

<https://doi.org/10.1038/s42003-026-09706-8>

Exploring microbial diversity using cell-size fractionated enrichment incubations from subsurface aquifers at Äspö, Sweden

Check for updates

George Westmeijer^{1,3}, Stephanie Turner¹, Patrik Hevele¹, Maliheh Mehrshad², Stefan Bertilsson² & Mark Dopson¹✉

The continental subsurface hosts energy-constrained groundwaters with a high diversity of ecologically elusive microorganisms adapted to the prevailing low-energy conditions. This study explored potential interactions among microbes using anaerobic enrichment incubations with three types of groundwater of contrasting hydrochemistry from the Äspö Hard Rock Laboratory, Sweden. Removing cells larger than 0.45 μm from the inoculum resulted in incubations enriched in populations characterized by very small genomes, including Patescibacteria, Nanobdellota, and Omnitrophota. These incubations had a higher diversity than non-fractionated incubations. However, cell numbers and community structure of the fractionated incubations did not change over an incubation period up to four months, despite high microbial diversity and experimental amendments with either simple (acetate) or more complex (cell lysate) carbon sources. In addition, network analysis on the groundwaters revealed multiple co-occurrences between populations affiliated with the Patescibacteria and the Desulfobacterota. Overall, these findings support that a considerable part of microbial diversity has a small cell size in these low energy groundwaters and strong co-occurrences among populations as an important survival strategy.

The continental deep subsurface, here defined as the bedrock below the soil horizon, contains a considerable proportion of the Earth's total biomass and hosts groundwaters that make up one of Earth's major groundwater reservoirs¹. Research on the deep subsurface is essential for understanding how microorganisms maintain biogeochemical processes while faced with resource limitation as a consequence of isolation from solar energy inputs²⁻⁴. The diversity of these subsurface microorganisms and their metabolism is strongly influenced by the degree of isolation from the surface, host rock lithology, and the availability of electron donors, such as organic carbon or hydrogen gas^{5,6}. Deeper, more ancient groundwaters may be isolated from the surface for up to millions of years and typically accommodate chemolithotrophs that frequently have the ability to fix carbon dioxide and utilize hydrogen as an energy source^{7,8}. In contrast, groundwaters characterized by a lower degree of isolation and a lower hydrological retention time are heavily influenced by the infiltration of surface-derived organic carbon, thereby sustaining heterotrophic populations⁹⁻¹¹.

Patescibacteria (synonym candidate phyla radiation or CPR clade), Omnitrophota, and the archaeal DPANN clade are frequently detected in

suboxic and anoxic groundwaters, possibly due to mobilization from soils and their adaptation to low energy conditions¹²⁻¹⁴. These bacteria and archaea are characterized as having a cell size as small as 0.1 μm , small genomes with limited metabolic capacity, and are often described as epibionts or parasites^{12,15}. Recent studies on these lineages provide insights into their host-associated lifestyle, such as the epibiosis of *Southlalkia epibionticum* on *Actinomyces israelii*¹⁶, *Candidatus* Yanofskybacteriaceae and *Ca. Minisyncoccaceae* parasitizing on methanogenic archaea¹⁷, *Ca. Absconditicoccus praedator* parasitizing on *Halorhodospira halophila*¹⁸, and the symbiotic interaction between *Ca. Micrarchaeota* (ARM-1) and *Metallosphaera* sp. AS-7¹⁹. The abundance and diversity of these ultra-small bacteria and archaea have likely been underestimated, especially in early genomic studies, due to their capacity to pass through a 0.2 μm pore that is typically used for cell capture^{20,21}. These types of bacteria and archaea have previously been identified as abundant in subsurface groundwaters and to be involved in biogeochemical cycling²².

The Äspö Hard Rock Laboratory (HRL), Sweden is an underground research facility hosted in Paleoproterozoic granitoids²³ that consists of

¹Centre for Ecology and Evolution in Microbial model Systems (EEMiS), Linnaeus University, Stuvaregatan 4, Kalmar, Sweden. ²Department of Aquatic Sciences and Assessment, Science for Life Laboratory, Swedish University of Agricultural Sciences, Uppsala, Sweden. ³Present address: Department of Chemistry, Umeå University, Umeå, Sweden. ✉e-mail: mark.dopson@lnu.se

boreholes intersecting anoxic groundwater-filled fractures between 70 and 460 m below sea level. Although some groundwaters are influenced by infiltrating water from the overlying land and the Baltic Sea, others are thought to be long-term isolated from surface waters, with lower recharge rates coupled with higher residence times^{9,23–26}. Microbial diversity and the concentration of dissolved organic carbon (DOC) tend to decrease with depth, and the detected organic carbon is typically refractory in nature^{11,24,27}. The composition of the microbial communities residing in these groundwaters has been well described^{11,26} and genome-resolved metagenomics has revealed a substantial diversity of populations affiliated with the Patescibacteria, Omnitrophota, and the DPANN clade. These phyla tend to co-occur and are mainly detected in groundwaters with higher recharge rates^{28,29}. However, it is unclear how populations affiliated with these phyla meet their energy and carbon demands in these groundwaters, how they respond to allochthonous organic carbon, as well as the identity of their metabolic partners or potential host(s).

In this study, anaerobic incubations were used to enrich for uncharacterized clades and investigate co-occurrences and potential host-symbiont interactions. Studying these co-occurrences provides insights into possible metabolic dependencies and how co-existing populations fulfill their carbon demands. To do so, three subsurface aquifers intersected by boreholes within the Äspö HRL, ranging in a depth from 70 to 460 m below sea level, were sampled under anoxic conditions and enriched using either a medium containing a single carbon substrate (acetate) or a more complex organic carbon mixture (cell lysate). The former medium was adapted to the genomic potential of uncharacterized clades detected in the groundwater types under scrutiny, while the latter was used to represent necrotrophy that is suggested to be an important strategy for obtaining energy and nutrients in these aquifers³⁰. Furthermore, larger cells (diameter > 0.45 µm) were removed from the inoculum during size fractionation to enrich for ultra-small bacteria and archaea (e.g., Patescibacteria) while adding acetate or lysed cells. The primary research question was to explore how co-occurring populations responded to allochthonous organic carbon. We hypothesized that the populations in the fractionated incubations would especially benefit from the cell lysate as it contained a more diverse mixture of organic carbon compounds. The microbial communities in the incubations were first characterized using 16S rRNA gene sequencing followed by metagenomic sequencing of incubations selected based on their taxonomic composition.

Results and discussion

Characterization of the groundwaters used as inoculum

The planktonic microbial communities in the meteoric (KR0015B), marine (SA1420A-1), and saline (SA2600A-1) groundwaters intersected by boreholes in the Äspö HRL were characterized using 16S rRNA gene amplicon sequencing (Fig. 1). The extraction control for the groundwater sampling contained 376 amplicon sequencing variants (ASVs; Supplemental Fig. S3) that were removed from these samples prior to the analysis (Supplemental Table S1). In total, 7030 ASVs were detected in the three groundwater types ($n = 18$), of which the majority were affiliated with Patescibacteria (1883), Chloroflexota (1124), and Desulfobacterota (749). Rarefaction curves of the ASV number versus sequencing depth supported that the majority of the microbial diversity was captured in the sequence libraries (Supplemental Fig. S1).

Even though these groundwaters have been previously studied^{11,24}, they were again characterized as part of this study before (October 2019) and after (April 2020) the collection of groundwater for the experiments to verify if the microbial communities and hydrochemistry were stable over time. The most pronounced difference between these groundwaters was the variable DOC concentration and $\delta^{18}\text{O}$ (Fig. 1 & Table 1). The meteoric groundwater was recharged by overlying soil groundwater and had a relatively high DOC content of 1.28 mM compared to the marine and saline groundwaters (0.55 and 0.08 mM, respectively). This meteoric groundwater had a much lower electrical conductivity compared to the marine and saline groundwaters (264 versus 945 and 3420 mS m^{-1}), mainly due to the chloride concentrations of 18.7 versus 83.5 and 389 mM, respectively. The marine

groundwater had a $\delta^{18}\text{O}$ of -7.5‰ that is more similar to the value reported for the Baltic Sea (-6‰)²³, implying seawater infiltration. However, despite this potential recharge, the microbial communities were more similar to the communities in the soil groundwaters at Äspö island and do not resemble those in the Baltic Sea^{9,24}. The saline groundwater had a $\delta^{18}\text{O}$ of -12.2‰ , suggesting a high degree of isolation from the Baltic Sea and overlying groundwaters and was thus, isolated from surface-derived organic carbon. This isolation likely contributed to the saline groundwater having the lowest microbial diversity (Shannon's H index = 4.1, standard deviation = 1.6, $n = 6$) compared to the meteoric (H = 5.9, sd = 0.79, $n = 6$) and marine (H = 5.6, sd = 1.0, $n = 6$) groundwaters (Fig. 1). The contrasting hydrochemical properties of these groundwaters were also reflected in microbial abundances, with cell numbers (according to epifluorescence microscopy) in the meteoric groundwater being an order of magnitude higher than the marine (range $6.8\text{--}8.9 \times 10^5$ cells mL^{-1} ($n = 2$)) compared to $5.9\text{--}7.4 \times 10^4$ mL^{-1} ($n = 2$)) and saline groundwater with $1.4\text{--}2.3 \times 10^4$ mL^{-1} ($n = 2$). These cell counts were of the same order of magnitude as other studies on deep subsurface groundwaters^{6,31–33}.

