Author Cover Page

Article submission to PeerJ

Manuscript category: Research Articles

Collection: "*Endless forms: Advances in evolutionary analyses of biodiversity*"

Title: High-throughput metabarcoding of soil and litter samples across Amazonia shows the highest biodiversity in areas with low organic carbon

Camila D. Ritter[1,2*], Alexander Zizka[1,2], Fabian Roger[3], Hanna Tuomisto[4], Christopher Barnes[5], R. Henrik Nilsson[1,2§], Alexandre Antonelli[1,2,6,7§]

[1] Gothenburg Global Biodiversity Centre, Box 461, SE-405 30 Göteborg, Sweden

[2] Department of Biological and Environmental Sciences, University of Gothenburg, Box 463, SE-405 30 Göteborg, Sweden

[3] Department of Marine Sciences, University of Gothenburg, Box 463, SE-405 30 Göteborg, Sweden

[4] Department of Biology, University of Turku, Turku, Finland

[5] Natural History Museum of Denmark, University of Copenhagen, Solvgade 83, DK-1307 Copenhagen, Denmark

[6] Gothenburg Botanical Garden, Carl Skottsbergs gata 22A, SE-41319 Göteborg, Sweden

[7] Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford St., Cambridge, MA 02138 USA

*Corresponding author: Camila Ritter, camila.ritter@bioenv.gu.se. Phone: +46 723924447, Carl Skottsbergs gata 22B, Box 461, SE 405 30 Göteborg, Sweden

§Equal contribution.

Abstract

Background. Amazonia is a biologically megadiverse region, but our knowledge on its biodiversity and environmental determinants stems almost exclusively from aboveground organisms, notably plants. In contrast, the environmental factors that drive diversity patterns for microorganisms remain elusive, despite the fact that they constitute the overwhelming majority of organisms in any given location, both in terms of diversity and abundance.

Methods. We used recently generated operational taxonomic units (OTU) inferred from high-throughput metabarcoding of 16S (prokaryotes) and 18S (eukaryotes) markers to estimate richness and community composition of prokaryotes and eukaryotes in soil and litter (i.e., leaves and above-ground debris) across Brazilian Amazonia. Together with novel data on soil chemical and physical properties, we identify abiotic correlations of soil microorganism richness and community structure using regression, ordination, and variance partitioning analysis.

Results. Soil organic carbon content was the strongest factor explaining OTU richness (negative correlation) and community composition across all datasets. We found important effects also for other soil variables, including pH. There was no significant correlation between OTU richness of litter and soil for eukaryotes, and only a weak correlation between OTU richness of soil and litter for prokaryotes.

Discussion. Our results provide a large-scale mapping of the physical and chemical correlations of soil and litter biodiversity in a longitudinal transect across the world's largest rainforest. Our methods help to understand links between soil compounds, OTU richness patterns, and community composition. The lack of strong correlation between litter and soil richness suggests the complementarity of these substrates, and highlights the importance to

include both in biodiversity assessments. Massive sequencing of soil and litter samples holds the potential to greatly complement traditional biological inventories in advancing our understanding of the factors affecting tropical diversity.

Key-words: Brazil; Eukaryotes; Operational Taxonomic Units (OTUs); Prokaryotes; Rainforest; Soil microorganisms

Introduction

Tropical rainforests are mega-diverse and environmentally heterogeneous biomes, and their biodiversity has been shown to vary considerably over space. In Amazonia, the world's largest rainforest that covers most of northern South America, geology and soil physicochemical compounds are often considered crucial to regulate the biotic dynamics, vegetation, and diversity patterns at local to regional scales (Vogel et al., 2009; Laurence et al., 2010; Higgins et al., 2011: Tuomisto et al., 2016).

Comentado [h1]: It is not in the bibliography

For plants, diversity patterns and community composition are associated with the availability of soil nutrients (Laurence et al., 2010). Soil cation concentration has been identified as the potentially most important factor determining plant species composition and turnover in Amazonia (e.g. Baldeck et al., 2016; Tuomisto et al., 2016; Cámara-Leret et al., 2017) and also shows a less prominent effect on species richness (Tuomisto et al., 2014). Soil chemistry, in particular phosphorus, is also known to affect the taxonomic composition of microbial communities (Buckley and Schmidt, 2001; Faoro et al., 2010; Navarrete et al., 2013). In addition, pH is known to shape microbial diversity (e.g. Osborne et al., 2011; Kuramae et al., 2012; Barnes et al., 2016). Finally, geology and soil physicochemical compounds affect all ecosystems directly and indirectly, via biotic interactions among animals, plants, and fungi (e.g. Tedersoo et al., 2016).

