Environmental Microbiology

ENVIRONMENTAL MICROBIOLOGY

BRIEF REPORT OPEN ACCESS

Large Filamentous Bacteria Isolated From Sulphidic Sediments Reveal Novel Species and Distinct Energy and Defence Mechanisms for Survival

Alexis Fonseca[1,2,3] | Thomas Ishoey[4] | Carola Espinoza[5] | Ian P. G. Marshall[1] | Lars Peter Nielsen[1] | Victor Ariel Gallardo[5]

[1]Center for Electromicrobiology, Department of Biology, Aarhus University, Aarhus, Denmark | [2]Department of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm, Sweden | [3]Baltic Sea Centre, Stockholm University, Stockholm, Sweden | [4]J. Craig Venter Institute, La Jolla, California, USA | [5]Department of Oceanography, University of Concepción, Concepción, Chile

Correspondence: Alexis Fonseca (alexis.fonseca@su.se) | Victor Ariel Gallardo (vagallar@udec.cl)

Received: 18 June 2024 | **Revised:** 25 January 2025 | **Accepted:** 4 March 2025

Funding: This work was supported by the Program 'Doctorado en el extranjero Becas Chile' from ANID (former CONICYT). Sloan Foundation Project 'Census of Marine Microbes' (ICoMM), chapter of 'Census of Marine Life' (CoML) program, under the direction of Jesse H. Ausubel, Director of the Program for the Human Environment, The Rockefeller University, New York, U.S.A. Agencia Nacional de Investigación y Desarrollo, 1070552; J. Craig Venter Institute.

Keywords: bacterial defence | bacterial genomics | CRISPR-Cas | large filamentous bacteria | sediments | sulfuretum | sulphur

ABSTRACT

Various morphotypes of large filamentous bacteria were isolated through micromanipulation from sulphidic sediment mats in the Bay of Concepción, central Chile. This study employed DNA amplification, whole-genome sequencing and bioinformatics analyses to unveil the taxonomic and genomic features of previously unidentified bacteria. The results revealed several novel genera, families and species, including three specimens belonging to Beggiatoales (Beggiatoaceae family), five to Desulfobacterales (Desulfobacteraceae family), two to the Chloroflexi phylum and one to the phylum Firmicutes. Metabolically, Beggiatoaceae bacteria exhibit a flexible and versatile genomic repertoire, enabling them to adapt to variable conditions at the sediment–water interface. All the bacteria demonstrated a mixotrophic mode, gaining energy from both inorganic and organic carbon sources. Except for the Firmicutes bacterium, all others displayed the ability to grow chemolithoautotrophically using H_2 and CO_2. Remarkably, the reverse tricarboxylic acid (rTCA) and Calvin–Benson–Bassham (CBB) pathways coexisted in one Beggiatoaceae bacterium. Additionally, various defence systems, such as CRISPR-Cas, along with evidence of viral interactions, have been identified. These defence mechanisms suggest that large filamentous bacteria inhabiting sulphidic sediments frequently encounter bacteriophages. Thus, robust defence mechanisms coupled with multicellularity may determine the survival or death of these large bacteria.

1 | Introduction

According to microbial community surveys based on 16S rRNA gene amplicon sequencing, the offshore anoxic seabed off central Chile exhibits remarkable bacterial diversity (Gallardo et al. 2016), primarily associated with S and H_2 metabolism, and dominated by Desulfobacterota, Proteobacteria and Chloroflexi (Gallardo et al. 2013; Fonseca et al. 2022). This region is within the Humboldt Sulfuretum (HS) system, which lies under the oxygen minimum zone (OMZ) off the Peru–Chile Undercurrent. This system constitutes a sulphur compound-rich (Gallardo et al. 2013; Gallardo et al. 2013) and geologically ancient

ecosystem that harbours extensive mats of large filamentous multicellular bacteria (Gallardo 1963; Maier and Gallardo 1984). While certain bacteria are unusually large in all dimensions, others are long and thin or consist of multiple cell assemblages, with cell sizes ranging from tens to hundreds of micrometres (Ionescu and Bizic 2019). This diversity prompted an ecological distinction: (a) 'macrobacteria', characterised by cells that are typically less than 10 μm in diameter, and (b) 'megabacteria', which have wider filaments and often possess relatively large vacuoles (Gallardo and Espinoza 2007). Examples include *Cand*. Marithioploca araucensis (Salman et al. 2011, corrected) and *Cand*. Venteria ishoeyi (Fonseca et al. 2017). Both mega- and macrobacteria are found in marine and continental biotopes, leaving their signatures in the fossil records (Schopf 2006).

Taxonomically, these giant bacteria span several phyla (Ionescu and Bizic 2019). For instance, the phylum Chloroflexi contains diverse filamentous multicellular bacteria such as the family Anaerolineaceae (Speirs et al. 2019). Similarly, the genus *Desulfonema* in the family Desulfobacteraceae (phylum Desulfobacterota) comprises three species of gliding, mesophilic and filamentous sulphate-reducing bacteria (Kuever, Rainey, and Widdel 2015). These include *Desulfonema magnum*, which degrades aromatic compounds (Widdel, Kohring, and Mayer 1983; Schnaars et al. 2021), *Desulfonema limicola*, which can grow chemolithoautotrophically with H_2 and CO_2 as the sole sources of energy and carbon, respectively, using the reductive Wood–Ljungdahl pathway (WLP) (Schnaars et al. 2021; Schauder et al. 1989), and *Desulfonema ishimotonii* (Fukui et al. 1999), isolated from marine sediments in Tokyo Bay, Japan, which displays a nutritional profile similar to that of *Desulfonema limicola*.

Among the most conspicuous giant bacteria are those belonging to Beggiatoales (phylum Pseudomonadota), that is, *Beggiatoa* spp. and *Cand*. Marithioploca (Salman et al. 2011, corrected), often referred to as 'colourless bacteria' (non-photosynthetic). These lithotrophs derive energy by oxidising sulphide to sulphate and are confined to the upper organic-rich sediment layer, where they capitalise on strong chemical gradients of sulphur and oxidants, such as nitrate and oxygen (Jørgensen, Findlay, and Pellerin 2019). In this regard, adaptive features have evolved with the benthic oxygenation process, enabling these originally anaerobic bacteria to cope with the increasing oxygen levels near the benthic environment. Some species have developed unique adaptations to address the spatial or temporal displacements of sulphide and oxidants, including motility and storage of nitrate in vacuoles and elemental sulphur (Schulz and Jørgensen 2001). The presence or absence of vacuoles and their variations contribute to their diverse widths.

From an ecological perspective, the large morphology of these organisms may make them susceptible to random encounters with bacteriophages, which are viruses that infect bacteria. For instance, *Cand*. Venteria ishoeyi, which was isolated from the HS, contains several CRISPR arrays that exhibit numerous spacer sequences and *cas* genes (Fonseca et al. 2017). These characteristics indicate that phage interactions with this bacterium are common. However, the roles of phages and defence systems in large filamentous bacteria have not been sufficiently studied. In contrast, large filamentous bacteria have been reported to resist grazer predation more effectively than regular bacteria (Justice et al. 2008).

Although a significant amount of research has been conducted on certain filamentous bacteria in the HS, such as *Cand*. Marithioploca, their taxonomy and lifestyle, particularly the uncultured so-called macrobacteria, remain unclear. This study aimed to provide insights into the taxonomy and major genomic features of 15 uncultured filamentous bacteria isolated from the HS. Genome analysis of these bacteria has revealed several new species, genera and a family across different taxonomic branches, as well as distinctive metabolic capacities and defence systems.

2 | Methods

2.1 | Sample Collection

Sulphidic sediment samples were collected on 6 February 2009, in the Bay of Concepción, off the central coast of Chile, from two sampling stations: Station 4 (36°38,1915 S—73°02,198 W) at a 29-m depth, and Station 7 (36°36,292 S–73°00,022 W) at 37-m depth (Figure 1). Sampling was carried out onboard the 'Otilia' motorboat using a gravity core sampler, equipped with a sampling tube of polymethyl methacrylate, measuring 1 m in length and 5 cm in diameter. The temperature of the overlying water at the sample sites was 15°C in Station 7 and 13.6°C in Station 4. The dissolved oxygen in the area, near the bottom, varied between approximately 0.2 mL L^{-1} during mid-summer and 1.3 mL L^{-1} in late winter. At the bottom, redox measurements indicated highly reduced conditions, with the notable presence of organoleptic hydrogen sulphide (Gallardo et al. 2013).

2.2 | Bacteria Isolation, MDA Amplification and DNA Sequencing

After rapid transportation to the J. Craig Venter Institute (JCVI), La Jolla, Ca., USA, 15 microbial filaments were carefully extracted from the sediment samples by micromanipulation (Ishoey et al. 2008). Filaments were isolated using an A1 Zeiss Axio Observer and Eppendorf Micromanipulation System. The selection of filament morphology determined the size of the microcapillaries used, which ranged from approximately 10 to 70 μm. The filaments were isolated, and whole-genome amplification was performed during two distinct sessions on freshly collected samples. Initially, the samples were diluted in sterile filtered seawater (Bay of Concepción) to reduce the cell densities. Subsequently, the selected filaments were isolated using micromanipulation, washing and cleaning in sterile filtered seawater (34.5 practical salinity units). The filaments were then transferred to 0.5-μL or 1-μL sterile phosphate-buffered saline (PBS) at potential hydrogen (pH) 7.4 in a 200-μL polymerase chain reaction (PCR) tube for whole-genome amplification (WGA). Multiple displacement amplification (MDA) was performed using a GenomiPhi HY kit (GE Healthcare, USA). Amplification was terminated after 6 h at 30°C. The WGA products were diluted 2-fold in Tris-EDTA (TE) buffer (pH 8.0) and stored at −20°C. A 20-fold dilution was prepared as a working solution for PCR analysis, quantification and assessment of

