

RESEARCH

Open Access



# Autotrophic biofilms sustained by deeply sourced groundwater host diverse bacteria implicated in sulfur and hydrogen metabolism

Luis E. Valentin-Alvarado<sup>1,2</sup>, Sirine C. Fakra<sup>3</sup>, Alexander J. Probst<sup>4,5</sup>, Jonathan R. Giska<sup>4,6</sup>, Alexander L. Jaffe<sup>1</sup>, Luke M. Oltrogge<sup>7,8</sup>, Jacob West-Roberts<sup>9</sup>, Joel Rowland<sup>4,10</sup>, Michael Manga<sup>4,11</sup>, David F. Savage<sup>2,7,8</sup>, Chris Greening<sup>12</sup>, Brett J. Baker<sup>13,14</sup> and Jillian F. Banfield<sup>2,4,9,14,15\*</sup>

## Abstract

**Background** Biofilms in sulfide-rich springs present intricate microbial communities that play pivotal roles in biogeochemical cycling. We studied chemoautotrophically based biofilms that host diverse CPR bacteria and grow in sulfide-rich springs to investigate microbial controls on biogeochemical cycling.

**Results** Sulfide springs biofilms were investigated using bulk geochemical analysis, genome-resolved metagenomics, and scanning transmission X-ray microscopy (STXM) at room temperature and 87 K. Chemolithotrophic sulfur-oxidizing bacteria, including *Thiothrix* and *Beggiatoa*, dominate the biofilms, which also contain CPR Gracilibacteria, Absconditabacteria, Saccharibacteria, Peregrinibacteria, Berkelbacteria, Microgenomates, and Parcubacteria. STXM imaging revealed ultra-small cells near the surfaces of filamentous bacteria that may be CPR bacterial episymbionts. STXM and NEXAFS spectroscopy at carbon K and sulfur L<sub>2,3</sub> edges show that filamentous bacteria contain protein-encapsulated spherical elemental sulfur granules, indicating that they are sulfur oxidizers, likely *Thiothrix*. Berkelbacteria and Moranbacteria in the same biofilm sample are predicted to have a novel electron bifurcating group 3b [NiFe]-hydrogenase, putatively a sulfhydrogenase, potentially linked to sulfur metabolism via redox cofactors. This complex could potentially contribute to symbioses, for example, with sulfur-oxidizing bacteria such as *Thiothrix* that is based on cryptic sulfur cycling. One Doudnabacteria genome encodes adjacent sulfur dioxygenase and rhodanese genes that may convert thiosulfate to sulfite. We find similar conserved genomic architecture associated with CPR bacteria from other sulfur-rich subsurface ecosystems.

**Conclusions** Our combined metagenomic, geochemical, spectromicroscopic, and structural bioinformatics analyses of biofilms growing in sulfide-rich springs revealed consortia that contain CPR bacteria and sulfur-oxidizing Proteobacteria, including *Thiothrix*, and bacteria from a new family within Beggiatoales. We infer roles for CPR bacteria in sulfur and hydrogen cycling.

**Keywords** Candidate phyla radiation, Groundwater microbiome, Synchrotron-based spectromicroscopy

\*Correspondence:

Jillian F. Banfield

jbanfield@berkeley.edu

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Sulfur is the fifth most abundant element on earth, and the sulfur cycle is a key component of Earth's interlinked biogeochemical cycles [1, 2]. In natural ecosystems, sulfur exists in several oxidation states,  $-2$ ,  $0$ ,  $+2$ ,  $+4$ , and  $+6$  being the most common, in the forms of polysulfide ( $\text{HS}_x$  or  $\text{S}_x^{2-}$ ;  $-2$ ,  $0$ ), thiosulfate ( $\text{S}_2\text{O}_3^{2-}$ ;  $-1$ ,  $+5$ ), tetrathionate ( $\text{S}_4\text{O}_6^{2-}$ ;  $-2$ ,  $+6$ ), sulfite ( $\text{SO}_3^{2-}$ ;  $+4$ ), and sulfate ( $\text{SO}_4^{2-}$ ;  $+6$ ). Microbes play an important role in sulfur cycling in aqueous and soil environments.  $\text{H}_2\text{S}$  is a toxic compound that must be maintained at low levels for the sustained growth of microbial consortia; thus, microbial sulfide oxidation is beneficial at the community level.

Sulfide ( $\text{S}^{2-}$ ) is common in natural springs and can serve as a source of energy and reducing power for chemolithoautotrophic microorganisms. Chemolithoautotrophic microbial communities with members that carry out the oxidation, reduction, and disproportionation of sulfur compounds are found in environments such as hydrothermal vents [3, 4], water column oxic/anoxic interfaces [5–7], terrestrial caves [8–10], groundwater [11, 12], and activated sludges [13]. Beggiatoaceae and Thiotrichaceae that have been cultivated have been shown to use hydrogen sulfide either mixotrophically or heterotrophically [14–17]. *Beggiatoa* spp. are gliding filamentous bacteria that form intracellular spherical  $\text{S}^0$  granules that may oxidize to sulfate when  $\text{H}_2\text{S}$  supply becomes limited [18]. *Thiothrix* spp. are gliding bacteria that can grow as long multicellular filaments (cells in a microtubular sheath), can form rosettes, and are known to accumulate intracellular spherical  $\text{S}^0$  granules when in the presence of reduced sulfur compounds [13, 19] and organics or  $\text{CO}_2$  (carbon and energy source) [14]. Prior work [20–24] indicates that sulfur-oxidizing bacteria support communities by providing resources such as fixed carbon and nitrogen.

To date, most studies of sulfur-based chemoautotrophic ecosystems have investigated the roles of the relatively most abundant organisms. However, it is well understood that microbial biofilms are structured as networks of interacting organisms, some of which are fundamentally dependent on other community members. Of particular interest are candidate phyla radiation (CPR) bacteria (also known as Patescibacteria) [25–28] that can form symbioses with host organisms [29–31]. Prior surveys have documented CPR bacteria in sulfur-based communities [25, 32, 33], yet the nature of CPR-host relationships and the roles of CPR in sulfur-based communities remain under-explored.

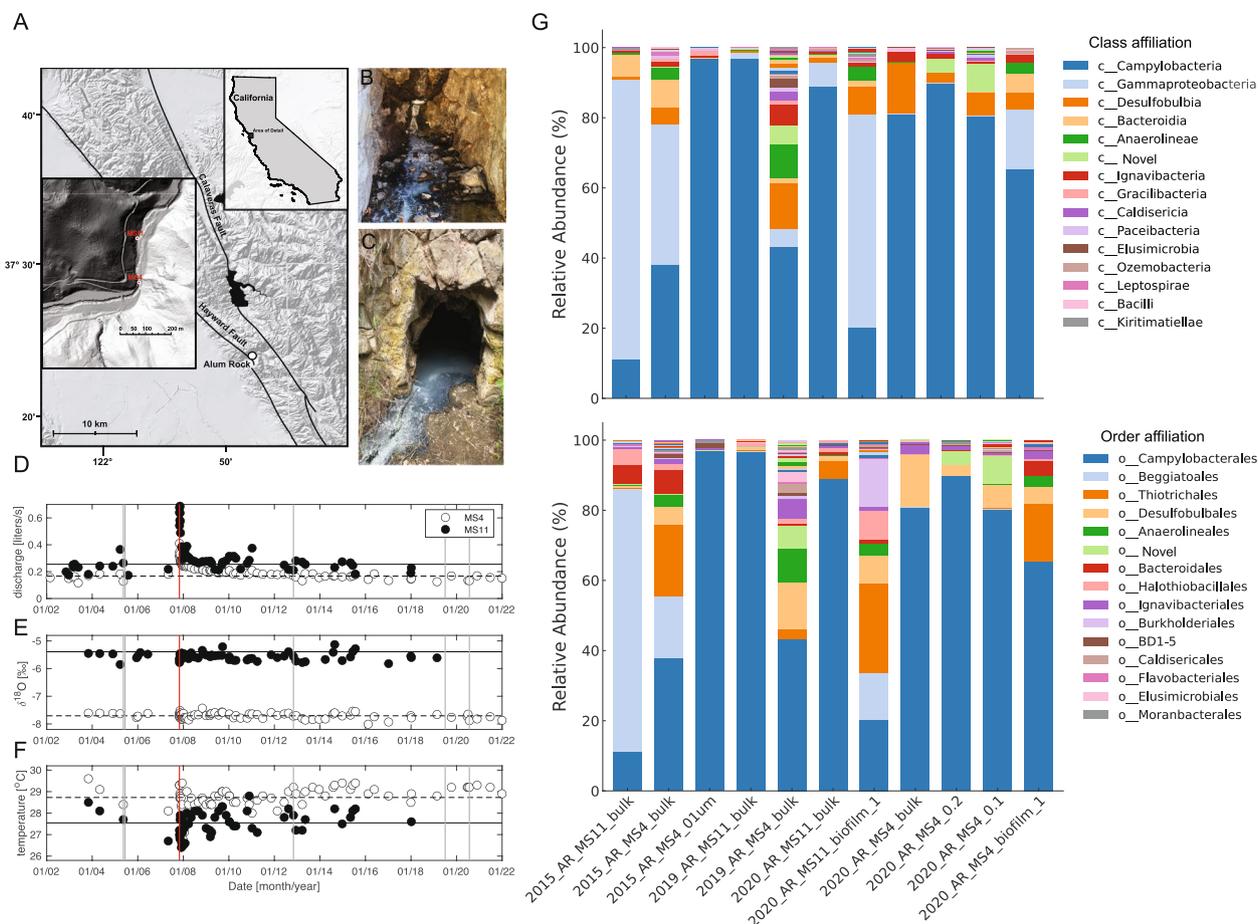
Here, we studied chemoautotrophic microbial communities sustained by sulfur metabolism in two mineral springs MS4 and MS11 [34] at Alum Rock Park, CA, USA, where sulfide-rich groundwater discharges along

the Hayward fault. We profiled oxygen isotopes, temperature, water composition, and spring discharge rates to constrain the sources of water. We then further combined genome-resolved metagenomics with X-ray spectromicroscopy and scanning electron microscopy to investigate metabolic capacities, interdependencies, and structure of the microbial biofilm community at these two springs. Synchrotron-based spectromicroscopy revealed a close association between ultra-small cells and sulfur-oxidizing bacteria. The possibility that these bacteria are CPR might indicate the existence of a cryptic chemoautotrophic ecosystem. We predict the contributions of the major community members to carbon, nitrogen, sulfur, and hydrogen cycling and investigate the potential roles of the abundant and diverse CPR bacteria in these consortia.

## Materials and methods

### Site description and microbial biomass collection

The spring system is located along Penitencia Creek in Alum Rock Park, San Jose, CA, USA ( $37^\circ 23' 57.7'' \text{N}$ ,  $121^\circ 47' 48.8'' \text{W}$ ) (Fig. 1A). The two sample sites, Mineral Springs 4 and 11 (MS4 and MS11), are located on opposite sides of the creek approximately 250 m from one another (Fig. 1B, C). Samples for geochemical analyses were taken in May 2005, during the dry season, and were filtered on-site using sterile  $0.2\text{-}\mu\text{m}$  filters. Biofilm samples for scanning electron microscopy were collected from both sites using sterile pipettes. Samples were then transported back to the laboratory on ice, and solutions for cation analyses were acidified with 3% nitric acid. Biofilm samples for metagenomic sequencing were collected on June 13, 2015, July 2, 2019, and July 24, 2020. Planktonic samples were collected on June 10, 2015 and July 24, 2020. Two sets of planktonic samples were taken by sequentially filtering 379 L and 208 L of water, respectively, from the MS4 spring onto  $0.65\text{-}\mu\text{m}$  and  $0.1\text{-}\mu\text{m}$  large volume filters (GraverTech 5-inch ZTEC-G filter). Filters were frozen on dry ice at the site and stored at  $-80^\circ \text{C}$  for genome-resolved metagenomic analyses. For synchrotron-based measurements (STXM and X-ray microprobe), thin white streamers were collected on June 13, 2015, with sterile tweezers at both sites and flash-frozen on site. A  $2\text{-}\mu\text{L}$  droplet of biofilm sample was deposited using a sterile pipette onto either a 3-mm diameter  $\text{Si}_3\text{N}_4$  window (SiMPore) or a TEM Cu grid (300 mesh, lacey carbon coated formvar, Ted Pella Inc.). The grids were manually blotted with filter paper (Grade 1 filter paper, Whatman®) and immediately plunged into liquid nitrogen using a portable  $\text{LN}_2$  plunger, and gas ethane (used for cryo-plunging) was not available at the time of sampling. Samples were not rinsed or spinned so as to preserve the structural integrity of the filaments and



**Fig. 1** **A** Shaded relief map showing the location of Alum Rock springs, CA, USA. Insets show the location of Alum Rock and of the MS4 and MS11 springs. Photographs of **(B)** MS4 and **(C)** MS11 biofilms. Thin white streamers (5–10 cm) are mostly found attached to the surfaces of rocks. Hydrogeological properties **(D)** discharge, **(E)**  $\delta^{18}\text{O}$ , and **(F)** temperature are steady over periods greater than a decade, except following large regional earthquakes. A discharge increase in late 2007 followed a magnitude 5.6 earthquake with an epicenter 4 km from the springs (vertical red line), neither  $\delta^{18}\text{O}$  nor the temperature changed indicating that fluid sources did not change. The horizontal lines show averages of plotted quantities over the entire sampling period, except discharge for which the average excludes the first 2 years after the earthquake. Vertical gray lines show dates of biofilm and planktonic sampling. **G** Microbial community composition at the class and order levels, respectively, highlighting the top 15 most abundant groups in each category. Each bar represents a sample collected from different biotopes (bulk/biofilm, 0.1- $\mu\text{m}$  filter) in the MS4 and MS11 springs over several years (2015, 2019, and 2020). The stacked bar plots illustrate the relative abundance of each microbial group, with each color corresponding to a different group, from top to bottom in decreasing order of overall abundance across all samples

preserve the CPR bacteria-bacteria-filaments spatial relationships. Samples were stored in a  $\text{LN}_2$  storage dewar (Taylor-Wharton, 34 L) until measurements.

**Geochemical analysis**

Water discharge (volume/time) was measured by diverting water into either a bucket or graduated cylinder to measure volume, and time was recorded with a stopwatch. Temperature was measured with a type K thermocouple until February 2008 and thereafter with a thermistor. Accuracy is 0.2 °C and 0.1 °C, respectively. Water for O and H isotope measurements was collected in 250-mL Nalgene bottles. Discharge and temperature

were not measured if outflow channels from the springs backed up to create pools of water. Cation analysis was performed on a PerkinElmer 5300 DV optical emission ICP with autosampler. Anion analysis was performed on-site using a HACH DR2010 spectrophotometer with protocols provided by the manufacturer. O and H isotopes were measured with a GV IsoPrime gas source mass spectrometer, with analytical precision of approximately 0.1 and 1 per mil, respectively.

