

<https://doi.org/10.1038/s42003-024-07027-2>

# *Candidatus Desulforudis audaxviator* dominates a 975 m deep groundwater community in central Sweden



George Westmeijer<sup>1,6</sup>  , Femke van Dam<sup>2</sup> , Riikka Kietäväinen<sup>3,4</sup> , Carolina González-Rosales<sup>1</sup>, Stefan Bertilsson<sup>5</sup> , Henrik Drake<sup>2</sup>  & Mark Dopson<sup>1</sup> 

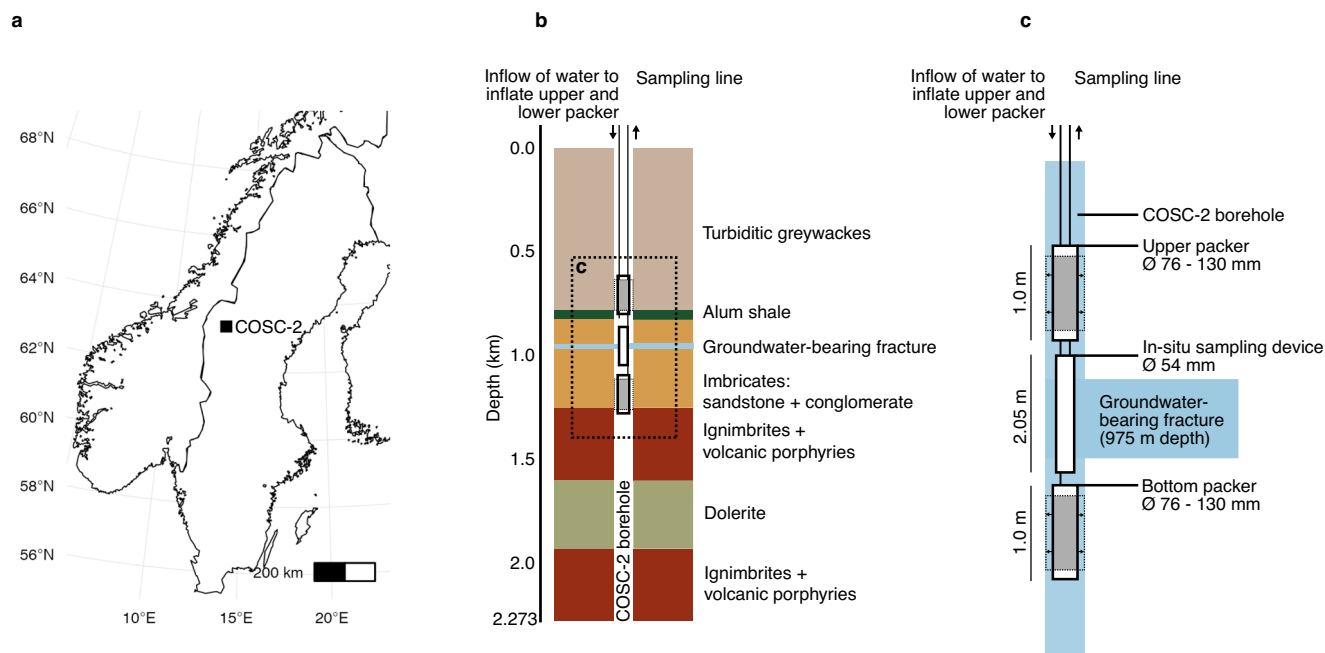
The continental bedrock contains groundwater-bearing fractures that are home to microbial populations that are vital in mediating the Earth's biogeochemical cycles. However, their diversity is poorly understood due to the difficulty of obtaining samples from this environment. Here, a groundwater-bearing fracture at 975 m depth was isolated by employing packers in order to characterize the microbial community via metagenomes combined with prokaryotic and eukaryotic marker genes (16S and 18S ribosomal RNA gene). Genome-resolved analyses revealed a community dominated by sulfate-reducing Bacillota, predominantly represented by *Candidatus Desulforudis audaxviator* and with Wood-Ljungdahl as the most prevalent pathway for inorganic carbon fixation. Moreover, the eukaryotic community had a considerable diversity and was comprised of mainly flatworms, chlorophytes, crustaceans, ochrophytes, and fungi. These findings support the important role of the Bacillota, with the sulfate reducer *Candidatus Desulforudis audaxviator* as its main representative, as primary producers in the often energy-limited groundwaters of the continental subsurface.

Microorganisms play a vital role in the deep subsurface, a largely unexplored biome up to several kilometers beneath the Earth's surface<sup>1</sup>. Despite being predominantly energy and nutrient limited, the deep subsurface hosts specialized populations that are involved in biogeochemical cycling by mediating the transformation of organic and inorganic matter<sup>2</sup>. Focusing on the continental subsurface, recent studies have used former mines<sup>3–6</sup>, boreholes<sup>7–11</sup>, and dedicated underground laboratories<sup>12–14</sup> to show that deep subsurface communities can be low diversity such that they are dominated by one bacterial population<sup>15</sup> or methanogenic archaea<sup>16</sup>, mixed bacterial communities active in sulfur cycling<sup>17</sup>, or be predominantly eukaryotic<sup>18</sup>. Depending on the host rock, geogenic hydrogen from the deep subsurface can be the main energy source for deep subsurface life<sup>19</sup> with chemolithotrophs such as sulfate reducers, methanogens, and acetogens competing for hydrogen to fuel reduction of sulfate or carbon dioxide. As geogenic hydrogen production by rock-water interactions is generally limiting<sup>20</sup>, groundwater circulation is key as it is thought to support these hydrogen-dependent populations<sup>16</sup>. Apart from certain sedimentary reservoirs, e.g., oil reservoirs<sup>21</sup>, organic carbon is usually scarce or of a refractory nature and autotrophic growth tends to be abundant, if not dominant, in the deep

subsurface<sup>22,23</sup>. The concentration of molecular oxygen dictates the mechanism of inorganic carbon fixation as, for example, the Wood-Ljungdahl pathway requires strict anoxia and is favored by prokaryotes in energy-limited conditions<sup>24</sup>. Additionally, it has been suggested that anaerobic fungi could produce hydrogen during respiration<sup>25</sup> and in line with this, consortia of fungi with hydrogenotrophic sulfate reducers<sup>26</sup> and methanogens<sup>27</sup> have been described in the continental subsurface.

The Precambrian craton, the Fennoscandian Shield, encompasses 900 to 3100 Ma old terranes in Norway, Sweden, Finland, and Russia. In terms of tectonic activity, the Fennoscandian Shield is among the most stable geological regions on Earth, with little tectonic activity since the formation of the Caledonian orogen in the Silurian<sup>28</sup>. Previous studies on subsurface groundwaters in the Fennoscandian Shield showed a decrease of microbial abundance<sup>29,30</sup> and a changing microbial community<sup>31,32</sup> with depth. For example, a study on groundwater-bearing fractures intersected by the Outokumpu borehole (Finland) revealed cell numbers to decrease with depth and a microbial community closer to the surface (0–300 m depth) dominated by Alpha- and Betaproteobacteria while the Bacillota (synonym Firmicutes) were abundant below 900 m depth<sup>29</sup>. In addition,

<sup>1</sup>Centre for Ecology and Evolution in Microbial model Systems (EEMiS), Linnaeus University, Stuvaregatan 4, Kalmar, Sweden. <sup>2</sup>Department of Biology and Environmental Sciences, Linnaeus University, Stuvaregatan 4, Kalmar, Sweden. <sup>3</sup>Geological Survey of Finland, Espoo, Finland. <sup>4</sup>Department of Geosciences and Geography, University of Helsinki, Helsinki, Finland. <sup>5</sup>Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, Uppsala, Sweden. <sup>6</sup>Present address: Department of Chemistry, Umeå University, Umeå, Sweden. ✉e-mail: [george.westmeijer@umu.se](mailto:george.westmeijer@umu.se)



**Fig. 1 | Field site and groundwater collection using inflatable packers.** **a** Location of the COSC-2 borehole within Sweden (Lat 63.3124° N, Lon 13.5265° E).

**b** Lithology of the borehole according to depth and the position of the packers used for isolating the groundwater-bearing fracture at 975 m below the surface (not drawn to scale). **c** Detail of the packers used during sampling. The packers (76 mm in

diameter) were inflated against the borehole wall (96 mm in diameter) by pressurizing the rubber shell (gray surface) with water, thereby preventing to sample groundwater from other fractures intersected by the borehole. The samples for hydrochemistry were collected using both pumped water and an in situ sampling device that allowed water collection at the desired depth.

thermochronology models suggest that habitable temperatures (below 122 °C) prevailed during the last ~300 million years for the upper few kilometers of most parts of this craton<sup>33</sup>. Microbiological investigations on lower reaches of thick Paleozoic sedimentary piles that overlay the Precambrian strata in certain areas of the Fennoscandian shield, such in the Caledonian thrust sheets, are however more or less absent.

