From Laurence Mouton
Université Claude Bernard Lyon 1
43 Bd du 11 Novembre 1918
Bat G. Mendel
UMR CNRS 5558 Biométrie et
Biologie Evolutive
69622 Villeurbanne Cedex
France

To the Recommender
Dr. Yuval Gottlieb
PCI Zoology

Villeurbanne, France, 11 February 2022

Dear Dr. Gottlieb,

Thank you very much for giving us the possibility to submit a revised version of our manuscript entitled "Analyses of symbiotic bacterial communities in the plant pest *Bemisia tabaci* reveal high prevalence of *Hemipteriphilus asiaticus* on the African continent" by L. Mouton, H. Henri, R. Romba, Z. Belgaidi, O. Gnankiné and F. Vavre (https://doi.org/10.1101/2021.10.06.463217) to *PCI Zoology*. We are most grateful to the reviewers for their thoughtful comments. We have considered all their comments and details of our responses are listed below (in blue). Please also note that we have slightly changed the title of the manuscript following a suggestion of the second reviewer (see below point 5 referee 2).

On behalf of all the authors,

Laurence Mouton

The managing board of *PCI Zoology*

1) Data must be available to readers, either in the text or through an open data repository such as Zenodo (free), Dryad (to pay) or some other institutional repository. Data must be reusable, thus metadata or accompanying text must carefully describe the data;
Data have been deposited in Dryad (https://doi.org/10.5061/dryad.547d7wm91).

2) Details on quantitative analyses (e.g., data treatment and statistical scripts in R, bioinformatic pipeline scripts, etc.) and details concerning simulations (scripts, codes) must be available to readers in the text, as appendices, or through an open data repository, such as Zenodo, Dryad or some other institutional repository. The scripts or codes must be carefully described so that they can be reused;
Scripts have been deposited in Dryad (https://doi.org/10.5061/dryad.547d7wm91).

3) Details on experimental procedures must be available to readers in the text or as appendices;
All the experimental procedures are included in the main text.

4) Authors must have no financial conflict of interest relating to the article. The article must contain a "Conflict of interest disclosure" paragraph before the reference section containing this sentence: "The authors of this article declare that they have no financial conflict of interest with the content of this article.";
This paragraph has been added.

5) This disclosure has to be completed by a sentence indicating, if appropriate, that some of the authors are PCI recommenders: "XY is one of the PCI Zool recommenders.".
We added this sentence.

The recommender, Dr. Yuval Gottlieb

Here are the main comments that need your attention:

1. Correction of the terminology of the method and the nomenclature of the symbionts.
Done.

2. Better defining of your sample material, and the sample size.
Done.

3. Reanalysis/classification of the data according to SILVA and the primers used.
Done. For the classification, reads from the Silva 138 SSURef NR99 reference database were extracted to match on the primer set 341F/805R.

4. Deposit the sequencing data in an accessible and accepted format (supplementary and SRA).
The process to depose sequencing data on Sequence Read Archive (SRA) is ongoing. Moreover all the data are already available in Dryad (https://doi.org/10.5061/dryad.547d7wm91).

5. Improving the phylogeny and the symbiont consortium figures.

About the phylogeny :

- We checked that the topologies of the trees for the individual loci are similar to the one of the concatenated tree. It is now indicated in the text.
- In the first version of the MS, only the maximum-likelihood method was used to construct the phylogenetic tree. In the revised version we added the use of the Bayesian inference method and found the same topologies for the two trees which indicated the robustness of our analysis.
- We did not add more sequences from members of the Rickettsiaceae family as suggested by referee 2 because this phylogenetic analysis has already been done by Bing et al. (2013) (Bing XL, Yang J, Zchori-Fein E, Wang XW, Liu SS. 2013. Characterization of a newly discovered symbiont of the whitefly *Bemisia tabaci* (Hemiptera : Aleyrodidae). Appl Env Microbiol 79 : 569-575). As the main point of this phylogenetic analysis in the present paper was to show the variability between strains of *Ca.* Hemipteriphilus asiaticus in the *Bemisia tabaci* species complex, we think that it would drown out the information.

