nature portfolio

Peer Review File

Assembly and maintenance of phyllosphere microbial diversity during rubber tree leaf senescence

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

The manuscript examines leaf senescence's impact on microbial communities' diversity and composition in endophytic and epiphytic layers of leaves. It investigates how leaf aging influences microbial diversity and identifies specific microorganism groups that are enriched or depleted in green and yellow leaves. Employing comprehensive statistical methods, the study analyzes the relationships between microbial communities and environmental factors, such as leaf nutrient content. This research contributes to understanding the dynamics of microbial communities during leaf senescence.
My overall impression of the work is generally positive. The outstanding feature of the study is a comprehensive statistical analysis that forms a coherent narrative, complemented by an abundance of supplementary materials, making the evidence more convincing.
However, my main critique relates to the order of presenting some information in the results and methods sections and the use of certain steps that seem incorrectly executed. Perhaps I misunderstood how exactly the authors performed a specific step, and the issue lies in my interpretation of this step, not its suboptimal processing. Nonetheless, this means some aspects should be articulated more clearly and explicitly to dispel any doubts in the reader's mind.
There is a structural issue within the manuscript regarding the presentation order of methodologies and results. For example, the linear mixed-effect regression (LMER) models, which analyze total nitrogen (TN), total potassium (TK), and organic matter (OM), are discussed in the methodology section before the generalized linear models (GLMs) that assess microbial OTU enrichment and depletion in yellow versus green leaves. This sequencing complicates the tracking and cross-referencing of methodologies and results, as the LMER models are presented earlier than the GLMs, yet the results section discusses the LMER findings after the microbial community analysis. Additionally, the integration of FAPROTAX and FUNGuild-related analyses within this sequence further complicates the narrative flow, as these are introduced as the third of the analyses in the methodology section, potentially confusing readers attempting to follow the logical progression of the study's analytical framework.
L86-93: I am confused regarding the sampling design described in the manuscript, despite multiple readings and an examination of the supplementary materials. The sampling was used across three plots within each of six sites (i.e. 18 plots), where at each plot, one mixed sample of green leaves and one mixed sample of yellow leaves were collected (from three trees and each cardinal direction) during both the rainy season (September 2019) and the dry season (December 2020) = 18 x 2 x 2 = 72. Is it correct? If yes, I am puzzled by the

statement that „A total of 72 leaves were collected in the two seasons (i.e., rainy season and dry season), and a total of 144 samples were collected throughout the experiment (September 2019 for rainy season and December 2020 for dry season)“ finding it confusing and possibly inaccurate, potentially doubling the information and misrepresenting the data (not 72 leaves but mixed leaf samples). The number 144 might pertain to samples from both endophytic and epiphytic communities, though this is not explicitly mentioned in the passage. Even if my understanding is incorrect, then still the description of the design should be rewritten for clarity and unambiguous understanding on the first read, especially regarding the number and labeling of mixed samples.

L95: I'm puzzled by the phrase "A portion of each sample was used to quantify." What exactly does this mean? If one mixed sample is supposed to contain leaves from three trees, from all four cardinal directions, that means at least 12 leaves per mixed sample. From this subset, were specific leaves selected for further analysis? If so, was this selection random, systematic (and if systematic, how was it determined — did it involve picking a certain number of leaves from each tree/cardinal direction)? Or was a certain part cut out from each leaf?

L114-L122: Why is the polymerase chain reaction (PCR) amplification process described twice? Typically, when scientific texts refer to previously published protocols, authors either directly adopt an existing protocol without changes or specify what modifications were made to the standard procedure. This approach makes it easier for readers to follow the methodology and understand how the current study extends or differs from previous work. It would be more logical if the authors first stated they are using a protocol described in the literature and then specified any modifications they made, to clarify how their method diverges from the standard. It's confusing that the authors first provide a detailed description of the PCR cycle and only then reference a previous study. This could lead to questions about why a detailed description is necessary if there's a standardized protocol that has already been published and referred to by the authors. Such a structure might raise questions among reviewers and readers about the reasons for this choice. Ideally, the authors should clearly explain in their text why they decided to present this procedure in this way. Or is it because the first description is for bacterial PCR and the second for fungal PCR? If so, this needs to be explicitly stated.

L135: Why use OTUs instead of ASVs? Utilizing Operational Taxonomic Units (OTUs) over Amplicon Sequence Variants (ASVs) might be seen as an outdated approach in sequencing data analysis. OTUs group sequences that are sufficiently similar into a single category, potentially leading to a loss of distinction between closely related but different sequences. ASVs, defined based on exact sequences regardless of a predefined similarity threshold, preserve complete sequence information. OTUs can be influenced by the choice of similarity threshold and clustering methods, whereas ASVs provide a unique identifier for

each sequence, improving reproducibility. ASVs thus allow for more detailed and accurate tracking of microbial community dynamics, crucial for studies focused on ecological interactions or monitoring environmental changes. Did the authors choose OTUs over ASVs for historical reasons, or to maintain consistency with previous data?
L142-L144: "Diversity indices (Chao and Shannon) (based on identified OTUs) were calculated" versus "Community α-diversity was estimated using the Chao and Shannon index." – why is this described twice? It likely refers to the same calculation and could be articulated more clearly for the reader.
L144: Given that epiphytic and endophytic microbiota were taken from the same sample, incorporating the identity of this sample as covariates in the Principal Coordinates Analysis (PCoA) and other analyses is entirely appropriate. This approach would help differentiate the microbial community variability specific to individual samples (which could be a major source of variability) from the differences between epiphytic and endophytic communities (or between green and yellow leaf communities). Partial PCoA (or, more broadly, any form of multivariate analysis that allows for the inclusion of covariates) facilitates the correction of this effect by using the sample identity as covariates.
L149: In PERMANOVA, 'strata' can be used to account for groups or blocks of samples that should be considered homogeneous groups during permutation tests. This is useful when we want to maintain structural dependencies in the data, such as the pairing of samples (yellow and green leaves from the same sample in this case). Using strata in PERMANOVA allows the analysis to respect this internal structure of the data while testing hypotheses about differences between groups. Moreover, the assumptions for PERMANOVA include the homogeneity of multivariate dispersions (i.e., variances) between groups. To verify this assumption, PERMDISP2 analysis is used. If PERMDISP2 indicates significant differences in dispersions between groups, we should be cautious in interpreting the results of PERMANOVA, as this may indicate a violation of assumptions.
L163: liner → linear, It's worth considering a Generalized Linear Mixed Model (GLMM) since both green and yellow leaves come from the same sample (the same trees from the same area), which could have a significant effect.
L168: I recommend incorporating the Sloan null model into author's analysis to help verify if the identified "core taxa" also surpasses model expectations as 'above prediction.' It's important to recognize that not all taxa present in over 60% of samples with >0.1% abundance are necessarily exceeding these expectations. Some may simply be ubiquitous species, not integral to specific ecosystems. Utilizing the Sloan model can assist in differentiating truly key taxa from those common and less essential, thereby enriching our understanding of microbial community structures and functions. Despite the Sloan model isn't designed for defining core taxa, it can significantly complement authors' methods to provide deeper insights into microbial dynamics and functionality.

L173: I must admit that I often use or see PICRUSt2 being used, preferred for its complexity and accuracy. PICRUSt2 is generally considered one of the most advanced tools for predicting functional capabilities based on taxonomic composition, thanks to its ability to reconstruct missing states and provide a wide range of functional predictions. It utilizes extensive databases like KEGG, allowing for detailed and comprehensive analysis of functional capabilities. Since the authors chose a less common approach, I would ask them to add a sentence or two in the discussion about why they made this choice.
L182: Beyond the explanation provided by the authors regarding values close to 0 (stochastic processes) and -1/+1 (deterministic processes), it can also be added that a value near -1 in the abundance-based beta null model suggests less dissimilarity between communities than expected, which could result from strong environmental filtering leading to community convergence under challenging conditions. Conversely, a value close to +1 indicates greater differences between communities, which may signal the importance of competition or specific environmental gradients in otherwise favorable conditions, supporting diversity. It might be interesting to interpret the results from this perspective, whether the values were closer to -1 or +1.
L210-215 and Figure 2 and corresponding parts in methods: Is this about comparing the absolute number of reads between samples? Comparing the absolute abundance of reads across different samples is extremely problematic, especially when data are generated using sequencing platforms like Illumina. Main issues include the number of reads being influenced by factors such as the amount of input DNA, PCR amplification efficiency, and the sequencing depth of each sample. The number of sequencing reads does not directly correspond to the number of bacterial cells or their biomass in the sample. Reads can be generated from various copies of rRNA genes within a single cell, which can lead to overestimation or underestimation of the actual abundance of certain bacterial species. Most studies focusing on comparing microbial communities prefer using relative abundance (the percentage representation of each taxon relative to the total number of sequenced reads) over absolute abundance. Relative abundance better reflects differences in microbial composition between samples and is less affected by sequencing technique variability. It's also advisable to complement such studies with additional methods that can quantify microbial communities independently of the number of generated sequencing reads, such as quantitative PCR (qPCR) for specific taxa.
L229-L235: In addition to the interpretation of results presented by the authors, another interesting observation is that the curves do not reach an asymptote, indicating that collecting more and more leaves would lead to a significantly higher number of OTUs. This lack of an asymptotic trend and the anticipated large diversity should be discussed in the discussion section. It suggests a rich and potentially unexplored microbial diversity within the leaf samples, underscoring the complexity of the ecosystem and the need for further

