ISME

www.nature.com/ismecomms

ARTICLE OPEN

Discovery of a novel bacterial class with the capacity to drive sulfur cycling and microbiome structure in a paleo-ocean analog

Adrien Vigneron [1,2,3,4]✉, Warwick F. Vincent [1,2,3,4] and Connie Lovejoy [1,3,4,5]

Uncultivated microbial taxa represent a large fraction of global microbial diversity and likely drive numerous biogeochemical transformations in natural ecosystems. Geographically isolated, polar ecosystems are complex microbial biomes and refuges of underexplored taxonomic and functional biodiversity. Combining amplicon sequencing with genome-centric metagenomic analysis of samples from one of the world's northernmost lakes (Lake A, Ellesmere Island, Canadian High Arctic), we identified a novel bacterial taxon that dominates in the bottom layer of anoxic, sulfidic, relict sea water that was isolated from the Arctic Ocean some 3000 years ago. Based on phylogenomic comparative analyses, we propose that these bacteria represent a new Class within the poorly described Electryoneota/AABM5-125-24 candidate phylum. This novel class, for which we propose the name Tariuqbacteria, may be either a relict of ancient ocean conditions or endemic to this High Arctic system, provisionally providing a rare example of high-taxonomy level endemism. Consistent with the geochemistry of the bottom water, the genetic composition of the *Candidatus* Tariuqbacter genome revealed a strictly anaerobic lifestyle with the potential for sulfate and sulfur reduction, a versatile carbon metabolism and the capability to eliminate competing bacteria through methylarsenite production, suggesting an allelochemical influence on microbiome structure by this planktonic microbe.

ISME Communications; https://doi.org/10.1038/s43705-023-00287-9

INTRODUCTION

Advances in DNA sequencing and computational approaches have accelerated the reconstruction of metagenome assembled genomes (MAGs) of uncultured microbial populations in multiple ecosystems [1]. This approach has revealed many novel lineages and expanded our vision of the tree of life and the ecological roles of uncultured microorganisms [2–5]. For example, the guild of potential sulfate and sulfite reducers has significantly expanded with genome-centric metagenomics [6]. As a result, the taxonomic diversity of potential sulfate-reducers now extends far away from taxa of culturable sulfate reducers to include numerous poorly characterized lineages such as *Schekmanbacteria*, *Rokubacteria*, *Zixibacteria*, and *Abyssubacteria* [6, 7]. Although these genomic results require physiological confirmation, they illustrate the power of genome-centric metagenomics to associate uncultured lineages with potential biogeochemical reactions and the need for revision of biogeochemical process models based on microbial community composition. Despite these significant advances, however, diversity surveys have suggested that many taxa remain underexplored and uncharacterized, including sulfate/sulfite reducers [8, 9].

With the contribution of metagenome assembled genomes (MAGs) the total numbers of partial and complete genomes in databases are growing exponentially with up to 317,000 and 907,000 curated genomes available in the Genome Taxonomy Database (GTDB; release 207) [10] and proGenomes v.3 database [11] respectively. Furthermore, around 1.5 million genomes are now deposited in NCBI (January 2023), making it possible to carry out extensive genomic comparisons, deep phylogenomic analysis, mapping of global distributions and exploration of community assembly [1, 12].

Extreme environments, such as hydrothermal vents [13–15], hypersaline basins [16], deep underground aquifers [3] and polar lakes [17] are major sources of previously unknown microbial biodiversity and are useful systems to investigate limits of life, microbial biogeography and the distribution of lineages in nature [18]. High Arctic Lake A (latitude 83 °N) is one of the northernmost lakes of the world and has many extreme features. Located at the far northern coast of Canada, Lake A is a perennially ice-covered, highly stratified, meromictic lake, with a bottom saline layer derived from Arctic Ocean seawater that was trapped by isostatic uplift around 3000 years ago [19]. Water column mixing is inhibited by the perennial ice cover over the lake, and by a surface 14-m thick layer of low-salinity water derived from snow and ice melt. This stratification ensures the isolation of the relict oceanic water that extends from 24 m to the bottom of the lake (128 m) and that is fully anoxic and sulfidic [19]. Microbial community composition along the steep geochemical gradients of Lake A has

[1]Département de Biologie, Université Laval, Québec, QC, Canada. [2]Centre d'études nordiques (CEN), Université Laval, Québec, QC, Canada. [3]Institut de Biologie Intégrative et des Systèmes, Université Laval, Québec, QC, Canada. [4]Takuvik Joint International Laboratory, Université Laval, Québec, QC, Canada. [5]Québec Océan, Université Laval, Québec, QC, Canada. ✉email: avigner0@gmail.com

Received: 7 March 2023 Revised: 21 July 2023 Accepted: 26 July 2023
Published online: 18 August 2023

been previously reported, and provided insights into aquatic biogeochemical cycles across a broad range of light, salinity, oxygen and sulfur conditions [7, 20, 21]. These studies revealed a vertical succession of freshwater, marine and deep-sea anoxic microbiomes, centered on sulfur transformations and biomass recycling down the water column [7, 20, 22].

In our previous work on Lake A, we detected an intriguing microbial community in the bottom waters of the lake dominated by a large fraction of unclassified microbial lineages. The unclassified lineages from 65 m largely predominated over other microbial taxa at that depth, including *Dehalococcoidia*, *Omnitrophicaeota*, *Atribacteria* and *Desulfobacterota*, which are typically recovered from deep sea and marine sediments [7]. In the present study, we combined a new analysis of 16 S rRNA (cDNA) and rRNA gene (DNA) sequence data, with genome-centric metagenomics with a specific focus on these unknown lineages. This allowed us to recover and characterize a novel bacterial Class-level lineage. The lineage was unique to the system and appears to be a major contributor to sulfur cycling in these relict Arctic Ocean waters.

METHODS

Site description and sample collection

Lake A is located on the northern coast of Ellesmere Island, Nunavut, in the Canadian High Arctic (latitude 83° 00′ N, longitude 75° 27′ W). The lake has an oxygenated freshwater layer (14.1 mg L^{-1} of oxygen; salinity <0.7 PSU) that extends from under the ice down to 12 m depth, a chemo- and halocline located below the freshwater down to 24 m, and an anoxic marine-derived saline deeper layer with oxygen below the detection limit of the oxygen profiler and salinity >30 PSU. Sulfate rich waters (>18 mM) are found below 24 m and sulfide concentration increase with depth (~0.5 mM at 65 m). Water samples were collected in summer 2017 from three separate 24 cm-diameter holes drilled through the ice in the center of the lake as previously detailed [7]. One liter of water was collected at eight different depths (2 m, 6 m, 14 m, 22 m, 34 m, 40 m, 55 m and 65 m) from the three holes. These depths span the oxygen, salinity and temperature gradients of the water column [7]. The triplicate water samples were filtered through separate 0.22 µm pore size Sterivex filters™ (Merck Millipore Ltd., Oakville, Canada) that were then stored below −50 °C until nucleic acid extraction.

Nucleic acid extraction

Nucleic acids (DNA and RNA) were co-extracted from two of the replicates per depth using Qiagen Allprep DNA/RNA Mini Kit with modifications (Qiagen, Hilden, Germany) [23]. The DNA extracts were stored at −20 °C until library preparation. For RNA extracts, two additional DNase steps (DNase I, Ambion, Foster City, CA, USA) were carried out to remove any trace of carried over DNA. The absence of DNA contamination was confirmed by amplification of 16 S rRNA genes with bacterial primers using the RNA extracts (undiluted and diluted ten times) as template, with no product detected after 35 PCR cycles. The RNA was then immediately converted to cDNA using a High-Capacity cDNA Reverse Transcription kit (Applied Biosystems, Foster City, CA, USA) and stored as cDNA at −20 °C until library preparation.