Referee 1

The manuscript by Mouton and colleagues explores the intracellular symbiotic community associated with Bemisia tabaci populations from Burkina Faso. The manuscript is well written clearly. Although the topic is very interesting, the study presents a few weaknesses that should be addressed before the manuscript is suitable for publication.

General comments:

The authors should use metabarcoding instead of metagenomics to avoid confusion.
Indeed, we made the changes.

Introduction:

The terms Primary symbiont (P-symbiont) and Secondary symbionts (S-symbiont) are more used with the meaning first and second than with their true meaning essential and not essential. This is particularly evident when the authors present nutritional symbiosis in cicadas and other auchenorrhyncha where the two symbionts (i.e. Sulcia and the co-symbiont) are both considered primary symbionts and are named co-primary symbionts. The authors should revise the introduction and the abstract to reflect that the terms primary and secondary are related to essentiality.
Mixing terms « facultative » and « secondary » may be confusing since some « secondary » symbionts may be, in some cases, « co-obligate ». We think that it is more clear if only one of these terms is employed, therefore we decided, as suggested by the third referee, to use the terms « primary » and « secondary » in the revised version, and deleted the notion of « facultative ».

The description of the characteristics of symbiosis in the introduction is specific to hemipteran. For instance, not all primary symbionts and intracellular (e.g., *Ishikawaella*, *Tachikawaea gelatinosa*; see lines 48-49). Moreover, not all endosymbionts have an extremely reduced genome (e.g., Sodalis pierantonius, Sodalis-like symbiont of spittlebugs). The author could generalize their introduction or specify that they are talking about symbioses in hemipterans.
Indeed, we added the term « hemipterans » (l.50).

Results

In the section Bacterial community characterization, the authors should add a figure to summarise the bacterial diversity associated with the different samples. They should also report within the paragraph how many samples were analyzed and the standard error associated with the average number of reads (line 120) since results are reported before the material and methods.

We added the number of individuals analyzed in the text (72 field samples) as well as the standard errors associated with the average numbers of reads. We also added a heatmap showing the relative abundance of the major taxa (Figure 1), and a taxa barplot done with all the sequences is available on Dryad (https://doi.org/10.5061/dryad.547d7wm91).

Lines 138-139: The authors should reference table 2 when mentioning the newly designed primers for the first time
Done.

Discussion

Line 224: replace "Since" with "Since then"
Done.

Materials and Methods

In the section "Sampling", it is unclear how the DNA extraction was carried out or if the DNA was already extracted during the work reported in Henri et al. If the DNA was extracted for the current work, the authors should report the extraction protocol.
All the DNA had been extracted in the study of Romba *et al.* 2018, therefore, for clarity, we deleted the sentence explaining the protocol used for extraction.

Line 308: Figure 1 should be Figure 2
Indeed, but we had another figure so the numbers of figures have changed.

In the section "Characterization of the bacterial community", the first paragraph should be split in two, separating the qPCR part from the metabarcoding part. In addition, the authors should also report exactly how many samples were analyzed using the metabarcoding approach and if the samples were pooled or not.
Indeed, the qPCR part should not be in this paragraph, it was a mistake, we deleted this part. Characterization of the bacterial community was done on individuals and not pools of individuals. We added this information in the revised version of the MS. The total number of individuals analyzed is indicated (72 field individuals plus 18 individuals from laboratory lines).

Line 375: "Three hundred four" should be corrected to "Three hundred and four"
Yes and we made a mistake: it was three hundred and thirty four. It has been corrected (l. 170).

The authors report some specific statistical analysis for which I do not have a sufficient background to comment on.

Referee 2

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.

We changed the nomenclature as suggested.

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.

We replaced *Portiera* by *Candidatus* Portiera aleyrodidarum in all the manuscript.

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

Done.

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

Here we mentioned genera, it is clarified in the revised version of the manuscript (l. 80).

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.

