Biogeosciences Discuss., 12, 13819–13857, 2015
www.biogeosciences-discuss.net/12/13819/2015/
doi:10.5194/bgd-12-13819-2015

Biogeosciences Discussions
Open Access

Estimation of microbial metabolism and co-occurrence patterns in fracture groundwaters of deep crystalline bedrock at Olkiluoto, Finland

M. Bomberg[1], **T. Lamminmäki**[2], **and M. Itävaara**[1]

[1]VTT Technical Research Centre of Finland, P.O. Box 1000, 02044 VTT, Espoo, Finland
[2]Posiva Oy, Olkiluoto, 27160 Eurajoki, Finland

Received: 7 July 2015 – Accepted: 22 July 2015 – Published: 26 August 2015

Correspondence to: M. Bomberg (malin.bomberg@vtt.fi)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

The microbial diversity in oligotrophic isolated crystalline Fennoscandian Shield bedrock fracture groundwaters is great but the core community has not been identified. Here we characterized the bacterial and archaeal communities in 12 water conductive fractures situated at depths between 296 and 798 m by high throughput amplicon sequencing using the Illumina HiSeq platform. The great sequencing depth revealed that up to 95 and 99 % of the bacterial and archaeal communities, respectively, were composed of only a few common species, i.e. the core microbiome. However, the remaining rare microbiome contained over 3 and 6 fold more bacterial and archaeal taxa. Several clusters of co-occurring rare taxa were identified, which correlated significantly with physicochemical parameters, such as salinity, concentration of inorganic or organic carbon, sulphur, pH and depth. The metabolic properties of the microbial communities were predicted using PICRUSt. The rough prediction showed that the metabolic pathways included commonly fermentation, fatty acid oxidation, glycolysis/gluconeogenesis, oxidative phosphorylation and methanogenesis/anaerobic methane oxidation, but carbon fixation through the Calvin cycle, reductive TCA cycle and the Wood-Ljungdahl pathway was also predicted. The rare microbiome is an unlimited source of genomic functionality in all ecosystems. It may consist of remnants of microbial communities prevailing in earlier conditions on Earth, but could also be induced again if changes in their living conditions occur. In this study only the rare taxa correlated with any physicochemical parameters. Thus these microorganisms can respond to environmental change caused by physical or biological factors that may lead to alterations in the diversity and function of the microbial communities in crystalline bedrock environments.

BGD
12, 13819–13857, 2015

1 Introduction

Identifying and understanding the core microbiome of any given environments is of crucial importance for predicting and assessing environmental change both locally and globally (Shade and Handelsman, 2012). So far, a core microbiome has not been identified in deep Fennoscandian subsurface environments. For example, in a previous study (Bomberg et al., 2015) we showed by 454 amplicon sequencing that the active microbial communities in Olkiluoto deep subsurface were strictly stratified according to aquifer water type. Nevertheless, more rigorous sequencing efforts and more samplings have shown that an archaeal core community consisting of the DeepSea Hydrothermal Vent Euryarchaeotal Group 6 (DHVEG-6), ANME-2D and Terrestrial Miscellaneous Group (TMEG) archaea may exsists in the anaerobic deep groundwater of Olkiluoto (Miettinen et al., 2015). The bacterial core groups in Olkiluoto deep groundwater include at least members of the Pseudomonadaceae, Commamonadaceae and Sphingomonadaceae (Bomberg et al., 2014, 2015; Miettinen et al., 2015). The relative abundance of these main groups varies at different depths from close to the detection limit to over 90 % of the bacterial or archaeal community (Bomberg et al., 2015; Miettinen et al., 2015). However, both the archaeal and the bacterial communities contain a wide variety of smaller bacterial and archaeal groups, which are distributed unevenly in the different water conductive fractures.

The rare biosphere is a concept describing the hidden biodiversity of an environment and has been suggested to be ancient (Sogin et al., 2006). The rare biosphere consists of microbial groups that are ubiquitously distributed in nature but rarely found. Due to modern high throughput sequencing techniques, however, the hidden diversity of rare microbiota has been revealed. These microorganisms are the basis for unlimited microbial functions in the environment and upon environmental change specific groups can readily activate and become abundant. Access to otherwise inaccessible nutrients activate specific subpopulations in the bacterial communities within hours of exposure (Rajala et al., 2015) and enrich distinct microbial taxa at the expense of the original

BGD
12, 13819–13857, 2015

microbial community in the groundwater (Kutvonen, 2015). Mixing of different groundwater layers due to e.g. breakage of aquifer boundaries and new connection of separated aquifers may cause the microbial community to change and activate otherwise dormant processes. This has previously been shown by Pedersen et al. (2013), who indicated increased sulphate reduction activity when sulphate-rich and methane-rich groundwater mixed. The stability of deep subsurface microbial communities in isolated deep subsurface groundwater fractures are assumed to be stable. However, there are indications that they may change over the span of several years as slow flow along fractures is possible (Miettinen et al., 2015; Sohlberg et al., 2015).

The microbial taxa present in an environment interact with both biotic and abiotic factors. Co-occurrence network analyses and metabolic predictions may help to understand these interactions. Barberan et al. (2012) visualised the co-occurrence networks of microbial taxa in soils and showed novel patterns connecting generalist and specialist species as well as associations between microbial taxa. They showed that specialist and generalist microbial taxa formed distinct and separate correlation networks, which also reflected the environmental settings. Metagenome predicting tools allows us to estimate microbial metabolic functions based on NGS microbiome data. Using the PICRUSt tool (Langille et al., 2013) Tsitko et al. (2014) showed that oxidative phosphorylation was the most important energy producing metabolic pathway throughout the 7 m depth profile of an Acidobacteria-dominated nutrient poor boreal bog. Cleary et al. (2015) showed that tropical mussel-associated bacterial communities could be important sources of bioactive compounds for biotechnology. This approach is nevertheless hampered by the fact that only little is so far known about uncultured environmental microorganisms and their functions and the PICRUSt approach is best applied for human microbiome for which it was initially developed (Langille et al., 2013). However, metagenomic estimations may give important indications of novel metabolic possibilities even in environmental microbiome studies.

Using extensive high throughput amplicon sequencing in this study we aimed to identify the core microbiome in the deep crystalline bedrock fractures of Olkiluoto Island

BGD
12, 13819–13857, 2015

and also to identify the rare microbiome. We aimed to show the interactions between the taxa of the rare biosphere and the surrounding environmental parameters in order to validate the factors that determine the distribution of the rare taxa. Finally we aimed to estimate the prevailing metabolic activities that may occur in the deep crystalline bedrock environment of Olkiloto, Finland.

2 Materials and methods

2.1 Background

The Olkiluoto site has previously been extensively described (Posiva, 2013) and is only briefly described here. The Island of Olkiluoto situating on the west coast of Finland has approximately 60 drillholes drilled for research and monitoring purposes. Studies on the chemistry and microbiology of the groundwater have been on-going since the 1980s. The groundwater is stratified with a salinity gradient extending from fresh to brackish water to a depth of 30 m and the highest salinity concentration of 125 $g L^{-1}$ total dissolved solids (TDS) at 1000 m depth (Posiva, 2013). The most abundant salinity causing cations are Na^{2+} and Ca^{2+} and anions Cl^{-}. Between 100 and 300 m depths, the groundwater originates from ancient (pre-Baltic) seawater and has high concentrations of SO_4^{2-}. Below 300 m the concentration of methane in the groundwater increases and SO_4^{2-} is almost absent. A sulphate-methane transition zone (SMTZ), where sulphate-rich fluid replaces methane-rich fluid, is located at 250–350 m depth. Temperature rises linearly with depth, from ca. 5–6 °C at 50 m to ca. 20 °C at 1000 m depth (Ahokas et al., 2008). The pH of the groundwater is slightly alkaline throughout the depth profile. Multiple drillholes intersect several groundwater-filled bedrock fractures, including larger hydrogeological zones such as HZ20 or HZ21 (Table 1). The bedrock of Olkiluoto consists mainly of micagneiss and pegmatitic granite type rocks (Kärki and Paulamäki, 2006).

BGD
12, 13819–13857, 2015

This study focused on 12 groundwater samples from water conductive fractures situated at between 296 and 798 m below sea level bsl and originating from 11 different drillholes in Olkiluoto. The samples represented brackish sulphate waters and saline waters (as classified in Posiva, 2013). The samples were collected between December 2009 and January 2013 (Table 1). The physicochemical parameters of the groundwater samples have been reported by reported by Miettinen et al. (2015), but have for clarity been collected here (Table 1).

2.2 Sample collection

The collection of samples was described previously (Bomberg et al., 2015; Miettinen et al., 2015; Sohlberg et al., 2015). The samples were obtained from both permanently packered drillholes and open drillholes, with which removable inflatable packers were used. Shortly, in order to obtain indigenous fracture fluids, the packer-isolated fracture zones were purged by removing stagnant drillhole water by pumping for a minimum of four weeks before the sample water was collected. The water samples were collected directly from the drillhole into an anaerobic glove box (MBRAUN, Germany) via a sterile, gas-tight poly acetate tube (8 mm outer diameter). Microbial biomass DNA extraction was concentrated from 1000 mL samples by filtration on cellulose acetate filters (0.2 µm pore size, Corning) by vacuum suction inside the glove box. The filters were immediately extracted from the filtration funnels and frozen on dry ice in sterile 50 mL cone tubes (Corning). The frozen samples were transported on dry ice to the laboratory where they were stored at −80 °C until use.

2.3 Nucleic acid isolation

Community DNA was isolated directly from the frozen cellulose-acetate filters with the PowerSoil DNA extraction kit (MoBio Laboratories, Inc., Solana Beach, CA), as previously described (Bomberg et al., 2015). Negative DNA isolation controls were included

BGD
12, 13819–13857, 2015

in the isolation protocol. The DNA concentration of each sample was determined using the NanoDrop 1000 spectrophotometer.

2.4 Estimation of microbial community size

The size of the microbial community was determined by epifluorescence microscopy of 4′,6 diamidino-2-phenylindole dihydrochloride (DAPI) (Sigma, MO, USA) stained cells as described in Purkamo et al. (2013). The size of the bacterial population was determined by 16S rRNA gene targeted quantitative PCR (qPCR) as described by Tsitko et al. (2014) using universal bacterial 16S rRNA gene-targeting primers fD1 (Weisburg et al., 1991) and P2 (Muyzer et al., 1993), which specifically target the V1–V3 region of the bacterial 16S rDNA gene. The size of the archaeal population in the groundwater was determined by using primers ARC344f (Bano et al., 2004) and Ar744r (reverse compliment from Barns et al., 1994) flanking the V4–V6 region of the archaeal 16S rRNA gene.

The qPCR reactions were performed in 10 µL reaction volumes using the KAPA 2 × Syrb® FAST qPCR-kit on a LightCycler480 qPCR machine (Roche Applied Science, Germany) on white 96-well plates (Roche Applied Science, Germany) sealed with transparent adhesive seals (4titude, UK). Each reaction contained 2.5 µM of relevant forward and reverse primer and 1 µL DNA extract. Each reaction was run in triplicate and no-template control reactions were used to determine background fluorescence in the reactions.