**Scanning electron and confocal fluorescence microscopy**

Biofilm samples for scanning electron microscopy were fixed for 2 h in a 2% glutaraldehyde solution (in 0.1-M

sodium cacodylate buffer) according to a standard protocol, then vacuum aspirated onto 0.22- $\mu\text{m}$  polycarbonate filters (Osmonics, poretics, 47 mm, catalog number K02CP04700), and rinsed three times in 0.1-M sodium cacodylate buffer. The samples were then dehydrated in successive ethanol baths of increasing concentration and dried using a Tousimis Autosamdri 815 Critical Point Dryer for approximately 1 h. Specimens were mounted on gold stubs and sputter coated with a gold/palladium mix. Imaging was performed on a Hitachi S-5000 scanning electron microscope at 10 keV at UC Berkeley.

Biofilm samples for confocal fluorescence microscopy were prepared with fluorescence dyes. Cell membranes and nucleic acids in the MS4 and MS11 biofilms were stained by adding simultaneously the lipophilic dye F4-64 (20  $\mu\text{g}/\text{mL}$ ) (Thermo Fisher, Grand Island, NY, USA) and SYTOX Blue (20  $\mu\text{g}/\text{mL}$ ) (Thermo Fisher, Grand Island, NY, USA), respectively. Samples were incubated at room temperature for 20 min, and then a 10- $\mu\text{L}$  aliquot of biofilm was deposited onto a glass slide. Samples were imaged using a Leica Stellaris 5 confocal fluorescence microscope. Images were acquired using a 63 $\times$  oil immersion objective with laser excitation for SYTOX Blue (480 nm) and FM4-64 (520 nm), keeping the same parameters for both biofilms. Unmixed images were combined and false colored using the Leica Application Suite X (LAS X). All data was collected at the Innovative Genomics Institute, UC Berkeley.

### Scanning transmission X-ray microscopy (STXM)

STXM and near-edge X-ray absorption fine structure (NEXAFS) spectroscopy measurements were performed on the soft X-ray undulator beamline 11.0.2 [35] of the Advanced Light Source (ALS), Berkeley, CA, USA. Data were recorded with the storage ring operating in top-off mode at 500 mA, 1.9 GeV. Frozen samples were thawed right before STXM-NEXAFS measurements at ambient temperature under He at pressure < 1 atm. A Fresnel zone plate lens (40-nm outer zones) was used to focus a monochromatic soft X-ray beam onto the sample. The sample was raster scanned in 2D through the fixed beam, and transmitted photons were detected with a phosphor scintillator-photomultiplier assembly; incident photon counts were kept below 10 MHz. The imaging contrast relies on the excitation of core electrons by X-ray absorption [36–38]. STXM images recorded at energies just below and at the elemental absorption edge (S  $L_3$  and C K) were converted into optical density (OD) images where the OD for a given energy can be expressed from the Beer-Lambert law as  $OD = -\ln(I/I_0) = \mu \rho t$ , where  $I$ ,  $I_0$ ,  $\mu$ ,  $\rho$ , and  $t$  are the transmitted intensity through the sample, incident intensity, mass absorption coefficient, density, and sample thickness, respectively. Protein,

carbon, and elemental sulfur maps were obtained by taking the difference of OD images at 280 and 288.2 eV, at 280 and 305 eV, and at 162 and 163.9 eV, respectively. Image sequences (“stacks”) recorded at energies spanning the S  $L_{2,3}$  edges (160–180 eV) with steps of 0.3 eV around the  $L_3$ -edge and C K-edge (280–305 eV) with steps of 0.12 eV around the K-edge were used to obtain NEXAFS spectra from specific regions. S  $L_{2,3}$  edges NEXAFS spectra are affected by spin-orbit coupling (multiplet structure) and provide information on the oxidation state of sulfur. Beam-induced radiation damage was carefully checked.

Additionally, STXM-NEXAFS measurements at 87 K were performed on frozen-hydrated samples deposited on  $\text{Si}_3\text{N}_4$  windows so as to preserve sample morphology and chemical integrity [39] and minimize beam-induced radiation damage [40]. We made sure to analyze exclusively fully frozen-hydrated regions of the samples by first imaging the entire window. The samples were cryo-transferred through a specimen chamber (< 100 mTorr) into an  $\text{LN}_2$ -cooled stage (87 K) inside the STXM operated with a scanning Fresnel zone plate lens (60-nm outer zones), under vacuum ( $10^{-6}$  Torr). With this setup, the sample is not rastered-scanned through the fixed beam so as to minimize sample vibrations; instead, the zone plate is scanned in 2D. Note that sulfur  $L_{2,3}$  edges could not be accessed with this setup due to geometrical constraints.

At least two different sample regions were analyzed at each elemental edge. The theoretical spectral and spatial resolutions during measurements were  $\pm 100$  meV: 40 nm (at room temperature) and 60 nm (at 87 K) respectively. The photon energy was calibrated at the C K-edge using the Rydberg transition of gaseous  $\text{CO}_2$  at 292.74 eV (C 1 s  $\rightarrow$  3 s ( $\nu=0$ )). Sulfur spectra were calibrated using the S  $2p_{3/2}$  edge of elemental sulfur set at 163.9 eV. An elemental sulfur standard spectrum was kindly provided by Geraldine Sarret [41] (Universite Grenoble Alpes). A lipid standard compound (PE) was kindly provided by Susan Glasauer (University of Guelph). Carbon spectra were pre-edge background subtracted using a linear function and normalized at 300 eV. All data was processed with the aXis2000 software version 06 Jul 2021 (<http://unicorn.mcmaster.ca/aXis2000.html>). Sulfur spectra were further pre-edge background subtracted and post-edge normalized using Athena version 0.9.26 (Demeter package [42]).

### X-ray fluorescence microprobe (XFM)

Synchrotron XFM measurements were performed in cryogenic conditions (95 K) at ALS XFM beamline 10.3.2 [43], with the storage ring operating in top-off mode at 500 mA, 1.9 GeV. Micro-focused X-ray fluorescence ( $\mu\text{XRF}$ ) elemental mapping was performed on

flash-frozen hydrated samples oriented at 45° to the incident X-ray beam; frozen biofilm samples were mounted onto a TEM cartridge in a liquid nitrogen bath and cryo-transferred into a LN<sub>2</sub>-cooled apparatus following methods described elsewhere [44]. All data were recorded using a single-element XR-100 silicon drift detector (Amptek, Be window).

XRF maps were recorded at 4138 eV (100 eV above the Ca K-edge) using a beam spot size of 3 μm × 4 μm, 2 × 2 μm pixel size, and 70-ms dwell time/pixel. Micro-XRF spectra were recorded simultaneously on each pixel of the maps. All maps were then deadtime corrected and decontaminated using custom LabVIEW 2018 (National Instruments, Austin, TX, USA) software available at the beamline. Maps were then processed using a custom MATLAB R2020b program (MathWorks, Natick, MA, USA) available at the beamline.

#### DNA extraction and metagenomic sequencing

Approximately, 200 μl of biofilm was extracted using MoBio PowerSoil DNA extraction kit (MoBio Laboratories, Inc., CA, USA) according to the manufacturer's protocol, with the bead-beating time reduced to less than 1 min. This DNA extract was then gel purified and quantified using a low-mass ladder (Promega).

Total genomic DNA for metagenomic sequencing (150-bp or 250-bp reads) for both biofilm and planktonic samples (20% of each filter) was extracted using MoBio PowerMax Soil DNA extraction kit. Cells were extracted from 20% of each filter by adding 15 ml of lysis buffer and vortexing for 10 min. Lysis of cells was modified by heating to 65 °C for 30 min and 1 min of bead beating. DNA was eluted in Milli-Q water, and ethanol precipitation was performed (70% EtOH, 3-M sodium acetate, incubation for 24 h at 4 °C).

#### Illumina sequencing, assembly, binning and sequence curation

Shotgun genomic reads were assembled using IDBA-UD [45]. Contigs larger than 2.5 kb were retained, and sequencing reads from all samples were mapped against each resulting assembly utilizing Bowtie 2 [46]. Differential coverage profiles, filtered with a 95% read identity threshold, were then used for genome binning using a suite of binning tools (MetaBAT2 [47], VAMB [48], MaxBin2 [49], Abawaca (<https://github.com/CK7/abawaca>), with the final bin choice determined by DAS Tool [50]. Draft genomes consisting of scaffolds ≥ 1 kbp in length were binned using ggKbase manual binning tools based on a combination of GC content, coverage, single-copy gene content, phylogenetic profile, and patterns of organism abundance over samples. The phylogenetic profile was established using representative genomes from the

GTDB database. In some cases, scaffold sequences from groups of bins were used to construct emergent self-organizing maps in which the structure was established using tetranucleotide composition (tetra-ESOMs). For scaffolds > 6 kb, scaffolds were subdivided into 3-kb segments and treated separately in the ESOM analysis. In cases where the majority of segments from the same scaffold did not group together in the ESOM, the scaffolds were evaluated manually (based on gene content and other information) to resolve their placement or assign them to unbinned. The scaffold set defined based on ESOM analysis was then used to generate a draft genome bin that was again checked for consistent binning signals (as above). As ESOMs only used scaffolds > 3 kb in length, scaffolds from the original bins were added if they had a tightly defined GC, coverage, and the expected phylogenetic profile. CheckM2 [51] was used for estimation of genome completeness, strain heterogeneity, and contamination. Curated genomes with less than 5 duplicated single-copy genes (some of which occur because genes are split at scaffold ends) and with ≥ 95% of the expected single-copy marker gene were used for completeness estimation: at least 42 of 51 single-copy genes used for preliminary bin evaluation in ggKbase were classified as near complete. Genomes with > 5 duplicated single-copy genes were classified as partial, regardless of other indicators of bin completeness.

#### Phylogenetic analyses

The concatenated ribosomal protein tree was generated using 16 syntenic genes that have been shown to undergo limited lateral gene transfer (rpL2, 3, 4, 5, 6, 14, 15, 16, 18, 22, 24 and rpS3, 8, 10, 17, 19) [52]. We obtained branch support with the ultrafast bootstrap [53] implemented in iQ-TREE v1.6.12 [54] with the following parameters: –bb 1000 –m LG+F+G4. Trees were visualized using iTOL v6.3.2 [55]. Amino acid alignments of the individual ribosomal proteins were generated using MAFFT v7.304 [56] and trimmed using trimAL [57] with the following setting: –gt 0.1.

To verify the presence of biogeochemically relevant genes, phylogenetic trees were constructed. We used gene markers for sulfur (DsrAB, Pdo), carbon metabolism (RuBisCO), and energy conservation ([NiFe]-hydrogenases). Sequences were obtained using GOOSOS and aligned using MAFFT v7.304. All other phylogenies were generated using iQ-TREE v1.6.12 using the ultrafast bootstrap and parameters specified previously.

Hydrogenase sequences from Alum Rock genomes were obtained using HMMs [58]. The phylogenetic classification was performed using reference sequences obtained from [58] and HydDB [59]. Verification of hydrogenase loci was performed via inspection of nearby

genes and the presence of required hydrogenase accessory genes. Genome context diagrams were generated using Clinker [60].

### Metagenomics metabolic pathways analysis

Preliminary functional annotations were established using METABOLIC [61], and collections of metabolic capacities in genome bins were overviewed using ggK-base tools [62]. In addition, metabolic profiling was done by mapping ORFs to KEGG ortholog groups (KOs) using an HMM database that was compiled as previously described [63]. This HMM database was used to scan the metagenomic bins, and ORFs were assigned the KO of the best-scoring HMM, providing it was above the noise threshold. In addition, we profiled metabolic capacities with KEGG functional annotation using METABOLIC [61].

### Protein structure prediction

Protein structures were predicted for the putative complexes of the nitrate reductase (Nrx), dioxygenase/rhodanase, and group 3b [NiFe]-hydrogenase using AlphaFold2 in multimer mode. Specifically, for the 3b [NiFe]-hydrogenase complexes, AlphaFold2 was used in multimer mode for the HyhL (hydrogenase large subunit), HyhS (hydrogenase small subunit), HyhG (diaphorase catalytic subunit), and HyhB (diaphorase electron transfer subunit) [64, 65]. In all cases, the average per residue confidence scores (pLDDT) exceeded 90, a level that is empirically shown to produce highly accurate local structural models. The best-scoring models were aligned to related protein complexes in PyMol.

## Results

### Groundwater of mixed origin hosts biofilms dominated by filamentous bacteria

We measured the flow rate, pH, and concentrations of ionic species (Supplementary Table S1) in the MS4 and MS11 groundwater. The MS11 spring has a higher flow rate, ionic strength, alkalinity, and sulfide levels than the MS4 spring. H and O stable isotope compositions of the waters, combined with salinity measurements, indicate that spring waters are mixtures of meteoric input and pore waters from the host Miocene Monterey Group shales and cherts and possibly deeper Cretaceous sediments of the Great Valley Group. MS4 water is more diluted by meteoric input than MS11. Long-term monitoring of these two springs shows they experience small seasonal fluctuations in temperature, and that they are generally hydrologically and geochemically stable (Fig. 1D–F). Water temperatures of 27–29 °C are well above the mean annual surface temperature of 15.1 °C. The salinity of the springs is 1.8 and 2.3‰ for MS4 and

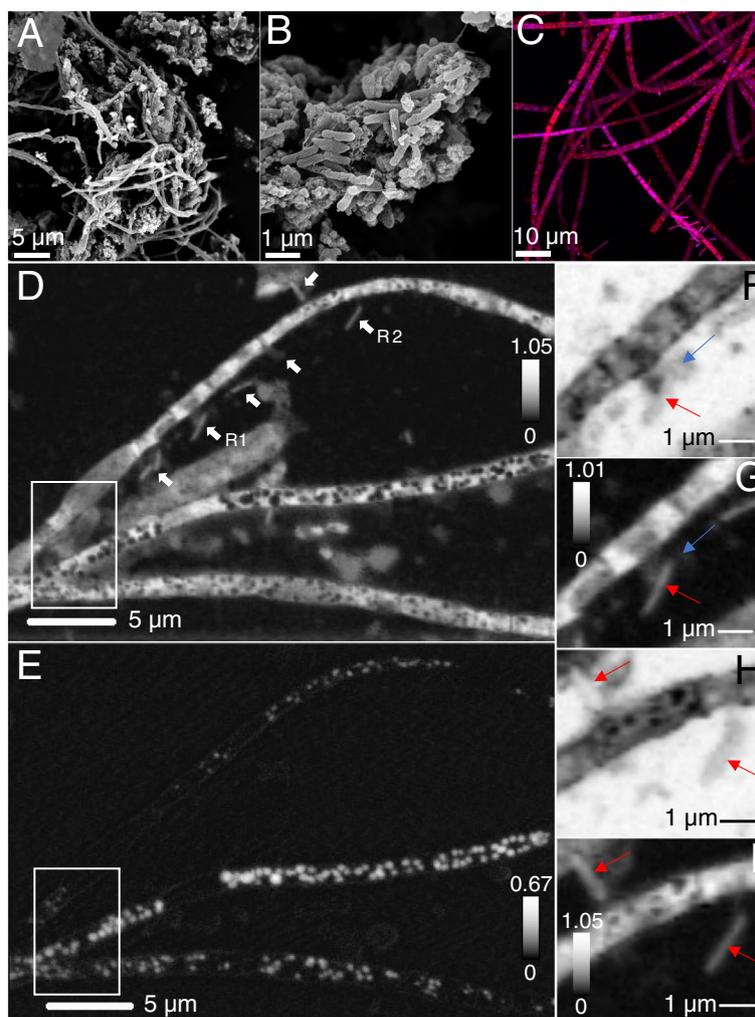
MS11, respectively. The sulfide levels (within the zone of oxygenation) range up to ~9 and 69 μmol/L at MS4 and MS11, respectively.

