Response to the reviews:

(page and line numbers correspond to the revised manuscript document)

Metagenomic prediction and relative abundance of certain metabolic pathways: Referees #1 and #2 brought up a question about the reliability of metagenomic prediction with PICRUSt. As the PICRUSt analysis is depending on the partial 16S rRNA sequence data, this approach may only be used as a prediction of the properties of the metagenome. The high weighted NSTI (nearest sequenced taxon index) values ranging from 0.07 to 0.3 also indicate the uncertainty of the prediction. However, we did not infer major conclusions about the predicted metagenome although we feel that this analysis gives an overview of the top-level functionality of the metagenome of this environment. Nevertheless, we omitted the results of the HUMANn analysis (p.10, l. 19→, p. 16, l. 24→, p. 17, l. 8→), in which we used these predicted metagenomes as starting material. As this analysis depends ultimately to the partial 16S rRNA sequences, stating that certain pathways are complete may not be feasible. Furthermore, the dominance of heterotrophy in this ecosystem can be inferred from the taxonomy of the dominating OTUs, as the referee #1 stated.

Qualification of certain species as keystones: Referee #1 also wanted to know why we have named OTUs belonging to Burkholderiales and Clostridia as the keystone species besides the fact that this was the result of the network analysis. We based this statement to the suggestion by Steele et al. 2011, who claimed that the highly connected nodes in an OTU network would be analogous to microbial keystone species. Keystone species are described as organisms that have more significant role in the ecosystem that their abundance usually refers and if these organisms would be removed from the ecosystem, it would be subjected to dramatic changes. If we consider the possible role of species belonging Burkholderiales in the deep subsurface, they are most likely involved in hydrogen oxidation and thus are the potential primary producers in this oligotrophic environment. Clostridial species on the other hand could be occupying the niche for heterotrophic carbon assimilation, possibly providing H_2 produced in fermentation processes. Burkholderiales and clostridial species have been found from other serpentinite-hosted deep subsurface aquifers (Brazelton et al. 2012, Tiago et al. 2012). These studies have suggested that Burkholderiales and clostridial species play a significant role in carbon and energy exchange between one another and other bacterial groups in these

environments. These notions have now been added to the manuscript (p. 18, l. 13- p. 19, l.1). What comes to the abundance of these species/OTUs, we argue that they are in fact quite abundant at least in all other total fracture communities except at 1820 m. However, we have toned down the title of the manuscript towards more descriptive, as the editor and the referee #1 suggested.

Staphylococci as keystone species: The members of Staphylococcus as a keystone species was not expected. We believe that this result could be explained by contamination for two reasons: 1) staphylococci are common in human body, and 2) the amount of sequences affiliating with Staphylococcus were low but they were discovered in all samples. However, we can't ascertain the physiology of an OTU from it's taxonomic affiliation, and it can be that these bacteria with Staphylococcus-like 16S rRNA may be true members of the deep biosphere with unknown metabolic properties. This has been added to the manuscript as well (p. 19, l. 6-11).

Specific comments, referee #1:

Origin of the samples: The information about the samples coming from deep groundwater is now added to the title of the manuscript, to the abstract (p. 2, l. 5) and to the beginning of the discussion part (p. 18, l. 4, 7).

E. coli plasmids: Information about the E. coli plasmids containing 16S rRNA gene has been added (p. 7, l. 25).

PCR conditions and bias introduced by the nested PCR approach: First PCR was done with 30 cycles in both archaeal and mcrA –targeted PCR (p. 8, l. 22). As the number of archaea and especially methanogens have been shown to be low in comparison to bacteria both in previous studies (Purkamo et al. 2013, Nyyssönen et al. 2014) as well as in this study, it was imperative to choose the nested PCR approach in order to obtain community data of these microbial groups although the nested approach is likely to introduce some error to the results.

The length limit of mcrA sequences: The length limit of 100 bp was chosen because of the relatively low amount of longer good-quality reads. This may affect the phylogenetic analysis of mcrA sequences towards less reliable. However, this was necessary in order to obtain information about the methanogens in the fractures of Outokumpu.

The coverage of archaeal and sulphate reducer diversity: We calculated average estimate of coverage (ACE) for the dsrB dataset, and overall the observed species values were ranging from 50 to 80% of the estimated number of species. However, lower coverage was detected from the samples, where the number of sequences were low (180 m DNA and RNA, 1820 m DNA) and where one OTU dominated the reads (500 m RNA). With the archaeal dataset, all rare species were singletons, so ACE could not be calculated. However, several archaeal samples contained less than 10 observed species (500 m, 2260 m and 2300 m), although the amount of reads in these samples ranged from less than 300 to over 14000. Therefore, we estimate that the interpretation of archaeal community composition is likely correct.

Amount of cells: The range of cell numbers has been added to the text (p. 11, 1-2).

Table 1: Errors in the Table 1 heading that has now been corrected. The concentrations of cations and anions are in mg/L, not mg/ml. Also, alkalinity was measured in mmol/L, not mmol/ml. The naturally high salinity of the samples complicated the geochemical analysis, which reflects to the raised detection limit of several cations and anions.

Table 2: The lower cell counts than 16S rRNA gene copy numbers in the upper samples may be explained by multiple 16S rRNA gene operons in one cell, but the inversed trend for the deeper samples may be for example due to lower DNA extraction yields in the deeper fractures, where salinity is higher. Last sentence of the table heading has been corrected.

Specific comments, referee #2:

Title and the information about the sample origin: We have changed the title of the manuscript according to referees' and editor's suggestions. Also information about the samples originating from groundwater has been added (p. 2, l. 5, p. 18, l. 4, 7).

DNase treatment of only some of the RNA extracts: DNase treatment was used only to the RNA extracts that were verified to contain residual DNA in order to avoid the loss of material, as the amount of retained RNA was low already in the beginning, and during the DNase treatment we would have lost more of the precious material.

The coverage of archaeal primers: Primers used in the archaeal nested PCR are reported to have frequent mismatch rate to some archaeal groups, such as SAGMEGs (Hadesarchaea)(Teske et al. 2008, Wang & Qian 2009). This might lead to underestimation of these groups in the final results. However, we did find a large amount of SAGMEG sequences

in one depth of the bedrock, further reinforcing the results from previous study (Purkamo et al. 2013). Using the Ribosomal Database Project's probe match platform, we detected that archaeal primers used in the first step of the nested PCR matched 39% of the archaeal 16S rRNA genes in the database, and the primers used in the second step 31%. Nevertheless, phylum Thermococci, in which newly described Hadesarchaea (SAGMEGs) belong to, is covered much better as the overall archaea, as the coverage range of the used primers is from 85-91%. We have added this primer coverage data to manuscript (p. 12, l. 3-9).

PCR-conditions of the 16S rRNA genes: Quantitative PCR conditions for bacterial 16S rRNA genes are described in the manuscript in p. 7, l. 9→). Archaeal 16S rRNA -targeted qPCR conditions are previously described in Bomberg et al. (2015): "The qPCR reactions were performed in 10µL reaction volumes using the KAPA 2 × Sybr 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 1µL 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 15s at 95 °C, 30s at 55 °C and 30s 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.". Amplicon library construction for sequencing is described in p. 8, lines 7-31.

The occurrence of phototrophs in over 2 km depth: In this sample, 4,5% of the sequences affiliated with Streptophyta, likely representing plant material. These are likely contaminants that could be originating from the drilling operations, where some peat material was used in the drilling mud as a stabilizer during the drilling. We have added a detailed explanation to the manuscript (p. 21, l. 2-7).

New information of SAGMEG archaea: Recently published data about the archaeal group SAGMEG, newly named as Hadesarchaea, provided us more detailed information about their possible energy and carbon metabolism. We have added a discussion part about the role of these archaea in the deep terrestrial biosphere of Outokumpu into the manuscript (p. 22, l.15-30). Also Supplementary Table 3 has been changed accordingly.

Difference between archaeal DNA and RNA community structures in 967 m: The number of reads in the samples originating from the fracture zone in 967 m depth was highly similar, so this is probably not the reason for different result of RNA and DNA –based analyses. As different microbes have different amount of ribosomes depending on the activity of the microbe and the need for production of proteins, this could explain the difference between total and active archaeal community structure at 967 m. There are reports from deep subsurface environments with similar results (Posiva report by Bomberg & Itävaara 2012 http://www.posiva.fi/files/3096/WR_2012-27.pdf, Miettinen et al. 2015 doi: 10.3389/fmicb.2015.01203). In addition, it was demonstrated in recent study that only certain microbes in the total population are active but their numbers might be so low that they are not even detectable via DNA-based analysis (Rajala et al. 2015).

Low archaeal diversity due to the low coverage of primers and PCR bias: We agreed with referee #2 that the low archaeal diversity detected can be due to the low coverage of the primer pairs used as well as be a result from nested PCR approach (p.23, l.14-15).

Microbial co-occurrence patterns in deep Precambrian bedrock fracture fluids

L. Purkamo[1], M. Bomberg[1], R. Kietäväinen[2], H. Salavirta[1], M. Nyyssönen[1], M. Nuppunen-Puputti[1], L. Ahonen[2], I. Kukkonen[2,*] and M. Itävaara[1]

[1]{VTT Technical Research Centre of Finland Ltd., Espoo, Finland}

[2]{Geological Survey of Finland (GTK), Espoo, Finland}

[*]{now at: University of Helsinki, Helsinki, Finland}

Correspondence to: L. Purkamo (lotta.purkamo@gmail.com)

Poistettu: The keystone species of

Muotoiltu: Fontti:(Oletus) Times New Roman

Poistettu: deep

Muotoiltu: Fontti:(Oletus) Times New Roman

Poistettu: biosphere belong to *Burkholderiales* and *Clostridiales*

Muotoiltu: Fontti:(Oletus) Times New Roman

Abstract

The bacterial and archaeal community composition and the possible carbon assimilation processes and energy sources of microbial communities in oligotrophic, deep, crystalline bedrock fractures is yet to be resolved. In this study, intrinsic microbial communities from groundwater of six fracture zones from 180-2300 m depths in Outokumpu bedrock were characterized using high-throughput amplicon sequencing and metagenomic prediction. *Comamonadaceae*-, *Anaerobrancaceae*- and *Pseudomonadaceae*-related OTUs form the core community in deep crystalline bedrock fractures in Outokumpu. Archaeal communities were mainly composed of *Methanobacteraceae* -affiliating OTUs. The predicted bacterial metagenomes showed that pathways involved in fatty acid and amino sugar metabolism were common. In addition, phylotypes with mixed metabolic properties as well as organotrophs were more abundant than lithotrophic phylotypes. This indicates that heterotrophic carbon metabolism is more important for microbial communities of the fracture zones. Network analysis based on co-occurrence of OTUs revealed possible "keystone" genera of the microbial communities belonging to *Burkholderiales* and *Clostridiales*. Bacterial communities in fractures resemble those found from oligotrophic, hydrogen-enriched environments. Serpentinization reactions of ophiolitic rocks in Outokumpu assemblage may provide a source of energy and organic carbon compounds for the microbial communities in

Poistettu: relative

Poistettu: ce

Poistettu: of genes coding the enzymes of autotrophic carbon fixation pathways in predicted metagenomes was low

Muotoiltu: Korosta

Poistettu: assimilation

Poistettu: the

the fractures. Sulfate reducers and methanogens form a minority of the total microbial communities, but OTUs forming these minor groups are similar to those found from other deep Precambrian terrestrial bedrock environments.

1 Introduction

The microbial communities in deep terrestrial subsurface biosphere contribute significantly to the overall biomass on Earth (Whitman et al. 1998, McMahon and Parnell 2014). It is essential to understand the metabolic capacity and energy sources of the microbial communities in deep biosphere in order to evaluate their role in global biogeochemical cycles, assess the risks these communities might cause to for example geological long-term storage of nuclear waste, and even to estimate the possibility of microbial life in deep subsurface of other planetary bodies. In general, chemolithoautotrophic organisms are thought to be the primary producers in deep crystalline rock environments, into which sunlight, organic carbon or oxygen produced in photosynthesis, do not penetrate (Gold 1992, Pedersen 1997, 2000). Therefore, energy and carbon sources for deep biosphere have to be geochemical. The most important source of reducing power in deep subsurface is H_2. It is produced in abiotic reactions such as through radiolysis of H_2O, water-rock interactions such as serpentinization, but also by microbial activity (Pedersen 2000, Lin et al. 2005, McCollom 2013, Szponar et al. 2013). Carbon sources for microbes in deep subsurface are usually in the form of CO_2, CH_4 or other small hydrocarbons. Abiotic synthesis of organic carbon may take place through Fischer-Tropsch type reactions and provide a photosynthesis-independent carbon source for

Poistettu: or the

heterotrophic organisms in deep terrestrial biosphere (Proskurowski et al. 2008, McCollom et al. 2010, Etiope and Sherwood Lollar 2013, Kietäväinen and Purkamo 2015). This process may be triggered and enhanced by continuous H_2 flux provided by for example serpentinization. Numerous studies have characterized microbial communities of deep Precambrian rock formations (eg. Pedersen et al.1996, 2008, Hallbeck and Pedersen 2008, 2012, Lin et al. 2006, Gihring et al. 2006, Silver et al. 2010, Itävaara et al. 2011, Nyyssönen et al. 2012, Purkamo et al. 2013, 2015, Osburn et al. 2014, Bomberg et al. 2015a, b). Although some of these studies have explored the energy and carbon sources or electron accepting processes in these environments, attention has been focused mainly on chemoautotrophic organisms utilizing H_2 and CO_2. After all, abiotic synthesis of organic carbon could also provide a photosynthesis-independent source of carbon and thus support heterotrophic organisms in deep biosphere (Amend and Teske 2005, Schrenk et al. 2013). However, heterotrophic involvement to the carbon cycling and energy production in the deep continental bedrock biosphere has been rather neglected (Amend and Teske 2005), although it was recently suggested that heterotrophy might play a significant role in deep fluids of Fennoscandian crystalline rock (Purkamo et al. 2015).

While the microbial communities in deep marine subsurface environments have been intensively characterized within the last decade with next-generation sequencing methods (Sogin et al. 2006, Biddle et al. 2008, 2011, Brazelton et al. 2012), high-throughput (HTP) sequencing techniques have only recently emerged in characterization of the terrestrial deep subsurface microbial communities (Nyyssönen et al. 2014, Bomberg et al. 2014, 2015a, b, Lau et al. 2014, Mu et al. 2014). Vast amount of data obtained from HTP sequencing studies can be used to estimate ecological measures such as species richness, abundance and β-diversity, but it also allows the exploration of significant relationships between microbial taxa and their coexistence in a specific environment (Zhou et al. 2011, Barberan et al. 2012, Lupatini et al. 2014). These co-occurrence patterns, i.e. interactions between different microbial taxa and the complexity of the microbial communities can significantly contribute to the processes that will take place in the ecosystem (Zhou et al. 2011). In addition, keystone organisms can be identified from co-occurrence patterns of the community (Steele et al. 2011). Keystone organisms often have a greater role in the ecosystem functionality than their abundance refers (Power et al. 1996). For example, in hydrogen-driven lithoautotrophic ecosystems, autotrophic methanogens can be responsible of primary production of the whole

Poistettu: .

ecosystem (Pedersen 2000, Nealson et al. 2005). Moreover, these diverse minority groups with low abundance, i.e. the so-called rare biosphere, can be an almost infinite source of genetic potential to be distributed through the microbial populations via gene transfer (Sogin et al. 2006).