sampling to fully understand its biodiversity.
L238-L242: What is meant here by the abundance of Actinobacteriota and why study the relationship between this abundance and the Chao and Shannon indices? A greater sequencing depth leads to a larger number of reads for a given taxon but also increases the chance of capturing more types of OTUs, which directly contributes to a higher Shannon index.
L253-L255: The observed paradox, where the relative abundance of epiphytic and endophytic pathogens in yellow leaves is higher than in green leaves, yet the overall relative abundance of plant pathogens is lower in yellow leaves, could indeed be due to the categorization within FUNGuild. The "plant pathogens" category might encompass a broader spectrum of pathogens than just epiphytic and endophytic ones. It's possible that other pathogen groups, not classified as epiphytic or endophytic, show lower representation in yellow leaves, leading to a decrease in the overall relative abundance of plant pathogens despite an increase in epiphytic and endophytic pathogens. This discrepancy might be a suitable topic for discussion, suggesting a nuanced interpretation of pathogen dynamics.
L261: „abundance Actinobacteriota“ → „abundance of…“
L280: Discussing the results of the abundance-based beta null model in detail, specifically whether the outcomes lean more towards -1 or +1, could illuminate whether strong environmental filtering or competition plays a more significant role. Providing explicit values, not just using the values in a single graph, would enhance the discussion of subsequent points by offering a clearer understanding of the underlying ecological dynamics.
L346-L351: The apparent paradox you've described could be possibly related to the idea that deterministic processes dominate at the two extremes of a continuum (values -1 and +1), while stochastic processes prevail in the middle. It's conceivable that if a factor consistently shifts this value, for example, always to the right, it could in some instances move from -1 to 0 (from deterministic to stochastic processes) and in others from 0 to +1 (from stochastic to deterministic). Thus, the effect would be clear and consistent (a shift to the right on the gradient), even though the outcome might appear paradoxical or inconsistent.
L365 and further: The paragraph describes a process where the aging of leaves leads to reduced selective filtering and, consequently, increased microbial diversity. However, the mention of increased selective pressures on bacteria in aging leaves introduces potential confusion about how these pressures relate to the concept of selective filtering and their combined effects on microbial diversity. A clearer distinction or explanation of terms and processes would help resolve these inconsistencies. There's a seeming contradiction between stating that leaf senescence reduces selective filtering (leading to increased

diversity) and citing previous work that suggests bacterial populations face increasing selective pressures as leaves age (which could imply a decrease in diversity due to harsher survival conditions).

Reviewer #2 (Remarks to the Author):

Authors investigated the outcome of phyllosphere microbial communities (endophytes and epiphytes) during leaf senescence in the rubber tree. They evaluated fungal and bacteria diversities by metabarcoding (16 rRNA and ITS, respectively) and compared them between healthy (green) and senescent (yellow) leaves. Authors also attempted to predict microbial community functions from taxonomic assignment of microbial sequence OTUs identified in leaf samples, and to link those functions with the putative role of some microorganisms (especially Alphaproteobacteria and Actinobacteria) in relocating nutrient during leaf senescence.

General comment: The introduction is too short and do not support the proposed hypothesis, which are developed but seem to come from nowhere. Especially the section devoted to the roles of plant-associated microbes in leaf development is laconic and highly descriptive (L 44-50). Authors should develop this section, at least to justify the proposed hypothesis where a lot of principles and concepts are presented without prior explanation (L64-70). No enough detail is provided in different sections of the method, just the literature about the extensive list of tools cited, assuming that you did exactly and appropriately what is described in each paper, but more methodological information would help the reader to understand what you exactly test for and to evaluate the limits of the study. I have also serious concerns about the methodology used by the authors to analyze they samples (e.g. type of OTU/taxa abundance: relative or absolute? No rarefaction applied. Statistical models not considering interaction between factors, etc...). Detailed taxonomy of OTUs are not provided and conclusion are only drawn for major, yet ecologically, evolutionary and functionally very diverse, bacteria groups. Observations often rely on potential technical biases and obvious statistical effects (correlation between dependant variables) rather than actual biological effects. Causality is often invoked from observed correlations, without experimental evidence. Also, I have serious reservations in predicting functions from 16rRNA/ITS data. The discussion and conclusion are interesting but overall lacks experimental support and is strongly jeopardized by limits of the methods and over interpretation of the results. If no clarification can be made to overcome these numerous issues, I strongly suggest that authors consider reanalyzing their data from scratch and rewrite their paper with help from a microbial ecologist.

Detailled comments:

Introduction.

L. 41 – The phyllosphere refers as the surface of leaves (epiphyte microbes) and do not include endophyte microbes. Also, massive is a bit too vague.

L. 71-73 – This is a very vague description of your results. Please be more specific.

Method.

L. 89-93 – Why pooling leaves from different trees? Also I do not understand how you come to 72 leaves and 144 samples. You have 4 leaves per tree, 3 trees per plot, 18 plots, 2 seasons, 2 types of leaves (4x3x18x2x2 = 864 leaves) and to make your samples, you pooled leaves (yellow and green separately) within plots (18x2x2 = 72 samples). Did you collect leaves in sterile bags to prevent contamination? How long and in which conditions you transported and conserved them until processing?

L. 94-97 – Could you be more specific about the method you used to extract the epiphyte and endophyte microbes? What is the leaf mass of each sample? What proportion you use for nutrient measurement and microbial community analysis, respectively? How long and in which condition did you conserve your samples at -80°C to prevent from external microbial DNA degradation/contamination?

L. 104 – Please provide more details about how microbial cell were “dislodged” especially how you ensured to minimize host DNA contamination and microbial contamination. Also, were these 10-15 g of leaves also used to retrieve epiphyte microbes? What was the method to collect epiphyte microbes from the leaves?

L. 121 – Why PCR was performed in triplicates? Were PCR products pooled afterward?

L. 123 – Results from negative controls are not provided in the result section. Also, those are not controlling for sample process (from collecting to microbial cell collection), which could be a big source of contamination.

L. 130 – Not sure what sequence coverage stand for here.

L. 135 – Did you apply this threshold because you assumed that OTUs with more than 97% of similarity belonged to the same species? Please justify.

L. 138 – What this confidence statistics means? Please justify the threshold.

L. 144 – Why using two diversity indexes? What are their differences?

L138-141 – How many reads and OTUs did you typically get for each sample, before and after Mt DNA removing? A huge variation among samples (especially between yellow and green leaves) might bias greatly your results. Please specify that mitochondrial reads correspond to the host plant contamination. I also expect that you observe more host (degraded) DNA in yellow than in green leaf samples. Did the 3454998 reads included mitochondrial DNA or you already removed those from this number. Same question for OTUs. Here you mentioned "OTUs with 100% identify" but you said you used a 97% threshold. Please clarify. Also, did you evaluate the proportion of chimeric OTUs (pairing of reads from different organisms) and if so, how did you control for them?

L146 – Why did you perform this analysis at the OTU level? Again did you assume it's representative of species? Please justify.

L147-148 – I am not familiar with this diversity statistic using the cumulative number of OTUs. Could you please explain or at least provide a reference? It seems that you did not perform per-sample OTU rarefaction curves before analysis your samples. This is an essential step to normalize your samples because of the great variability in reads/OTUs expected among your samples, and also to ensure that sequencing was deep enough to capture captured all the microbial diversity. Without this step, your analyses will be strongly biased toward samples with higher microbial DNA yield and over-representation or rare OTUs, while samples with low yield will be highly sensitive to contamination and might induce a lot of variability in your subsequent analyses. For instance, the great differences you might observe between green and yellow leaf microbial communities, or between endosphere and phyllosphere, might just result in different microbial DNA yields between these different types of samples. These biases will be greatly amplified if you performed (as I suspect) all your subsequent analyses on absolute abundances rather than on relative abundances. Especially, I expect that treatment used to extract endophyte microbes generate more DNA due to cell (both microbe and host plant) degradation/contamination, and so more yield than for epiphyte sample. Similarly, I expect that cells are more degraded in yellow than in green leaves, and generate more DNA yield.