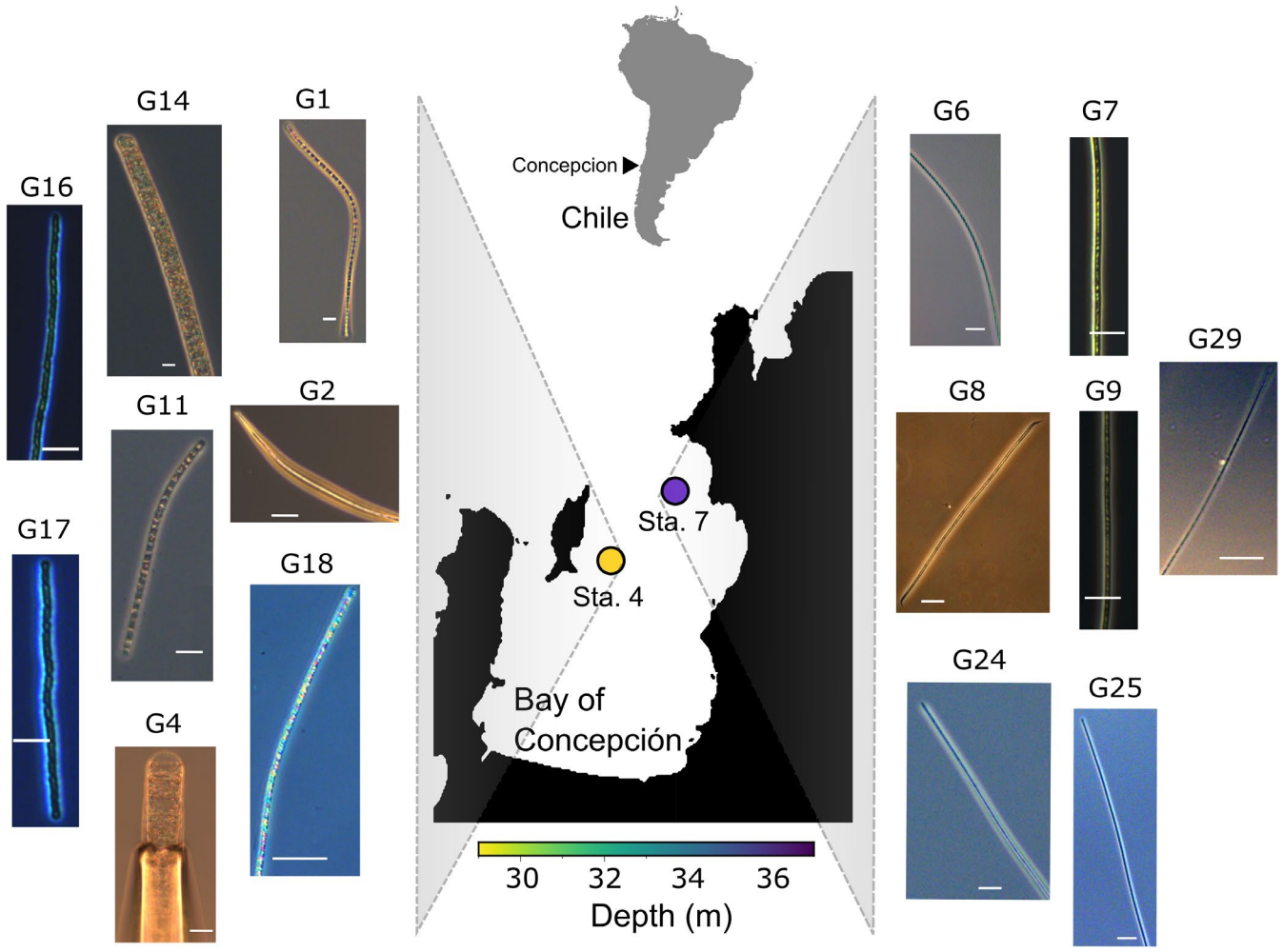

FIGURE 1 | Map of sampling points and microphotographs of the collected bacteria. The microphotographs show 15 filamentous multicellular bacteria collected from Station 4 (Sta. 4) at a 29-m depth and Station 7 (Sta. 7) at a 37-m depth, located in the Bay of Concepción, Chile. The white line in the images corresponds to a scale of 10 μm.

purity by direct sequencing of the 16S gene from the PCR product using universal primers. DNA samples with complete 16S rRNA gene sequences of interest from the WGA were sequenced on an Illumina MiSeq platform using the Nextera XT DNA Library, 300 bp paired-end.

2.3 | Data Analysis

2.3.1 | Preprocessing and Genome Assembly

After sequencing, the raw reads were subjected to quality control and preprocessing. Quality control was performed using FastQC v0.11.9 (FastQC. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), which gives a fast and comprehensible way to evaluate the quality of raw reads. The PRINSEQ-lite 0.20.4 software (Prinseq-lite. http://prinseq.sourceforge.net) was used to remove low-quality reads-bases (<20 Phred quality score) and short sequences (<100 bp), while adapters and primers were trimmed. The 15 genomes were built through *de novo* genome assembly using the program SPAdes v3.15.1 (Bankevich et al. 2012) with the options '—careful—sc -t 12 -k 21,33,55,77,99,111,127'.

2.3.2 | Genome Edition and Refinement

The binning was carried out using the Metabat 2 v2.12.1 software (Kang et al. 2019), following the removal of scaffolds <1 Kbp. Genome completeness and contamination were assessed using CheckM v.1.2.1 (Parks et al. 2015) and BUSCO v5.3.2 (Simão et al. 2015) programs with default parameters. After binning, the CheckM 'merge' method was used to evaluate bin pairs and merge them when completeness increased by at least 10% and contamination remained below 3%. The 'outliers' and 'taxon_profile' methods of the RefineM v.0.0.14 program (Parks et al. 2017) were used to assess scaffolds in bins that deviated from the GC mean, tetranucleotide signature, depth of coverage and taxonomy consensus. The genomes have been deposited under Bioproject PRJNA982541 in the NCBI database (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA982541/).

2.3.3 | Structural and Functional Annotation

Structural and functional annotation of the genomes was carried out using Prokka v1.14.6 (Seemann 2014) and GhostKOALA (Kanehisa, Sato, and Morishima 2016) from the Kyoto

Encyclopedia of Genes and Genomes (KEGG). In addition, metabolic pathways were identified using the mode 'find' (pathway and reaction finder) of the gapseq v1.2 program (Zimmermann, Kaleta, and Waschina 2021) and the MetaCyc database as the reference. Furthermore, the structural annotation was manually revised, and the InterProScan program (Jones et al. 2014) was used to complete the metabolic pathways.

The presence of CRISPR cassettes and associated genes was evaluated through the server 'CRISPRs webserver' (https://crisprcas.i2bc.paris-saclay.fr/), antiphage defence systems were analysed using the program DefenseFinder (Tesson et al. 2022), and prophages were analysed by 'Prophage Hunter' (Song et al. 2019).

Metabolic pathways detected across all bacteria (412 pathways) through gapseq were utilised to execute non-metric multidimensional scaling (NMDS) using the Euclidean distance matrix. A percentage similarity breakdown (SIMPER) was performed to identify primary contributors to the NMDS group. Both analyses were performed using the vegan package (Dixon 2003) in R v3.6.3.

2.4 | Phylogeny and Taxonomic Identification

The draft and complete genomes of Beggiatoales (28), Firmicutes (15), Chloroflexi (15) and Desulfobacterales (23) were obtained from the NCBI database. The marker protein sequences were identified, concatenated and aligned using the GTDB-Tk v2 program (Chaumeil et al. 2022) in the 'identify' and 'align' modes. A phylogenetic tree was constructed using the IQ-TREE v1.6.12 program (Minh et al. 2020) with the TESTNEW (TN + F + I + G4) model option and 500 bootstraps. The tree was visualised using FigTree v1.4.4. (https://github.com/rambaut/figtree/releases/). The taxonomy was evaluated using the Average Nucleotide Identity (ANI) index for whole-genome comparison with nearby reference genomes using the pyani v0.2.8 program (Pritchard, Cock, and Esen 2019). An ANI > 96% was used for species identification (Richter and Rosselló-Móra 2009). In addition, the maximum Average Amino Acid Identity (AAI) was determined using the Microbial Genomes Atlas (MIGA) server (Rodriguez-R et al. 2018) and taxonomic classification by the routine 'classify_wf' of the GTDB-Tk v2 program. Novel species, genera and families were proposed based on the results of the phylogenetic tree, AAI, ANI and 16S rRNA gene sequence identity (Yarza et al. 2014). Nomenclatural taxa were submitted to the Code of Nomenclature of Prokaryotes Described from Sequence Data (SeqCode) (Hedlund et al. 2022; Whitman et al. 2022) when the genome had ≥ 90% completeness and ≤ 5% contamination.

3 | Results and Discussion

3.1 | Novel Species, Genera and a Family in Four Phyla

The 15 bacterial specimens were identified and classified into four different taxonomic groups: (1) to the family Beggiatoaceae in the phylum Pseudomonadota, (2) to the family Desulfobacteraceae (phylum Desulfobacterota), (3) to the order Clostridiales (phylum Chloroflexi) and (4) to the family Clostridiaceae (phylum Firmicutes). Draft genome statistics are detailed in Table 1 and Appendix 1.

3.2 | Five Novel Species Within the Order Beggiatoales

The G18 bacterium was identified as *Candidatus* Venteria ishoeyi (Fonseca et al. 2017) (Figure 2), with an ANI of 99% (Figure S1A). *Candidatus* Venteria ishoeyi is a non-vacuolated chemolithotrophic sulphide-oxidising bacterium belonging to the Beggiatoaceae family. The G18 bacterium was proposed as the type species, named *Venteria ishoeyi*^{Ts}, in the SeqCode registry (refer to Supplementary 1 with the protologue), given its higher genome quality (98% completeness; 0.0% contamination) compared to the formerly described genome. The SeqCode registry uses genome sequence data for the typification of microorganisms, following rules similar to those of the ICNP for priority (Whitman et al. 2022).

The G4 bacterium was identified as *Candidatus* Marithioploca araucensis (Salman et al. 2011, corrected), placed in a monophyletic clade with an ANI of 97%. Over several decades, studies on sublittoral soft bottoms in the coastal upwelling ecosystem off central Chile have confirmed the presence of this bacterium as a component of extensive mats, primarily composed of *Thioploca* spp. (Zopfi, Böttcher, and Jørgensen 2008; Gallardo et al. 2013). *Candidatus* Marithioploca araucensis is still recognised in the order Thiotrichales; however, this taxonomic nomenclature is considered illegitimate. To avoid confusion and for practical reasons, we will hereafter label it as Beggiatoales. In addition, the addition of genomic information for this bacterium, despite its low genome quality, was still considered valuable.

According to phylogenetic and taxonomic analyses (Table S1), G1, G11 (are the most closely related with an ANI of 82%) and G14 (highest ANI of 78% with G4) represent novel species in the Beggiatoaceae family (Beggiatoales order). The G14 bacterium is a novel genus closely related to *Beggiatoa* spp. On the other hand, the 16S rRNA gene sequence places the G11 bacterium within the genus *Cand*. Halobeggiatoa (Grünke et al. 2012). The G11 bacterium was proposed as the novel species *Halobeggiatoa hulotii*^{Ts} (Table 1) in the SeqCode registry. This proposal aims to amend the condition of *Candidatus* of the Halobeggiatoa genus, which is currently described using only 16S rRNA gene sequences. The addition of a high-quality genome (98% completeness and 0.0% contamination) to the SeqCode registry supports this proposal. Additionally, this finding expands the geography of *Halobeggiatoa* from cold seeps and coastal sediments off Norway and Japan (Grünke et al. 2012) to the HS system off Chile.

3.3 | Six Novel Species and a Novel Genus Within the Order Desulfobacterales

Bacteria G2, G7, G8, G9, G16 and G17 are novel species in the family Desulfobacteraceae (synonym Desulfococcaceae), order Desulfobacterales. ANI values with the closest relatives from the databases ranged from 71% to 75% (Figure S1A). G2, G8 and G16

TABLE 1 | Nomenclature types of new species and main genome statistics. Genome size in millions of base pairs (Mbp), number of Contigs, N50 in kilobases (1 kbp = 1000 bp), guanine–cytosine content (GC%), number of coding sequences (CDSs), completeness of genome in percentage (Comp.); level of contamination (Cont.) and type of CRISPR-Cas system.