Comentado [h2]: It is not in the bibliography

Soils are complex systems, and different soil layers may show different patterns of biodiversity (Hinsinger et al., 2009). For instance, Porazinska et al. (2012) found that

taxonomic composition varies between mineral soil and organic matter (litter) across biomes. High-throughput amplicon-based analyses such as metabarcoding (Taberlet et al., 2012) allow detailed examination of soil diversity patterns (Bardgett and van der Putten, 2014). However, these studies are usually focused on one or a few taxonomic groups, which make it difficult to draw general conclusions on the effects of soil properties on the overall determinants of biodiversity (e.g. Faoro et al., 2010; Laurence et al., 2010; Navarrete et al., 2013; Barnes et al., 2016).

Comentado [h3]: It is not in the bibliography

Although several studies have reported on the importance of soil compounds for biodiversity patterns and community structure, no unified pattern has emerged. With the exception of a few studies reporting results on arthropods and microorganisms (Basset et al., 2012; Ramirez et al., 2014; Prober et al., 2015; Tedersoo et al., 2016), the diversity of most inconspicuous, less studied groups of organisms remains poorly understood. To understand the evolution, maintenance, ecosystem function, and distribution of organisms, a greater focus on the world's poorly known taxa such as fungi, insects, nematodes, and bacteria is warranted. After all, these organisms play key roles in mediating a wide range of biotic and abiotic processes (Falkowski et al., 2008; Stajich et al., 2009; Friesen et al., 2011). Understanding the role of soil physicochemical compounds in shaping organism richness and community composition in any location, but in particular in mega-diverse regions such as Amazonia, is therefore crucial.

In this study, we investigate the physio-chemical correlation of soil and litter richness and community composition on a west-to-east transect across Brazilian Amazonia, along the

Comentado [h4]: physico

Amazon River. We analyse richness estimates based on environmental DNA of the ribosomal 16S (prokaryote) and nuclear ribosomal 18S (eukaryote) genes. Specifically, we seek to answer two questions on the correlations of OTU richness and community structure for both organism groups: (1) Does the genetic diversity (measured as OTU richness) of the litter layer correlate with the OTU diversity of the underlying soil; and (2) Are eukaryote (18S) and prokaryote (16S) OTU richness and community composition correlated with physical and chemical soil properties? If so, what are the most important correlates of OTU diversity?

2. Materials and Methods

2.1. Sampling design and localities

We sampled four main localities along a longitudinal transect across the Amazon River (Fig. 1) following the sampling design of Tedersoo et al. (2014). Detailed locality descriptions are available in Ritter et al. (in review). Briefly, we sampled a total of 39 plots in different habitat types, which can be summarized as terra-firme forests, várzeas, igapós, and naturally open areas (e.g. campinas). These environments support distinct biota and are often associated with different kinds of soil. Terra-firme forests are unflooded and generally characterized by poor nutrient latosols (Falesi, 1984). In contrast, várzeas and igapós are seasonally flooded forests that remain submerged during parts of the year and they are differentiated by the type of the flooding water. Várzeas are flooded by white-water rivers and are nutrient rich areas (Junk et al., 2011). Igapós are flooded by black-water rivers and are characterized by large areas of white sands (podzols). Finally, the open areas of Amazonia are related to nutrient-impoverished sandy soils (Prance, 1996; Fine et al., 2005).

Our sampling was placed in areas that cover the different habitat types localized in: *Benjamin Constant* (9 plots), our westernmost locality, approximately 1,100 km west of Manaus in the upper Amazonas River (4.383° S, 70.017° W; Fig 1A); *Jaú national park* (6 plots; 1.850° S, 61.616°W; Fig 1B) and *Novo Airão* (3 plots; 2.620° S, 60.944°W; Fig 1C), on the west side of the Negro River; *Reserva do Cuieras* (6 plots; 2.609º S, 60.217º W; Fig 1D) and *Reserva da Campina* (3 plots; 2.592º S, 60.030º W; Fig 1E), on the east side of the Negro River; and *Caxiaunã* (12 plots), a national forest located 350 km west of Belém (1.7352° S, 51,463°W; Fig 1F), which constitutes our easternmost locality. All samples collection were authorized by Brazilian authorities: ICMBio (registration number 48185-2) and IBAMA (registration number 127341).

