Reference: bg-2023-147

Title: Continental scientific drilling and microbiology: (extremely) low biomass in crystalline bedrock of central Sweden

Kalmar, 31 October 2023

Dear editor,

Thank you for the constructive comments. We would hereby like to submit a revised version of the original manuscript. We have copied the comments below and added a point-by-point response. All comments were carefully considered, and we believe we have met the comments and suggestions.

Please note that the line numbers in the reply refer to the revised, annotated manuscript.

George Westmeijer *et al.*

Reviewer 1
General impression:
Westmeijer et al. address one of the major concerns for deep continental biosphere microbial ecologists: irreversible contamination of previously isolated, low biomass microbial communities by microbes from drilling fluid and subsequent contamination in the lab from nucleic acid isolation procedures and sequencing. The study emphasizes not only the importance of tracing potential sources of contamination (e.g., drilling fluids, reagents), but also testing multiple extraction methods to assess if there could be discrepancies between techniques that may yield differing results. A study like this is appreciated for its candor in publicizing a common yet far-too-often underreported reality of deep biosphere microbial ecology studies: often there just isn't enough biomass available to allow the use of widely used tools developed for higher biomass environments.
Reply: We are grateful for the appreciation of our manuscript.

The manuscript is generally well written, though I believe its clarity and general importance to the field would benefit from a more thorough discussion about the many steps in the "sample collection to data analysis" pipeline that are considered the most "at risk" for issues like contamination (by microbes or nucleases), user error, and the introduction of bias in deep biosphere microbial ecology research. Likewise, it would be incredibly beneficial to the community at large to be made aware of successful efforts to mediate (or at least be mindful of) these "sample integrity" bottlenecks. For example, there is an established literature on the pervasive issue of primer biases against autochthonous deep biosphere clades, particularly archaea (e.g., see the 2008 review by Teske and Sørensen – https://doi.org/10.1038/ismej.2007.90). The authors are clearly mindful of such challenges, as they go to efforts such as testing multiple DNA extraction procedures from both commercially available kits and more traditional phenol-chloroform protocols. However, I think the justification for these efforts, while implied, are not explicitly offered in the text, which seems like a missed opportunity. I would have liked to seen, for example, a presentation of the varying DNA yields from each of the attempted extraction methods, as well as a broader discussion speaking to the general bias of commercially available kits against isolating microbes from deep biosphere samples (e.g., Lloyd et al., 2020 and references therein https://doi.org/10.1128/AEM.00877-20).
Reply: Agreed. The reason why six extraction protocols and three PCR primers were used is now stated: "Biases are introduced during both DNA extraction and PCR amplification (Teske and Sørensen, 2008; Morono et al., 2014). To address this, six extraction protocols were tested, using methods based on phenol-chloroform adapted to low biomass rock samples and commercially available kits while maximizing input material (ground rock). To minimize amplification bias, three different primer pairs were tested, targeting both bacteria and archaea. Despite these measures, none of the extracts from the cores contained detectable DNA (< 0.05 ng μL^{-1})." Unfortunately, no quantifiable amounts of DNA were detected (lower limit 0.05 ng μL-1) and a comparison of the various extraction method and their yield could not be done. (lines 126-129)

I was surprised by the absence of representative fluorescence microscopy images. Of all the methodologies assessed, this one seemed the most successful at actually verifying the existence of biomass. It has the added benefit that the

methodology is simple and illustrates an essentially pristine and minimally processed sample. I think it should be highlighted.
Reply: An example of a fluorescent microscopy image was added with the caption: "Figure 3 Fluorescence microscopy image of microorganisms in drill core at 112 m depth. Microbial populations were visualized with the DNA-binding dye Syto9 (in green). The presence of microorganisms could not be confirmed with CARD-FISH as the fluorescent signals were deemed as unspecific. The scale bar represents 5 µm." (lines 160-165)