L148-151 – Unless I misunderstood (but then you should clarify that your statistical approach considered it), your PERMANOVA tested for season (S), location (L) and leaf stage (LS) as additive factors (S+L+LS), but do not test for interaction between those factors (SxLxLS), which is expected to be quite high. For instance, the effect of season on your microbial communities might be dependant of different local climatic conditions in locations (SxL interaction). Season and leaf stage might also covariate (SxLS) because the proportion of yellow leaves might be higher during dry season and you should see this effect on microbial community even if you sampled equally yellow and green leaves in each season.

L154 – Please clarify what VIF was used for.

L157 – Why did you use abundance of dominant phyla? What is your criteria for this dominance? Also, did you use relative of absolute abundance? Absolute abundance will be strongly biased by total microbial yield (see my previous comments) and will reflect technical more than biological effects. Why removing underrepresented phyla? Even rare, they might reveal important biological processes if your samples have enough read and are correctly normalized (see my previous comments). Also, why focusing of phyla and not OTUs? And why not intermediate taxonomic levels (Order, Family, Genera), which could increase your statistical power? The results you will obtain with phyla only might only reflect very long term evolutionary processes rather than more recent/specialized biological processes, as Actinobacteria and Proteobacteria, for instance, contain different orders and families of bacteria that diverged for billion years ago and having completely different evolutionary routes, ecological niches and thus, different functions.

L. 161 – Could you justify why you considered location and saisons as random effect?

L. 169 – Please justify why you consider taxa (phyla? justify) with abundance >0.1% and present in at least 60% of samples are core taxa?

L. 173-175 – I am very sceptical about functional predictions from rRNA16s/ITS. Those markers could just be used to infer broadly (as the genus level) the taxonomic composition of a microbial community but cannot summarize the incredible genomic complexity and variation in term of diversity and functions within a single microbial genera. Results from such functional predictions should be taken very cautiously by authors.

L. 176-177: Could you clarify what do you mean by stochasticity ratio, determinist/stochastic community assembly?

L. 193-194 – Why only focusing on those particular phyla and nutrients? .

L. 196 – Here and elsewhere, I think authors meant “environmental filtering”. Also, could you clarify what you were testing for here, and the method (FEAST) you used?

Results

L. 205 – What do you mean by “OTU level?” Diversity? Also, this is not clear if pvalues are associated to differences between leaf types or between compartments. Also, here and elsewhere, remind the statistical test associated to each p-value.

Figure 1 – One can observe some clustering of samples according to unidentified factors. Probably season and/or location? I would have appreciated the same PCA displayed with location and seasons.

Figure 2 – Did authors actually performed all the analyses based on microbial absolute abundance?

Table S2 – Again interactions between factors were not rested. The residual (unexplained) variance is not showed. With the exception of endophytic bacteria, variance explained by leaf stage is negligible compared to season and location.

L. 218 – Please remind the statistical test associated with each pvalue. Did you correct pvalues for multiple tests (Bonferroni)?

L. 223-224 & Table S7 – “the majority of core bacteria OTUs”, also please clarify in what proportion. Why not providing the detailed taxonomy of each OTU? As mentioned before, those two dominant classes probably arbour diverse genera of bacteria with very contrasted lifestyles. Phyllosphere-associated bacteria are very well studied and it is known that Alphaproteobacteria, for instance, are dominant in the phyllosphere (eg. Methylobacterium, Beijerinckia, Sphingomonadales…), but some are more associated for instance, with the Rhizosphere (many but not all Hyphomicrobiales).

L. 225-227 – Why not providing details for fungi, as you did for bacteria? Please justify.

L. 238-242 and Figure 3C – This is a statistical effect, not a biological observation! You expect that a diversity index is positively correlated with the abundance of taxa, and

especially with the most abundance ones, since this index is calculated from those abundances!

L. 244-251 – You did not test for gene but for functional enrichment. Again, you cannot predict the gene composition of an organism just from a single marker gene (16s or ITS). Your could just broadly and cautiously infer their function from strongly biased and incomplete databases of microbes with similar 16s/ITS sequences and with known genomes and metabolisms. Also “Epiphytic functional genes do not seem too different between the green and yellow leaves“ is a very vague statement.

L 251-252 and Figure 4B: “This indicates that the increase of saprophytic bacteria promotes the nutrient relocation for yellow leaves“ could you please clarify how you came to this conclusion? Also, you did not experimentally demonstrated the mechanism. Hence, this is a correlation, not a causality!

L. 264-269 and Figure 5C: Again, correlation between the most dominant taxa abundances and the most meaningful function abundances are expected since both measurement directly derive from OTU abundance per sample. This is a calculation effect, not a biological effect

L. 277-279 and Figure 6B – “a decrease in total nitrogen concentration in yellow leaves increases the abundance of α-Proteobacteria » - You did not experimentally demonstrated the mechanism. Hence, this is a correlation, not a causality!

L 282-289 and Figure 6C – Intuitively, I think those correlations just reflect statistical effects, no biological processes. All these statistics are dependant because calculated from microbial abundances so we could expect they are all strongly correlated together.

Discussion.

L. 322 – Actinobacteria are known to be saprophytic. Authors did not discuss it.

L. 327-328 – Concepts of stochasticity of community assembly and r-selected microbes were presented in hypothesis but not explained in the introduction

L. 335 – Some Hyphomicrobiales (syn. Rhizobiales) typical of the phyllosphere (Methylobacterium) were shown to have community shift during leaf development (doi.org/10.1128/mbio.03175)

Responses to reviewers' comments

Manuscript number: COMMSBIO-24-0195-T
Title: Assembly and maintenance of phyllosphere microbial diversity during rubber tree leaf senescence
Journal title: Communications Biology

Reviewer #1 (Remarks to the Author):

1. The manuscript examines leaf senescence's impact on microbial communities' diversity and composition in endophytic and epiphytic layers of leaves. It investigates how leaf aging influences microbial diversity and identifies specific microorganism groups that are enriched or depleted in green and yellow leaves. Employing comprehensive statistical methods, the study analyzes the relationships between microbial communities and environmental factors, such as leaf nutrient content. This research contributes to understanding the dynamics of microbial communities during leaf senescence.

   My overall impression of the work is generally positive. The outstanding feature of the study is a comprehensive statistical analysis that forms a coherent narrative, complemented by an abundance of supplementary materials, making the evidence more convincing. However, my main critique relates to the order of presenting some information in the results and methods sections and the use of certain steps that seem incorrectly executed. Perhaps I misunderstood how exactly the authors performed a specific step, and the issue lies in my interpretation of this step, not its suboptimal processing. Nonetheless, this means some aspects should be articulated more clearly and explicitly to dispel any doubts in the reader's mind.

Response: We thank the reviewer for the positive and constructive comments.

2. There is a structural issue within the manuscript regarding the presentation order of methodologies and results. For example, the linear mixed-effect regression (LMER) models, which analyze total nitrogen (TN), total potassium (TK), and organic matter (OM), are discussed in the methodology section before the generalized linear models (GLMs) that assess microbial OTU enrichment and depletion in yellow versus green leaves. This sequencing complicates the tracking and cross-referencing of methodologies and results, as the LMER models are presented earlier than the GLMs, yet the results section discusses the LMER findings after the microbial community analysis. Additionally, the integration of FAPROTAX and FUNGuild-related analyses within this sequence further complicates the narrative flow, as these are introduced as the third of the analyses in the methodology section, potentially confusing readers attempting to follow the logical progression of the study's analytical framework.

Response: We apologize for the confusion of the presentation. We have now restructured the presentation to straighten out the flow of presentation. In the revision we removed the sections about linear mixed-effect regression (LMER) as per the suggestion of reviewer 2, which we agree may not effectively elucidate biological

processes.

