This is a "preproof" accepted article for *Geo-Bio Interfaces*.
This version may be subject to change during the production process.
DOI: 10.1180/gbi.2025.10004

Microbial community composition and function in groundwater systems of the Deccan Traps

Avishek Dutta[1,2,3], Sourav Mukhopadhyay[1], Debarshi Mukherjee[1], Rajendra Prasad Sahu[1], Srimanti Duttagupta[2], Abhishek Gupta[1], Himadri Bose[1], Abhijit Mukherjee[4], and Pinaki Sar[1*]

[1]Environmental Microbiology and Genomics Laboratory, Department of Bioscience and Biotechnology, Indian Institute of Technology Kharagpur, Kharagpur, 721302, India

[2]Department of Geology, University of Georgia, Athens, GA 30602, USA

[3]Savannah River Ecology Laboratory, University of Georgia, Aiken, SC 29802, USA

[4]Department of Geology and Geophysics, Indian Institute of Technology Kharagpur, Kharagpur, 721302, India

*Corresponding author:

Pinaki Sar, Department of Bioscience and Biotechnology, Indian Institute of Technology Kharagpur, Kharagpur, 721302, WB, India; Email: psar@bt.iitkgp.ac.in, Phone: +913222 283754, Fax: +913222 255303

Abstract:

Microbial diversity and functions within the Earth's deep subsurface remain pivotal in the Earth's major biogeochemical activities. Microbial communities of groundwater systems hosted by ~65-million-year-old Deccan basalts are investigated to delineate their characteristics, biogeochemical functions, and environmental control. Quantitative PCR-based bacterial cell counts suggest 4.3×10^2 to 3.9×10^3 cells/mL. 16S rRNA gene sequence analysis shows considerable bacterial diversity and the existence of a core microbiome [16 amplicon sequence variants (ASVs) out of a total of 2,020 ASVs] across the groundwater samples. Members of *Burkholderiaceae* and *Moraxellaceae* are predominant taxa within the groundwater. In comparison, the spring water and surface water microbiomes are significantly

distinct. Non-metric multidimensional scaling highlights that the basaltic groundwater communities are influenced by local environmental parameters. Analysis of whole metagenomes indicates that the Calvin-Benson-Bassham cycle (CBB cycle) is a primary mode of C-fixation in the subsurface water system of the Deccan traps. Metagenome-assembled genomes are affiliated to the genera *Limnohabitans* and *Methylotenera*, among others. Together with the presence of sulfate and nitrate in the groundwater environment, the presence of genes involved in dissimilatory nitrate and sulfate reduction indicates the prevalence of anaerobic/facultative anaerobic lifestyles among the microorganisms in this system. Amplicon and whole metagenome sequence-based analyses suggest the presence of microbial populations involved in local biogeochemical cycling. This study on the geomicrobiology of the water systems of Deccan traps elucidates microbial community composition and biogeochemical function in the igneous rock-hosted deep biosphere.

Keywords

Groundwater, Deccan Traps, Microbial community, Metagenome-assembled genomes; Chemoautotrophy, N- and S- cycling

1. Introduction

Microbial diversity and function within the deep Earth subsurface remain pivotal in understanding the major biogeochemical activities of our planet. Despite significant research efforts on microbial life that resides beneath the Earth's surface, the deep biosphere continues to represent one of the “least understood environments till date” (Lopez-Fernandez et al., 2018; McMahon and Parnell, 2014). Although a broad knowledge base on microbial diversity and their patterns of distribution is created through several continental and ocean drilling initiatives covering varied geological settings and ages, a number of fundamental questions on deep life remain poorly resolved. Major questions pertaining to the sustenance of life at greater depths fairly concentrate on two major aspects: (a) Who inhabits the deep subterranean environment, and what putative metabolisms support their sustenance in such an aphotic, oligotrophic, multi-extreme habitat? and, (b) How do geochemical parameters (i.e., various electron acceptors, electron donors, etc.) influence microbial metabolisms and, thereby, shape the microbial community structure in deep environments? (Amils et al., 2023; Atencio et al., 2023; Bell et al., 2020; Bomberg et al., 2015; Herzig et al., 2024; Kadnikov et al., 2020; Kieft, 2016; Lollar et al., 2019; Momper et al., 2023; Osburn et al., 2014; Pedersen et al., 2014; Purkamo et al., 2015; Simkus et al., 2016). Deep subterranean habitats are diverse with respect to geochemical

parameters and subsequent microbial colonization; consequently, the answers to the questions are subjected to the habitat specifications (Amils et al., 2023; Atencio et al., 2023; Bell et al., 2018, 2020; Hernsdorf et al., 2017; Ijiri et al., 2018; Ino et al., 2016; Kadnikov et al., 2020; Momper et al., 2023).

Although different deep biosphere habitats manifest different features, these subterranean environments are often characterized by oligotrophy, absence of sunlight, and photosynthetically derived organic matter, where chemolithoautotrophic organisms are thought to be the primary producers (Pedersen, 2000). Previous studies suggest that geogenic and/or biogenic hydrogen (H_2) facilitates microbial metabolism within these energy-starved subsurface environments (Amils et al., 2023; Hernsdorf et al., 2017; Ijiri et al., 2018; Nealson et al., 2005; Nyyssönen et al., 2014; Stevens, 1997; Stevens and McKinley, 2000). Under the natural conditions that prevailed in the deep Earth crust, geogenic hydrogen may be produced through serpentinization, radiolysis of water, degassing of magma-hosted systems, etc. or biologically via fermentation, anaerobic methane oxidation, etc. (Beaver and Neufeld, 2024; Gregory et al., 2019; Nealson et al., 2005). Along with hydrogen metabolism, the interplay of dynamic carbon, nitrogen, and sulfur cycles with the involvement of diverse microbial groups make the overall system a complex paradigm (Hernsdorf et al., 2017; Momper et al., 2023). Complex processes such as autotrophic carbon fixation with hydrogen/sulfur oxidation (as a source of electrons/energy) coupled with nitrate reduction are reported to be one of the putative metabolisms plausible in anoxic deep biosphere (Lau et al., 2016). Oxidation of methane and other reduced inorganic molecules, such as ammonia and nitrite, and reduction of sulfate or iron (Fe^{3+}) or manganese (Mn^{4+}) as terminal electron acceptors while assimilating carbon via auto/hetero/mixotrophic processes are also frequently observed in these environments (Casar et al., 2021; Dutta et al., 2018; Momper et al., 2017; Purkamo et al., 2013; Vaccarelli et al., 2021). Interestingly, heterotrophy, fueled by the metabolic intermediates or products of chemolithoautotrophic metabolism, is also detected in many such environments (Purkamo et al., 2015; Sahu et al., 2022). Deciphering the biogeochemical processes constituted by the complex networks of metabolic transformation by the resident microbial members of any particular deep subsurface environment is identified as one of the major approaches for understanding the different modes of sustenance of life, such as the energetic edge (Ino et al., 2016; X. Wu et al., 2016).

A number of deep biosphere studies are conducted in the subterranean water systems across the continents (Hallbeck and Pedersen, 2008; Herzig et al., 2024; Itävaara et al., 2011; Momper

et al., 2017; Nyyssönen et al., 2014; Purkamo et al., 2016; Sahl et al., 2008; X. Wu et al., 2016). One of the first studies that gave an overview of microbial diversity as well as its function from a genomic point of view was conducted in deep crystalline rock environments of the Fennoscandian shield (Nyyssönen et al., 2014). Taxonomically and functionally diverse populations are observed among different deep subsurface horizons of the Fennoscandian shield, which vary in response to the prevailing lithology and hydrochemistry. Subsequent studies within other deep crystalline rock and groundwater environments gave an extensive overview of the microbial processes that govern the terrestrial subsurface environments (Amils et al., 2023; Bell et al., 2020; Momper et al., 2017; Probst et al., 2018; X. Wu et al., 2016). Many of these studies employed genome reconstruction from metagenomes to get better insights into individual genomes of microbes in a microbial community of deep terrestrial subsurface (Atencio et al., 2023; Bell et al., 2018, 2020; Coskun et al., 2024; Hernsdorf et al., 2017; Kadnikov et al., 2020; Momper et al., 2023; Thieringer et al., 2023). Although past research has provided important information microbial communities inhabiting various continental subsurface environments, however the progressively hot and seismically active igneous province of deep continental biosphere within Deccan Traps, is comparatively underexplored. Albeit limited, studies initiated by our laboratory (Dutta et al., 2018; Dutta, Peoples, et al., 2019; Dutta, Sar, et al., 2019), have provided some insights into the endolithic microbial life and their genomic potential. Nevertheless, nature and function of deep biosphere within the aquifers hosted by the Deccan Traps, and comparison of such microbial life to the rock hosted one remains uncharacterized unlike other subsurface environments.

The volcanic province of the deep subsurface of the Deccan Traps is considered to be an unusual extreme environment for life with progressive increase in temperature and lithostatic pressure (at a rate of ~15 °C/km and ~26 MPa/km, respectively), low organic carbon (< 50 mg TOC/kg) and presence of multiple heavy metals (Dutta et al., 2018; Goswami et al., 2024; Roy & Rao, 2000; Shukla et al., 2022). In the present study, microorganisms inhabiting the aquifers deep-seated within the one of the largest continental flood basalt, i.e., the Deccan Traps, and their interaction with surrounding rock matrix is investigated through a combination of geochemistry, metataxonomy, metagenomics followed by genome reconstruction techniques. Aquifer inhabiting planktonic and rock inhabiting endolithic microorganisms has been investigated to identify the core microorganisms. The knowledge base developed through this study, can be used for further gathering information on global subsurface microbial abundance, diversity and metabolism.

The Deccan Traps in western India (also known as Deccan Volcanic Province) represents one of the largest continental flood-basalt provinces with an approximate area coverage of > 0.5 million km^2 and a total thickness of >2 km near the eruptive center (Bhaskar Rao et al., 2017; Schoene et al., 2015). The ~65 My old Deccan basalts are underlain directly by granitic basement rocks of Precambrian age (~2700 Ma) with a thin intermediate, weathered zone. Apart from its geologic importance, this volcanic province has gained considerable scientific attention owing to the well-known Reservoir Triggered Seismicity within the Koyna Seismogenic Zone (KSZ) (Roy, 2017). The KSZ has experienced recurrent seismic activity during the past five and half decades, starting soon after the impoundment of the Shivajisagar (Koyna) water reservoir on the Koyna River in 1962 (Gupta et al., 2015). Geological mapping combined with geophysical and geochemical studies carried out in the past provide crucial information about the subsurface fault-fracture systems in this seismogenic zone (Goswami et al., 2019, 2020, 2024). However, the characteristics of microbial life that exist in the deep biosphere of this igneous province remain unexplored, mostly due to difficulty in accessing the deep subsurface samples. Recent scientific drilling (up to several kilometers below the surface) enabled geomicrobiologists to explore deep life hosted by the igneous crust using the rock cores recovered. These studies conducted on the basaltic and granitic rocks provided evidence of bacterial and archaeal life within this extreme realm, including the characterization of piezophilic bacteria and various chemolithotrophic bacterial populations thriving in the deep biosphere (Dutta et al., 2018; Dutta, Peoples, et al., 2019; Dutta, Sar, et al., 2019; Mandal et al., 2022; Saha et al., 2024; Sahu et al., 2022). In comparison to these, microbial life within the deep groundwater of this igneous province remained uncharted, until the aquifer samples were available for geomicrobiological analysis, post establishment of the deep boreholes. Therefore, the present study is conducted to investigate the microbial community structure and function of the groundwater systems of Deccan traps, identify the local geochemical factors likely to control the microbial communities, and ascertain if the groundwater system represents a distinct ecological niche as compared to the surface water system of this igneous province. Overall, the study aims to provide a better understanding of the nature and function of deep microbial life as guided by the physicochemical factors of the Deccan Traps.

2. Materials and Methods

2.1. Sample collection

Water samples were collected from three groundwater sources (bore wells), one natural spring, and two surface water sources (Koyna River and Koyna Reservoir) of the Koyna-Warna region

(Figure S1, Table 1). The sites used for water sampling were : (i) Koyna pilot borehole KFD1 in Gothane, drilled at the center of a 2 km × 2 km seismic cluster in the northcentral part of the Koyna seismogenic zone, in the proximity of Donichawadi fault (sample obtained from a depth of 1027m, referred to as DBW in this manuscript), (ii) a tube well adjacent to the borehole KBH-5 in Phansavale, located in the vicinity of a seismic cluster in the southwestern fringes of the Koyna seismogenic zone, adjoining the western face of the Western Ghats escarpment (Depth: 100 m, referred to as BW in this manuscript), and (iii) groundwater well near Karad, about 40 km to the east of the Koyna seismogenic zone (Depth: 100 m, referred to as BGRL in this manuscript). For groundwater sampling, water from each of the wells was flushed out via continuous pumping (for 15-30 min) till the dissolved oxygen level reached a steady state (as monitored by the Thermo Orion multi-parameter probe). The spring water was sampled from the water tank, which received the slow-flowing water from the spring. For the surface water, multiple sub-samples were collected from the Koyna River and Koyna Reservoir. From each site, two to four-liter water was collected in sterile containers and stored at 4 °C for shipment. The samples were stored at 4 °C in the laboratory until further analysis.

2.2. Hydrochemical analysis

Measurement of major elements was conducted using ICP-MS (iCAP Q, Thermo Scientific) and Cavity Ring-Down Spectroscopy (CRDS), L2130i Picarro was used to measure stable isotopes of oxygen ($\delta^{18}O$) and hydrogen ($\delta^{2}H$). Major anions (Cl^-, NO_2^-, SO_4^{2-}, NO_3^-, Br^-, and PO_4^{3-}) were measured using Dionex ICS 2100 (Thermo Scientific). Total organic carbon (TOC), total inorganic carbon (TIC), and total carbon (TC) were measured on an OI analytical TOC analyzer. The alkalinity of the water samples was measured using USEPA method 310.2. Measurement of pH, dissolved oxygen (DO), and conductivity was done using highly sensitive probes (Orion) fitted with an Orion multimeter (Thermoelectron Corporation, Beverly, MA). For measurements of organic acids, 20 µL sample was injected into the Ultimate High-Performance Liquid Chromatography (Dionex® 3000 U-HPLC, ThermoFisher Scientific). C18- reverse-phase analytical column was used with a mobile phase consisting of a binary mixture of solvents (a) water (30%) and (b) acetonitrile (70%). The separation was performed at room temperature, and the flow rate was maintained at 1 mL/min. The compounds were monitored at the following wavelengths: 254 nm, 265 nm, 286 nm, and 295 nm. The calibration curves consisted of five standard solutions ranging from 0.1 to 50 mg/L. Chromeleon 6.8 software was used for data acquisition and integration.