Comentado [h5]: was

2.2. PHYSICOCHEMICAL SOIL ANALYSES

We determined the physicochemical and nutrient profiles of the first five centimetres of three soil samples from each plot, totalling 117 samples. The samples were analysed for several chemical and textural variables. pH was measured in water (ratio 1:2.5). The exchangeable concentrations were measured for sodium (Na), potassium (K), and phosphorus (P) using Mehlich-1 extraction (unit mg/dm^3) and for calcium and magnesium (Ca, Mg) using KCl (1 mol/L) extraction (unit cmolc/dm^3). The sum of all exchangeable bases (SB; which comprises K_+, Ca^{2}_+, Mg^{2}_+, and Na_+; unit cmolc/dm^3) was then calculated. We also estimated exchangeable aluminium (Al and H+Al; unit cmolc/dm^3) extracted with calcium acetate (0.5 mol/L at pH 7.0), aluminium saturation index (m; unit %), and Base Saturation Index (V; unit %). The effective cation exchange capacity (t) as well as the cation exchange capacity (T)

were measured at pH 7.0 (unit cmolc/dm^3). The C (organic carbon), and organic matter (M.O) = C (organic carbon) x 1,724 - Walkley-Black was quantified (unit g/kg). Soil texture was characterized by the percentage of fine (0.05 – 0.2 mm), thick (0.2 – 2 mm), and total sand (0.05 – 2 mm) as well as the silt (0.002 – 0.05 mm) and clay (< 0.002 mm) fraction of the soil weight. All analyses were commissioned from EMBRAPA Ocidental (Brazil), following the protocol described in Donagema et al. (2011). Afterwards, we used the mean of the three soil samples from the same plot to obtain a unique value for the measurement of each variable for each plot.

2.3. DNA EXTRACTION, AMPLIFICATION, AND SEQUENCING

The details of laboratory procedures are described in Ritter et al. (in review). We extracted soil and litter using the PowerMax® Soil DNA Isolation Kit (MO BIO Laboratories, USA), following the manufacturer's instructions. The amplification of 16S was performed by Macrogen (Republic of Korea) following standard protocols, and sequencing was performed using the Illumina MiSeq 2x300 platform. For metabarcoding of the 18S gene, sequencing preparation was performed at the laboratory of the University of Gothenburg as described in Ritter et al. (in review) and the amplicons were sequenced at SciLifeLab (Stockholm, Sweden) using an Illumina MiSeq 2x250 machine.

2.4. SEQUENCE ANALYSES

We used the USEARCH/UPARSE v9.0.2132 Illumina paired reads pipeline (Edgar, 2013) to filter out poor-quality sequences, de-replicate and sort reads by abundance, infer operational taxonomic units (OTUs; Blaxter et al. 2005), and remove singletons. We inferred OTUs at the

97% sequence similarity level. We used the SINA v1.2.10 for ARB SVN (revision 21008; Pruesse, Peplies, and Glöckner, 2012) reference dataset for both markers and used SILVAngs 1.3 for taxonomic assignments (Quast et al., 2013).

2.5. Statistical analyses

We rarefied all samples to the lowest number of reads (number of sequences in each OTU, which is analogous to abundance of each "species") obtained from a single plot (22,209 for 16S and 1,359 for 18S; Fig S2) to standardize the sampling effort per plot. An average value was computed to calculate local diversity using the function "rarefy" in the package vegan v. 2.4-3 (Oksanen et al., 2007) in R v3.3.2 (R Development Core Team, 2017). We subsequently transformed the OTU tables to presence/absence for both prokaryote (16S) and eukaryote (18S) data.

For soil compounds, we first normalized all soil variables to zero mean and unit variance using the "scale" function of vegan. Afterwards, we performed two principal component analyses (PCAs). In the first analysis we used the chemical variables phosphorus (P), exchangeable bases (Na, K, Ca, and Mg), the sum of all exchangeable bases (SB), exchangeable aluminium (Al and the cation H+Al), Saturation Index by Aluminium (m), Base Saturation Index (V), effective cation exchange capacity (t), and cation exchange capacity (T). In the second analysis we used silt, clay, and three sand fractions (fine, thick, and total). Since the first PCA axis explained 66% of both chemical and physical variables, we used the first PCA axis of each analysis in the subsequent analyses. Given their importance in regulating the soil biota, we also used soil organic carbon content (Nielson et al., 2011) and pH (Lauber et al., 2009) as independent variables.

Comentado [h6]: It's S1

Comentado [h7]: Lines 73 and 97 in your R script. I would recommend changing the label parameter to label = FALSE

Comentado [h8]: I don't understand the sentence, if you rarefied to the lowest sample why do you do an average?
It is confusing, you should say something like: An average value was computed to calculate the richness of each location after using ...

