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

Title

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

Running title: *Sulfurimonas* pan-genomic analysis

Author names and affiliation

Corinne Biderre-Petit[1]*, Damien Courtine[1], Claire Hennequin[1], Pierre E. Galand[2], Stefan Bertilsson[3], Didier Debroas[1], Arthur Monjot[1], Cécile Lepère[1], Anna-Maria Divne[4], Corentin Hochart[2]

[1]Université Clermont Auvergne, CNRS, Laboratoire Microorganismes : Génome et Environnement, F-63000, Clermont-Ferrand, France

[2]Sorbonne Universités, CNRS, Laboratoire d'Ecogéochimie des Environnements Benthiques (LECOB), Observatoire Océanologique de Banyuls, Banyuls sur Mer, France

[3]Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences and Science for Life Laboratory, Sweden

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

*For correspondence: Corinne Biderre-Petit, 1 Impasse Amélie Murat -Bat BioA 63178 Aubière (France). E-mail: corinne.petit@uca.fr, Tel: +33(0)473405139; Fax+33 (0)473407670.

Abstract

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

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

Introduction

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

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

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

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

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

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

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

Materials and methods

Sample collection

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

Single amplified genomes (SAGs)

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

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

Metagenome sequencing, assembly, and binning

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

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

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

Quality evaluation and MAG taxonomic analysis

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

(mcpl) and the mean contamination (mctn) values to classify genome quality according to established community standards (Bowers et al., 2017). Finally, quality scores were calculated as follows: *Q-score = mcpl − 5 × mctn;* and *Q'-score = mcpl − 5 × mctn + mctn × (Checkm strain heterogeneity)/100 + 0.5 × log[N50]*.

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

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

Sulfurimonas MAGs and SAG abundance

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

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

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

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

Analyses of acquired resistance genes were performed by ResFinder v4.1 (Zankari et al., 2012). Insertion sequence (IS) elements were manually annotated according to the ISfinder database

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

Pan-genomics analysis

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

Phylogeny

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

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

Results

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

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

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

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

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

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

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

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

Metabolic potential of the LP strains

Lake Pavin SAG and MAGs were assessed for the presence of specific genome-encoded functional traits involved in numerous processes, including mediation of carbon, nitrogen, sulfur, and dihydrogen (H_2) cycles, oxidative stress, motility, vitamin biosynthesis and transport (Table S8, Fig. 4).

Sulfur metabolism.

Bin58 had the potential to use bisulfide (HS^-) as electron donor *via* either sulfide:quinone reductase (Sqr) or flavocytochrome c sulfide dehydrogenase (Fcc) pathways. Like in most *Sulfurimonas* representatives, Type II (SqrB), IV (SqrD) and VI (SqrF) (Fig. S3, Table S9) were identified, with two copies of Type II as in *S. marina*, *S. hongkongensis* and *S. gotlandica.* Moreover, the presence of a FccB homolog suggests that this pathway could also be used to catalyze initial sulfide oxidation (Fig. 5A). The less complete Bin76 genome had only one Type II and one Type IV Sqr. Only the *soxCDY$_2$Z$_2$H* gene cluster for oxidation of reduced sulfur compounds was found in both LP strains. In *S. xiamenensis* CVO and *S. denitrificans*, that display this phenotype, S^0 oxidation correlates to high expression of an OprD-like porin (CVO_00715 and Suden_1917, respectively, both sharing 57% amino acid sequence identity), for which a homolog (55% amino acid similarity) was found in Bin58. The elemental sulfur reduction in LP strains could also be mediated by the polysulfide reductase-like molybdoenzymes (psrBCA) (Fig. 5A, Table S8).

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

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

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

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

***Nitrogen metabolism*.**

All components required for the denitrification pathway were detected in LP strains (Fig. 5B, Table S11), with a periplasmic nitrate reductase catalytic subunit (NapA) organized in a well conserved operon. Only *S. pluma* is currently known to lack this operon (Molari et al., 2023). Moreover, as previously shown, NapA from hydrothermal vent strains represent a monophyletic group distinct from genomes obtained from other ecosystems (Fig. S7). The LP strains also contain genes encoding ammonium (*atmB1* and *atmB2* genes) and nitrate (*narK* homolog) uptake. These compounds can subsequently be funneled into amino acid metabolism *via* the ferredoxin-nitrite reductase (*nirA*) and the GS-GOGAT/GDH pathways (Fig. 5B). Unlike Bin58 and most other *Sulfurimonas* genomes, Bin76 contains a high-affinity nitrate ABC transporter (NrtABC) for nitrate/nitrite uptake for which homologs have only been found in *Ca. S. marisnigri*, *S. litotrophica, S. marina, S. pluma* and *S. gotlandica*. Bin76 genes also encode key proteins for N_2 fixation (*nifHD, nifA, nifT*) with homologs only detected in *Ca. S. marisnigri, Ca. S. baltica*, *S. litotrophica* and *S. marina*. However, while both *Candidatus* strains can fix N_2 (Henkel et al., 2021), *S. marina* was seemingly unable to grow with N_2 as the sole nitrogen source (Z. Wang et al., 2022).

***H_2, O_2, oxidative stress and central carbon metabolisms*.**

Most *Sulfurimonas* can grow with H_2 as an energy source. The LP strains contained [NiFe]-hydrogenases of groups 1a and 1b that mediate respiratory H_2 uptake, and group 4e, representing a multisubunit membrane-bound Energy-Conserving Hydrogenase (Ech) that couples ferredoxin oxidation to H_2 evolution (Table S8). Unlike group 1 [NiFe]-hydrogenases, which is ubiquitous in *Sulfurimonas spp.*, group 4e is present only in marine pelagic species (*S. gotlandica, Candidatus S. marisnigri* and *Ca. S. baltica)* and in both LP strains.

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

encodes a DNA-binding protein belonging to the ferritin-like diiron-carboxylate protein superfamily (known as Dps). This gene is found in almost half of the *Sulfurimonas* genomes (Fig. S8). In contrast, a second member within this gene superfamily (*i.e.* a gene encoding for a 24-subunit non-heme ferritin) was only found in four *Sulfurimonas* reference genomes (Fig. S8). In addition to the ferritin, LP genomes have genes encoding a superoxide dismutase (SOD), producing H_2O_2 and O_2 as byproducts, and a superoxide reductase (SOR), producing only H_2O_2 (Table S8, Fig. 4A), both designed to specifically minimize stress and cell damage from reactive oxygen species (ROS) (Johnson & Hug, 2019; Lushchak, 2011).

In assessing the potential for aerobic respiration, both LP strains have a gene coding for the cbb3-like high-O_2 affinity cytochrome, as all other *Sulfurimonas* spp. This cytochrome is active in anaerobic processes such as dinitrogen (N_2) fixation and denitrification (Bell et al., 2020). In addition, two hemerythrins, non-heme intracellular O_2-binding proteins, were present in LP strains, one with a single-domain and one with a multi-domain that belongs to the hemerythrin-GGDEF family (Table S8). We also detected an Aer-PAS domain primary energy sensor for motility as well genes coding for a flagellum and chemotaxis (Fig. 4A, Table S8). Finally, both LP strains contain all essential genes for carbon fixation *via* the reverse tricarboxylic acid (rTCA) cycle, a common autotrophic carbon fixation pathway of all *Sulfurimonas* spp. (Fig. 4A).

Mobilome and defense mechanism analysis

MGE in LP strains

MGE is a generic term for a multitude of genomic sequences (plasmid, prophage, ICE, IME, 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 subseafloor (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. paralvinellae*) 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***

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

Pan-genome analysis of the genus *Sulfurimonas*

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

Among the shell GCs identified in LP strains, 340 to 946 belonged to the shell GCs when analyzing the 17 *Sulfurimonas* reference genomes while 96 to 206 belonged to the cloud GCs (Fig. 7B). This is what is generally observed with the addition of new genomes in pan-genomic analyses: certain GCs firstly identified as cloud are subsequently shared by more than one genome after adding additional genomes resulting in their updated classification as shell GCs. In this study, 316 distinct GCs classified as cloud GCs when analyzing the 17 *Sulfurimonas* reference genomes were reclassified into shell GCs when adding LP genomes (Table S13). The prediction of their function *via* assignment to the Clusters of Orthologous Groups (COGs) classification shows that 39% (123 GCs) could not be categorized due to the absence of common features with other proteins. The 61% remaining clusters fell into 18 COGs with the

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

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

Discussion

In this comparative study of Lake Pavin microbiota, we identified two uncharacterized *Sulfurimonas* species inhabiting the bottom anoxic zone using both MAGs and SAGs data. The clear difference we observed in level of completeness between SAG and MAGs affiliated to this genus is congruent with that was previously observed by Alneberg et al. (2018) and attributed this to constraints related to the WGA during single-genome amplification. The MAG quality was ensured by low level of redundancy, rejecting the possibility of contaminant metagenome contigs included in the binning and assembly. Moreover, our abundance profiles suggest that the two LP strains may differ in their ecological preferences related to oxygen (O_2) and electron acceptor/donor availability. Such coexistence of *Sulfurimonas* species within a single ecosystem with overlapping niches has never been described before.

The presence of various genes in the vicinity of the ribosomal genes for the LP Lake Pavin strain 1 implies multiple copies per genome, as observed for most *Sulfurimonas* members (Table S1). Furthermore, phylogenomic analysis as well as the diversity and distribution pattern of specific genes (*e.g. dsrE* genes) suggest that LP strains form, with other freshwater representatives, a distinct freshwater clade within this genus. Regarding all the other *Sulfurimonas* strains analysed in this study, *S. sediminis* S2-6 and S. sp. strain SWIR-19 exhibit a high degree of genome relatedness. This finding allows us to propose renaming the latter to *S. sediminis* strain SWIR-19, similarly to the recent renaming of *S.* sp. strain CVO (renamed *S. xiamenensis* strain CVO (Henkel et al., 2021)). The high genetic similarity, also observed with the *S. hydrogeniphila* strain, could be explained by the fact that they all originate from the Indian Ocean (Table S1) with the most closely related *S. sediminis* S2-6 and SWIR-19 isolated from the same area (GPS coordinates 37.78S 49.64E and 37.28S 49.65E, respectively (S. Wang, Shao, et al., 2021)), while the third, *S. hydrogeniphila*, was recovered from a more distant 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 *soxCDY$_2$Z$_2$H* genes and not the *soxXY$_1$Z$_1$AB* genes
520 (Lahme et al., 2020). The products of this gene cluster are involved in S^0 oxidation with
521 thiosulfate as metabolic end product. Cell attachment to S^0 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 S^0 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 S^0 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

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

The presence of various systems for intracellular iron availability control in LP strain genome suggests its potential role as electron donors. Also, though essential, iron can be a harmful element for the cell by catalyzing the formation of ROS by reacting with H_2O_2 in a Fenton reaction. Almost half of the *Sulfurimonas* genomes, including LP strains, encode a Dps-like protein known to minimize damage from such oxidative stress. Indeed, this protein quenches both Fe^{2+} and H_2O_2 while physically protecting genomic DNA and hence protects the cell against Fenton-reaction (Bradley et al., 2020).

Numerous genes implicated in biofilm formation and motility are present in LP strain genomes. Some of these genes were found only in a very limited number of species (*e.g. pga*, hemerythrin genes). With regard to one of the two hemerythrins found, the GGDEF domain located at its C terminus is believed to be involved in the synthesis of a second messenger that contributes to the control of motility, virulence and biofilm formation (French et al., 2008; Römling et al., 2013). This suggests a link between O_2 levels and the cyclic-di-GMP signaling network in LP strains. However, how the interaction of the hemerythrin-like domain with O_2 alters the activity of the associated signaling domain is still not known and should be investigated. Furthermore, the importance of O_2 in *Sulfurimonas* strains motility is reinforced by the presence, in the few strains where this hemerythrin where detected (including LP strains), of an Aer-PAS domain primary energy sensor for motility, known to respond to changes in O_2 concentration. Hence, in cooperation with chemotaxis, such traits could confer an ability for the cell to position themselves in a microenvironment appropriate for their metabolic needs (Maschmann et al., 2022).

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

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

The LP strains also harbor a broad arsenal of virulence and resistance determinants that may demonstrate their capacity for genetic recombination and enable rapid and efficient adaptation to the local environment. Among the resistance determinants, the FA operon shows an organization that is consistent with the finding that Gram-negative bacteria 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 (≈60m, S: 3.3µM) to slightly sulfidic
598 (≈70m, H_2S: 15µM). 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

Acknowledgements

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

References

Alneberg, J., Bjarnason, B. S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., Lahti, L., Loman, N. J., Andersson, A. F., & Quince, C. (2014). Binning metagenomic contigs by coverage and composition. *Nature Methods*, *11*(11), 1144-1146. https://doi.org/10.1038/nmeth.3103

Alneberg, J., Karlsson, C. M. G., Divne, A.-M., Bergin, C., Homa, F., Lindh, M. V., Hugerth, L. W., Ettema, T. J. G., Bertilsson, S., Andersson, A. F., & Pinhassi, J. (2018). Genomes from uncultivated prokaryotes: A comparison of metagenome-assembled and single-amplified genomes. *Microbiome*, *6*(1), 173. https://doi.org/10.1186/s40168-018-0550-0

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403-410. https://doi.org/10.1016/S0022-2836(05)80360-2

Anantharaman, K., Brown, C. T., Hug, L. A., Sharon, I., Castelle, C. J., Probst, A. J., Thomas, B. C., Singh, A., Wilkins, M. J., Karaoz, U., Brodie, E. L., Williams, K. H., Hubbard, S. S., & Banfield, J. F. (2016). Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nature Communications*, *7*(1), 13219. https://doi.org/10.1038/ncomms13219

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, *19*(5), 455-477. https://doi.org/10.1089/cmb.2012.0021

Bell, E., Lamminmäki, T., Alneberg, J., Andersson, A. F., Qian, C., Xiong, W., Hettich, R. L., Frutschi, M., & Bernier-Latmani, R. (2020). Active sulfur cycling in the terrestrial deep subsurface. *The ISME Journal*, *14*(5), 1260-1272. https://doi.org/10.1038/s41396-020-0602-x

Berben, T., Overmars, L., Sorokin, D. Y., & Muyzer, G. (2019). Diversity and Distribution of Sulfur Oxidation-Related Genes in Thioalkalivibrio, a Genus of Chemolithoautotrophic and Haloalkaliphilic Sulfur-Oxidizing Bacteria. *Frontiers in Microbiology*, *10*, 160. https://doi.org/10.3389/fmicb.2019.00160