The qPCR conditions consisted of an initial denaturation at 95 °C for 10 min followed by 45 amplification cycles of 15 s at 95 °C, 30 s at 55 °C and 30 s at 72 °C with a quantification measurement at the end of each elongation. A final extension step of three minutes at 72 °C was performed prior to a melting curve analysis. This consisted of a denaturation step for 10 s at 95 °C followed by an annealing step at 65 °C for one minute prior to a gradual temperature rise to 95 °C at a rate of 0.11 °C s^{-1} during which the fluorescence was continuously measured. The number of bacterial 16S rRNA genes was determined by comparing the amplification result (Cp) to that of a ten-fold dilu-

BGD
12, 13819–13857, 2015

tion series (10^1–10^7 copies μL^{-1}) of *Escherichia coli* (ATCC 31608) 16S rRNA genes in plasmid for bacteria and a dilution series of genomic DNA of *Halobacterium salinarum* (DSM 3754) for archaea. The lowest detectable standard concentration for the qPCRs was 10^2 gene copies/reaction. Inhibition of the qPCR by template tested by adding 2.17×10^4 plasmid copies containing fragment of the morphine-specific Fab gene from *Mus musculus* gene to reactions containing template DNA as described in Nyyssönen et al. (2012). Inhibition of the qPCR assay by the template DNA was found to be low. The average Crossing point (Cp) value for the standard sample (2.17×10^4 copies) was 28.7 (±0.4 SD), while for the DNA samples Cp was 28.65–28.91 (±0.03–0.28 SD). Nucleic acid extraction and reagent controls were run in all qPCRs in parallel with the samples. Amplification in these controls was never higher than the background obtained from the no template controls.

2.5 Amplicon library preparation

This study is part of the Census of Deep Life initiative, which strives to obtain a census of the microbial diversity in deep subsurface environment by collecting samples around the world and sequencing the 16S rRNA gene pools of both archaea and bacteria. The extracted DNA samples were sent to the Marine Biological Laboratory in Woods Hole, MA, USA, for preparation for HiSeq sequencing using the Illumina technology. The protocol for amplicon library preparation for both archaeal and bacterial 16S amplicon libaries can be found at http://vamps.mbl.edu/resources/faq.php. Shortly, amplicon libraries of the V6 region of both the archaeal and bacterial 16S rRNA genes were produced. For the archaea, primers A958F and A1048R containing Truseq adapter sequences at their 5′ end were used, and for the bacteria primers B967F and B1064R for obtaining 100 nt long paired end reads.

BGD
12, 13819–13857, 2015

2.6 Sequence processing and analysis

Contigs of the paired end fastq files were first assembled with mothur v 1.32.1 (Schloss et al., 2009). Analyzes were subsequently continued using QIIME v. 1.8. (Caporaso et al., 2010). Only sequences with a minimum length of 50 bp were included in the analyses. The bacterial and archaeal 16S rRNA sequences were grouped into OTUs (97 % sequence similarity) using both the open reference and closed reference OTU picking strategy and classified using the GreenGenes 13_8 16S reference database (DeSantis et al., 2006). The sequencing coverage was evaluated by rarefaction analysis and the estimated species richness and diversity indices were calculated. For comparable α- and β-diversity analyses the data sets were normalized by random subsampling of 17 000 sequences/sample for archaea and 140 000 sequences/sample for bacteria. Microbial metabolic pathways were estimated based on the 16S rRNA gene data from the closed OTU picking method using the PICRUSt software (Langille et al., 2013) on the web based Galaxy application (Goecks et al., 2010; Blankenberg et al., 2010; Giardine et al., 2005). The sequence data has been submitted to the Sequence Read Archive (SRA, http://www.ncbi.nlm.nih.gov/sra) under study SRP053854, Bioproject PRJNA275225.

2.7 Statistical analyses

Pearson's *r* correlation between biological and geochemical/-physical factors and canonical correspondence analysis was calculated using PAST3 (Hammer and Harper, 2001). Pearson's *r* correlation between the occurrence and abundance of the archaeal and bacterial genera in each sample was tested using the out.association command in mothur. Correlation pairs of microbial genera with $p < 0.01$ were read into the Gephi software (Bastian et al., 2009) for visualization of a correlation network between the significant biological groups.

BGD
12, 13819–13857, 2015

3 Results

3.1 Microbial community size

The total number of microbial cells detected by epifluorescence microscopy of DAPI stained cells was between 2.3×10^4 and 4.2×10^5 cells mL^{-1} groundwater (Fig. 1, Table 1). The concentration of bacterial 16S rRNA gene copies mL^{-1} varied between 9.5×10^3 and 7.0×10^5 and that of the archaea 2.6×10^1 and 6.3×10^4 (Fig. 1, Table 1).

3.2 Sequence statistics, diversity estimates and sequencing coverage

The number of bacterial v6 sequence reads from the 12 samples varied between 1.4–7.8×10^5 reads, with a mean sequencing depth of 2.9×10^5 ($\pm 1.8 \times 10^5$ standard deviation) reads/sample (Table 2). The archaeal v6 sequence reads ranged from 0.17–12.1×10^5 reads with a mean sequencing depth of 4.1×10^5 ($\pm 3.5 \times 10^5$ standard deviation) reads/sample. Compared to the actual numbers of observed operational taxonomic units (OTUs), on average 82.6 % (±12.5 %) of the Chao1 and 78.1 % (±13.4 %) of the ACE estimated numbers of bacterial OTUs were detected (Table 2a, b). The coverage between samples ranged from 57.3–99.4 and 51.6–97.6 % of the Chao1 and ACE estimated OTU richness, respectively. The archaeal communities were slightly better covered, with on average 88.5 % (±11.5 %) of the Chao1 and 84.8 % (±12.6 %) of the ACE estimated number of OTUs detected. The coverage between samples ranged between 61.6–99.4 and 56.8–98.1 % of the Chao1 and ACE estimated OTU richness, respectively. Shannon diversity index H′, calculated from 140 000 and 17 000 random sequence reads per sample for the bacteria and archaea, respectively, was high for both bacterial and archaeal communities. High H′ values and climbing rarefaction curves (Fig. 2) indicated high diversity in the microbial communities in the different deep groundwater fracture zones of Olkiluoto. The bacterial H′ was on average 13 (±0.74), ranging from 11 to 14 between the different samples. The archaeal H′ was on average 11 (±1.2) ranging from 9 to 12 between the samples.

BGD
12, 13819–13857, 2015

From the bacterial v6 sequences 49 different bacterial Phyla were detected (Supplement 1). These phyla included 165 bacterial classes, 230 orders, 391 families and 651 genera. The greatest number of sequences, between 21.83 and 47.94 % per sample, clustered into an undetermined bacterial group (Bacteria, Other), which may be due the fact that sequences of poorer quality may be difficult to classify, especially as the sequences are short.

Only 31 of the identified genera represented at least 1 % of the bacterial sequence reads in any sample (Table 3).

The archaea were represented by two identified phyla, the Euryarchaeota and the Crenarchaeota (Supplement 2). These included 21 classes, 38 orders, 61 families and 81 genera. Between 4.7 and 35.0 % of the archaeal sequences of each sample were classified to unassigned Archaea, with a general increase in unassigned archaeal sequences with increasing depth. 15 archaeal genera were present at a minimum of 1 % relative abundance in any of the samples (Table 4).

3.3 Core communities

The bacterial core community, i.e. the taxa occurring in all the tested samples, constituted of 95 out of 651 identified bacterial genera (Supplement 3). These genera accounted for 80.78–95.81 % of all the bacterial sequence reads in the samples. The archaeal core community consisted of 25 of the 81 identified genera and accounted for 95.05–99.75 % of the total number of sequence reads in each sample (Supplement 4).

3.4 Impact of environmental parameters on the microbial communities

The general bacterial and archaeal communities in the different samples corresponded mostly to sulphate and total sulphur concentrations at depths between 328 and 423 m as shown by the canonical correspondence analysis (Fig. 3). The community for the deepest sampling depth at 798 m corresponded mostly to K and Fe, but also to salinity. In general the communities corresponded most strongly to the type of groundwater they

BGD
12, 13819–13857, 2015

originated from. When examining the different microbial groups most of the detected bacterial and archaeal genera did not correlate significantly with any of the measured parameters, because most of the genera were present at very low relative abundance and distributed evenly throughout the depth profile (Supplements 1 and 2). However, of the dominating (> 1 % relative abundance) genera Actinobacteria/Other correlated positively and significantly ($r > 0.7$, $p < 0.005$) with N_{tot}, NPOC, Fe(II) (Tables 5–8). The genus *Clostridium* also correlated positively and significantly with N_{tot} and NPOC, while the unassigned Gammaproteobacteria/Other correlated positively and significantly with Fe(II). Enterobacteriaceae/Other and Escherichia correlated positively and significantly with N_{tot}. *Microbacterium* was the only major detected genus that correlated positively and significantly with pH and Flavobacteriaceae/Other the only major bacterial genus to correlate positively and significantly with Fe_{tot}. Caulobacteraceae/Other and *Sulfuricurvum* correlated positively and significantly with sulphide.

Of the major archaeal genera Methanobacteriales/Other, MSBL1/Other and SAGMEG-1 correlated positively and significantly with sampling depth, EC, TDS, Ca, Cl and Na (Tables 5–8). In addition, Methanobacteriales/Other correlated positively and significantly with N_{tot}, while MSBL1/Other and SAGMEG-1 together with Crenarchaeota/Other correlated positively and significantly with bicarbonate, DIC, and Fe_{tot}. Thermoplasmata E2/Other correlated positively and significantly with Alkalinity and Fe(II).

The concentration of sulphide correlated significantly with the greatest number of different bacterial genera, while the archaeal community did not show any correlation with sulphide (Table 8). Instead different minor groups of ANME archaea correlated positively and significantly with the concentration of sulphate and S_{tot}. Deltaproteobacterial SRB correlated greatly with Fe(II). N_{tot} and NPOC affected numerous bacterial and archaeal clades.

BGD
12, 13819–13857, 2015

3.5 Co-occurrence network

Of the 651 bacterial and 81 archaeal genera (or equivalent groups) identified in this study 42 bacteria and 59 archaeal genera showed any significant correlation with other genera. These groups all represented rare biosphere groups. These were divided into 7 distinct communities (Fig. 4). The majority of the co-occurring rare groups showed positive correlation to the same environmental factors clustering into communities occurring in relation to total N and organic C concentrations (community 1), sulphur, sulphate and Fe(II) (communities 2, 3), salinity (communities 3–7), inorganic C (communities 3, 4, 6, 7), organic C (community 5) and depth (communities 5, 6).

3.6 Predicted metabolic functions of the deep subsurface microbial communities

The putative metabolic functions of the microbial communities at different depth was predicted using the PICRUSt software, which compares the identified 16S rRNA gene sequences to those of known genome sequenced species thereby estimating the possible gene contents of the uncultured microbial communities. The analysis is only an approximation, but may give an idea of the possible metabolic activities in the deep biosphere. In order to evaluate the soundness of the analysis a nearest sequenced taxon index (NSTI) for each of the bacterial and archaeal communities was calculated by PICRUSt. An NSTI value of 0 indicates high similarity to the closest sequenced taxon while NSTI = 1 indicates no similarity. The NSTI of the bacterial communities at different depths varied between 0.045 in sample OL-KR44 and 0.168 in sample OL-KR13 (Table 9a). The NSTI for archaea were much higher ranging from 0.141 in sample OL-KR9 at depth of 432 m and 0.288 in OL-KR44. This indicates that the metagenomic estimates are very rough. The estimated microbial metabolism did not differ noticeably between the different depths (Table 9b). The most important predicted metabolic pathways included membrane transport in both bacterial and archaeal communities. The most common pathways for carbohydrate metabolism were the butanoate, propi-

BGD
12, 13819–13857, 2015

onate, glycolysis/gluconeogenesis and pyruvate metabolism pathways for the bacteria and glycolysis/gluconeogenesis and pyruvate metabolism pathways for the archaea. The most abundant energy metabolic pathway in the bacterial communities was the oxidative phosphorylation while for the archaea the methane metabolism was the most important.