In this study, we used high-throughput amplicon sequencing, metagenome prediction and co-occurrence analysis 1) to describe the microbial community structure, 2) to detect key microbial genera of deep fracture fluids in Outokumpu, 3) to evaluate the possible carbon assimilation processes taking place in deep bedrock and ultimately, 4) to understand the origin of carbon and energy sources in Outokumpu Palaeoproterozoic deep bedrock and to establish links between microbial communities and the geology and geochemistry in Outokumpu crystalline rock biosphere. Groundwater samples were collected from six different fracture zones located at depths ranging from 180 m to 2300 m, and we characterized bacterial and archaeal communities in these fractures by their 16S rRNA genes and transcripts. In addition, we characterized two functional groups carrying out important electron accepting processes in deep subsurface, namely sulphate reduction and methanogenesis by dissimilatory sulfite reductase and methyl-coenzyme M reductase genes, respectively.

Poistettu: by

2 Materials and Methods

2.1 Sample collection and geochemistry

Deep subsurface fracture fluids were collected during years 2009-2011 from the Outokumpu Deep Drill Hole, Eastern Finland. The sampling was conducted from overall six depths, 180, 500, 967, 1820, 2260 and 2300 m, as described previously (Purkamo et al. 2013). Shortly, 967 m and shallower depths were packer-isolated and purged for 21-42 days, and deeper fractures were sampled with slow continuous pumping of the fluid from the fracture depth for 9-63 days in order to flush the drill hole. Care was taken to ensure that the pumping rate did not exceed the rate of inflow from the fracture zone. The hydrogeological characteristics of these fluids differ with depth (Table 1). Kietäväinen et al. (2013) described five different water types in Outokumpu, and the fracture zones in this study represent the types I (180 m), II (500 and 967 m), IV (1820 and 2260 m) and V (2300 m). The type I water is characterized with high pH (around 10) and higher alkalinity than other water types in Outokumpu. High pH in the drill hole water column probably originates from cementation within the uppermost 200 m of the drill hole, while during long-term pumping of the 180 m fracture zone, pH dropped to

the level of 8.5. Water type II contains the highest amount of dissolved gases in the whole water column, of which approximately 75% (22-32 mmol/l) is methane. Distinctive greenish colour and unpleasant 'rotten egg' odour are typical for water type IV, indicating presence of reduced sulfur compounds. Water type V also has special features, such as high K and Li concentration due to the interaction with surrounding granitic rocks. In addition, the dominant dissolved gases in the two deepest water types IV and V are He and H_2, in contrast to the CH_4 -dominated water types above 2 km depth (Kietäväinen et al. 2013).

The fluid from each fracture zone was collected in the field into sterile, acid-washed glass bottles (Schott) in an anaerobic chamber (MBraun, Germany). The anaerobic conditions in the chamber were achieved as previously described (Purkamo et al., 2013). The biomass for RNA and DNA extraction was collected on nitrocellulose acetate filters (Corning Inc., NY, USA) from 3 × 1 l and 3 × 0.5 l of fracture fluid by vacuum suction. The filter was cut from the filter funnel with sterile scalpel and placed immediately to dry ice in a sterile 50-ml plastic tube (Corning Inc., NY, USA). In the laboratory, the samples were preserved at -80 °C before processing. In addition, duplicate 100 ml fluid samples for microbial cell enumeration were obtained from each fracture zone. Sterile, acid-washed 120-ml serum bottles were flushed with a small amount of fracture fluid in the anaerobic cabinet and subsequently filled with 100 ml of the sample fluid, capped with butyl rubber stoppers, sealed with aluminium crimp caps and kept refridgerated until further processing in the laboratory within five days after the sampling.

2.2 Enumeration of the total amount of microbes

In order to calculate the total amount of microbes in fracture fluids, microbes were stained with 4'-6-diamidino-2-phenylindole (DAPI). Preparation of the duplicate samples for examination by microscopy was conducted as in Purkamo et al. (2013). Stained microbes were collected from 5 ml fracture fluid samples by filtering, rinsed and filter was placed on microscopy slide. The total cell number in the samples was based on the sum of counted cells and the effective area of the filter divided by volume of filtrated sample, number of randomly selected microscopy fields and the surface area of the field at 100 × magnification.

2.3 Nucleic acids preparation

DNA and RNA were extracted from the biomass with PowerSoil DNA or PowerWater RNA extraction kit (MO BIO Laboratories, Inc., CA, USA) as previously described (Purkamo et al.

2013). An additional DNase treatment was applied to RNA extracts that had DNA contamination detected by PCR performed with with P1 and P2 primers for bacterial 16S rRNA gene (Muyzer et al. 1993) as previously described in Purkamo et al. (2013). RNA was reverse-transcribed in triplicate reactions with random hexamers using the Superscript III Reverse Transcriptase kit (Invitrogen, ThermoFisherScientific, MA, USA) as described in Purkamo et al. (2013). The triplicate reactions were pooled and subsequent cDNA as well as DNA were stored at -80 °C. Negative controls for reagents were included in every extraction and translation step.

2.4 Quantitative estimation of bacterial and archaeal communities

The bacterial and archaeal numbers in each fracture were estimated with quantitative PCR from DNA extracts. 16S rRNA gene copy number was used as a proxy of the quantity of bacteria and archaea. In addition, qPCR was conducted to calculate the abundance of genes representing the key metabolic processes in anaerobic subsurface environments, namely sulfate reduction and methanogenesis with dissimilatory sulphite reductase (*dsr*B) and methyl-coenzyme M reductase (*mcr*A) genes, respectively. Bacterial 16S rRNA gene copy numbers were determined with V3 region-targeted primers P1 and P2 (Muyzer et al. 1993) resulting in a 190-bp product. A 370 bp fragment of *dsrB* gene and transcript was amplified with the primer pair DSRp2060f and DSR4R (Wagner et al. 1998, Geets et al. 2006). Archaeal 16S rRNA genes were amplified with ARC344f (Bano et al., 2004) and Ar774r (modified from Barns et al. 1994) primers producing a 430 bp product. A 330-bp fragment of *mcr*A gene was amplified with the primer pair ME1 and ME3rc (Hales et al. 1996, Nyyssönen et al. 2012).

Bacterial 16S rRNA gene-targeted qPCR was performed in triplicate reactions of each sample with KAPA™ SYBR® Fast 2 × Master mix for Roche LightCycler 480 (Kapa Biosystems, Inc., MA, USA) and 0.3 μM each of forward and reverse primer. The qPCR was performed on a Roche LightCycler 480 (Roche Applied Science, Germany) on white 96-well plates (4titude, UK) and sealed with transparent adhesive seals (4titude, UK). 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 and a final extension step of 3 min at 72 °C. After the quantification analysis, the melting curves for each reaction were determined. The melting curve analysis consisted of a denaturation step for 10 s at 95 °C followed by an

annealing step at 65 °C for 1 min 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. Amplification of *dsr*B, archaeal 16S rRNA and *mcr*A genes were performed in triplicate for each sample as described in Purkamo et al. (2013) and in Nyyssönen et al. (2014), respectively. The gene copy numbers were calculated by comparing the amplification result to a standard dilution series. Bacterial 16S rRNA and *dsr*B gene copy numbers were determined in each sample by comparing the amplification result to a standard dilution series ranging from 0 to 10^7 of plasmid DNA containing 16S rRNA gene of *Esherichia coli* ATCC 31608 or from 1.5×10^1 to 1.5×10^7 copies of *Desulfobulbus propionicus* DSM 2554 *dsr*B gene, respectively. Archaeal 16S rRNA and *mcr*A gene copy numbers were determined by comparing the amplification result to a dilution series of genomic DNA of *Halobacterium salinarum* DSM 3754 or to 5 to 5×10^6 copies of *Methanothermobacter thermautotrophicus* DSM 1053 *mcr*A gene, respectively. No-template controls as well as nucleic acid extraction and translation reagent controls were analysed with the corresponding samples in the same run. The inhibition effect of the samples was evaluated by mixing a specified amount of standard dilution to each sample DNA or cDNA. Spiked reactions were then subsequently amplified using the same protocols as described above. The inhibition in each sample could be evaluated by comparing the amplification efficiency of the sample-spiked standard DNA to the corresponding standard dilution quantity in the standard curve. Inhibition was found to be low in all samples (data not shown).

2.5 High-throughput amplicon sequencing

PCR amplicon libraries from hypervariable region V1-V3 of bacterial 16S rRNA gene were generated with barcoded 8f and P2 primers (Edwards et al. 1989, Muyzer et al. 1993). Amplification of *dsr*B gene fragment for dissimilatory sulphate reduction was done with 2060f and 4R primers with barcode sequences (Wagner et al., 1998, Geets et al., 2006). Archaeal libraries were produced with nested PCR method, first using A109f and A915r primers (Großkopf et al., 1998 and Stahl and Amann, 1991, respectively) to amplify ca. 800 bp long fragment of the archaeal 16S rRNA gene and using the resulting product as template in PCR reaction with barcoded A344f and A744 primers (Bano et al., 2004 and modified from Barns et al., 1994, respectively). *Mcr*A amplicons were also produced with nested PCR, first applying mcrA463f and mcrA1615r primers (Nyyssönen et al., 2012) and secondly barcoded primers Me1 and Me3 (modified from Hales et al., 1996). PCR reaction mix composed of one

unit of proofreading Phusion DNA Polymerase (ThermoScientific), 1 × high fidelity buffer and dNTP mix (2.5 mM each), filled to 50 µl with molecular biological grade H_2O. Dimethylsulfoxide was used in all PCR reactions to enhance the template availability to polymerase. The amplification cycle consisted of an initial denaturation at 98 °C for 30 s, 35 (bacteria and *dsr*B), 30 (for the first PCR) or 40 (for the second PCR) (archaea and *mcr*A) times repetition of 10 s at 98 °C, 15 s at 55 °C and 30 s at 72 °C, and a final extension step of 5 min at 72 °C. Three samples were used for each fracture zone community (RNA or DNA) and two amplification reactions of each sample replicate, thus resulting to maximum of six positive reactions (verified with agarose gel electrophoresis). Succesful reactions were pooled prior to sequencing. PCR reactions were performed also for nucleic acid extraction and reagent control samples. The sequencing of the 180 m samples was performed at Research and Testing Laboratory (Texas, USA) and the rest of the samples, at the institute of Biotechnology (Helsinki, Finland) using the FLX 454 Titanium – platform (454 Life Sciences, Branford, CT, USA).

Poistettu: or

Poistettu: 40

2.6 Quality control, classification and phylogenetic analysis of sequences

Sequences were analyzed using Mothur (v. 1.32.1) (Schloss et al. 2009) and QIIME programs (MacQIIME v. 1.7.0), Caporaso et al. 2010). The QIIME pipeline was used with 16S rRNA gene sequences and Mothur with the functional gene sequences. In QIIME, sequences were compared against Greengenes representative OTU set version gg_13_8 with 97% similarity and the taxonomy was assigned with RDP. With 16S rRNA sequences, the quality score window was set to 50 and sequences shorter than 360 and longer than 450 base pairs were discarded. The proximal primer sequences were allowed to have two or six nucleotide mismatches for bacterial and archaeal sequences, respectively. The high mismatch rate allowed for archaeal sequences was due to an extra guanine nucleotide in the primer sequence of the 500 m sample. Sequences of the functional genes representing sulphate reducers (*dsr*B) and methanogens (*mcr*A) were analyzed with Mothur. Raw flowgrams were denoised with the PyroNoise algorithm to reduce PCR and sequencing noise in the data (Quince et al. 2009). All *dsr*B sequences shorter than 200 bp were discarded and no mismatches in the forward primer sequence were allowed. The length limit for *mcr*A sequences was set to 100 bp and four mismatches in primer sequence were allowed due to ambiguous bases in the primer sequence. The resulting sequences were further aligned with model alignments of *dsr*B and *mcr*A sequences from the Fungene repository (Fish et al. 2013) and sequences were assigned

to OTUs with nearest neighbour clustering method. Final phylogeny of the representative OTUs was done using the Geneious Pro software package, version 6.1.7 (Biomatters Inc., New Zealand) and blastn and blastx for comparison of the representative OTU sequences to NCBI's databases (Altschul et al. 1990). All sequence data were uploaded to ENA database with accession numbers ERS846377-ERS846388 (bacteria), ERS846389-ERS846397 (archaea), ERS846399- ERS846407 (dsrB) and ERS846408- ERS846414(mcrA).

2.7 Ecological indices and statistical analyses

Chao1 richness estimates were calculated for the bacterial and archaeal communities with 97% sequence similarity using the alpha.diversity.py command in QIIME. The estimates of diversity, richness and rarefaction were calculated from random subsample of 3030 sequences per sample for bacteria and 270 sequences per sample for archaea. Same α-diversity estimates for *dsr*B and *mcr*A datasets were calculated in Mothur from subsamples of 115 and 1712 sequences, respectively. Due to the low amount of *dsr*B sequences (47) retrieved from the 180 m fracture, these data were not subsampled. The bacterial and archaeal OTUs with resolved taxonomy were compared to the hydrogeochemical data as well as to the lithology of each fracture zone. Canonical correspondence analysis was performed with Past3 to the environmental metadata and taxonomical OTU matrix with all archaeal OTUs and bacterial OTUs with more than 0.1 % abundance in the fracture communities (Hammer and Harper, 2001).

2.8 Prediction of functionality and co-occurrence analysis

De novo OTUs were removed from the 16S rRNA OTU taxonomy file (.biom-table) prior to uploading to the Galaxy pipeline for PICRUSt analysis (Goecks et al. 2010, Blankenberg et al. 2010, Giardine et al. 2005, Langille et al. 2013). PICRUSt compares 16S rRNA marker gene data to reference genomes and provides a prediction of the metagenome of a sample. Data in the biom-file was normalized with 16S rRNA gene copy number. Prediction of the functionality of the metagenome of each sample was done by multiplying the normalized abundance of each OTU by each predicted functional feature abundance. A weighted nearest sequenced taxon index (NSTI) was calculated for all samples. The NSTI value describes the average branch length that separates each OTU in the sample from a reference genome, weighted by the abundance of that OTU in the sample. For example, the NSTI value of 0.03 means that the OTUs in the sample are on average 97% similar to the genomes in the

database.

Co-occurrence of OTUs in total and active microbial communities of Outokumpu bedrock fractures was analyzed with the otu.association command in Mothur. Based on pairwise Pearson correlations with significant *p*-value (<0.01), visualization of the co-occurrence network was constructed using Fruchtermann-Feingold layout in the Gephi program (Bastian et al. 2009). The possible "keystone" OTUs were revealed with the betweenness centrality calculation and the connectivity of the network with the closeness centrality estimate (Brandes, 2001). Modular structure of the community was evaluated with the modularity index calculation (Blondel et al. 2008; Lambiotte et al. 2009).

