nature microbiology

Article

https://doi.org/10.1038/s41564-023-01342-w

A hydrogenotrophic *Sulfurimonas* is globally abundant in deep-sea oxygen-saturated hydrothermal plumes

Received: 22 February 2022

Accepted: 10 February 2023

Published online: 9 March 2023

Check for updates

Massimiliano Molari[1,2] ✉, Christiane Hassenrueck[3], Rafael Laso-Pérez[1,4,7], Gunter Wegener[1,4], Pierre Offre[5], Stefano Scilipoti[1,6] & Antje Boetius[1,2,4]

Members of the bacterial genus *Sulfurimonas* (phylum Campylobacterota) dominate microbial communities in marine redoxclines and are important for sulfur and nitrogen cycling. Here we used metagenomics and metabolic analyses to characterize a *Sulfurimonas* from the Gakkel Ridge in the Central Arctic Ocean and Southwest Indian Ridge, showing that this species is ubiquitous in non-buoyant hydrothermal plumes at Mid Ocean Ridges across the global ocean. One *Sulfurimonas* species, ^{U}*Sulfurimonas pluma*, was found to be globally abundant and active in cold (<0–4 °C), oxygen-saturated and hydrogen-rich hydrothermal plumes. Compared with other *Sulfurimonas* species, ^{U}*S. pluma* has a reduced genome (>17%) and genomic signatures of an aerobic chemolithotrophic metabolism using hydrogen as an energy source, including acquisition of A2-type oxidase and loss of nitrate and nitrite reductases. The dominance and unique niche of ^{U}*S. pluma* in hydrothermal plumes suggest an unappreciated biogeochemical role for *Sulfurimonas* in the deep ocean.

The genus *Sulfurimonas* belongs to the phylum Campylobacterota (former class Epsilonproteobacteria). It was originally proposed after the isolation of *Sulfurimonas autotrophica* from sediments collected at a deep-sea hydrothermal vent[1]. Since then, 12 distinct *Sulfurimonas* species have been isolated from oxygen-deficient environments[2–11]. Bassed on 16S rRNA gene sequences, this mesophilic and chemolithoautotrophic bacterial genus is ubiquitous and a dominant member of microbial communities inhabiting redoxcline environments[12], including sulfidic environments at deep-sea hydrothermal vents[13–17]. The described members of the genus *Sulfurimonas* occupy habitats defined by moderate temperatures, elevated hydrogen sulfide concentrations and low oxygen concentrations (<40 µM) compared with the habitats of other hydrothermal Campylobacterota members (that is, *Sulfuruvum*[16]) and marine sulfur oxidizers (that is, SUP05[18,19]). Yet, abundant *Sulfurimonas* 16S rRNA gene sequences have also been reported in non-buoyant stage of hydrothermal plumes[14,20–24]. Hydrothermal plumes occur wherever hot anoxic hydrothermal fluids emitted from the seabed mix with cold oxygenated seawater. They can rise to hundreds of metres off the seafloor and disperse thousands of kilometres away from their source[25]. At the non-buoyant stage, hydrothermal plumes consist mostly of cold and oxygen-saturated seawater with highly dilute admixtures of hydrothermal fluid (<0.01%)[25,26]. For this reason, non-buoyant hydrothermal plumes have not been considered a permanent niche and habitat for *Sulfurimonas*. The repeated detection of *Sulfurimonas* sequences in such plumes was explained by passive transport from seafloor and subseafloor environments[26]. However, no study has directly tested whether non-buoyant plumes provide a suitable environment for growth of specific members of

[1]Max Planck Institute for Marine Microbiology, Bremen, Germany. [2]Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany. [3]Leibniz Institute for Baltic Sea Research Warnemünde (IOW), Rostock, Germany. [4]MARUM Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany. [5]Department of Marine Microbiology and Biogeochemistry, NIOZ, Royal Netherlands Institute for Sea Research, Den Burg, the Netherlands. [6]Center for Electromicrobiology, Department of Biology, Aarhus University, Aarhus, Denmark. [7]Present address: Systems Biology Department, Centro Nacional de Biotecnología (CNB-CSIC), Madrid, Spain. ✉e-mail: mamolari@mpi-bremen.de

https://doi.org/10.1038/s41564-023-01342-w

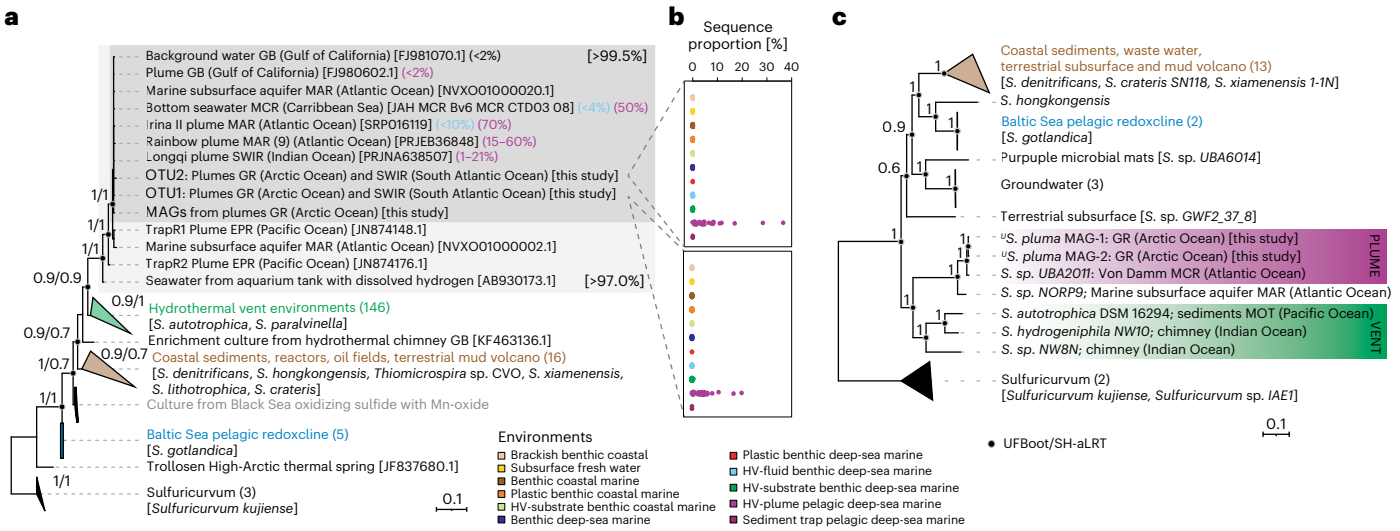


Fig. 1 | Phylogeny and environmental distribution of hydrothermal plume *Sulfurimonas* sp. a, Rooted phylogenetic tree of 16S rRNA gene sequences of *Sulfurimonas* species (*S.*) and closest relatives, including isolates and environmental sequences, with *Sulfuricurvum* as outgroup. The integer numbers and the percentage in parentheses indicate the number of sequences in a given branch and the contribution of *Sulfurimonas* sequences to the total number of sequences in Illumina amplicon sequencing datasets, respectively. In square brackets, the percentage of 16S rRNA gene identity is reported for the plume *Sulfurimonas* cluster. **b**, The two plots show the contribution of the hydrothermal plume *Sulfurimonas* ecotypes (see Extended Data Fig. 3 for details) to the total number of sequences. **c**, Rooted outgroup phylogenetic tree based on concatenated SCG = 258 of *Sulfurimonas* and *Sulfuricurvum* using partition substitution models. Hydrothermal vent (HV) environments include: chimney, sediments, fluids and animal body/nest. The scale bar represents the expected number of changes per nucleotide position. UFBoot and SH-aLRT values are based on 1,000 replicates. Best substitution model for 16S rRNA gene tree: TVMe+I + G4. GB: Guaymas Basin; MAR: Mid Atlantic Ridge; MCR: Mid Cayman Rise; GR: Gakkel Ridge; SWIR: Southwest Indian Ridge; EPR: East Pacific Rise.

Sulfurimonas. The hydrothermal plumes contain substantial amounts of inorganic reduced gases (H_2S, CH_4 and H_2) and metals (Fe, Mn, Cu, Zn and Co)[27], which have considerable impact on ocean chemistry[28]. Thus, the identification and elucidation of the physiology of microorganisms growing in the plume are crucial to understanding the ocean's biogeochemistry.

In this study, we investigated the distribution and function of *Sulfurimonas* in the hydrothermal plumes. We studied its ribotypes, genotypes and metabolism in two vent plumes of Gakkel Ridge and in one plume of the Southwest Indian Ridge (SWIR), and compared these with publicly available data from other vent plumes of Mid Ocean Ridges and other environments hosting *Sulfurimonas* sp. Our hypothesis is that non-buoyant hydrothermal plumes are a suitable environment for specific members of *Sulfurimonas*.

Results

An uncharacterized *Sulfurimonas* in hydrothermal plumes

We investigated microbial community compositions in deep-sea water samples from the Gakkel Ridge and the Southwest Indian Ridge (SWIR). Campylobacterota accounted for 70–79% and 9–19% of all 16S rRNA sequences of non-buoyant hydrothermal plumes of Aurora (3,360–3,575 m depth) and Polaris (2,574–2,846 m depth) mounds[29], respectively, and up to 16% of total microbial cells (Extended Data Fig. 1a). Almost all Campylobacterota-affiliated 16S rRNA sequences (>99%) in the non-buoyant hydrothermal plumes of Gakkel Ridge and in seawater from a ridge valley of the SWIR belonged to the genus *Sulfurimonas* (Supplementary Table 1 and Extended Data Fig. 2). In addition, more than 97% of the *Sulfurimonas* sequences of these three remote sites on ultraslow spreading ridges belonged to two closely related operational taxonomic units (OTU1 and OTU2), with a similarity of 99.5%. Fluorescence in situ hybridization using both a Campylobacterota-specific rRNA probe and tailored highly specific probes for the two detected *Sulfurimonas* OTUs confirmed these results (Extended Data Fig. 1b–f).

On the basis of environmental DNA retrieved from hydrothermal plume samples from Gakkel Ridge, we obtained two near-complete (93.6 – 99.95% completeness) and high-quality draft[30] *Sulfurimonas* metagenome-assembled genomes (MAG-1 and MAG-2; Supplementary Table 2). These two MAGs have an average nucleotide identity (ANI) of 98.9%, confirming that the Gakkel Ridge plumes host two closely related strains of the same *Sulfurimonas* species. The 16S rRNA gene sequences from the MAGs are 99.5–100% identical with the dominant OTUs of the 16S rRNA gene amplicons, pointing toward the prevalence of the same *Sulfurimonas* strains in the Gakkel Ridge and the SWIR water samples, which are more than 15,000 km apart and in different deep-water current systems.

The *Sulfurimonas* species that is most closely related to the hydrothermal plume ribotype and genotype is *Sulfurimonas autotrophica*, with 94% 16S rRNA gene identity and an ANI of only 74.2%. *S. autotrophica* has been isolated from Pacific Ocean hydrothermal sediments[1]. Yet, phylogenetic analyses based on 16S rRNA genes assigned the plume-hosted *Sulfurimonas* sequences to an independent clade which included neither *S. autotrophica* nor *Sulfurimonas* sequences from other hydrothermal vent environments (Fig. 1a). Instead, our *Sulfurimonas* phylotypes form a well-supported clade together with sequences derived from hydrothermal plumes in the Atlantic Ocean (Mid Atlantic Ridge, Mid Cayman Ridge), in the Gulf of California (Guaymas Basin), in the Pacific Ocean (East Pacific Rise) and from an oxic subsurface aquifer (Mid Atlantic Ridge). This phylogenetic placement is supported by phylogenomic trees based on concatenated single-copy genes (SCGs) (Fig. 1c). The comparative analysis of *Sulfurimonas* genomes identified 7,569 clusters of orthologous genes. About 13% of these represent the core genome of this genus (Extended Data Fig. 4a). Similar results were recently reported in a more comprehensive pangenomic survey of the genus *Sulfurimonas*[8]. Altogether, these findings support the occurrence of a previously uncharacterized *Sulfurimonas* taxon in non-buoyant hydrothermal plumes. According to the standards for microbial

Table 1 | Genomic comparison of ^{U}*S. pluma* with isolated *Sulfurimonas* strains

	Sd	Sh	Sg	Sa	Sp1	(Putative) characteristics
Genome size (Mbp)	2.20	2.30	2.95	2.15	1.77	The Sp1 genome is 17% to 40% reduced compared with other *Sulfurimonas* strains.
Hydrogen oxidation						
(NiFe)-hydrogenase	+	+	+	+	+	Sp1 encodes only the oxygen-sensitive type (group 1b).
Dissimilatory sulfur oxidation						
Flavocytochrome c sulfide dehydrogenase	+	+	+	–	+	In Sp1 putative primary enzyme of sulfide oxidation pathway.
Sulfide:quinone reductase	+	+	+	+	(+)	Canonical enzyme for sulfide oxidation in *Sulfurimonas*. Sp1 contains only type VI (missing types I–V).
Dissimilatory nitrogen reduction						
Nitrate reductase	+	+	+	+	–	Niche-specific enzymes for *Sulfurimonas*.
Nitrite reductase	+	+	+	+	–	
Nitric oxide reductase	+	+	+	+	–	
Carbon fixation						
Oxoglutarate:ferredoxin oxidoreductase	+	+	+	+	+	Sp1 encodes a five-subunit version of these enzymes compared with typical four-subunit enzyme, potentially resulting in a higher oxygen tolerance of the enzyme.
Pyruvate:ferredoxin oxidoreductase	+	+	+	+	+	
Oxygen reduction						
cbb3-type cytochrome c oxidase	+	+	+	+	+	Canonical enzyme for *Sulfurimonas*. High affinity for oxygen, thus optimized for microaerobic environments.
caa3-type cytochrome c oxidase	–	–	–	–	+	Canonical enzyme for aerobic bacteria.
Oxidative stress						
Superoxide dismutase	+	+	+	+	–	Primary antioxidant system in aerobic organisms.
Periplasmic superoxide reductase	–	–	–	–	+	Putative role in scavenging the periplasmic superoxide.

Sd: *S. denitrificans*; Sh: *S. hongkongensis*; Sg: *S. gotlandica*; Sa: *S. autotrophica*; Sp1: ^{U}*S. pluma* MAG-1. For details, see Supplementary Table 5.

uncultivated species[31,32], we propose to name this uncultivated (U) taxon '^{U}*Sulfurimonas pluma*'.

^{U}*S. pluma* inhabits oxygen-saturated hydrothermal plumes

To identify the environmental factors shaping the habitat of the here described ^{U}*S. pluma* taxon, we compared the hypervariable V3–V4 region of the 16S rRNA genes of *Sulfurimonas* obtained from the Arctic and from SWIR to those found in other environments (Extended Data Fig. 5), and assigned environmental characteristics (Supplementary Table 3). The studied 308 samples contained 1,389 *Sulfurimonas* oligotypes (amplicon sequence variants distinguished on the basis of information entropy determined for each nucleotide position[33]; Supplementary Fig. 1 and Table 4). On the basis of the presence and proportion of oligotypes within and among the environmental categories, they formed 149 ecotypes, 38 of which were specific to hydrothermal plumes (Extended Data Fig. 3). Two of these oligotypes represented between 28% and 97% of the *Sulfurimonas* sequences in the hydrothermal plume samples. These oligotypes were identical to the dominant OTUs in Gakkel and SWIR plumes (OTU1 and OTU2). These results confirm the existence of non-buoyant plume-specific *Sulfurimonas* ecotypes that are very rare in other hydrothermal environments (sequence proportions <1%; Fig. 1b). The ^{U}*S. pluma* cluster of closely related sequences (>99.5% identity) dominated hydrothermal plumes across the ridge systems of the Central Arctic, Atlantic and Indian/Southern Oceans (Fig. 1a). The same ribotype was also found in the plume and the surrounding water column of the Guaymas Basin in the Gulf of California[34], but with low proportions to the total bacterial community (Fig. 1a).

