

Interactive comment on “Determining the hierarchical order by which the variables of sampling season, dust outbreaks occurrence, and sampling location, can shape the airborne bacterial communities in the Mediterranean basin” by Riccardo Rosselli et al.

Riccardo Rosselli et al.

squart@unipd.it

Received and published: 12 March 2021

Anonymous Referee #2 Review bg-2020-324 : . General comment Rosselli et al. (2020) present a paper entitled “Determining the hierarchical order by which the variables of sampling season, dust outbreaks occurrence, and sampling location, can shape the airborne bacterial communities in the Mediterranean basin”. The presented results are of great interest as there is an urgent need of data about bioaerosol's

C1

biodiversity and transport together with the understanding of the parameters shaping the air microbiome. The case studies presented in this paper are very well chosen as they consider different situations, all taken in Sardinia with sampling on two different sites: Sassari (North Sardinia) exposed to European influence, Calgari (South Sardinia) exposed to African influence. Also two time periods have been considered: May and September, as well as dust and clear sky events. NGS-based taxonomic analysis has been carried out in all these samples and compared using various pertinent statistical tools. The results allowed to propose a hierarchy of the variables determining the biodiversity of the collected bioaerosols: sampling season » ongoing meteorological events > sampling site. The sampling period had clearly the major impact. This paper is worth publishing in Biogeosciences when the authors answer some major remarks and possibly make new complementary analyses.

ANSWER: We appreciate the very positive remarks expressed and the praise to the topic interest and overall report's soundness and to add that the paper is suited for a publication in Biogeosciences.

Major concerns My major concern lies on the fact that the authors did not compare in an extensive way the results obtained in this study with those previously published by Rosselli et al (Sci Rep. 2015). In my opinion they should have included these first results obtained at the same location sites (Calagari and Sassari), under dust and clear sky conditions but at a different season (February). The same data sets recorded by the same authors are available and could be compared directly with the present one using the same statistical tools. This comparison would bring strong arguments to generalize the hierarchy of the variables influencing the structure of the microbial communities and make the conclusions clearer.

ANSWER: We are grateful for recognizing the relevance of our 2015 published Scientific Reports article; the comment is acknowledged and will be fully addressed below. As regards the first advice (to include those data and results in the present manuscript and join the datasets to include the published ones as well as a whole analysis) the

C2

limiting guideline is the correct policy of Biogeosciences, which prescribes as requirements: "The work submitted for publication has not been published before, except in the form of abstracts, preprints, published lectures, theses, discussion papers, or similar formats that have not undergone full journal peer review, and it is not under consideration for peer-reviewed publication elsewhere" Another caveat, from the technical side, in relation to comparing datasets from different experiments is that, even when all methods were the same, our two sets of samples were not sequenced in the same flow cell run, which per se brings about a possible batch effect that could introduce some variation bias. The same problem applies in fact to any attempt to compare data with those existing in literature which were obtained by either one's own research or studies from other authors. Nonetheless, we had quoted the key aspects of our prior paper and commented the consistent points as follows: "In our prior work (Rosselli et al., 2015) we had studied community composition in the same Sardinian stations in a short period of winter (in late February) during and after a single dust-carrying event. In that study the main feature evidenced was the existence of a conserved core microbiome, encompassing 86-95 % of the taxa, to which the incoming dust would cause some detectable diversity variation but on a rather limited proportional scale. Such minor effect of the dust-lifting storms observed in winter is in fact confirmed in the present work in which the time of the year factor appears as the variable of major order in shaping community structure and richness." And we have now added further discussion on that, as we will outline in the answers to the next queries.

The discussion presented in the present paper has to be consistent with the one published earlier to be really acceptable and sound. For instance, the present paper (Rosselli et al. 2020, Table 3, Page 17) clearly shows that dust events have little influence of the biodiversity indexes (Simpson 1-D, Shannon H), this is quite contradictory with what is claimed in Rosselli et al. 2015 (Table 2, Page 4).