3. L86-93: I am confused regarding the sampling design described in the manuscript, despite multiple readings and an examination of the supplementary materials. The sampling was used across three plots within each of six sites (i.e. 18 plots), where at each plot, one mixed sample of green leaves and one mixed sample of yellow leaves were collected (from three trees and each cardinal direction) during both the rainy season (September 2019) and the dry season (December 2020) = 18 x 2 x 2 = 72. Is it correct? If yes, I am puzzled by the statement that A total of 72 leaves were collected in the two seasons (i.e., rainy season and dry season), and a total of 144 samples were collected throughout the experiment (September 2019 for rainy season and December 2020 for dry season)“ finding it confusing and possibly inaccurate, potentially doubling the information and misrepresenting the data (not 72 leaves but mixed leaf samples). The number 144 might pertain to samples from both endophytic and epiphytic communities, though this is not explicitly mentioned in the passage. Even if my understanding is incorrect, then still the description of the design should be rewritten for clarity and unambiguous understanding on the first read, especially regarding the number and labeling of mixed samples.

Response: Apologies for the lack of clarity in describing the sampling design. “A total of 72 leaves” should be “A total of 72 leaf samples” (line 100, page 5). After this modification, the entire sampling method should become clear. Yes, the number 144 corresponds to samples collected from both endophytic and epiphytic communities. We have also revised this section to enhance clarity in the updated manuscript.

4. L95: I'm puzzled by the phrase "A portion of each sample was used to quantify." What exactly does this mean? If one mixed sample is supposed to contain leaves from three trees, from all four cardinal directions, that means at least 12 leaves per mixed sample. From this subset, were specific leaves selected for further analysis? If so, was this selection random, systematic (and if systematic, how was it determined — did it involve picking a certain number of leaves from each tree/cardinal direction)? Or was a certain part cut out from each leaf?

Response: Thank you for your comments. We randomly selected several leaves from each sample for quantifying the water content and other data. (line 108, page 5).

5. L114-L122: Why is the polymerase chain reaction (PCR) amplification process described twice? Typically, when scientific texts refer to previously published protocols, authors either directly adopt an existing protocol without changes or specify what modifications were made to the standard procedure. This approach makes it easier for readers to follow the methodology and understand how the current study extends or differs from previous work. It would be more logical if the authors first stated they are using a protocol described in the literature and then specified any modifications they made, to clarify how their method diverges from the standard. It's confusing that the authors first provide a detailed description of the PCR cycle and only then reference a previous study. This could lead to questions

about why a detailed description is necessary if there's a standardized protocol that has already been published and referred to by the authors. Such a structure might raise questions among reviewers and readers about the reasons for this choice. Ideally, the authors should clearly explain in their text why they decided to present this procedure in this way. Or is it because the first description is for bacterial PCR and the second for fungal PCR? If so, this needs to be explicitly stated.

Response: I apologize for the oversight in describing the PCR process twice in the manuscript. We have now revised it accordingly based on your comments (lines 130-141, page 6).

6. L135: Why use OTUs instead of ASVs? Utilizing Operational Taxonomic Units (OTUs) over Amplicon Sequence Variants (ASVs) might be seen as an outdated approach in sequencing data analysis. OTUs group sequences that are sufficiently similar into a single category, potentially leading to a loss of distinction between closely related but different sequences. ASVs, defined based on exact sequences regardless of a predefined similarity threshold, preserve complete sequence information. OTUs can be influenced by the choice of similarity threshold and clustering methods, whereas ASVs provide a unique identifier for each sequence, improving reproducibility. ASVs thus allow for more detailed and accurate tracking of microbial community dynamics, crucial for studies focused on ecological interactions or monitoring environmental changes. Did the authors choose OTUs over ASVs for historical reasons, or to maintain consistency with previous data?

Response: As per your suggestion regarding the advantages of ASVs over OTUs, we have now replaced OTUs by ASVs in the revised study. (lines 158-160, page 7)

7. L142-L144: "Diversity indices (Chao and Shannon) (based on identified OTUs) were calculated" versus "Community α-diversity was estimated using the Chao and Shannon index." – Why is this described twice? It likely refers to the same calculation and could be articulated more clearly for the reader.

Response: There are indeed the same calculation. Sorry for the confusion. We have now cleaned up the presentation in the revision (lines 165-167, page 7).

8. L144: Given that epiphytic and endophytic microbiota were taken from the same sample, incorporating the identity of this sample as covariates in the Principal Coordinates Analysis (PCoA) and other analyses is entirely appropriate. This approach would help differentiate the microbial community variability specific to individual samples (which could be a major source of variability) from the differences between epiphytic and endophytic communities (or between green and yellow leaf communities). Partial PCoA (or, more broadly, any form of multivariate analysis that allows for the inclusion of covariates) facilitates the correction of this effect by using the sample identity as covariates.

Response: We performed PLS-DA (Partial least squares Discriminant Analysis) and OPLS-DA (Orthogonal Partial least squares Discriminant Analysis) using the ropls package in R environment. However, the results can not separate epiphytic and

endophytic samples. Therefore, we still performed PCoA analysis in the revised manuscript, we also performed the PCoA for different sites and different seasons of rubber tree leaf samples. Please see Figure S3. (Supporting Information: page 5)

9. L149: In PERMANOVA, 'strata' can be used to account for groups or blocks of samples that should be considered homogeneous groups during permutation tests. This is useful when we want to maintain structural dependencies in the data, such as the pairing of samples (yellow and green leaves from the same sample in this case). Using strata in PERMANOVA allows the analysis to respect this internal structure of the data while testing hypotheses about differences between groups. Moreover, the assumptions for PERMANOVA include the homogeneity of multivariate dispersions (i.e., variances) between groups. To verify this assumption, PERMDISP2 analysis is used. If PERMDISP2 indicates significant differences in dispersions between groups, we should be cautious in interpreting the results of PERMANOVA, as this may indicate a violation of assumptions.

Response: In the adonis2 analysis, the parameter "strata" was selected. Additionally, PERMDISP2 analysis was employed to validate the results obtained from the PERMANOVA analysis (lines 176-178, page 7). Please see Table S3. (Supporting Information: page 13)

10. L163: liner → linear, It's worth considering a Generalized Linear Mixed Model (GLMM) since both green and yellow leaves come from the same sample (the same trees from the same area), which could have a significant effect.

Response: We only aimed to assess whether microbial ASVs on yellow leaves exhibited significant enrichment or depletion compared to those on green leaves. To achieve this and also following the suggestion of the other reviewer, we only employed a generalized linear model without incorporating any additional factors into the analysis.

11. L168: I recommend incorporating the Sloan null model into author's analysis to help verify if the identified "core taxa" also surpasses model expectations as 'above prediction.' It's important to recognize that not all taxa present in over 60% of samples with >0.1% abundance are necessarily exceeding these expectations. Some may simply be ubiquitous species, not integral to specific ecosystems. Utilizing the Sloan model can assist in differentiating truly key taxa from those common and less essential, thereby enriching our understanding of microbial community structures and functions. Despite the Sloan model isn't designed for defining core taxa, it can significantly complement authors' methods to provide deeper insights into microbial dynamics and functionality.

Response: Thanks for the suggestion. In the revised study, we utilized the Sloan model to define the key taxa and estimate the ecological processes for core ASVs (lines 187-194, page 8). Please also see Figure 5, Table S5 and Table S6.

12. L173: I must admit that I often use or see PICRUSt2 being used, preferred for its complexity and accuracy. PICRUSt2 is generally considered one of the most

advanced tools for predicting functional capabilities based on taxonomic composition, thanks to its ability to reconstruct missing states and provide a wide range of functional predictions. It utilizes extensive databases like KEGG, allowing for detailed and comprehensive analysis of functional capabilities. Since the authors chose a less common approach, I would ask them to add a sentence or two in the discussion about why they made this choice.

Response: Thanks for letting us know that PICRUSt2 is a more powerful and accurate tool. In the revision, we changed to use PICRUSt2 to predict the bacterial KEGG functions in yellow and green rubber tree leaves (lines 184-186, page 8). Also see the results shown in Figures 3 and 4.

13. L182: Beyond the explanation provided by the authors regarding values close to 0 (stochastic processes) and -1/+1 (deterministic processes), it can also be added that a value near -1 in the abundance-based beta null model suggests less dissimilarity between communities than expected, which could result from strong environmental filtering leading to community convergence under challenging conditions. Conversely, a value close to +1 indicates greater differences between communities, which may signal the importance of competition or specific environmental gradients in otherwise favorable conditions, supporting diversity. It might be interesting to interpret the results from this perspective, whether the values were closer to -1 or +1.

Response: A very good point that makes the interpretation clearer. We have now revised the description following your suggestion and provided a more detailed explanation of the results (lines 364-369, page 14).