Based on 16S rRNA gene amplicon sequencing, the communities in both the meteoric and marine groundwater types were characterized by a high abundance of the phyla Patescibacteria, Desulfobacterota, Chloroflexota, Omnitrophota, and Nitrospirata. Overall, the Patescibacteria was the most abundant phylum, especially in the meteoric and marine groundwaters. Classes ABY1, Paceibacteria, and Microgenomatia were the most abundant groups affiliated with this phylum, accounting for 16, 9, and 4%, respectively of the total sequence reads for the three groundwaters (Supplemental Fig. S2). These Patescibacteria potentially metabolize the terrigenous, refractory DOC present in these groundwaters to meet their carbon needs, as they may prefer more complex carbon compounds^{12,27,34}. Desulfobacterota was mainly represented by the typically heterotrophic orders Syntrophales and the Desulfobulbales. The Anaerolineales were the most abundant order within the Chloroflexota, and this group has been described as heterotrophic obligate anaerobes with most representatives having the capacity for both fermentation and oxidation of more complex carbon compounds^{35,36}. Omnitrophota was mainly represented by the orders Gygaellales and Duberdicusellales. While a comprehensive characterization of these specific orders is currently lacking, Omnitrophota typically have a small cell size (similar to the Patescibacteria) and have a diverse lifestyle including mixotrophy, fermentation, and acetogenesis^{14,37}. Nitrospirata predominantly contained sulfate-reducing clades affiliated with the Thermodesulfovibrionales that has been previously reported in the continental deep subsurface³⁸. The main difference in the communities of the meteoric and marine groundwater types was the presence of the Campylobacterota in the meteoric groundwater with the sulfur-oxidizing genus *Sulfuricurvum*³⁹ as its main representative, while this phylum was not detected in the marine groundwater that could be due to a lower concentration of reduced sulfur species (H_2S and HS^-) in the marine groundwater (Table 1). The saline groundwater type was characterized by a high abundance of Atribacterota and Acidobacteriota. The Atribacterota had only one genus represented (34–128), which is a lineage affiliated with the class JS1 suggested to mediate anaerobic hydrocarbon degradation⁴⁰. Based on the community composition according to the most abundant phyla in this study and previous studies^{11,24}, the microbial community seemed to be stable over a three-year period.

Initial enrichment incubations

The meteoric groundwater was used as an inoculum and the microbial abundance was quantified over a ten-week period using real-time PCR, targeting the 16S rRNA gene with both bacterial and archaeal primers. ASVs detected in the extraction control (358 ASVs) and in the acetate (141 and 183 ASVs) and lysate (233 and 220 ASVs) media controls were removed prior to the analyses. Certain ASVs from the controls affiliated with the Spirochaetota and the Acidobacteriota were abundant in the fractionated acetate incubations and the non-fractionated lysate incubations, respectively. Although this affected the community structure of

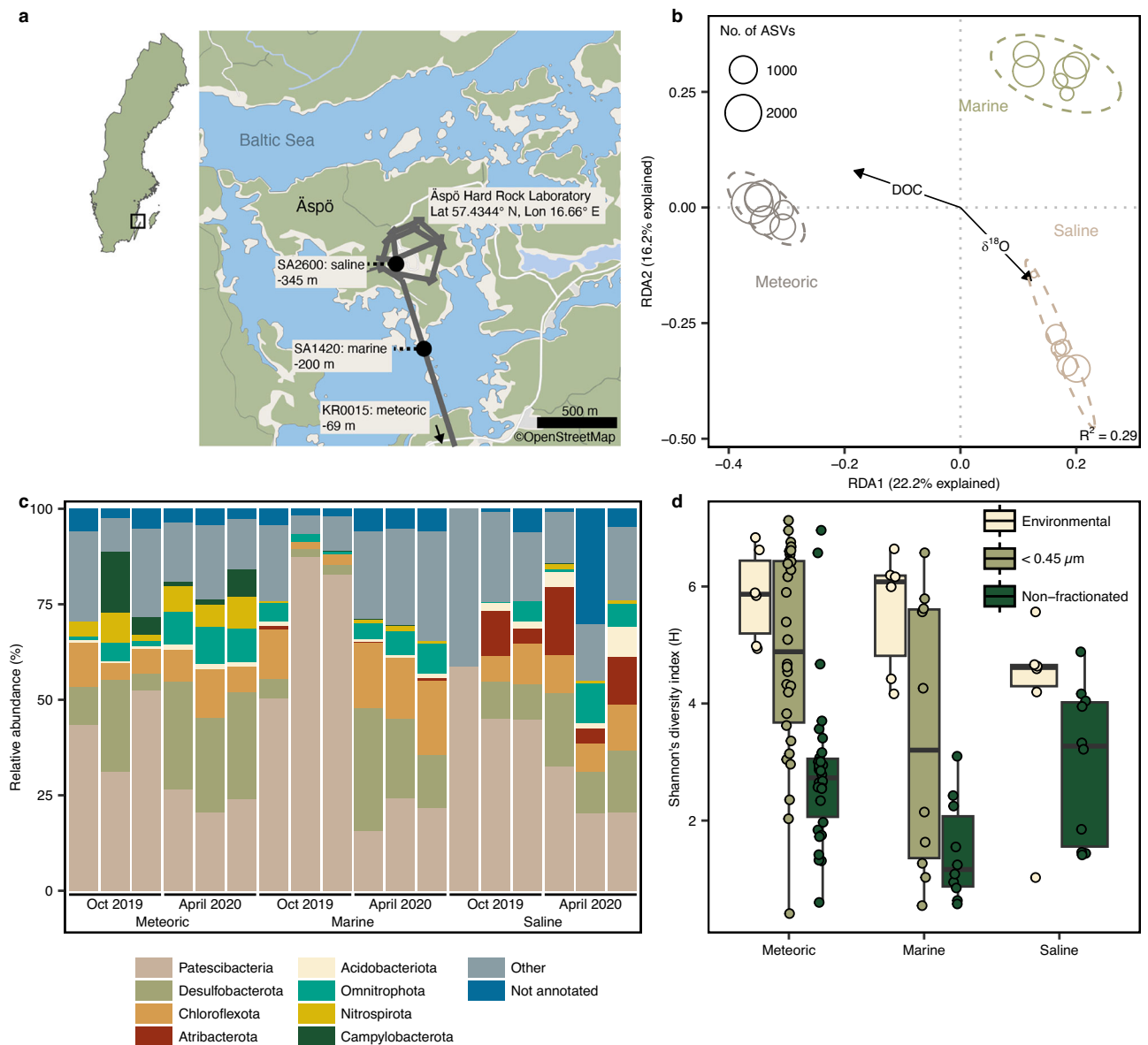


Fig. 1 | Characterization of the microbial community in the three groundwater types.

a Location of the Åspö Hard Rock Laboratory within Sweden. The three sampled aquifers are marked on the map with an approximation of the borehole's location. **b** Redundancy analysis (RDA) of the microbial communities combined with hydrochemistry showing only non-redundant variables (Pearson's $p < 0.9$) after including the values listed in Table 1. The ellipse depicts a 95% confidence interval for each groundwater type as a proxy for the variation in the ordination between the samples ($n = 6$). **c** Community composition on the level of phylum, including the eight most abundant phyla (based on summed relative abundance)

while grouping remaining phyla as "Other" and including ASVs without an assigned phylum as "Not annotated". The groundwaters were sampled in triplicates before (October 2019) and after (April 2020) groundwater collection for inoculation of the enrichment incubations. **d** Alpha diversity according to the Shannon H index for the environmental samples (the groundwater, $n = 18$), all size-fractionated incubations ($n = 40$), and all of the incubations without size fractionation ($n = 50$). The box represents the first and third quartiles with the whiskers extending to 1.5 times the inter-quartile range.

these incubations, this did not alter the overall interpretations of the results. In the non-fractionated incubations ($n = 20$), the copy number stabilized for both media types (acetate and lysate) around 10^6 and 10^7 gene copies mL^{-1} (Fig. 2). Most size fractionated incubations (i.e., cells $< 0.45 \mu\text{m}$; $n = 20$) did not grow during the incubation period, despite an initial 16S rRNA gene copy number of 10^4 and 10^3 gene copies mL^{-1} for bacteria and archaea, respectively. The only increase in copy number for the fractionated incubations was observed when Spirochetes (acetate) or Desulfobacterota (cell lysate) persisted after filtration ($0.45 \mu\text{m}$ pore size) of the inoculum (Fig. 3). Fractionated incubations had a significantly higher alpha diversity compared to the non-fractionated incubations, despite having received part of the inoculum

(Shannon's H 5.2 and 3.2, respectively; Mann-Whitney test p -value 6.9×10^{-5} , $n = 40$). Especially the fractionated incubations enriched with lysate at weeks 6, 7, and 9 had a high diversity (Shannon's H 6.7, 6.4, and 6.4, respectively) and this diversity was similar to that in the original groundwater (Shannon's H 5.9). This suggested a substantial part of the original diversity in the groundwater to be present in these fractionated incubations, despite incubation periods of six weeks or more.

