Microbiome

RESEARCH

Open Access

A glimpse of the paleome in endolithic microbial communities

Carl-Eric Wegner[1], Raphaela Stahl[2], Irina Velsko[2], Alex Hübner[2], Zandra Fagernäs[2], Christina Warinner[2,3,4], Robert Lehmann[5], Thomas Ritschel[5], Kai U. Totsche[4,5] and Kirsten Küsel[1,4,6*]

Abstract

Background The terrestrial subsurface is home to a significant proportion of the Earth's microbial biomass. Our understanding about terrestrial subsurface microbiomes is almost exclusively derived from groundwater and porous sediments mainly by using 16S rRNA gene surveys. To obtain more insights about biomass of consolidated rocks and the metabolic status of endolithic microbiomes, we investigated interbedded limestone and mudstone from the vadose zone, fractured aquifers, and deep aquitards.

Results By adapting methods from microbial archaeology and paleogenomics, we could recover sufficient DNA for downstream metagenomic analysis from seven rock specimens independent of porosity, lithology, and depth. Based on the extracted DNA, we estimated between 2.81 and 4.25×10^5 cells $\times g^{-1}$ rock. Analyzing DNA damage patterns revealed paleome signatures (genetic records of past microbial communities) for three rock specimens, all obtained from the vadose zone. DNA obtained from deep aquitards isolated from surface input was not affected by DNA decay indicating that water saturation and not flow is controlling subsurface microbial survival. Decoding the taxonomy and functional potential of paleome communities revealed increased abundances for sequences affiliated with chemolithoautotrophs and taxa such as *Cand.* Rokubacteria. We also found a broader metabolic potential in terms of aromatic hydrocarbon breakdown, suggesting a preferred utilization of sedimentary organic matter in the past.

Conclusions Our study suggests that limestones function as archives for genetic records of past microbial communities including those sensitive to environmental stress at modern times, due to their specific conditions facilitating long-term DNA preservation.

Keywords Subsurface, Metagenomics, Endolithic, DNA damage, Chemolithotrophy

*Correspondence:
Kirsten Küsel
kirsten.kuesel@uni-jena.de
[1] Aquatic Geomicrobiology, Institute of Biodiversity, Friedrich Schiller University Jena, Dornburger Str. 159, 07743 Jena, Germany
[2] Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany
[3] Department of Anthropology, Harvard University, Cambridge, MA, USA
[4] Cluster of Excellence Balance of the Microverse, Friedrich Schiller University Jena, Jena, Germany
[5] Hydrogeology, Institute of Geosciences, Friedrich Schiller University Jena, Burgweg 11, 07749 Jena, Germany
[6] German Center for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Puschstraße 4, 04103 Leipzig, Germany

Background

The subsurface harbors a significant portion of the Earth's microbial biomass and contributes to global biogeochemical cycling [1–3]. The difficulty of access impairs estimating global subsurface biomass, activity, and biodiversity, especially in the continental biosphere. A comprehensive compilation of cell count measurements suggested that there are approximately 2 to 6×10^{29} cells in the continental subsurface [3]. Biomass estimates are exposed to significant uncertainties due to poorly understood parameters such as the ratio of surface-attached to pelagic groundwater cells, for which assumptions range

between 1 and 10,000. Total organic carbon content and groundwater cellular abundances have been shown to be poor predictors for biomass and biodiversity [1–4].

Groundwater and other aqueous sample material are keys [4–8] for studying subsurface microbiomes but only provide limited information regarding surface-attached or endolithic microbes inhabiting rock matrix pores. Microbial communities inhabiting the subsurface have been studied predominantly in porous, unconsolidated sediments, for example, alluvial aquifer systems [9–12]. The bedrock itself has been rarely investigated [13–15]. Similarly, the vadose zone, the shallow unsaturated bedrock zone more connected to surface habitats [16, 17], has received little attention. Water saturation and nutrient supply, both controlled by relief position, rock properties (i.e., porosity, permeability, fracture network, composition), and groundwater quality and circulation patterns, control subsurface microbial life [16, 18].

The subsurface endolithic microbiome consists of subsurface specialists that prefer particular lithologies [19, 20], long-term descendants of organisms that colonized sediments during deposition [21], surface immigrants transported by fluid flow over geological time [21, 22], and invaders introduced as a result of human activities, such as drilling or flooding. Continental subsurface habitats, not fuelled by liquids and gases from active and hot regions of the Earth's crust, are often viewed as energy-starved systems, as the total organic carbon (TOC) can be refractory and scarce, especially in crystalline rock settings. For most sedimentary subsurface environments, organic carbon contents are not a limiting factor. Ancient sedimentary carbon might represent a significant source of carbon for microorganisms in subsurface rock environments [23–25]. Part of these carbon compounds can be still metabolized, diffuse from aquitards into aquifers [26] and from less permeable into more permeable layers where they drive microbial activity [9, 11]. Nonetheless, the availability of carbon and energy sources can be very heterogeneous, and their provenance can differ as well. We have shown that chemolithoautotrophy is highly relevant in groundwater from such environments [27] and modulated by the overall limited availability of inorganic electron donors and acceptors [28, 29]. Because of the low amounts of microbial biomass and the challenge of recovering it, 16S rRNA gene amplicon studies from rock core material have been the primary means of investigating microbial community composition. These surveys provide limited insights into the metabolic potential of organisms and by default do not allow discrimination between living, potentially active, and dead cell material. Advances in paleomicrobiology, achieved through distinguishing "ancient" and "modern" DNA by high-throughput sequencing and DNA damage pattern analysis, are potential door openers for subsurface microbiology. Similar to hard tissue samples (bone, teeth, shells) [30–33], carbonate rocks contain calcium carbonates and calcium phosphates, which could adsorb or encase extracellular DNA, released postmortem, by neutralizing negative charges present in the DNA backbone and the mineral surface by bivalent calcium cations [34]. We hypothesized that limestone/marlstone parent material would allow the recovery of metagenomic DNA (mgDNA), which could be sufficient to gain insights into the genomic potential of endolithic microbes.

In this study, we adapted methods from microbial archaeology and paleogenomics for mgDNA recovery, used comprehensive wet- and dry-lab control measures to minimize the risk of contamination, applied metagenomics, and analyzed DNA damage patterns. The goal was to assess endolithic microbial biomass and use DNA damage as a proxy to distinguish DNA from intact and potentially alive cells from the paleome, the genetic remains of past microbial communities [30] including damaged cells stressed during environmental fluctuations at modern times. In addition, we aimed for decoding the taxonomic compositions and metabolic potentials of the endolithic microbiomes to understand how these communities are or were adapted to a life in consolidated rocks.

Methods

Bedrock sampling and sample preparation

We collected fractured bedrock from Upper Muschelkalk marine deposits (Germanic Triassic) in the groundwater recharge area of the Hainich low mountain range, as well as from isolated equivalents in the center of the Thuringian Syncline (both Central Germany). Sampling was done during the construction of groundwater monitoring wells (Hainich CZE: 2011, 2014; samples: H13-17, H22-8, H22-30, KS36-H32, CM1-H32) and during the INFLUINS exploratory drilling (EF1/2012: 2013; samples: INF-MB2, INF-MB3). The INFLUINS cores were extracted using drilling mud based on local drinking water supplies and bentonite. The Hanich CZE cores were recovered using local groundwater (Kammerforst deep well: KS36-H32, CM1-H32; Hainich CZE well H51: H13-17, H22-8, H22-30). Drilling fluids used to recover the INFLUINS cores contain dominant taxa [17], which were not detected in analyzed rock core samples, showing that bacteria in the water used for drilling did not contaminate the inner parts of the cores that were used for subsequent DNA extractions. Measures to minimize contaminations included utilization of washed, de-rusted, steam-cleaned drill pipes, and ethanol-washed PVC liners in the Hainich CZE. Selected core segments of drill cores, recovered with rotary drill rigs (mud-rotary wireline), were

immediately wrapped in sterile plastic bags and transported on dry ice until storage in deep freezers (−80 °C). Subsamples of bedrock matrices for DNA extraction that were not in contact with drilling fluid and other potential sources of contamination were prepared by fast hydraulic splitting of still frozen drill cores, under removal of the outer parts of the core segments. Subsamples for X-ray micro-computed tomography (X-ray μCT) analysis (13-mm plugs, vertical orientation) were prepared with a drill press. Samples for carbon analysis were extracted from directly adjacent rock and also used for rock typing.

Rock typing/characterization, pore classification, and analysis of carbon fractions

The rocks were classified based on stereoscope inspection, carbonate test (HCl 10%), and analytical carbon measurements by applying the Dunham [35] and a mudrock classification scheme [36]. Porosity types and pore sizes were classified as described previously [37, 38]. Milieu indicators, including weathering colors and secondary pore mineralizations (Munsell colors), and derived oxicity rating were determined by stereoscope inspection and also contrasted against characteristics of the core segment and borehole/well. The contents of total carbon and organic carbon (TOC) of the rock samples were determined on homogenized duplicate subsamples (~1.6 mg) of ~30-g ground rock using an elemental analyzer (Euro EA, HEKAtech, Wegberg, Germany). The OC was calculated as the difference from total carbon measurements released under combustion at 950 °C and 600 °C.

X-ray micro-computed tomography (X-ray μCT)

The three-dimensional structure of the plugs was assessed nondestructively by X-ray μCT (Xradia 620 Versa, Zeiss, Jena, Germany). Each plug was scanned in 1601 projections to give a full 360° rotation at 0.4× magnification and an exposure time of 2 s per step using X-rays produced with 80 kV and 126 μA. Tomographic reconstruction yielded a three-dimensional grayscale image with 1024^3 voxels at a resolution of 25.99 μm with automated removal of ring artifacts and beam hardening. Images were cropped to remove boundaries, denoised with nonlocal means filtering [39] and binarized into pore space and solid by manual thresholding using Fiji (ImageJ v. 1.51) [40]. The pore sizes were calculated from binarized images using the maximum inscribed sphere algorithm implemented in the BoneJ plug-in [41] in Fiji. The volumetric pore size distribution was derived from the histogram of resulting images, while total X-ray μCT visible porosity was derived directly from the histogram of binarized images. Connected pore space of binarized images was assessed assuming 26 connectivity and visualized by randomly assigning a color to each set of voxels belonging to the same region.

Protocols for DNA extraction and sequencing library preparation

We adapted protocols routinely used for ancient DNA preparation for downstream sequencing, which are all available from protocols.io (https://dx.doi.org/10.17504/protocols.io.bvt9n6r6). We reference the respective protocols in the following sections and describe them for the sake of completeness. The bench protocols available on protocols.io include detailed lists with respect to needed equipment and reagents, as well as necessary precautions.