Specific comments:
Lines 79-80: "the available data are patchy at best" is informal phrasing and frankly dismissive of very significant efforts that have been made to elucidate the diversity, abundance, and distribution of microbes inhabiting the deep subsurface biosphere (e.g., Lau et al., 2016 https://doi.org/10.1073/pnas.1612244113; Lloyd et al., 2018 https://doi.org/10.1128/msystems.00055-18; Magnabosco et al., 2018 doi: 10.1038/s41561-018-0221-6, Bar-On et al., 2019 https://doi.org/10.1073/pnas.1711842115, Meyer-Dombard and Malas 2022 https://doi.org/10.3389/fmicb.2022.891528). I would strongly encourage the authors to reconsider this phrasing.
Reply: Thank you. Done. This sentence now reads: "Despite the concerns regarding contamination, studies on this vast biome are necessary for understanding the role of the deep subsurface microbiome in carbon and nutrient cycling and would contribute to the growing body of literature on the continental subsurface." (lines 79-80)

Lines 93-94: "the drilling fluid contained on average 335 +/- 396 cells/mL". This is a very precise number and also a large standard deviation implying that the minimum quantifiable limit of cells in the procedural blank was 0 cells/mL. Is this the case?
Reply: Correct. Using black membranes, molecular grade reagents, and preparing the slides in the dark allowed to quantify cell densities as low as 1 cell mL^{-1}. Due to the high standard deviation, the range (1 - 1553 cell mL^{-1}) was also provided. (line 95)

Line 100: It might be easier to interpret Figure 1B if the sampled depths were provided on the temperature vs. depth line rather than in the lithology stack. There is also a discrepancy between the depths presented as sampled in Figure 1B and the depths listed in Table 2. Related: was temperature measured continuously in the borehole or is the line in Figure 1B extrapolated from discrete measurements? Clarification on the methodology would be appreciated.
Reply: As requested, the sampled fractures are now shown on the temperature vs. depth line and colored according to lithology. Temperature was measured continuously as part of the logging and over 200 thousand temperature measurements in function of depth were used in Fig. 1b. (lines 100-105)

Line 150: In Table 2, are the DNA quantities for both the fracture sample and the drilling fluid presented as total ng? Or is it ng normalized per unit mass or volume (e.g., per gram, per mL, or per µL)?
Reply: DNA quantities are now shown as concentration per gram of input material (ng μL^{-1} g^{-1}). (line 175)

Line 233: It is not mentioned what mass of material was removed from the fractures, and since it is not clear if/how the DNA recovery amounts are normalized per unit mass, I do not have a clear frame of reference for how much material was processed that resulted in the reported biomass amounts.
Reply: The amount of sampled material from the fracture (2-3 g) is now specified. (line 303)

Technical comments:
Please either put spaces between paragraphs or indent new paragraphs to make the manuscript easier to read.
Reply: Thank you. Done.

The phrase "quenching" of fluorescent beads is used several times throughout the manuscript but not defined.
Reply: Thank you. "Extinguished" is used instead of "Quenched" throughout the manuscript.

Reviewer 2

Overall assessment

The manuscript „Continental scientific drilling and microbiology: (extremely) low biomass in crystalline bedrock of central Sweden" by Westmeijer et al. addresses a very important aspect of deep biosphere microbiology and environmental microbiology dealing with low biomass samples in general: the ratio of contaminating DNA sequences to the DNA signal of the target microbial community. The manuscript is well written, and the presented figures and tables are clear. Just as my fellow reviewer Dr. Harris, I am thankful to the authors for addressing this important yet underreported and sometimes even willingly ignored obstacle on our way to understanding deep subsurface microbial life. I have no major criticism for the manuscript. However, I have some minor comments and suggestions, which I hope will make the presented results even more insightful and useful to the environmental microbiology community.
Reply: We are grateful for the appreciation of our manuscript.