The referee is right. We changed *Hemipteriphilus asiaticus* by *Candidatus* Hemipteriphilus asiaticus in the entire MS including the title.

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.

Indeed, we made the changes.

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).

The fact that we did not find *Candidatus* Fritschea sp. is not surprising since this symbiont has been rarely found except in MEAM1, New World and MED species, and, in the worldwide distributed MED species, it has been detected only in China yet (Kanakala & Ghanim, 2019 : ref 35). Moreover, in a previous survey done in West Africa (Gnankiné *et al.*, IDCV, 2013 : ref. 27), although the presence of *Candidatus* Fritschea sp. was searched using specific PCR primers, it was not found. We added this information in the revised version (l. 133).

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.

We added a heatmap showing the relative abundance of the major taxa in the revised version of the MS (Figure 1) and a taxa barplot done with all the data is available on Dryad (https://doi.org/10.5061/dryad.547d7wm91).

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

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?
Information has been added.

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?
No.

Line 152: I could not find *Orientia tsutsumagushi* on your tree.
Indeed. We deleted *Orientia tsutsumagushi* in the text.

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

Line 220: Here and elsewhere, "metabarcoding" and not "metagenomic".
We made the changes everywhere in the MS.

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.
Indeed, we changed the term « confirm » by « reveal ».

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

Line 281: You used "Clearly" twice in three sentences. It is redundant.
We deleted the second « indeed ».

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.
We rephrased by: "In summary, these data confirmed the variability of the symbiotic community in the *B. tabaci* complex species despite its high temporal stability in populations from Burkina Faso, and reveal the presence of another player whose role deserves to be studied." (l. 311-313).

Line 308: I believe you referred here to Figure 2.
Indeed, we changed the number of the figure but we added a figure thus numbers of figures have changed.

Line 322: I believe this is the primer 341F and not 319F.
Indeed, it was a mistake, thank you, we changed the name of the primer.

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.
It has been done.

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

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.
In the revised version, we used, as suggested, the Silva 138 SSURef NR99 reference database covering the primer set 341F/805R used to obtain the amplicons (since 319F was a mistake, see comment above).

Line 371: How did you choose this substitution model?
We used the « model Testing tool » of CLC DNA Workbench 8.0. We have added this information in the revised MS (l. 396).

Line 375: If I add the values presented in Figure 3, I find 334 individuals, not 304. Please check where the mistake came from.
Thank you! We forgot the word « thirty » in the text. Symbiont composition have been determined for 334 individuals.

Line 394: Please provide the version of R.
Done.

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.
The process to depose sequencing data on Sequence Read Archive (SRA) is ongoing.

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.
Chloroplast sequences have now been removed from the analysis.

Line 519: 32 and not 329.
Indeed, we made the change.

Line 594: 2018 and not 2108

Indeed, we made the change.

Figure 1: First, I could not find *Orientia tsutsumagushi* in your tree but you mentioned it in the legend.

Indeed, we deleted *Orientia tsutsumagushi* 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?

Topologies of the trees for *GroEL* and *GltA* are similar to the one of the concatenated tree (16srRNA is not informative since the 483bp sequence of *Ca.* Hemipteriphilus asiaticus is 100% similar to Rickettsia). We added this information in the revised version (l. 156-158)

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.

This phylogenetic analysis has already been done by Bing et al. (2013) (Bing XL, Yang J, Zchori-Fein E, Wang XW, Liu SS. 2013. Characterization of a newly discovered symbiont of the whitefly *Bemisia tabaci* (Hemiptera : Aleyrodidae). Appl Env Microbiol 79 : 569-575). We think that, in the present study, it would drown out the information since the main point of this figure 1 is to show the variability between strains of *Ca.* Hemipteriphilus asiaticus in the *Bemisia tabaci* species complex.

- 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.

We constructed a tree using Bayesian inference and found the same topologies as for the tree based on the maximum likelihood method. These two analyses have been indicated in the revised version but only the ML tree has been provided (l. 397-400).

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 percentages are now written horizontally (now Figure 4). Unfortunately, the tool does not allow adding more graduations on the plots.