PCR amplification, sequencing and analysis

Microbial community composition of the samples was determined by high throughput sequencing of bacterial 16 S rRNA (cDNA) and 16 S rRNA genes (DNA) using primers targeting the bacterial V4-V5 region (S-D-Bact-0516-a-S-18/S-D-Bact-0907-a-A-20; 460 bp product). Amplifications were carried out in duplicate using Brilliant III master Mix (Agilent Technologies, Santa Clara, CA, USA), 500 nM of each primer and 1 ng of DNA/cDNA template on 25 µl reaction volume. PCR conditions were as follows: initial denaturation at 95 °C for 2 min, then 30 cycles of denaturation at 95 °C for 15 s, annealing at 58 °C for 30 s and extension at 72 °C for 30 s; and a final extension at 72 °C for 7 min. Duplicate PCR products were pooled and purified on agarose gels using the QIAquick Gel Extraction Kit (QIAGEN), then Illumina MiSeq adaptors and sample barcodes were added during a second 10 cycles-PCR step as detailed in [24]. Samples were pooled equimolarly, then sequenced along with negative controls for nucleic acids extraction, transcription and PCRs steps, using an Illumina MiSeq v3 kit at the IBIS/Laval University, Plate-forme d'Analyses Génomiques (Québec, QC). Reads were assembled into single paired-end sequences using FLASH v2.2.0, curated and clustered into OTUs (97% sequence similarity) using VSEARCH v2.3.4. as detailed in a GitHub repository (https://github.com/CruaudPe/MiSeq_Multigenique). Taxonomic affiliations of the reads were determined with mothur [25] using both the mothur version of the Bayesian classifier and BLAST against Silva database release 132 and 138 [26]. For the BLAST algorithm, the 5 best hits of BLASTn for each sequence were collected and taxonomic affiliation was defined by the lowest classification shared by the 5 hits. In addition, the sequences of unclassified bacterial OTUs were also compared against NCBI-nr (access date: Nov. 2022), RDP (Release 11.5, taxonomy 18), IMG-16S (access date: Nov. 2022), EzBioCloud (Release 2021-07) and Greengenes v.13.5 databases.

Metagenomic library preparation, sequencing and analysis

One metagenomic library per sampled depth was prepared using Illumina Nextera XT kit and sequenced in two Illumina MiSeq (2*300 bp) runs and one Illumina NexSeq run (2*150 bp) at the Institut de Biologie Integrative et des Systèmes (IBIS) sequencing platform (Université Laval, Canada) and at the CGEB – Integrated Microbiome Resource (Dalhousie University, Canada), respectively. Datasets were quality filtered using Trimmomatic v.0.39 [27] keeping both R1 and R2 reads when reads overlapped. Reads of ribosomal small subunit (rRNA) were extracted from metagenomic reads using Infernal v.1.1.4 [28], and taxonomic affiliation of the extracted 16 S rRNA reads longer than 100 bp were determined using BLAST against the Silva database release 138 as reference [26]. Full length 16 S rRNA genes were reconstructed from extracted 16 S rRNA reads using SPADES v.3.15.4 [29] (careful mode) and metagenomic reads were mapped against the complete 16 S rRNA genes using BBMAP v.38.86 (perfectmode option) [30] to determine depth distribution of 16 S rRNA genes. For metagenome assembled genome reconstruction, all quality filtered sequences were pooled and co-assembled using MEGAHIT v.1.2.9 [31]. Read coverage of the contigs was carried out using bwa-mem (http://bio-bwa.sourceforge.net), followed by contig binning using MetaBAT-2 [32] with contigs longer than 2000 bp. The completeness and contamination level of the MAGs were evaluated using CheckM v.1.1.5 [33]. Relative abundance of the MAGs was estimated with the average coverage of the contigs determined using bwa-mem and averaged using the jgi_summarize_bam_contig_depths script. Taxonomic affiliation of the MAGs was carried out using GTDB-tk v.2.2.6 [34]. Genomes taxonomically close to our bin of interest (bin 685) on the phylogenomic tree were downloaded and investigated for the presence of the dsr operon as previously described [6]. Amino acid identity percentage between bin 685 and these genomes was determined using AAI calculator [35]. The genetic composition of genomic bins was explored using KofamScan with the KEGG database v103.0 [36] and METABOLIC v.4 [37]. The results were manually checked for the presence of specific pathways.

For phylogenetic analysis of dsrA gene, amino acid sequences of dsrA were recovered from the 65 m metagenome and only sequences longer than 260 AA were analysed. Recovered amino acid sequences were compared against public databases (NCBI and IMG/MR) using BLASTp. Sequences with the best BLAST hits were downloaded and aligned with the metagenomic sequences using MAFFT v.7.471 [38]. Maximum likelihood tree of 180 sequences with 491 amino acid sites was constructed using IQTree v.2 with 1000 bootstraps and the LG + I + G4 model [39] and visualized using iTOL v.6 [40]. A similar methodology was used for the phylogenetic analysis of the *arsM* gene of *Candidatus* Tariuqbacter. The sequence was compared against NCBI amino-acid sequence database using BLASTp and aligned with the best hits of BLAST using MAFFT. A maximum likelihood tree of 80 sequences with 302 amino acid sites was constructed using IQTree v.2 with 1000 bootstraps and the JTT + I + G4 model, and was visualized using iTOL v.6.

RESULTS AND DISCUSSION

A single unclassified microorganism predominates in the bottom seawater

The microbial survey using 16 S rRNA gene and rRNA amplicon sequencing of the Lake A water column revealed the predominance of a single OTU within bottom waters of the lake (16.5% of the 16 S rRNA genes at 65 m, Fig. 1). This single OTU, represented up to 36% of the 16 S rRNA sequences (from cDNA) amplified from the same samples, suggesting that these microorganisms were metabolically active (Fig. 1). The 16 S ribosomal sequence of the OTU was unclassified at the phylum level (Unclassified Bacteria)

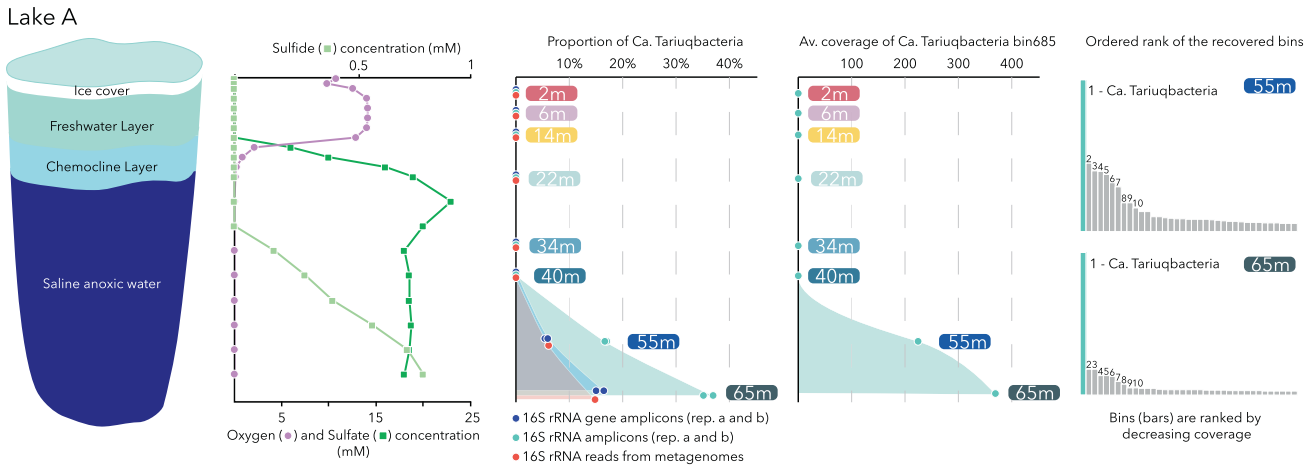

Fig. 1 Depth distribution of *Candidatus* Tariuqbacter arcticus in the Lake A water column, based on 16 S rRNA and rRNA gene sequencing and metagenomic sequencing. Numbers in the histogram of ranked coverage refers to the following lineages: For the 55 m sample: 1 Ca. Tariuqbacteria (222 rpb), 2 Desulfaltia (99 rpb), 3 Atribacterota (88 rpb), 4 Dehalococcoidales (88 rpb), 5 Marinisomatales (83 rpb),6 Dehalococcoidales (70 rpb), 7 Marinisomatales (65 rpb),8 Aminicenantaceae (41 rpb),9 Cyanobiaceae (41 rpb),10 Dehalococcoidales (34 rpb). For the 65 m sample: 1 Ca. Tariuqbacteria (377 rpb), 2 Atribacteria (68 rpb), 3 Desulfaltia (67 rpb), 4 Dehalococcoidales (51 rpb), 5 Dehalococcoidales (51 rpb),6 Marinisomatales (48 rpb), 7 Aminicenantaceae (36 rpb),8 Ca. JABMQX01 (28 rpb),9 Desulfatibia (19 rpb),10 Bipolaricaulia (18 rpb).

