

Interactive comment on “The keystone species of Precambrian deep bedrock biosphere belong to Burkholderiales and Clostridiales” by L. Purkamo et al.

Anonymous Referee #1

Received and published: 9 December 2015

In this manuscript the authors analyzed the community structure of bacteria and archaea as well as sulfate reducers and methanogens from different depths within the Outokumpu bedrock using amplicon pyrosequencing. They used metagenomic predictions based on the 16S rRNA-based taxonomy and co-occurrence networks to elucidate potential metabolic pathways and keystone species across the different samples. They found that Comamonadaceae-, Anaerobrancaceae- and Pseudomonadaceae-related OTUs formed the core community with keystone genera related to Burkholderiales and Clostridiales. Moreover, their results suggested only a minor contribution of autotrophic carbon dioxide fixation pathways in the investigated subsurface system. The manuscript addresses an interesting topic and is generally well written, and the

C8363

introduction covers all the important aspects relevant for this study. The authors also clearly state in what way their study is different from previous studies addressing terrestrial subsurface microbiology, and their objectives are clearly described. However, I have some concerns regarding the data analysis and data interpretation that the authors should address: The first question addresses the reliability of the metagenomic prediction. Taxonomic affiliation of the bacterial 16S rRNA reads is based on a rather limited fragment length and a rather small number of reads per sample. For many bacterial groups it is known that genomes can differ substantially even though 16S rRNA gene sequence similarity may be 97% or higher, and that physiologies may differ within a family. So this first step of the analysis already comes along with a certain degree of uncertainty. I wonder if the completion of metabolic pathway modules based on these predicted metagenomes as described in the method section is justified. Could the authors comment on the error or probability of prediction associated with this approach? Can you really give a quantitative estimation of the relevance of certain pathways if the original data is only 16S rRNA gene-based information? Given all the uncertainties associated with this analysis, can you really state that a pathway is "complete" (p. 18122, l. 27-28)? In addition, it is questionable whether this approach is really needed to support the key statements of this manuscript. The suggested dominance of heterotrophic over autotrophic metabolisms might just as well have been inferred from the taxonomic affiliation of the community members alone, without doing the metagenomic prediction. I would also suggest that the keystone species concept should be discussed in more depth. For what reasons are the named species qualified as keystone species besides the fact that this was the outcome of network analysis? The explanation of the metabolic flexibility of the Burkholderiales species is not convincing. If these traits enable them to be well adapted to their environment, why are they then not more abundant? Was it expected that members of the genus *Staphylococcus* would be keystone species of subsurface communities?

Specific comments: Until the end of the introduction part, the reader does not learn whether the investigations of subsurface microbiota are based on rock core material

C8364

or on groundwater obtained from deep aquifers. Please state clearly earlier in the discussion and also in the abstract that the microbial communities investigated here originated from groundwater. p. 18110, l. 16: Please add information that plasmids contained the 16S rRNA gene of *E. coli* ATCC 31608. p. 18111, l. 4-19: Archaeal 16S rRNA genes and *mcrA* genes were amplified in a nested PCR while bacterial 16S rRNA genes and *dsrB* genes were not. It is not clear from the description of the PCR cycling conditions how many PCR cycles were used in total in the nested approach. How do the authors address the potentially higher PCR bias associated with the nested PCR? The sequencing of archaeal 16S rRNA genes and *mcrA* genes might be subject to a different extent of error compared to the bacterial 16S rRNA and *dsrB* genes. p. 18112, l. 16: Why was the length limit of *mcrA* sequences set to 100 bp? Please comment on how the limited size of the fragment may affect phylogenetic analysis of *mcrA* sequences. p. 18113, l. 2-9: The number of sequences per samples was rather low, especially for archaea and *dsrB*. What was the coverage of diversity for these samples? To what extent could comparisons of community composition be biased due to the low read numbers? p. 18114, l. 21-23: Please give the range of numbers also here in the text. Table 1: Less than 200 mg mL⁻¹ nitrate is a rather rough measurement. Why did the authors not use a more precise method to determine the concentration of this important electron acceptor? Is it likely that units should rather be mg L⁻¹ and not mg mL⁻¹? Table 2: Total cell counts were lower than 16S rRNA gene abundances in the upper samples (which could be explained by multiple 16S rRNA gene operons) while the trend was inversed for the deep samples. Please comment on that.

Interactive comment on Biogeosciences Discuss., 12, 18103, 2015.