We are extremely thankful for the very detailed comments on our manuscript. We acknowledge that these days it is very difficult to find reviewers who will take the time to carefully review a paper. We will similarly respond, as best as possible, to all the reviewers' comments.

Responses to reviewers' comments are below in italics.

Overall comment:

This study presents the fungal endophyte diversity and composition within the Rubiaceae family. Fungal endophytes communities are compared across host species, between contrasting locations, between different tissue types, and development stages. This manuscript is very interesting, and the findings shall contribute to generate interest among the readers, especially considering that knowledge on fungal endophytes in tropical rain forests is still scarce. Having said that, I think that there is space for improving the manuscript and I offer my comments/recommendations.

R/ Thank you for the positive comment and I believe that with the reviewers' suggestions, the manuscript will be much improved.

Evaluation of the different components of the article:

Title:

The article title could be enhanced regarding the hypotheses explored: "Exploring Rubiaceae fungal endophytes across contrasting tropical forests, tree tissues and development stages."

R/ Appreciate the suggestion for an enhanced and more comprehensive title. We've changed it to: Exploring Fungal Endophytes in Tropical Rubiaceae: Effects of Forest Regions, Tissue Types, and Developmental Stages. (L1-2).

Abstract

The abstract presents the main goal of the study and the associated results. I have a few minor remarks:

L29-30: This sentence could be deleted.

R/ Changed.

L30: I suggest to detail more the method as such: "Sapwood, mature and young leaves were collected from 47 Rubiaceae species in old-growth forests in Golfito and Guanacaste, Costa Rica. Fungal diversity and composition was assessed using metabarcoding of the ITS2 nrDNA region."

R/ A modified version of this suggestion was included. See L28-30.

Introduction

The introduction explains the motivation for the study, even though an accent could be made on why the authors chose to study fungal endophytes. Endophytes are broader than just fungi, maybe could add details to this point. Hypotheses need also to be refined to enhance clarity.

R/Thank you for the feedback, we did refine the hypotheses and provided a more detailed explanation or rationale for studying endophytes. Many of those changes are highlighted in the revised

L50-51: It is worth mentioning here that there are also studies on root fungal endophytes, not only concerning mycorrhizal and free-living forms: "but traditionally understudied compared to root fungal endophytes, mycorrhizal and free-living forms".

R/We added the statement. L55-56

L53-54: While the introduction on the role of endophytes is smoothly progressing, the mention here feels somewhat abrupt and it needs to be more developed.

R/We reorganized the ideas in the first paragraph of the Introduction to enhance clarity and coherence while keeping the dual aspects of their importance without making the transition feel so abrupt. Hopefully, these are clearer.

L56: More details are needed on horizontal transfer and why it is considered the main colonization process (for the authors) since vertical transfer is extremely important as well.

R/ We agree that both horizontal and vertical transfer are important in endophytes, even though those that we studied are mostly horizontally transmitted. To not extend the Introduction or distract the readers from the main topic of the study, we removed that claim and rephrased this section to address other relevant remarks also suggested by the other reviewer's comments.

L72: There is a jump from Rubiaceae family to a specific genera, and more details are needed to explain why the focus on Palicourea?

R/The last comprehensive inventory of Rubiaceae in Costa Rica we were able to find: a handbook published 10 years ago, reported 458 species and 89 genera. Recent updates have revealed that the genus Palicourea actually includes 91 species, rather than the 44 reported in the handbook, highlighting that the earlier data is outdated. We used Palicourea as an illustrative example of this discrepancy. Unfortunately, we could not find additional updated examples. We have revised the paragraph to clarify this point and hope it addresses your concern.

L77-85: Emphasize that you are studying fungal endophytes. Endophytes are much broader than only fungi. Why fungal endophytes can be more clearly enhance through out the introduction.

R/We have added the term 'fungal' in line 80 to clarify our focus. The emphasis on fungal endophytes is now established in the title and explained further throughout the Introduction. We have tried not to imply that our study extends to endophytes in a broader sense, except in cases where we make statements that do apply to endophytes in general.

L78: The first hypothesis is clear and interesting. What is the expect outcome, ex higher diversity in Golfito compared to Guanacaste due to higher rainfalls in this region?

R/ General comment about our hypotheses: We sought to explore a broad range of potential outcomes by testing various factors, so to maintain clarity and focus, we formulated two-tailed hypotheses that predict differences without specifying the direction. Detailed expectations and the rationale behind these hypotheses are thoroughly discussed in the Discussion section, where we provide a nuanced interpretation of our research approach and findings.