Berg, J. S., Jézéquel, D., Duverger, A., Lamy, D., Laberty-Robert, C., & Miot, J. (2019). Microbial diversity involved in iron and cryptic sulfur cycling in the ferruginous, low-sulfate waters of Lake Pavin. *PLOS ONE*, *14*(2), e0212787. https://doi.org/10.1371/journal.pone.0212787

Biderre-Petit, C., Boucher, D., Kuever, J., Alberic, P., Jézéquel, D., Chebance, B., Borrel, G., Fonty, G., & Peyret, P. (2011). Identification of sulfur-cycle prokaryotes in a low-sulfate lake (Lake Pavin) using *aprA* and 16S rRNA gene markers. *Microbial Ecology*, *61*(2), 313-327. https://doi.org/10.1007/s00248-010-9769-4

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 database of bacterial integrative and conjugative elements. *Nucleic Acids Research*, *47*(D1), D660-D665. https://doi.org/10.1093/nar/gky1123

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Soto-Jimenez, L., Estrada, K., & Sanchez-Flores, A. (2014). GARM: Genome Assembly, Reconciliation and Merging Pipeline. *Current Topics in Medicinal Chemistry*, *14*(3), 418-424. https://doi.org/10.2174/1568026613666131204110628

Takai, K., Suzuki, M., Nakagawa, S., Miyazaki, M., Suzuki, Y., Inagaki, F., & Horikoshi, K. (2006). *Sulfurimonas paralvinellae* sp. Nov., a novel mesophilic, hydrogen- and sulfur-oxidizing chemolithoautotroph within the *Epsilonproteobacteria* isolated from a deep-sea hydrothermal vent polychaete nest, reclassification of *Thiomicrospira denitrificans* as *Sulfurimonas denitrificans* comb. Nov. And emended description of the genus *Sulfurimonas*. *International Journal of Systematic and Evolutionary Microbiology*, *56*(8), 1725-1733. https://doi.org/10.1099/ijs.0.64255-0

Tanaka, Y., Yoshikaie, K., Takeuchi, A., Ichikawa, M., Mori, T., Uchino, S., Sugano, Y., Hakoshima, T., Takagi, H., Nonaka, G., & Tsukazaki, T. (2020). Crystal structure of a YeeE/YedE family protein engaged in thiosulfate uptake. *Science Advances*, *6*(35), eaba7637. https://doi.org/10.1126/sciadv.aba7637

Taviti, A. C., & Beuria, T. K. (2019). Bacterial Min proteins beyond the cell division. *Critical Reviews in Microbiology*, *45*(1), 22-32. https://doi.org/10.1080/1040841X.2018.1538932

Taylor, G. T., Suter, E. A., Pachiadaki, M. G., Astor, Y., Edgcomb, V. P., & Scranton, M. I. (2018). Temporal shifts in dominant sulfur-oxidizing chemoautotrophic populations across the Cariaco Basin's redoxcline. *Deep Sea Research Part II: Topical Studies in Oceanography*, *156*, 80-96. https://doi.org/10.1016/j.dsr2.2017.11.016

Thompson, C. C., Chimetto, L., Edwards, R. A., Swings, J., Stackebrandt, E., & Thompson, F. L. (2013). Microbial genomic taxonomy. *BMC Genomics, 23*(14), 913. https://doi.org/ 10.1186/1471-2164-14-913

Tian, H., Gao, P., Chen, Z., Li, Y., Li, Y., Wang, Y., Zhou, J., Li, G., & Ma, T. (2017). Compositions and Abundances of Sulfate-Reducing and Sulfur-Oxidizing Microorganisms in Water-Flooded Petroleum Reservoirs with Different Temperatures in China. *Frontiers in Microbiology*, *8*, 143. https://doi.org/10.3389/fmicb.2017.00143

Tully, B. J., Wheat, C. G., Glazer, B. T., & Huber, J. A. (2018). A dynamic microbial community with high functional redundancy inhabits the cold, oxic subseafloor aquifer. *The ISME Journal*, *12*(1), 1-16. https://doi.org/10.1038/ismej.2017.187

van Vliet, D. M., von Meijenfeldt, F. A. B., Dutilh, B. E., Villanueva, L., Sinninghe Damsté, J. S., Stams, A. J. M., & Sánchez-Andrea, I. (2021). The bacterial sulfur cycle in expanding dysoxic and euxinic marine waters. *Environmental Microbiology*, *23*(6), 2834-2857. https://doi.org/10.1111/1462-2920.15265

Wacey, D., Kilburn, M. R., Saunders, M., Cliff, J., & Brasier, M. D. (2011). Microfossils of sulphur-metabolizing cells in 3.4-billion-year-old rocks of Western Australia. *Nature Geoscience*, *4*(10), 698-702. https://doi.org/10.1038/ngeo1238

Wang, L., Cheung, M. K., Liu, R., Wong, C. K., Kwan, H. S., & Hwang, J.-S. (2017). Diversity of Total Bacterial Communities and Chemoautotrophic Populations in Sulfur-Rich Sediments of Shallow-Water Hydrothermal Vents off Kueishan Island, Taiwan. *Microbial Ecology*, *73*(3), 571-582. https://doi.org/10.1007/s00248-016-0898-2

Wang, S., Jiang, L., Hu, Q., Cui, L., Zhu, B., Fu, X., Lai, Q., Shao, Z., & Yang, S. (2021). Characterization of *Sulfurimonas hydrogeniphila* sp. Nov., a Novel Bacterium Predominant in Deep-Sea Hydrothermal Vents and Comparative Genomic Analyses of the Genus *Sulfurimonas*. *Frontiers in Microbiology*, *12*, 626705. https://doi.org/10.3389/fmicb.2021.626705

Wang, S., Jiang, L., Xie, S., Alain, K., Wang, Z., Wang, J., Liu, D., & Shao, Z. (2023). Disproportionation of Inorganic Sulfur Compounds by Mesophilic Chemolithoautotrophic *Campylobacterota*. *MSystems*, *8*(1), e00954-22. https://doi.org/10.1128/msystems.00954-22

Wang, S., Shao, Z., Lai, Q., Liu, X., Xie, S., Jiang, L., & Yang, S. (2021). *Sulfurimonas sediminis* sp. Nov., a novel hydrogen- and sulfur-oxidizing chemolithoautotroph isolated from a hydrothermal vent at the Longqi system, southwestern Indian ocean. *Antonie van Leeuwenhoek*, *114*(6), 813-822. https://doi.org/10.1007/s10482-021-01560-4

Wang, X., Preston, J. F., & Romeo, T. (2004). The *pgaABCD* Locus of *Escherichia coli* Promotes the Synthesis of a Polysaccharide Adhesin Required for Biofilm Formation. *Journal of Bacteriology*, *186*(9), 2724-2734. https://doi.org/10.1128/JB.186.9.2724-2734.2004

Wang, Z., Wang, S., Lai, Q., Wei, S., Jiang, L., & Shao, Z. (2022). Sulfurimonas marina sp. Nov., an obligately chemolithoautotrophic, sulphur-oxidizing bacterium isolated from a deep-sea sediment sample from the South China Sea. *International Journal of Systematic and Evolutionary Microbiology*, *72*(10), 005582. https://doi.org/10.1099/ijsem.0.005582

Waskom, M. (2021). seaborn: Statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. https://doi.org/10.21105/joss.03021

Whelan, S., & Goldman, N. (2001). A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. *Molecular Biology and Evolution*, *18*(5), 691-699. https://doi.org/10.1093/oxfordjournals.molbev.a003851

Wiedenheft, B., Mosolf, J., Willits, D., Yeager, M., Dryden, K. A., Young, M., & Douglas, T. (2005). An archaeal antioxidant: Characterization of a Dps-like protein from *Sulfolobus solfataricus*. *Proceedings of the National Academy of Sciences*, *102*(30), 10551-10556. https://doi.org/10.1073/pnas.0501497102

Xie, Y., Wei, Y., Shen, Y., Li, X., Zhou, H., Tai, C., Deng, Z., & Ou, H.-Y. (2018). TADB 2.0: An updated database of bacterial type II toxin–antitoxin loci. *Nucleic Acids Research*, *46*(D1), D749-D753. https://doi.org/10.1093/nar/gkx1033

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F. O., Ludwig, W., Schleifer, K.-H., Whitman, W. B., Euzéby, J., Amann, R., & Rosselló-Móra, R. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nature Reviews Microbiology*, *12*(9), 635-645. https://doi.org/10.1038/nrmicro3330

Yeo, C. C. (2018). GNAT toxins of bacterial toxin-antitoxin systems: Acetylation of charged tRNAs to inhibit translation: GNAT toxins of bacterial toxin-antitoxin systems. *Molecular Microbiology*, *108*(4), 331-335. https://doi.org/10.1111/mmi.13958

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F. M., & Larsen, M. V. (2012). Identification of acquired antimicrobial resistance genes. *Journal of Antimicrobial Chemotherapy*, *67*(11), 2640-2644. https://doi.org/10.1093/jac/dks261

Zeng, X., Alain, K., & Shao, Z. (2021). Microorganisms from deep-sea hydrothermal vents. *Marine Life Science & Technology*, *3*(2), 204-230. https://doi.org/10.1007/s42995-020-00086-4

Data accessibility

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

ORCID

Corinne Biderre-Petit https://orcid.org/0000-0001-7962-4171
Arthur Monjot https://orcid.org/0000-0002-6978-4785
Cécile Lepère https://orcid.org/0000-0003-4767-0477
Didier Debroas https://orcid.org/ 0000-0002-9915-1268
Pierre Galand https://orcid.org/0000-0002-2238-3247

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

Conflict of interests

The authors declare that they have no known competing interests.

Author contributions

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

Figure titles and legends

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Table S8: Specific functions involved in numerous processes, including: carbon, nitrogen, sulfur, hydrogen (H_2) cycling, oxidative stress, motility, vitamin biosynthesis, transport, etc, identified in Lake Pavin strains.

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

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

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

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

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

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

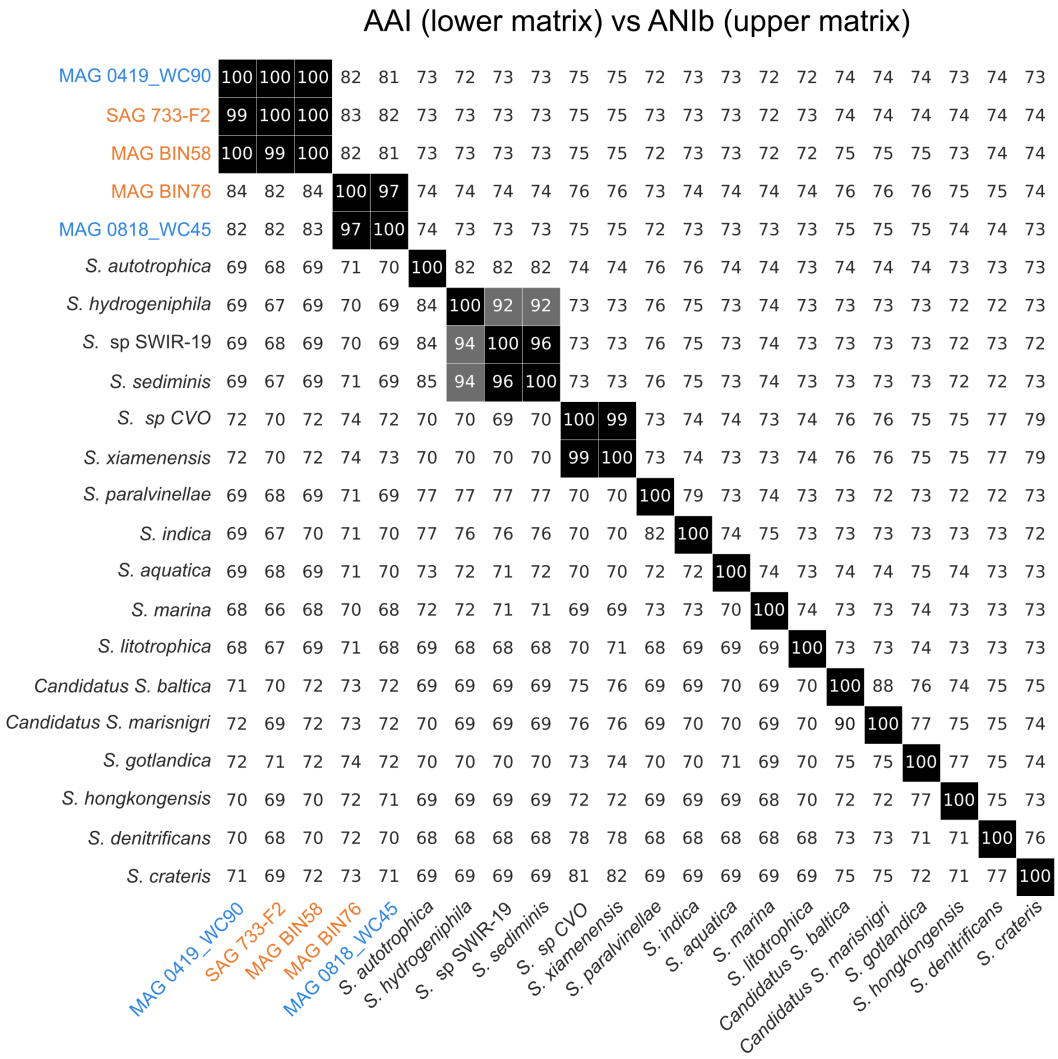

AAI (lower matrix) vs ANIb (upper matrix)
MAG 0419_WC90 100 100 100 82 81 73 72 73 73 75 75 72 73 73 72 72 74 74 74 73 74 73
SAG 733-F2 99 100 100 83 82 73 73 73 73 75 75 73 73 73 72 73 74 74 74 74 74 74
MAG BIN58 100 99 100 82 81 73 73 73 73 75 75 72 73 73 72 72 75 75 75 73 74 74
MAG BIN76 84 82 84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