14. L210-215 and Figure 2 and corresponding parts in methods: Is this about comparing the absolute number of reads between samples? Comparing the absolute abundance of reads across different samples is extremely problematic, especially when data are generated using sequencing platforms like Illumina. Main issues include the number of reads being influenced by factors such as the amount of input DNA, PCR amplification efficiency, and the sequencing depth of each sample. The number of sequencing reads does not directly correspond to the number of bacterial cells or their biomass in the sample. Reads can be generated from various copies of rRNA genes within a single cell, which can lead to overestimation or underestimation of the actual abundance of certain bacterial species. Most studies focusing on comparing microbial communities prefer using relative abundance (the percentage representation of each taxon relative to the total number of sequenced reads) over absolute abundance. Relative abundance better reflects differences in microbial composition between samples and is less affected by sequencing technique variability. It's also advisable to complement such studies with additional methods that can quantify microbial communities independently of the number of generated sequencing reads, such as quantitative PCR (qPCR) for specific taxa.

Response: Yes, we used the absolute number of reads in the previous version but have changed that to relative abundance in the revised study (Figure 1). Other results about

absolute abundance have been removed.

15. L229-L235: In addition to the interpretation of results presented by the authors, another interesting observation is that the curves do not reach an asymptote, indicating that collecting more and more leaves would lead to a significantly higher number of OTUs. This lack of an asymptotic trend and the anticipated large diversity should be discussed in the discussion section. It suggests a rich and potentially unexplored microbial diversity within the leaf samples, underscoring the complexity of the ecosystem and the need for further sampling to fully understand its biodiversity.

Response: Thanks for the observation. We have discussed this in the discussion section. In the revision, we state that "The accumulated ASVs curves do not reach an asymptote, indicating a rich and potentially unexplored microbial diversity within the leaf samples, highlighting the complexity of the ecosystem and emphasizing the necessity for further sampling to gain a comprehensive understanding of its biodiversity." (lines 411-414, page 16)

16. L238-L242: What is meant here by the abundance of Actinobacteriota and why study the relationship between this abundance and the Chao and Shannon indices? A greater sequencing depth leads to a larger number of reads for a given taxon but also increases the chance of capturing more types of OTUs, which directly contributes to a higher Shannon index.

Response: Our original idea is to use the relationship to indicate how the abundance of Actinobacteriota affect community diversity. In the revision, we have deleted this section because we feel this would not reveal the biological process.

17. L253-L255: The observed paradox, where the relative abundance of epiphytic and endophytic pathogens in yellow leaves is higher than in green leaves, yet the overall relative abundance of plant pathogens is lower in yellow leaves, could indeed be due to the categorization within FUNGuild. The "plant pathogens" category might encompass a broader spectrum of pathogens than just epiphytic and endophytic ones. It's possible that other pathogen groups, not classified as epiphytic or endophytic, show lower representation in yellow leaves, leading to a decrease in the overall relative abundance of plant pathogens despite an increase in epiphytic and endophytic pathogens. This discrepancy might be a suitable topic for discussion, suggesting a nuanced interpretation of pathogen dynamics.

Response: We are honestly not so sure what caused the problem. As you point out that the FUNGuild classification is rather coarse and may not classify fungi to epiphytic and endophytic. This could be a problem. We have now discussed this issue in the revision (lines 334-341, page 13).

18. L261: „abundance Actinobacteriota" → „abundance of…"

Response: Thanks. Fixed.

19. L280: Discussing the results of the abundance-based beta null model in detail,

specifically whether the outcomes lean more towards -1 or +1, could illuminate whether strong environmental filtering or competition plays a more significant role. Providing explicit values, not just using the values in a single graph, would enhance the discussion of subsequent points by offering a clearer understanding of the underlying ecological dynamics.

Response: Excellent suggestion. We have revised the discussion following this comment (lines 364-369, page 14). We further discussed the ecological processes of core ASVs derived from the Sloan model (lines 387-390, page 14-15). Also see Tables S5 and S6.

20. L346-L351: The apparent paradox you've described could be possibly related to the idea that deterministic processes dominate at the two extremes of a continuum (values -1 and +1), while stochastic processes prevail in the middle. It's conceivable that if a factor consistently shifts this value, for example, always to the right, it could in some instances move from -1 to 0 (from deterministic to stochastic processes) and in others from 0 to +1 (from stochastic to deterministic). Thus, the effect would be clear and consistent (a shift to the right on the gradient), even though the outcome might appear paradoxical or inconsistent.

    Response: Excellent suggestion. It was worth noting that the MST for leaf endophytic bacteria and fungi show a trend from 0 to 1with leaf senescence, while leaf surface shows a trend from 1 to 0, these may imply that as the leaves senescence, epiphytic and endophytic microbial communities undergo a different assembly process. (lines 369-373, page 14)

21. L365 and further: The paragraph describes a process where the aging of leaves leads to reduced selective filtering and, consequently, increased microbial diversity. However, the mention of increased selective pressures on bacteria in aging leaves introduces potential confusion about how these pressures relate to the concept of selective filtering and their combined effects on microbial diversity. A clearer distinction or explanation of terms and processes would help resolve these inconsistencies. There's a seeming contradiction between stating that leaf senescence reduces selective filtering (leading to increased diversity) and citing previous work that suggests bacterial populations face increasing selective pressures as leaves age (which could imply a decrease in diversity due to harsher survival conditions).

Response: Sorry for causing the confusion. In the revision, we substantially revised this section by proposing that epiphytic bacteria on yellow leaves were more likely to penetrate into the inner layers compared to green leaves, or vice versa. (lines 295-299, page 11) . To avoid the confusion, we no longer refer to this process as selective filtering.

Reviewer #2 (Remarks to the Author):

1. Authors investigated the outcome of phyllosphere microbial communities (endophytes and epiphytes) during leaf senescence in the rubber tree. They

evaluated fungal and bacteria diversities by metabarcoding (16 rRNA and ITS, respectively) and compared them between healthy (green) and senescent (yellow) leaves. Authors also attempted to predict microbial community functions from taxonomic assignment of microbial sequence OTUs identified in leaf samples, and to link those functions with the putative role of some microorganisms (especially Alphaproteobacteria and Actinobacteria) in relocating nutrient during leaf senescence.

Response: Thanks you for your positive comments.

General comment:

2. The introduction is too short and do not support the proposed hypothesis, which are developed but seem to come from nowhere. Especially the section devoted to the roles of plant-associated microbes in leaf development is laconic and highly descriptive (L 44-50). Authors should develop this section, at least to justify the proposed hypothesis where a lot of principles and concepts are presented without prior explanation (L64-70). No enough detail is provided in different sections of the method, just the literature about the extensive list of tools cited, assuming that you did exactly and appropriately what is described in each paper, but more methodological information would help the reader to understand what you exactly test for and to evaluate the limits of the study. I have also serious concerns about the methodology used by the authors to analyze they samples (e.g. type of OTU/taxa abundance: relative or absolute? No rarefaction applied. Statistical models not considering interaction between factors, etc…). Detailed taxonomy of OTUs are not provided and conclusion are only drawn for major, yet ecologically, evolutionary and functionally very diverse, bacteria groups. Observations often rely on potential technical biases and obvious statistical effects (correlation between dependent variables) rather than actual biological effects. Causality is often invoked from observed correlations, without experimental evidence. Also, I have serious reservations in predicting functions from 16rRNA/ITS data. The discussion and conclusion are interesting but overall lacks experimental support and is strongly jeopardized by limits of the methods and over interpretation of the results. If no clarification can be made to overcome these numerous issues, I strongly suggest that authors consider reanalyzing their data from scratch and rewrite their paper with help from a microbial ecologist.

Response: Thank you very much for your comments. We have rewritten the introduction and methods sections according to your feedback. Additionally, we have reanalyzed our data and removed content discussing the correlation between dependent variables. We have also revised the discussion and conclusion to avoid over interpretation of the results.

Detailed comments:

Introduction.

3. L. 41 – The phyllosphere refers as the surface of leaves (epiphyte microbes) and do not include endophyte microbes. Also, massive is a bit too vague.

Response: Thanks you for your comments. We have revised it as “The leaf surface harbors a large number of microbial communities” (lines 42, page 3)

4. L. 71-73 – This is a very vague description of your results. Please be more specific.

Response: Thanks you very much for your comments. This study provides empirical evidence for the consequences of rubber leaf aging, thereby providing a theoretical basis for fertilization and other management of rubber trees. (lines 80-82, page 4)

Method.

5. L. 89-93 – Why pooling leaves from different trees? Also I do not understand how you come to 72 leaves and 144 samples. You have 4 leaves per tree, 3 trees per plot, 18 plots, 2 seasons, 2 types of leaves (4x3x18x2x2 = 864 leaves) and to make your samples, you pooled leaves (yellow and green separately) within plots (18x2x2 = 72 samples). Did you collect leaves in sterile bags to prevent contamination? How long and in which conditions you transported and conserved them until processing?