The specificity of ^{U}*S. pluma* to hydrothermal plumes is also supported by recruitment analysis of metagenomic and metatranscriptomic reads from different hydrothermal environments. MAG-1 of ^{U}*S. pluma* recruited preferentially metagenomic reads from hydrothermal plumes rather than from benthic hydrothermal environments (Extended Data Fig. 4b). ^{U}*S. pluma* MAG-1 recruited transcripts from all seawater metatranscriptomes, yet the transcription of more than 60% of all its gene clusters was limited to samples from Gakkel Ridge non-buoyant plumes and from a diffusive fluid sample from Juan de Fuca Ridge (Extended Data Fig. 4b). The recruitment of few metagenomic and metatranscriptomic reads from Guaymas Basin hydrothermal plumes is consistent with previous studies reporting low proportions of *Sulfurimonas* sequences in the plume[34,35] (Fig. 1a). As to the plume samples from Beebe vent at Mid Cayman Rise, the recorded water depth (4,900 m) suggests that water samples were collected in the deepest part of the rising plume, where the microbial community is still mostly influenced by benthic communities, yet comprises few microbes growing within the hydrothermal non-buoyant plume[36,37].

^{U}*S. pluma* lost typical *Sulfurimonas* genes

The ^{U}*S. pluma* MAGs have a size of 1.68–1.77 Mbp, which is considerably smaller than the genomes of other *Sulfurimonas* strains, including *S. gotlandica* GD1 isolated from a pelagic redox environment (Extended Data Fig. 4a). Despite their small size, the MAGs of ^{U}*S. pluma* are as complete as the closed genome of *S. autotrophica*, based both on reference SCGs and machine learning algorithm methods (Supplementary Table 2). The smaller size is partly explained by the high gene density (that is, number of genes per kbp) of ^{U}*S. pluma* MAGs, which exceeds those of all other *Sulfurimonas* genomes (Extended Data Fig. 4a). This suggests partial deletion of non-coding DNA sequences, resulting in a streamlined genome. Nonetheless, the ^{U}*S. pluma* MAGs lack several genes that code for core and apparent niche functions of isolated *Sulfurimonas* strains, such as for sulfide oxidation and nitrate/nitrite reduction[8,12] (Table 1 and Supplementary Table 5). We investigated

https://doi.org/10.1038/s41564-023-01342-w

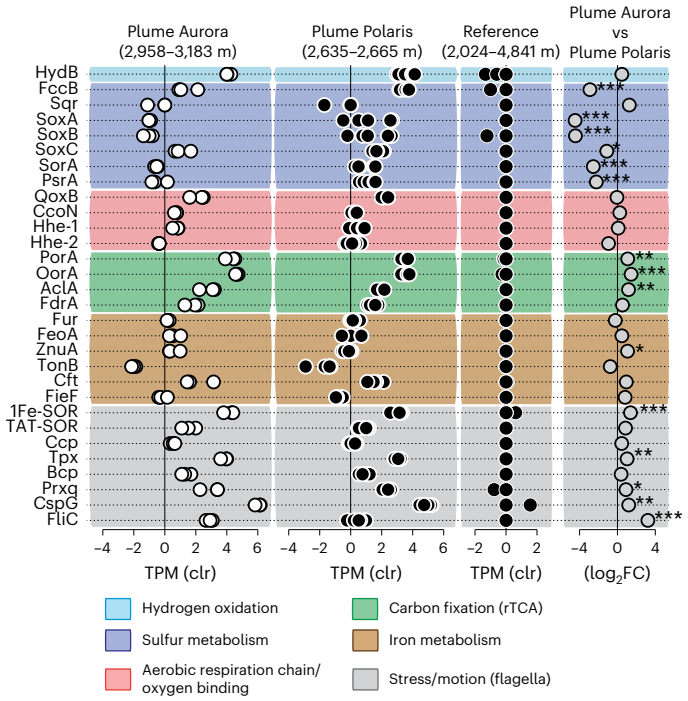


Fig. 2 | Transcriptome of ^{U}*S. pluma*. The expression of marker genes for the main metabolic pathways of *Sulfurimonas* sp. from Gakkel Ridge plumes (Aurora and Polaris) and reference seawater. The gene expression is centred log-ratio transformed (clr). Differential expression of genes between the plumes of Aurora ($n = 3$) and Polaris ($n = 6$) is reported as log2-fold-change ($\log_2$FC). The pairwise statistical test is based on quantile-adjusted conditional maximum-likelihood (qCML) method and the likely value was adjusted by applying FDR. *$P_{adjust} < 0.05$; **$P_{adjust} < 0.01$; ***$P_{adjust} < 0.001$. hydB: (NiFe)-hydrogenase Group 1b, large subunit; fccB: flavocytochrome c sulfide dehydrogenase; sqr: sulfide:quinone reductase; soxA, soxB, soxC: sulfur oxidation proteins; sorA: sulfite dehydrogenase; psrA: polysulfide reductase, subunit I; qoxB: cytochrome c oxidase, caa3-type, subunit I; ccoN: cytochrome c oxidase, cbb3-type, subunit I; hhe-1 and hhe-2: bacteriohemerythrins; porA: pyruvate:ferredoxin oxireductase, subunit I; oorA: oxoglutarate:ferredoxin oxireductase, subunit I; aclA: ATP-dependent citrate lyase, subunit I; fdrA: fumarate reductase, subunit I; fur: iron uptake regulation; feoA: Fe^{2+} uptake; znuA: Mn^{2+}/Zn^{2+} uptake; tonB: siderophore transport; cft: ferritin; 1Fe-SOR and TAT-SOR: superoxide reductases; ccp: cytochrome c peroxidase; tpx, bcp and prxq: peroxiredoxins; cspG: cold shock-like protein; fliC: flagellin; rTCA: reductive tricarboxylic acid cycle.

whether misassembly or binning errors caused the absence of these genes. None of the 16 individual genomic bins that were the basis for the consensus ^{U}*S. pluma* MAGs contained the missing functional genes. Furthermore, contigs obtained from the individual assemblies ($n = 12$) and the co-assembly ($n = 1$) of Gakkel Ridge seawater metagenomes did not contain additional *Sulfurimonas*/Campylobacterota genes that did not bin to the ^{U}*S. pluma* MAGs. The alignment of ^{U}*S. pluma* MAGs with the *S. autotrophica* genome showed that only single genes or operons were missing in conserved genomic regions located in the largest contigs (>40 kpb). These results excluded that the observed genome reduction was an artefact of assembly and binning procedures.

^{U}*S. pluma* is a hydrogen oxidizer

The MAGs of ^{U}*S pluma* encode a membrane-bound (NiFe)-hydrogenase, which is the highest expressed catabolic gene in all metatranscriptomes of both Gakkel Ridge plumes (Fig. 2). (NiFe)-hydrogenases are ubiquitous in *Sulfurimonas* genomes, yet an expression of the encoding genes by *Sulfurimonas* in hydrothermal plumes has not been reported before (Supplementary Note 1). The fact that the hydrogenase was >13 to >500 times higher expressed than genes for sulfur oxidation suggests that hydrogen is a critical energy source to sustain the growth of ^{U}*S. pluma* in the Aurora plume (Fig. 2), where it was most abundant and active (Supplementary Table 1 and Extended Data Fig. 2). Laboratory experiments with cultures of *S. denitrificans* also found that this species grows more efficiently when supplied with hydrogen than with thiosulfate as electron donor[38], suggesting that hydrogen can be an important energy substrate for the genus *Sulfurimonas*.

The genome of ^{U}*S. pluma* contains all genes for the oxidation of zero-valent sulfur (S^0), or thiosulfate as typically described for the genus *Sulfurimonas*, but it misses the genes for sulfide:quinone reductases (*sqr*), a functional-core gene for this genus (Fig. 3 and Supplementary Table 5). Notably, ^{U}*S. pluma* encodes a non-canonical flavocytochrome c sulfide dehydrogenase (*fcc*) that might replace *sqr* in sulfide oxidation (Extended Data Fig. 6 and Supplementary Note 2). At both Gakkel Ridge plumes, the expression of genes for canonical sulfur oxidation was substantially lower compared with that of membrane-bound (NiFe)-hydrogenase (Fig. 2).

In previous studies, reduced manganese and iron species were suggested to be additional potential electron donors for microorganisms in hydrothermal plumes[25]. In the MAGs of ^{U}*S. pluma*, we found no evidence for known genes encoding enzymes for extracellular Fe (II) and Mn (II) oxidation, such as outer-membrane embedded cytochrome c (Cyc2 or Cyt572) and porin–cytochrome c protein complexes[39,40]. However, ^{U}*S. pluma* encodes and expresses different genes for the uptake and control of the intercellular availability of iron and manganese (Fig. 3). Among them, the ^{U}*S. pluma* genomes encode for an iron storage protein such as ferritin, which is absent in the genomes of *S. gotlandica* and *S. autotrophica* (Supplementary Table 5). Ferritin is an intracellular protein that oxidizes excess ferrous iron and can store up to 2,000 non-reactive ferric iron atoms[41]. During this reaction, ferretins also consume cytoplasmic O_2 and remove H_2O_2[42] (Fig. 3). The high transcription of the ferritin gene in the hydrothermal plume metatranscriptomes of the Gakkel Ridge (Fig. 2) suggests a potential role in mitigating oxidative stress and in iron storage. Future studies should focus on the iron cycling pathways of ^{U}*S. pluma* because they may be important elements of the 'microbial Fe pump' between hydrothermal vents and the ocean[26].

Adaptations to high oxygen concentrations

^{U}*S. pluma* differs from cultivated *Sulfurimonas* strains in the lack of genes for nitrate and nitrite reductases and by coding for caa3-type (A2-type) cytochrome c oxidase (Table 1). All other *Sulfurimonas* sp. genomes encode only cbb3-type oxidase[12], which has a high affinity for oxygen and allows growth at microaerobic conditions[42]. Yet, at higher oxygen concentrations (that is, >20%), the cbb3-type oxidase becomes inefficient, resulting in impaired growth[9,12]. In fact, the cultured *Sulfurimonas* strains grow optimally at an O_2 concentration of 1–8%, and become inactive at O_2 concentrations higher than 20%[1–5,9,11]. Moreover, previous studies found *Sulfurimonas* predominantly in environments subject to strong fluctuations in O_2 concentrations (that is, benthic and pelagic redoxclines[12]; Supplementary Table 3). The cold polar waters studied here are oxygen-saturated and the diluted hydrothermal fluids do not substantially lower their oxygen contents. Hence, ^{U}*S. pluma* is permanently exposed to high oxygen concentrations (ca. 300 µM; Supplementary Table 3). We hypothesize that the acquisition of caa3-type (A2-type) cytochrome c oxidase allows an efficient respiration of ^{U}*S. pluma* in this fully oxic environments. This cytochrome c oxidase is present in many aerobic bacteria and it has strong homology to the mitochondrial cytochrome oxidase (A1-type)[43]. Of note, within the phylum of Campylobacterota, we found all four subunits of caa3-type oxidase in the genome of *Sulfurovum* sp. AR derived from aerobic Arctic sediments[44]. This oxidase has an amino acid identity of 70% to that of ^{U}*S. pluma*. However, this caa3-type oxidase cannot be misassembled in the ^{U}*S. pluma* MAGs because *Sulfurovum* sequences are rare in the Gakkel seawater (Supplementary Table 1), and the synteny analysis

https://doi.org/10.1038/s41564-023-01342-w

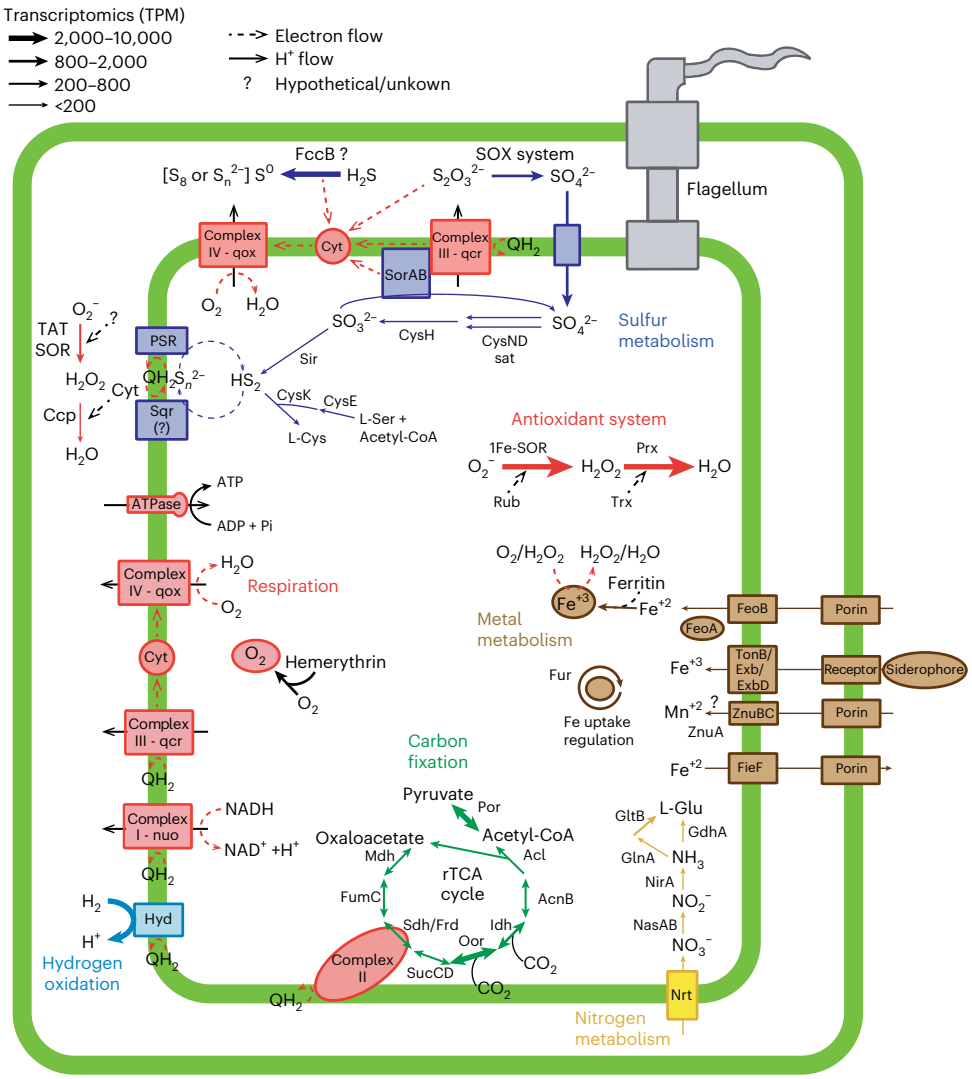


Fig. 3 | Metabolic map of [U]*S. pluma*. Metabolic scheme and gene transcription levels of genes involved in aerobic chemolithoautotrophy of [U]*S. pluma* MAG-1. The average gene expression of Aurora and Polaris plumes is reported as TPM. For enzymes with multiple subunits, the transcription of the catalytic subunit is reported. Steps with more than one arrow indicate that several operons encoding different enzymes catalysing that reaction are present in the genome.

of contigs encoding for this enzyme points toward an acquisition by horizontal gene transfer (Supplementary Fig. 2).

