

Interactive comment on “Microbial functional signature in the atmospheric boundary layer” by Romie Tignat-Perrier et al.

Anonymous Referee #1

Received and published: 23 September 2020

General comments: In this paper, the authors present results of metagenomic sequencing of air filter samples collected at nine different locations around the world with functional profiles for fungi and bacteria. The authors compare their data with data sets from various other ecosystems. The main aim of the study was to characterize the functional potential of the airborne community and to identify potential atmosphere-specific signatures. The results indicate that the functional potential of fungi and bacteria is not specific for the atmosphere but similar to the underlying ecosystem. The manuscript is well written but can be further improved as suggested below.

Specific comments: Real-Time PCR analysis: The authors give a method description for qPCR (starting L133) on air filter samples but the results for the air samples appear to be already published somewhere else as stated in L242/243 and caption of Table S4.

C1

Table S4 gives also some qPCR results for soil samples. Where are these data from? Can the authors provide details in the method section about collecting/extracting/qPCR for the soil samples or a reference? To which air samples/location do these soil samples belong? Were they taken at the same time/location as air sampling was done? It is also not clear to which unit (L, m³, gram?) the gene copy numbers given in Table S4 refer and how the authors calculated the cell concentrations as stated in L133, L146, L243/244 and where these values can be found. The authors refer in L243 to Table S4 for cell concentration ratios, but this table only includes gene copy numbers and their ratios. Overall, the motivation for the qPCR seems not clear in terms of the purpose of the study. The qPCR is not mentioned in the abstract or introduction/motivation and the discussion is confusing as the term “gene copy numbers” seems to be used also as “cell concentrations”. Given the multicopy nature of ribosomal genes and different copy numbers in different organisms this should be corrected in the text.

Metagenomic data analysis: The bioinformatic analysis appears to be focused only on fungi and bacteria, but there are also other microbial organisms such as e.g., Protozoa, Archaea, Algae in the atmosphere. Can the authors add information about the numbers of non-fungal and non-bacterial reads and explain how and why they were excluded although they are/many of them are microorganisms (see title!). Overall, it might be straighter to separate fungi and bacteria in all figures as they belong to different domains of life. For example, Fig 4a and Fig 5a seem not to provide any additional value to panels b and c, if panels a include only fungi plus bacteria but no other microorganisms.

2.2.3/2.3.2: How can the data be normalized to 10000 sequences (L225) when the filtering cut-off was 6000 sequences (L173) before?

Fig2: The numbers in the figure are hard to read. The grey for the air samples appears not to be in the figure. The order seems to be by % of fungal and bacterial sequences. To me it seems more useful to display the different sample groups (air, soil, water,...) so that one might be able to see trends. Does this figure display all sequences for

C2

each of the samples/sample sets i.e., different total numbers of reads/sequences per sample/sample/set or is this figure based on rarefied sequences (6000, 2000)?

L204ff: The authors selected specific stress-related functions with the purpose to identify a specific atmospheric functional potential signature. Stresses like e.g., UV, desiccation, however, are not limited to the atmosphere. Also soil bacteria and microorganisms living on e.g., plant or building surfaces are exposed to these stresses. This might help to explain that the authors did not find a specific signature with their selected genes in the airborne fraction.

L322/323: The authors state that the methane mono-oxygenase-related functional proteins per 10000 sequences were only detectable when considering all sequences. As all sequences are the sum of fungal and bacterial sequences I wonder why they can only detect it when they sum up the sequences. If they have sequences in the sum, they must have had them for fungi and/or bacteria before. Can the authors clarify?

L470/471: Please correct the statement that concentration of fungal spores and fungal hyphae fragments in air are unknown. For example, numbers of spore and hyphae concentration can be found in Després et al., 2012 and references therein.

Després, V.R., Huffman, J.A., Burrows, S.M., Hoose, C., Safatov, A.S., Buryak, G., Fröhlich-Nowoisky, J., Elbert, W., Andreae, M.O., Pöschl, U., Jaenicke, R., 2012. Primary biological aerosol particles in the atmosphere: a review. *Tellus B* 64. <http://dx.doi.org/10.3402/tellusb.v64i0.15598>

Table S1: site should be capitalized, abbreviations in first column should be explained; what is meant with "same hour"? There is no time information in this table.

Figure S1: This is a nice figure, but is only mentioned once and it seems not be used for discussion in the text. I suggest to consider this figure when discussing Fig.3, as it supports the results shown in Fig 3.

Figure S2: This figure is very hard to read as a lot of text overlaps

C3