Response: Sorry for the error, “72 leaves” should be “72 leaf samples”. (lines 100, page 5)

We have revised the methods section as follows: In order to analyze the physicochemical properties of the leaves, we collected leaves from three rubber trees to ensure they are representative. (lines 95-96, pages 4-5)

The leaves were collected in sterile bags to prevent contamination. The collected leaves were then placed in sterilized, sealed bags and stored in a dry ice box. After completing field sampling, the samples were sent to a nearby research station for processing within 2 hours. (lines 105-106, page 5)

6. L. 94-97 – Could you be more specific about the method you used to extract the epiphyte and endophyte microbes? What is the leaf mass of each sample? What proportion you use for nutrient measurement and microbial community analysis, respectively? How long and in which condition did you conserve your samples at -80°C to prevent from external microbial DNA degradation/contamination?

Response: Each sample had a leaf mass of approximately 200g, with around 185-190g allocated for nutrient measurement and the remaining portion utilized for microbial community analysis. (lines 107-108, page 5)

The method we used to extract the epiphyte and endophyte microbes please see lines 115-125, pages 5-6.

7. L.104–Please provide more details about how microbial cell were “dislodged” especially how you ensured to minimize host DNA contamination and microbial contamination. Also, were these 10-15 g of leaves also used to retrieve epiphyte microbes? What was the method to collect epiphyte microbes from the leaves?

Response: To collect the epiphytic suspensions, intact leaves (to minimize host DNA contamination) weighing 10-15 grams were individually placed into centrifuge tubes containing 1 ml of PBS and shaken at 300 rpm for 15 minutes at 4°C. Subsequently, the same leaves were washed three times in fresh PBS, transferred to new microcentrifuge

tubes containing 1 ml of PBS, and then subjected to sonication in a sonication bath for 1 minute at 4°C to collect the phyllosphere suspensions. (lines 117-123, page 5)
Following surface sterilization, the same leaves were placed in sealed plastic bags for endophytic microbes (line 122-123, page 5).
The phyllosphere suspensions were then filtered through a 0.2 μm filter membrane to isolate epiphytic microbes (line 123-124, page 5)

8. L. 121 – Why PCR was performed in triplicates? Were PCR products pooled afterward?

Response: Thank you for your comments. There was a descriptive error in the statement. Specifically, only three parallel samples (biological replicates) underwent PCR amplification three times each. It is important to note that the PCR products from these replicates were not pooled together. As a result of this clarification, the erroneous sentence has been removed from the document.

9. L. 123 – Results from negative controls are not provided in the result section. Also, those are not controlling for sample process (from collecting to microbial cell collection), which could be a big source of contamination.

Response: Thank you for your comments. Results from negative controls please see Figure S2.
We strictly follow the relevant standards for microbial collection, including disinfection, dry ice transportation, and timely sample processing. (lines 104-106, pages 5)

10. L. 130 – Not sure what sequence coverage stand for here.

Response: Coverage also known as sequencing depth. (lines 152, pages 6)

11. L. 135 – Did you apply this threshold because you assumed that OTUs with more than 97% of similarity belonged to the same species? Please justify.

Response: In the revised manuscript, we utilize ASVs at 100% identity instead of OTUs. (lines 158-160, page 7)

12. L. 138 – What this confidence statistics means? Please justify the threshold.

Response: This threshold is a recognized threshold; please refer to the references 54. Sorry for that perhaps I do not fully comprehend your question.

13. L. 144 – Why using two diversity indexes? What are their differences?

Response: The Chao index measures ASV richness, whereas the Shannon index incorporates both species richness and evenness within a community.

14. L138-141 – How many reads and OTUs did you typically get for each sample, before and after Mt DNA removing? A huge variation among samples (especially between yellow and green leaves) might bias greatly your results. Please specify that mitochondrial reads correspond to the host plant contamination. I also expect that you observe more host (degraded) DNA in yellow than in green leaf samples.

Did the 3454998 reads included mitochondrial DNA or you already removed those from this number. Same question for OTUs.

Response: Nonbacterial ASVs (1719ASVs of 70693 reads were assigned to Mitochondria or Chloroplast) were then removed. Overall, paired-end sequencing resulted in 7491619 high-quality reads, and these reads could be assembled into 13911 bacterial ASVs at 100% identity. (lines 160-164, page 7)

15. Here you mentioned "OTUs with 100% identify" but you said you used a 97% threshold. Please clarify. Also, did you evaluate the proportion of chimeric OTUs (pairing of reads from different organisms) and if so, how did you control for them?

Response: Thank you for your comments. Sorry for the error. It should be "ASVs with 100% identify".

We employed the denoising approach of DADA2 in the Amplicon Sequence Variant (ASV) analysis, which seamlessly combines the denoising procedure with chimera removal. (lines 145-147, page 6)

Consequently, it becomes challenging to accurately assess the proportion of chimeric ASVs. However, we approximated that following the application of DADA2 in the ASV analysis, around 3.0-8.0% of sequences were eliminated.

16. L146- Why did you perform this analysis at the OTU level? Again did you assume it's representative of species? Please justify.

Response: Notably, the differences between yellow and green leaves are most pronounced at general level. Because we focus on the difference in leaf stage, there we conducted the analysis at the genus level in the revised manuscript. (lines 170, page 7)

17. L147-148 – I am not familiar with this diversity statistic using the cumulative number of OTUs. Could you please explain or at least provide a reference? It seems that you did not perform per-sample OTU rarefaction curves before analysis your samples. This is an essential step to normalize your samples because of the great variability in reads/OTUs expected among your samples, and also to ensure that sequencing was deep enough to capture captured all the microbial diversity. Without this step, your analyses will be strongly biased toward samples with higher microbial DNA yield and over-representation or rare OTUs, while samples with low yield will be highly sensitive to contamination and might induce a lot of variability in your subsequent analyses. For instance, the great differences you might observe between green and yellow leaf microbial communities, or between endosphere and phyllosphere, might just result in different microbial DNA yields between these different types of samples. These biases will be greatly amplified if you performed (as I suspect) all your subsequent analyses on absolute abundances rather than on relative abundances. Especially, I expect that treatment used to extract endophyte microbes generate more DNA due to cell (both microbe and host plant) degradation/contamination, and so more yield than for epiphyte sample. Similarly, I expect that cells are more degraded in yellow than in green leaves, and generate more DNA yield.

Response: Thank you very much for your comments. In the revised manuscript, we have detailed the cumulative number of ASVs (total ASVs across all samples for either green or yellow leaves, please see Figure S5).
Additionally, we conducted rarefaction analyses for each sample, with the curves demonstrating that the sequencing depth was sufficient to capture the microbial diversity (lines 233-235, page 9, also see Figure S4.).
Most importantly, we also utilized relative abundance as the basis for our data analysis.

18. L148-151 – Unless I misunderstood (but then you should clarify that your statistical approach considered it), your PERMANOVA tested for season (S), location (L) and leaf stage (LS) as additive factors (S+L+LS), but do not test for interaction between those factors (SxLxLS), which is expected to be quite high. For instance, the effect of season on your microbial communities might be dependent of different local climatic conditions in locations (SxL interaction). Season and leaf stage might also covariate (SxLS) because the proportion of yellow leaves might be higher during dry season and you should see this effect on microbial community even if you sampled equally yellow and green leaves in each season.

Response: We have revised the manuscript based on your feedback. The interactions between these factors were analyzed, and the results are presented in Table S2.

19. L154 -Please clarify what VIF was used for.

Response: This section has been removed from the revised manuscript due to its inadequacy in explaining the biological processes of leaf aging.

20. L157 – Why did you use abundance of dominant phyla? What is your criteria for this dominance? Also, did you use relative of absolute abundance? Absolute abundance will be strongly biased by total microbial yield (see my previous comments) and will reflect technical more than biological effects. Why removing underrepresented phyla? Even rare, they might reveal important biological processes if your samples have enough read and are correctly normalized (see my previous comments). Also, why focusing of phyla and not OTUs? And why not intermediate taxonomic levels (Order, Family, Genera), which could increase your statistical power? The results you will obtain with phyla only might only reflect very long term evolutionary processes rather than more recent/specialized biological processes, as Actinobacteria and Proteobacteria, for instance, contain different orders and families of bacteria that diverged for billion years ago and having completely different evolutionary routes, ecological niches and thus, different functions.