Now specifically about hypothesis i.: Our anticipation was that fungal endophytic diversity would be higher in Golfito, a tropical rainforest, compared to Guanacaste, which features a seasonal dry forest and other distinct biogeographical factors. This expectation arises from a combination of unique biogeographical and environmental conditions of each region, beyond just precipitation. We have elaborated on these aspects in the Methods and Discussion sections to provide a more comprehensive understanding of the expected outcomes.

L80: The second hypothesis could be more refined, and more details on the expectation can be given. It can be broken down into more than one. First on plant and trees, second on tissue type and development stage, for example.

R/ We have revised our hypotheses to better address the distinct aspects. Hypothesis ii now specifically concerns differences between plants, while Hypothesis iii addresses tissue types and developmental stages, including the specific expectation of tissue-specific distributions (L86). Additionally, we included relevant studies starting in line 60 to support our hypotheses; these new references support the notion that factors such as tissue type and growth stage influence endophyte composition.

L82-83: No hypotheses were given concerning fungal host specialization and this could be enhanced. To what question or hypotheses do the CLAM test and indicator taxa respond to?

R/ We used these tests to identify which fungal groups contribute to the differences observed in our study. We have revised the text to clarify that these analyses were intended to investigate the presence of specialist or generalist endophytes and their associations with specific environmental conditions or host characteristics.

L84: What is meant by "processes"? If it is diversity and composition, these are not processes. Community processes refer to selection, dispersion, drift, speciation. Perhaps a more precise word is needed here.

R/ We originally used "processes" to refer to the factors shaping community assemblages. But we understand the need for precision in terminology. To clarify, we have replaced "processes" with "factors" to more accurately reflect the host and environmental characteristics influencing endophytic communities (L80).

Materials and methods

This section can be more detailed and here are my specific comments:

L88 Study site: Geographical coordinates can be added to the text and also to the Table1. It would also be helpful to provide a brief description of the Figure 1 in the text.

R/ Because we have over 20 sets of GPS coordinates, we thought this would clutter the text and not significantly aid comprehension. However, we now added all the GPS locations to Figure 1. We also added a very brief description of the figure in lines 94 -96.

L100: To be more precise, talk about individuals instead of plants

R/ Fixed (L107).

L100: The selection suggests a broad sampling approach, how were the 47 species selected? What was the criteria?

R/ Our sampling approach involved selecting plant species from old-growth forests, which provided a diverse and extensive pool of potential species. We did not have control over the exact locations of plant species; rather, we sampled plants as we encountered them within these forests. The variability in plant distribution—sometimes finding several individuals in close proximity, other times walking for kilometers without encountering any—was inherent to the natural setting.

To maximize the diversity of endophytes, we aimed to include as many different plant species as possible, thereby increasing the range of ecological niches. Our sampling process was designed to be random within these constraints, adhering to appropriate scientific study design principles.

For sapwood collection, we specifically targeted woody trees with a minimum trunk diameter of approximately 10 cm. This was the only criterion we applied in this regard and the reason why we only have 23 sapwood samples.

We hope this explanation clarifies the selection process. We have added some more details (L110-116) in the text to provide a clearer sampling approach and address your comment.

L101: Here "plants" means "species" no? Start the sentence such as "The taxonomy was confirmed by... "

R/ Yes, all the individuals/plants we sampled. We changed the sentence per your suggestion. See L109-110.

L103: Were the branched sampled in the upper or lower crown of the tree?

R/ We sampled three different branches from each tree, selected randomly. Our criteria required that each branch have at least three nodes to ensure we could collect both young and mature leaves. We aimed to sample from the upper, middle, and lower crown, as well as from different sides of the tree (left, right, front, and back) to maximize representation. However, there were limitations in some cases, particularly with very tall trees that we could not reach using a pole pruner or by climbing. We added a statement to address your comment directly in the text, in line 112.

L104-105: For the differences in development stages, why would three nodes be sufficient to separate new and mature leaves?

R/We believe three nodes are sufficient to distinguish young from mature leaves due to branch structure. New leaves are on the most recent node at the tip, while mature leaves are on older nodes further back. Per Chen et al. (2016), leaves on the leftmost child nodes are older, while those on non-leftmost nodes closer to the tip are younger. See: https://doi.org/10.1016/j.ejc.2004.12.008.