DNA extraction

DNA extraction from rock samples was performed by modifying a protocol originally designed for recovering ancient DNA from dental calculus (https://dx.doi.org/10.17504/protocols.io.bidyka7w). Metagenomic DNA was extracted from either 2.5 g of rock powder obtained using a dental drill or 2.5 g of rock pieces obtained by chipping rock material. To decalcify the samples, the rock material was rotated in EDTA (0.5 M, pH 8.0) for up to 10 days (rock pieces, rock powder 5 days) at 37 °C before being concentrated down to a volume of 1 mL using Amicon® ultra centrifugal filtering units (MWCO 30 kDa and 10 kDa). Concentrated samples were mixed with 1 mL of extraction buffer (EDTA pH 8.0, 0.45 M; Proteinase K 0.025 mg/mL) and rotated overnight at 37 °C. Samples were spun down and subsequently mixed with 10 mL of binding buffer (guanidine hydrochloride, 4.77 M; isopropanol, 40% [v/v]) and 400-μL sodium acetate (3 M, pH 5.2). Samples were transferred to a high pure extender assembly from the High Pure Viral Nucleic Acid Large Volume kit (Roche, Mannheim, Germany) and centrifuged for 8 min with 1500 rpm at room temperature. The column from the high pure extender assembly was removed, placed in a new collection tube and dried by being centrifuged for 2 min with 14,000 rpm at room temperature. A total of 450 μL of wash buffer (High Pure Viral Nucleic Acid Large Volume kit) were added, and samples were centrifuged for 1 min at 8000×g at room temperature. This washing step was repeated once, and columns were dried afterwards by centrifugation. DNA was eluted into a siliconized tube by adding 50 μL of TET (0.04% Tween 20 in 1×Tris–EDTA [pH 8.0]), incubating samples for 3 min at room temperature, and centrifugation for 1 min 14,000 rpm at room temperature. The elution step was repeated once, and the pooled eluate was stored at −20 °C until further processed. All outlined steps were carried out in the ancient DNA lab of Max Planck Institute for the Science of Human History

(MPI-SHH) to reduce the risk of contamination with modern environmental DNA. Blank extractions were carried out alongside the sample extractions, using identical steps, with the exception that water instead of rock material was used as input material. DNA concentrations were determined using a Qubit® fluorometer and the DNA high-sensitivity assay (ThermoFisher, Schwerte, Germany). Cell number estimates were calculated by dividing the amount of extracted DNA per gram rock material by the approximate mass of one prokaryotic genome, assuming a molecular weight per base pair of 618 Da (g/mol) [42] and a genome length of 3 Mbp.

Library preparation

Anticipating that extracted metagenomic DNA could contain both severely fragmented ancient DNA and high-molecular-weight modern DNA, we first used a Covaris M220 ultrasonicator to shear any high-molecular-weight DNA present to a maximum length of 500 bp prior to library construction. This ensured that all DNA present in the DNA extract would be suitable for library construction. We then used a library construction protocol (https://dx.doi.org/10.17504/protocols.io.bakricv6) that is specifically designed to be compatible with degraded and ultrashort DNA fragments [43]. Metagenomic DNA samples were blunt end repaired by mixing 10 µL of DNA with 40 µL of a mastermix containing NEB buffer no. 2 (1×), ATP 1 mM, BSA 0.8 mg/mL, dNTPs 0.1 mM, T4 PNK 0.4 U, and T4 polymerase 0.024 U. Samples were incubated for 20 min at 25 °C, followed by a 10-min incubation step at 12 °C. Blunt-end repaired samples were subsequently purified using the MinElute Reaction Clean-up Kit (Qiagen, Hilden, Germany). Samples were finally eluted in 20 µL of the elution buffer containing 0.05% Tween20. 18 µL of eluted samples were mixed with 21 µL of a mastermix containing Quick Ligase buffer (final concentration 1×) and a mix of adapters (0.25 µM). Next, 1 µL of Quick Ligase (5 U) was added, and libraries were incubated at 22 °C for 20 min. Reactions were again purified using the MinElute Reaction Clean-up Kit. Samples were eluted using 22-µL elution buffer. The adapter fill-in reaction was performed in a final volume of 40 µL. The reaction mix consisted of a 20-µL eluate and a 20-µL mastermix containing isothermal buffer (final concentration 1×), dNTPs (0.125 mM each), and Bst polymerase (0.4 U). Reactions were incubated for 30 min at 37 °C, before being incubated at 80 °C for additional 10 min to inactivate the polymerase. Before being further processed, libraries were quality checked by quantitative PCR (qPCR). Dilutions of the libraries (1:10) were mixed (1-µL template), with 19 µL of a mixture containing DyNAmo mastermix (final concentration 1×) and IS7 and IS8 primers (1 µM). The thermal profile was 10 min at 95 °C, 40 cycles of 30 s at 95 °C, 1 min at 60 °C, 30 s at 72 °C, followed by a melting curve (60–95° C). Libraries were subsequently indexed (https://dx.doi.org/10.17504/protocols.io.bvt8n6rw) and amplified (https://dx.doi.org/10.17504/protocols.io.beqkjduw) as outlined in the referenced protocols. Libraries were equimolarly pooled and sequenced on an Illumina NextSeq 500 instrument in paired-end mode (2×150 bp) using v. 2.5 chemistry. The sequencing depth ranged between 2.24 and 4.81 Gbp (Table S1). All outlined steps were carried out in the ancient and modern DNA clean rooms of the MPI-SHH to reduce the possibility of contamination. Library blanks were prepared alongside the sample extractions, using identical steps, with the exception that water instead of rock material eluate was used as input material.

Sequence data preprocessing

Quality parameters of raw sequencing data were assessed using *FastQC* (v0.11.8) [44]. Adapter and quality trimming were done with *bbduk* (v38.22) [45] (settings: qtrim=rl trimq=20 ktrim=r k=25 mink=11) using its included set of common sequence contaminants and adapters. Trimmed sequences were subsequently subjected to taxonomic profiling and metagenome assembly and binning.

Taxonomic profiling

Trimmed sequences were taxonomically profiled using *kaiju* (v1.7.3) [46] and *diamond* (v2.0.7.145) [47, 48]. *Diamond* was used for the taxonomic assignment of trimmed, and paired-end assembled (with *vsearch* (v2.14.1) [49]) sequences, while *kaiju* was used for the taxonomic assignment of assembled contigs. For *kaiju*, sequences were translated into open-reading frames, which were used for string matching with the implemented backward-search algorithm based on the one that is part of the Burrows-Wheeler transform [50, 51]. *Kaiju* was run in greedy mode with up to 5 allowed mismatches (-a greedy -e 5). *Diamond* searches were done in sensitive mode applying an *E*-value threshold of 0.0001 (-e 0.00001 -c 1 –sensitive). Database hits were annotated making use of the LCA algorithm implemented in *megan* (v6.21.1) [52, 53] with default settings. NCBI nr [54] was used as the reference database for taxonomic profiling (*kaiju*, nr_euk release 2020–05; *diamond*, custom built database based on NCBI nr retrieved from NCBI in 2020–03). Taxa representing contaminants on different taxonomic levels were identified using taxonomic profiles obtained from *diamond* and *decontam* (v1.1.1) [55] based on prevalence and frequency in true samples and extraction and library blanks.

Metagenome assembly and binning

Metagenome coverage was estimated based on k-mer redundancy using *nonpareil* (v3.303) (-T kmer) [56, 57]. Trimmed sequences were assembled into contigs with *megahit* (v1.2.9) (default settings) [58] and *metaSPADES* (v3.13.0) (–only-assembler) [59, 60]. Due to better performance, we used the megahit assemblies for all subsequent steps. Contigs longer than 1 kb were kept, and quality-controlled sequences were mapped onto these contigs using *bowtie2* (v2.3.4.1) [61] (–no-unal). Resulting.sam files were converted into.bam files and indexed with *samtools* (v1.7) [62]. Contigs and indexed mapping files were used for manual metagenomic binning using *anvio* (v. 6.2) [63] based on sequence composition and differential abundance. The completeness, redundancy, and heterogeneity of bins were assessed with *checkm* (v1.1.2) [64]. Bins were taxonomically assigned using *gtdb-tk* (v0.3.2) [65].

Functional annotation

Functional profiling of trimmed sequences was done with *humann* (v3.0) [66] using precompiled Uniref50 and Uniref90 protein databases (release 2019–01) and applying default settings. The resulting gene families table was regrouped to KEGG orthologies, normalized to copies per million (CoPM) and summarized with respect to pathways and functions of interest using a custom *python* script (available in the supplementary material and from the Open Science Framework [OSF] repository mentioned under "Availability of data and material").

DNA damage pattern analysis

Using assembled contigs and the output from mapping trimmed sequences back onto them, DNA damage patterns were identified and analyzed using *mapdamage* (v2.2.1) [67, 68] and *pydamage* (v0.50alpha) [69]. The output from *mapdamage* was ultimately used as it provides metrics with respect to all possible DNA damage-related substitutions. DNA damage pattern analysis was also done for selected subsets of the assembled contigs based on taxonomy (assigned with *kaiju*).

Figure generation

Figures were prepared using the R packages *ggplot2* (part of *tidyverse*) (v1.3.1) [70] and *ggpubr* (v0.4.0) (https://rpkgs.datanovia.com/ggpubr/index.html) and finalized with *inkscape* (https://inkscape.org/).

Results

General sample characteristics and porosity analysis

We analyzed five bedrock samples from the vadose zone of a low-mountain range groundwater recharge area (Hainich Critical Zone Exploratory (CZE)) from depths between 9 and 33 m below ground level (mbgl) and two samples from deep isolated aquitards with similar stratigraphic position and lithology (INFLUINS deep drilling) from depths 285 and 296 mbgl. The rock samples, representing the thin-bedded marine alternations of mixed carbonate-siliciclastic rock that form widely distributed fractured-rock aquifers, range from argillaceous marlstones to bioclastic limestones with a broad range of porosity (Table 1). Three samples showed pores bigger than 0.02 mm (Fig. S1) with volumetric fractions of 0.9% (INF-MB3), 2.4% (H22-30), or 8.9% (H13-17). INF-MB3 (Fig. S2) showed a distribution of pores within 0.02–0.28 mm, which appeared at homogeneously distributed, but disconnected locations. The pore space in H22-30 (Fig. S3) also shows several disconnected pores but includes fractures and carbonate dissolution features that span large parts of the entire sample. The pore size distribution is slightly higher in the range of 0.02–0.52 mm. With pores in the size of 0.02–1.58 mm, H13–17 featured large macropores (Fig. 1) from intensive carbonate dissolution that connect most of the internal pore space. H22-8 consisted of dense rock and showed only a single fracture in a size near the µCT limit of detection (Fig. 1) that impeded a meaningful quantification. The other three rock samples did not show any pores above 0.02 mm, reflecting very dense rock matrices (Figs. S4–S6). The macroscopic inspection revealed the presence of secondary Fe-minerals in large dissolution pores in two limestone specimens, also representing connected matrix habitats in the main aquifer (Trochitenkalk formation) (Fig. 1A + E, Table 1). The total carbon content ranged between 5.53 ± 0.18 (H22-8) and 12.39 ± 0.17% (H32-KS36). The organic carbon content was, with the exception of CM1-H32 (8.17 ± 1.49%), below 3%.

Recovery of metagenomic DNA (mgDNA) independent from the specimen

We were able to extract mgDNA from all rock specimens. DNA extractions yielded higher amounts from rock pieces than from ground rock powder (Fig. 2A) with concentrations ranging between 0.011 and 0.051 $ng \times \mu L^{-1}$ (0.033 ± 0.013 $ng \times \mu L^{-1}$) for pieces and 0.019 ± 0.005 $ng \times \mu L^{-1}$ from powder. The latter was in the range of the extraction blanks (0.017 ± 0.007 $ng \times \mu L^{-1}$). The quantitation of prepared sequencing libraries by quantitative PCR yielded results in line with the results from DNA extraction (Fig. 2A).