Highly isolated fluids at 755 m below sea level in the Paleozoic carbonate aquifer in the Death Valley (southeastern California), within Archaean metasediments at Mponeng, the Beatrix and Tau Tona gold mines in the Witwatersrand Basin (South Africa), as well as a borehole drilled into a Cretaceous aquifer in the Siberian artesian mega-basin have all been shown to hold a microbial community dominated by *Candidatus Desulfuridis audaxviator*, with very low diversity<sup>9,15,34</sup>. This sulfate-reducing bacterium is of particular relevance for subsurface microbial ecology as it has a highly similar nucleotide identity (>99%) across continents (Africa, Eurasia, and North America) and has been hitherto exclusively described in deep subsurface groundwaters.

The COSC-2 (Collisional Orogeny in the Scandinavian Caledonides) borehole, located in central Sweden, was drilled through an upper 1250 m vertical section of early Paleozoic bedrock thrust sheets of the easternmost parts of the Caledonian orogen range that overlays deeper Precambrian crystalline bedrock (porphyries of ~1.8 Ga, and sections of dolerite intrusions) of the Fennoscandian Shield<sup>35</sup>. The drilling took place during spring to summer 2020 and extends 2250 m below the surface (elevation field site 320 m), thereby intersecting multiple natural, groundwater-bearing fractures. To accurately characterize this subsurface microbial diversity, it is desirable to isolate the groundwater-bearing fracture from other such fractures as these subsurface environments have been long-term undisturbed and to prevent the microbial communities to mix. It is also important to avoid intrusion of liquid from the main open borehole water pillar into the sampling connection, as this liquid is likely to contain contaminant microbial populations from the drilling operation<sup>36</sup>. One promising strategy to collect groundwater from a certain section in the borehole while reducing contamination from stagnant water in the main borehole, is to use inflatable packers to isolate the fracture of interest<sup>37</sup>.

In this study, the microbial community in a groundwater-bearing fracture intersected by the COSC-2 borehole was characterized using metagenome sequencing combined with 16S and 18S ribosomal RNA gene amplicons. Microbial abundance was assessed using a combination of real-time PCR and flow cytometry. The main goal was to characterize this potentially long-term isolated subsurface microbial community in conjunction with characterization of the eukaryotic community.

## Results and discussion

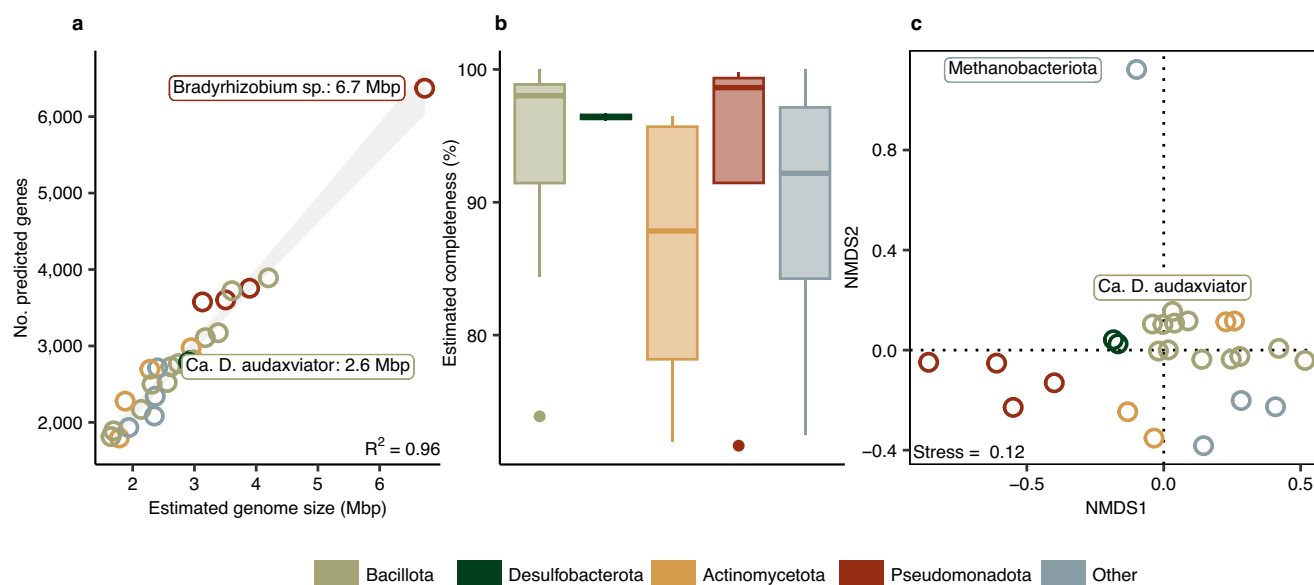
### Groundwater characteristics

A groundwater-bearing fracture at 975 m depth located in a lower Paleozoic conglomerate rock section, intersected by the COSC-2 borehole, was isolated by employing inflatable packers (Fig. 1). The pH of the groundwater at this depth was 9.8 and the oxygen concentration in the range of 1.6 and 2.0 mg L<sup>-1</sup> ( $n = 2$ , Thomas Wiersberg, pers. comm.). Given the large depth of the borehole and the dominance of anaerobic bacterial populations, the low but detectable amount of oxygen indicated a possible contamination from either the drilling operation (drilled in 2020), sampling and sample treatment, or as a product of microbial activity<sup>10</sup>. The alkalinity was 24 mg CaCO<sub>3</sub> L<sup>-1</sup>, the electrical conductivity 11.1 mS cm<sup>-1</sup>, and the chloride to bromide mass ratio 142:1 (Cl<sup>-</sup> 3600 mg L<sup>-1</sup>, Br<sup>-</sup> 57 mg L<sup>-1</sup>). An overview of the hydrochemistry is provided in Supplemental Table 1. The groundwater intersected by the COSC-2 borehole most resembled groundwaters from Forsmark (Sweden), recharged by brackish glacial meltwater, at a depth of 409 to 549 m below sea level<sup>38</sup> (Supplemental Fig. 1). In contrast, the pH, chloride and magnesium concentrations, and the chloride to bromide mass ratio from the COSC-2 groundwater differed from groundwaters intersected by boreholes at the Paleoproterozoic granitoid-hosted Äspö Hard Rock Laboratory (Sweden) at depths between 69 and 467 m<sup>30,31</sup>, and groundwaters intersected by Outokumpu borehole (Finland) at 500, 1000, and 1500 m depth (hosted by wall rocks Proterozoic mica schist and black schist)<sup>29</sup>. Hence, comparison of this and other deep subsurface groundwaters located in the Fennoscandian Shield, based on chloride and magnesium concentration and the chloride to bromide mass ratio, indicated a non-marine origin of the studied groundwater.

**Table 1 | Details of groundwater sampling**

Filter	Groundwater filtered (L)	DNA yield (ng L <sup>-1</sup> )	Bacterial gene copies (mL <sup>-1</sup> )	Archaeal gene copies (mL <sup>-1</sup> )	No. ASVs / bins
1 + 2	4.1	20.1	$3.3 \times 10^4 \pm 153$	$27 \pm 1.2$	177/24
3 + 4	2.9	11.9	$1.7 \times 10^4 \pm 1.1 \times 10^3$	$27 \pm 3.6$	134/26
5 + 6	2.6	8.2	$1.5 \times 10^4 \pm 523$	$24 \pm 1.6$	218/25
7	0 (control)	< 0.05	BD	BD	43

Groundwater samples were collected and filtered on 3 September 2022. Two filters were combined during DNA extraction to increase the yield, except for the negative control sample. The DNA yield is presented as the amount of DNA extracted standardized by the volume of groundwater filtered (ng L<sup>-1</sup>). The reported bacterial and archaeal abundances (mean  $\pm$  sd,  $n = 3$ ) were quantified using qPCR, targeting the 16S rRNA gene. The number of gene copies in the control sample were below detection limit (BD;  $\Delta$ Cq with no-template controls below 3).



**Fig. 2 | General characteristics of the de-replicated genomes.** **a** Number of predicted genes versus estimated genome size showing a high association ( $R^2 = 0.96$ ). **b** Estimated genome completeness as quantified using 104 marker genes. Out of a total of 26 de-replicated genomes, eighteen had a completeness above 90% and eight

had a completeness between 70 and 90%. **c** Divergence among genomes based on functional orthologous genes (KO functional orthologs, accounting for multiple gene copies) using non-metric dimensional scaling (NMDS).

### Sequencing output

To characterize the subsurface community, cells from a total of 9.6 L groundwater were captured by membrane filtration ( $n = 6$ , pore size 0.1  $\mu$ m). Two membranes were pooled during DNA extraction, resulting in three DNA extracts that contained 20, 12, and 8 ng L<sup>-1</sup> genomic DNA per volume groundwater filtered (Table 1). Sequencing of 16S rRNA gene amplicons produced a total of 257 amplicon sequencing variants (ASVs) from  $78 \times 10^3$  sequencing reads (standard deviation  $23 \times 10^3$ ,  $n = 3$ , Supplemental Table 2). In addition, sequencing of the 18S rRNA gene produced a total of 2982 ASVs from  $136 \times 10^3$  (sd  $43 \times 10^3$ ,  $n = 3$ ). Rarefaction curves were asymptotic and clearly demonstrated sufficient sequencing depth to capture the diversity in the sequencing libraries (Supplemental Fig. 2). Contamination during either sample collection, DNA extraction, or DNA amplification is a major concern as biomass in the deep subsurface is usually several orders of magnitude lower compared to samples retrieved from surface environments<sup>39</sup>. The DNA concentration of the negative control collected at the field site (blank filter) was below detection limit ( $0.05 \text{ ng } \mu\text{L}^{-1}$ ) and PCR amplification did not yield a visible product during gel electrophoresis. Nevertheless, this library was sequenced and featured 42 ASVs from a total of 2183 reads. The community composition in the control sample did not resemble the community in the samples of interest (Supplemental Fig. 2). Only two ASVs were overlapping between the communities and those were removed from downstream analyses.