ANSWER: In the present paper, as regards the diversity or evenness indexes, the only differences which have been remarked and claimed as statistically significant are the

C3

ones across the two seasons (Formerly Fig.5B and Fig 5C and currently Fig S9B and S9C, Supplementary); having respectively $p < 0.00584$ and $p < 0.026$. As regards the same indexes, Table 3 of the present manuscript lists the values of each sampling but not the statistical significance comparison among each. In that case the dust vs. control condition did not result significantly differing. If this outcome has to be compared with the prior report, it needs to be noticed that, in that case the data shown in Table 2 of that publication, were instead not subjected to statistical significance analysis and the apparent difference of higher diversity in relation to the first and single dust outbreak analyzed at that time, in one of the localities (Cagliari) was even showing opposite results when using Shannon or Simpson indexes (increasing in one case and decreasing with the other). Therefore: 1) The present data cannot be considered in disagreement with the prior report as in that case the statistical analysis on those ecological indexes values fluctuations had not been performed; 2) The present data are instead backed up by appropriate analyses, repeated for two independent dust outbreaks, and consequently deemed as much more robust.

Also this dust influence is illustrated in PCA and Cluster dendrograms presented Figure 4 (Page 5) and 5 (Page 6) of the results published in 2015 (Rosselli et al.), again this is contradictory with figures 4, 5 and 6 (Pages 18, 19 and 21 respectively) of the present manuscript. Could the authors comment on these results and possibly merge the data of the two papers with new statistical analyses integrating all the data. It is quite important to clarify the influence of not of dust events on the microbiome compositions.

ANSWER: In the 2015 paper the PCA (which was the only multivariate analysis in that occasion) and the dendrogram, had been plotted by ordering data according to the on Euclidean distance, with average linkage method based on the identified genera. Notice also that in that old paper the percentage of variation explained by each axis was not shown (and it was actually rather low). In the current manuscript instead, only the cluster analysis (Fig.3) made use of that method. While for multivariate analyses

C4

(object of this query): a) the PCA was obtained using the Bray-Curtis distances which are much more appropriate than the Euclidean one in multivariate ordinations when dealing with data in which some taxa feature zero presence in some samplings. b) the percent of explained variability of the two axes shown is reported, amounting to 35%+14% for the PCA of Fig. 4, and to 54% +14% for the PCoA of Fig. 5. (c) raw data had been transformed by the combination of total sum scaling with square root transformation which is the renowned 'Hellinger transformation'. Such transformation has been praised as a preferable choice in ecological community comparisons, (Legendre and Legendre, 1998, Numerical ecology, 2nd English edn. Elsevier, Amsterdam) as it offers the best trade-off between linearity and resolution in comparison to chi-square metrics and other approaches. It is also recognized as more balanced for the weight given to rare species. For this reason it is also the first recommended choice of data transformation in the Calypso webtool suite (Zakrzewski, et al., Bioinformatics 33, 782–783,2017), which has to date used and cited by 292 articles indexed in Web of Science. As regards figs 5 and 6 of the present manuscript, those analyses (Principal Coordinate Analysis and Discriminant Analysis of the principal components) had not been run in the Rosselli et al., 2015 published data and the comparison for their outcome is not feasible. In essence we view the present data as much more reliable than those stemmed from the previous report and at the same time not showing conflicts of compliance with them due to the more limited methods used in the prior paper and the absence of some statistical significance tests thereby used.

Another point concerns the evidence of a “major conserved core microbiome” that could be considered as a “global Sardinian air microbiome” (Figure 3 Page 5, Figure 5 Page 6, discussion Page 6) published in Rosselli et al. (2015). Again it would be very interesting to merge the data obtained in 2020 and 2015 to confirm the presence of such a conserved core microbiome. Could the authors make this analysis with the integrated data.

ANSWER: while, as mentioned, we could not in the first place use the published cam-

C5

paign's analyses due to the journal's policy, there are other important considerations that can be made to reconcile any apparent doubt between these two reports and we are grateful to the reviewer for invoking this clarification which leads to this addition to the manuscript's text in the discussion section, commenting on the full consistency between the two reports: “In that prior analysis of ours the existence of a common core microbiome of the investigated area was one of the suggested evidences. That concept was stemming from the analysis run in February, therefore towards the end of a winter period throughout which Europe experiences its minima in terms of temperature-driven air turbulence events and as consequence receives more limited influxes of air travelling from seas to land. In the present analysis, we observe that, in spite of the major changes brought about by the temporal factor, the two sampling stations at opposite corners of the 270 km-long island shared the closest level of community composition when they were compared at the same time (see CA Ctrl vs. SS Ctrl in Fig.3 Fig.4, Fig. S9, and Table 5). Moreover, this similarity was maintained in May even though the two control stations were compared after the dust outbreak. Foremost, those two distant sites achieved the maximum of community overlap in September, when controls were compared right before the next outbreak, after a 109 days-long period without such events. During that time the air microbiome of the whole area appears to have changed profoundly, but in a concerted fashion, leading to a high uniformity across the island. These data confirm the view of the prevalence of a core microbiome, as emerged in our 2015 report and add the evidence that such extended core community undergoes also a temporally related concerted turnover. Whether or not this could be also a seasonal (regularly recurrent) phenomenon, will have to be demonstrated by further research on this subject.”