Based on the amount of extracted DNA from the processed samples, we crudely estimated the number of cells potentially present in the rock material. Taking into account the molecular weight of one base pair and using

Table 1 Origin and contextual data with respect to processed rock samples. Pandora DB refers to the internal sample database of the MPI SHH/MPI EVA, pwd refers to rock powder samples, and pc refers to rock pieces samples

Sample ID	Pandora DB ID	Sampling depth (mbsf)	Aquifer type	Water saturation (in situ)	Rock type[a], genetic porosity[b], pore size classes[c]	Milieu indicators (plug, Munsell colors)	Oxicity	Porosity[d] (%)	Pore size range (mm)	Total carbon (%)	Organic carbon (%)	Estimated cell concentration (g^{-1})
H13-17	SET004.B0103 (pwd) SET004.B0203 (pc)	14.34–14.48	Fracture/karst	Saturated	Limestone (packstone); D (e); mc, sms-lms	Matrix: 2.5Y 6/2; secondary Fe-minerals in pores	Oxic	8.9	0.02–1.58	12.00 ± 0.02	2.95 ± 0.69	3.62E+05
H22-8	SET005.A0103 (pwd) SET005.A0203 (pc)	13.61–13.70	Fracture	Unsaturated	A: Delithified argillaceous marlstone; F, D (e); mc B: Delithified calcareous mudstone; F, D (e); mc	A, matrix: 2.5Y 7/4; B, matrix: 2.5Y 7/2	Oxic to sub-oxic	Na	Na	A: 5.52 ± 0.18 B: 8.60 ± 0.10	A: 1.42 ± 0.47 B: 2.12 + 0.26	4.25E+05
H22-30	SET004.A0103 (pwd) SET004. A0203 (pc)	32.55–32.80	Fracture	Unsaturated	Limestone (oolithic packstone); D (e), F; mc, sms-lms	Matrix: 2.5Y 6/2; secondary Fe-minerals in pores	Oxic	2.4	0.02–0.52	11.92 ± 0.08	0.73 ± 0.09	2.38E+05
KS36-H32	SET001.B0104 (pwd) SET001.B0203 (pc)	8.92–9.06	Fracture	Unsaturated	Limestone (packstone); D (e); mc, sms-lms	Matrix: 5Y 5/1; secondary Fe-minerals in pores	Oxic	Na	Na	12.39 ± 0.17	0.59 ± 0.14	3.52E+05
CM1-H32	SET003.A0103 (pwd) SET003. A0203 (pc)	21.9–22.0	Fracture	Saturated	Calcareous mudstone to limestone (wackestone); D (r); mc	Matrix: 10Y 3.5/1	Oxygen deficient	Na	Na	10.74 ± 0.05	8.17 ± 1.49	4.12E+05
INF-MB2	SET001.A0103 (pwd) SET001. A0203 (pc)	285.44–285.62	Aquitard	Saturated	Calcareous mudstone; D (r); mc	Matrix: 5GY 2.5/1	Anoxic	Na	Na	8.61 ± 0.09	2.93 ± 0.07	2.94E+05
INF-MB3	SET002.A0103 (pwd) SET002. A0203 (pc)	295.71–295.87	Aquitard	Saturated	Limestone (packstone to grainstone); D (r); mc	Matrix: 5Y 6/1; bioclasts: N3	Anoxic	0.9	0.02–0.28	12.28 ± 0.01	0.86 ± 0.01	2.94E+05

[a] Dunham limestone classification after Wright (1992) [35] and mudrock classification (after Hennissen et al. (2017) [36], modified)

[b] (Visible) carbonate porosity class after Ahr et al. (2005) [37]: apparent genetic factors: S (depositional; *i*, interparticle), D (diagenetic; *d*, dissolution; *p*, replacement; *r*, reduced; *e*, enhanced), F (fracture)

[c] Pore size classes after Choquette and Pray (1970) [71]: mc (micropores, < 1/16 mm, macroscopically invisible), sms (small mesopores, 1/16–1/2 mm), lms (large mesopores, 1/2–4 mm), smg (small megapores, 4–32 mm)

[d] Based on µCT analysis for pore sizes >0.26 µm, Na = no pores < 0.26 µm

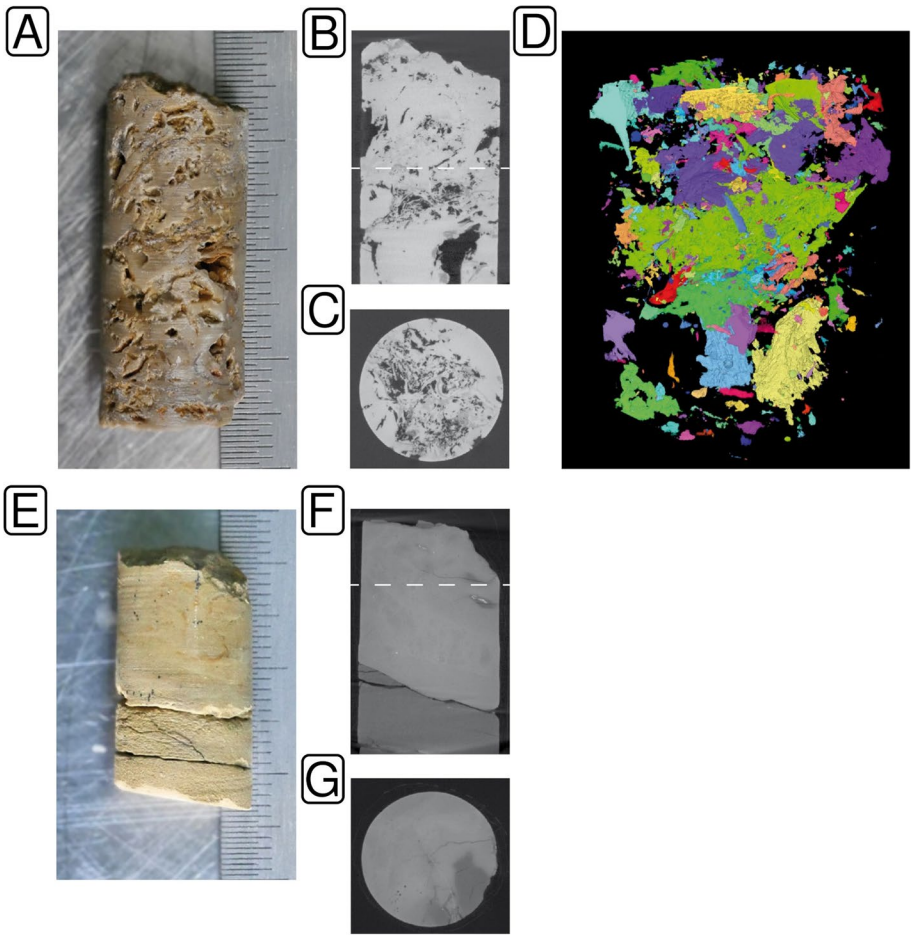

Fig. 1 Pore space characteristics of samples H13-17 (**A**–**D**) and H22-8 (**E**–**G**) determined by µCT analysis. Moldic pores (up to large mesopores) dominate the packstone. Scale: 0.5 mm. Plug diameter 13 mm (**A**). Vertical section shows considerable porosity. The dashed line marks the position of C (**B**). Horizontal section (**C**). Reconstructed pore space. Colors mark parts of the pore system that are each connected by throats > 26 µm (**D**). The plug comprises delithified siliceous marlstone (lower part) and delithified calcareous mudstone (upper part). Scale: 0.5 mm. Plug diameter 13 mm (**E**). Vertical section shows thin fractures (micropores). The dashed line marks the position of G (**F**). Horizontal section showing fine fractures and rare micro- to small mesopores. The matrix exhibits no pores connected by throats > 26 µm (**G**)

a length of three million base pairs as proxy for a prokaryotic genome, we estimated between 2.81 and 4.25×10^5 cells $\times g^{-1}$ processed rock material.

Taxonomic profiling

Sequence data pre-processing (Table S1, Fig. S7) indicated that the length distribution was generally skewed towards shorter lengths (Fig. 2B). Consequently, the proportion of taxonomically assigned reads was rather low and varied between 6.2 and 18.6% (Fig. 2C, Table S1). k-mer-based redundancy analysis (Fig. S8) suggested that our data covered more than 90% of the anticipated diversity based on recovered mgDNA. Decontamination analysis identified in total 31 contaminants, one on phylum level (Spirochaetes), two on class level (Epsilonproteobacteria, Chlamydia), 9 on family, and 19 on genus level (Table S2). Principal component analysis on phylum level (Fig. 3A) showed that H22-8, H22-30, and KS36-H32 were separated from blank data sets, independent from decontamination. The remaining four data sets were grouped together with some of the library and extraction blanks, independent of sample type. Decontamination made datasets more distinguishable from blanks, which was for instance evident in the case of CM1-H32 and H13-17. On family level (Fig. 3B), decontaminated data sets could be clearly distinguished from blanks. For the subsequent taxonomic profiling, pieces and powder data sets have been pooled.

Taxonomic profiles were characterized by inverse abundance patterns that divided the data sets into two groups. Group (1) included H22-8, H22-30, and KS36-H32 and group (2) H13-17, CM1-H32, INF-MB2, and INF-MB3. Acidobacteria (3.93–11.48%), *Cand.* Rokubacteria (8.28–17.09%), Chloroflexi (4.07–14.74%),

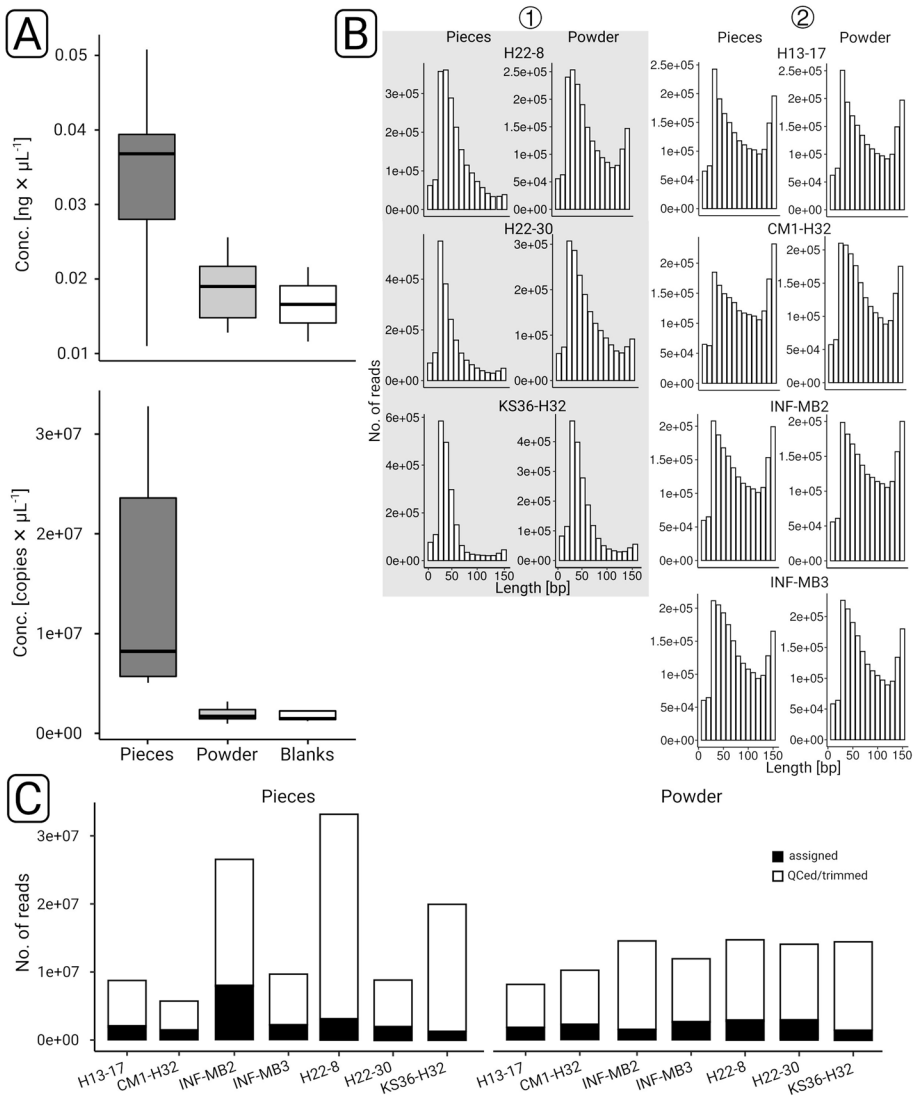

Fig. 2 Overview of data (pre-)processing. Samples were quantified by fluorometry and quantitative PCR after DNA extraction (upper panel) and library preparation (lower panel) (**A**). Sequence length histograms were generated after quality control and trimming based on subsampled ($n = 1$ million read pairs) data sets. The gray shading highlights three data sets for which the read length distribution was skewed to the left. Based on taxonomic profiling (see main text), we summarized these three data sets in two groups: (1) and (2) (**B**). The proportion of quality-controlled and trimmed sequences that could be assigned taxonomically was determined based on database queries with *diamond* against NCBI nr