General points

While the authors convincingly show that the contaminations are unlikely to be introduced during the drilling process, e.g. by drilling fluid, they stay vague about the exact sources of contamination and suggestions for minimizing it. To improve this part of the discussion I would like to suggest following changes and additions to the authors:

1) It would be helpful to place descriptions of sterile working practices (now LL. 292-294) right at the beginning of the DNA extraction methods section. You could also make them more detailed. How controlled was the environment for solution preparation or DNA extraction? Was a cleanroom used? You could even write such handling details as aliquoting practices of mol. grade solutions.
Reply: Agreed. The text was moved, and more details was added: "Extractions were performed in a room dedicated to extraction of nucleic acids while working in a fume hood that was sterilized with both bleach and 70 % ethanol. Maximum eight samples (including controls) were extracted simultaneously, and the reagents were aliquoted according to the volume needed for processing the specific number of samples. For all methods described, reagents used were molecular grade, all plastic consumables (tips, centrifuge tubes) were PCR grade and placed for 10 min in a UV hood before moving the material to the fume hood. Glassware was baked at 160 °C for 3 h, and the lysis solution was filter sterilized (0.1 µm pore size) prior to autoclaving. The autoclave was cleaned prior to use and filled with fresh deionized water." lines (345-351)

2) Assuming the contaminations stem from the reagents, it would be useful to narrow it down a little bit: which reagents or labware are the most likely sources of contamination? Have you sequenced just a blank clean-up reaction vs. blank extraction + clean-up? While I assume that molecular-grade certified reagents can be trusted and that you were extremely diligent while preparing solutions, I am skeptical with respect to clean-up kits, especially such broad application ones like the Norgen Clean-All. Typical commercial kits have already been shown as a source of background contamination in previous studies (Salter et al. 2014 which is also cited by the authors). Such kits are made for broad use with mostly high biomass samples (cultures, gut, soil, bioreactor sludge) where kit contaminants stay "invisible" in the sequence data. It is not clear how kit columns are sterilized or certified as clean after their assembly. They contain great surfaces for contaminating microorganisms to attach to: plastic tubes, membranes, sealing rings. However, I am aware that we need these clean-up kits for PCR to work on most soil, sediment, or rock extracts. Could kits based on magnetic beads be cleaner? Are there special ultra-clean kits for low biomass already on the market?
Reply: Unfortunately no blank clean-up control was sequenced. More information with a suggestion on the origin of the contamination was added as: "These contaminants, known to be notorious reagent contaminants, particularly in commercially available kits, were most likely introduced during DNA extraction despite strict measurements to prevent this. Most reagents used were high-quality including molecular grade phenol-chloroform, ethanol, and co-precipitant, suggesting the contaminants originated from either the column-based commercial kit used for purifying the DNA extract or the lysis solution, despite combining sterile filtering (0.1 µm pore size) with autoclaving of the latter. To better adapt the DNA extraction protocol to microbial biomass, it could be desirable to subsample material to be extracted in order to quantify cell density (for example, using fluorescence microscopy) prior to starting molecular work." (lines 213-219)

[As a personal recommendation from my own experience: I found magnetic beads-based kits to be cleaner. They also exist in versions with low elution volumes (10 µl), which helps concentrating nucleic acids after clean-up.]

3) The removal of ASVs found in the negative controls does not seem to have worked, as typical contaminant taxa were still left among the sequences from your samples. Could you discuss it in a little bit more detail, since it is the standard approach many scientists are relying on. Should we remove genus-level taxa instead of precise ASVs? What is your take on why it happened: was it an imperfection of the ASV algorithm as it failed to group sequences of contaminating organisms correctly?
Reply: Indeed, it seems removal of the contaminant ASVs from the samples of interest was insufficient as the remaining taxa in the fracture samples were affiliated with notorious contaminants. Removal of contaminant taxa on a genus level would partially solve this issue. The following sentences were added to discuss this suggestion: "Removing contaminants based on taxonomy (e.g., genus) instead of ASVs would partially resolve this issue. However, this led to only 306 ASVs (3.7 % of the sequence counts) being retained compared to 1,304 ASVs (11 %) when using ASVs. Either way, both approaches indicated the dataset was predominantly comprised of contaminant clades and unsuitable for ecological interpretations." (lines 189-210)