Response: In the revised manuscript, we have removed the section pertaining to Linear Mixed-Effect Regression (LMER) due to its inadequacy in elucidating the biological process effectively. As you rightly pointed out, phyla is not proper taxonomic level instrumental in revealing biological processes. Therefore, we have identified the core ASVs and estimated their ecological processes (lines 187-194, page 8). Please see Table S5 & Table S6. Furthermore, we have replaced the use of absolute abundance with

relative abundance.

21. L. 161 – Could you justify why you considered location and seasons as random effect?

    Response: We have removed this section (see above). However, Principal coordinates analysis (PCoA) of taxonomical similarity based on Bray-Curtis distances for different season and different site of rubber tree leaves were performed in the revised manuscript. Please Figure S3.

22. L. 169 – Please justify why you consider taxa (phyla? justify) with abundance >0.1% and present in at least 60% of samples are core taxa?

Response: Recognizing that some rare species may play a crucial role in the overall community, we have employed the Sloan model, as recommended by Reviewer 1, to define the core ASVs for bacteria and fungi. The use of the Sloan model aids in distinguishing truly pivotal taxa from those that are common yet less essential, thereby enhancing our comprehension of microbial community structures and functions. (lines 187-194, page 8).

23. L. 173-175- I am very skeptical about functional predictions from rRNA16s/ITS. Those markers could just be used to infer broadly (as the genus level) the taxonomic composition of a microbial community but cannot summarize the incredible genomic complexity and variation in term of diversity and functions within a single microbial genera. Results from such functional predictions should be taken very cautiously by authors.

Response: Thank you for your insightful comments. In the revised manuscript, we have utilized PICRUSt2 to predict the bacterial KEGG pathways in both yellow and green rubber tree leaves (line 184, page 8). Please see Figure 3 and Figure 4.

As for fungal functions, despite the coarse nature of FUNGuild in predicting fungal functions, we currently do not have an alternative. Nonetheless, we will exercise caution in interpreting the results.

24. L. 176-177: Could you clarify what do you mean by stochasticity ratio, determinist/stochastic community assembly?

Response: The Modified Stochasticity Ratio (MST) is an index developed with 50% serving as the boundary point, distinguishing communities with more deterministic assembly processes (less than 50%) from those with more stochastic assembly processes (greater than 50%). This index is used to simulate community processes by considering factors such as abiotic filtering, competition, environmental noise, and spatial scales. (lines 197-199, page 8)

25. L. 193-194 – Why only focusing on those particular phyla and nutrients?

Response: In the revised manuscript, we have removed the section pertaining to SEM due to its inadequacy in elucidating the biological process effectively.

However, the relationships between bacterial functional gene abundance (at KEGG

level2 as inferred from PICRUSt2) the nutrient contents were analyzed. Please See Figure 3&4.

26. L. 196-Here and elsewhere, I think authors meant "environmental filtering". Also, could you clarify what you were testing for here, and the method (FEAST) you used?

Response: Thank you for your comments. We used FEAST to estimate the percentage of epiphytic (endophytic) microorganisms originated from the leaf interior (surface). Thereby, we aim to find the reason why the differences in diversity between green and yellow rubber trees leaves. (lines 204-208, page 8)

Results

27. L. 205 – What do you mean by "OTU level?" Diversity? Also, this is not clear if p-values are associated to differences between leaf types or between compartments. Also, here and elsewhere, remind the statistical test associated to each p-value.

Response: Apologies for the previous lack of clarity. In the revised manuscript, we have analyzed PCoA analysis based on the bacterial and fungal compositions at a general level. For the *p*-values associated with leaf stages and compartments, please refer to Table S2.

27. Figure 1 – One can observe some clustering of samples according to unidentified factors. Probably season and/or location? I would have appreciated the same PCA displayed with location and seasons.

Response: We have revised Figure 1. Additionally, we have performed the Principal Coordinates Analysis (PCoA) to examine the variations related to locations and seasons, which are presented in Figure S3.

29. Figure 2 – Did authors actually performed all the analyses based on microbial absolute abundance?

Response: In the revised manuscript, sections about microbial absolute abundance have been updated to relative abundance instead.

30. Table S2 –Again interactions between factors were not rested. The residual (unexplained) variance is not showed. With the exception of endophytic bacteria, variance explained by leaf stage is negligible compared to season and location.

Response: Interactions between various factors were analyzed, and the findings, including the residual (unexplained) variance, are detailed in Table S2.

31. L. 218 – Please remind the statistical test associated with each p-value. Did you correct p-values for multiple tests (Bonferroni)?

Response: We correct p-values for multiple tests by using Bonferroni method. (line 182, page 8)

32. L. 223-224 & Table S7 – "the majority of core bacteria OTUs", also please clarify in what proportion. Why not providing the detailed taxonomy of each OTU? As mentioned before, those two dominant classes probably harbor diverse genera of

bacteria with very contrasted lifestyles. Phyllosphere-associated bacteria are very well studied and it is known that Alphaproteobacteria, for instance, are dominant in the phyllosphere (eg. Methylobacterium, Beijerinckia, Sphingomonadales…), but some are more associated for instance, with the Rhizosphere (many but not all Hyphomicrobiales).

Response: "a majority (more than 70%) of them being members of α-Proteobacteria and Actinobacteriota". (line 285, page 11).

In the revised manuscript, Table S7 has been replaced with Table S5. Additionally, detailed taxonomic classifications for each ASV have been included.

33. L. 225-227 – Why not providing details for fungi, as you did for bacteria? Please justify.

Response: Thank you for your comments. The details for core fungal ASVs are shown in Table S6.

34. L. 238-242 and Figure 3C – This is a statistical effect, not a biological observation! You expect that a diversity index is positively correlated with the abundance of taxa, and especially with the most abundance ones, since this index is calculated from those abundances!

Response: Thanks for your comments. We have deleted Figure 3C.

35. L. 244-251 – You did not test for gene but for functional enrichment. Again, you cannot predict the gene composition of an organism just from a single marker gene (16s or ITS). You could just broadly and cautiously infer their function from strongly biased and incomplete databases of microbes with similar 16s/ITS sequences and with known genomes and metabolisms. Also "Epiphytic functional genes do not seem too different between the green and yellow leaves" is a very vague statement.

Response: Thanks for your comments. We just predicted the potential gene function based on the sequences. We have revised it according your comments. In the revised version, PICRUSt2 was utilized to predict the bacterial KEGG pathways in both yellow and green rubber tree leaves. (lines 184-185, pages 8)

36. L 251-252 and Figure 4B: "This indicates that the increase of saprophytic bacteria promotes the nutrient relocation for yellow leaves "could you please clarify how you came to this conclusion? Also, you did not experimentally demonstrated the mechanism. Hence, this is a correlation, not a causality!

Response: Thanks for your comments. As you indicated, we did not conduct experiments to support this conclusion. Therefore, we have removed this section from the manuscript.

37. L. 264-269 and Figure 5C: Again, correlation between the most dominant taxa abundances and the most meaningful function abundances are expected since both measurement directly derive from OTU abundance per sample. This is a calculation effect, not a biological effect.

Response: Thanks for your comments. As you said that this is not a biological effect.

We have deleted it.
In the revised version, PICRUSt2 was utilized to predict the bacterial KEGG pathways in both yellow and green rubber tree leaves. (lines 184-185, pages 8)
Subsequently, we examined the relationship between nitrogen concentration (and organic matter) and the gene abundance of metabolic functions. Our aim was to elucidate the biological processes underlying leaf senescence in rubber trees. (Please see Figure 3 and Figure 4)

38. L. 277-279 and Figure 6B – "a decrease in total nitrogen concentration in yellow leaves increases the abundance of α-Proteobacteria - You did not experimentally demonstrated the mechanism. Hence, this is a correlation, not a causality!

Response: As you indicated, the findings only demonstrated a correlation between nitrogen concentration and the abundance of α-Proteobacteria. Consequently, we have amended the manuscript in accordance with your comments.

39. L 282-289 and Figure 6C – Intuitively, I think those correlations just reflect statistical effects, no biological processes. All these statistics are dependent because calculated from microbial abundances so we could expect they are all strongly correlated together. Response: Thanks for your comments. We have deleted Figure 6C. In the revised manuscript, we scrutinize the factors contributing to the higher bacterial diversity observed within yellow leaves compared to green leaves from other perspective. For instance, the outcomes of *FEAST* analysis reveal that during leaf senescence, there is an increased likelihood of bacterial ingress into the leaf, potentially attributed to the breakdown of the leaf cell wall. (lines 295-297, pages 11)

Discussion.