4 Discussion

The phenotypic characteristics of the Fennoscandian Shield deep subsurface microbial communities are still largely unknown although specific reactions to introduced environmental stimulants have been shown (e.g. Pedersen et al., 2013, 2014; Rajala et al., 2015; Kutvonen, 2015). Nevertheless, the connection of these microbial responses to specific microbial groups is still only in an early phase. Metagenomic and gene specific analyses of deep subsurface microbial communities have revealed prominent metabolic potential of the microbial communities, which appear to be associated with the prevailing lithology and physicochemical parameters (Nyyssönen et al., 2014; Purkamo et al., 2015). It has also been shown with fingerprinting methods with ever increasing efficiency that the bacterial and archaeal communities are highly diverse in the saline anaerobic Fennoscandian deep fracture zone groundwater (Bomberg et al., 2014, 2015; Nyyssönen et al., 2012, 2014; Pedersen et al., 2014; Miettinen et al., 2015; Sohlberg et al., 2015). Nevertheless, the concentration of microbial cells in the groundwater is quite low (Fig. 1, Table 1). Most of the microbial communities at different depth in Olkiluoto bedrock fractures consist of bacteria. However, at specific depths (328, 423 m) the archaea may contribute with over 50 % of the estimated 16S rRNA gene pool (Table 1). The major archaeal group present at these depths were the ANME-2D archaea indicating that nitrate-mediated anaerobic oxidation of methane may be especially common (Haroon et al., 2013).

Previously, using 454 amplicon sequencing, we have observed OTU numbers of approximately 800 OTUs per sample covering approximately 550 bacterial genera (or

BGD
12, 13819–13857, 2015

equivalent groups) and approximately 350 archaeal OTUs including approximately 80 different genera (or equivalent groups) (Miettinen et al., 2015). The OTUs in the study above were defined at 97 % sequence homology and the number of sequence reads per sample was at most in the range of 10^4. However, in this study the number of sequence reads was 10- to 100-fold higher and the number of OTUs per sample in general 100-fold higher. This indicates that a greater sequencing depth increases the number of taxa detected from the subsurface environment and allows us a novel view of the so far hidden rare biosphere. Nevertheless, in comparison to the high number of OTUs detected the number of identified genera, 651 and 81 bacterial and archaeal genera, respectively, seems low. On the other hand this indicates that the sequencing depth has been sufficient to detect most of the prokaryotic groups present.

In general, the microbial communities at different depth grouped loosely into clusters according to the groundwater chemistry (Fig. 3). The clearest clustering was observed for samples derived from between 328 and 423 m depth where total sulphur and sulphate concentrations influenced the population. Individual bacterial and archaeal groups on the other hand showed strong positive correlation to the different geochemical parameters (Tables 5–8). However, most of the bacterial and archaeal groups correlating with any of the measured geochemical parameters belonged to the rare biosphere, i.e. low abundance and sporadic appearance in the bacterial or archaeal communities. The core communities, defined as taxa present in all the studied samples, accounted for between 80–97 % and 95–> 99 % of the bacterial and archaeal communities, respectively. This is a considerable frequency of common microbial taxa. Nevertheless, the number of rare taxa detected from the sample set was 3.3 to 6.8 fold higher than the number of core taxa on genus level. Our results agree with Sogin et al. (2006), who showed that a relatively small number of taxa dominate deep-sea water habitats, but a rare microbiome consisting of thousands of taxonomically distinct microbial groups are detected at low abundances. What this means for the functioning of the deep subsurface is that the microbial communities have the capacity to respond and change due to changes in environmental conditions. For example, Peder-

BGD
12, 13819–13857, 2015

sen et al. (2014) showed that by adding sulphate to the sulphate-poor but methane-rich groundwater in Olkiluoto the bacterial population changed over the span of 103 days from a non-SRB community to a community dominated by SRB. In addition, a change in the geochemical environment induced by H_2 and methane impacted the size, composition and functions of the microbial community and ultimately led to acetate formation (Pedersen et al., 2012, 2014; Pedersen, 2013).

The metabolic pathways predicted by PICRUSt are far from certain when uncultured and unculturable deep subsurface microbial communities are concerned. The NSTI values for both the bacterial and well as the archaeal communities were great indicating that no closely related species have yet been sequenced. However, on higher taxonomical level common traits for specific groups of microorganisms may be revealed.

4.1 Energy metabolism

Deep subsurface environments are often declared energy deprived environments dominated by autotrophic microorganisms (Hoehler and Jorgensen, 2013). However, recent reports indicate that heterotrophic microorganisms play a greater role than the autotrophic microorganisms in Fennoscandian deep crystalline subsurface environments (Purkamo et al., 2015). Heterotrophic communities with rich fatty acid assimilation strategies have been reported to fix carbon dioxide on the side of e.g. fermenting activities in order to replenish the intracellular carbon pool, which otherwise would be depleted. Our results agree with Purkamo et al. (2015) that a greater proportion of the microbial community is involved in carbohydrate and fatty and organic acid oxidation than in fixation of inorganic carbon. Nevertheless, autotrophic carbon fixation pathways were predicted in the analysis with PICRUSt, indicating that both the archaeal and bacterial communities include autotrophic members, although these microorganisms might not be obligate autotrophs.

Several carbon fixation pathways were predicted in the archaeal and bacterial communities. The Calvin cycle and the reductive TCA (rTCA) cycle were found in both the archaeal and the bacterial communities. The Wood-Ljungdahl pathway is consid-

BGD
12, 13819–13857, 2015

ered the most ancient autotrophic carbon fixation pathway in bacteria and archaea (Fuchs, 1989; Martin et al., 2008; Berg et al., 2010; Hügler and Sievert, 2011). Despite the long isolation of the ancient groundwater of Olkiluoto the Wood-Ljungdahl pathway was only predicted in the bacterial community. In the archaeal community the Calvin cycle and the rTCA were especially pronounced in the samples from 296 m, 405–423 m and somewhat lower at 510–527 m depth. The bacterial communities are predicted to fix CO_2 at almost all depths with the exception of 405 and 559 m depth. The predicted methane metabolism (methane and methyl compound consumption) and oxidative phosphorylation were equally strong in the bacterial community. Sulphur metabolism was not a common pathway for energy in either the archaeal or the bacterial communities, but bacteria with either assimilative or dissimilative sulphate reduction were present. No sulphur oxidation through the *sox* system was predicted. Ammonia and nitrate appear not to be taken up by the microorganisms or used for energy.

Oxidative phosphorylation was one of the most prominent energy generating metabolic pathways in the bacterial community. This indicates that ATP is generated by electron transfer to a terminal electron acceptor, such as oxygen, nitrate or sulphate. In the archaeal community the oxidative phosphorylation was not as strongly indicated, but this may be due to missing data on archaeal metabolism in the KEGG database.

The main energy metabolism of the archaeal communities appeared to be the methanogenesis, especially at 296 and 405 m. Methanogenesis was common also at all other depths except 330–347, 415 and 693–798 m. Methane is produced from CO_2-H_2 and methanol, and from acetate, although evidence for the acetate kinase enzyme was lacking. Methanogenesis from methylamines may also be possible, especially at 296 and 405 m. Methane oxidation using methane monoxygenases and methanol dehydrogenases does not occur in either bacterial or archaeal communities.

4.2 Carbohydrate metabolism

Glycolysis/gluconeogenesis is one of the most common carbohydrate-metabolizing pathways predicted for both the archaeal and bacterial communities. Pyruvate from

BGD
12, 13819–13857, 2015

glycolysis is oxidized to acetyl-CoA by both archaea and bacteria and used in the TCA cycle. The TCA cycle provides for example raw material for many amino acids, such as lysine and glutamate. The butanoate and propanoate metabolisms were also common in the bacterial communities, indicating fermentative metabolism and capability of fatty acid oxidation.

4.3 Amino acid metabolism

Non-essential amino acids, such as alanine, aspartate and glutamate are produced from ammonia and pyruvate or oxaloacetate especially in the archaeal populations. In the archaeal population proline appears to be produced from glutamate. Despite the low use of sulphate as energy source in the microbial communities sulphate and other sulphur compounds are taken up for the production of the amino acids cysteine and methionine by both the archaeal and the bacterial communities. A higher predicted relative abundance of genes involved in aromatic amino acid synthesis (phenylalanine, tyrosine, tryptophane) was seen in the archaeal than in the bacterial communities. Both the archaeal and the bacterial communities synthesise branched chained amino acids (isoleucine, leucine and valine), but only the bacteria degrade them. Especially proteobacteria have been shown to be able to use the branched chained amino acids (isoleucine, leuscine and valine) and short chained fatty acids (acetate, butyrate, propionate) as sole energy and carbon source (Kazakov et al., 2009). The branched chained amino acids function as raw material in the biosynthesis of branched chained fatty acids, which regulate the membrane fluidity of the bacterial cell. In salt stress conditions, the proportion of branch-chained fatty acids in the membranes decreases.

4.4 Nucleotide metabolism

The estimated number of genes for both the purine and pyrimidine metabolism was more than two times higher in the archaeal community than in the bacterial community.

BGD
12, 13819–13857, 2015

4.5 Membrane transport

According to the predicted metagenomes, the microbial cells transport sulphate into the cell, but do not take up nitrate. Nitrogen is taken up as glutamate but not as urea. Iron is taken up by an Fe(III) transport system and an iron complex transport system in the bacterial communities, but generally only by the iron complex transport system in archaea. However, Fe(III) transport system may also exist in the archaeal communities at 405 to 423 m depth, where also some manganese/iron transport systems could be found. Molybdate and phosphate is transported into the cell by molybdate and phosphate ATPases, respectively. Nickel is taken up mainly by a nickel/peptide transport system but also to some extent by a cobalt/nickel transport system. Zink is taken up to some extent by a zink transport system, but transport systems for manganese, manganese/iron, manganese/zink/iron, or iron/zink/copper are negligent. Ammonia is taken up by an Amt transport system.

5 Conclusions

The wide diversity of microbial groups in the deep Fennoscandian groundwater at the Olkiluoto site revealed that the majority of the microbial community present belong to only a few microbial taxa while the greatest part of the microbial diversity is represented by low abundance and rare microbiome taxa. The core community was present in all tested samples from different depths, but the relative abundance of the different taxa varied in the different samples. Specific rare microbial groups formed tight co-occurrence clusters that corresponded to different environmental conditions and these may become more abundant if the environmental conditions change. Fermentation or oxidation of fatty acids was a common carbon cycling and energy harvesting metabolic pathways in the bacterial communities whereas the archaea may either produce or consume methane. Glycolysis/gluconeogenesis was predicted to be common in both the archaeal and bacterial communities. In addition both the bacterial and archaeal

BGD
12, 13819–13857, 2015

communities were estimated to contain different common carbon fixation pathways, such as the Calvin cycle and the reductive TCA, while only the bacteria contained the Wood-Ljungdahl pathway.