Cyanobacteria (0.56–2.71%), NC10 (1.49–4.18%), Nitrospirae (1.33–3.01%), and Thaumarchaeota (0.69–2.75%) (Fig. 3C, Table S3) featured increased abundances in group (1). In comparison, the relative abundances of for instance Firmicutes (up to 15.34%), *Cand.* Saccharibacteria (2.68–9.08%), and Bacteroidetes (15.01–20.08%) were higher in group (2). Some of the mentioned taxa were also detected in the blanks. Bacteroidetes reached abundances up to 36%, while *Cand.* Saccharibacteria were only detected in one blank (4.7%). The relative abundances of Acidobacteria and Chloroflexi did not exceed 2 and 1.5%, respectively. Cyanobacteria abundances were comparable between data sets and blanks. Nitrospirae were only found in two blanks, and the abundances were below 0.5% (Table S3). Proteobacteria were highly abundant in all data sets (up to 70.81%) and partially much more abundant in the blanks (up to 85.3%).

Decontamination did not lead to major changes in the taxonomic profiles (Fig. 3C). Lesser abundant phyla increased in relative abundance. Examples include *Cand.* Eisenbacteria (KS36-H32), *Cand.* Jorgensenbacteria (H22-30), *Cand.* Levybacteria (H22-8), and *Cand.* Omnitrophica (KS36-H32, H22-8). Taxonomic profiles

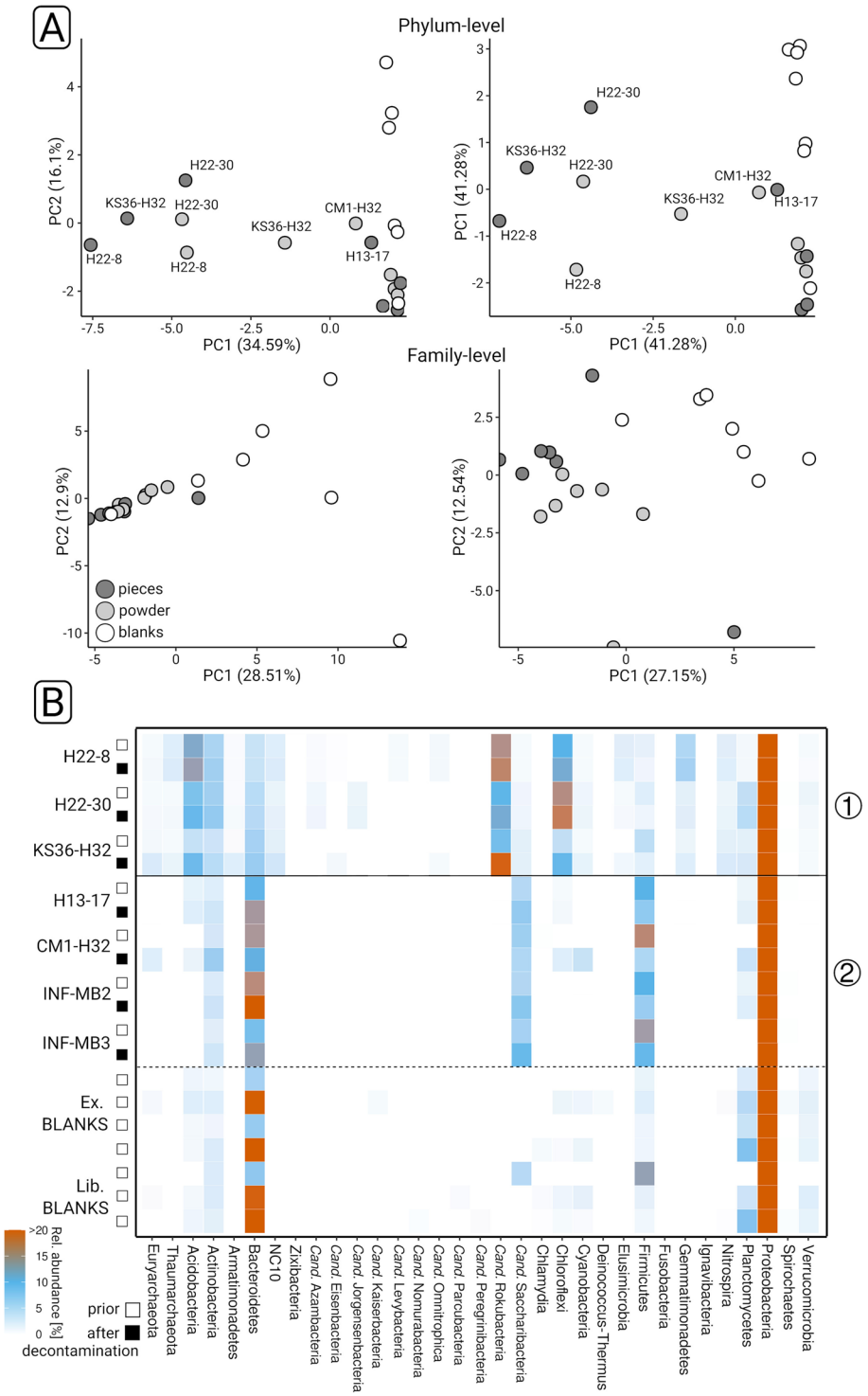

Fig. 3 Taxonomic profiling of rock endolithic microbial communities. Principle component analyses were carried out based on phylum-level (**A**) and family-level (**B**) taxonomic profiles, prior to (left) and after (right) decontamination. The color coding indicates the sample type. Phylum-level taxonomic profiles were visualized as heatmap (**C**). (1) and (2) indicate two groups of samples (see main text for details). White and black boxes indicate if the corresponding profile is based on decontaminated data. Ex. and Lib. BLANKS refer to extraction and library blanks, respectively

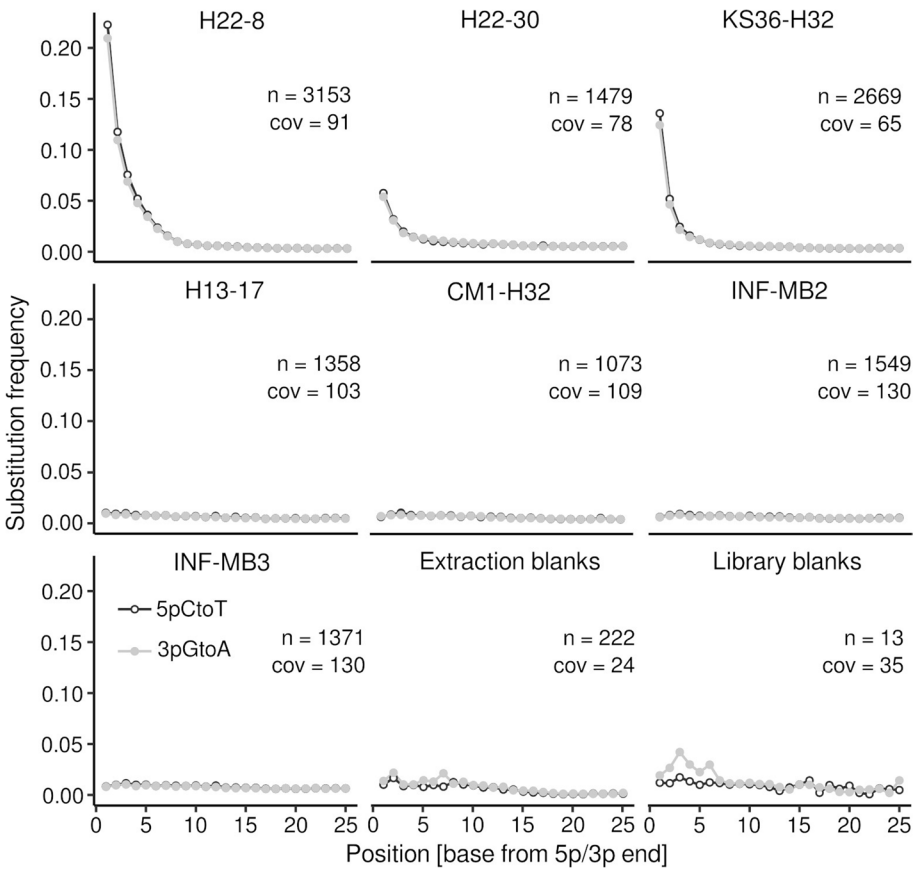

Fig. 4 DNA damage pattern analysis. Quality-controlled sequence reads were mapped onto assembled contigs (> 1 kbp). The damage pattern analysis was carried out with *mapdamage* (v.2.2.1) [69]. The plots show the substitution frequency (5pCtoT [5′ cytosine to thymine substitutions], 3pGtoA [3′ guanine to adenine substitutions]) versus the relative position (from the 5p and 3p end). n, number of contigs > 1 kbp considered for the analysis; cov, mean coverage of the contigs

at deeper levels are not described as the assignment rate dropped beyond phylum level.

Metagenome assembly and DNA damage pattern analysis

For DNA damage pattern analysis, we co-assembled data sets from rock pieces and rock powder from all sites. We compared two different assemblers, *megahit* [58] and *metaSPADES* [60], and ultimately settled on the *megahit* assembly. The assemblies obtained from *metaSPADES* featured larger total assembly lengths, but N50 values and maximum contig lengths were significantly larger when using *megahit* (Fig. S9). From none of the assemblies, we obtained more than 3153 contigs longer than 1 kbp (1.07–3.15 contigs, 1.81 ± 0.78 [mean ± SD]). The N50 values and the maximum lengths of these contig subsets were rather low, 1.69 ± 0.16 and 16.79 ± 5.43 kbp, respectively. The proportion of recruited reads (after quality control) to the individual assemblies ranged between 6.8 and 23.8% (average 17%), which indicated that our assemblies are only representative for a small part of the generated sequencing data (Table S4). We used the mapping files from read recruitment analysis to determine mgDNA fragment lengths (Fig. S10), which showed that fragment sizes were, with the exception of H22-30, shorter for group (1) samples.