- The x-axis could have the full name of the bacteria just by rotating the labels at 45°. This would improve the readability.

We agree and we tried to indicate the full names of the bacteria. However, it takes a lot of space and it overloads the figure. Therefore we decided to let letters for references to the bacteria in the figure.

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.

Numbers were not the same for the two modalities and we indicated the number of samples in the legend of the figure in the revised version. We tried to draw the box plot with a log scale but it didn't improve the data visualization. On the other hand, the width of the box plots now reflects the number of samples for each modality (Figure 5 in the revised version).

Review F. Renoz

The Mouton et al. study provides a complete picture of the bacterial symbiotic community associated with the *B. tabaci* complex. For this they used a metagenomic approach complemented by a more specific diagnostic on a wider range of samples. The authors point out in particular the presence of *Hemipteriphilus*, a secondary symbiont recently detected in Asian populations and whose biological significance in *B. tabaci* remains unknown. I really enjoyed reading this study. It is very well written and the ideas are clearly established. The weakness of the study is that it is quite descriptive and we are left wanting more. But it has the merit of being technically sound and of establishing new perspectives. I have no major comments to make and I find the study suitable for publication. I do, however, have some minor comments and suggestions.

I find it inappropriate to see in the same idea "facultative" symbiont (L61) and "co-obligate" symbiont ("S-symbionts that complement the metabolic network" L64). For me it is not the same "symbiotic object", not the same evolutionary history. Or it should be clearly mentioned that there are secondary or additional symbionts in *Bemisia* and that some are co-obligate partners. But don't use the term facultative (co-obligate symbionts are additional but not facultative in essence). The terms "secondary" and "facultative" are often used as synonyms. But in my opinion, it would be more relevant to use "secondary" for "additional" and "facultative" when these additional symbionts are not mandatory for the development of the host.
We agree and we used the term « primary » and « secondary » in the revised version, and deleted the notion of « facultative ».

Concerning *Hemipteriphilus*, I think it is a pity that the authors did not try to determine the tissue and cellular tropism of this symbiont. Indeed, since the biological significance of the presence of this bacterium is still unknown in *Bemisia*, determining whether it is present in bacteriocytes and its location in relation to the obligate symbiont *Portiera* could provide valuable information on its status. Indeed, several studies show that in di-symbiotic systems where symbionts complement each other metabolically, the bacteria show a very close physical proximity in the bacteriome with highly nested bacteriocytes [1–7] (and sometimes a nested location with a symbiont living in another symbiont [8]). In my opinion the *Bemisia* model is super interesting to tackle these developmental aspects. If the authors still have samples of whole individuals, it would be worthwhile to examine them (at least for the MED-Q1 biotype).
We fully agree! But, unfortunately, it is not possible because the individuals were collected in the field in Burkina Faso and they were not conserved in a way compatible with FISH studies.

 A comment for further works: I also find that analyzing the density of the obligate symbiont in relation to the presence of additional symbionts is a very good idea. This is an aspect that is missing in many studies and that allows to eventually show the impact of the additional symbiont on the ancestral primary symbiont. Such data along with a fine analysis of tissue tropism by FISH during host development could shed light on many aspects in these multi-partner interactions.

L285: delete "clearly": the word is already used L278.
Done.

L313: "Mediterranean" and not "Meditteranean"
We made the change.

The authors used Greengenes. I don't have a problem with that (but is that database still updated?). Otherwise there are other databases like kraken2 and SILVA.
In the revised version we used the Silva 138 SSURef NR99 reference database as suggested.

I would have appreciated a presentation of the metagenomic results such as a table summarizing the sequencing data, the taxonomic assignment of the identified bacteria, a graph showing the proportion of reads for each identified taxa, the relative abundance of bacterial taxa in the form of a heat map. I think this data can be put in Supplemental Material (which I don't see here).
We added a heatmap showing the relative abundance of the major taxa in the revised version (Figure 1) and a taxa barplot done with all the data is available on Dryad (https://doi.org/10.5061/dryad.547d7wm91).