The Supplement related to this article is available online at doi:10.5194/bgd-12-13819-2015-supplement.

Acknowledgements. The Illumina sequencing data were made possible by the Deep Carbon Observatory's Census of Deep Life supported by the Alfred P. Sloan Foundation. Illumina sequencing was performed at the Marine Biological Laboratory (Woods Hole, MA, USA) and we are grateful for the assistance of Mitch Sogin, Susan Huse, Joseph Vineis, Andrew Voorhis, Sharon Grim, and Hilary Morrison at MBL and Rick Colwell, OSU. MB was supported by the Academy of Finland (project 261220).

References

Bano, N., Ruffin, S., Ransom, B., and Hollibaugh, J. T.: Phylogenetic composition of Arctic ocean archaeal assemblages and comparison with Antarctic assemblages, Appl. Environ. Microbiol., 70, 781–789, 2004.

Barns, S. M., Fundyga, R. E., Jeffries, M. W., and Pace, N. R.: Remarkable archaeal diversity detected in Yellowstone National Park hot spring environment, P. Natl. Acad. Sci. USA, 91, 1609–1613, 1994.

Barberán, A., Bates, S. T., Casamayor, E. O., and Fierer, N.: Using network analysis to explore co-occurrence patterns in soil microbial communities, ISME J., 6, 343–351, 2012.

Bastian, M., Heymann, S., and Jacomy, M. Gephi: an open source software for exploring and manipulating networks, ICWSM, 8, 361–362, 2009.

Berg, I. A., Kockelkorn, D., Ramos-Vera, W. H., Say, R. F., Zarzycki, J., Hügler, M., Alber, B. E., and Fuchs, G.: Autotrophic carbon fixation in archaea, Nat. Rev. Microbiol., 8, 447–460, 2010.

BGD
12, 13819–13857, 2015

Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A., and Taylor, J.: Galaxy: a web-based genome analysis tool for experimentalists, Curr. Protoc. Mol. Biol., 19, 19.10.1-21, doi:10.1002/0471142727.mb1910s89, 2010.

Bomberg, M., Nyyssönen, M., Nousiainen, A., Hultman, J., Paulin, L., Auvinen, P., and Itävaara, M.: Evaluation of molecular techniques in characterization of deep terrestrial biosphere, Open J. Ecol., 4, 468–487, 2014.

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

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Gonzalez Pena, A., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Turnbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko, T., Zaneveld, J., and Knight, R.: QIIME allows analysis of high-throughput community sequencing data, Nat. Methods, 7, 335–336, 2010.

Cleary, D. F., Becking, L. E., Polónia, A. R., Freitas, R. M., and Gomes, N. C.: Composition and predicted functional ecology of mussel-associated bacteria in Indonesian marine lakes, Anton. Leeuw. Int. J. G., 107, 821–834, 2015.

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., Dalevi, D., Hu, P., and Andersen, G. L.: Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB, Appl. Environ. Microb., 72, 5069–5072, 2006.

Fuchs, G.: Alternative pathways of autotrophic CO_2 fixation, in: Autotrophic Bacteria, edited by: Schlegel, H. G. and Bowien, B., Springer, Berlin, 365–382, 1989.

Giardine, B., Riemer, C., Hardison, R. C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., Miller, W., Kent, W. J., and Nekrutenko, A.: Galaxy: a platform for interactive large-scale genome analysis, Genome Res., 15, 1451–1455, 2005.

Goecks, J., Nekrutenko, A., Taylor, J., and the Galaxy Team: Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences, Genome Biol., 11, R86, doi:10.1186/gb-2010-11-8-r86, 2010.

Hallbeck, L. and Pedersen, K.: Culture-dependent comparison of microbial diversity in deep granitic groundwater from two sites considered for a Swedish final repository of spent nuclear fuel, FEMS Microbiol. Ecol., 81, 66–77, 2012.

BGD
12, 13819–13857, 2015

Hammer, Ø., Harper, D. A. T., and Ryan, P. D.: PAST: paleontological statistics software package for education and data analysis, Palaeontol. Electron., 4, 1–31, available at: http://palaeo-electronica.org/2001_1/past/issue1_01.htm (last access: 22 August 2015), 2001.
Haroon, M. F., Hu, S., Shi, Y., Imelfort, M., Keller, J., Hugenholtz, P., Yuan, Z., and Tyson, G. W.: Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage, Nature, 500, 567–570, 2013.
Hoehler, T. M. and Jørgensen, B. B.: Microbial life under extreme energy limitation, Nat. Rev. Microbiol., 11, 83–94, 2013.
Hügler, M. and Sievert, S. M.: Beyond the Calvin cycle: autotrophic carbon fixation in the ocean, Mar. Sci., 3, 261–289, doi:10.1146/annurev-marine-120709-142712, 2011.
Kärki, A. and Paulamäki, S.: Petrology of Olkiluoto, Posiva, Olkiluoto, 2006-02, 2006.
Kazakov, A. E., Rodionov, D. A., Alm, E., Arkin, A. P., Dubchak, I., and Gelfand, M. S.: Comparative genomics of regulation of fatty acid and branched-chain amino acid utilization in proteobacteria, J. Bacteriol., 191, 52–64, 2009.
Kutvonen, H.: Nitrogen-cycling bacteria in groundwater of the low and medium active nuclear waste repository in Olkiluoto, Finland, M.Sc. Thesis, Faculty of Agriculture and Forestry, University of Helsinki, Finland, 2015 (in Finnish).
Langille, M. G. I., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., Clemente, J. C., Burkepile, D. E., Vega Thurder, R. L., Knight, R., Beiko, R. G., and Huttenhower, C.: Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences, Nature Biotechnol., 31, 814–821, 2013.
Martin, W., Baross, J., Kelley, D., and Russell, M. J.: Hydrothermal vents and the origin of life, Nat. Rev. Microbiol., 6, 805–814, 2008.
Miettinen, H., Bomberg, M., Nyyssönen, M., Salavirta, H., Sohlberg, E., Vikman, M., and Itävaara, M.: The Diversity of Microbial Communities in Olkiluoto Bedrock Groundwaters 2009–2013. Olkiluoto, Finland, Posiva Oy, Posiva Working Report 2015-12, Olkiluoto, 1–160, 2015.
Muyzer, G., de Waal, E. C., and Uitterlinden, A. G.: Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA, Appl. Environ. Microb., 59, 695–700, 1993.
Nyyssönen, M., Bomberg, M., Kapanen, A., Nousiainen, A., Pitkänen, P., and Itävaara, M.: Methanogenic and sulphate-reducing microbial communities in deep groundwater of crystalline rock fractures in Olkiluoto, Finland, Geomicrobiol. J., 29, 863–878, 2012.

BGD
12, 13819–13857, 2015

Nyyssönen, M., Hultman, J., Ahonen, L., Kukkonen, I. T., Paulin, L., Laine, P., Itävaara, M., and Auvinen, P.: Taxonomically and functionally diverse microbial communities in deep crystalline rocks of the Fennoscandian shield, ISME J., 8, 126–138, 2014.

Pedersen, K.: Subterranean microbial populations metabolize hydrogen and acetate under in situ conditions in granitic groundwater at 450 m depth in the Äspö Hard Rock Laboratory, Sweden, FEMS Microbiol. Ecol., 81, 217–229, 2012.

Pedersen, K.: Metabolic activity of subterranean microbial communities in deep granitic groundwater supplemented with methane and H_2, ISME J., 7, 839–849, 2013.

Pedersen, K., Arlinger, J., Eriksson, S., Hallbeck, A., Hallbeck, L., and Johansson, J.: Numbers, biomass and cultivable diversity of microbial populations relate to depth and borehole-specific conditions in groundwater from depths of 4–450 m in Olkiluoto, Finland, ISME J., 2, 760–775, 2008.

Pedersen, K., Bengsson, A., Edlund, J., and Eriksson, L.: Sulphate-controlled diversity of subterranean microbial communities over depth in deep groundwater with opposing gradients of sulphate and methane, Geomicrobiol. J., 31, 617–631, 2014.

Posiva: Olkiluoto site description 2011, Posiva 2011-02, Olkiluoto, 2013.

Purkamo, L., Bomberg, M., Nyyssönen, M., Kukkonen, I., Ahonen, L., Kietäväinen, R., and Itävaara, M.: Retrieval and analysis of authentic microbial communities from packer-isolated deep crystalline bedrock fractures: evaluation of the method and time of sampling, FEMS Microbiol. Ecol., 85, 324–337, 2013.

Purkamo, L., Bomberg, M., Nyyssönen, M., Kukkonen, I., Ahonen, L., and Itävaara, M.: Heterotrophic communities supplied by ancient organic carbon predominate in deep Fennoscandian bedrock fluids, Microb. Ecol., 69, 319–332, 2015.

Rajala, P., Bomberg, M., Kietäväinen, R., Kukkonen, I., Ahonen, L., Nyyssönen, M., and Itävaara, M.: Deep subsurface microbes rapidly reactivate in the presence of C-1 compounds, Microorganisms, 3, 17–33, 2015.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J., and Weber, C. F.: Introducing MOTHUR: open-source, platform-independent, community-supported software for describing and comparing microbial communities, Appl. Environ. Microb., 75, 7537–7541, 2009.

Shade, A. and Handelsman, J.: Beyond the Venn diagram: the hunt for a core microbiome, Environ. Microbiol., 14, 4–12, 2012.

BGD
12, 13819–13857, 2015

Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., Arrieta, J. M., and Herndl, G. J.: Microbial diversity in the deep sea and the underexplored "rare biosphere", P. Natl. Acad. Sci. USA, 103, 12115–12120, 2006.

Sohlberg, E., Bomberg, M., Miettinen, H., Nyyssönen, M., Salavirta, H., Vikman, M., and Itävaara, M.: Revealing the unexplored fungal communities in deep groundwater of crystalline bedrock fracture zones in Olkiluoto, Finland, Front. Microbiol., 6, 573, doi:10.3389/fmicb.2015.00573, 2015.

Tsitko, I., Lusa, M., Lehto, J., Parviainen, L., Ikonen, A. T. K., Lahdenperä, A.-M., and Bomberg, M.: The variation of microbial communities in a depth profile of an acidic, nutrient-poor boreal bog in southwestern Finland, Open J. Ecol., 4, 832–859, 2014.

Weisburg, W. G., Barns, S. M., Pelletier, D. A., and Lane, D. J.: 16S ribosomal DNA amplification for phylogenetic study, J. Bacteriol., 173, 697–703, 1991.

BGD
12, 13819–13857, 2015

Table 1. Geochemical and microbiological measurements from 12 different water conductive fractures in the bedrock of Olkiluoto, Finland. The different drillholes are presented at the top of the table. The data is compiled from Posiva (2013).