The biofilms at both MS4 and MS11 sites (Fig. 2A, B) are mainly composed of thin white streamers (~5–10 cm long) that are primarily attached to rocks. Scanning electron microscopy (SEM), confocal fluorescence microscopy (CFM), and scanning transmission X-ray microscopy (STXM) revealed that MS4 biofilms consist of filaments and cells distributed among the filaments (Fig. 2), whereas the MS11 biofilm consists mainly of filamentous bacteria (Fig. S2).

### Filamentous bacteria have encapsulated elemental sulfur granules and episymbionts

Sulfur μXRF distribution maps at 95 K evidenced the presence of sulfur across MS4 filamentous bacteria (Fig. S1). STXM sulfur maps (Fig. 2) and S L<sub>2,3</sub> NEXAFS spectra showed that the filaments contain spherical S<sup>0</sup> granules (Fig. S3) encapsulated in protein-rich compartments (Fig. 3A, B). The spherical granules in the middle filament (Fig. 2E) are roughly 380-nm average diameter, as estimated from 76 granules. The width of these MS4 filaments is < 1.6 μm. Filaments exhibit septa and longitudinal cell envelopes as observed by STXM (Fig. S4) and CFM (Fig. 2). Rod-shaped, curved-shaped, and few coccoid-shaped cells were found near the filaments in MS4 biofilms, as well as ultra-small cells (Fig. 2, Fig. S4D), often found within extracellular polymeric substances (EPS). Higher-resolution protein maps of MS4 and MS11 filaments (Fig. 3) suggest that sulfur granules are surrounded by proteins, as further confirmed on maps recorded at 87 K (Fig. 4).

Carbon K-edge NEXAFS spectra at 87 K of filamentous bacteria in MS4 and MS11 biofilms (Fig. 4) exhibit similar spectra (Supplementary Table S2), with a major peak at 288.2 eV (amide carbonyl groups in proteins [66]), a peak at 285.2 eV attributed mainly to aromatic groups in proteins, and peaks at 286, 286.6, 287.4, and 289.4 eV that can be attributed to nucleic acids [67–69] (Fig. S5). One prior report [69] suggests that spectra of nucleotide bases with peaks occurring within 284.7–286.9 eV are likely associated with π\*(C=C), whereas peaks in the 287.3–287.4 eV range are likely associated with π\*(C=N). Resonances are more defined than in prior studies at room temperature [37, 44, 70, 71], due to reduced Debye–Waller thermal disorder at low temperatures. This in turn allows us to unequivocally detect the presence of nucleic acids in the filaments, in addition to proteins. Copious extracellular polymeric substances surrounding MS4 and MS11 filaments exhibit a main peak at 288.7 eV, which can be attributed to carboxyl groups in acidic polysaccharides

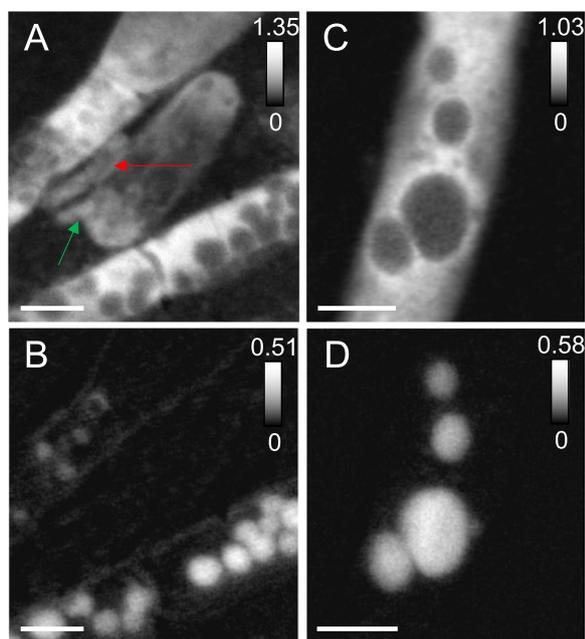


**Fig. 2** Microscopic characterization of MS4 biofilms. **A–B** Scanning electron microscopy of filamentous bacteria and associated cells, small cells are pointed by arrows **(C)** confocal fluorescence microscopy of cells treated with SYTOX (blue) for nucleic acid and F-64 (red) for membrane. Scanning transmission x-ray microscopy: **(D)** carbon map of filaments and associated cells (white arrows). **E** Corresponding distribution map of S<sup>0</sup> evidencing spherical elemental sulfur granules within the compartments of the filaments. The top, middle, and bottom filament widths are  $1.23 \pm 0.5 \mu\text{m}$ ,  $1.01 \pm 0.2 \mu\text{m}$ , and  $1.33 \pm 0.3 \mu\text{m}$ , respectively. **F** An ultra-small cell  $\sim 480 \text{ nm}$  long,  $\sim 270 \text{ nm}$  wide, (blue arrow) in contact with an apparently episymbiotic cell  $1.86 \pm 0.1 \mu\text{m}$  long,  $\sim 360 \text{ nm}$  wide (red arrow), imaged at 280 eV (region R1, **D**) and corresponding **(G)** carbon map. **H** Two apparently episymbiotic cells (red arrows) connected to filaments, imaged at 280 eV (region R2, **D**) and corresponding **(I)** carbon map. The intensity scales correspond to optical density

(Supplementary Table S2). The spectra of MS11 cells are also similar to those of filamentous bacteria, with a major peak at 288.2 eV (amide bonds) and other peaks associated with nucleic acids. The spectrum of a surface-attached cell on an MS4 filament exhibits a major peak at 288.2 eV (amide bonds) and nucleic acid-associated peaks; the 288.2 and 285.2 eV peaks are broader likely due to the presence of EPS surrounding this cell. Cells, filaments, and EPS all exhibited a shifted carbonate peak at 290.7 eV that corresponds to either organic

carbonates or carbonate minerals [72] and originates mainly from dissolved carbonates and carbonate precipitates present in the groundwater at circumneutral pH (Supplementary Table S1 and S2)

Strikingly, small cells were found along the surfaces of the filaments in MS4 (Fig. 2F, G, Fig. S4) and MS11 (Fig. S4A) biofilms; these cells are typically about 480 nm long, 250 nm wide, as estimated from STXM images. Ultra-small cells ( $290 \pm 20 \text{ nm}$  long,  $120 \pm 15 \text{ nm}$  wide) were also found in close proximity to the filaments in MS4 (Fig. S4D).



**Fig. 3** Scanning transmission X-ray microscopy of MS4 and MS11 biofilms. **A** Protein map and corresponding **(B)** distribution map of  $S^0$  in MS4 biofilms (in white boxed area of Fig. 2). Cells that are  $908 \pm 32$  nm long,  $370 \pm 30$  nm wide (red arrow),  $687 \pm 34$  nm long,  $244 \pm 33$  nm wide (green arrow), seen in close contact with filaments. **C** Protein map and corresponding **(D)** distribution map of  $S^0$  in MS11 biofilms, showing the presence of sulfur granules (up to  $1.08 \pm 0.12$   $\mu$ m in diameter) in a small area of a long filament. Sulfur  $L_{2,3}$  edge spectra of the granules can be found in Fig. S2. The intensity scale corresponds to the optical density. Scale bars: 1  $\mu$ m

### Biofilms contain diverse bacteria including CPR bacteria

To determine the bacterial community composition across biofilm samples collected at MS4 and MS11 sites over the years, we analyzed the relative abundance of bacterial taxa at the phylum, class, and order levels. We observed a diverse range of bacterial phyla across the samples (Fig. 1G, see also supplementary material for details). The most abundant phylum across the samples was Campylobacterota, representing >50% of the total bacteria, followed by Proteobacteria and Desulfobacterota, respectively. Data on the MS11 biofilms in the year 2020 indicate a possible shift in the bacterial community structure during this period. A closer look at the class level revealed that Gammaproteobacteria are predominant in most samples from the MS4 site, representing a significant portion of the bacterial community. However, distinct patterns of class-level diversity were evident in different samples, showcasing the dynamic nature of biofilm communities over the years at both sites.

We used genome-resolved metagenomics to investigate microbial consortia, metabolisms, and microbial interactions that underpin the Alum Rock communities. In

total, we recovered 212 nonredundant genomic bins from the MS4 and MS11 samples (57 from MS11 and 155 from the biofilm + planktonic samples from MS4). Of these, 38 were classified as near complete (>95%, Supplementary Table S3). Taxonomic affiliations of all of the bacterial genomes were established based on concatenated ribosomal protein trees (Fig. 5A and Supplementary Table S4).

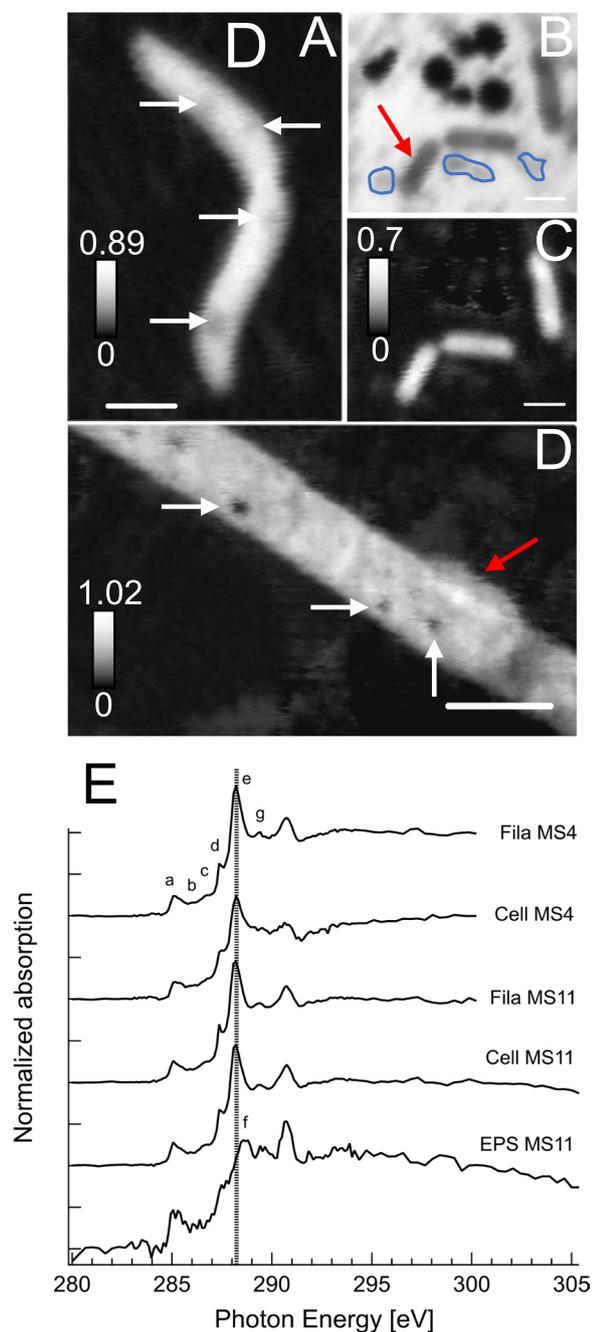
Genomically represented groups in the biofilms and planktonic fractions from both sites include Gammaproteobacteria (Thiotrichales, Chromatiales, Beggiotales), Campylobacterota (Campylobacterales), Betaproteobacteria (including *Thiomonas*), Deltaproteobacteria (specifically Desulfobacterales), Bacteroidota, Chloroflexota, Ignavibacteria, Spirochaetes, Lentisphaerae, Riflebacteria, Verucomicrobia, Acidobacteria, KSB1, Caldiseri-cota, Planctomycetota, Edwardsbacteria, Dependientia (TM6), and Margulisbacteria. Diverse groups of CPR are present, including Uhrbacteria (OP11), Gracilibacteria (BD1-5), Peregrinibacteria (PER), Moranbacteria (OD1), Woesebacteria (OP11), Roizmanbacteria and Gottesmanbacteria (OP11), Saccharibacteria (TM7), Falkowbacteria (OD1), Absconditabacteria (SR1), Berkelbacteria, and Doudnabacteria and Dojkabacteria (WS6) (see <https://ggkbase.berkeley.edu/alumrock-genomes/organisms>).

To estimate the relative abundance of organisms in the two springs (independent of binning), we calculated the DNA read coverage of all of the genomic bins for each spring (Fig. S6). The MS4 spring is dominated by Halothiobacillales, Beggiatoales, Thiotrichales, and Campylobacterales based on relative abundance among genomes (Supplementary Table S5). The most abundant species in MS4 shares genome-wide average 51% amino acid similarity with the sulfur-oxidizer *Thiothrix nivea* [73]. The MS11 spring is dominated by a single *Beggiatoa* sp. (*Beggiatoa*-related\_37\_1401).

### Diverse bacteria are implicated in sulfur cycling

We used a list of high-confidence functional gene assignments for all draft or better-quality genomes to resolve the capabilities central to the metabolisms of the dominant bacteria in the springs. Not surprisingly, genes encoding sulfur cycling are common in the most abundant organisms at both sites (Supplementary Table S5A, B).

