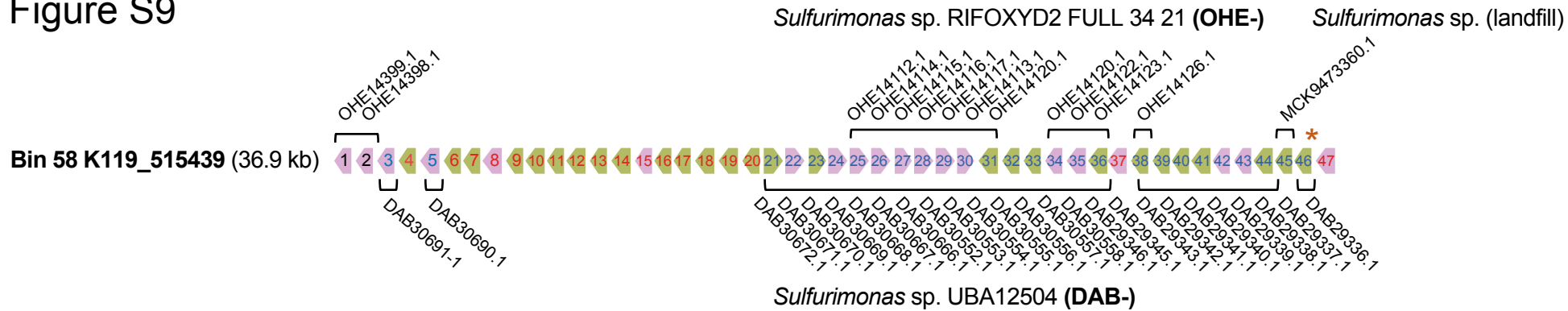
0.05

Figure S8



0.2

Figure S9



▶ Hypothetical proteins
 ★ Only *Sulfurimonas* MAG sequences in NR database (no *Sulfurimonas* species or other taxa sequences)

▶ Known proteins

Nb	Protein name
1	chaperonin GroL
2	co-chaperonin GroES
3	LexA family transcriptional repressor
5	DNA adenine methylase
8	lytic transglycosylase domain-containing protein
15	phage tail tape measure protein
22	phage protease

Nb	Protein name
24	major capsid protein
25	DUF1320 family protein
26	DUF1804 family protein
27	phage terminase large subunit
28	DUF935 family protein
29	phage minor head protein
30	phage virion morphogenesis protein

Nb	Protein name
34	regulatory protein GemA / DUF1018
35	helicase DNaB
37	DUF2528 family protein
42	DNA transposition protein
43	DDE-type integrase/transposase/recombinase
47	phage antirepressor N-terminal domain-containing protei