Microbial communities and their predicted metabolic characteristics in deep fracture groundwaters of the 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, FIN-02044 VTT, Finland

[2] Posiva Oy, Olkiluoto, 27160 Eurajoki, Finland

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

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 m and 798 m by high throughput amplicon sequencing using the Illumina HiSeq platform. Between 1.7×10^4 – 1.2×10^6 bacterial or archaeal sequence reads per sample was obtained. These sequences revealed that up to 95% and 99% of the bacterial and archaeal sequences obtained, respectively, belonged to 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. The metabolic properties of the microbial communities were predicted using PICRUSt. The approximate estimation 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 environmental conditions, but could also be induced again if changes in their living conditions occur.

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). 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 less abundant groups, which are distributed unevenly in the different water conductive fractures.

The rare biosphere is a concept describing the hidden biodiversity of an environment (Sogin et al., 2006). The rare biosphere consists of microbial groups that are ubiquitously distributed in nature but often present at low relative abundance and may thus stay below the limit of detection. 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 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. In deep subsurface groundwater the biomass concentration is often low and the sampling efforts may not yield enough biomass for exstensive metagenomic analysis of the microbial communities. Tools for predicting metabolic pathways may help to establish a consensus of the microbial metabolic characteristics present in an environment and the possible interactions of the microbial communities with the abiotic environment. Tools, such as PICRUSt (Langille *et al.,* 2013), allows us to estimate microbial metabolic functions based on NGS microbiome data. For example, 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 and also to study the rare microbiome. In addition, 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 1,000 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 & Paulamäki, 2006). The *in situ* temperature at 300 m depth in the Olkiluoto bedrock is stable at approximately 10°C and increases linearily to approximately 16°C at 800 m depth (Sedighi et al., 2013).

This study focused on 12 groundwater samples from water conductive fractures situated at between 296 m and 798 m below sea level bsl and originating from 11 different drillholes in Olkiluoto (Figure 1). 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 occurred between December 2009 and January 2013 (Table 1) as described previously (Bomberg et al., 2015; Miettinen et al., 2015; Sohlberg et al., 2015). The samples were obtained from 11 different permanently packered or open drillholes equipped with removable inflatable packers. The position and direction of the drillholes are indicated in Figure 1. 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 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 10µL 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 minutes followed by 45 amplification cycles of 15 seconds at 95 °C, 30 seconds at 55 °C and 30 seconds 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 seconds 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 dilution 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 for completely overlapping paired-end sequencing of the V6 region of both the archaeal and bacterial 16S rRNA genes were produced as previously described (Eren et al., 2013). 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 (https://vamps.mbl.edu/resources/primers.php). The sequencing was performed using a HiSeq 1000 system (Illumina).

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 core acrchaea and bacteria communities were identified from the OTU tables with the compute_core_microbiome.py function in QIIME using default values, with exception of the minimum number of samples where an OTU must be detected, which was set to 80%. 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 predicted KO numbers were plotted on KEGG pathway maps (http://www.genome.jp/kegg/) separately for the bacterial and archaeal predicted metagenomes, with a threshold of a minimum of 100 genes in total estimated from all samples. 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 and data visualization

The similarity of the archaeal and bacterial communities between the different samples was tested by principal coordinates analysis (PCoA) using the Phyloseq package in R (McMurdie and Holmes, 2014; R Core Team, 2013). The analysis was performed using the raw OTU tables outputted by QIIME. In addition, a PCoA analysis showing the effect of library size on the ordination of the samples was calculated using vegan (Oksanen et al., 2016). The Bray-Curtis distance model was used for both analyses. The samples were hierarchically clustered in a UPGMA tree based on the raw OTU counts using the heatmap function of phyloseq in R.

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 (Figure 2, 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 (Figure 2, Table 1).

3.2 Sequence statistics, diversity estimates and sequencing coverage

The number of bacterial v6 sequence reads from the 12 samples varied between $1.4 - 7.8 \times 10^5$ 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 $0.17 – 12.1 \times 10^5$ reads with a mean sequencing depth of 4.1×10^5 ($\pm 3.5 \times 10^5$ standard deviation) reads/sample. The numbers of observed operational taxonomic units (OTUs) represented on average 82.6% (± 12.5%) of the Chao1- and 78.1 % (± 13.4%) of the ACE-estimated numbers of bacterial OTUs (Table 2ab). 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. 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 (Figure S1) 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. A total of 468,684 archaeal and 301,458 bacterial OTUs were obtained in this study.

3.3 Microbial communities

From the bacterial v6 sequences 49 different bacterial Phyla were detected (Appendix 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.

The archaea were represented by two identified phyla, the Euryarchaeota and the Crenarchaeota (Appendix 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.

The archaeal and bacterial core communities were determined as OTUs present in at least 80% of the samples. Of the more than 4.6×10^5 archaeal OTUs the core community consisted of 82 OTUs belonging to three archaeal orders, the E2 of the Thermoplasmatales, the Methanobacteriales and the Methanosarcinales (Figure 3). Additionally, a great proportion of the OTUs of the core community did not receive any other taxonomic identity than Archaea. The most common archaeal family of the core community was the ANME-2D belonging to the Methanosarcinales. The bacterial core community consisted of only 26 OTUs, compared to more than 3.0×10^5 bacterial OTUs in total (Figure 3). These OTUs belonged to six different families, the Alteromonadales, Burkholderiales, SB-45, Sphingomonadales, Syntrophobacterales and Thiobacterales. In addition, a great portion of the core community OTUs were classified only as unassigned bacteria. The most abundant of the bacterial core community OTUs belonged to Thiobacteriaceae and Comamonadaceae. In both the archaeal and bacterial sequence data a great proportion of the sequence reads were only identified as archaea or bacteria, without more detailed taxonomic assignments. The core OTUs were distributed with different abundance in the different samples (Figure 3). Most of the OTUs detected were present in less than 20% of the samples (Figure 4).

3.4 Environmental parameters and the microbial communities

The microbial community profiles of the different samples were clustered in a UPGMA tree based on the OTU tables and the Bray-Curtis distance model (Figure 5). The archaeal and bacterial communities clustered according to the OTUs detected in the samples, but not clearly according to any physicochemical parameter. In the PCoA plots, the samples clustered into three groups (Figures 6a-d). However, no clear environmental factor was identified to drive the communities. In these analyses the microbial communities, with 4.6×10^5 archaeal and 3.0×10^5 bacterial OTUs, were the strongest clustering forces and thus even quite different communities obtained similar coordinates on the PCoA plots. Coordinate 1 determined 14.9% and Coordinate 2 12.6% of the variance in the archaeal communities and 20% and 17.2%, respectively, of the variance in the bacterial communities (Figure 6e,f).

3.5 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 (Figure 7). 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 only indicative. The estimated microbial metabolism did not differ noticeably between the different depths (Figure 8a and b). 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, propionate, glycolysis/gluconeogenesis and pyruvate metabolism pathways for the bacteria and glycolysis/gluconeogenesis and pyruvate metabolism pathways for the archaea (Figure 9). Glucose is converted into pyruvate and further to Acetyl-CoA by both bacteria and archaea. The bacterial community may produce and utilize acetate. Both the bacterial and archaeal communities fix carbon via the Wood-Ljungdal (WL) reverse Citric acid cycle (rTCA) and Calvin pathways. Methane is produced from methylamines, CO_2 and methanol by the methanogenic archaea. In addition to the strong evidence of methanogenesis in the archaeal community the reverse methanogenesis, i.e. anaerobic methane oxidation by the ANME-2D archaea is possible. Based on the predicted metagenomes the bacterial community is not able to oxidize methane or hydrolyze methanol, but the methylotrophs present may use formic acid and trimethylamines.

The most abundant energy metabolic pathway in the bacterial communities was the oxidative phosphorylation (Figure S3) while for the archaea the methane metabolism was the most important (Figure 8, 9). Utilization of propanoate and butanoate (Figure 8) by the bacterial communities as well as well covered fatty acid biosynthesis and degradation pathways indicate that the bacterial community is capable of fermentation (Figure S4a and b). Nitrate is reduced both through dissimilatory nitrate reduction to ammonia and through denitrification to nitrous oxide by the bacteria (Figure S5). In addition, nitrogen is fixed to ammonia by both archaea and bacteria. The ammonia is then used as raw material for L-glutamate synthesis (Figure S5). Sulfur metabolism was not a major pathway in either the bacterial or the archaeal communities according to the predicted number of genes. However, assimilatory sulphate reduction was indicated in both the bacterial and archaeal communities, while dissimilatory sulphate reduction and sulphur oxidation was indicated only in the bacterial communities (Figure S6).

Several amino acid synthesis pathways were predicted (Figure 8), of which the most prominent were the alanine, aspartate and glutamate synthesis, arginine and proline synthesis, cysteine and methionine synthesis, glycine, serine and threonine synthesis, phenylalanine, tyrosine and tryptophan synthesis and the valine, leucine and isoleucine synthesis pathways.