Based on the community composition of MS4 and MS11 biofilms, we further focused our analysis on the metabolic pathways of MS4 bacteria where we detected ultra-small and surface-attached cells on filamentous bacteria implicated in sulfur oxidation. The most abundant organism in MS4, which is closely related to the filamentous bacterium *Thiothrix nivea*, encodes genes (*soxABC*, periplasmic thiosulfate-oxidizing; *aprAB*, adenylylsulfate reductase; *dsrAB*, reverse dissimilatory sulfite



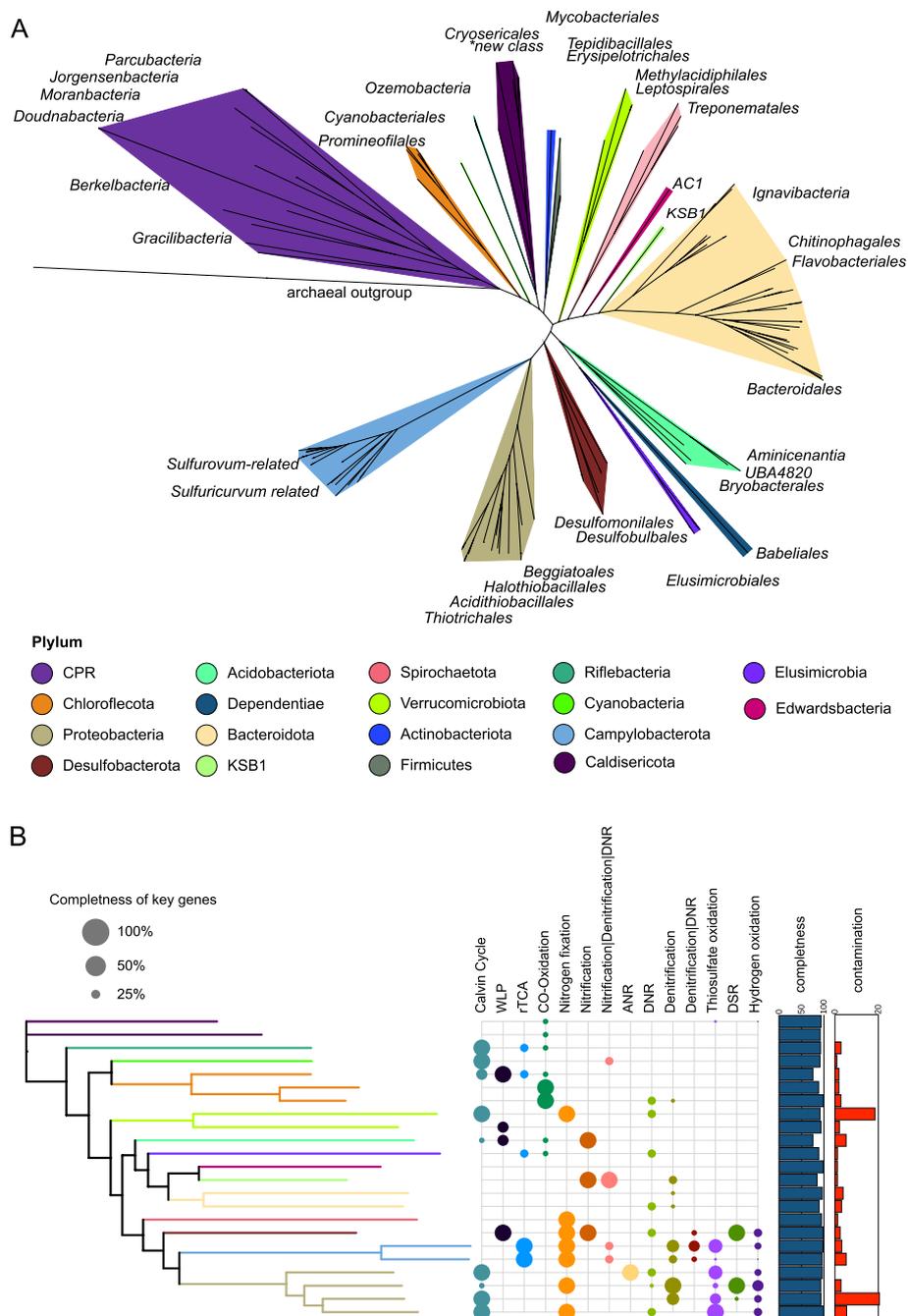
**Fig. 4** Scanning transmission x-ray spectromicroscopy at 87 K of frozen-hydrated MS4 and MS11 biofilms. **A** Protein map of an MS11 small filament ( $700 \text{ nm} \pm 120 \text{ nm}$  in diameter),  $\text{S}^0$  granules are pointed by white arrows. **B** Extracellular  $\text{S}^0$  granules ( $\sim 300$  to  $850 \text{ nm}$  in diameter) and MS11 cells imaged at  $288.2 \text{ eV}$  (amide group in proteins) and corresponding **(C)** protein map. **D** Protein map of an MS4 filament ( $1.46 \pm 0.16 \mu\text{m}$  in diameter) with a surface-attached cell ( $1.9 \pm 0.21 \mu\text{m}$ , red arrow). **E** Carbon K-edge NEXAFS spectra of filamentous bacteria ( $\text{S}^0$  granule-free areas) exhibiting a major peak at  $288.2 \text{ eV}$  and other peaks mainly associated with nucleic acids (see Table S2). Spectra of MS11 and MS4 cells (red arrow) exhibit a main peak at  $288.2 \text{ eV}$  (peptide bond), and associated extracellular polymeric substances (EPS, circled in blue) show a main peak at  $288.7 \text{ eV}$  (carboxyl groups in acidic polysaccharides). See Table S2 for further details. Dashed line is at  $288.2 \text{ eV}$ . The intensity scale corresponds to the optical density. Scale bars:  $1 \mu\text{m}$  (**A–C**) and  $2 \mu\text{m}$  (**D**)

population of *Sulfurovum* bacteria encode *sqr* genes and thus likely oxidize sulfide to  $\text{S}^0$ . Some *Sulfurovum* bacteria in both communities have genomes that also encode *soxCDYZ* complexes, suggesting they mediate thiosulfate oxidation (potentially coupled to nitrate reduction, e.g., via *narG* and *napA*). *Sulfuricurvum* species are also relatively abundant in MS4 and encode genes for sulfur and thiosulfate oxidation, in line with culture-based studies [76]. The genomes of Chloroflexota encode the capacity for thiosulfate disproportionation via thiosulfate reductase/polysulfide reductase (*phsA*) and sulfide oxidation via flavocytochrome *c* sulfide dehydrogenase. Two low abundance Gammaproteobacteria species related to *Acidithiobacillus* have the capacity for thiosulfate oxidation. Several genomes from moderately abundant Halothiobacillales have the metabolic capacity for sulfide and thiosulfate oxidation via *fccB*, *dsrAB*, and *soxBCY*, respectively (Supplementary Table S6A and B).

Some bacteria from MS4 spring also potentially mediate dissimilatory sulfate reduction. Specifically, the genomes of some Desulfobacterales belonging to the families of Desulfatiglandaceae, Syntrophobacterales, Desulfurivibrionaceae, and Desulfarculales encode the capacity to reduce sulfate back to sulfide via Dsr genes, likely coupled to oxidation of organic carbon or  $\text{H}_2$ . Some rare Desulfocapsaceae from MS4 that are related to bacteria of the genus *Desulfocapsa* have thiosulfate reductase, group 3b [NiFe] (Hyd; possibly sulfhydrogenase), and SAT (Sulfate adenylyltransferase) and APR (adenylylsulfate reductase) for the oxidation of sulfite to sulfate. Thus, it appears these bacteria are involved in sulfur disproportionation, whereby  $\text{S}^0$ , thiosulfate, and sulfite are converted to  $\text{H}_2\text{S}$  and sulfate, as has been demonstrated in cultures of bacteria from this genus [77]. Other *Desulfocapsa* spp. have tetrathionate reductase genes, suggesting they are capable of converting tetrathionate to

reductase) to convert sulfide to thiosulfate, elemental sulfur, and sulfate (Fig. 5B). The absence of *dsrD* genes indicates that the Dsr complex operates in the sulfide oxidation direction (i.e., rDsr pathway). This *Thiothrix* bacterium also lacks any *soxC* genes, which in bacterial genomes has been associated with the accumulation of sulfur granules or polysulfide [74, 75].

MS4 contains various other bacteria capable of oxidizing reduced sulfur compounds. A subdominant



**Fig. 5** Phylogenetic analysis and metabolism of bacteria represented by MAGs from the MS4 and MS11 sites. **A** The tree is based on 16 concatenated ribosomal proteins (rpL2, 3, 4, 5, 6, 14, 15, 16, 18, 22, 24, and rpS3, 8, 10, 17, 19) generated using iQ-TREE. An archaeon, *Thermococcus alcaliphilus*, was used as the outgroup. **B** The metabolic capacities for generalized biogeochemical pathways in Alum Rock genomes are represented by colored circles. A pathway is present if the core KEGG orthologs encoding that pathway are identified in each genome. Abbreviations are as follows: WLP, Wood–Ljungdahl pathway, rTCA, reductive tricarboxylic acid cycle; ANR, assimilatory nitrate reduction; DNRA, dissimilatory nitrate reduction to ammonia; thiosulfate oxidation by SOX complex; DSR, dissimilatory sulfate reduction; hydrogen oxidation, [NiFe]-hydrogenases, and NAD-reducing hydrogenase

thiosulfate. The *Desulfocapsa*-related bacteria also contain *dsrABD* genes, which fall within the reductive cluster closely related to those from *Desulfocapsa sulfexigens*. We

infer that the *Desulfocapsa*-related bacteria are capable of S disproportionation, as previously reported [78], and the presence of the *dsrD* functional marker protein suggests

that these species in the springs are capable of both S disproportionation and sulfite reduction. Only members of the candidate phylum Riflebacteria, family Ozemobacteraceae, have the capacity of anaerobic sulfite reduction via anaerobic sulfite reductase system (*asrABC*). A bacterium from a new class of *Caldithrix* from the MS4 spring is predicted to perform sulfur oxidation via dissimilatory sulfite reductase, sulfite oxidation, sulfate reduction, and thiosulfate disproportionation (Supplementary Table S6A). We also identified abundant bacteria from novel families of Bacteroidetes, which generally encode thiosulfate reductase genes (*phS*) and adenylylsulfate reductase (*aprA*) involved in thiosulfate disproportionation and sulfate reduction.

Surprisingly, we identified persulfide dioxygenase (*sdo*) and rhodanase (thiosulfate sulfurtransferase) in the genomes of Elusimicrobia, Riflebacteria, and Oscillatoriophycidae and in a novel family of Syntrophales (Fig. 6A). These enzymes are also present in some heterotrophic bacteria, where they play important roles in the detoxification of intracellular sulfide and sulfur assimilation respectively [79, 80]. We also found a putative sulfur dioxygenase encoded in a Doudnabacteria genome that clusters with protein sequences of other CPR bacteria from public data. In the operon, a sulfur transferase is adjacent, suggesting its potential function in thiosulfate oxidation (Fig. 8). This is interesting because persulfide dioxygenase has not been previously linked to CPR bacteria. Modeling of the persulfide dioxygenase from Doudnabacterium using AlphaFold2 indicates that it has structural homology with the biochemically characterized persulfide dioxygenase (Fig. 6B–D). Furthermore, we identified these two adjacent genes in the genomes of several other CPR from high sulfide environments, including Kaiserbacteria (groundwater from California), Pacebacteria (wastewater), Moranbacteria, and Gracilibacteria (Crystal Geyser aquifer). Thus, we suggest that these genes may enable a variety of CPR bacteria to grow and generate energy from sulfur oxidation.

In contrast, the most highly sampled genomes in MS11 spring are from a *Beggiatoa* species including a novel family within this group (Supplementary Tables S6A). As expected, these *Beggiatoa* genome encodes the Dsr genes (*dsrABCHJKMOPR*); *dsrD* genes were not identified, and we conclude that the Dsr genes are operational in a reverse Dsr pathway (rDsrABs) [81]. The genome also encodes AprAB (adenylylsulfate reductases) and Sat (sulfate adenylyltransferase) for the oxidation of sulfide to sulfate, sulfide-quinone oxidoreductase (*Sqr*), and sulfide dehydrogenase (*fccB*) genes for the oxidation of hydrogen sulfide to S<sup>0</sup>. The genomes contain a partial set of sulfur-oxidizing sox pathway genes, but *soxDXYZ* were identified. Given the lack of *soxC*, we conclude that like

*Thiothrix*, the primary role of *Beggiatoa* in the community is the conversion of sulfide to thiosulfate, elemental sulfur, and sulfate. The absence of *soxCD* in bacterial genomes has been previously associated with the accumulation of sulfur granules or polysulfide [74, 75].

#### Sulfur-oxidizing bacteria also contribute to nitrogen cycling

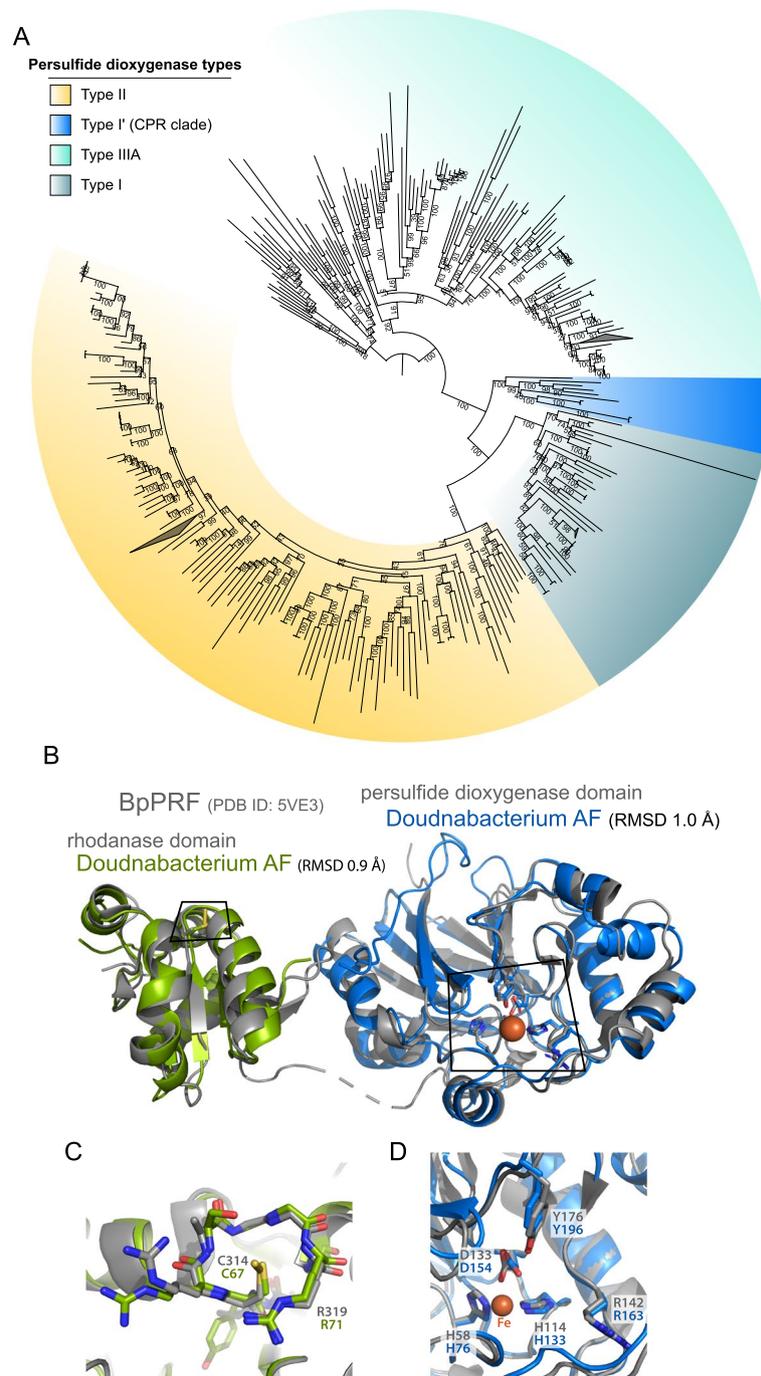
The dominant bacteria in MS4 and MS11 springs are predicted to mediate nitrogen fixation and denitrification processes. In both MS4 and MS11, genes encoding nitrogenase implicated in N<sub>2</sub> fixation are widespread in Proteobacteria, including in the dominant *Thiothrix*, *Beggiatoa*, and *Sulfurovum*, Verrucomicrobia. Other organisms with this capacity include other Gammaproteobacteria, Chromatiales, Campylobacteriales, *Sulfuricum*, Ignavibacteria, *Sulfurospirillum*, Spirochaetes, *Desulfocapsa*, and potentially Lentisphaerae.