We chose a 3 nodes separation standard as a pragmatic approach for practical consistency across species with varying growth rates and phyllochron. This fixed node spacing effectively captures the transition between early and late stages of development without much overlap. Field observations confirmed differences in size, color, and texture for most sampled species. This approach is also practical as not all plants have more than 3-4 nodes, ensuring representative sampling.

Other studies use a 3-node separation to sample new and mature leaves, for example, Lin et al (2022) used the third and sixth leaves (https://doi.org/10.3389/fmicb.2022.872034)

L106: Again, individuals?

R/ Rephrased.

L110: Was this protocol for leaf surface sterilization developed specifically for this study or is it a standard protocol?

The protocol is well-established and standard procedure employed by this and other research groups. We have included a reference to a recent publication where this protocol was utilized, which also cites other studies that have adopted the same method. L123.

L113: Less details are provided concerning the sapwood procedure. I'm curious to know why the sapwood sampled were singed? (slightly burned?) I have never worked with sapwood sampled before.

R/The purpose of surface sterilization is to eliminate contaminants and external microbes, to ensure that we only capture those microorganisms that inhabit the endosphere. The main difference between the protocols for leaves and sapwood arises from their inherent characteristics: leaves are external and exposed to the environment throughout their development, while sapwood is an internal tissue, initially protected by bark. We acknowledge that this distinction might not be immediately clear to everyone and have added further details in the methods section (L123-125):

The singeing process involves briefly exposing the excised sapwood samples to a flame to eliminate potential surface contaminants introduced during handling. It is performed as a cautionary measure. This is comparable to practices in microbiology where microscopy slides, culture media plates, and laboratory tools are singed or briefly heated to remove contaminants.

L116: Even if the experimental procedures were outsourced or used in kits for the DNA extraction, it would be helpful to give the key steps for each procedure: DNA extraction, the PCR, what were the negative and positive controls, which reverse and forward primers were used. Why were primers fITS7 and ITS4 suitable?

R/Although it is not standard practice to add the methods that these companies use (such as Naturalis, Novogene or Psomagen, for example), we were able to get some basic information from them regarding these methods. We have added more details in lines 134-147. Regarding the protocol for DNA extraction, we wrote "using the manufacturer's instructions" because we did not modify anything from those. We also added the information on positive and negative controls.

L131-133: What cutoffs were used during the curation process to filter quality, trim and remove low abundance reads? Provide specifics about the quality thresholds applied.

R/As suggested, we have revised the methods section to specify the cutoffs and quality thresholds used for data curation, including quality filtering, trimming, and removing low-abundance reads. Details in lines 153-158 have been updated to reflect these changes.

L141: Which version of R was used?

R/L152-153 already describe the R and RStudio version: R v. 4.0.3 in RStudio v. 2022.12.0+353. We removed 'in RStudio' from now line 158 to avoid the repetition and potential confusion.

L142: Give more details to why alpha diversity indices were transformed to Hill effective species numbers. Perhaps cite that this was shown to provide more robust estimates of diversity (Jost et al., 2010)

R/L171 now includes a reference to a couple of papers discussing the importance of transforming the indices to Hill numbers. We also added a bit more explanation (L171-174)

L147: Following a PERMANOVA, was the homogeneity of dipersion checked to ensure that differences in PERMANOVA were not due to differences in disperion? It can easily be checked using the betadisper function in the vegan package.

R/ As noted later on, we did check for homogeneity of dispersion. We have added further details to the methods section to outline our workflow, including the main functions used. Please refer to lines 179-182 for these additional details.

L148-151: The sentence is dense, perhaps break it down. Restate also what were the 6 categories. Chose either categories or habitat.

R/Thank you for the feedback. Per your suggestion, we have broken down the long sentence to enhance clarity. We have chosen 'category' as it more precisely describes the different levels or types within each factor, avoiding the ecological connotations associated with "habitats." We also clarified the six categories.

Results

Consider structuring the results with sub-section headings to enhance the cohesion and flow, maybe according to the hypotheses? Otherwise, yes the results were described and interpreted correctly.

R/Thank you for the suggestion. We appreciate the idea of using sub-section headings to enhance cohesion and flow. However, we believe that the results section, as currently structured, presents the observations clearly and is easy to follow. We will accept your suggestion to use subheadings in the discussion section, where it is crucial to contextualize our approach and results.

L162 the first line of the results section could easily be moved at the end of the materials & method, at the end of the paragraph ASV classification and taxa assigment.