Poistettu: In order to obtain complete metabolic pathway modules from the predicted metagenomes, the KEGG abundance data from the PICRUSt analysis was used as input for HUMAnN v. 0.99 (Abubucker et al., 2012), which was modified to include modules M00597 (Anoxygenic photosystem II), M00598 (Anoxygenic photosystem I), M00595 (Thiosulfate oxidation by SOX complex, thiosulfate => sulfate), K16952 (sulfur oxygenase/reductase), M00596 (Dissimilatory sulfate reduction, sulfate => H_2S), M00567 (Methanogenesis, CO_2 => CH_4), M00528 (Nitrification, ammonia => nitrite), and M00563 (Methanogenesis, methylamine/dimethylamine/trimethylamine => CH_4).

3 Results

3.1 Microbial density in the fracture zones

The total microbial cell numbers were highest in the 180 m fracture (2.97×10^5) and declined according to the depth (Fig. 1)(Table 2). A similar trend was observed with the copy numbers of bacterial 16S rRNA gene ranging from 5.13×10^6 in shallowest fracture to 9.00×10^2 in the deepest fracture. Archaeal 16S rRNA gene copy numbers varied more, but the highest number of archaeal 16S rRNA gene copies was detected from the fracture at 180 m depth (6.24×10^3) (Table 2).

Poistettu:

The number of *dsr*B and *mcr*A gene copies, used as an estimate for the amount of sulphate reducing bacteria and methane producing archaea, respectively, was assessed with quantitative PCR. The copy numbers were quantified also from RNA in order to estimate the activity of sulphate reduction and methanogenesis. The *dsr*B copy numbers varied between 3-6 $\times 10^2$ copies ml^{-1} in most fractures with the exceptions of the 500 m and 967 m fractures where the *dsr*B copy number was 7.4×10^3 copies ml^{-1} and 1.5×10^1 copies ml^{-1}, respectively (Supplementary Table 1). As a proxy of active transcription of *dsr*B genes, the number of mRNA transcripts was also quantified. The highest *dsr*B gene transcription was observed at 1820 m, where the number of *dsr*B transcripts was more than 6.0×10^2 transcripts ml^{-1}. All other fractures had below 1.0×10^2 transcripts ml^{-1}. Methanogenesis marker gene copies were detected only from the upper three fractures (180 m, 500 m and 967 m). The *mcr*A gene copy numbers were just above the detection limit of the assay, i.e. less than 4.0×10^1 copies ml^{-1} in all. *Mcr*A gene transcripts were detected only from the 967 m fracture, where the copy number was just above 1.0×10^2 ml^{-1}.

Poistettu:

3.2 The structure of the microbial communities and correlation to geochemistry

The sequencing data acquired from DNA were used as a representation of the total microbial community present in the fracture fluid samples whereas the microbial communities derived from RNA were used as a proxy of an active community. The microbial communities differed between the nucleic acid fractions as well as the sampling depth (Fig. 2). Pyrosequencing of the total and active bacterial communities based on 16S rRNA gene resulted to 268 identified OTUs representing a total of 157 families in six fractures analysed (Supplementary Table 2). Sulfate reducing communities (based on *dsr*B gene) were successfully sequenced from all fracture fluid samples except the active community in the 2260 m fracture and total and active communities in 2300 m. The archaeal communities were overall less diverse than the bacterial communities. Archaeal sequences (16S rRNA gene) were retrieved from all fractures except the one at 1820 m. Only 17 different OTUs could be divided to 11 families (Supplementary Table 3). Total methanogen communities were detected from 180, 500, 2260 and 2300 m fractures and active methanogen communities from 500 and 967 m fractures.

Primers used in the archaeal nested PCR are reported to have frequent mismatch rate to some archaeal groups, such as SAGMEGs, leading to underestimation of these groups in the community structure (Teske et al. 2008, Wang & Qian 2009). Using the Ribosomal Database Project's probe match platform, we detected that although the overall coverage of the used archaeal primers in this study ranged from 33 to 31 % (first and second PCR, respectively), the coverage of the primers for phylum Thermococci, in which SAGMEG archaea belong to, was much higher, 81-91%.

The 180 m fracture

The bacterial communities in the upmost fracture zone analysed in this study were dominated by OTUs resembling *Comamonadaceae* (Fig. 2). This β-proteobacterial family constituted over 70% of the OTUs in total and active communities in the 180 m fracture (Supplementary Table 2). The estimated richness of the community was 69% of the observed OTUs of the 180 m bacterial communities (Supplementary Table 4a). *Desulfatirhabdum* (54% relative abundance) and *Desulfotomaculum* (98% relative abundance) were most abundant in total and active sulfate reducing communities, respectively (Fig. 2). The archaeal community in the fracture zone at 180 m was dominated by OTUs affiliated with *Methanobacteriaceae* and *Methanoregula*, while *Methanosarcina* and methylotrophic *Methanolobus* OTUs represented

minor groups (Supplementary Table 3). This depth hosted the most diverse archaeal communities (the Shannon diversity index H' 2.4 and 2.1 for total and active archaeal communities, respectively) (Supplementary Table 4b). Similarly, the methanogen community was the most diverse at this fracture, and the dominating groups were similar to unclassified, uncultured methanogen sequences retrieved from wetland soil (LW-25) and acidic peat bog (MB04-15a).

The 500 m fracture

The total bacterial community in the 500 m fracture was dominated by *Comamonadaceae* (70%)(Fig. 2, Supplementary Table 2). The dominating OTUs in active bacterial community affiliated with α-proteobacterial order *Rhodobacterales* with 38% relative abundance, otherwise this community comprised of OTUs affiliating to *Comamonadaceae* (23%), *Dietzia* (23%) and *Pseudomonas* (6%). The amount of observed OTUs captured 77-86% richness of the communities according to the Chao1 estimate at this depth (Supplementary Table 4a). *Desulfotomaculum* and *Pelotomaculum* -affiliating OTUs were the most dominant sulfate reducers in this fracture (Fig. 2). The total and active archaeal communities comprised almost solely of methanogenic *Methanobacteriaceae*, while *Methanobrevibacter* and *Methanosarcina* dominated the communities detected with methanogen-specific marker gene (*mcr*A) (Fig. 2).

The 967 m fracture

The total bacterial community in 967 m fracture zone comprised of *Natrananaerobiales, Clostridiales* and other *Firmicutes* in addition to mollicute *Acholeplasma*. In the active bacterial community, peptococcal *Syntrophobotulus* dominated and otherwise the community resembled the total community (Fig. 2). The observed richness was 84 or 88% of the estimated richness of the total and active communities, respectively (Supplementary Table 4a). Based on the Shannon diversity index H' (2.3) the total archaeal community in the 967 m fracture was among the most diverse of the archaeal communities. It was dominated by OTUs affiliating with SAGMEG-1 Euryarchaeota. In the active archaeal community in this fracture *Methanobacteraceae* dominated and SAGMEG OTUs represented only a minority of the OTUs. *Methanosarcina* dominated the active methanogen community in 967 m fracture (Fig. 2).

The 1820 m fracture

The number of observed bacterial OTUs was among the highest in both total and active bacterial communities in the fracture zone at 1820 m. *Pseudomonadales* (29% relative abundance), *Burkholderiales* (22%) with *Comamonadaceae* and *Oxalobacteriaceae*, *Clostridiales* (13%) comprised mainly of *Dethiosulfatibacter* and other Firmicutes with unresolved phylogeny dominated the total community in this fracture zone. In the active community OTU 86 belonging to Firmicutes was dominant with 39% relative abundance (Fig. 2). The sequenced communities at this depth were estimated to have captured in average 80% of the richness of the total communities (Supplementary Table 4a). The total SRB community in this fracture was entirely composed of *Desulfovibrio* -affiliating OTU. The estimated diversity was low because only 115 sequences were retained. On the other hand, the active SRB community was diverse, with OTUs affiliating with *Desulfatirhabdum*, *Desulfobulbus* and *Desulfoarculus*. Amplification of archaeal and methanogen communities was not succesfull from this fracture indicating low abundance of these groups.

The 2260 m fracture

The fracture at 2260 m hosted a bacterial community mainly comprising of actinobacterial OBP41 class (53% relative abundance) and *Burkholderiales* (34%) (Fig. 2). The active community in this fracture had the highest amount of observed OTUs of the whole dataset and best success in capturing the richness (91%) of the community. OTUs belonging to α-proteobacterial *Bradyrhizobium* (20%) and *Rickettsiales* (11%) in addition to Firmicutes and Actinobacteria dominated this active community. *Desulfotomaculum* and *Desulforudis* -affiliating OTUs dominated the total SRB community at this fracture and *Methanobacterium* dominated both archaeal and methanogen communities (Fig. 2).

The 2300 m fracture

The most frequent OTUs in the bacterial communities in the fracture zone at 2300 m represented *Burkholderiales* (31% of the OTUs) and *Pseudomonadales* (25%) such as *Moraxellaceae* and *Pseudomonadaceae*. In addition, OTUs belonging to other Firmicutes, *Clostridiales*, *Actinomycetales* and *Natranaerobiales* were detected (Fig. 2). The sequenced DNA community covered 86% of the estimated richness. The active community of this fracture–mainly composed of unclassifiable OTUs: only 4% of the community could be determined to more specifically than to phylum level, while half of the community could be determined only to phylum level (Firmicutes) leaving the rest of the community, 46%

unresolved. This reflected also to the richness and coverage indicator values: only 51% of the richness was captured and the coverage was barely half (51%) of the total abundance of the community.

Sulfate reducers were not detected at this depth, and *Methanobacterium* dominated the total archaeal and methanogen communities.

The core microbial community in Outokumpu bedrock fractures

Only a few OTUs that were present in all communities constituted the core community in the Outokumpu deep bedrock. *Pseudomonas* and *Dethiosulfatibacter* in addition to two OTUs with uncertain taxonomic classification (Firmicutes OTU 86 and bacterial OTU1) were detected in all total and active bacterial communities. When observing only the total bacterial communities, most abundant members of the core community were *Comamonadaceae*, *Dethiobacter* and *Pseudomonas*.

The relationship of the microbial community structure to geochemistry

Microbial OTUs formed three loose clusters in canonical correspondence analysis based on the depth where fracture fluid samples were retrieved (Fig. 3). A cluster of bacterial OTUs belonging to orders *Burkholderiales* and *Rhodobacterales* plotted near 180 and 500 m depths and correlated with biotite gneiss and concentration of its major elemental components, iron and magnesium. Microbial OTUs affiliating with the most abundant groups in the 967 m fracture (*Peptococcaceae*, *Anaerobrancaceae*, *Thermoanaerobacterales,* SAGMEG archaea) grouped loosely around the 967 m depth with sulfur concentration pointing to this ordination. The depths of 1820 and 2300 m correlated with sulfide concentration and defining rock types were black schist and pegmatititic granite. Clostridial *Dethiosulfatibacter* and other Firmicutes -affiliating OTUs in addition to several *Burkholderiales* -affiliating OTUs clustered close to these depths.

3.3 The functionality estimation of the microbial communities

The physiology of the members of the microbial communities was estimated from classified OTUs based on the prevalent physiology of the cultured representatives of each OTU at family level according to the Prokaryotes handbook (Rosenberg et al. 2014) (Fig. 4). Bacterial physiotypes with capacity to use versatile metabolic pathways for carbon assimilation and energy production were characteristic in the fracture communities at shallower depths (180 m

and 500 m), while lithotrophic bacterial physiotypes are more frequently detected in 967 and 1820 m fractures. Overall, physiotypes with unknown metabolism became more frequent in the communities at fractures located deeper in the bedrock due to the lack of exact taxonomic classification of the OTUs detected in these fractures (Fig. 4a). In the archael communities the most dominant archaeal physiotype was hydrogenotrophic methanogenesis except in the 967 m fracture (Fig. 4b). The total archaeal community in this fracture was dominated by SAGMEG –affiliating OTUs with undetermined physiology.

Predicted bacterial metagenomes

The metagenomes of the microbial communities of different fracture zones representing six different biotopes were predicted from the 16S rRNA gene sequences, i.e. from the different OTUs with resolved taxonomy. In order to evaluate the accuracy of the prediction of metagenomes, nearest sequenced taxon index (NSTI) was calculated for each sample (Supplementary Table 5). The NSTI's varied between the bacterial communities from 0.07 (the communities in 1820 m and 2300 m fractures) to 0.30 for the community in 2260 m fracture. The archaeal communities represented NSTI's from 0.04 to 0.07 with the exception of the total community in 967 m fracture of which the NSTI was 0.29. Overall, the predicted metagenomes of the total communities did not vary greatly from the active community metagenomes. Top-level functionality estimates revealed differences between bacterial and archaeal communities. The average values for cellular processes and environmental information processing were more abundant in the predicted bacterial metagenomes than in the archaeal metagenomes. In contrast, genetic information processing and unknown features were more abundant in the predicted archaeal metagenomes (Fig. 5).

The most abundant group of bacterial predicted on the basis of PICRUSt analysis were those involved in amino acid metabolism (21-22% of all metabolism genes), carbohydrate metabolism (19-21%) and energy metabolism (11-13% of all genes involved in metabolism) (Table 3a). The predicted bacterial metagenomes differed mostly between 180 m and other fractures. In all fractures, the most abundant amino acid metabolism genes were amino acid related enzymes and arginine and proline metabolism genes (11-15 % and 11-13 % respectively) (Supplementary Table 6a). Gene predictions on carbohydrate utilization revealed a highly similar pattern in all other fracture communities than those at 180 m (Supplementary Table 6a). Amino sugar metabolism genes were more abundant in the 180 m,

Poistettu: In the predicted bacterial metagenome in the180 m fracture, genes involved in phenylalanine, tyrosine, tryptophan and lysine biosynthesis were more abundant than in other fractures. On the other hand, branched-chain amino acid (valine, leucine and isoleucine) degradation represented 7-11% of the predicted amino acid metabolism genes in all other bacterial communities than those at 180 m fracture, where it was approximately only half of this (4-5%).

while in other fractures propanoate and butanoate metabolism genes were dominant. The most abundant energy metabolism genes were involved in oxidative phosphorylation in the predicted metagenomes of the bacterial communities (22-26% in total bacterial communities, and 20-28% in active communities) (Supplementary Table 6a). Genes involved in carbon fixation pathways in prokaryotes were almost as common, in addition to methane metabolism genes.

Predicted archaeal metagenomes

Similar to bacteria, half of the genes in the predicted archaeal metagenomes of each fracture zone were involved in metabolism (Fig. 5b). Energy metabolism genes were most common (18-25%) in addition to amino acid metabolism and carbohydrate metabolism genes (21-25% and 17-20%, respectively) (Table 3b). The predicted metagenome of the total archaeal community in the 967 m fracture zone had the highest abundance of amino acid metabolism genes (25% of all metabolism genes) and the lowest abundance of energy metabolism genes (18%) and thus differed from the community metagenome derived from RNA in the same fracture. Otherwise the predicted metabolism genes were similar in the different archaeal communities in the fracture zones. The predicted metagenome of the total archaeal community of the 967 m fracture differed also in predicted amino acid usage, carbohydrate metabolism and in the energy metabolism gene predictions. As expected, genes representing methane metabolism were the most dominant in all archaeal communities (Supplementary Table 6b).