The metagenomes ( $n = 3$ ) contained a minimum of 16.1 and a maximum 25.2 million paired-end reads and produced 26 unique de-replicated

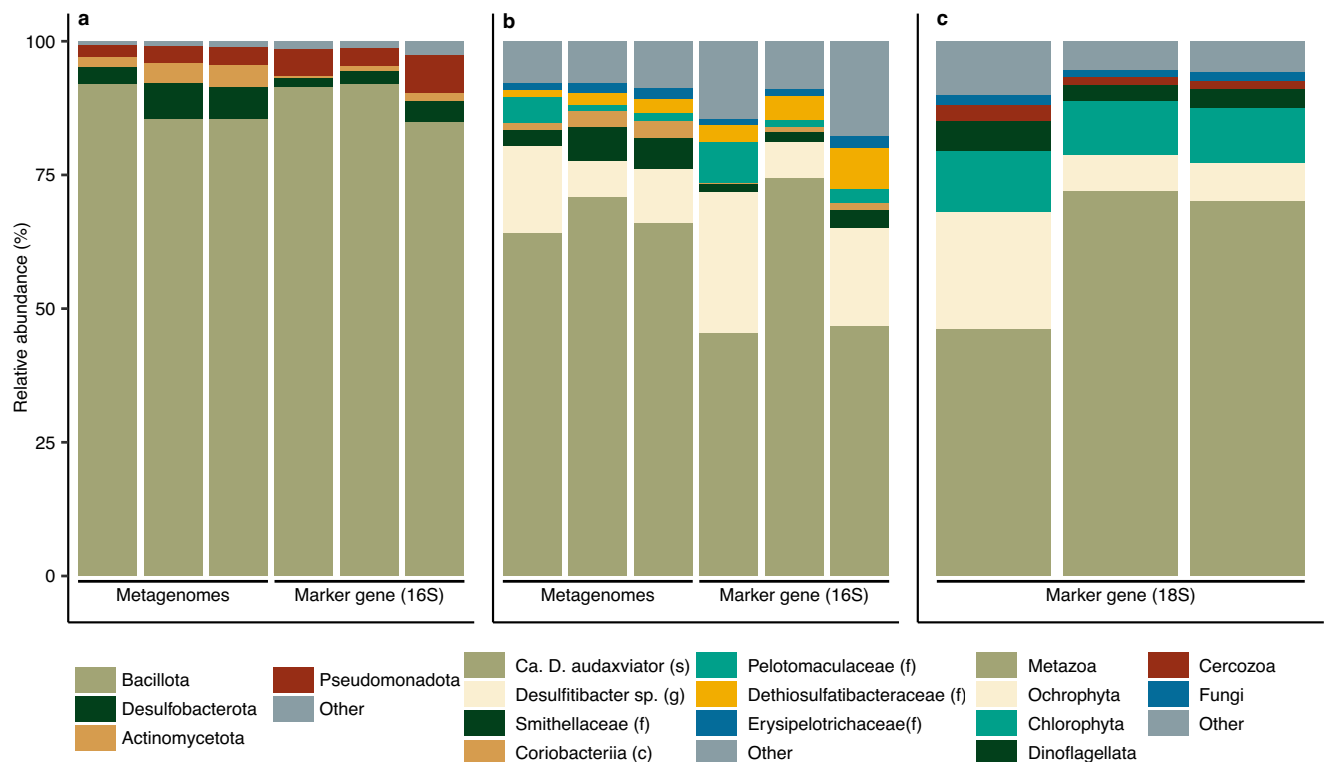
metagenome-assembled genomes (MAGs; hereafter termed reconstructed genomes). All were bacterial except one archaeal (Fig. 2). Eighteen genomes had an estimated completeness above 90% and the remaining eight had an estimated completeness between 70 and 90%<sup>40</sup>.

### Microbial abundance

Bacterial abundance (Table 1) was  $22 \times 10^3$  16S rRNA gene copies mL<sup>-1</sup> (sd  $9.0 \times 10^3$ ,  $n = 18$ ) while the archaeal abundance was much lower with only 26 gene copies mL<sup>-1</sup> (sd 2.4,  $n = 18$ ). Based on flow cytometry, the microbial abundance was estimated to be in the range of 22.4 and  $22.8 \times 10^3$  cells mL<sup>-1</sup> ( $n = 2$ ). These numbers were in the same range ( $1.3 \times 10^4$  to  $4.4 \times 10^5$  cells mL<sup>-1</sup>) as previous epifluorescence microscopy-based estimates from a groundwater-bearing fracture at 967 m depth, intersected by the Outokumpu borehole, also located in the Fennoscandian Shield<sup>37</sup>. In general, multiple studies on the continental deep subsurface report cell density to be in the order of  $10^4$  cells mL<sup>-1</sup> groundwater<sup>12,30,41</sup>.

### Microbial community structure

The subsurface community was dominated by the Bacillota (synonym Firmicutes), with 12 of the 26 reconstructed genomes being affiliated with this phylum and accounting for 88% of the metagenomic read coverage (relative abundance was expressed as percentage read coverage; Fig. 3). More specifically, *Candidatus Desulforudis audaxviator* and *Desulfitibacter* sp. (both representatives of the Bacillota) were highly abundant and



**Fig. 3 | Overview of the microbial community structure.** **a** Bacterial community composition depicted at the level of phylum. Bacillota (synonym Firmicutes) dominated both the metagenomes and the marker-gene dataset in terms of relative abundance. Phyla are arranged according to abundance while grouping low-abundant phyla as “Other”. **b** Bacterial community composition depicted at the

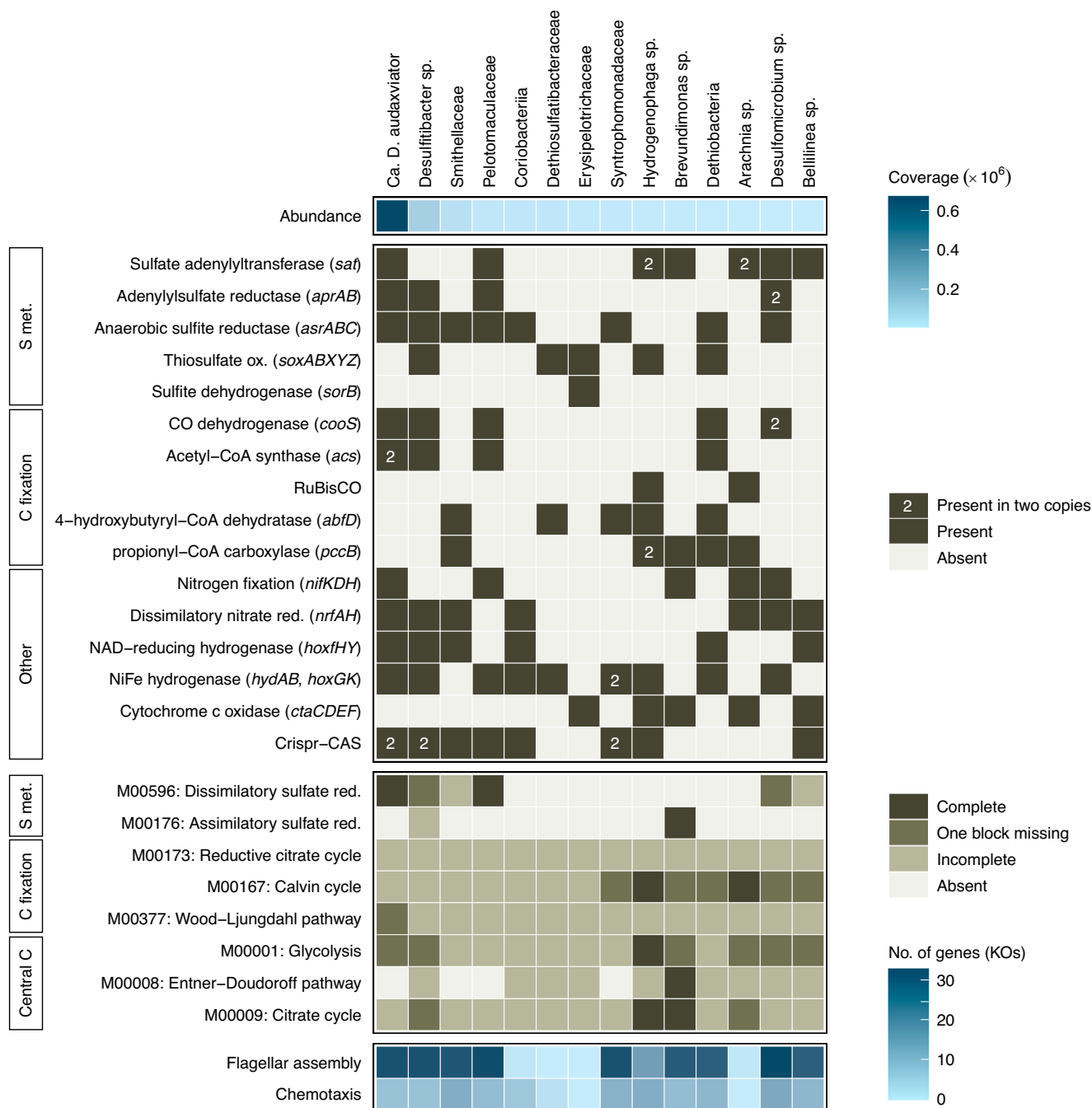
highest taxonomic resolution possible, i.e., class (c), family (f), genus (g), or species (s). **c** Eukaryotic community composition based on 18S rRNA gene amplicons, depicted at the level of phylum. Phyla are arranged according to abundance while grouping low-abundant phyla as “Other”.