Genome	Taxa	Genome size (Mbp)	Contigs (Scaffolds)	N50 (Kbp)	GC%	CDSs	Proteins of known function	Comp.%	Cont.%	CRISPR-Cas type
G1	Beggiatoales	2.2	465 (457)	6.9	34	1970	858	76	0	—
G4	Beggiatoales (Thiotrichales)	2.3	1035 (1030)	2.5	40	1857	619	38	0	III-B
G11	*Halobeggiatoa hulotii*[Tsa]	5.2	332 (326)	25.1	34	3909	1600	98	0	II-B/I-F
G14	Beggiatoales	6.6	549 (531)	16.9	36	4560	1665	87	1	III-B
G18	*Venteria ishoeyi*[Tsa]	4.4	126 (115)	67.7	43	3528	1701	98	0	I-F/I-E
G2	Desulfobacterales	5.6	334 (288)	30.9	47	4516	1394	68	0	III-A/I-B
G7	*Desulfofilum molinae*	5.0	169 (164)	43.6	43	4118	1763	62	0	I-D
G8	*Desulfonema wilhelmii*[Tsa]	6.0	412 (389)	22.8	46	4692	1656	90	0	III-A
G9	*Desulfofilum molinae*[Tsb]	6.6	206 (197)	50.8	43	5436	2344	93	0	III
G16	Desulfobacterales	4.8	412 (408)	24.1	42	3771	1689	83	0	I-E
G17	Desulfobacterales	4.9	952 (947)	9.0	40	4024	1806	64	2	III
G6	*Sedimentiflexa bahamondei*	4.6	145 (130)	64.2	46	3351	1345	85	0	III-B
G24	*Sedimentiflexa* sp. HS	4.0	212 (206)	35.1	46	2901	1178	75	0	III-B
G25	*Sedimentiflexa bahamondei*[Tsc]	4.9	165 (143)	55.3	46	3576	1496	90	0	III-B
G29	*Filamentum carolae*[Tsa]	2.1	268 (266)	13.6	28	1929	1134	91	3	—

[a]Nomenclatural types proposed for novel species submitted to the SeqCode registry. These genomes meet the quality requirements to generate a novel nomenclatural type (> 90% completeness; < 5% contamination).
[b]Novel genus (submitted to SeqCode).
[c]Novel family: Sedimentiflexaceae (order Anaerolineales). Submitted to SeqCode.

form novel species-level clades in the genus *Desulfonema*. G8 was proposed as *Desulfonema wilhelmii*[Ts], whereas G7 and G9 are novel genus and species, similar to *Desulfonema* (Figure 2, Figure S1B and Table S1), proposed as *Desulfofilum molinae*[Ts] (both submitted to SeqCode), with G9 as the SeqCode-type species. The G17 bacterium is proposed as a novel species within the genus *Desulfobacula*, based on its 16S rRNA gene identity sequence (Table S1) and the phylogenetic tree. Interestingly, the description of *Desulfobacula* in the literature does not align with the morphology of the G17 bacteria. *Desulfobacula* is described as consisting of single or pairs of oval-coccoid or slightly curved cells (with dimensions of 1.5 × 1.2–3.0 μm) (Galushko and Kuever 2015). In contrast, the G17 bacterium comprises oval cells arranged in rows, which suggests that it may represent a novel species within the genus *Desulfobacula*. Considering the low contamination of the G17 genome (2%) and its evolutionary relationship, this hypothesis presents a reasonable possibility.

3.4 | Three Novel Species, a Novel Genus and a Novel Family in the Phylum Chloroflexi

According to the phylogenetic tree, G6, G24 and G25 demonstrated evolutionary proximity to the Caldilineaceae and Anaerolineaceae families (Figure 2) within the phylum Chloroflexi. Although their genome similarity with these reference bacteria is low, the 16S rRNA gene sequence identity falls below the cut-off for family identification (86%) (Yarza et al. 2014). The 16S rRNA gene sequences are < 91% identity with undescribed Chloroflexi bacteria from the NCBI database and ANI values of < 82% (Figure S1). The results indicated that G6 and G25 belong to the same novel genus and species (ANI = 98%) (Figure S1C), which was proposed as *Sedimentiflexa bahamondei*[Ts] in the SeqCode registry. The G24 bacterium belongs to the same novel genus with a 16S rRNA gene identity = 97%, ANI = 82% with G6 and G25 and is proposed as

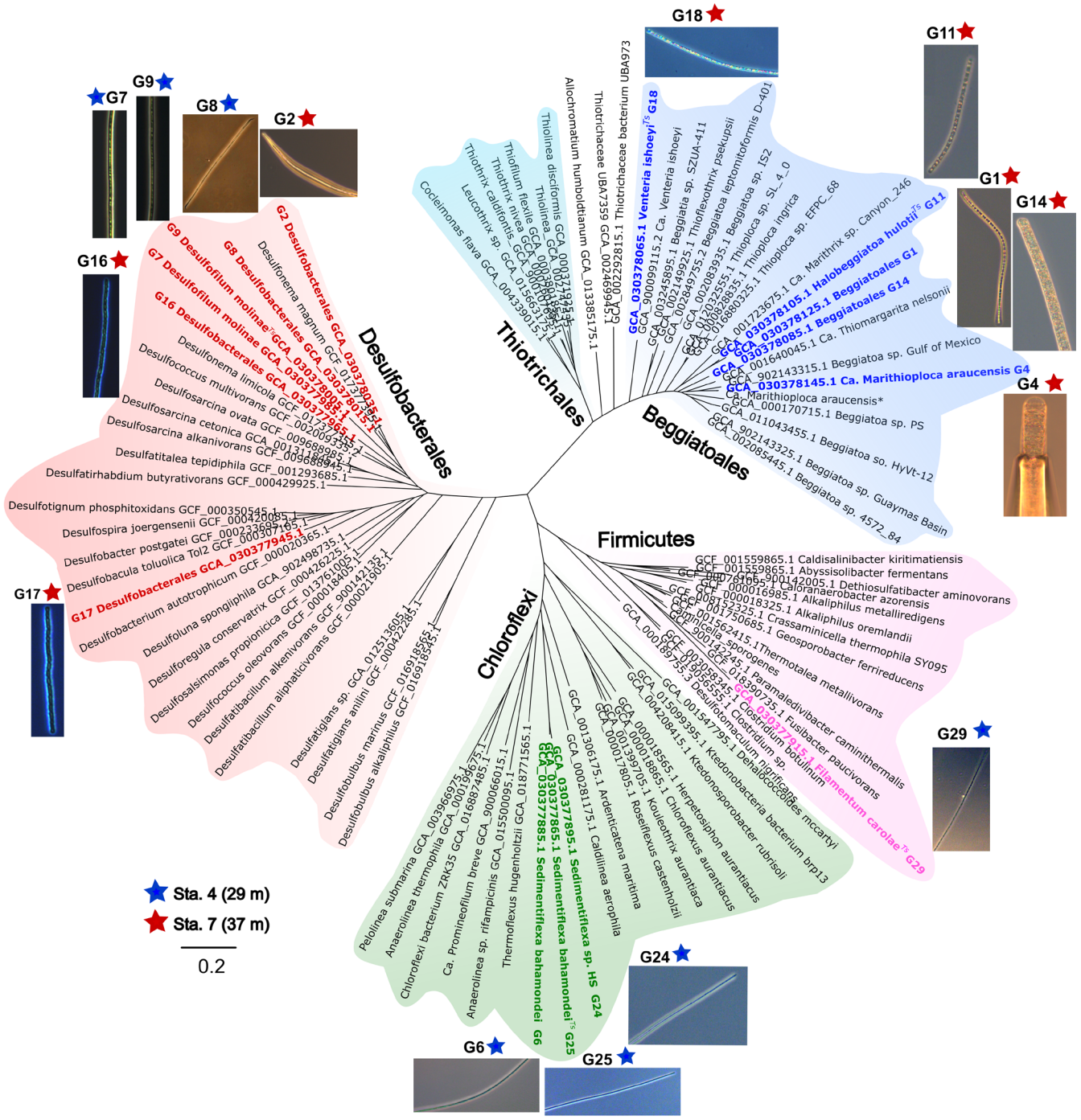

FIGURE 2 | Phylogenetic tree. The tree included 28 Beggiatoales and Thiotrichales, 15 Firmicutes, 15 Chloroflexi and 23 Firmicutes draft and complete genomes collected from the NCBI database, along with the 15 query genomes. Marker protein sequences in the genomes were identified, concatenated and aligned using the GTDB-Tk. The IQTree program was used to build the phylogenetic tree using the model TESTNEW (TN + F + I + G4) and 500 bootstraps. The red and blue stars represent sampling stations.

Sedimentiflexa sp. HSG24 but not submitted to the SeqCode registry, as its genome is < 90% complete. These three bacteria form a novel family, proposed here as Sedimentiflexaceae, and submitted to the SeqCode registry within the phylum Chloroflexi (Supplementary 1). Notably, the phylum Chloroflexi has been reported as the third most abundant phylum in the sediments of the sublittoral region under study, with numerous undescribed forms (Fonseca et al. 2022). Thus, the present results shed light on the previously unknown forms of Chloroflexi in the HS system.

3.5 | A Novel Genus and Species of Filamentous Bacteria Within the Phylum Firmicutes

The G29 bacterium was placed in the same clade as *Fusibacter paucivorans* and closely related to *Clostridium* within the phylum Firmicutes. The anaerobic chemoorganotrophic bacterium *Crassaminicella* sp. SY095 (Family Clostridiaceae) exhibits the highest genome identity with the G29 bacterium (ANI = 72%), whereas *Fusibacter paucivorans* and the G29 have an ANI of 69% (Figure S1D, Table S1 and Appendix 1). Furthermore, the

16S rRNA gene sequence of G29 was classified as Clostridiales using the RDP classifier database. Thus, the results suggest that the G29 is a novel genus and species, proposed here as *Filamentum carolae*^{Ts} (submitted to SeqCode registry), within the family Clostridiaceae and phylum Firmicutes (synonym Bacillota). Filamentous Firmicutes bacteria are commonly associated with fish symbionts such as *Epulopiscium* spp. (within the family *Lachnospiraceae*) found in surgeonfish and *Metabacterium polyspora* (Ionescu and Bizic 2019), which are unusually large bacteria found in the intestinal tract of guinea pigs (Chatton and Pérard 1913). Similarly, filamentous Firmicutes bacteria with streptococcus-like morphology are occasionally observed in activated sludge plants (Eikelboom and Geurkink 2002) (refer to Appendix 1 for more details). Interestingly, free-living filamentous forms of the phylum Firmicutes have been poorly described in the literature. Thus, *Filamentum carolae*^{Ts} appears to be an unusual species within the family Clostridiaceae and phylum Firmicutes, particularly in environmental sulphidic sediments.

3.6 | Metabolic Capacities at the Sediment–Water Interface

The bacterial genomes under study show potential for organoheterotrophic metabolism and utilisation of a wide range of organic matter. The results of the central carbohydrate metabolism, acetate consumption, fermentation and sugar biosynthesis are shown in Figures 3 and S2, Appendix 2 and Appendix 3. Figure S3 and Appendix 4 describe the genes in the main metabolic pathways using KEGG categories.

The glycolysis pathway (including glycolysis EMP), except in G4, was identified in all Beggiatoales genomes, along with Desulfobacterales G9 and G16, Chloroflexi G6 and G25, and Firmicutes G29 (Figures 3 and 4). The analysed bacteria presented different forms of glucose phosphorylation using polyphosphate and/or ATP, showing diverse kinase activities. Although there is little evidence that large sulphur bacteria grow on sugars, genes encoding glycolysis are commonly observed in large Beggiatoaceae bacteria that use polyphosphate glucokinase or ATP glucokinase, such as in *Cand*. T. nelsonii Thio36 and *Cand*. Maribeggiatoa sp., *Beggiatoa alba* B18LD and *Thioploca ingrica* (Winkel et al. 2016). In contrast, the TCA cycle, except in the genomes of the Desulfobacterales G7 and Firmicutes G29 bacteria, was identified in all studied bacteria. Conversely, β-oxidation of fatty acids was sparse and only present in the genomes of Desulfobacterales bacteria (85% completeness), whereas some scattered genes were identified in Chloroflexi genomes (refer to Appendix 3 and 8).