3.4 Co-occurrence of microbial OTUs in deep crystalline bedrock

From all detected microbial OTUs in deep crystalline fractures, only 15% showed positive correlation ($r > 0.8$) with other members of the microbial communities. Only three significantly negative correlations ($p \leq 0.01$, $r > 0.8$) were detected among the total microbial communities and none in the active communities. The network analysis of the total microbial community divided significantly correlating OTUs into 8 modules with number of nodes ranging from 4 to 41 (Fig. 6). The closeness of centrality (CC) values varied only slightly between most of the OTUs indicating that the network had high connectance between different members (variance 0.5) (Table 4). The "keystone" OTUs were detected based on their above 300 betweenness of centrality (BC) value and these belonged to *Burkholderiaceae*-like OTU, *Desulfitobacter* and *Clostridiaceae*-affiliating OTU, all of which

Poistettu:)(

Poistettu: To evaluate the operational capacity of different metabolic pathways detected with PICRUST, HUMAnN analysis was performed on the predicted metagenomes. Genes needed for several amino acid biosynthesis pathways and transport systems to function were present. The relative abundance of the genes involved in carbon fixation pathways such as Arnon-Buchanan, Wood-Ljungdahl and Calvin cycle were low according to the HUMAnN analysis (Supplementary Table 7). Calvin cycle genes were only present in the predicted metagenomes of the active bacterial communities in 1820 and 2260 m fractures. The coverage of pathways involved in carbohydrate metabolism (KEGG modules 1-4, 6-9,11) was high. ... [1]

Poistettu:)(

Poistettu: 3 b

Poistettu: However, oxidative phosphorylation genes were twice as frequent (16% relative abundance) in the metagenome based on the total community in the 967 m fracture as in all other predicted metagenomes. Additionally, the community in the 180 m fracture differed from the other archaeal communities in predicted energy metabolism: less methane metabolism genes were detected in this community (61% in the total community and 64% in the active community), and genes involved in carbon fixation pathways were more abundant in the community in this fracture (16% in DNA community and 14% in RNA community).

Poistettu: The methanogenesis pathway from CO_2 to CH_4 was present and complete in all archaeal communities according to the HUMAnN analysis (Supplementary Table 8). Methanogenesis pathway from methylamines or methanol was detected and likely operational only in the communities in the 180 m fracture.

had relatively low abundance (0-3%) in the communties, in addition to *Dethiobacter* with 0-5% relative abundance in other total bacterial communities except in the 967 m fracture, in which the relative abundance of this OTU in the total bacterial community was higher, 25% (Fig. 2a and 6). The most connected OTUs in the network belonged to *Rhodococcus* and *Herbaspirillum* (48 connections each) in addition to OTUs resembling *Renibacterium*, *Gemellaceae*, *Trabulsiella* and *Novosphingobium* (46-47 connections each). The positively correlating OTUs of the active microbial community divided into 8 clusters with number of nodes ranging from 2 to 64 (Fig. 7). The active community network had also small variation in the CC values (variance 0.3). The "keystone" genera of the active microbial community network were *Comamonas, Curvibacter* and *Sphingomonas*, with BC values above 470 each. *Comamonadaceae* –affiliating OTU was determined to be part of the core community in this Fennoscandian deep subsurface site, as it was frequently found in all depths with relatively high abundance ranging from 7-72% of the total community. *Curvibacter* and *Sphingomonas* were both present in the active communities, but with a very low relative abundance.

4 Discussion

The bacterial community structure in the groundwater of Outokumpu fracture zones varies between the different depths. In addition, the structure of the total and active communities differs within the fracture zones. The core bacterial community of deep crystalline bedrock fracture fluids in Outokumpu was composed of few OTUs found from all microbial communities in the fractures. Most abundant of these were *Comamonadaceae*-, *Firmicutes*-, *Anaerobrancaceae*- and *Pseudomonadaceae* -affiliating OTUs. However, the majority of the bacterial OTUs discovered in this study could be regarded as members of the so-called "rare biosphere" with their relatively low abundance and uneven distribution throughout the fracture zones (Sogin et al. 2006).

A keystone species has greater impact on its community or living environment than would be expected from its relative abundance or total biomass (Paine 1995). Steele et al. (2011) suggested that the highly connected nodes in co-occurrence network would be analogous to microbial "keystone species". The most connected nodes of both the active and the total microbial community were representatives of *Burkholderiales* (e.g. *Comamonas*, *Curvibacter*, *Oxalobacter*, *Herbaspirillum*, *Pelomonas*, *Cupriavidus*), thus representing the possible "keystone species" of deep bedrock fracture fluid communities in Outokumpu. In addition,

Poistettu: Several keystone genera

Poistettu:).

some clostridial phylotypes were among the highly connected nodes. Several of these "keystone" genera detected with the co-occurrence networks were members of the rare biosphere, thus providing further evidence for the significance of the less abundant microorganisms for the whole community (Sogin et al. 2006, Brown et al. 2009). Hence, we propose that these have a major role in the functionality of the network. The species belonging to *Burkholderiales* are most likely involved in hydrogen oxidation in the deep subsurface and thus are the potential primary producers in this oligotrophic environment. Clostridial species on the other hand could be occupying the niche for heterotrophic carbon assimilation, possibly providing H_2 produced in fermentation processes. *Burkholderiales* and clostridial species have been found from other serpentinite-hosted deep subsurface aquifers (Brazelton et al. 2012, Tiago et al. 2012). These studies have suggested that *Burkholderiales* and clostridial species play a significant role in carbon and energy exchange between one another and other bacterial groups in these environments. The metabolic flexibility of *Burkholderiales* species, for example the ability to use both autotrophic and heterotrophic carbon fixation mechanisms is beneficial in isolated deep biosphere environments where concentrations of different carbon substrates fluctuate over time (Moser et al. 2005, Magnabosco et al. 2015).

Poistettu: keystone genera.

One of the most connected nodes in active microbial community was *Staphylococcus*. We assume that this can be explained by contamination. Staphylococci are common in human body, and while the amount of sequences affiliating with Staphylococcus were low, they were discovered in all samples. However, we can't ascertain the physiology of an OTU from it's taxonomic affiliation and these Staphylococcus-like OTUs might be true members of the deep biosphere.

Serpentinization as a source for energy in Outokumpu

Hydrogen oxidizing, facultatively chemolithotrophic members of the *Comamonadaceae*-family were dominating the 180 m and 500 m fracture zone communities. These microbes are commonly found from hydrogen-enriched subsurface environments. These include findings of *Ralstonia* and *Hydrogenophaga* in Lost City Hydrothermal Field and Tablelands Ophiolite serpentinite springs, respectively (Brazelton et al. 2012, 2013). Likewise, *Comamonadaceae*-affiliating and clostridial sequences formed the majority of the bacterial community in serpentinization-driven subsurface aquifer in Portugal (Tiago and Veríssimo 2013).

Comamonadaceae were also dominant in the drill hole water communities in Outokumpu at shallow depths (0-100 m) (Itävaara et al. 2011), and at 200 and 2300 m depths (Nyyssönen et al. 2014). Additionally, *Comamonadaceae* were detected from other depths in the drill hole water column, such as at 1100-1500 m depth that is characterized by ophiolithic rock sequence (Nyyssönen et al. 2014). However, Kietäväinen et al. (2013) detected substantial amounts of H_2 only in fractures below 1500 m in Outokumpu, which might indicate that the source for H_2 for abundant hydrogen-oxidizers is in the bedrock aquifer that 180 and/or 500 m fractures intersect or lead into. The seismic reflectors of Outokumpu bedrock demonstrate ophiolite-derived rock types in the vicinity of the drill hole, some of which are located also at shallow depths (Kukkonen et al. 2011). Thus, we assume that these may affect the two shallowest fracture fluids and explain the similarity of the microbial results with other ophiolithic, serpentinizing environments.

The bacterial community of the 967 m fracture zone also resembles those of serpentinizing environments. One major member of the total bacterial community at 967 m fracture was hydrogen-oxidizing *Dethiobacter* that has previously been detected also from groundwaters associated with ophiolithic rock sequence with active serpentinization processes (Tiago and Veríssimo 2013). Brazelton et al. (2013) detected acetyl-CoA synthase gene affiliating with *Dethiobacter* from bacterial shotgun-sequenced metagenomes from bacterial Winter House Canyon (WHCB) samples from Tablelands Ophiolite. Moreover, similarities between serpentinizing environments and the deep biosphere of Outokumpu bedrock include the detection of abundant clostridial phylotypes such as *Anaerobrancaceae* from the deeper parts of the bedrock (Itävaara et al. 2011, Brazelton et al. 2013, Purkamo et al. 2013, Nyyssönen et al. 2014). The "keystone species" of the total and active microbial communities detected in this study reflect the similarity between the serpentinizing environments and Outokumpu deep biosphere with *Comamonadaceae*, *Burkholderiaceae*, *Clostridiaceae* and *Dethiobacter* as the recognized "keystone" families.

Overall, the spatial distribution of *Burkholderiales* at shallower depths and *Clostridiales* in the fractures located deeper in the bedrock can be explained to some level with the availability of electron acceptors. Both of these groups are able to use H_2 as electron donor, but *Comamonadaceae* are mainly aerobic chemoorganotrophs using a vide variety of different organic carbon compounds for energy and using oxygen as terminal electron acceptor (Willems et al. 1991), while *Dethiobacter* is a strict anaerobe that reduces sulphur compounds

but not sulphate (Sorokin et al. 2008). However, as their co-occurrence in the 1820 m fracture suggests, these organisms can prevail in same environment. In Outokumpu, low concentrations of oxygen were measured during the pumping of the fracture fluids (< 0.1 mg l^{-1}) (Purkamo et al. 2013). While oxygen can be introduced in water due to the atmospheric contamination, small amounts of oxygen can be produced in radiolysis of water in bedrock (Pedersen 1997, Lin et al. 2005). This might explain the detection of aerobic bacteria such as *Comamonadaceae* in the deep fluids of Outokumpu.

Higher hydrogen concentration in the two deepest fracture fluids (Kietäväinen et al. 2013) could indicate that something else than the electron donor is limiting the communities at these depths. Abundant bacterial groups of the communities in the 2260 m fracture belonged to Actinobacteria and α-proteobacteria. Little is known about metabolic capacities of the OPB41 candidate phylum, but as an actinobacterial phylotype, they may be chemoorganotrophs, while *Bradyrhizobiaceae* are mixotrophs with the capacity to oxidize H_2. In this sample, 4.5% of the sequences affiliated with Streptophyta, representing likely plant material. Kietäväinen et al. (2013) reported that 2260 m depth contained up to 30 % of residual drilling fluids, thus offering a possible explanation for these sequences. Source of this material can be peat that was used in the drilling mud as stabilizer during the drilling of shallower depths (<1000 m) (Ahonen et al. 2011). The drilling fluids are circulated in the system throughout the drilling operation, thus a small fraction of this peat material may be present still at this depth.

The fractures located below 2 km depth, mostly dominated by unclassifiable phylotypes, might suggest that these depths harbour life that differs considerably from the currently known microorganisms.

Comparison of the microbial community structure and functionality between different Precambrian deep subsurface sites

Bacterial and archaeal communities from the Outokumpu fracture zones resemble the drill hole water communities described in a previous study (Nyyssönen et al. 2014). This is not surprising, as drill hole fluids are a mixture of the different fracture fluids emanating to the drill hole from the fractures at different depths of the bedrock. *Comamonadaceae* form a major part of the bacterial community at most depths in the drill hole, as they are abundant in 180, 500, 2260 and 2300 m fracture communities. At 967 and 1820 m fracture communities,

phylotypes affiliating with *Clostridiales* are dominating, while clostridial phylotypes represent a major fraction of the drill hole communities at 1000-1500 m. Many sulfate reducer phylotypes detected in this study were similar to those detected from the drill hole fluids (Itävaara et al. 2011, Purkamo et al. 2015) and from fracture zones (Purkamo et al. 2013) with DGGE. These included *Desulfotomaculum* and *Desulfovibrio*. Archaeal communities in the fractures and in the drill hole are mainly dominated by *Methanobacterium*, but SAGMEG archaea are also abundant in the drill hole communities at 1000 m and above (Nyyssönen et al. 2014), possibly originating from the 967 m fracture where SAGMEG archaea were dominating the total archaeal community.

Members of *Comamonadaceae* and *Pseudomonadaceae* belong to the core microbial community in Outokumpu. These were also detected from all studied microbial communities in another Fennoscandian crystalline bedrock environment in Olkiluoto fracture zones, (Bomberg et al. 2015a). To emphasize the importance of these microbes to the total communities in deep crystalline bedrock environments, members of *Comamonadaceae* were recognized as the "keystone genera" of the active microbial community in the Outokumpu deep biosphere.

The microbial communities in Outokumpu deep crystalline bedrock share common features with those in the deep ecosystems in Witwatersrand Basin, South Africa. Clones affiliating with *Comamonadaceae* have been found from a deep drill hole outlet in Driefontein gold mine in South Africa. In addition, clostridial sulphate reducers, such as *Desulfotomaculum,* are dominating the SRB communities in Outokumpu as well as the deep borehole communities in Driefontein (Baker et al. 2003, Moser et al. 2003, 2005, Silver et al. 2010). Candidatus *Desulforudis audaxviator* was a minor component (with 1-35% relative abundance) of the SRB communities in Outokumpu fractures at depths of 180-2260 m, in contrast to a microbial community where *D. audaxviator* formed a single-species ecosystem in deep bedrock fracture in Mponeng mine (Chivian et al. 2008).

The detection of a SAGMEG-dominated archaeal community in the 967 m fracture reinforces the results from previous study of Outokumpu deep fracture zone microbial communities (Purkamo et al. 2013). These archaea were first discovered from gold mines in South Africa (Takai et al. 2001) and recently named as Hadesarchaea (Baker et al. 2016). The first clues about carbon metabolism of these archaea came from a study of deeply buried ocean

sediment, where SAGMEG archaea were part of the mainly heterotrophic archaeal community (Biddle et al. 2006). A recent study of four genomic bins of Hadesarchaea revealed the metabolic diversity of this new archaeal lineage. Although they share several characteristics with other Euryarchaeota, they lack the marker genes for methanogenesis (Baker et al. 2016). Correspondingly, the predicted metagenome of the total archaeal community at 967 m with 94% of Hadesarchaea in our study demonstrated significantly lower amount of methane metabolism genes than the other fracture fluid communities. The most propable energy harvesting method for these archaea is thought to be the oxidation of carbon monoxide coupled with H_2O reduction (Baker et al. 2016). Thus, we assume that in Outokumpu subsurface, these archaea may fill the niche of carbon monoxide utilizers and produce H_2 for hydrogenotrophic part of the microbial community.

Furthermore, this study supports the biogeographical trend that methanogens in different deep Precambrian sites are similarly distributed within depth (Kietäväinen and Purkamo 2015). Methanogens with wider substrate range were found in the fractures located at shallower depths both in Outokumpu and several deep subsurface sites in South Africa. On the other hand, obligately hydrogenotrophic methanogens were detected in the fracture zones located deeper (Moser et al. 2005, Gihring et al. 2006, Lin et al. 2006). Archaeal communities represeted much less diversity, and interestingly, while almost all fracture zones were dominated by methanogenic archaea, the archaeal fracture community in 180 m fracture was the most diverse, where for example archaea belonging to Miscellaneous Crenarchaeal Group (MCG) and Terrestrial Miscellaneous Euryarchaeal Group (TMEG) were detected. In Olkiluoto, the highest archaeal species richness was detected from a fracture at a depth of 296 m (Bomberg et al. 2015a). In another study of archaeal communities in Olkiluoto, some indication of correlation between increasing depth and decreasing diversity could be detected (Bomberg et al. 2015b). Low archaeal diversity may be due to the coverage of the primer pairs that were used in the nested PCR approach as well as the two-step PCR itself.