The Thiotrichales genomes encode numerous genes for the reduction of nitrate and nitrite; however, only the dominant *Thiothrix* species have the capability to reduce nitrite to nitrous oxide via *nirS* and *norBC* genes. Some Chromatiales bacteria at both springs also appear to be capable of dissimilatory nitrite oxidation to ammonia. The sulfur-oxidizing Campylobacteriales that we identified in both MS4 and MS11 springs have numerous genes implicated in the reduction of nitrate (*napAB*) and nitric oxide (*norBC*). Two low abundance Acidithiobacillales in MS4 that are predicted to perform thiosulfate oxidation have ammonia monooxygenase (*amoA*) genes, suggesting they may be involved in ammonia oxidation and nitrite ammonification. Chloroflexota populations at both springs have the capacity for nitrite reduction via nitrite reductase (*nirK*), nitric oxide reduction (*norBC*), and nitrite ammonification. A novel *Caldithrix* species from MS4 has the potential of nitric oxide reduction via the nitric oxide reductase *norBC* and nitrite reduction via the periplasmic nitrate reductase NapA (Fig. 5B).

In addition to being the most abundant sulfur oxidizers in the MS11 spring, *Beggiatoa* are metabolically versatile with regard to nitrogen cycling. Their genomes encode genes with similarity to nitrate reductase (*narABG*), nitrite reductase (*nirS*), nitric oxide reductase (*norBC*), and nitrous-oxide reductase (*nosZ*) for the complete reduction of nitrate to N<sub>2</sub>. They also contain *nrfA* potentially for dissimilatory nitrite reduction to ammonia (DNRA) or nitrite ammonification. Thus, these bacteria can likely couple sulfur oxidation to nitrate reduction, in line with prior studies [3, 12].

#### Extensive links between hydrogen and sulfur metabolism

To gain insight into the role of hydrogen metabolism in the Alum Rock springs, we analyzed the distribution of



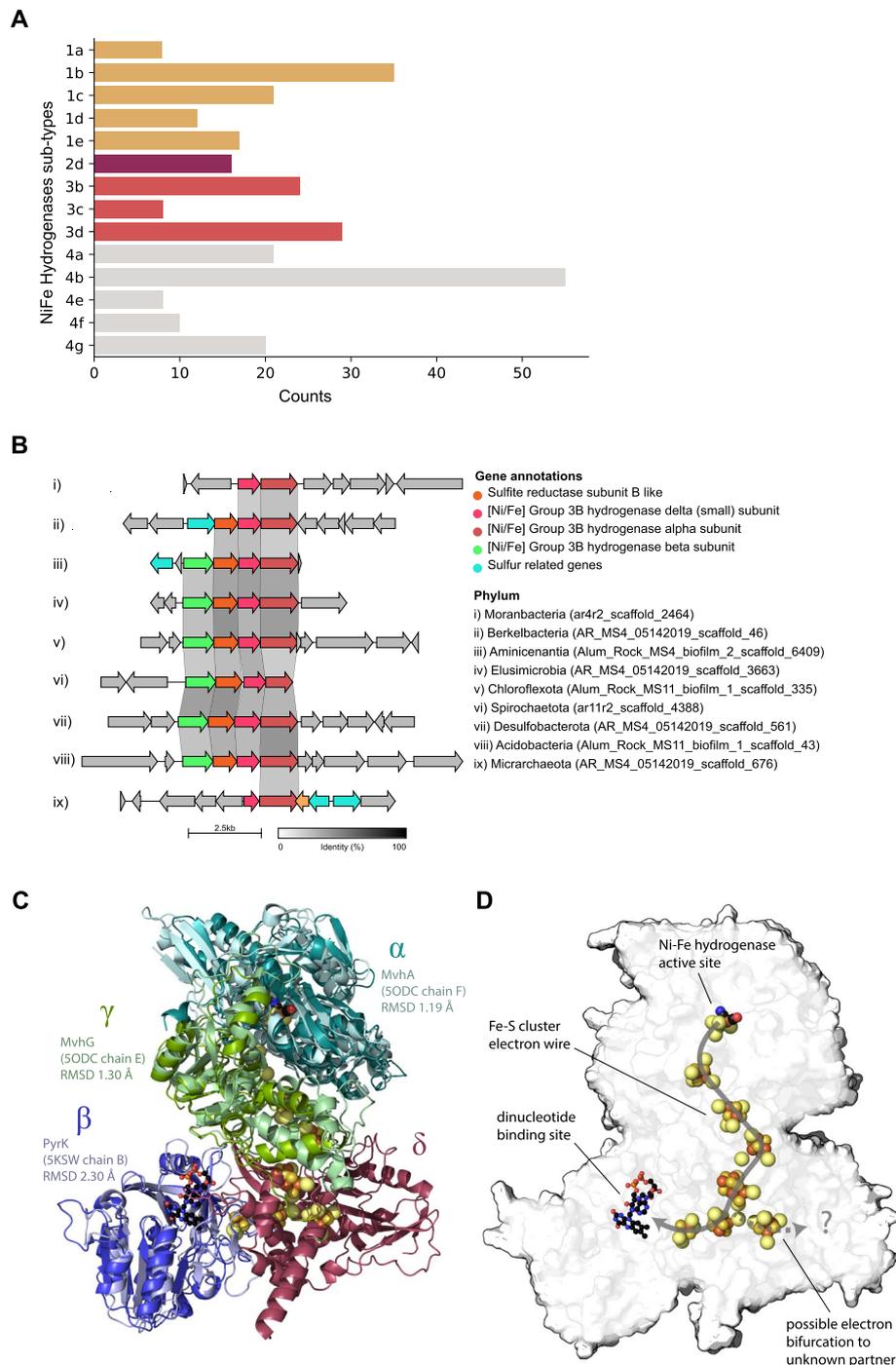
**Fig. 6** Novel persulfide dioxygenase within CPR bacteria. **A** Phylogenetic analyses of persulfide dioxygenase proteins from the Alum Rock genomic bins. The blue monophyletic clade shows the persulfide dioxygenase found in CPR bacteria from sulfur-rich environments. **B** AlphaFold models of *Doudnabacterium* putative rhodanase (green) and persulfide dioxygenase (blue) aligned with the corresponding domains of the characterized natural fusion protein BpRF (PDB ID: 5VE3). **C** and **D** Zoomed views of the active sites of the aligned structures reveal a strong coincidence of the key residues

hydrogenases and associated enzymes in the genomes. There was considerable capacity for fermentative  $H_2$  production using nicotinamides (via group 3b and 3d

[NiFe]-hydrogenases), ferredoxin (via group A [FeFe]-hydrogenases and group 4 [NiFe]-hydrogenases), and formate (via formate hydrogenases) as electron donors

(Fig. 7A). Some putative H<sub>2</sub> producers are likely to be metabolically flexible bacteria such as *Sulfurospirillum* and Flavobacteriales, which can switch to fermentation

when limited for respiratory electron acceptors based on previous reports [59, 82]. CPR bacteria, TA06, and Spirochaetes with group 3b and 3d [NiFe]-hydrogenases are



**Fig. 7** Hydrogenases distribution in Alum Rock genomes and structural insights of group 3b [NiFe]-hydrogenase complex. **A** Total distribution of hydrogenases from the Alum Rock spring. **B** Genomic organization of novel group 3b [NiFe] hydrogenases from different organisms present in the springs. **C** and **D** AlphaFold multimeric model for the Berkelbacterium putative group 3b [NiFe]-hydrogenase complex with the closest known structural matches aligned to each protein

likely to be obligate fermenters given they apparently lack terminal reductases (Supplementary Table S7). The gene arrangements of the group 3b [NiFe]-hydrogenases in the genomes of the CPR bacteria Berkelbacteria and Moranbacteria (Fig. 7B) are similar to the biochemically characterized hydrogenase and sulfhydrogenase of *Pyrococcus furiosus* [83] and those previously reported in other CPR bacteria [25, 84], suggesting that these hydrogenases may be capable of reversible oxidation of hydrogen or capable of reducing sulfur compounds like polysulfide. We modeled the complex from Berkelbacteria genome using AlphaFold; this model suggests a hydrogenase module ( $\alpha$ - and  $\gamma$ -subunits) with an electron wire of FeS clusters connecting to a nucleotide reducing module ( $\beta$  subunit) (Fig. 7C). The  $\delta$  subunit has no close structural analogues but contains an additional FeS cluster and may accommodate an additional electron-accepting partner (Fig. 7D). Based on this structural analysis, there are two separate paths for the electrons suggesting this 3b [NiFe]-hydrogenase complex is potentially an electron-bifurcating hydrogenase.

Numerous bacteria in the Alum Rock springs are predicted to consume  $H_2$  for energy generation. Most of these hydrogenotrophs are predicted to use  $H_2$  to reduce sulfate (via group 1b and 1c [NiFe]-hydrogenases; primarily Deltaproteobacteria), elemental sulfur (via group 1e [NiFe]-hydrogenases; primarily Gammaproteobacteria), or heterodisulfides (via group 3c [NiFe]-hydrogenases; various lineages including Acidobacteria). The most abundant Gammaproteobacteria and Campylobacteria likely oxidize both  $H_2$  and sulfur compounds either mixotrophically or alternatively autotrophically. The hydrogenase repertoire of these organisms includes the oxygen-tolerant group 1b and 1d [NiFe]-hydrogenases [85, 86].

### Organic carbon cycling and fermentation

The ability to fix inorganic carbon ( $CO_2$ ) is a common predicted capacity for bacteria from both sites (Supplementary Table S6A and B). The dominant *Thiothrix*, *Beggiatoa*, and Chromatiales-related bacteria have type II RuBisCO genes that function in the Calvin-Benson-Bassham (CBB) cycle (Fig. S7). One Absconditabacteria genome has a RuBisCO that phylogenetic analysis places within the form II/III CPR clade, as reported previously [25, 87]; these enzymes are inferred to function in a nucleoside salvage pathway in which  $CO_2$  is added to ribulose-1,5-bisphosphate to form 3-phosphoglycerate [88]. Elusimicrobia and Campylobacterota, including species related to Sulfurimonadaceae, have ATP citrate lyase genes that encode the critical enzyme for  $CO_2$  fixation via the reverse TCA (rTCA) cycle. We also identified rTCA genes in a novel Bacteroidetes organism (Supplementary

Table S6A and B). Genes of the Wood–Ljungdahl carbon fixation pathway (*cooS/acsA*, *acsB*, and *acsE*) were widespread in both springs, including in members of the Bacteroidetes, Desulfocapsa, Lentisphaerae, Chloroflexi, and Aminicenantia with the potential of oxidation of small organic compounds.

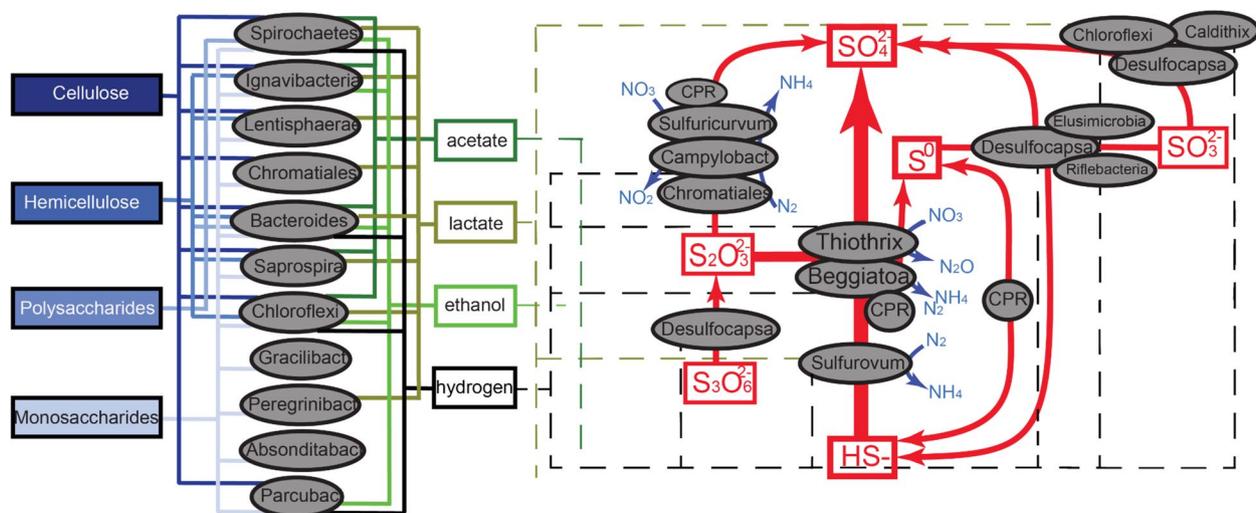
We used marker genes involved in carbohydrate metabolism to infer polymer biomass degradation capacity of the microbial communities. Many bacteria in both springs have the capacity to hydrolyze complex organic molecules to produce a variety of electron donors such as acetate, hydrogen, and lactate (Fig. 8A). Of the organisms in the community, Bacteroidetes and Ignavibacteria contain the most glycosyl-hydrolase genes, and thus, they likely play important roles in polysaccharide degradation. Notably, one Bacteroidetes from MS11 has 66 glycoside hydrolase genes. This organism is the only bacterium that appears to be capable of degrading cellulose, hemicellulose, polysaccharides, and monosaccharides. Gammaproteobacteria, Spirochaetes, Bacilli, and Lentisphaerae also contain genes for the degradation of a variety of complex carbohydrates, but these genes are at relatively low abundance in the sulfur-oxidizing Proteobacteria.