2.3. Extraction of DNA

For each sample, 1000 mL of water was filtered using a sterile 0.22 μm filter, and total DNA was extracted from the filters using a Metagenomic DNA Isolation kit (Epicentre) following the manufacturer's protocol. The quality of the extracted DNA and its concentration were measured using a NanoDrop 2000 spectrophotometer, followed by fluorometric quantitation using Qubit (Thermo-Fisher Scientific).

2.4. Quantitative polymerase chain reaction (qPCR)

Quantification of the copy number of structural (bacterial and archaeal specific 16S rRNA gene) and functional gene (*dsrB*) has been done through qPCR following the methods reported in Dutta et al. (2018), (Dutta et al., 2018). Details of qPCR primers used are provided in Table S1. All the reactions were set in triplicate. Bacterial 16S rRNA gene copy numbers were determined in each sample by comparing the amplification result to a standard dilution series ranging from 10^2 to 10^{10} of plasmid DNA containing the 16S rRNA gene of *Achromobacter* sp. MTCC 12117. Genes encoding archaeal 16S rRNA and *dsrB* were PCR amplified from the metagenome, cloned into a TA cloning vector, and plasmid DNAs for each gene targets with copy numbers 10^2 to 10^{10} were used as standards for quantitation purposes. Quantitative PCR has been performed in Quant Studio 5 with Power SYBR green PCR master-mix (Invitrogen), with a primer concentration of 5 pM. Melting curve analysis was run after each assay to check PCR specificity.

2.5. 16S rRNA gene amplification and sequencing

Following the quality check, metagenomes were subjected to amplification of the V4 region of the 16S rRNA gene using primers 515F/806R (Bates et al., 2011) (Table S2). Upon purification of the PCR products, further sequencing was done on the Ion S5 platform (Thermo Fish er Scientific). Refer to Table S2 for a detailed procedure. The sequenced reads were submitted to the short read archive under BioProject ID PRJNA445925.

2.6. Amplicon data analysis and contamination removal

Raw data obtained from Ion S5 were analysed through QIIME 2 amplicon-2024.10 (Bolyen et al., 2019) bioinformatics pipeline. Default parameters were used with each module, unless specified otherwise. Along with the reads from the samples, a reagent control (Accession No.:) was also included in the amplicon data analysis. This control sample was created by subjecting nuclease-free water to the same DNA extraction and amplicon sequencing procedure as the samples. All datasets were quality filtered, clustered, denoised, and chimera filtered using DADA2 with the parameters “denoise-single --p-trunc-len 270 --p-trim-left 30 --p-pooling-

method pseudo --p-min-fold-parent-over-abundance 2" (Callahan et al., 2016). The ASVs (Amplicon Sequence Variants) obtained were assigned taxonomy using "feature-classifier classify-sklearn" module with the naive Bayes classifier "GG-2024.09.backbone.v4.nb.sklearn-1.4.2" pre-trained on Greengenes2 2024.09 (which is in turn is based on GTDB r220) (Bokulich et al., 2018; McDonald et al., 2024; Pedregosa et al., 2011). Based on recommendations by Sheik et al. (2018) and techniques used by Purkamo et al. (2020), contamination removal was performed by removing all ASVs unassigned to domains, and those affiliated to Eukarya, mitochondria, chloroplast, *Streptococcaceae*, *Propionibacteriaceae*, *Lactobacillaceae*, *Escherichia-Shigella*, and *Staphylococcaceae*. The reagent control was used to run "quality-control" modules with "decontam-identify" and "decontam-remove" parameters (Davis et al., 2018). Finally, singletons (ASVs with a total of only one read across samples) were removed as suspected contaminants, and the retained ASVs were used for determining relative abundance and core ASVs. Diversity parameters were determined by rarifying all samples to 180,000 reads and then using the "diversity alpha" module.

2.7. Statistical analysis and phylogenetic tree construction

The relative abundance of core ASVs of the groundwater samples (BGRL, BW, DBW) was determined at the genus level, and used to calculate Spearman's rs correlation with the geochemical characteristics of the respective samples. Euclidean distances between the correlations were used to plot the UPGMA (Unweighted Pair Group Matrix Analysis) dendrogram. These statistical analyses were performed with PAST v4.12b (Hammer et al., 2001). Non-metric multidimensional scaling (NMDS) based on Bray-Curtis dissimilarity matrix was calculated for the microbial communities at family level. Families having greater than 1% percentage average abundance across all the samples were considered for this analysis. The function envfit was used to fit environmental vectors to the ordination plot with environmental variables using the vegan package (Oksanen et al., 2025) in R. NMDS analysis of microbial community structure was conducted using R and R Studio. Ten of the ASVs with the highest relative abundance (>2% on average) in groundwater samples were subjected to phylogenetic analysis. National Center for Biotechnology Information (NCBI) databases were queried with the blastn suite to find the closest matches for these ASVs from extreme environments. The ASVs were grouped by their order and aligned with ClustalX2 (Larkin et al., 2007). A neighbour-joining tree was created from the alignment using MEGA 11, with the bootstrap method (n=1,000), Jukes-Cantor substitution model, gamma distribution (G=1.00)

rates among sites, and pairwise deletion treatment (Tamura et al., 2021). The phylogenetic tree was visualized with the ggtree v3.12.0 R package (Xu et al., 2022).

2.8. Whole metagenome sequencing and analysis

Metagenomes of three water samples (two groundwater samples and one surface water sample) were considered for whole metagenome sequencing. Following metagenomic extraction, DNA was enzymatically digested, and size selection was made using the E-Gel electrophoresis system (Invitrogen, ThermoFisher Scientific). After size selection, DNA has been ligated with Ion Xpress barcode adapters and pooled together for emulsion PCR using the Ion OneTouch™ 2 System (ThermoFisher Scientific). The PCR product has been sequenced using the Ion S5 platform. Quality checking of the trimmed paired-end reads was conducted using FastQC (Version 0.11.7). Trimmed sequences were assembled using SPAdes assembler (Bankevich et al., 2012) using default parameters. Quast 4.3 was used for evaluations of metagenomic assemblies and different assembly statistics (Gurevich et al., 2013). Best k-mer was selected on the basis of longest contigs and N50 value for different samples and was annotated using Genomes OnLine Database (GOLD) v.6 (Mukherjee et al., 2017). KEGG-based annotation was selected for further analysis. The annotated data are available under IMG GOLD Analysis Project ID Ga0247434 (PBW), Ga0247435 (DBW), and Ga0247436 (BGRL). Heatmaps were created to compare metagenomic inventories of rock and water samples with respect to carbon, sulfur, and nitrogen cycles using the pheatmap package in R (Kolde and Kolde, 2015). The annotated data for rocks were based on the analysis reported by Dutta et al., 2018(Dutta et al., 2018).

For reconstructing metagenome-assembled genomes, single-end metagenome sequencing raw reads obtained were reverse complemented in silico with seqtk-v1.4 (r122) (H. Li, 2016). Read quality filtering, human contamination removal, assembly, binning, bin refinement, and reassembly were performed with MetaWRAP v1.3.2 pipeline and the tools therein using default settings unless specified otherwise (Uritskiy et al., 2018). Filtered reads were concatenated for co-assembly with MEGAHIT v1.2.9 (D. Li et al., 2015). Binning was performed in single-end, “universal” (bacteria and archaea) mode using MetaBAT 2 v2.12.1, MaxBin 2.0 v2.2.7 and CONCOCT v1.1.0 (Alneberg et al., 2014; Kang et al., 2019; Y.-W. Wu et al., 2016). Bin refinement was performed using MetaWRAP’s bin_refinement module with a minimum completion threshold of 50% and maximum contamination of 10%. The resulting bins were used to determine relative abundance in terms of genome copies per million filtered reads using the quant_bins module, and then refined bins were reassembled using identical

thresholds. The samples were clustered via the UPGMA (Unweighted Pair Group Method with Arithmetic mean) algorithm using the Bray-Curtis similarity index and a bootstrap of 1000, with the help of PAST v4.12b (Bray and Curtis, 1957; Hammer et al., 2001). Completeness of the bins based on the presence of lineage-specific marker genes was calculated using CheckM v1.0.18, and quality was inferred based on standards delineated by Bowers et al. (2017) (Bowers et al., 2017; Parks et al., 2015). Taxonomy assignment was performed using GTDB-Tk v2.4.0 (database r220) (Chaumeil et al., 2020). Functional annotation was performed in metagenome mode using eggNOG-mapper v2.1.12 with eggNOG DB v5.0.2, employing Prodigal v2.6.3 for gene prediction while searching via MMseqs2 v15.6f452, with default settings (Cantalapiedra et al., 2021; Huerta-Cepas et al., 2019; Hyatt et al., 2010; Steinegger and Söding, 2017). DRAM 1.4.0 was used for functional annotation of MAGs, assessment of pathway completeness, and categorizing different microbial metabolisms (Shaffer et al., 2020).

3. Results

3.1. Hydrochemical characteristics of water samples

The hydrochemical characteristics of the water samples were analyzed to ascertain the local physicochemical conditions available to the inhabiting microbial communities (Table 1). Groundwater samples were found to be slightly alkaline in nature (pH 7.8 - 8.2), having low dissolved oxygen (1.7 - 4.7 mg/L) and organic carbon (1.9 - 2.2 mg/L). The inorganic carbon content of these samples, however, was relatively higher (3.0 - 7.1 mg/L), particularly compared to their surface water counterpart. Diverse concentrations of the anions (Cl^- 9.9 - 595.1 mg/L, NO_2^- 0.6 - 4.6 mg/L, NO_3^- 0 - 61 mg/L, and PO_4^{3-} 0 - 52 mg/L) and elements (Ca 10 -32.1 mg/L, Fe 106.4 - 177.8 mg/L, Mg 5.8 - 27.8 mg/L, K 9.3 - 16.0 mg/L and Na 10.4 - 17.2 mg/L) were detected. An overall higher concentration of anions and elements was detected in groundwater samples than in surface water. However, the concentration of Fe was found to be higher in surface water. Among the groundwater samples, DBW has shown relatively higher concentrations of Cl^-, $SO_4^{2-,}$ and NO_3^-. The values of $\delta^{18}O$ and δ^2H in the samples showed variation but indicated their meteoric origin (Das et al., 2020; Hosono et al., 2020). The DBW sample showed the lowest values of $\delta^{18}O$ and δ^2H. Non-metric multidimensional scaling (NMDS) based on the geochemical characteristics of the samples showed distinct patterns of the partitioning of groundwater and surface water samples (Figure S2). It was highlighted that the groundwater ecosystems of the Deccan Traps were mainly constrained by pH, dissolved organic carbon (TOC) and total inorganic carbon (TIC), Mg, Ca, NO_3^-, Fe, and SO_4^{2-}. Among the organic acids tested, oxalic and lactic acids were detected at relatively higher concentrations

(oxalic acid 0.12 - 0.20 mg/L and lactic acid 0.03 - 0.20 mg/L), followed by acetic acid (0.01 - 0.05 mg/L) and succinic acid (0.02 - 0.10 mg/L). Principal coordinate analysis (PCoA) based on major hydrochemical parameters, along with organic acids detected in the samples, also showed the distinctive nature of groundwater compared to its surface water counterpart (Figure S3).

3.2. Quantitative estimation of microbial cells and dissimilatory sulfite reductase gene (*dsrB*)

Quantitative PCR-based estimation of bacterial 16S rRNA and *dsrB* genes was possible in all but two samples (KD and BW). The estimated bacterial cell counts, as obtained through qPCR, showed 4.3×10^2 to 3.9×10^3 cells/mL (Table 1) [assuming an average of 4.2 16S rRNA gene copies/genome of bacteria (Stoddard et al., 2015)]. Archaeal 16S rRNA gene was detected only in the BGRL, showing a presence of around 8.5×10^2 cells/L. Gene fragments of the *dsrB* gene were detected in the same four samples whose bacterial 16S rRNA copy number was estimated. Deep subsurface water samples (BGRL and DBW) yielded 1.08×10^3 - 3.34×10^3 copies/L, while the surface water sample KR showed a relatively lower abundance of *dsrB* (1.48×10^2 copies/L).

3.3. 16S rRNA gene sequence data and alpha diversity analysis

After quality filtering, a total number of 939,455 and 691,998 reads were obtained from the groundwater and surface water samples, respectively. These quality-filtered reads were grouped into 2,020 and 1,596 ASVs for groundwater and surface water, respectively (Table S3). The average numbers of filtered reads and ASV s obtained per groundwater sample were 313,152 and 705, respectively, while 345,999 reads and 847 ASVs were obtained for surface water samples on average. Among the groundwater samples, the highest number of ASVs was obtained for BGRL (1,466 ASVs). Most of the ASVs were affiliated to bacteria (>98.86% for all samples) and a small fraction (<1.14 % for all samples) to archaea. The Shannon and the Simpson indices indicated that microbial communities of the groundwater were relatively less diverse and uneven compared to their surface water counterpart. Interestingly, only 16 out of 2,020 ASVs were shared among the three samples (Table S3). These shared ASVs representing the core community of the groundwater covered 4.69%-36.95%, with the highest in the sample from the deepest sample (DBW, 1027 mbs).

3.4. Microbial community analysis

Taxonomic affiliation showed the presence of 61 bacterial and four archaeal phyla. Pseudomonadota was the most abundant phylum across the samples (Fig. 1a). Based on

average abundance, Pseudomonadota accounted for 88.41-98.93 % of the groundwater communities, whereas, in the surface water, it covered only 30.08% (KD) and 35.64 % (KR). However, spring water (PBW) showed an abundance of Proteobacteria (65.69%) and Deinococcota (33.34%). The rest of the major populations constituted by Actinomycetota, Deinococcota, Bacteroidota, Planctomycetota, Verrucomicrobiota, and Cyanobacteriota in the bacterial domain and Thermoproteota and Nanoarchaeota in the archaeal domain. Upon examining KD and KR, it was interesting to note that, Actinomycetota, Bacteroidota, Planctomycetota, and Verrucomicrobiota corresponded to significantly higher proportions in these surface water samples in addition to Pseudomonadota.

Among the most abundant families, *Burkholderiaceae*_A_595421 was detected as the predominant one across the three deep groundwater samples (13.93-58.65 %), followed by PIVM01 (in class Gammaproteobacteria, 0-43.46%) *Moraxellaceae* (2.87-12.49 %) and *Nevskiaceae* (0.02-26.51 %) (Fig. 1b, Table S4). Among the other major taxa in groundwater samples, *Sphingomonadaceae*_486827, *Methylophilaceae*, *Pseudomonadaceae*, along with *Burkholderiaceae*_A_566089, *Minwuiaceae*, and *Caulobacteraceae*, were detected in fewer groundwater samples. Surface water microbial communities (for sample KD and KR) were dominated by *Nanopelagicaceae* (1 2.69-21.75 %), *Burkholderiaceae*_A_595421 (12.37-14.57%), *Pseudomonadaceae* (0.009-17.57 %), UBA1268 (under order Pirellulales, 15.39-8.33 %), *Chitinophagaceae* (5.21- 7.68%), and *Cyanobiaceae* (1.86- 10.56%). The spring water community was found to be quite interesting, as it was mainly composed of members of only three families, each present in nearly equal proportions [*Burkholderiaceae*_A_595421 (36.15%), *Deinococcaceae* (33.34%), Pseudomonadaceae (28.56%)] and together covering 98. 05%**.**

Among the taxa which could be identified at the genera level PIVM01, *Hydrocarboniphaga*, *Methylotenera*_A_557637, *Novosphingobium*, *Fluviicoccus*, *Pseudomonas*_E_648040, *Comamonas*_F_589250, *Sphingorhabdus*_C (*Nevskiaceae*), *Acinetobacter*, and *Limnobacter* were most prominent in the groundwater. *Cavicella*, *Pararheinheimera*, *Caulobacter*_487784 etc., were the other prominent genera present in these samples. In comparison to the groundwater, the surface water samples showed a characteristic abundance of unclassified *Nanopelagicus*, *Rhodoferax*_C, *Prosthecobacter*, *Aquirufa*_904070, *Caulobacter*_487784, *Priestia*_289346, *Limnohabitans*_A, and *Parasediminibacterium*. *Deinococcus*_B, *Pseudomonas*_E_647464, and *Janthinobacterium*_571526 members were the major genus (covering 79.77% of the community) represented in the spring water community. Phylogenetic

analysis of the predominant ASVs from the major orders showing high sequence identity and resemblance with organisms of similar taxonomic groups reported from deep subsurface, igneous crust, or volcanic or other extreme environments corroborated their strong ecological, physiological, and metabolic attributes (Figure S4).

Nonmetric multidimensional scaling was used to describe the distribution of different samples according to the relative abundance of microbial taxa in relation to geochemical parameters (Fig. 2a). The figure highlighted the geochemical variability among the samples and demonstrated their relative control on the microbial community composition. A set of environmental parameters, especially pH, NO_2^-, PO_4^{3-}, Ca, K, Mg, explained their control in shaping groundwater community mostly through regulating the abundance of *Burkholderiaceae*_A_595421, *Nevskiaceae*, *Sphingomonadaceae*_486827, PIVM01 (in class Gammaproteobacteria), *Methylophiliaceae*, and *Moraxellaceae*. Relatively higher TOC, DO, HCO_3^- and conductivity segregated the surface water samples and the spring water with major contributions from *Spirosomataceae, Verrucomicrobiaceae, Cyanobiaceae, Chitinophagaceae, Flavobacteriaceae,* and *Bacillaceae*_H_289398.

3.5. Analysis of the core community of the groundwater

Sixteen ASVs of the groundwater community constituted the core microbiome (shared ASV s across three groundwater samples) (Table S6). Identification of these compositional core members was considered to be imperative in terms of providing the foundation for understanding the unique contribution of dominant, rare, and common taxonomic groups (Shade and Handelsman, 2012). Taxonomic analysis of the shared ASVs shows their affiliation with three classes and eight families. A maximum number of the core- ASVs (11 of 16 ASVs) were affiliated with Gammaproteobacteria (Major family: *Burkholderiaceae, Moraxellaceae*), followed by Alphaproteobacteria (4 of 16 ASVs, *Sphingomonadaceae* and *Caulobacteraceae*) and Bacteroidia (one ASV in the core microbiome, *Chitinophagaceae*).

The distribution patterns of the core ASVs (at class level) correlated with the geochemical parameters, and they clustered in five groups mainly according to the geochemistry of the water samples (Fig. 2b). *Hydrocarboniphaga*, *Aquabacterium*_A_591266, and *Phenylobacterium* (members of Group 1) were positively correlated with TOC. In contrast, Limnobacter and *Sediminibacterium* (members of Group 5) were negatively correlated with TOC. Group 2 members (*Acinetobacter*, uncultured *Burkholderiaceae*_A_595421, *Sphingorhabdus*_C, *Cavicella*, and *Ramlibacter*_588642) were positively correlated with pH, whereas Group 3

(*Caulobacter*_487784, *Novosphingobium*, u_*Moraxellaceae*) members were negatively correlated with pH. *Pseudomonas*_E_648040 (member of Group 4) was found to be positively correlated with SO_4^{2-}, Cl-, Na, K, HCO_3^-, PO_4^{3-}, and NO_3^-. Members of group 2 were negatively correlated with Conductivity, total carbon, dissolved oxygen (DO), and total inorganic carbon (TIC), whereas members of group 3 were positively correlated with the same environmental variables.

3.6. Analysis of metagenome-assembled genomes (MAGs)

Whole metagenome data from two groundwater (BGRL and DBW) and the spring water (PBW) yielded six high-quality (>90% completeness and <5% contamination) and ten medium-quality (>50% completeness and <10% contamination) bins (Table S7). Quality thresholds suggested by Bowers et al. (Bowers et al., 2017) were followed. These genomic bins were used for detailed annotation and analysis (Fig. 3 and Table S8). The selected medium and high-quality draft bins showed affiliation to diverse bacterial taxa spanning the families *Alteromonadaceae* (genus *Pararheinheimera*), *Burkholderiaceae* (*Rhodoferax*, *Janthinobacterium*, *Undibacterium*, and *Limnohabitans*), *Deinococcaceae* (*Deinococcus*), *Methylophilaceae* (*Methylotenera*), *Moraxellaceae* (*Acinetobacter*, *Fluviicoccus*, *Perlucidibaca*, and JAJNDL01), *Nevskiaceae* (*Hydrocarboniphaga*) and *Sphingomonadaceae* (*Sphingorhabdus* and *Novosphingobium*). Among these, all but *Deinococcus* were members of phylum Pseudomonadota (formerly Proteobacteria). Of the Pseudomonadota, 13 bins were of class Gammaproteobacteria and 2 of Alphaproteobacteria.

Abundances of the bins across the samples were compared, and it was found that all bins were present in all the samples, though their abundance varied considerably, from <1 to 491 GPM (genome copies per million filtered reads), thus suggesting a ubiquity of these members. The most abundant bins in BGRL were bin nos. 9, 3, 15, 12, and 11, affiliated to the genera *Limnohabitans*, *Methylotenera*, *Fluviicoccus*, *Deinococcus* and *Limnohabitans*, respectively. All of which were present in the other samples in low to moderate abundances. PBW was dominated by bin nos. 6, 1, 12, 2, and 11 (*Rhodoferax*, *Janthinobacterium*, *Deinococcus*, *Undibacterium*, and *Limnohabitans*, respectively) of which bin nos. 11 and 12 are shared with BGRL. While bin nos. 4, 7, 14, 10, and 5 (*Hydrocarboniphaga*, *Moraxellaceae* genus JAJNDL01, *Methylotenera*, *Pararheinheimera*, and *Novosphingobium*, respectively), were predominant in DBW, all of which showed low abundance in the other two samples. UPGMA clustering of metagenome samples on the basis of bin abundances using the Bray-Curtis

dissimilarity showed that both groundwater (BGRL and DBW) could be grouped together, while the spring water (PBW) grouped separately (Fig. 3c).

3.7. Metabolic potentials of MAGs

Analyses of the genes related to carbon metabolism provided a broad overview of the genomic potential for C-fixation and its utilization via catabolic pathways (Fig. 3 and Table S8). Genes encoding inorganic C fixation (HCO_3^-) by six pathways were present in the reconstructed genomes. The most complete pathway for C-fixation was the Calvin-Benson-Bassham cycle (CBB cycle), followed by the reductive tricarboxylic acid cycle (rTCA or Arnon-Buchanan cycle), the dicarboxylate-hydroxybutyrate cycle (DC/4-HB cycle), the Wood-Ljungdahl pathway (WLP), the 3-hydroxypropionate bi-cycle (3-HP bi-cycle), and the 3-hydroxypropionate/4-hydroxybutyrate cycle (3-HP/4-HB cycle). Among these, genes for rTCA and DC/4-HB cycle were present in all the bins, though pathway completeness was low. The observed lack of completeness could be due to the limitation of the sequencing technique instead of the actual absence of genes, and it might be possible that at least some members of these microbial communities are capable of lithoautotrophic metabolism. Other carbon metabolism pathways that were detected include the hexose monophosphate shunt (HMP shunt or pentose phosphate pathway), tricarboxylic acid cycle (TCA cycle), and Entner-Doudoroff pathway (ED pathway), which were complete in some of the bins. Since the presence of complete pathways was seen in bins with higher completeness, it is likely that the absence of genes is due to bin incompleteness. Incomplete genes were detected for the Embden-Meyerhof-Parnas pathway (EMP pathway), glyoxylate cycle, and methanogenesis.

The bins were further assessed for the abundance of genes associated with biogeochemical cycles and stress tolerance (Fig. 3 and Table S9). Based on the incompleteness of the reconstructed bins and their pathways, the absence of genes was not regarded as evidence of a complete lack of function. Furthermore, the presence of genes was compared in terms of the abundance of gene copies to determine the preponderance of specific metabolic functions. All bins showed the presence of nitrate reduction genes, while oxidative genes were not detected. Specifically, nitrate, nitrite, and nitric oxide reductases were detected. Similarly, sulphur reduction gene copies outweighed oxidative genes. However, iron oxidation genes were detected from most bins. Genes conferring tolerance against radiation, thermal, and osmotic stresses were detected from all bins, but desiccation tolerance genes were not detected from most bins.

3.8. Metagenomic comparison among the rock and groundwater systems

Comparing the metagenomic data sets revealed differences in the predominant carbon and energy metabolism pathways between groundwater and crustal (rock) microbiome (Fig. 4), as previously published (Dutta et al., 2018). With respect to C-fixation pathways, CBB was found to be the most abundant pathway across the groundwater system. Genes for WLP (*acsB*) were found to be the highest in one of the groundwater samples (BGRL) compared to other rocks and water samples. Genes responsible for methylotrophy were not detected in the water samples, whereas genes related to methylotrophy were detected in basalts and transition zone rock samples. Interestingly, genes responsible for methanogenesis (*mcrA*) were ubiquitous in the rocks and were absent in the groundwater ecosystem.

With respect to nitrogen and sulphur metabolism, groundwater and rock systems showed stark differences. The gene encoding for nitrogenase (*nifH*; plays an important role in nitrogen fixation), was ubiquitous in the rocks but only observed in one of the water samples (DBW). However, genes related to the conversion of nitrate to nitrite (both cytoplasmic [*nar*] and periplasmic [*nap*] nitrate reductase) were present in higher abundances in the water samples compared to the rocks. Genes involved in sulfate respiration (*apr* and *dsr*) were found to be most abundant in one of the groundwater samples (BGRL) compared to other rocks and water samples. Interestingly, genes involved in dissimilatory sulfate reduction were not observed in any other water samples except for BGRL. In the sulfur cycle, major distinctions across water and rock samples were observed with respect to sulfur oxidation and sulfate transport. Genes involved in sulfur oxidation and sulfate transport were higher in water samples than in rock samples.

4. Discussion

4.1. Geochemical characteristics and microbial abundance

Our geomicrobiological investigation clearly demonstrated a strong stratification of microbial community composition and function within the groundwater and surface water systems. Results indicated that different environmental factors, including nutrient sources present in a particular habitat, shaped microbial ecosystems. Deep subsurface aquifers hosted by crystalline igneous crusts were mostly isolated from surface activities and devoid of surface-derived carbon and energy sources (Kieft, 2016). The oligotrophic, often extreme ecosystems were predominantly driven by chemolithoautotrophy utilizing locally available inorganic electron donors and energy sources. Microorganisms occupying these habitats were found to be

endowed with abilities for utilizing the local resources as nutrient substrates (Beaver and Neufeld, 2024; Kazy et al., 2021; Kieft, 2016; Momper et al., 2023).

Geochemical data obtained in this study presented the characteristic features of the igneous rock-hosted groundwater of the Deccan Traps, indicating the possible role of these local factors in impacting the inhabitant microbial communities. Among the major parameters tested, groundwater ecosystems of the Deccan Traps were found to be constrained by pH, organic carbon (TOC), inorganic carbon (TIC), Mg^{2+}, Ca^{2+}, NO_3^-, $Fe^{3+,}$ and SO_4^{2-}. In comparison to various other continental subsurface water (*e.g.*, Fennoscandian Shield, Precambrian bedrock of Witwatersrand basin, etc.,) groundwater (Bell et al., 2018; Lau et al., 2016; Magnabosco et al., 2016; Nyyssönen et al., 2014; Purkamo et al., 2016) from the Deccan Traps basaltic crust was found to be deprived of organic carbon considerably (3.5-6.4 fold lower than others), and corroborated well with the TOC level of deep basaltic rocks of the studied region reported earlier (Dutta et al., 2018). The moderately alkaline nature of the groundwater of the Deccan Traps, along with elevated levels of Ca and Mg, were in line with the abundance of aluminum silicate or carbonate-rich minerals in the host rocks, including Ca/Mg bearing Mg-phyllosilicates, Mg-carbonates, Clinopyroxene, calcic plagioclase, etc. (Dutta et al., 2018). The alkalinity and abundance of Ca and Mg seemed to be more like a characteristic chemical property of igneous crust hosted deep water as also reported earlier in multiple studies (Itävaara et al., 2011; Kieft et al., 2005; Magnabosco et al., 2016; Nyyssönen et al., 2012; Purkamo et al., 2016). Distinctly higher concentrations of Mg, Ca, NO_3^-, and PO_4^{3-} in two groundwater (BGRL and DBW) and the spring water (PBW) could be inferred as an outcome of subsurface rock-water interactions and different active biogeochemical processes, sourced by breakdown/weathering of various rock minerals affecting the water chemistry. Low bacterial abundance (up to 10^4 cells/ml) of the groundwater compared to other igneous crust-hosted water systems of comparable depths portrayed another characteristic property, which could be linked with the overall thermodynamic and nutritional constraints (including lack of readily metabolizable C and energy sources, reduced level of electron acceptors, etc.) of the igneous crust hosted groundwater (Kieft et al., 2005; Purkamo et al., 2017; Sahl et al., 2008; X. Wu et al., 2016).

4.2. Microbial ecology of the water systems

Microbial communities retrieved from different parts of the water system of the igneous province of the Deccan Traps were characterized as having low microbial abundance, dominated by bacteria with limited species richness and abundance. Microbial communities in

the groundwater were less diverse than the microbial communities of the surface water. Among the samples from the deep subsurface, BGRL showed higher abundances of archaeal presence compared to other groundwater samples (more archaeal ASVs corroborate with qPCR-based detection of archaea-specific 16S rRNA reads). In general, the predominance of bacteria over archaea in deep aquifer ecosystems has been reported to be a characteristic phenomenon (Kadnikov et al., 2018). Nevertheless, a few archaea-based investigations (including the one from the Deccan Traps) indicated their distinct presence and highlighted biogeochemically relevant functions coupling Fe and S oxidation and C metabolism (Dutta, Sar, et al., 2019; Probst et al., 2018; Purkamo et al., 2016; X. Wu et al., 2016). One of the most interesting findings was the prevalence of 16 ASVs as members of the core populations of this subsurface habitat; however, their total relative abundances varied across different groundwater samples (BGRL: 4.69%, BW: 12.17%, DBW: 36.95%). The presence of a few ASVs as core communities of deep igneous crust ecosystems has been previously observed and identified to be of central importance in deep biosphere community function (Boada et al., 2021; Sahu et al., 2022; Yan et al., 2020).

Our 16S rRNA gene-based community profiling provided an overview of the aquatic microbial ecosystems of the Deccan Traps. The presence of Pseudomonadota as the single most abundant phylum in groundwater samples, with its members known to be capable of performing various important biogeochemical transformations (of C, N, S, etc.), necessary for providing essential metabolic resources to the inhabitant microorganisms, was noticed. The predominance of Pseudomonadota (formerly Proteobacteria) corroborated previous reports on the cosmopolitan distribution of these bacteria in various subsurface water systems (groundwater, fracture-fault water, etc.) of Colorado, the Fennoscandian Shield, and the Witwatersrand basin (Lau et al., 2016; Nyyssönen et al., 2014; Onstott et al., 2003; Purkamo et al., 2016; Sahl et al., 2008). The abundance of Pseudomonadota members also highlighted their metabolic versatility and major environmental importance within the deep biosphere ecosystems. Niche-specific distinct community composition across the subsurface environment (compared to their surface counterparts) indicated the role of local geochemical factors. Family-level community data, as well as the core ASV-based correlation analyses, confirmed the patterns of community assemblages and their relation with geochemical parameters. Higher values of pH, NO_2^-, NO_3^-, and PO_4^{3-} influenced the abundance of *Burkholderiaceae*_A_595421, *Nevskiaceae*, *Sphingomonadaceae*_486827, PIVM01 (in class Gammaproteobacteria), *Methylophiliaceae*, and *Moraxellaceae*. The distinct patterns of microbial assemblages in groundwater (with the

dominance of *Moraxellaceae*, *Methylophilaceae,* and *Sphingomonadaceae* along with *Burkholderiaceae* and *Nevskiaceae*) could be inferred as their characteristic lineages with deep aquatic ecosystems, as also evidenced by several investigations on other subterranean water samples (Itävaara et al., 2011; Jangir et al., 2019; Nyyssönen et al., 2014; Purkamo et al., 2016). Particularly, *Moraxellaceae*, *Burkholderiaceae*, etc., could be designated as resident bacterial taxa of the deep biosphere in crystalline groundwater.

Further analysis of the metabolic properties of various taxa detected in the deep subsurface of the Deccan Traps portrayed the broad functional abilities and their biogeochemical significance. One of the unifying themes that emerged was the primary role of the chemolithoautotrophic mode of metabolism, along with nitrate reduction. In line with the observations made by Deja-Sikora *et al.* (Deja-Sikora et al., 2019), our taxonomic data clearly indicated that inorganic carbon fixation coupled with sulfur/iron oxidation and nitrate reduction constituted the community's energy-circuit under the prevailing (relatively) nitrate-rich anoxic condition. Other energy metabolism pathways (e.g., H_2 or Fe^{2+} oxidation) could also be involved. Among the predominant bacterial taxa, *Acinetobacter* and *Methylotenera* were detected. Oxidation of sulfur compounds and denitrification have been well reported in other major taxa of *Moraxellaceae* (*Acinetobacter*) and *Methylophilaceae* (*Methylotenera*) (Y. Liu et al., 2011; Song et al., 2008; Su et al., 2017). Utilization of C1 compounds (e.g. methane, methanol, formate, etc.), other hydrocarbons (alkanes), and biomineralization (of carbonates) under oligotrophic settings by these taxa were very relevant as these compounds can be generated in deep igneous systems through geogenic or biogenic synthesis (Chen et al., 2021; Y. Liu et al., 2011). The presence of hydrocarbon degrading capability of subsurface microbial communities is further supported by the presence of *Novosphingobium, Hydrocarboniphaga,* and *Sphingorhabdus* in the groundwater samples (Jeong et al., 2016; H. Liu et al., 2016; Segura et al., 2017).

In contrast to the prevalence of the chemolithoautotrophic organisms in the deep groundwater, surface water ecosystems could be characterized through heterotrophic bacterial activities. The most abundant taxa therein, namely, *Nanopelagicaceae, Pseudomonadaceae,* UBA1268 (affiliated to Planctomycetota), *Chitinophagaceae,* etc., were reported earlier to occupy freshwater and eutrophic reservoirs, water column/streams, and wetlands (Aizenberg-Gershtein et al., 2012; Andrei et al., 2019; Hosen et al., 2017; Neuenschwander et al., 2018; Newton and McLellan, 2015; Serra Moncadas et al., 2024; Wang et al., 2025). The presence of cyanobacterial members supported their role in providing photosynthetically fixed carbon

that aids the growth of heterotrophic and chemoorganotrophic microbial groups. The spring water (PBW) microbiome displayed the presence of extreme stress-tolerant *Deinococcaceae* (*Deinococcus*), which corroborates well with our recent findings on similar organisms in deep granitic crusts underneath the Deccan Traps (Sahu et al., 2022).

Core communities have been considered to represent the stable fraction of a community. Identifying these members and their geochemical drivers was considered to be imperative in gaining insights into the foundation (Shade and Handelsman, 2012; Yan et al., 2020; Sahu et al., 2022). Earlier studies on the endolithic microbial communities inhabiting the Archean granitoid basement underneath the Deccan Traps, India, identified the presence of microbial communities whose abundances varied with the prevailing lithology and geochemistry (Sahu et al., 2022). The variability of the abundances of core ASVs across different deep groundwater samples further indicated niche partitioning of the subsurface water system. Correlation analysis of these core ASVs with groundwater hydrochemistry alluded to presence of both autotrophy and heterotrophy in the groundwater environment. In an oligotrophic deep subsurface environment, chemoautotrophic CO_2 fixation, with energy derived through oxidation of inorganic substrate coupled with NO_3 reduction, might be a plausible metabolic route to drive in the deep biosphere (Lau et al., 2017). The products of autotrophy could be utilized by different heterotrophs in the community. The presence of autotrophy could be further validated by the fact that major taxonomic members of the groundwater samples have manifested similar metabolisms in previous reports. Iron or sulfur oxidation coupled nitrate reduction-driven autotrophic carbon fixation has been reported in deep groundwater settings (Lau et al., 2016; Lopez-Fernandez et al., 2018).

4.3. Comparison of metagenomic bins and assessing the biogeochemical potential

A comparison of bin abundances across the metagenome samples revealed that groundwater communities were significantly different from the spring water. This difference could be attributed to the distinct physicochemical conditions that prevailed in these two habitats. Since most bins were present in all the samples, inter-mixing of ground and spring water would be highly likely, allowing the presence of a few common microbial species. Nevertheless, the observed differences in the abundance of individual bins probably arise because each sample forms a microenvironment that allows the selection of a particular section of the total community.

Upon comparing the taxonomic lineages of the bins from metagenome data with 16S rRNA amplicon sequencing results, members of Pseudomonadota (formerly phylum Proteobacteria)

remained the most prominent phylum, with representatives of class Gammaproteobacteria and Alphaproteobacteria. Similarly, metagenome data also corroborated the status of Deinococcota (formerly phylum Deinococcus-Thermus) as the next most abundant phylum. The overall genome copies per million filtered reads were highest for BGRL, then PBW, and DBW. This trend did not match what was seen in amplicon data, but the differences could be explained in light of practical limitations of sequencing technique rather than actual absence. Major families of groundwater samples as detected from whole metagenome analysis (*Comamonadaceae*, *Moraxellaceae*, *Methylophilaceae*, *Nevskiaceae*, *Sphingomonadaceae*) were detected from amplicon analysis as well. On the other hand, preponderant surface water families (*Sporichthyaceae, Chitinophagaceae, Verrucomicrobioaceae, Acidimicrobiaceae, and Flavobacteriaceae*) were not represented in the genomes reconstructed from metagenome analysis. The taxonomy of reconstructed genomes thus corroborated the 16S rRNA gene amplicon-based analysis. *Burkholderiaceae* and *Alteromonadaceae* were low or sparse in amplicon data but had metagenomic bins affiliated to them, highlighting the superiority of whole metagenome-based community analysis. At a lower taxonomic level, metagenome data retained multiple genera (*viz., Limnohabitans, Perlucidibaca, Acinetobacter, Hydrocarboniphaga, Methylotenera*) also detected in 16S rRNA gene amplicon sequencing, thus confirming their presence in the groundwater and spring water. Major taxa detected as core community members of the groundwater microbiome were represented by the reconstructed genomes as well, further validating our observations of the subsurface aquifer microbial communities.

Functional annotation of reconstructed genomes demonstrated a prevalence of chemoautotrophic CO_2 fixation pathways. The presence of chemoautotrophy is to be expected, considering the aphotic subsurface ecosystems of the deep biosphere have a minimal supply of photosynthetically produced organic carbon. The low organic carbon status of the igneous crust-hosted aquifers necessitates chemoautotrophy as the only metabolic route for primary production in sub-terrestrial food webs (Overholt et al., 2022; Templeton and Caro, 2023). The specific pathways that are employed for carbon fixation vary across habitats, but CBB and rTCA cycles are frequently reported from groundwater studies, alongside WLP (Atencio et al., 2023; Y. Li et al., 2024; Momper et al., 2017; Overholt et al., 2022; Probst et al., 2018; X. Wu et al., 2016). Pathways involved in fermentation were found to be complete and abundant. Notably, it has been previously reported that CO_2 fixation pathways like the CBB cycle can be used by mixotrophic bacteria in conjunction with sugar fermentation pathways to utilize the

CO_2 liberated, which minimizes loss of organic carbon. Recent studies have shown that organic carbon lean deep terrestrial subsurface microbial communities have predominance of multiple heterotrophic (glycolysis, respiratory and fermentation) pathways (along with chemolithoautotrophy) responsible for facilitating oxidation of locally available small and complex (even recalcitrant) organic compounds of both autochthonous and allochthonous types (Atencio et al., 2023; Coskun et al., 2024; Ijiri et al., 2018; Momper et al., 2023; Shimizu et al., 2015; Valentin-Alvarado et al., 2024). Our previous studies on oligotrophic, deep (up to 3000 mbs) crystalline igneous rock-hosted microbial communities within and below the Deccan Traps also showed the presence of both chemoautotrophic and heterotrophic pathways, including acetate metabolizing and other fermentative metabolisms (Dutta et al., 2018; Mandal et al., 2022; Sahu et al., 2022). Based on these reports and results obtained in the present study, we infer that deep aquifer microbiomes of the Deccan Traps are capable of an overall mixotrophic mode of metabolism, allowing complete cycling of carbon within this subterranean ecosystem through autotrophic as well as heterotrophic pathways.

The exploration of genes associated with redox reactions of inorganic compounds was able to support only a few of the metabolic functions expected from the assessment of the microbial community, as inferred from amplicon sequence analysis. But this may just be due to inadequate metagenome sampling depth. In spite of this limitation, the presence of a near-complete reductive half of the nitrogen cycle, represented by nitrite reductase, nitrate reductase, and nitric oxide reductase, was metagenomically confirmed. The information gathered from metagenome analysis can be regarded as a preliminary overview of the preferred choice of electron acceptors and donors for these microbial communities, implicating sulphate and oxides of nitrogen as potential electron acceptors. Such prevalence of nitrate and sulfate reducers and iron oxidizers have been reported from other studies of groundwater microbial communities as well, suggesting that these bacteria are key players in subterranean mineral mobilization and geochemical cycling (Atencio et al., 2023; Y. Li et al., 2024; Momper et al., 2017; Overholt et al., 2022; Probst et al., 2018; X. Wu et al., 2016). Since amplicon and metagenome-based studies only allow predictions of such functions, in vitro validation of microbial capacity to carry out the various biogeochemical reactions would be the conclusive proof that can be sought in future studies into the microbial diversity of this region. In addition, the depth of sequencing might also impact the detection of certain genes that are not abundant in the samples.

When stress tolerance genes were investigated, a paucity of desiccation tolerance genes was observed. This may be because the microflora was adapted for growth in water and not frequently exposed to water scarcity, or due to sequence incompleteness. However, the presence of radiation tolerance genes could confer desiccation stress resistance since radiation tolerance genes tend to be associated with desiccation stress resistance as well (Khan et al., 2024). Thermotolerance genes detected may arise from deeper groundwater samples as temperature increases with greater depths (Bregnard et al., 2023). Osmotolerance could be relevant to near-surface and deep subsurface microorganisms as a sudden influx of meteoric water or dissolution of rock minerals can cause sharp and/or localized changes in osmotic pressure (Or et al., 2007; Sokol et al., 2022).

Though differences in microbial communities and their functions were observed across subsurface rock (Dutta et al., 2018) and water systems in the Deccan traps, their foundations were very similar. These similarities in the bioenergetics of the system were likely due to the result of an exchange of resident microflora taking place during rock-water interactions and the prevalence of similar conditions in the subsurface igneous province. Such interactions were regarded as a major driver of nutrient cycling in the oligotrophic deep biosphere. Rocks provide the mineral solutes that are solubilized in water to serve as nutrients and energy sources for the geochemically active microorganisms inhabiting these niches (Meyer-Dombard and Malas, 2022). The overall data represent preferential recruitment of major metabolism across the aquatic and crustal provinces of the Deccan Traps.

Acknowledgements

This study was supported by the Ministry of Earth Sciences (MoES), Government of India, Project ID:MoES/P.O.(Seismo)/1(181)/2013 dated April 7, 2014, for which we are grateful to the Secretary, MoES. We would like to extend our gratitude and special thanks to Dr. Sukanta Roy for his immense help and support for field sampling and for guiding us throughout this study. We thank the Director and all the participants and investigators from CSIR-National Geophysical Research Institute, Hyderabad, engaged in the KBH-5 scientific drilling operations in the Koyna-Warna region of Deccan traps. We are grateful to Harsh Gupta (NGRI), Shailesh Nayak (MoES) for unstinted support. Special thanks to Dr Arun Gupta (MoES) and his office for their consistent administrative support. We thank Ajoy Roy, Balaram Mohapatra, and Surajit Misra for their help and discussion during the field trip. Next generation sequencing facility created at PS's laboratory was funded by the Indian Institute of Technology Kharagpur Challenge Grant (IIT/SRIC/BT/ODM/2015-16/141). AD and SDG gratefully

acknowledge the fellowship provided by IIT Kharagpur. SM gratefully acknowledges the fellowship provided by the Ministry of Earth Sciences (MoES), Government of India. RPS and DM gratefully acknowledge the financial assistance provided by the Council of Scientific and Industrial Research, Government of India. AG thanks the Department of Biotechnology, Government of India for providing fellowship under DBT-JRF category. HB gratefully acknowledges the fellowship provided by the Science and Engineering Board, Department of Science and Technology (SERB-DST) Government of India. This work used the high-performance computing facility, Param Shakti, of IIT Kharagpur established under National Supercomputing Mission (NSM), Government of India, and supported by Centre for Development of Advanced Computing (CDAC), Pune, India.

Author Contribution

PS: Conceptualization (Lead), Data curation (Lead), Formal analysis (Lead), Funding acquisition (Lead), Investigation (Lead), Methodology (Lead), Project administration (Lead), Resources (Lead), Software (Lead), Supervision (Lead), Validation (Lead), Visualization (Lead), Writing – original draft (Lead), Writing – review & editing (Lead), **AD**: Conceptualization (Equal), Data curation (Lead), Formal analysis (Lead), Investigation (Lead), Methodology (Lead), Software (Lead), Validation (Lead), Visualization (Lead), Writing – original draft (Lead), Writing – review & editing (Lead); **SM:** Formal analysis (Supporting), Investigation (Supporting), Methodology (Supporting), Software (Equal), Validation (Equal), Visualization (Equal), Writing – original draft (Equal), Writing – review & editing (Equal); **DM:** Formal analysis (Supporting), whole metagenome sequence analysis (Equal), re-analysis of 16S rRNA amplicon data for validation (Supporting), Writing – review & editing (Supporting); **RPS:** Data curation (Supporting), Formal analysis (Supporting), Methodology (Supporting), Software (Supporting), Writing – original draft (Supporting), Writing – review & editing (Supporting); **SDG:** Data curation (Equal), Formal analysis (Equal), Investigation (Supporting), Methodology (Equal), Software (Equal), Writing – original draft (Supporting), Writing – review & editing (Supporting); **AG:** Data curation (Equal), Formal analysis (Equal), Investigation (Supporting), Methodology (Equal), Software (Equal), Writing – original draft (Supporting), Writing – review & editing (Supporting); **HB:** Formal analysis (Supporting), Software (Supporting), Visualization (Supporting), Writing – original draft (Supporting), Writing – review & editing (Supporting); **AM:** Formal analysis (Supporting), Investigation (Supporting), Methodology (Supporting), Resources (Equal), Supervision (Supporting), Validation (Supporting), Writing – review & editing (Supporting);

Conflict of interest

The authors declare no conflict of interests

Data availability

The raw 16S rRNA sequencing data is available in the NCBI SRA database under BioProject ID PRJNA445925, and whole metagenome sequence data is available at IMG GOLD Analysis Project ID IMG GOLD Analysis Project ID Ga0247434, Ga0247435, and Ga0247436.

References

Aizenberg-Gershtein, Y., Vaizel-Ohayon, D., and Halpern, M. (2012). Structure of bacterial communities in diverse freshwater habitats. *Canadian Journal of Microbiology*, **58(3)**, 326–335. https://doi.org/10.1139/W11-138

Alneberg, J., Bjarnason, B. S., De Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., Lahti, L., Loman, N. J., Andersson, A. F., and Quince, C. (2014). Binning metagenomic contigs by coverage and composition. *Nature Methods*, **11(11)**, 1144–1146. https://doi.org/10.1038/nmeth.3103

Amils, R., Escudero, C., Oggerin, M., Puente Sánchez, F., Arce Rodríguez, A., Fernández Remolar, D., Rodríguez, N., García Villadangos, M., Sanz, J. L., Briones, C., Sánchez-Román, M., Gómez, F., Leandro, T., Moreno-Paz, M., Prieto-Ballesteros, O., Molina, A., Tornos, F., Sánchez-Andrea, I., Timmis, K., … Parro, V. (2023). Coupled C, H, N, S and Fe biogeochemical cycles operating in the continental deep subsurface of the Iberian Pyrite Belt. *Environmental Microbiology*, **25(2)**, 428–453. https://doi.org/10.1111/1462-2920.16291

Andrei, A. Ş., Salcher, M. M., Mehrshad, M., Rychtecký, P., Znachor, P., and Ghai, R. (2019). Niche-directed evolution modulates genome architecture in freshwater Planctomycetes. *ISME Journal*, **13(4)**, 1056–1071. https://doi.org/10.1038/s41396-018-0332-5

Atencio, B., Geisler, E., Rubin-blum, M., Bar-zeev, E., Adar, E. M., Ram, R., Ronen, Z., and Oceanographic, I. (2023). Metabolic adaptations and aquifer connectivity underpin the high productivity rates in the relict subsurface water. *Scientific Reports*, 1–34. https://doi.org/10.1038/s41598-024-68868-9

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., and Pevzner, P. A. (2012). SPAdes: A new

genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, **19(5)**, 455–477. https://doi.org/10.1089/cmb.2012.0021

Bates, S. T., Berg-Lyons, D., Caporaso, J. G., Walters, W. A., Knight, R., and Fierer, N. (2011). Examining the global distribution of dominant archaeal populations in soil. *The ISME Journal*, **5(5)**, 908–917. https://doi.org/10.1038/ismej.2010.171

Beaver, R. C., and Neufeld, J. D. (2024). Microbial ecology of the deep terrestrial subsurface. *ISME Journal*, **18(1)**. https://doi.org/10.1093/ismejo/wrae091

Bell, E., Lamminmäki, T., Alneberg, J., Andersson, A. F., Qian, C., Xiong, W., Hettich, R. L., Balmer, L., Frutschi, M., Sommer, G., and Bernier-Latmani, R. (2018). Biogeochemical cycling by a low-diversity microbial community in deep groundwater. *Frontiers in Microbiology*, **9(SEP)**, 1–17. https://doi.org/10.3389/fmicb.2018.02129

Bell, E., Lamminmäki, T., Alneberg, J., Andersson, A. F., Qian, C., Xiong, W., Hettich, R. L., Frutschi, M., and Bernier-Latmani, R. (2020). Active sulfur cycling in the terrestrial deep subsurface. *ISME Journal*, **14(5)**, 1260–1272. https://doi.org/10.1038/s41396-020-0602-x

Bhaskar Rao, Y. J., Sreenivas, B., Vijaya Kumar, T., Khadke, N., Kesava Krishna, A., and Babu, E. V. S. S. K. (2017). Evidence for Neoarchean basement for the Deccan Volcanic flows around Koyna-Warna region, western India: Zircon U-Pb age and Hf-isotopic results. *Journal of the Geological Society of India*, **90(6)**, 752–760. https://doi.org/10.1007/s12594-017-0787-4

Boada, E., Santos-Clotas, E., Cabrera-Codony, A., Martín, M. J., Bañeras, L., and Gich, F. (2021). The core microbiome is responsible for volatile silicon and organic compounds degradation during anoxic lab scale biotrickling filter performance. *Science of the Total Environment*, **798**. https://doi.org/10.1016/j.scitotenv.2021.149162

Bokulich, N. A., Kaehler, B. D., Rideout, J. R., Dillon, M., Bolyen, E., Knight, R., Huttley, G. A., and Caporaso, J. G. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome*, **6(1)**, 90. https://doi.org/10.1186/s40168-018-0470-z

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, **37(8)**, 852–857. https://doi.org/10.1038/s41587-019-0209-9

Bomberg, M., Lamminmäki, T., and Itävaara, M. (2015). Estimation of microbial metabolism and co-occurrence patterns in fracture groundwaters of deep crystalline bedrock at Olkiluoto, Finland. *Biogeosciences Discussions*, **12(16)**, 13819–13857. https://doi.org/10.5194/bgd-12-13819-2015

Bowers, R. M., Kyrpides, N. C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T. B. K., Schulz, F., Jarett, J., Rivers, A. R., Eloe-Fadrosh, E. A., Tringe, S. G., Ivanova, N. N., Copeland, A., Clum, A., Becraft, E. D., Malmstrom, R. R., Birren, B., Podar, M., Bork, P., … Woyke, T. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature Biotechnology*, **35(8)**, 725–731. https://doi.org/10.1038/nbt.3893

Bray, J. R., and Curtis, J. T. (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*, **27(4)**, 325–349. https://doi.org/10.2307/1942268

Bregnard, D., Leins, A., Cailleau, G., Vieth-Hillebrand, A., Eichinger, F., Ianotta, J., Hoffmann, R., Uhde, J., Bindschedler, S., Regenspurg, S., and Junier, P. (2023). Unveiling microbial diversity in deep geothermal fluids, from current knowledge and analogous environments. *Geothermal Energy*, **11(1)**. https://doi.org/10.1186/s40517-023-00269-z

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, **13(7)**, 581–583. https://doi.org/10.1038/nmeth.3869

Cantalapiedra, C. P., Herṇandez-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, **38(12)**, 5825–5829. https://doi.org/10.1093/molbev/msab293

Casar, C. P., Momper, L. M., Kruger, B. R., and Osburn, M. R. (2021). Iron-Fueled Life in the Continental Subsurface : Deep Mine Microbial Observatory, South Dakota, USA. *Applied and Environmental Microbiology*, **87(20)**, 1–14.

Chaumeil, P. A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2020). GTDB-Tk: A toolkit to classify genomes with the genome taxonomy database. *Bioinformatics*, **36(6)**, 1925–1927. https://doi.org/10.1093/bioinformatics/btz848

Chen, P., Zhou, H., Huang, Y., Xie, Z., Zhang, M., Wei, Y., Li, J., Ma, Y., Luo, M., Ding, W., Cao, J., Jiang, T., Nan, P., Fang, J., and Li, X. (2021). Revealing the full biosphere structure and versatile metabolic functions in the deepest ocean sediment of the

Challenger Deep. *Genome Biology*, **22(1)**, 1–28. https://doi.org/10.1186/s13059-021-02408-w

Coskun, Ö. K., Gomez-Saez, G. V., Beren, M., Doğacan, Ö., Günay, S. D., Elkin, V., Hoşgörmez, H., Einsiedl, F., Eisenreich, W., and Orsi, W. D. (2024). Quantifying genome-specific carbon fixation in a 750-meter deep subsurface hydrothermal microbial community. *FEMS Microbiology Ecology*, **100(5)**. https://doi.org/10.1093/femsec/fiae062

Das, K., Debnath, P., Duttagupta, S., Sarkar, S., Agrahari, S., and Mukherjee, A. (2020). Implication of submarine groundwater discharge to coastal ecology of the Bay of Bengal. *Journal of Earth System Science*, **129(1)**. https://doi.org/10.1007/s12040-019-1317-0

Davis, N. M., Proctor, Di. M., Holmes, S. P., Relman, D. A., and Callahan, B. J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome*, **6(1)**, 1–14. https://doi.org/10.1186/s40168-018-0605-2

Deja-Sikora, E., Gołębiewski, M., Kalwasińska, A., Krawiec, A., Kosobucki, P., and Walczak, M. (2019). Comamonadaceae OTU as a Remnant of an Ancient Microbial Community in Sulfidic Waters. *Microbial Ecology*, **78(1)**, 85–101. https://doi.org/10.1007/s00248-018-1270-5

Dutta, A., Dutta Gupta, S., Gupta, A., Sarkar, J., Roy, S., Mukherjee, A., and Sar, P. (2018). Exploration of deep terrestrial subsurface microbiome in Late Cretaceous Deccan traps and underlying Archean basement, India. *Scientific Reports*, **8(1)**, 17459. https://doi.org/10.1038/s41598-018-35940-0

Dutta, A., Peoples, L. M., Gupta, A., Bartlett, D. H., and Sar, P. (2019). Exploring the piezotolerant/piezophilic microbial community and genomic basis of piezotolerance within the deep subsurface Deccan traps. *Extremophiles*, **23(4)**, 421–433. https://doi.org/10.1007/s00792-019-01094-8

Dutta, A., Sar, P., Sarkar, J., Dutta Gupta, S., Gupta, A., Bose, H., Mukherjee, A., and Roy, S. (2019). Archaeal Communities in Deep Terrestrial Subsurface Underneath the Deccan Traps, India. *Frontiers in Microbiology*, **10**. https://doi.org/10.3389/fmicb.2019.01362

Goswami, D., Akkiraju, V. V., and Roy, S. (2024). Heat flow and thermal structure in the Koyna seismic zone, western India: Implications for recurrent reservoir triggered seismicity. *Tectonophysics*, **873(March 2023)**, 230216.

https://doi.org/10.1016/j.tecto.2024.230216

Goswami, D., Hazarika, P., and Roy, S. (2020). In Situ Stress Orientation From 3 km Borehole Image Logs in the Koyna Seismogenic Zone, Western India: Implications for Transitional Faulting Environment. *Tectonics*, **39(1)**, 1–17. https://doi.org/10.1029/2019TC005647

Goswami, D., Roy, S., and Akkiraju, V. V. (2019). Delineation of Damage Zones From 3 km Downhole Geophysical Logs in the Koyna Seismogenic Zone, Western India. *Journal of Geophysical Research: Solid Earth*, **124(6)**, 6101–6120. https://doi.org/10.1029/2018JB017257

Gregory, S. P., Barnett, M. J., Field, L. P., and Milodowski, A. E. (2019). Subsurface microbial hydrogen cycling: Natural occurrence and implications for industry. *Microorganisms*, **7(2)**. https://doi.org/10.3390/microorganisms7020053

Gupta, H., Purnachandra Rao, N., Roy, S., Arora, K., Tiwari, V. M., Patro, P. K., Satyanarayana, H. V. S., Shashidhar, D., Mallika, K., Akkiraju, V. V., Goswami, D., Vyas, D., Ravi, G., Srinivas, K. N. S. S. S., Srihari, M., Mishra, S., Dubey, C. P., Raju, D. C. V., Borah, U., … Nayak, S. (2015). Investigations related to scientific deep drilling to study reservoir-triggered earthquakes at Koyna, India. *International Journal of Earth Sciences*, **104(6)**, 1511–1522. https://doi.org/10.1007/s00531-014-1128-0

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, **29(8)**, 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Hallbeck, L., and Pedersen, K. (2008). Characterization of microbial processes in deep aquifers of the Fennoscandian Shield. *Applied Geochemistry*, **23(7)**, 1796–1819. https://doi.org/10.1016/j.apgeochem.2008.02.012

Hammer, Ø., Harper, D. A. T. a. T., and Ryan, P. D. (2001). PAST: Paleontological Statistics Software Package for Education and Data Analysis. In *Palaeontologia Electronica: Vol. 4(1)* (Issue 1, pp. 1–9). https://doi.org/10.1016/j.bcp.2008.05.025

Hernsdorf, A. W., Amano, Y., Miyakawa, K., Ise, K., Suzuki, Y., Anantharaman, K., Probst, A., Burstein, D., Thomas, B. C., and Banfield, J. F. (2017). Potential for microbial H2 and metal transformations associated with novel bacteria and archaea in deep terrestrial subsurface sediments. *ISME Journal*, **11(8)**, 1915–1929. https://doi.org/10.1038/ismej.2017.39

Herzig, M., Hyötyläinen, T., Vettese, G. F., Law, G. T. W., Vierinen, T., and Bomberg, M. (2024). Altering environmental conditions induce shifts in simulated deep terrestrial

subsurface bacterial communities—Secretion of primary and secondary metabolites. *Environmental Microbiology*, **26(1)**, 1–18. https://doi.org/10.1111/1462-2920.16552

Hosen, J. D., Febria, C. M., Crump, B. C., and Palmer, M. A. (2017). Watershed Urbanization Linked to Differences in Stream Bacterial Community Composition. *Frontiers in Microbiology*, **8**, 1452.

Hosono, T., Yamada, C., Manga, M., Wang, C. Y., and Tanimizu, M. (2020). Stable isotopes show that earthquakes enhance permeability and release water from mountains. *Nature Communications*, **11(1)**, 1–9. https://doi.org/10.1038/s41467-020-16604-y

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., Von Mering, C., and Bork, P. (2019). EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, **47(D1)**, D309–D314. https://doi.org/10.1093/nar/gky1085

Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, **11(1)**, 119. https://doi.org/10.1186/1471-2105-11-119

Ijiri, A., Inagaki, F., Kubo, Y., Adhikari, R. R., Hattori, S., Hoshino, T., Imachi, H., Kawagucci, S., Morono, Y., Ohtomo, Y., Ono, S., Sakai, S., Takai, K., Toki, T., Wang, D. T., Yoshinaga, M. Y., Arnold, G. L., Ashi, J., Case, D. H., … Yoshida, N. (2018). Deep-biosphere methane production stimulated by geofluids in the Nankai accretionary complex. *Science Advances*, **4(6)**, 1–16. https://doi.org/10.1126/sciadv.aao4631

Ino, K., Konno, U., Kouduka, M., Hirota, A., Togo, Y. S., Fukuda, A., Komatsu, D., Tsunogai, U., Tanabe, A. S., Yamamoto, S., Iwatsuki, T., Mizuno, T., Ito, K., and Suzuki, Y. (2016). Deep microbial life in high-quality granitic groundwater from geochemically and geographically distinct underground boreholes. *Environmental Microbiology Reports*, **8(2)**, 285–294. https://doi.org/10.1111/1758-2229.12379

Itävaara, M., Nyyssönen, M., Kapanen, A., Nousiainen, A., Ahonen, L., and Kukkonen, I. (2011). Characterization of bacterial diversity to a depth of 1500m in the Outokumpu deep borehole, Fennoscandian Shield. *FEMS Microbiology Ecology*, **77(2)**, 295–309. https://doi.org/10.1111/j.1574-6941.2011.01111.x

Jangir, Y., Karbelkar, A. A., Beedle, N. M., Zinke, L. A., Wanger, G., Anderson, C. M., Reese, B. K., Amend, J. P., and El-Naggar, M. Y. (2019). In situ Electrochemical Studies of the Terrestrial Deep Subsurface Biosphere at the Sanford Underground

Research Facility, South Dakota, USA. *Frontiers in Energy Research*, **7(November)**. https://doi.org/10.3389/fenrg.2019.00121

Jeong, H. I., Jin, H. M., and Jeon, C. O. (2016). Complete genome sequence of Sphingorhabdus sp. M41, a versatile hydrocarbon degrader, isolated from crude oil-contaminated costal sediment. *Journal of Biotechnology*, **227**, 41–42. https://doi.org/10.1016/j.jbiotec.2016.04.016

Kadnikov, V. V., Mardanov, A. V., Beletsky, A. V., Banks, D., Pimenov, N. V., Frank, Y. A., Karnachuk, O. V., and Ravin, N. V. (2018). A metagenomic window into the 2-km-deep terrestrial subsurface aquifer revealed multiple pathways of organic matter decomposition. *FEMS Microbiology Ecology*, **94(10)**, 1–16. https://doi.org/10.1093/femsec/fiy152

Kadnikov, V. V., Mardanov, A. V., Beletsky, A. V., Karnachuk, O. V., and Ravin, N. V. (2020). Microbial Life in the Deep Subsurface Aquifer Illuminated by Metagenomics. *Frontiers in Microbiology*, **11(September)**, 1–15. https://doi.org/10.3389/fmicb.2020.572252

Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., and Wang, Z. (2019). MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*, **2019(7)**, 1–13. https://doi.org/10.7717/peerj.7359

Kazy, S. K., Sahu, R. P., Mukhopadhyay, S., Bose, H., Mandal, S., and Sar, P. (2021). Microbial Life in Deep Terrestrial Continental Crust. In A. Pandey and A. Sharma (Eds.), *Extreme Environments*. CRC Press. https://doi.org/10.1201/9780429343452

Khan, A., Liu, G., Zhang, G., and Li, X. (2024). Radiation-resistant bacteria in desiccated soil and their potentiality in applied sciences. *Frontiers in Microbiology*, **15(June)**. https://doi.org/10.3389/fmicb.2024.1348758

Kieft, T. L. (2016). Microbiology of the Deep Continental Biosphere. In C. J. Hurst (Ed.), *Their World: A Diversity of Microbial Environments* (pp. 225–249). Springer International Publishing. https://doi.org/10.1007/978-3-319-28071-4_6

Kieft, T. L., McCuddy, S. M., Onstott, T. C., Davidson, M., Lin, L.-H., Mislowack, B., Pratt, L., Boice, E., Lollar, B. S., and Lippmann-Pipke, J. (2005). Geochemically generated, energy-rich substrates and indigenous microorganisms in deep, ancient groundwater. *Geomicrobiology Journal*, **22(6)**, 325–335.

Kolde, R., and Kolde, M. R. (2015). *Package 'pheatmap.'* R Package. https://cran.r-project.org/web/packages/pheatmap/pheatmap.pdf

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., Mcgettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, **23(21)**, 2947–2948. https://doi.org/10.1093/bioinformatics/btm404

Lau, M. C. Y., Kieft, T. L., Kuloyo, O., Linage-Alvarez, B., Van Heerden, E., Lindsay, M. R., Magnabosco, C., Wang, W., Wiggins, J. B., Guo, L., Perlman, D. H., Kyin, S., Shwe, H. H., Harris, R. L., Oh, Y., Yi, M. J., Purtschert, R., Slater, G. F., Ono, S., … Onstott, T. C. (2016). An oligotrophic deep-subsurface community dependent on syntrophy is dominated by sulfur-driven autotrophic denitrifiers. *Proceedings of the National Academy of Sciences of the United States of America*, **113(49)**, E7927–E7936. https://doi.org/10.1073/pnas.1612244113

Li, D., Liu, C. M., Luo, R., Sadakane, K., and Lam, T. W. (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, **31(10)**, 1674–1676. https://doi.org/10.1093/bioinformatics/btv033

Li, H. (2016). *Toolkit for processing sequences in FASTA/Q formats*. https://github.com/lh3/seqtk

Li, Y., Zhu, D., Niu, L., Zhang, W., Wang, L., Zhang, H., Zou, S., and Zhou, C. (2024). Carbon-fixing bacteria in diverse groundwaters of karst area: Distribution patterns, ecological interactions, and driving factors. *Water Research*, **261(March)**. https://doi.org/10.1016/j.watres.2024.121979

Liu, H., Yao, J., Yuan, Z., Chen, H., Wang, F., Masakorala, K., and Choi, M. M. F. (2016). Degradation of hydrocarbons by indigenous microbial communities from two adjacent oil production wells in one block. *Energy Sources, Part A: Recovery, Utilization and Environmental Effects*, **38(23)**, 3423–3434. https://doi.org/10.1080/15567036.2014.943853

Liu, Y., Song, X.-F., Jiang, J.-T., Liu, Y.-H., Xu, C.-J., Li, H., and Liu, Z.-P. (2011). Hydrocarboniphaga daqingensis sp. nov., isolated from a freshwater lake. *International Journal of Systematic and Evolutionary Microbiology*, **61(2)**, 408–411.

Lollar, G. S., Warr, O., Telling, J., Osburn, M. R., and Lollar, B. S. (2019). 'Follow the Water': Hydrogeochemical Constraints on Microbial Investigations 2.4 km Below Surface at the Kidd Creek Deep Fluid and Deep Life Observatory. *Geomicrobiology Journal*, **36(10)**, 859–872. https://doi.org/10.1080/01490451.2019.1641770

Lopez-Fernandez, M., Simone, D., Wu, X., Soler, L., Nilsson, E., Holmfeldt, K., Lantz, H.,

Bertilsson, S., and Dopson, M. (2018). Metatranscriptomes reveal that all three domains of life are active but are dominated by bacteria in the fennoscandian crystalline granitic continental deep biosphere. *MBio*, **9(6)**, 1–15. https://doi.org/10.1128/mbio.01792-18

Magnabosco, C., Ryan, K., Lau, M. C. Y., Kuloyo, O., Sherwood Lollar, B., Kieft, T. L., Van HeerDen, E., and Onstott, T. C. (2016). A metagenomic window into carbon metabolism at 3 km depth in Precambrian continental crust. *ISME Journal*, **10(3)**, 730–741. https://doi.org/10.1038/ismej.2015.150

Mandal, S., Bose, H., Ramesh, K., Sahu, R. P., Saha, A., Sar, P., and Kazy, S. K. (2022). Depth wide distribution and metabolic potential of chemolithoautotrophic microorganisms reactivated from deep continental granitic crust underneath the Deccan Traps at Koyna, India. *Frontiers in Microbiology*, **13(3)**, 1–23. https://doi.org/10.3389/fmicb.2022.1018940

McDonald, D., Jiang, Y., Balaban, M., Cantrell, K., Zhu, Q., Gonzalez, A., Morton, J. T., Nicolaou, G., Parks, D. H., Karst, S. M., Albertsen, M., Hugenholtz, P., DeSantis, T., Song, S. J., Bartko, A., Havulinna, A. S., Jousilahti, P., Cheng, S., Inouye, M., … Knight, R. (2024). Greengenes2 unifies microbial data in a single reference tree. *Nature Biotechnology*, **42(5)**, 715–718. https://doi.org/10.1038/s41587-023-01845-1

McMahon, S., and Parnell, J. (2014). Weighing the deep continental biosphere. *FEMS Microbiology Ecology*, **87(1)**, 113–120. https://doi.org/10.1111/1574-6941.12196

Meyer-Dombard, D. R., and Malas, J. (2022). Advances in Defining Ecosystem Functions of the Terrestrial Subsurface Biosphere. *Frontiers in Microbiology*, **13(June)**, 1–19. https://doi.org/10.3389/fmicb.2022.891528

Momper, L., Casar, C. P., and Osburn, M. R. (2023). A metagenomic view of novel microbial and metabolic diversity found within the deep terrestrial biosphere at DeMMO: A microbial observatory in South Dakota, USA. *Environmental Microbiology*, **25(12)**, 3719–3737. https://doi.org/10.1111/1462-2920.16543

Momper, L., Jungbluth, S. P., Lee, M. D., and Amend, J. P. (2017). Energy and carbon metabolisms in a deep terrestrial subsurface fluid microbial community. *ISME Journal*, **11(10)**, 2319–2333. https://doi.org/10.1038/ismej.2017.94

Mukherjee, S., Stamatis, D., Bertsch, J., Ovchinnikova, G., Verezemska, O., Isbandi, M., Thomas, A. D., Ali, R., Sharma, K., Kyrpides, N. C., and Reddy, T. B. K. (2017). Genomes OnLine Database (GOLD) v.6: Data updates and feature enhancements. *Nucleic Acids Research*, **45(D1)**, D446–D456. https://doi.org/10.1093/nar/gkw992

Nealson, K. H., Inagaki, F., and Takai, K. (2005). Hydrogen-driven subsurface

lithoautotrophic microbial ecosystems (SLiMEs): Do they exist and why should we care? *Trends in Microbiology*, **13(9)**, 405–410. https://doi.org/10.1016/j.tim.2005.07.010

Neuenschwander, S. M., Ghai, R., Pernthaler, J., and Salcher, M. M. (2018). Microdiversification in genome-streamlined ubiquitous freshwater Actinobacteria. *ISME Journal*, **12(1)**, 185–198. https://doi.org/10.1038/ismej.2017.156

Newton, R. J., and McLellan, S. L. (2015). A unique assemblage of cosmopolitan freshwater bacteria and higher community diversity differentiate an urbanized estuary from oligotrophic Lake Michigan. *Frontiers in Microbiology*, **6(SEP)**, 1–13. https://doi.org/10.3389/fmicb.2015.01028

Nyyssönen, M., Bomberg, M., Kapanen, A., Nousiainen, A., Pitkänen, P., and Itävaara, M. (2012). Methanogenic and Sulphate-Reducing Microbial Communities in Deep Groundwater of Crystalline Rock Fractures in Olkiluoto, Finland. *Geomicrobiology Journal*, **29(10)**, 863–878. https://doi.org/10.1080/01490451.2011.635759

Nyyssönen, M., Hultman, J., Ahonen, L., Kukkonen, I., Paulin, L., Laine, P., Itävaara, M., and Auvinen, P. (2014). Taxonomically and functionally diverse microbial communities in deep crystalline rocks of the Fennoscandian shield. *ISME Journal*, **8(1)**, 126–138. https://doi.org/10.1038/ismej.2013.125

Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Borman, T., Carvalho, G., Chirico, M., De Caceres, M., … Weedon, J. (2025). *vegan: Community Ecology Package*. https://cran.r-project.org/package=vegan

Onstott, T. C., Moser, D. P., Pfiffner, S. M., Fredrickson, J. K., Brockman, F. J., Phelps, T. J., White, D. C., Peacock, A., Balkwill, D., Hoover, R., Krumholz, L. R., Borscik, M., Kieft, T. L., and Wilson, R. (2003). Indigenous and contaminant microbes in ultradeep mines. *Environmental Microbiology*, **5(11)**, 1168–1191. https://doi.org/10.1046/j.1462-2920.2003.00512.x

Or, D., Smets, B. F., Wraith, J. M., Dechesne, A., and Friedman, S. P. (2007). Physical constraints affecting bacterial habitats and activity in unsaturated porous media - a review. *Advances in Water Resources*, **30(6–7)**, 1505–1527. https://doi.org/10.1016/j.advwatres.2006.05.025

Osburn, M. R., LaRowe, D. E., Momper, L. M., and Amend, J. P. (2014). Chemolithotrophy in the continental deep subsurface: Sanford underground research facility (SURF), USA. *Frontiers in Microbiology*, **5(NOV)**, 1–14. https://doi.org/10.3389/fmicb.2014.00610

Overholt, W. A., Trumbore, S., Xu, X., Bornemann, T. L. V., Probst, A. J., Krüger, M., Herrmann, M., Thamdrup, B., Bristow, L. A., Taubert, M., Schwab, V. F., Hölzer, M., Marz, M., and Küsel, K. (2022). Carbon fixation rates in groundwater similar to those in oligotrophic marine systems. *Nature Geoscience*, **15(7)**, 561–567. https://doi.org/10.1038/s41561-022-00968-5

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, **25(7)**, 1043–1055. https://doi.org/10.1101/gr.186072.114

Pedersen, K. (2000). Exploration of deep intraterrestrial microbial life: current perspectives. *FEMS Microbiology Letters*, **185(1)**, 9–16.

Pedersen, K., Bengtsson, A. F., Edlund, J. S., and Eriksson, L. C. (2014). Sulphate-controlled Diversity of Subterranean Microbial Communities over Depth in Deep Groundwater with Opposing Gradients of Sulphate and Methane. *Geomicrobiology Journal*, **31(7)**, 617–631. https://doi.org/10.1080/01490451.2013.879508

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, **12(Oct)**, 2825–2830.

Probst, A. J., Ladd, B., Jarett, J. K., Geller-Mcgrath, D. E., Sieber, C. M. K., Emerson, J. B., Anantharaman, K., Thomas, B. C., Malmstrom, R. R., Stieglmeier, M., Klingl, A., Woyke, T., Ryan, M. C., and Banfield, J. F. (2018). Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. *Nature Microbiology*, **3(3)**, 328–336. https://doi.org/10.1038/s41564-017-0098-y

Purkamo, L., Bomberg, M., Kietäväinen, R., Salavirta, H., Nyyssönen, M., Nuppunen-Puputti, M., Ahonen, L., Kukkonen, I., and Itävaara, M. (2016). Microbial co-occurrence patterns in deep Precambrian bedrock fracture fluids. *Biogeosciences*, **13(10)**, 3091–3108. https://doi.org/10.5194/bg-13-3091-2016

Purkamo, L., Bomberg, M., Nyyssönen, M., Ahonen, L., Kukkonen, I., and Itävaara, M. (2017). Response of deep subsurface microbial community to different carbon sources and electron acceptors during ~2 months incubation in microcosms. *Frontiers in Microbiology*, **8(FEB)**. https://doi.org/10.3389/fmicb.2017.00232

Purkamo, L., Bomberg, M., Nyyssönen, M., Kukkonen, I., Ahonen, L., and Itävaara, M.

(2015). Heterotrophic Communities Supplied by Ancient Organic Carbon Predominate in Deep Fennoscandian Bedrock Fluids. *Microbial Ecology*, **69(2)**, 319–332. https://doi.org/10.1007/s00248-014-0490-6

Purkamo, L., Bomberg, M., Nyyssönen, M., Kukkonen, I., Ahonen, L., Kietäväinen, R., and Itävaara, M. (2013). Dissecting the deep biosphere: retrieving authentic microbial communities from packer-isolated deep crystalline bedrock fracture zones. *FEMS Microbiology Ecology*, **85(2)**, 324–337. https://doi.org/10.1111/1574-6941.12126

Purkamo, L., Kietäväinen, R., Nuppunen-Puputti, M., Bomberg, M., and Cousins, C. (2020). Ultradeep microbial communities at 4.4 km within crystalline bedrock: Implications for habitability in a planetary context. *Life*, **10(1)**. https://doi.org/10.3390/life10010002

Roy, S. (2017). Scientific drilling in Koyna region, Maharashtra. *Current Science*, **112(11)**, 2181.

Saha, P. K., Sahu, R. P., Sar, P., and Kazy, S. K. (2024). Deep subsurface igneous rocks from the Deccan traps harbor H2 and CO2 utilizing chemolithoautotrophic bacteria. *Biologia*, **79(5)**, 1553–1569. https://doi.org/10.1007/s11756-024-01631-7

Sahl, J. W., Schmidt, R., Swanner, E. D., Mandernack, K. W., Templeton, A. S., Kieft, T. L., Smith, R. L., Sanford, W. E., Callaghan, R. L., Mitton, J. B., and Spear, J. R. (2008). Subsurface microbial diversity in deep-granitic-fracture water in Colorado. *Applied and Environmental Microbiology*, **74(1)**, 143–152. https://doi.org/10.1128/AEM.01133-07

Sahu, R. P., Kazy, S. K., Bose, H., Mandal, S., Dutta, A., Saha, A., Roy, S., Dutta Gupta, S., Mukherjee, A., and Sar, P. (2022). Microbial diversity and function in crystalline basement beneath the Deccan Traps explored in a 3 km borehole at Koyna, western India. *Environmental Microbiology*, **24(6)**, 2837–2853. https://doi.org/10.1111/1462-2920.15867

Schoene, B., Samperton, K. M., Eddy, M. P., Keller, G., Adatte, T., Bowring, S. A., Khadri, S. F. R., and Gertsch, B. (2015). U-Pb geochronology of the Deccan Traps and relation to the end-Cretaceous mass extinction. *Science*, **347(6218)**, 182–184. https://doi.org/10.1126/science.aaa0118

Segura, A., Hernández-Sánchez, V., Marqués, S., and Molina, L. (2017). Insights in the regulation of the degradation of PAHs in Novosphingobium sp. HR1a and utilization of this regulatory system as a tool for the detection of PAHs. *Science of the Total Environment*, **590–591**, 381–393. https://doi.org/10.1016/j.scitotenv.2017.02.180

Serra Moncadas, L., Hofer, C., Bulzu, P. A., Pernthaler, J., and Andrei, A. S. (2024). Freshwater genome-reduced bacteria exhibit pervasive episodes of adaptive stasis.

Nature Communications, **15(1)**, 1–15. https://doi.org/10.1038/s41467-024-47767-7

Shade, A., and Handelsman, J. (2012). Beyond the Venn diagram: The hunt for a core microbiome. *Environmental Microbiology*, **14(1)**, 4–12. https://doi.org/10.1111/j.1462-2920.2011.02585.x

Shaffer, M., Borton, M. A., McGivern, B. B., Zayed, A. A., La Rosa, S. L. 0003 3527 8101, Solden, L. M., Liu, P., Narrowe, A. B., Rodríguez-Ramos, J., Bolduc, B., Gazitúa, M. C., Daly, R. A., Smith, G. J., Vik, D. R., Pope, P. B., Sullivan, M. B., Roux, S., and Wrighton, K. C. (2020). DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Research*, **48(16)**, 8883–8900. https://doi.org/10.1093/nar/gkaa621

Sheik, C. S., Reese, B. K., Twing, K. I., Sylvan, J. B., Grim, S. L., Schrenk, M. O., Sogin, M. L., and Colwell, F. S. (2018). Identification and removal of contaminant sequences from ribosomal gene databases: Lessons from the Census of Deep Life. *Frontiers in Microbiology*, **9(APR)**, 1–14. https://doi.org/10.3389/fmicb.2018.00840

Shimizu, R., Dempo, Y., Nakayama, Y., Nakamura, S., Bamba, T., Fukusaki, E., and Fukui, T. (2015). New insight into the role of the Calvin cycle: reutilization of CO 2 emitted through sugar degradation. *Scientific Reports*, **5**, 11617.

Simkus, D. N., Slater, G. F., Lollar, B. S., Wilkie, K., Kieft, T. L., Magnabosco, C., Lau, M. C. Y., Pullin, M. J., Hendrickson, S. B., Wommack, K. E., Sakowski, E. G., Heerden, E. van, Kuloyo, O., Linage, B., Borgonie, G., and Onstott, T. C. (2016). Variations in microbial carbon sources and cycling in the deep continental subsurface. *Geochimica et Cosmochimica Acta*, **173**, 264–283. https://doi.org/10.1016/j.gca.2015.10.003

Sokol, N. W., Slessarev, E., Marschmann, G. L., Nicolas, A., Blazewicz, S. J., Brodie, E. L., Firestone, M. K., Foley, M. M., Hestrin, R., Hungate, B. A., Koch, B. J., Stone, B. W., Sullivan, M. B., Zablocki, O., Trubl, G., McFarlane, K., Stuart, R., Nuccio, E., Weber, P., … Pett-Ridge, J. (2022). Life and death in the soil microbiome: how ecological processes influence biogeochemistry. *Nature Reviews Microbiology*, **20(7)**, 415–430. https://doi.org/10.1038/s41579-022-00695-z

Song, J., Choo, Y.-J., and Cho, J.-C. (2008). Perlucidibaca piscinae gen. nov., sp. nov., a freshwater bacterium belonging to the family Moraxellaceae. *International Journal of Systematic and Evolutionary Microbiology*, **58(1)**, 97–102.

Steinegger, M., and Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, **35(11)**, 1026–1028. https://doi.org/10.1038/nbt.3988

Stevens, T. O. (1997). Subsurface lithoautotrophic microbial ecosystems (SLMEs) in igneous rocks: prospects for detection. In R. B. Hoover (Ed.), *Instruments, Methods, and Missions for the Investigation of Extraterrestrial Microorganisms* (Vol. 3111, Issue July 1997, pp. 358–365). https://doi.org/10.1117/12.278790

Stevens, T. O., and McKinley, J. P. (2000). Abiotic controls on H2 production from Basalt - Water reactions and implications for aquifer biogeochemistry. *Environmental Science and Technology*, **34(5)**, 826–831. https://doi.org/10.1021/es990583g

Stoddard, S. F., Smith, B. J., Hein, R., Roller, B. R. K., and Schmidt, T. M. (2015). rrnDB: Improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Research*, **43(D1)**, D593–D598. https://doi.org/10.1093/nar/gku1201

Su, J. feng, Shi, J. xin, and Ma, F. (2017). Aerobic denitrification and biomineralization by a novel heterotrophic bacterium, Acinetobacter sp. H36. *Marine Pollution Bulletin*, **116(1–2)**, 209–215.

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*, **38(7)**, 3022–3027. https://doi.org/10.1093/molbev/msab120

Templeton, A. S., and Caro, T. A. (2023). The Rock-Hosted Biosphere. *Annual Review of Earth and Planetary Sciences*, **51(1)**, 493–519. https://doi.org/10.1146/annurev-earth-031920-081957

Thieringer, P. H., Boyd, E. S., Templeton, A. S., and Spear, J. R. (2023). Metapangenomic investigation provides insight into niche differentiation of methanogenic populations from the subsurface serpentinizing environment, Samail Ophiolite, Oman. *Frontiers in Microbiology*, **14(July)**, 1–18. https://doi.org/10.3389/fmicb.2023.1205558

Uritskiy, G. V., DiRuggiero, J., and Taylor, J. (2018). MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome*, **6(1)**, 1–13. https://doi.org/10.1186/s40168-018-0541-1

Vaccarelli, I., Matteucci, F., Pellegrini, M., Bellatreccia, F., and Del Gallo, M. (2021). Exploring Microbial Biosignatures in Mn-Deposits of Deep Biosphere: A Preliminary Cross-Disciplinary Approach to Investigate Geomicrobiological Interactions in a Cave in Central Italy. *Frontiers in Earth Science*, **9(March)**, 1–10. https://doi.org/10.3389/feart.2021.590257

Valentin-Alvarado, L. E., Fakra, S. C., Probst, A. J., Giska, J. R., Jaffe, A. L., Oltrogge, L. M., West-Roberts, J., Rowland, J., Manga, M., Savage, D. F., Greening, C., Baker, B. J.,

and Banfield, J. F. (2024). Autotrophic biofilms sustained by deeply sourced groundwater host diverse bacteria implicated in sulfur and hydrogen metabolism. *Microbiome*, **12(1)**, 1–20. https://doi.org/10.1186/s40168-023-01704-w

Wang, T., Zhang, P., Anantharaman, K., Wang, H., Zhang, H., Zhang, M., and Xu, J. (2025). Metagenomic analysis reveals how multiple stressors disrupt virus–host interactions in multi-trophic freshwater mesocosms. *Nature Communications*, **16(1)**. https://doi.org/10.1038/s41467-025-63162-2

Wu, X., Holmfeldt, K., Hubalek, V., Lundin, D., Åström, M., Bertilsson, S., and Dopson, M. (2016). Microbial metagenomes from three aquifers in the Fennoscandian shield terrestrial deep biosphere reveal metabolic partitioning among populations. *ISME Journal*, **10(5)**, 1192–1203. https://doi.org/10.1038/ismej.2015.185

Wu, Y.-W., Simmons, B. A., and Singer, S. W. (2016). MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics*, **32(4)**, 605–607. https://doi.org/10.1093/bioinformatics/btv638

Xu, S., Li, L., Luo, X., Chen, M., Tang, W., Zhan, L., Dai, Z., Lam, T. T., Guan, Y., and Yu, G. (2022). Ggtree: A serialized data object for visualization of a phylogenetic tree and annotation data. *IMeta*, **1(4)**, 5–8. https://doi.org/10.1002/imt2.56

Yan, L., Herrmann, M., Kampe, B., Lehmann, R., Totsche, K. U., and Küsel, K. (2020). Environmental selection shapes the formation of near-surface groundwater microbiomes. *Water Research*, **170**. https://doi.org/10.1016/j.watres.2019.115341

Table 1. Hydrochemical characteristics of the water samples. “ND” denotes “not detected”.

Sample Details		KD [SW]	KR [SW]	PBW [GWS]	BW [GW]	BGRL [GW]	DBW [GW]
Location		17°24'34.4"N 73°44'38.4"E	17°21'03.2"N 73°52'43.8"E	17°18'15.9"N 73°47'36.1"E	17°09'01.0"N 73°40'03.5"E	17°17'52.0"N 74°07'42.0"E	17°17'56.5"N 73°44'18.6"E
Depth (m)		0	0	2	100	100	1027
Elements (mg/L)	Ca	12.8	11.0	21.2	29.3	32.1	10.0
	Fe	142.1	177.8	113.5	170.7	135.0	106.4
	K	16.0	15.7	9.3	14.3	15.5	16.0
	Mg	8.6	6.7	16.9	25.0	27.8	5.8
	Na	17.1	16.9	10.4	15.5	16.6	17.2
Carbon (mg/L)	TOC	2.8	2.2	2.5	2.1	1.9	2.2
	TIC	3.6	5.7	2.8	3.0	7.1	5.2
	TC	7.9	9.4	6.7	6.6	10.4	8.9
Organic acid (µg/L)	Acetate	0.05	0.03	0.04	0.02	0.01	0.05
	Succinate	0.09	0.05	0.08	0.05	0.02	0.10
	Pyruvate	0.00	0.00	0.00	0.00	0.00	0.00
	Citrate	0.00	0.00	0.00	0.00	0.00	0.00
	Oxalate	0.20	0.16	0.19	0.15	0.12	0.20
	Lactate	0.19	0.11	0.17	0.10	0.03	0.20
Anions (mg/L)	Cl-	232.4	330.6	295.6	9.9	500.6	595.1
	SO_4^{2-}	156.6	312.7	140.6	0.0	209.3	266.3
	NO_3^-	24.0	36.0	21.0	0.0	50.6	61.0
	NO_2^-	2.4	1.6	0.6	3.1	1.3	4.6
	PO_4^{3-}	14.0	22.6	6.6	0.0	41.3	52.0
	HCO_3^-	49.2	2.2	52.0	4.8	10.9	13.6
Others	$\delta^{18}O$	-4.1	-4.7	-6.3	-4.0	-3.9	-2.5
	$\delta^{2}H$	-44.0	-33.6	-45.2	-34.0	-34.6	-23.4
	pH	7.5	7.8	6.4	8.2	7.8	8.1
	DO (ppm)	5.0	3.1	3.5	1.7	4.7	4.3
	Conductivity (µS/cm)	352.0	403.0	285.0	220.0	325.0	296.0
No. of Bacterial cells (per L)		ND	547,000	3,930,000	ND	3,090,000	433,000
No. of Archaeal cell count (per L)		ND	ND	ND	ND	850	ND
*dsr*B copy number (per L)		ND	148	1270	ND	3340	1080

Fig. 1. Microbial community composition of the water samples. (a) Relative abundance of major taxa (at Phylum level) present in the microbial communities; b and c. Heat maps displaying the relative percentage abundance of major microbial families (b) and genera (c).