Many of the abundant bacterial groups in Outokumpu bedrock are organotrophic with capacity to use a wide range of substrates for biosynthesis and either fermentation or anaerobic respiration for energy conservation. Hence, depending on the available sources of energy and carbon, these organisms can switch to the best energy mechanism currently available. With the low relative abundance of the genes involved in autotrophic carbon fixation pathways in the predicted metagenomes, we propose that in Outokumpu,

Poistettu: The archaeal community in the 967 m fracture was dominated by SAGMEG archaea that were first discovered from gold mines in South Africa (Takai et al. 2001). The predicted metagenome of the archaeal community at this depth showed notably higher amount of genes involved in oxidative phosphorylation than other fracture communities. Whether this is a bias induced by the absence of complete genomes of closely related species in the database or an indication of oxidative phosphorylation genes having a specific role in SAGMEGs remains to be resolved.

heterotrophic carbon metabolism is important also for the intrinsic fracture zone communities as it is for the drill hole water column communities (Purkamo et al. 2015). Archaeal communities in Outokumpu fracture zones are mainly methanogenic, using most likely the hydrogenotrophic methanogenesis pathway from CO_2 to CH_4.

5 Conclusions

The microbial communities of Outokumpu Precambrian crystalline bedrock fractures share features with serpenization-driven microbial communities in alkaline springs and subsurface aquifers. These include members belonging to *Burkholderiales* and Clostridia. Additionally, these phylotypes were regarded as the "keystone species" in Outokumpu deep biosphere. Additionally, *Comamonadaceae* are part of the core microbial community in Fennoscandian crystalline bedrock environments. Sulfate reducing microbes and methanogens are present, but they represent marginal groups of the microbial communities. The dominating taxa of the sulphate reducing communities observed in this study are similar to the ones detected from the deep subsurface of Driefontein mine in South Africa. Similarly, the total archaeal community at 967 m fracture is dominated by SAGMEG archaea, initially described at deep gold mines of South Africa. Otherwise hydrogenotrophic methanogens, mainly *Methanobacterium*, dominate the archaeal communities. The fracture zone at 180 m in Outokumpu hosts the most diverse archaeal communities. Many probable "keystone species" of Outokumpu deep biosphere belong to rare biosphere, with low abundance but a wide range of carbon substrates and a capacity for H_2 oxidation. Metagenomic predictions of the bacterial communities revealed that heterotrophy is also important in the deep fracture zones in Outokumpu.

Author contributions

LP planned and conducted the research, analysed the data and is the corresponding author. MB aided in planning, sampling, analysis of the data and writing the manuscript. RK took part in sampling, provided hydrogeochemistry data and participated in writing the manuscript. HS assisted in analysing the data and writing the manuscript. MN took part in sampling and writing the manuscript. MNP assisted in sampling and aided in laboratory research, LA assisted in planning and executing the sampling and hydrogeochemistry analysis and writing the manuscript. IK provided access to Outokumpu Deep Drill Hole, supported sampling,

writing the manuscript and provided funding and MI took part in sampling and writing the manucript and provided funding.

Acknowledgements

Mirva Pyrhönen is acknowledged for the excellent laboratory skills and help in preparing the sampling trips. Arto Pullinen, Leea Ojala, Pauliina Rajala and Sirpa Jylhä are acknowledged for the professional sampling assistance at the Outokumpu field site. This study was funded by Kone Foundation, The Academy of Finland (DEEP LIFE project, grant number 133348/2009), The Foundation for Research of Finnish Natural Resources (grant number 1718/09) and Finnish research program on nuclear waste management (KYT, grants GEOMOL, KABIO, SALAMI and RENGAS).

References

Ahonen, L., Kietäväinen, R., Kortelainen, N., Kukkonen, I. T., Pullinen, A., Toppi, T., Bomberg, M., Itävaara, M., Nousiainen, A., Nyyssönen, M. and Öster, M.: Hydrogeological characteristics of the Outokumpu Deep Drill Hole. Edited by Ilmo T. Kukkonen, Geological Survey of Finland, Special Paper 51, 151–168, 2011

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J.: Basic local alignment search tool. J. Mol. Biol. 215:403-410, 1990

Abubucker, S., Segata, N., Goll, J., Schubert, A. M., Izard, J., Cantarel, B. L., Rodriguez-Mueller, B., Zucker, J., Thiagarajan, M. and Henrissat, B.: Metabolic reconstruction for metagenomic data and its application to the human microbiome, PLoS Comput Biol, 8, e1002358, 2012.

Amend, J. P. and Teske, A.: Expanding frontiers in deep subsurface microbiology, Palaeogeogr. , Palaeoclimatol. , Palaeoecol., 219, 131-155, 2005.

Baker, B. J., Moser, D. P., MacGregor, B. J., Fishbain, S., Wagner, M., Fry, N. K., Jackson, B., Speolstra, N., Loos, S. and Takai, K.: Related assemblages of sulphate-reducing bacteria associated with ultradeep gold mines of South Africa and deep basalt aquifers of Washington State, Environ. Microbiol., 5, 267, 2003.

Baker, B. J., Saw, J. H., Lind, A. E., Lazar C. S., Hinrichs, K-U., Teske, A. P., Ettema, T. J. G.: Genomic inference of the metabolism of cosmopolitan subsurface Archaea, Hadesarchaea, Nature Microbiol, 16002, 2016.

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 a Yellowstone National Park hot spring environment, Proc. Natl. Acad. Sci. U. S. A., 91, 1609-1613, 1994.

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

Biddle, J. F., Lipp, J. S., Lever, M. A, Lloyd, K. G., Sørensen, K. B., Anderson, R., Fredricks, H. F., Elvert, M., Kelly, T. J., Schrag, D. P., Sogin, M.L., Brenchley, J. E., Teske, A., House, C. H., Hinrichs, K-U.: Heterotrophic Archaea dominate sedimentary subsurface ecosystems off Peru, Proc. Natl. Acad. Sci. U S A., 10, 3846-3851, 2006.

Biddle, J. F., Fitz-Gibbon, S., Schuster, S. C., Brenchley, J. E. and House, C. H.: Metagenomic signatures of the Peru Margin subseafloor biosphere show a genetically distinct environment, Proceedings of the National Academy of Sciences, 105, 10583, 2008.

Biddle, J. F., White, J. R., Teske, A. P. and House, C. H.: Metagenomics of the subsurface Brazos-Trinity Basin (IODP site 1320): comparison with other sediment and pyrosequenced metagenomes, The ISME Journal, 2011.

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.

Blondel V.D., Guillaume, J-L, Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks, Journal of Statistical Mechanics: Theory and Experiment 2008, 1000

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 Journal of Ecology, 2014, 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 Research International, 2015, 2015a.

Bomberg, M., Lamminmäki, T., and Itävaara, M.: Estimation of microbial metabolism and co-occurrence patterns in fracture groundwaters of deep crystalline bedrock at Olkiluoto, Finland. Biogeosciences, 2015, 2015b

Brandes, U.: A Faster Algorithm for Betweenness Centrality, Journal of Mathematical Sociology 25(2):163-177, 2001

Brazelton, W. J., Morrill, P. L., Szponar, N. and Schrenk, M. O.: Bacterial communities associated with subsurface geochemical processes in continental serpentinite springs, Appl. Environ. Microbiol., 79, 3906-3916, 10.1128/AEM.00330-13; 10.1128/AEM.00330-13, 2013.

Brazelton, W. J., Nelson, B. and Schrenk, M. O.: Metagenomic evidence for H_2 oxidation and H_2 production by serpentinite-hosted subsurface microbial communities, Front. Microbiol., 2, 268, 10.3389/fmicb.2011.00268 [doi], 2012.

Brown, M. V., Philip, G. K., Bunge, J. A., Smith, M. C., Bissett, A., Lauro, F. M., Fuhrman, J. A., and Donachie, S. P.: Microbial community structure in the North Pacific ocean, *ISME J.* 3, 1374–1386. doi: 10.1038/ismej.2009.86, 2009

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Pena, A. G., Goodrich, J. K. and Gordon, J. I.: QIIME allows analysis of high-throughput community sequencing data, Nature methods, 7, 335-336, 2010.

Chivian, D., Brodie, E. L., Alm, E. J., Culley, D. E., Dehal, P. S., DeSantis, T. Z., Gihring, T. M., Lapidus, A., Lin, L. H., Lowry, S. R., Moser, D. P., Richardson, P. M., Southam, G., Wanger, G., Pratt, L. M., Andersen, G. L., Hazen, T. C., Brockman, F. J., Arkin, A. P. and Onstott, T. C.: Environmental genomics reveals a single-species ecosystem deep within Earth, Science, 322, 275-278, 10.1126/science.1155495 [doi], 2008.

Edwars, U., Rogall, T., Blocker, H., Emde, M., and E. C. Bottger, E.C.: Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA, Nucleic Acids Research, 17, 7843–7853, 1989.

Etiope, G. and Sherwood Lollar, B.: Abiotic methane on Earth, Rev. Geophys., 51, 276-299, 2013.

Fish, J. A., B. Chai, Q. Wang, Y. Sun, C. T. Brown, J. M. Tiedje, and J. R. Cole.: FunGene: the Functional Gene Pipeline and Repository. Front. Microbiol. **4**:291; doi: 10.3389/fmicb.2013.00291, 2013.

1 Geets, J., Borremans, B., Diels, L., Springael, D., Vangronsveld, J., van der Lelie, D. and
2 Vanbroekhoven, K.: *Dsr*B gene-based DGGE for community and diversity surveys of sulfate-
3 reducing bacteria, J. Microbiol. Methods, 66, 194-205, 2006.

4 Giardine, B., Riemer, C., Hardison, R.C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y.,
5 Blankenberg, D., Albert, I., Taylor, J., Miller, W., Kent, W.J., Nekrutenko, A.: Galaxy: a
6 platform for interactive large-scale genome analysis. Genome Res., 15, 1451-5, 2005.

7 Gihring, T., Moser, D., Lin, L. H., Davidson, M., Onstott, T., Morgan, L., Milleson, M., Kieft,
8 T., Trimarco, E. and Balkwill, D.: The distribution of microbial taxa in the subsurface water
9 of the Kalahari Shield, South Africa, Geomicrobiol. J., 23, 415-430, 2006.

10 Goecks, J., Nekrutenko, A., Taylor, J. and The Galaxy Team: Galaxy: a comprehensive
11 approach for supporting accessible, reproducible, and transparent computational research in
12 the life sciences. Genome Biol. 11, R86, 2010.

13 Gold, T.: The deep, hot biosphere, Proc. Natl. Acad. Sci. U. S. A., 89, 6045-6049, 1992.

14 Grosskopf, R., Janssen, P. H. and Liesack, W.: Diversity and structure of the methanogenic
15 community in anoxic rice paddy soil microcosms as examined by cultivation and direct 16S
16 rRNA gene sequence retrieval, Appl. Environ. Microbiol., 64, 960-969, 1998.

17 Hales, B. A., Edwards, C., Ritchie, D. A., Hall, G., Pickup, R. W. and Saunders, J. R.:
18 Isolation and identification of methanogen-specific DNA from blanket bog peat by PCR
19 amplification and sequence analysis, Appl. Environ. Microbiol., 62, 668-675, 1996.

20 Hallbeck, L. and Pedersen, K.: Characterization of microbial processes in deep aquifers of the
21 Fennoscandian Shield, Appl. Geochem., 23, 1796-1819, 2008.

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

Hammer, Ø., Harper, D. and Ryan, P.: Past: Paleontological Statistics Software Package for education and data analysis. Paleontología Electrónica 4: 1-9, URL: http://palaeo-electronica.org/2001_1/past/issue1_01.html, 2001.

Itävaara, M., Nyyssönen, M., Kapanen, A., Nousiainen, A., Ahonen, L. and Kukkonen, I.: Characterization of bacterial diversity to a depth of 1500 m in the Outokumpu deep borehole, Fennoscandian Shield, FEMS Microbiol. Ecol., 77, 295-309, 2011.

Kietäväinen, R., Ahonen, L., Kukkonen, I. T., Hendriksson, N., Nyyssönen, M. and Itävaara, M.: Characterisation and isotopic evolution of saline waters of the Outokumpu Deep Drill Hole, Finland–Implications for water origin and deep terrestrial biosphere, Appl. Geochem., 32, 37-51, 2013.

Kietäväinen, R. and Purkamo, L.: The origin, source and cycling of methane in deep crystalline rock biosphere, Frontiers in Microbiology, 6, 725, 2015.

Kukkonen, I. T., Rath, V., Kivekäs, L., Šafanda, J. and Čermák, V.: Geothermal studies of the Outokumpu deep drill hole, Outokumpu Deep Drilling Project 2003–2010. Edited by Ilmo T. Kukkonen, Geological Survey of Finland, Special Paper 51, 181–198, 2011

Lambiotte, R., Delvenne, J-C., Barahona, M.: Laplacian Dynamics and Multiscale Modular Structure, Networks 2009.

Langille, M. G., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., Clemente, J. C., Burkepile, D. E., Thurber, R. L. V. and Knight, R.: Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences, Nat. Biotechnol., 31, 814-821, 2013.

Lau, M. C., Cameron, C., Magnabosco, C., Brown, C. T., Schilkey, F., Grim, S., Hendrickson, S., Pullin, M., Lollar, B. S. and van Heerden, E.: Phylogeny and phylogeography of functional genes shared among seven terrestrial subsurface metagenomes reveal N-cycling and microbial evolutionary relationships, Frontiers in microbiology, 5, 2014.

Lin, L., Hall, J., Lippmann-Pipke, J., Ward, J. A., Sherwood Lollar, B., DeFlaun, M., Rothmel, R., Moser, D., Gihring, T. M. and Mislowack, B.: Radiolytic H_2 in continental crust:

nuclear power for deep subsurface microbial communities, Geochem. Geophys. Geosyst., 6, 2005.

Lin, L., Hall, J., Onstott, T., Gihring, T., Lollar, B. S., Boice, E., Pratt, L., Lippmann-Pipke, J. and Bellamy, R. E.: Planktonic microbial communities associated with fracture-derived groundwater in a deep gold mine of South Africa, Geomicrobiol. J., 23, 475-497, 2006.

Lupatini, M., Suleiman, A. K., Jacques, R. J., Antoniolli, Z. I., de Siqueira Ferreira, A., Kuramae, E. E. and Roesch, L. F.: Network topology reveals high connectance levels and few key microbial genera within soils, Frontiers in Environmental Science, 2, 10, 2014.

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

McCollom, T. M.: Laboratory simulations of abiotic hydrocarbon formation in Earth's deep subsurface. Rev. Mineral. Geochem. 75, 467–494. doi: 10.2138/rmg.2013.75.15, 2013.

McCollom, T. M., Lollar, B. S., Lacrampe-Couloume, G. and Seewald, J. S.: The influence of carbon source on abiotic organic synthesis and carbon isotope fractionation under hydrothermal conditions, Geochim. Cosmochim. Acta, 74, 2717-2740, 2010.