Figure 1

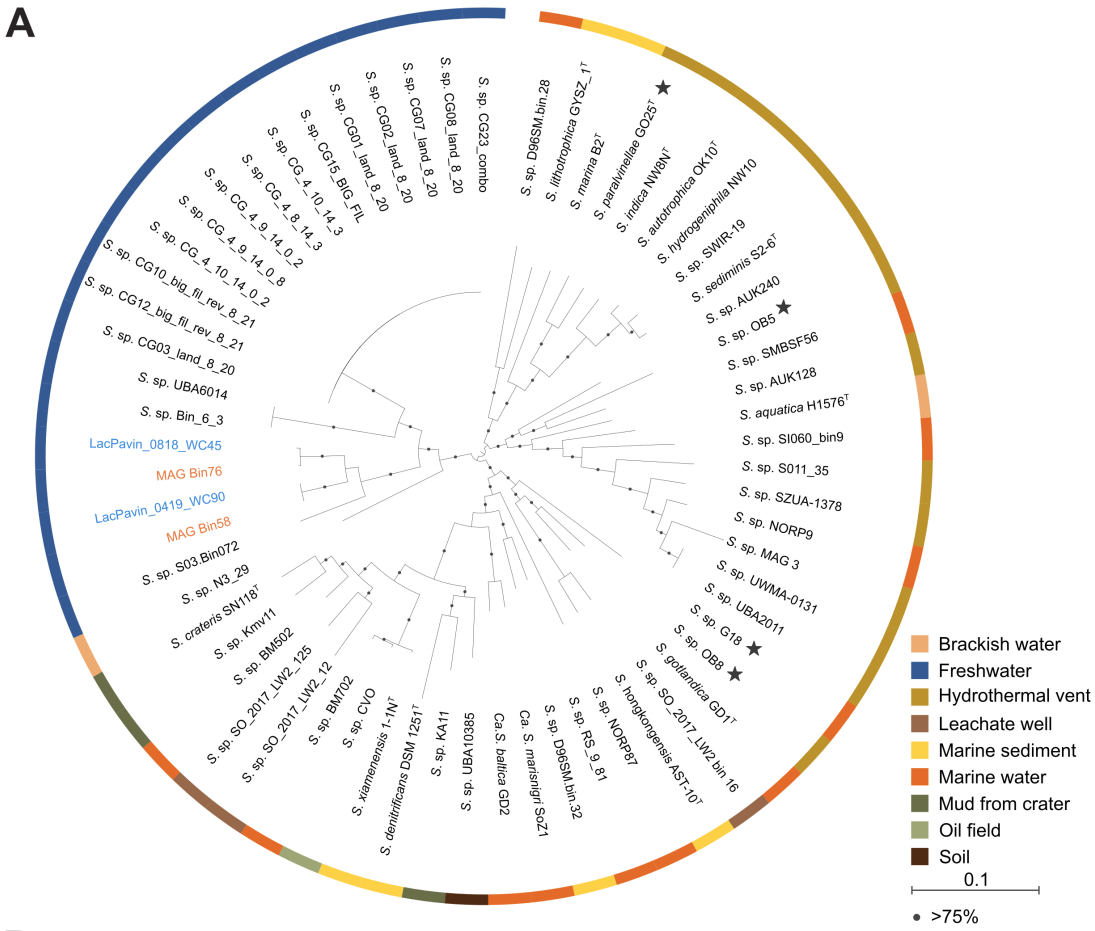

A
S. sp. CG23_combo
S. sp. CG08_land_8_20
S. sp. CG07_land_8_20
S. sp. CG02_land_8_20
S. sp. CG01_land_8_20
S. sp. CG15_BIG_FIL
S. sp. CG_4_10_14_3
S. sp. CG_4_8_14_3
S. sp. CG_4_9_14_0_2
S. sp. CG_4_9_14_0_8
S. sp. CG_4_10_14_0_2
S. sp. CG10_big_fil_rev_8_21
S. sp. CG12_big_fil_rev_8_21
S. sp. CG03_land_8_20
S. sp. UBA6014
S. sp. Bin_6_3
LacPavin_0818_WC45
MAG Bin76
LacPavin_0419_WC90
MAG Bin58
S. sp. S03Bin072
S. sp. N3_29
S. crateris SN118T
S. sp. Kmv11
S. sp. BM502
S. sp. SO_2017_LW2_125
S. sp. SO_2017_LW2_12
S. sp. BM702
S. sp. CVO
S. xiamenensis 1-1NT
S. denitrificans DSM 1251T
S. sp. KA11
S. sp. UBA10385
Ca. S. baltica GD2
Ca. S. marisnigri SoZ1
S. sp. D96SM.bin.32
S. sp. RS_9_81
S. sp. NORP87
S. hongkongensis AST-10T
S. sp. SO_2017_LW2.bin.16
S. gotlandica GD1T
S. sp. OB8
S. sp. G18
S. sp. UBA2011
S. sp. UWMA-0131
S. sp. MAG 3
S. sp. NORP9
S. sp. SZUA-1378
S. sp. S011_35
S. sp. SI060_bin9
S. aquatica H1576T
S. sp. AUK128
S. sp. SMBSF56
S. sp. OB5
S. sp. AUK240
S. sediminis S2-6T
S. sp. SIWIR-19
S. hydrogeniphila NW10
S. autotrophica OK10T
S. indica NW8T
S. paralivinellae GO25T
S. marina B2T
S. lithotrophica GYSZ_1T
S. sp. D96SM.bin.28
Brackish water
Freshwater
Hydrothermal vent
Leachate well
Marine sediment
Marine water
Mud from crater
Oil field
Soil
0.1
>75%
B
Bin58
Metagenomes MetaG2
SAG 733-F2
K119_1813739
7 798 bp
Contig_65m_8143
7 534 bp
Contig_65m_181779
896 bp
Contig_70m_91318
1 445 bp
Contig_80m_26660
1 059 bp
Contig_65m_62274
1 861 bp
23S 3' end
Contig_80m_9714
1 908 bp
23S 5' end
Contig_80m_48135
738 bp
23S middle
Contig_65m_88545
1 523 bp
Contig_70m_20155
4 040 bp
Contig_80m_6164
2 469 bp
Node 37
8 649 bp
5S
23S
16S
1/1'- leucyl/phenylalanyl-tRNA-protein transferase
2- ClpA protease
3- ClpS protease
4- dethiobiotin synthetase
5- aspartate carbamoyltransferase, catalytic chain
6- aminodeoxychorismate synthase component I
7- two-component sensor histidine kinase
8- hypothetical protein
9- transposase IS30 (with a frameshift)
10- hypothetical protein
a/b- ATP citrate (pro-S) lyase
c- DUF6056
d- aminodeoxychorismate/anthranilate synthase component II
e- ABC transporter permease
f- flagellar hook-associated protein (FlgK)

Figure 2

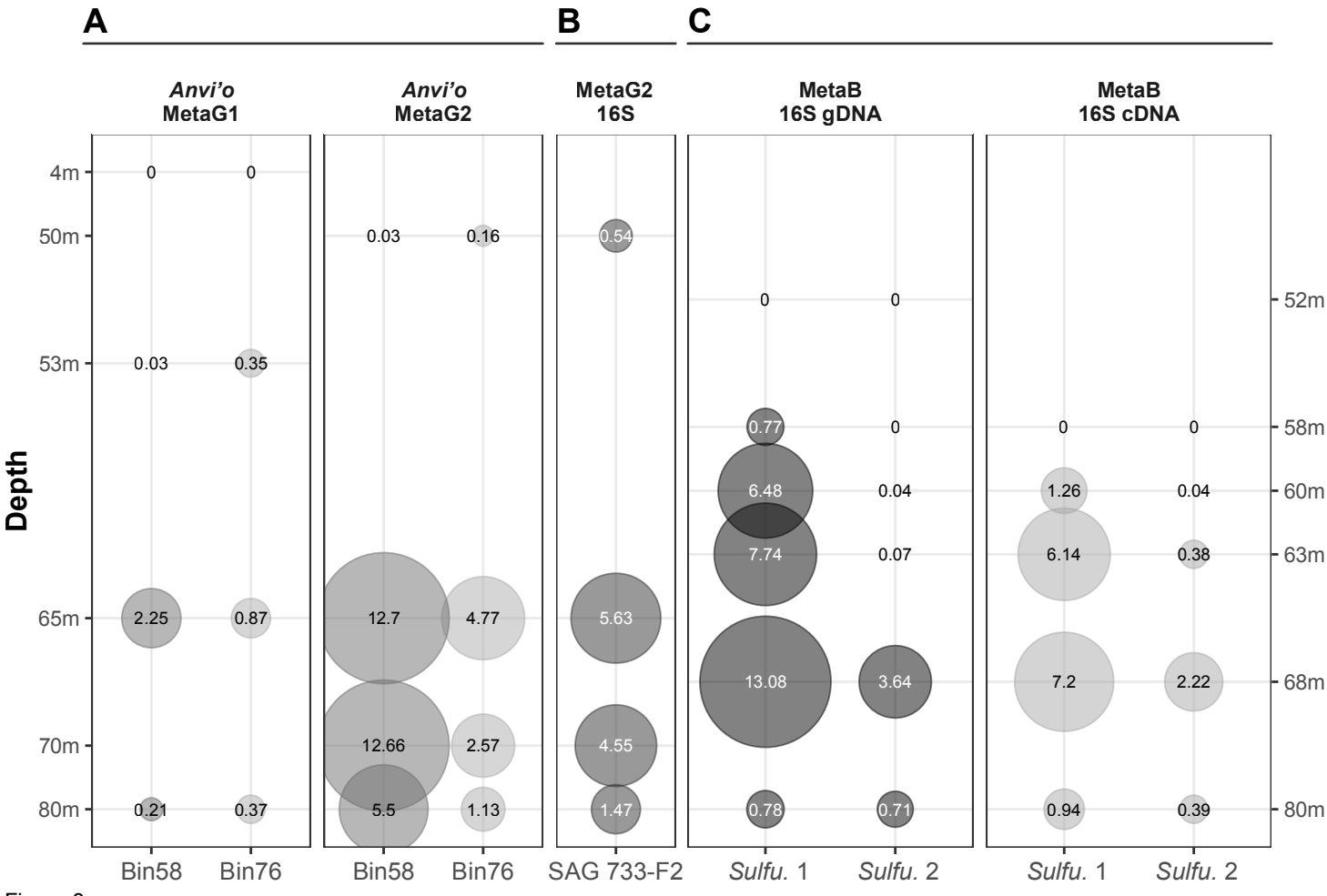
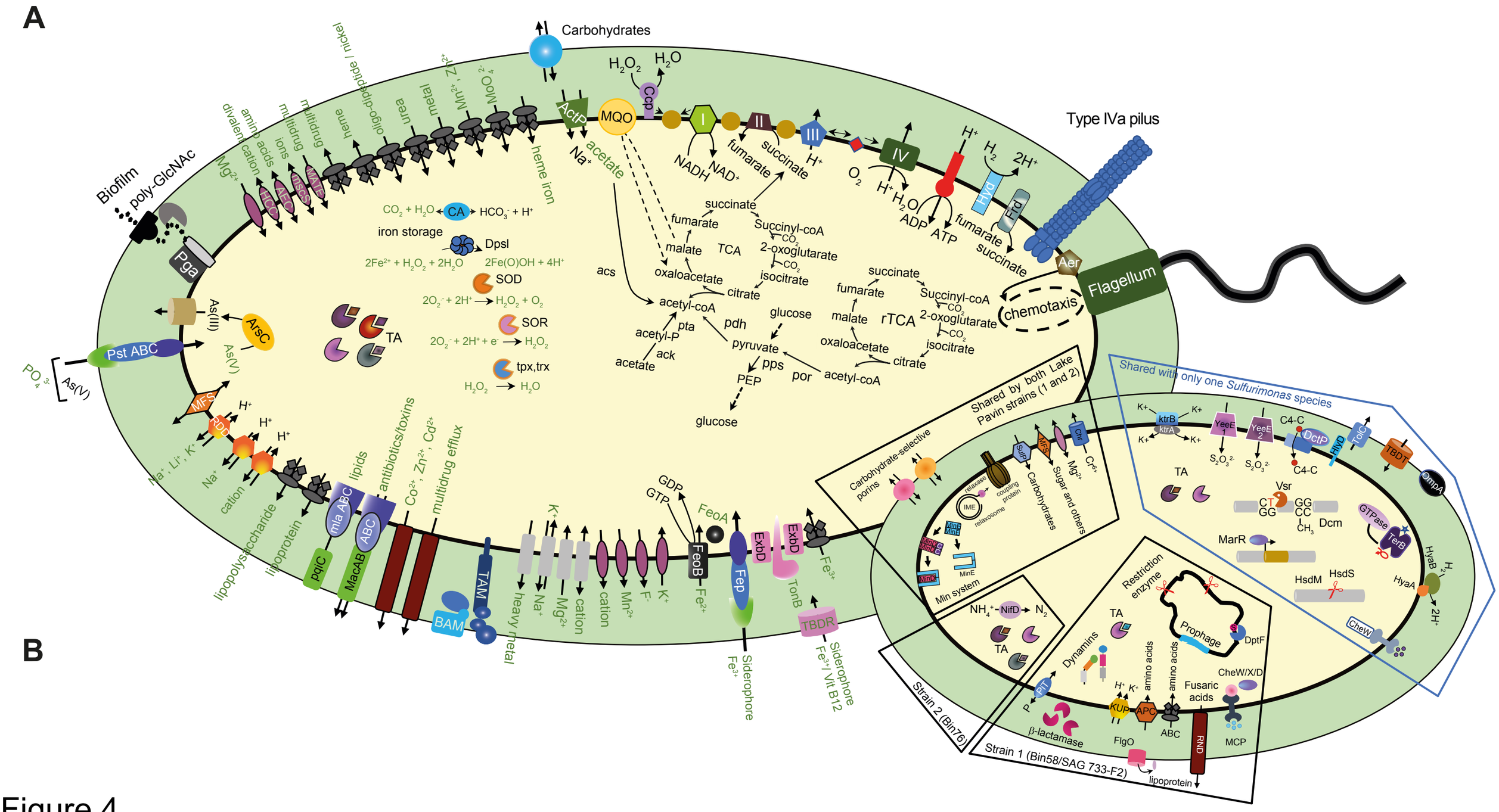
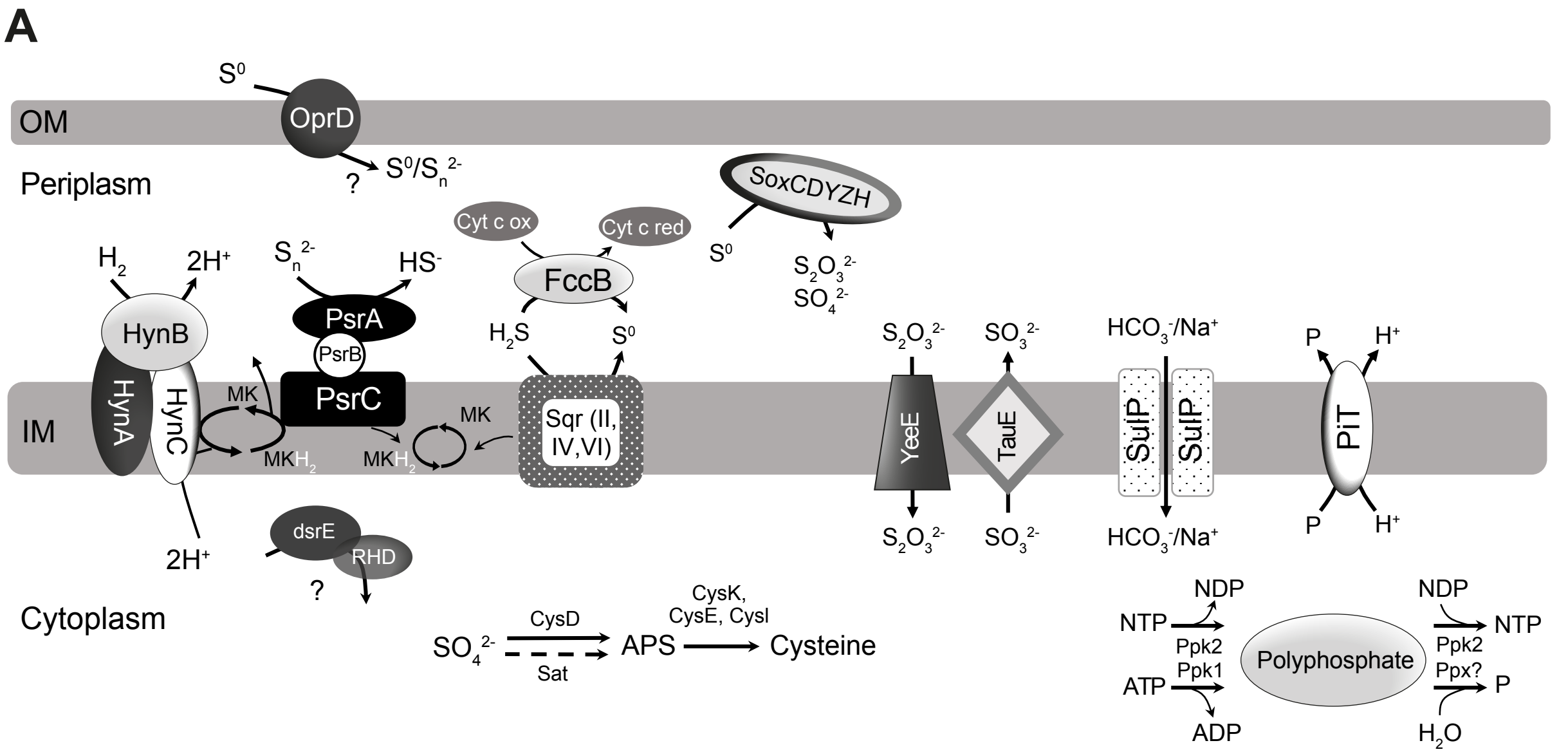