Drillhole	OL-KR13	OL-KR6	OL-KR3	OL-KR23	OL-KR5	OL-KR49	OL-KR9	OL-KR9	OL-KR2	OL-KR1	OL-KR44	OL-KR29
Sampling date	3 Nov 2010	18 May 2010	29 Aug 2011	15 Dec 2009	16 Oct 2012	14 Dec 2009	31 Oct 2011	29 Aug 2011	27 Jan 2010	26 Jan 2010	15 Jan 2013	18 May 2010
Depth (m)	296	328	340	347	405	415	423	510	559	572	693	798
Alkalinity mEq/L	2.19	0.37	0.47	0.05	0.27	0.16	0.18	0.13	0.29	0.23	0.49	0.13
Ec mS/m	897	1832	1047	2190	2240	2670	2300	2960	4110	3770	6690	7820
pH	7.9	7.9	7.9	7.5	7.9	8.1	7.7	8.1	8.6	7.8	7.5	7.3
NPOC mg L^{-1}	10	0	12	5.1	19	3	5.1	6.6	11	5	110	10
DIC mg L^{-1}	27	4.1	4.1	3.9	0	3	3	0	3.75	3.75	6.5	81
HCO_3 mg L^{-1}	134	22.6	25	17.1	16	9.8	11.6	7.3	17.7	14	30	424
N_{tot} mg L^{-1}	0.71	0	1.1	0.42	1.2	0.16	0.38	0.66	1.1	0.41	10	3.1
NH_4^+ mg L^{-1}	0.07	0.03	0.03	0	0	0	0.05	0	0.02	0.04	0.08	0.08
S_{tot} mg L^{-1}	31	130	12	1.7	1.7	0	4.8	0	0	0	4	0
SO_4^{2-} mg L^{-1}	79.5	379	32	2.9	3	1.4	13.7	0.9	0.5	0.5	9.6	2
S_2^- mg L^{-1}	5.1		0.38	0.62	2	0.02	0.36	0	0.02	0.13	0.02	0.02
Fe_{tot} mg L^{-1}	0.0042	0.0037	0.022	0.062	0.2	0.71	0.036	0.02	0	0.49	1.2	560
Fe(II) mg L^{-1}	0	0	0.02	0.08	0.21	0.53	0.06	0.02	0.02	0.04	1.2	0.46
TDS mg L^{-1}	4994	10 670	5656	12 710	12 880	15 900	13 430	18 580	25 500	23 260	37 410	53 210
K mg L^{-1}	8.2	9.3	7.6	8.3	18	27	12	17	19	20	24	27
Mg mg L^{-1}	35	77	17	55	68	19	32	41	18	52	33	136
Ca mg L^{-1}	460	1100	290	2100	1750	2700	2260	2930	4600	3700	7680	10 000
Cl mg L^{-1}	2920	6230	3400	7930	7950	9940	8220	11 500	15 700	14 600	22 800	33 500
Na mg L^{-1}	1320	2800	1850	2530	2990	3110	2790	3970	4980	4720	6570	9150
TNC mL^{-1}	4.2×10^5	1.0×10^5	2.4×10^5	2.5×10^5	2.1×10^5	1.5×10^4	na	2.9×10^4	5.9×10^4	8.7×10^4	5.5×10^4	2.3×10^4
16S qPCR mL^{-1}												
bacteria	7.0×10^5	9.5×10^3	2.0×10^4	3.6×10^5	4.9×10^4	1.3×10^4	7.2×10^4	1.5×10^5	1.4×10^5	1.9×10^4	3.2×10^4	1.5×10^4
archaea	5.8×10^3	2.0×10^4	9.9×10^3	6.3×10^4	6.2×10^3	1.5×10^2	4.4×10^4	5.2×10^2	7.5×10^2	3.0×10^3	2.6×10^1	2.8×10^2

BGD
12, 13819–13857, 2015

Table 2. The total number of sequence reads, observed and estimated (Chao1, ACE) number of OTUs, number of singleton and doubleton OTUs, and Shannon diversity index per sample of the bacterial 16S rRNA gene data set. The analysis results are presented for both the total number of sequence reads per sample as well as for data normalized according tot he sample with the lowest number of sequence reads, i.e. 140 000 random sequences per sample.

(a) Bacteria Sample	All sequences							Normalized to 140 000 sequences					
	Number of sequence reads	Observed OTUs	Chao1	ACE	Singles	Doubles	Shannon species	Observed	Chao1	ACE	Singles	Doubles	Shannon
OLKR13/296 m	786 346	79 527	87 188	91 360	18 025	21 203	13	37 045	74 288	84 530	22 445	6762	13
OLKR3/318 m	345 433	52 381	53 238	54 961	5789	19 557	14	39 309	57 793	64 021	19 287	10 061	13
OLKR6/328 m	188 812	29 411	35 018	37 269	9209	7561	13	26 442	34 964	37 626	10 420	6369	13
OLKR23/347 m	485 154	33 257	37 175	38 895	8000	8166	11	20 494	34 268	37 305	10 641	4109	11
OLKR49/415 m	184 052	38 275	49 758	53 525	14 799	9535	13	34 117	48 804	52 938	15 372	8043	13
OLKR9/423 m	175 295	36 412	44 452	47 571	12 357	9494	14	33 596	44 496	48 161	13 489	8345	14
OLKR5/435 m	141 886	40 445	70 520	78 340	22 166	8167	14	40 145	70 288	78 232	22 086	8090	14
OLKR9/510 m	241 312	41 545	51 348	54 535	14 251	10 357	13	33 208	49 115	53 631	15 592	7640	13
OLKR2/559 m	257 789	45 456	72 269	78 325	22 550	9481	13	32 600	62 318	69 573	19 071	6118	12
OLKR1/572 m	210 659	29 804	35 362	37 491	9197	7607	12	25 703	34 934	37 682	10 650	6142	12
OLKR44/750 m	303 058	31 410	31 589	32 188	2005	11 200	12	25 937	33 448	36 295	10 346	7124	12
OLKR29/798 m	221 524	37 989	45 126	48 042	11 991	10 071	13	31 911	44 957	48 533	14 078	7594	13

BGD
12, 13819–13857, 2015

Table 2. The total number of sequence reads, observed and estimated (Chao1, ACE) number of OTUs, number of singleton and doubleton OTUs, and Shannon diversity index per sample of the archaeal 16S rRNA gene data set. The analysis results are presented for both the total number of sequence reads per sample as well as for data normalized according tot he sample with the lowest number of sequence reads, i.e. 17 000 random sequences per sample.

(b)

Archaea		All sequences						Normalized to 17 000 sequences					
Sample	Number of sequence reads	Observed OTUs	Chao1	ACE	Singles	Doubles	Shannon OTUs	Observed	Chao1	ACE	Singles	Doubles	Shannon
OLKR13/296 m	507 373	27 111	29 516	30 699	5835	7076	10	3957	13 380	15 062	2867	435	10
OLKR3/318 m	271 699	25 491	32 299	34 231	9205	6221	11	4955	15 044	17 238	3546	622	10
OLKR6/328 m	446 380	21 597	22 930	23 781	3861	5588	10	3776	11 705	14 020	2748	475	9
OLKR23/347 m	395 339	20 800	22 403	23 214	4083	5199	10	3919	11 855	13 323	2755	477	9
OLKR49/415 m	210 545	22 600	23 372	24 004	2975	5733	12	7023	17 088	19 874	4738	1114	12
OLKR9/423 m	697 360	22 014	22 527	23 082	2381	5520	9	3180	9,617	10 586	2224	383	9
OLKR5/435 m	769 026	21 127	22 235	23 078	3515	5574	9	2596	10 114	10 078	1852	227	9
OLKR9/510 m	169 142	12 709	12 782	12 960	713	3488	11	4879	11 205	13 215	3148	782	11
OLKR2/559 m	100 101	15 359	24 950	27 026	7840	3203	11	5119	14 497	16 488	3548	670	11
OLKR1/572 m	1 213 360	28 884	33 207	34 832	7846	7118	9	2273	9233	9923	1631	190	9
OLKR44/750 m	17 716	6436	8748	9750	2890	1805	12	6325	8743	9804	2921	1763	12
OLKR29/798 m	98 770	15 641	16 720	17 483	3158	4617	12	6951	14 655	17 184	4483	1303	12

BGD
12, 13819–13857, 2015

Table 3. The bacterial taxa representing at least 1 % of the sequence reads in any of the samples.

Phylum	Group	Relative abundance	Occurrence at depths
Actinobacteria	Actinobacteria/Other	7.6 %	693 m
	Actinomycetales/Other	1.1 %	415 m
	Microbacteriaceae/Other	1.7–7.3 %	405–798 m
	Microbacterium	1.0–24.2 %	296–572 m
	Propionibacerium	2.4 %	340 m
Bacteroidetes	Flavobacteriales/Other	2.2 %	510 m
	Flavobacteriaceae/Other	1.1–15.8 %	328–510, 572, 798 m
Chlorobi	SM1H02	1.0 %	423 m
Firmicutes	Staphylococcus	1.2 %	340 m
	Clostridium	3.3 %	693 m
Nitrospirae	Thermodesulfovibrionaceae/Other	1.2 %	340 m
OP9	SB-45	1.2–8.9 %	296–798 m
Proteobacteria	Other	1.2–9.7 %	296–798 m
	Caulobacteraceae/Other	1.3–6.7 %	296, 347–405, 559 m
	Erythrobacteriaceae/Other	1.3–5.7 %	405–510 m
	Betaproteobacteria/Other	5.7–12.3 %	296–328, 347–572, 798 m
	Burkholderiales/Other	2.1 %	347 m
	Comamonadaceae/Other	2.0–14.5 %	296–798 m
	Hydrogenophaga	1.2–1.3 %	340, 423 m
	Thiobacterales/Other	1.2–2.8 %	328, 415, 510 m
	Thiobacteraceae/Other	2.7–3.0 %	415, 510 m
	Thiobacteraceae/Unclassified	2.4–2.8 %	415, 510 m
	Sulfuricurvum	1.4 %	296 m
	Gammaproteobacteria/Other	3.5–16.1 %	296–798 m
	Pseudoalteromnas	2.9–15.3 %	296–328, 347–572, 798 m
	Enterobacteriaceae/Other	5.0 %	693 m
	Escherichia	1.1 %	693 m
	Acinetobacter	1.9–16.3 %	340, 693 m
	Pseudomonadaceae/Other	1.2–3.0 %	405–415, 572–693 m
	Pseudomonas	1.2–12.8 %	347–415, 559–798 m
Spirochaetes	PL-11810	2.0 %	296 m

BGD
12, 13819–13857, 2015

Table 4. The archaeal taxa representing at least 1 % of the sequence reads in any of the samples.