The bacterial genomes under investigation exhibit a versatile repertoire of metabolic pathways capable of metabolising organic matter. However, these capabilities have distinct limitations. For instance, Desulfobacterales appear to be able to utilise fatty acid catabolism to produce acetyl-CoA, which they likely use as a substrate for the TCA cycle, unlike other bacteria. Thus, managing a broad repertoire would be advantageous when the microenvironment changes the availability of labile organic matter and inorganic energy sources are scarce or incapable of oxidising it. In addition, these bacteria may play a crucial role in the biogeochemical cycling of carbon, phosphorus and sulphur, contributing to the breakdown of organic matter and the recycling of nutrients within the sediment ecosystem.

Interestingly, Chloroflexi G6 and G25 and Firmicutes G29 bacteria harbour the *tresS* gene, which encodes trehalose synthase/amylase TreS, involved in the production of trehalose. Furthermore, it was found that several genomes of Beggiatoales, Desulfobacterales and Chloroflexi encode glycogen synthase (UDP–glucose–glycogen glucosyltransferase), polyphosphate kinase and phosphate transporter proteins, which are involved in the formation of polyphosphate inclusions (Figure S8). These bacteria also engage in sulphur reduction via polysulphides (Figure 3). This potential energy source may be utilised during famine periods, as has been proposed for hadal Chloroflexi (Liu et al. 2022). They employ a 'feast-or-famine' metabolic strategy, synthesising energy storage compounds such as trehalose and polyphosphate (for more details, refer to Appendix 3). Additionally, regulatory modules respond to changes in nutrient conditions, allowing them to utilise the stored energy (Liu et al. 2022). Moreover, internally stored sulphur granules can be utilised by Beggiatoaceae bacteria via the rDsrAB genes (Mußmann et al. 2007; Dahl, Friedrich, and Kletzin 2008). This is consistent with the observation that Thioploca may migrate to nitrate-rich layers to access the stored energy (Zopfi, Böttcher, and Jørgensen 2008). Thus, most of the studied specimens across all bacterial groups would have the potential to store and use internal energy sources during famine or migration periods.

All bacterial genomes examined in this study, except Chloroflexi G24 and Firmicutes G29, may utilise acetyl-coenzyme A synthetase (encoded by *acsA*) to incorporate acetate into their overall metabolism. Acetate oxidation has been used as an indicator of organic carbon decomposition in anoxic sediments (Sørensen, Christensen, and Jørgensen 1981; Finke, Vandieken, and Jørgensen 2007). Conversely, the genomes of Firmicutes G29 and Chloroflexi G6–G25 bacteria exhibit the ability to produce acetate through fermentation (via phosphate acetyltransferase and acetate kinase). These findings indicate that the majority of the studied bacteria exhibit chemoorganotrophic behaviour and use acetate as a potential energy source for carbon. Firmicutes G29 and Chloroflexi G6–G25 bacteria, however, would generate acetate through fermentation (Figure 4).

3.7 | Carbon Fixation: Chemolithoautotrophy via the rTCA and CBB Pathways in Beggiatoales and Chloroflexi and WLP in Desulfobacterales

CO_2 fixation pathways were identified in Beggiatoales, Chloroflexi and Desulfobacterales. The Calvin–Benson–Bassham (CBB) pathway was present in all Beggiatoales bacteria (Figure 3), except in G4, with 70%–84% completeness and key genes. Interestingly, the genome of Beggiatoales G18 harbours the reverse TCA (rTCA) pathway, though incomplete. A similar pattern was observed for Chloroflexi G6 and G25. The genomes of Beggiatoales G11, G14 and G18 harbour the *cbbM* gene, which encodes ribulose bisphosphate carboxylase/

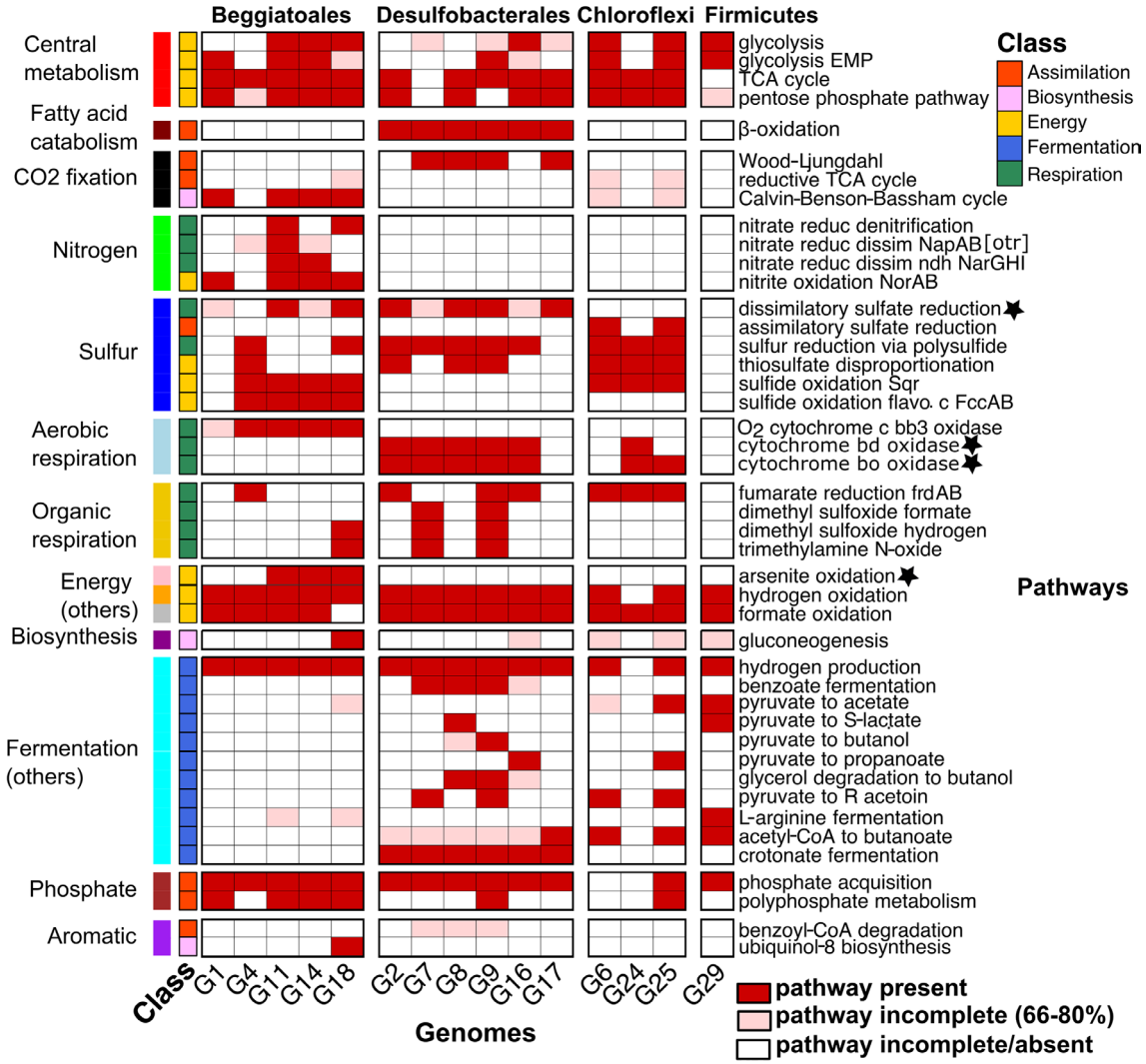

FIGURE 3 | Metabolic pathways at different levels of completeness. When pathways contained reactions larger than 80% in the MetaCyc database, they were considered present and functional (red). Similar to pathways with more than 66% completeness and key reactions. Pathways with reactions ranging between 66% and 80% without key reactions were classified as incomplete (pale red colour). Pathways with reactions below 66% were considered absent or non-functional (white tiles). A black star in the dissimilatory sulphate reduction pathways indicates that Beggiatoales genomes have genes for the canonical dissimilatory sulphate reduction pathway (DSR), which probably works in reverse and includes oxidative DsrAB types. The presence of black stars in arsenite oxidation indicates that Beggiatoales genomes encode enzymes for arsenite oxidation. However, the purpose of this process is unclear, whether it is for energy gain or detoxification, as well as for cytochrome bd and bo in Desulfobacterales and Chloroflexi.

oxygenase (RubisCO) form II (Figure S4), which catalyses the first major step in carbon fixation. Furthermore, the *cfxP* gene, encoding phosphoribulokinase, was identified in the genomes of Beggiatoales G1, G11, G14 and G18 (Figure S4). Similarly, the CBB pathway was 76% complete in the G6 and G25 Chloroflexi bacteria; however, genes encoding RuBisCO were not identified, which are essential for fixing CO_2 via CBB. Notably, only a few bacteria are known to fix CO_2 via both the rTCA and CBB cycles. These two alternative pathways have been observed primarily in sulphur-oxidising organisms, tubeworm symbionts and in the large sulphur bacteria *Cand*. Thiomargarita Thio36 (Winkel et al. 2016; Rubin-Blum, Dubilier, and Kleiner 2019). However, the sole presence of the CBB cycle has been reported in filamentous Chloroflexi bacteria found in activated sludge wastewater (Speirs et al. 2019; McIlroy et al. 2016), whereas the simultaneous presence of both pathways for CO_2 fixation remains unknown.

The WLP for CO_2 fixation was identified in Desulfobacterales: G7, G8, G9 and G17, maintaining 70%–90% pathway completeness (Figure 3). Indeed, genes such as *acsE*, encoding methyltetrahydrofolate:corrinoid/iron–sulphur protein methyltransferase, and *cooS*, encoding carbon monoxide dehydrogenase, were present only in Desulfobacterales genomes (Figure S4 and Appendix 2). These bacteria likely utilise WLP for acetate oxidation, encoding Acetyl-coenzyme A synthetase for the reversible conversion of acetate to acetyl-CoA, which has been observed previously (Gardner et al. 2006). Indeed, sulphate-reducing bacteria, such as *Desulfonema limicola* and *Desulfonema magnum*, utilise WLP in reverse (Schnaars et al. 2021). In this process, they couple the endergonic acetate oxidation to H_2 and CO_2 with the exergonic reduction of sulphate to sulphide (Ragsdale and Pierce 2008). Therefore, Desulfobacterales could utilise both the oxidative and reducing sides of the WLP to obtain energy, depending on substrate availability.