using both Silva (v132 and v138), RDP (Release 11.5, taxonomy 18), EzBioCloud (Release 2021-07) and GreenGenes v.13.5 databases with both RDP classifier and BLAST classification algorithms. After metagenomic sequencing of the samples, a full length 16 S rRNA gene with 100% similarity with the unclassified 16 S rRNA OTU was assembled. Mapping of the metagenomic reads sequenced from 2 m, 6 m, 14 m, 22 m, 34 m, 40 m, 55 m and 65 m samples against the novel 16 S rRNA gene sequence confirmed its absence (not-detected) in upper waters and the predominance of this lineage in the anoxic and marine bottom waters of Lake A, where it represented 15% of the 16 S rRNA metagenomic reads (Fig. 1). Comparison of the full 16 S rRNA gene sequence assembled from the metagenome dataset and the 16 S V4-V5 region amplified by PCR against the NCBI-nr database indicated that only one sequence (KM018918), amplified from the brine-seawater interface of Erba Deep brine pool samples (Red Sea), shared a percentage of identity higher than 90% over the full 16 S rRNA gene sequence (91.7%) and V4-V5 PCR-amplified region (93.3%). In addition, we then searched publicly available metagenomes through the Integrated Microbial Genomes and Microbiomes (IMG/M) portal. Only a few sequences, all recovered from the sulfidic waters of meromictic Lake Kivu (Congo, Rwanda), were found to share a percentage of identity higher than 90% (max: 91.8%) with the novel 16 S rRNA gene sequence.

Co-assembly and binning of the contigs from Lake A enabled the reconstruction of 250 draft metagenome-assembled genomes (MAGs), as reported previously [7, 20, 22]. The novel MAG was estimated to be 79.91% complete (contamination: 2.01%), with high coverage (up to 377 reads per base) for the 55 m and 65 m metagenomes (Fig. 1), and a 16 S rRNA gene sequence matching the unclassified rRNA gene OTU. Taxonomic affiliation of the MAG using the genome classifier tool in the proGenomes database (900,000 genomes) resulted in no match, whereas phylogenetic analyses of the MAG using GTDB-tk pipeline [34] revealed a deep branching within the Fibrobacterota, Chlorobiota and Bacteroidota (FCB) group, basal to Zixibacteria and Calditrichota phyla and within the "AABM5-125-24" phylum (Fig. 2, Supplementary Fig. 1). A few members of the AABM5-125-24 phylum, also named *Candidatus* Electryoneota [41] have been identified as rare members of Ace Lake, Antarctica [17], which is a meromictic marine-derived lake like Lake A. However, analysis of the average amino acid identity (AAI, Fig. 3) indicated a low similarity with most of Electryoneota/AABM5-125-24 genomes (<46%), including the Ace Lake representatives and a maximum of 57.5% of amino acid identity percentage with a genome recovered from Black Sea brackish waters (Fig. 2). Together these results support the classification of this MAG as a new bacterial class, order, family, genus and species within the Electryoneota/AABM5-125-24 phylum. We propose that this candidate taxon be named *Candidatus* Tariuqbacter arcticus, family Tariuqbacteraceae, order Tariuqbacterales and class Tariuqbacteria with "Tariuq" meaning saltwater in Inuktitut, the language of the Inuit, including in the community of Aujuittuq (Grise Fiord, Nunavut, Canada), the nearest human settlement to Lake A.

The low sequence similarity with reference genomes from all investigated databases suggests that Tariuqbacteria represented by *Tariuqbacter arcticus* may potentially be endemic to High Arctic Lake A and similar relict Arctic Ocean seawater habitats (e.g., lakes B, C1 and C2 in the same region [42]) that remain to be explored. The inferred endemicity of a lineage is based on the molecular markers that were used, and strongly depends on the reference databases and their taxonomic resolution [43]. Reported microbial endemism often refers to genotypes and sub-species [44–46], therefore this lineage would represent a rare example of bacterial endemism at a high taxonomic level [43]. Similar results have been previously reported in isolates of geothermal samples from Antarctica, where a single unclassified archaeon predominated the microbial community [18], suggesting that habitats in polar regions are refuges for ancient lineages that have evolved in isolation, resulting in novel biodiversity and potentially endemic populations with unusual metabolic features. However, perennially ice-covered polar lakes are largely under-sampled and little studied compared to non-polar systems, which could explain the absence of taxonomically close species in databases. In this context, the addition of the *Tariuqbacter* genome to public databases may enhance the detection of related Tariuqbacteria in future environmental studies, as for other recently described novel lineages [4].

Tariuqbacter drives the sulfur cycle in relict Arctic Ocean water

The predominance of the highly abundant population of *Tariuqbacter* in Lake A suggests that members of this taxon have outcompeted other microbes and are uniquely adapted to the extreme biogeochemical conditions in the lower water column of