Different types of membrane transport (ABC transporters) was identified where sulphate and iron (III) were taken up by the bacteria and tungstate, molybdate, proline, zink, cobalt and nickel was taken up by both archaea and bacteria (Figure S7). 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 (Figure 8a and b).

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 (Figure 2, Table 1). In accordance with other

Fennoscandian deep subsurface environments (Purkamo et al., 2016), most of the microbial communities at different depth in Olkiluoto bedrock fractures consist of bacteria. Archaea have in general been shown to constitute at most approximately 1% of the Fennoscandian deep bedrock groundwater (Purkamo et al., 2016). However, at specific depths in Olkiluoto (328 m, 423 m) the archaea contributed 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). The high abundance of archaea in Olkiluoto is special for this environment. Archaea have also been quantified from the Outokumpu deep scientific borehole (Purkamo et al., 2016), but unlike the situation in Olkiluoto the archaeal community was less than 1% of the total community at best.

Previously, using 454 amplicon sequencing, we have observed OTU numbers of approximately 800 OTUs per sample covering approximately 550 bacterial genera (or equivalent groups) and approximately 350 archaeal OTUs including approximately 80 different genera (or equivalent groups) (Miettinen et al., 2015). Miettinen et al. (2015) defined the OTUs 97% sequence homology and the number of sequence reads per sample was at most in the range of 10^4. In contrast, our sequence read numbers were 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. Nevertheless, the obtained numbers of OTUs per sample in this study were huge (Table 2). This may reflect the high level of variability in the short sequence reads of the v6 region used in this study. As discussed by Huse et al. (2008), short sequence reads very often match several different full-length 16S rRNA reads. As shown in our study taxonomic assignments, such as 'Proteobacteria_other' were common and may be due to multiple matches for the individual sequence reads obtained in the identification step of the analysis.

In general, the microbial communities at different depth grouped loosely into clusters (Figure 6). Although no clear environmental factor seemed to drive the microbial communities at different depths, the core communities appeared to be more similar in samples from similar depths, especially in the bacterial communities (Figure 3). OTUs belonging to both sulphate reducers (Desulfobacteraceae) and sulphur oxidizers (Thiobacteraceae) were present in the bacterial core community. The archaeal core community consisted mostly of methane-oxidizing ANME-2D archaea. Interestingly, however, their abundance was higher in the deeper samples. Previuos studies on the Finnish deep biosphere has shown that the microbial communities at different sites vary strongly from each other. Purkamo et al. (2015) investigated the bacterial and archaeal communities of different fracture zones of the Outokumpu deep scientific borehole and found that the majority of the bacterial populations at depths between 180 m and 500 m depth consist of Betaproteobacteria belonging to the *Commamonadaceae* and the archaeal communities consist of *Methanobacteriaceae* and *Methanoregula*.

The core communities, defined as OTUs present in all the studied samples, accounted for between 0.2% and 11.7% of the archaeal and 0.4% - 4.1% of the bacterial communities, respectively. This is surprisingly low and indicates an obviously high microbial diversity in these deep fracture water samples. Nevertheless, the short read length and high sequence variability within the v6 region may affect these results. Nevertheless, genus-level identification of the sequence data showed that between 95 – >99% of the archaeal sequence reads fell in to 25 genera, which were present in all samples. Likewise for the bacterial communities, 80 – 97% of the sequene reads belonged to 95 bacterial genera that were detected in all samples. The number of OTUs that were not present in more than 80% of the samples far outnumbered the number of core OTUs both in the archaeal and the bacterial communities. Only approximately 800 archaeal OTUs and less than 600 bacterial OTUs were present in at least 50% of the samples. Interestingly, 20% of the samples share only approximately 10,000 archael and 30,000 bacterial OTUs, indicating that the proportion of rare OTUs present in only 1 or 2 samples is huge.

Our results agree with Sogin et al. (2006) and Magnabosco et al. (2014), who showed that a relatively small number of taxa dominate deep-sea water and deep groundwater habitats, respectively, 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, Pedersen 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; Pedersen, 2013; Pedersen et al., 2014).

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 high indicating that closely related species to those found in our deep groundwater have yet to be sequenced. This is in accordance with Langille et al. (2013), who showed that environments containing a high degree of unexplored microbiota also tend to have high NSTI values. Staley et al. (2014) also showed in a comparison between PICRUSt and shot gun metagenomic sequencing of riverine microbial communities that PICRUSt may not be able to correctly assess rare biosphere functions. Nevertheless, Langille et al. (2013) showed that PICRUSt may predict the metagenomic content of a microbial community more reliably than shallow metagenomic sequencing. Although PICRUSt does not give as detailed results as metagenomics or genomic analyses may give, it is still a useful tool for predicting functions in microbial communities when the possibility for metagenomics analysis may be impossible, e.g. due to low biomass in the samples.

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. Wu et al. (2015) also found by metagenomic analyses that fermentation was a major metabolic activity in the microbial community of Swedish deep groundwater. 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. It was also noted that even though evidence for methane oxidation could not be inferred from the PICRUSt predictions (no *pmo*A genes), the bacterial community may oxidize formate, which is in agreement with the findings reported by Wu et al. (2015).

Several carbon fixation pathways were predicted in the metagenomes, the Calvin cycle, reductive TCA (rTCA) cycle and Wood-Ljungdahl (WL) pathway. The WL-pathway is considered 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) and was found in both the bacterial and the archaeal communities. 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 m and 559 m depth. Nevertheless, our results agree with Nyyssönen et al. (2014), who showed my metagenomic analysis that the microbial communities at different depth of the Outokumpu scientific deep drill hole may fix carbon in several ways, of which the rTCA, the WL pathway and the Calvin cycle were identified. Magnabosco et al. (2016) showed that the WL pathway was the dominating form of carbon fixation in metagenomes of 3 km deep Precambrian crust biospheres in South Africa. Dong et al. (2014) also suggested that microorganisms in low-energy deep subsurface environment may have several strategies for e.g. carbon fixation, as shown in the *Halomonas sulfidaeris*, in order to access as many resources as possible. The predicted methane metabolism (methane and methyl compound consumption) and oxidative phosphorylation were equally strong in the bacterial community. Sulphur metabolism was not predicted to be a common pathway for energy in either the archaeal or the bacterial communities. However, PICRUSt predicted bacteria with either assimilative or dissimilative sulphate reduction to be present. Sulphur oxidation through the *sox* system was in general not predicted, but the *sox*D gene was predicted and oxidation of thiosulphate to sulphate may be possible (Figure S6). Nitrate is reduced both through dissimilatory nitrate reduction to ammonia and through denitrification to nitrous oxide by the bacteria. In addition, nitrogen is fixed to ammonia by both archaea and bacteria. The ammonia is then used as raw material for L-glutamate synthesis.

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 m and 405 m. Methanogenesis was common also at all other depths except 330 m – 347m, 415 m and 693 m – 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 m and 405 m. Methane oxidation using methane monoxygenases and methanol dehydrogenases does not occur in either bacterial or archaeal communities. It should be noted, however, that the ANME-2D archaea are likely to use the methanogenesis pathway in the reverse for oxidizing methane anaerobically to carbon dioxide (Haroon et al., 2013). The produced carbon dioxide may be fixed by the same archaea and turned in to acetate, which may serve as carbon substrate and electron donor and acceptor for a large variety of microorganisms in the groundwater.

Carbohydrate metabolism. Glycolysis/gluconeogenesis is one of the most common carbohydrate-metabolizing pathways predicted for both the archaeal and bacterial communities (Figure 9). Pyruvate from 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.

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.

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 m 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. Nevertheless, the proportion of OTUs found in only a small propostion (e.g. 20%) of the samples far surpassed the number of OTUs included in the core communities. 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 communities were estimated to contain several different common carbon fixation pathways, such as the Calvin cycle and the reductive TCA and the Wood-Ljungdahl pathway.

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. Proc. Natl. Acad. Sci. USA 91, 1609-1613, 1994.

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.

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, 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., Article ID 979530, 2015. http://dx.doi.org/10.1155/2015/979530.

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. Nature 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. Antonie van Leeuwenhoek, 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. Microbiol., 72, 5069-72, 2006.

Dong, Y., Kumar, C.G., Chia, N., Kim, P.-J., Miller, P.A., Price, N.D., Cann, I.K.O., Flynn, T.M., Sanford, R.A., Krapac, I.G., Locke, R.A.II, Hong, P.-Y., Tamaki, H., Liu, W.-T., Mackie, R.I., Hernandez, A.G., Wright, C.L., Mikel, M.A., Walker, J.L., Sivaguru, M., Fried, G., Yannarell, A.C., Fouke, B.W. Halomonas sul- fidaeris-dominated microbial community inhabits a 1.8 km-deep subsurface Cambrian Sandstone reservoir. Environ. Microbiol. 16, 1695-1708, 2014.

Eren, A.M., Maignien, L., Sul, W.J., Murphy, L.G., Grim, S.L., Morrison, H.G. and Sogin, M.L. Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. Meth. Ecol. Evol., 4, 1111-1119, 2013.

