

Responses to Reviewer #1.

Specific comments:

Lines 118-120: Are there references available for these primers? If yes, please add the citation. References for these primers were included: (Klindworth et al. 2012; Caporaso et al. 2011). See line 142 in revised manuscript.

Line 125: "... sequences were processed using QIIME2 v.2020.8 (cite) as follows" please add a reference for QIIME2.

Reference for QIIME2 was added: (Bolyen et al. 2019). See line 148 in revised manuscript.

Line 126: Please add a reference for DADA2.

Reference added: (Callahan et al. 2016). See line 150 in revised manuscript.

Lines 129-132: It is not clear from the manuscript text and the tables (Table 3, S2-S6) how the ASVs were annotated. If the taxonomic assignment was based on searching the GreenGenes database what confidence thresholds were used? Which database do the classification data originate from (e.g. for family and order level designation)? What level of similarity/confidence thresholds were used for class, order, and genus level assignments? What if the best hits were contradictory such as in the cases of ASV 17 and ASV 18 in Table 3? Were these treated as 'unclassified' ASVs? For instance, one of the closest hits to ASV17 (Table 3), Tremblaya is solely found as an obligate endosymbiont in mealybugs, and its phylogenetic relationships with other Proteobacteria are still uncertain. Tremblaya shows up as an unclassified genus in NCBI and a member of Burkholderiaceae in EZbioCloud. Tremblayales and Tremblayaceae taxa are listed in Table S5 and S4, which are only denoted in the GreenGenes database. However, according to Table 3, the closest hit to ASV17 was a Clostridium sequence in the GreenGenes database. This shows that the taxonomic assignment of ASV17 is uncertain, but it is not clear to the reader if this ASV, for instance, was treated as unclassified or maybe as a representative of 'Tremblayaceae' and a member of Tremblaya.

ASV taxonomy labels used for composition analysis, graphs and supplementary tables came from the GreenGenes database annotations using the QIIME2 default confidence level (≥ 0.7 or 70%) for this database. This clarification was added to lines 152-153 in the Methods section. According to QIIME2 support, this confidence value can be applied only for the deepest taxonomic level assigned for each ASV (you can see forum here: <https://forum.qiime2.org/t/confidence-values-taxonomic-assignment/13199>). In our process of taxonomic annotation, only for those ASVs where GreenGenes was unable to assign a Genus label, we used NCBI-BLAST and EzBiocloud to identify/assign Genus based on best-hits and identity threshold of 97% for both methods. When NCBI-BLAST and EzBiocloud resulted in contradictory Genus best-hits at $\geq 97\%$ identify each, the original GreenGenes identification taxon level was maintained. This was also clarified in the Method section, lines 156-158.

Line 142: PCoA and NMDS abbreviations should be described here where they are first used. Full names of PCoA and NMDS were included. (see lines 169-170 in revised manuscript).

Lines 166-168: The sentence states 'visualized with PCoA and NMDS plotting (Fig. 1C).', however, NMDS plotting is only shown in Figure 1D, which is not cited in the text. The sentence needs to be corrected, and Fig 1D should be referenced.

Both, PCoA and NMDS plots were referenced in the main text as: (Fig. 1C,D). See line 243 in revised manuscript.

Lines 188-190 (and Table 3, Tables S2-S6): It is valuable that the authors included tables summarizing the percental distributions of ASVs at different taxonomic ranks. However, it is not clear to the reader how the ASVs were annotated and what are the classification data based on (please see my comments on the method section).

The methodological steps used for assignments of taxonomic labels was clarified in the Methods section (see lines 152-153 and 156-158) as also described for another request/comment above in this review.

Lines 217-218: Seven or 19 ASVs (fig 2D) were unique to nymphs? Or seven ASV were found among nymph samples exclusively besides the 40 ASVs universally present in all nymph samples? Please, rephrase the sentence to make statements on the unique ASVs clear. It might be good to add whether all ASV found in adult samples were also found in at least one nymph sample, or if there were ASVs unique to adult samples.

To make this statement more clear in terms of the identification of the core microbiota, we have removed the information concerning the "unique" ASVs and have paraphrased the initial statement in this part. As stated in the original paragraph, the line about unique ASVs was not essential to the description of the core microbiota. See now lines 303-307 in the revised manuscript.

Lines 223-225: According to the methods section GreenGenes was the primary source of taxonomic assignment, but this database is not mentioned in the sentence. Please, clarify.

The use of GreenGenes as the primary source for taxonomic annotations has been indicated in these lines (311-312 in revised manuscript).