The MAGs of [U]*S. pluma* contain all genes of the reverse tricarboxylic acid (rTCA) cycle (Fig. 2), which is the common autotrophic carbon fixation pathway of all cultured *Sulfurimonas* species[12]. Its key genes 2-oxaglutarate:ferredoxin oxidoreductase (OOR), pyruvate:ferredoxin oxidoreductase (POR) and ATP-dependent citrate lyase (ACL) were highly expressed, suggesting the rTCA cycle as the main pathway for autotrophic growth (Fig. 3). However, carbon fixation via rTCA usually tolerates only trace amounts of oxygen because OOR and POR are oxygen sensitive[45]. Our study suggests that [U]*S. pluma* is capable of using rTCA under fully and permanently oxic conditions. Fixing carbon via the rTCA pathway under these conditions requires specific adaptations of the genes involved as recently reported for a marine nitrite-oxidizing bacterium[46]. Indeed, the gene operon encoding the OOR and POR enzymes in [U]*S. pluma* has five subunits instead of the four subunits present in other *Sulfurimonas* strains (Table 1). The phylogenetic and synteny analysis of this operon showed that the catalytic subunits of OOR (*oorA*) and POR (*porA*) are most closely related with members of the Aquificales and that the operon has the same organization as the members of this bacterial order (Fig. 4 and Extended Data Fig. 7). The Aquificales member *Hydrogenobacter thermophilus* encodes two OOR with either two or five subunits. Under anoxic conditions, it synthesizes the two-subunit form, but when exposed to oxygen, it switches to the five-subunit form. This strategy allows carbon fixation via the rTCA cycle even at 40% oxygen[47,48]. This five-subunit OOR has a lower specific activity compared with two-subunit anaerobic enzymes[47], which probably requires production of large amounts of this oxygen-tolerant form[49]. Accordingly, the [U]*S. pluma* genes encoding for OOR and POR subunits were among the most highly expressed genes in the plumes (Fig. 2).

[U]*S. pluma* contains multiple oxygen-sensitive enzymes, electron carriers and enzymes that form reactive oxygen species, hence it requires solutions to reduce intracellular O_2 concentrations. One of the most effective solutions is aerobic respiration. The cbb3-type and caa3-type are both highly expressed in the plumes (Fig. 2), suggesting that under sustained growth, their activity can be combined to efficiently reduce intracellular O_2 concentrations and/or simultaneously use multiple electron sources for energy conservation.

Highly expressed bacterial hemerythrin-like genes may help the activity of oxidases (Fig. 3). Hemerythrins are non-heme intracellular

https://doi.org/10.1038/s41564-023-01342-w

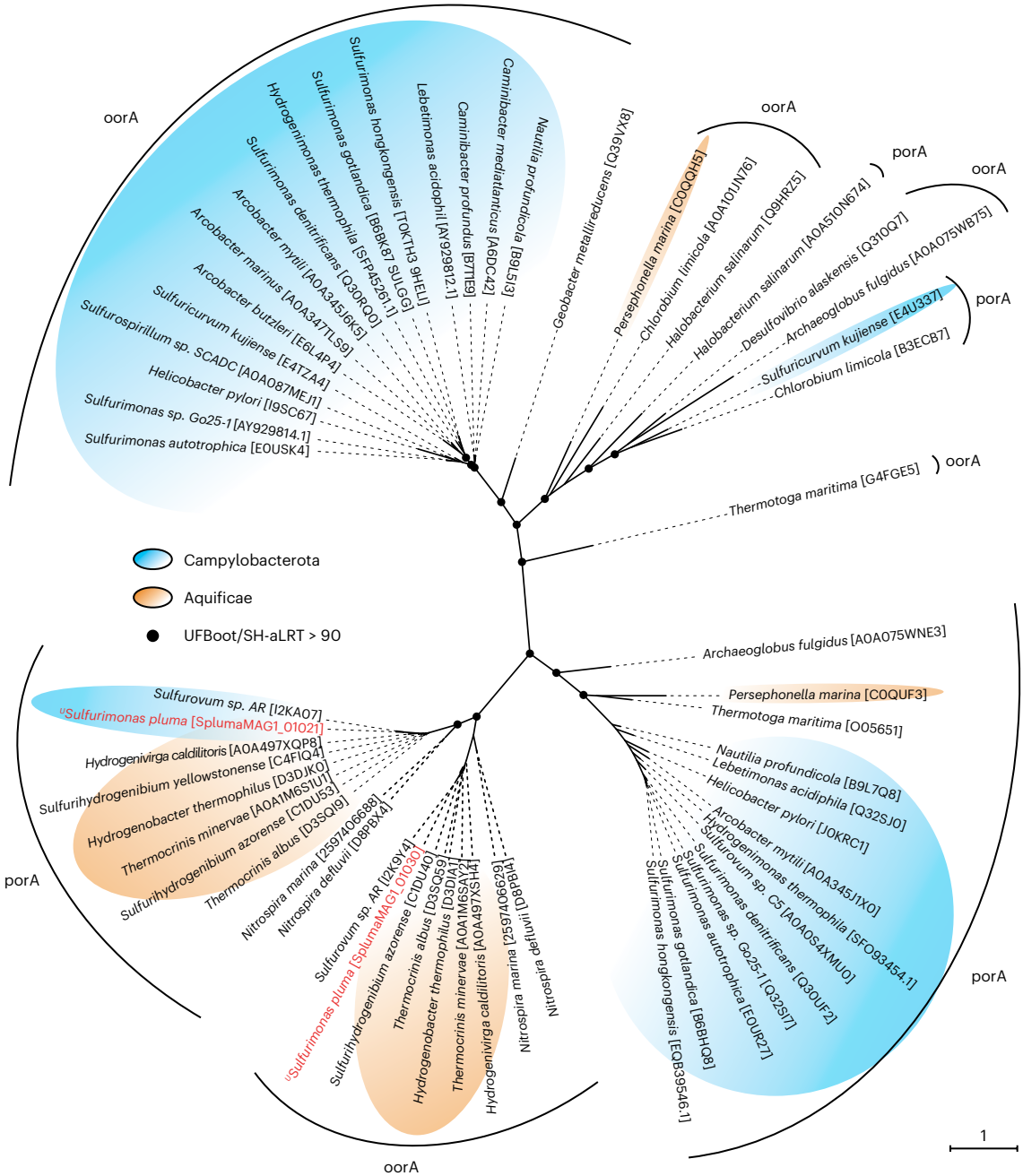


Fig. 4 | Phylogenetic tree of rTCA cycle ferredoxin oxidoreductases. Phylogenetic relationships between the alpha subunit of 2-oxoglutarate: ferredoxin oxidoreductase (oorA) and the alpha subunit of pyruvate:ferredoxin oxidoreductase (porA). The text in red shows oorA and porA genes of ᵁ*S. pluma*. The scale bar represents the expected number of changes per amino acid position. UFBoot and SH-aLRT values are based on 1,000 replicates. Best substitution model: LG + R4.

oxygen-binding proteins. The two hemerythrin-like sequences found in genomes of ᵁ*S. pluma* have one single hemerythrin domain, similar to those found in other *Sulfurimonas* species (Table 1) and other Campylobacterota and Aquificae[50]. Putative functions of bacterial hemerythrins include oxygen storage, oxygen detection or oxygen detoxification[50]. In motile microaerophilic or anaerobic bacteria, it has been proposed that hemerythrins act as an oxygen sensor for aerotaxis[51,52], yet there are evidences for oxygen detoxification in microaerophilic Campylobacterota species[52] and the enhancement of oxygen respiration in methane oxidizers[53]. ᵁ*S. pluma* is probably motile, as indicated by the presence and expression of genes for the flagellum (Fig. 3), thus we cannot exclude a sensing role of its hemerythrins.

A further strategy to reduce intracellular O_2 concentration in hydrothermal plume ᵁ*S. pluma* is the loss of superoxide dismutases (Table 1), an enzyme that catalyses the transformation of superoxide (O_2^-) into molecular oxygen and hydrogen peroxide[54]. This is the primary antioxidant system in aerobic organisms, and it is also present in all cultivated members of *Sulfurimonas* (Table 1). In ᵁ*S. pluma*, the antioxidant system includes superoxide reductase (SOR) and peroxidases (Table 1), which sequentially reduce O_2^- to H_2O_2 and then to H_2O using electrons from other reduced substrates, respectively[54] (Fig. 3). This system contributes to the decrease in intracellular O_2 concentration and also consumes reductants in the cytosol, thereby lowering the formation of harmful reactive oxygen species[54]. In addition to the canonical/cytoplasmic core gene for 1Fe-SOR, ᵁ*S. pluma* showed

Table 2 | Comparison of hydrothermal systems and vent fluid types for non-buoyant hydrothermal plumes hosting chemolithotrophic ^{U}*S. pluma* and SUP05 cluster (family Thioglobaceae)

Ridge system	Vent field	Vent site	Rock hosting system	Fluid type	^{U}*S. pluma*	SUP05	Reference
					%	%	
Mid Atlantic Ridge	Logatchev	Irina I	Ultramafic	Hydrogen-metal-rich	up to 70	<10	22,135
	Rainbow	–	Ultramafic	Hydrogen-metal-rich	up to 60	<10	23,136,137
SouthWest Indian Ridge	Longqi	DFF1, DFF3, DFF5, DFF6	Basalt	Sulfide-metal-rich	up to 21	<5	24,138
Guaymas Basin	–	–	Basalt and organic-rich sediments	Methane-ammonia-rich	<2	30	34,55
Gakkel Ridge	Aurora	–	Basaltic/ultramafic	Hydrogen-metal-rich	up to 79[a]	<10[a]	this study[29,58]
	Polaris	–	Basaltic	Hydrogen-rich	<18[a]	up to 40[a]	this study[59]
Kermadec Arc	Macauley, Brothers	–	Submarine volcano	Sulfide-metal-rich	<4	up to 80	56

[a]Based on 16S rRNA sequences from metatranscriptomes (Extended Data Fig. 2). Percentages refer to the proportion of 16S rRNA gene sequences in sample from non-buoyant hydrothermal plumes.

high transcription of a gene encoding for a 1Fe-SOR with translocation signal peptide (TAT-1Fe-SOR; Fig. 3), which is not present in any other *Sulfurimonas* genomes (Table 1). The presence of a TAT motif suggests a periplasmic location for this enzyme, and therefore a putative role in scavenging the periplasmic superoxide. This is further evidence for a lifestyle adapted to the oxygen-saturated non-buoyant plumes of deep hydrothermal vents and chemically similar aquatic systems.

Discussion

In this comparative study of Arctic Ocean and global deep-sea microbiota, we identified an uncharacterized aerobic *Sulfurimonas* species inhabiting non-buoyant hydrothermal plumes and fluids, and subsurface aquifers. The genome of ^{U}*S. pluma* shows substantial genome reduction (17–40%), including the loss of genus-specific functional genes (that is, for denitrification) and acquisition of genes allowing growth in pelagic oxygen-saturated environments.

Previous studies described SUP05 as a widespread and dominant chemolithoautotroph in hydrothermal plumes, using both sulfur compounds and hydrogen as energy source[25,55]. The ability of ^{U}*S. pluma* and SUP05 to rely on the same substrates for energy to grow and their variable co-occurrence in hydrothermal plumes (Table 2) suggest that they could be competitors, as described for marine pelagic redox gradients[19]. Due to the low amounts of substrates (that is, tens to hundreds of nM) and the lack of a redoxcline in non-buoyant hydrothermal plumes, the mechanisms controlling niche partitioning between *Sulfurimonas* and SUP05 in this environment might be different from those proposed for pelagic redoxclines of the Baltic Sea[19]. Our results suggest hydrogen as an essential energy substrate for the growth of ^{U}*S. pluma* in the studied hydrothermal plumes. This agrees with its dominance in other hydrogen- and metal-rich non-buoyant hydrothermal plumes (Table 2). Of note, '*Candidatus* Sulfurimonas marisnigri' could grow at atmospheric oxygen concentrations, but only when MnO_2 was supplied as a terminal electron acceptor[10]. It remains to be clarified whether metals and other hydrothermal compounds favour ^{U}*S. pluma* growth and contribute to niche differentiation, and how ^{U}*S. pluma* adapts to growth at temperatures close to the freezing point.

The global presence of a hydrogenotrophic *Sulfurimonas* species in transient environments such as non-buoyant hydrothermal plumes opens new paradigms in the microbial ecology of this and other aquatic habitats. So far, it has been postulated that microbes growing in the plume, such as the sulfur-oxidizing Gammaproteobacteria of the SUP05 clade and mixotrophic SAR324 deltaproteobacteria, are derived primarily from ambient seawater[26,56]. These microorganisms are also abundant and active in other marine pelagic environments (for example, surface and deep ocean, oxygen minimum zones), suggesting that their habitat is not exclusively hydrothermal plume but is widespread in the oceans[25]. Our results showed that non-buoyant hydrothermal plumes are a suitable environment for the growth of microorganisms typically inhabiting hydrothermal vents such as *Sulfurimonas*. We suggest that the hydrothermal plume does not act exclusively as a vector for dispersing microorganisms from benthic hydrothermal environments, but it might also support ecological connections between pelagic and seafloor/subsurface habitats. The phylogenetic analysis suggests that the ^{U}*S. pluma* lineage could have been derived from a hydrothermal vent-associated ancestor (probably by sympatric speciation), which acquired higher oxygen tolerance and then spread across the oceans. However, it remains to be further investigated whether ^{U}*S. pluma* originates from vent-associated environments or from background seawater. On one hand, the presence of a very similar ribotype (>99.5% 16S rRNA gene sequence similarity) in hydrothermal plumes across the globe (Fig. 1) suggests that the *Sulfurimonas* cluster, including ^{U}*S. pluma*, is part of the ocean microbial seed bank, and therefore that background seawater might be the source of ^{U}*S. pluma*. On the other hand, it may be that ^{U}*S. pluma* enters into the hydrothermal plumes from populations living on seafloor vent-associated environments, which due to oxygen tolerance have a higher dispersal potential than benthic *Sulfurimonas* species, resulting in higher global connectivity[17]. Future studies on uncultivated *Sulfurimonas* species described here will be needed to verify these hypotheses, and to shed light on environmental and ecological forces that shape the connections and composition of microbial communities between different environments such as subsurface aquifers, diffusive flow and hydrothermal plumes.

Methods

Sample collection

Water samples were collected in the valley of the Southwest Indian Ridge (SWIR segment 10°–17° E; Atlantic sector), and at the Aurora and Polaris vent sites of the Gakkel Ridge (Central Arctic Ocean) during RV *Polarstern* expeditions PS81 (9 November–16 December 2013)[57], PS86 (7 July–3 August 2014)[58] and PS101 (9 September–23 October 2016)[59], respectively. A conductivity-temperature-depth (CTD) rosette equipped with 24 12-l Niskin bottles was used to collect water samples for DNA extraction at all three sites, and for cell counts at Aurora and Polaris sites. Microbial biomass inside and outside the hydrothermal plumes at the Aurora and Polaris sites was collected by filtering in situ a large volume of seawater (ca. 200 l per sample) using in situ pumps

(WTS-LV04; McLane) attached to the CTD-frame and CTD-wire and equipped with polycarbonate filters (142 mm diameter, 0.2 µm pore size; Millipore). The pumps were programmed to operate at maximum pump rate for 90 min. The hydrothermal plume signal was recorded by oxygen, sulfide, turbidity and Koichi-type redox sensors mounted on the CTD, by miniature autonomous plume recorders (MAPR supplied by the PMEL Earth-Ocean observation programme of the National Oceanic and Atmospheric Administration (NOAA)) and by custom-made sensors for temperature, redox, pH and H_2S mounted on an ocean floor observation system. The typical physico-chemical signatures for hydrothermal vent plumes were not found in the water column of the SWIR segment investigated, yet turbidity and redox anomalies were recorded in bottom water[60]. Non-buoyant hydrothermal plumes were identified both at Aurora[58] and Polaris[59] locations. Both plumes contained CH_4, H_2 and H_2S, the major difference being the high abundance of particles (that is, polymetallic sulfides) in the Aurora plume[58] compared with the Polaris plume[59]. At these sites, seawater samples were collected in the hydrothermal plume and in the surrounding water: above the plume, below the plume, bottom water and background water (that is, seawater without physico-chemical signatures for hydrothermal plume). Seawater samples were also collected at reference stations (that is, not affected by hydrothermal plume) located 2 km (inside the ridge; 'internal reference') and 56 km (outside the ridge; 'external reference') away from the Aurora field, and 16 km (inside the ridge; 'internal reference') and 190 km (outside the ridge; 'external reference') away from the Polaris field. The stations are listed in Supplementary Table 6.