Fuchs, G.: Alternative pathways of autotrophic CO_2 fixation. In: Autotrophic Bacteria, ed.

HG Schlegel, B Bowien, pp. 365–82. Berlin: Springer, 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., Nekrutenko, A.: Galaxy: a platform for interactive large-scale genome analysis. Genome Res., 15, 1451-5, 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, 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.

Hammer, Ø., Harper, D.A.T. and Ryan, P.D.: PAST: Paleontological statistics software package for education and data analysis. Palaeontologia Electronica 4: 9pp. http://palaeo-electronica.org/2001_1/past/issue1_01.htm, 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, 2011.

Huse, S.M., Dethlefsen, L., Huber, J.A., Welch, D.M., Relman, D.A. and Sogin, M.L. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. PLoS Genet, 4, p.e1000255, 2008.

Kärki, A. and Paulamäki, S.: Petrology of Olkiluoto. Posiva 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 (In Finnish) M.Sc. Thesis, Faculty of Agriculture and Forestry, University of Helsinki, Finland, 2015.

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.

Magnabosco, C., Tekere, M., Lau, M. C., Linage, B., Kuloyo, O., Erasmus, M., Cason, E., van Heerden, E., Borgonie, G. and Onstott, T. L. Comparisons of the composition and biogeographic distribution of the bacterial communities occupying South African thermal springs with those inhabiting deep subsurface fracture water. Front. Microbiol, 5, 10-3389, 2014.

Magnabosco, C., Ryan, K., Lau, M.C., Kuloyo, O., Lollar, B.S., Kieft, T.L., van Heerden, E. and Onstott, T.C. A metagenomic window into carbon metabolism at 3 km depth in Precambrian continental crust. ISME J., 10, 730-741, 2016.

Martin, W., Baross, J., Kelley, D. and Russell, M.J.: Hydrothermal vents and the origin of life. Nat. Rev. Microbiol. 6, 805–14, 2008.

McMurdie, P.J.; Holmes, S.: Shiny-phyloseq: Web Application for Interactive Microbiome Analysis with Provenance Tracking. Bioinformatics 31, 282–283, 2015.

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, 1-160, 2015. www.posiva.fi

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. Microbiol. 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.

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.

Oksanen, J.; Blanchet, F.G.; Friendly, M.; Kindt, R.; Legendre, P.; McGlinn, D.; Minchin, P.R.; O'Hara, R.B.; Simpson, G.L.; Solymos, P.; Stevens, M.H.H.; Szoecs, E.; Wagner, H.: vegan: Community Ecol- ogy Package. R package version 2.4-0. 2016. http://CRAN. R-project.org/package=vegan.

Osburn, M.R., LaRowe, D.E., Momper, L.M. and Amend, J.P. Chemolithotrophy in the continental deep subsurface: Sanford Underground Research Facility (SURF), USA. Front. Microbiol., *5*, 610, 2014.

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.

Pedersen, K.: Metabolic activity of subterranean microbial communities in deep granitic groundwater supplemented with methane and H2. ISME J., 7, 839-849, 2013.

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.

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

Purkamo, L., Bomberg, M., Nyyssönen, M., Kukkonen, I., Ahonen, L., Kietäväinen, R., 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.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org/.

Rajala, P., Bomberg, M., Kietäväinen, R., Kukkonen, I., Ahonen, L., Nyyssönen, M., 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., Weber, C.F.: Introducing MOTHUR: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol., 75, 7537–41, 2009.

Sedighi, M., Bennett, D., Masum, S.A., Thomas, H.R., Johansson, E.: Analysis of temperature data at the Olkiluoto. Posiva Working Report 2013-58, 1-78, 2013

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

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”. Proc. 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, 2015.

Staley, C., Gould, T.J., Wang, P., Phillips, J., Cotner, J.B., Sadowsky, M.J. Core functional traits of bacterial communities in the Upper Mississippi River show limited variation in response to land cover. *Front. Microbiol*. 5,.414, 2014.

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.

Wu, X., Holmfeldt, K., Hubalek, V., Lundin, D., Åström, M., Bertilsson, S. and Dopson, M. Microbial metagenomes from three aquifers in the Fennoscandian shield terrestrial deep biosphere reveal metabolic partitioning among populations. ISME J., on line.

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) and Miettinen et al. (2015)

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	3/11/2010	18/5/2010	29/8/2011	15/12/2009	16/10/2012	14/12/2009	31/10/2011	29/8/2011	27/1/2010	26/1/2010	15/1/2013	18/5/2010
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	10670	5656	12710	12880	15900	13430	18580	25500	23260	37410	53210
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	10000
Cl mg L^{-1}	2920	6230	3400	7930	7950	9940	8220	11500	15700	14600	22800	33500
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

Table 2a. 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.

Bacteria		All sequences						Normalized to 140,000 sequences					
Sample	Number of sequence reads	Observed OTUs	Chao1	ACE	Singletons	Doubletons	Shannon	Observed OTUs	Chao1	ACE	Singletons	Doubletons	Shannon
OLKR13/296m	786,346	79,527	87,188	91,360	18,025	21,203	13	37,045	74,288	84,530	22,445	6,762	13
OLKR3/318m	345,433	52,381	53,238	54,961	5,789	19,557	14	39,309	57,793	64,021	19,287	10,061	13
OLKR6/328m	188,812	29,411	35,018	37,269	9,209	7,561	13	26,442	34,964	37,626	10,420	6,369	13
OLKR23/347m	485,154	33,257	37,175	38,895	8,000	8,166	11	20,494	34,268	37,305	10,641	4,109	11
OLKR49/415m	184,052	38,275	49,758	53,525	14,799	9,535	13	34,117	48,804	52,938	15,372	8,043	13
OLKR9/423m	175,295	36,412	44,452	47,571	12,357	9,494	14	33,596	44,496	48,161	13,489	8,345	14
OLKR5/435m	141,886	40,445	70,520	78,340	22,166	8,167	14	40,145	70,288	78,232	22,086	8,090	14
OLKR9/510m	241,312	41,545	51,348	54,535	14,251	10,357	13	33,208	49,115	53,631	15,592	7,640	13
OLKR2/559m	257,789	45,456	72,269	78,325	22,550	9,481	13	32,600	62,318	69,573	19,071	6,118	12
OLKR1/572m	210,659	29,804	35,362	37,491	9,197	7,607	12	25,703	34,934	37,682	10,650	6,142	12
OLKR44/750m	303,058	31,410	31,589	32,188	2,005	11,200	12	25,937	33,448	36,295	10,346	7,124	12
OLKR29/798m	221,524	37,989	45,126	48,042	11,991	10,071	13	31,911	44,957	48,533	14,078	7,594	13

Table 2b. 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.

Archaea		All sequences						Normalized to 17,000 sequences					
Sample	Number of sequence reads	Observed OTUs	Chao1	ACE	Singletons	Doubletons	Shannon	Observed OTUs	Chao1	ACE	Singletons	Doubletons	Shannon
OLKR13/296m	507,373	27,111	29,516	30,699	5,835	7,076	10	3,957	13,380	15,062	2,867	435	10
OLKR3/318m	271,699	25,491	32,299	34,231	9,205	6,221	11	4,955	15,044	17,238	3,546	622	10
OLKR6/328m	446,380	21,597	22,930	23,781	3,861	5,588	10	3,776	11,705	14,020	2,748	475	9
OLKR23/347m	395,339	20,800	22,403	23,214	4,083	5,199	10	3,919	11,855	13,323	2,755	477	9
OLKR49/415m	210,545	22,600	23,372	24,004	2,975	5,733	12	7,023	17,088	19,874	4,738	1,114	12
OLKR9/423m	697,360	22,014	22,527	23,082	2,381	5,520	9	3,180	9,617	10,586	2,224	383	9
OLKR5/435m	769,026	21,127	22,235	23,078	3,515	5,574	9	2,596	10,114	10,078	1,852	227	9
OLKR9/510m	169,142	12,709	12,782	12,960	713	3,488	11	4,879	11,205	13,215	3,148	782	11
OLKR2/559m	100,101	15,359	24,950	27,026	7,840	3,203	11	5,119	14,497	16,488	3,548	670	11
OLKR1/572m	1,213,360	28,884	33,207	34,832	7,846	7,118	9	2,273	9,233	9,923	1,631	190	9
OLKR44/750m	17,716	6,436	8,748	9,750	2,890	1,805	12	6,325	8,743	9,804	2,921	1,763	12
OLKR29/798m	98,770	15,641	16,720	17,483	3,158	4,617	12	6,951	14,655	17,184	4,483	1,303	12