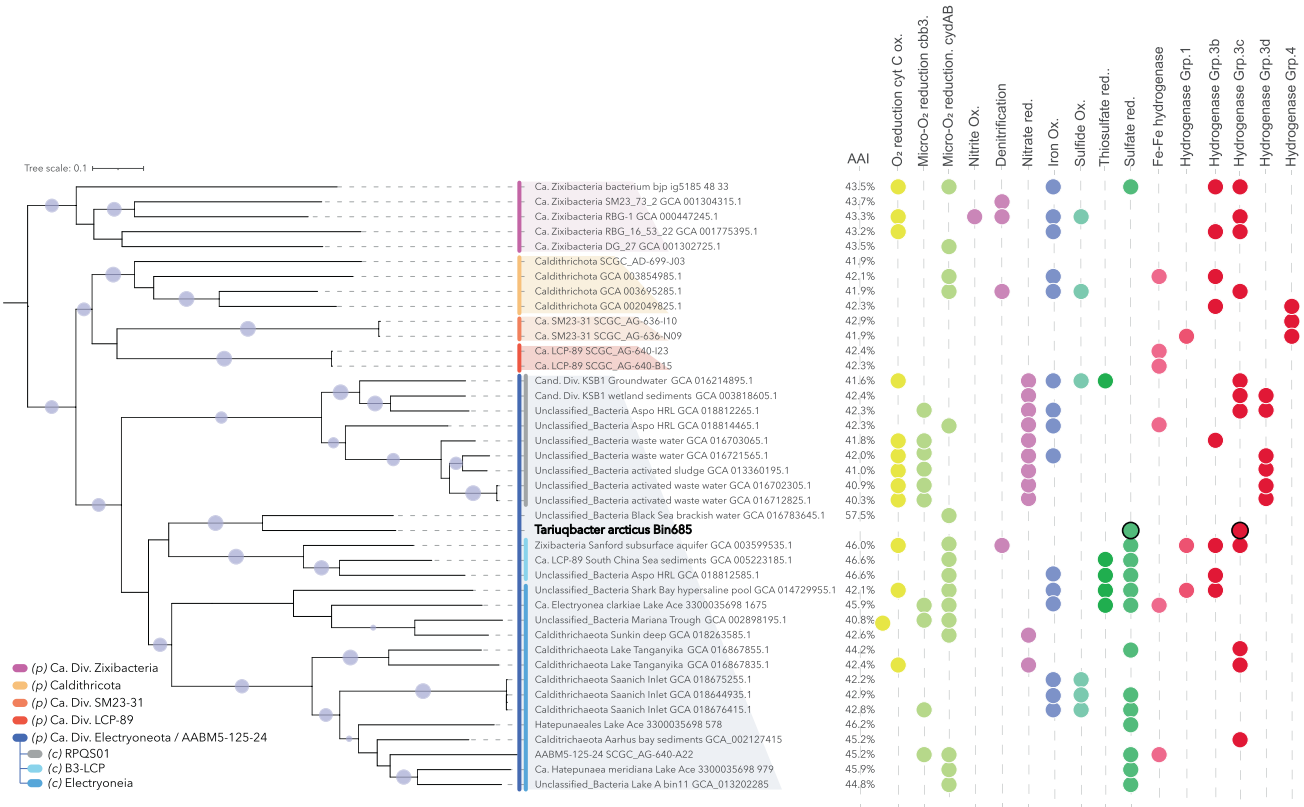

Fig. 2 Phylogenetic tree of the Electryoneota/AABM5-125-24 phylum (blue line), including reference genomes from phylum Caldithrichota (yellow line) and candidate phylum Zixibacteria (pink line), SM23-31 (orange line), LCP-89 (red line). Only bootstraps higher than 80% are represented by purple dots. Dots indicate the detection of marker genes for specific metabolic pathways. Yellow shade: oxygen reduction, purple: nitrogen cycling, blue: iron oxidation, green: sulfur cycling and red: hydrogen metabolism. Average amino-acid identity (AAI) was calculated for each genome against the tariuqbacterial MAG.

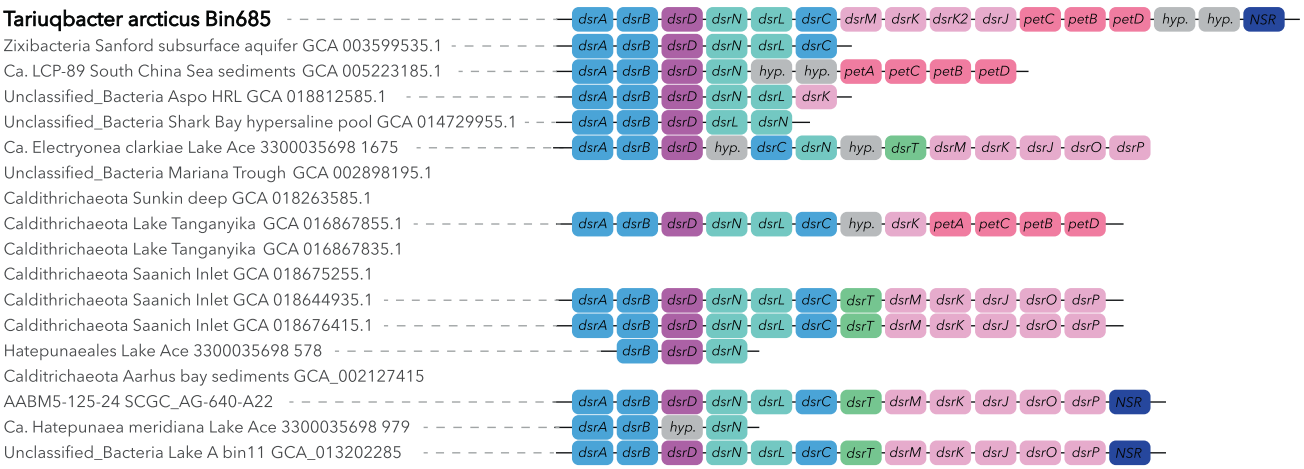

Fig. 3 Structure of the dsr operon identified in Tariuqbacter arcticus and related genomes from the Electryoneota/AABM5-125-24 phyla. Dsr operons were represented only when at least two genes of the dsr operon including dsrA or dsrB were identified in the genomes. Representation of the operon stops where genes identified in the contigs were not related to the dsr operon.

Lake A. To decipher the underlying evolutionary and ecological forces driving this limited distribution, predicted proteins encoded in the tariuqbacterial MAG were compared against the Kegg database (release 103.0) (Supplementary Table 1) and using the METABOLIC v.4 pipeline [37], which uses numerous curated HMM profiles. A complete sulfate reduction pathway was identified, including genes coding sulfate adenyltransferase (*sat*), adenylsulfate reductase (*aprAB*) and the QmoABC membrane complex that provides AprAB electrons, and dissimilatory sulfite reductase (*dsrABC*) (Fig. 3). This capability to reduce sulfate is particularly relevant under the environmental conditions near the bottom of the lake, since sulfate concentrations of around 18 mM and higher have been measured in these fully anoxic waters (Fig. 1) [19]. Analysis of the genes identified in the vicinity of *dsrABC* revealed the presence of a complex dsr operon in *Tariuqbacter* (Fig. 3). Two divergent copies of *dsrK* gene, coding for the trisulfide reductase that act on DsrC and release hydrogen sulfide from the sulfate reduction pathway [47] were detected. Genes of periplasmic DsrO

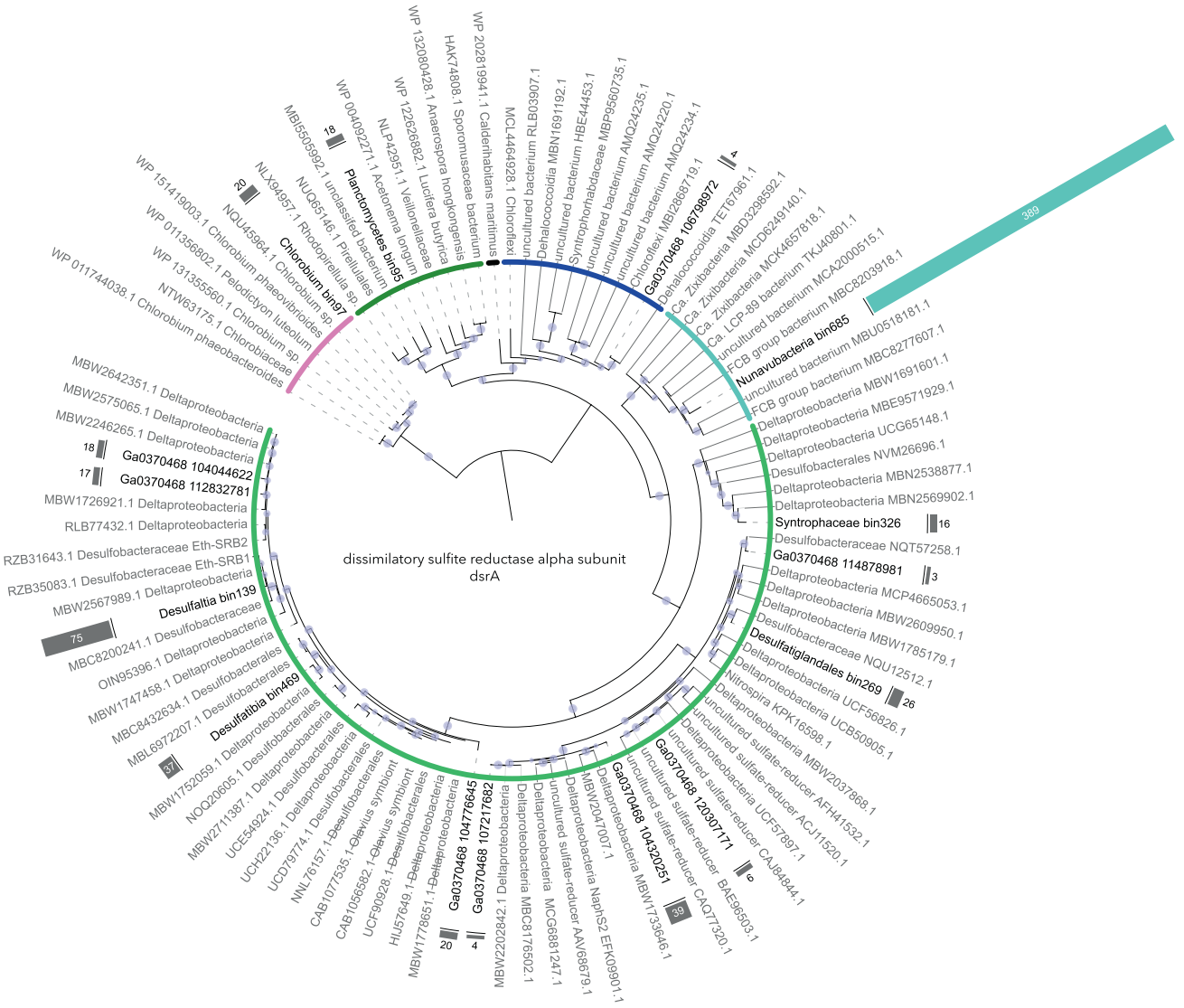

Fig. 4 Taxonomic tree of the alpha subunit of the dissimilatory sulfite reductase (dsrA) gene (sequences longer than 250 amino-acids) recovered from the metagenome of the 65 m sample. Bars in front of the labels indicate the estimated number of copies of the sequence, indicated in or next to the bars.