Analysis of community differentiation (beta diversity) revealed that the incubation controls (containing sterile ultrapure water instead of inoculum, $n = 4$) were dissimilar to the incubations (Fig. 3, details in Supplemental Fig. S3). Moreover, the size-fractionated incubations ($n = 20$) often remained similar to the initial groundwater community, whereas non-

Table 1 | Characteristics groundwater used as inocula

Borehole	KR0015B	SA1420A-1	SA2600A-1
Origin ^a	Meteoric	Marine	Saline
Depth (m)	69.0	200.6	345.0
pH	7.6	7.5	7.5
T (°C)	10.0	11.4	13.7
δ ¹⁸ O (‰)	-10.5	-7.45	-12.2
DOC	1.28	0.549	0.0832
SO ₄ ²⁻	0.628	3.27	6.56
H ₂ S and HS ⁻	1.5 × 10 ⁻²	3.3 × 10 ⁻³	BD
δ ³⁴ S (‰)	27.4	31.1	12.4
Fe ²⁺	7.0 × 10 ⁻³	1.5 × 10 ⁻²	3.0 × 10 ⁻³
NO ₃ ⁻	BD	BD	BD
NH ₄ ⁺	0.0179	0.106	2.4 × 10 ⁻³
EC (mS m ⁻¹)	264	945	3420
Cl ⁻	18.7	83.5	389

Concentrations in mM. ^aGroundwater origin according to Mathurin et al.²⁵. Data from SKB's geochemical monitoring program, sampled (November 2019) as close as possible to microbiological sampling (October 2019 and April 2020). *BD* below detection. *DOC* dissolved organic carbon. *EC* electrical conductivity.

fractionated incubations ($n = 20$) markedly diverged over the ten-week period (Fig. 3). For example, in the acetate-enriched fractionated incubations from week five, members of the Breznakiellaceae (affiliated with the Spirochaetota), which are known anaerobic fermenters⁴¹, became the most abundant group. These spirochetes likely endured filtration (0.45 μm pore size) due to their spiral cell structure and possibly fermented the supplemented acetate as an energy source. In two fractionated incubations enriched with the cell lysate, *Humidusulfobacterium* sp. (phylum Desulfobacterota) emerged as the dominant community member. In contrast, nine fractionated incubations enriched with either acetate ($n = 3$) or lysate ($n = 6$) exhibited neither an increase in microbial abundance nor a change in community composition. These incubations were dominated by members of the phyla Patescibacteria (with Paceibacteria and ABY1 as the most abundant classes), Nanobdellota (previously classified as Nanoarchaeota and dominated by the order Pacearchaeales), and Omnitrophota (Koll11). Among these, the Patescibacteria displayed the highest diversity, comprising 20 classes and 97 orders. In comparison, the Nanobdellota contained only a single class and seven orders, underscoring that a substantial proportion of the microbial diversity in the fractionated incubations was affiliated with the Patescibacteria. Despite being size fractionated and thus not containing the complete inoculum, these incubations were characterized by a high diversity with an average of 2050 ASVs (sd 455, $n = 9$) in each culture. Except for the Omnitrophota, all these phyla were also represented in the metagenome-assembled genomes (MAGs, Fig. 4).

In summary, fractionated incubations had a substantially higher diversity than non-fractionated incubations and were enriched in Patescibacteria, Nanobdellota, and Omnitrophota. These fractionated incubations did not change (regarding gene copy number and diversity) over the incubation period of ten weeks, despite the amendment of a single organic carbon source (acetate) or a more complex carbon source (cell lysate). The stable high diversity of fractionated incubations enriched with phyla, such as Patescibacteria, combined with the constant gene copy numbers during the incubation period, suggests that these populations were potentially lacking a host or metabolic partner.

Incubations with meteoric, marine, and saline groundwater

After the initial experiment, all three groundwater types were used as inoculum and cell abundance was quantified over an incubation period of 17 weeks using epifluorescence microscopy. An approximate ten-fold increase of cell abundance in all of the non-fractionated incubations ($n = 6$) was

detected, with the meteoric incubations increasing from the initial 8×10^5 to approximately 10^7 cells mL⁻¹, the incubations with marine groundwater increasing from the initial 6.5×10^4 to 10^6 cells mL⁻¹, and the saline incubations increasing from the initial 2×10^4 to 10^5 cells mL⁻¹ after enriching.

Similar to the initial experiment, most size fractionated incubations had a largely stable community structure (based on 16S rRNA gene sequencing) and very little increase in cell numbers for the meteoric and marine groundwaters (Fig. 2 and Supplemental Fig. S4). Also for these incubations, ASVs detected in the extraction and media controls were removed prior to any analysis. A ten-fold increase in cell number was observed for a size fractionated marine culture, but this was due to the growth of a representative of the Bacillota (family Acholeplasmataceae, genus UBA2284) persisting after 0.45 μm filtration of the inoculum (Supplemental Fig. S5). This was similar to the initial experiment, whereby the only increase in cell numbers in fractionated incubations was observed for populations affiliated with either the Spirochaetota or the Desulfobacterota. Size fractionation of the saline groundwater failed as no cells were detected in the incubations, possibly due to a lower cell number in this groundwater combined with a slightly lower abundance of ultra-small cells in this groundwater (Fig. 1).

The alpha diversity was significantly higher (Mann-Whitney test p -value 0.01, $n = 50$) in the fractionated incubations (Shannon's H 3.8, sd = 2.2) compared to the non-fractionated incubations (H 2.2, sd = 1.1). A further comparison of all fractionated incubations (from both experiments, $n = 40$) with the non-fractionated incubations ($n = 40$) confirmed the former to have a higher diversity (Mann-Whitney test p -value 4.6×10^{-6}). Comparing the fractionated incubations with the meteoric and marine groundwater (used as inocula) indicated no significant differences in diversity (p -values 0.25 and 0.06, respectively). In fact, 25 fractionated incubations with meteoric and marine groundwater had a very similar diversity to the original groundwater (Fig. 1, range Shannon's H 4.3–7.1 versus 4.2–6.8 in the groundwater). The phyla Patescibacteria, Nanobdellota, and Omnitrophota were typically abundant in these high-diversity incubations. Overall, this relatively high diversity in the 0.1–0.45 μm fraction suggested that a substantial part of the microbial diversity was contained within this fraction for the meteoric and marine groundwater types.

In general, microbial abundance (both qPCR and epifluorescence microscopy) and 16S rRNA gene amplicon sequencing showed that the non-fractionated incubations were strongly enriched in either Bacillota, Spirochaetota, or Desulfobacterota. These non-fractionated incubations differentiated over time as the cell abundance increased (approximately ten-fold) and the diversity decreased, suggesting that populations affiliated with these phyla were capable of cell division (Supplemental Fig. S5). In contrast, approximately half of the fractionated incubations were characterized by a relatively high diversity similar to the groundwater used as inocula. Patescibacteria, Nanobdellota, and Omnitrophota were dominant in these incubations and responsible for up to 90% of the sequence reads, after being enriched with a 0.45 μm filtration of the inoculum. Despite the high diversity of populations affiliated with these phyla, these populations did not demonstrate any increase in cell numbers during the incubation period, despite the amendment of single (acetate) and more complex (lysate) carbon sources.

Metabolic potential

Seven size fractionated (3 enriched with acetate, 4 with cell lysate) and eight non-fractionated incubations (5 enriched with acetate, 3 with lysate) were selected for metagenomic sequencing. These incubations were selected based on community composition (according to ASVs) and the presence of populations affiliated with the Patescibacteria, Nanobdellota, or Omnitrophota. The metagenomes yielded 35 de-replicated metagenome-assembled genomes (MAGs) with a completeness over 50% and less than 5% contamination (Supplemental Table S2) of which 26 had an estimated completeness over 90%. None of the reconstructed genomes had an isolated species representative and 21 of the 35 reconstructed genomes were identified on genus level using alphanumeric placeholder labels while 3 of the 35

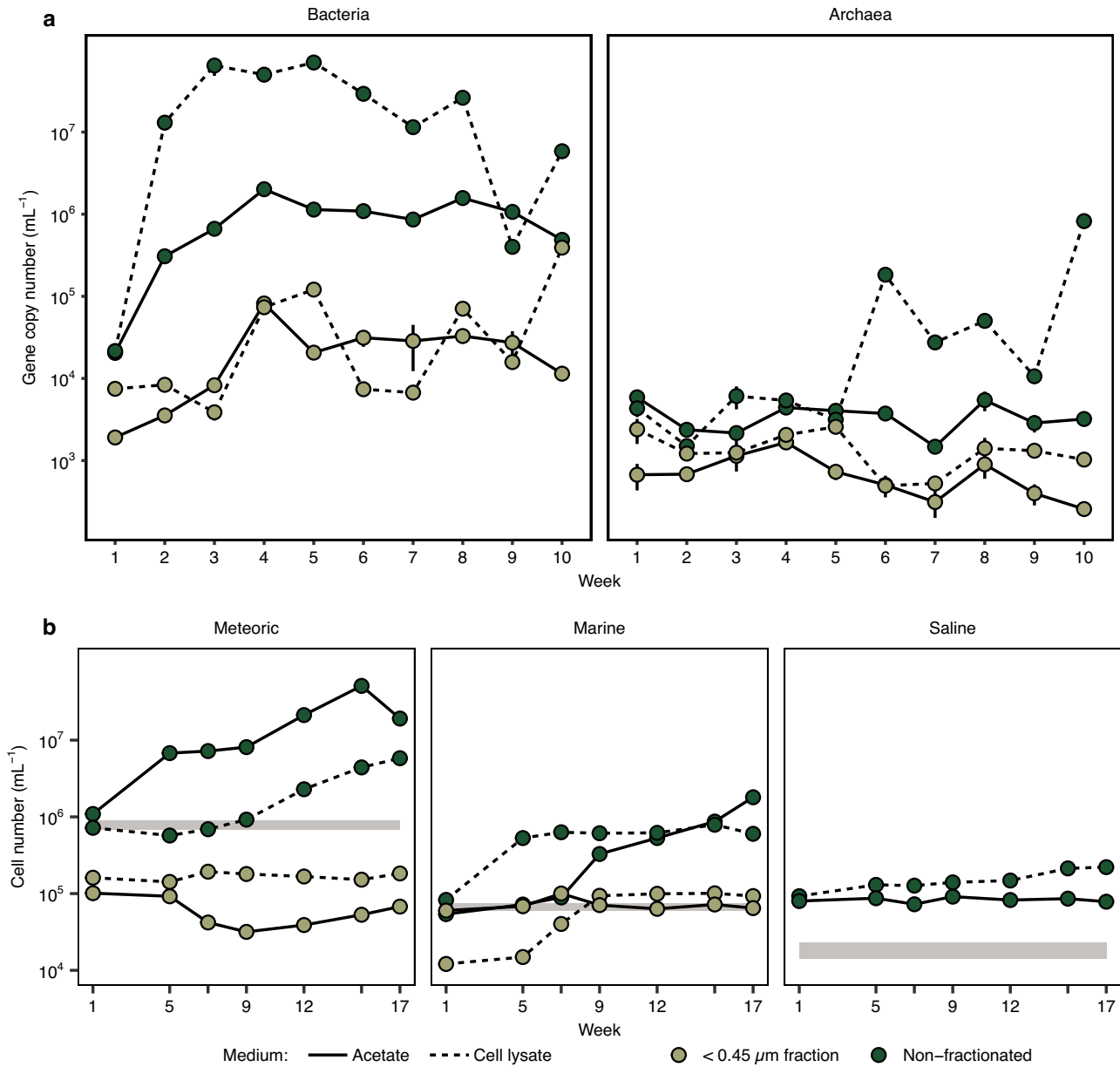


Fig. 2 | Growth of the bacterial and archaeal populations in the incubations. **a** Gene copy numbers (16S rRNA) according to real-time PCR of the incubations using meteoric groundwater as inoculum. Each week number represents a separate culture. Points represent mean values and the vertical bars indicate ± 2 standard deviations. The gene copy number was determined for ten incubations for each

media and size fractionation combination (total $n = 40$). **b** Cell numbers according to epifluorescence microscopy for the incubations using meteoric ($n = 4$), marine ($n = 4$), and saline ($n = 2$) groundwater. The gray bar depicts the range of the cell numbers of the groundwater at t_0 ($n = 2$).