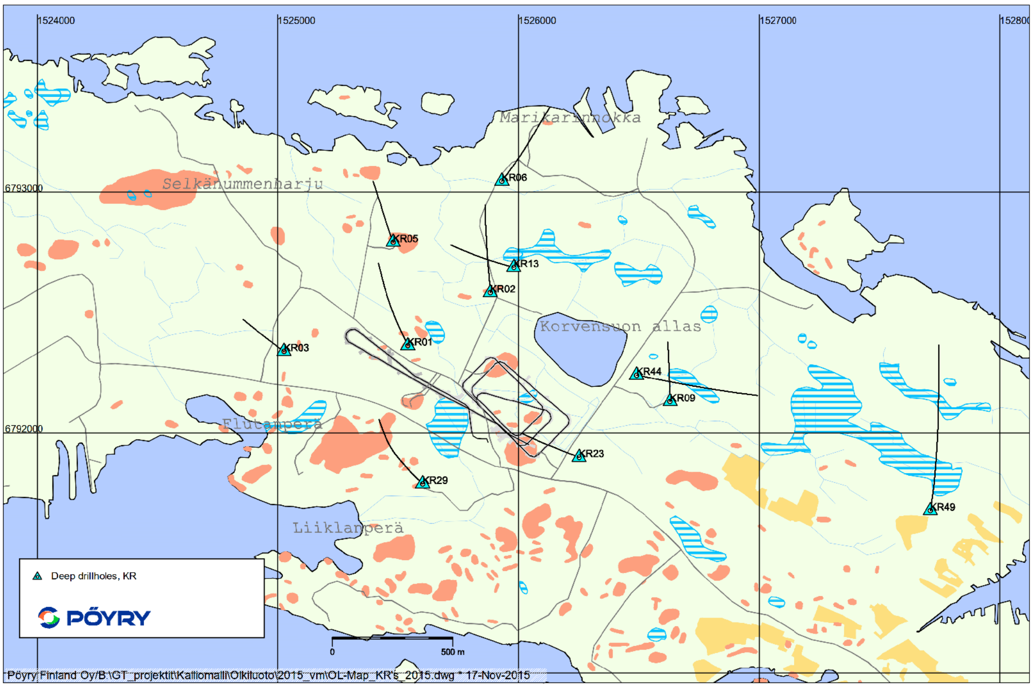

Figure 1. Map of Olkiluoto. The boreholes used in this study are marked with a turquoise triangle and the attached black line depicts the direction of the borehole. (with courtesy of Pöyry Oy, Nov 17th, 2015 by Eemeli Hurmerinta)

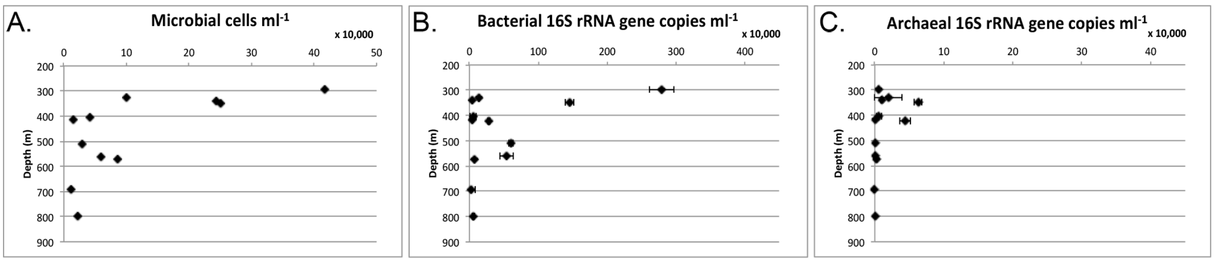

Figure 2. The concentration of A) microbial cells mL^{-1} determined by epifluorescence microscopy and the estimated concentration of B) bacterial and C) archaeal 16S rRNA gene copies mL^{-1} groundwater determined by qPCR in water conductive fractures situated at different depths in the Olkiluoto bedrock.

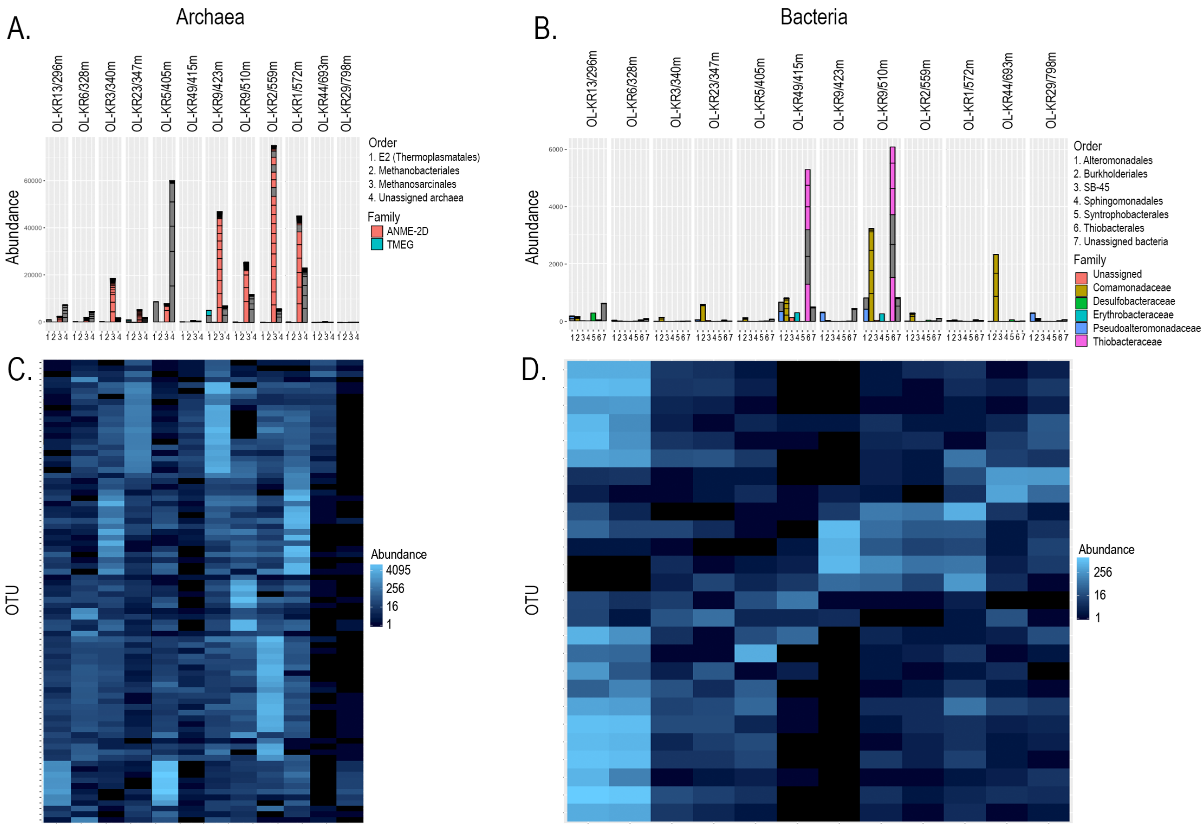

Figure 3. The core A) archaeal and B) bacterial community OTUs detected from at least 80% of the samples with heatmaps on the abundance of the C) archaeal and D) bacterial core community profiles. In A) and B) the OTUs are stacked in the columns according to the number of sequence reads, with the most abundant OTUs at the bottom of the columns. The OTU segments of the columns are colored according to the family to which they belong. Each Order is presented as a separate column for each sample.

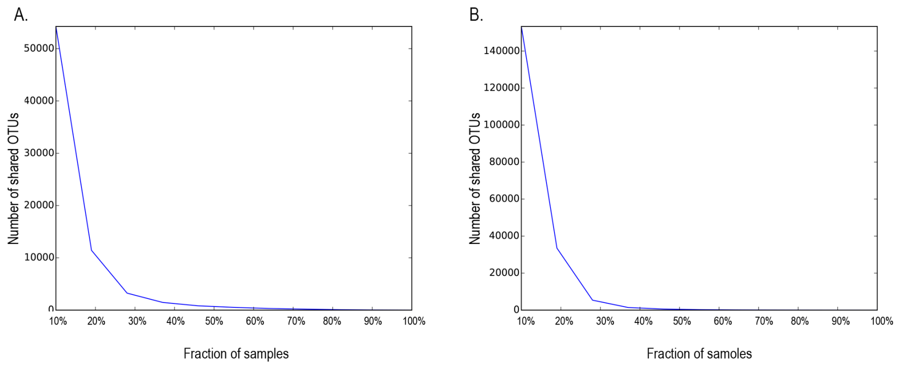

Figure 4. The number of shared A) archaeal and B) bacterial OTUs in the different samples. The number of shared OTUs is shown on the Y-axis and the proportion of samples on the X-axis.