Phylum	Group	Relative abundance	Occurrence at depths
Crenarchaeota	Other	1.2–5.5 %	328–347, 415, 510, 693–798 m
	Thaumarchaeota/Other	1.5–11.8 %	296–559, 693 m
Euryarchaeota	Other	12.9–63.3 %	296–798 m
	Methanobacteriales/Other	1.2–3.1 %	415, 693–798 m
	MSBL1/Other	1.4 %	798 m
	SAGMEG-1	1.4–6.7 %	415, 693–798 m
	Methanobacteriaceae/Other	2.4 %	510 m
	Methanobacteiaceae/Unclassified	7.6 %	510 m
	Methanobacterium	3.6 %	340 m
	Methanomicrobia/Other	1.4–5.1 %	296–328, 347–559, 693–798 m
	Methanosarcinales/Other	2.3–13.3 %	296–693 m
	ANME-2D	4.9–56.1 %	296–798 m
	E2/Other	1.9–21.2 %	296–798 m
	Marine group II/Other	1.5–3.3 %	510–559 m
	Marine group II/Unclassified	1.6 %	347 m

BGD
12, 13819–13857, 2015

Table 5. Bacterial and archaeal taxa identified to genus level, if possible, showing positive and significant Pearson correlation ($r > 0.7$, $p < 0.01$) to depth, alkalinity, bicarbonate concentration, pH and conductivity. Taxa presented in bold text showed highest significance ($p < 0.005$), taxa presented in italic text showed high correlation ($r > 0.7$) and $p = 0.01$. O indicates "Other" and UC indicates "Unclassified" according to Greengenes taxonomy.

(a) Bacteria Depth	Alkalinity	Bicarbonate	pH	Conductivity
Sphingomonadaceae, O	Mycobacterium	C. Rhodoluna	Holophagales, O	Acidobacteria_iii1-15, O
Alcaligenaceae, O	Zhouia	CL0-1	Sva0725	Chitinophagaceae, UC
Variovorax	WCHB1-03	Cryomorphaceae, O	Parascardovia	Alcaligenaceae, O
	Caldithrixales, O	Fluviicola	**Microbacterium**	
	Roseiflexales	Mesonia	Emticicia	
	S085	Sejongia	Lentisphaeraceae	
	SL56	Chitinophagaceae, UC	GOUTA19	
	Sporosarcina	Arcicella	OM27	
	Fusobacteria, UC	NC10_wb1-A12	Acidithiobacillus	
	Gemm-2	Kordiimonadaceae	LD19	
	Lentisphaeria, O	Variovorax		
	FW_4-29	Polynucleobacter		
	Pedomicrobium	Methylobacillus		
	Rhodospirillaceae, UC	IndB3-24		
	Zymomonas	Haliangium		
	AEGEAN_112	Rickettsiella		
	Pseudoalteromonadaceae, O	Ulr1583		
	Halothiobacillus	EC214		
	Thioalkalimicrobium			
	SGSH944			
	ZA3648c, O			
	Opitutaceae			
	Chthoniobacterales, O			
	Sediment-1			
Archaea Depth	Alkalinity	Bicarbonate	pH	Conductivity
Archaea, Other	**E2, O**	Archaea, Othe	Halobacteriaceae, O	Archaea, Othe
MCG_pGrfC26	E2, UC	**Crenarchaeota, O**	Ferroplasma	MCG_pGrfC26
Methanobacteriales, O		*Crenarchaeota*, UC		**Methanobacteriales, O**
MSBL1, O		MBGA, O		**MSBL1, O**
SAGMEG-1		MBGB		**SAGMEG-1**
F99a103		MCG, UC		F99a103
Methanocellaceae, O		MCG_pGrfC26		Methanocellaceae, O
		MSBL1, O		
		SAGMEG-1		
		F99a103		
		Micrarchaeales		

BGD
12, 13819–13857, 2015

Table 5. Bacterial and archaeal taxa identified to genus level, if possible, showing positive and significant Pearson correlation ($r > 0.7$, $p < 0.01$) to concentrations of dissolved organic carbon (DIC), non-purgeable organic carbon (NPOC), total nitrogen (N) and ammonium. Taxa presented in bold text showed highest significance ($p < 0.005$), taxa presented in italic text showed high correlation ($r > 0.7$) and $p = 0.01$. O indicates "Other" and UC indicates "Unclassified" according to Greengenes taxonomy.

(b)

Bacteria

DIC	NPOC	N	Ammonium
C. Rhodoluna	**Actinobacteria, O**	**Actinobacteria, O**	MVS-40
CL0-1	Arsenicicoccus	Arsenicicoccus	Williamsia
Cryomorphaceae, O	Sanguibacter	Sanguibacter	Salinimicrobium
Fluviicola	Chitinophagaceae, O	Chitinophagaceae, O	Elusimicrobia, O
Mesonia	Pedobacter	Pedobacter	Clostridiales, O
Sejongia	VC38	VC38	Erysipelotrichaceae, O
Chitinophagaceae, UC	Clostridia, O	Clostridia, O	AEGEAN_112
Arcicella	Clostridiaceae, O	Clostridiaceae, O	Polaromonas
NC10_wb1-A12	**Clostridium**	**Clostridium**	Geobacter
Kordiimonadaceae	OD1, O	OD1, O	Sediment-1
Variovorax	OD1_ZB2	OD1_ZB2	
Polynucleobacter	Brevundimonas	Brevundimonas	
Methylobacillus	Phenylobacterium	Phenylobacterium	
IndB3-24	Rhodoplanes	Rhodoplanes	
Haliangium	Achromobacter	Sphingomonadaceae, O	
NB1-j	Hydrogenophilales, O	Alcaligenaceae, O	
Rickettsiella	Hydrogenophilaceae, O	Achromobacter	
Ulr1583	Thiobacillus	Hydrogenophilaceae, O	
EC214	Desulfobulbaceae, UC	Thiobacillus	
	Azorhizophilus	Methylobacillus	
	Spirochaetes, O	**Enterobacteriaceae, O**	
	Spirochaetes, O	**Escherichia**	
	Sphaerochaetales, O	Salmonella	
	Sphaerochaetaceae, O	Azorhizophilus	
	Sphaerochaetaceae, UC	Spirochaetes, O	
	Mollicutes, O	Spirochaetes, O	
	Acholeplasmataceae	Sphaerochaetales, O	
		Sphaerochaetaceae, O	
		Sphaerochaetaceae, UC	
		Mollicutes, O	
		Acholeplasmataceae	

Archaea

DIC	NPOC	N	Ammonium
Archaea, Othe	MCG_B10	MCG_B10	Crenarchaeota, UC
Crenarchaeota, O	DSEG_104A5	DSEG_104A5	DHVEG-1
Crenarchaeota, UC	pMC1A4	**Methanobacteriales, O**	
MBGA, O	Methanocellaceae, O	pMC1A4	
MBGB	Methanocellaceae, UC	Methanocellaceae, O	
MCG, UC	Methanomicrobiales, O	Methanocellaceae, UC	
MCG_pGrfC26	Methanoregulaceae, O	Methanomicrobiales, O	
MSBL1, O	Methanoregula	Methanoregulaceae, O	
SAGMEG-1	YLA114	Methanoregula	
F99a103		YLA114	
Micrarchaeales			

BGD
12, 13819–13857, 2015

Table 6. Bacterial and archaeal taxa identified to genus level, if possible, showing positive and significant Pearson correlation ($r > 0.7$, $p < 0.01$) to concentration of total dissolved solids (TDS), calcium, chloride, sodium and magnesium. Taxa presented in bold text showed highest significance ($p < 0.005$), taxa presented in italic text showed high correlation ($r > 0.7$) and $p = 0.01$. O indicates "Other" and UC indicates "Unclassified" according to Greengenes taxonomy.

Bacteria				
TDS	Calcium	Chloride	Sodium	Magnesium
C. Rhodoluna	CL0-1	C. Rhodoluna	CL0-1	C. Rhodoluna
CL0-1	Mesonia	CL0-1	Mesonia	CL0-1
Mesonia	Sejongia	Mesonia	Sejongia	Mesonia
Sejongia	Chitinophagaceae, UC	Sejongia	Chitinophagaceae, UC	Sejongia
Chitinophagaceae, UC	Alcaligenaceae, O	Chitinophagaceae, UC	Alcaligenaceae, O	Chitinophagaceae, UC
Arcicella	Kordiimonadaceae	Arcicella	*Arcicella*	Arcicella
Kordiimonadaceae	**Comamonadaceae, O**	Kordiimonadaceae	Kordiimonadaceae	Kordiimonadaceae
Alcaligenaceae, O	IndB3-24	Alcaligenaceae, O	Variovorax	Nautella
Variovorax	NB1-j	***Comamonadaceae, O***	IndB3-24	Variovorax
Polynucleobacter	*Rickettsiella*	Variovorax	NB1-j	Polynucleobacter
IndB3-24	EC214	Polynucleobacter	Rickettsiella	IndB3-24
NB1-j		IndB3-24	EC214	NB1-j
Rickettsiella		NB1-j		Rickettsiella
EC214		Rickettsiella		EC214
		EC214		
Archaea				
TDS	Calcium	Chloride	Sodium	Magnesium
Archaea, Othe	Archaea, Othe	Archaea, Othe	Archaea, Othe	MCG_pGrfC26
MBGB	MBGB	MBGB	*MCG, UC*	F99a103
MCG, UC	MCG, UC	MCG, UC	MCG_pGrfC26	Micrarchaeales
MCG_pGrfC26	MCG_pGrfC26	MCG_pGrfC26	**Methanobacteriales, O**	
Methanobacteriales, O	**Methanobacteriales, O**	**Methanobacteriales, O**	**MSBL1, O**	
MSBL1, O	**MSBL1, O**	**MSBL1, O**	**SAGMEG-1**	
SAGMEG-1	**SAGMEG-1**	**SAGMEG-1**	F99a103	
F99a103	F99a103	F99a103	Methanocellaceae, O	
Methanocellaceae, O	Methanocellaceae, O	Methanocellaceae, O	Micrarchaeales	
Micrarchaeales	Micrarchaeales	Micrarchaeales		

BGD
12, 13819–13857, 2015

Table 7. Bacterial taxa identified to genus level, if possible, showing positive and significant Pearson correlation ($r > 0.7$, $p < 0.01$) to concentrations of iron (Fe), ferrous iron (Fe(II)), sulphate, sulphide and total sulfur. Taxa presented in bold text showed highest significance ($p < 0.005$), taxa presented in italic text showed high correlation ($r > 0.7$) and $p = 0.01$. O indicates "Other" and UC indicates "Unclassified" according to Greengenes taxonomy.