genomes lacked a genus-level annotation entirely, underscoring the potential for the isolation of novel taxa in these subsurface groundwaters. The majority of these reconstructed genomes were affiliated with the phyla Desulfobacterota ($n = 7$), Bacillota (7), Pseudomonadota (6), Bacteroidota (4), Patescibacteria (3), and Nanobdellota (2). These MAGs are discussed collectively in order to compare their metabolic potential.

The reconstructed genomes affiliated with the Spirochaetota or Desulfobacterota had an average estimated genome size of 4.6 million base pairs (Mb, $sd = 0.78$ Mb, $n = 8$) that were considerably larger than the average genome size of reconstructed genomes affiliated with either Patescibacteria (range 0.4 - 0.9 Mb, $n = 3$) or Nanobdellota (0.8–1.2 Mb, $n = 2$). Despite their small genomes, a considerable variation in coding density was observed for both the Patescibacteria (0.88, 0.91, and 0.95) and the Nanobdellota (0.85 and 0.90), in line with the variation described in Castelle

et al.¹². In terms of metabolic potential, the ordination plot (using NMDS) based on KEGG orthologs revealed a clear differentiation of the genomes affiliated with the Patescibacteria or the Nanobdellota compared to the other populations. This differentiation was reflected in the lack of genomic potential for organic carbon oxidation (including glycolysis and fermentation), carbon fixation, and aerobic respiration in the genomes affiliated with the Patescibacteria and the Nanobdellota. Marker genes for aerobic respiration (COX1 and *coxSML*) were detected in 16/35 and 18/35 of the genomes, respectively, with all representatives of the Pseudomonadota ($n = 6$) having at least one of the two genes encoded that possibly enabled these populations to scavenge molecular oxygen. The potential for aerobic respiration could also indicate that these populations originated from the overlying soil groundwaters at Äspö island²⁴. All genomes affiliated with the Pseudomonadota encoded genes involved in either anaerobic respiration

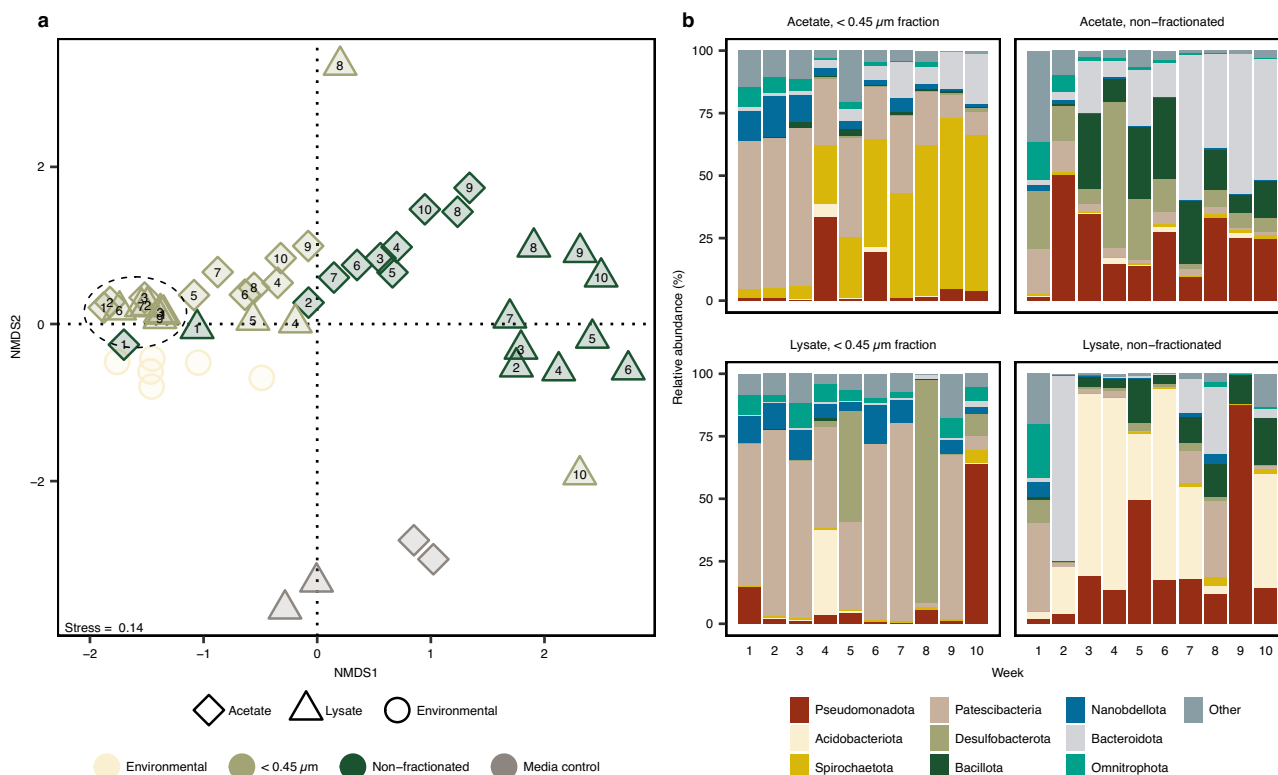


Fig. 3 | Beta diversity and composition of the communities in both the groundwaters and the incubations. **a** NMDS of microbial communities in the meteoric groundwater ($n = 6$) and the initial enrichment incubations using meteoric groundwater as inoculum for both fractionated ($n = 20$) and non-fractionated incubations ($n = 20$). The numbers refer to the incubation period in weeks. The

ordination was based on Bray-Curtis dissimilarities, and the relative abundance of the ASVs within a sample were used to calculate this metric. **b** Community composition of the corresponding incubations with meteoric groundwater showing the nine most abundant phyla while grouping low-abundant phyla as ‘Other’. Each bar represents one unique and independent culture (total $n = 40$).

(for example, fermentation) and aerobic respiration, suggesting these lineages were either facultative anaerobes or aerobes. In addition, *Rhodoferrax* sp. and JAHJQQ01 sp., both representatives of the Gammaproteobacteria (phylum Pseudomonadota), also had the potential for carbon fixation via the Calvin cycle. All lineages affiliated with the Desulfobacterota had a broad metabolic potential, with the possibility to fix inorganic carbon (Wood-Ljungdahl pathway), oxidize a broad range of organic compounds, and sulfur metabolism (Fig. 5). Representatives of the Bacillota were divided among the Clostridia and the Bacilli classes and the former contained the acetogens (*Acetobacterium* sp.) with the capacity for carbon fixation using the Wood-Ljungdahl pathway. The latter contained four lineages with smaller genomes (between 1.4 and 2.0 Mb) such as 4572 – 104 sp. and UBA2284 sp. with seemingly fermentative, strict anaerobic lifestyles and no capacity for aerobic respiration, sulfur metabolism, or carbon fixation.

The choice of using acetate as an organic component builds on previous investigations into metabolic features, revealing that certain populations affiliated with Patescibacteria had the potential to metabolize acetate (acetyl-CoA synthetase, *acs*, Supplemental Table S3)²⁹. However, despite such metabolic potential that was confirmed to be present in the incubations and an abundance of acetate in the medium, populations affiliated with the ABY1 class (Patescibacteria) did not grow. An explanation for this could be that the concentration of acetate as a carbon source in the enrichment incubations (1.2 mM) was too high for these groups that typically thrive in low carbon and energy environments, potentially causing a reduced growth efficiency under high energy conditions⁴². Most bacteria with a genome size smaller than 0.8 Mb and not affiliated with the Patescibacteria have been described as obligate symbionts¹². Possibly, the reduced genome size of these lineages prevented them from dividing in these fractionated incubations due to a limited biosynthetic capability and coupled dependence on a host or metabolic partner for cell division.

The three MAGs affiliated with the Patescibacteria included genes encoding peptidoglycan synthesis for cell wall formation (*murA-G*, *femX*, *mraY*) and cell division (*ftsZ*). Furthermore, this group encoded genes related to motility (*pilT*), pilus assembly (*pilCDMW*), production of extracellular polymeric substances (*gspG*), DNA uptake (*comEA*), and protein translocation (*secADEFGY*), that could all play a role in membrane transport or the physical attachment to adjacent cells. Genes related to organic carbon metabolism were limited to pyruvate metabolism (*porA*, *ppsA*), degradation of polysaccharides (*malZ*, *treS*), and acetyl-CoA production (*ackA*). With the caveat that the reconstructed genomes were incomplete, none of the reconstructed Patescibacteria genomes encoded a complete glycolysis pathway (nor the core module comprising five enzymes), citrate cycle, nor gluconeogenesis. Only one gene involved in fatty acid biosynthesis (*fabG*) was detected in one out of the three MAGs, suggesting an extracellular source of fatty acids or lipids was required to synthesize their cell membrane¹².