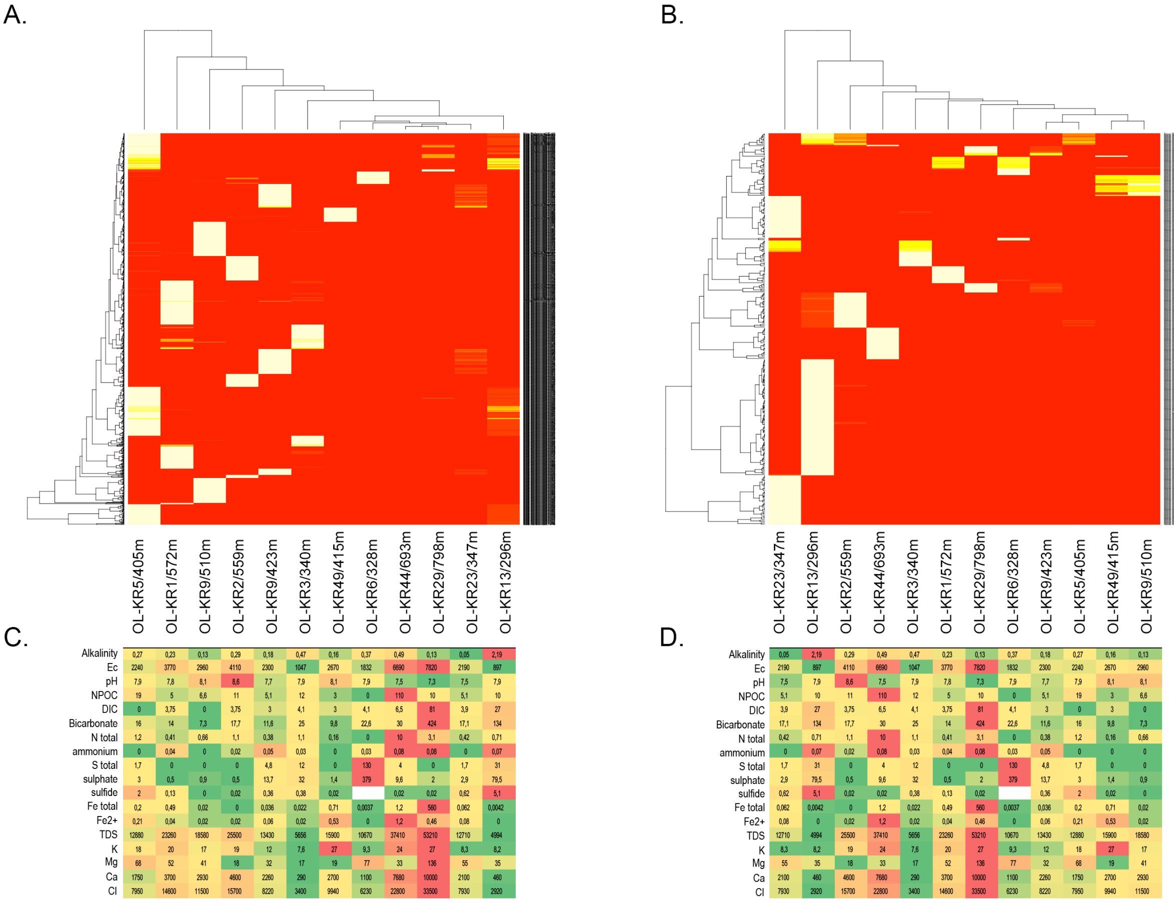

Figure 5. A UPGMA cladogram clustering the samples based on the A) archaeal and B) bacterial OTU profile according to the Bray Curtis distance model. Red colour indicates low abundance and yellow colour indicates high abundance. C) and D) show the corresponding physicochemical parameters as shown in Table 1, with the lowest values in green, medium values in yellow and high values in red.

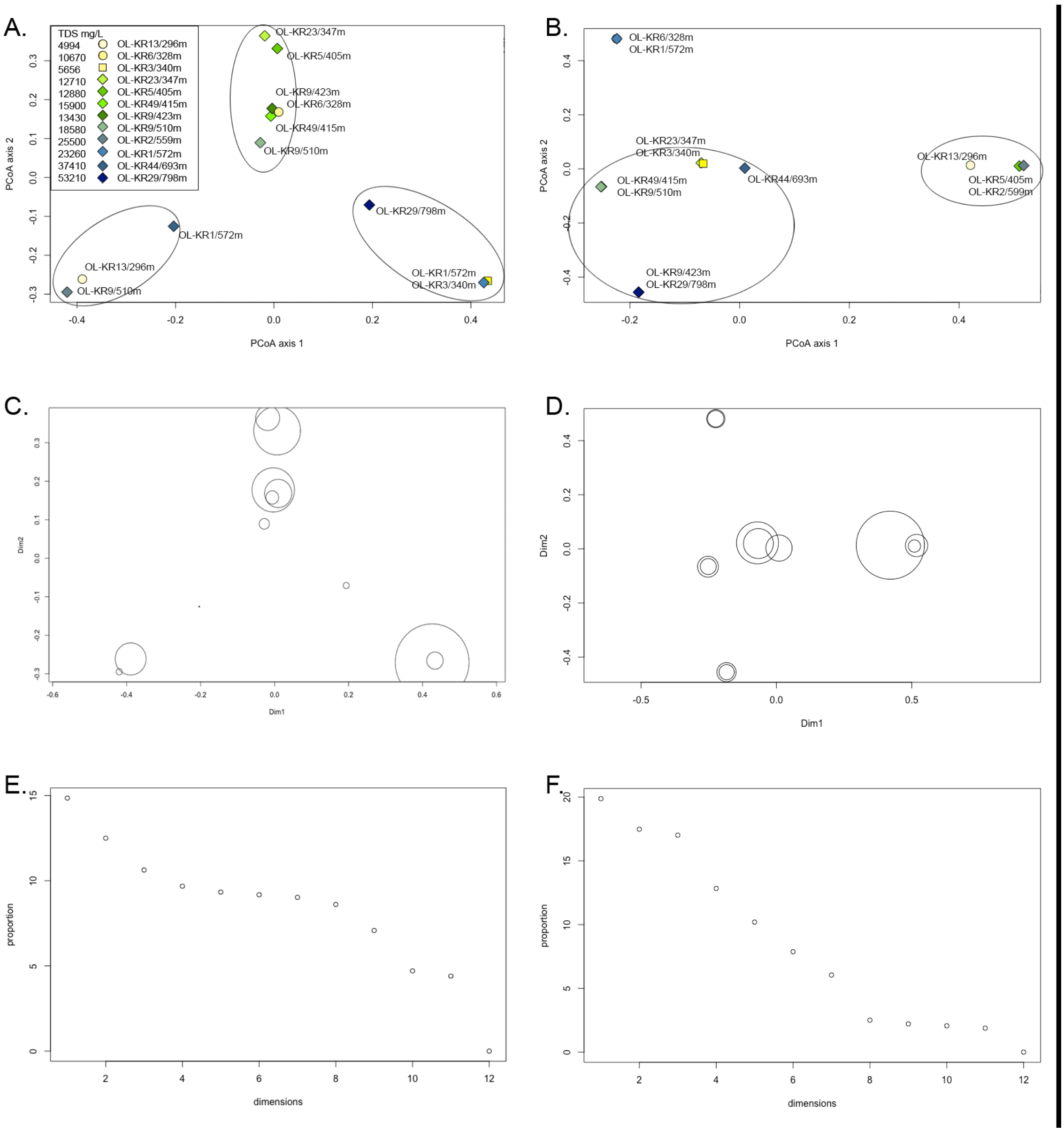

Figure 6. Principal coordinates analysis (PCoA) on the whole OTU profiles of the different samples based on Bray Curtis distance model. The analyses for archaea are shown in A) C) and E, and the bacteria in B), D) and F). In A) and B) the points indicate water type, where the circle is for brackish sulphate-rich water, the square is for brackish chloride-rich water and the diamond is for saline water. The colouring of the points are according to concentration of total solids as indicated in the upper left corner of figure A. In C) and D) the different sizes of the points describe the library size and the positions are the same as in A) and B). E) and F) display the proportion of variance for 12 dimensions, of which 1 and 2 were used for plotting the PCoA.