Bacteria Fe	Fe(II)	Sulfate	Sulfide	S
C. Rhodoluna	**Actinomycetales, O**	Acidobacteria_iii1-15, O	MVS-40	Acidobacteria_iii1-15, O
Cryobacterium	Demequina	Knoellia	Mycobacterium	Knoellia
CL0-1	Arthrobacter	bOHTK-109	Flavobacteriales, UC	bOHTK-109
Cryomorphaceae, O	Williamsia	Flavobacteria, O	Zhouia	Flavobacteria, O
Fluviicola	Salinimicrobium	Tenacibaculum	WCHB1-03	Tenacibaculum
Flavobacteriaceae, O	Winogradskyella	Flammeovirgaceae, A4	Caldithrixales, O	Flammeovirgaceae, A4
Mesonia	Flammeovirgaceae, O	Elusimicrobia, O	Ignavibacteria, O	Elusimicrobia, O
Sejongia	Chlorobi, O	Carnobacteriaceae, O	WCHB1-05	Carnobacteriaceae, O
Chitinophagaceae, UC	Clostridiales, O	Nitrospirales, O	Roseiflexales	Nitrospirales, O
Arcicella	Acetobacterium	Leptospirillaceae, UC	S085	Leptospirillaceae, UC
NC10_wb1-A12	Peptococcaceae, O	OD1_ABY1	SL56	OD1_ABY1
Brocadiales, O	Erysipelotrichaceae, O	OM190, O	Synechococcophycideae, O	OM190, O
Kordiimonadaceae	Psychrilyobacter	Rhodospirillales, O	Prochlorococcus	Rhodospirillales, O
Comamonas	Thermodesulfovibrio	DTB120	Elusimicrobiales, O	Desulfococcus
Variovorax	*JS1*	Desulfococcus	Planococcaceae, O	Campylobacterales, O
Polynucleobacter	Methylobacteriaceae, O	Campylobacterales, O	Fusobacteria, UC	Moritella
Methylobacillus	Xanthobacter	Moritella	Gemm-2	Legionella
IndB3-24	AEGEAN_112	Legionella	Lentisphaeria, O	Vibrio
Haliangium	Polaromonas	Vibrio	FW_4-29	
NB1-j	*Schlegelella*		LCP-6	
Rickettsiella	Desulfocapsa		OP8, UC	
Ulr1583	Desulfovibrionaceae, O		OM190, UC	
EC214	Desulfovibrio		OM190_agg27	
	Desulfuromonas		Caulobacterales, O	
	Geobacter		**Caulobacteraceae, O**	
	Helicobacteraceae, UC		Bosea	
	Gammaproteobacteria, O		Pedomicrobium	
	Idiomarinaceae, UC		Roseococcus	
	Oleibacter		Rhodospirillaceae, UC	
	Spirochaetes, UC		Rickettsiales, O	
	Sphaerochaetaceae, UC		*Pelagibacteraceae*	
	Spirochaeta		Zymomonas	
	Persicirhabdus		Rhodocyclaceae, O	
	LD1-PA13		Desulfobulbaceae, O	
	SSS58A		Desulfobacterium	
	Sediment-1		**Sulfuricurvum**	
			Pseudoalteromonadaceae, O	
			Halothiobacillus	
			Thioalkalimicrobium	
			AB16, O	
			A714017, O	
			A714017, UC	
			SGSH944	
			SargSea-WGS	
			ZA3312c	
			ZA3648c, O	
			VHS-B5-50	
			PL-11B10	

BGD
12, 13819–13857, 2015

Table 8. Archaeal taxa identified to genus level, if possible, showing positive and significant Pearson correlation ($r > 0.7$, $p < 0.01$) to concentrations of iron (Fe), ferrous iron (Fe(II)), sulphate, sulphide and total sulfur. Taxa presented in bold text showed highest significance ($p < 0.005$), taxa presented in italic text showed high correlation ($r > 0.7$) and $p = 0.01$. O indicates "Other" and UC indicates "Unclassified" according to Greengenes taxonomy.

Archaea				
Fe	Fe(II)	Sulfate	Sulfide	S
Archaea, Other	Crenarchaeota, UC	Euryarchaeota, UC		Euryarchaeota, UC
Crenarchaeota, O	AK31	ANME-1		ANME-1
MBGA, O	Cenarchaeales	D-C06		D-C06
MBGB	Euryarchaeota, UC	ANME-2a-2b		ANME-2a-2b
MCG, UC	Halalkalicoccus	ANME-2c		ANME-2c
MCG_pGrfC26	MSBL1, UC			
Halococcus	RCII			
MSBL1, O	**E2, O**			
SAGMEG-1	CCA47			
F99a103	DHVEG-1			
Micrarchaeales	WCHD3-02			
	Picrophiliaceae, O			
	Thermogymnomonas			

BGD
12, 13819–13857, 2015

Table 9. The relative abundance of predicted genes of the most abundant pathways identified in the bacterial **(a)** and the archaeal **(b)** populations in the PICRUSt analysis. The pathways are according to KEGG. The bar in each cell visualizes the relative abundance and the number in each cell indicates the relative abundance of predicted genes belonging to each pathway in each sample. The NSTI number indicates the distance to the nearest sequenced taxon to which the obtained bacterial 16S rRNA gene sequences have been compared.

A.

Borehole	OL-KR13	OL-KR6	OL-KR3	OL-KR23	OL-KR5	OL-KR49	OL-KR9	OL-KR9	OL-KR2	OL-KR1	OL-KR44	OL-KR29	
Depth	296 m	328 m	340 m	347 m	405 m	415 m	423 m	510 m	559 m	572 m	693 m	798 m	
NSTI	0.17	0.15	0.12	0.08	0.11	0.12	0.12	0.11	0.16	0.14	0.05	0.15	Kegg pathway
Flagellar assembly	1.08	1.04	0.80	0.90	1.15	1.02	0.95	1.03	0.99	0.95	0.90	0.96	Cell Motility
ABC transporters	2.87	2.93	3.32	3.55	3.18	3.00	2.82	3.07	3.01	3.38	3.93	2.83	Membrane Transport
Transporters	8.26	8.17	5.04	5.35	6.41	8.46	8.18	6.57	4.66	4.94	5.98	8.31	
Bacterial secretion system	0.92	0.81	0.84	1.18	0.76	0.88	0.87	0.87	0.96	0.85	0.96	0.85	
Secretion system	2.21	2.01	2.09	2.63	2.13	2.18	2.07	2.19	2.34	2.07	2.31	2.07	
Two-component system	2.61	2.51	2.43	2.81	3.09	2.69	2.33	2.67	2.55	2.41	2.79	2.34	Signal Transduction
Alanine, aspartate and glutamate metabolism	0.86	0.83	0.85	0.74	0.72	0.85	0.87	0.85	0.88	0.87	0.77	0.92	Amino Acid Metabolism
Amino acid related enzymes	1.26	1.25	1.21	1.02	1.21	1.22	1.22	1.23	1.23	1.25	1.08	1.27	
Arginine and proline metabolism	1.06	1.00	1.07	1.07	0.90	1.06	1.11	1.06	1.02	1.11	1.17	1.14	
Cysteine and methionine metabolism	0.76	0.71	0.71	0.66	0.70	0.70	0.71	0.70	0.78	0.74	0.67	0.71	
Glycine, serine and threonine metabolism	0.73	0.75	0.80	0.79	0.77	0.79	0.83	0.79	0.89	0.85	0.83	0.88	
Phenylalanine, tyrosine and tryptophan biosynthesis	0.74	0.80	0.73	0.59	0.68	0.70	0.71	0.71	0.71	0.75	0.63	0.74	
Valine, leucine and isoleucine biosynthesis	0.68	0.68	0.68	0.60	0.66	0.69	0.71	0.69	0.71	0.72	0.63	0.75	
Valine, leucine and isoleucine degradation	0.85	0.99	0.97	1.06	0.94	0.96	1.22	0.96	1.00	0.98	1.02	1.06	
Butanoate metabolism	1.15	1.40	1.22	1.34	1.15	1.17	1.37	1.17	1.18	1.27	1.16	1.20	Carbohydrate Metabolism
Citrate cycle (TCA cycle)	0.87	0.87	0.86	0.82	0.74	0.81	0.81	0.81	0.81	0.85	0.68	0.87	
Glycolysis / Gluconeogenesis	1.01	0.98	0.97	1.05	0.82	0.90	0.93	0.89	0.92	0.96	0.80	0.98	
Glyoxylate and dicarboxylate metabolism	0.81	0.82	0.82	1.16	0.90	0.96	0.89	0.95	0.94	0.82	0.99	0.84	
Propanoate metabolism	0.99	1.15	1.08	1.10	1.00	1.00	1.21	1.00	0.99	1.10	1.05	1.09	
Pyruvate metabolism	1.12	1.18	1.17	1.16	1.07	1.10	1.12	1.11	1.11	1.12	1.03	1.14	
Carbon fixation pathways in prokaryotes	1.34	1.55	1.26	1.09	1.19	1.15	1.33	1.16	1.18	1.38	1.01	1.23	Energy Metabolism
Methane metabolism	1.24	1.42	1.13	1.19	1.03	1.15	1.15	1.14	0.95	1.24	0.93	1.08	
Nitrogen metabolism	0.86	0.85	0.83	1.06	1.10	0.88	0.79	0.88	0.83	0.80	0.83	0.79	
Oxidative phosphorylation	1.47	1.35	1.39	1.24	1.29	1.43	1.39	1.43	1.50	1.39	1.17	1.57	
Sulfur metabolism	0.31	0.29	0.30	0.26	0.32	0.32	0.32	0.32	0.32	0.29	0.30	0.33	
Unclassified energy metabolism	1.12	1.04	0.98	1.21	1.12	1.06	1.00	1.06	0.92	0.99	0.97	0.96	
Peptidases	1.34	1.28	1.31	1.27	1.23	1.31	1.36	1.31	1.39	1.32	1.28	1.40	Enzyme Families
Fatty acid metabolism	0.76	0.98	0.87	0.85	0.78	0.86	1.09	0.87	0.86	0.89	0.84	0.91	Lipid Metabolism
Porphyrin and chlorophyll metabolism	1.00	1.06	0.88	1.01	1.10	1.00	0.98	0.98	0.90	0.96	0.91	0.94	Metabolism of Cofactors and Vitamins
Purine metabolism	1.85	1.82	1.85	1.66	1.82	1.90	1.83	1.90	1.86	1.87	1.72	1.92	Nucleotide Metabolism
Pyrimidine metabolism	1.29	1.27	1.26	1.08	1.22	1.28	1.24	1.29	1.27	1.31	1.09	1.32	
Other ion-coupled transporters	0.98	0.87	1.10	0.87	0.94	0.88	0.89	0.87	0.89	0.92	1.27	0.86	Cellular Processes and Signaling
Protein folding and associated processing	0.88	0.96	0.86	0.85	0.94	0.97	0.92	0.95	0.80	0.87	0.82	0.86	Genetic Information Processing
Function unknown	1.68	1.67	1.73	1.78	1.63	1.75	1.66	1.75	1.54	1.65	1.89	1.65	Poorly Characterized
General function prediction only	3.40	3.56	3.48	3.33	3.24	3.42	3.42	3.42	3.10	3.46	3.55	3.27	

B.