From a taxonomic perspective, the assembled contigs were skewed towards few taxa that assembled well. Contigs from group (2) data sets are dominated by Actinobacteria and Proteobacteria, with combined relative abundances above 95% (Table S5). The contigs from group (1) data sets were taxonomically more diverse but also dominated by Actinobacteria and Proteobacteria, with combined relative abundances of 76% or more. Taxa that were highly abundant based on profiling quality-controlled sequences were underrepresented. For instance, no more than 0.8% of the assembled contigs were affiliated with *Cand.* Rokubacteria (H22-30), and we only obtained contigs from this taxon from group (1) data sets (Table S5).

Mapping metagenomic sequence reads onto assembled contigs larger than 1 kb revealed a pronounced deamination signal in the case of group (1) samples. We detected substitution frequencies partially above 20% (Fig. 4). 5' cytosine to thymine substitutions (5pCtoT) and 3' guanine to adenine substitutions (3pGtoA) were comparable for group (1) data sets. Substitution frequencies were negligible for the remaining data sets. The average coverage of the contigs considered for damage analysis was between 65 and 130× but substantially lower for extraction and library blanks, 24 and 35×, respectively. Extraction and library blanks indicated in comparison to group (1) data sets weak damage signals, with discrepancies between 5pCtoT and 3pGtoA frequencies. The library blanks featured over the first five positions up to 4.2% 3pGtoA, while 5pCtoT did not exceed 1.7% (Fig. 4). We subsampled contigs affiliated with *Cand.* Rokubacteria and detected substitution frequencies between 24 and 32% (Fig. S11).

Metagenome binning

Metagenome binning led to the reconstruction of 12 bins with a completeness of at least 20% (Table S6); five of the reconstructed bins were more than 50% complete. The redundancy of the reconstructed bins was generally low and did not exceed 3.95%, while the heterogeneity reached values of up to 100% (Table S6). Nine bins were assigned to *Actinomyces*. Two of the bins belonged to the Acidiferrobacterales (one Sulfurifustaceae [H228_bin5], one Acidiferrobacteraceae [KS36MB2_bin3]). One bin was assigned to UBA9968 (Table S6). All of the bins were highly fragmented (no. of contigs > 390), and N50 values did not exceed 4 kbp. In most cases, N50 values were below 2 kbp. The relative abundance of Acidiferrobacterales based on profiling quality-controlled sequences did not exceed 0.28%. They were only detected in H22-8 and KS36-H32. We wanted to compare the two Acidiferrobacterales bins to bins recovered from the Hainich CZE groundwater [27], where this taxon is thought to be involved in sulfur cycling [29], but phylogenomic and ANI (average nucleotide identity)-based comparisons were impossible for the lack of a shared set of single-copy marker genes and the high degree of fragmentation.

Functional profiling

Taking into account that our assemblies recruited only small proportions of the quality-controlled sequences, we used the latter for functional profiling using *humann* [66]. Using only the reads mapped onto assembled contigs would provide a biased view and not adequately cover the genomic potential of the respective taxonomic groups. Between 57.3 (INF-MB2) and 85.5% (KS36-H32) (Fig. 5, "UNMAPPED") of the sequences did not yield database hits. We regrouped the output from *humann* into KEGG orthologies and summarized the normalized data (copies per million, CoPM) for KEGG pathways (Table S7) based on the sequences with database hits. We subsequently focused on pathways that differed between group (1) and group (2) data sets (Table S8), in particular functions in the context of carbon fixation, chemolithotrophy, anaerobic respiration, and aromatic hydrocarbon breakdown (Fig. 5).

Calvin cycle-related sequences were only detected for group (1) data sets (H22-8, KS36-H32) that showed pronounced DNA damage. The corresponding logCoPM values were 5.35 and 5.51, respectively. Similarly, evidence for the chemolithotrophic oxidation of sulfur and ammonia was only found in that group, with the exception of H13-17 from group (2). Evidence for nitrification was only found in H22-8. Sequences linked to the reductive TCA cycle were found in all data sets.

Sequences related to aromatic hydrocarbon breakdown were detected in all data sets, with group (1) data sets showing a broader metabolic potential to utilize these substrates, in particular H22-8 (Table S7 + Table S8). Matched sequences were affiliated with the breakdown of diverse compound classes, including among others toluene and polycyclic aromatic hydrocarbons (PAH) (Fig. 5). Group (2) data sets featured comparable narrow metabolic capabilities, including the potential for the breakdown of benzoate and related compounds (Table S7).

Discussion

We were able to recover sufficient mgDNA from all seven rock specimens for metagenomic analysis of endolithic microbial communities using protocols adapted from paleogenomics. The amounts of recovered DNA were extremely small. Extractions from rock pieces were more efficient than from powdered samples. The heat released during powdering may have led to a reduced DNA yield. Estimated cell numbers were within a narrow range of 2.81 and 4.25×10^5 cells $\times g^{-1}$ rock, independent from sampling depth and rock characteristics. The subsurface cell count database assembled by Magnabosco and colleagues [3] includes 3787 analyses, of which 2439 were linked to core samples. The database does not include cell counts from limestone but from rock material classified as carbonate from Lake Van [72] from depths between 0 and 100 mbls (meters below land surface). Our estimated cell numbers lie within the reported broad range (1.27×10^3–$4.18 \times 10^7 \times g^{-1}$ lake core material). Filling this knowledge gap is important, as carbonate rocks represent approximately 10% of the global drinking water supplies [73]. The provision of clean drinking water is considered to be the most important ecosystem service that the subsurface provides to us humans. This service

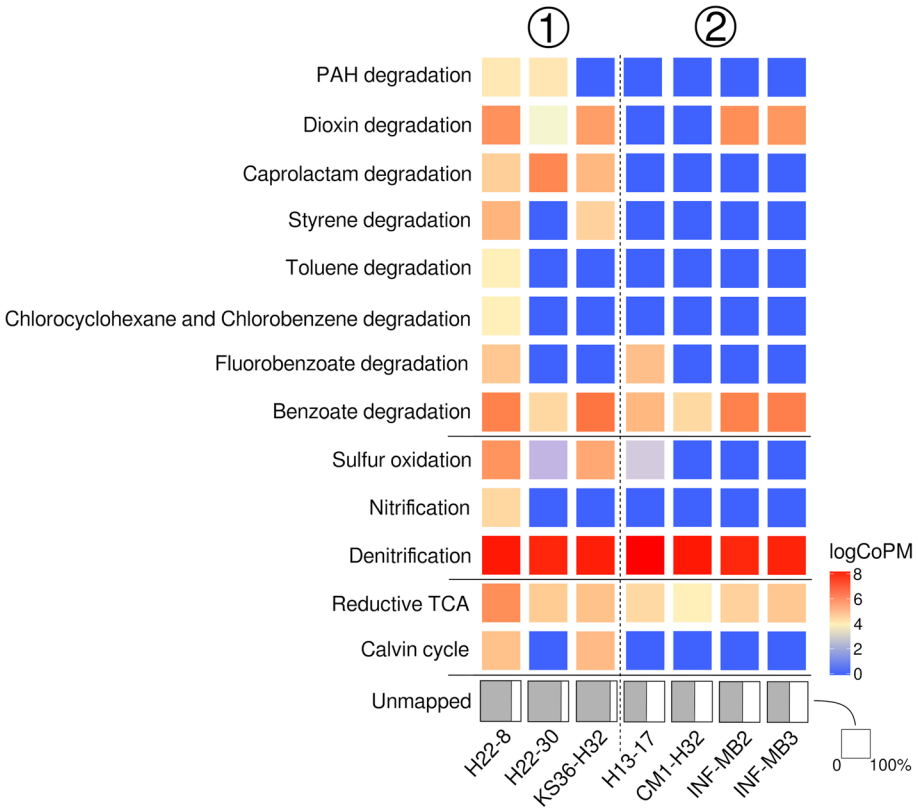

Fig. 5 Functional profiling of rock endolithic microbial communities. Profiles were generated based on output from *humann* regrouped into KEGG orthologies. KEGG orthologies were summarized based on pathways and selected functions as described in the methods. Unmapped indicates the proportion of sequences that did not yield any hits against the pre-compiled UniRef databases shipped with *humann*. logCoPM, log copies per million

is very susceptible to anthropogenic and climatic impacts [74]. Increased risks of droughts will significantly alter water availability with consequences for drinking water supplies also in Germany, which has been not affected by water stress so far [75].

The majority of the data assembled by Magnabosco and colleagues [3] were based on microscopic counts derived from surface fracture samples after desorbing cells, which might reflect the endolithic community. For obtaining microscopic counts, fluorescent stains like acridine orange or DAPI (4′,6-diamidino-2-phenylindole) are commonly used, which cannot distinguish between dead cells and those with an intact membrane, which are presumably alive. The other fraction of the cell counts was based on qPCR targeting the 16S rRNA gene [3], which is by default also not suited to differentiate between dead and live cells or extracellular DNA. A differentiation between past and potentially alive and active subsurface microbiome members provides relevant information that helps to assess the quality and potential risks associated with groundwater resources.

DNA damage pattern analysis is commonly used in the context of paleogenomics for distinguishing "modern" from degraded "ancient" DNA, which is crucial when studying prehistoric populations of humans, plants, animals, or (pathogenic) microbes [76, 77]. Determined DNA fragment sizes, DNA damage pattern analysis, and taxonomic and functional profiling and set apart the group (1) samples. The pronounced damage patterns indicate that DNA obtained from H22-8, H22-30, and KS36-H32 had undergone chemical degradation, which occurs postmortem. We are unable to determine the age of the cells of the group (1) samples.The most common forms of DNA damage are depurination, strand breakage, and cytosine deamination on single-stranded overhangs, which occur in sequence during DNA decay [78, 79]. Cytosine deamination occurs at the end of DNA fragments and can be identified by determining the frequency of 5′ cytosine to thymine transitions (3′ guanine to adenine transitions on the reverse complement strand) by mapping metagenomic sequence reads onto metagenome assemblies [68]. We detected substitution

frequencies partially above 20%, which is expected for highly degraded DNA from dead organisms [69, 77]. We cannot rule out that all these microbes were already dead when transported into rock matrix pores. It is more likely that they died after being disconnected from energy and water fluxes either in modern times due to environmental stress or over longer time scales. Environmental conditions such as low temperature, high ionic strength, pH, and protection by adsorption can delay the decay of DNA [80–82]. The different forms of crystalline carbonates present in the thin-bedded, alternating mixed carbonate-/siliciclastic bedrock of the Hainich CZE and the INFLUINS site might have favored DNA preservation through neutralizing negative charges, similar to the situation in hard tissue samples (bone, teeth, shells) [30–34]. We propose to consider the genetic records from these three samples as rock paleome signatures, signatures of past microbial communities [83].

Unlike sample materials commonly studied for paleogenomics, such as dental calculus, bones, and shells, subsurface microbial communities are not necessarily isolated due to being encased by a mineral matrix. Our -μCT analyses showed that the pore size of the rock specimens is generally large enough to allow for the transport of water along with cells. Therefore, microbial communities in any specific pore domain could constantly exchange with those in other domains and are prone to stress, e.g., groundwater fluctuations. The subsurface has to be considered as an open system, a giant biogeoreactor with constant or intermittent connection to fluid flow and matter transport, including living microbes [84]. Consequently, subsurface paleome microbiome signatures could originate from both ancient and dead modern DNA, which affects substitution rates and DNA damage patterns. A constant or regular influx of wounded or dead microbes would translate into a steady supply of utilizable resources for alive subsurface microbes, which should temporarily fuel metabolism. As a result, the DNA damage signal should get diluted, and we should have picked up "modern" DNA from alive and potentially active microbes. The DNA substitution rates detected for group (1) data sets stress the dominance of decayed DNA in these rock specimens, likely caused by temporary or spatial isolation.