R/We believe this sentence is part of the results and would prefer to keep it in the results section.

L178-179: English needs rephrasing such as: "Regarding the development stage, the taxo richness (q=0) was significantly higher in mature leaves compared to new leaves.

R/Fixed.

L188: It is hear mentioned that dispersion homogeneity was check but not mentioned in the materials and methods.

R/We have revised the methods section.

L198: Perhaps the paragraph on the taxonomy could be merged after L167, at the end of the first paragraph.

R/We would prefer to keep the sections as they are to maintain the narrative flow of our findings. Merging the taxonomy paragraph might disrupt what we believe is a logical progression of the results and would then require significant restructuring of the last paragraphs. I hope this is OK with the reviewer.

Tables and figures

The tables and figures are overall very nicely done.

R/Thank you.

Figure 1: L792: Is it necessary to add "Made per request.." Perhaps add it to the materials and methods text or in the acknowledgments. There is no legend concerning the colors of the climatic subregions, eventhough some are described in the legend text.

R/The map shows 24 climatic subregions across Costa Rica, but our sampling was limited to the subregions described in the caption to guide the reader in relation to our study. We have rephrased the acknowledgment of the figure's creation at the end of the caption for clarity. We believe it's important to maintain this note to credit and accurately reflect the role of the National Meteorological Institute in creating the figure as requested.

Figure 2: L799: More details in the legend could be given: "... (A) different locations, Golfito and Guanacaste, (B) different tissues, mature and new leaves and sapwood, ..."

R/ We are unclear on what the reviewer specifically wants here. We believe that the figure legend indicates what each subfigure is, and then there are legends on the bottom right that show the Rubiaceae tribe, geographic location, tissue, and developmental stage.

Figure 3: L806: Great figure, very clear! Just a suggestion to make it even more clearer: I found the CLAM test legend confusing. I would only say : "habitat specialization is indicated by the CLAM test". Because no colors means habitat generalization, no? Also, make the legend key wider for all legend text to spell out correctly.

R/Thank you for the feedback. You are correct that no color indicates habitat generalization, so we have updated the legend as suggested. Additionally, we noticed that the figure's size changed

when pasted, affecting the format and text clarity. This issue has been corrected, and the legend key has been adjusted to ensure all text is properly displayed.

Table 1: Geographical coordinates can be added to the text and also to the Table1.

R/See response above. We added all the GPS locations to Figure 1. Here we also fixed a mistake regarding the number of sites. Instead of 24 in Guanacaste, there are 23.

Discussion

I overall would suggest the structure of the discussion to be devided with sub-section titles to help the flow of the discussion. I suggest to first summarize the main findings and then address clearly each hypothesis in the order it was presented in the introduction. I also did not see how the first part of the second hypothesis (concerning the difference between plants and trees) was assessed and discussed, but perhaps I missed something.

R/ We have divided the Discussion into several sections. The first section summarizes the main findings per the reviewer's suggestion. What we did was to move the statements from the Conclusions to this first paragraph of the Discussion.

L247: I don't fully agree with this statement since species identity could drive the community, even though the results in this study did not find fungal endophyte communities to cluster according to taxonomy. It does not mean that plant taxonomy does not shape the community, the interpretation here is too abrupt.

R/ We have revised the statement to incorporate your feedback and ensure a smoother transition between points (291-295). Additionally, we have added references to the main text figure showing the NMDS at the tribe level, and to supplemental figure S5, which presents data at the genus and subfamily levels. We agree that species identity could influence community structure and have adjusted the discussion to reflect that, potential effects at finer scales, such as within genera, cannot be ruled out. Unfortunately, we do not have enough replicates at the species level to make such inferences. And while previous studies have observed community differences at broader taxonomic levels, our study focused on a single family.

L254: I don't really understand the link between the results concerning H1 and the temporal distribution of endophytes.

R/Thank you for pointing this out. We have revised this section to improve the flow of ideas and clarify the link between our hypothesis and the fact that endophytes can exhibit a temporal distribution. (L297-306)

L260: "higher organisms" do you mean macro-organisms? Maybe a more precise word is needed here.

R/The term is indeed ambiguous and can be confusing. However, "macro-organisms" is often associated with macrofauna, macroinvertebrates or macrobenthic communities. In any case, your feedback led us to reconsider the relevance of this sentence, and we decided to remove it during our revision of that paragraph.

L262-64: How are the environemental conditions related to higher variability in fungal endophyte communities. A clearer link is needed here to understand the interpretation, not just citing the Table 1 and Figure 1.