Line 245-253: Consider moving part of these to the Introduction section. It would be great to highlight that only a few studies have investigated the bacterial symbionts of Miridae plant bugs and to mention these previous studies already in the introduction.

We agree with this observation. These lines have been moved to the Introduction as suggested. See lines 106-112 in the revised manuscript.

Line 334: 'by the potential gut-associated bacteria genera' – These genera probably include non-gut-associated bacteria as well; consider changing it to e.g. 'by potential gut-associated members of the genera'.

We adopted this suggestion. Accordingly, changes have been included in the line 463 of the revised manuscript.

Table 3: The authors show different taxonomic levels for assignment. I think it would be better to show the closest hit in each database with their accession numbers. It is not clear from the

manuscript and the table which taxonomic assignment did the authors conclude for each ASV, and how family/genus level assignments were done. I suggest adding a column to indicate the taxonomic assignment inferred for each ASV and denote the ASVs as 'unclassified' when that is uncertain e.g. in the case of ASV17 and ASV18.

Taking into consideration these comments and those above about the need to clarify taxonomic assignments, we decided to modify Table 3. In the new revised table we include only the concluded taxonomic label at the Genus rank according to the described taxonomic annotation procedure (in the Methods) that uses GreenGenes as primary source or BLAST/BioCloud when needed. The new table shows a unique genus label for each ASV (or "Undetermined") and includes a short text explaining the origin/source of the annotations. We think this will help to avoid any confusion about the final taxonomic assignments.

Figure 1. All plots (A-D) show a comparison between the nymph and adult life stages of *M. velezangeli*. Rephrase the title of Figure 1 to indicate this, e.g. 'Diversity and community structure of the bacterial microbiota in adult and nymph life stages of *Monalonion velezangeli*'. Suggested modification for Fig. 1 caption was done to include these details. See revised fig.1 caption (lines 814-820).

Does Figure 1D show the result of the ANOSIM test? Please mention the statistical test relevant to Fig 1D in the caption of the figure, too. What does the Stress value mean in Fig 1D? This value is not mentioned in the manuscript when describing the statistical tests.

The statistical tests PERMANOVA and ANOSIM are independent of the PCoA and NMDS plots. These tests complement the plots. To avoid any misunderstanding from readers we decided to eliminate the test results from the plots and modify figure caption (see revised Fig1). On the other hand for Fig1D, the stress value represents the Goodness-of-fit of the NMDS. As a general rule for stress value in NMDS plots, values below 0.05 are considered as "excellent", so interpretations or inferences can be extracted from a NMDS distribution plot. Stress values over 0.2 are considered "poor". A short explanation of the stress values was also included in the figure caption for clarity of the readers.

Figure 2. All plots compare the nymph and adult life stages of *M. velezangeli*. Consider modifying the caption to make this clear, e.g. 'Taxonomic composition of the bacterial microbiota in nymph and adult life stages of *M. velezangeli*'.

Suggested modification for Fig. 2 caption was done to include these details. See revised fig.2 caption (lines 823-827).

Figure 3. Similarly, to figures 1 and 2, the title could be modified to indicate that adult and nymph samples of *M. velezangeli* are summarized on the plots, e.g. 'Relative abundances of the bacterial microbiota with the exclusion of *Wolbachia*-associated sequences in nymph and adult samples of *M. velezangeli*'.

Suggested modification for Fig. 3 caption was done to include these details. See revised fig.3 caption (lines 830-832).

Responses to Reviewer #2:

In this manuscript, Navarro-Escalante et al. identified the bacterial microbiota of an insect pest, the polyphagous plant bug *Monalonion velezangeli*. The authors performed DNA extraction from whole insects followed by 16S rRNA amplification and sequencing and then analyzed the bacterial clades present in two life stages: nymphs and adults. They highlighted the high prevalence of *Wolbachia* symbionts in both life stages and the presence of a core bacterial microbiota shared by the two life stages composed of 13 bacterial genera.

This study is interesting because it provides new information on insect microbiota, especially in the Miridae insect family for which the knowledge on the diversity of bacterial symbionts is very limited. However, I suggest to incorporate additional experiments in this study to precise the nature of the symbionts, especially for the prevalent *Wolbachia*. Identification of *Wolbachia* as the most prevalent bacteria in *M. velezangeli* symbionts is one of the key findings of this study. However, the authors did not present a comprehensive analysis of this bacterial type, despite the recent studies revealing the presence of *Wolbachia* in other Heteroptera. A full identification of the *Wolbachia* strain (or strains) identified in *M. velezangeli* should be performed, with the use of specific gene sequencing (*wsp* or *ftsZ*) followed by a phylogenetic analysis. Data from this study need to be compared with those of recent studies on other Heteroptera with *Wolbachia* endosymbionts (Gerroidea superfamily in Conjard et al, 2022 PLoS One ; Aphelocheiridae family in Kaczmarczyk-Ziemba & Krepinski, 2020 Entomological Science ; Miridae family in Kikuchi & Fukatsu, 2003 Applied and Environmental Microbiology).