and membrane protein DsrP involved in electron transfer in the dsrKMOP complex [6] were not identified using Kegg nor METABOLIC dsr operon HMM profiles. Instead, petBCD genes coding for Rieske proteins and the cytochrome b_6f complex were detected and specific domains of these proteins were confirmed by sequence analysis using Interpro [48]. These results suggest that electron transfer to F-type ATPase could be mediated by the cytochrome b_6f complex. Genes coding the heterodisulfide reductase-[NiFe]-hydrogenase (HdrABC-MvhAGD) frequently found in sulfate-reducing bacteria and methanogens were also identified [49]. This complex reduces disulfide with the reduction of ferredoxin in an energy-conserving flavin-based electron bifurcation process [50]. In addition, the dsr operon in *Tariuqbacter* also included a NAD(P)H elemental sulfur oxidoreductase (NSR) suggesting that it could also reduce S_0 (Fig. 3) [51]. A sulfidic smell and the yellow-orange color of the water below 22 m was noted during sampling, indicating the presence of polysulfides and aqueous elemental sulfur in the water, and supporting the S_0-reducing potential of *Tariuqbacter*. Mining of the 28 publicly available genomes affiliated to "Electryoneota/AABM5-125-24" phylum failed to detect the sulfate-reducing potential in the taxonomically closest genome (57.5% AAI). However, the analysis identified 12 MAGs with the dsr operon, including one (bin11; Fig. 2) that was also recovered from Lake A deep waters, supporting the sulfate reduction potential of some Electryoneota lineages [5] (Figs. 2 and 3). However, the structure of the operon and more particularly the genes involved in electron transfer were variable, and included, in two other MAGs, genes coding the cytochrome b_6f instead of dsrJOP. In addition to the potential for sulfate reduction, previous studies on Electryoneota lineages have reported a potential capability for aerobic respiration using high affinity cytochromes (cbb3-type and cydAB) and sulfide oxidation genes, suggesting a facultative anaerobic lifestyle [17]. Analysis of energetic pathways found in publicly available draft genomes of Electryoneota revealed that genes for aerobic respiration at low oxygen levels are frequent in the phylum (75% of the MAGs). However, none of these aerobic pathways were detected in the *Tariuqbacter* genome, suggesting a strictly anaerobic metabolism that differs from Antarctic and other Electryoneota lineages. This result is consistent with the detection of *Tariuqbacter* only in the deeper anoxic water of Lake A, while Electryoneota populations were detected in the oxic-anoxic interface of Ace Lake. The

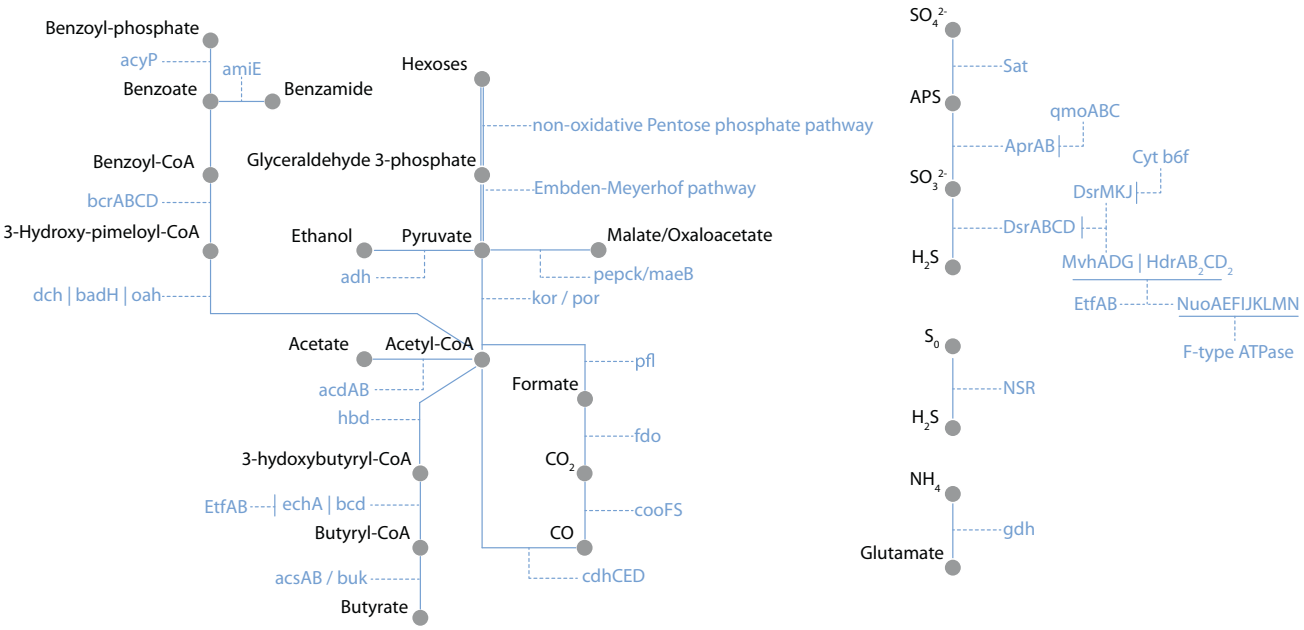

Fig. 5 Metabolic reconstruction of *Tariuqbacter arcticus* based on genes identified using the Kegg database (release 103.0).

Tariuqbacter MAG also differed from other Electryoneota genomes by the apparent absence of nitrate and thiosulfate reduction or iron and sulfide oxidation pathways that were identified in the RPQS01 class and in other Electryoneota genomes (Fig. 2). The unique energetic pathway gene combination of *Tariuqbacter* supports its position as a novel class within the Electryoneota phylum.

Taxonomic analysis of the drsA gene recovered from the metagenomic dataset of the 65 m sample confirmed the taxonomic placement of Tariuqbacteria within the FCB group as well as the strong predominance of this lineage in the deep water of Lake A (Fig. 4). Up to 389 copies of the *Tariuqbacter* dsrA gene were identified at 65 m, representing 69% of all dsrA genes identified at this depth whereas the second most abundant dsrA gene, related to *Desulfaltia*, (MAG number 3 in Fig. 1 – 65 m) had 75 copies (Fig. 4). These results suggest that *Tariuqbacter* plays a major role in the sulfur cycle in the relict Arctic Ocean water of Lake A. Utilisation of intermediates of the sulfur cycle and sulfides has been proposed to shape a large fraction of the Lake A microbiome throughout the water column [7], and therefore *Tariuqbacter arcticus* may be a major structuring species for this extreme ecosystem.