Fluorescence in situ hybridization (FISH)

Seawater samples (ca. 500 ml) were fixed with formaldehyde (2% final concentration) for 8 h at 4 °C and filtered onto 0.22 µm polycarbonate filters (47 mm diameter; Millipore). The filters were washed with sterile-filtered seawater and with 70% ethanol in MQ-water, dried and stored at −20 °C until further processing. Catalysed reporter deposition FISH (CARD-FISH) was performed according to ref. [61]. For the hybridization of ribosomal RNA, the oligonucleotide EPSY914 5′–GG TCCCCGTCTATTCCTT–3′ (35% formamide[62]; synthetized by Biomers) was used to target members of the *Campylobacterota* phylum, and the oligonucleotide NON338 5′–ACTCCTACGGGAGGCAGC–3′ (synthetized by Biomers) was used as negative control (35% formamide)[63]. All hybridizations were conducted for 2.5 h at 46 °C. Cells were imaged at ×1,000 magnification with an epifluorescence microscope (Axiophot II Imaging, Zeiss). Cells were counter-stained with 4′,6-diamidino-2-phenylindole (DAPI; Sigma-Aldrich) and for each filter, a minimum of 800–1,000 DAPI-stained cells from 20–30 different fields of view were counted.

Oligonucleotide probes specific for ^{U}*S. pluma* were designed on the basis of sequences retrieved through amplicon sequencing of the hypervariable V3–V4 region of the 16S rRNA gene of samples collected at the Aurora vent site and using the ARB software package[64]. The Probe-Match function of ARB was used to test in silico the coverage and specificity of the probes against the SILVA rRNA reference database (release 132)[65] complemented with 16S rRNA gene sequences from this study. Almost all the *Sulfurimonas* sequences clustered into two groups (within similarity >99%), representing the *Sulfurimonas* OTU1 and OTU2 identified by the analysis of 16S rRNA gene amplicon sequences (described in the section 'Illumina 16S rRNA gene sequencing'). We designed specific probes for OTU1 (SLFM-A484 5′–GCTTATTCATAGG CTACC–3′; 15% formamide) and OTU2 (SLFM-B484 5′–GCTTATTCATAT GCTACC–3′; 20% formamide), both synthetized by Biomers. Due to the high similarity between these two oligonucleotides (one mismatch for G and T), each probe was used in a mix together with the other (non-labelled) probe as competitor oligonucleotide. To check the coverage and specificity of ^{U}*S. pluma*'s probes in the environmental samples, double CARD-FISH hybridizations were carried out using the *Campylobacterota* probe (EPSY914) as a positive control.

DNA and RNA extraction

Seawater samples for 16S rRNA gene analysis were filtered immediately after the retrieval of the Niskin bottles. The filtration was carried out in a temperature-controlled room (2 °C) in the dark and did not exceed 1–1.5 h. Three liters of seawater were filtered onto a 0.22 µm polycarbonate filter (47 mm diameter; Millipore, Merck) during PS81 and PS86 using a vacuum pump (N 022 AN.18; KNF), and 10 liters through a 0.22 µm Sterivex filter (Millipore, Merck) during PS101 using a peristaltic pump (Masterflex; Cole Parmer). All filters were stored and transported at a temperature between −20 and −80 °C. Genomic bacterial DNA was isolated in a combined chemical and mechanical lysis procedure using the PowerWater DNA Isolation kit (MO BIO Laboratories). Before DNA isolation, the Sterivex cartridges of the 0.22 µm membranes were cracked open to place the filters in the kit-supplied bead beating tubes. The isolation was continued according to the manufacturer's instructions, and genomic DNA was stored at −20 °C.

DNA and RNA for metagenomic and metatranscriptomic analyses, respectively, were extracted from the microbial biomass retrieved with the in situ pumps. After the recovery of the pumps, filters were immediately cut into 6 pieces, transferred to a screw-cap tube, frozen in liquid nitrogen for 5 min, and then stored and transported at −80 °C. RNA was extracted from three sections of filters in an RNase-free tube with the mirVana mRNA Isolation kit (Ambion). RNA extracts were treated with the TURBO DNA-free kit (Ambion) to remove co-extracted DNA, and purified and concentrated using the RNeasy MinElute kit (Qiagen). RNA extracted from each filter section was pooled together and stored at −80 °C. The genomic DNA was extracted from three pieces of the filter using the PowerWater DNA Isolation kit (MO BIO Laboratories) following the manufacturer's instructions. The extracted genomic DNA from each filter was pooled and stored at −20 °C.

Illumina 16S rRNA gene sequencing

The hypervariable V3–V4 region of the bacterial 16S rRNA gene was amplified using bacterial primers S-D-Bact-0341-b-S-17 (5′-CCTACGGGNGGCWGCAG-3′) and S-D-Bact-0785-a-A-21 (5′-GACTACHVGGGTATCTAATCC-3′)[66]. Sequences were obtained on the Illumina MiSeq platform in a 2 × 300 bp paired-end run aiming for >50,000 reads per sample (CeBiTec), following the standard instructions of the 16S metagenomic sequencing library preparation protocol (Illumina). The workflow and scripts used in this study for the quality cleaning, merging, clustering and annotation of the sequences can be found in ref. [67]. Briefly, only reverse and forward reads with quality score higher than 20 (applying a sliding window of 4) were merged, clustering of sequences into OTUs was done using the programme swarm (v2.2.2)[68], and the taxonomic classification was based on the SILVA rRNA reference database (release 132)[65].

Metatranscriptomes

RNA quantification, library preparation and sequencing were carried out at CeBiTec. Total RNA quantity and integrity were assessed using the Aligent 2100 bioanalyzer with the Prokaryote total RNA Pico assay (Agilent RNA 6000 Pico kit). Only RNA samples with an integrity number >7 were used for sequencing. The TruSeq Stranded Total RNA kit (Illumina) was used for RNA library preparation. The rRNA depletion step was omitted. Of the total RNA, 80 ng (in 5 µl volume) was combined with 13 µl of 'Fragment, Prime and Finish mix' for the RNA fragmentation step according to the Illumina TruSeq stranded mRNA sample preparation guide. Subsequent steps were performed as described in the sample preparation guide. The library was sequenced on a HiSeq1500 platform (Illumina) in a 1 × 150 bp single-end run generating >20 million reads per sample. The resulting reads were pre-processed, including removal of adaptors and quality trimming (slidingwindow:4:21 minlen:100) using bbduk v34 from the BBMAP package[69] and Trimmomatic software v0.35[70], respectively. The trimmed reads were sorted into ribosomal RNA (rRNA) and non-ribosomal RNA (non-rRNA) reads using

SortMeRNA software v2.0[71]. A random subset of 1 million rRNA reads per sample was taxonomically classified with phyloFlash software v3.0 beta 1[72] based on the SILVA database (release 132)[65].

Oligotyping of *Sulfurimonas* 16S rRNA amplicon sequences

In addition to the data presented in this study, *Sulfurimonas* V3–V4 16S sequences were obtained from previous studies conducted in environments likely to host *Sulfurimonas* sp. bacteria (Supplementary Table 7). We searched the European Nucleotide Archive (ENA) and PubMed using the following search strategies: (1) The accession numbers of all amplicon datasets generated on the Illumina platform (paired-end) from ENA were extracted, filtering the results for ecological metagenomes from coastal and marine regions likely to contain *Sulfurimonas* sp. (that is, redoxcline environments; accessed 5 March 2020). (2) PubMed was searched using the keywords 'Sulfurimonas, bacteria, 16S sequencing, Illumina', filtering the results on the basis of the amplified region. (3) Further unpublished datasets were directly obtained from the authors (Supplementary Table 7).

Given the inconsistent state of the sequences obtained from public archives or the authors directly, several bioinformatic pre-processing steps were necessary to ensure a uniform data set as input for oligotyping. If required, sequences were demultiplexed and primer-clipped using cutadapt v1.9.1[73]. Paired-end sequences were merged with PEAR v0.9.6[74] and quality trimmed with Trimmomatic v0.35[70] using a sliding window of 4 bp with an average quality of 15. Only sequences between 380 bp and 450 bp that did not contain any ambiguous bases were retained. If minor modifications affecting at maximum 2 bp of the V3–V4 primers used here were employed in the previous studies, the resulting amplicons were trimmed to matching start and end positions, assuming that any primer bias regarding *Sulfurimonas* would be negligible.

Before oligotyping via minimum entropy decomposition (MED)[33], non-*Sulfurimonas* sequences were filtered out using a blastn search against a custom *Sulfurimonas* database consisting of 246 curated full-length 16S *Sulfurimonas* sequences obtained from the SILVA reference database (v138)[65] and assembled from the metagenomes presented in this study. Sequences with an alignment identity of at least 87%, an alignment length of at least 380 bp, a query start at or before position 10, a query end at or after position 380, a subject start between 280 and 380, and a subject end at or after position 690 were retained. An additional taxonomic cross-check was conducted with SINA v1.2.11[75] using the full SILVA reference database (v138)[65]. To reduce computational requirements, sequences were first clustered with Swarm v2.2.2[68] and only the seed sequence of each swarm was taxonomically classified. Only sequences associated with swarms whose seed representative was classified as *Sulfurimonas* at a minimum alignment quality of 90 were used for oligotyping.

MED uses the information content (entropy) at each base position in a sequence alignment to partition the data set into oligotypes[33]. It is therefore able to resolve even single nucleotide differences between closely related sequences if these are likely to originate from a biological signal and not from random noise. Here, MED was used to identify oligotypes among the *Sulfurimonas* sequences generated from samples collected worldwide. MED (decompose v2.1) was run with the following parameters using default values if not otherwise indicated: minimum entropy of 0.0965 (-m), one discriminant for decomposition (-d), a minimum substantive abundance of 50 (-M) and with outlier relocation (-R). After MED, the oligotype representative sequences were again aligned against the SILVA reference database (v138)[65] and only oligotypes classified as *Sulfurimonas* were retained for the subsequent analysis.

Before we further explored the diversity of *Sulfurimonas* 16S amplicons and attempted the detection of environment-specific *Sulfurimonas* ecotypes, we excluded samples where *Sulfurimonas* amplicons were likely to constitute a contamination from adjacent environments, that is, we assumed that in substrate-associated and water samples, proportions of *Sulfurimonas* 16S sequences lower than 0.1% and 1%, respectively, constituted a contamination. Additionally, samples with less than 1,000 sequences in total were excluded from subsequent analyses to avoid bias caused by the low sequencing depth. A total of 308 samples with 2,082,428 sequences represented in 1,389 *Sulfurimonas* oligotypes were retained for further analysis. A full list of samples, their total and *Sulfurimonas* sequence numbers, and number of *Sulfurimonas* oligotypes is included in Supplementary Table 4.

To identify ecotypes, samples were categorized by salinity (marine, brackish, fresh water), zone (pelagic, benthic, subsurface), water depth (coastal, deep-sea), hydrothermal vent influence and artificial environment. On the basis of these categories, the samples were grouped in 11 environments (Supplementary Table 8).

Ecotypes were defined as oligotypes occurring in at least 50% of the samples of a specific group, with 90% of the values in any of the outgroups being lower than half of the 10% quantile of the non-zero sequence proportions in the ingroup. Gamma diversity of *Sulfurimonas* oligotypes was assessed as rarefaction curve of oligotype numbers depending on sampling effort. For each possible number of samples, 100 samples were randomly selected to determine oligotype numbers.

Metagenome sequencing, assembly and binning

Paired-end libraries were prepared with the TruSeq DNA PCR-Free Sample Prep kit (Illumina) and sequencing of libraries was performed on a MiSeq instrument (Illumina; 2 × 300 bp paired reads) using the v3 sequencing chemistry (CeBiTec). Reads had adapters and contaminants removed with the bbduk tool of the BBMap suite (k = 27; mink = 12)[69] and trimmed with Trimmomatic (slidingwindow:4:20; minlen:100; v0.35)[70]. Forward and reverse reads from 12 metagenomes (Supplementary Table 6) were de novo assembled following two strategies: (1) co-assembly with MEGAHIT ('–min-contig-len 1000'; v1.1.2)[76] and (2) individual assembllies with metaSPAdes ('–meta') within the SPAdes suite ('-k 21,33,55,77,99,127'; v3.9.0)[77]. Reads were mapped and indexed using bwa (v0.7.12)[78] and SAMtools (v1.5)[79], and the contigs were binned with CONCOCT (v1.1.0)[80]. A total of 661 bins were de-replicated using dRep (v2.3.2)[81] on the basis of Mash distance[82] and genome-wide ANI[83]. Only 7% of the bins passed length (>50 kpb), completeness (>75%) and redundancy (<25%) filtering, and a total of 19 de-replicated bins (ANI > 99%) were obtained. *Sulfurimonas* bins were identified and refined using Anvi'o interactive interface (v6.2)[84] after the Anvi'o contig database was built to calculate *k*-mer frequencies to identify open reading frames using Prodigal (v2.6.3)[85] and single-copy genes using HMMER (v3.2.1)[86], and to classify the bins on the basis of single-copy gene taxonomy of GTDB[87] using DIAMOND (v0.9.14)[88]. Sequences of 16S rRNA genes were extracted with RNAmmer (v1.2)[89]. Refined *Sulfurimonas* bins were repeatedly re-assembled using BBmap (99% similarity) and SPAdes, removing contigs smaller than 1 kb after each re-assembly step to extend contigs and reduce the size of genome gaps. Completeness and redundancy of the final *Sulfurimonas* MAGs were evaluated using CheckM (v1.2.1; based on 104 bacterial single-copy genes)[90], CheckM2 (v0.1.3; based on machine learning algorithm)[91] and BUSCO (v5.2.2; based on 628 Campylobacterales single-copy genes)[92]. The number of transfer RNAs was identified using ARAGORN (v1.2.36)[93]. We obtained two almost complete *Sulfurimonas* MAGs, named MAG-1 and MAG-2 (Supplementary Table 2). These two MAGs represent consensus MAGs, which are based on 16 individual bins produced from different environmental samples. Proteins from the final *Sulfurimonas* MAGs were predicted and annotated using Prokka (v1.11)[94]. The Prokka-predicted proteins were additionally annotated with Pfam (release 30)[95] and TIGRFAM (release 14)[96] profiles using HMMER searches (v3.1b2)[86] and by the identification of KEGG Orthology numbers with the GhostKOALA webserver[97]. The proteins were also assigned to clusters of orthologous groups (COGs)[98] using the software COGsoft (v4.19.2012)[99] and transmembrane motifs were identified

using TMHMM (v2.0)[100]. On the basis of the various annotation tools, the annotation of proteins of specific interest was manually refined. The sequences of hydrogenases were classified using HydDB[101]. Iron-related genes were identified using FeGenie's tool and database[102]. RedoxyBase[103] and SORGOds[104] were used to identify classes of peroxidase and types of superoxide reductase, respectively.

The quality-filtered short metatranscriptomic reads were mapped to annotated MAG-1 and quantified using the programme kallisto (v0.43.1)[105]. To compare expression levels between genes, gene expression is reported as transcripts per million (TPM)[106], while the effective counts were used for the differential expression of MAG-1 genes between Aurora and Polaris plumes.

Gene prediction from metagenomic assemblies

The genes were predicted on contigs using Prodigal (v2.6.3)[85] and the proteins clustered using MMseq2 (v13.45111;–min-seq-id 0.95 -c 0.95–cov-mode 1–cluster-mode 2)[107]. Proteins with length <50 amino acids were removed (ca. 15% of total proteins). The functional annotation of proteins was conducted using KEGG[108] (release 101.0; with diamond v2.0.15[88]), Pfam (v35.0[109]; with hmmsearch HMMER v3.3.2[86]) and NCycDB[110], a curated database for nitrogen cycling genes (with diamond v2.0.15[88]). Taxonomic affiliation of the contigs was done with CAT (v5.2.3)[111] using GTDB (release 207)[87] as reference database. The metagenomic and metatranscriptomic reads were mapped to the gene sequences using bwa-mem2 (v2.2.1)[112] and converted to counts per gene with htseq-count script from HTSeq (v2.0.2)[113]. On average, 80% and 87% of reads from Polaris and Aurora metagenomes, respectively, were retrieved by the catalogue of genes.