Similarly, many bacteria other than the sulfur-oxidizing Proteobacteria (and CPR) have indications of the capacity for beta-oxidation pathway of saturated fatty acids to acetyl-CoA. Many CPR bacteria have a few glycosyl hydrolase genes, which is significant given the scarce indications of other metabolic capacities in these microorganisms. Methane oxidation is predicted to be a capacity of members of Verrucomicrobia, specifically members of the Methylocidiphilales. This reaction involves particulate methane monooxygenase (pMMO-ABC), the genes for which were identified and classified phylogenetically.

One of the more interesting organisms present in the MS4 spring is a Gracilibacteria, which is predicted to have minimal metabolic capacities beyond glycolysis, production of peptidoglycan, and generation of formate, which may be available for use by other community members. Other capacities predicted for this bacterium are the production of riboflavin, amino sugars, RNA degradation, 1C by folate, interconversion of purines and pyrimidines, and biosynthesis of a few amino acids (Fig. 8).

### Discussion

Some springs are hotspots where resources associated with deeply sourced water can sustain chemoautotrophic ecosystems independent of sunlight. We studied two closely spaced but different sites that discharge a mixture of deeply sourced and shallow groundwater, providing microorganisms with reduced compounds and oxygen. Our research integrated geochemical, synchrotron-based



**Fig. 8** Inference of partitioning of carbon, sulfur, and nitrogen cycling in the Alum Rock springs. Based on the gene content of genomes reconstructed from the springs. Arrows indicate metabolic capacities reconstructed from metagenomes recovered from MS4 and MS11 springs. The dashed lines represent potential electron donors for anaerobic respiration processes

spectromicroscopy, metatranscriptomics, and genome-resolved metagenomic data to resolve the network of microorganisms that define the ecosystems. This approach provided insights into organism metabolic capacities and their associations, including those that involve CPR bacteria and the biogeochemical processes that sustain autotrophic ecosystems in the context of their spring-based hydrological setting.

Analysis of the metabolisms of the dominant bacteria in the springs revealed that genes implicated in sulfur cycling are common at both sites (Fig. 8). As expected, the primary energy source is reduced sulfur in the form of sulfide. Overall, the most common sulfur metabolisms are sulfide oxidation, thiosulfate disproportionation, sulfur oxidation, and less commonly sulfite oxidation and sulfate reduction. Sulfide can be oxidized aerobically and in some cases, anaerobically, coupled with nitrate reduction. Our metagenomic analyses suggest that intermediate sulfur compounds and sulfate and sulfide are actively cycled by Campylobacterota (*Sulfurovum*, *Thiovulum*) and Gammaproteobacteria (Thiotrichales and Beggiotales) in the spring communities, probably coupled to nitrogen compound reduction in some microhabitats. Elemental sulfur serves as an energy source stored as sulfur granules as observed by STXM, in *Beggiatoa* [89]. Interestingly, elemental sulfur-bearing granules may serve as an energy source for the growth of *Beggiatoa* and/or *Thiothrix*. The sulfur oxidizers are the primary source of fixed carbon and nitrogen in the ecosystem.

A higher flow rate and a higher concentration of sulfate were observed at MS11 compared to MS4, and the communities have distinct microbial community

composition. The MS4 ecosystem is highly diverse and dominated by abundant sulfide-oxidizing Gammaproteobacteria (*Thiothrix*, *Sulfurovum*) and sulfate-reducing Desulfobacterales. The MS11 spring has relatively low diversity and is highly dominated by Campylobacterota (*Sulfurovum*, *Thiovulum*) and Gammaproteobacteria (Thiotrichales and Beggiotales). Our findings are consistent with predictions from studies that indicate that filamentous Campylobacterota dominate biofilms with high sulfide/oxygen (>150) ratios, whereas Gammaproteobacteria (*Beggiatoa* like) prefer lower (<75) ratios [9].

We further investigated the metabolic capacities of several CPR bacteria within these communities, as their roles in sulfur-based chemoautotrophic ecosystems remain poorly known. CPR bacteria are often characterized by small genomes and minimal anaerobic fermentative metabolism [90]; however, recent studies have shown auxiliary metabolisms such as the presence of hydrogenases [25, 84], rhodopsin [91], nitrite reductases [92], and F-type ATPase [93], which may contribute to alternative energy conservation and adaptations to different environments and host associations. Notably, we identified genes potentially involved in elemental sulfur reduction (sulfhydrylase) and thiosulfate oxidation (persulfide dioxygenase and rhodonase) in the genomes of several CPR bacteria, suggesting a potential new energy generation mechanism for these bacteria. We found that other CPR bacteria from high sulfur environments have the same predicted potential for thiosulfate oxidation, suggesting an important general adaptation of CPR bacteria in sulfur-rich environments.

The most interesting aspect of the current study regards interactions involving CPR bacteria and their host microorganisms. CPR-host associations have rarely been documented, with the exception of oral microbiome-associated Saccharibacteria (TM7) [29, 94], and Actinobacteria further laboratory studies [95] have validated genomic predictions of metabolic interdependency [84]. One study suggested the presence of CPR cells on the surfaces of their Actinobacteria hosts via SEM and showed them to be rod shaped and  $<0.2 \mu\text{m}$  in diameter and  $\sim 0.5 \mu\text{m}$  in length [96]. Another study linked *Vampirococcus* with anoxygenic photosynthetic Gammaproteobacteria [97]. Two studies suggest links between Parcubacteria and Archaea, in one case *Methanosaeta* [98] and *Methanothrix* [98]. In the case of the CPR Neelsonbacteria associated with *Methanosaeta*, cryo-TEM imaging indicated that *Methanosaeta*-attached cells are  $\sim 0.5 \mu\text{m}$  in diameter. Other cultivation-independent studies have verified that CPR cells are ultra-small and can be better analyzed via filtration through a  $0.2\text{-}\mu\text{m}$  pre-filtering [90]. Cryo-TEM imaging and tomographic analyses have documented ultra-small cells directly associated with CPR cells and host bacteria [31, 90]. Generally, these data indicate that CPR cells are a fraction of a micron in length and diameter, consistent with the size for filament-associated ultra-small cells reported here ( $<650 \text{ nm}$  long,  $\sim 250 \text{ nm}$  wide, the smallest being  $290 \pm 20 \text{ nm}$  long,  $120 \pm 15 \text{ nm}$  wide). In the MS4 biofilms, ultra-small cells were found associated with the surfaces of long filamentous bacteria containing relatively large  $\text{S}^0$  granules as evidenced by STXM. Given that the only abundant filamentous bacteria in these samples are sulfide-oxidizing *Thiothrix*, we predict that some of these tiny epibiotic cells are CPR bacteria. CPR identifications include Gracilibacteria, Berkelbacteria, Moranbacteria, or Doudnabacteria, based on microbial community abundance information. Unfortunately, our attempts to perform fluorescent in situ hybridization to determine CPR bacterial identity were unsuccessful due to low amount of material, so this inference remains tentative. Future laboratory co-cultivation of *Thiothrix* and their epibionts may be required to identify the CPR types, so as to better understand the nature of their association (e.g., mutualistic, parasitic). If confirmed, and given the prediction that some CPR bacteria have putative sulfhydrogenases that may produce  $\text{H}_2\text{S}$  [83, 99], these epibionts may be involved in cryptic sulfur cycling that also involves sulfur-oxidizing bacteria.

Hydrogen is an important resource in many environments [100], yet little is known about the distribution and importance of hydrogenases in sustaining groundwater microbiomes. The most common chemolithoautotrophs

in the Alum Rock spring biofilms are  $\text{H}_2$ -oxidizing bacteria, which use  $\text{H}_2$  as an energy source via the enzyme hydrogenase. Specifically, group 3b [NiFe]-hydrogenases are widely distributed in the genomes of many of the microbial community members. These complexes may mediate hydrogen metabolism or the direct hydrogenation of elemental sulfur to hydrogen sulfide [99]. Other hydrogenases of the microbial community members are implicated in hydrogen production and oxidation. Together, these findings suggest that most bacteria in Alum Rock springs cycle hydrogen gas and sulfur compounds, reactions that underpin the biology and geochemistry of this ecosystem.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40168-023-01704-w>.

**Additional file 1: Supplementary Fig. 1.** Micro-focused X-ray fluorescence distribution map of sulfur in MS4 filaments at 95 Kelvin. The color intensity scale represents X-ray fluorescence counts.

**Additional file 2: Supplementary Fig. 2.** Confocal fluorescence microscopy of MS11 biofilms showing cells treated with SYTOX (blue) for nucleic acid and F-64 (red) for membrane.

**Additional file 3: Supplementary Fig. 3.** Sulfur  $\text{L}_{2,3}$ -edges NEXAFS spectra of sulfur granules in MS4 and MS11 biofilms from filaments shown in Fig. 3, compared to an elemental sulfur standard.

**Additional file 4: Supplementary Fig. 4.** Scanning transmission x-ray microscopy of filamentous bacteria and small cells in MS11 (A) and MS4 (B-D) biofilms. A) Image at 174 eV (at S  $\text{L}_{2}$ -edge) of an ultra-small cell  $477 \pm 23 \text{ nm}$  long,  $218 \pm 18 \text{ nm}$  wide (blue arrow) in close contact with a cell (black arrow) on the surface of a filament. B) Image at 300 eV (above C K-edge) showing curved-shaped cells near filaments, red arrow points to a cell  $764 \pm 24 \text{ nm}$  long,  $309 \pm 14 \text{ nm}$  wide. C) Cell  $656 \pm 24 \text{ nm}$  long,  $284 \pm 22 \text{ nm}$  wide (red arrow) with extracellular polymeric substances and sulfur granules (white arrow), imaged at 200 eV (above S  $\text{L}_{2,3}$ -edges). D) Epibiotic cell  $\sim 1.1 \mu\text{m}$  long,  $\sim 280 \text{ nm}$  wide (red arrow), with copious extracellular polymeric substances, imaged at 163.9 eV (at S  $\text{L}_{3}$ -edge), see Fig. 2F-G. An ultra-small cell  $\sim 480 \text{ nm}$  long,  $\sim 270 \text{ nm}$  wide (blue arrow) seen in contact with an apparently epibiotic cell  $\sim 2 \mu\text{m}$  long,  $\sim 360 \text{ nm}$  wide (black arrow). Inset: "Ultra-small" cell  $290 \pm 20 \text{ nm}$  long,  $120 \pm 15 \text{ nm}$  wide (blue arrow),  $1 \mu\text{m}$  away from another filament ( $\sim 3.5 \mu\text{m}$  wide). The intensity scale corresponds to the optical density. Pixel sizes:  $25 \times 25 \text{ nm}$  (A-B, D Inset),  $40 \times 40 \text{ nm}$  (C),  $100 \times 100 \text{ nm}$  (D).

**Additional file 5: Supplementary Fig. 5.** Carbon standard compounds. Bovine serum albumin (BSA, protein), alginate (acidic polysaccharide), agarose (neutral polysaccharide), deoxyribonucleic acid (DNA), 1,2-dipalmitoyl-sn-glycero-3-phosphoethanolamine (PE lipid), sodium bicarbonate ( $\text{NaHCO}_3$ ). Spectra were normalized at 300 eV. Peaks a: 285.2, b: 288.2, c: 288.6, d: 289.3, e: 285, f: 286, g: 286.7, h: 287.4, i: 289.4, j: 285.1, k: 287.7, l: 288.5, m: 290.2. Vertical dashed line is at 288.2 eV. (see table S2 and references herein).

**Additional file 6: Supplementary Fig. 6.** DNA read coverage of all of the genomic bins for each spring.

**Additional file 7: Supplementary Fig. 7.** Phylogenetic analyses of RuBisCO-like proteins from the Alum Rock genomic bins. Red circles represent the Alum Rock springs sequences.

**Additional file 8: Supplementary Table 1.** Geochemical parameters of the two Alum Rock springs in June 2005. Flow rates given in mL/s, temperature in  $^{\circ}\text{C}$ , and concentrations in mg/L.

**Additional file 9: Supplementary Table 2.** Major carbon functional groups present in MS4 and MS11 biofilms, the peaks were assigned according to prior work [37, 68, 101–103].

**Additional file 10: Table S3.** Nonredundant genomic bins classified as near complete.

**Additional file 11: Table S4.** Taxonomic affiliations based on GTDB.

**Additional file 12: Table S5.** Genomes coverages (CoverM).

**Additional file 13: Table S6.** A. Metabolic expanded.

**Additional file 14: Table S7.** Hydrogenases.

### Acknowledgements

We thank Christine He, Raphaël Méheust, Susan Mullen, Ben Rubin, Jordan Hoff, Lily Law, and Haridha Shivram for their assistance with one sampling trip and Rohan Sachdeva, Shufei Lei, Alex Crits-Christoph, and Adair Borges for helpful discussions and comments on the manuscript. We thank Prof. Arash Komeili and Carson Bickley for providing the FM4-64 reagent and advice on visualizing membranes using confocal fluorescence microscopy. We thank Alum Rock Park for granting sampling permits. We thank Grayson Chadwick for his guidance on the CPR Fluorescence in situ hybridization (FISH) probe design and for providing technical assistance throughout the project.

### Authors' contributions

L.E.V-A was involved in metagenome sample preparation, genomic and metabolic reconstruction, phylogenetic and protein structures analyses, data integration and writing main manuscript text. S.C.F was involved in STXM and X-ray fluorescence microprobe sample preparation, data analysis, data integration and writing the paper. C.G. contributed to the hydrogenases analyses and writing the paper. A. J. P. contributed to analyses of the genomic data. A.L.J. contributed to analyses of the CPR metabolism. J.W-R. contributed to phylogenetic analyses. M.M. helped conceive the study, collected water samples and measurements, and analyzed the geochemical data. J.R. collected the geochemical data. L.M.O. contributed to protein structure modeling and analyses. B.J.B. contributed to phylogenetic and metabolic reconstruction and analyses. D.F.S. contributed financial support. J.F.B. conceived the study, was involved in writing the paper and analyses of the genomic data. All the authors read and approved the final manuscript.

### Funding

We gratefully acknowledge the Innovative Genomics Institute and sequencing resources. This work was partly supported by a NASA Astrobiology Institute, a DOE Carbon Cycle/Kbase grant, and a Sloan Fellowship in Ocean Science to B. J. B. This research used resources of the Advanced Light Source, a US DOE Office of Science User Facility under contract no. DE-AC02-05CH11231. This work was supported by NIH grants 1R01GM12763 and RM1HG009490 to D. F. S. Hydrogeological sampling and analysis were supported by NSF grants 0909701, 1344424, 1724986, and 2116573 to M. M and Chan Zuckerberg Biohub and the Innovative Genomics Institute to J. F. B. LVA was supported by the University of California Dissertation-Year Fellowship. AJP was supported by a scholarship from the German Research Foundation DFG PR-1603/1-1.