Overall, the MAGs affiliated with the Patescibacteria appeared to have an anaerobic lifestyle with only 1/3 genomes encoding a marker gene for aerobic respiration (cytochrome *c* oxidase). These MAGs seemed to lack most of the genomic potential for central carbon metabolism (e.g., glycolysis or citrate cycle) and may meet their carbon demands by metabolizing acetate, pyruvate, or polysaccharides, such as starch. The limited metabolic potential of these MAGs confirmed a symbiotic or at least an associative lifestyle, as in line with existing literature^{12,15,22}.

Co-occurrence analysis

To explore potential host-symbiont relationships, a co-occurrence analysis based on ASVs was performed using both previously published data on 24 groundwaters intersected by Äspö HRL ($n = 72$)¹¹, combined with the environmental data generated within this study ($n = 18$) and the non-

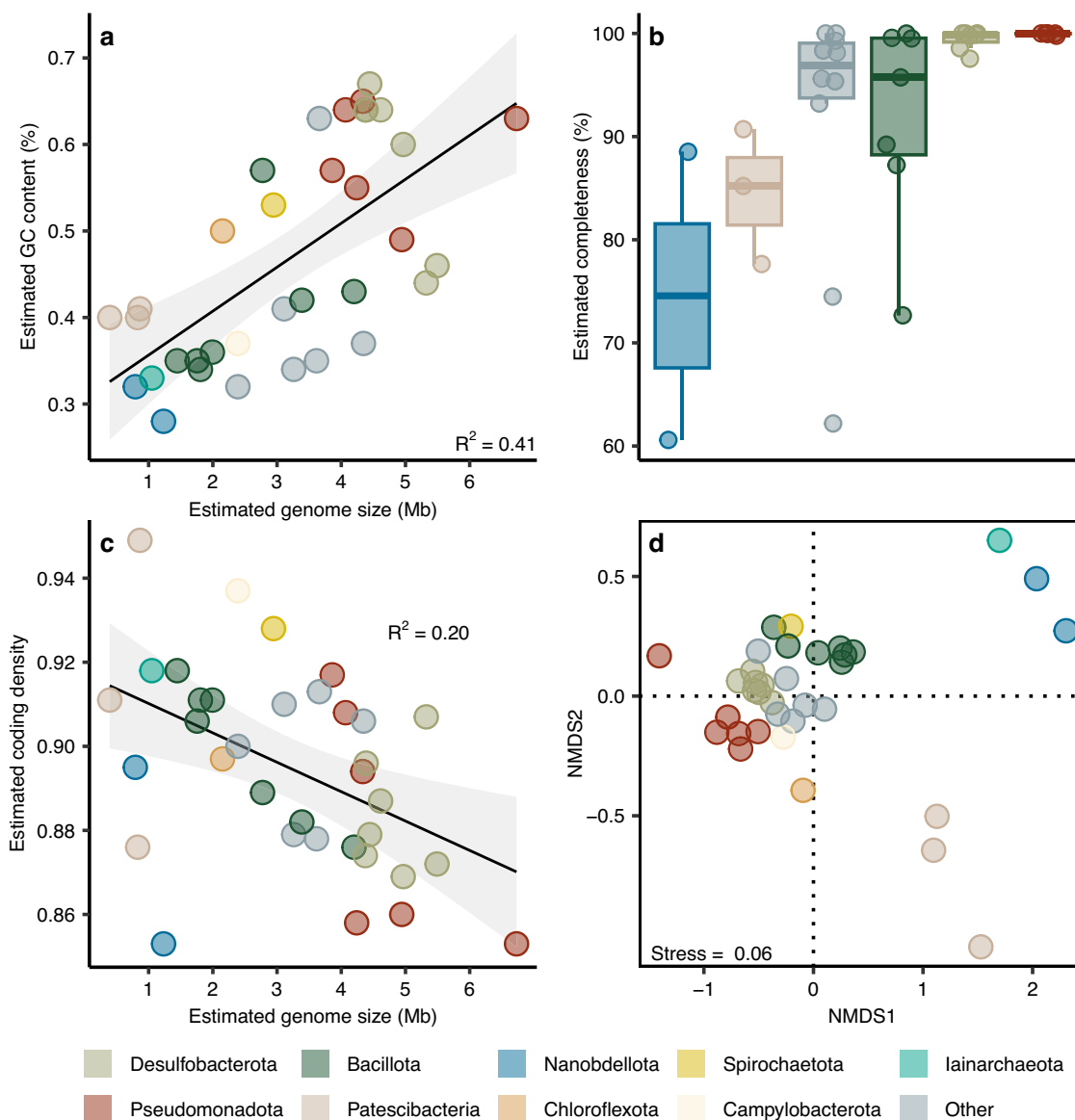


Fig. 4 | Features of the individual metagenome-assembled genomes. Each point represents a MAG ($n = 35$), colored according to phylum while grouping low-abundant phyla (based on 16S rRNA gene amplicons from the incubations) as “Other”. **a** Estimated genome size (x) versus estimated GC content (y). The black line represents a linear fit with the gray shading depicting a 95% confidence interval. **b** Estimated completeness, categorized according to phylum. The box represents the

first and third quartiles with the whiskers extending to 1.5 times the inter-quartile range. **c** Genome size (x) versus coding density (y). A higher coding density implies that coding sequences (open reading frames) take up a larger proportion of the total genome. **d** NMDS based on the repertoire and copy number of KEGG orthologs in the individual MAGs.

fractionated incubations ($n = 50$). This analysis revealed a co-occurrence (Spearman’s $\rho > 0.75$ and at least ten co-occurrences) of ASVs predominantly affiliated with the Patescibacteria (142 ASVs or nodes), Desulfobacterota (48), Bacteroidota (25), Omnitrophota (17), and Chloroflexota (10) in the groundwaters (Fig. 6). The network of the groundwater samples contained 206 co-occurring ASVs in total while the network of the incubations contained only 52 nodes. The low number of nodes (ASVs) in the network of the incubations was possibly due to different ASVs present in different incubations, thereby preventing robust co-occurrence patterns. To further investigate co-occurrences, the ASVs were grouped on phylum and order level (Supplemental Fig. S6 and S7, respectively) for both groundwater and incubation samples. This revealed a co-occurrence between the Patescibacteria and the Omnitrophota as well as certain orders affiliated with the Chloroflexota. For the Omnitrophota, this pattern was weak for the groundwater samples ($R^2\text{-adj} = 0.14$, $p\text{-value} = 0.05$, $df = 19$), but stronger for the incubations ($R^2\text{-adj} = 0.59$, $p\text{-value} = 3.4 \times 10^{-8}$, $df = 34$). Among

these Chloroflexota, it was especially the order Limnocyndrales that co-occurred with the orders UBA9983 and the Paceibacterales (Supplemental Fig. S7), both affiliated with the Paceibacteria class. However, this co-occurrence could not be confirmed with the incubations, as the Limnocyndrales were not abundant in the enrichment incubations. At order level there was a co-occurrence among the Dehalococcoidales (Chloroflexota) and the Magasanikbacterales (Patescibacteria) in a limited number of incubations ($R^2\text{-adj} = 0.86$, $p\text{-value} = 1.1 \times 10^{-9}$, $df = 19$). Both Chloroflexota and Omnitrophota have been reported to co-occur with Patescibacteria in suboxic or anoxic groundwaters^{43–45}. However, further studies are needed to investigate whether these co-occurrences are cell-cell interactions or a consequence of similar preferences for environmental conditions.

Conclusions

In contrast to the incubations with the complete inoculum, most incubations from size fractionated inoculant did not feature an increase in cell



Fig. 5 | Metabolic potential of the individual metagenome-assembled genomes. The colored bar depicts which phylum each MAG represents with the estimated genome completeness beneath it. **a** KEGG module completeness based on KEGG

orthologs. **b** Metabolic potential according to METABOLIC-c. **c** Marker genes involved in oxidative phosphorylation or aerobic respiration.

numbers over incubation periods up to 17 weeks. These incubations where cells larger than 0.45 μm were removed had a higher diversity (both ASV richness and Shannon H index) compared to non-fractionated incubations and included Patescibacteria, Nanobdellota, and Omnitrophota. Furthermore, the diversity of the fractionated incubations was similar to the original groundwater, suggesting a considerable part of the microbial diversity in these groundwaters was contained in the 0.1 - 0.45 μm fraction. While reconstructed genomes from the incubations affiliated with the Patescibacteria and Nanobdellota had the potential to e.g., metabolize acetate, and the incubations were enriched with this carbon source, no increase in cells numbers was observed. This was potentially due to a symbiotic or parasitic lifestyle. However, it should be noted that even in the non-fractionated incubations with presence of putative hosts, there was no systematic and consistent increase in the abundance of Patescibacteria and Nanobdellota.

Co-occurrence analysis based on ASVs using the environmental data ($n = 90$) and the non-fractionated incubations ($n = 50$) suggested an association between Patescibacteria and populations affiliated with the Desulfobacterota, Chloroflexota, and the Omnitrophota, and while it cannot be confidently stated that this is a host-symbiont relationship, this merits further study.