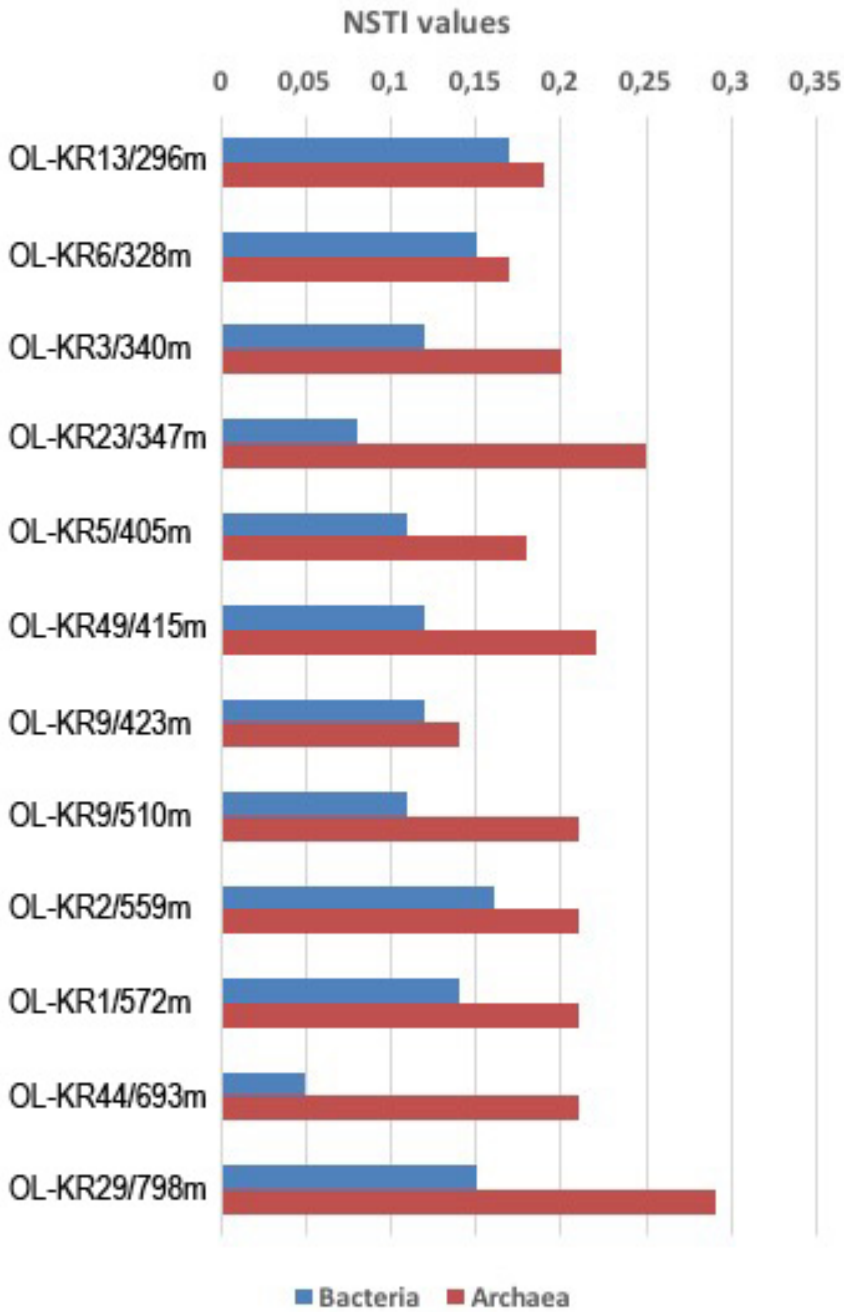

Figure 7. The nearest sequenced taxon index (NSTI) values calculated by PICRUSt for the bacterial (blue) and archaeal (red) connumities. The NSTI value describes the sum of phylogenetic distances of each OTU to its nearest relative with a sequenced reference genome, and measures substitutions per site in the 16S rRNA gene and the weighted the frequency of the each OTU in a sample dataset. A higher NSTI value indicates greater distance to the closest sequenced relatives of the OTUs in each sample.

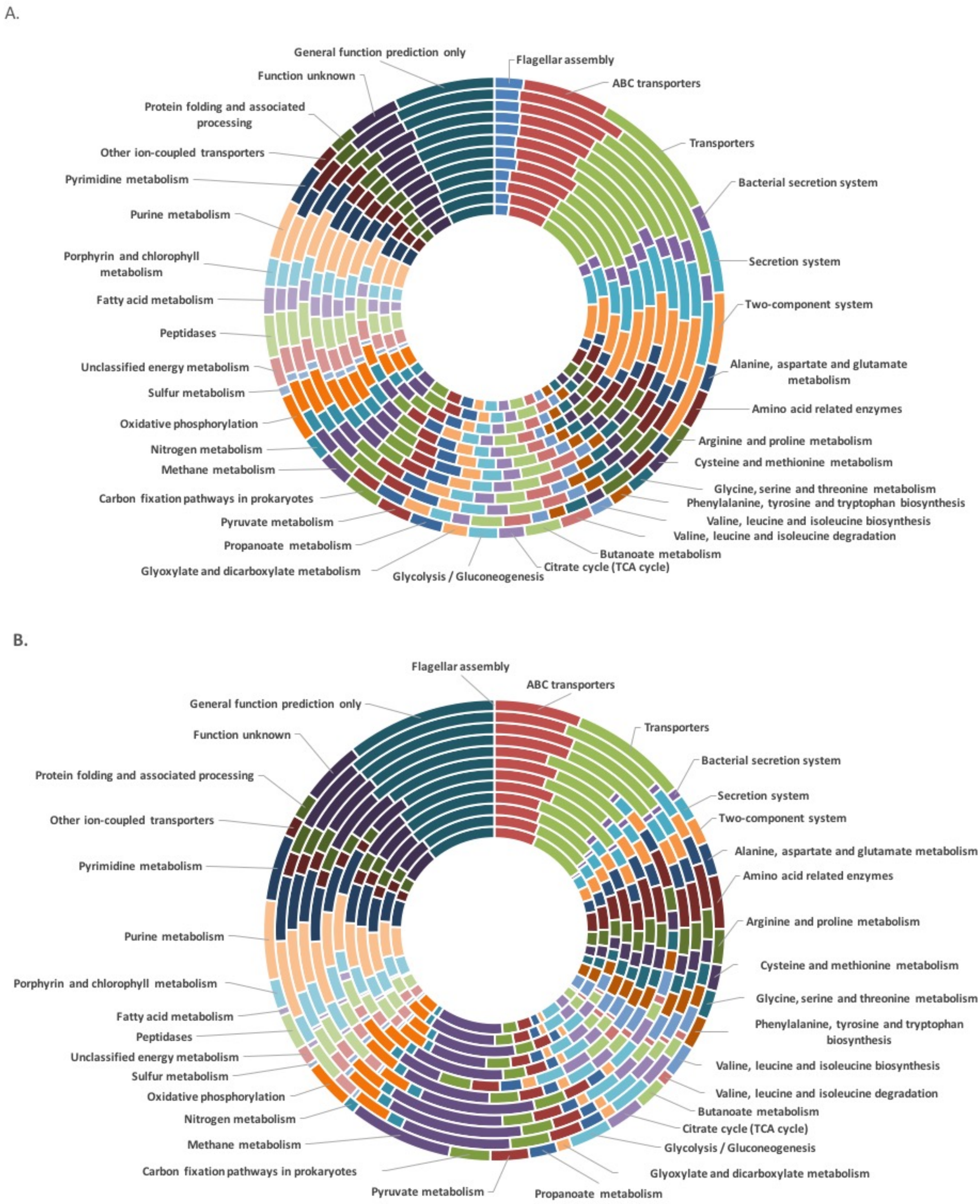

Figure 8. The relative abundance of predicted genes of the most abundant pathways identified in the A) bacterial and B) archaeal populations in the PICRUSt analysis. The pathways are presented according to KEGG. The samples are ordered according to depth, with OL-KR13/296m as innermost and OL-KR29/798m as the outermost sample.