R/ That sentence was revised (L286-287). In addition, we further explain: Guanacaste sampling points encompass various subclimates (with different environmental conditions), as described in the Methods section and illustrated in Figure 1, whereas all sampling points in Golfito were from 1 type of subclimate. We believe this variability contributes to the observed differences in fungal endophyte communities. Table 1 reinforces this by showing that the standard errors (SE) for precipitation (± 45.03) and elevation (± 41.72) in Guanacaste, are significantly higher than those in Golfito, indicating greater environmental variability across Guanacaste's sampling points. We hope that the clear presentation of these data and figures will enable readers to draw these associations from the references provided, without the need for repetition or overly explicit explanations. We also tried to explain this in the next paragraph (starting on L311).

L266-270: Perhaps other unmeasured environmental factors could explain the distribution ? (ex. Humidity, wind, soil characteristics, forest structure and composition…)

R/You are correct. At the end of the discussion, we have added 2 paragraphs on the limitations of our study where we include other potential predictors of fungal endophytic communities that we did not measure.

L268 Do you mean variation of composition?

R/Indeed, we have added the clarification.

L269: neutral processes are not just dispersion limitation, but also consider ecological drift and diversification (Vellend et al., 2010; Nemergut et al., 2013)

R/ Yes, we are aware there are other factors. Our intention was to illustrate one example of a neutral process relevant to the ideas we are exploring. We incorporated ecological drift, as it aligns well with the concepts presented in that paragraph. L319

L270: I do not completely agree, there is also vertical transmission. The plant does not germinate without an initial microbiota. The future horizontal colonizing endophytes have also to compete with already present microorganisms.

R/ We have revised the paragraph to better balance the discussion of horizontal and vertical transmission.

L274-75: I suggest to keep all the idea about distance decay in one unique paragraph.

R/We revised and restructured the paragraph.

L282-288: The paragraph is overall confusing with the first sentence saying that both locations were dominated by host generalists, but the second part mentions a comparison with higher specialists. The mention "host generalist" refers to the trees or fungi? It is confusing. Revision of english wording is needed too. I suggest to develop the idea of host generalist, their potential influence/interaction with microorganisms, compared to host specialists if this is the idea of the paragraph. Or develop the biology, lifestyle, physiology of fungi generalists vs specialists.

R/We have revised the text for clarity.

L298: What do the two different compartment leaf vs spawood offer in terms of niche requirements for the endophytes to thrive?

R/We have dedicated one section to this topic ("Tissue type predicts endophyte community variations"). This entire section of the discussion addresses the niche requirements provided by the different compartments. We discussed key factors such as the chemical environment, nutrient availability, oxygen levels, entry points, dispersal mechanisms, and tissue structure. If you have any additional factors or specific suggestions that we might have overlooked, we would appreciate your input and are happy to consider including them.

Category	Foliar Tissues	Sapwood Tissues
Chemical Environment	Rich in bioactive compounds; variable	May have chemical defenses related to xylem sap
Nutrient Availability	Variable; influenced by leaf metabolism and external sources	High nutrient capacity stored in parenchyma cells and xylem sap but hard to access
Oxygen Levels	Generally adequate; high due to atmospheric exposure	Lower; limited within xylem cells
Entry Points	Stomata, trichomes, epidermal structures, wounds	Natural fissures, lenticels, entry through damage
Dispersal Mechanisms	Wind, rain, throughfall	Rain, environmental conditions, grooves, and fissures
Lifespan/Turnover	Shorter; periodic shedding of leaves	Longer; part of the tree's structural system

Tissue Structure	Complex; includes epidermal, mesophyll, and vascular tissues	Structured for water transport; includes living and dead xylem cells
Defense Mechanisms	Bioactive compounds (phenolics, alkaloids, terpenes); self-cleaning due to cuticle	Reaction zones, complex polymers (lignin, suberin); physical barriers (xylem vessels)
Pressure	Atmospheric pressure; changes with weather conditions	Internal negative pressure within xylem can limit movement
Exposure	Direct exposure to environmental conditions (temperature, UV light, etc.)	More sheltered from direct environmental exposure; primarily involved in internal processes
Disruption Risk	Higher risk due to frequent environmental interactions and physical damage	Lower risk from direct environmental disruption but affected by internal factors and pathogens
Microbial Interactions	High potential for interactions with a variety of external microorganisms	Potential for interactions with internal and external microorganisms but influenced by structural barriers

L342-345: Small contradiction here: It was mentioned that leaves are considered oligotrophic environments (L289-290), and here is written high nutrient availability.