40. L. 322 – Actinobacteria are known to be saprophytic. Authors did not discuss it.
Response: Thanks for your comments. We have discussed it in the discussion section. Actinobacteria are a group of saprophytic bacteria widely distributed in soil, water and plants, and play an important ecological role in recycling substances in the nature. (lines 328-330, pages 12-13)

41. L. 327-328 – Concepts of stochasticity of community assembly and r-selected microbes were presented in hypothesis but not explained in the introduction.
Response: Thanks for your comments. We have explained stochasticity (lines 50-51, page 3) and *r*-selected microbes (lines 75-76, page 4) in the introduction section. Please see the revised manuscript.

42 L.335 Some Hyphomicrobiales (syn. Rhizobiales) typical of the phyllosphere (Methylobacterium) were shown to have community shift during leaf development https://journals.asm.org/doi/epub/10.1128/mbio.03175-21
Response: Thank you for recommending the paper; it has enhanced our understanding of leaf microbiota. (lines 391-395, page 15)
We have included this reference in our main text. (line 650, page 21)

REVIEWERS' COMMENTS:

Reviewer #1 (Remarks to the Author):

I appreciate that the authors have addressed most of my comments extensively and humbly. However, a few comments remain, mostly stemming from the changes they made. I am not entirely satisfied with these changes, and I believe some reorganization or further clarification might be necessary:

L177: When describing the use of the PERMDISP2 algorithm for testing multivariate homogeneity of group dispersions, this should precede the PERMANOVA as it verifies their assumptions. Alternatively, it could follow the PERMANOVA as a post hoc verification of their assumptions. It is therefore inappropriate to insert a discussion on Generalized Linear Models (GLMs) between the PERMANOVA and the PERMDISP2 algorithm.

L175 and L178: When discussing GLM models, I notice another inconsistency in the authors' description. They mention Generalized Linear Models on line L175, but only introduce the acronym on line L178, as if for the first time. More importantly, GLMs typically employ distributions other than the normal distribution. Which distribution was used: gamma, Poisson, negative binomial, binomial? This should be explicitly stated.

L184-L186: This sentence does not make much sense as it seems to mix bacterial and fungal sequences. I suggest the sentence could be revised as follows: "PICRUSt2 (version 2.2.0) was used to predict bacterial functions [59-60]. Fungal functional classification in the samples and the abundance of each functional type in different stages were obtained using FUNGuild." It might be necessary for a native speaker to review the newly added text.

L209-L213: I do not fully understand the purpose of this paragraph. It constitutes a very short subchapter containing only one paragraph, so it might be better integrated into the preceding section. Additionally, this paragraph merely summarizes what has been mentioned in the previous paragraphs, making it seem redundant.

L226-L227: It is crucial to include the p-value obtained from PERMDISP2. If I understand correctly, both PERMDISP2 and PERMANOVA yielded significant results. However, the manuscript later states on lines L251-L252 that PERMDISP2 confirmed the PERMANOVA results, which is a misunderstanding of how PERMDISP2 works. If PERMDISP2 is significant, it actually questions the PERMANOVA results, suggesting that the significant PERMANOVA results might be due to differences in beta diversity rather than group

composition. Therefore, the p-value for PERMDISP2 must be explicitly stated, and the results of PERMDISP2 and PERMANOVA should be thoroughly discussed in relation to each other.

L388-L395 and other parts of newly added text: The text seems to contain grammatical errors (e.g., "become more importance" should be "more important", "senesce" should be "senescent"). The newly added sections should be carefully reviewed for English language accuracy.

Reviewer #2 (Remarks to the Author):

Authors have significantly improved their manuscript. I enjoyed reading this new version and have no major comment at this point.

Minor comment:

L-394: Methylorubrum is not a valid genus name (but some databases like SILVA and NCBI are not updated). Please use Methylobacterium when referring to Methylobacterium-Methylorubrum.

Responses to reviewers' comments

Manuscript number: COMMSBIO-24-0195-A
Title: Assembly and maintenance of phyllosphere microbial diversity during rubber tree leaf senescence
Journal title: Communications Biology

Reviewer #1

I appreciate that the authors have addressed most of my comments extensively and humbly. However, a few comments remain, mostly stemming from the changes they made. I am not entirely satisfied with these changes, and I believe some reorganization or further clarification might be necessary:

1. L177: When describing the use of the PERMDISP2 algorithm for testing multivariate homogeneity of group dispersions, this should precede the PERMANOVA as it verifies their assumptions. Alternatively, it could follow the PERMANOVA as a post hoc verification of their assumptions. It is therefore inappropriate to insert a discussion on Generalized Linear Models (GLMs) between the PERMANOVA and the PERMDISP2 algorithm.

Response: Thanks. We have now removed this section to next paragraph (Lines 382-383).

2. L175 and L178: When discussing GLM models, I notice another inconsistency in the authors' description. They mention Generalized Linear Models on line L175, but only introduce the acronym on line L178, as if for the first time. More importantly, GLMs typically employ distributions other than the normal distribution. Which distribution was used: gamma, Poisson, negative binomial, binomial? This should be explicitly stated.

Response: Sorry for the overlook in describing the Generalized Linear Models. For the models on L175, it should be Linear Models (LMs) not Generalized Linear Models.

For L178: A generalized linear model (GLM) based on negative binomial distribution was used to detect if microbial ASVs on yellow leaves were significantly enriched or depleted compared to those on green leaves (Lines 383-385).

3. L184-L186: This sentence does not make much sense as it seems to mix bacterial and fungal sequences. I suggest the sentence could be revised as follows: "PICRUSt2 (version 2.2.0) was used to predict bacterial functions [59-60]. Fungal functional classification in the samples and the abundance of each functional type in different stages were obtained using FUNGuild." It might be necessary for a native speaker to review the newly added text.

Response: Thank you. Revised as suggested (Lines 390-392).

4. L209-L213: I do not fully understand the purpose of this paragraph. It constitutes a very short subchapter containing only one paragraph, so it might be better integrated into the preceding section. Additionally, this paragraph merely summarizes what has been mentioned in the previous paragraphs, making it seem redundant.

Response: We added this paragraph in accordance with the journal's formatting requirements. But we agree it is better to merge it with the previous section, as we do now in the revision.

5. L226-L227: It is crucial to include the p-value obtained from PERMDISP2. If I understand correctly, both PERMDISP2 and PERMANOVA yielded significant results. However, the manuscript later states on lines L251-L252 that PERMDISP2 confirmed the PERMANOVA results, which is a misunderstanding of how PERMDISP2 works. If PERMDISP2 is significant, it actually questions the PERMANOVA results, suggesting that the significant PERMANOVA results might be due to differences in beta diversity rather than group composition. Therefore, the p-value for PERMDISP2 must be explicitly stated, and the results of PERMDISP2 and PERMANOVA should be thoroughly discussed in relation to each other.

Response: Thanks for the very helpful comment. We have now added the p-value and revised to make sure the statement for the results of PERMDISP2 and PERMANOVA are correct.

Additional PERMANOVA results (Endophytic: $R^2 = 0.07$, $p = 0.001$, Epiphytic: $R^2 = 0.03$, $p < 0.001$) and PERMDISP2 analysis (Endophytic: $p = 0.535$, Epiphytic: $p = 0.531$) indicated significant differences in the bacterial community between green and yellow leaves, but no significant difference was observed in the fungal endophytic community ($p < 0.001$ for both PERMANOVA and PERMDISP2 tests) (Table S2 and Table S4). See Lines 95-99.

The findings from the PCoA ordination and PERMANOVA analysis ($R^2 = 0.11$, $p < 0.001$) of functional gene compositions (KEGG level 2) revealed significant differences in endophytic bacterial functional gene compositions between yellow and green leaves (Figure S6A and Table S4), while no differences were observed in epiphytic compositions. The PERMDISP2 analysis for endophytic bacteria ($p = 0.340$) also confirmed these results (Table S3). Lines 120-125.

6. L388-L395 and other parts of newly added text: The text seems to contain grammatical errors (e.g., "become more importance" should be "more important", "senesce" should be "senescent"). The newly added sections should be carefully reviewed for English language accuracy.

Response: Thank you for pointing out the grammatical errors. Corrected. We also have taken extra care to minimize the grammatical errors and typos throughout the ms (see those changes in blue color.

Reviewer #2:

Authors have significantly improved their manuscript. I enjoyed reading this new version and have no major comment at this point.

Minor comment:

1. L-394: Methylorubrum is not a valid genus name (but some databases like SILVA and NCBI are not updated). Please use Methylobacterium when referring to Methylobacterium-Methylorubrum.

Response: Thank you. Changed.