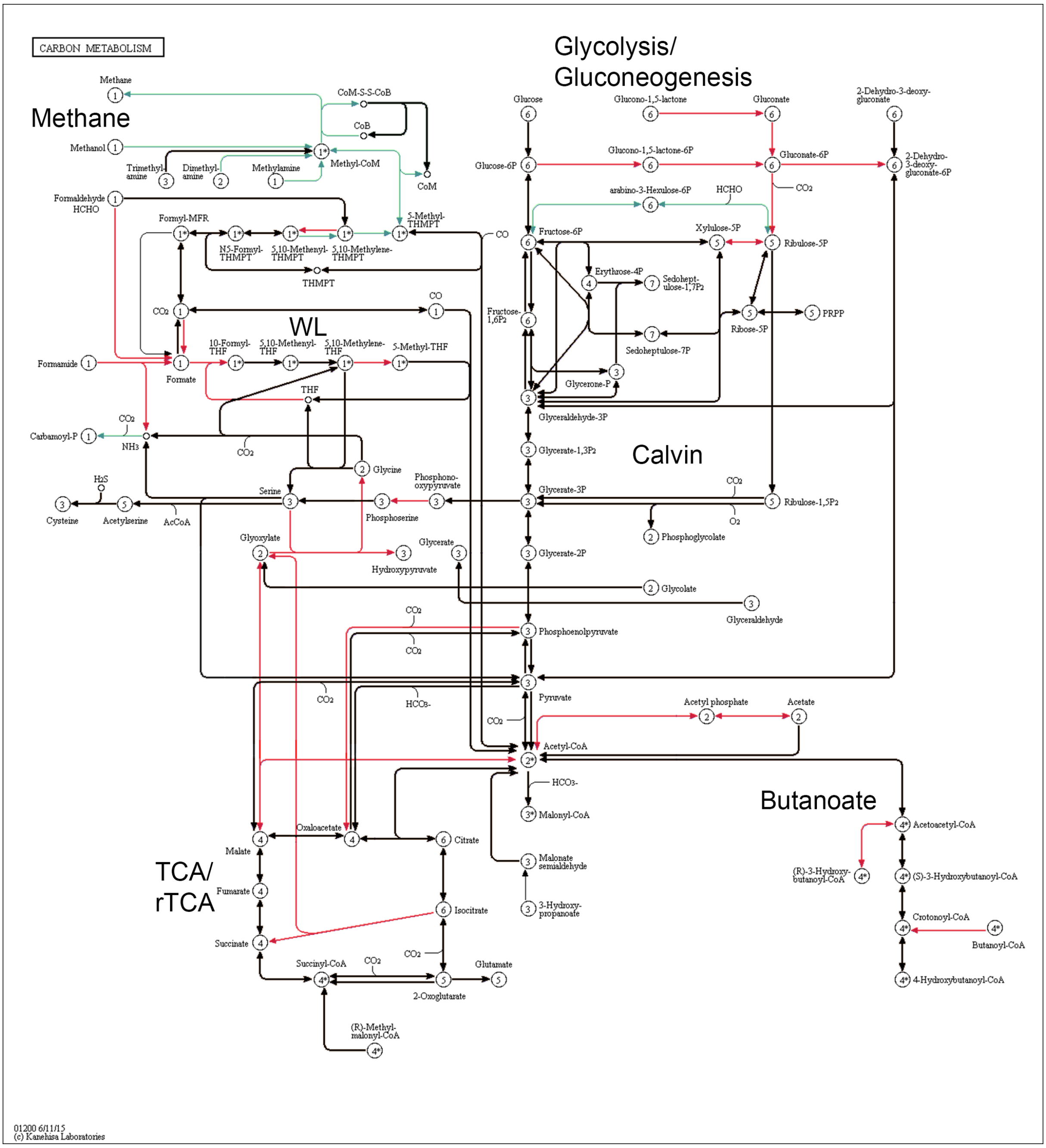

Figure 9. The microbial carbon metabolism pathway according to KEGG. The predicted genes combined from all samples were plotted on the map. Green arrows indicate enzymes predicted only in the archaeal communities, red arrows indicate genes predicted only in the bacterial communities, black arrows show enzymes predicted in both the archaeal and bacterial communities.

Supplementary figures

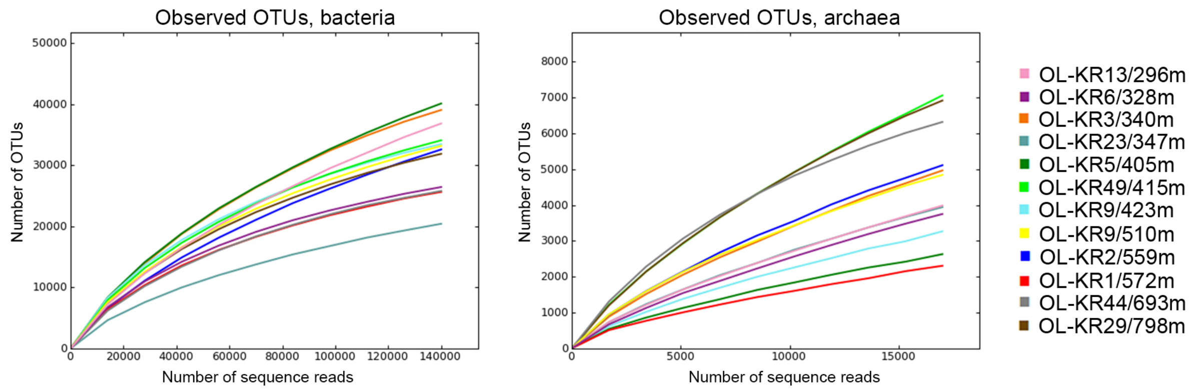

Figure S1. The rarefaction curves of observed bacterial (left pane) and archaeal (right pane) OTUs in each sample generated on sequence data normalized to 140,000 reads for bacteria and 17,000 reads for archaea.

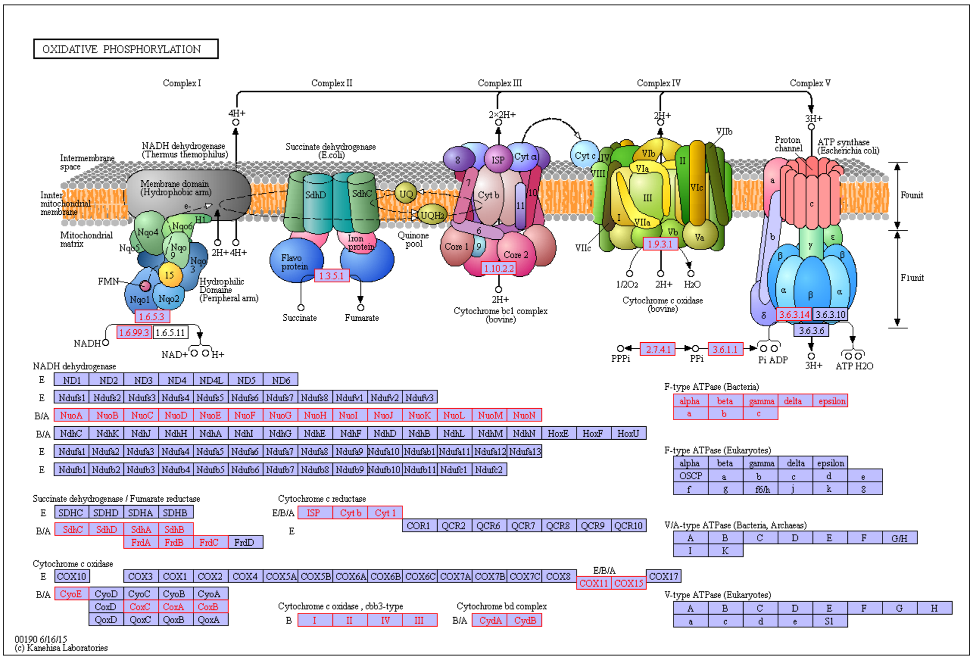

Figure S2. Bacterial oxidative phosphorylation according to KEGG. The predicted genes from the bacterial communities belonging to the oxidative phosphorylation are shown in pink.

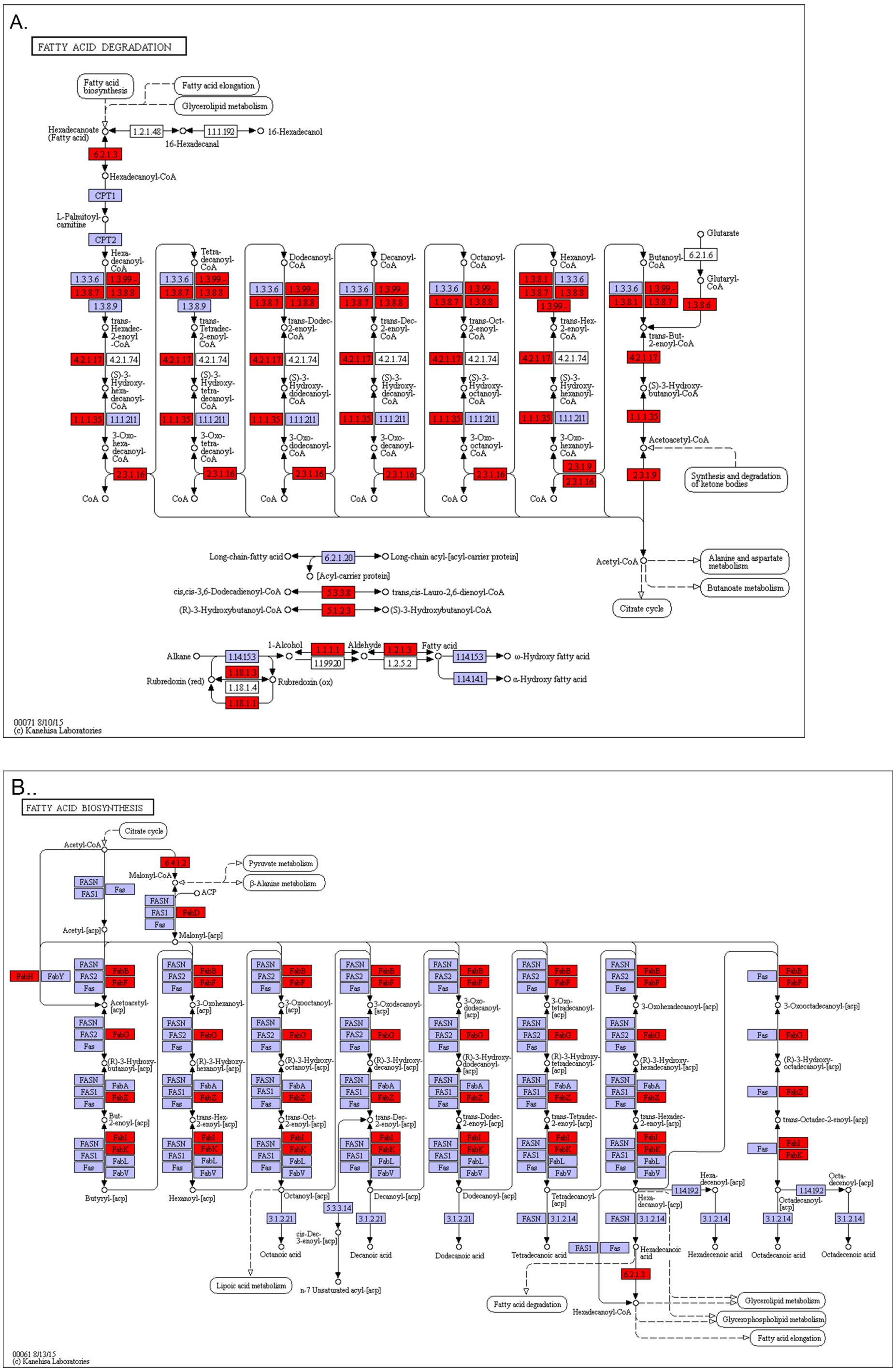

Figure S3. Predicted genes shown in red of the bacterial A) fatty acid degradation and B) fatty acid biosynthesis pathways, combined from all samples.

