Decision: Major revision advised

Referee: 1

Comments to the Author

In this article, Welsh and Eisenhofer explore a critical question in Phyllosphere and, more widely, microbial community studies: the lack of control use. The recent explosion of phyllosphere studies in the two last decades, resulting from revolution in the sensibility of sequencing methods and analysis, now allow to – finally – study properly this unfairly qualified "poor" environment. The major issue is the usually low DNA yield obtained after leaf sample processing and DNA extraction, challenging the proper characterization of an amazing and underestimated diversity, and the risk to consider traces of contaminant DNA as true Phyllosphere inhabitants. The use of negative and positive controls (mock and/or spike in) should be indeed systematized in microbial community studies, and these controls must pass through all the processes (sampling, DNA extraction, PCR amplification, sequencing, bioinformatic analysis and data availability). I can't be more agree with authors useful guidelines. Although supported by a nice meta-analysis on recent phyllosphere papers, I have however few concerns regarding the methodology and potential biases in this study. See my minor comments below

L87 – There are 3 papers from 2023 listed in your supp. table 1
In this case these articles are listed as being published in 2023, however, were available and published online in 2022, this allowed them to be found using our search criteria and thus would also be found in replicated searches. For this reason we decided they did not warrant exclusion and kept them in the analysis. We have adjusted this to line (L88) to account for this as well as added a note to supp. table 1 explaining these dates.

L 100 – It is unclear what the 2 reviewers did; were they in charge of screening papers in this meta-analysis? Please clarify
Both authors were in charge of screening papers and harvesting data. We have further clarified this in the methods on LNs 108 & 112.

L120 – China and the US represent 60% of papers reviewed here. This is not surprising and commonly observed in research. Does this bias correlate with your observations and would you be able to correct for them? Also I don't have access to the detailled list of studies but for instance, you have 6 studies coming from Canada and 5 of them come from people from or having worked in the same lab. Would you

consider including a bias correction against obvious correlations between methodology/practice and country/lab of origin?
We didn't wish to point fingers at any particular country, and additionally, such an analysis would be underpowered due to the highly uneven sample sizes. Rather, the goal of our study is to highlight these issues across phyllosphere research as a whole, as such, if certain countries or labs produce more data for the field (phyllosphere microbiome) this should be reflected within our results and reported on as such.

L127 - Only 3 studies, of which 2 come from the same lab... this is a strong bias (see my previous comment). Also, I am myself biased because, of course, I checked if a couple articles I knew were present in your study. One, although passing all your criteria, was not cited, and this study came from the same lab, and included both positive and negative controls. This is just an example but then I wonder how many other studies did not pass the filters and why.
We don't believe that studies coming from the same lab represent a bias, as the goal of this methodological review is to survey control usage across the phyllosphere research as a whole.
As with any systematic review surveying hundreds of published papers, a few may be missed. In such a case, we don't believe that this will impact the major findings and message of our paper — that control usage in phyllosphere research is currently underutilised and in need of improvement.

L187 – Systematic randomization of samples at different steps (processing, extracting, PCR, sequencing…) is also a good and complementary way to mitigate the effect of contamination (and very useful when you don't have controls)
This is an excellent point, and we have incorporated it into the maintext: LN=212

L234 – 32/143 ? Or maybe your excluded papers from 2023? Please clarify
Thank you for pointing that out. We did already tussle with the inclusion of the 2023 papers, however, determined to keep them as they were successfully retrieved by the search criteria and did not warrant exclusion. This is definitely an artifact of us potentially removing those three articles and has been adjusted to fit our final decision on this issue.

Figure 5: Why don't replacing number by a curve? I'd like to see positive controls in this figure. Is the observed trend for blue and red line significative? Anyway those trends have to be corrected by the total number of papers per year because it also increases with time (2-9 per year until 2019, then 17 in 2020, 30 in 2021 and 46 in 2022).
We have adjusted this figure to use proportions instead of counts as suggested, and have also included positive control usage.