4) Based on your qPCR data, how many cells went into DNA extraction? What was the qPCR value of the DNA extraction + clean-up blank (i.e. how many copies / g would you have back-calculated with an average amount of sample input)? Can you make a recommendation for the minimum amount of cells needed for DNA extraction with current methods in order to have a good signal to contaminant ratio? How do your results compare to the 104 cells identified as a threshold by Salter et al. (2014)? Would you recommend assessing cell densities before deciding on DNA extraction protocol in future deep subsurface studies (and other studies of low biomass environments)?
Reply: On average, there were $9.6 \times 10^4 \pm 1.0 \times 10^5$ gene copies per 2 g of fracture sample (mean ± sd, n = 9). As 2 g of input material was used for the extraction, ~ 10^5 cells went into each DNA extraction (on average). The qPCR value for the extraction blanks (~ 10^3 gene copies g^{-1}) is now shown in Table 2. The following sentence was added: "The high abundance of contaminant populations was also reflected in the qPCR data, with the gene copies in the fractures ranging from 10^3 to 10^5 copies g^{-1} while the extraction controls contained 10^3 copies g^{-1}, giving a ratio as low as 1 to 10 for six out of seven samples." (lines 210-212) A suggestion on estimating microbial abundance prior to DNA extraction was added as: "To better adapt the DNA extraction protocol to the microbial biomass in the samples environment, it could be desirable to subsample material to be extracted in order to quantify cell density (for example, using fluorescence microscopy) prior to starting molecular work." (lines 217-219)

Specific comments:

Table 1: Are there cell counts from fracture samples? Are DNA amounts total amounts in ng or concentrations in ng/µl?
Reply: As the amount of fracture sample (~ 2 g) was limited, it was decided to use all material for DNA extraction and therefore cells were not counted. Instead, real-time PCR was done on the DNA extract to have an estimate on microbial abundance. DNA amounts in former Table 1 (now Table 2) are now shown in ng μl^{-1} g^{-1} fracture sample. (lines ...)

Table 2: You could add qPCR data of extraction blanks (i.e. how many copies / g would you have back-calculated with an average amount of sample input)
Reply: The extraction controls were added to Table 2 (former Table 1, line 175)

Figure 3: If all Burkholderiales are Ralstonia, I would specify it in the figure. Burkholderiales is a very diverse order and some (mostly uncultured) families of Burkholderiales are typical chmolithoautotrophic inhabitants of oligotrophic freshwater environments.
*Reply: Agreed. A nuance has been added to the caption of Figure 4 (former Figure 3) and now reads: "The four most abundant orders were **mainly** represented by genera described as notorious reagent contaminants such as Ralstonia (Burkholderiales), Sphingomonas (Sphingomonadales), Corynebacterium (Mycobacteriales) and Agrobacterium (Rhizobiales)." (line 240). The relative abundance (% of sequence counts) has been added to the most abundant genera. (line 187).*

Figure 3 or Figure 4: Would it make sense to add a bar chart of only the "contamination-free" community, even if it does not carry an ecological signal/pattern?
Reply: Clarification was added to Figure 4 (former Figure 3) that the community composition was visualized after removing the ASVs present in the drilling fluid and controls from the fracture samples. (line 239)

Missing figures: I agree with Dr. Harris, fluorescence microscopy pictures of potential cell signals and false-positive signal / no signal / particle autofluorescence could be shown as an additional figure.
Reply: Agreed, a fluorescence microscopy image was added as Figure 3 while mentioning the signal could not be confirmed with CARD-FISH in the caption. (lines 162-165)

Supplements: Can an ASV table be uploaded as a supplement (excel or tab-delimited file)? Just to fully understand what is going on with point 3 mentioned above.
Reply: Together with the R Markdown, the annotated ASVs (asvs_annotated.tsv) and metadata (metadata.tsv) is now uploaded to Github (https://github.com/geweaa/fractures/).