accounted for 78% of the metagenomic read coverage. The remaining reconstructed genomes were affiliated with the phyla Desulfobacterota (5.4% read coverage), Actinomycetota (synonym Actinobacteria; 3.4%), Pseudomonadota (synonym Proteobacteria; 2.9%), Chloroflexota (0.4%), Bacteroidota (0.4%), and Methanobacteriota (0.1%). The community structure according to the metagenomes strongly resembled the community based on the PCR-based 16S ribosomal RNA although no archaea were detected in the latter assay. This implied that the primers used adequately captured the bacterial subsurface diversity in this groundwater even if the primers were originally designed for marine ecosystems. It is not unusual for populations affiliated with the Bacillota to be abundant or even dominate groundwater communities in the deep subsurface<sup>3,9,15</sup> and this has been associated with the long residence time of the groundwater with the ability of Bacillota to thrive in bedrock fractures isolated from recent surface recharge<sup>3</sup>. Together with the stability (low tectonic activity), hydro-chemistry, and age (Precambrian) of the studied bedrock<sup>28</sup>, the dominance of Bacillota suggested this groundwater-bearing fracture at 975 m depth to be long-term isolated from surface recharge.

The taxonomic resolution of the eukaryotic ASVs (according to 18S rRNA gene amplicon sequencing) was similar to the prokaryotic ASVs, with 91% of the eukaryotic ASVs characterized on the level of phylum, 82% on order, and 53% on genus compared to 94, 83, and 59% for the prokaryotic ASVs, respectively. Furthermore, the eukaryotic community comprised 297 eukaryotic genera while 126 prokaryotic genera were detected. Whether this eukaryotic diversity originated from surface water recharge or leached from the soil was unclear. The eukaryotic community was dominated by the flatworm *Praeconvoluta castinea* that was responsible for 45% of the amplicon read counts (Fig. 3). Other abundant groups were the chlorophyte *Desmodesmus* sp. (6.9%), the crustacean *Acartia tonsa* (6.6%), and the ochrophyte *Spumella elongata* (3.7%). Borgonie et al.<sup>18</sup> demonstrated the presence of eukaryotic groups such as flatworms, fungi, and arthropods in the

fracture waters on the deep subsurface and suggested that food instead of dissolved oxygen was the limiting factor for growth. Additional studies found fungi and chlorophytes to be abundant in deep subsurface groundwaters<sup>42</sup>. Thus, all the major eukaryotic groups presented here have been detected before in groundwaters of the deep subsurface. A considerable proportion of the eukaryotic community presumably respire aerobically (for example, *Desmodesmus* sp. or *Acartia tonsa*) and their origin and function in the groundwater under scrutiny here is largely unknown. Possibly, some eukaryotic groups originated from the main borehole pillar, despite using packers for a more targeted groundwater retrieval. However, as the eukaryotic community was reconstructed from the same DNA sample as the prokaryotic community, one would expect aerobic clades also to be abundant in the latter community while the abundance of aerobic prokaryotes was rather low.

Fungi were also detected and were mainly represented by the Cryptomycota (1.2%), Oomycota (0.8%), and Chytridiomycota (0.4%). ASVs affiliated with these fungal clades ( $n = 193$ ) were, in addition to the PR<sup>2</sup> database<sup>43</sup>, aligned with entries in the RefSeq NCBI database<sup>44</sup> but did not result in additional taxonomic delineation. Previously the Cryptomycota have been detected in the continental deep biosphere where they can inhabit sedimentary rock matrices at around 250 m depth<sup>42</sup> and subsurface groundwaters at around 10 m depth<sup>45</sup>. Literature on this fungal clade in the deep subsurface in general and deep fracture fluids is either scarce or absent, respectively. This phylum was only recently discovered using a genetic approach and the Cryptomycota are unique as fungal representatives in lacking the chitin-rich cell wall<sup>46</sup>. Representatives of the Cryptomycota are described in the context of an endo- or epibiotic lifestyle with the potential host affiliated with the phyla Oomycota or Chytridiomycota<sup>47</sup>. The presence of the Cryptomycota, Oomycota, and Chytridiomycota in this groundwater-bearing fracture at 975 m depth hints at a putative unrecognized role of fungi as potential decomposers of refractory organic matter in the continental deep biosphere<sup>26,27</sup>.



**Fig. 4 | Genomic potential of the most abundant clades.** Included in the heat map are (from top to bottom) abundance, genomic potential based on marker genes (present / absent), completeness of selected pathways (the module identifier referring

to KEGG modules), and pathways of flagellar assembly and chemotaxis (no. of gene orthologs). Only genomes with a metagenomic read coverage above 1% are included. The EC references for the marker genes are provided in the method section.

### Genomic potential

A substantial portion of the reconstructed genomes contained genes involved in carbon fixation (9 out of 26 genomes, metagenomic read coverage 85%), nitrogen fixation (6, 72%), hydrogen oxidation (16, 94%), sulfate reduction (3, 72%), and sulfur oxidation (14, 93%).

Both the presence of individual marker genes and pathway completeness in the reconstructed genomes were used to identify how the detected subsurface populations could fix carbon. Of the major fixation pathways, the reductive acetyl-CoA (Wood-Ljungdahl) pathway was the most prevalent (Fig. 4), and the genomes of the two most abundant populations (together accounting for 78% metagenomic read coverage) contained carbon monoxide dehydrogenase (*cooS*) and acetyl-CoA

synthase (*acsA*). The Wood-Ljungdahl pathway is reported to be used by prokaryotes living in strict anoxic, energy-limited environments that are close to the thermodynamic limit of life, and is observed in e.g., sulfate-reducing bacteria, acetogenic bacteria, and methanogenic archaea<sup>24</sup>. Propionyl-CoA carboxylase (key enzyme of the 3-hydroxypropionate bi-cycle) was also abundant. This cycle differs from the Wood-Ljungdahl pathway in that it uses bicarbonate as a carbon source instead of carbon dioxide and also has a higher energetic cost while requiring less metals and coenzymes<sup>24,48</sup>. Additionally, 4-hydroxybutyryl-CoA dehydratase (key enzyme of the dicarboxylate-hydroxybutyrate cycle) was detected in multiple genomes. This enzyme is a key enzyme in the oxygen-insensitive hydroxypropionate-hydroxybutyrate cycle; however, comparing pathway



completeness revealed that it was more likely that the oxygen-sensitive dicarboxylate-hydroxybutyrate cycle was used. The latter cycle, together with the 3-hydroxypropionate bi-cycle, are described as using bicarbonate as an inorganic carbon source and this may be beneficial under the prevailing alkaline conditions<sup>24</sup>. Regarding the reductive citrate cycle, acetyl-CoA is formed from CO<sub>2</sub> and the further conversion of acetyl-CoA requires three enzymes (fumarate reductase, 2-oxoglutarate oxidoreductase, and ATP-citrate lyase) that are characteristic of this cycle<sup>49</sup>. As only two out of three of these enzymes were detected, with ATP-citrate lyase not present in any of the genomes, the fumarate reductase and 2-oxoglutarate oxidoreductase were most likely involved in the conventional citrate cycle rather than the reverse cycle<sup>49</sup>.