Metabolic versatility of *Tariuqbacter arcticus*

The potential for degradation of hexoses to pyruvate through the non-oxidative pentose phosphate pathway and the Embden-Meyerhof pathway was detected in the *Tariuqbacter* MAG (Fig. 5), suggesting a sugar-based metabolism. The TCA cycle was incomplete in the MAG. However, an alcohol dehydrogenase gene was identified providing the Lake A *Tariuqbacter* with a pathway for pyruvate fermentation. Genes coding for pyruvate ferredoxin oxidoreductase (por) and 2-oxoglutarate ferredoxin oxidoreductase (kor), converting pyruvate into acetyl-CoA were detected as well as a gene coding for the acetyl-CoA synthetase (ADP-forming) allowing acetate production and ATP generation from acetyl-CoA. This feature, usually found in fermentative Archaea, has also been observed in Zixibacteria a taxonomically close phyla of Electryoneota, where it was proposed to catalyse the reverse reaction, enabling acetate utilisation [52]. Genes coding for pyruvate formate lyase and activating enzymes (plfACEX) were also detected, suggesting a potential for formate utilisation [53]. Formate could also be converted to acetyl-CoA through CO_2 and CO by formate dehydrogenase (fdo), anaerobic carbon-monoxide dehydrogenase (cooFS) and the acetyl-CoA decarbonylase/synthase complex (cdhCED) (Fig. 5). Formate is a volatile fatty acid frequently associated with metabolic exchanges and electron transfer, notably under anaerobic conditions [54, 55]. These results suggest that *Tariuqbacter* could be involved in cooperative interspecies interactions.

A pathway for the anaerobic degradation of aromatic compounds (benzoate and aminobenzoate) to acetyl-CoA was identified (Fig. 5). Genome analysis also suggested the capability of butyrate-fermentation in Tariuqbacteria, as reported in Electryoneia and Zixibacteria [56]. In this pathway, butyryl-CoA is converted from acetyl-CoA by four enzymes (Hbd, Crt, Ech and Bcd) and the electron transfer flavoprotein complex (Etf). The butyryl-CoA dehydrogenase-EtfAB complex provide another cytoplasmic flavin-based electron bifurcation mechanism for energy conservation and hydrogen production in *Tariuqbacter* [50]. Butyryl-CoA is then converted to butyrate by butyrate kinase (Buk) or acetyl-CoA synthetase (Acs) resulting in the generation of ATP by substrate-level phosphorylation (Fig. 5) [52, 56]. All of these processes can be reversible under certain conditions and may thus serve in alcohol, acetate, carbon monoxide and butyrate utilization, indicating the versatility of anaerobic metabolism in *Tariuqbacter*, and potentially explaining its predominance in Lake A. In addition, autotrophic abilities through reverse TCA cycle have been previously proposed for Electryoneota lineages recovered from Ace Lake. We did not detect genes coding acetate citrate lyase, fumarate reductase and citryl-CoA synthetase or lyase (*aclAB*, *frdABCDE*, *ccsAB* and *ccl*, respectively), nor did we detect other CO_2 fixation pathways. Although we can not exclude the possibility that key genes were not detected due to the incompleteness of the MAG, the lack of CO_2 fixation pathways suggests that *Tariuqbacter* has a strictly heterotrophic lifestyle.

Toxic success of *Tariuqbacter*?

Heavy metals such as mercury and arsenic occur naturally in the environment and are also conveyed from anthropogenic sources to the Arctic region by atmospheric circulation [57], leading to a bioaccumulation of these pollutants in Arctic aquatic ecosystems. For example, 0.1 µg of Hg per gram of sediment was detected in

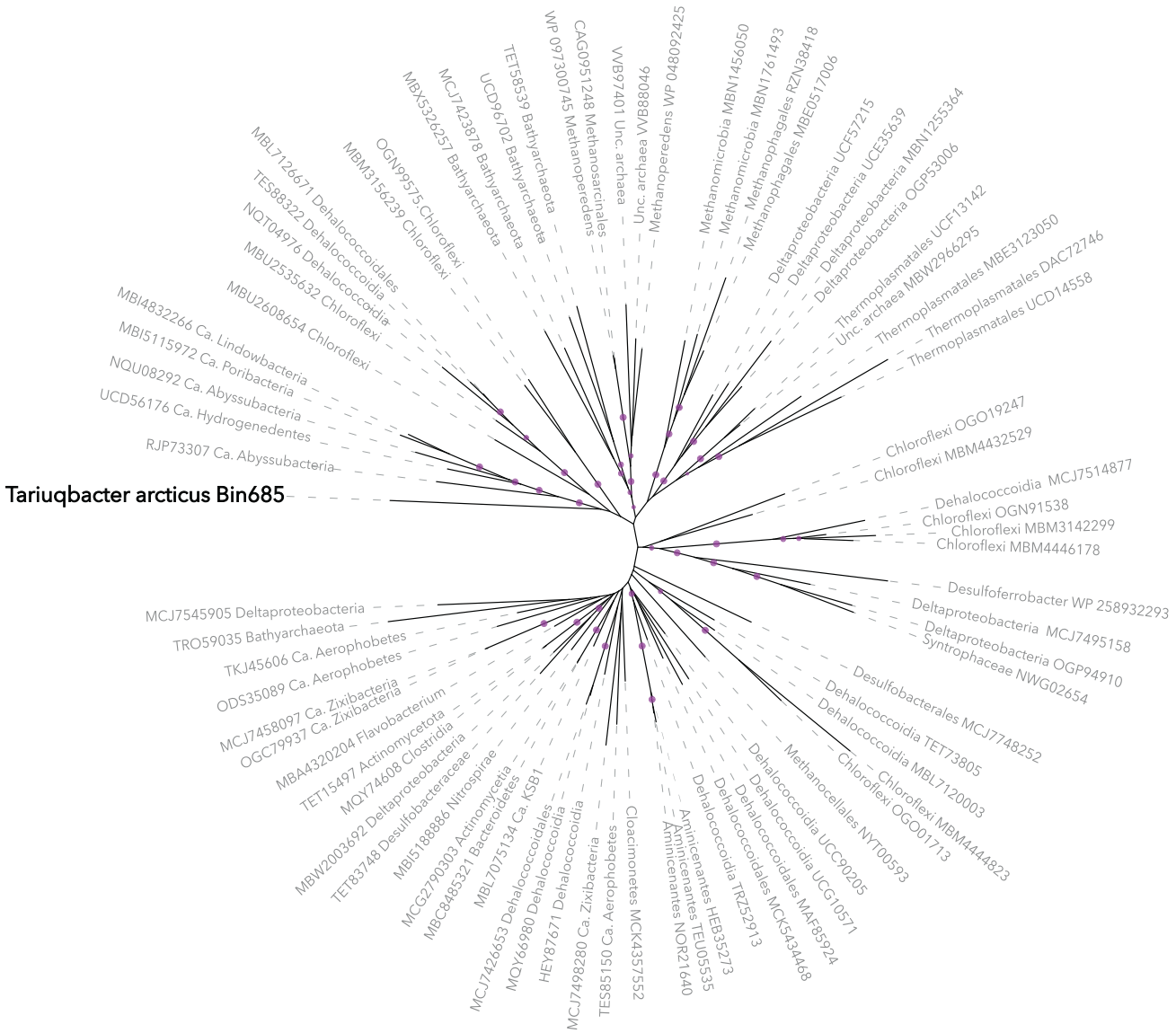

Fig. 6 Unrooted taxonomic tree of closest representatives for the arsenite S-adenomethyltransferase (ArsM) gene recovered from the *Tariuqbacteria* MAG.