Moser, D., Onstott, T., Fredrickson, J., Brockman, F., Balkwill, D., Drake, G., Pfiffner, S., White, D., Takai, K. and Pratt, L.: Temporal shifts in microbial community structure and geochemistry of an ultradeep South African gold mine borehole, Geomicrobiol. J., 20, 517-548, 2003.

Moser, D. P., Gihring, T. M., Brockman, F. J., Fredrickson, J. K., Balkwill, D. L., Dollhopf, M. E., Lollar, B. S., Pratt, L. M., Boice, E., Southam, G., Wanger, G., Baker, B. J., Pfiffner, S. M., Lin, L. H. and Onstott, T. C.: Desulfotomaculum and Methanobacterium spp. dominate a 4- to 5-kilometer-deep fault, Appl. Environ. Microbiol., 71, 8773-8783, 71/12/8773 [pii], 2005.

Mu, A., Boreham, C., Leong, H. X., Haese, R. R. and Moreau, J. W.: Changes in the deep subsurface microbial biosphere resulting from a field-scale CO_2 geosequestration experiment, Frontiers in microbiology, 5, 2014.

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., Paulin, L., Laine, P., Itävaara, M. and Auvinen, P.: Taxonomically and functionally diverse microbial communities in deep crystalline rocks of the Fennoscandian shield, The ISME journal, 2013.

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, Frontiers in microbiology, 5, 2014.

Paine, R.T.: A Conversation on Refining the Concept of Keystone Species, Conservation Biology 9, 962–964, 1995.

Pedersen, K.: Exploration of deep intraterrestrial microbial life: current perspectives, FEMS Microbiol. Lett., 185, 9-16, 2000.

Pedersen, K.: Microbial life in deep granitic rock, FEMS Microbiol. Rev., 20, 399-414, 1997.

Pedersen, K.: Investigations of subterranean bacteria in deep crystalline bedrock and their importance for the disposal of nuclear waste, Can. J. Microbiol., 42, 382-391, 1996.

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, The ISME journal, 2, 760-775, 2008.

Power, M. E., Tilman D., Estes, J. A., Menge, B. A., Bond, W. J., Mills, L. S., Gretchen, D., Castilla, J. C., Lubchenco, J., and Paine, R. T.: Challenges in the quest for keystones, BioScience, 4, 609–620, 1996.

Proskurowski, G., Lilley, M. D., Seewald, J. S., Fruh-Green, G. L., Olson, E. J., Lupton, J. E., Sylva, S. P. and Kelley, D. S.: Abiogenic hydrocarbon production at lost city hydrothermal field, Science, 319, 604-607, 10.1126/science.1151194; 10.1126/science.1151194, 2008.

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., 1-14, 2014.

Purkamo, L., Bomberg, M., Nyyssönen, M., Kukkonen, I., Ahonen, L., Kietäväinen, R. and Itävaara, M.: Dissecting the deep biosphere: retrieving authentic microbial communities from packer-isolated deep crystalline bedrock fracture zones, FEMS Microbiol. Ecol., 85, 324-337, 2013.

Quince, C., Lanzén, A., Curtis, T. P., Davenport, R. J., Hall, N., Head, I. M., Read, L. F. and Sloan, W. T.: Accurate determination of microbial diversity from 454 pyrosequencing data, Nature methods, 6, 639-641, 2009.

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. and Robinson, C. J.: Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities, Appl. Environ. Microbiol., 75, 7537-7541, 2009.

Silver, B., Onstott, T., Rose, G., Lin, L. H., Ralston, C., Sherwood-Lollar, B., Pfiffner, S., Kieft, T. and McCuddy, S.: In situ cultivation of subsurface microorganisms in a deep mafic sill: implications for SLiMEs, Geomicrobiol. J., 27, 329-348, 2010.

Sogin, M. L., Morrison, H. G., Huber, J. A., Mark Welch, D., 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. U. S. A., 103, 12115-12120, 0605127103 [pii], 2006.

Stahl, D.: Development and application of nucleic acid probes, Nucleic acid techniques in bacterial systematics, 1991.

Steele, J.A., Countway, P.D., Xia, L., Vigil, P.D., Beman, J.M., Kim, D.Y., Chow, C.E., Sachdeva, R., Jones, A.C., Schwalbach M.S., Rose, J.M., Hewson, I., Patel, A., Sun, F., Caron, D.A. and Fuhrman, J.A.: Marine bacterial, archaeal and protistan association networks reveal ecological linkages. ISME J., 9, 1414-1425, 2011.

Szponar, N., Brazelton, W. J., Schrenk, M. O., Bower, D. M., Steele, A. and Morrill, P. L.: Geochemistry of a continental site of serpentinization, the Tablelands Ophiolite, Gros Morne National Park: a Mars analogue, Icarus, 224, 286-296, 2013.

Teske, A., Sørensen, K.,B.: Uncultured archaea in deep marine subsurface sediments: have we caught them all? ISME J., 2, 3-18, 2008

Tiago, I. and Veríssimo, A.: Microbial and functional diversity of a subterrestrial high pH groundwater associated to serpentinization, Environ. Microbiol., 15, 1687-1706, 2013.

Wang, Y., & Qian, P.-Y.: Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. PLoS ONE, 4, e7401, 2009

Wagner, M., Roger, A. J., Flax, J. L., Brusseau, G. A. and Stahl, D. A.: Phylogeny of dissimilatory sulfite reductases supports an early origin of sulfate respiration, J. Bacteriol., 180, 2975-2982, 1998.

Willems, A., De Ley, J., Gillis, M. and Kersters, K.: Comamonadaceae, a New Family Encompassing the Acidovorans rRNA Complex, Including Variovorax paradoxus gen. nov. , comb. nov. for Alcaligenes paradoxus (Davis 1969), Int. J. Syst. Bact, 445-450, 1991.

Zhou, J., Deng, Y., Luo, F., He, Z. and Yang, Y.: Phylogenetic molecular ecological network of soil microbial communities in response to elevated CO_2, MBio, 2, 10.1128/mBio.00122-11. Print 2011, 10.1128/mBio.00122-11 [doi], 2011.

Table 1. Hydrogeochemical characteristics of six fracture zones of Outokumpu.

Concentrations for cations and anions are given in mg L^{-1}, EC in mS m^{-1} and alkalinity in mmol L^{-1}.

Depth m	Prevalent rock type	Ca	Fe	Mg	Na	S	Br	Cl	SO_4	NO_3	Sulfide	TOC	DOC	TIC	DIC	pH	EC[1]	Alkalinity
180	Mica schist, biotite gneiss	1060	0.34	16.7	1070	1.27	23	3280	1.5	<20	0.057	12.8	9	0.7	0.6	7.4	1060	0.31
500	Chlorite-sericite schist	2250	<0.03	12.9	1810	3.49	<50	8180	1.0	<100	b.d	b.d	b.d	b.d	b.d	8.3	1900	0.19
967	Mica schist, chlorite-sericite schist	2000	<0.03	0.8	1770	17.1	62.2	5790	0.6	<40	b.d	6.93	6.4	<0.2	<0.2	8.9	1740	0.29
1820	Mica and black schist, granite	11800	0.03	15.1	3820	44.4	159	30300	2.6	<200	0.87	30.33	29.7	0.4	0.51	9.0	6930	0.37
2260	Biotite gneiss	8130	0.03	21	2630	4.8	<1000	16400	<2	<2000	b.d	b.d	b.d	b.d	b.d	8.2	4890	0.25
2300	Mica schist, granite	9480	<0.03	18.7	3120	7.42	123	24500	<2	<200	0.086	34.33	34	0.4	<0.2	8.6	4370	0.29

[1]Electrical conductivity at 25°C
b.d = below detection limit

Poistettu: ml

Poistettu: ml

Table 2. The total number of cells and the 16S rRNA gene copy numbers of microbial communities in six fractures of Outokumpu. Values are given in mL^{-1}.

Fracture depth	Total cell amount		Bacterial 16S rRNA gene		Archaeal 16S rRNA gene	
m	cell number	SEM[1]	copy number	SEM[1]	copy number	SEM[1]
180	2,97E+05	6,25E+04	5,13E+06	1,49E+05	6,24E+03	1,25E+00
500	5,72E+04	3,04E+03	1,88E+06	2,99E+05	8,62E+01	1,23E+00
967	1,00E+04	8,91E+02	1,26E+05	2,15E+04	4,90E+02	1,24E+00
1820	4,74E+03	1,17E+03	9,05E+02	2,29E+01	b.d	n.a
2260	1,51E+03	3,52E+02	9,01E+02	3,72E+01	2,32E+01	1,07E+00
2300	6,30E+03	1,89E+03	9,00E+02	2,07E+01	b.d	n.a

[1] SEM = standard error of mean
b.d = below detection limit
n.a = not available

Poistettu: Numbers
Poistettu: are
Poistettu: ml

Table 3. The most abundant metabolism-related A) bacterial and B) archaeal genes in the predicted metagenomes.

A)	180 m		500 m		967 m		1820 m		2260 m		2300 m	
	DNA	RNA	DNA	RNA	DNA	RNA	DNA	RNA	DNA	RNA	DNA	RNA
Carbohydrate Metabolism	19%	21%	20%	19%	20%	19%	19%	20%	21%	19%	19%	19%
Amino Acid Metabolism	21%	21%	22%	23%	22%	22%	21%	22%	21%	21%	22%	21%
Energy Metabolism	13%	13%	12%	11%	11%	13%	11%	12%	13%	13%	11%	11%
Metabolism of Cofactors and Vitamins	9%	0%	9%	9%	8%	10%	8%	8%	9%	8%	8%	9%
Lipid Metabolism	6%	0%	7%	8%	8%	7%	8%	8%	7%	7%	8%	7%
Nucleotide Metabolism	8%	0%	7%	6%	7%	8%	6%	7%	7%	7%	6%	8%
Xenobiotics Biodegradation and Metabolism	4%	0%	7%	8%	7%	6%	9%	6%	7%	9%	9%	7%
Metabolism of Terpenoids and Polyketides	4%	0%	4%	5%	5%	4%	4%	4%	5%	5%	4%	4%
Metabolism of Other Amino Acids	4%	0%	4%	4%	4%	3%	4%	4%	4%	4%	4%	4%
Enzyme Families	4%	0%	3%	3%	4%	4%	4%	4%	3%	3%	4%	4%
Glycan Biosynthesis and Metabolism	7%	0%	3%	3%	4%	4%	4%	3%	3%	3%	4%	4%
Biosynthesis of Other Secondary Metabolites	2%	0%	2%	2%	2%	2%	1%	1%	1%	2%	2%	1%

B)	180 m		500 m		967 m		1820 m		2260 m		2300 m	
	DNA	RNA	DNA	RNA	DNA	RNA	DNA	RNA	DNA	RNA	DNA	RNA
Energy Metabolism	23%	23%	25%	25%	18%	24%			24%	24%	24%	
Amino Acid Metabolism	21%	21%	21%	21%	25%	22%			22%	22%	22%	
Carbohydrate Metabolism	18%	20%	17%	17%	19%	17%			17%	17%	17%	
Nucleotide Metabolism	11%	10%	11%	11%	13%	11%			11%	11%	11%	
Metabolism of Cofactors and Vitamins	10%	10%	11%	11%	10%	10%			10%	10%	10%	
Xenobiotics Biodegradation and Metabolism	3%	4%	3%	3%	2%	3%			3%	3%	3%	
Enzyme Families	3%	3%	3%	3%	3%	3%	n.d	n.d	3%	3%	3%	n.d
Metabolism of Terpenoids and Polyketides	3%	3%	3%	3%	3%	3%			3%	3%	3%	
Glycan Biosynthesis and Metabolism	2%	2%	3%	3%	2%	2%			2%	2%	2%	
Biosynthesis of Other Secondary Metabolites	2%	2%	2%	2%	2%	2%			2%	2%	2%	
Lipid Metabolism	2%	1%	1%	1%	2%	1%			1%	1%	1%	
Metabolism of Other Amino Acids	1%	1%	1%	1%	1%	1%			1%	1%	1%	

n.d = not detected

Table 4. The "keystone genera" of the total microbial communities in Outokumpu fractures.

Keystone OTUs	Closeness Centrality	Betweenness Centrality	Degree	Family	Relative abundance[1]
Other *Burkholderiaceae*	2,0	394	41	*Burkholderiacaea*	1%
Desulfitobacter	2,4	302	4	*Peptococcaceae*	6%
Other *Clostridiaceae*	1,7	302	20	*Clostridiaceae*	2%
Dethiobacter	2,4	302	26	*Anaerobrancaceae*	25%
Herbaspirillum	1,8	248	48	*Oxalobacteracaea*	6%
Pelomonas	1,8	218	41	*Comamonadaceae*	72%
Novosphingobium	1,1	162	46	*Sphingomonadaceae*	2%
Comamonas	1,8	151	29	*Comamonadaceae*	72%
Average	1,4	21	25		

[1]Highest relative abundance in the family level in the community

Table 5. The "keystone genera" of the active microbial communities in Outokumpu fractures.

Keystone OTUs	Closeness Centrality	Betweenness Centrality	Degree	Family	Relative abundance[1]
Curvibacter	1,8	797	41	*Comamonadaceae*	71%
Comamonas	1,8	476	4	*Comamonadaceae*	71%
Sphingomonas	1,7	474	20	*Sphingomonadaceae*	0,1%
Bacilli OTU87	1,1	294	26	*Bacilli**	1%
Flavobacterium	1,8	276	48	*Flavobacteraceae*	<0,1%
Williamsia	1,1	256	41	*Williamsiacaea*	<0,1%
Staphylococcus	2,5	189	46	*Staphylococcaceae*	1%
Oxalobacteraceae	1,8	172	29	*Oxalobacteraceae*	0,3%
Herbaspirillum	1,9	155	57	*Oxalobacteraceae*	0,3%
Average	1,3	21	37		

[1]Highest relative abundance in the family level in the community
*only identified to class level

Figure legends:

Figure 1. Total number of microbial cells in Outokumpu fracture fluids.

Figure 2. Microbial community structure at six different fractures in Outokumpu Precambrian crystalline bedrock. On the center a schematic presentation of the Outokumpu Deep Drill hole lithology (blue mica schist, green ophiolitic rocks, pink pegmatitic granite) with arrows pointing to the depths of the fractures studied. The composition of the total communities on the left and the active communities on the right side. The taxonomic classification of only the most abundant OTUs is shown. B: bacteria, SRB: sulphate reducing bacteria, A: archaea, MG: methanogens.

Figure 3. Canonical correspondence plot of the microbial OTUs (blue, letters before the OTU name denoting the origin and the domain, i.e. D: derived from DNA, R: derived from RNA, B: Bacteria, A: Archaea), depths (black) and the geochemical parameters (red). Horizontal axis explains the 35 % of the variance of the data with statistical significance ($p < 0.01$), as vertical axis explains 27% of the variance of the data ($p = 0.11$).

Figure 4. Binned A) bacterial and B) archaeal physiotypes according to the predominant metabolism of OTUs in the family level.

Figure 5. The average predicted functionality of all A) bacterial and B) archaeal metagenomes reconstructed from 16S rRNA gene sequences.

Figure 6. The co-occurrence network of the total (A) and active (B) microbial community in Outokumpu bedrock. The size of each node corresponds to the betweenness of centrality value of the OTU. The modules are represented by different colours.

