The study of Mouton et al focused on the bacterial community associated with the plant pest *Bemisia tabaci* in Bukina Faso. Combining Illumina metabarcoding, Sanger sequencing and quantitative PCR, they report for the first time the presence and high prevalence of *Candidatus* Hemipteriphilus asiaticus in African samples. They also quantified the prevalence and co-occurrence of seven bacterial symbionts in 334 insect samples encompassing 4 biotypes.

Overall the article is well written and the experimental design is sound. This study provides new information about the distribution and diversity of *Candidatus* Hemipteriphilus asiaticus, a bacterial symbiont relatively under-studied so far.

However, I believe that the quality of this manuscript could be improved (and get a bigger impact) by improving the analysis of the data (see my comments below, especially about the phylogeny and the visualization of the metabarcoding data).

I also invite the authors to follow the International Code of Nomenclature of Prokaryotes (https://doi.org/10.1099/ijsem.0.000778) and to revise the manuscript accordingly (use of *Candidatus*, use of italicized characters only when appropriate, etc...). Moreover I found it very misleading that the authors used the bacterial genus to mention bacterial strain or species.

You will find more specific details and suggestions below.

line 73: Genome size is estimated for a strain, not for a genus. So you should replace *Portiera* by *P. aleyrodidarum*. Similar to *Hemipteriphilus asiaticus*, you might also want to speak about *Candidatus* Portiera aleyrodidarum since there is no isolate in a reference collection.

Lines 78-79: sp. should not be italicized. Please correct throughout the entire manuscript.

Lines 84-85: Here again I believe you are mentioning strains and not genera. If so, please correct.

Line 92: Is there an isolated strain of *Hemipteriphilus asiaticus* deposited in a reference collection? If not, it should be called *Candidatus* Hemipteriphilus asiaticus. This would apply for the entire manuscript, including the title.

I also believe that it is misleading and inaccurate to use the genus name to mention a species, especially since the present study reveals the existence of different members within this genus.

Line 96: If I understand correctly, you used a metabarcoding (PCR-based) approach, not a metagenomic (PCR free) approach. Please correct throughout the entire manuscript.

Line 127: The absence of *Candidatus* Fritschea sp. could also be explained by the specificity of your primers (a quick analysis on the SILVA database reveals only 80% coverage and 30% specificity for *Candidatus* Fritschea with your primers).

Line 133: I found it a bit frustrating that this section is not illustrated with a figure. Maybe a barplot showing the relative abundance of the major taxa in the different host genetic groups could be useful to support your findings.

Line 141: By convention, if you use *Candidatus* Hemipteriphilus asiaticus, only “Candidatus” should be italicized. Please correct throughout the entire manuscript.

Line 141: Can you please provide values for these similarities? Was it based on a blast search? Can you also provide the date of this online analysis?

Line 149 and line 157: With your data, is it possible to infer species or strain delimitation? In other words, are we looking at new strains or new species?

Line 152: I could not find *Orientia tsutsumagushi* on your tree.

Line 208: "Damage" is an uncountable singular noun. "is huge and results in..". Please correct.

Line 220: Here and elsewhere, "metabarcoding" and not "metagenomic".

Line 243: Based on your current phylogenetic analysis, you are not confirming (or maybe I missed something) but your are showing/revealing the existence of polymorphism.

Line 277: "did not reveal any significant difference…"

Line 281: You used "Clearly" twice in three sentences. It is redundant.

Line 295: When I look at the Figure 3, I see stability but I also see a lot of variability. You should rephrase to improve clarity.

Line 308: I believe you referred here to Figure 2.

Line 322: I believe this is the primer 341F and not 319F.

Line 326: Please replace "*16SrDNA* gene" by 16S rRNA gene. Also note that by convention "16S rRNA" is never italicized when it refers to the gene. The same applies for the 23S rRNA gene (and 18S or 28S rRNA gene). Please correct throughout the entire manuscript, including the tables.

Line 328: Please provide more details about the amplification procedure (PCR conditions, volume of the reaction, …).

Line 341: I am not familiar with this specific classifier but since your amplicons encompassed the V3-V4 region (position 319F-805R), why did you use a reference database covering only the 515F/806R positions? Is there any chance that you missed some information or get an inaccurate classification?
On a side note, the naïve Bayesian classifier (from RDP) and BLAST usually provide more accurate classification (see fro instance, https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-018-0470-z). Lastly, even though it is still widely, Greengenes database is outdated (last update in 2013…). Nowadays, Silva is usually recommended for 16S rRNA gene classification.

Line 371: How did you choose this substitution model?

Line 375: If I add the values presented in Figure 3, I find 334 individuals, not 304. Please check where the mistake came from.

Line 394: Please provide the version of R.

Line 403-404: I highly appreciated to be able to access all the data on Dryad, with a clear and complete description of the files. However I believe that the Illumina data should submitted to an official repository such as the Sequence Read Archive (SRA) (https://www.ncbi.nlm.nih.gov/sra) and that a accession number should be provided in the manuscript.
One more thing about the Illumina data. I thank the reviewers for sharing the taxa-bar-plots.qzv file, it was useful to explore the data and it improves transparency. Looking at this file, I noticed that your sequences included some chloroplast sequences (classified as: k__Bacteria;p__Cyanobacteria;c__Chloroplast;). Although they are not abundant, those sequences still need to be removed from your analysis of the bacterial community.

Line 519: 32 and not 329.

Line 594: 2018 and not 2108

Figure 1: First, I could not find *Orientia tsutsumagushi* in your tree but you mentioned it in the legend.
Second, the use of the concatenated alignment of three individual loci rises the question of the congruence of these loci for the phylogenetic inference, mostly because these loci have different histories and evolutionary rates. Was this aspect evaluated in your analysis? If so, could you please document it?
Third, although this relatively simple phylogenetic tree is enough to classify your sequences, I believe that a more meticulous phylogenetic analysis could really increase the impact and the significance of this article. Indeed, this tree supports one of the main findings of the study.
Here are some suggestions:
- you could add more sequences in the tree, in order to have more representative members of the Rickettsieae family. Because various genomes are available, it should be possible to extract these 3 genes. This would allow you to maybe clarify the position of your sequences within this family.
- you could provide a consensus tree based on different methods (eg. parsimony, maximum likelihood, Bayesian inference...) with different estimations of the node robustness (eg. bootstraps, Approximate likelihood-ratio tests, aBayes…). Most of these analyses can now be easily done with online tools. Such analyses would help to better evaluate the robustness of your analysis.

Figure 3: I think that these “Mondrian plots” are an elegant and simple way to visualize co-occurrence species data. I just would like to make some suggestions to help the reader:
- The y-axis could be easier to read with more graduations, written horizontally and also with a title explaining what we are looking at.
- The x-axis could have the full name of the bacteria just by rotating the labels at 45°. This would improve the readability.

Figure 4: Do you have the same number of samples in both conditions? Maybe you could provide this information in the figure or at least in the legend of the figure.
One last suggestion here, maybe a log-scale on the y-axis would improve the data visualization.