Referee: 2

Comments to the Author

Throughout the manuscript the author has performed a systematic analysis and made technical points of improving quality standard that is dealt with contaminations for phyllosphere microbiome data. The provided strategy in the manuscript can be applied to nix poor quality of the phyllosphere microbiome data. While the author nicely described the major concern, there are additional comments to have rooms for improvements of the manuscript.

Line 35: Please address the surface of a plant's shoots to the epiphytic compartment on phyllosphere

Done.

Line 36-38: Phyllosphere research gains an acess to understandings in several disciplines including plant microbiota, and plant physiology, in particular plant immunity, please mention and elaborate this aspect:

For plant microbiota

1) Vorholt JA. Microbial life in the phyllosphere. Nat Rev Microbiol. 2012 Dec;10(12):828-40. doi: 10.1038/nrmicro2910. PMID: 23154261.

For plant immunity

2) Sara Shakir, Syed Shan-e-Ali Zaidi, Franciska T. de Vries, Shahid Mansoor, Plant Genetic Networks Shaping Phyllosphere Microbial Community, Trends in Genetics, Volume 37, Issue 4, 2021.

Added.

Line 42: Please note that the contaminant DNA (for phyllosphere microbial profiling) is not only derived from external DNA materials, but also mitochondrial and chloroplast DNA from plants. Importantly, 16S rRNA (plastid and mitochondria) originating from plant tissues can introduce >80% of the standard amplicon sequences, and this is a high risk of a contamination that significantly interferes microbial composition and can be considered as an internal contamination, please mention and elaborate this aspect in the review:

1) Bulgarelli, D., Rott, M., Schlaeppi, K. et al. Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. Nature 488, 91–95 (2012). https://doi.org/10.1038/nature11336

2) Lundberg, D., Lebeis, S., Paredes, S. et al. Defining the core Arabidopsis thaliana root microbiome.Nature 488, 86–90 (2012). https://doi.org/10.1038/nature11237

We thank the reviewer for bringing up this very important point. We have expanded on this in the manuscript — principally in relation to how it can confound the estimation of the limit of detection. LN: 227

Line 54: Please indicate what organisms or plant compartments that the author compared with. It is unclear to me, because a low amount of biomass is prone to have DNA contamination, especially a case that microbial contaminants with a low titer can contribute this problem on phyllosphere, but not in soil and root.
This study surveyed the literature of the plant phyllosphere, specifically, the epiphytic compartment.

Line 92-93: The author harvested research phyllosphere microbiome studies from 2021 and 2022 across database by using keywords: phyllosphere OR epiphytic AND microbiome OR microbial communities OR fungal communities OR bacterial communities OR communities. Finally, the author obtained 143 papers. Although the kappa statistic indicates that the selected publications for this “negative control” assessment is unbiased from the applied keywords. But there are many missing parameters to be considered here, and it seems that this step is a bottleneck step to harvest publications.
We spent a lot of time carefully choosing and testing our keywords in order to capture as many relevant studies (16S rRNA gene amplicon sequencing of phyllosphere/epiphytic communities) as possible. Reviewers 1 & 3 had no issue with this aspect. Additionally, the kappa statistic did not assess the biases of the applied keywords, however, it reports on the two reviewers' ability to rate articles independently with significant agreeance (this has been further clarified on line 108). We agree that the search thread acts as a ‘bottleneck’, as such we did work on producing a search thread that would be both broad enough to encapsulate a wide array of phyllosphere research, but specific enough that target methodologies could be found (the broad nature of the search can be seen by the large number of rejected articles (n = 307)). This has been reiterated in the text on Line 91.

As far as I understand the amplicon sequencing offers a highly sensitive detection to determine microbial composition both from epiphytic and endophytic compartment of the plants. This has been yet challenging to precisely determine epiphytic microbial community alone. By using previously reported bacterial genus as a common contaminant on epiphyte is problematic with several reasons:
1). First, the author cannot precisely dissect what type of bacterial communities, specifically Sphingomonas, and Pseudomonas can be a potential contaminant, because most of the genera are considered as a core microbiota member (Bai, Y., Müller, D., Srinivas, G. et al. Functional overlap of the Arabidopsis leaf and root microbiota. Nature 528, 364–369 (2015). https://doi.org/10.1038/nature16192). For example, Sphingomonas and Pseudomonas are a core member of phyllosphere microbiota in Arabidopsis at the intrastrain level where the amplicon sequencing alone cannot resolve those closely related strain due to bacterial 16S polymorphism.