Poistettu: different

Poistettu: in

Poistettu: Figure 7. The co-occurrence network of the active microbial community in Outokumpu bedrock. The size of each node corresponds to the betweenness of centrality value of the OTU. Modules are represented by different colors. ... [2]

To evaluate the operational capacity of different metabolic pathways detected with PICRUST, HUMAnN analysis was performed on the predicted metagenomes. Genes needed for several amino acid biosynthesis pathways and transport systems to function were present. The relative abundance of the genes involved in carbon fixation pathways such as Arnon-Buchanan, Wood-Ljungdahl and Calvin cycle were low according to the HUMAnN analysis (Supplementary Table 7). Calvin cycle genes were only present in the predicted metagenomes of the active bacterial communities in 1820 and 2260 m fractures. The coverage of pathways involved in carbohydrate metabolism (KEGG modules 1-4, 6-9,11) was high.

The genes needed for coding of the enzymes in sulfur and sulphate reduction pathways were only covered in the communities in the 180 m fracture. Aerobic methane oxidation pathway on the other hand was considered operational in communities from 500 m fracture and other fractures below this (Supplementary Table 7).

Figure 7. The co-occurrence network of the active microbial community in Outokumpu bedrock. The size of each node corresponds to the betweenness of centrality value of the OTU. Modules are represented by different colors.

Figures:

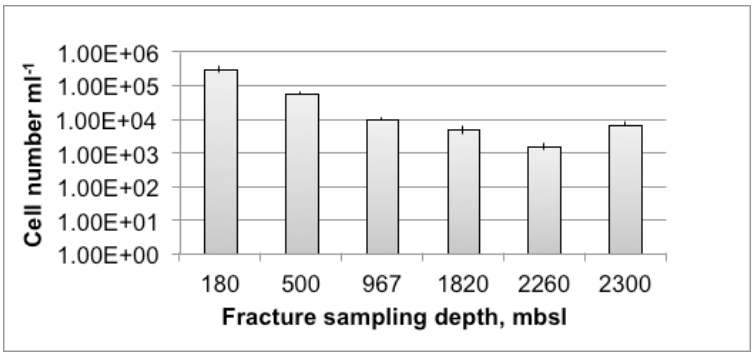


Figure 1.

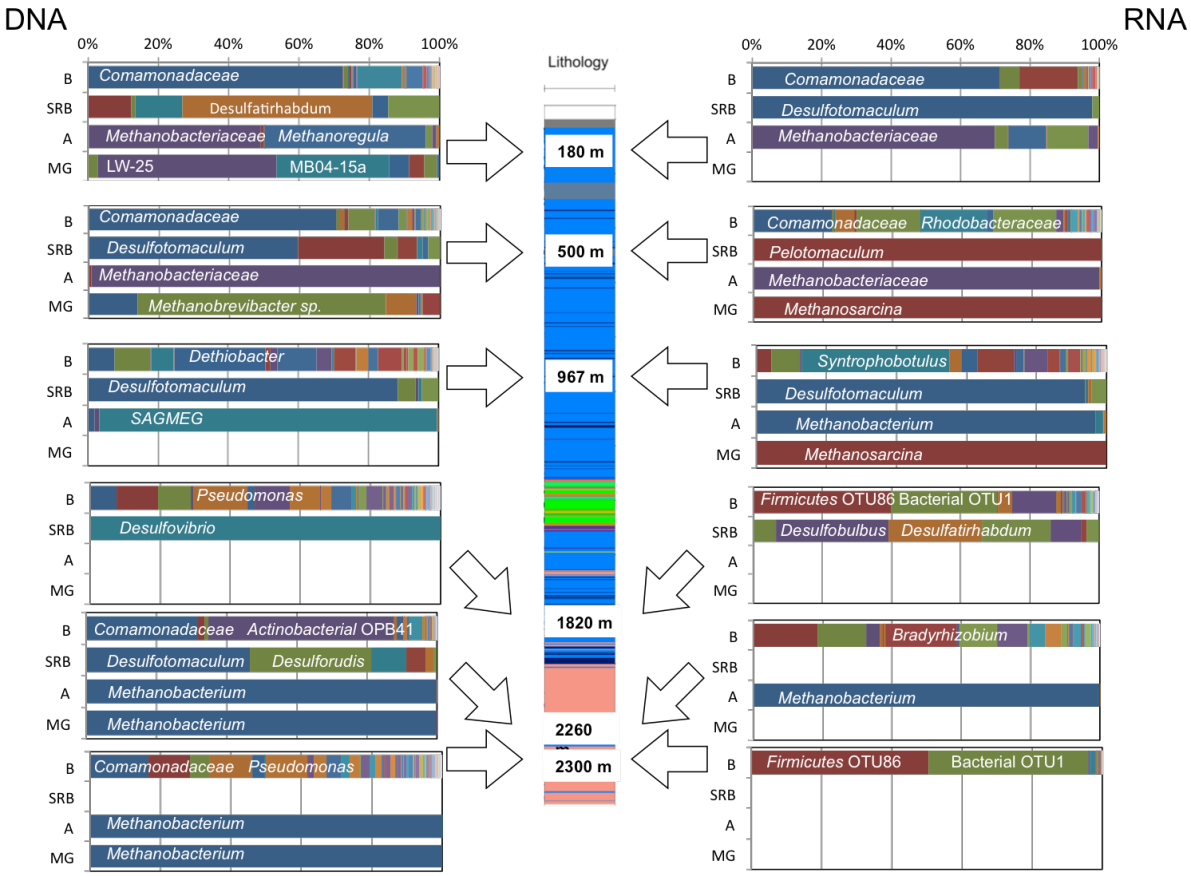

DNA
RNA
Lithology
0%
20%
40%
60%
80%
100%
B
SRB
A
MG
Comamonadaceae
Desulfatirhabdum
Methanobacteriaceae
Methanoregula
LW-25
MB04-15a
180 m
Desulfotomaculum
Methanobacteriaceae
Comamonadaceae
Rhodobacteraceae
Pelotomaculum
Methanosarcina
500 m
Methanobrevibacter sp.
Dethiobacter
SAGMEG
967 m
Syntrophobotulus
Methanobacterium
Pseudomonas
Desulfovibrio
Firmicutes OTU86
Bacterial OTU1
Desulfobulbus
1820 m
Comamonadaceae
Actinobacterial OPB41
Desulforudis
Bradyrhizobium
2260
2300 m

Figure 2.

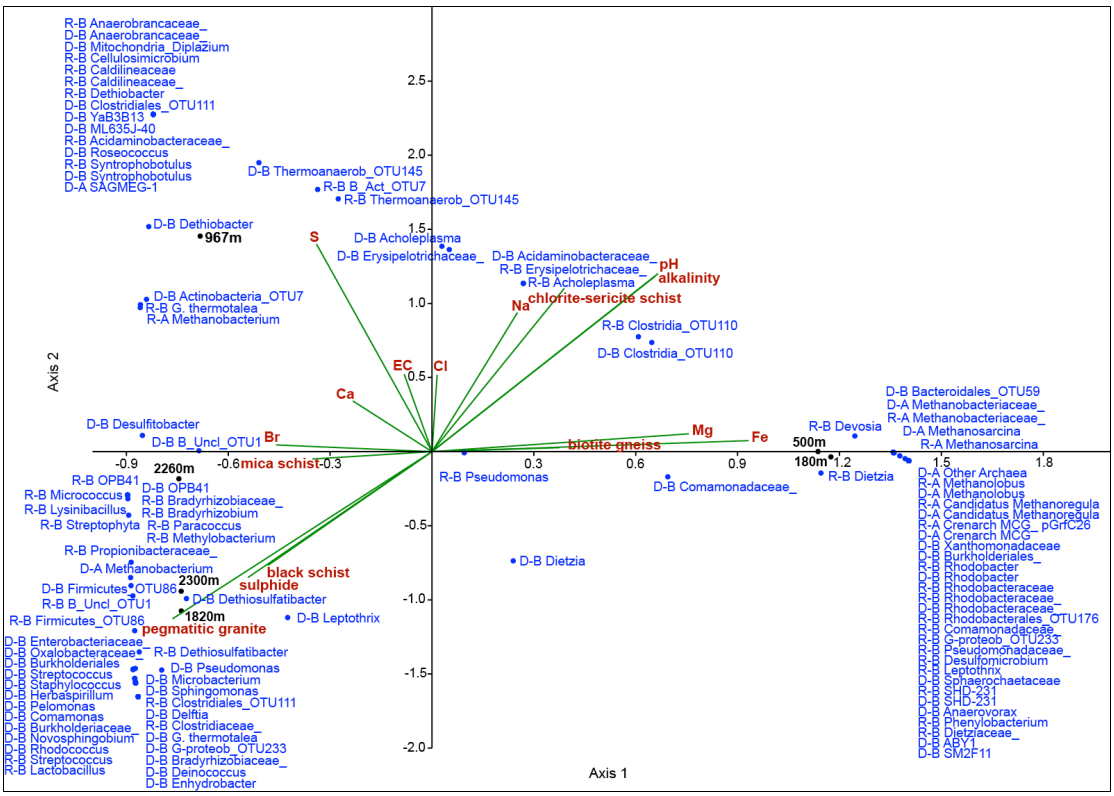
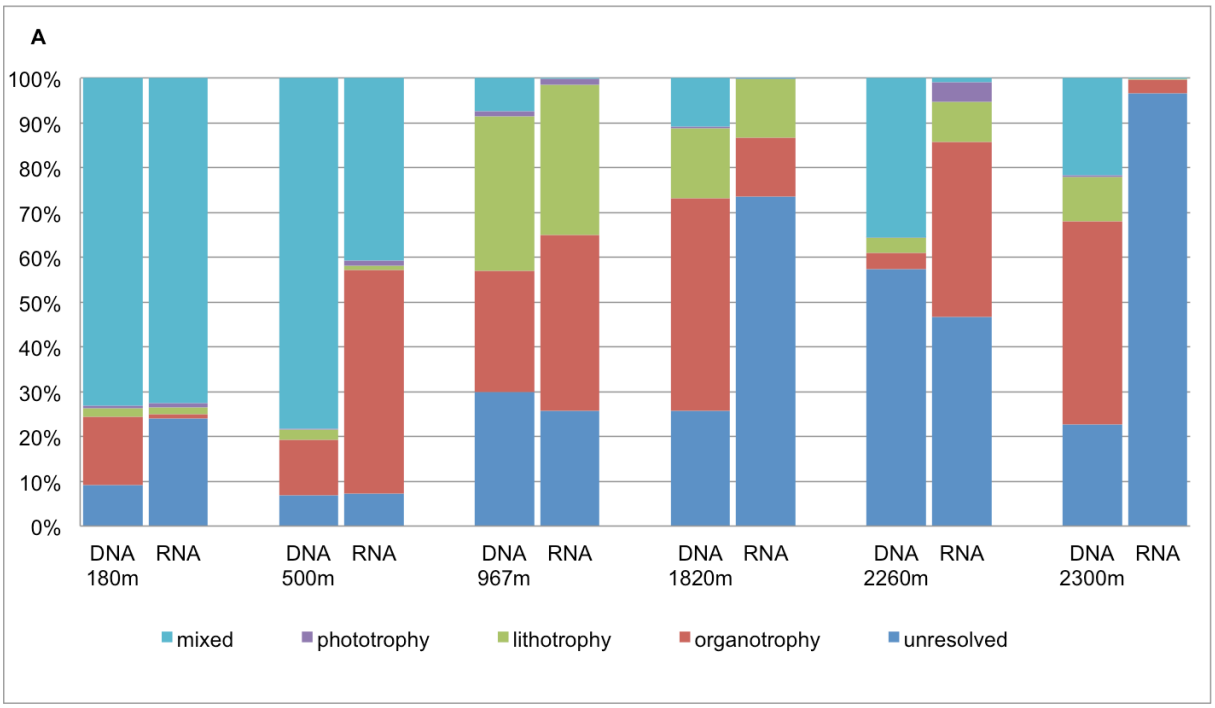

Axis 1
Axis 2
S
pH
alkalinity
Na
chlorite-sericite schist
EC
Cl
Ca
Br
Mg
Fe
biotite gneiss
mica schist
black schist
sulphide
pegmatitic granite
967m
2260m
2300m
1820m
500m
180m
R-B Anaerobrancaceae_
D-B Anaerobrancaceae_
D-B Mitochondria_Diplazium
R-B Cellulosimicrobium
R-B Caldilineaceae
R-B Caldilineaceae_
R-B Dethiobacter
D-B Clostridiales_OTU111
D-B YaB3B13
D-B ML635J-40
R-B Acidaminobacteraceae_
D-B Roseococcus
R-B Syntrophobotulus
D-B Syntrophobotulus
D-A SAGMEG-1
D-B Thermoanaerob_OTU145
R-B B_Act_OTU7
R-B Thermoanaerob_OTU145
D-B Dethiobacter
D-B Acholeplasma
D-B Erysipelotrichaceae_
D-B Acidaminobacteraceae_
R-B Erysipelotrichaceae_
R-B Acholeplasma
D-B Actinobacteria_OTU7
R-B G. thermotalea
R-A Methanobacterium
R-B Clostridia_OTU110
D-B Clostridia_OTU110
D-B Bacteroidales_OTU59
D-A Methanobacteriaceae_
R-A Methanobacteriaceae_
D-A Methanosarcina
R-A Methanosarcina
D-B Desulfitobacter
D-B B_Uncl_OTU1
R-B Devosia
R-B Pseudomonas
D-B Comamonadaceae_
R-B Dietzia
D-A Other Archaea
R-A Methanolobus
D-A Methanolobus
R-A Candidatus Methanoregula
D-A Candidatus Methanoregula
R-A Crenarch MCG_ pGrfC26
D-A Crenarch MCG
D-B Xanthomonadaceae
D-B Burkholderiales_
D-B Rhodobacter
D-B Rhodobacter
R-B Rhodobacteraceae
R-B Rhodobacteraceae_
R-B Rhodobacteraceae_
R-B Rhodobacterales_OTU176
R-B Comamonadaceae
R-B G-proteob_OTU233
R-B Pseudomonadaceae_
R-B Desulfomicrobium
D-B Leptothrix
D-B Sphaerochaetaceae
R-B SHD-231
D-B SHD-231
D-B Anaerovorax
R-B Phenylobacterium
R-B Dietziaceae_
D-B ABY1
D-B SM2F11
R-B OPB41
D-B OPB41
R-B Micrococcus
R-B Lysinibacillus
R-B Streptophyta
R-B Bradyrhizobiaceae_
R-B Bradyrhizobium
R-B Paracoccus
R-B Methylobacterium
R-B Propionibacteraceae_
D-A Methanobacterium
D-B Firmicutes_OTU86
R-B B_Uncl_OTU1
R-B Firmicutes_OTU86
D-B Dethiosulfatibacter
D-B Leptothrix
D-B Enterobacteriaceae_
D-B Oxalobacteraceae
D-B Burkholderiales
D-B Streptococcus
D-B Staphylococcus
D-B Herbaspirillum
D-B Pelomonas
D-B Comamonas
D-B Burkholderiaceae_
D-B Novosphingobium
D-B Rhodococcus
R-B Streptococcus
R-B Lactobacillus
D-B Pseudomonas
D-B Microbacterium
D-B Sphingomonas
R-B Clostridiales_OTU111
D-B Delftia
R-B Clostridiaceae_
D-B G. thermotalea
D-B G-proteob_OTU233
D-B Bradyrhizobiaceae_
D-B Deinococcus
D-B Enhydrobacter
D-B Dietzia

Figure 3.