R/ To clarify, the term "relatively oligotrophic" in line 344 refers to the surfaces of the phyllosphere, which are generally low in nutrients compared to other environments. In contrast, lines 393-399 refer to the leaf interior, where endophytes thrive due to high moisture and sufficient nutrient availability. Although nutrient availability in leaves may be lower than in some other environments, it is still adequate to support endophyte growth. We have added a comma in line 398 to distinguish between "high moisture content" and "nutrient availability" as separate factors, ensuring that "high" does not imply it applies to both factors simultaneously.

Discussion: do they talk about roots, other microbes? Endophtyes are not only fungi, maybe broader with bacteria, protists etc.. and talk about roots? An importance source of horizontal transfer is from the soil to the roots to the leaves.

R/In the added statements at the end of the discussion contemplating the limitations of our study we do mention soil and root endophytes.

L381: I would summarize the main results as done in this paragraph, at the very beginning of the discussion. And perhaps end the discussion opening to the challenges in the conservation /agriculture .

R/We accept the good suggestion.

L346-47: This is the expectation worth mentioning in the introduction for the second hypothesis.

R/ As mentioned earlier, we would prefer to keep our hypotheses two-tailed and elaborated on the specific expectations in the discussion. This approach allows for a more flexible interpretation of our results and aligns with practical considerations in our analysis.

L346-52: I would simplify this sentence and nuance the meaning behing since this study did not evaluate changes in genetic traits, tissue chemistry, topology, plant hormonal, physiological properties, physical characteristics to see the direct impact on the community composition.

R/We understand your concern, so we've polished the statement to improve flow and clarity. However, we believe it's important to mention these factors. When we chose to sample tissues at different developmental stages, we based our hypothesis on defense mechanisms, as literature suggests that young tissues are typically more chemically protected. However, since we did not conduct a metabolomic analysis of our samples, we had to rely on existing literature about tropical plants. Acknowledging these limitations is crucial because, just as chemical defenses can influence endophyte colonization, other tissue traits might also affect them. We included this to caution readers when interpreting our results.

L360: It is really true that plants in the tropics do not necessarily experience strong selective pressures to develop sophisticated chemical defenses? It may be worth checking this assumption again. It's not my area of expertise so I'm probably learning something new.

R/Our expectation was based on the optimal defense theory, which suggests that younger leaves might exhibit a different metabolic profile compared to mature leaves, potentially with higher chemical defenses. However, much of the literature supporting this theory focuses on temperate or boreal biomes. After reviewing studies specifically on tropical plants, we found that tropical plants do not necessarily produce higher quantities of chemical defenses against biotic threats. It is important to clarify that our reference to "higher" pertains to quantity rather than complexity or sophistication. Consequently, in tropical contexts, younger leaves may not necessarily have more secondary metabolites for defense than mature leaves. This finding aligns with the broader understanding of defense mechanisms in tropical plants as discussed in that paragraph.

L361-63: This is actually a very interesting results that could maybe be related to the notion of a core microbiome of the Rubiaceae family: a stable fungal community within the host across habitats, taxonomy species, and development stages. Maybe an idea to explore.

R/ We also believe this would be an interesting idea to explore. Unfortunately, we did not go any further with the analyses as we believe the manuscript is already quite extensive and has covered an adequate number of hypotheses and questions.

L381: I would not include leaves and sapwood as part of the phyllosphere, I would dissociate them.

R/Agreed. Also, to avoid confusion, we moved the Conclusions section to the first paragraph of the Discussion, which is where we summarized the main findings.

References

L411 and L533: Check the reference format.

Review by anonymous reviewer 1, 29 Jun 2024 13:59

The authors have produced a manuscript which details an important and understudied component of tropical ecosystems. By collecting a significant number of specimens across two broad geographic regions, at different leaf ages and tissue types they are able to show the importance of these regions for endophyte diversity and begin to show how this diversity is structured. It is well written and contains clear support for points made with the use of well put together figures in addition to supplementary methods. Although I think this study is worthy of publication, it could be significantly improved by adding in a little more detail with regards to how they reach their hypotheses, a few areas of methodology lacking clarity and an acknowledgment of where this study is limited. These points are detailed further below.

R/ Thank you. Hopefully, we have addressed all your comments in the revised manuscript.

Title and abstract are clear.