We could not date the DNA due to the tiny amounts recovered. DNA in geological records is in most cases not preserved for more than 10^5 years [85–89], and 10^6 years is considered the maximum period over which DNA survival is sufficient for recovery and analysis [90]. The detected paleome signatures cannot reflect the metabolic potentials of microbes colonizing sediments about 240 million years ago, when the Upper Muschelkalk and Lower Keuper (lithostratigraphic subgroups of the Middle Triassic) were formed [91]. Our paleome signatures cannot be considered as biosignatures from ancient microbial life over geological time periods, as those identified in calcite and pyrite veins across the Precambrian Fennoscandian shield by isotopic and molecular analyses [92]. Rather, carbonate bedrocks represent DNA archives that can be used to learn more about the near biological past. We argue that distinguishing paleome from non-paleome signatures is a useful approach to identify more recent communities and their functions from those that did contribute to subsurface functioning in the past.

We are confident that the H22-8, H22-30, and KS36-H32 data sets are robust. Their taxonomic profiles differed from the laboratory blanks, and they exhibited high DNA fragmentation and higher levels of cytosine deamination than laboratory blanks, indicating that the DNA from group (1) samples disproportionately derives from dead organisms. The remaining "modern" group (2) samples did not feature any pronounced DNA damage and likely originate from alive or recently living organisms.

The paleome signatures of the group (1) samples were all obtained from vadose zone habitats in the low-mountain groundwater recharge area [17]. These shallow bedrock habitats are characterized by spatially and temporally limited water and nutrient supply via seepage from the surface, which can lead to more pronounced starvation especially in disconnected pores compared to saturated habitats. The "modern" signatures of group (2), except H13-17, were obtained from the permanently water-saturated phreatic zone of a fractured aquifer (Hainich CZE) and from ~ 300-m-deep aquitard samples (INFLUINS deep drilling) with similar matrix permeabilities, but without fracture networks [17]. The resulting isolation from the surface did not appear to be critical to the potential survival of endolithic microorganisms in the deep aquitard samples. However, our sample size is too small to conclusively explain the recovery of paleome and non-paleome signatures based on environmental factors or rock characteristics.

Endolithic microbiomes from both groups seem to rely on a bottom-up, chemolithotrophy food web driven by taxa such as *Cand.* Rokubacteria, Gemmatimonadetes, NC10, Nitrospirae, Thaumarchaeota, and Euryarchaeota. Remarkably, we found an increased abundance of chemolithoautotrophs in the paleome signatures coinciding with more detected sequences linked to carbon fixation, nitrification, and sulfur oxidation.

Metagenome assemblies were skewed towards taxa that did assemble well with consequences for DNA damage patterns. Therefore, we also carried DNA damage pattern analysis for only *Cand.* Rokubacteria contigs and could show that these contigs did feature DNA damage as well, supporting that this taxon was a member of the paleome

community. *Cand.* Rokubacteria was hypothesized to use nitrite oxidation to build a proton motive force [93]. *Cand.* Rokubacteria genomes were previously shown to contain early-branching *dsrAB* genes [94]. They possess motility genes, genes for sensor proteins for diverse stimuli, and genes for respiration (aerobic and anaerobic), fermentation, nitrogen respiration, and nitrite oxidation underline metabolic flexibility and the ability to actively move, which might favor survival in connected rock pore networks.

The phylum NC10, including *Cand. Methylomirabilis oxyfera*, is known to couple anaerobic methane oxidation to nitrite reduction [95]. Nitrospirae and Thaumarchaeota are both well known for nitrification [96, 97], including COMAMMOX in case of the former [98]. Nitrospirae are overall poorly characterized and mostly associated with nitrite oxidation. A candidate genus identified in rice paddy soil, "Candidatus Sulfobium," was associated with sulfur respiration [99]. Euryarchaeota include the majority of the known methanogens and the Methanosarcinales-related ANME (anaerobic methane oxidizing archaea) clades [100]. However, we cannot make more concrete statements about their specific role in the studied subsurface habitats.

Group (1) data sets showed a broader metabolic potential with respect to sedimentary organic carbon breakdown in the context of aromatic hydrocarbons. The use of sedimentary organic matter by pelagic groundwater microbes of the Hainich CZE was recently shown by DIC isotope pattern analyses [24, 101]. Group (1) samples also featured increased abundances of Acidobacteria. Ubiquitous in soils, Acidobacteria are characterized by a versatility relating to the utilization of (complex) carbohydrates [102] and as K-strategists [103]. Acidobacteria, Bacteroidetes, and *Cand.* Saccharibacteria are known as potential degraders for complex polysaccharides [102–105]. The latter two were more abundant in group (2) samples. These taxonomic and metabolic differences suggest a stronger adaptation of the paleome community to the harsh conditions of the endolithic habitat dominated by inorganic electron donors and CO_2 as carbon source, whereas modern communities might profit from a more constant supply of biomass rich in proteins and carbohydrates under water-saturated conditions, which could be derived from plants but also microbial biomass.

Detected endolithic Cyanobacteria, which have been more prevalent in group (1) samples, could have made use of their fermentative capabilities [106], feeding on available organic carbon, maybe preprocessed by other community members. A study targeting the Iberian Pyrite Belt Mars showed that Cyanobacteria were highly abundant, and they seemed to consume hydrogen [15]. Hydrogenotrophy might be a physiological trait in Cyanobacteria dating back to nonphotosynthetic ancestors [107]. Using mgDNA, we detected candidate phyla radiation (CPR) taxa in both groups. We previously hypothesized that CPR taxa are ideally suited to invade and colonize endolithic environments due to their small cell size [17] and their preference to be translocated with seepage water from soil into the vadose zone and finally into groundwater [108]. This would not apply to episymbiotic CPR with tight relationships with partner organisms. In the paleome, we detected increased abundances of *Cand.* Eisenbacteria, *Cand.* Jorgensenbacteria, and *Cand.* Levybacteria. *Cand.* Eisenbacteria were recently found [109] to possess a potential for secondary metabolite biosynthesis. Because of primer bias of the 16S rRNA gene [110], some CPR may have been missed in many subsurface gene surveys, similar to our previous study of endolithic bacteria from the Hainich CZE [17].

Conclusion

DNA damage patterns can be used as a proxy to distinguish DNA from intact and potentially alive cells from paleome signatures. Limestone rocks seem to represent ideal archives for genetic records of past microbial communities, including those sensitive to environmental stress at modern times, due to their specific conditions facilitating long-term DNA preservation. Neither the amount of extractable DNA nor the status of the endolithic microbiome were indicated by porosity. Water saturation, but not groundwater flow, might be key for microbial survival, as all paleome signatures were detected in the shallow vadose zone, whereas DNA obtained even from deep aquitards, isolated from surface input, did not show any DNA decay. Taxonomic and functional profiling highlighted the importance of hydrocarbon utilization and chemolithotrophy linked to sulfur cycling, the latter presumably driven by *Cand.* Rokubacteria in the paleome. Our study shows that carbonate rocks harbor microbial biomass, but that a large portion of the microbes detected by metagenomic sequencing are likely echoes of past microbial communities. The challenge for future research is now to answer the question of how old these dead cells are. Metagenomics and the distinction between "modern" and "ancient" DNA can pave the way to a deeper understanding of the subsurface geomicrobiological history and its changes over time.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40168-023-01647-2.

Additional file 1: Supplementary figures: Fig. S1. Pore size distribution of limestone samples with connected pore space (throats >26 µm). LOD = limit of detection. **Fig. S2.** Pore space characteristics of sample H22-30 determined by µCT analysis. a) Moldic pores (up to large mesopores)

dominate over fine fractures in the oolithic packstone. Scale: 0.5 mm. Plug diameter 13 mm. b) Vertical section shows porosity >26 µm. The dashed line marks the position of c. c) Horizontal section shows a moldic pore in a gastropod fossil. d) Reconstructed pore space. Colors mark parts of the pore system connected by throats >26 µm. **Fig. S3.** Pore space characteristics of sample KS36-H32 determined by µCT analysis. a) The packstone shows minimal alteration (Fe-mineral-stained spots, not visible). Scale: 0.5 mm. Plug diameter 13 mm. b) In the vertical section, connected pores up to ~26 µm (resolution limit) are assumed. The dashed line marks the position of c. c) Horizontal section showing the tight matrix that exhibits no pores connected by throats >26 µm. **Fig. S4.** Pore space characteristics of sample CM1-H32 determined by µCT analysis. a) The wackestone lacks alteration. Scale: 0.5 mm. Plug diameter 13 mm. b) In the vertical section, connected pores up to ~26 µm (resolution limit) are assumed. The dashed line marks the position of c. c) Horizontal section showing the tight matrix that exhibits no pores connected by throats >26 µm. **Fig. S5.** Pore space characteristics of sample INF-MB2 determined by µCT analysis. a) The calcareous mudstone lacks alteration. Scale: 0.5 mm. Plug diameter 13 mm. b) In the vertical section, connected pores up to ~26 µm (resolution limit) are assumed. The dashed line marks the position of c. c) Horizontal section showing the tight matrix that exhibits no pores connected by throats >26 µm. **Fig. S6.** Pore space characteristics of sample INF-MB3 determined by µCT analysis. a) The packstone/grainstone sample lacks alteration. Scale: 0.5 mm. Plug diameter 13 mm. b) Vertical section shows minor porosity >26 µm. The dashed line marks the position of c. c) Reconstructed pore space. Colors mark parts of the pore system connected by throats >26 µm. **Fig. S7.** Number of bp and sequences before and after sequence trimming visualized as a combined box plot and violin plot. Statistical significance was tested using Wilcoxon signed-rank test. **Fig. S8.** Estimated coverage of metagenome data sets based on k-mer based redundancy using nonpareil [56, 111]. The dashed red line indicates 95% coverage. **Fig. S9.** Assembly statistics for megahit (v1.2.9) [58] and metaspades (v3.13.0) [60] assemblies. **Fig. S10.** DNA fragment size distribution. Fragment sizes were deduced from 100k sampled read pairs mapped onto assembled contigs (> 1 kbp). **Fig. S11.** DNA damage pattern analysis of Cand. Rokubacteria contigs. Contigs were subsampled based on the taxonomic affiliation, which was determined with kaiju 5. Quality-controlled sequence reads were mapped onto assembled contigs (> 1 kbp). The damage pattern analysis was carried out with mapdamage (v2.2.1) 6. The plots show the substitution frequency (5pCtoT [5′ cytosine to thymine substitutions], 3pGtoA [3′ guanine to adenine substitutions]) versus the relative position (from the 5p and 3p end). n = number of contigs > 1kbp considered for the analysis, cov = mean coverage of the contigs.

Additional file 2: Supplementary tables: Table S1. Statistics sequence data processing. pwd = powdered sample, pc = rock pieces sample, QC = quality control. **Table S2.** Identified contaminants for different taxonomic ranks. Freq = frequency, prev = prevalence, p.freq = tail probability at value R, p.prev = tail probability of the chi-square distribution for the respective taxon based on presence/absence in true samples and negative controls, p = p-value from Fisher's exact test, NA = not available. Please see [112] for details regarding the mentioned metrics. **Table S3.** Phylum-level taxonomic profiles. lib_blk = library blank, ex_blank = extraction blank, pc = rock pieces sample, pwd = powdered sample. **Table S4.** Basic assembly statistics and results from read recruitment. **Table S5.** Phylum-level taxonomic profiles of assembled contigs (> 1 kbp). **Table S6.** Taxonomy and quality information regarding recovered genome bins based on checkm [64] output. **Table S7.** Functional profile based on KEGG Lvl3 Orthologies. Abundances are given as CoPM. CoPM = copies per million. **Table S8.** Functional profile based on a subset of KEGG pathways. CoPM = copies per million, logCoPM = log copies per million.