The observed completeness of the carbon fixation pathways was low. Despite high estimated genome completeness with on average 96% (sd 7.0,  $n = 14$ ) for the more abundant populations included in this analysis (read coverage > 1%), it was challenging to robustly determine whether the low pathway completeness was due to genome completeness or rather due to a reduced genome size<sup>50</sup>. Of all the major carbon fixation pathways included in this analysis, only two genomes (*Arachnia* sp. and *Hydrogenophaga* sp.) contained all the genes comprising one of these pathways, i.e., the Calvin cycle. Interestingly, aside from containing all the genes for the energy-demanding Calvin cycle, the genome of *Hydrogenophaga* sp. also contained the aerobic type of carbon monoxide dehydrogenase (*cutSML*) and the key enzymes of the 3-hydroxypropionate bi-cycle (*abfD*) and the dicarboxylate-hydroxybutyrate cycle (*pccAB*). This genus is repeatedly detected in subsurface groundwaters<sup>3,10,29</sup> and cultured representatives have been described to be both chemoorganotrophic and chemolithotrophic<sup>51</sup>. The genome of *Hydrogenophaga* sp. was different from the majority of reconstructed genomes as it contained several genes (RuBisCO, *cutSML*, and *abfD*) involved in multiple carbon fixation pathways. This was unlike most genomes, which encoded only one carbon fixation pathway (Fig. 4). Having the genomic potential for aerobic respiration (NADH ubiquinone oxidoreductase, the cytochrome *bc*<sub>1</sub> complex, and cytochrome *c* oxidase) and aerobic carbon fixation suggested that *Hydrogenophaga* sp. could act as an oxygen scavenger coupled to hydrogen oxidation, while having the genomic potential for thiosulfate oxidation to sulfate (*glpE*, *sox*), and oxidation of sulfur to sulfite (*sdo*) in the absence of oxygen. In general, based on the presence of key enzymes, and especially when taking the abundance of the individual populations into account, the Wood-Ljungdahl pathway was the most prevalent carbon fixation pathway in the community, supporting the trend that this pathway is selected for in low-energy ecosystems such as the continental deep subsurface.

Regarding nitrogen cycling, genes involved in nitrogen fixation (*nifDHK*) and dissimilatory nitrate reduction to ammonium (*nrfADH*) were prominent in abundant populations affiliated with the Bacillota. Genes coding for nitrification (*amoABC*) and anammox (*hzsA*, *hzsA*) were not detected while those encoding for nitrate and nitrite reduction (*napAB*, *narGH*, *nirKS*, *nirBD*) were solely detected in low-abundant genomes affiliated with the Pseudomonadota and Actinomycetota. This suggested that populations affiliated with the Bacillota mediated nitrogen fixation and dissimilatory nitrate reduction to ammonium (DNRA) in this subsurface groundwater, while denitrification was mainly encoded by less abundant populations affiliated with the Pseudomonadota and the Actinomycetota.

For sulfur cycling, 14 out of 26 reconstructed genomes (read coverage 93%) encoded genes involved in sulfur metabolism. These populations were affiliated with Bacillota (eight reconstructed genomes), Pseudomonadota (4), and Desulfobacterota (2). The high abundance of populations potentially involved in sulfate reduction (*aprAB*, *sat*) and the presence of populations containing genes involved in sulfur oxidation (*sdo*, *sor*) suggested that sulfur cycling could occur in this groundwater. Sulfur isotope measurements on the dissolved sulfate (0.5 mM, total sulfur pool 1.0 mM) in the groundwater yielded a  $\delta^{34}\text{S}$  value of 12.4‰ VCDT (sd 0.14) that was similar or slightly lower (by 10–20‰) compared to Paleozoic marine sulfate  $\delta^{34}\text{S}$ <sup>52</sup>, meaning that the sulfate may originate from such connate waters.

### *Candidatus Desulforudis audaxviator*

Based on the metabolic weight score, a metric that takes the genomes' metabolic potential and coverage into account, *Ca. Desulforudis audaxviator* contributed to most of the metabolic functions related to sulfur cycling, carbon fixation, and nitrogen fixation (Supplemental Fig. 3). To date, *Ca. Desulforudis audaxviator* has been exclusively detected in the continental deep subsurface and has been reported to dominate fracture waters at ~3 km depth in the Mponeng (pH 8.6) and Tau Tona (pH 8.6) gold mines (South Africa)<sup>3,15</sup>, a 2-km-deep groundwater (pH 8.2) in Siberia<sup>34</sup>, and a groundwater (pH 8.2) in North America at 750 m depth<sup>9</sup>. The success of the sulfate-reducing lineage *Ca. Desulforudis audaxviator* in these deep subsurface and slightly alkaline ecosystems has been ascribed to the ability to switch between an autotrophic and heterotrophic lifestyle with concomitant potential to fix nitrogen<sup>15</sup>. Indeed, its genome (estimated to be 2.6 Mbp in size, GC content 60.9%, 98% complete, 2728 predicted genes) contained the key genes necessary for dissimilatory sulfate reduction, nitrogen fixation, autotrophic carbon fixation, and flagellar assembly (Fig. 4). The genome also contained two clustered regularly interspaced short palindromic repeat (CRISPR) and adjacent CRISPR-associated genes, potentially involved in viral defense. Comparison of the average nucleotide identity (ANI) of the reconstructed genomes generated in this study ( $n = 2$ ) with those from South Africa<sup>3,6,53</sup> ( $n = 7$ ) and Russia<sup>34</sup> ( $n = 1$ ) resulted in an average ANI of 98.8% (sd 0.22, minimum 98.3,  $n = 10$ ; Fig. 5) as compared to 99.5% (sd 0.35, minimum 98.8,  $n = 8$ ) when omitting the reconstructed genomes from the present study. These results suggested that the reconstructed genomes generated with this study are less similar compared to those from South Africa and Russia. However, it should be noted that the variations in DNA extraction methods possibly influenced this relatively low average ANI compared to the minimum of 99.2% reported by Becraft et al.<sup>9</sup>

### Conclusion

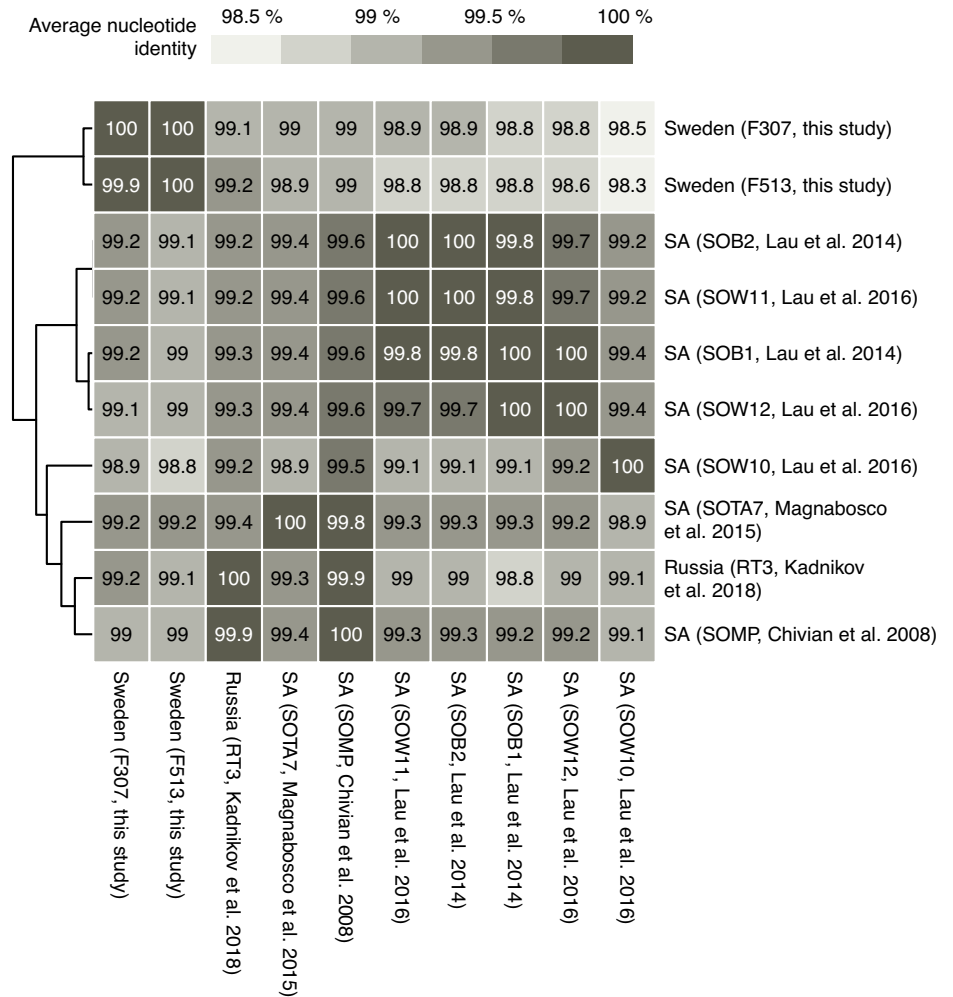
The genome-resolved metagenomics revealed a predominantly bacterial community with a microbial abundance typical for low-energy, deep subsurface groundwaters and mainly consisting of Bacillota (88% of the metagenomic read counts). The Bacillota were largely represented by the sulfate-reducer *Ca. Desulforudis audaxviator* and contained genes enabling survival under anaerobic, low-energy conditions, such as the carbon monoxide dehydrogenase as part of the Wood-Ljungdahl pathway for the fixation of inorganic carbon. *Ca. Desulforudis audaxviator* being abundant in deep subsurface, anoxic groundwaters has been reported in several independent studies, implicating this taxon as well-adapted to low-energy conditions. Finally, the high diversity of the eukaryotic community, including presumably aerobic populations, suggested an unexplored role of prokaryotes in this subsurface ecosystem.

