

Manuscript title: " Diversity of bacterial symbionts associated with the tropical plant bug *Monalonion velezangeli* (Hemiptera: Miridae) revealed by high-throughput 16SrRNA sequencing"

<https://doi.org/10.21203/rs.3.rs-2022560/v2>

Please note, that I made my comments based on version 2 of the manuscript available at the DOI link above.

The paper describes a high-throughput amplicon sequencing-based analysis of the bacterial community associated with the tropical plant bug *Monalonion velezangeli* (Hemiptera: Miridae), an important agriculture pest damaging coffee plants. The study uses 3-3 independent sample sets of adult and nymphal life stages to reveal the composition of the bacterial microbiota of *M. velezangeli*.

The abstract summarizes the aim, methods, and results of the study well. The introduction provides a solid overview of bacterial symbionts of insects and those among the plant sap-feeding and phytophagous insects within the order Hemiptera. Although the family Miridae is the most species-rich group of true bugs, little is known about the bacterial symbionts associated with these insects, which - I think - could be emphasized more in the introduction to highlight the importance of the present study. The authors could also summarize previous studies on the bacterial symbionts of Miridae insects in the introduction, without going into much detail on their results before the discussion.

Most of the description of the methods is straightforward and informative. However, the taxonomic assignment of Amplicon Sequence Variants (ASVs) should be described in more detail. Although taxonomy results make up the backbone of the present study, it is not evident how the ASVs were annotated. It is not clear what confidence/sequence similarity thresholds were used for different levels of taxonomic ranks and which database the classification data are based on.

The result section is well-structured. The discussion takes into account past studies when making conclusions on the data. The results are adequately discussed and not overstated. The reference list needs formatting to have a unified layout. Please add the journal volume and page numbers of the articles whenever possible.

The figures in the manuscript are informative and nicely done. Figure 1D is not referenced in the manuscript text. I suggest listing the taxonomic assignment concluded for each ASV of the core microbiota besides the best hits in different databases in Table 3.

Please find my specific comments below.

Specific comments:

Lines 118-120: Are there references available for these primers? If yes, please add the citation.

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

Line 126: Please add a reference for DADA2

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*.

Line 142: PCoA and NMDS abbreviations should be described here where they are first used.

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.

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).

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.

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.

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.

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'

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.

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*'. 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.

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*'.

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*'.