Phylogeny

The backbone phylogenetic tree for 16S rRNA gene sequences was calculated using *Sulfurimonas* sequences from SILVA SSU r138 RefNR ($n = 19$), *Thiomicrospira* sp. CVO (original name for a *Sulfurimonas denitrificans* strain) from SILVA SSU r138 RefNR (U46506), *Sulfurimonas crateris* (MK859925), Gakkel Ridge plume *Sulfurimonas* MAGs (this study) and closest related 16S rRNA gene sequences to ^{U}*S. pluma* MAGs (blastn hits filtered by identity >98 and coverage >97%: JN874148.1 and JN874176.1; GeneBank nucleotide; accessed May 2020). The sequences of *Sulfuricurvum kujiense* from SILVA SSU r138 RefNR ($n = 3$) were used as outgroup. Sequences were aligned with MAFFT using the L-INS-i method with default settings[114], and the alignment was cleaned with BMGE with default setting[115]. Both programmes were used on the Galaxy platform[116]. A maximum-likelihood-based tree was constructed using W-IQ-TREE[117], first searching for the best substitution model[118] before evaluating branch support using 1,000 ultrafast boostrap (UFBoot) and SH-aLRT branch test replicates. Evolutionary placement algorithm (EPA) in RAxML (v8.2.4)[119] was applied to add 253 partial *Sulfurimonas* 16S rRNA gene sequences (250–1,400 bp retrieved from GenBank nucleotide database; data accessed May 2020) to the tree without changing its topology. Further partial 16S rRNA gene sequences of *Sulfurimonas* sp. obtained from previous next-generation sequencing studies conducted in deep-sea hydrothermal fluids (JAH_MCR_Bv6_MCR_CTD03_08; JAH_AXV_Bv6v4_FS788; downloaded from vamps.mbl.edu) and plumes (PRJEB36848; SRP016119; PRJNA638507) were likewise added to the tree.

Phylogenetic trees for functional genes (that is, alpha subunit of 2-oxoglutarate:ferredoxin oxidoreductase, alpha subunit of pyruvate:ferredoxin oxidoreductase, beta subunit of flavocytochrome c sulfide dehydrogenase and sulfide:quinone oxidoreductase) were constructed following the workflow described for the backbone 16S rRNA gene tree, with amino acids as coding sequence and MAFFT alignment method set to 'auto'. Sequences were obtained from GenBank nucleotide and UniProt databases (data accessed May 2020). The sulfide:quinone oxidoreductase sequence of ^{U}*S. pluma* was added to the backbone tree with EPA. All phylogenetic trees were visualized and refined with iTOL[120].

Comparative genomic analysis

We used Anvi'o v6.2[84] as the main tool to carry out comparative genomic analysis, and the overall workflows applied in this section can be found at http://merenlab.org/2016/11/08/pangenomics-v2/, http://merenlab.org/2017/06/07/phylogenomics/ and http://merenlab.org/data/prochlorococcus-metapangenome/.

Pangenomic analysis

We retrieved 26 *Sulfurimonas* and 2 *Sulfuricurvum* genomes (completeness >85% and redundancy <5%) from GenBank (accessed January 2020). Supplementary Table 9a reports information for each isolate genome and MAG. DNA and amino acid sequences of the genomes, including ^{U}*S. pluma* MAG-1 and MAG-2, were stored in an Anvi'o's genome database (programme 'anvi-gen-genomes-storage'). From the genome database, we computed the pangenome to identify the gene clusters (programme 'anvi-pan-genome') representing sequences of one or more predicted open reading frames (Prodigal v2.6.3)[85] grouped together on the basis of their homology at the translated DNA sequence level. For multiple sequences alignments, Anvi'o used MUSCLE (v3.8.1551)[121], the MCL algorithm to identify clusters in amino acid sequence similarity[122] and the programme 'anvi-run-ncbi-cogs' to annotate genes with functions by searching them against the COG database (October 2019 release)[98] using blastp v2.9.0+[123]. ANI was computed for all *Sulfurimonas* species and MAGs representative for different environments (that is, hydrothermal vent and plume, marine pelagic, marine oxic aquifer, costal and terrestrial) with the anvi'o programme 'anvi-compute-genome-similarity'.

Phylogenomic analysis

We visualized the distribution of gene clusters across genomes with 'anvi-display-pan' and manually selected 258 SCGs present in all genomes. Concatenated amino acid sequences and partition file for SCG were extracted with 'anvi-get-sequences-for-gene-clusters' and used for phylogenomic analysis. The phylogenomic tree was built with IQ-Tree (v2.0)[124], first identifying the partition substitution models[125] and then constructing a consensus tree using a maximum-likelihood approach with 1,000 UFBoot branch test replicates.

Metapangenomic analysis

Here we refer to 'metapangenome' as defined in ref. [126], that is, 'the outcome of the analysis of pangenomes in conjunction with the environment where the abundance and prevalence of gene clusters and genomes are recovered through shotgun metagenomes'. We have also extended the analysis to include gene cluster expression as retrieved through shotgun metatranscriptomes. The Anvi'o pangenome was computed as described above but including only *Sulfurimonas* isolate genomes and ^{U}*S. pluma* MAG-1, and using the programme 'anvi-import-functions' to import annotations from other databases, including Pfam[95], eggNOG[127], TIGRFAM[96] and the script by E. Graham (https://github.com/edgraham/GhostKoalaParser) to import GhostKOALA/KEGG annotations[97]. To investigate the environmental distribution and gene transcription patterns of ^{U}*S. pluma* MAG-1, metagenomic and metatranscriptomic reads from hydrothermal plumes and benthic vent environments were mapped to MAG-1 and reads recruitment information was imported to the Anvi'o pangenome. To test the specificity of read recruitments, the genomes from isolated *Sulfurimonas* species were also included in the metapangenomic analysis. We downloaded hydrothermal vent (that is, plume, seawater, fluid, chimney, mineral deposit and sediment; $n = 63$) and background seawater ($n = 3$) metagenomes from ENA (data accessed February 2020; Supplementary Table 9b). Hydrothermal plume ($n = 2$), seawater ($n = 6$), fluid ($n = 6$) and

background seawater ($n = 2$) metatranscriptomes were downloaded from ENA (data accessed May 2020) and the metagenomics RAST server (MG-RAST; data accessed May 2020; Supplementary Table 9b). Low-quality reads were removed from metagenomes and metatranscriptomes with Trimmomatic (slidingwindow:4:20 and minlen:65; v0.35)[70]. The quality-filtered short metagenomic reads were mapped to the concatenated *Sulfurimonas* genomes and ^{U}*S. pluma* MAG-1 with Bowtie2 (default setting and '–no-unal' flag; v2.3.2)[128]. An Anvi'o database was created for storing contigs information (DNA sequence, GC content, tetranucleotide frequency and open reading frames) of *Sulfurimonas* genomes and ^{U}*S. pluma* MAG-1 with the programme 'anvi-gen-contigs-database'. Then the programme 'anvi-profile' was used to process BAM files and generate an Anvi'o profile database that contained the coverage statistics of each *Sulfurimonas* genome in a given metagenome, summarizing gene coverage values across metagenomes and percentage read recruitment per genome with 'anvi-summarize'. This percentage refers to the contribution of reads recruited by a genome to total reads recruited by all genomes for each metagenome. The table was then imported to the *Sulfurimonas* pangenome in Anvio'o using the programme 'anvi-import-misc-data' as 'layers'. The quality-filtered short metatranscriptomic reads were mapped to ^{U}*S. pluma* MAG-1 gene clusters (extracted from Anvi'o's *Sulfurimonas* pangenome with 'anvi-get-sequences-for-gene-clusters') with the programme kallisto (v0.43.1)[105]. The expression of gene clusters is reported in TPM, correcting for the effective length of the transcript. The TPM of ^{U}*S. pluma* MAG-1 genes were imported to the *Sulfurimonas* pangenome using the programme 'anvi-import-misc-data' as 'items'.

Statistical analysis

All statistical analyses were conducted in R using the core distribution (v3.5.2)[129], the package vegan[130] for ecotype analysis and the package edgeR[131] for testing differences in the genes' expression between Aurora and Polaris plumes, applying false discovery rate (FDR) criterion proposed by Benjamini and Hochberg[132]. Unless specified otherwise, plots were made using ggplot2[133] and pooled, refined and labelled with Adobe Illustrator CS5.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The sequences generated in this study and the ^{U}*Sulfurimonas pluma* genomes have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under Bioproject PRJEB48226. All the sequences were archived using the data brokerage service of the German Federation for Biological Data (GFBio)[134]. The accession numbers for raw sequences and ^{U}*S. pluma* MAGs are listed in Supplementary Table 6 and Supplementary Table 2, respectively.

Sulfurimonas V3-V4 16S rRNA gene sequences were extracted from metabarcoding studies obtained from ENA (https://www.ebi.ac.uk/ena/browser/home), and the studies and sequences accession numbers are reported in Supplementary Tables 4 and Table 7. *Sulfurimonas* and *Sulfuricurvum kujiense* full-length 16S rRNA gene sequences used in this study were obtained from SILVA RefNR database (v138; https://www.arb-silva.de) and NCBI Genebank (https://www.ncbi.nlm.nih.gov/nucleotide; JN874148.1 and JN874176.1). Partial *Sulfurimonas* 16S rRNA gene sequences (250–1,400 bp) were retrieved from NCBI GenBank (https://www.ncbi.nlm.nih.gov/nucleotide; data accessed May 2020), and from metabarcoding studies obtained from ENA (https://www.ebi.ac.uk/ena/browser/home) and VAMPS (vamps.mbl.edu). Sequences and studies accession numbers are reported in Fig. 1. Functional gene sequences used for phylogenetic tree analysis were obtained either from UniProt (https://www.uniprot.org/uniprotkb?query=*; data accessed May 2020) or NCBI GenBank (https://www.ncbi.nlm.nih.gov/nucleotide; data accessed May 2020), and their accessions are provided in Fig. 4 and Extended Data Fig. 6. *Sulfurimonas* and *Sulfuricurvum* genomes used in this study are available via NCBI GenBank (https://www.ncbi.nlm.nih.gov/nucleotide; data accessed January 2020) and accessions are reported in Supplementary Table 9a. Metagenomic and metranscitptomic data used in metapangenomic analysis were obtained either from ENA (https://www.ebi.ac.uk/ena/browser/home; data accessed May 2020) or MG-RAST (https://www.mg-rast.org; data accessed May 2020), and study accession number is provided in Supplementary Table 9b. Source data are provided with this paper.

References

1. Inagaki, F., Takai, K., Kobayashi, H., Nealson, K. H. & Horikoshi, K. *Sulfurimonas autotrophica* gen. nov., sp. nov., a novel sulfur-oxidizing e-proteobacterium isolated from hydrothermal sediments in the Mid-Okinawa Trough. *Int. J. Syst. Evol. Microbiol.* **53**, 1801–1805 (2003).
2. Timmer-Ten Hoor, A. A new type of thiosulphate oxidizing, nitrate reducing microorganism: *Thiomicrospira denitrificans* sp. nov. *Neth. J. Sea Res.* **9**, 344–350 (1975).
3. Cai, L., Shao, M. & Zhang, T. Non-contiguous finished genome sequence and description of *Sulfurimonas hongkongensis* sp. nov., a strictly anaerobic denitrifying, hydrogen- and sulfur-oxidizing chemolithoautotroph isolated from marine sediment. *Stand. Genom. Sci.* **9**, 1302–1310 (2014).
4. Wang, S., Jiang, L., Liu, X., Yang, S. & Shao, Z. *Sulfurimonas xiamenensis* sp. nov. and *Sulfurimonas lithotrophica* sp. nov., hydrogen- and sulfur-oxidizing chemolithoautotrophs within the Epsilonproteobacteria isolated from coastal sediments, and an emended description of the genus *Sulfurimonas*. *Int. J. Syst. Evol. Microbiol.* **70**, 2657–2663 (2020).
5. Takai, K. et al. *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*. *Int. J. Syst. Evol. Microbiol.* **56**, 1725–1733 (2006).
6. Hu, Q., Wang, S., Lai, Q., Shao, Z. & Jiang, L. *Sulfurimonas indica* sp. nov., a hydrogen- and sulfur-oxidizing chemolithoautotroph isolated from a hydrothermal sulfide chimney in the Northwest Indian Ocean. *Int. J. Syst. Evol. Microbiol.* **71**, 1466–5034 (2021).
7. Wang, S. et al. *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**, 813–822 (2021).
8. Wang, S. et al. Characterization of *Sulfurimonas hydrogeniphila* sp. nov., a novel bacterium predominant in deep-sea hydrothermal vents and comparative genomic analyses of the genus *Sulfurimonas*. *Front. Microbiol.* **12**, 626705 (2021).
9. Labrenz, M. et al. *Sulfurimonas gotlandica* sp. nov., a chemoautotrophic and psychrotolerant epsilonproteobacterium isolated from a pelagic redoxcline, and an emended description of the genus *Sulfurimonas*. *Int. J. Syst. Evol. Microbiol.* **63**, 4141–4148 (2013).
10. Henkel, J. V. et al. 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. *Syst. Appl. Microbiol.* **44**, 126155 (2021).
11. Ratnikova, N. M. et al. *Sulfurimonas crateris* sp. nov., a facultative anaerobic sulfur-oxidizing chemolithoautotrophic bacterium isolated from a terrestrial mud volcano. *Int. J. Syst. Evol. Microbiol.* **70**, 487–492 (2020).