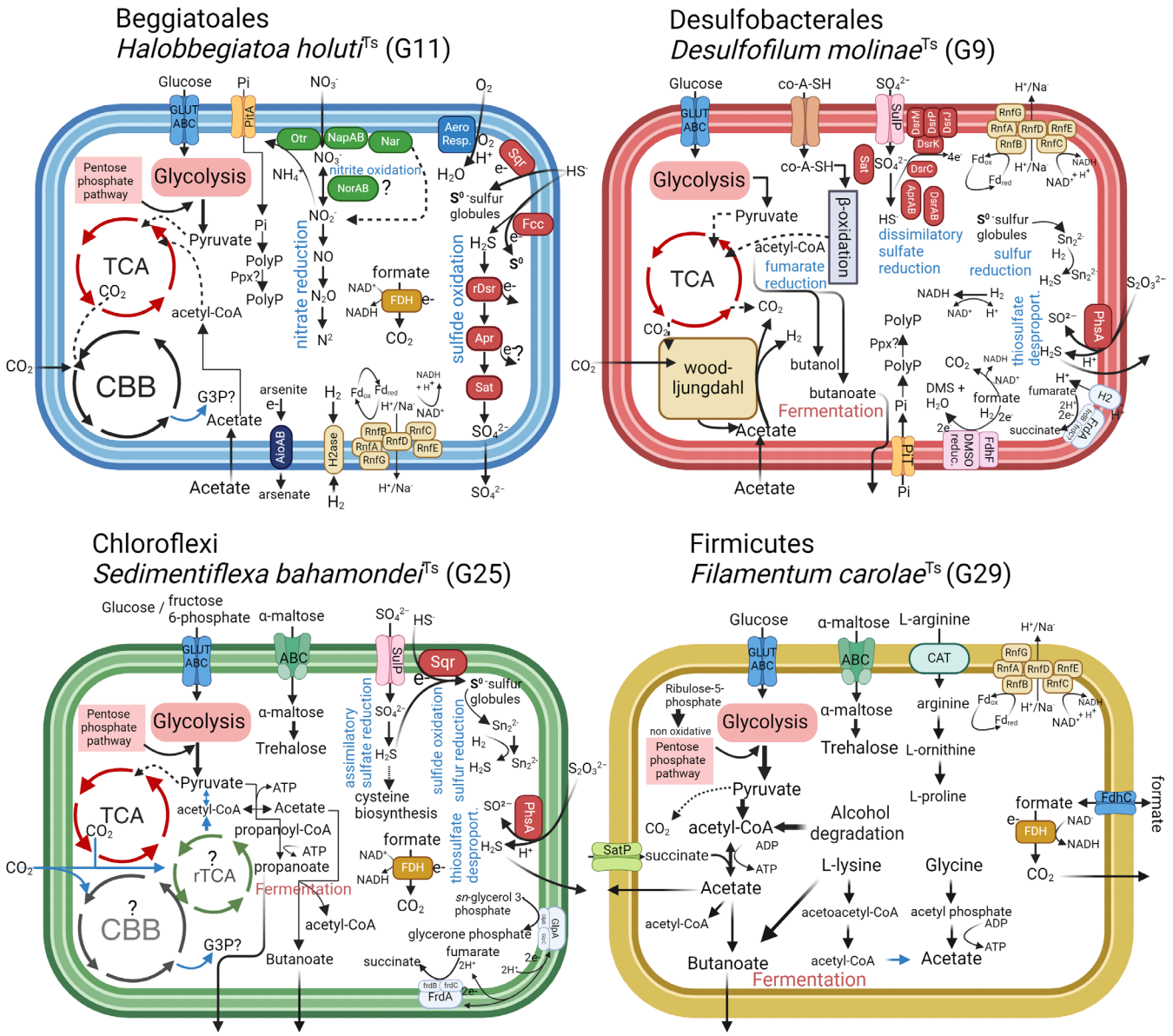

FIGURE 4 | Metabolic models for the novel, more complete bacterial genomes of Beggiatoales, Desulfobacterales, Chloroflexi and Firmicutes. ABC, ATP-binding cassette; Acetyl-CoA, acetyl-coenzyme A; Aero Resp., cbb3-type cytochrome c oxidase complex; AioAB, arsenite oxidase; Apr, phosphosulphate reductase; CAT, cationic amino acid transporters; CBB, Calvin–Benson–Bassham cycle; Co-A-SH, coenzyme A; DMSO reduc., dimethyl sulphoxide reductase; Dsr, dissimilatory sulphite reductase; Fcc, flavocytochrome c sulphide dehydrogenase; FDH, formate dehydrogenase; FdhC, formate transporter; FdhF, formate dehydrogenase H; FrdABC, fumarate reductase; G3P, glyceraldehyde 3-phosphate; GlpABC, anaerobic glycerol-3-phosphate dehydrogenase; H2, hydrogen; H2ase, hydrogen reductase; NapAB, Periplasmic nitrate reductase; Nar, membrane-bound respiratory nitrate reductase; NorAB, nitric oxide reductase; Otr, octaheme cytochrome; PhsA, thiosulphate reductase; Pi, inorganic phosphate; PitA, Low-affinity inorganic phosphate transporter; PpX, exopolyphosphatase; rDsr, reverse dissimilatory sulphite reductase; RnfABCDEG, ion-translocating oxidoreductase complex; rTCA, reverse tricarboxylic acid cycle; S0, sulphur globules; Sat, sulphate adenylyl transferase/ATP sulphurylase; SatP, succinate-acetate permease; Sqr, sulphide:quinone oxidoreductase; SulP, sulphate transporter; TCA, tricarboxylic acid cycle; Wood–Ljungdahl, reductive acetyl-coenzyme A (acetyl-CoA) pathway.

3.8 | Inorganic Electron Donors: Large Filamentous Bacteria Use Several Inorganic Sources of Energy

3.8.1 | Sulphur

The findings suggest that bacteria from Beggiatoales and Chloroflexi are capable of sulphur oxidation. Genomes of Beggiatoales bacteria were found to harbour genes from the canonical dissimilatory sulphate reduction (DSR) pathway with 50%–100% completeness (Figure 3). However, it probably functions in reverse, as oxidative DsrAB types were identified through phylogenetic analysis utilising data from Müller et al. (2015). This reverse behaviour of DSR has been previously reported in large sulphur bacteria (Mußmann et al. 2007). In addition, the Beggiatoales G18 genome encodes for thiosulphate

oxidation, featuring the Sox pathway with the soxXYZAB genes (Figure S5). Other Beggiatoales and Desulfobacterales genomes analysed contained scattered Sox genes (Appendix 5). Sulphide oxidation is catalysed by sulphide-quinone reductase (Sqr), which converts hydrogen sulphide to elemental sulphur. Sqr was identified in the Beggiatoales G4, G11, G14 and G18 genomes as well as in all Chloroflexi genomes. Additionally, Beggiatoales genomes contain genes encoding the sulphur dehydrogenase system, composed of flavocytochrome c. This system consists of a large sulphide-binding flavoprotein (FccB) and a small cytochrome c (FccA) dedicated to sulphide oxidation (Figure S5). Thus, sulphur oxidation was a key factor (Appendix 6, Figure S9B–D) in clustering Beggiatoales in the NMDS analysis (Figure S9A). Furthermore, Beggiatoales bacteria exhibit greater flexibility, as they encode several enzymes for sulphide oxidation, which may be advantageous in accommodating variable environmental conditions, such as at the sediment–water interface. Meanwhile, the presence of the sulphide:quinone oxidoreductase (gene *sqr*) type II family in Chloroflexi has been previously reported (Bryant et al. 2012; Thiel et al. 2016; Liu et al. 2022), along with the ability to oxidise sulphide to polysulphides, which can even support autotrophic growth (Madigan, Petersen, and Brock 1974), which would be a fundamental source of energy for filamentous Chloroflexi bacteria in the HS.

3.8.2 | Alternative Inorganic Electron Donors

Hydrogen oxidation pathways were identified in all bacteria, except for Chloroflexi G24 (Figure 3). All Desulfobacterales and G29 genomes encode Hnd hydrogenase ($NADP^+$-reducing), which permits use of H_2 as an electron donor and energy source under anaerobic conditions, similar to that reported in *Desulfovibrio* (Badziong, Thauer, and Zeikus 1978). Moreover, Desulfobacterales genomes also harbour hydAB genes, which encode periplasmic [NiFe] hydrogenase subunits, whereas all Desulfobacterales and Firmicutes G29 genomes maintain almost or completely *hndABCD* operons encoding NADP-reducing hydrogenases. In contrast, Beggiatoales G1, G11 and G14 encoded the putative cytochrome b type Ni/Fe-hydrogenase 2 (Hyb complex), while G18, along with Chloroflexi G6 and G25, hosted a subunit of the NAD-reducing hydrogenase HoxS, making H_2 a widespread potential energy source for all studied bacterial groups. According to previous observations, in large sulphur bacteria, H_2 seems common as an energy source (Winkel et al. 2016). In this regard, we hypothesised that the Chloroflexi and Desulfobacterales bacteria in the present study could use H_2 as an electron donor for CO_2 reduction (in WLP).

Strikingly, the G11, G14 and G18 Beggiatoales bacteria harbour *aioAB* genes, associated with arsenite oxidation. The same has been previously reported for the large sulphide-oxidising *Candidatus* Thiomargarita nelsonii (Flood et al. 2016). However, whether these bacteria gain energy or perform arsenic detoxification requires further clarification.

Formate oxidation capacity was observed in all bacteria, except G18. Beggiatoales G1, G11 and G14 and Chloroflexi bacteria harbour the *fdnH* gene, which facilitates formate oxidation (formate dehydrogenase, nitrate-inducible, iron–sulphur). This enzyme would enable Beggiatoales bacteria to utilise formate as a major electron donor during anaerobic respiration, with nitrate serving as the electron acceptor. Nonetheless, it is unclear whether Chloroflexi bacteria can oxidise formate using this enzyme. In contrast, the remaining genomes harboured *fdhAB* genes, which are catalytic and small subunits. Formate oxidation has been previously reported in filamentous Firmicutes, such as *Trichococcus flocculiformis*, as well as in Chloroflexi bacteria (Fincker 2020). However, formate reductases in Beggiatoales, particularly in large sulphide-oxidising bacteria, have not been widely reported.

Additionally, genes encoding enzymes for nitrite oxidation were present in the Beggiatoales G1, G11, G14 and G18, facilitated by the nitrate reductase (cytochrome c): NirT subunit and *norAB* genes, which encode nitrite oxidoreductases. In this context, it has been postulated that Beggiatoa uses nitrite as an electron donor (Mußmann et al. 2007). Thus, these results suggest that Beggiatoales may gain energy through nitrite oxidation via nitrite oxidoreductase and nitrate reductases, using a reverse mechanism, analogous to nitrite-oxidising bacteria.

3.9 | Electron Acceptors: Anaerobic Respiration Is Dominant, While Beggiatoales Alternatively Uses Oxygen

3.9.1 | Nitrate

Nitrate respiration pathways were exclusively identified in the genomes of Beggiatoales. Among them, the G11, G14 and G18 genomes exhibit 75%–100% completeness in the nitrate respiration pathway for nitrate reduction through denitrification (Figure 3). This includes the presence of *napAB* genes encoding periplasmic nitrate reductases, *narGH* genes for membrane-bound nitrate reductase, *nirS* gene for nitrite reductases, *norBC* genes for nitric oxide reductases, cytochrome c and *nosZ* genes for nitrous oxide reductases (Figure S6). The G11 genome showed 100% completeness for denitrification and dissimilatory nitrate reduction pathways (Figure 4), harbouring the *napAB* genes and an *otr* homologous gene (90% identity). The *otr* gene encodes an octaheme c-type cytochrome that catalyses the reduction of nitrite to ammonium in *Shewanella oneidensis* (Atkinson et al. 2007). In contrast, Desulfobacterales genomes harbour the *narG* gene, which encodes the respiratory nitrate reductase subunit beta, and genomes G2, G9 and G17, along with Chloroflexi G6 and G25, contain the *nosZ* gene encoding nitrous oxide reductase. From an ecological perspective, nitrate is one of the preferred alternative electron acceptors mainly because of its low redox potential and high concentration in sediments (Sørensen, Jørgensen, and Revsbech 1979). Previous reports have indicated that large sulphur bacteria exhibit contrasting behaviours. Although *Thioploca* spp. are known to reduce intracellular NO_3^- to NH_4^+ rather than undergo denitrification (Zopfi et al. 2001), there is also evidence suggesting that denitrification activity may be attributed to *Beggiatoa* spp (Sweerts et al. 1990). What is certain is that nitrate, abundant in the sediments of the HS, would be able to sustain the respiration and proliferation of Beggiatoales bacteria when oxygen is unavailable.