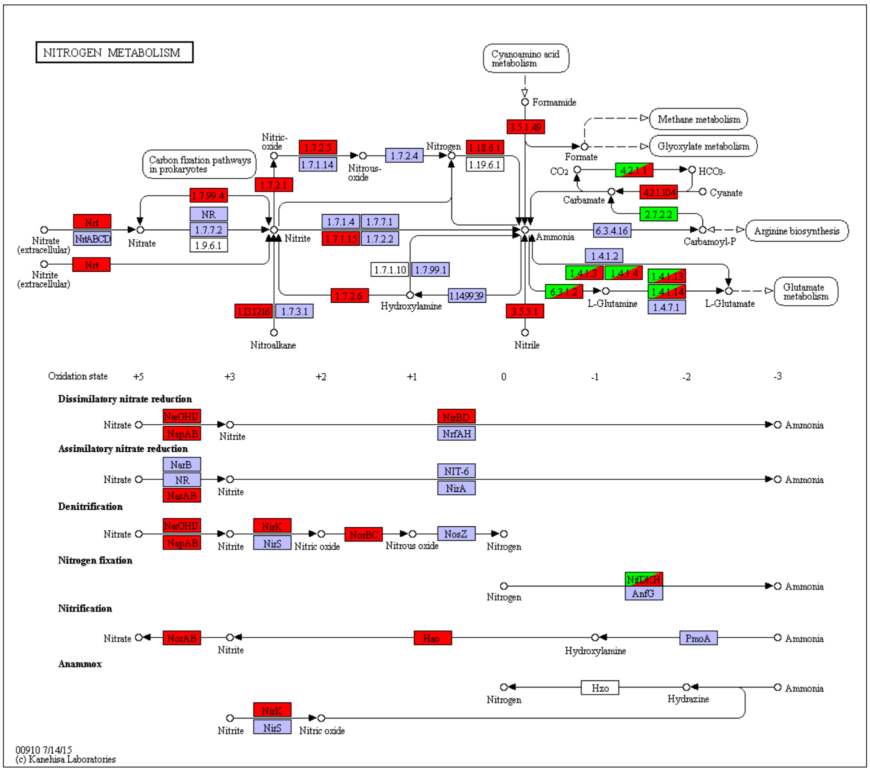

Figure S4. The predicted genes of enzymes included in the microbial nitrogen metabolism according to KEGG. Enzymes predicted from the bacterial communities are shown in red, archaeal communities in green and enzymes predicted from both archaeal and bacterial communities in green/red. Enzymes not predicted from either community are shown in blue or white.

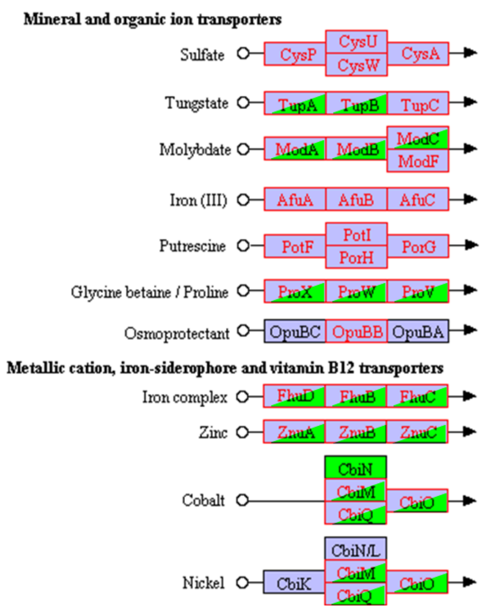

Figure S5. The genes of ABC transporters predicted from the bacterial (pink), archaeal (green) or both (pink/green) communities. Genes not predicted in any of the communities are shown in blue.

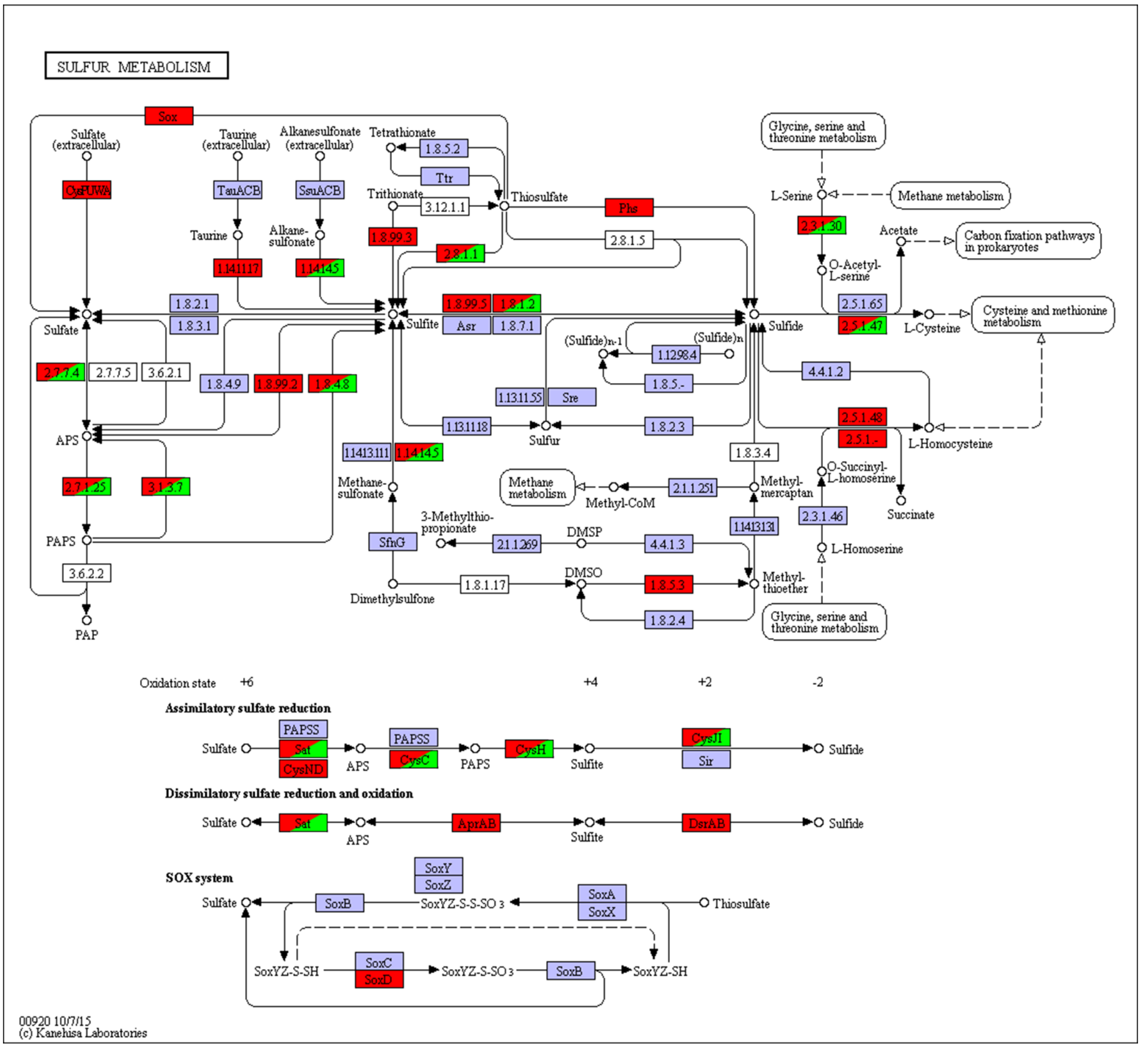

Figure S6. The predicted genes of enzymes included in the microbial sulphur metabolism according to KEGG. Enzymes predicted from the bacterial communities are shown in red, archaeal communities in green and enzymes predicted from both archaeal and bacterial communities in green/red. Enzymes not predicted from either community are shown in blue or white.