12. Han, Y. & Perner, M. The globally widespread genus *Sulfurimonas*: versatile energy metabolisms and adaptations to redox clines. *Front. Microbiol.* **6**, 989 (2015).
13. López-garcía, P. et al. Bacterial diversity in hydrothermal sediment and epsilonproteobacterial dominance in experimental microcolonizers at the Mid-Atlantic Ridge. *Environ. Microbiol.* **5**, 961–976 (2003).
14. Nakagawa, S. et al. Distribution, phylogenetic diversity and physiological characteristics of epsilon-Proteobacteria in a deep-sea hydrothermal field. *Environ. Microbiol.* **7**, 1619–1632 (2005).
15. Huber, J. A. et al. Isolated communities of Epsilonproteobacteria in hydrothermal vent fluids of the Mariana Arc seamounts. *FEMS Microbiol. Ecol.* **73**, 538–549 (2010).
16. Meier, D. V. et al. Niche partitioning of diverse sulfur-oxidizing bacteria at hydrothermal vents. *ISME J.* **11**, 1545–1558 (2017).
17. Mino, S. et al. Endemicity of the cosmopolitan mesophilic chemolithoautotroph Sulfurimonas at deep-sea hydrothermal vents. *ISME J.* **11**, 909–919 (2017).
18. Akerman, N. H., Butterfield, D. A., Huber, J. A., Huber, J. A. & Paul, J. B. Phylogenetic diversity and functional gene patterns of sulfur-oxidizing subseafloor Epsilonproteobacteria in diffuse hydrothermal vent fluids. *Front. Microbiol.* **4**, 185 (2013).
19. Rogge, A., Vogts, A., Voss, M. & Labrenz, M. Success of chemolithoautotrophic SUP05 and *Sulfurimonas* GD17 cells in pelagic Baltic Sea redox zones is facilitated by their lifestyles as K- and r -strategists. *Environ. Microbiol.* **19**, 2495–2506 (2017).
20. German, C. R. et al. Diverse styles of submarine venting on the ultraslow spreading Mid-Cayman Rise. *Proc. Natl Acad. Sci. USA* **107**, 14020–14025 (2010).
21. Sylvan, J. B., Pyenson, B. C., Rouxel, O., German, C. R. & Edwards, K. J. Time-series analysis of two hydrothermal plumes at 9°50′ N East Pacific Rise reveals distinct, heterogeneous bacterial populations. *Geobiology* **10**, 178–192 (2012).
22. Perner, M. et al. In situ chemistry and microbial community compositions in five deep-sea hydrothermal fluid samples from Irina II in the Logatchev field. *Environ. Microbiol.* **15**, 1551–1560 (2013).
23. Haalboom, S. et al. Patterns of (trace) metals and microorganisms in the Rainbow hydrothermal vent plume at the Mid-Atlantic Ridge. *Biogeosciences* **17**, 2499–2519 (2020).
24. Li, J. et al. Distribution and succession of microbial communities along the dispersal pathway of hydrothermal plumes on the Southwest Indian Ridge. *Front. Mar. Sci.* **7**, 581381 (2020).
25. Dick, G. J. et al. The microbiology of deep-sea hydrothermal vent plumes: ecological and biogeographic linkages to seafloor and water column habitats. *Front. Microbiol.* **4**, 124 (2013).
26. Dick, G. J. The microbiomes of deep-sea hydrothermal vents: distributed globally, shaped locally. *Nat. Rev. Microbiol.* **17**, 271–283 (2019).
27. German, C. R. & Seyfried, W. E. in *Treatise on Geochemistry* 2nd edn (eds Holland, H. D. & Turekian, K. K.), **8**, 191–233 (Elsevier, 2014).
28. Kadko, D., Baross, J. & Alt, J. The magnitude and global implications of hydrothermal flux. *Geophys. Monogr. Ser.* **91**, 446–466 (1995).
29. German, C. R. et al. Volcanically hosted venting with indications of ultramafic influence at Aurora hydrothermal field on Gakkel Ridge. *Nat. Commun.* **13**, 6517 (2022).
30. Bowers, R. M. et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **35**, 725–731 (2017).
31. Konstantinidis, K. T., Rosselló-móra, R. & Amann, R. Uncultivated microbes in need of their own taxonomy. *ISME J.* **11**, 2399–2406 (2017).
32. Murray, A. E. et al. Roadmap for naming uncultivated Archaea and Bacteria. *Nat. Microbiol.* **5**, 987–994 (2020).
33. Eren, A. M. et al. Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *ISME J.* **9**, 968–979 (2015).
34. Dick, G. J. & Tebo, B. M. Microbial diversity and biogeochemistry of the Guaymas Basin deep-sea hydrothermal plume. *Environ. Microbiol.* **12**, 1334–1347 (2010).
35. Lesniewski, R. A., Jain, S., Anantharaman, K., Schloss, P. D. & Dick, G. J. The metatranscriptome of a deep-sea hydrothermal plume is dominated by water column methanotrophs and lithotrophs. *ISME J.* **6**, 2257–2268 (2012).
36. Sheik, C. S. et al. Spatially resolved sampling reveals dynamic microbial communities in rising hydrothermal plumes across a back-arc basin. *ISME J.* **9**, 1434–1445 (2014).
37. Reed, D. C. et al. Predicting the response of the deep-ocean microbiome to geochemical perturbations by hydrothermal vents. *ISME J.* **9**, 1857–1869 (2015).
38. Han, Y. & Perner, M. The role of hydrogen for *Sulfurimonas denitrificans*' metabolism. *PLoS ONE* **9**, 8–14 (2014).
39. Ilbert, M. & Bonnefoy, V. Insight into the evolution of the iron oxidation pathways. *Biochim. Biophys. Acta* **1827**, 161–175 (2013).
40. Yu, H. & Leadbetter, J. R. Bacterial chemolithoautotrophy via manganese oxidation. *Nature* **583**, 453–458 (2020).
41. Andrews, S. C. Iron storage in bacteria. *Adv. Microb. Physiol.* **40**, 281–351 (1998).
42. Pitcher, R. S. & Watmough, N. J. The bacterial cytochrome cbb 3 oxidases. *Biochim. Biophys. Acta* **1655**, 388–399 (2004).
43. Sousa, F. L. et al. The superfamily of heme–copper oxygen reductases: types and evolutionary considerations. *Biochim. Biophys. Acta* **1817**, 629–637 (2012).
44. Park, B. et al. Cultivation of autotrophic ammonia-oxidizing archaea from marine sediments in coculture with sulfur-oxidizing bacteria. *Appl. Environ. Microbiol.* **76**, 7575–7587 (2010).
45. Fuchs, G. Alternative pathways of carbon dioxide fixation: insights into the early evolution of life? *Annu. Rev. Microbiol.* **65**, 631–658 (2011).
46. Bayer, B. et al. Metabolic versatility of the nitrite-oxidizing bacterium *Nitrospira marina* and its proteomic response to oxygen-limited conditions. *ISME J.* **15**, 1025–1039 (2021).
47. Yamamoto, M., Arai, H., Ishii, M. & Igarashi, Y. Role of two 2-oxoglutarate: ferredoxin oxidoreductases in *Hydrogenobacter thermophilus* under aerobic and anaerobic conditions. *FEMS Microbiol. Lett.* **263**, 189–193 (2006).
48. Yamamoto, M., Ikeda, T., Arai, H., Ishii, M. & Igarashi, Y. Carboxylation reaction catalyzed by 2-oxoglutarate:ferredoxin oxidoreductases from *Hydrogenobacter thermophilus*. *Extremophiles* **14**, 79–85 (2010).
49. Berg, I. A. Ecological aspects of the distribution of different autotrophic CO_2 fixation pathways. *Appl. Environ. Microbiol.* **77**, 1925–1936 (2011).
50. French, C. E., Bell, J. M. L. & Ward, F. B. Diversity and distribution of hemerythrin-like proteins in prokaryotes. *FEMS Microbiol. Lett.* **279**, 131–145 (2008).
51. Isaza, C. E., Silaghi-dumitrescu, R., Iyer, R. B., Kurtz, D. M. & Chan, M. K. Structural basis for O_2 sensing by the hemerythrin-like domain of a bacterial chemotaxis protein: substrate tunnel and fluxional n terminus. *Biogeochemistry* **45**, 9023–9031 (2006).
52. Kendall, J. J., Barrero-tobon, A. M., Hendrixson, D. R. & Kelly, D. J. Hemerythrins in the microaerophilic bacterium *Campylobacter jejuni* help protect key iron–sulphur cluster enzymes from oxidative damage. *Environ. Microbiol.* **16**, 1105–1121 (2014).

53. Nariya, S. & Kalyuzhnaya, M. G. Hemerythrins enhance aerobic respiration in *Methylomicrobium alcaliphilum* 20Z R, a methane-consuming bacterium. *FEMS Microbiol. Lett.* **367**, fnaa003 (2020).
54. Sheng, Y. et al. Superoxide dismutases and superoxide reductases. *Chem. Rev.* **114**, 3854–3918 (2014).
55. Anantharaman, K., Breier, J. A., Sheik, C. S. & Dick, G. J. Evidence for hydrogen oxidation and metabolic plasticity in widespread deep-sea sulfur-oxidizing bacteria. *Proc. Natl Acad. Sci. USA* **110**, 330–335 (2013).
56. Dede, B. et al. Niche differentiation of sulfur-oxidizing bacteria (SUP05) in submarine hydrothermal plumes. *ISME J.* **16**, 1479–1490 (2022).
57. Schlindwein, V. (ed.) *The Expedition of the Research Vessel 'Polarstern' to the Antarctic in 2013 (ANT-XXIX/8)*. Reports on polar and marine research, Bremerhaven, Alfred Wegener Institute for Polar and Marine Research, **672**, 111 (2014); https://doi.org/10.2312/BzPM_0672_2014
58. Boetius, A. *The Expedition PS86 of the Research Vessel POLARSTERN to the Arctic Ocean in 2014*. Reports on polar and marine research, Bremerhaven, Alfred Wegener Institute for Polar and Marine Research, **685**, 133 (2015); https://doi.org/10.2312/BzPM_0685_2015
59. Boetius, A. & Purser, A. *The Expedition PS101 of the Research Vessel POLARSTERN to the Arctic Ocean in 2016*. Reports on polar and marine research, Bremerhaven, Alfred Wegener Institute for Polar and Marine Research, **706**, 230 (2017); https://doi.org/10.2312/BzPM_0706_2017
60. Varliero, G., Bienhold, C., Schmid, F., Boetius, A. & Molari, M. Microbial diversity and connectivity in deep-sea sediments of the South Atlantic Polar Front. *Front. Microbiol.* **10**, 665 (2019).
61. Pernthaler, A., Pernthaler, J. & Amann, R. Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Appl. Environ. Microbiol.* **68**, 3094–3101 (2002).
62. Alm, E. W., Oerther, D. B., Larsen, N., Stahl, D. A. & Raskin, L. The oligonucleotide probe database. *Appl. Environ. Microbiol.* **62**, 3557–3559 (1996).
63. Amann, R. I. et al. Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl. Environ. Microbiol.* **56**, 1919–1925 (1990).
64. Ludwig, W. et al. ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**, 1363–1371 (2004).
65. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, 590–596 (2013).
66. Klindworth, A. et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* **41**, e1 (2013).
67. Hassenrück, C., Quast, C., Rapp, J. & Buttigieg, P. *Amplicon* (GitHub, accessed 15 April 2019); https://github.com/chassenr/NGS/tree/master/AMPLICON
68. Mahé, F., Rognes, T., Quince, C., de Vargas, C. & Dunthorn, M. Swarm: robust and fast clustering method for amplicon-based studies. *PeerJ* **2**, e593 (2014).
69. Bushnell, B. *BBMap: A Fast, Accurate, Splice-Aware Aligner. United States* (2014). https://www.osti.gov/servlets/purl/1241166
70. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
71. Kopylova, E., Noe, L. & Touzet, H. Sequence analysis SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217 (2012).
72. Gruber-vodicka, H. R., Seah, B. K. & Pruesse, E. phyloFlash: rapid small-subunit rRNA profiling and targeted assembly from metagenomes. *mSystems* **5**, e00920 (2020).
73. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* https://doi.org/10.14806/ej.17.1.200 (2011).
74. Zhang, J., Kobert, K., Fluori, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614–620 (2014).
75. Pruesse, E., Peplies, J. & Glöckner, F. O. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**, 1823–1829 (2012).
76. Li, D., Liu, C., Luo, R., Sadakane, K. & Lam, T. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
77. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).
78. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
79. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
80. Alneberg, J. et al. Binning metagenomic contigs by coverage and composition. *Nat. Methods* **11**, 1144–1146 (2014).
81. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J.* **11**, 2864–2868 (2017).
82. Ondov, B. D. et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **17**, 132 (2016).
83. Varghese, N. J. et al. Microbial species delineation using whole genome sequences. *Nucleic Acids Res.* **43**, 6761–6771 (2015).
84. Eren, A. M. et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* **3**, e1319 (2015).
85. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
86. Wheeler, T. J. & Eddy, S. R. nhmmer: DNA homology search with profile HMMs. *Bioinformatics* **29**, 2487–2489 (2013).
87. Parks, D. H. et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018).
88. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
89. Lagesen, K. et al. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**, 3100–3108 (2007).
90. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
91. Chklovski, A., Parks, D. H., Woodcroft, B. J. & Tyson, G. W. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. Preprint at *bioRxiv* https://doi.org/10.1101/2022.07.11.499243 (2022).
92. Manni, M., Berkeley, M. R., Seppey, M. & Zdobnov, E. M. BUSCO: assessing genomic data quality and beyond. *Curr. Protoc.* **1**, e323 (2021).
93. Laslett, D. & Canback, B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* **32**, 11–16 (2004).
94. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).

95. El-Gebali, S. et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).
96. Haft, D. H. et al. TIGRFAMs: a protein family resource for the functional identification of proteins. *Nucleic Acids Res.* **29**, 41–43 (2001).
97. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **248**, 726–731 (2015).
98. Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**, 33–36 (2000).
99. Kristensen, D. M. et al. A low-polynomial algorithm for assembling clusters of orthologous groups from intergenomic symmetric best matches. *Bioinformatics* **26**, 1481–1487 (2010).
100. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
101. Søndergaard, D., Pedersen, C. N. S. & Greening, C. HydDB: a web tool for hydrogenase classification and analysis. *Sci. Rep.* **6**, 34212 (2016).
102. Garber, A. I. et al. FeGenie: a comprehensive tool for the identification of iron genes and iron gene neighborhoods in genome and metagenome assemblies. *Front. Microbiol.* **11**, 37 (2020).
103. Passardi, F. et al. PeroxiBase: the peroxidase database. *Phytochemistry* **68**, 1605–1611 (2007).
104. Lucchetti-miganeh, C., Goudenège, D., Thybert, D., Salbert, G. & Barloy-hubler, F. SORGOdb: superoxide reductase gene ontology curated database. *BMC Microbiol.* **11**, 105 (2011).
105. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 4–8 (2016).
106. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
107. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
108. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–D462 (2016).
109. Mistry, J. et al. Pfam: the protein families database in 2021. *Nucleic Acids Res.* **49**, D412–D419 (2021).
110. Tu, Q., Lin, L., Cheng, L., Deng, Y. & He, Z. NCycDB: a curated integrative database for fast and accurate metagenomic profiling of nitrogen cycling genes. *Bioinformatics* **35**, 1040–1048 (2019).
111. Li, H. et al. A cross-species alignment tool (CAT). *BMC Bioinformatics* **8**, 349 (2007).
112. Vasimuddin, M., Misra, S., Li, H. & Aluru, S. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS),* Rio de Janeiro, Brazil, 314–324, doi: 10.1109/IPDPS.2019.00041 (2019); https://ieeexplore.ieee.org/document/8820962
113. Putri, G. H., Anders, S., Pyl, P. T., Pimanda, J. E. & Zanini, F. Analysing high-throughput sequencing data in Python with HTSeq 2.0. *Bioinformatics* **38**, 2943–2945 (2022).
114. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
115. Criscuolo, A. & Gribaldo, S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* **10**, 210 (2010).
116. Jalili, V. et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. *Nucleic Acids Res.* **48**, 395–402 (2020).
117. Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A. & Minh, B. Q. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **44**, W232–W235 (2016).
118. Kalyaanamoorthy, S. et al. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
119. Berger, S. A., Krompass, D. & Stamatakis, A. Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood. *Syst. Biol.* **60**, 291–302 (2011).
120. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **2**, W256–W259 (2019).
121. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
122. van Dongen, S. & Abreu-goodger, C. in *Bacterial Molecular Networks: Methods and Protocols, Methods in Molecular Biology* (eds van Helden, J. et al.) 281–295 (Springer, 2012).
123. Altschup, S. F., Gish, W., Pennsylvania, T. & Park, U. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
124. Nguyen, L., Schmidt, H. A., Haeseler, A., Von & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2014).
125. Chernomor, O., von Haeseler, A. & Minh, B. Q. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst. Biol.* **65**, 997–1008 (2016).
126. Delmont, T. O. & Eren, A. M. Linking pangenomes and metagenomes: the *Prochlorococcus* metapangenome. *PeerJ* **6**, e4320 (2018).
127. Jensen, L. J. et al. eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic Acids Res.* **36**, 250–254 (2008).
128. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–360 (2012).
129. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2013).
130. Oksanen, J. et al. vegan: Community Ecology Package. R package version 2.6-4. https://CRAN.R-project.org/package=vegan (2022).
131. Robinson, M. D., Mccarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
132. Reiner-Benaim, A. FDR control by the BH procedure for two-sided correlated tests with implications to gene expression data analysis. *Biom. J.* **49**, 107–126 (2007).
133. Villanueva, R. A. M. & Chen, Z. J. ggplot2: elegant graphics for data analysis (2nd ed.). *Meas. Interdiscip. Res. Perspect.* **17**, 160–167 (2019).
134. Diepenbroek, M. et al. Towards an integrated biodiversity and ecological research data management and archiving platform: the German Federation for the Curation of Biological Data (GFBio). In *Informatik 2014 – Big Data Komplexität meistern* Proc. 232 (eds Plödereder, E. et al.) 1711–1725 (Gesellschaft für Informatik, 2014).
135. Schmidt, K., Koschinsky, A., Garbe-Schönberg, D., de Carvalho, L. M. & Seifert, R. Geochemistry of hydrothermal fluids from the ultramafic-hosted Logatchev hydrothermal field, 15°N on the Mid-Atlantic Ridge: temporal and spatial investigation. *Chem. Geol.* **242**, 1–21 (2007).
136. Perner, M. et al. The influence of ultramafic rocks on microbial communities at the Logatchev hydrothermal field, located 15 degrees N on the Mid-Atlantic Ridge. *FEMS Microbiol. Ecol.* **61**, 97–109 (2007).