A
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
DNA 180m
RNA
DNA 500m
RNA
DNA 967m
RNA
DNA 1820m
RNA
DNA 2260m
RNA
DNA 2300m
RNA
mixed
phototrophy
lithotrophy
organotrophy
unresolved

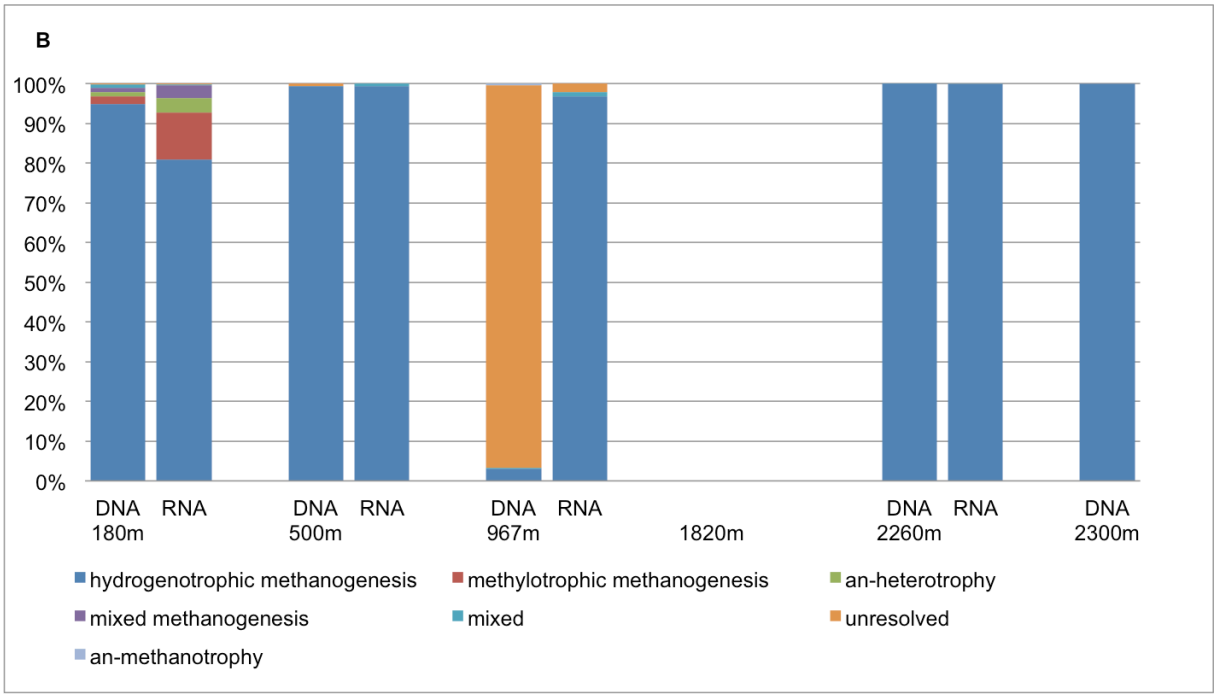
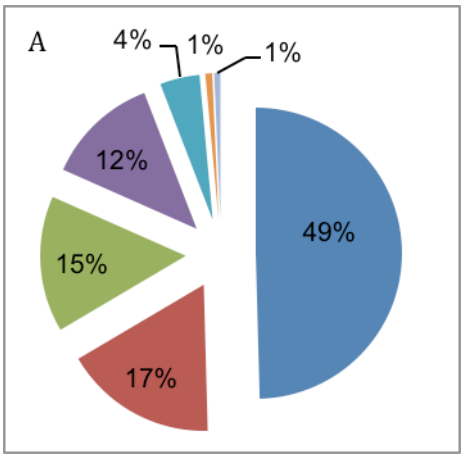

B
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
DNA 180m
RNA
DNA 500m
RNA
DNA 967m
RNA
1820m
DNA 2260m
RNA
DNA 2300m
hydrogenotrophic methanogenesis
methylotrophic methanogenesis
an-heterotrophy
mixed methanogenesis
mixed
unresolved
an-methanotrophy

Figure 4.


A
4%
1%
1%
12%
49%
15%
17%

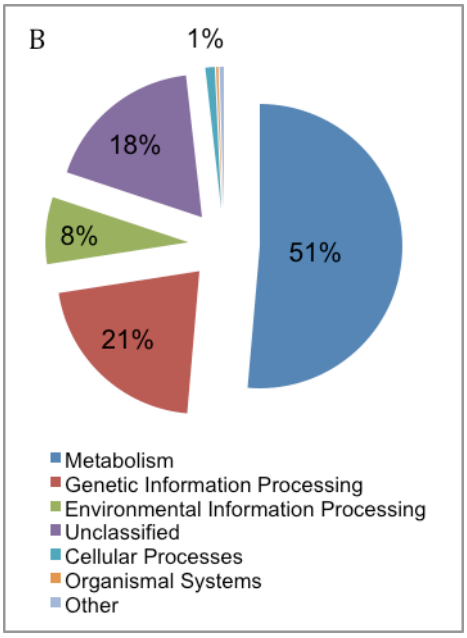


Figure 5.

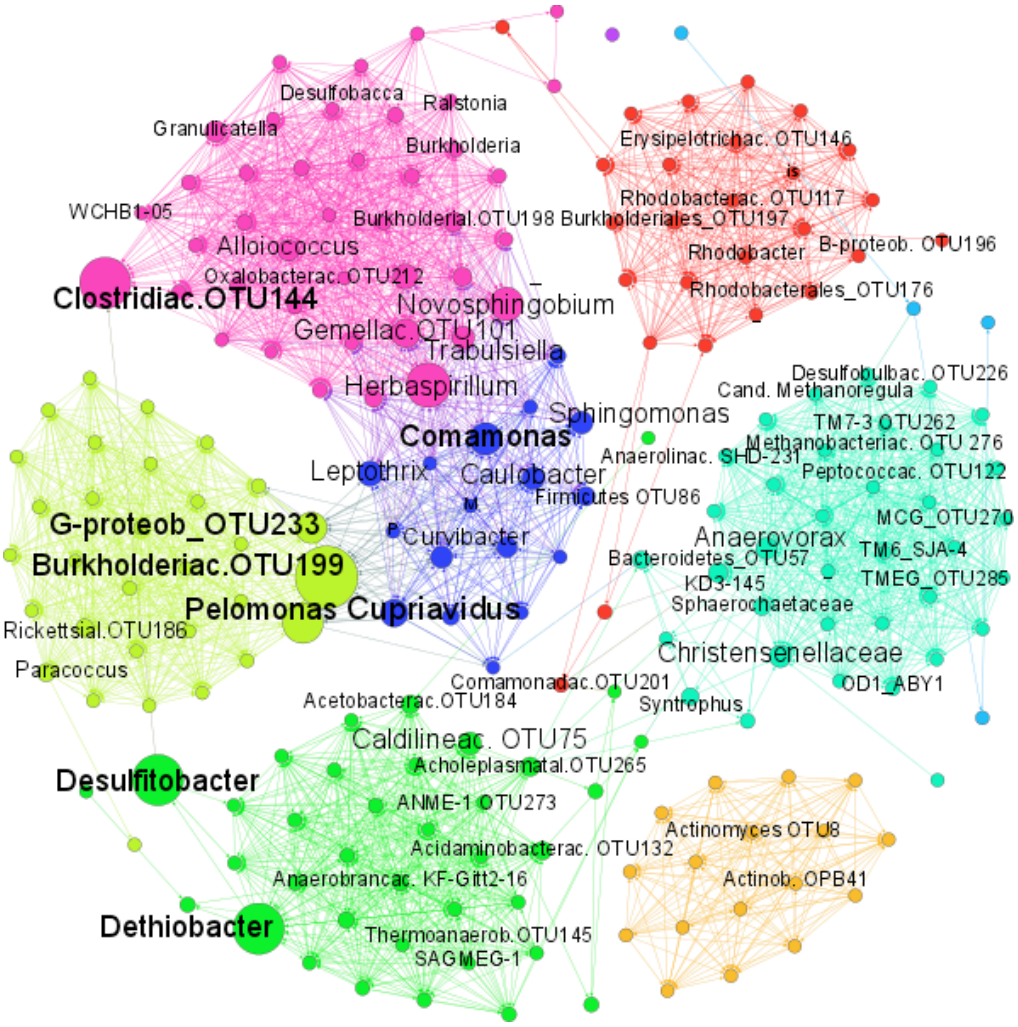


Figure 6.

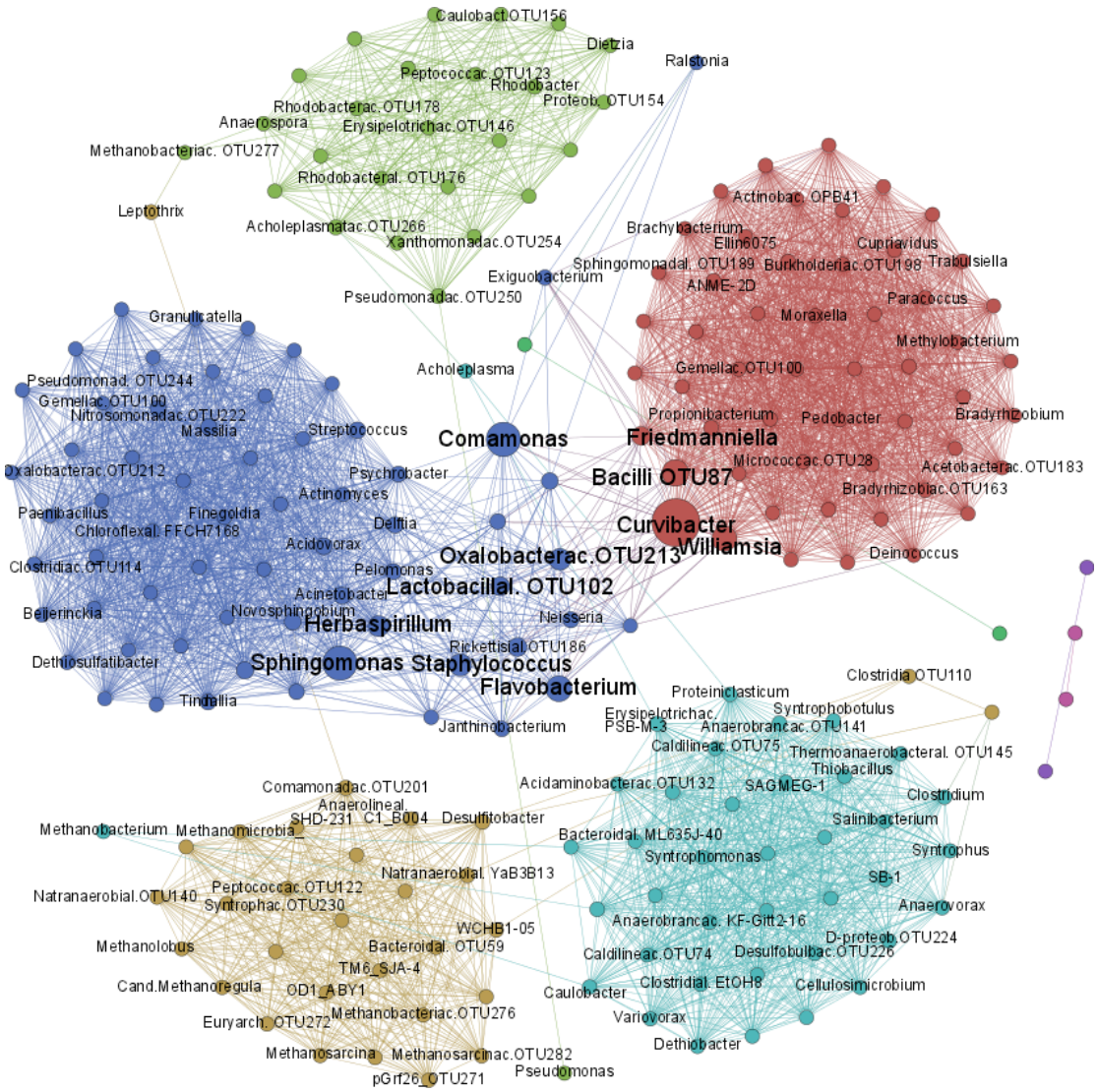

Caulobact.OTU156
Dietzia
Ralstonia
Peptococcac.OTU123
Rhodobacter
Proteob_OTU154
Rhodobacterac.OTU178
Anaerospora
Erysipelotrichac.OTU146
Methanobacteriac. OTU277
Rhodobacteral. OTU176
Leptothrix
Acholeplasmatac.OTU266
Xanthomonadac.OTU254
Pseudomonadac.OTU250
Exiguobacterium
Actinobac. OPB41
Brachybacterium
Ellin6075
Cupriavidus
Sphingomonadal. OTU189
Burkholderiac.OTU198
Trabulsiella
ANME-2D
Paracoccus
Moraxella
Methylobacterium
Gemellac.OTU100
Granulicatella
Acholeplasma
Pseudomonad. OTU244
Gemellac.OTU160
Nitrosomonadac.OTU222
Massilia
Streptococcus
Propionibacterium
Pedobacter
Bradyrhizobium
Comamonas
Friedmanniella
Micrococcac.OTU28
Acetobacterac.OTU183
Oxalobacterac.OTU212
Psychrobacter
Bacilli OTU87
Bradyrhizobiac.OTU163
Actinomyces
Paenibacillus
Finegoldia
Chloroflexal. FFCH7168
Delftia
Curvibacter
Acidovorax
Williamsia
Deinococcus
Clostridiac.OTU114
Oxalobacterac.OTU213
Pelomonas
Acinetobacter
Lactobacillal. OTU102
Bellerinckia
Novosphingobium
Herbaspirillum
Neisseria
Rickettisial.OTU186
Dethiosulfatibacter
Sphingomonas Staphylococcus
Flavobacterium
Tindallia
Janthinobacterium
Clostridia.OTU110
Proteiniclasticum
Syntrophobotulus
Erysipelotrichac.
PSB-M-3
Anaerobrancac.OTU141
Caldilineac.OTU75
Thermoanaerobacteral. OTU145
Thiobacillus
Comamonadac.OTU201
Acidaminobacterac.OTU132
SAGMEG-1
Clostridium
Anaerolineal.
SHD-231
C1_B004
Desulfitobacter
Salinibacterium
Methanobacterium
Methanomicrobia_
Bacteroidal. ML635J-40
Syntrophomonas
Syntrophus
Natranaerobial. YaB3B13
SB-1
Natranaerobial.OTU140
Peptococcac.OTU122
Syntrophac.OTU230
Anaerovorax
Anaerobrancac. KF-Gitt2-16
WCHB1-05
D-proteob.OTU224
Methanolobus
Bacteroidal. OTU59
Caldilineac.OTU74
Desulfobulbac.OTU226
TM6_SJA-4
Clostridial. EtOH8
Cellulosimicrobium
Cand.Methanoregula
OD1_ABY1
Caulobacter
Methanobacteriac.OTU276
Variovorax
Euryarch. OTU272
Dethiobacter
Methanosarcina
Methanosarcinac.OTU282
Pseudomonas
pGrf26_OTU271

Figure 7.