Methods

Initially, one groundwater (KR0015B) intersected by a borehole from the Äspö HRL was used to inoculate anaerobic enrichment media containing either a single organic carbon source (acetate) or a more complex organic carbon composition (cell lysate). The enrichments were incubated for ten weeks and after each week, the majority of a culture was sacrificed for extracting DNA. For each combination of media type and incubation

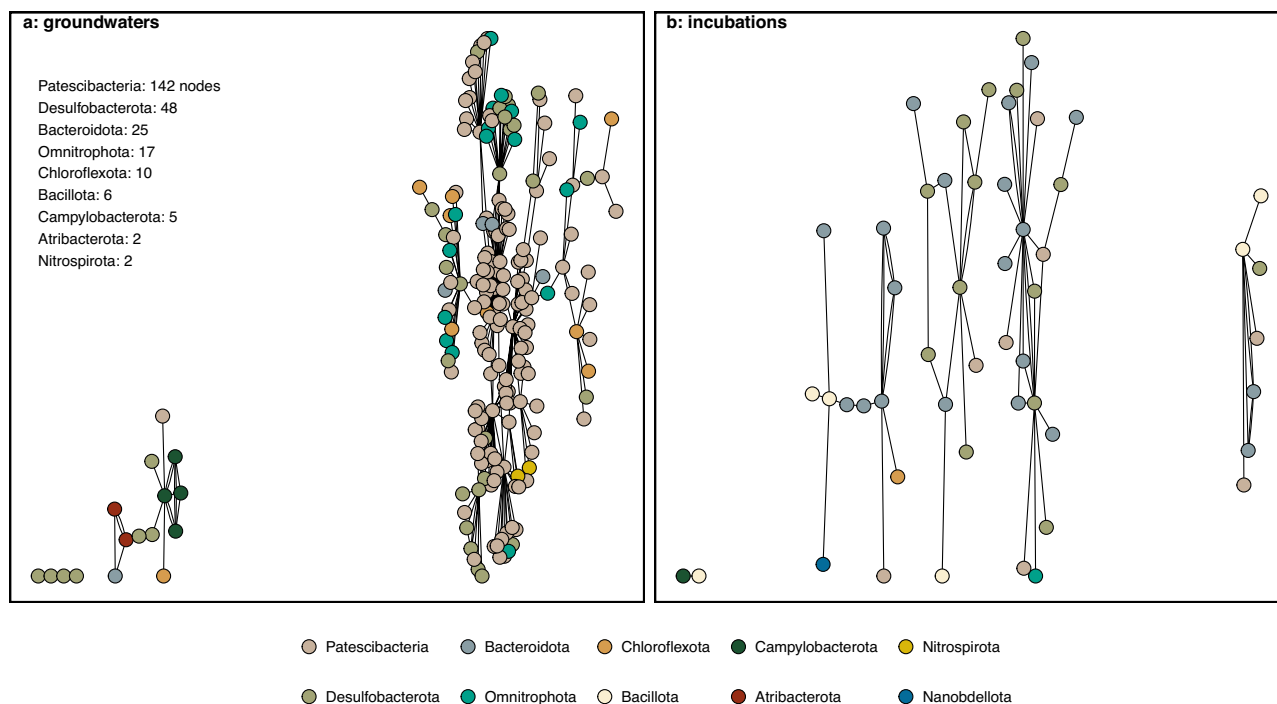


Fig. 6 | Network analysis of co-occurring ASVs. Each node depicts one ASV, connected by edges that represent a correlation according to Spearman's $\rho > 0.75$. **a** Co-occurring ASVs in the groundwater samples ($n = 72$) with 206 nodes in total. **b** Co-occurring ASVs in the non-fractionated incubations ($n = 50$) with 48 nodes in total.

period, two incubations were made as one culture was size fractionated by removing cells with a diameter larger than $0.45 \mu\text{m}$ from the inoculum, while the inoculum of the other culture was unmodified. This required 20 incubations to cover the incubation period for each media type, resulting in 40 incubations. Growth was assessed indirectly based on real-time PCR. Subsequently, three groundwater types (KR0015B, SA1420A-1, and SA2600A-1) were used as inoculum with identical enrichment media but quantifying cell numbers using epifluorescence microscopy. For each media type and size fraction, five replicate incubations were made (total $n = 50$) and incubated for 17 weeks. The enriched populations were identified using 16S rRNA gene sequencing, and genomic potential of populations of special interest was explored using genome-resolved metagenomics.

Groundwater sampling

Planktonic cells in the groundwater were captured before (October 2019) and after (April 2020) groundwater collection for the enrichment incubations using a high-pressure filter holder (Millipore) connected to the borehole. These boreholes are part of the Äspö HRL that is owned and operated by the Swedish Nuclear Fuel and Waste Management Company (SKB). Sampling connections were flushed with the groundwater of interest with three section volumes to reduce contamination risk from stagnant water. After flushing, the filter holder was equipped with a sterile polyvinylidene fluoride membrane 47 mm in diameter (Durapore, pore size $0.1 \mu\text{m}$)¹¹. Depending on the borehole, between 48 and 229 L of groundwater was filtered at a flow rate between 60 and 120 mL min^{-1} before retrieving the filter (Supplemental Table S1). The side of the filter containing the cells (facing the water source) was always facing inwards when stored in a tube, thereby likely releasing more cells from the filter during subsequent DNA extraction. The filters were aseptically collected, stored in liquid nitrogen during transport, and frozen at $-80 \text{ }^\circ\text{C}$ upon arrival at the home laboratory.

Groundwater collection

Groundwaters to be used as inoculum were collected in acid-washed, sterile glass bottles flushed with $100\% \text{ N}_2$ gas and kept under slight over-pressure by using 20 mm thick butyl stoppers. Identical to the groundwater sampling,

sampling tubes and connections were flushed with a minimum of three section volumes prior to water collection. Water was collected by puncturing the butyl stopper with a sterile needle that was connected to the borehole via the sampling tube. While maintaining over-pressure in the sampling bottle during water collection and allowing the bottle to partly fill up, the stopper was punctured with a second needle to prevent pressure buildup in the sample bottle. The water samples were transported to the laboratory at $4 \text{ }^\circ\text{C}$ where they were transferred to an anaerobic hood (Mbraun Labstar, H_2O and O_2 purity below 1 ppm) for use as inoculum on the same day as sampling.

Enrichment media

Two media were designed for enrichments, containing either acetate or a cell lysate as the sole organic carbon source. The acetate medium targeted sulfate-reducing bacteria and was adapted to the metabolic potential of clades of special interest that were previously detected at Äspö HRL while the cell lysate medium represented a complex medium not targeting any specific group. For selected incubations, the inoculum was size fractionated by applying a $0.45 \mu\text{m}$ filter, thereby removing the majority of larger cells from the inoculum. Resazurin was used in all incubations as a redox indicator as the resorufin/hydroresorufin redox couple is colorless at a redox potential below -110 mV , verifying that the medium was sufficiently anoxic.

Preparation of the cell lysate was based on Wu et al.⁴⁶. Briefly, an axenic culture of *Pseudomonas aeruginosa* was grown in $1 \times \text{LB}$ medium in dH_2O at $30 \text{ }^\circ\text{C}$ for 48 h while shaking. The cells were harvested by centrifuging 40 mL aliquots at $6000 \times g$ for 20 min at room temperature. After decanting, the pellet was washed twice with ultrapure water (MilliQ) before dissolving the pellet in 10 mL ultrapure water. The liquid was autoclaved at $121 \text{ }^\circ\text{C}$ for 20 min, sonicated for 20 min (130 W , 20 kHz , $100 \mu\text{m}$ amplitude; VCX 130, Sonics), followed by centrifugation at $6000 \times g$ for 10 min at room temperature. The supernatant was filtered ($0.2 \mu\text{m}$ pore size) and a subsample of the liquid was diluted (1:10) in ultrapure water to measure the total organic carbon content. Organic carbon content was measured using a spectrophotometer (DR 5000, Hach-Lange) combined with the LCK381 kit (Hach) with a detection range of 60 to 735 mg L^{-1} . The cell lysate (4 mL) was supplemented with (L^{-1}): 2 mL vitamin solution (Wolf's solution), 1 mL

trace element solution (SL-10), 1 mg resazurin, 0.3 g cysteine-HCl, 2.7 g NaHCO₃. Except for the vitamin and the cysteine-HCl solutions, the medium was dispensed into serum flasks and autoclaved for 20 min at 121 °C. Prior to inoculation, the vitamin (0.1 µm sterile filtered) and the cysteine-HCl solutions were added, the medium was purged for 30 min with N₂ gas (100%), and the pH was adjusted to 7.5 at 25 °C. Similar to the acetate medium, the serum flask was plugged with a butyl stopper and sealed with an aluminum crimp top.

Inoculation, incubation, and sampling

While working in an anaerobic hood (Mbraun Labstar), 40 mL of the acetate medium (pH 7.6 at 25 °C, components and preparation described in Supplemental Table S4) was aseptically dispensed into sterile serum flasks (capacity 100 mL). Groundwater inoculum (40 mL) was added and the serum flask was closed using a butyl rubber stopper (20 mm thick) combined with an aluminum crimp top while working in an anaerobic hood. Two incubation controls were made for each media that were identical to the incubations except for containing sterile ultrapure water (0.2 µm sterile filtered) instead of the inoculum (groundwaters). These controls were processed identically to the other incubations, including during the downstream molecular work. For the fractionated incubations, the inoculum (groundwater) was filtered with a syringe filter (Swinnex, Millipore) containing a membrane (25 mm in diameter, 0.45 µm pore size) while adding it to the serum flask. A pore size of 0.45 µm was chosen to select for small bacteria and archaea (e.g., Patescibacteria, Nanobdellota, and Omnitrophota). Instead, a pore size of 0.2 or 0.6 µm could be either too restrictive (leading to removal of most cells) or too permissive (not selective enough for the clades of interest), respectively. The culture flasks were incubated for either ten weeks (initial experiment) or 17 weeks (experiment using three groundwater types). The flasks were kept dark at 15 °C to resemble ambient temperatures in the meteoric (10.0 °C), marine (11.4 °C), and saline (13.7 °C) groundwater types. Growth was monitored using either real-time PCR or fluorescence microscopy. For real-time PCR, 40–60 mL of the incubations were subsampled in the anaerobic hood, and the cells were captured using a syringe filter holder (Millipore, 0.1 µm pore size) containing a sterile polyvinylidene fluoride membrane (25 mm in diameter). DNA was extracted directly after harvesting the cells. The extracted DNA was used both for real-time PCR and for PCR followed by 16S ribosomal RNA gene amplicon sequencing. For microscopy, the surface of the butyl stopper was sterilized by dipping the stopper in a 70% ethanol-soaked cloth and briefly moving the stopper through a flame. The incubations were sampled with a sterile needle after flushing the dead volume with 100% N₂ gas.