137. Douville, E. et al. The rainbow vent fluids (36°14′N, MAR): the influence of ultramafic rocks and phase separation on trace metal content in Mid-Atlantic Ridge hydrothermal fluids. *Chem. Geol.* **184**, 37–48 (2002).
138. Ji, F. et al. Geochemistry of hydrothermal vent fluids and its implications for subsurface processes at the active Longqi hydrothermal field, Southwest Indian Ridge. *Deep Sea Res. I* **122**, 41–47 (2017).

Acknowledgements

We thank the captain, crew and science party of RV *Polarstern* expeditions PS81, PS86 and PS101; H. Tegetmeyer for support with Illumina sequencing (CeBiTec laboratory; HGF-MPG Joint Research Group on Deep Sea Ecology and Technology); A. Mera and M. Philippi for helping with design and testing of the specific probes for ^{U}*S. pluma*; M. Meiners, W. Stiens, R. Stiens, J. Barz, F. Schramm and A. Nordhausen (HGF-MPG Joint Research Group for Deep-Sea Ecology and Technology) for technical support; and the Anvi'o team and community for making microbial omics analyses accessible to every scientist. This work was funded by the Helmholtz Association (Alfred Wegener Institute Helmholtz Center for Polar and Marine Research, Bremerhaven), Max Planck Society (MPG), the ERC Advanced Investigator Grant ABYSS (294757) to A.B., and Germany's Excellence Initiative of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through the Clusters of Excellence 'The Ocean in the Earth System' (EXC-309- 49926684) and 'The Ocean Floor-Earth's Uncharted Interface' (EXC-2077–390741603).

Author contributions

M.M. conceived the study. A.B. led the expeditions and the discovery and sampling of hydrothermal vents in the Arctic Ocean. M.M. and G.W. performed field measurements and sampling activities as well as chemical analyses. M.M. supervised the design of gene probes and CARD-FISH analysis. C.H. and M.M. selected datasets and analysed 16S rRNA TAG sequences. S.S. and M.M. assembled Gakkel plume MAGs. M.M. performed phylogenetic analyses, comparative genomic, metatranscriptomic and differential gene expression analyses. P.O. provided support for the phylogenetic analyses, and annotation and identification of functional marker genes. R.L.-P. and M.M. reconstructed the metabolic map of ^{U}*S. pluma*. M.M. wrote the draft of the manuscript with input from all co-authors.

Funding

Open access funding provided by Max Planck Society.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41564-023-01342-w.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41564-023-01342-w.

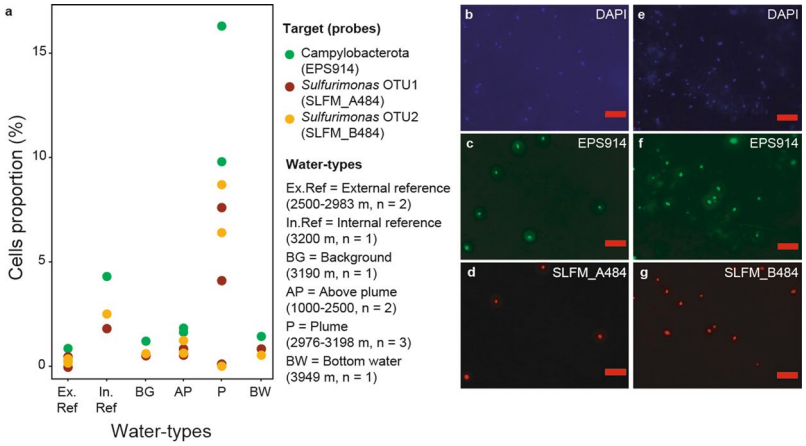
Correspondence and requests for materials should be addressed to Massimiliano Molari.

Peer review information *Nature Microbiology* thanks Karthik Anantharaman and Mirjam Perner for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

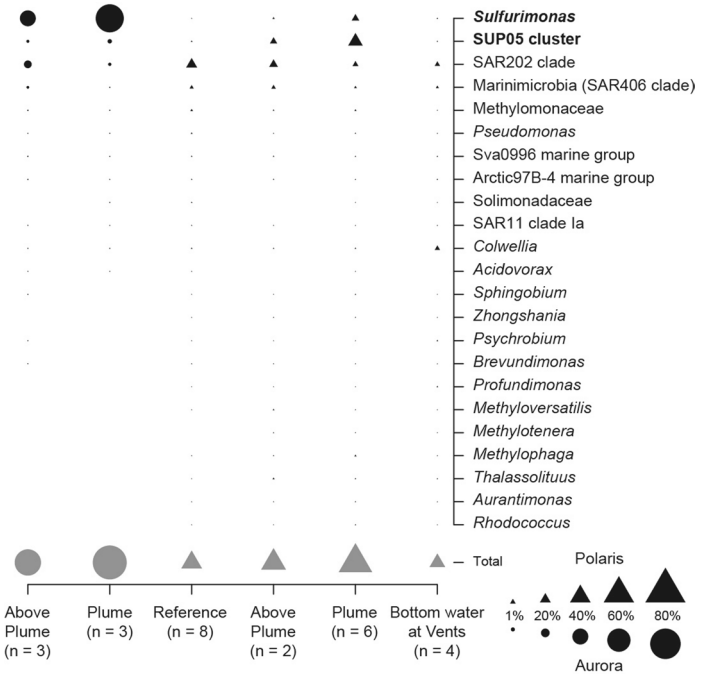
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

https://doi.org/10.1038/s41564-023-01342-w



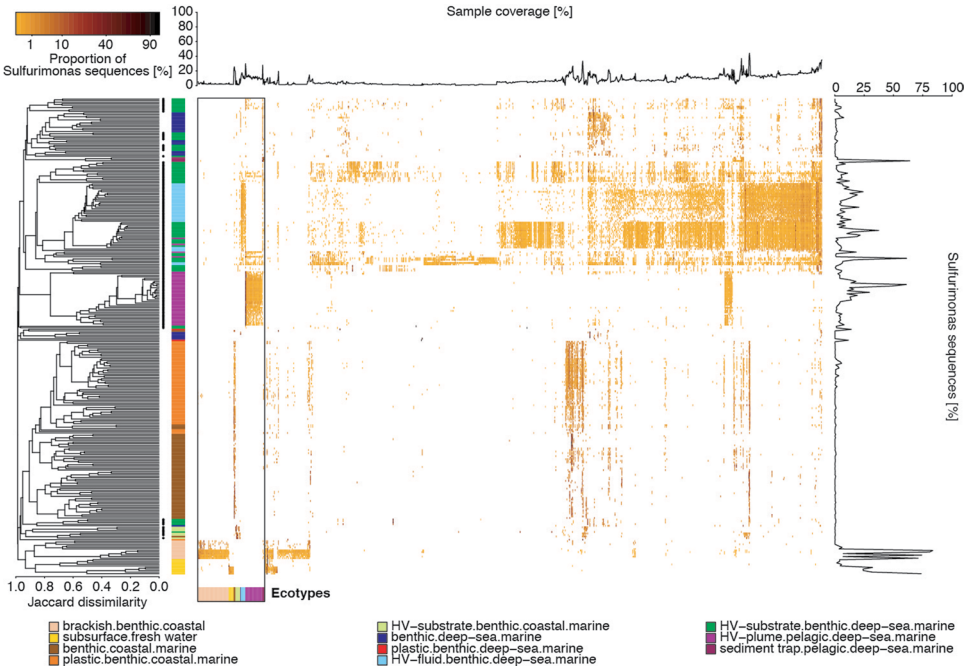
Extended Data Fig. 1 | Abundance of Campylobacterota and hydrothermal plume *Sulfurimonas* sp. cells at Arctic sites. a, Proportion of Campylobacterota and hydrothermal plume *Sulfurimonas* sp. to DAPI-stained cell counts in different water types at Aurora vent and reference sites (water depth and number of samples are reported in parenthesis); **b** – **g**, fluorescence micrographs of microbes in the Aurora plume (station PS86-55) stained with DAPI (blue; **b** and **e**) and simultaneously (double CARD-FISH) with specific probes for Campylobacterota (green; **c** and **f**) and for Arctic *Sulfurimonas* OTU1 (red; **d**) and OTU2 (red; **g**). Panels b–g represents observations made from independent samples (n = 3). Scale bars 10 µm (b–g).

https://doi.org/10.1038/s41564-023-01342-w



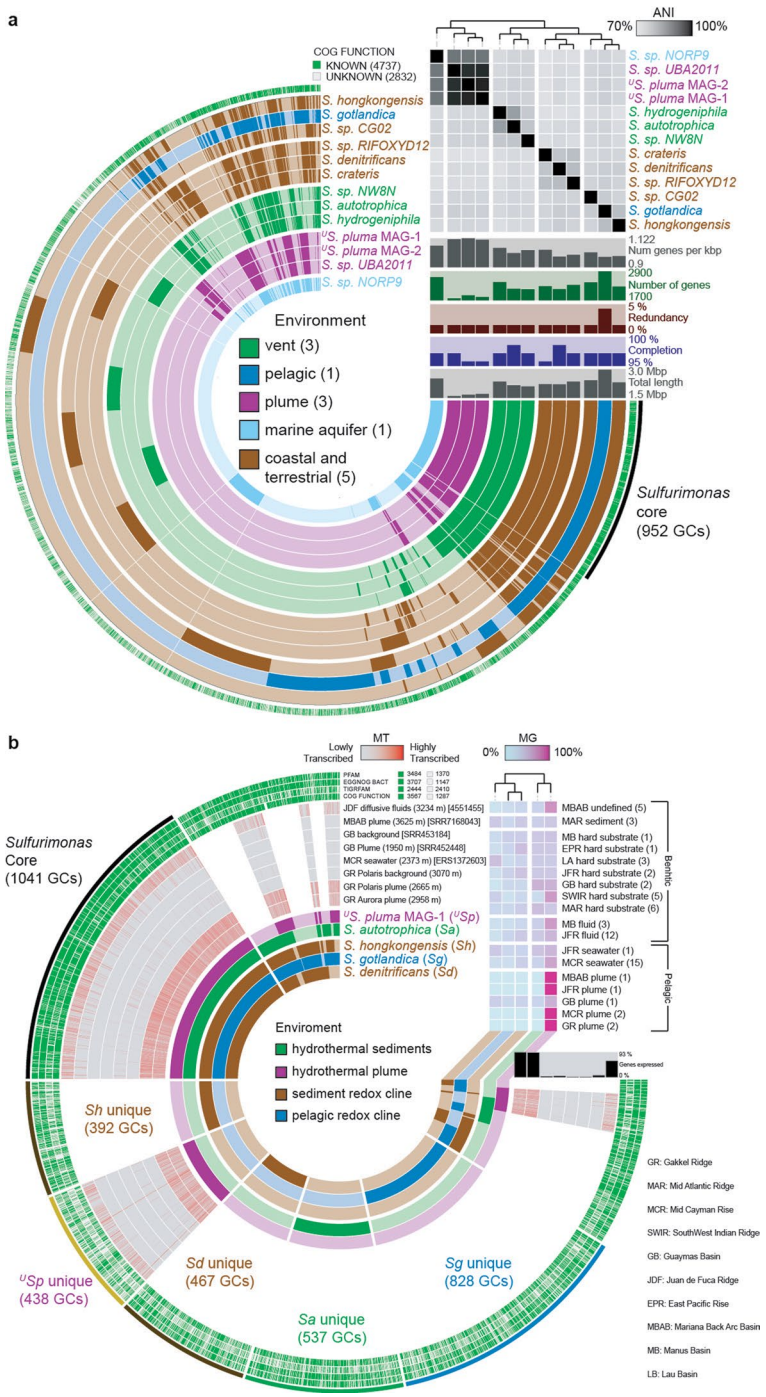
Extended Data Fig. 2 | Active bacterial community structure at Aurora and Polaris water types. Dominant active bacterial taxa (cut-off ≥1.5 %) as detected by 16 S rRNA sequencing (cDNA). Average of sequences proportion is reported. The most dominant and active chemolithotrophs are in bold. Number of observations (n) is reported in parentheses.

https://doi.org/10.1038/s41564-023-01342-w



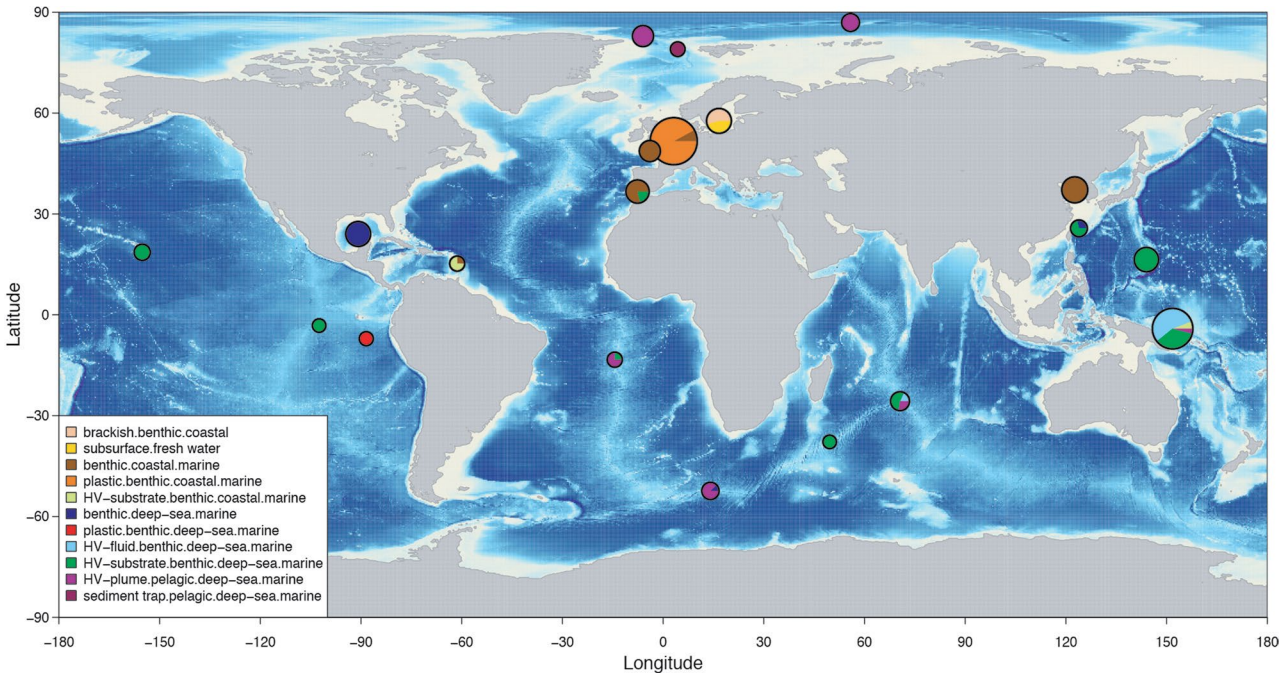
Extended Data Fig. 3 | Heatmap of the distribution of *Sulfurimonas* oligotypes across samples. Samples are ordered based on average linkage clustering of Jaccard dissimilarities (left panel). Asterisks indicate samples of hydrothermal origin. Top panel: percentage of samples an oligotype occurs in (sample coverage). Right panel: percentage of *Sulfurimonas* sequences of total sequences per sample. Bottom panel: Ecotypes per environmental category (brackish pelagic and benthic, as well as pelagic deep-sea and hydrothermal plume were pooled).