R/ We made slight modifications which were suggested by another reviewer.

Introduction

57 -would benefit from more info on environmental and geographical factors influencing this. You mention these in your hypotheses but it would be good to highlight studies where distance or topography have been revealed as potentially important in structuring endophyte communities. For example (Suryanarayanan, Murali, & Venkatesan, 2002; Suryanarayanan, Venkatesan, & Murali, 2003; Zimmerman & Vitousek, 2012; Cordier et al., 2012; Izuno, Kanzaki, Artchawakom, Wachrinrat, & Isagi, 2016). Many of these feature in your discussion but they could be clarified here to guide the reader to how you reach your hypotheses

R/ Thank you for the provided references, we made some additions and revised our statement following your suggestion (L61-64).

72-79 - your hypotheses line up with the literature but appear somewhat out of nowhere, can you refer to other studies where authors have looked at diversity patterns of endophytes within different plant species e.g Arnold et al. 2000; Donald et al. 2020 or different tissues Carroll 1988.

R/ We have revised the introduction to incorporate relevant studies, as suggested. This revision helps frame our hypotheses within the context of existing research and demonstrates how our study builds on previous work. Additionally, we have included a statement in lines 82-83 that directly addresses your concern.

Methods

100-101 - very limited replication within species, I.e. only one tree per species. In addition, the combination of species is not equivalent across sites. This means you cannot test for a species effect on diversity as all you would show here is differences between individual trees. In addition, any inference of differences in location could be down to the different combination of host tree species but your analysis cannot account for this. Your results are still of interest for such an understudied region, but this limitation could result in an over inference of importance. This needs acknowledging in the discussion, but also reappraising in your introduction and hypotheses. In your analysis and results, you cluster at the family level where you would have more replication.

R/ You are correct about the limitations related to species replication. We acknowledge that we used the term "species" interchangeably with "plant individuals," which caused confusion and have fixed that. Our primary goal was to explore diversity within the Rubiaceae family, encompassing various species from different genera, tribes, and subfamilies, rather than testing species-level effects. As such, our analyses focus on broader, non-taxonomic categories, where we have sufficient samples, and our results reflect that. We have revised our hypotheses to better reflect the scope and objectives of our study. Thank you for highlighting the need for this clarification.

Regarding the locations, we now refer to them as 'forest regions'. It is important to note that the plants were selected randomly and the combinations we present in the manuscript are confounded with the regions, as they arise from the biogeographic and environmental conditions of each area. Climate and the environment shape plant diversity and, by extension, the microorganisms that interact with them. While it would be valuable to study identical plant species in the same amount across very different forest regions, we were limited by their availability in the sites we chose to sample, we simply couldn't find them, which is typical in highly diverse tropical forests.

105 - can you clarify for people like me not good at maths, how many leaves per tree?

R/ We included the number in L113.

113- more detail required on "singed" method

R/There are limited steps involved in the singeing method, which, by definition, consists of briefly exposing the tissue to a hot flame to achieve a superficial burn. We have extended our initial statement (L122-124) to clarify the purpose of this step.

123- did you have any prc or extraction replication from your individual leaf samples?

R/Due to costs and workload constraints, we included only biological replicates, with each leaf section representing one replicate (six per individual in total), and no technical replicates were performed. We added a statement on L130-131 regarding this.

131-134 - you mention numerous filtering steps but there is no discussion of what thresholds these were set at, or a record of how many reads / asv s were removed at each step. This numbers should be stated since they give an indication as to the quality of your dataset. Did you have any blanks / extraction or pcr controls? Aside from sequence errors such as chimeras, how does this method account for potential contaminants?

R/We have revised the methods section (L152-164) to specify the cutoffs and thresholds used during the curation process, providing a clearer indication of our dataset's quality. We appreciate the comment about including details on positive and negative controls; these details were added in line 138 and then mentioned again in L164.

156- can you clarify what you mean by variable classes? This would help the reader understand what you mean when you refer to a specialist or indicator

R/Thank you for pointing this out. Just a few lines above, we clarified that the six categories refer to the 2 forest regions (Golfito and Guanacaste), developmental stages (young and mature), and tissue types (leaf and sapwood). We used these categories to determine habitat affinity and identify indicator taxa among the most abundant fungal families. We have reworded the statement to "from the six categories" in line 191 to better reference the categories previously mentioned in the text, avoiding unnecessary duplication. We hope this improves the clarity.