A
B
C
Anvi'o MetaG1
Anvi'o MetaG2
MetaG2 16S
MetaB 16S gDNA
MetaB 16S cDNA
Depth
4m
50m
53m
65m
70m
80m
52m
58m
60m
63m
68m
80m
0
0
0.03
0.35
2.25
0.87
0.21
0.37
0.03
0.16
12.7
4.77
12.66
2.57
5.5
1.13
0.54
5.63
4.55
1.47
0
0
0.77
0
6.48
0.04
7.74
0.07
13.08
3.64
0.78
0.71
0
0
1.26
0.04
6.14
0.38
7.2
2.22
0.94
0.39
Bin58
Bin76
Bin58
Bin76
SAG 733-F2
Sulfu. 1
Sulfu. 2
Sulfu. 1
Sulfu. 2

Figure 3


A
Carbohydrates
Type IVa pilus
Biofilm
poly-GlcNAc
Pga
Pst ABC
As(III)
As(V)
ArsC
MFS
RDD
Na+, Li+, K+
Na+
cation
H+
lipopolysaccharide
lipoprotein
mla ABC
pqiC
MacAB
ABC
lipids
antibiotics/toxins
Co2+, Zn2+, Cd2+
multidrug efflux
TAM
BAM
heavy metal
K+
Na+
Mg2+
cation
cation
Mn2+
F-
K+
GDP
GTP
FeoA
FeoB
Fe2+
Fep
Siderophore
Fe3+
ExbD
ExbD
TonB
Fe3+
TBDR
Siderophore
Fe3+/Vit B12
divalent cation
Mg2+
amino acids
HCO3-
AEC
multidrug
multidrug
ions
MATE
nrsS
heme
oligo-dipeptide / nickel
urea
metal
Mn2+, Zn2+
MoO4 2-
heme iron
ActP
acetate
Na+
MQO
Cpo
H2O2
H2O
I
II
III
IV
NADH
NAD+
fumarate
succinate
H+
O2
H+
H2O
ADP
ATP
H+
H2
2H+
Hyd
Fdx
fumarate
succinate
Aer
Flagellum
chemotaxis
CO2 + H2O
CA
HCO3- + H+
iron storage
Dpsl
2Fe2+ + H2O2 + 2H2O
2Fe(O)OH + 4H+
SOD
2O2- + 2H+
H2O2 + O2
SOR
2O2- + 2H+ + e-
H2O2
tpx,trx
H2O2
H2O
TA
acs
succinate
fumarate
malate
TCA
oxaloacetate
citrate
isocitrate
2-oxoglutarate
CO2
Succinyl-coA
CO2
acetyl-coA
acetyl-P
pta
acetate
ack
pdh
glucose
pyruvate
pps
por
PEP
glucose
succinate
fumarate
malate
rTCA
oxaloacetate
citrate
isocitrate
2-oxoglutarate
CO2
Succinyl-coA
CO2
acetyl-coA
PO4 3-
As(V)
Shared by both Lake Pavin strains (1 and 2)
Carbohydrate-selective porins
relaxase
coupling protein
IME
relaxosome
MinE
MinD
MinC
Min system
SulP
MFS
ChrA
Carbohydrates
Sugar and others
Mg2+
Cr2+
NH4+
NifD
N2
TA
Restriction enzyme
TA
Prophage
DptF
Dynamins
PIT
P
β-lactamase
H+
K+
KUP
APC
amino acids
amino acids
ABC
Fusaric acids
CheW/X/D
MCP
FlgO
RND
lipoprotein
Strain 2 (Bin76)
Strain 1 (Bin58/SAG 733-F2)
Shared with only one Sulfurimonas species
K+
ktrB
ktrA
K+
K+
K+
YeeE 1
YeeE 2
C4-C
C4-C
DctP
HlyD
TolC
TBDT
OmpA
TA
Vsr
CT
GG
GG
CC
CH3
Dcm
MarR
GTPase
TerB
HsdM
HsdS
HyaB
HyaA
H2
2H+
CheW
S2O3 2-
S2O3 2-

B

Figure 4


A
OM
Periplasm
Cytoplasm
IM
OprD
S0
S0/Sn 2-
?
SoxCDYZH
S2O3 2-
SO4 2-
Cyt c ox
Cyt c red
FccB
H2S
S0
H2
2H+
HynB
HynA
HynC
Sn 2-
HS-
PsrA
PsrB
PsrC
MK
MKH2
Sqr (II, IV,VI)
dsrE
RHD
S2O3 2-
SO3 2-
YeeE
TauE
HCO3-/Na+
SulP
SulP
P
H+
PiT
CysD
Sat
CysK, CysE, CysI
SO4 2-
APS
Cysteine
NDP
NTP
Ppk2
Ppk1
ATP
ADP
Polyphosphate
Ppx?
H2O

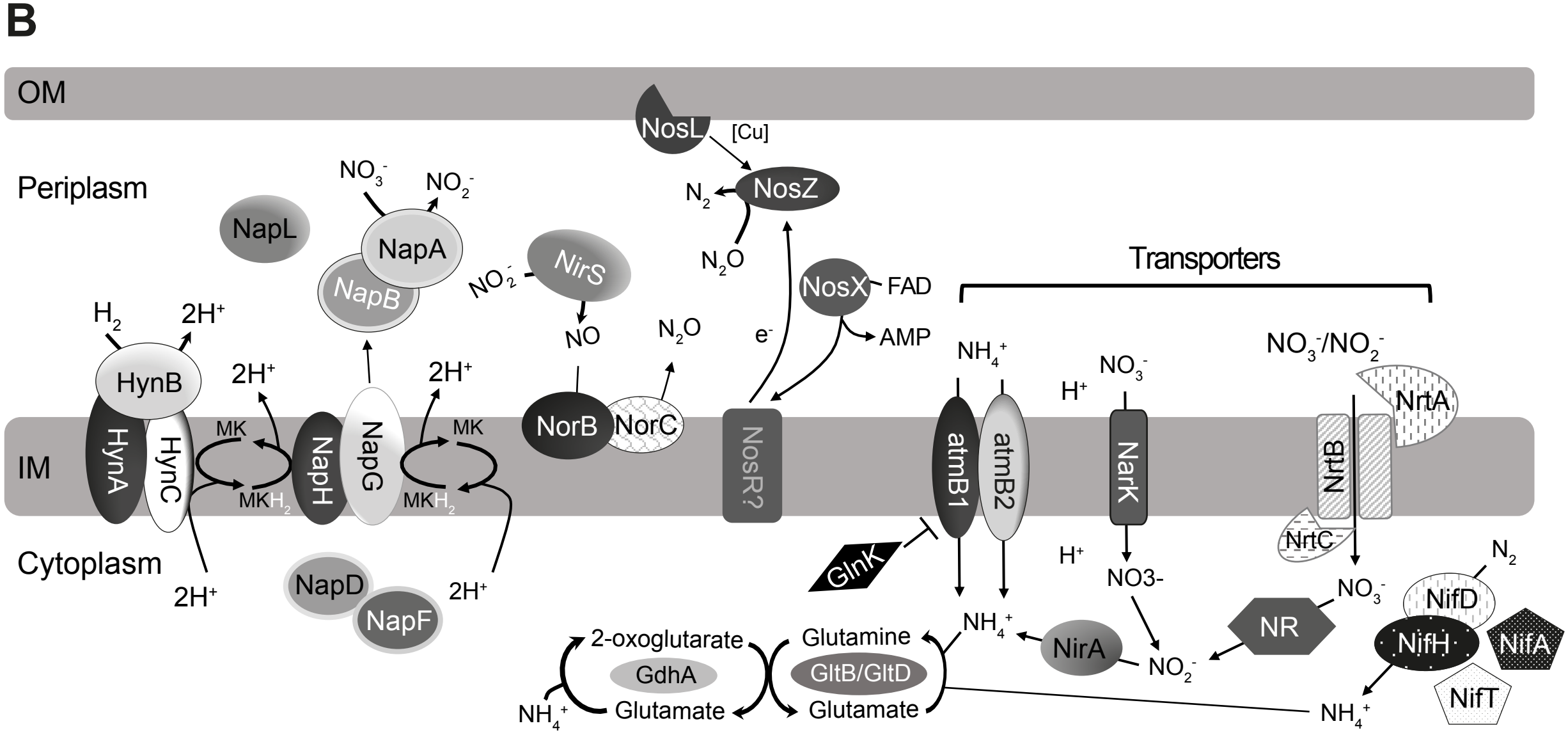
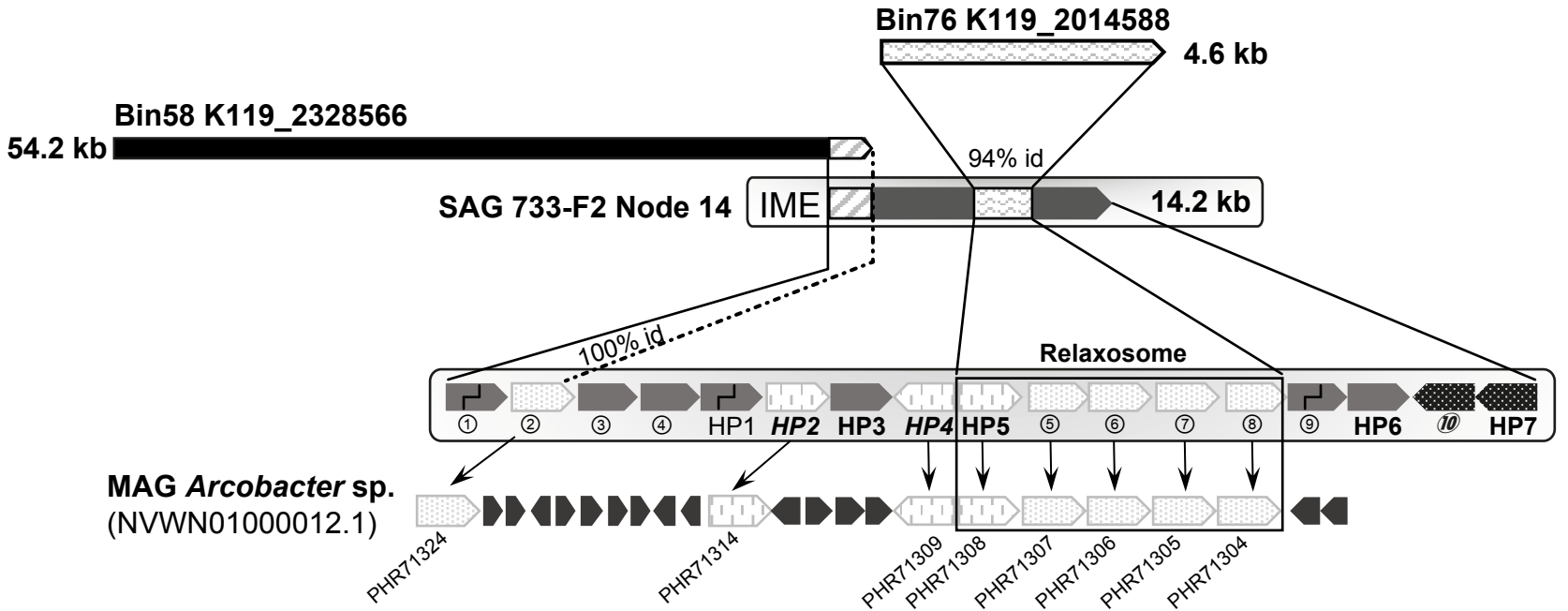