2). Secondly, Sphingomonas seems to appear almost everywhere on leaf of many plant species and in soil with its ability to colonize plants (Lundberg, D., de Pedro Jové, R., Pramoj Na Ayutthaya, P. et al. Contrasting patterns for microbial dominance in the Arabidopsis thaliana phyllosphere PNAS (2022). https://doi.org/10.1073/pnas.221188111), but Pseudomonas seems to specifically appear on Brassicaceae plants. Sphingomonas can also play an ecological role in a context of non-mutualistic relationship with plants. In contrast, Pseudomonas have been demonstrated at strain level to possess Pseudomonad-Arabidopsis interacitons in gnotobiotic systems (Karasov TL, Almario J, Friedemann C, Ding W, Giolai M, Heavens D, Kersten S, Lundberg DS, Neumann M, Regalado J, Neher RA, Kemen E, Weigel D. Arabidopsis thaliana and Pseudomonas Pathogens Exhibit Stable Associations over Evolutionary Timescales. Cell Host Microbe. 2018 Jul 11;24(1):168-179.e4. doi: 10.1016/j.chom.2018.06.011. PMID: 30001519; PMCID: PMC6054916. & Shalev, O., Karasov, T.L., Lundberg, D.S. et al. Commensal Pseudomonas strains facilitate protective response against pathogens in the host plant. Nat Ecol Evol 6, 383–396 (2022). https://doi.org/10.1038/s41559-022-01673-7). With these observations, it seems to me that is not clear to consider those genus as a contaminant on epiphyte.
The reviewer highlights an important issue that we briefly mention in our study — some bacteria that have been associated with plants (e.g. *Pseudomonas*, *Sphingomonas*) have also been highly prevalent in the negative controls of other studies. We never claimed that such genera are contaminants, and we actually mentioned that *Pseudomonas syringae* is a well-studied plant pathogen (see paragraph from LN:63). We think this point underscores the importance of the community employing more rigorous controls — particularly due to the limited resolution of amplicon sequencing as mentioned by the reviewer.

3). However, to support the assumption made by the author, I would suggest after the removal of possible bacterial contaminants, one could calculate beta-diversity index or Bray-Curtis dissimilarity across all sources from selected publications with an outgroup to determine whether the amplicon sequences after the genus removal give the similar or different index as a control analysis.
This is a very interesting idea worth pursuing. However, given that our paper is not about whether these genera are contaminants or not, we believe it beyond the scope of what we can include.

4). There are some publications using a term such as ITS1 rRNA from Regalado, J., Lundberg, D.S., Deusch, O. et al. Combining whole-genome shotgun sequencing and rRNA gene amplicon analyses to improve detection of microbe–microbe interaction networks in plant leaves. ISME J 14, 2116–2130 (2020). https://doi.org/10.1038/s41396-020-0665-8. Please also take this example into account of your searching criterion.

Our search criteria did return papers containing ITS1, but it seems not this one that the reviewer pointed out. Given the nature of systematic reviewers (and the scale of the literature surveyed), there are always some papers that will be missed. We have hesitations about adding papers post-hoc to our analyses (it's counter to the concept of reproducibility in systematic reviewers), however, we think that our search has captured most of the 16S/ITS phyllosphere studies to date.

5). As mentioned in Line 52, the major concern via 16S rRNA or ITS1 rRNA method is mitochondrial and chloroplast DNA that interrupts microbiota composition determinations in plants (Lundberg, D., Yourstone, S., Mieczkowski, P. et al. Practical innovations for high-throughput amplicon sequencing. Nat Methods 10, 999–1002 (2013). https://doi.org/10.1038/nmeth.2634 and Mayer, T., Mari, A., Almario, J., Murillo-Roos, M., Syed M. Abdullah, H., Dombrowski, N., Hacquard, S., Kemen, E.M. and Agler, M.T. (2021), Obtaining deeper insights into microbiome diversity using a simple method to block host and nontargets in amplicon sequencing. Mol Ecol Resour, 21: 1952-1965. https://doi.org/10.1111/1755-0998.13408). Two major research has tackled this problem by creating a small molecule to block both mitochondrial and chloroplast amplification in plants. The first one developed peptide nucleic acid (namely clamp) to anneal specifically at V3V4 region (with use of 341F – 799R 16S), and the second one developed blocking oligo primers that block V3V4 (341F – 799R) and V5V6V7 (with use of 799F – 1192R). Please consider this development and elaborate the blocking issue in the manuscript.
This is a great suggestion, and we have added it to our recommendations; LN: 227.