DNA isolation

The DNeasy PowerWater kit (Qiagen) was used for extracting DNA from cells captured on the filter. To do so, the manufacturer's instructions were followed apart from eluting the nucleic acids in 50 µL instead of 100 µL elution buffer in order to increase the concentration. One negative extraction control from the groundwater sampling and one from the sampling of the incubations were included (both blank filters) and processed simultaneously with the other samples. DNA yield was measured using a Qubit 2.0 fluorometer (Life Technologies) with a lower detection limit of 0.05 ng µL⁻¹ (Supplemental Table S5 and S6).

16S rRNA gene amplification and sequencing

The V3–V4 region of the 16S rRNA gene was amplified with the primer pair 341 F and 805R⁴⁷, using a maximum of 10 ng DNA as template. A maximum of 10 µL of the DNA extract was added if the DNA concentration was below or just above the lower detection limit of the fluorometer. The product of the first PCR (20 cycles) served as template for a second amplification step (12 cycles) containing the library-specific sequencing barcodes (Nextera DNA Dual-Indexes). The amplified product was purified after each PCR using the AMPure XP bead-based reagent (Beckman Colter). For the negative extraction controls as for low biomass samples in general, the

amplified product was typically not detected after gel electrophoresis and the volume of the entire purified PCR product (approximately 35 µL) was used for equimolar pooling (i.e. 39 ng of each library). After equimolar pooling, the concentration of the pooled library was measured (Qubit 2.0) and the fragment size distribution was assessed using automated gel electrophoresis (Agilent TapeStation 4150). The libraries were sequenced at the Science for Life Laboratory, Sweden, on an Illumina MiSeq platform, producing 2 × 300 bp paired-end reads. The concentration of the library pool was verified at the sequencing facility with a qPCR targeting the sequencing barcodes. After sequencing, demultiplexing of the data was done by the sequencing facility with the user-provided spreadsheet containing the sequencing barcodes for each library.

Real-time PCR and epifluorescence microscopy

The qPCR was run as previously described⁴⁸. Briefly, the 16S rRNA gene was amplified using the bacterial primers 908F-mod and 1075R⁴⁹ and the archaeal primers 915 F and 1059R⁵⁰ on a LightCycler 480 instrument (Roche Diagnostics). Cycling conditions were 2 min at 95 °C, 15 s at 95 °C and 30 s at 60 °C, followed by a melt curve analysis to assess product 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). The efficiency (calculated as $10^{-1/\text{slope}} - 1 \times 100$) was 92.8 (range 90.2–94.8%, $n = 5$) and 92.5% (range 91.0–94.1%, $n = 5$) for the bacterial and archaeal assays, respectively. Standards ($n = 7$), samples of interest, and no-template controls were run in triplicates and the former two were 1:10, 1:100, and 1:1000 diluted in nuclease-free water to account for potential inhibitors in the DNA extracts.

For the microscopy, 1 mL of the culture was fixed with 2.2% formaldehyde (vol/vol) and 1 × TE-buffer (molecular grade, pH 8.0) to a final volume of 10 mL, incubated at 4 °C for 10 min, and brought on a black polycarbonate membrane (pore size 0.2 µm, Whatman) under a vacuum pressure of 10 to 15 kPa and briefly placed in a sterile petri dish to dry. Meanwhile, a staining solution was prepared consisting of 30% 100 × SYBR Green I (Invitrogen), 10% 10 g L⁻¹ p-phenylenediamine (sterile filtered, pore size 0.1 µm), and 60% 1:1 glycerol in PBS (sterile filtered, pore size 0.2 µm). 10 µL of the staining solution was brought on a glass slide and cover slip before adding the filter and incubating in the dark for 5 min. 300 cells or ten fields of view (5 × 5 raster) were counted in the dark using a 100-fold objective on an epifluorescence microscope (Olympus BX50) combined with a light source (Olympus) that illuminated the DNA/dye complex at a wavelength of 497 nm. A validation of the obtained cell counts was done by quantifying the cell numbers in an axenic culture (*Pseudomonas aeruginosa*) with a light microscope (Zeiss Primo Star), diluting the culture to 10⁶ cells mL⁻¹, followed by quantification with epifluorescence microscopy.

Metagenomics

Libraries of incubations of special interest based on community composition ($n = 15$) were prepared for metagenomic sequencing using the Tecan MagicPrep while adjusting the cycling conditions (7–12 cycles, details in Supplemental Table S7) to the amount of template DNA added (10 to 80 ng) according to the manufacturer's instructions. The concentration of the libraries was measured using a Qubit 3 fluorometer (Life Technologies). After equimolar pooling (65.1 ng DNA of each library), the fragment length distribution of the pooled libraries was checked using agarose gel electrophoresis. The metagenomes were sequenced at the Science for Life Laboratory (Sweden) on an Illumina NovaSeq platform equipped with a SP flowcell, producing 2 × 150 bp paired-end reads.

Bioinformatics and data analysis

Raw sequencing reads were processed using the nf-core/ampliseq pipeline^{51,52} (v2.12.0) that relied on Nextflow (v24.04.4), Cutadapt (v4.6), FastQC (v0.12.1), DADA2 (v4.3.2), and the SBDI Sativa curated 16S GTDB database (release 220)⁵³. The nf-core/ampliseq pipeline was run

with default settings, except for the trimming of the primers whereby reads not containing the primer sequence or containing two copies were discarded from downstream analysis. ASVs present in the DNA extraction controls of the groundwater sampling and the incubations were removed from the respective samples prior to the analysis. In addition, ASVs from the acetate and lysate controls (four controls in total) were removed from the incubation samples that were enriched with these media. For the network analysis, the relative abundance of each ASV within a sample was correlated to every other ASV using Spearman's ρ in the groundwaters ($n = 18$), previously published data on 24 groundwaters intersected by Äspö HRL ($n = 72$)¹¹, and the non-fractionated incubations ($n = 50$) while filtering out ASVs with a read count below 0.1%. A minimum correlation of 0.75 (Spearman's ρ) was chosen as a threshold for the network graph, while setting the minimum co-occurrences of two ASVs to 10 and 3 for the groundwaters and incubations, respectively. The R libraries *igraph* and *tidygraph* were used to construct and visualize the networks^{54,55}. For the co-occurrence analysis on order and phylum level, the relative abundances of ASVs affiliated with Patescibacteria and the order or phylum of interest were summed within each sample, followed by a square root transformation. These summed abundances were used in a linear model, followed by scoring the goodness of fit with the R-squared value.

Raw sequences from the metagenomes ($n = 15$) were assembled, binned, and annotated using the *nf-core/mag* pipeline^{52,56} (v2.2.1). Adapter and quality trimming was done with *fastp* (v0.23.2), reads were assembled into contigs 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 *eggNOG-mapper* (v2.1.9), 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 *CheckM2* (v1.0.2). De-replication was done using *dRep* (v3.4.2), setting the maximum contamination to 5% and the minimum completeness to 50%. *GTDB-Tk* (v2.4.0) was used for taxonomic assignment together with the *GTDB* release 220 as a reference database. Next, the coverage of all the reconstructed genomes was assessed by mapping the quality-trimmed reads (the output of *fastp*) to the de-replicated bins with *CoverM* (v0.6.1). Finally, to explore the metabolic potential of the reconstructed genomes, the de-replicated bins combined with the quality-trimmed reads were analyzed using *METABOLIC-c* (v4)⁵⁷.

Statistics and reproducibility

Data analysis, statistics, and data visualization were performed in R (4.4.1). A compiled version of the Quarto document together with the number of sequence reads throughout the bioinformatic pipeline are provided on GitHub at <https://doi.org/10.5281/zenodo.16270880>. Curated pipelines were used that were run on mainly default settings for processing both the amplicon and metagenome data.

Reporting summary

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

Data availability

All sequencing data is publicly available at the European Nucleotide Archive under project reference PRJEB71769. The nucleotide sequences of the MAGs were deposited in Figshare under the accession 10.6084/m9.figshare.31228549.

Code availability

The R code generated to analyze the data and to create the figures has been deposited at <https://doi.org/10.5281/zenodo.16270880>.