B
OM
Periplasm
Cytoplasm
IM
NosL
[Cu]
N2
NosZ
N2O
NosX
FAD
AMP
e-
NO3-
NO2-
NapL
NapA
NapB
NirS
NO
N2O
NorB
NorC
NosR?
Transporters
NH4+
atmB1
atmB2
H+
NO3-
NarK
NO3-/NO2-
NrtA
NrtB
NrtC
GlnK
NapH
NapG
NapD
NapF
2H+
MK
MKH2
2-oxoglutarate
GdhA
Glutamate
Glutamine
GltB/GltD
NirA
NO2-
NR
N2
NifD
NifH
NifA
NifT

Figure 5



① PAS domain containing protein (frameshift)

② Site-specific integrase

③ DNA-binding protein

④ **Cysteine hydrolase**

Relaxosome

⑤ **MobC family plasmid mobilization relaxosome protein**

⑥ **Relaxase/mobilization nuclease domain-containing protein**

⑦ **Type IV secretory system conjugative DNA transfer family protein**

⑧ **DnaJ domain-containing protein**

⑨ Recombinase (frameshift)

⑩ PIN domain-containing protein

HP: Hypothetical protein

Frameshift

Figure 6

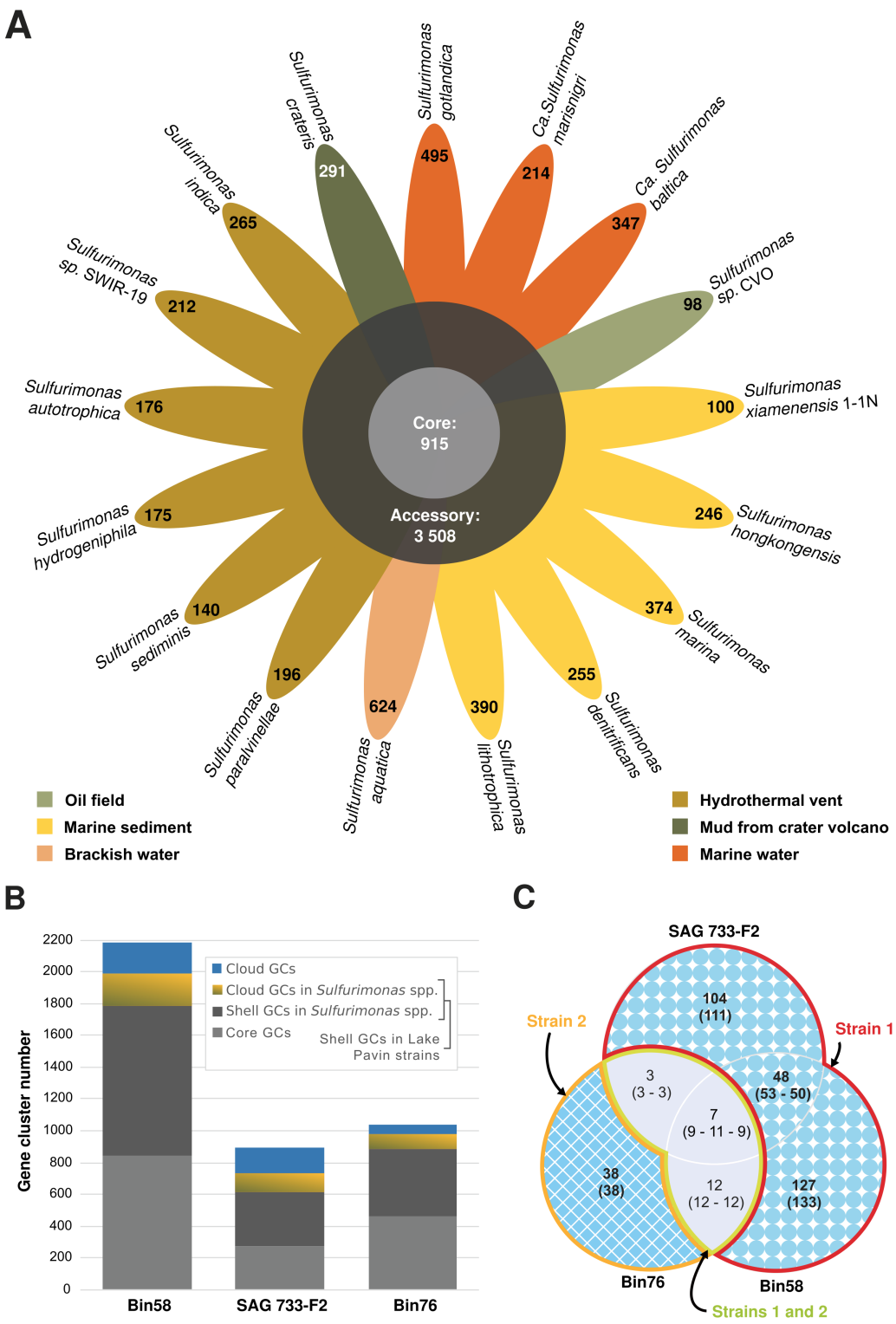

A
Sulfurimonas gotlandica
Ca.Sulfurimonas marisnigri
Ca. Sulfurimonas baltica
Sulfurimonas sp. CVO
Sulfurimonas xiamenensis 1-1N
Sulfurimonas hongkongensis
Sulfurimonas marina
Sulfurimonas denitrificans
Sulfurimonas lithotrophica
Sulfurimonas aquatica
Sulfurimonas paralvinellae
Sulfurimonas sediminis
Sulfurimonas hydrogeniphila
Sulfurimonas autotrophica
Sulfurimonas sp. SWIR-19
Sulfurimonas indica
Sulfurimonas crateris
495
214
347
98
100
246
374
255
390
624
196
140
175
176
212
265
291
Core: 915
Accessory: 3 508
Oil field
Marine sediment
Brackish water
Hydrothermal vent
Mud from crater volcano
Marine water
B
Gene cluster number
2200
2000
1800
1600
1400
1200
1000
800
600
400
200
0
Cloud GCs
Cloud GCs in Sulfurimonas spp.
Shell GCs in Sulfurimonas spp.
Core GCs
Shell GCs in Lake Pavin strains
Bin58
SAG 733-F2
Bin76
C
SAG 733-F2
Strain 2
Strain 1
104 (111)
3 (3 - 3)
48 (53 - 50)
7 (9 - 11 - 9)
38 (38)
12 (12 - 12)
127 (133)
Bin76
Bin58
Strains 1 and 2

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		
	General function prediction only	R	7	1	2

Figure 8

Table 1

Family	Toxin	Antitoxin	Cell process affected	Stage blocked	*S. paralvinellae*	*S. autotrophica*	*S. indica*	*S. hydrogeniphila*	*S. sediminis* ‡	*S.* sp_SWIR19‡	*S. aquatica*	*S. gotlandica*	*S. Candidatus baltica*	*S. Candidatus marisnigri*	*S. marina*	*S. denitrificans*	*S. hongkongensis*	*S. litotrophica*	*S. xiamenensis* 1-1N*	*S. xiamenensis* CVO*	*S. crateris*	SAG 733-F2	Bin58	Bin76
Loci nber¥					1 + 1	10	9	18	22	19	5 + 5	3	25	11	1	13	13	9	7	7	28	8	9	12
Type II	RelE	HTH	Translation	RNAse ribosome-dependent targeting mRNA	●				●	●						★								
		RHH							■				●				●				★	●	●	●
		Phd/YefM					♦	★	★	●			★			●	●				●		●	●
		CopG						●			●							●						
		XRE						●	●	●	●		●					●						
		HigA								●				●										
		PHD																						
		CcdA																						
		AM						●					★	●			●							
		DUF											●			●								
		HP				★		★	★	●	●		■	★		■	♦	●		★	⁂	★	●	★
		Orphan			○									○				○						
	YafQ	RelB/DinJ								●												★	●	★
		CopG						●																
		PHD						●	●	●								●			●			
		HP															●							
		Orphan															○							
	Txe/ YoeB	PHD						●																
		Phd/YefM							●	●								●				●	●	●
		HP													●									
	HigB	TR				●																		
		HigA																						
		DNAb						●															●	
		HTH																	●		●			
		HP																			●			
	BrnT	BrnA				●		●	●							●								
		CopG				●			●	●						●	★							
		HTH														●								
		HP							●	●	●										●			
	PemK/ MazF	CopG		RNAse ribosome-independent targeting mRNA, rRNA and tmRNA (hybrid m- and tRNA)		●													●	●			●	
		Phd/YefM				●			●															
		MazE				●		●																
		DNAb							●															
		HP					●			●							●		●	●				
		Orphan					○		○															
	VapC	CopG																	●				●	●
		Phd/YefM																						●
	PIN	CopG			●																			
		HTH								●														
		HP											●	●							●	●		
	HicA	HicB															●							
	HipA	HTH		Elongation						●	★/●		●								★			●
		XRE											●											
		RR											●											
		HP									●		●	●										
	GNAT	RR											●	●	●									
	NTase	HEPN		Elongation?		●		●																●
		DUF86					♦	★	■	■			●	●			●	●			❖			
		HP					●	●	●	★				●										
	Fic	HP	Replication	DNA Replication	●																			
	Zeta	HP	Cell wal synthesis	Peptidoglycan				●	●												●			
	YaiI/ YqxD	GIY-YIG	nd	nd							●	●	●	●					●	●	●			
		Orphan																						
	MMP	HTH	nd	nd								●		●							●			
		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: Addiction 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
¥: black: chromosomal location, pink: plasmid location

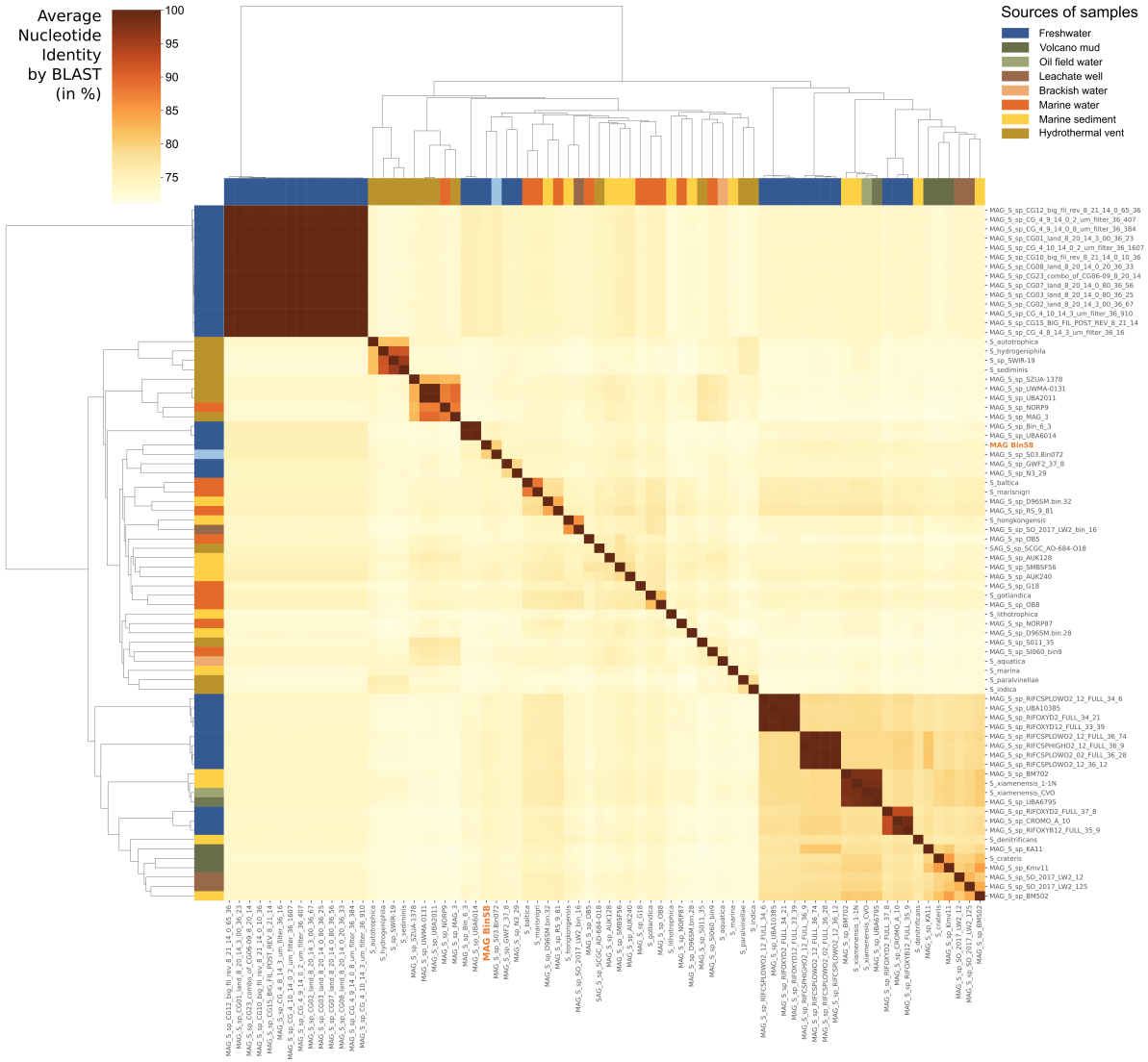

Average Nucleotide Identity by BLAST (in %)
100
95
90
85
80
75
Sources of samples
Freshwater
Volcano mud
Oil field water
Leachate well
Brackish water
Marine water
Marine sediment
Hydrothermal vent
MAG Bin58