Because of this reason, in supplementary table 1 the author should revise the strategy to select publications that applied the blocking in their protocol and compare with publications that did not also apply the blocking. Also, the author should check an impact on the blocking to increase accuracy for the compositional plant microbiota by statistical comparisons (similar to 3) section), I think this can be one of the major concerns for the section "RECCOMMENDATIONS FOR FUTURE PHYLLOSPHERE RESEARCH".
We have added this important point to our main recommendations in the RECOMMENDATIONS FOR FUTURE PHYLLOSPHERE RESEARCH section (LN: 227). However, we think that both re-reviewing the literature for this issue, and performing statistical comparisons on plant microbiota compositions is well beyond the scope of the paper, and would dilute its core message (exogenous DNA contamination).

Line 347-348: Figure 3 is not insightful, in my opinion, and can be improved by providing lists of bacterial genus contaminants, laboratory sources (whether the similar contaminations came from the same or different laboratory), and regions of 16S rRNA that were amplified, this can provide whether the potential contamination

has a systematic or random error made by possible causes of the contaminations offered by the author.
We agree that learning more about the distribution of contaminant taxa would be insightful. However, we chose not to go down this path for multiple reasons.
1) we noticed that the reporting of taxonomic compositions is highly heterogeneous from study to study (some report only genera, others only families, other phyla, etc.)
2) studies vary substantially in how they annotate taxonomy (and the databases employed; e.g. SILVA, greengenes, GTDB)
3) not all data (or metadata) is publicly available (as we identified in the main text), and the standardized processing of all this data would be a monumental task
The reviewer raises a very interesting idea, but we believe it would be better explored in a separate, more focused paper.

I hope the author finds my comments will be useful, and please address the issue. Thank you very much.
We sincerely thank the reviewer for their constructive comments, which we believe have strengthened the manuscript further.

Referee: 3

Comments to the Author

This is a timely article. The inclusion of controls is clearly important to ensure microbiome studies of low biomass environments, including the phyllosphere are sound. The paper presents a systematic review which examines evidence for inclusion of both negative and positive controls in selected literature since 2012. The way in which the review has been conducted is robust, and the data is discussed nicely in the context of the different controls which have been used, and recommendations for controls to include in future studies are provided. There is a brief narrative on a number of differences in sample collection and processing that the authors have identified- this is a little limited but still identifies some important issues (esp >20 % of studies with no deposited raw sequence files). I have a few comments the authors may wish to consider

1. The authors identify no use of controls in the majority of studies. However I think this underestimates actual use of controls. As co-author of one of the studies which formed the basis for the evidence presented, and good knowledge of a second, I can confirm that negative controls (extraction and PCR) were used in both studies, but this was not reported in the paper. Instead these controls were used at an early stage in the project as a routine activity to confirm data was robust and not subject to contamination. This reality should be recognised. Part of the issue is therefore reporting, rather than actual use of controls, and this should be acknowledged and the message through the article should be about reporting rather than solely use of controls. The recommendation is therefore that controls should not only be done but the date generated in the controls and way in which this data has been used for quality control should also be reported in the paper, together with deposition of contaminant sequences.

This is an excellent point and suggestion, and we have made adjustments to the paper to address it. See LN: 135, 244 and 276

2. Fig 5. Trend here does look like increased use of both controls since 2012 ie counter to text suggestion at line 145-6- it may be this date needs to be corrected to proportions of studies rather than number of studies.

We have adjusted figure 5 to use proportions, rather than counts to account for the number of studies.