### Methods

#### Groundwater sampling

The COSC-2 borehole (Lat 63.3124° N, Lon 13.5265° E) has a depth of 2276 m from the surface (elevation 320 m) and was drilled in clastic sedimentary rocks and the crystalline basement of the Fennoscandian Shield. A groundwater-bearing fracture within a lower Paleozoic conglomerate rock section at 975 m depth was isolated from other fractures using inflatable rubber and stainless-steel packers (Lapela Technology Oy, Rauma, Finland) positioned below and above the fracture (Fig. 1). The temperature in the borehole at this depth was 25 °C (data from Lorenz et al.<sup>54</sup>). The packers (76 mm in diameter) were inflated against the borehole wall (96 mm in diameter) using a 4 mm polyamide tube (Toppi Oy, Espoo, Finland) that was pressurized using a manual pressure pump filled with tap water. These packers were operated with a custom-made winch trailer (Lapela Technology Oy) and secured with a fiber rope 5 mm in diameter. Samples for the hydrochemistry were taken both from pumped fluid (pH, alkalinity, salinity, sulfur isotope) and from a custom-made in situ fluid sampling device (Lapela Technology Oy) placed below the upper packer. Electrical conductivity and pH were measured in the field using a portable device (WTW). Alkalinity was determined the same day by end-point titration to pH 4.5

**Fig. 5 | Genome comparison of *Ca. Desulforudis audaxviator*.** The reconstructed genomes ( $n = 10$ ) were compared based on average nucleotide identity. Rows were clustered according to Euclidean distances. Included genomes originated from Chivian et al. (2008)<sup>15</sup>, Lau et al. (2014)<sup>53</sup>, Magnabosco et al. (2016)<sup>3</sup>, Lau et al. (2016)<sup>6</sup>, and Kadnikov et al. (2018)<sup>66</sup>. SA; South Africa.



with 0.16 M sulfuric acid using a digital titrator (Hach, model 16900). Gas samples were taken from the in situ sampling device and gas analyses were conducted at GFZ Potsdam using a Quadrupole Mass Spectrometer (QMS, OmniStar™ GSD 300, Pfeiffer Vacuum, Germany). Elemental analyses were done with inductively coupled plasma mass spectrometry (Neptune MC-ICP-MS; ThermoFisher). Anions (Br, Cl, F, SO<sub>4</sub>, and NO<sub>3</sub>) were analyzed with ion chromatography. All geochemical laboratory analyses were conducted by an external laboratory (Eurofins Environment Testing Finland Oy, Lahti, Finland). Sulfur isotopes (<sup>32</sup>S and <sup>34</sup>S) were analyzed on a high-resolution multicollector (MC-ICP-MS; Nu Instruments). Ion exchange was performed with a cation exchange resin (AG 50W-X8, 200 to 400 μm mesh). Sulfur isotope analyses were conducted by the Geological Survey of Finland. Prior to sampling, the sampling connections, including ca. 1000 m of polyamide tubing 10 mm in diameter, were flushed overnight with three section volumes (3 × 314 L) to avoid contamination from stagnant water. Pumped water samples for anions and alkalinity (2 × 250 mL) were unfiltered and subsampled (60 mL) for the sulfur isotope analysis. This subsample was filtered with a 0.45 μm pore size. Pumped water samples for cations (100 mL) were filtered (0.45 μm pore size) and acidified with 0.5 mL of 65% ultrapure HNO<sub>3</sub>. Cells were captured under a flow rate of 1.5–2.1 L h<sup>-1</sup> using a high-pressure filter holder (Millipore) containing a sterile polyvinylidene fluoride membrane with a 0.1 μm pore size (Merck Durapore, diameter 47 mm). The membranes ( $n = 7$ ) were aseptically collected in extraction tubes and immediately frozen in a mobile -80 °C freezer that was also used to transport the samples to the laboratory where the samples were stored at -80 °C. The side of the filter containing the cells (facing the water source) faced inwards in the collection tube to maximize

yield during DNA extraction. Samples for flow cytometry (10 mL,  $n = 2$ ) were fixed in 2.2% (vol/vol) formaldehyde on site and stored at 4 °C.

### DNA extraction and amplification

Extraction was performed in a room dedicated to the extraction of nucleic acids while working in a fume hood that was cleaned with both bleach and 70% ethanol prior to use. All plastic consumables (tips, centrifuge tubes, multiwell plates) were PCR grade and UV treated for 10 min before placing the material in the fume hood. DNA was extracted using the DNeasy PowerWater kit (Qiagen), following the manufacturer's protocol apart from eluting the nucleic acids in 50 μL instead of 100 μL of the included elution buffer. For every extraction, two filters were pooled to increase the DNA yield, except for the control sample. The V3-V4 region of the 16S rRNA gene was amplified using the primer pair 341 F and 805R<sup>55</sup> while the 18S rRNA gene (V4-V5 region) was amplified with the 454 F 981 R primer pair<sup>56</sup>. The product of the first PCR served as template for a second amplification step containing the sample-specific sequencing barcodes. In total, the PCR programs contained 32 and 37 cycles for the 16S and 18S assays, respectively. The amplified product was purified after each PCR using the AMPure XP reagent (Beckman Coulter). To identify potential contaminants introduced during field sampling or molecular work, a blank filter was collected at the sampling site and processed simultaneously with the other samples. Prior to equimolar pooling, the concentration and fragment length distribution of the amplified products were measured using a Qubit 2.0 fluorometer (Life Technologies) and automated gel electrophoresis (Agilent TapeStation 4150), respectively. Sequencing was done

at the Science for Life Laboratory, Sweden on an Illumina MiSeq (v3-600), producing  $2 \times 301$  bp paired-end reads.

Libraries for metagenomic sequencing were prepared using the Tecan MagicPrep while adjusting the cycling conditions (10 to 12 cycles) to the amount of template DNA added (10 to 20 ng) according to the manufacturer's instructions. The concentration of the libraries was measured using a Qubit 3 fluorometer (Life Technologies). After equimolar pooling, the fragment length distribution of the pooled libraries was checked using agarose gel electrophoresis. The metagenomes were sequenced at Science for Life Laboratory (SNP&SEQ platform) on an Illumina NovaSeq platform equipped with a SP flowcell, producing  $2 \times 150$  bp paired-end reads.

### Real-time PCR and flow cytometry

Microbial abundance was assessed according to 16S rRNA gene copies using a real-time PCR (qPCR) on a LightCycler 480 (Roche Diagnostics). The reaction volume was 10  $\mu$ L and consisted of 5  $\mu$ L Platinum SYBR Green qPCR SuperMix-UDG with ROX (Thermo Fisher Scientific), 0.4  $\mu$ L 10  $\mu$ M primer (Eurofins), 3.2  $\mu$ L nuclease-free water (Thermo Scientific), and 1.0  $\mu$ L template. Bacterial and archaeal gene fragments were amplified using the primers 908F\_mod/1075R<sup>57</sup> and 915 F/1059R<sup>58</sup>, respectively. Cycling conditions were 2 min at 95 °C, 40 cycles of 15 s at 95 °C and 30 s at 60 °C, followed by a melt curve analysis to assess primer specificity. Standard curves were generated with a dilution series of purified PCR product using genomic DNA of pure cultures as template (i.e., *Acidiphilium cryptum* JF-5 for bacteria and *Ferroplasma acidiphilum* BRGM4 for archaea). Standards ( $n = 7$ ), samples of interest ( $n = 3$ ), and no-template controls ( $n = 1$ ) were run in triplicates and the former two were 1:10 and 1:100 diluted in nuclease-free water to account for inhibition of the polymerase. The gene copy numbers were reported in gene fragments  $\text{mL}^{-1}$  after correcting for groundwater volume filtered for DNA extraction and the elution volume. Reaction efficiency for the bacterial and archaeal assays were 89.9% and 87.5%, respectively. A minimum number of three quantification cycles ( $\Delta\text{Cq}$ ) between samples and no-template controls was maintained as a limit of detection.

In addition to qPCR, microbial abundance was also assessed using a flow cytometer (Cytotflex, Beckman Coulter) equipped with two lasers (blue and violet) and a microwell plate autosampler. The fixed samples (2.2% formaldehyde,  $n = 2$ ) were stained using SYBR Green I that binds to DNA, thereby visualizing both heterotrophic and autotrophic populations. Sterile-filtered ultrapure water (Milli-Q, pore size 0.2  $\mu\text{m}$ ) was used as a reference to distinguish cells of interest from any background signal.

### Bioinformatics

Raw sequencing reads from the 16S rRNA gene amplicon ( $n = 4$ ) were processed using the ampliseq pipeline<sup>59</sup> (v2.3.0) from the nf-core framework that relied on Nextflow (v21.10.6), Cutadapt (v3.4), FastQC (v0.11.9), DADA2 (v1.22.0), and the SBDI Sativa curated 16S GTDB database<sup>60</sup> (release 207). The ampliseq pipeline was run with default settings, except for the trimming of the primers whereby reads not containing the primer sequence or containing double copies were discarded from downstream analysis. The raw reads from the 18S rRNA gene amplicon ( $n = 3$ ) were processed identically to the 16S amplicons, except from using the Protist Ribosomal Reference database<sup>43</sup> (v4.14.0) for taxonomy assignment.