Finally, in this paper (2020), Rosselli et al. efficiently exploit wind rose graphs integrating wind speed and direction, temperature and humidity (Figures SM9, 10, 11, 12, 13, 14) to explain some of their results. Actually these wind rose graphs are presented for March, April, May, June, July, August, September, October and November. Unfortunately, data are not presented for February, a time period of interest for the ex-

C6

periments reported in Rosselli et al. (2015). Could the authors add these data and comment about the results of 2015.

ANSWER: The plot for the quadrant of February, also in line with the above new paragraph, is actually rather similar to that of March as during the December through March winter months southern Europe as well as general Europe air circulations and corresponding climate tend to be relatively stationary. Therefore, we have checked, but we find that the February 2014 image would not actually add particular interpretive clues to the big picture. Actually, we find that a much more predictive comparison element is the day air mass backward trajectories calculated by the NOAA HYSPLIT model shown in Fig. 2 of the present manuscript and in Fig. 1 of the prior paper.

Other comments It would be interesting to give the total number of cells present in the various samples as it is also a very important indicator describing the air microbiome. Does this number differ depending on the situations (sampling site, season, wind..)?

ANSWER: The number of microbial propagules, cells, clusters of them, spores or other quantitative aspects of this type is not possible to assess as, even under microscopy-aided operations as it would imply counting objects which are not visually distinguishable from their background of particulate dust and any non-biotic debris. Even under epifluorescence by vital stains, the problem lies in the stratified layers of target and non-target objects that physically mask the visibility of the former and do not allow the unobstructed two-dimensional counting that would be required for such task.

Fig S1 and S4: the authors present data concerning the amount of PM10, do they have data on PM 2.5? It would be interesting to look at them and see if there is a variation of their concentrations with the seasons, locations, dust events ...etc.

ANSWER: PM10 is in fact including also PM 2.5 as the class is measured as <10 μm diameter size. The choice of the former is that the latter would not encompass the majority of airborne microbial cells and PM10 is therefore deemed as more informative for such information.

C7

The authors refer to the importance of the “daytime height of the planetary boundary layer over Europe “ (Page 26, line 543). This is indeed an important factor that can shape the air microbiome. Do you have data on the height of the boundary layer at the sampling sites and during the air mass trajectories? It would be of great interest to add it to this manuscript and take it into account.

ANSWER: That value is not a uniform measure that could be easily taken nor officially found for a defined given area as it fluctuates continuously in relation to the uprise of the sun towards the zenith and it is affected at local microscale by factors as scattered clouds resulting in ground shadows that are receiving less radiation and so on. Moreover, in the context of the interpretation given in the text we are not referring to the condition occurring just during the days of air sampling but throughout the whole mid-to-late summer period which is however always bound to result in averagely higher values of hot air convective turbulence in comparison to the colder spring times. The comment was essentially about the observation that late summer in Southern Europe constitutively carries every year higher loads of uplifted microbial cells.

Finally, I found some mistakes in the citation of the references:

*Gleick et al (1993, Page 2 line 48) and Shine et al. (2000, Page 3 line 69) are not in the reference list. *Page 3 line 64 “Polymenakou “(and not “Polimenakou”). *Some references in the list are not cited in the text: Harland et al, 2008 (page 34, line 748) Koenig et al, 2010 (page33 line 706) Kramer et al, 2006 (Page 35, line 768) Latif et al, 2014 (Page 35, line771) Poschl, 2006 (page 37, line 808) Shao et al, 2011 (page 38, line 834) Shiklomanov et al, 1993 (page 38, line 837) Wainwright et al, 2003 (page 39, line 867)

ANSWER: We thank you for the careful attention. The two missing references have been added and the errors fixed. Koenig was not a reference on its own but a carriage return line of the Caporaso et al.