Received: 21 July 2025; Accepted: 3 February 2026;

Published online: 14 February 2026

References

- Magnabosco, C. et al. The biomass and biodiversity of the continental subsurface. *Nat. Geosci.* **11**, 707–717 (2018).
- Pedersen, K. Microbial life in deep granitic rock. *FEMS Microbiol. Rev.* **20**, 399–414 (1997).
- 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).
- Bell, E. et al. Active sulfur cycling in the terrestrial deep subsurface. *ISME J.* **14**, 1260–1272 (2020).
- Wu, X. et al. Potential for hydrogen-oxidizing chemolithoautotrophic and diazotrophic populations to initiate biofilm formation in oligotrophic, deep terrestrial subsurface waters. *Microbiome* **5**, 37 (2017).
- Lin, L.-H. et al. Long-term sustainability of a high-energy, low-diversity crustal biome. *Science* **314**, 479–482 (2006).
- Chapelle, F. H. et al. A hydrogen-based subsurface microbial community dominated by methanogens. *Nature* **415**, 312–315 (2002).
- Chivian, D. et al. Environmental genomics reveals a single-species ecosystem deep within Earth. *Science* **322**, 275–278 (2008).
- Hubalek, V. et al. Connectivity to the surface determines diversity patterns in subsurface aquifers of the Fennoscandian shield. *ISME J.* **10**, 2447–2458 (2016).
- 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).
- Lopez-Fernandez, M., Astrom, 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).
- Castelle, C. J. et al. Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. *Nat. Rev. Microbiol.* **16**, 629–645 (2018).
- Herrmann, M. et al. Predominance of *Cand.* Patescibacteria in groundwater is caused by their preferential mobilization from soils and flourishing under oligotrophic conditions. *Front. Microbiol.* **10**, 1407 (2019).
- Seymour, C. O. et al. Hyperactive nanobacteria with host-dependent traits pervade Omnitrophota. *Nat. Microbiol.* **8**, 727–744 (2023).
- Brown, C. T. et al. Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523**, 208–211 (2015).
- Wang, Y. et al. Genetic manipulation of Patescibacteria provides mechanistic insights into microbial dark matter and the epibiotic lifestyle. *Cell* **186**, 4803–4817 (2023).
- Kuroda, K. et al. Microscopic and metatranscriptomic analyses revealed unique cross-domain parasitism between phylum *Candidatus* Patescibacteria/candidate phyla radiation and methanogenic archaea in anaerobic ecosystems. *mBio* **15**, e0310223 (2024).
- Yakimov, M. M. et al. Cultivation of a vampire: '*Candidatus* Absconditicoccus praedator'. *Environ. Microbiol.* **24**, 30–49 (2022).
- Sakai, H. D. et al. Insight into the symbiotic lifestyle of DPANN archaea revealed by cultivation and genome analyses. *Proc. Natl. Acad. Sci. USA* **119**, e2115449119 (2022).
- Luef, B. et al. Diverse uncultivated ultra-small bacterial cells in groundwater. *Nat. Commun.* **6**, 6372 (2015).
- Nakai, R. Size Matters: ultra-small and filterable microorganisms in the environment. *Microbes Environ.* **35**, ME20025 (2020).
- Wrighton, K. C. et al. Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science* **337**, 1661–1665 (2012).
- Mathurin, F. A., Astrom, M. E., Laaksoharju, M., Kalinowski, B. E. & Tullborg, E. L. Effect of tunnel excavation on source and mixing of groundwater in a coastal granitoidic fracture network. *Environ. Sci. Technol.* **46**, 12779–12786 (2012).
- Westmeijer, G. et al. Connectivity of Fennoscandian Shield terrestrial deep biosphere microbiomes with surface communities. *Commun. Biol.* **5**, 37 (2022).

25. Alakangas, L. J., Mathurin, F. A. & Åström, M. E. Diverse fractionation patterns of Rare Earth Elements in deep fracture groundwater in the Baltic Shield—Progress from utilisation of Diffusive Gradients in Thin-films (DGT) at the Äspö Hard Rock Laboratory. *Geochim. Cosmochim. Acta* **269**, 15–38 (2020).
26. Pedersen, K., Arlinger, J., Ekendahl, S. & Hallbeck, L. 16S rRNA gene diversity of attached and unattached bacteria in boreholes along the access tunnel to the Äspö hard rock laboratory, Sweden. *FEMS Microbiol. Ecol.* **19**, 249–262 (1996).
27. Osterholz, H. et al. Terrigenous dissolved organic matter persists in the energy-limited deep groundwaters of the Fennoscandian Shield. *Nat. Commun.* **13**, 4837 (2022).
28. Mehrshad, M. et al. Energy efficiency and biological interactions define the core microbiome of deep oligotrophic groundwater. *Nat. Commun.* **12**, 4253 (2021).
29. Dopson, M. et al. Novel candidate taxa contribute to key metabolic processes in Fennoscandian Shield deep groundwaters. *ISME Comm.* **4**, ycae113 (2024).
30. Lopez-Fernandez, M. et al. Investigation of viable taxa in the deep terrestrial biosphere suggests high rates of nutrient recycling. *FEMS Microbiol. Ecol.* **94**, fty121 (2018).
31. 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).
32. 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).
33. Westmeijer, G. et al. *Candidatus* Desulforudis audaxviator dominates a 975m deep groundwater community in central Sweden. *Commun. Biol.* **7**, 1332 (2024).
34. Wrighton, K. C. et al. Metabolic interdependencies between phylogenetically novel fermenters and respiratory organisms in an unconfined aquifer. *ISME J.* **8**, 1452–1463 (2014).
35. Liu, R. et al. Novel *Chloroflexi* genomes from the deepest ocean reveal metabolic strategies for the adaptation to deep-sea habitats. *Microbiome* **10**, 75 (2022).
36. Mehrshad, M. et al. Hidden in plain sight—highly abundant and diverse planktonic freshwater Chloroflexi. *Microbiome* **6**, 176 (2018).
37. Perez-Molphe-Montoya, E., Küsel, K. & Overholt, W. A. Redefining the phylogenetic and metabolic diversity of phylum Omnitrophota. *Environ. Microbiol.* **24**, 5437–5449 (2022).
38. Lindsay, M. R. et al. Species-resolved, single-cell respiration rates reveal dominance of sulfate reduction in a deep continental subsurface ecosystem. *Proc. Natl. Acad. Sci. USA* **121**, e2309636121 (2024).
39. Kodama, Y. & Watanabe, K. *Sulfuricurvum kujiense* gen. nov., sp. nov., a facultatively anaerobic, chemolithoautotrophic, sulfur-oxidizing bacterium isolated from an underground crude-oil storage cavity. *Int. J. Syst. Evol. Microbiol.* **54**, 2297–2300 (2004).
40. Liu, Y.-F. et al. Anaerobic hydrocarbon degradation in candidate phylum ‘Atribacteria’ (JS1) inferred from genomics. *ISME J.* **13**, 2377–2390 (2019).
41. Brune, A., Song, Y., Oren, A. & Paster, B. J. A new family for ‘termite gut treponemes’: description of *Breznakiellaceae* fam. nov., *Gracilinema caldarium* gen. nov., comb. nov., *Leadbettera azotonutricia* gen. nov., comb. nov., *Helmutkoenigia isoptercolens* gen. nov., comb. nov., and *Zuelzera stenostrepta* gen. nov., comb. nov., and proposal of *Rectinemataceae* fam. nov. *Int. J. Syst. Evol. Microbiol.* **72**, 005439 (2022).
42. Ishida, Y., Imai, I., Miyagaki, T. & Kadota, H. Growth and uptake kinetics of a facultatively oligotrophic bacterium at low nutrient concentrations. *Microb. Ecol.* **8**, 23–32 (1982).
43. Chaudhari, N. M. et al. The economical lifestyle of CPR bacteria in groundwater allows little preference for environmental drivers. *Environ. Microbiome* **16**, 24 (2021).
44. Chaudhari, N. M., Pérez-Carrascal, O. M., Overholt, W. A., Totsche, K. U. & Küsel, K. Genome streamlining in Parcubacteria transitioning from soil to groundwater. *Environ. Microbiome* **19**, 41 (2024).
45. Wang, J., Zhong, H., Chen, Q. & Ni, J. Adaption mechanism and ecological role of CPR bacteria in brackish-saline groundwater. *npj Biofilms Microbiomes* **10**, 141 (2024).
46. Wu, X. et al. Culturing of “unculturable” subsurface microbes: natural organic carbon source fuels the growth of diverse and distinct bacteria from groundwater. *Front. Microbiol.* **11**, 3171 (2020).
47. 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).
48. Westmeijer, G. et al. Continental scientific drilling and microbiology: (extremely) low biomass in bedrock of central Sweden. *Biogeosciences* **21**, 591–604 (2024).
49. Ohkuma, M. & Kudo, T. Phylogenetic analysis of the symbiotic intestinal microflora of the termite *Cryptotermes domesticus*. *FEMS Microbiol. Lett.* **164**, 389–395 (1998).
50. 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).
51. 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).
52. Ewels, P. A. et al. The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* **38**, 276–278 (2020).
53. Lundin, D. & Andersson, A. SBDI Sativa curated 16S GTDB database. *SciLifeLab. Dataset*. <https://doi.org/10.17044/scilifelab.14869077.v4> (SBDI, 2021).
54. Pedersen, T. L. tidygraph: a tidy API for graph manipulation. *R package version* <https://CRAN.R-project.org/package=tidygraph> (CRAN, 2024).
55. Csardi, G. et al. igraph: network analysis and visualization in R. *R package version* **1**, 10–5281 (2023).
56. 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).
57. 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).

Acknowledgements

The authors thank the Swedish Nuclear Fuel and Waste Management Co (SKB) for access to the Äspö HRL and the Sicada database. The authors acknowledge Emelie Nilsson (LNU) and Daniel Lundin (LNU) for their advice and support during the bioinformatic processing, and Matthias Hoetzing (SLU) and Prune Leroy (SLU) for their help in preparing the metagenomic sequencing libraries. The study was supported by The Swedish Research Council (contracts 2018-04311, 2017-04422, and 2024-04607). MD thanks the Crafoord Foundation (contract 20180599). SB acknowledges financial support from the Swedish Research Council and Science for Life Laboratory. 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 (NAISS 2024/22-1165 and 2024/6-279). The computations were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) and the Swedish National Infrastructure for Computing (SNIC) at UPPMAX, Uppsala University, partially funded by the Swedish Research Council through grant agreement nos. 2022-06725 and 2018-05973.

Author contributions

M.D., S.B., and M.M. conceived the study; G.W. and S.T. sampled the groundwaters, G.W. and P.H. inoculated and maintained the anaerobic

incubations; G.W., P.H., and S.T. performed the molecular work and processed the sequencing data; G.W. performed ecological interpretations on the data; G.W. and M.D. wrote the manuscript with comments from all authors. M.D. and S.B. provided funding. All authors read and improved the final version of the manuscript.

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/s42003-026-09706-8>.

Correspondence and requests for materials should be addressed to Mark Dopson.

Peer review information *Communications Biology* thanks the anonymous reviewers 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) 2026