As suggested by the reviewer, we tested the presence of *Wolbachia* in individual insect samples using PCR with *wsp* primers. Several *Wolbachia* *wsp* amplicons were sequenced using Sanger, analyzed using phylogenetic clustering and compared against NCBI *Wolbachia* sequences using BLAST (see new lines in Methods: 179-220). These results are presented in the revised manuscript (lines 318-344) and discussed accordingly (see lines 376-376 and 395-400).

A similar phylogenetic confirmation for the 12 other members of the core microbiota (when data on other insect species are available) would strengthen the data presented in this manuscript.

Unfortunately any other additional confirmation for remaining bacteria in the core microbiota is technically difficult since no specific molecular tools exist for these symbionts.

Since the study was performed using whole insect DNA, a confirmation by means of adult or nymph gut dissection followed by DNA extraction, sequencing and analysis would clearly reinforce the author's conclusions on the core microbiota.

After this study was performed, we did not have access to newly fresh insect material for the suggested analysis. Several attempts to dissect intact guts from the very few insect samples in alcohol that we still had resulted in bad quality tissues for analysis.

A recent study has determined that partial sequencing of the 16S sequence could lead to imprecise determination of bacterial clades (Johnson et al, 2019 Nature Communication),

however authors have focused on the V3-V4 region of 16S rRNA. A possible bias through this partial sequencing should be discussed.

Based on this suggestion, we have included a paragraph in the Discussion section describing these limitations and bias in reference to our results and interpretations. See lines 362-371.

Precisions on the Material and Methods section:

Lines 110-112: Were the nymphs and adults collected on the same plant parts? It should be indicated.

We have clarified that insect samples were collected from multiple coffee plants. See line 130 in revised version.

Line 118 : Address for Invitrogen is missing

Address for Invitrogen was included. See line 139.

Line 125: Reference for the QIIME2 software is missing

The corresponding reference was included. See line 148.

Precision on the Result section:

Figure 2C and Figure 3B: The heatmap results presented here for Romboutsia are not clear. In figure 2C, Romboutsia is abundant for the first adult sample and in figure 3C this is the opposite (abundant for the two other adult samples).

The variations in the relative abundance of Romboutsia and other genera per sample between figures 2C and 3B is due to the effect in the total read counts in each sample caused by removing the Wolbachia-associated reads. Due to the significant number of Wolbachia-associated reads present in all samples, this resulted in differences in the relative read counts for ASVs across the samples. For example, for adult samples, total raw count per sample before Wolbachia removal, were 69,782; 90,168 and 90,009 reads respectively. Then after Wolbachia removal, total reads were 9,226; 5,388 and 4,478, respectively. However, despite these disparities, the statistical similarities in the overall relative abundance of each genus between adults and nymphs remained unchanged after the removal of Wolbachia reads. For better clarify this in the manuscript, a statement describing these changes in relative abundances has been included in the Results section (lines 294-295).

Suggestion on the Result section: The two types of green colors used in Figure 3 for Chlorobi and Tenericutes are very similar and could lead to misinterpretation of a similar profile between two samples: Adult1 and Nymph2 when in fact the two samples are very different. I suggest to use more distinct shades of green or different colors.

Figure 3 has been edited to clearly differentiate Tenericutes from Chlorobi as suggested. See revised figure 3.

Precision on the Supplementary data:

The tables S1 to S6 present in the column 8 the P-values obtained with the Mann-Whitney test for Adult vs Nymphs. These P-values have variable numbers after the dot and some P-values

are equal to 1 (is it an approximate value?). For all P-values, it would be important to normalize the number of decimals after the dot for all the bacterial clades.

All p-values in supplementary tables have been rounded to 3 decimals. P-values of 1 are the actual number produced by the Mann-Whitney U test run on the MicrobiomeAnalyst tool that was used in these analyses.

Genus/species names should all be in italics in the columns 3/4/5 of Table 3, in the figure legend of Figure 1 and in the first column of Table S6.

Genus names in revised Table 3, figures and table S6 have been changed to italics.