Raw sequences from the metagenomes ( $n = 3$ ) were assembled, binned, and annotated on default settings if not specified using the curated mag pipeline<sup>61</sup> (v2.2.1) from the nf-core framework<sup>62</sup>. Adapter plus quality trimming was done with fastp (v0.23.2), reads were co-assembled by grouping the three metagenomes using MEGAHIT (v1.2.9), followed by evaluation of the contigs with Quast (v5.0.2). Protein-coding genes were predicted using Prodigal (v2.6.3), open reading frames were functionally annotated with Prokka (v1.14.6), and the metagenomes were binned with MetaBAT2 (v2.15). The bins were refined with DAS Tool (v1.1.4) after which the quality of the binned genomes was evaluated with CheckM (v1.1.3). dRep (v3.4.2) was used for de-replication, setting the maximum contamination to 5% and the minimum completeness to 70%, followed by

taxonomic assignment with GTDB-Tk (v2.1.1). The read coverage of the bins was normalized to transcripts per million (tpm) after mapping the quality-trimmed reads from the three metagenomes (the output of fastp) to the de-replicated bins with CoverM (v0.6.1). The de-replicated genomes were additionally functionally annotated with eggNOG-mapper (v2.1.9) and the obtained KEGG orthologs were used to identify the completeness of KEGG modules (KEGG mapper) by scoring them as complete, one block (gene) missing, incomplete, or absent. The presence of genes involved in viral defense was checked using the CRISPRCasFinder<sup>63</sup>, setting the minimum evidence level to 3 and requiring both CRISPR and CRISPR-associated genes to be present.

Marker genes included for constructing the heat map on genomic potential were sulfate adenylyltransferase (*sat*; EC:2.7.7.4), adenylylsulfate reductase (*aprAB*; EC:1.8.99.2), anaerobic sulfite reductase (*asrABC*; K16950, K16951, K00385), L-cysteine S-thiosulfotransferase (*soxABXYZ*; EC:2.8.5.2), sulfite dehydrogenase (*sorAB*; EC:1.8.2.1), nitrogenase (*nifDHK*; EC:1.18.6.1), nitrite reductase (*nrfAH*; EC:1.7.2.2), anaerobic carbon monoxide dehydrogenase (*cooS*; EC:1.2.7.4), CO methylating acetyl-CoA synthase (EC:2.3.1.169), ribulose-bisphosphate carboxylase (rubisco; EC:4.1.1.39), 4-hydroxybutanoyl-CoA dehydratase (*acsA*; EC:4.2.1.120), propionyl-CoA carboxylase (*pccAB*; EC:6.4.1.3), hydrogen dehydrogenase (*hoxHY*; EC:1.12.1.2), NiFe hydrogenase (*hoxGK*; EC:1.12.99.6, *hydAB*; EC:1.12.2.1), and cytochrome *c* oxidase (*ctaCDEF*, EC:1.9.3.1). The functional annotation from both Prokka and eggNOG-mapper were queried based on the EC reference or gene name and the output was manually curated to verify if the gene of interest was obtained. The de-replicated reconstructed genomes combined with the quality-trimmed reads (the output of fastp) were analyzed using METABOLIC<sup>64</sup> (v4), thereby producing the metabolic weight scores for each reconstructed genome. Finally, the genomes affiliated with *Ca. Desulfurudis audaaxviator* were compared with reconstructed genomes reported in literature based on average nucleotide identity using FastANI<sup>65</sup> (v1.34).

### Statistics and reproducibility

Statistics and data visualization were performed in R (v4.2.1). For the amplicon data, absolute counts were standardized within a sample to relative abundances by dividing an ASVs count by the total number of counts within a sample.

### Reporting summary

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

### Data availability

All sequencing data has been made publicly available at the European Nucleotide Archive under project reference PRJEB68186. A compiled version of the R Markdown document, together with the number of sequence reads (16S and 18S rRNA gene) throughout the bioinformatic pipeline, are provided on GitHub at <https://doi.org/10.5281/zenodo.13882042>.

Received: 8 November 2023; Accepted: 7 October 2024;

Published online: 15 October 2024

### References

- Gold, T. The deep, hot biosphere. *Proc. Natl Acad. Sci. USA* **89**, 6045–6049 (1992).
- McMahon, S. & Parnell, J. Weighing the deep continental biosphere. *FEMS Microbiol. Ecol.* **87**, 113–120 (2014).
- Magnabosco, C. et al. A metagenomic window into carbon metabolism at 3 km depth in Precambrian continental crust. *ISME J.* **10**, 730–741 (2016).
- Lollar, G. S., Warr, O., Telling, J., Osburn, M. R. & Lollar, B. S. Follow the water<sup>1</sup>: hydrogeochemical constraints on microbial investigations 2.4 km below surface at the Kidd Creek deep fluid and deep life observatory. *Geomicrobiol. J.* **36**, 859–872 (2019).