### Availability of data and materials

Genomes presented in this manuscript are also made available at <https://ggkbase.berkeley.edu/alumrock-genomes>.

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

JFB is a co-founder of Metagenomi.

### Author details

<sup>1</sup>Graduate Group in Microbiology, University of California, Berkeley, CA, USA. <sup>2</sup>Innovative Genomics Institute, University of California, Berkeley, CA, USA. <sup>3</sup>Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. <sup>4</sup>Earth and Planetary Science, University of California, Berkeley, CA, USA. <sup>5</sup>Environmental Metagenomics, Research Center One Health Ruhr of the University Alliance Ruhr, Faculty of Chemistry, University of Duisburg-Essen, Essen, Essen, Germany. <sup>6</sup>Cleaner Air Oregon Program, Oregon Department of Environmental Quality, Portland, USA. <sup>7</sup>Department of Molecular and Cell Biology, University of California, Berkeley, CA, USA. <sup>8</sup>Howard Hughes Medical Institute, University of California, Berkeley, CA 94720, USA. <sup>9</sup>Environmental Science, Policy and Management, University of California, Berkeley, CA, USA. <sup>10</sup>Earth and Env. Sciences Division, Los Alamos National Laboratory, Los Alamos, NM, USA. <sup>11</sup>University of Duisburg-Essen, Universitätsstraße 5, 45141 Essen, Germany. <sup>12</sup>Department of Microbiology, Biomedicine Discovery Institute, Monash University, Clayton, Australia. <sup>13</sup>Department of Integrative Biology, University of Texas, Austin, USA. <sup>14</sup>Department of Marine Science, University of Texas, Austin, USA. <sup>15</sup>Energy Geoscience Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA.

Received: 30 November 2022 Accepted: 18 October 2023

Published online: 26 January 2024

### References

1. Sheik CS, Anantharaman K, Breier JA, Sylvan JB, Edwards KJ, Dick GJ. Spatially resolved sampling reveals dynamic microbial communities in rising hydrothermal plumes across a back-arc basin. *ISME J*. 2015;9:1434–45. <https://doi.org/10.1038/ismej.2014.228>.
2. Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, et al. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat Commun*. 2016;7:13219. <https://doi.org/10.1038/ncomms13219>.
3. Kalanetra KM, Huston SL, Nelson DC. Novel, attached, sulfur-oxidizing bacteria at shallow hydrothermal vents possess vacuoles not involved in respiratory nitrate accumulation. *Appl Environ Microbiol*. 2004;70:7487–96. <https://doi.org/10.1128/AEM.70.12.7487-7496.2004>.
4. Takai K, Campbell BJ, Cary SC, Suzuki M, Oida H, Nunoura T, et al. Enzymatic and genetic characterization of carbon and energy metabolisms by deep-sea hydrothermal chemolithoautotrophic isolates of Epsilon-proteobacteria. *Appl Environ Microbiol*. 2005;71:7310–20. <https://doi.org/10.1128/AEM.71.11.7310-7320.2005>.
5. Yakushev EV, Pollehne F, Jost G, Kuznetsov I, Schneider B, Umlauf L. Analysis of the water column oxic/anoxic interface in the Black and Baltic seas with a numerical model. *Mar Chem*. 2007;107:388–410. <https://www.sciencedirect.com/science/article/pii/S0304420307001314>.
6. Grote J, Schott T, Bruckner CG, Glöckner FO, Jost G, Teeling H, et al. Genome and physiology of a model Epsilonproteobacterium responsible for sulfide detoxification in marine oxygen depletion zones. *Proc Natl Acad Sci U S A*. 2012;109:506–10. <https://doi.org/10.1073/pnas.1111262109>.
7. Madrid VM, Taylor GT, Scranton MI, Chistoserdov AY. Phylogenetic diversity of bacterial and archaeal communities in the anoxic zone of the Cariaco Basin. *Appl Environ Microbiol*. 2001;67:1663–74. <https://doi.org/10.1128/AEM.67.4.1663-1674.2001>.
8. Macalady JL, Lyon EH, Koffman B, Albertson LK, Meyer K, Galdenzi S, et al. Dominant microbial populations in limestone-corroding stream biofilms, Frasassi cave system, Italy. *Appl Environ Microbiol*. 2006;72:5596–609. <https://doi.org/10.1128/AEM.00715-06>.
9. Macalady JL, Dattagupta S, Schaperdoth I, Jones DS, Druschel GK, Eastman D. Niche differentiation among sulfur-oxidizing bacterial populations in cave waters. *ISME J*. 2008;2:590–601. <https://doi.org/10.1038/ismej.2008.25>.
10. Engel AS, Porter ML, Stern LA, Quinlan S, Bennett PC. Bacterial diversity and ecosystem function of filamentous microbial mats from aphotic (cave) sulfidic springs dominated by chemolithoautotrophic "Epsilon-proteobacteria." *FEMS Microbiol Ecol*. 2004;51:31–53. <https://doi.org/10.1016/j.femsec.2004.07.004>.

11. Sarbu SM, Kinkle BK, Vlasceanu L, Kane TC, Popa R. Microbiological characterization of a sulfide-rich groundwater ecosystem. *Geomicrobiol J.* 1994;12:175–82. <https://doi.org/10.1080/01490459409377984>.
12. Sharrar AM, Flood BE, Bailey JV, Jones DS, Biddanda BA, Ruberg SA, et al. Novel large sulfur bacteria in the metagenomes of groundwater-fed chemosynthetic microbial mats in the Lake Huron basin. *Front Microbiol.* 2017;8:791. <https://doi.org/10.3389/fmicb.2017.00791>.
13. Williams TM, Unz RF. Filamentous sulfur bacteria of activated sludge: characterization of *Thiothrix*, *Beggiatoa*, and *Eikelboom* type 021N strains. *Appl Environ Microbiol.* 1985;49:887–98. <https://doi.org/10.1128/aem.49.4.887-898.1985>.
14. Nelson DC, Jørgensen BB, Revsbech NP. Growth pattern and yield of a chemoautotrophic *Beggiatoa* sp. in oxygen-sulfide microgradients. *Appl Environ Microbiol.* 1986;52:225–33. <https://doi.org/10.1128/aem.52.2.225-233.1986>.
15. Nielsen PH, de Muro MA, Nielsen JL. Studies on the in situ physiology of *Thiothrix* spp. present in activated sludge. *Environ Microbiol.* 2000;2:389–98. <https://doi.org/10.1046/j.1462-2920.2000.00120.x>.
16. Hinck S, Neu TR, Lavik G, Mussmann M, de Beer D, Jonkers HM. Physiological adaptation of a nitrate-storing *Beggiatoa* sp. to diel cycling in a phototrophic hypersaline mat. *Appl Environ Microbiol.* 2007;73:7013–22. <https://doi.org/10.1128/AEM.00548-07>.
17. Ehrlich HL, Newman DK. *Geomicrobiology*, Fifth Edition. Taylor & Francis; 2008. Available from: <https://play.google.com/store/books/details?id=5D0-mwEACAAJ>.
18. Sweerts J-PRA, Beer DD, Nielsen LP, Verdouw H, Van den Heuvel JC, Cohen Y, et al. Denitrification by sulphur oxidizing *Beggiatoa* spp. mats on freshwater sediments. *Nature.* 1990;344:762–3. <https://www.nature.com/articles/344762a0> Available from Cited 2022 Nov 11.
19. Rossetti S, Blackall LL, Levantesi C, Uccelletti D, Tandoi V. Phylogenetic and physiological characterization of a heterotrophic, chemolithoautotrophic *Thiothrix* strain isolated from activated sludge. *Int J Syst Evol Microbiol.* 2003;53:1271–6. Available from: <https://doi.org/10.1099/ijs.0.02647-0>.
20. Inagaki F, Takai K, Neelson KH, Horikoshi K. *Sulfurovum lithotrophicum* gen nov., sp. nov., a novel sulfur-oxidizing chemolithoautotroph within the epsilon-Proteobacteria isolated from Okinawa Trough hydrothermal sediments. *Int J Syst Evol Microbiol.* 2004;54:1477–82. <https://doi.org/10.1099/ijs.0.03042-0>.
21. Jones DS, Albrecht HL, Dawson KS, Schaperdoth I, Freeman KH, Pi Y, et al. Community genomic analysis of an extremely acidophilic sulfur-oxidizing biofilm. *ISME J.* 2012;6:158–70. <https://doi.org/10.1038/ismej.2011.75>.
22. Hamilton TL, Jones DS, Schaperdoth I, Macalady JL. Metagenomic insights into S(0) precipitation in a terrestrial subsurface lithoautotrophic ecosystem. *Front Microbiol.* 2014;5:756. <https://doi.org/10.3389/fmicb.2014.00756>.
23. Rossmassler K, Hanson TE, Campbell BJ. Diverse sulfur metabolisms from two subterranean sulfidic spring systems. *FEMS Microbiol Lett.* 2016;363. <https://doi.org/10.1093/femsle/fnw162>.
24. Meziti A, Nikouli E, Hatt JK, Konstantinidis KT, Kormas KA. Time series metagenomic sampling of the Thermopyles, Greece, geothermal springs reveals stable microbial communities dominated by novel sulfur-oxidizing chemoautotrophs. *Environ Microbiol.* 2021;23:3710–26. <https://doi.org/10.1111/1462-2920.15373>.
25. Wrighton KC, Thomas BC, Sharon I, Miller CS, Castelle CJ, VerBerkmoes NC, et al. Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science.* 2012;337:1661–5. <https://doi.org/10.1126/science.1224041>.
26. Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, et al. Unusual biology across a group comprising more than 15% of domain bacteria. *Nature.* 2015;523:208–11. <https://doi.org/10.1038/nature14486>.
27. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, et al. A new view of the tree of life. *Nat Microbiol.* 2016;1:16048.
28. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol.* 2018;36:996–1004. <https://doi.org/10.1038/nbt.4229>.
29. He X, McLean JS, Edlund A, Yooseph S, Hall AP, Liu S-Y, et al. Cultivation of a human-associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. *Proc Natl Acad Sci U S A.* 2015;112:244–9. <https://doi.org/10.1073/pnas.1419038112>.
30. Bor B, Bedree JK, Shi W, McLean JS, He X. Saccharibacteria (TM7) in the human oral microbiome. *J Dent Res.* 2019;98:500–9. <https://doi.org/10.1177/0022034519831671>.
31. He C, Keren R, Whittaker ML, Farag IF, Doudna JA, Cate JHD, et al. Genome-resolved metagenomics reveals site-specific diversity of epibiotic CPR bacteria and DPANN archaea in groundwater ecosystems. *Nat Microbiol.* 2021;6:354–65. <https://doi.org/10.1038/s41564-020-00840-5>.
32. Vigneron A, Cruaud P, Culley AI, Couture R-M, Lovejoy C, Vincent WF. Genomic evidence for sulfur intermediates as new biogeochemical hubs in a model aquatic microbial ecosystem. *Microbiome.* 2021;9:46. <https://doi.org/10.1186/s40168-021-00999-x>.
33. Hahn CR, Farag IF, Murphy CL, Podar M, Elshahed MS, Youssef NH. Microbial diversity and sulfur cycling in an Early Earth Analogue: from ancient novelty to modern commonality. *MBio.* 2022;13:e0001622. <https://doi.org/10.1128/mbio.00016-22>.
34. Rowland JC, Manga M, Rose TP. The influence of poorly interconnected fault zone flow paths on spring geochemistry. *Geofluids.* 2008;8:93–101. <https://doi.org/10.1111/j.1468-8123.2008.00208.x>.
35. Kilcoyne ALD, Tyliczszak T, Steele WF, Fakra S, Hitchcock P, Franck K, et al. Interferometer-controlled scanning transmission X-ray microscopes at the advanced light source. *J Synchrotron Radiat.* 2003;10:125–36. <https://doi.org/10.1107/s0909049502017739>.
36. Stöhr J. NEXAFS spectroscopy [Internet]. Springer Science & Business Media; 1992. Available from: <https://play.google.com/store/books/details?id=N5NBD0393ZYC>.
37. Ade H, Zhang X, Cameron S, Costello C, Kirz J, Williams S. Chemical contrast in X-ray microscopy and spatially resolved XANES spectroscopy of organic specimens. *Science.* 1992;258:972–5. <https://doi.org/10.1126/science.1439809>.
38. Kirz J, Jacobsen C, Howells M. Soft X-ray microscopes and their biological applications. *Q Rev Biophys.* 1995;28:33–130. <https://doi.org/10.1017/s0033583500003139>.
39. Comolli LR, Downing KH. Dose tolerance at helium and nitrogen temperatures for whole cell electron tomography. *J Struct Biol.* 2005;152:149–56. <https://doi.org/10.1016/j.jsb.2005.08.004>.
40. Beetz T, Jacobsen C. Soft X-ray radiation-damage studies in PMMA using a cryo-STXM. *J Synchrotron Radiat.* 2003;10:280–3. <https://doi.org/10.1107/s0909049503003261>.
41. Sarret G, Connan J, Kasrai M, Bancroft GM, Charrié-Duhaut A, Lemoine S, et al. Chemical forms of sulfur in geological and archeological asphaltenes from Middle East, France, and Spain determined by sulfur K- and L-edge X-ray absorption near-edge structure spectroscopy. *Geochim Cosmochim Acta.* 1999;63:3767–79. Available from: <https://www.sciencedirect.com/science/article/pii/S0016703799002057>.
42. Ravel B, Newville M. ATHENA, ARTEMIS, HEPHAESTUS: data analysis for X-ray absorption spectroscopy using IFFFIT. *J Synchrotron Radiat.* 2005;12:537–41. <https://doi.org/10.1107/S0909049505012719>.
43. Marcus MA, MacDowell AA, Celestre R, Manceau A, Miller T, Padmore HA, et al. Beamline 10.3.2 at ALS: a hard X-ray microprobe for environmental and materials sciences. *J Synchrotron Radiat.* 2004;11:239–47. <https://doi.org/10.1107/S0909049504005837>.
44. Fakra SC, Luef B, Castelle CJ, Mullin SW, Williams KH, Marcus MA, et al. Correlative cryogenic spectromicroscopy to investigate selenium bioreduction products. *Environ Sci Technol.* 2018;52:503–12. <https://doi.org/10.1021/acs.est.5b01409>.
45. Peng Y, Leung HCM, Yiu SM, Chin FYL. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics.* 2012;28:1420–8.
46. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9:357–9. <https://doi.org/10.1038/nmeth.1923>.
47. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ.* 2019;7:e7359. <https://doi.org/10.7717/peerj.7359>.
48. Nissen JN, Johansen J, Allesøe RL, Sønderby CK, Armenteros JJA, Grønbech CH, et al. Improved metagenome binning and assembly using deep variational autoencoders. *Nat Biotechnol.* 2021;39:555–60. <https://doi.org/10.1038/s41587-020-00777-4>.