LL. 131-134: The PFA fixation is not recommended for Gram-positive bacteria, as it makes the cell wall even more impermeable. Gram-positive and Archaea fixation protocols usually use 50% ethanol (Roller et al. 1994 https://doi.org/10.1099/00221287-140-10-2849, Amann & Fuchs 2008 https://doi.org/10.1099/00221287-140-10-2849). While this could be a reason for CARD-FISH failure, it doesn't seem that you have many Gram-positives in your samples. "Over-fixation" is also said to lead to decreased permeability, although I cannot find a reference systematically showing it right now. How long was your fixation? Overnight at 4°C? At room temperature it should be not more than 1-2 h.
Reply: Fixation was done for 2 h at room temperature. Permeabilization of the cell wall with lysozyme and acromopectidase should be sufficient to detect Gram-positive bacteria by CARD-FISH. However, as stated in the manuscript, we cannot rule out that the choice of permeabilization methods caused the lack of a signal during the CARD-FISH.

LL. 163 – 165: Do you have data from previous studies of similar samples, with more biomass, that can be used as a reference, a "positive control" of what taxa can be expected? Groundwater, rock from the same region, some habitat where the organisms you expect to find could be sourced/seeded from? If yes, it would be worth mentioning here or showing in Fig. 3
Reply: Studies done on Outokumpu borehole, Olkiluoto island, and Äspö Hard Rock Laboratory all characterized microbial ecology in groundwaters of the deep subsurface. Often with higher biomass (detectable amounts of DNA). Purkamo et al. (https://doi.org/10.3390/life10010002) included DNA extractions from ground rock from the Fennoscandian Shield containing ultra-low biomass. However, the authors also acknowledge that the data derived from ultra-low biomass samples had a high abundance of known kit contaminants (section 4.4).

L. 244: "1000-fold magnification": Are you referring to combined 10x eyepiece plus 100x objective magnification? It is more common just to state the 100-fold objective.
Reply: This has now been rephrased as: "Finally, 300 cells or ten fields of view (5 × 5 raster) were counted in the dark using a 100-fold objective on an epifluorescence microscope (Olympus BX50) combined with a light source (Olympus) that illuminated the DNA/dye complex at a wavelength of 497 nm." (lines 314-316)

L. 293: 10 min in the UV hood is not a lot. Consider that there are bacteria able to survive high doses of radiation.
Reply: Agreed. The UV hood was not used to sterilize the consumables but rather for sterilizing the outside of, for example, a tip box. By doing so, we aimed to reduce the contamination risk before placing the material in the thoroughly cleaned fume hood. This was phrased as: "For all methods described, reagents used were molecular grade, all plastic consumables (tips, centrifuge tubes) were PCR grade and placed for 10 min in a UV hood before moving the material to the fume hood." (lines 348-349)

L. 296: What was the input material amount for fracture samples?
Reply: The input material is now stated: "Fracture samples (n = 29) were extracted using the custom protocol based on Lever et al. (2015a) described above, adding the entire sample (2 to 3 g)." (lines 379-380)

L. 342: Which program was used for phylogenetic tree calculation?
Reply: The R libraries (phangorn and ape) used for constructing the phylogenetic tree are now added. (line 434)

Technical comments:

L. 50: Rephrase: “without isolated representatives for which the ecology is not fully understood” Is it the ecology of the isolates that is not fully understood? Or the ecology of “the large proportion of diversity” that does not have representative isolates?
Reply: Done. This sentence now reads: "Genomic approaches have greatly increased the understanding of microbial diversity in the continental subsurface and revealed a large proportion of diversity for which the ecology is not fully understood (Hug et al., 2016)." (lines 48-50)

LL. 83-84: Which workshop are you referring to?
Reply: The reference to the workshop was removed and this sentence now reads: "Pursuing answers to these questions, two subsurface scientific drilling projects termed Collisional Orogeny in the Scandinavian Caledonides (COSC-1 and COSC-2), located in western central Sweden, were performed in 2014 and 2020 (Lorenz et al., 2015, 2022)." (lines 83-85)

L. 269: Move the description of DNA extraction here: after sample material preparation and before post-extraction assessments of DNA yield etc.
Reply: Done. (lines 345-351)