5. Momper, L., Jungbluth, S. P., Lee, M. D. & Amend, J. P. Energy and carbon metabolisms in a deep terrestrial subsurface fluid microbial community. *ISME J.* **11**, 2319–2333 (2017).
6. Lau, M. C. Y. et al. An oligotrophic deep-subsurface community dependent on syntrophy is dominated by sulfur-driven autotrophic denitrifiers. *Proc. Natl Acad. Sci. USA* **113**, E7927–E7936 (2016).
7. Suzuki, Y. et al. Biogeochemical signals from deep microbial life in terrestrial crust. *PLoS ONE* **9**, e113063 (2014).
8. Nuppenen-Puputti, M. et al. Rock surface fungi in deep continental biosphere—exploration of microbial community formation with subsurface in situ biofilm trap. *Microorganisms* **9**, 64 (2020).
9. Becraft, E. D. et al. Evolutionary stasis of a deep subsurface microbial lineage. *ISME J.* **15**, 2830–2842 (2021).
10. Ruff, S. E. et al. Hydrogen and dark oxygen drive microbial productivity in diverse groundwater ecosystems. *Nat. Commun.* **14**, 3194 (2023).
11. Nuppenen-Puputti, M. et al. Epilithic microbial community functionality in deep oligotrophic continental bedrock. *Front. Microbiol.* **13**, 826048 (2022).
12. Bomberg, M., Lamminmäki, T. & Itävaara, M. Microbial communities and their predicted metabolic characteristics in deep fracture groundwaters of the crystalline bedrock at Olkiluoto, Finland. *Biogeosciences* **13**, 6031–6047 (2016).
13. Wu, X. et al. Microbial metagenomes from three aquifers in the Fennoscandian Shield terrestrial deep biosphere reveal metabolic partitioning among populations. *ISME J.* **10**, 1192–1203 (2016).
14. Casar, C. P., Momper, L. M., Kruger, B. R. & Osburn, M. R. Iron-fueled life in the continental subsurface: deep Mine Microbial Observatory, South Dakota, USA. *Appl. Environ. Microbiol.* **87**, e0083221 (2021).
15. Chivian, D. et al. Environmental genomics reveals a single-species ecosystem deep within Earth. *Science* **322**, 275–278 (2008).
16. Chapelle, F. H. et al. A hydrogen-based subsurface microbial community dominated by methanogens. *Nature* **415**, 312–315 (2002).
17. Bell, E. et al. Active sulfur cycling in the terrestrial deep subsurface. *ISME J.* **14**, 1260–1272 (2020).
18. Borgonie, G. et al. Eukaryotic opportunists dominate the deep-subsurface biosphere in South Africa. *Nat. Commun.* **6**, 8952 (2015).
19. Pedersen, K. Exploration of deep intraterrestrial microbial life: current perspectives. *FEMS Microbiol. Lett.* **185**, 9–16 (2000).
20. Anderson, R. T., Chapelle, F. H. & Lovley, D. R. Evidence against hydrogen-based microbial ecosystems in basalt aquifers. *Science* **281**, 976–977 (1998).
21. Kadnikov, V. V., Mardanov, A. V., Beletsky, A. V., Karnachuk, O. V. & Ravin, N. V. Microbial life in the deep subsurface aquifer illuminated by metagenomics. *Front. Microbiol.* **11**, 572252 (2020).
22. Pedersen, K. Microbial life in deep granitic rock. *FEMS Microbiol. Rev.* **20**, 399–414 (1997).
23. Osterholz, H. et al. Terrigenous dissolved organic matter persists in the energy-limited deep groundwaters of the Fennoscandian Shield. *Nat. Commun.* **13**, 4837 (2022).
24. Berg, I. A. et al. Ecological aspects of the distribution of different autotrophic CO<sub>2</sub> fixation pathways. *Appl. Environ. Microbiol.* **77**, 1925–1936 (2011).
25. Ivarsson, M., Bengtson, S. & Neubeck, A. The igneous oceanic crust - Earth's largest fungal habitat? *Fungal Ecol.* **20**, 249–255 (2016).
26. Drake, H. et al. Anaerobic consortia of fungi and sulfate reducing bacteria in deep granite fractures. *Nat. Commun.* **8**, 55 (2017).
27. Drake, H. et al. Fossilized anaerobic and possibly methanogenesis-fueling fungi identified deep within the Siljan impact structure, Sweden. *Commun. Earth Environ.* **2**, 34 (2021).
28. Dahl-Jensen, T. et al. Evidence for early Proterozoic plate tectonics from seismic reflection profiles in the Baltic shield. *Nature* **348**, 34–38 (1990).
29. Itävaara, M. et al. Characterization of bacterial diversity to a depth of 1500m in the Outokumpu deep borehole, Fennoscandian Shield. *FEMS Microbiol. Ecol.* **77**, 295–309 (2011).
30. Lopez-Fernandez, M., Åström, M., Bertilsson, S. & Dopson, M. Depth and dissolved organic carbon shape microbial communities in surface influenced but not ancient saline terrestrial aquifers. *Front. Microbiol.* **9**, 2880 (2018).
31. Westmeijer, G. et al. Connectivity of Fennoscandian Shield terrestrial deep biosphere microbiomes with surface communities. *Commun. Biol.* **5**, 37 (2022).
32. Hubalek, V. et al. Connectivity to the surface determines diversity patterns in subsurface aquifers of the Fennoscandian shield. *ISME J.* **10**, 2447–2458 (2016).
33. Drake, H. & Reiners, P. W. Thermochronologic perspectives on the deep-time evolution of the deep biosphere. *Proc. Natl Acad. Sci. USA* **118**, e2109609118 (2021).
34. Karnachuk, O. V. et al. Domestication of previously uncultivated *Candidatus Desulforudis audaxviator* from a deep aquifer in Siberia sheds light on its physiology and evolution. *ISME J.* **13**, 1947–1959 (2019).
35. Lorenz, H. et al. COSC-2 - Drilling the basal décollement and underlying margin of palaeocontinent Baltica in the Paleozoic Caledonide Orogen of Scandinavia. *Sci. Drill.* **30**, 43–57 (2022).
36. Westmeijer, G. et al. Continental scientific drilling and microbiology: (extremely) low biomass in crystalline bedrock of central Sweden. *Biogeosci. Discuss.* **Sep**, 1–22 (2023).
37. Purkamo, L. et al. Dissecting the deep biosphere: retrieving authentic microbial communities from packer-isolated deep crystalline bedrock fracture zones. *FEMS Microbiol. Ecol.* **85**, 324–337 (2013).
38. Gimeno, M. J., Tullborg, E.-L., Nilsson, A.-C., Auqué, L. F. & Nilsson, L. Hydrogeochemical characterisation of the groundwater in the crystalline basement of Forsmark, the selected area for the geological nuclear repositories in Sweden. *J. Hydrol.* **624**, 129818 (2023).
39. Starnawski, P. et al. Microbial community assembly and evolution in subseafloor sediment. *Proc. Natl Acad. Sci. USA* **114**, 2940–2945 (2017).
40. Musto, H. et al. Genomic GC level, optimal growth temperature, and genome size in prokaryotes. *Biochem. Biophys. Res. Commun.* **347**, 1–3 (2006).
41. Lin, L.-H. et al. Long-term sustainability of a high-energy, low-diversity crustal biome. *Science* **314**, 479–482 (2006).
42. Saitoh, Y., Hirano, S., Nagaoka, T. & Amano, Y. Genetic survey of indigenous microbial eukaryotic communities, mainly fungi, in sedimentary rock matrices of deep terrestrial subsurface. *Ecol. Genet. Genom.* **12**, 100042 (2019).
43. Guillou, L. et al. The Protist Ribosomal Reference database (PR<sup>2</sup>): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.* **41**, D597–D604 (2013).
44. O'Leary, N. A. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).
45. Perkins, A. K. et al. Highly diverse fungal communities in carbon-rich aquifers of two contrasting lakes in Northeast Germany. *Fungal Ecol.* **41**, 116–125 (2019).
46. Jones, M. D. M. et al. Discovery of novel intermediate forms redefines the fungal tree of life. *Nature* **474**, 200–203 (2011).
47. Gleason, F. H., Carney, L. T., Lilje, O. & Glockling, S. L. Ecological potentials of species of *Rozella* (Cryptomycota). *Fungal Ecol.* **5**, 651–656 (2012).
48. Berg, I. A. et al. Autotrophic carbon fixation in archaea. *Nat. Rev. Microbiol.* **8**, 447–460 (2010).
49. Rubin-Blum, M., Dubilier, N. & Kleiner, M. 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**, e00394–18 (2019).
50. Eisenhofer, R., Odriozola, I. & Alberdi, A. Impact of microbial genome completeness on metagenomic functional inference. *ISME Commun.* **3**, 12 (2023).

51. Willems, A. & Gillis, M. *Hydrogenophaga*. in *Bergey's Manual of Systematics of Archaea and Bacteria* 1–15 (Wiley, 2015). <https://doi.org/10.1002/9781118960608.gbm00947>.
  52. Strauss, H. The isotopic composition of sedimentary sulfur through time. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **132**, 97–118 (1997).
  53. Lau, M. C. Y. et al. Phylogeny and phylogeography of functional genes shared among seven terrestrial subsurface metagenomes reveal N-cycling and microbial evolutionary relationships. *Front. Microbiol.* **5**, 531 (2014).
  54. Lorenz, H. et al. COSC-2 operational report - Operational data sets. *GFZ Data Services* <https://doi.org/10.5880/ICDP.5054.003> (2021).
  55. Herlemann, D. P. R. et al. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J.* **5**, 1571–1579 (2011).
  56. Stoeck, T. et al. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol. Ecol.* **19**, 21–31 (2010).
  57. Ohkuma, M. & Kudo, T. Phylogenetic analysis of the symbiotic intestinal microflora of the termite *Cryptotermes domesticus*. *FEMS Microbiol Lett.* **164**, 389–395 (1998).
  58. Yu, Y., Lee, C. & Hwang, S. Analysis of community structures in anaerobic processes using a quantitative real-time PCR method. *Water Sci. Technol.* **52**, 85–91 (2005).
  59. Straub, D. et al. Interpretations of environmental microbial community studies are biased by the selected 16S rRNA (gene) amplicon sequencing pipeline. *Front. Microbiol.* **11**, 550420 (2020).
  60. Lundin, D. & Andersson, A. SBDI Sativa curated 16S GTDB database. *SciLifeLab. Dataset.* <https://doi.org/10.17044/scilifelab.14869077.v4> (2021).
  61. Krakau, S., Straub, D., Gourel, H., Gabernet, G. & Nahnsen, S. nf-core/mag: a best-practice pipeline for metagenome hybrid assembly and binning. *NAR Genom. Bioinform.* **4**, lqac007 (2022).
  62. Ewels, P. A. et al. The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* **38**, 276–278 (2020).
  63. Couvin, D. et al. CRISPRCasFinder, an update of CRISPRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* **46**, W246–W251 (2018).
  64. Zhou, Z. et al. METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. *Microbiome* **10**, 33 (2022).
  65. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **9**, 5114 (2018).
  66. Kadnikov, V. V. et al. A metagenomic window into the 2-km-deep terrestrial subsurface aquifer revealed multiple pathways of organic matter decomposition. *FEMS Microbiol. Ecol.* **94**, fiy152 (2018).
- Foundation (to HD). SB acknowledges financial support from the Swedish Research Council and Science for Life Laboratory. High-throughput sequencing was carried out at the National Genomics Infrastructure hosted by the Science for Life Laboratory. Bioinformatics analyses were carried out utilizing the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) at Uppsala University (projects NAISS 2023/22-893 & 2023/6-261). The computations were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX partially funded by the Swedish Research Council through grant agreement no. 2016-07213. Funding for open-access publishing provided by Linnaeus University.

### Author contributions

H.D., M.D., S.B., and G.W. designed the research; H.D., F.D., G.W., and R.K. performed field work; R.K. performed hydrochemistry measurements; G.W. and F.D. performed the molecular work; G.W. and C.G.R. processed and performed ecological interpretations of the sequencing data; G.W. and M.D. wrote the manuscript with comments from all authors. M.D. and H.D. provided funding. All authors read and improved the final version.

### Funding

Open access funding provided by Linnaeus University.

### 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/s42003-024-07027-2>.

**Correspondence** and requests for materials should be addressed to George Westmeijer.

**Peer review information** *Communications Biology* thanks Nikolai V. Ravin, Magdalena R. Osburn and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Tobias Goris. A peer review file is available.

**Reprints and permissions 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.

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

© The Author(s) 2024

### Acknowledgements

The authors acknowledge Henning Lorenz from Uppsala University and Arto Pullinen from the Geological Survey of Finland for their support during the field sampling, Emelie Nilsson for her advice and support for the bioinformatic processing, and Maliheh Mehrshad for running the flow cytometry. Tabo V. Jain is acknowledged for his help in preparing the 18S sequencing library. Matthias Hoetzing and Prune Leroy are acknowledged for their help in preparing the metagenomic sequencing libraries. The study was supported by The Swedish Research Council (Vetenskapsrådet contracts 2018-04311 to MD plus 2017-05186 and 2021-04365 to HD), FORMAS (2020-01577 to HD), Olle Engkvist's Foundation (contract 216-0462 to MD), the Crafoord foundation (contract 20210524 to HD), J. Richt.