3.9.2 | Sulphur

Sulphate reduction pathways were observed in all Desulfobacterales and Chloroflexi genomes (Figure 3). The canonical DSR pathway, which produces hydrogen sulphide through four main reactions, is prevalent in Desulfobacterales genomes with > 75% completeness. All Desulfobacterales genomes, except G2, contained *dsrAB* genes encoding sulphite reductases (Figure S7). However, G7 and G16 were found to lack the adenylsulphate reductase enzymes. In this regard, the canonical DSR pathway is widely described in the family Desulfobacteraceae and couples the oxidation of organic molecules with the reduction of sulphate (SO_4^{2-}) to sulphide (S^{2-}) (Widdel 1988). Remarkably, the G6 and G25 Chloroflexi genomes exhibited 100% completeness in assimilatory sulphate reduction (Figure 4). Thus, Desulfobacteraceae bacteria demonstrate the capacity to utilise sulphur for respiration through DSR, which has been extensively reported previously. In addition, Chloroflexi exhibit a high affinity and capacity for utilising sulphur compounds, which are abundantly available in the HS system.

3.9.3 | Organic Electron Acceptors

The presence of the *frdB* gene, which encodes the cytoplasmic catalytic domain of fumarate reductase from the fumarate reduction complex, was discovered in Beggiatoales G4, Desulfobacterales G2, G9 and G16, as well as in Chloroflexi genomes. The *hybAB* genes, which encode hydrogen:quinone oxidoreductase, which catalyses the electron transfer between hydrogen and fumarate, were identified in the Desulfobacterales G9 genome. The genomes of Desulfobacterales G9 and Chloroflexi contain the *glpABC* genes, which encode anaerobic glycerol-3-phosphate dehydrogenase (GlpABC respiratory enzyme) and allow growth with glycerol and fumarate. Additionally, the capacity for dimethyl sulfoxide (DMSO) respiration was identified in Beggiatoales G18, Desulfobacterales G7 and G9 (Figure 4), with the presence of the *dmsAB* gene, which encodes the catalytic subunit of the dimethyl sulfoxide reductase complex. Lastly, the ability for trimethylamine N-oxide (TMAO) respiration, facilitated by C-type cytochromes, was identified in G18 (Beggiatoales), G7 and G9 bacteria (Desulfobacterales).

Under oxygen-depleted conditions, bacteria are compelled to utilise alternative electron acceptors in place of oxygen (Jørgensen, Findlay, and Pellerin 2019). The anaerobic respiration capacity appears to be predominant across the analysed bacteria, employing various compounds, including sulphur and nitrate, which are largely available in the HS and organic compounds. This observation aligns with the oxygen depletion in the sediment and overlying waters within the study area.

3.10 | Aerobic Respiration

The key genes responsible for oxygen respiration, *ccoN* and *ccoP*, encode the terminal oxidase and facilitate the transfer of electrons from donor cytochrome c (part of the cbb3-type cytochrome c oxidase complex), which were identified in all Beggiatoales. Additionally, Beggiatoales harbour genes encoding the proton-pumping NADH:ubiquinone oxidoreductase respiratory complex I (Appendix 5). Oxygen has been reported to be an electron acceptor in large sulphur-oxidising bacteria; encoding a high-affinity cytochrome under microoxic conditions (Mußmann et al. 2007). On the other hand, Chloroflexi genomes contain the *nuo* and *qno* genes that encode cytochromes for NADH:ubiquinone oxidoreductase (Complex I). However, they do not encode cytochrome c oxidase. In this connection, the acquisition of genes involved in aerobic respiration has been consistently observed in the Chloroflexi classes (Ward et al. 2018). In contrast, Desulfobacterales genomes carried the *cydAB* and *appC* genes, encoding cytochrome bd-I and II ubiquinol oxidases (Figure 3), which reduce molecular oxygen even at sub-micromolar concentrations. Although it is unclear whether these genes are used for aerobic respiration or for O_2 detoxification in both bacterial groups (Ward et al. 2018; Cypionka 2000). Thus, according to the genomic features found here and previous reports, Beggiatoales use aerobic and anaerobic respiration pathways in a mixotrophic mode, which allows them to withstand periods of exposure to oxygen and anaerobic conditions, whereas in Desulfobacterales and Chloroflexi bacteria, this is unclear.

4 | Defence Systems and Multicellularity May Be Essential for the Survival of Large Filamentous Bacteria

Bacteria have evolved multiple lines of defence that collectively form what can be called the 'prokaryotic immune system' (Bernheim and Sorek 2020). These defences include mechanisms against bacteriophages, transposable elements and plasmids. Notably, several defence systems are prevalent in the analysed bacterial genomes, including CRISPR-Cas, restriction-modification systems (R-M systems), toxin-antitoxin systems (TA systems) and DNA phosphothiolation systems (DND).

The CRISPR-Cas system provides adaptive immune protection against foreign mobile genetic elements (MGEs). It consists of CRISPR arrays, with direct repeats separated by spacer sequences. Different types of CRISPR-Cas systems have been identified (Figure 5A), even among bacteria from the same taxonomic branch. This diversity suggests various origins and modes of incorporating these defence systems, possibly through horizontal gene transfer. The genome of Desulfobacterales G8 contains 239 CRISPR arrays and the maximum number of total spacer sequences (799 spacers), whereas the lowest number was found in the Firmicutes G29 (Figure 5A). The average bacterial count according to literature is approximately 50 (Garrett 2021). The most extensive single CRISPR array was identified in Beggiatoales G18 bacterium, with 151 repeat sequences (Figure 5C) co-located with CRISPR-Cas I-E type. All bacterial genomes carried *cas* gene sequences, except for the Firmicutes G29 genome (Figure 5B), and together form 'crRNA-effectors', which monitor cells for target nucleic acids (Brouns et al. 2008). Furthermore, CRISPR-Cas systems showed different gene arrangements across the genomes analysed (Figure 5C). Some were interrupted by uncharacterised gene sequences such as CRISPR-Cas III-B in G11 and G14. Likewise, CRISPR-Cas I-F had distinct arrangements in Beggiatoales G11 and G18, and CRISPR-Cas I-E varied in G18 and G16.

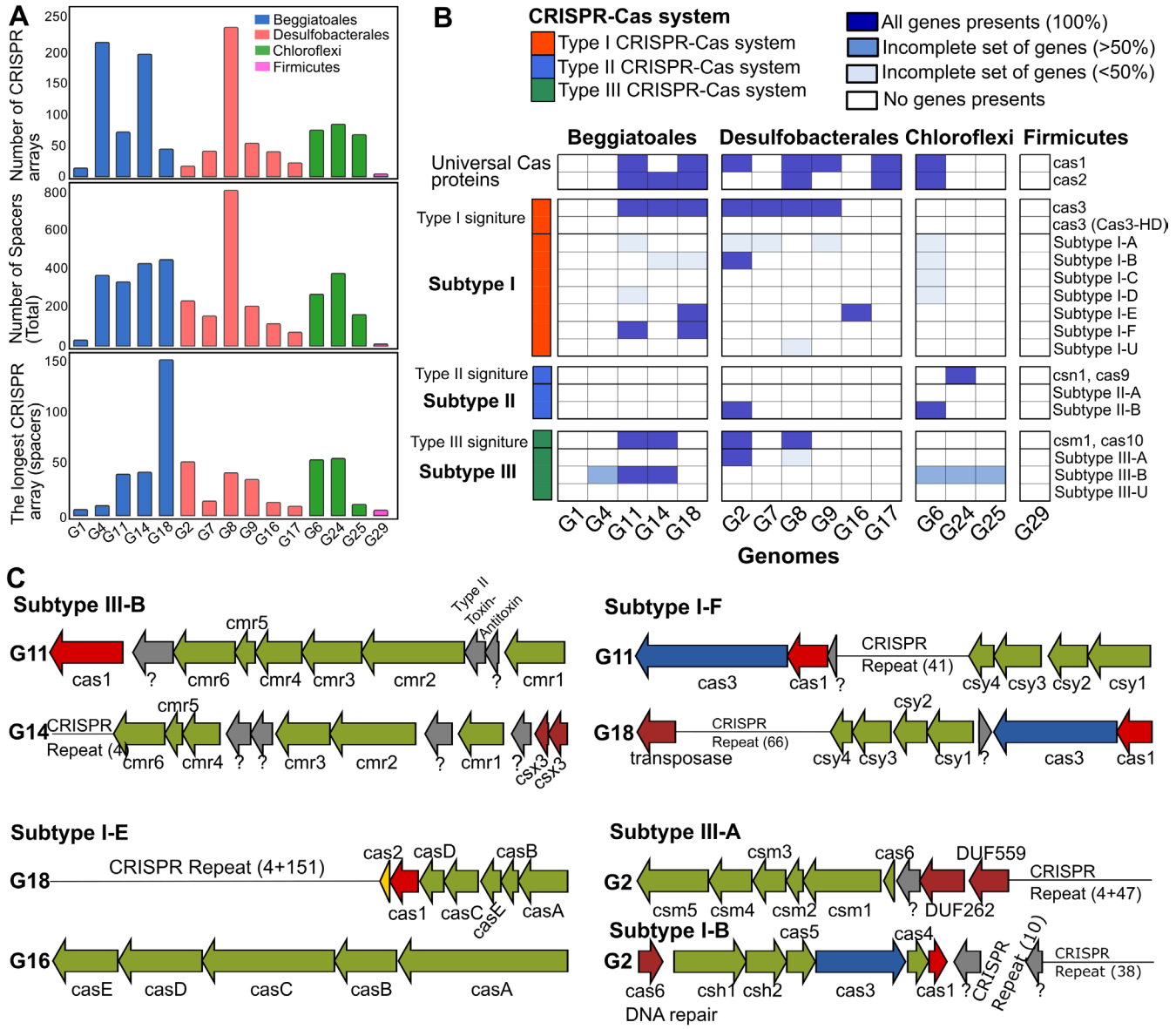

FIGURE 5 | CRISPR-Cas arrays. (A) Bar graph depicting the number of CRISPR arrays, total number of spacers and the longest single CRISPR array (measured in number of spacers) for each genome. (B) Chart illustrating the presence and organisation of different types of CRISPR-Cas systems by genome, according to signature genes. (C) Diagram showing the organisation of complete CRISPR-Cas systems in several bacterial genomes.

Bacteriophages modulate the abundance and diversity of bacterial communities. Previous investigations in sediments of the HS have demonstrated that virus-induced mortality corresponds to 44%–138% of the net bacterial production (Middelboe and Glud 2006). It is highly probable that bacteriophages exert a significant selective pressure on large bacteria in sedimentary environments. The atypical size, dense populations and extensive motility of filamentous multicellular bacteria may be the primary factors that facilitate random encounters with viruses. In this context, one indicator is the notably high frequency of spacer sequences in CRISPR-Cas arrays in G18 (*Venteria Ishoeyi*^{Ts}) and G8. Spacer sequences are foreign genetic elements incorporated into the host genome, such as viral sequences, after an infection event. Similarly, other large multicellular bacteria such as the sulphide-oxidising *Cand*. Electronema aureum GS (Kjeldsen et al. 2019) harbour an impressive number of 393 spacers in a single CRISPR array. This is exceptionally high compared with typical arrays of bacteria and archaea. According to previous studies, bacteria contain fewer than 50 and archaea fewer than 100 (Pourcel et al. 2020). Furthermore, *Desulfonema limicola* harbours over 50 defence systems (Tesson et al. 2022).