Figure S1

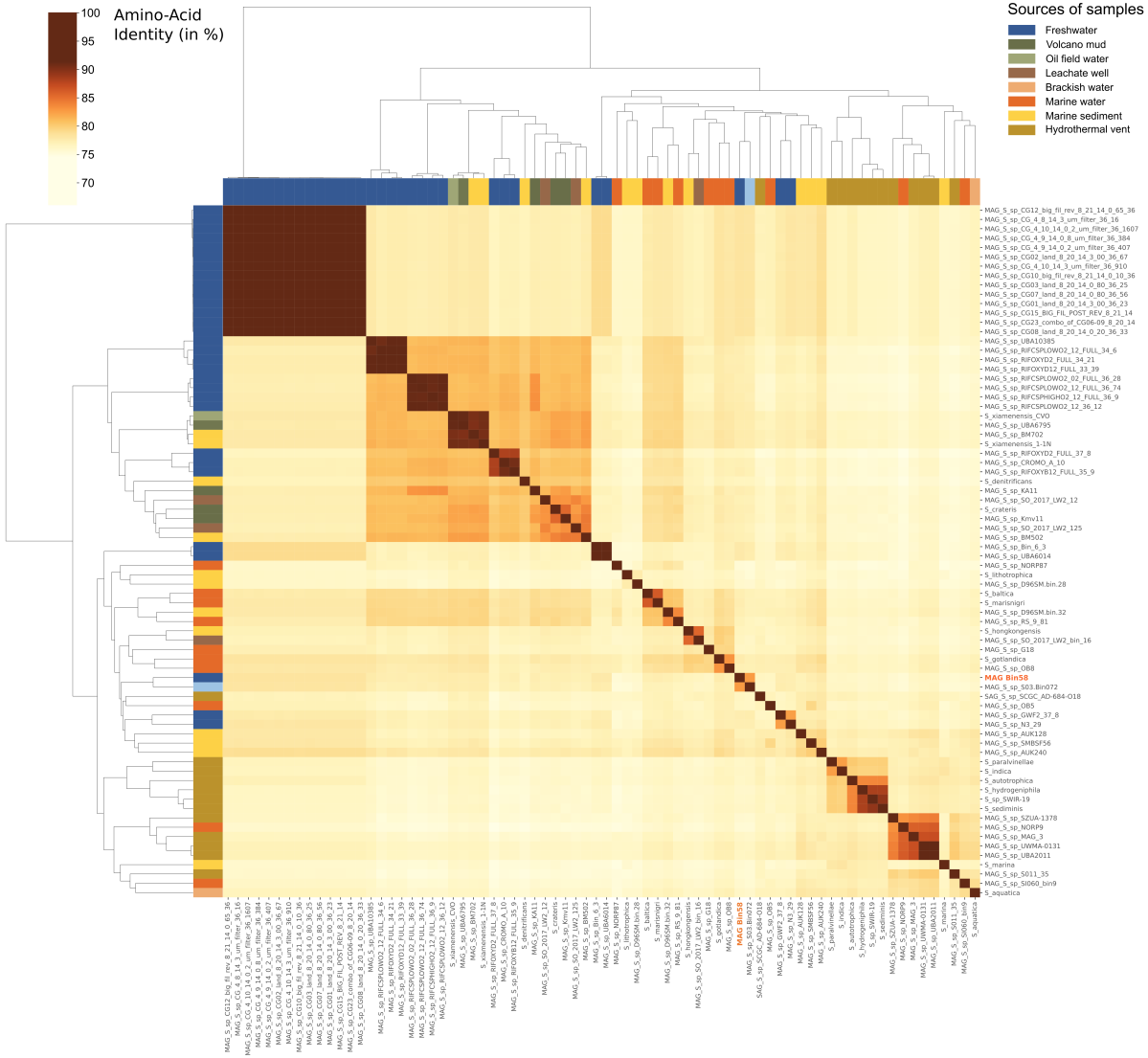

Amino-Acid Identity (in %)
100
95
90
85
80
75
70
Sources of samples
Freshwater
Volcano mud
Oil field water
Leachate well
Brackish water
Marine water
Marine sediment
Hydrothermal vent
MAG Bin58
MAG Bin39

Figure S2

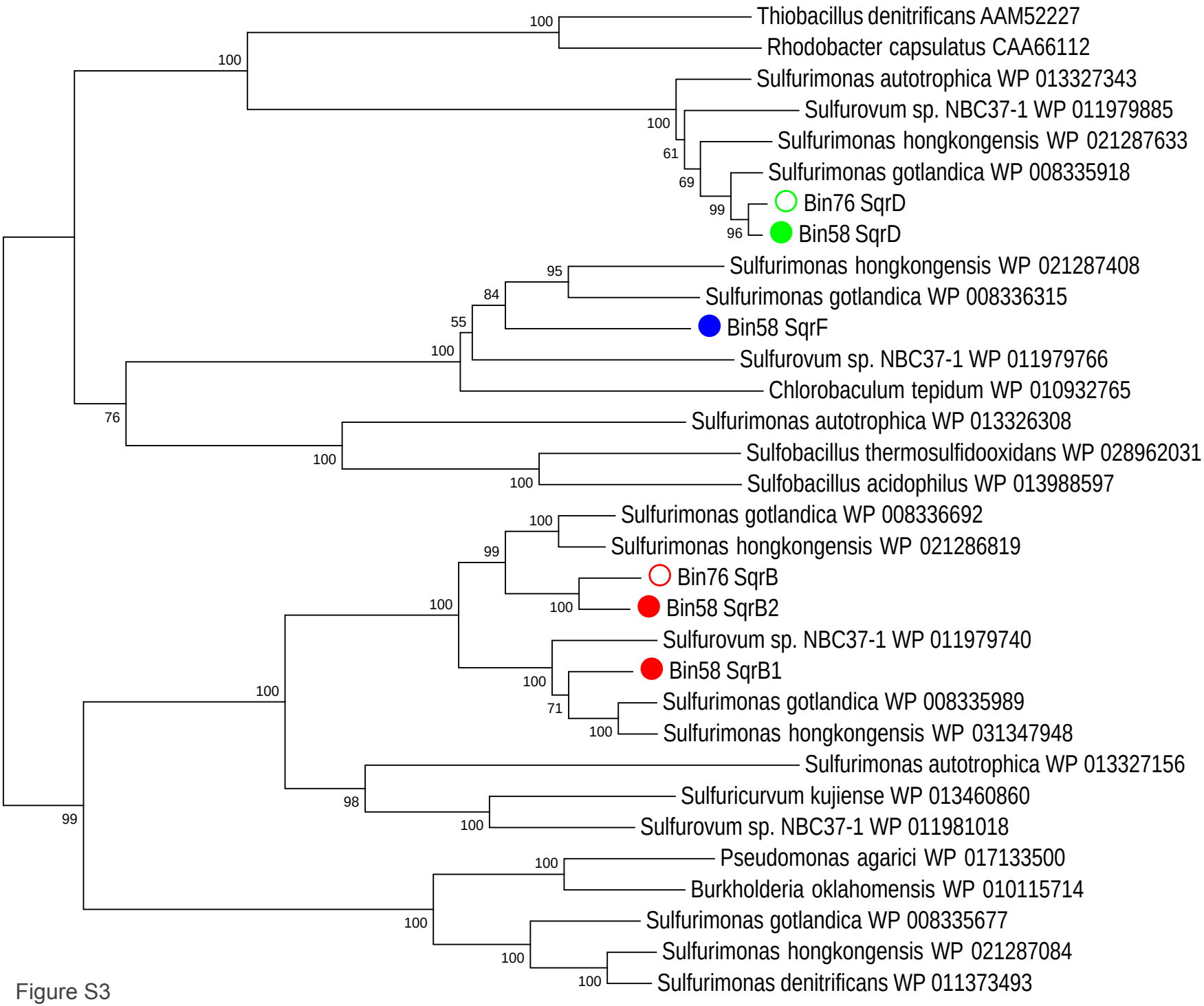
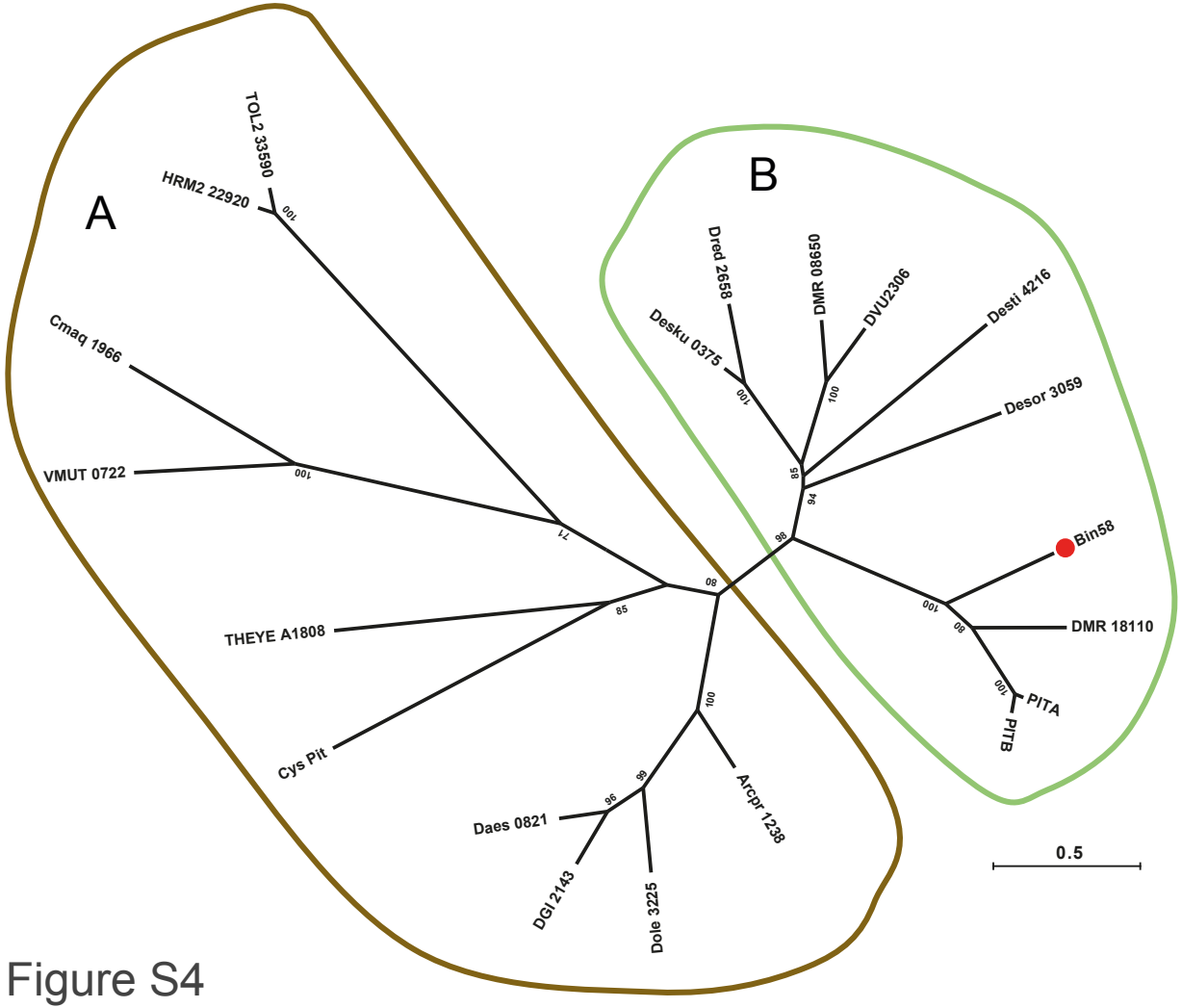

Thiobacillus denitrificans AAM52227
Rhodobacter capsulatus CAA66112
Sulfurimonas autotrophica WP 013327343
Sulfurovum sp. NBC37-1 WP 011979885
Sulfurimonas hongkongensis WP 021287633
Sulfurimonas gotlandica WP 008335918
Bin76 SqrD
Bin58 SqrD
Sulfurimonas hongkongensis WP 021287408
Sulfurimonas gotlandica WP 008336315
Bin58 SqrF
Sulfurovum sp. NBC37-1 WP 011979766
Chlorobaculum tepidum WP 010932765
Sulfurimonas autotrophica WP 013326308
Sulfobacillus thermosulfidooxidans WP 028962031
Sulfobacillus acidophilus WP 013988597
Sulfurimonas gotlandica WP 008336692
Sulfurimonas hongkongensis WP 021286819
Bin76 SqrB
Bin58 SqrB2
Sulfurovum sp. NBC37-1 WP 011979740
Bin58 SqrB1
Sulfurimonas gotlandica WP 008335989
Sulfurimonas hongkongensis WP 031347948
Sulfurimonas autotrophica WP 013327156
Sulfuricurvum kujiense WP 013460860
Sulfurovum sp. NBC37-1 WP 011981018
Pseudomonas agarici WP 017133500
Burkholderia oklahomensis WP 010115714
Sulfurimonas gotlandica WP 008335677
Sulfurimonas hongkongensis WP 021287084
Sulfurimonas denitrificans WP 011373493
100
100
100
100
61
69
99
96
95
84
55
100
76
100
100
100
99
100
100
71
100
100
98
100
99
100
100
100
100
100
0.2

Figure S3


A
B
TOL2 33590
HRM2 22920
Cmaq 1966
VMUT 0722
THEYE A1808
Cys Pit
Daes 0821
DGI2143
Dole 3225
Arcpr 1238
Dred 2658
Desku 0375
DMR 08650
DVU2306
Desti 4216
Desor 3059
Bin58
DMR 18110
PITA
PITB
0.5