Borehole	OL-KR13	OL-KR6	OL-KR3	OL-KR23	OL-KR5	OL-KR49	OL-KR9	OL-KR9	OL-KR2	OL-KR1	OL-KR44	OL-KR29	
Depth	296 m	328 m	340 m	347 m	405 m	415 m	423 m	510 m	559 m	572 m	693 m	798 m	
NSTI	0.19	0.17	0.20	0.25	0.18	0.22	0.14	0.21	0.21	0.21	0.29	0.28	Kegg pathway
Flagellar assembly	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	Cell Motility
ABC transporters	3.56	3.31	3.15	2.81	3.48	2.65	3.12	2.53	3.33	3.36	2.83	3.24	Membrane Transport
Transporters	4.53	4.20	4.19	3.68	4.42	3.66	3.92	3.49	4.32	4.35	4.04	4.29	
Bacterial secretion system	0.27	0.29	0.27	0.43	0.28	0.37	0.33	0.29	0.28	0.27	0.28	0.37	
Secretion system	1.09	1.03	1.03	0.95	1.06	0.82	0.93	0.88	1.05	1.06	1.13	0.91	
Two-component system	1.13	1.19	1.03	1.06	1.15	1.15	1.23	1.00	1.11	1.08	0.97	1.00	Signal Transduction
Alanine, aspartate and glutamate metabolism	1.06	1.10	1.06	1.29	1.08	1.23	1.15	1.09	1.10	1.06	1.14	1.23	Amino Acid Metabolism
Amino acid related enzymes	1.66	1.77	1.69	2.10	1.67	2.09	1.84	1.82	1.73	1.67	1.77	1.99	
Arginine and proline metabolism	1.19	1.19	1.17	1.25	1.20	1.30	1.23	1.12	1.19	1.18	1.16	1.31	
Cysteine and methionine metabolism	0.86	0.90	0.87	1.06	0.89	1.16	1.04	0.92	0.89	0.86	0.92	0.97	
Glycine, serine and threonine metabolism	0.77	0.83	0.81	0.99	0.80	1.23	0.91	0.93	0.83	0.79	0.88	1.06	
Phenylalanine, tyrosine and tryptophan biosynthesis	1.09	1.07	1.01	1.28	1.10	0.99	1.13	0.96	1.06	1.05	0.87	1.19	
Valine, leucine and isoleucine biosynthesis	1.15	1.20	1.14	1.18	1.17	1.35	1.28	1.20	1.16	1.15	1.06	1.21	
Valine, leucine and isoleucine degradation	0.21	0.31	0.27	0.40	0.25	0.73	0.37	0.42	0.28	0.23	0.30	0.47	
Butanoate metabolism	0.85	0.92	0.80	0.94	0.88	1.29	0.98	0.83	0.89	0.82	0.79	1.10	Carbohydrate Metabolism
Citrate cycle (TCA cycle)	1.00	1.06	1.11	1.25	1.02	1.49	1.10	1.32	1.08	1.06	1.32	1.38	
Glycolysis / Gluconeogenesis	1.56	1.54	1.50	1.38	1.58	1.51	1.44	1.42	1.54	1.53	1.48	1.57	
Glyoxylate and dicarboxylate metabolism	0.45	0.49	0.48	0.50	0.47	0.76	0.56	0.57	0.49	0.46	0.53	0.54	
Propanoate metabolism	0.68	0.76	0.72	0.85	0.70	1.14	0.76	0.85	0.75	0.70	0.91	1.04	
Pyruvate metabolism	1.27	1.30	1.22	1.23	1.27	1.48	1.24	1.24	1.29	1.25	1.29	1.44	
Carbon fixation pathways in prokaryotes	1.29	1.38	1.35	1.50	1.32	1.67	1.45	1.50	1.35	1.33	1.55	1.52	Energy Metabolism
Methane metabolism	5.54	4.90	5.07	3.48	5.39	2.24	4.14	5.39	5.07	5.56	4.76	3.82	
Nitrogen metabolism	0.53	0.53	0.60	0.51	0.54	0.56	0.57	0.67	0.54	0.57	0.65	0.49	
Oxidative phosphorylation	1.55	1.65	1.47	1.74	1.59	2.01	1.80	1.45	1.60	1.50	1.44	1.73	
Sulfur metabolism	0.25	0.26	0.23	0.19	0.26	0.23	0.27	0.16	0.25	0.23	0.18	0.19	
Unclassified energy metabolism	0.87	0.79	0.88	0.63	0.85	0.63	0.77	0.81	0.83	0.88	0.89	0.66	
Peptidases	1.15	1.19	1.28	1.34	1.16	1.46	1.21	1.44	1.22	1.22	1.46	1.35	Enzyme Families
Fatty acid metabolism	0.14	0.19	0.16	0.23	0.15	0.44	0.22	0.21	0.18	0.15	0.23	0.27	Lipid Metabolism
Porphyrin and chlorophyll metabolism	1.57	1.53	1.46	1.31	1.59	1.11	1.70	1.39	1.47	1.53	1.07	1.08	Metabolism of Cofactors and Vitamins
Purine metabolism	2.54	2.63	2.71	3.13	2.56	2.93	2.66	2.96	2.65	2.63	3.01	2.92	Nucleotide Metabolism
Pyrimidine metabolism	2.29	2.32	2.42	2.61	2.29	2.42	2.30	2.56	2.34	2.35	2.52	2.46	
Other ion-coupled transporters	0.88	0.84	0.87	0.85	0.86	0.57	0.76	0.81	0.85	0.88	0.94	0.77	Cellular Processes and Signaling
Protein folding and associated processing	1.13	1.03	1.13	0.99	1.09	0.78	0.96	1.13	1.08	1.14	1.15	0.95	Genetic Information Processing
Function unknown	2.80	2.58	2.94	2.48	2.74	2.08	2.53	3.07	2.73	2.90	2.92	2.26	Poorly Characterized
General function prediction only	5.79	5.53	5.92	5.69	5.91	4.84	5.43	6.07	5.79	5.88	5.83	5.52	

BGD
12, 13819–13857, 2015

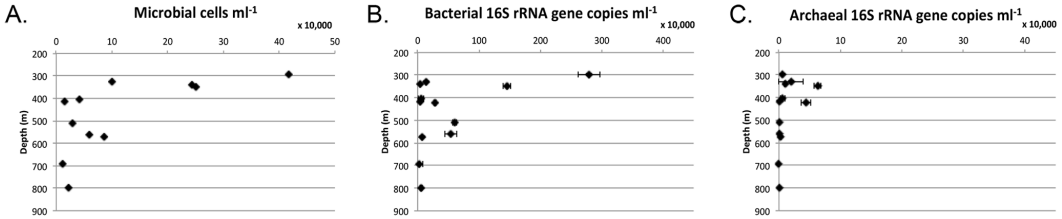

Figure 1. The concentration of **(a)** microbial cells mL^{-1} determined by epifluorescence microscopy and the estimated concentration of **(b)** bacterial and **(c)** archaeal 16S rRNA gene copies mL^{-1} groundwater determined by qPCR in water conductive fractures situated at different depths in the Olkiluoto bedrock.

BGD
12, 13819–13857, 2015

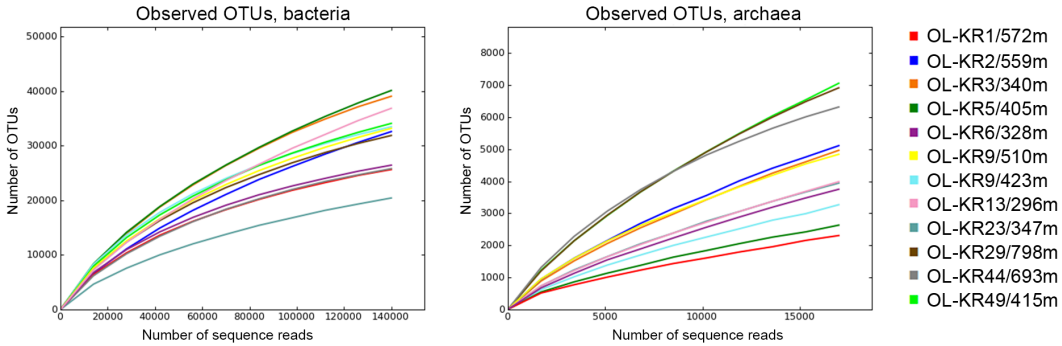

Figure 2. The rarefaction curves of observed bacterial (left pane) and archaeal (right pane) OTUs in each sample generated on sequence data normalized to 140 000 reads for bacteria and 17 000 reads for archaea.

BGD
12, 13819–13857, 2015

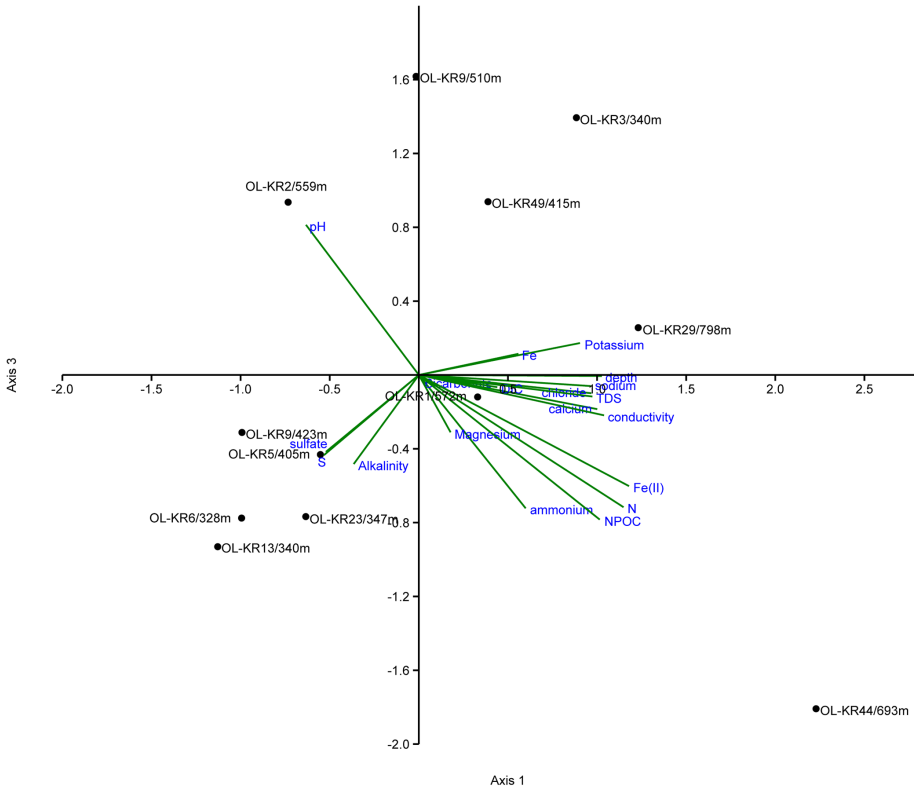

Figure 3. A canonical correspondence plot showing the distribution of the different samples according to the relative abundance of bacterial and archaeal taxa in relation to geochemical parameters. The triplot (green) indicates directionality of the environmental variables.

BGD
12, 13819–13857, 2015

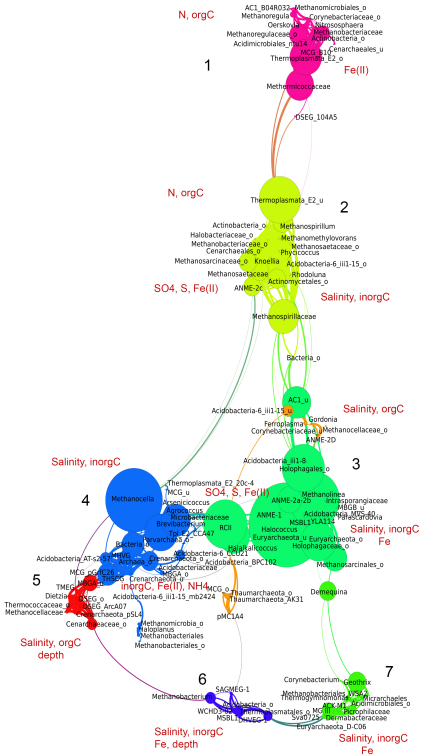

Figure 4. Network of co-occurring rare microbial taxa based on significant correlation ($R > 0.7$, $p < 0.001$) between different taxa. Each circle (node) represents a taxon and the size of the node is proportional to the number of connections of the node. The different clusters of microbial taxa, i.e. communities' (1–7) are shown in different colour. Geochemical parameters to which the community members showed strong correlation are indicated in the figure.

BGD
12, 13819–13857, 2015