In addition to adaptive immune CRISPR-Cas systems, most analysed bacterial genomes harbour genes encoding four types of R-M systems, TA systems and DND defence systems (Appendix 7, Tables S2 and S3). Furthermore, several potentially active prophages were found across all bacteria (Table S4).

A comparative analysis of the filamentous bacteria from the present study with 16 non-filamentous bacteria (genome completeness 93%–100%), randomly selected from the same taxonomic branches and environment, was conducted using the DefenseFinder program to detect antiphage systems. The results revealed that filamentous bacteria harbour more antiphage

defence systems than non-filamentous bacteria, with an average of 24 (maximum of 55 in G14 Beggiatoales) versus nine (maximum of 22 in GCF_013385175.1 Allochromatium humboldtianum) (Table S5 and Figure S10). Although more extensive investigations are necessary, including larger databases, these results suggest that large filamentous bacteria experience high phage predation pressure and require a greater number and diversity of antiphage defence systems than non-filamentous bacteria in sedimentary environments.

According to a mathematical model, viruses contribute to the evolution of multicellularity at least in part (Iranzo et al. 2014). Thus, multicellularity of large bacteria could be an evolutionary defence strategy. One possible mechanism is that the infection eventually ceases and does not spread throughout the filaments. Therefore, while some filament cells die and are disrupted by the infection, others survive because the spread of the virus is impaired; for example, diffusion is no longer possible. Although these hypotheses require further validation, it is plausible to postulate that the presence of robust defence mechanisms in conjunction with multicellularity may determine the distinction between survival and mortality in large filamentous bacteria.

5 | Conclusions

Our results revealed novel genera and species within the families Desulfobacteraceae and Beggiatoaceae, and the phyla Firmicutes and Chloroflexi, which also include a novel family. These findings show that the HS environment is a rich and diverse source of unknown large filamentous bacteria, in particular, of the so-called 'macrobacteria'. Metabolically, large bacteria from the HS exhibited significant diversity, including a mixotrophic lifestyle that combined chemolithotrophy and chemoheterotrophy, in addition to aerobic and anaerobic respiration. Furthermore, various defence systems with distinct characteristics coexist in these large bacteria, which, in conjunction with their multicellularity nature, may represent the ultimate evolutionary adaptation to resist MGEs and phages.

Author Contributions

Alexis Fonseca: conceptualization, investigation, writing – original draft, methodology, visualization, writing – review and editing, data curation. **Thomas Ishoey:** methodology, writing – review and editing. **Carola Espinoza:** methodology, writing – review and editing. **Ian P. G. Marshall:** conceptualization, methodology, writing – review and editing, supervision. **Lars Peter Nielsen:** funding acquisition, writing – review and editing, supervision. **Victor Ariel Gallardo:** investigation, writing – review and editing, funding acquisition, conceptualization.

Acknowledgements

We acknowledge the fundamental financial contributions of both the National Agency for Research and Development of Chile (ANID—former CONICYT), Project N°1070552 (2008), and the International Sloan Foundation Project 'Census of Marine Microbes' (ICoMM), chapter of 'Census of Marine Life' (CoML) program, under the direction of Jesse H. Ausubel, Director of the Program for the Human Environment, The Rockefeller University, New York, USA. Next, we deeply acknowledge the most critical ('sine qua non') collaboration, which under the terms of our standing JCVI/UdeC MOU with the 'J. Craig Venter Institute', secured micromanipulations, DNA extraction, amplification and sequencing of the selected filamentous bacteria, besides a great hospitality. Last, but not least, we thank the Department of Biology, Aarhus University, Denmark, particularly the expert local laboratory technicians Mes. Susanne Nielsen, Britta Poulsen and Lykke Beinta Bjærge Bamdali. The authors acknowledge funding from the Program 'Doctorado en el extranjero Becas Chile' from ANID (former CONICYT).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The genomes generated and used for the analysis are publicly available in the Gene Data Bank repository with the accession Bioproject code PRJNA982541 https://www.ncbi.nlm.nih.gov/bioproject/PRJNA982541/.

References

Atkinson, S. J., C. G. Mowat, G. A. Reid, and S. K. Chapman. 2007. "An Octaheme c-Type Cytochrome From *Shewanella Oneidensis* Can Reduce Nitrite and Hydroxylamine." *FEBS Letters* 581, no. 20: 3805–3808.

Badziong, W., R. K. Thauer, and J. G. Zeikus. 1978. "Isolation and Characterization of Desulfovibrio Growing on Hydrogen Plus Sulfate as the Sole Energy Source." *Archives of Microbiology* 116: 41–49.

Bankevich, A., S. Nurk, D. Antipov, et al. 2012. "SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing." *Journal of Computational Biology* 19, no. 5: 455–477.

Bernheim, A., and R. Sorek. 2020. "The Pan-Immune System of Bacteria: Antiviral Defense as a Community Resource." *Nature Reviews Microbiology* 18, no. 2: 113–119.

Brouns, S. J., M. M. Jore, M. Lundgren, et al. 2008. "Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes." *Science* 321, no. 5891: 960–964.

Bryant, D. A., Z. Liu, T. Li, et al. 2012. "Comparative and Functional Genomics of Anoxygenic Green Bacteria From the Taxa Chlorobi, Chloroflexi, and Acidobacteria." *Functional Genomics and Evolution of Photosynthetic Systems* 33: 47–102. https://doi.org/10.1007/978-94-007-1533-2_3.

Chatton, E., and C. Pérard. 1913. "Schizophytes du caecum du cobaye. 1. Oscillospira guilliermondi ng, n. sp." *Comptes rendus des séances de la Société de Biologie* 74: 1159–1162.

Chaumeil, P. A., A. J. Mussig, P. Hugenholtz, and D. H. Parks. 2022. "GTDB-Tk v2: Memory Friendly Classification With the Genome Taxonomy Database." *Bioinformatics* 38, no. 23: 5315–5316.

Cypionka, H. 2000. "Oxygen Respiration by Desulfovibrio Species." *Annual Reviews in Microbiology* 54, no. 1: 827–848.

Dahl, C., C. Friedrich, and A. Kletzin. 2008. "Sulfur Oxidation in Prokaryotes." In *Encyclopedia of Life Sciences (ELS)*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470015902.a0021155.

Dixon, P. 2003. "VEGAN, a Package of R Functions for Community Ecology." *Journal of Vegetation Science* 14, no. 6: 927–930.

Eikelboom, D. H., and B. Geurkink. 2002. "Filamentous Micro-Organisms Observed in Industrial Activated Sludge Plants." *Water Science and Technology* 46, no. 1–2: 535–542.

Fincker, M. 2020. *Genomic and Physiological Features of Slow-Growing Chloroflexi*. Stanford University.

Finke, N., V. Vandieken, and B. B. Jørgensen. 2007. "Acetate, Lactate, Propionate, and Isobutyrate as Electron Donors for Iron and Sulfate Reduction in Arctic Marine Sediments, Svalbard." *FEMS Microbiology Ecology* 59, no. 1: 10–22.

Flood, B. E., P. Fliss, D. S. Jones, et al. 2016. "Single-Cell (Meta-) Genomics of a Dimorphic Candidatus Thiomargarita Nelsonii Reveals Genomic Plasticity." *Frontiers in Microbiology* 7: 603.

Fonseca, A., C. Espinoza, L. P. Nielsen, I. P. Marshall, and V. A. Gallardo. 2022. "Bacterial Community of Sediments Under the Eastern Boundary Current System Shows High Microdiversity and a Latitudinal Spatial Pattern." *Frontiers in Microbiology* 13: 1–13.

Fonseca, A., T. Ishoey, C. Espinoza, et al. 2017. "Genomic Features of "Candidatus Venteria Ishoeyi", a New Sulfur-Oxidizing Macrobacterium From the Humboldt Sulfuretum Off Chile." *PLoS One* 12, no. 12: e0188371.

Fukui, M., A. Teske, B. Aßmus, G. Muyzer, and F. Widdel. 1999. "Physiology, Phylogenetic Relationships, and Ecology of Filamentous Sulfate-Reducing Bacteria (Genus Desulfonema)." *Archives of Microbiology* 172: 193–203.

Gallardo, V. A. 1963. "Notas sobre la densidad de la fauna bentónica en el sublitoral del norte de Chile." *Gayana* 10: 3–15.

Gallardo, V. A., and C. Espinoza. 2007. "New Communities of Large Filamentous Sulfur Bacteria in the Eastern South Pacific." *International Microbiology* 10, no. 2: 97 ©.

Gallardo, V. A., C. Espinoza, A. Fonseca, and S. Musleh. 2013. "Las grandes bacterias del Sulfureto de Humboldt." *Gayana (Concepción)* 77, no. 2: 136–170 (b).

Gallardo, V. A., A. Fonseca, C. Espinoza, N. Ruiz-Tagle, and S. Musleh. 2016. "Bacteria of the Humboldt Sulfuretum Comply With Unifying Macroecological Principles." *Marine Biodiversity* 46: 399–406.

Gallardo, V. A., A. Fonseca, S. S. Musleh, and C. Espinoza. 2013. "Extrapolations of Standing-Stocks of Big Bacteria in Humboldt Eastern Boundary Current Ecosystem (HEBCE)." *Oceanography* 1, no. 110: 2 (a).

Galushko, A., and J. Kuever. 2015. "Desulfobacula." In *Bergey's Manual of Systematics of Archaea and Bacteria*, 1–4. Wiley.

Gardner, J. G., F. J. Grundy, T. M. Henkin, and J. C. Escalante-Semerena. 2006. "Control of Acetyl-Coenzyme A Synthetase (AcsA) Activity by Acetylation/Deacetylation Without NAD+ Involvement in *Bacillus Subtilis*." *Journal of Bacteriology* 188, no. 15: 5460–5468.

Garrett, S. C. 2021. "Pruning and Tending Immune Memories: Spacer Dynamics in the CRISPR Array." *Frontiers in Microbiology* 12: 664299.

Grünke, S., A. Lichtschlag, D. de Beer, et al. 2012. "Mats of Psychrophilic Thiotrophic Bacteria Associated With Cold Seeps of the Barents Sea." *Biogeosciences* 9, no. 8: 2947–2960.

Hedlund, B. P., M. Chuvochina, P. Hugenholtz, et al. 2022. "SeqCode: a nomenclatural code for prokaryotes described from sequence data." *Nature Microbiology* 7, no. 10: 1702–1708.

Ionescu, D., and M. Bizic. 2019. *Giant Bacteria. eLS*, 1. John Wiley & Sons, Ltd.

Iranzo, J., A. E. Lobkovsky, Y. I. Wolf, and E. V. Koonin. 2014. "Virus-Host Arms Race at the Joint Origin of Multicellularity and Programmed Cell Death." *Cell Cycle* 13, no. 19: 3083–3088.

Ishoey, T., T. Woyke, R. Stepanauskas, M. Novotny, and R. S. Lasken. 2008. "Genomic Sequencing of Single Microbial Cells From Environmental Samples." *Current Opinion in Microbiology* 11, no. 3: 198–204.