Lake A [58], supporting the presence of heavy metals in the ecosystem. Arsenic has never been quantified in Lake A. However, 118 MAGs over the 250 good-quality MAGs (44.8%) recovered from Lake A samples from all depths included arsenic resistance and detoxification systems, such as arsenate reductase and arsenite transporters [59], suggesting that arsenic is likely present in the system. The *Tariuqbacter* MAG includes arsenite transporters (ACR3/arsB) and intracellular arsenate reductase (ArsC) that catalyzes the reduction of arsenate to arsenite. Interestingly, the *Tariuqbacter* MAG also includes the arsenite S-adenomethyltransferase gene (*arsM*), conferring the capability to methylate arsenite into highly toxic trivalent methylated species such as methylarsenite and dimethylarsenite [60, 61]. Under aerobic conditions, this enzyme is considered as a major detoxification pathway since trivalent methylated species are rapidly and abiotically oxidized to non-toxic pentavalent species methylarsenate, dimethylarsenate and volatile trimethylarsine [60]. However, under anoxic conditions such as at the bottom waters of Lake A, methylarsenite and dimethylarsenite are thermodynamically stable and are highly toxic antimicrobial compounds [62]. This gene has been previously detected in a few genomes related to the sister phyla Zixibacteria as well as in other anaerobic marine lineages (Fig. 6). However, phylogenetic analysis of the arsM sequence revealed that *Tariuqbacter* gene was more related to sequences recovered form *Candidate* phylum Hydrogenedentes/Abyssubacteria bacteria (Fig. 6), illustrating the numerous horizontal transfer events that characterize this gene [60]. Analysis of the sequence confirmed the presence of a SAM-binding domain, as well as the four conserved cysteine residues allowing arsenite binding and supporting its activity. A gene coding for a methylarsenite efflux permease (*arsP*) [63], which is frequently observed in anaerobic arsenite methylators [60], was also identified in the *Tariuqbacter* MAG. This enzyme protects the producing bacteria by pumping out toxic methylarsenite, leading to the death of neighboring bacteria [62]. This system has been proposed as an important selective mechanism that potentially favored cyanobacterial blooms in the anoxic paleo-ocean [62]. Known methylarsenite detoxification pathways so far are limited to oxic and nitrate-reducing conditions [64], and were only identified at the upper oxic layer of Lake A. Therefore, in Lake A, which might be considered a geochemical analog of the anoxic paleo-ocean, this pathway could confer a significant fitness

advantage for *Tariuqbacter*, potentially explaining their ecological success and their predominance in the anoxic marine waters of this ecosystem.

CONCLUSIONS

Isolated by perennial ice-cover at the northern limit of Canada, meromictic Lake A is a compelling model to explore microbial metabolism and adaptation to environmental extremes. Our genomic analysis of the deep anoxic and sulfidic water of the lake revealed the predominance of a novel candidate bacterial taxon that likely drives the sulfur cycle of the ecosystem and that may have a controlling influence on overall microbiome structure. Culturing or activity measurements are needed to confirm the ecophysiology of Tariuqbacteria. However, our results indicate that these bacteria may outcompete other microorganisms by relying on a combination of metabolic versatility and biotoxicity. Although additional explorations of polar lakes are needed to better estimate bacterial distribution in these remote ecosystems, this lineage may potentially represent a rare example of high taxonomy level endemism outside of hydro-geothermal areas. Our observations support the hypothesis that isolated polar habitats host an unexplored microbial diversity with lineages that escape global dispersion and that may avoid competition with cosmopolitan taxa.

DATA AVAILABILITY

Assembled metagenome data are available in IMG/MR (https://img.jgi.doe.gov/mer/) under the following accession numbers: 3300033443, 3300033444, 3300033445, 3300033439, 3300033411, 3300033473, 3300033474, 3300033495. Co-assembly is also available on IMG/MR under accession number 3300033064. Raw amplicon sequences and bin files were deposited in the NCBI public database under Bioproject PRJNA616293. The Tariuqbacter MAG is available in Supplementary Material as well as in GTDB-tk and GenBank/NCBI databases under accession number GCA_013202315.1/SAMN14944631. In addition, the full length 16 S rRNA gene sequence of *Ca*. Tariuqbacter is deposited in GenBank under accession number OQ709077. The list of Kegg Orthologies identified in the Tariuqbacter MAG is available in Supplementary Material. In-house scripts used in this study are available on GitHub/CruaudPe. Environmental metadata were previously published [7, 19, 42] and additional data for Lake A limnology and local climate are available in the Nordicana D database (https://nordicana.cen.ulaval.ca/).

REFERENCES

1. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, et al. Recovery of nearly 8000 metagenome-assembled genomes substantially expands the tree of life. Nat Microbiol. 2017;2:1533.
2. Castelle CJ, Banfield JF. Major new microbial groups expand diversity and alter our understanding of the tree of life. Cell. 2018;172:1181–97.
3. 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.
4. Baker BJ, De Anda V, Seitz KW, Dombrowski N, Santoro AE, Lloyd KG. Diversity, ecology and evolution of Archaea. Nat Microbiol. 2020;5:887–900.
5. Youssef NH, Farag IF, Hahn CR, Jarett J, Becraft E, Eloe-Fadrosh E, et al. Genomic characterization of candidate division LCP-89 reveals an atypical cell wall structure, microcompartment production, and dual respiratory and fermentative capacities. Appl Environ Microbiol. 2019;85:e00110–19.
6. 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.
7. 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.
8. Vigneron A, Cruaud P, Alsop E, de Rezende JR, Head IM, Tsesmetzis N. Beyond the tip of the iceberg; a new view of the diversity of sulfite- and sulfate-reducing microorganisms. ISME J. 2018;12:2096–9.
9. Spang A, Caceres EF, Ettema TJG. Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. Science. 2017;357:eaaf3883.
10. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Res. 2022;50:D785–94.
11. Fullam A, Letunic I, Schmidt TSB, Ducarmon QR, Karcher N, Khedkar S, et al. proGenomes3: approaching one million accurately and consistently annotated high-quality prokaryotic genomes. Nucleic Acids Res. 2023;51:D760–6.
12. 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.
13. De Anda V, Chen L-X, Dombrowski N, Hua Z-S, Jiang H-C, Banfield JF, et al. Brockarchaeota, a novel archaeal phylum with unique and versatile carbon cycling pathways. Nat Commun. 2021;12:2404.
14. McKay LJ, Dlakić M, Fields MW, Delmont TO, Eren AM, Jay ZJ, et al. Co-occurring genomic capacity for anaerobic methane and dissimilatory sulfur metabolisms discovered in the Korarchaeota. Nat Microbiol. 2019;4:614–22.
15. Power JF, Carere CR, Lee CK, Wakerley GLJ, Evans DW, Button M, et al. Microbial biogeography of 925 geothermal springs in New Zealand. Nat Commun. 2018;9:2876.
16. Vavourakis CD, Mehrshad M, Balkema C, van Hall R, Andrei A-Ş, Ghai R, et al. Metagenomes and metatranscriptomes shed new light on the microbial-mediated sulfur cycle in a Siberian soda lake. BMC Biol. 2019;17:69.
17. Williams TJ, Allen MA, Panwar P, Cavicchioli R. Into the darkness: the ecologies of novel 'microbial dark matter' phyla in an Antarctic lake. Environ Microbiol. 2022;24:2576–603.
18. Herbold CW, Lee CK, McDonald IR, Cary SC. Evidence of global-scale aeolian dispersal and endemism in isolated geothermal microbial communities of Antarctica. Nat Commun. 2014;5:3875.
19. Gibson JAE, Vincent WF, Van Hove P, Belzile C, Wang X, Muir D. Geochemistry of ice-covered, meromictic Lake A in the Canadian High Arctic. Aquat Geochem. 2002;8:97–119.
20. Vigneron A, Cruaud P, Lovejoy C, Vincent WF. Genomic evidence of functional diversity in DPANN archaea, from oxic species to anoxic vampiristic consortia. ISME Commun. 2022;2:4.
21. Comeau AM, Harding T, Galand PE, Vincent WF, Lovejoy C. Vertical distribution of microbial communities in a perennially stratified Arctic lake with saline, anoxic bottom waters. Sci Rep. 2012;2:604.
22. Vigneron A, Cruaud P, Lovejoy C, Vincent WF. Genomic insights into cryptic cycles of microbial hydrocarbon production and degradation in contiguous freshwater and marine microbiomes. Microbiome. 2023;11:104.
23. Cruaud P, Vigneron A, Fradette MS, Charrette S, Rodriguez M, Dorea CC, et al. Open the SterivexTM casing: an easy and effective way to improve DNA extraction yields. Limnol Oceanogr Methods. 2017;15:1015–20.
24. Cruaud P, Vigneron A, Fradette M-S, Dorea CC, Culley AI, Rodriguez MJ, et al. Annual bacterial community cycle in a seasonally ice-covered river reflects environmental and climatic conditions. Limnol Oceanogr. 2019;65:S21–37.
25. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75:7537–41.
26. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 2007;35:7188–96.
27. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.
28. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 2013;29:2933–2935.
29. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19:455–77.
30. Bushnell B. BBMap: A fast, accurate, splice-aware aligner. 2014. https://www.osti.gov/biblio/1241166.
31. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;31:1674–6.
32. Kang DD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ. 2015;3:e1165.
33. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25:1043–55.
34. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics. 2020;36:1925–7.
35. Rodriguez-R LM, Konstantinidis KT. Bypassing cultivation to identify bacterial species. Microbe. 2014;9:111–8.
36. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, et al. KEGG for linking genomes to life and the environment. Nucleic Acids Res. 2008;36:D480–4.