Acknowledgements
Not applicable.

Authors' contributions
CEW carried out data processing, data analysis, and wrote and revised the manuscript based on input from all co-authors. RS, supported by ZF, was responsible for rock sample processing, testing and adapting protocols, DNA extractions, and sequencing library preparation. IV and AH contributed to sequence data preprocessing, decontamination analysis, and data interpretation. RL coordinated the sampling, acquired permits, and characterized sampled rock material. TR performed µCT analysis. KUT coordinated the sampling, acquired permits, acquired funding, and contributed to data interpretation. CW conceptualized the research, contributed to data interpretation, and acquired funding. KK conceptualized the research, contributed to data interpretation, and acquired funding.

Funding
Open Access funding enabled and organized by Projekt DEAL. This work was supported financially by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy — EXC 2051 — Project-ID 390713860, the Collaborative Research Centre AquaDiva (CRC 1076 AquaDiva — Project-ID 218627073) of the Friedrich Schiller University Jena, and the Max Planck Society.

Availability of data and materials
Sequence data were deposited at the European Nucleotide Archive under BioProject number PRJEB52959 (https://www.ebi.ac.uk/ena/browser/view/PRJEB52959). Detailed information about key aspects of our data processing and analysis can be accessed at the following OSF repository: https://osf.io/v8gsd/.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 12 October 2022 Accepted: 9 August 2023
Published online: 25 September 2023

References

1. Whitman WB, Coleman DC, Wiebe WJ. Prokaryotes: the unseen majority. Proc Natl Acad Sci U S A. 1998;95:6578–83.
2. Bar-On YM, Phillips R, Milo R. The biomass distribution on Earth. Proc Natl Acad Sci U S A. 2018;115:6506–11.
3. Magnabosco C, Lin L-H, Dong H, Bomberg M, Ghiorse W, Stan-Lotter H, et al. The biomass and biodiversity of the continental subsurface. Nat Geosci. 2018;11:707–17 (Nature Publishing Group).
4. Pedersen K. Microbial life in deep granitic rock. FEMS Microbiol Rev. 1997;20:399–414.
5. Zhang G, Dong H, Jiang H, Xu Z, Eberl DD. Unique microbial community in drilling fluids from Chinese continental scientific drilling. Geomicrobiol J Taylor Francis. 2006;23:499–514.
6. Suzuki S, Ishii S, Wu A, Cheung A, Tenney A, Wanger G, et al. Microbial diversity in the Cedars, an ultrabasic, ultrareducing, and low salinity serpentinizing ecosystem. Proc Natl Acad Sci U S A. 2013;110:15336–41.
7. Momper L, Kiel Reese B, Zinke L, Wanger G, Osburn MR, Moser D, et al. Major phylum-level differences between porefluid and host rock bacterial communities in the terrestrial deep subsurface. Environ Microbiol Rep. 2017;9:501–11.
8. Purkamo L, Kietäväinen R, Nuppunen-Puputti M, Bomberg M, Cousins C. Ultradeep microbial communities at 4.4 km within crystalline bedrock: implications for habitability in a planetary context. Life [Internet]. 2020;10. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7175195/.
9. Krumholz LR, McKinley JP, Ulrich GA, Suflita JM. Confined subsurface microbial communities in Cretaceous rock. Nature. 1997;386:64–6 (Nature Publishing Group).
10. Fredrickson JK, McKinley JP, Bjornstad BN, Long PE, Ringelberg DB, White DC, et al. Pore-size constraints on the activity and survival of subsurface bacteria in a late cretaceous shale-sandstone

sequence, northwestern New Mexico. Geomicrobiol J Taylor Francis. 1997;14:183–202.
11. Krumholz LR. Microbial communities in the deep subsurface. Hydrogeol J. 2000;8:4–10.
12. Giongo A, Haag T, Medina-Silva R, Heemann R, Pereira LM, Zamberlan PM, et al. Distinct deep subsurface microbial communities in two sandstone units separated by a mudstone layer. Geosci J. 2020;24:267–74.
13. Zhang G, Dong H, Xu Z, Zhao D, Zhang C. Microbial diversity in ultra-high-pressure rocks and fluids from the Chinese continental scientific drilling project in China. Appl Environ Microbiol. 2005;71:3213–27.
14. Dutta A, Dutta Gupta S, Gupta A, Sarkar J, Roy S, Mukherjee A, et al. Exploration of deep terrestrial subsurface microbiome in Late Cretaceous Deccan traps and underlying Archean basement. India Sci Rep. 2018;8:17459.
15. Puente-Sánchez F, Arce-Rodríguez A, Oggerin M, García-Villadangos M, Moreno-Paz M, Blanco Y, et al. Viable cyanobacteria in the deep continental subsurface. Proc Natl Acad Sci U S A. 2018;115:10702–7.
16. Ben Maamar S, Aquilina L, Quaiser A, Pauwels H, Michon-Coudouel S, Vergnaud-Ayraud V, et al. Groundwater isolation governs chemistry and microbial community structure along hydrologic flowpaths. Front Microbiol. 2015;6:1457.
17. Lazar CS, Lehmann R, Stoll W, Rosenberger J, Totsche KU, Küsel K. The endolithic bacterial diversity of shallow bedrock ecosystems. Sci Total Environ. 2019;679:35–44.
18. Lehmann R, Totsche KU. Multi-directional flow dynamics shape groundwater quality in sloping bedrock strata. J Hydrol. 2020;580:124291.
19. Walter Anthony KM, Anthony P, Grosse G, Chanton J. Geologic methane seeps along boundaries of Arctic permafrost thaw and melting glaciers. Nat Geosci. 2012;5:419–26 (Nature Publishing Group).
20. Jones AA, Bennett PC. Mineral microniches control the diversity of subsurface microbial populations. Geomicrobiol J Taylor & Francis. 2014;31:246–61.
21. Kieft TL, Murphy EM, Haldeman DL, Amy PS, Bjornstad BN, McDonald EV, et al. Microbial transport, survival, and succession in a sequence of buried sediments. Microb Ecol. 1998;36:336–48.
22. Amy PS, Haldeman DL, Ringelberg D, Hall DH, Russell C. Comparison of identification systems for classification of bacteria isolated from water and endolithic habitats within the deep subsurface. Appl Environ Microbiol. 1992;58:3367–73.
23. Schwab VF, Herrmann M, Roth VN, Gleixner G, Lehmann R, Pohnert G, et al. Functional diversity of microbial communities in pristine aquifers inferred by PLFA- and sequencing-based approaches. Biogeosciences. 2017;14:2697–714.
24. Nowak ME, Schwab VF, Lazar CS, Behrendt T, Kohlhepp B, Totsche KU, et al. Carbon isotopes of dissolved inorganic carbon reflect utilization of different carbon sources by microbial communities in two limestone aquifer assemblages. Hydrol Earth Syst Sci. 2017;21:4283–300.
25. Fredrickson JK, Balkwill DL. Geomicrobial processes and biodiversity in the deep terrestrial subsurface. Geomicrobiol J Taylor Francis. 2006;23:345–56.
26. McMahon PB, Chapelle FH. Microbial production of organic acids in aquitard sediments and its role in aquifer geochemistry. Nature. 1991;349:233–5 (Nature Publishing Group).
27. Overholt WA, Trumbore S, Xu X, Bornemann TLV, Probst AJ, Krüger M, et al. Carbon fixation rates in groundwater similar to those in oligotrophic marine systems. Nat Geosci. 2022;15:561–7 (Nature Publishing Group).
28. Herrmann M, Rusznyák A, Akob DM, Schulze I, Opitz S, Totsche KU, et al. Large fractions of CO2-fixing microorganisms in pristine limestone aquifers appear to be involved in the oxidation of reduced sulfur and nitrogen compounds. Appl Environ Microbiol. 2015;81:2384–94.
29. Wegner C-E, Gaspar M, Geesink P, Herrmann M, Marz M, Küsel K. Biogeochemical regimes in shallow aquifers reflect the metabolic coupling of elements of nitrogen, sulfur and carbon. Appl Environ Microbiol. 2018. https://doi.org/10.1128/AEM.02346-18.
30. Brundin M, Figdor D, Sundqvist G, Sjögren U. DNA binding to hydroxyapatite: a potential mechanism for preservation of microbial DNA. J Endod. 2013;39:211–6.
31. Del Valle LJ, Bertran O, Chaves G, Revilla-López G, Rivas M, Casas MT, et al. DNA adsorbed on hydroxyapatite surfaces. J Mater Chem B Mater Biol Med. 2014;2:6953–66.
32. Der Sarkissian C, Pichereau V, Dupont C, Ilsøe PC, Perrigault M, Butler P, et al. Ancient DNA analysis identifies marine mollusc shells as new metagenomic archives of the past. Mol Ecol Resour. 2017;17:835–53.
33. Sullivan AP, Marciniak S, O'Dea A, Wake TA, Perry GH. Modern, archaeological, and paleontological DNA analysis of a human-harvested marine gastropod (Strombus pugilis) from Caribbean Panama. Mol Ecol Resour. 2021;21:1517–28.
34. Romanowski G, Lorenz MG, Wackernagel W. Adsorption of plasmid DNA to mineral surfaces and protection against DNase I. Appl Environ Microbiol. 1991;57:1057–61.
35. Wright VP. A revised classification of limestones. Sediment Geol. 1992;76:177–85.
36. Hennissen JAI, Hough E, Vane CH, Leng MJ, Kemp SJ, Stephenson MH. The prospectivity of a potential shale gas play: an example from the southern Pennine Basin (Central England, UK). Mar Pet Geol. 2017;86:1047–66.
37. Ahr WM, Allen D, Boyd A, Bachman HN, Ramamoorthy R. Confronting the carbonate conundrum. Oilfield Review unknown. 2005;17:18–29.
38. Philip W. Choquette (2) Lloyd C. P. Geologic nomenclature and classification of porosity in sedimentary carbonates. Am Assoc Pet Geol Bull. American Association of Petroleum Geologists AAPG/Datapages; 1970;54. Available from: http://search.datapages.com/data/doi/10.1306/5D25C98B-16C1-11D7-8645000102C1865D.
39. Buades A, Coll B, Morel J-M. A non-local algorithm for image denoising. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). 2005. p. 60–5 vol. 2.
40. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012;9:676–82.
41. Doube M, Kłosowski MM, Arganda-Carreras I, Cordelières FP, Dougherty RP, Jackson JS, et al. BoneJ: free and extensible bone image analysis in ImageJ. Bone. 2010;47:1076–9.
42. Muddiman DC, Anderson GA, Hofstadler SA, Smith RD. Length and base composition of PCR-amplified nucleic acids using mass measurements from electrospray ionization mass spectrometry. Anal Chem. 1997;69:1543–9.
43. Meyer M, Kircher M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb Protoc. 2010;2010:db.prot5448.
44. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
45. Bushnell B. BBMap short read aligner. 2016. Available from: https://www.sourceforge.net/projects/bbmap/.
46. Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun. 2016;7:1–9 (Nature Publishing Group).
47. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12:59–60.
48. Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods. 2021;18:366–8 (Nature Publishing Group).
49. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. PeerJ. 2016;4:e2584–e2584.
50. M. Burrows DJW. A block-sorting lossless data compression algorithm. Technical report 124 , Palo Alto, CADigital Equipment Corporation. 1994. Available from: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.3.8069.
51. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.
52. Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. Genome Res. 2007;17:377–86.
53. Huson DH, Beier S, Flade I, Gorska A, El-Hadidi M, Mitra S, et al. MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. PLoS Comput Biol. 2016;12:1–12.
54. Pruitt KD, Tatusova T, Brown GR, Maglott DR. NCBI reference sequences (RefSeq): current status, new features and genome annotation policy. Nucleic Acids Res. 2012;40:D130–5.

55. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome. 2018;6:226.
56. Rodriguez-R LM, Konstantinidis KT. Nonpareil: a redundancy-based approach to assess the level of coverage in metagenomic datasets. Bioinformatics. 2014;30:629–35.
57. Rodriguez-R LM, Gunturu S, Tiedje JM, Cole JR, Konstantinidis KT. Nonpareil 3: fast estimation of metagenomic coverage and sequence diversity. mSystems. 2018;3. Available from: https://doi.org/10.1128/mSystems.00039-18.
58. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2014;31:1674–6.
59. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19:455–77.
60. Nurk S, Meleshko D, Korobeynikov A, Pevzner P. metaSPAdes: a new versatile de novo metagenomics assembler. Genome Res. 2016;27:824–34.
61. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.
62. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078–9.
63. Murat Eren A, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ PeerJ Inc. 2015;3:e1319.
64. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW, Parks DH, et al. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25:1043–55.
65. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics. 2019; Available from: https://doi.org/10.1093/bioinformatics/btz848.
66. Beghini F, McIver LJ, Blanco-Míguez A, Dubois L, Asnicar F, Maharjan S, et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. Elife. 2021;10. Available from: https://doi.org/10.7554/eLife.65088.
67. Ginolhac A, Rasmussen M, Gilbert MTP, Willerslev E, Orlando L. mapDamage: testing for damage patterns in ancient DNA sequences. Bioinformatics. 2011;27:2153–5.
68. Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics. 2013;29:1682–4.
69. Borry M, Hübner A, Rohrlach AB, Warinner C. PyDamage: automated ancient damage identification and estimation for contigs in ancient DNA de novo assembly. PeerJ. 2021;9:e11845.
70. Wickham H. ggplot2. WIREs Comp Stat. 2011;3:180–5.
71. Choquette PW, Pray LC. Geologic nomenclature and classification of porosity in sedimentary carbonates. AAPG Bull GeoScienceWorld. 1970;54:207–50.
72. Kallmeyer J, Grewe S, Glombitza C, Kitte JA. Microbial abundance in lacustrine sediments: a case study from Lake Van. Turkey Int J Earth Sci. 2015;104:1667–77.
73. Stevanović Z. Karst waters in potable water supply: a global scale overview. Environ Earth Sci. 2019;78:662.
74. Gleeson T, Befus KM, Jasechko S, Luijendijk E, Cardenas MB. The global volume and distribution of modern groundwater. Nat Geosci. 2015;9:161–7 (Nature Publishing Group).
75. Wunsch A, Liesch T, Broda S. Deep learning shows declining groundwater levels in Germany until 2100 due to climate change. Nat Commun. 2022;13:1221.
76. Key FM, Posth C, Krause J, Herbig A, Bos KI. Mining metagenomic data sets for ancient DNA: recommended protocols for authentication. Trends Genet. 2017;33:508–20.
77. Warinner C, Herbig A, Mann A, Fellows Yates JA, Weiß CL, Burbano HA, et al. A robust framework for microbial archaeology. Annu Rev Genomics Hum Genet. 2017;18:321–56.
78. Briggs AW, Stenzel U, Johnson PLF, Green RE, Kelso J, Prüfer K, et al. Patterns of damage in genomic DNA sequences from a Neandertal. Proc Natl Acad Sci U S A. 2007;104:14616–21.
79. Orlando L, Allaby R, Skoglund P, Sarkissian CD, Stockhammer PW, Ávila-Arcos MC, et al. Ancient DNA analysis. Nat Rev Methods Primers. 2021;1:1–26 (Nature Publishing Group).
80. Ogram A, Sayler GS, Gustin D, Lewis RJ. DNA adsorption to soils and sediments. Environ Sci Technol. 1988;22:982–4.
81. Nielsen KM, Johnsen PJ, Bensasson D, Daffonchio D. Release and persistence of extracellular DNA in the environment. Environ Biosafety Res. 2007;6:37–53.
82. Pietramellara G, Ascher J, Borgogni F, Ceccherini MT, Guerri G, Nannipieri P. Extracellular DNA in soil and sediment: fate and ecological relevance. Biol Fertil Soils. 2009;45:219–35.
83. Inagaki F, Okada H, Tsapin AI, Nealson KH. Microbial survival: the paleome: a sedimentary genetic record of past microbial communities. Astrobiology. 2005;5:141–53.
84. Küsel K, Totsche KU, Trumbore SE, Lehmann R, Herrmann M, Steinhäuser C, et al. How deep can surface signals be traced in the critical zone? Merging biodiversity with biogeochemistry research in a Central German Muschelkalk landscape. Front Earth Sci Chin. 2016;4:1–18.
85. Linderholm A. Palaeogenetics: dirt, what is it good for? Everything Curr Biol. 2021;31(16):R993-5.
86. Edwards ME. The maturing relationship between quaternary paleoecology and ancient sedimentary DNA. Quat Res. 2020;96:39–47 (Cambridge University Press).
87. Massilani D, Morley MW, Mentzer SM, Aldeias V, Vernot B, Miller C, et al. Microstratigraphic preservation of ancient faunal and hominin DNA in Pleistocene cave sediments. Proc Natl Acad Sci U S A. 2022;119. Available from: https://doi.org/10.1073/pnas.2113666118.
88. Vernot B, Zavala EI, Gómez-Olivencia A, Jacobs Z, Slon V, Mafessoni F, et al. Unearthing Neanderthal population history using nuclear and mitochondrial DNA from cave sediments. Science. 2021;372. Available from: https://doi.org/10.1126/science.abf1667.
89. Zavala EI, Jacobs Z, Vernot B, Shunkov MV, Kozlikin MB, Derevianko AP, et al. Pleistocene sediment DNA reveals hominin and faunal turnovers at Denisova Cave. Nature. 2021;595:399–403.
90. van der Valk T, Pečnerová P, Díez-Del-Molino D, Bergström A, Oppenheimer J, Hartmann S, et al. Million-year-old DNA sheds light on the genomic history of mammoths. Nature. 2021;591:265–9.
91. Kozur HW, Bachmann GH. Correlation of the Germanic Triassic with the international scale. Albertiana. 2005;32:21–35.
92. Drake H, Roberts NMW, Reinhardt M, Whitehouse M, Ivarsson M, Karlsson A, et al. Biosignatures of ancient microbial life are present across the igneous crust of the Fennoscandian shield. Commun Earth Environ. 2021;2:1–13 (Nature Publishing Group).
93. Becraft ED, Woyke T, Jarett J, Ivanova N, Godoy-Vitorino F, Poulton N, et al. Rokubacteria: genomic giants among the uncultured bacterial phyla. Front Microbiol. 2017;8:1–12.
94. Anantharaman K, Hausmann B, Jungbluth SP, Kantor RS, Lavy A, Warren LA, et al. Expanded diversity of microbial groups that shape the dissimilatory sulfur cycle. ISME J. 2018;12:1715–28.
95. Haroon MF, Hu S, Shi Y, Imelfort M, Keller J, Hugenholtz P, et al. Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage. Nature. 2013;500:567–70 (Nature Publishing Group).
96. Pester M, Schleper C, Wagner M. The Thaumarchaeota: an emerging view of their phylogeny and ecophysiology. Curr Opin Microbiol. 2011;14:300–6.
97. Offre P, Spang A, Schleper C. Archaea in biogeochemical cycles. Annu Rev Microbiol. 2013;67:437–57.
98. Koch H, van Kessel MAHJ, Lücker S. Complete nitrification: insights into the ecophysiology of comammox Nitrospira. Appl Microbiol Biotechnol. 2018; Available from: https://doi.org/10.1007/s00253-018-9486-3.
99. Zecchin S, Mueller RC, Seifert J, Stingl U, Anantharaman K, von Bergen M, et al. Rice paddy Nitrospirae carry and express genes related to sulfate respiration: proposal of the new genus "Candidatus Sulfobium." Appl Environ Microbiol. 2018;84. Available from: https://doi.org/10.1128/AEM.02224-17.
100. Knittel K, Boetius A. Anaerobic oxidation of methane: progress with an unknown process. Annu Rev Microbiol. 2009;63:311–34.
101. Schwab VF, Nowak ME, Elder CD, Trumbore SE, Xu X, Gleixner G, et al. 14C-free carbon is a major contributor to cellular biomass in

geochemically distinct groundwater of shallow sedimentary bedrock aquifers. Water Resour Res. 2019;55:2104–21.
102. Eichorst SA, Trojan D, Roux S, Herbold C, Rattei T, Woebken D. Genomic insights into the Acidobacteria reveal strategies for their success in terrestrial environments. Environ Microbiol. 2018;20:1041–63.
103. Kielak AM, Barreto CC, Kowalchuk GA, van Veen JA, Kuramae EE. The ecology of acidobacteria: moving beyond genes and genomes. Front Microbiol. 2016;7:1–16.
104. Grondin JM, Tamura K, Déjean G, Abbott DW, Brumer H. Polysaccharide utilization loci: fuelling microbial communities. J Bacteriol. 2017;199:JB.00860-16.
105. Kindaichi T, Yamaoka S, Uehara R, Ozaki N, Ohashi A, Albertsen M, et al. Phylogenetic diversity and ecophysiology of candidate phylum Saccharibacteria in activated sludge. FEMS Microbiol Ecol. 2016;92:fiw078.
106. Stal LJ, Moezelaar R. Fermentation in cyanobacteria1Publication 2274 of the Centre of Estuarine and Coastal Ecology, Yerseke, the Netherlands.1. FEMS Microbiol Rev. 1997;21:179–211.
107. Mulkidjanian AY, Koonin EV, Makarova KS, Mekhedov SL, Sorokin A, Wolf YI, et al. The cyanobacterial genome core and the origin of photosynthesis. Proc Natl Acad Sci U S A. 2006;103:13126–31.
108. Herrmann M, Wegner C-E, Taubert M, Geesink P, Lehmann K, Yan L, et al. Predominance of Cand. Patescibacteria in groundwater is caused by their preferential mobilization from soils and flourishing under oligotrophic conditions. Front Microbiol. 2019;10:1407.
109. Sharrar AM, Crits-Christoph A, Méheust R, Diamond S, Starr EP, Banfield JF. Bacterial secondary metabolite biosynthetic potential in soil varies with phylum, depth, and vegetation type. MBio. 2020;11. Available from: https://doi.org/10.1128/mBio.00416-20.
110. Brown CT, Hug LA, Thomas BC, Sharon I, Castelle CJ, Singh A, et al. Unusual biology across a group comprising more than 15% of domain bacteria. Nature. 2015;523:208–11.
111. Rodriguez-R LM, Gunturu S, Tiedje JM, Cole JR, Konstantinidis KT. Nonpareil 3: fast estimation of metagenomic coverage and sequence diversity ABSTRACT. mSystems. 2018;3(3). https://doi.org/10.1128/mSystems.00039-18.
112. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome. 2018;6(1). https://doi.org/10.1186/s40168-018-0605-2.

Publisher's Note