Jones, P., D. Binns, H. Y. Chang, et al. 2014. "InterProScan 5: Genome-Scale Protein Function Classification." *Bioinformatics* 30, no. 9: 1236–1240.

Jørgensen, B. B., A. J. Findlay, and A. Pellerin. 2019. "The Biogeochemical Sulfur Cycle of Marine Sediments." *Frontiers in Microbiology* 10: 849.

Justice, S. S., D. A. Hunstad, L. Cegelski, and S. J. Hultgren. 2008. "Morphological Plasticity as a Bacterial Survival Strategy." *Nature Reviews Microbiology* 6, no. 2: 162–168.

Kanehisa, M., Y. Sato, and K. Morishima. 2016. "BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences." *Journal of Molecular Biology* 428, no. 4: 726–731.

Kang, D. D., F. Li, E. Kirton, et al. 2019. "MetaBAT 2: An Adaptive Binning Algorithm for Robust and Efficient Genome Reconstruction From Metagenome Assemblies." *PeerJ* 7: e7359.

Kjeldsen, K. U., L. Schreiber, C. A. Thorup, et al. 2019. "On the Evolution and Physiology of Cable Bacteria." *National Academy of Sciences of the United States of America* 116, no. 38: 19116–19125.

Kuever, J., F. A. Rainey, and F. Widdel. 2015. "Desulfonema." In *Bergey's Manual of Systematics of Archaea and Bacteria*, 1–7. Wiley.

Liu, R., X. Wei, W. Song, et al. 2022. "Novel Chloroflexi Genomes From the Deepest Ocean Reveal Metabolic Strategies for the Adaptation to Deep-Sea Habitats." *Microbiome* 10, no. 1: 1–7.

Madigan, M. T., S. R. Petersen, and T. D. Brock. 1974. "Nutritional Studies on Chloroflexus, a Filamentous Photosynthetic, Gliding Bacterium." *Archives of Microbiology* 100: 97–103.

Maier, S., and V. A. Gallardo. 1984. "*Thioploca Araucae* Sp. Nov. and *Thioploca Chileae* Sp. Nov." *International Journal of Systematic and Evolutionary Microbiology* 34, no. 4: 414–418.

McIlroy, S. J., S. M. Karst, M. Nierychlo, et al. 2016. "Genomic and In Situ Investigations of the Novel Uncultured Chloroflexi Associated With 0092 Morphotype Filamentous Bulking in Activated Sludge." *ISME Journal* 10, no. 9: 2223–2234.

Middelboe, M., and R. N. Glud. 2006. "Viral Activity Along a Trophic Gradient in Continental Margin Sediments Off Central Chile." *Marine Biology Research* 2, no. 1: 41–51.

Minh, B. Q., H. A. Schmidt, O. Chernomor, et al. 2020. "IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era." *Molecular Biology and Evolution* 37, no. 5: 1530–1534.

Müller, A. L., K. U. Kjeldsen, T. Rattei, M. Pester, and A. Loy. 2015. "Phylogenetic and Environmental Diversity of DsrAB-Type Dissimilatory (Bi) Sulfite Reductases." *ISME Journal* 9, no. 5: 1152–1165.

Mußmann, M., F. Z. Hu, M. Richter, et al. 2007. "Insights Into the Genome of Large Sulfur Bacteria Revealed by Analysis of Single Filaments." *PLoS Biology* 5, no. 9: e230.

Parks, D. H., M. Imelfort, C. T. Skennerton, P. Hugenholtz, and G. W. Tyson. 2015. "CheckM: Assessing the Quality of Microbial Genomes Recovered From Isolates, Single Cells, and Metagenomes." *Genome Research* 25, no. 7: 1043–1055.

Parks, D. H., C. Rinke, M. Chuvochina, et al. 2017. "Recovery of Nearly 8,000 Metagenome-Assembled Genomes Substantially Expands the Tree of Life." *Nature Microbiology* 2, no. 11: 1533–1542.

Pourcel, C., M. Touchon, N. Villeriot, et al. 2020. "CRISPRCasdb a Successor of CRISPRdb Containing CRISPR Arrays and Cas Genes From Complete Genome Sequences, and Tools to Download and Query Lists of Repeats and Spacers." *Nucleic Acids Research* 48, no. D1: D535–D544.

Pritchard, L., P. Cock, and Ö. Esen. 2019. "pyani v0. 2.8: average nucleotide identity (ANI) and related measures for whole genome comparisons."

Ragsdale, S. W., and E. Pierce. 2008. "Acetogenesis and the Wood–Ljungdahl Pathway of CO2 Fixation." *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* 1784, no. 12: 1873–1898.

Richter, M., and R. Rosselló-Móra. 2009. "Shifting the Genomic Gold Standard for the Prokaryotic Species Definition." *National Academy of Sciences of the United States of America* 106, no. 45: 19126–19131.

Rodriguez-R, L. M., S. Gunturu, W. T. Harvey, et al. 2018. "The Microbial Genomes Atlas (MiGA) Webserver: Taxonomic and Gene Diversity Analysis of Archaea and Bacteria at the Whole Genome Level." *Nucleic Acids Research* 46, no. W1: W282–W288.

Rubin-Blum, M., N. Dubilier, and M. Kleiner. 2019. "Genetic Evidence for Two Carbon Fixation Pathways (The Calvin-Benson-Bassham Cycle

and the Reverse Tricarboxylic Acid Cycle) in Symbiotic and Free-Living Bacteria." *Msphere* 4, no. 1: e00394-18.

Salman, V., R. Amann, A. C. Girnth, et al. 2011. "A Single-Cell Sequencing Approach to the Classification of Large, Vacuolated Sulfur Bacteria." *Systematic and Applied Microbiology* 34, no. 4: 243–259.

Schauder, R., A. Preuss, M. Jetten, and G. Fuchs. 1989. "Oxidative and Reductive Acetyl-CoA/Carbon Monoxide Dehydrogenase Pathway in *Desulfobacterium Autotrophicum*. 2. Demonstration of the Enzymes of the Pathway and Comparison of CO Dehydrogenase." *Archives of Microbiology* 151, no. 1: 84–89.

Schnaars, V., L. Wöhlbrand, S. Scheve, C. Hinrichs, R. Reinhardt, and R. Rabus. 2021. "Proteogenomic Insights Into the Physiology of Marine, Sulfate-Reducing, Filamentous Desulfonema Limicola and *Desulfonema Magnum*." *Microbial Physiology* 31, no. 1: 36–56.

Schopf, J. W. 2006. "Fossil Evidence of Archaean Life." *Philosophical Transactions of the Royal Society, B: Biological Sciences* 361, no. 1470: 869–885.

Schulz, H. N., and B. B. Jørgensen. 2001. "Big Bacteria." *Annual Reviews in Microbiology* 55, no. 1: 105–137.

Seemann, T. 2014. "Prokka: Rapid Prokaryotic Genome Annotation." *Bioinformatics* 30, no. 14: 2068–2069.

Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov. 2015. "BUSCO: Assessing Genome Assembly and Annotation Completeness With Single-Copy Orthologs." *Bioinformatics* 31, no. 19: 3210–3212.

Song, W., H. X. Sun, C. Zhang, et al. 2019. "Prophage Hunter: An Integrative Hunting Tool for Active Prophages." *Nucleic Acids Research* 47, no. W1: W74–W80.

Sørensen, J., D. Christensen, and B. B. Jørgensen. 1981. "Volatile Fatty Acids and Hydrogen as Substrates for Sulfate-Reducing Bacteria in Anaerobic Marine Sediment." *Applied and Environmental Microbiology* 42, no. 1: 5–11.

Sørensen, J., B. B. Jørgensen, and N. P. Revsbech. 1979. "A Comparison of Oxygen, Nitrate, and Sulfate Respiration in Coastal Marine Sediments." *Microbial Ecology* 5: 105–115.

Speirs, L. B., D. T. Rice, S. Petrovski, and R. J. Seviour. 2019. "The Phylogeny, Biodiversity, and Ecology of the Chloroflexi in Activated Sludge." *Frontiers in Microbiology* 10: 2015.

Sweerts, J.-P. R. A., D. De Beer, L. P. Nielsen, et al. 1990. "Denitrification by Sulphur Oxidizing Beggiatoa Spp. Mats on Freshwater Sediments." *Nature* 344, no. 6268: 762–763.

Tesson, F., A. Hervé, E. Mordret, et al. 2022. "Systematic and Quantitative View of the Antiviral Arsenal of Prokaryotes." *Nature Communications* 13, no. 1: 2561.

Thiel, V., J. M. Wood, M. T. Olsen, et al. 2016. "The Dark Side of the Mushroom Spring Microbial Mat: Life in the Shadow of Chlorophototrophs. I. Microbial Diversity Based on 16S rRNA Gene Amplicons and Metagenomic Sequencing." *Frontiers in Microbiology* 7: 919.

Ward, L. M., J. Hemp, P. M. Shih, S. E. McGlynn, and W. W. Fischer. 2018. "Evolution of Phototrophy in the Chloroflexi Phylum Driven by Horizontal Gene Transfer." *Frontiers in Microbiology* 9: 260.

Whitman, W. B., M. Chuvochina, B. P. Hedlund, et al. 2022. "Development of the SeqCode: A Proposed Nomenclatural Code for Uncultivated Prokaryotes With DNA Sequences as Type." *Systematic and Applied Microbiology* 45, no. 5: 126305.

Widdel, F. 1988. "Microbiology and Ecology of Sulfate-and Sulfur-Reducing Bacteria." *Biology of Anaerobic Microorganisms*: 469–585.

Widdel, F., G. W. Kohring, and F. Mayer. 1983. "Studies on Dissimilatory Sulfate-Reducing Bacteria That Decompose Fatty Acids: III. Characterization of the Filamentous Gliding *Desulfonema Limicola* Gen. Nov. Sp. Nov., and *Desulfonema Magnum* Sp. Nov." *Archives of Microbiology* 134: 286–294.

Winkel, M., V. Salman-Carvalho, T. Woyke, et al. 2016. "Single-Cell Sequencing of Thiomargarita Reveals Genomic Flexibility for Adaptation to Dynamic Redox Conditions." *Frontiers in Microbiology* 7: 964.

Yarza, P., P. Yilmaz, E. Pruesse, et al. 2014. "Uniting the Classification of Cultured and Uncultured Bacteria and Archaea Using 16S rRNA Gene Sequences." *Nature Reviews Microbiology* 12, no. 9: 635–645.

Zimmermann, J., C. Kaleta, and S. Waschina. 2021. "Gapseq: Informed Prediction of Bacterial Metabolic Pathways and Reconstruction of Accurate Metabolic Models." *Genome Biology* 22, no. 1: 1–35.

Zopfi, J., M. E. Böttcher, and B. B. Jørgensen. 2008. "Biogeochemistry of Sulfur and Iron in Thioploca-Colonized Surface Sediments in the Upwelling Area Off Central Chile." *Geochimica et Cosmochimica Acta* 72, no. 3: 827–843.

Zopfi, J., T. Kjaer, L. P. Nielsen, and B. B. Jørgensen. 2001. "Ecology of Thioploca Spp.: Nitrate and Sulfur Storage in Relation to Chemical Microgradients and Influence of Thioploca Spp. on the Sedimentary Nitrogen Cycle." *Applied and Environmental Microbiology* 67, no. 12: 5530–5537. https://doi.org/10.1128/AEM.67.12.5530-5537.2001.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.