https://doi.org/10.1038/s41564-023-01342-w



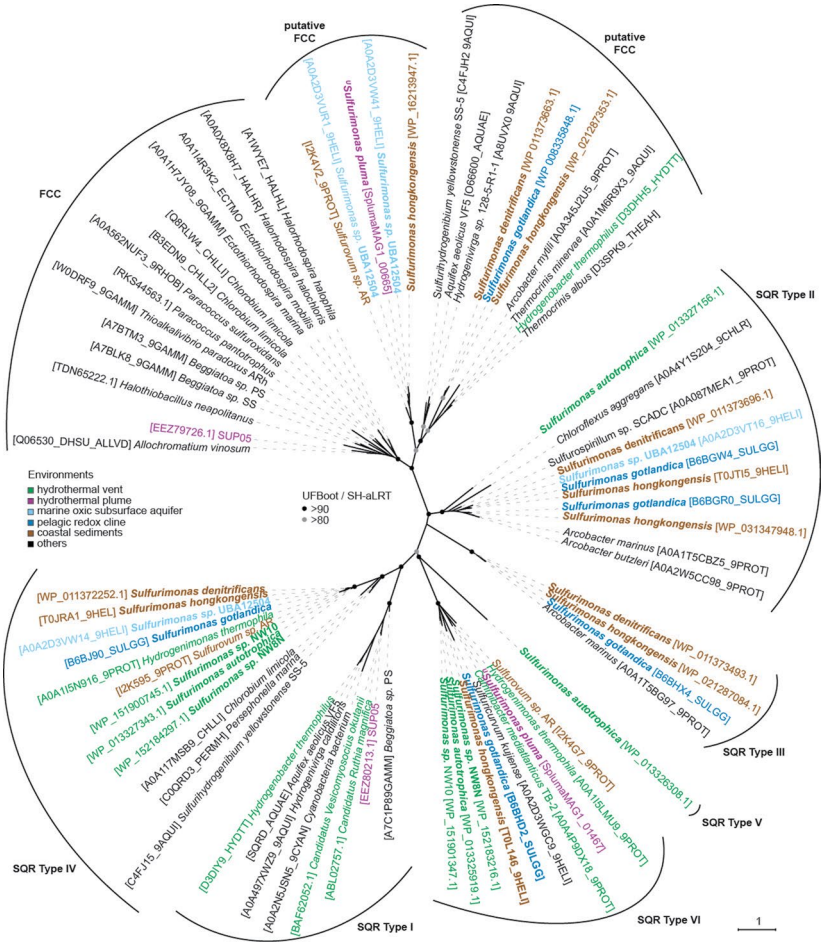
Extended Data Fig. 4 | Pangenomic analyses of hydrothermal plume ᵁ*Sulfurimonas pluma*. Pangenome analysis of *Sulfurimonas* isolate genomes and MAGs from different environments. Starting from the centre, the first 13 circles show the occurrence of gene clusters in a given *Sulfurimonas* genome. The external circle reports the gene clusters in which at least one gene was functionally annotated to COGs. The 7569 gene clusters (GCs) are organized based on their distribution across the genomes. The genomes are organized based on gene clusters they share (using Euclidian distance and Ward clustering). The barplots show some general features of each genome, the heatmap shows the ANI between genomes. **b**, Metapangenome analysis of *Sulfurimonas* isolate genomes (*S. denitrificans*, *S. hongkongebsis*, *S. gotlantica*, *S. autotrophica*) and hydrothermal plume ᵁ*S. pluma*. Starting from the centre: the first 5 circles show the occurrence of gene clusters in a given *Sulfurimonas* genomes; then the following 8 circles report the expression of gene clusters of ᵁ*S. pluma* in hydrothermal plume, seawater and fluid as inferred from metatransciptomic (MT) read recruitment (only representatives for available ridge systems are reported); finally, the external circles report the gene clusters in which at least one gene was functionally annotated (green and grey for known and unknown functions, respectively). The 4854 gene clusters (GCs) and the genomes are organized based on their distribution across the genomes and on gene clusters they share (using Euclidian distance and Ward clustering), respectively. The barplot shows the percentage of genes expressed, the heatmap shows the percentage of *Sulfurimonas* reads recruited by five genomes (inner circle) from metagenomes (MG; average value is reported for habitats having more than one metagenome available per ridge system; the number of metagenomes is reported in parenthesis). In both panels the dendrograms represents the hierarchical clustering of genomes based on the occurrence of gene clusters.

https://doi.org/10.1038/s41564-023-01342-w



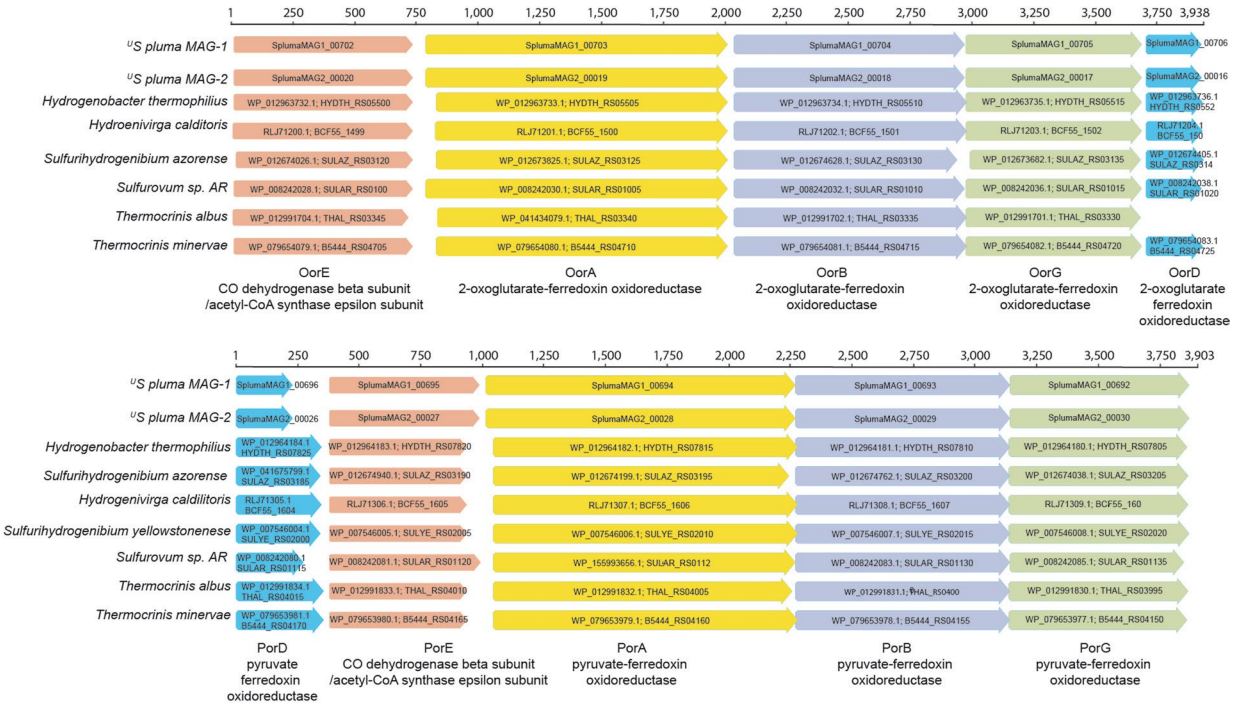
Extended Data Fig. 5 | Samples' location for oligotyping. World map showing the geographic origin, and environment of the samples used for the analysis of *Sulfurimonas* oligotypes. The point size is proportional to the number of samples collected at each location. HV: hydrothermal vent.

https://doi.org/10.1038/s41564-023-01342-w



Extended Data Fig. 6 | Synteny analysis of operon encoding the OOR and POR enzymes. Gene cluster organization of the five subunits of 2-oxaglutarate-ferrodoxin oxidoreductase and pyruvate-ferrodoxin oxidoreductase in the genomes of *[U]Sulfurimonas pluma*, *Sulfurovum AR*, and Aquificales' representatives.

https://doi.org/10.1038/s41564-023-01342-w



Extended Data Fig. 7 | Phylogenetic tree of sulfide oxidizing enzymes. Phylogenetic relationships between the beta subunit of Flavocytochrome c sulfide dehydrogenase (FCC) and Sulfide:quinone oxidoreductase (SQR). The scale bar represents the expected number of changes per amino acids position. UFBoot and SH-aLRT values are based on 1000 replicates. Best substitution model: LG + R4. *Sulfurimonas* species are in bold.

nature portfolio

Corresponding author(s): Massimiliano Molari

Last updated by author(s): Jan 10, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No softwares were used for data collection
Data analysis	R v 3.5.2; IQ-Tree v2.0; Anvi'o v6.2; iTOL v6; RAxML v8.2.4; ARB v6.0.6; bbduk v34; TRIMMOMATIC v0.35; SortMeRNA v2.0; phyloFlash v3.0 beta 1; cutadapt v1.9.1; PEAR v0.9.6; swarm v.2.2.2; decompose v2.1; MEGAHIT v1.1.2; SPAdes v3.9.0; bwa v0.7.12; SAMtools v1.5; CONCOCT v1.1.0; dRep v2.3.2; Prodigal v2.6.3; HMMER v3.2.1; DIAMOND v0.9.14; RNAmmer v1.2; bwa v0.7.12; SAMtools v1.5; CheckM v.1.2.1; ARAGORN v1.2.36; Prokka v. 1.11; kallisto v0.43.1; MUSCLE v3.8.1551; blastp v2.9.0+; Bowtie2 v2.3.2; CheckM2 v.0.1.3; BUSCO v.5.2.2; MMseq2 v13.45111; CAT v.5.2.3; bwa-mem2 v.2.2.1; HTSeq v.2.0.2; Galaxy platform (https://galaxyproject.org); https://github.com/edgraham/GhostKoalaParser; edgeR v3.24.3; Adobe Illustrator CS5; Vegan v.2.5.-6; ggplot2 3.2.1.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The sequences generated in this study and the USulfurimonas pluma genomes have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under Bioproject PRJEB48226. All the sequences were archived using the data brokerage service of the German Federation for Biological Data (GFBio).

Sulfurimonas V3-V4 16S rRNA gene sequences were extracted from metabarcoding studies obtained from ENA (https://www.ebi.ac.uk/ena/browser/home), and the studies accession numbers are reported in Supplementary Table 5. Sulfurimonas and Sulfuricurvum kujiense full-length 16S rRNA gene sequences used in this study were obtained from SILVA RefNR database (version 138; https://www.arb-silva.de) and NCBI Genebank (https://www.ncbi.nlm.nih.gov/nucleotide; JN874148.1 and JN874176.1). Partial Sulfurimonas 16S rRNA gene sequences (250–1400 bp) were retrieved from NCBI GenBank (https://www.ncbi.nlm.nih.gov/nucleotide; data accessed May 2020), and from metabarcoding studies obtained from ENA (https://www.ebi.ac.uk/ena/browser/home) and VAMPS (vamps.mbl.edu). Sequences and studies accession numbers are reported in Fig.1. Functional genes sequences used for phylogenetic tree analysis were obtained either from UniProt (https://www.uniprot.org/uniprotkb?query=*; data accessed May 2020) or NCBI GenBank (https://www.ncbi.nlm.nih.gov/nucleotide; data accessed May 2020), and their accessions are provided in the Figure 4 and Extended Data Figure 6. Sulfurimonas and Sulfuricurvum genomes used in this study are available via NCBI GenBank (https://www.ncbi.nlm.nih.gov/nucleotide; data accessed January 2020) and accessions are reported in Supplementary Table 6a. Metagenomic and metranscitptomic data used in metapangenomic analysis were obtained either from ENA (https://www.ebi.ac.uk/ena/browser/home; data accessed May 2020) or MG-RAST (https://www.mg-rast.org; data accessed May 2020), and study accession number is provided in Supplementary Table 6b.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Using omics' approaches, we compared the diversity and distribution of Sulfurimonas from Gakkel Ridge hydrothermal plumes with those from different Mid Ocean Ridges and other Sulfurimonas-hosting environments.
Research sample	Microbes studied in this study come form seawater samples collected in the valley of the South-West Indian Ridge (SWIR segment 10°–17°E; Atlantic sector), and at the Aurora and Polaris vent sites of the Gakkel Ridge (Central Arctic Ocean) during RV Polarstern expeditions PS81 (9th November – 16th December 2013), PS86 (7th July – 3rd August 2014), and PS101 (9th September – 23rd October 2016). Existing 16S rRNA and proteins sequences and genomes were retrieved from ENA, PubMed, SILVA SSU r138 RefNR, GeneBank, vamps.mbl.edu, UniProt, NCBI, MG-RAST.
Sampling strategy	Seawater samples were collected in the hydrothermal plume and in the surrounding waters: above the plume, below the plume, bottom water and background water (i.e. seawater without physico-chemical signatures for hydrothermal plume). Seawater samples were also collected at reference stations (i.e. not affected by hydrothermal plume) located 2 km (inside the ridge; "Internal Reference") and 56 km (outside the ridge; "External Reference") away from Aurora field, and 16 km (inside the ridge; "Internal Reference") and 190 km (outside the ridge; "External Reference") away from Polaris field.
Data collection	DNA and RNA were extracted in MPI laboratory using Powerwater DNA isolaiton Kit (Mo Bio) and mirVAna mRNA Isolation Kit (Ambion), respectively. RNA extracts were treated with the TURBO DNA-free Kit (Ambion) to remove co-extracted DNA, and purified and concentrated using RNeasy MinElute Kit (Qiagen), and only RNA with integrity number > 7 was used for sequencing. The hypervariable V3–V4 region of the bacterial 16S rRNA gene was amplified using bacterial primers S-D-Bact-0341-b-S-17 and S-D-Bact-0785-a-A-21. Sequences were obtained on the Illumina MiSeq platform in a 2 × 300 bp paired-end run with a number of reads per sample >50 000 (CeBiTec Bielefeld, Germany), following the standard instructions of the 16S Metagenomic Sequencing Library Preparation protocol (Illumina). For metagenomes paired-end libraries were prepared with the TruSeq DNA PCR-Free Sample Prep Kit (Illumina) and sequencing of libraries was performed on a MiSeq 2500 instrument (Illumina; 2 × 300 paired reads) using the v3 sequencing chemistry (CeBiTec laboratory). For metatranscriptomes the libraries were sequenced on a HiSeq1500 platform (Illumina), in a 1×150 bases single-end run and with a total number of reads per sample > 20 million.
Timing and spatial scale	Deep-sea seawater samples were collected in the summer and/or early autumn in the Arctic and Antarctic regions, as this is the time of year when there is the lowest sea ice coverage and better marine weather conditions, allowing for ship navigation and collection of deep water samples. Samples were collected based on the presence/absence of hydrothermal plume chemical-physical signal (at scale of tens of hundreds of metres), and at reference stations (from kilometres to hundreds of kilometres).
Data exclusions	No data were excluded from the analyses
Reproducibility	"Methods" section contains all information to reproduce this work, including: location of sampling sites and devices, type and number of seawater samples, protocols for sample conservation, DNA and RNA extractions and sequencing, detailed description of workflow applied for data analysis and all the softwares, and references and/or accession numbers for sequences downloaded from public databases.
Randomization	In this comparative study of Arctic Ocean and global deep-sea microbiota, we identified a new aerobic Sulfurimonas species inhabiting neutrally buoyant hydrothermal plumes and fluids, and subsurface aquifers. Therefore randomization is not applicable to this study.
Blinding	It is not applicable here, as this study does not include clinical trials.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	Seawater samples were collected at water depth > 2000 m, and the parameters mostly affecting the samples were those related to presence of hydrothermal plumes (i.e. anomalies in temperature, turbidity and redox). Sampling activities were mostly affected by marine weather conditions at South Atlantic Ocean, and by sea-ice drift in the Arctic Central Ocean.
Location	Seawater samples were collected at South-West Indian ridge (LAT. from -52.646 to -52.233, LONG. from 12.526 to 15.738; water depth 2500-4320 m) and at the Gakkel Ridge in Aurora (LAT. from 82.896 to 83.105, LONG. -6.377 to -2.466; water depth 2000-3955 m) and Polaris (LAT. from 85.292 to 86.984, LONG. from 55.575 to 60.184; water depth 2051-4841 m) hydrothermal fields.
Access & import/export	All samples obtained are from the High Seas, except the samples from Auora from Greenland EEZ for wich diplomatic permission was obtained. No further permission was needed to collect and use the samples. The research vessel complies with all environmental standards in polar seas, and the participants sign a compliance clause for sustainable sampling and data management. There were no specific import clauses for the water samples obtained.
Disturbance	Sampling activities included exclusively in situ physiochemical measurements by sensors, collection of seawater (tens of litres) and in situ seawater filtration (hundreds of litres), therefore causing only limited disturbance to the environment and organisms.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging