MICROBIOLOGY

MICROBIOLOGY SOCIETY

OPEN ACCESS

A global perspective on bacterial diversity in the terrestrial deep subsurface

A. Soares[1,2,3]†, A. Edwards[2,4,*], D. An[5], A. Bagnoud[6], J. Bradley[7], E. Barnhart[8,9], M. Bomberg[10], K. Budwill[11], S. M. Caffrey[12], M. Fields[9,13], J. Gralnick[3], V. Kadnikov[14], L. Momper[15], M. Osburn[16], A. Mu[17,18,19], J. W. Moreau[20], D. Moser[21], L. Purkamo[10,22,23], S. M. Rassner[1,4], C. S. Sheik[24], B. Sherwood Lollar[25], B. M. Toner[26], G. Voordouw[5], K. Wouters[27] and A. C. Mitchell[1,4,*]

Abstract

While recent efforts to catalogue Earth's microbial diversity have focused upon surface and marine habitats, 12–20% of Earth's biomass is suggested to exist in the terrestrial deep subsurface, compared to ~1.8% in the deep subseafloor. Metagenomic studies of the terrestrial deep subsurface have yielded a trove of divergent and functionally important microbiomes from a range of localities. However, a wider perspective of microbial diversity and its relationship to environmental conditions within the terrestrial deep subsurface is still required. Our meta-analysis reveals that terrestrial deep subsurface microbiota are dominated by *Betaproteobacteria*, *Gammaproteobacteria* and *Firmicutes*, probably as a function of the diverse metabolic strategies of these taxa. Evidence was also found for a common small consortium of prevalent *Betaproteobacteria* and *Gammaproteobacteria* operational taxonomic units across the localities. This implies a core terrestrial deep subsurface community, irrespective of aquifer lithology, depth and other variables, that may play an important role in colonizing and sustaining microbial habitats in the deep terrestrial subsurface. An *in silico* contamination-aware approach to analysing this dataset underscores the importance of downstream methods for assuring that robust conclusions can be reached from deep subsurface-derived sequencing data. Understanding the global panorama of microbial diversity and ecological dynamics in the deep terrestrial subsurface provides a first step towards understanding the role of microbes in global subsurface element and nutrient cycling.

DATA SUMMARY

The 16S rRNA gene sequencing data utilized in the present study are available on NCBI under the following project accessions: PRJNA262938, PRJNA268940, PRJNA248749, PRJNA251746, PRJNA375701, PRJEB1468 and PRJEB10822. The code used for the processing and data analysis of the datasets is available at: https://github.com/GeoMicroSoares/mads_scripts.

INTRODUCTION

Understanding the distribution of microbial diversity is pivotal for advancing our knowledge of deep subsurface global biogeochemical cycles [1, 2]. Subsurface biomass is suggested to have exceeded that of the Earth's surface by an order of magnitude (~45% of Earth's total biomass) before land plants evolved, ca. 0.5 billion years ago [3]. Integrative modelling of cell count and quantitative PCR (qPCR) data and geophysical factors indicated in late 2018 that the bacterial and archaeal biomass found in the global deep subsurface may range from 23 to 31 petagrams of carbon (PgC) [4]. These values halved estimates from efforts earlier that year but maintained the notion that the terrestrial deep subsurface holds ca. 5-fold more bacterial and archaeal biomass than the deep marine subsurface [4, 5]. Further, it is expected that 20–80% of the possible $2-6\times10^{29}$ prokaryotic cells present in the terrestrial subterranean biome exist as biofilms and play crucial roles in global biogeochemical cycles [4, 6, 7].

Cataloguing microbial diversity and functionality in the terrestrial deep subsurface has mostly been achieved by means of marker gene and metagenome sequencing from aquifers associated with coals, sandstones, carbonates and clays, as well as deep igneous and metamorphic rocks [8–19]. Only recently has the first comprehensive database of 16S rRNA gene-based studies targeting terrestrial subsurface environments been compiled [4]. This work focused on updating estimates for bacterial and archaeal biomass, and cell numbers across the terrestrial deep subsurface, but also linked the identified bacterial and archaeal phylum-level compositions to host-rock type, and to 16S rRNA gene region primer targets [4]. While highlighting *Firmicutes* and Proteobacterial dominance in the bacterial component of the terrestrial deep subsurface, no further taxonomic insights emerged. Genus-level identification remains an important niche necessary for understanding community composition, inferred metabolism and hence microbial contributions of distinct community members to biogeochemical cycling in the deep subsurface [15, 20–22]. Indeed, such genus-specific traits have been demonstrated to be critical for understanding crucial biological functions in other microbiomes, and genus-specific functions of relevance for deep subsurface biogeochemistry are clear [23–25].

So far, the potential biogeochemical impacts of microbial activity in the deep subsurface have been inferred through shotgun metagenomics, as well as from incubation experiments of primary geological samples amended with molecules or minerals of interest [16, 17, 19, 26–29]. Recent studies of deep terrestrial subsurface microbial communities further suggest that these are metabolically active, often associated with novel uncultured phyla, and potentially directly involved in carbon and sulphur cycling [30–36]. Concomitant advancements in subsurface drilling, molecular methods and computational techniques have aided exploration of the subsurface biosphere, but serious challenges remain, mostly related to deciphering sample contamination by drilling methods, community interactions with reactive casing materials and sample transportation to laboratories for processing [37, 38]. The logistical challenges inherent in accessing and recovering *in situ* samples from hundreds to thousands of metres below the surface complicate our view of terrestrial subsurface microbial ecology [39].

In this study, we capitalize on the increased availability of 16S rRNA gene amplicon data from multiple studies of the terrestrial deep subsurface conducted over the last decade. We apply bespoke bioinformatics scripts to generate insights into the microbial community structure and controls upon bacterial microbiomes of the terrestrial deep subsurface across a large distribution of habitat types on multiple continents. The deep biosphere is as-yet undefined as a biome – elevated temperature, anoxic conditions, varying levels of organic carbon, and measures of isolation from the surface photosphere are some of the criteria used, albeit without a consensus. For this work a more general approach has been taken to define the terrestrial deep subsurface for the purposes of this initial examination as the zone at least 100 m from the surface [7, 40, 41].

METHODS

Data acquisition

The Sequence Read Archive database of the National Center for Biotechnology Information (SRA-NCBI) was queried for 16S rRNA-based deep subsurface datasets (excluding marine and ice samples, as well as any human-impacted samples); available studies were downloaded using the SRA Run Selector. Studies were selected considering their metadata and information on sequencing platform used – i.e. only samples derived from 454 pyrosequencing and Illumina sequencing were considered. Due to a lack of public availability for Illumina datasets targeting environments of interest, only 454 pyrosequencing datasets were retained. Analysis of related literature resulted in the detection of other deposited studies that previous search efforts in NCBI-SRA failed to detect. Further private contacts allowed access to unpublished data included in this study. The final list of NCBI accession numbers, totalling 222 samples, was downloaded using *fastq-dump* from the SRA toolkit (https://hpc.nih.gov/apps/sratoolkit.html)

As seen in Table 1, required metadata included host-rock lithology, general and specific geographical locations, depth of sampling, DNA extraction method, sequenced 16S rRNA gene region and sequencing method. Any samples for which the above-mentioned metadata could not be found were discarded and not considered for downstream analyses.

Pre-processing of 16S rRNA gene datasets

A customized pipeline was created in *bash* language making use of *python* scripts developed for QIIME v1.9.1, to facilitate bioinformatic analyses in this study (see https://github.com/GeoMicroSoares/mads_scripts for scripts) [42]. Briefly, demultiplexed FASTQ files were processed to create an operational taxonomic unit (OTU) table. Quality control steps involved trimming, quality-filtering and chimera checking by means of USEARCH 6.1 [43]. Sequence data that passed quality control were then subjected to closed-reference (CR) OTU-picking on a per-study basis using UCLUST and reverse strand matching against the SILVA v123 taxonomic references (https://www.arb-silva.de/documentation/release-123/) [43]. CR OTU picking excludes OTUs whose taxonomy has not been found in the 16S rRNA gene database used. Although this limits the recovery of prokaryotic diversity to that recorded in the database, cross-study comparisons of bacterial communities generated by different 16S rRNA gene primers are made possible. This conservative approach classified OTUs in each study individually to the common 16S rRNA gene reference database from the merging of all classification outputs. A single BIOM (Biological Observation Matrix) file was generated using QIIME's *merge_otu_tables.py* script. The BIOM file was then filtered to exclude samples represented by fewer than two OTUs using *filter_samples_from_otu_table.py,* as well as OTUs represented by one sequence (singleton OTUs) by using *filter_otus_from_otu_table.py*. In an attempt to reduce the impacts of potential contaminant OTUs from the dataset, the post-singleton filtered dataset was further filtered to include only OTUs represented by at least 500 sequences and present in at least 10 samples overall using *filter_otus_from_otu_table.py*.

Data analysis

All downstream analyses were conducted using the *phyloseq* (https://github.com/joey711/phyloseq) package within R, which allowed for simple handling of metadata and taxonomy and abundance data [44–46]. Merged and filtered BIOM files were imported into R using internal *phyloseq* functions, which allowed further filtering, transformation and plotting of the dataset (see https://github.com/GeoMicroSoares/mads_scripts for scripts). Briefly, following a general assessment of the number of reads across samples and OTUs, *tax_glom* (*phyloseq*) allowed the agglomeration of the OTU table at the phylum level. For the metadata category-directed analyses, the function *merge_samples* (*phyloseq*) created averaged OTU tables, which permitted testing of hypotheses for whether geology or depth had significant impacts on bacterial community structure and composition. Computation of a Jensen–Shannon divergence PCoA (principal coordinate analysis) was achieved with ordinate (*phyloseq*), which makes use of metaMDS (*vegan*) [47, 48]. All figures were plotted via the *ggplot2* R package (https://github.com/tidyverse/ggplot2), except for the UpsetR plot in Fig. S4, which was plotted with the package *UpSetR* (https://github.com/hms-dbmi/UpSetR).

RESULTS

A total of 233 publicly available subsurface samples targeting multiple 16S rRNA gene hypervariable regions originating in nine countries were originally downloaded from the NCBI SRA database. These accounted for 24632035 chimera-checked sequences [11, 27, 49–53], which underwent SILVA 123-aided CR OTU-picking. The discovery of 46 OTUs classified as Chloroplast (*Cyanobacteria*) and phototrophic members of the phyla *Chloroflexi* and *Chlorobi* as well as orders *Rhodospirillales* and *Chromatiales* (*Alpha-* and *Gammaproteobacteria* classes, respectively) justified the use of additional stricter contamination-aware filtering (see Methodology, Table S1 for differences in numbers of reads between methods).

The final dataset consisted of 70 samples and 2207 OTUs (513929 sequences). Seventeen aquifers were included that were associated with either sedimentary- or crystalline-host rocks, from depths spanning 94–2300 m below the land surface, targeting mostly groundwater across five countries (Table S2). Nine DNA extraction techniques were used in these studies, ranging from standard and modified kit protocols (e.g. MOBIO PowerSoil) to phenol–chloroform and CTAB/NaCl-based methods [50, 51, 54–57]. Six different primer pair amplified regions of the 16S rRNA gene with 454 pyrosequencing technology were used to generate the datasets (see Fig. S1). Metadata variables that were unavailable for all samples in the dataset were excluded from the statistical analyses. All studies followed aseptic sample handling protocols and included DNA extraction and PCR controls (for further information see Methods sections of the papers enumerated in Table 1) as per recommended guidelines for the subsurface microbiology community [38, 58].

Among a total of 45 detected bacterial phyla, *Proteobacteria* were seen to dominate most deep subsurface community profiles in this dataset (Fig. 1). The most abundant proteobacterial classes (*Alpha-*, *Beta-*, *Delta-*, *Gammaproteobacteria*) represented 57.2% of the total number of reads. *Betaproteobacteria*, chiefly represented by the order *Burkholderiales*, accounted for 26.1% of all reads in the dataset. The order *Burkholderiales* was the main component of some host-rocks, accounting for up to 59.5 and 92.7% of host-rock-level relative abundance profiles for biotite-gneiss and chlorite-sericite-schist (see Fig. S2 for standard deviations of Fig. 1) and co-dominated others. *Gammaproteobacteria* and *Clostridia* (*Firmicutes*) were key components of other

Table 1. Metadata table for the studies utilized in this meta-analysis (see Table S3, Fig. S3 for more details)

NA, Not available. The dataset unavailable through SRA is available in JGI GOLD Study ID Gs0047444.

SRA accession	DOI	Year(s) of sampling	Final no. of samples	Location	Depth gradient (m)	Host rock	Final no. of sequences	Final no. of OTUs	Reference
PRJNA262938	10.3389/fmicb.2014.00610	2013/2014	6	South Dakota, USA	243.84–1478.28	Sulphide-rich schists	170 364	1367	[50]
PRJNA268940	NA	2007–2011	8	California, USA; Ontario, Canada	94–2383	Dolomite, tuff, rhyolite/tuff-breccia	68 071	741	NA
PRJNA248749	NA	2011	6	Minnesota, USA	730	Haematite iron formation	69 757	511	NA
PRJNA251746	NA	2009	6	Washington, USA	393–1135	Basalt	30 121	613	NA
PRJNA375701	10.1016/j.coal.2016.05.001	2013	4	Montana, USA	109–114.7	Sub-bituminous coal, shale, sandstone, siltstone	35 926	154	[51]
NA	10.1128/AEM.01737–15	2009	27	Alberta, Canada	140–1064.4	Sub-bituminous coal, volatile bituminous coal	100 618	5110	[52]
PRJEB1468	10.1111/1574–6941.12171	NA	6	Mol, Belgium	217–232	Kaolinite/illite and smectite	47 123	497	[27]
PRJEB10822	10.5194/bg-13-3091-2016	2009–2011	7	Outokumpu, Finland	180–2300	Mica-schist, biotite-gneiss, chlorite–sericite-schist	12 290	177	[11]

	Sandstone	Dolomite	Siltstone	Shale	Kaolinite/illite and smectite	Biotite-gneiss	Hematite	Mica-schist	Chlorite-sericite-schist	Sulphide-rich schists	Basalt	Tuff	Rhyolite-tuff-breccia	Volatile bituminous coal	Sub-bituminous coal
Alphaproteobacteria	2.4	2.2	0.4	5.9	17.7	0	3.8	0	0.2	1.8	11.2	8.8	1.9	11.6	1.9
Burkholderiales	48.6	10.7	34.4	27.4	52	59.5	0.2	43	92.7	3.8	22.8	15.9	1.1	2.1	36.4
Hydrogenophilales	0	0.1	0	0	0	0	0	0	0	2.5	0	9.9	0.5	0	0.2
Methylophilales	0	0	0	0	3.1	0	0	0	0	0.8	8.7	0	0	0.1	0.2
Nitrosomonadales	1.4	0	0	0	0	0	19.3	0	0	15.9	0	0	0	0	0.1
Rhodocyclales	0.6	1.1	1.2	1	5.7	0	0	0	0	1.8	2.6	4.9	0	0.9	16.8
Other Betaproteobacteria	0	0.2	0.5	0.5	0	0	0	0.5	0	0	0.6	0	0.4	0.1	0
Deltaproteobacteria	6	5.3	0	0.1	0.2	0	8.4	0	0	13.7	18.5	1.3	0	7	1.1
Gammaproteobacteria	5.6	0.6	4.4	7.6	3.6	1.1	40.2	44.1	0.6	16.5	17.2	16.8	2.8	50.2	33.6
Other Proteobacteria	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0
Bacilli	20.7	0.2	37.3	21.5	0	0.1	0	1.1	0	1.1	0.4	2.5	25	2.1	4.3
Clostridia	5.1	61.4	19	31.9	1	0.7	19.9	7.8	6.6	5.8	0.5	11	0.5	11	3.2
Other Firmicutes	0	0	0	0	0.7	0	0	0	0	0	2.6	0.3	0	0.1	0
Actinobacteria	8.1	0.8	0.4	0.1	2.1	38.6	1.5	1.3	0	1.4	7.8	0.2	58.9	3.1	0.4
Other (<3%)	0	0.4	0	0	0.3	0	0.2	0	0	0.8	0.1	0.7	0.2	0.3	0

Averaged Relative Abundance (%)
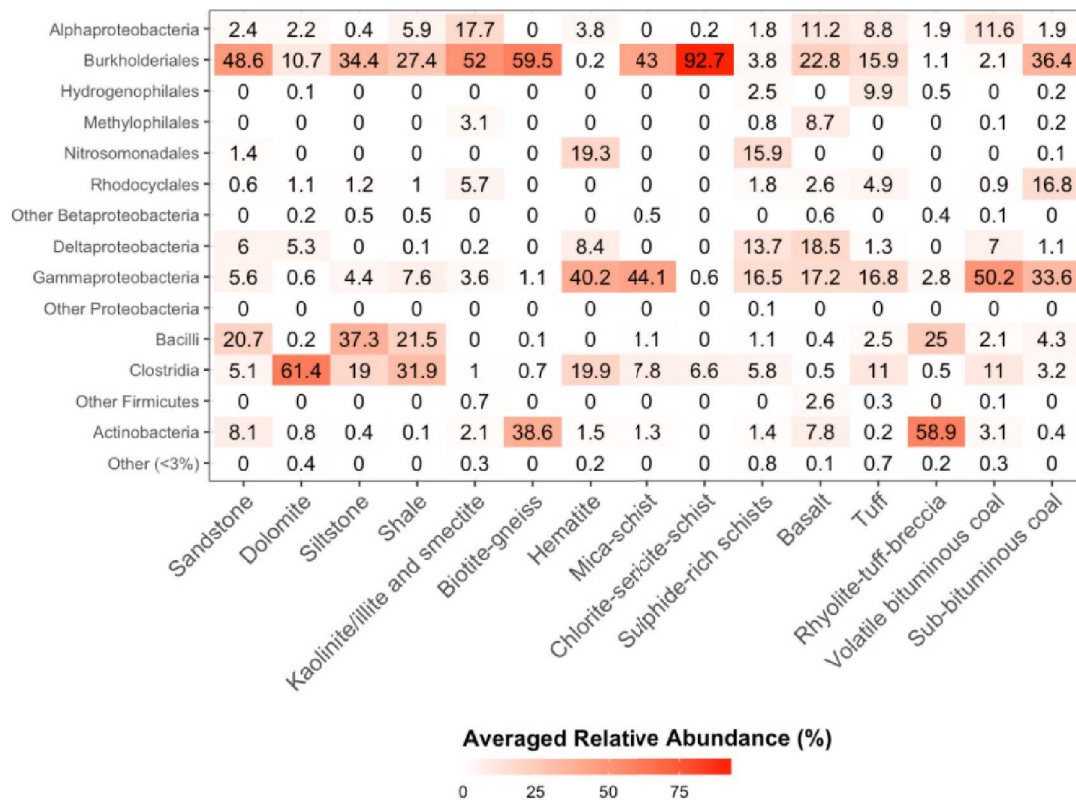
0   25   50   75

Fig. 1. Averaged relative abundances (coloured by increasing percentage abundance) of the most abundant taxonomic groups (*y*-axis) across the dataset across all analysed aquifer lithologies (*x*-axis).

profiles. *Clostridia* and other *Firmicutes* accounted for large fractions of sedimentary host-rocks (dolomite, siltstone and shale) and a haematite iron formation. Finally, *Actinobacteria* was the most abundant taxonomic group in rhyolite-tuff-breccia.

Analysis of prevalence across the dataset revealed that seven OTUs, all affiliated with the genus *Pseudomonas*, were present in more than 25 and up to 41 samples, accounting for 18149 reads (3.5% of the total reads, see Fig. 2, Table S3). Other bacterial orders, namely *Burkholderiales*, *Alteromonadales* and *Clostridiales* (*Betaproteobacteria*, *Gammaproteobacteria*, *Clostridia*) were also highly prevalent throughout. Network analysis (Table 2) highlighted a *Pseudomonas* OTU highly connected to other OTUs in the dataset. Furthermore, BLAST results indicated that recovered sequences for OTUs affiliated with this genus were generally
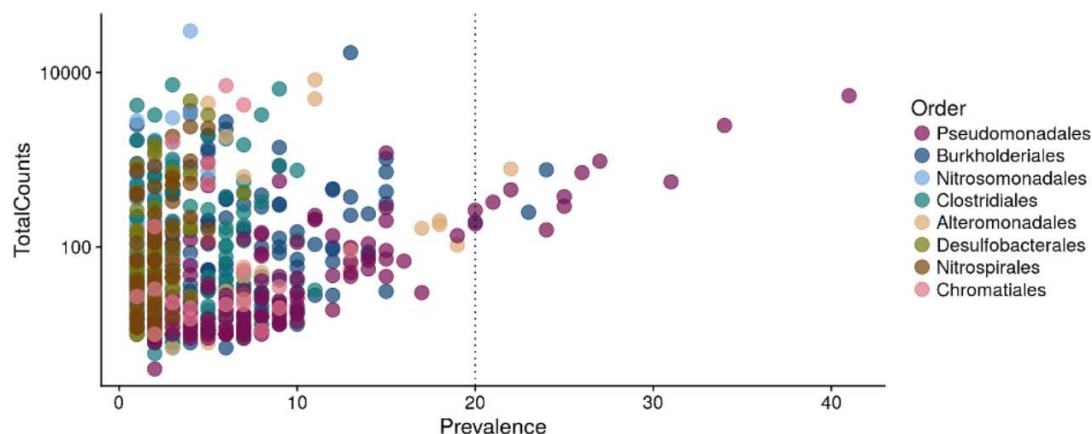
Fig. 2. Prevalence (number of samples in which an OTU is present, *x*-axis) of OTUs across the dataset and associated reads (*y*-axis). Colours depict classification of OTUs at the order level. The vertical line is at 20 samples on the *x*-axis to highlight OTUs present in 20 or more samples.

Table 2. Top 10 most central OTUs (Proteobacterial classes highlighted) in the Jaccard distances network (as defined by eigenvector centrality scores, or the scored value of the centrality of each connected neighbour of an OTU) and corresponding closeness centrality (scores of shortest paths to and from an OTU to all the remaining OTUs in a network) and degree (number of directly connected edges, or OTUs) values

OTU ID	OTU classification	Centrality	Closeness	Degree
EF554871.1.1486	*Proteobacteria;* ***Gammaproteobacteria****; Pseudomonadales; Pseudomonadaceae; Pseudomonas*	1.0000000	2.13e-05	38
HH792638.1.1492	*Proteobacteria;* ***Betaproteobacteria****; Rhodocyclales; Rhodocyclaceae; Thauera*	0.9753542	2.13e-05	36
HQ681977.1.1496	*Proteobacteria;* ***Betaproteobacteria****; Burkholderiales; Comamonadaceae; Diaphorobacter*	0.9445053	2.13e-05	34
KF465077.1.1336	*Proteobacteria;* ***Betaproteobacteria****; Burkholderiales; Comamonadaceae; Acidovorax*	0.8887751	2.13e-05	30
JQ072853.1.1348	*Proteobacteria;* ***Betaproteobacteria****; Rhodocyclales; Rhodocyclaceae; Thauera*	0.8808435	2.13e-05	30
KM200734.1.1449	*Proteobacteria;* ***Alphaproteobacteria****; Rhizobiales; Rhizobiaceae; Rhizobium*	0.8716886	2.13e-05	31
KC758926.1.1392	*Proteobacteria;* ***Betaproteobacteria****; Burkholderiales; Comamonadaceae; Acidovorax*	0.8662805	2.13e-05	29
FJ032194.1.1456	*Proteobacteria;* ***Betaproteobacteria****; Burkholderiales; Comamonadaceae; Rhodoferax*	0.8662805	2.13e-05	29
EU771645.1.1366	*Firmicutes; Bacilli; Bacillales; Planococcaceae; Planomicrobium*	0.8476970	2.13e-05	30
JN245782.1.1433	*Proteobacteria;* ***Alphaproteobacteria****.; Rhodobacterales; Rhodobacteraceae; Defluviimonas*	0.8356655	2.13e-05	29

associated with marine and terrestrial soil and sediments (see Fig. S3, Table S4) [59]. Four OTUs affiliated to *Burkholderiales* (*Betaproteobacteria*), the second most prevalent order in the dataset, were also found to be connected to up to 34 other OTUs. The genus *Thauera* (*Betaproteobacteria, Rhodocyclales*), represented by a single OTU, was the second most central to the dataset.

While relative abundance patterns across the dataset (Fig. 1) indicate that lithology could influence microbial community composition and structure, sample sizes for each host-rock in the final dataset were insufficient to provide robust statistical support of that hypothesis. Despite this, host-rocks (10 out of 15) presented, on average, more unique OTUs than they shared with other host-rocks (Fig. S4). In particular, in sulphide-rich schists, 73% of the OTUs were, on average, unique to the host-rock. Sub-bituminous and volatile bituminous coals shared a total of 143 OTUs; this was the strongest interaction between host-rocks in the dataset. No significant correlations were found for the presence of the most abundant clades in the dataset and depth, *Actinobacteria* being the only major taxonomic group to have a positive, albeit weak, correlation with depth (Pearson's $r=0.42$, $P<0.01$, Fig. S5). Proportions of *Beta-* and *Gammaproteobacteria* generally decreased with depth (Pearson's $r=-0.29$ and -0.093, respectively), but no other major clades were shown to correlate.

Ordination of the final dataset further suggests 50.6% of Jensen–Shannon distances were significantly explained by aquifer lithology (ADONIS/PERMANOVA, F-statistic=4.65, $P<0.001$, adjusted Bonferroni correction $P<0.001$). Other environmental features such as absolute depth and medium-scale location (i.e. state, region of the sampling site) explained only 3.08 and 2.78% of the significant metadata-driven variance in bacterial community structure, respectively (ADONIS/PERMANOVA, F-statistic=3.95, 3.57, $P<0.001$, adjusted Bonferroni correction $P<0.001$). Finally, no evidence was found for DNA extraction or 16S rRNA gene region significantly affecting bacterial community structure in this meta-analysis (ADONIS/PERMANOVA, F-statistic=3.85, 3.23, $P<0.01$, adjusted Bonferroni correction $P<0.001$).

DISCUSSION

The deep biosphere is an active, diverse biome still largely under-investigated in terms of the Earth's biogeochemistry [12, 60–62]. In this study, publicly available 16S rRNA gene data revealed a prevalence of *Betaproteobacteria* and *Gammaproteobacteria* in the deep biosphere that may be explained by the diverse metabolic capabilities of taxa within these clades. The families *Gallionellaceae, Pseudomonadaceae, Rhodocyclaceae* and *Hydrogeniphillaceae* within *Betaproteobacteria* and *Gammaproteobacteria* are suggested to play critical roles in deep subsurface iron, nitrogen, sulphur and carbon cycling across the world [50, 61, 63]. The relative abundance of the order *Burkholderiales* (*Betaproteobacteria*) in surficial soils has previously been correlated ($R^2=0.92$, ANOVA $P<0.005$) with mineral dissolution rates, while the genus *Pseudomonas* (*Gammaproteobacteria*) is widely known to play a key role in hydrocarbon degradation, denitrification and coal solubilization in different locations [64–66]. The dominance of *Betaproteobacteria* and *Gammaproteobacteria* in coals builds on culture-based evidence of widespread degradation of coal-associated complex organic compounds by these classes [67–70].

The metabolic plasticity of the orders *Pseudomonadales* and *Burkholderiales* has been demonstrated and may be a catalyst for their apparent centrality across the terrestrial deep subsurface microbiomes analysed in this study [71–74]. These bacterial orders may represent important keystone taxa in microbial consortia responsible for providing critical substrates to other colonizers in deep subsurface environments [75, 76]. In particular, given the number of highly central *Pseudomonas*-affiliated OTUs and

the prevalence of this genus in the dataset, we suggest that this genus may play a central role in establishing conditions for microbial colonization in many terrestrial subsurface environments. The genus *Pseudomonas* and possibly several members of *Burkholderiales* may therefore comprise an important component of the global core terrestrial deep subsurface bacterial community [11]. Geographically comprehensive RNA-based approaches should in the future investigate the potential roles of the genus *Pseudomonas* and order *Burkholderiales* in this biome.

The class *Clostridia* was found to be prevalent across the dataset and to dominate in sedimentary host-rocks (dolomite, siltstone and shale) in this study. This class includes anaerobic hydrogen-driven sulphate reducers also known to sporulate and metabolize a wide range of organic carbon compounds [77]. Previously, members of *Clostridia* have also been identified as dominant components in extremely deep subsurface ecosystems beneath South Africa, Siberia and California (USA) from metabasaltic and metasedimentary lithologies [78]. Adaptation to extreme environments in this class has been associated with diverse metabolic capabilities that include sporulation ability and capacity for CO_2- or sulphur-based autotrophic H_2-dependent growth [21, 79]. In this study, network analysis and prevalence values suggested roles of putative importance for the classes *Betaproteobacteria*, *Gammaproteobacteria* and *Clostridia* in the deep terrestrial subsurface (Fig. 2, Table S3). Their maintained presence in this biome across strikingly dissimilar host-rocks and depth, among others, could be indicative of higher metabolic plasticity, providing physiological advantages over other members of microbial communities.

Lithotrophic microbial metabolisms and mineralogy-driven microbial colonization of relatively inert lithologies have previously been demonstrated with low abundance but more reactive minerals within rock matrices often cited as key controls on community structure [80–83]. Limiting factors for life in the terrestrial deep subsurface such as pressure and temperature are more closely correlated with depth. Growth of bacterial isolates from the deep subsurface has been documented at up to 48 MPa and 50 °C and has been associated with production of extracellular polymeric substances (EPS) [84, 85]. However, robust conclusions on the effects of lithology or depth on the structure and composition of microbial communities across Earth's crust have presented a widespread challenge for science, as in this study due to the small and varied sample sizes resulting from the contamination-aware filtering process and the limited number of comparable lithology types. Large-scale evidence for the roles of eukaryotes, bacteria, archaea and viruses in the deep terrestrial subsurface and the environmental controls over their occurrence in this biome is still lacking. We recommend a field-level research strategy to gain insights into these aspects of life within Earth's crust. Larger scale collation of data from samples collected and processed using unified, reproducible workflows will be cognizant of significant potential for contamination and ultimately allow robust insights on wide-ranging microbial metabolic processes in the terrestrial subsurface.

Collecting contamination-free samples from the deep subsurface is difficult but important for cataloguing the authentic microbial diversity of the terrestrial subsurface. This study follows recent recommendations for downstream processing of contaminant-prone samples originated in the deep subsurface (Census of Deep Life project – https://deepcarbon.net/tag/census-deep-life), where physical, chemical and biological, but also *in silico* bioinformatics strategies to prevent erroneous conclusions have been highlighted [38, 86, 87]. This study also follows frequency-based OTU filtration techniques similar to those recommended previously [38] and designed to remove possible contaminants introduced during sampling or during the various steps related to sample processing [38]. The pre-emptive quality control steps hereby undertaken support a non-contaminant origin for the taxa analysed in this dataset. As such, the predominance of typically contaminant taxa affiliated, for example, to the genus *Pseudomonas* is accepted as a representative trend in reflecting the microbial ecology of the terrestrial deep subsurface.

Standardizing sampling, DNA extraction, sequencing and bioinformatics methods and strategies across the subsurface research community would help further reduce methodology-based variations. This would more efficiently permit re-analyses after collection, where methodological variations would be controlled, and robust wide-ranging overarching conclusions would more easily be achieved. Despite this, host-rock matrices and local geochemical conditions often pose unique challenges that require particular protocol adjustments [88]. In the near future, the advent of recently developed techniques for primer bias-free long read 16S rRNA and 16S rRNA-ITS gene amplicon long-read-based sequencing may initiate a convergence of molecular methods from which the deep subsurface microbiology community would benefit greatly [89, 90]. The future of large-scale, collaborative deep subsurface microbial diversity studies should encompass not only an effort towards standardization of several molecular biology techniques but also the long-term archival of samples [91]. Finally, the ecology of domains Eukarya and Archaea across the terrestrial deep subsurface remains generally under-characterized and requires future attention. This study presents an important first step towards characterizing bacterial community structure and composition in the terrestrial deep biosphere.

A global-scale meta-analysis addressing the available 16S rRNA gene-based studies of the deep terrestrial subsurface revealed a dominance of *Betaproteobacteria*, *Gammaproteobacteria* and *Firmicutes* across this biome. Evidence for a core terrestrial deep subsurface microbiome population was recognized through the prevalence and centrality of the genus *Pseudomonas* (*Gammaproteobacteria*) and several other genera affiliated with the class *Betaproteobacteria*. The adaptable metabolic capabilities associated with the above-mentioned taxa may be critical for colonizing the deep subsurface and sustaining communities therein. The terrestrial deep subsurface is a hard-to-reach, complex ecosystem crucial to global biogeochemical cycles. Efforts by multiple teams of investigators to sequence subsurface ecosystems over the last decade were hereby consolidated to characterize the 12–20%

of global biomass this biome represents [5]. The strict contamination-aware filtering process applied whittled down the publicly available datasets representing terrestrial subsurface bacterial diversity to just 70 samples from two continents, indicating the need for systematic exploration of biodiversity within this major component of the biosphere. As a first step, this study consolidates a global-scale understanding of taxonomic trends underpinning a major component of terrestrial deep subsurface microbial ecology and biogeochemistry.

Author contributions
A.R.S. developed the methodology, collated and analysed the data, and wrote the manuscript. A.E. and A.M. conceived the study, supervised A.S. and helped write the manuscript. Other authors provided data from field sites used in the global meta-analysis. All authors contributed, edited and approved the final manuscript.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. **Jelen BI**, **Giovannelli D**, **Falkowski PG**. The role of microbial electron transfer in the coevolution of the biosphere and geosphere. *Annu Rev Microbiol* 2016;70:45–62.

2. **Falkowski PG**, **Fenchel T**, **Delong EF**. The microbial engines that drive earths biogeochemical cycles. *Science* 2008;320:1034–1039.

3. **McMahon S**, **Parnell J**. The deep history of earth's biomass. *J Geol Soc London* 2018;175:716–720.

4. **Magnabosco C**, **Lin L-H**, **Dong H**, **Bomberg M**, **Ghiorse W**, *et al*. The biomass and biodiversity of the continental subsurface. *Nature Geosci* 2018;11:707–717.

5. **Bar-On YM**, **Phillips R**, **Milo R**, **Falkowski PG**. The biomass distribution on Earth. *Proc Natl Acad Sci U S A* 2018;115:6506–6511.

6. **Flemming H-C**, **Wuertz S**. Bacteria and archaea on Earth and their abundance in biofilms. *Nat Rev Microbiol* 2019;17:247–260.

7. **Smith HJ**, **Zelaya AJ**, **De León KB**, **Chakraborty R**, **Elias DA**, *et al*. Impact of hydrologic boundaries on microbial planktonic and biofilm communities in shallow terrestrial subsurface environments. *FEMS Microbiol Ecol* 2018;94.

8. **Miettinen H**, **Kietäväinen R**, **Sohlberg E**, **Numminen M**, **Ahonen L**, *et al*. Microbiome composition and geochemical characteristics of deep subsurface high-pressure environment, Pyhäsalmi mine Finland. *Front Microbiol* 2015;6:1–16.

9. **Bagnoud A**, **Chourey K**, **Hettich RL**, **de Bruijn I**, **Andersson AF**, *et al*. Reconstructing a hydrogen-driven microbial metabolic network in Opalinus Clay rock. *Nat Commun* 2016;7:12770.

10. **Dong Y**, **Kumar CG**, **Chia N**, **Kim P-J**, **Miller PA**, *et al*. Halomonas sulfidaeris-dominated microbial community inhabits a 1.8 km-deep subsurface Cambrian Sandstone reservoir. *Environ Microbiol* 2014;16:1695–1708.

11. **Purkamo L**, **Bomberg M**, **Kietäväinen R**, **Salavirta H**, **Nyyssönen M**, *et al*. Microbial co-occurrence patterns in deep Precambrian bedrock fracture fluids. *Biogeosciences* 2016;13:3091–3108.

12. **Long PE**, **Williams KH**, **Hubbard SS**, **Banfield JF**. Microbial metagenomics reveals climate-relevant subsurface biogeochemical processes. *Trends Microbiol* 2016;24:600–610.

13. **Anantharaman K**, **Brown CT**, **Hug LA**, **Sharon I**, **Castelle CJ**, *et al*. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat Commun* 2016;7:1–11.

14. **Hug LA**, **Thomas BC**, **Sharon I**, **Brown CT**, **Sharma R**, *et al*. Critical biogeochemical functions in the subsurface are associated with bacteria from new phyla and little studied lineages. *Environ Microbiol* 2016;18:159–173.

15. **Purkamo L**, **Kietäväinen R**, **Miettinen H**, **Sohlberg E**, **Kukkonen I**, *et al*. Diversity and functionality of archaeal, bacterial and fungal communities in deep Archaean bedrock groundwater. *FEMS Microbiol Ecol* 2018;94.

16. **Baker BJ**, **Moser DP**, **MacGregor BJ**, **Fishbain S**, **Wagner M**, *et al*. Related assemblages of sulphate-reducing bacteria associated with ultradeep gold mines of South Africa and deep basalt aquifers of Washington State. *Environ Microbiol* 2003;5:267–277.

17. **Chivian D**, **Brodie EL**, **Alm EJ**, **Culley DE**, **Dehal PS**, *et al*. Environmental genomics reveals a single-species ecosystem deep within Earth. *Science* 2008;322:275–278.

18. **Gihring TM**, **Moser DP**, **Lin L-H**, **Davidson M**, **Onstott TC**, *et al*. The distribution of microbial taxa in the subsurface water of the Kalahari Shield, South Africa. *Geomicrobiol J* 2007;23:415–430.

19. **Moser DP**, **Gihring T**, **Fredrickson JK**, **Brockman FJ**, **Balkwill D**. *Desulfotomaculum* spp. and *Methanobacterium* spp. dominate 4-5 km deep fault. *Appl Environ Microbiol* 2005;71:8773–8783.

20. **Ueno A**, **Shimizu S**, **Tamamura S**, **Okuyama H**, **Naganuma T**, *et al*. Anaerobic decomposition of humic substances by Clostridium from the deep subsurface. *Sci Rep* 2016;6:2016.

21. **Sousa DZ**, **Visser M**, **van Gelder AH**, **Boeren S**, **Pieterse MM**, *et al*. The deep-subsurface sulfate reducer *Desulfotomaculum kuznetsovii* employs two methanol-degrading pathways. *Nat Commun* 2018;9:239.

22. **Brazelton WJ**, **Morrill PL**, **Szponar N**, **Schrenk MO**. Bacterial communities associated with subsurface geochemical processes in continental serpentinite springs. *Appl Environ Microbiol* 2013;79:3906–3916.

23. **Nuppunen-Puputti M**, **Purkamo L**, **Kietäväinen R**, **Nyyssönen M**, **Itävaara M**, *et al*. Rare biosphere archaea assimilate acetate in precambrian terrestrial subsurface at 2.2 km depth. *Geosciences (Basel)* 2018;8:418.

24. **Hubalek V**, **Wu X**, **Eiler A**, **Buck M**, **Heim C**, *et al*. Connectivity to the surface determines diversity patterns in subsurface aquifers of the Fennoscandian shield. *ISME J* 2016;10:2447–2458.

25. Vieira-Silva S, Falony G, Darzi Y, Lima-Mendez G, Garcia Yunta R, et al. Species–function relationships shape ecological properties of the human gut microbiome. Nat Microbiol 2016;1:16088.

26. Emerson JB, Thomas BC, Alvarez W, Banfield JF. Metagenomic analysis of a high carbon dioxide subsurface microbial community populated by chemolithoautotrophs and bacteria and archaea from candidate phyla. Environ Microbiol 2016;18:1686–1703.

27. Wouters K, Moors H, Boven P, Leys N. Evidence and characteristics of a diverse and metabolically active microbial community in deep subsurface clay borehole water. FEMS Microbiol Ecol 2013;86:458–473.

28. Unal B, Perry VR, Sheth M, Gomez-Alvarez V, Chin K-J, et al. Trace elements affect methanogenic activity and diversity in enrichments from subsurface coal bed produced water. Front Microbiol 2012;3:1–14.

29. Hernsdorf AW, Amano Y, Miyakawa K, Ise K, Suzuki Y, et al. Potential for microbial H_2 and metal transformations associated with novel bacteria and archaea in deep terrestrial subsurface sediments. ISME J 2017;11:1915–1929.

30. Lopez-Fernandez M, Broman E, Turner S, Wu X, Bertilsson S, et al. Investigation of viable taxa in the deep terrestrial biosphere suggests high rates of nutrient recycling. FEMS Microbiol Ecol 2018;94.

31. Lloyd KG, Steen AD, Ladau J, Yin J, Crosby L. Phylogenetically novel uncultured microbial cells dominate Earth microbiomes. mSystems 2018;3:e00055-18.

32. Sherwood Lollar B, Lacrampe-Couloume G, Slater GF, Ward J, Moser DP, et al. Unravelling abiogenic and biogenic sources of methane in the Earth's deep subsurface. Chemical Geology 2006;226:328–339.

33. Lin L-H, Wang P-L, Rumble D, Lippmann-Pipke J, Boice E, et al. Long-term sustainability of a high-energy, low-diversity crustal biome. Science 2006;314:479–482.

34. Li L, Wing BA, Bui TH, McDermott JM, Slater GF, et al. Sulfur mass-independent fractionation in subsurface fracture waters indicates a long-standing sulfur cycle in Precambrian rocks. Nat Commun 2016;7:13252.

35. Lollar GS, Warr O, Telling J, Osburn MR, Sherwood Lollar B. (n.d.) 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;In review

36. Gilbert A, Sherwood Lollar B, Musat F, Giunta T, Chen S, et al. Intramolecular isotopic evidence for bacterial oxidation of propane in subsurface natural gas reservoirs. Proc Natl Acad Sci U S A 2019;116:6653–6658.

37. Griffin WT, Phelps TJ, Colwell FS, Fredrickson JK. Methods for obtaining deep subsurface microbiological samples by drilling. In: Microbiology of the Terrestrial Deep Subsurface. CRC Press, 2018. pp. 23–44.

38. Sheik CS, Reese BK, Twing KI, Sylvan JB, Grim SL, et al. Identification and removal of contaminant sequences from ribosomal gene databases: Lessons from the Census of Deep Life. Front Microbiol 2018;9:840.

39. Wilkins MJ, Daly RA, Mouser PJ, Trexler R, Sharma S, et al. Trends and future challenges in sampling the deep terrestrial biosphere. Front Microbiol 2014;5:1–8.

40. Penny SA, Dana LH. The Microbiology of the Terrestrial Deep Subsurface. CRC Press, 2018. pp. 75–102.

41. Russell BF, Phelps TJ, Griffin WT, Sargent KA. Procedures for sampling deep subsurface microbial communities in unconsolidated sediments. Groundwater Monitoring & Remediation 1992;12:96–104.

42. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 2010;7:335–336.

43. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 2010;26:2460–2461.

44. McMurdie PJ, Holmes S. Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE 2013;8:e61217.

45. . Wickham H, Chang W. ggplot2: An Implementation of the Grammar of Graphics.

46. R Development Core Team. (n.d.) R: A language and environment for statistical computing.

47. Dixon P. VEGAN, a package of R functions for community ecology. J Veg Sci 2003;14:927–930.

48. Fuglede B, Topsoe F. Jensen-Shannon divergence and Hilbert space embedding. In: International Symposium OnInformation Theory, 2004. ISIT 2004. Proceedings. IEEE. Chicago, Illinois, USA, 2004. p. 30.

49. Frank YA, Kadnikov VV, Gavrilov SN, Banks D, Gerasimchuk AL, et al. Stable and variable parts of microbial community in Siberian deep subsurface thermal aquifer system revealed in a long-term monitoring study. Front Microbiol 2016;7:1–15.

50. Osburn MR, LaRowe DE, Momper LM, Amend JP. Chemolithotrophy in the continental deep subsurface: Sanford Underground Research Facility (SURF). Front Microbiol 2014;5:1–14.

51. Barnhart EP, Weeks EP, Jones EJP, Ritter DJ, McIntosh JC, et al. Hydrogeochemistry and coal-associated bacterial populations from a methanogenic coal bed. International Journal of Coal Geology 2016;162:14–26.

52. Lawson CE, Strachan CR, Williams DD, Koziel S, Hallam SJ, et al. Patterns of endemism and habitat selection in coalbed microbial communities. Appl Environ Microbiol 2015;81:7924–7937.

53. Bomberg M, Lamminmäki T, Itävaara M. Microbial communities and their predicted metabolic characteristics in deep fracture groundwaters of the crystalline bedrock at Olkiluoto, Finland. Biogeosciences 2016;13:6031–6047.

54. Andreou LV. Preparation of genomic DNA from bacteria. In: Methods in Enzymology. Hoboken, NJ, USA: John Wiley & Sons, Inc, 2013. pp. 143–151.

55. Foght J, Aislabie J, Turner S, Brown CE, Ryburn J, et al. Culturable bacteria in subglacial sediments and ice from two southern hemisphere glaciers. Microb Ecol 2004;47:329–340.

56. Tillett D, Neilan BA. Xanthogenate nucleic acid isolation from cultured and environmental cyanobacteria. Journal of Phycology 2001;36:251–258.

57. Leuko S, Goh F, Ibáñez-Peral R, Burns BP, Walter MR, et al. Lysis efficiency of standard DNA extraction methods for Halococcus spp. in an organic rich environment. Extremophiles 2008;12:301–308.

58. Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R, et al. Contamination in low microbial biomass microbiome studies: issues and recommendations. Trends Microbiol 2019;27:105–117.

59. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403–410.

60. Bomberg M, Nyyssönen M, Pitkänen P, Lehtinen A, Itävaara M. Active microbial communities inhabit sulphate-methane interphase in deep bedrock fracture fluids in Olkiluoto, Finland. Biomed Res Int 2015;2015:979530.

61. Rajala P, Bomberg M. Reactivation of deep subsurface microbial community in response to methane or methanol amendment. Front Microbiol 2017;8:431.

62. Lollar GS, Warr O, Telling J, Osburn MR, Lollar BS. 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 2019;36:859–872.

63. Blanco Y, Rivas LA, García-Moyano A, Aguirre J, Cruz-Gil P, et al. Deciphering the prokaryotic community and metabolisms in South African deep-mine biofilms through antibody microarrays and graph theory. PLoS ONE 2014;9:e114180.

64. Singh PK, Singh AL, Kumar A, Singh MP. Control of different pyrite forms on desulfurization of coal with bacteria. Fuel 2013;106:876–879.

65. Machnikowska H, Pawelec K, Podgórska A. Microbial degradation of low rank coals. *Fuel Processing Technology* 2002;77–78:17–23.

66. Lepleux C, Turpault MP, Oger P, Frey-Klett P, Uroz S. Correlation of the abundance of betaproteobacteria on mineral surfaces with mineral weathering in forest soils. *Appl Environ Microbiol* 2012;78:7114–7119.

67. Gründger F, Jiménez N, Thielemann T, Straaten N, Lüders T, *et al*. Microbial methane formation in deep aquifers of a coal-bearing sedimentary basin, Germany. *Front Microbiol* 2015;6:200.

68. Yagi JM, Sims D, Brettin T, Bruce D, Madsen EL. The genome of *Polaromonas naphthalenivorans* strain CJ2, isolated from coal tar-contaminated sediment, reveals physiological and metabolic versatility and evolution through extensive horizontal gene transfer. *Environ Microbiol* 2009;11:2253–2270.

69. Kostka JE, Prakash O, Overholt WA, Green SJ, Freyer G, *et al*. Hydrocarbon-degrading bacteria and the bacterial community response in gulf of Mexico beach sands impacted by the deepwater horizon oil spill. *Appl Environ Microbiol* 2011;77:7962–7974.

70. Posman KM, DeRito CM, Madsen EL. Benzene degradation by a variovorax species within a coal tar-contaminated groundwater microbial community. *Appl Environ Microbiol* 2017;83:e02658-16.

71. Alhasawi A, Costanzi J, Auger C, Appanna ND, Appanna VD. Metabolic reconfigurations aimed at the detoxification of a multi-metal stress in *Pseudomonas fluorescens*: Implications for the bioremediation of metal pollutants. *J Biotechnol* 2015;200:38–43.

72. Raiger Iustman LJ, Tribelli PM, Ibarra JG, Catone MV, Solar Venero EC, *et al*. Genome sequence analysis of *Pseudomonas extremaustralis* provides new insights into environmental adaptability and extreme conditions resistance. *Extremophiles* 2015;19:207–220.

73. Uroz S, Calvaruso C, Turpault M-P, Frey-Klett P. Mineral weathering by bacteria: ecology, actors and mechanisms. *Trends Microbiol* 2009;17:378–387.

74. Rosenberg E. The family *Comamonadaceae*. In: *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria*. 2013. pp. 1–1012.

75. Onstott TC, Moser DP, Pfiffner SM, Fredrickson JK, Brockman FJ, *et al*. Indigenous and contaminant microbes in ultradeep mines. *Environ Microbiol* 2003;5:1168–1191.

76. Davidson MM, Silver BJ, Onstott TC, Moser DP, Gihring TM, *et al*. Capture of planktonic microbial diversity in fractures by long-term monitoring of flowing boreholes, evander basin, South Africa. *Geomicrobiology Journal* 2011;28:275–300.

77. Dürre P. Clostridia. In: *Encyclopedia of Life Sciences*. Chichester, UK: John Wiley & Sons, Ltd, 2015. pp. 1–11.

78. Moser DP, Gihring TM, Brockman FJ, Fredrickson JK, Balkwill DL, *et al*. Desulfotomaculum and *Methanobacterium* spp. dominate a 4- to 5-kilometer-deep fault. *Appl Environ Microbiol* 2005;71:8773–8783.

79. Aüllo T, Ranchou-Peyruse A, Ollivier B, Magot M. *Desulfotomaculum* spp. and related gram-positive sulfate-reducing bacteria in deep subsurface environments. *Front Microbiol* 2013;4:1–12.

80. Mitchell AC, Lafrenière MJ, Skidmore ML, Boyd ES. Influence of bedrock mineral composition on microbial diversity in a subglacial environment. *Geology* 2013;41:855–858.

81. Lapanje A, Wimmersberger C, Furrer G, Brunner I, Frey B. Pattern of elemental release during the granite dissolution can be changed by aerobic heterotrophic bacterial strains isolated from Damma Glacier (central Alps) deglaciated granite sand. *Microb Ecol* 2012;63:865–882.

82. Cockell CS, Voytek MA, Gronstal AL, Finster K, Kirshtein JD, *et al*. Impact disruption and recovery of the deep subsurface biosphere. *Astrobiology* 2012;12:231–246.

83. Boyd ES, Cummings DE, Geesey GG. Mineralogy influences structure and diversity of bacterial communities associated with geological substrata in a pristine aquifer. *Microb Ecol* 2007;54:170–182.

84. Fichtel K, Logemann J, Fichtel J, Rullkötter J, Cypionka H, *et al*. Temperature and pressure adaptation of a sulfate reducer from the deep subsurface. *Front Microbiol* 2015;6:1078.

85. Booker AE, Hoyt DW, Meulia T, Eder E, Nicora CD, *et al*. Deep-subsurface pressure stimulates metabolic plasticity in shale-colonizing *Halanaerobium* spp. *Appl Environ Microbiol* 2019;85:e00018-19.

86. Sogin M, Edwards K. Deep subsurface microbiology and the deep carbon observatory. In: *DCO Deep Life Workshop*. Catalina Island, California, 2010.

87. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* 2018;6:226.

88. Direito SOL, Marees A, Röling WFM. Sensitive life detection strategies for low-biomass environments: Optimizing extraction of nucleic acids adsorbing to terrestrial and Mars analogue minerals. *FEMS Microbiol Ecol* 2012;81:111–123.

89. Karst SM, Dueholm MS, McIlroy SJ, Kirkegaard RH, Nielsen PH, *et al*. Retrieval of a million high-quality, full-length microbial 16S and 18S rRNA gene sequences without primer bias. *Nat Biotechnol* 2018;36:190–195.

90. Martijn J, Lind AE, Spiers I, Juzokaite L, Bunikis I, *et al*. Amplicon sequencing of the 16S-ITS-23S rRNA operon with long-read technology for improved phylogenetic classification of uncultured prokaryotes. *Microbiology* 2017.

91. Fierer N, Cary C. Don't let microbial samples perish. *Nature* 2014;512:253.

Edited by: J. Cavet