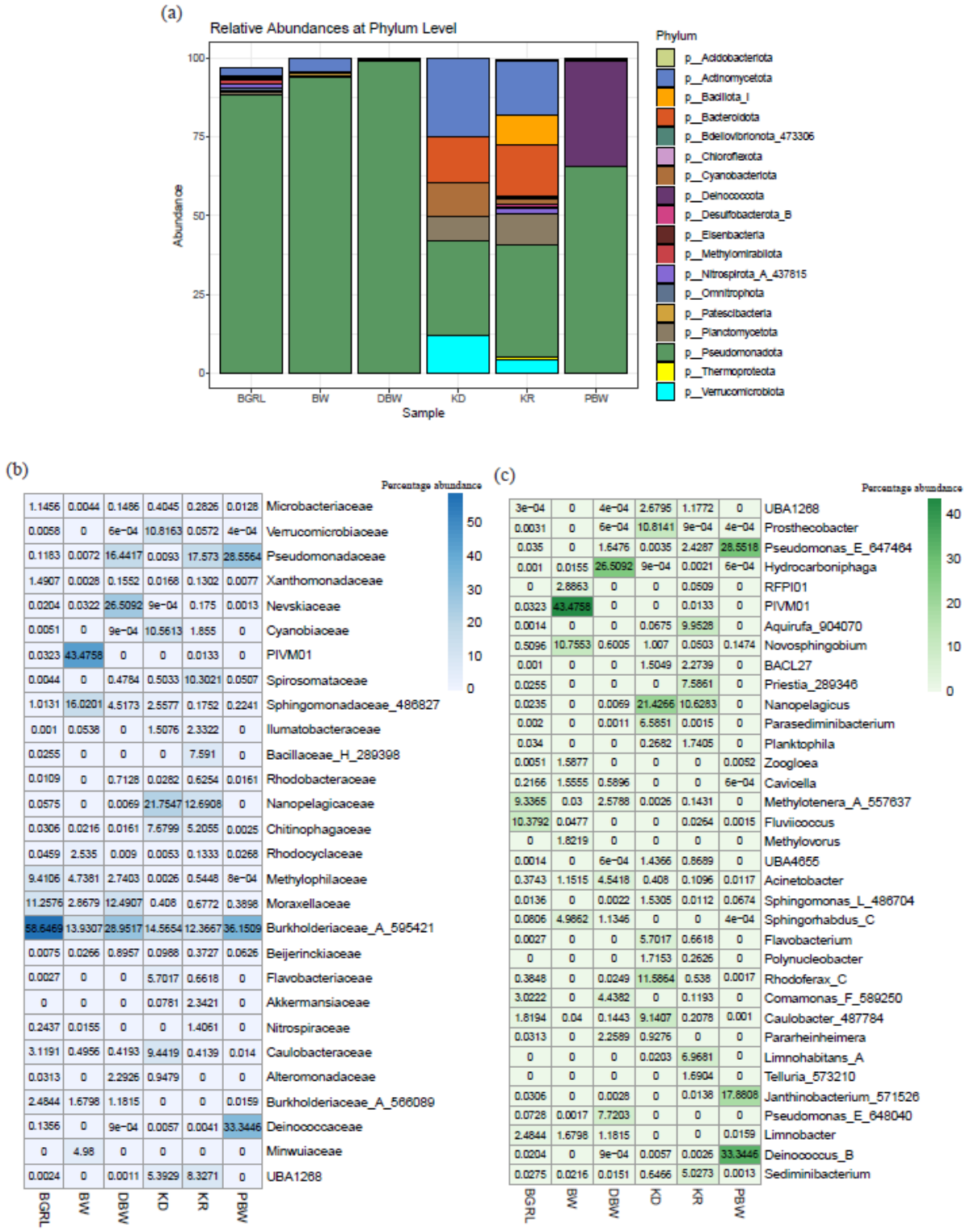

Fig. 2. (a) Non-metric multidimensional scaling (NMDS) plot of major microbial families with geochemical parameters (Stress: 2.740543 $\times 10^{-5}$). In the NMDS plot, the circles represent samples, the squares represent families, and the vector represents geochemical parameters. GW - Groundwater, GWS - Groundwater Seepage, SW - Surface Water (b) Heatmap of the Spearman correlations between core ASVs with the geochemical parameters. Dendrogram displaying the unweighted pair group method with arithmetic mean (UPGMA) based clustering of microbial classes and the geochemical parameters.

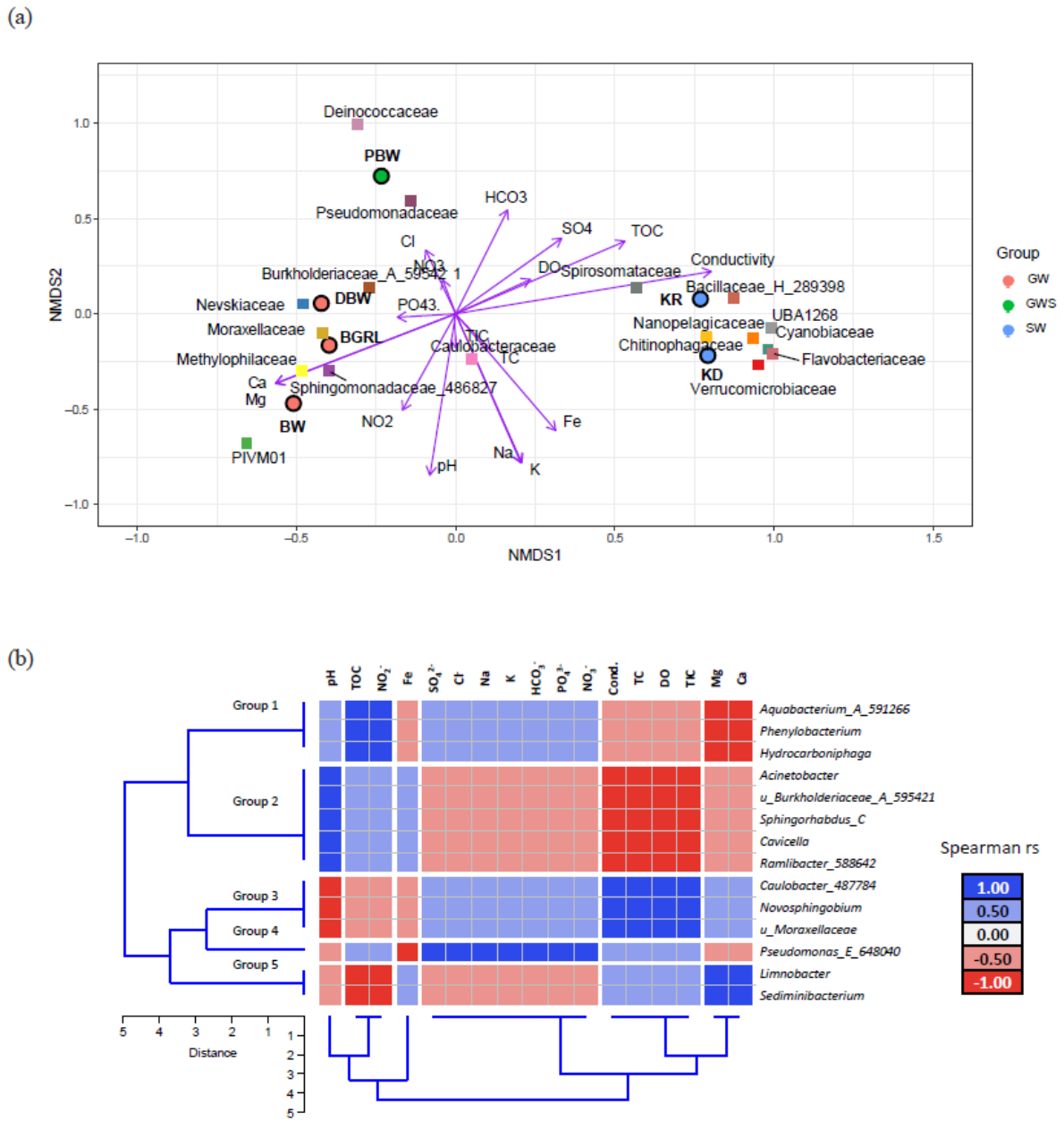

Fig. 3. Heatmap of (a) percent completeness of carbon metabolism pathways in each bin; (b) abundance of genes involved in biogeochemical cycling and stress tolerance in terms of copies per bin; and (c) bin abundance in each sample expressed as genome copies per million filtered reads (GPM). The samples were clustered via UPGMA according to bin abundances based on Bray-Curtis similarity index with a bootstrap of 1000

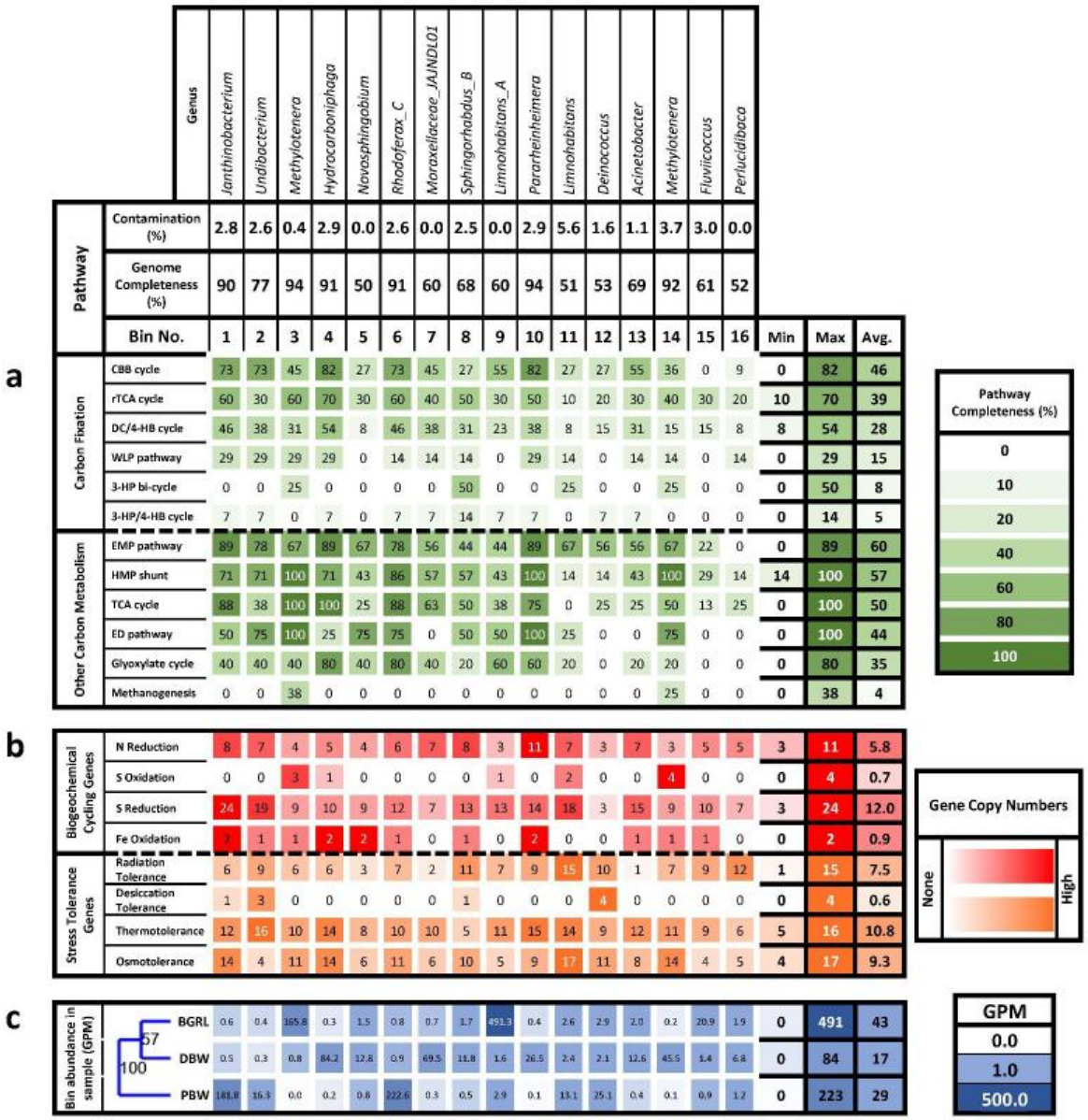

Fig. 4. Heatmap showing occurrences (min-max scaled based on rows) of genes related to (a) carbon cycle, (b) nitrogen cycle, and (c) sulfur cycle across three rock samples and three water samples. The number represented in each cell of the heatmap denotes the relative percentage occurrences of genes, which are calculated based on the occurrence of a particular gene / total number of genes predicted × 100. The values are rounded to seven decimal points. The color codes signify a particular gene's highest and lowest abundance across six samples for comparing rock and water samples. Nitrate_Reduction_A is a group of genes needed for the reduction of nitrate to nitrite, whereas Nitrate_Reduction_B is a group of genes involved in dissimilatory nitrate reduction to ammonium. CF prefix in the carbon cycle heatmap signifies carbon fixation pathways. CF_3HP: 3-Hydroxypropionate bi-cycle; CF_CBB: Calvin–Benson–Bassham (CBB) cycle; CF_HPHB: Hydroxypropionate-hydroxybutyrate cycle; CF_rTCA: reverse tricarboxylic acid cycle; CF_WLP: Wood-Ljungdahl pathway. * Gene found in CF_3HP and CF_HPHB pathways