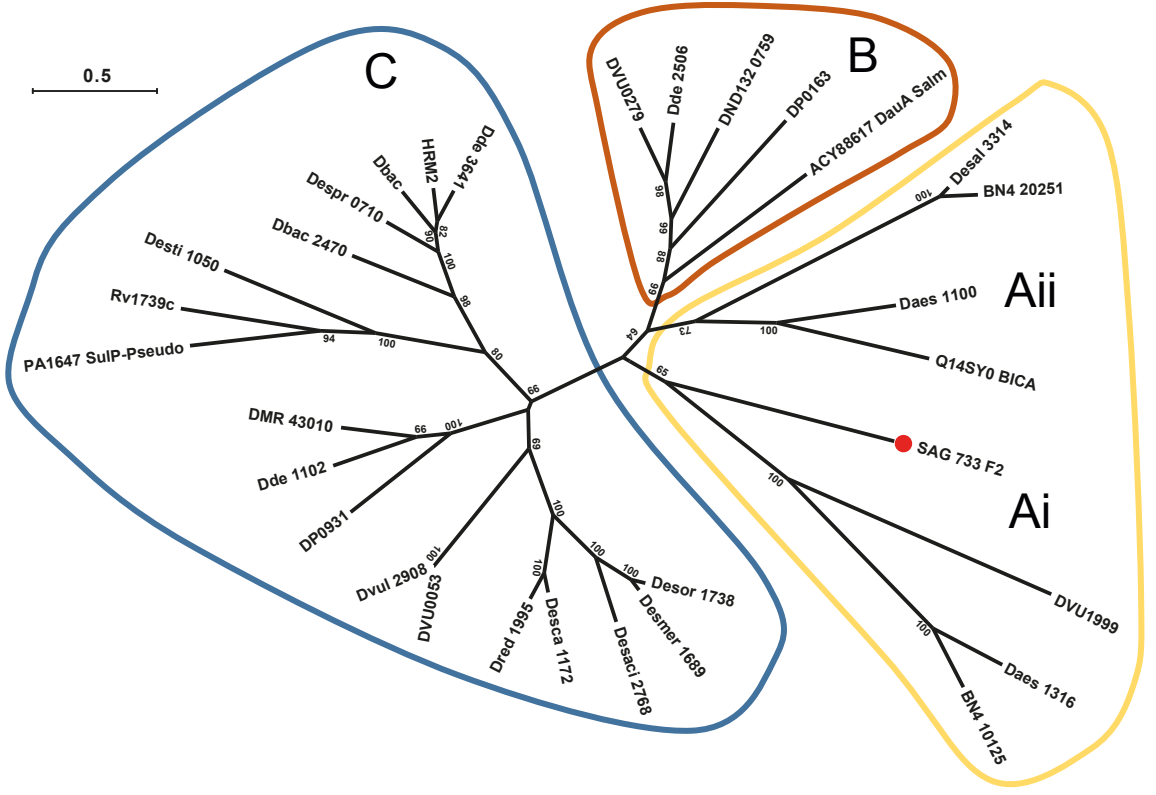
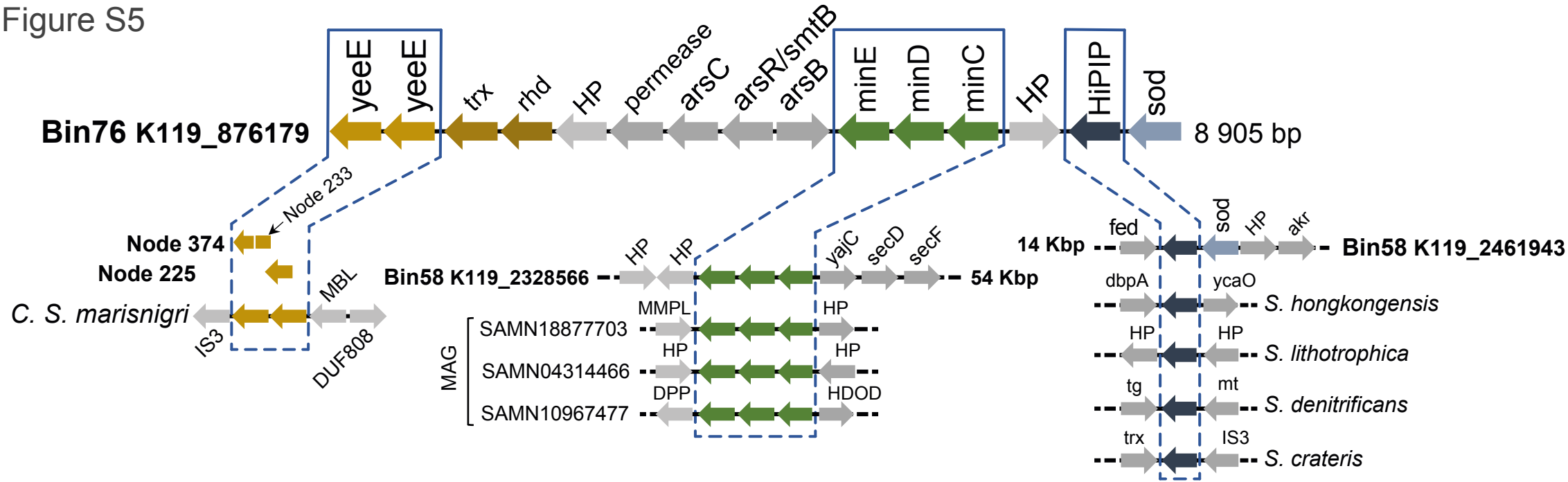

0.5
C
B
Aii
Ai
Dbac
HRM2
Dde 3641
Despr 0710
Dbac 2470
Desti 1050
Rv1739c
PA1647 SulP-Pseudo
DMR 43010
Dde 1102
DP0931
Dvul 2908
DVU0053
Dred 1995
Desca 1172
Desaci 2768
Desmer 1689
Desor 1738
DVU0279
Dde 2506
DND132 0759
DP0163
ACY88617 DauA Salm
Desal 3314
BN4 20251
Daes 1100
Q14SY0 BICA
SAG 733 F2
DVU1999
Daes 1316
BN4 10125

Figure S4

Figure S5



akr: aldo-keto reductase, fed: ferredoxin, ars: arsenate reductase genes, dbpA: decorin-binding protein A, DPD: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, rdh: 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.

Figure S6


A
N-terminus
TK90 27 TPHRCATPFYLGAVMASMDVDVY DsrE3B
Atc_2345 24 TAHRCATPFYLGALLSSMDAEV
LCX93_RS03235 36 IVYQCDFPDIQRIHLMLNTINNA DsrEd
ENK67_03485 36 VVYQCDFADPKRVHLMLNTLNNA
LCX93_03195 21 VVFDCSSGNASYIKSRMWLVGKT DsrEf
SUN_RS08545 20 AVFDCSSSDADYIKSRMWLVGKT
AS592_RS10230 24 AVFDCSSDNADYIKSRMWLVGKT
FJR48_RS04845 21 AVFDCAAGDMKFVLTRMKLIDMT DsrEg
JXK04_04865 23 AVFDCAAKDMKFVASRMFLIEES
LCX_RS03235 71 INVVALGPCLQYMMKDFKGTG DsrEd
ENK67_03485 71 IDIVALGPCLQYVMKDFKDTG
LCX93_03195 57 VALTLHGSCVAMVANDYDMIV DsrEf
SUN_RS08545 56 FVITLHGGCVSMVSKEYKEFV
AS592_RS10230 60 FAITLHGGCVSMVSKEYEMIV
FJR48_RS04845 56 FVLTIHSKCTPIVNKKSDNKD DsrEg
JXK04_04865 58 FVLTIHSGCTPIAVEEEGDDE
Ahos : Acidianus hospitalis W1, Alvin : Allochromatium vinosum DSM 180, AS592 : Sulfurovum riftiae strain 1812E, Atc : Acidithiobacillus caldus SM-1, b3344 : Escherichia coli str. K-12, B649 : Candidatus Sulfuricurvum sp. RIFRC-1, ENK67 : Flavobacteriia bacterium isolate HyVt-598, FHQ18 : Deferribacter autotrophicus, I6M62 : Acinetobacter courvalinii, IMZ28 : Sulfurovum indicum strain ST-419, JXK04 : Campylobacterales bacterium isolate ZodW, Lferr : Acidithiobacillus ferrooxidans ATCC 53993, Mcup : Metallosphaera cuprina, M0P43 : Arcobacteraceae bacterium isolate SO_2017_CLC, Ppha : Pelodictyon phaeoclathratiforme strain DSM 5477, PSEBR : Pseudomonas brassicacearum, 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.
✱ : active cysteine residue common to all DsrE-like proteins
C-terminus
Bin58 58 NDLVEMLKKMYASGIKLQACGTCNARCGIFK
Bin76 58 YDLVEMLKKMYASGINLQACGTCNARCGIFK
FHQ18_10630 61 QDLVQMLKELYHSGVELKVCGTCQARCGVNK DsrEa
LCX93_RS10685 58 YDLAEMLQTMYKDGISLQVCGTCNARCGIFK
YH65_RS07735 58 YDLLEMLKKMYTRGINLQVCGTCNARCGLFK
B649_02020 58 HDLVDMLKKLYEEGISLQACGTCNARCGLFK
GJV85_RS05995 56 SDLQERVKVVQATGVKFNACVVCTDDYRVSK DsrEb
IMZ28_RS04185 52 EALCETVKKVQKSGVTFSCCIVCSDDYGVTE
SULKU_RS10025 52 AELQTMVTKVLETGVKASVCVVCSDDYGVTE
LCX93_RS07285 84 KETAVRVDALMQYDVEFVACGNTMRTLHIKE DsrEc
BM227_RS06525 83 KKIAVRVRSLQTYDVEFIACGNTMRTLKIKP
LCX93_RS03235 109 VSRFKGLKQLGGDNIHFFACGNTMKKKHVKD DsrEd
ENK67_03485 109 ASRIKGLKQLGGDNIKMFACNNTMDKKNVKP
LCX93_RS10690 92 ELQKRIASMADTYDVEFLMCKASMPKNHLEA DsrEe
IMZ28_RS08640 93 ILAKRIDSLIKTYDVEFLMCGAAMVKNKLQA
AS592_RS11895 94 DLAKRIASMAKNYDVEFLMCGAAMPRNKLEA
LCX93_03195 87 QSQQYLKKLAKRKNVTITACAMSLAHNAIEK DsrEf
SUN_RS08545 86 KAQEHLKTLATRKKVKVIVCAMSLDANAIER
AS592_RS10230 90 KAQEYLRVLAEKRGVKVIVCAMSLASNAIEQ
FJR48_RS04845 79 AIHQKLTQLKREHNVKIEACKIATNKFGYKK DsrEg
JXK04_04865 82 GISQRLKKLHAEYGVKIEACEIALERFGIEK
FJR48_RS09400 84 TSIKDMLSSLVKKGGAVVVCGMNAKFQEVKA DsrEh
M0P43_03615 85 KSIRDMLISLSKNGGTVMLCGMCASFEEITT
KGZ62_05410 112 ATIPELRELCQEAEVKFIACQMTVELFGFEH DsrE2A
Alvin_2601 100 ASIEELRELCQEAEVKMIACQMTVDLFDMPK
Lferr_2186 97 ASIEELRELCVEAEVRMIACQMTVDLFEFDT
Atc_2353 97 ASLEELRELCVEAEVRMIACQMTVDLFEFDT
Ppha_2326 111 ATIEQLRGMCQEFGVRFIACQMTMEVFGFEK
Ahos_1698 71 TWDQLIMQAKEVGEVKVYACSTTMEFFGLKR DsrE2B
Mcup_0682 80 MWHQLVKEAKDVGEVKVYACSTTMEFFGIKK
Saci_0337 74 NPFIRFFDMAKENGVKMYVCVQSLKDMCHMN DsrE3A
Mcup_0681 74 NPFIHFFDMAKDNNVKMYVCVQSLKDMCHMN
Ahos_1699 68 NPFIHFFDMARDNGVKMYVCIQSLKDMCHMN
TK90_0639 75 KLIKDFIQDAKRAGVRLHVCQPALPGYRIDA DsrE3B
Atc_2345 72 KKVIEFIRDAKNAGVEIHVCRPALPGYEIPA
Saci_0332 67 DFTGLFKKAKKSGKVKIYACSYASKLFNYSK DsrE4
Mcup_1724 66 NVFEHFKKAKKSGKLKILACSYASKIHGLDK
Mcup_1706 62 KYYLDNLFELAGEDLEITACEFGMRVKDVHE DsrE5
I6M62_RS02065 53 DDQRNLKQEWQKLSIRLPVCVSAALARGITD DsrE/TusD
PSEBR_a3552 58 DLPQQWRTFVSDNQLDGVVCIAAALRRGVLN
Alvin_1253 59 HIVNRWAELAEQYELDMVVCVAAAQRRGIVD
b3344 59 DLVRAWQQLNAQHGVALNICVAAALRRGVVD
✱

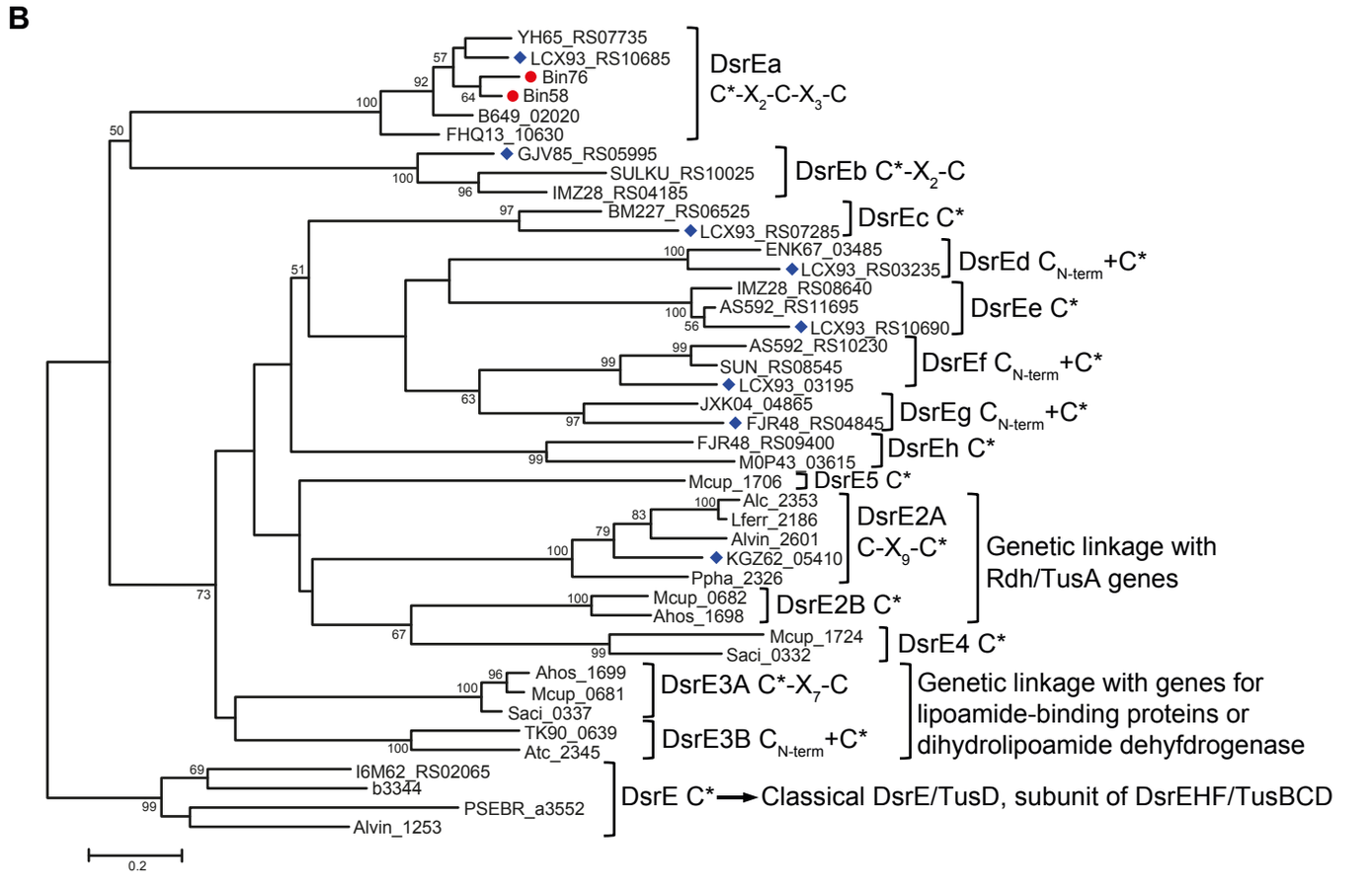

B
YH65_RS07735
LCX93_RS10685
Bin76
Bin58
B649_02020
FHQ13_10630
DsrEa
C*-X2-C-X3-C
GJV85_RS05995
SULKU_RS10025
IMZ28_RS04185
DsrEb C*-X2-C
BM227_RS06525
LCX93_RS07285
DsrEc C*
ENK67_03485
LCX93_RS03235
DsrEd CN-term+C*
IMZ28_RS08640
AS592_RS11695
LCX93_RS10690
DsrEe C*
AS592_RS10230
SUN_RS08545
LCX93_03195
DsrEf CN-term+C*
JXK04_04865
FJR48_RS04845
DsrEg CN-term+C*
FJR48_RS09400
M0P43_03615
DsrEh C*
Mcup_1706
DsrE5 C*
Alc_2353
Lferr_2186
Alvin_2601
KGZ62_05410
Ppha_2326
DsrE2A
C-X9-C*
Genetic linkage with
Rdh/TusA genes
Mcup_0682
Ahos_1698
DsrE2B C*
Mcup_1724
Saci_0332
DsrE4 C*
Ahos_1699
Mcup_0681
Saci_0337
DsrE3A C*-X7-C
TK90_0639
Atc_2345
DsrE3B CN-term+C*
Genetic linkage with genes for
lipoamide-binding proteins or
dihydrolipoamide dehyfdrogenase
I6M62_RS02065
b3344
PSEBR_a3552
Alvin_1253
DsrE C*→Classical DsrE/TusD, subunit of DsrEHF/TusBCD
0.2