49. Wu Y-W, Simmons BA, Singer SW. MaxBin 20: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics*. 2016;32:605–7. <https://doi.org/10.1093/bioinformatics/btv638>.
50. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol*. 2018;3:836–43. <https://doi.org/10.1038/s41564-018-0171-1>.
51. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat Methods*. 2023;20:1203–12. <https://doi.org/10.1038/s41592-023-01940-w>.
52. Sorek R, Zhu Y, Creevey CJ, Francino MP, Bork P, Rubin EM. Genome-wide experimental determination of barriers to horizontal gene transfer. *Science*. 2007;318:1449–52. <https://doi.org/10.1126/science.1147112>.
53. Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol*. 2013;30:1188–95. <https://doi.org/10.1093/molbev/mst024>.
54. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32:268–74. <https://doi.org/10.1093/molbev/msu300>.
55. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res*. 2021;49:W293–6. <https://doi.org/10.1093/nar/gkab301>.
56. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30:772–80. <https://doi.org/10.1093/molbev/mst010>.
57. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25:1972–3. <https://doi.org/10.1093/bioinformatics/btp348>.
58. Matheus Carnevali PB, Schulz F, Castelle CJ, Kantor RS, Shih PM, Sharon I, et al. Hydrogen-based metabolism as an ancestral trait in lineages sibling to the Cyanobacteria. *Nat Commun*. 2019;10:463. <https://doi.org/10.1038/s41467-018-08246-y>.
59. Søndergaard D, Pedersen CNS, Greening C. HydDB: a web tool for hydrogenase classification and analysis. *Sci Rep*. 2016;6:34212. <https://doi.org/10.1038/srep34212>.
60. Gilchrist CLM, Chooi Y-H. Clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics*. 2021; Available from: <https://doi.org/10.1093/bioinformatics/btab007>.
61. Zhou Z, Tran PQ, Breister AM, Liu Y, Kieft K, Cowley ES, et al. METABOLIC: high-throughput profiling of microbial genomes for functional traits, biogeochemistry, and community-scale metabolic networks. *bioRxiv*. 2020;761643. <https://doi.org/10.1101/761643v2> Available from Cited 2022 Feb.
62. Raveh-Sadka T, Thomas BC, Singh A, Firek B, Brooks B, Castelle CJ, et al. Gut bacteria are rarely shared by co-hospitalized premature infants, regardless of necrotizing enterocolitis development. *Elife*. 2015;4. <https://doi.org/10.7554/eLife.05477>.
63. Castelle CJ, Brown CT, Anantharaman K, Probst AJ, Huang RH, Banfield JF. Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. *Nat Rev Microbiol*. 2018;16:629–45. <https://doi.org/10.1038/s41579-018-0076-2>.
64. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596:583–9. <https://doi.org/10.1038/s41586-021-03819-2>.
65. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, et al. Protein complex prediction with AlphaFold-Multimer. *bioRxiv*. 2021;10(04):463034. <https://doi.org/10.1101/2021.10.04.463034v2> Available from Cited 2022 Jul 25.
66. Stewart-Ornstein J, Hitchcock AP, Hernández Cruz D, Henklein P, Overhage J, Hilpert K, et al. Using intrinsic X-ray absorption spectral differences to identify and map peptides and proteins. *J Phys Chem B*. 2007;111:7691–9. <https://doi.org/10.1021/jp0720993>.
67. Boese J, Osanna A, Jacobsen C, Kirz J. Carbon edge XANES spectroscopy of amino acids and peptides. *J Electron Spectrosc Relat Phenom*. 1997;85:9–15. <https://www.sciencedirect.com/science/article/pii/S0368204897000327>.
68. Samuel NT, Lee C-Y, Gamble LJ, Fischer DA, Castner DG. NEXAFS characterization of DNA components and molecular-orientation of surface-bound DNA oligomers. *J Electron Spectrosc Relat Phenom*. 2006;152:134–42. <https://www.sciencedirect.com/science/article/pii/S0368204806000375>.
69. Zubavichus Y, Shaporenko A, Korolkov V, Grunze M, Zharnikov M. X-ray absorption spectroscopy of the nucleotide bases at the carbon, nitrogen, and oxygen K-edges. *J Phys Chem B*. 2008;112:13711–6. Available from: <https://www.google.com/url?q=https://pubs.acs.org/doi/https://doi.org/10.1021/jp802453u&sa=D&source=docs&ust=1669789373089348&usg=AOvVaw2Z-opOAKZ6bwpTr2gyK1cS>.
70. Benzerara K, Yoon TH, Tylliszczak T, Constantz B, Spormann AM, Brown GE. Scanning transmission X-ray microscopy study of microbial calcification. *Geobiology*. 2004;2:249–59. <https://doi.org/10.1111/j.1472-4677.2004.00039.x>.
71. Toner BM, Fakra SC, Manganini SJ, Santelli CM, Marcus MA, Moffett JW, et al. Preservation of iron(II) by carbon-rich matrices in a hydrothermal plume. *Nat Geosci*. 2009 2:197–201. <https://www.nature.com/articles/ngeo433>. Available from Cited 2022 Oct 30.
72. Brandes JJ, Wirick S, Jacobsen C. Carbon K-edge spectra of carbonate minerals. *J Synchrotron Radiat*. 2010;17:676–82. <https://doi.org/10.1107/S0909049510020029>.
73. Lapidus A, Nolan M, Lucas S, Glavina Del Rio T, Tice H, Cheng J-F, et al. Genome sequence of the filamentous, gliding *Thiothrix nivea* neotype strain (JP2(T)). *Stand Genomic Sci*. 2011;5:398–406. <https://doi.org/10.4056/sigs.2344929>.
74. Friedrich CG, Bardischewsky F, Rother D, Quentmeier A, Fischer J. Prokaryotic sulfur oxidation. *Curr Opin Microbiol*. 2005;8:253–9.
75. Frigaard N-U, Dahl C. Sulfur metabolism in phototrophic sulfur bacteria. *Adv Microb Physiol*. 2009;54:103–200. [https://doi.org/10.1016/S0065-2911\(08\)00002-7](https://doi.org/10.1016/S0065-2911(08)00002-7).
76. Handley KM, Bartels D, O'Loughlin EJ, Williams KH, Trimble WL, Skinner K, et al. The complete genome sequence for putative H<sub>2</sub>- and S-oxidizer *Candidatus Sulfuricurvum* sp, assembled de novo from an aquifer-derived metagenome. *Environ Microbiol*. 2014;16:3443–62. <https://doi.org/10.1111/1462-2920.12453>.
77. Finster K, Liesack W, Thamdrup B. Elemental sulfur and thiosulfate disproportionation by *Desulfocapsa sulfooxigenis* sp. Nov., a new anaerobic bacterium isolated from marine surface sediment. *Appl Environ Microbiol*. 1998;64:119–25. <https://doi.org/10.1128/AEM.64.1.119-125.1998>.
78. Finster KW, Kjeldsen KU, Kube M, Reinhardt R, Musmann M, Amann R, et al. Complete genome sequence of *Desulfocapsa sulfexigenis*, a marine deltaproteobacterium specialized in disproportionating inorganic sulfur compounds. *Stand Genomic Sci*. 2013;8:58–68. <https://doi.org/10.4056/sigs.377412>.
79. Motl N, Skiba MA, Kabil O, Smith JL, Banerjee R. Structural and biochemical analyses indicate that a bacterial persulfide dioxxygenase-rhodanese fusion protein functions in sulfur assimilation. *J Biol Chem*. 2017;292:14026–38. <https://doi.org/10.1074/jbc.M117.90170>.
80. Zhang J, Liu R, Xi S, Cai R, Zhang X, Sun C. A novel bacterial thiosulfate oxidation pathway provides a new clue about the formation of zero-valent sulfur in deep sea. *ISME J*. 2020;14:2261–74. <https://doi.org/10.1038/s41396-020-0684-5>.
81. Anantharaman K, Hausmann B, Jungbluth SP, Kantor RS, Lavy A, Warren LA, et al. Expanded diversity of microbial groups that shape the dissimilatory sulfur cycle. *ISME J*. 2018;12:1715–28. <https://doi.org/10.1038/s41396-018-0078-0>.
82. Berney M, Greening C, Conrad R, Jacobs WR Jr, Cook GM. An obligately aerobic soil bacterium activates fermentative hydrogen production to survive reductive stress during hypoxia. *Proc Natl Acad Sci U S A*. 2014;111:11479–84. <https://doi.org/10.1073/pnas.1407034111>.
83. Pedroni P, Della Volpe A, Galli G, Mura GM, Pratesi C, Grandi G. Characterization of the locus encoding the [Ni-Fe] sulfhydrogenase from the archaeon *Pyrococcus furiosus*: evidence for a relationship to bacterial sulfite reductases. *Microbiology*. 1995;141(2):449–58. <https://doi.org/10.1099/13500872-141-2-449>.
84. Jaffe AL, Castelle CJ, Matheus Carnevali PB, Gribaldo S, Banfield JF. The rise of diversity in metabolic platforms across the candidate phyla radiation. *BMC Biol*. 2020;18:69. <https://doi.org/10.1186/s12915-020-00804-5>.

85. Olson JW, Maier RJ. Molecular hydrogen as an energy source for *Helicobacter pylori*. *Science*. 2002;298:1788–90. <https://doi.org/10.1126/science.1077123>.
86. Fritsch J, Scheerer P, Frielingsdorf S, Kroschinsky S, Friedrich B, Lenz O, et al. The crystal structure of an oxygen-tolerant hydrogenase uncovers a novel iron-sulphur centre. *Nature*. 2011;479:249–52. <https://doi.org/10.1038/nature10505>.
87. Wrighton KC, Castelle CJ, Varaljay VA, Satagopan S, Brown CT, Wilkins MJ, et al. RubisCO of a nucleoside pathway known from Archaea is found in diverse uncultivated phyla in bacteria. *ISME J*. 2016;10:2702–14. <https://doi.org/10.1038/ismej.2016.53>.
88. Sato T, Atomi H, Imanaka T. Archaeal type III RuBisCOs function in a pathway for AMP metabolism. *Science*. 2007;315:1003–6. <https://doi.org/10.1126/science.1135999>.
89. Schwedt A, Kreutzmann A-C, Polerecky L, Schulz-Vogt HN. Sulfur respiration in a marine chemolithoautotrophic bebbiata strain. *Front Microbiol*. 2011;2:276. <https://doi.org/10.3389/fmicb.2011.00276>.
90. Luef B, Frischkorn KR, Wrighton KC, Holman HYN, Birarda G, Thomas BC, et al. Diverse uncultivated ultra-small bacterial cells in groundwater. *Nat Commun*. 2015;6:6372. <https://doi.org/10.1038/ncomms7372>.
91. Jaffe AL, Konno M, Kawasaki Y, Kataoka C, Bèjà O, Kandori H, et al. Saccharibacteria harness light energy using type-1 rhodopsins that may rely on retinal sourced from microbial hosts. *ISME J*. 2022;16:2056–9. <https://doi.org/10.1038/s41396-022-01231-w>.
92. Danczak RE, Johnston MD, Kenah C, Slattery M, Wrighton KC, Wilkins MJ. Members of the candidate phyla radiation are functionally differentiated by carbon- and nitrogen-cycling capabilities. *Microbiome*. 2017;5:112. <https://doi.org/10.1186/s40168-017-0331-1>.
93. Chaudhari NM, Overholt WA, Figueroa-Gonzalez PA, Taubert M, Bornemann TLV, Probst AJ, et al. The economical lifestyle of CPR bacteria in groundwater allows little preference for environmental drivers. *Environ Microbiome*. 2021;16:24. <https://doi.org/10.1186/s40793-021-00395-w>.
94. Cross KL, Campbell JH, Balachandran M, Campbell AG, Cooper CJ, Griffen A, et al. Targeted isolation and cultivation of uncultivated bacteria by reverse genomics. *Nat Biotechnol*. 2019;37:1314–21. <https://doi.org/10.1038/s41587-019-0260-6>.
95. Tian J, Utter DR, Cen L, Dong P-T, Shi W, Bor B, et al. Acquisition of the arginine deiminase system benefits epiparasitic Saccharibacteria and their host bacteria in a mammalian niche environment. *Proc Natl Acad Sci U S A*. 2022;119. <https://doi.org/10.1073/pnas.2114909119>.
96. Bor B, Collins AJ, Murugkar PP, Balasubramanian S, To TT, Hendrickson EL, et al. Insights obtained by culturing Saccharibacteria with their bacterial hosts. *J Dent Res*. 2020;99:685–94. <https://doi.org/10.1177/0022034520905792>.
97. Moreira D, Zivanovic Y, López-Archilla AI, Iniesto M, López-García P. Reductive evolution and unique predatory mode in the CPR bacterium *Vampirococcus lugosii*. *Nat Commun*. 2021;12:2454. <https://doi.org/10.1038/s41467-021-22762-4>.
98. Kuroda K, Yamamoto K, Nakai R, Hirakata Y, Kubota K, Nobu MK, et al. Symbiosis between *Candidatus Patescibacteria* and Archaea discovered in wastewater-treating bioreactors. *MBio*. 2022;e0171122. <https://doi.org/10.1128/mbio.01711-22>.
99. Ma K, Schicho RN, Kelly RM, Adams MW. Hydrogenase of the hyperthermophile *Pyrococcus furiosus* is an elemental sulfur reductase or sulfhydrogenase: evidence for a sulfur-reducing hydrogenase ancestor. *Proc Natl Acad Sci U S A*. 1993;90:5341–4. <https://doi.org/10.1073/pnas.90.11.5341>.
100. Chapelle FH, O'Neill K, Bradley PM, Methé BA, Ciuffo SA, Knobel LL, et al. A hydrogen-based subsurface microbial community dominated by methanogens. *Nature*. 2002;415:312–5. <https://doi.org/10.1038/415312a>.
101. Hitchcock AP, Morin C, Heng YM, Cornelius RM, Brash JL. Towards practical soft X-ray spectromicroscopy of biomaterials. *J Biomater Sci Polym Ed*. 2002;13:919–37. <https://doi.org/10.1163/156856202320401960>.
102. Kaznacheyev K, Osanna A, Jacobsen C, Plashkevych O, Vahtras O, Ågren, et al. Innershell absorption spectroscopy of amino acids. *J Phys Chem A*. 2002;106:3153–68. <https://doi.org/10.1021/jp013385w>.
103. Yabuta H, Uesugi M, Naraoka H, Ito M, Kilcoyne ALD, Sandford SA, et al. X-ray absorption near edge structure spectroscopic study of Hayabusa category 3 carbonaceous particles. *Earth Planets Space*. 2014;66:1–8. <https://doi.org/10.1186/s40623-014-0156-0> Available from Cited 2022 Oct.

## Publisher's Note

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