Results

176- how confident can you be in differences in endophytes from different tissues when you use different methods to conduct the extractions. The tissues themselves may be trickier to extract rna from?

R/ While we acknowledge that DNA extraction from wood tissues can be challenging due to the presence of lignin and secondary metabolites like phenolics, we used the same Qiagen DNeasy Plant Mini Kit® to standardize the method and avoid biases due to different components in the kits. This widely used commercial kit has also been used for bark samples in our previous studies. In addition, the sapwood samples were not too thick, and we performed thorough tissue disruption to prevent the presence of solid fragments, as outlined in L127-130. After disrupting the tissue in the FastPrep, the sample became fully powdery.

209-210 do you have enough replicates of host tree species to determine if they are specialists? See Novotny et al., 2002

R/Thank you for pointing this out. Unfortunately, in some statements we misused the term 'species' when referring to the individuals we sampled. We have made changes in the hypotheses and the methods always stated "plant partners available at the study sites" (L187-188) to clarify that it is not an aim of this study to draw conclusions at the species level of the plants. It is also not an

objective to provide insights into host specialization at specific levels as we acknowledge that we do not have enough individuals from the same species.

We do not discuss, for example, that specific fungal families are host specialists of specific plant species because, as you have pointed out, we lack enough replicates at that level to accurately determine that type of specialization. Our goal was not to draw broad-scale conclusions but to identify which fungal groups were driving the differences observed in our study.

The d' index analysis serves as a complement to both abundance, which explains the quantity of fungi present but doesn't tell us how this quantity is spread across different plants, and prevalence, which indicates the presence of fungi across sampled individuals but doesn't indicate the amount on each plant. The d' index combines both the occurrence (presence/absence) and frequency (how often fungi are found, which in this case is represented by abundance) to measure specialization.

By focusing on individual plants, we can observe specific interactions without needing multiple replicates. The d' index expresses how specialized our fungal families are in relation to the plant individuals we sampled, considering the abundance data. It allows us to determine if the fungi-plant interactions are random or exhibit specific patterns of specialization. If a fungus interacts with many different individual plants, it suggests a generalist strategy, indicating a cosmopolitan behavior. In contrast, a fungus that colonizes only one or a few individuals shows specialization at some degree.

While our results focus on the specific individuals we sampled from different species, we believe it is still relevant to present these findings, as the observed patterns could potentially be extrapolated to similar ecological contexts.

Discussion

264- I would caution against over inference of the role of environmental controls here given you only have two blocks of habitat sampled rather than any specific reported range of experienced environment with sufficient replication

R/We appreciate your caution and have carefully avoided over-interpreting the role of environmental factors. The original phrase "could be attributed to" was intentionally used to reflect this cautious approach. Given the known environmental differences between the two habitat blocks we tested, we consider these differences as one of several factors that might explain the variation in endophyte communities. We have revised the sentence (L309-310) to further ensure caution and to avoid any potential misinterpretations.

You need to have a section which acknowledges the limitations of this study. Although you reveal lots of great patterns relating to differences across two regions in CR, there should be

more caution or caveats made when referring to detected environmental effects or what your analysis o specialist / generalist taxa really can show.

R/We have added a statement acknowledging the limitations of our study and the need for caution when interpreting the patterns observed, particularly regarding the forest regions with different environmental factors (L434-455).

Regarding the analysis of specialist/generalist taxa, it is important to highlight that the CLAM test evaluates habitat affinity, as it is based on species relative abundance in two habitats. We consider we do have enough samples in each of the dual categories we tested (regions, tissues, developmental stages) to identify patterns of endophyte preference.

On the other hand, we intentionally did not delve into host specialization because, as mentioned before, it is not within the scope of this study to address this at any specific taxonomic level of the plant. With the d' index that also uses the term 'specialist' our goal was to go beyond abundance and prevalence, providing a nuanced view of the observed fungal-plant interactions.

In that case, although the analysis focuses on sampled individuals, the observed trends can still inform broader ecological patterns. For instance, consistent patterns among fungi can be extrapolated to similar ecological contexts within the Rubiaceae family.

More modifications to your methods to address the comments above will require the modification of your discussion to account for this. Otherwise, points are well supported with reference to a broad pool of relevant literature demonstrating how 5e research fits into the current state of art well.

R/We have revised our hypotheses and provided more details in the methods. We have carefully considered this reviewer's feedback, as well as the others, to ensure our study is robust and well-supported by the relevant literature. We hope these modifications adequately address all your concerns.