Figure S7

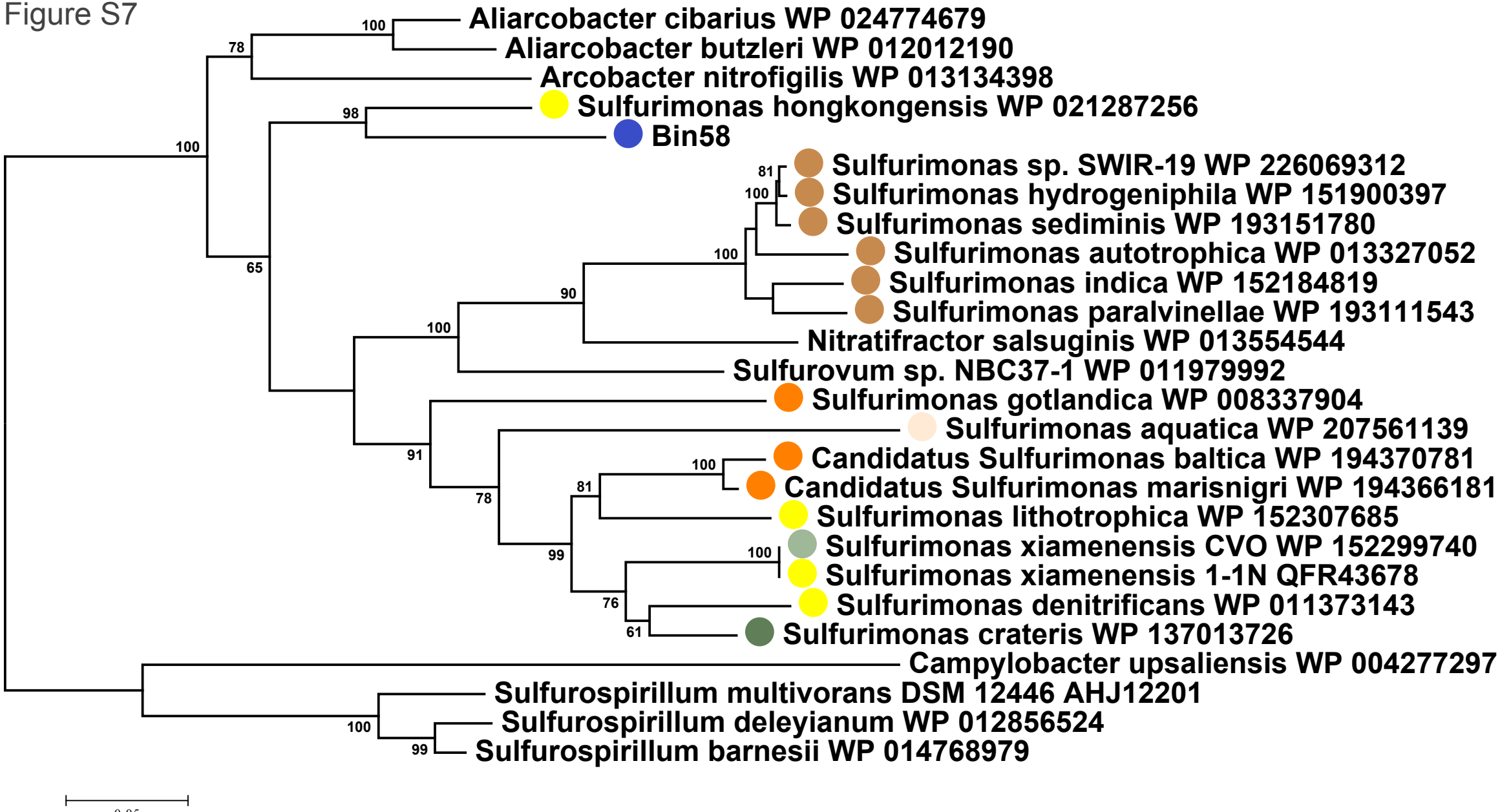

Aliarcobacter cibarius WP 024774679
Aliarcobacter butzleri WP 012012190
Arcobacter nitrofigilis WP 013134398
Sulfurimonas hongkongensis WP 021287256
Bin58
Sulfurimonas sp. SWIR-19 WP 226069312
Sulfurimonas hydrogeniphila WP 151900397
Sulfurimonas sediminis WP 193151780
Sulfurimonas autotrophica WP 013327052
Sulfurimonas indica WP 152184819
Sulfurimonas paralvinellae WP 193111543
Nitratifractor salsuginis WP 013554544
Sulfurovum sp. NBC37-1 WP 011979992
Sulfurimonas gotlandica WP 008337904
Sulfurimonas aquatica WP 207561139
Candidatus Sulfurimonas baltica WP 194370781
Candidatus Sulfurimonas marisnigri WP 194366181
Sulfurimonas lithotrophica WP 152307685
Sulfurimonas xiamenensis CVO WP 152299740
Sulfurimonas xiamenensis 1-1N QFR43678
Sulfurimonas denitrificans WP 011373143
Sulfurimonas crateris WP 137013726
Campylobacter upsaliensis WP 004277297
Sulfurospirillum multivorans DSM 12446 AHJ12201
Sulfurospirillum deleyianum WP 012856524
Sulfurospirillum barnesii WP 014768979
100
78
98
100
65
81
100
100
90
100
91
78
100
81
99
100
76
61
100
99
0.05

Figure S8

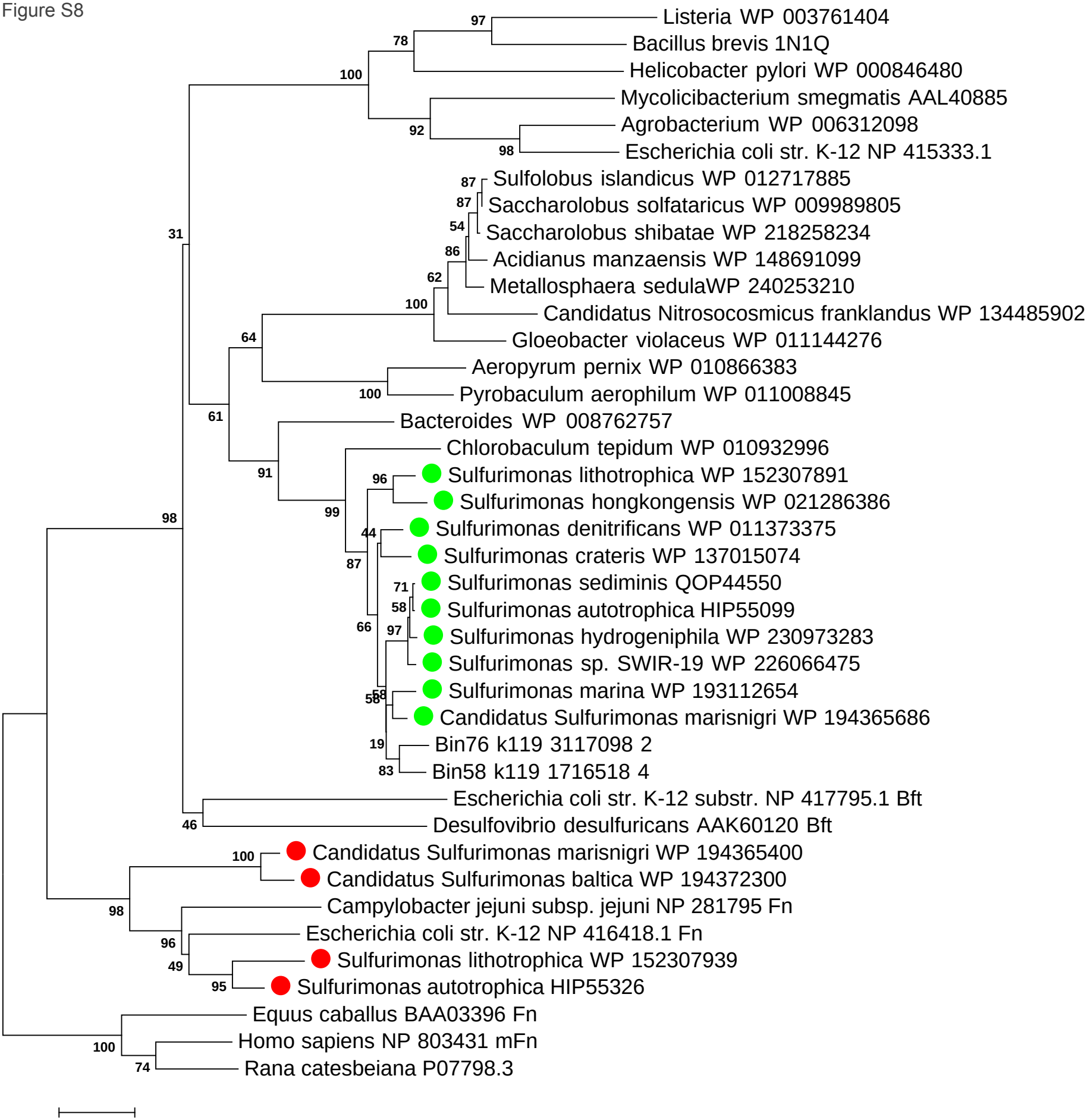
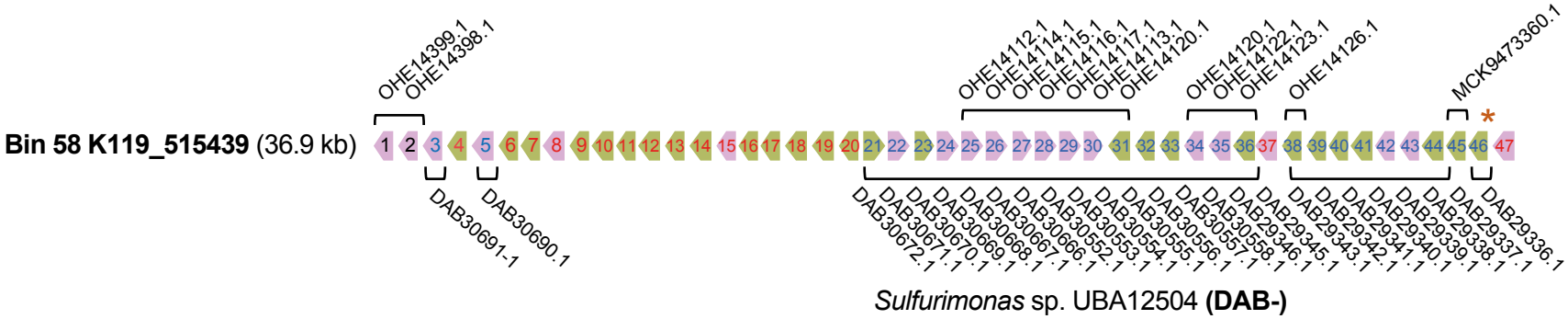

Listeria WP 003761404
Bacillus brevis 1N1Q
Helicobacter pylori WP 000846480
Mycolicibacterium smegmatis AAL40885
Agrobacterium WP 006312098
Escherichia coli str. K-12 NP 415333.1
Sulfolobus islandicus WP 012717885
Saccharolobus solfataricus WP 009989805
Saccharolobus shibatae WP 218258234
Acidianus manzaensis WP 148691099
MetallosphaeraWP 240253210
Candidatus Nitrosocosmicus franklandus WP 134485902
Gloeobacter violaceus WP 011144276
Aeropyrum pernix WP 010866383
Pyrobaculum aerophilum WP 011008845
Bacteroides WP 008762757
Chlorobaculum tepidum WP 010932996
Sulfurimonas lithotrophica WP 152307891
Sulfurimonas hongkongensis WP 021286386
Sulfurimonas denitrificans WP 011373375
Sulfurimonas crateris WP 137015074
Sulfurimonas sediminis QOP44550
Sulfurimonas autotrophica HIP55099
Sulfurimonas hydrogeniphila WP 230973283
Sulfurimonas sp. SWIR-19 WP 226066475
Sulfurimonas marina WP 193112654
Candidatus Sulfurimonas marisnigri WP 194365686
Bin76 k119 3117098 2
Bin58 k119 1716518 4
Escherichia coli str. K-12 substr. NP 417795.1 Bft
Desulfovibrio desulfuricans AAK60120 Bft
Candidatus Sulfurimonas marisnigri WP 194365400
Candidatus Sulfurimonas baltica WP 194372300
Campylobacter jejuni subsp. jejuni NP 281795 Fn
Escherichia coli str. K-12 NP 416418.1 Fn
Sulfurimonas lithotrophica WP 152307939
Sulfurimonas autotrophica HIP55326
Equus caballus BAA03396 Fn
Homo sapiens NP 803431 mFn
Rana catesbeiana P07798.3
0.2

Figure S9



Hypothetical proteins

Known proteins

* Only *Sulfurimonas* MAG sequences in NR database (no *Sulfurimonas* species or other taxa sequences)

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 capside 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