37. Zhou Z, Tran PQ, Breister AM, Liu Y, Kieft K, Cowley ES, et al. METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. Microbiome. 2022;10:33.
38. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30:3059–66.
39. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37:1530–4.
40. Letunic I, Bork P. Interactive tree of life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019;47:W256–W259.
41. Pallen MJ, Rodriguez-R LM, Alikhan N-F. Naming the unnamed: over 65,000 candidatus names for unnamed archaea and bacteria in the Genome Taxonomy Database. Int J Syst Evol Microbiol. 2022;72:005482.
42. Vincent WF, Fortier D, Lévesque E, Boulanger-Lapointe N, Tremblay B, Sarrazin D, et al. Extreme ecosystems and geosystems in the Canadian High. Arctic: Ward Hunt Island and vicinity. Écoscience. 2011;18:236–61.
43. Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH. Beyond biogeographic patterns: processes shaping the microbial landscape. Nat Rev Microbiol. 2012;10:497–506.
44. Oakley BB, Carbonero F, van der Gast CJ, Hawkins RJ, Purdy KJ. Evolutionary divergence and biogeography of sympatric niche-differentiated bacterial populations. ISME J. 2010;4:488–97.
45. Cho JC, Tiedje JM. Biogeography and degree of endemicity of fluorescent *Pseudomonas* strains in soil. Appl Environ Microbiol. 2000;66:5448–56.
46. Whitaker RJ, Grogan DW, Taylor JW. Geographic barriers isolate endemic populations of hyperthermophilic Archaea. Science. 2003;301:976–8.
47. Santos AA, Venceslau SS, Grein F, Leavitt WD, Dahl C, Johnston DT, et al. A protein trisulfide couples dissimilatory sulfate reduction to energy conservation. Science. 2015;350:1541–5.
48. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, et al. InterPro in 2022. Nucleic Acids Res. 2023;51:D418–27.
49. Pereira IAC, Ramos A, Grein F, Marques M, Da Silva S, Venceslau S. A comparative genomic analysis of energy metabolism in sulfate reducing bacteria and archaea. Front Microbiol. 2011;2:69.
50. Buckel W, Thauer RK. Flavin-based electron bifurcation, ferredoxin, flavodoxin, and anaerobic respiration with protons (Ech) or NAD+ (Rnf) as electron acceptors: a historical review. Front Microbiol. 2018;9:401.
51. Schut GJ, Bridger SL, Adams MW. Insights into the metabolism of elemental sulfur by the hyperthermophilic archaeon *Pyrococcus furiosus*: characterization of a coenzyme A- dependent NAD(P)H sulfur oxidoreductase. J Bacteriol. 2007;189:4431–41.
52. Castelle CJ, Hug LA, Wrighton KC, Thomas BC, Williams KH, Wu D, et al. Extraordinary phylogenetic diversity and metabolic versatility in aquifer sediment. Nat Commun. 2013;4:2120.
53. Zelcbuch L, Lindner SN, Zegman Y, Vainberg Slutskin I, Antonovsky N, Gleizer S, et al. Pyruvate formate-lyase enables efficient growth of *Escherichia coli* on acetate and formate. Biochemistry. 2016;55:2423–6.
54. Rotaru AE, Shrestha PM, Liu F, Ueki T, Nevin K, Summers ZM, et al. Interspecies electron transfer via hydrogen and formate rather than direct electrical connections in cocultures of *Pelobacter carbinolicus* and *Geobacter sulfurreducens*. Appl Environ Microbiol. 2012;78:7645–51.
55. Rabus R, Venceslau SS, Wöhlbrand L, Voordouw G, Wall JD, Pereira IAC. Chapter two - A post-genomic view of the ecophysiology, catabolism and biotechnological relevance of sulphate-reducing prokaryotes. Adv Microb Physiol. 2015;66:55–321.
56. Vital M, Howe AC, Tiedje JM. Revealing the bacterial butyrate synthesis pathways by analyzing (meta)genomic data. mBio. 2014;5:e00889–14.
57. Cheng MD, Hopke PK, Barrie L, Rippe A, Olson M, Landsberger S. Qualitative determination of source regions of aerosol in Canadian high Arctic. Environ Sci Technol. 1993;27:2063–71.
58. Muir DCG, Wang X, Yang F, Nguyen N, Jackson TA, Evans MS, et al. Spatial trends and historical deposition of mercury in Eastern and Northern Canada inferred from lake sediment Cores. Environ Sci Technol. 2009;43:4802–9.
59. Ben Fekih I, Zhang C, Li YP, Zhao Y, Alwathnani HA, Saquib Q, et al. Distribution of arsenic resistance genes in prokaryotes. Front Microbiol. 2018;9:2473.
60. Chen S-C, Sun G-X, Rosen BP, Zhang S-Y, Deng Y, Zhu B-K, et al. Recurrent horizontal transfer of arsenite methyltransferase genes facilitated adaptation of life to arsenic. Sci Rep. 2017;7:7741.
61. Ajees AA, Rosen BP. As(III) S-adenosylmethionine methyltransferases and other arsenic binding proteins. Geomicrobiol J. 2015;32:570–6.
62. Chen J, Yoshinaga M, Rosen BP. The antibiotic action of methylarsenite is an emergent property of microbial communities. Mol Microbiol. 2019;111:487–94.
63. Chen J, Madegowda M, Bhattacharjee H, Rosen BP. ArsP: a methylarsenite efflux permease. Mol Microbiol. 2015;98:625–35.
64. Chen J, Rosen BP. The arsenic methylation cycle: how microbial communities adapted methylarsenicals for use as weapons in the continuing war for dominance. Front Environ Sci. 2020;8:43.

ACKNOWLEDGEMENTS

We express our thanks to Larry Audlaluk for discussions in Grise Fiord about this work and the naming in Inuktitut. We thank Alexander Culley (Université Laval), Denis Sarrazin (Centre d'études nordiques) and other members of the Northern Ellesmere Island in the Global Environment (NEIGE) team for collaborating in the sampling effort; Perrine Cruaud for very helpful discussions; Parks Canada for access to facilities; Polar Continental Shelf Program and Canadian Helicopter Ltd. for the logistic support; and Compute Canada and Calcul Québec for computing resources and technical support.

AUTHOR CONTRIBUTIONS

AV and WFV designed the research, WFV obtained funding, logistics and permits for the research; WFV led the field sampling; CL provided laboratory and computational access; AV generated the molecular data and analyzed the data; AV wrote the paper with WFV and CL.

FUNDING

This work was supported by the program Sentinel North financed by the Canada First Research Excellence Fund (CFREF); the Natural Sciences and Engineering Research Council of Canada (NSERC); the CNR (Italy) – Université Laval UMI-MicroMeNu program; the fonds de Recherche du Quebec Nature et Technologies (FQRNT); and the Network of Centres of Excellence ArcticNet.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43705-023-00287-9.

Correspondence and requests for materials should be addressed to Adrien Vigneron.

Reprints and permission information is available at http://www.nature.com/reprints

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