Comentado [h9]: prokaryotic

Comentado [h10]: eukaryotic

Comentado [h11]: It's not in the bibliography

193

194 To address our first research question (Does genetic diversity (measured as OTU richness) of
195 the litter layer predict the OTU diversity of the underlying soil, for eukaryotes (18S) and
196 prokaryotes (16S)?), we analysed the relationship between litter and soil OTU richness for
197 prokaryote (16S) and eukaryote (18S) data using a linear regression model (the *lm* function in
198 R). We analysed community dissimilarity between the different substrates performing two
199 two-dimensional non-metric multidimensional scaling (NMDS) ordinations of the
200 presence/absence matrices. The first ordination was of litter and soil for prokaryotes (16S),
201 and the second for eukaryotes (18S). Both ordinations used the quantitative version of the
202 Jaccard dissimilarity index as implemented in the metaMDS function in the package vegan.
203 Additionally, we used the "envfit" method implemented in vegan to fit substrate type (litter or
204 soil) onto the NMDS ordination as a measure of the correlation of these factors with the
205 NMDS axes. To visualize the OTU community similarity among plots, we constructed a
206 similarity network with the qgraph v. 1.4.4 (Epskamp et al., 2012) package in R using a
207 similarity index (1/Jaccard dissimilarity).

208

209 To address our second research question (Are eukaryote (18S) and prokaryote (16S) OTU
210 richness and community composition correlated with physical and chemical soil properties? If
211 so, what are the most important soil compounds correlates with biodiversity?), we performed
212 four Bayesian general linear models (GLM) as implemented in the R-INLA v. 17.6.20 R
213 package (Rue et al., 2009) with the OTU richness of litter and soil from the prokaryote (16S)
214 and eukaryote (18S) data as response variables, and soil properties as explanatory variables.
215 We first tested the effect of spatial auto-correlations by comparing analyses of standard
216 GLMs, with GLM analysis using stochastic partial differential equations (SPDE) that

explicitly consider spatial correlation. Three plots were missing sand, silt, and clay information, and the corresponding values were therefore inferred through calculation of the regression weights from the observed data using the mice v. 2.30 R package (Buuren and Groothuis-Oudshoorn, 2011).

To investigate the effects of soil compounds on community composition, we performed four partial Mantel tests using the distance matrices of geographical distance, environmental (soil properties) distance, and community distance (using the quantitative Jaccard dissimilarity). We then performed variation partitioning using the quantitative Jaccard dissimilarity distance-based redundancy analysis on the litter and soil data for the prokaryote (16S) and eukaryote (18S) communities. Variation partitioning gives an indication of the unique and shared contribution of each explanatory variable to the total community variation (Legendre and Legendre, 1998). We used the "varpart" function of the vegan package and assessed the significance for each section of the variation partitioning approach using redundancy analysis.

Additional R packages we used for data curation and visualization were tidyverse v. 1.1.1 (Wickham, 2017), Hmisc v. 4.0-3 (Harrell Jr., 2016), ggfortify v. 0.1.0 (Tang et al., 2016), gridExtra v. 2.2.1 (Auguie, Antonov and Auguie 2016), ggplot2 (Wickham, 2016), and viridis v. 0.4.0 (Garnier, 2016). Scripts for all analyses are provided in the supplementary material.

3. Results

3.1. Correlation between litter and soil OTU richness (research question 1)

Comentado [h12]: Only taking into account richness not Evenness

Comentado [h13]: The reference is from 2008 and it's Harrell Jr. And Dupont.

Comentado [h14]: Not in bibliography

The mean of the rarefied numbers and standard deviations of OTUs for each plot for litter and soil separately are provided in Table S1, and the rarefaction curves are showed in Figure S1. We found a weak positive correlation between litter and soil OTU richness for prokaryotes (adj. R^2 0.25, $p < 0.001$; Fig. 2A). For eukaryotes, the correlation was not significant (Fig. 2B). However, we registered an outlier for the plot "CXNCAMP3" with very low soil OTU richness. Excluding this data point strengthened the correlation for prokaryotes (adj. R^2 0.46, $p < 0.001$; Fig. S2A), but not for eukaryotes (Fig. S2B).

The community composition of litter and soil had a weak separation on the two axes of the NMDS, even by habitat type (Fig. 3). The envfit test indicated weak but significant effects of substrate type on both the prokaryote (R^2 of 0.1, $p < 0.001$) and eukaryote (R^2 of 0.11, $p < 0.001$) community compositions and a strong effect on the habitat type on both the prokaryote (R^2 of 0.40, $p < 0.001$) and eukaryote (R^2 of 0.17, $p < 0.001$) community composition. The similarity network shows a weak separation between litter and soil for both the prokaryote and the eukaryote communities (Fig. S4). There is no clear taxonomic variation among groups in litter or soil, neither in the prokaryote nor in the eukaryote data (Fig. 4).

3.2. Correlations between soil properties and OTU composition (research question 2)

The first PCA axis represents a substantial proportion of the variation of the physical (66%) and chemical (66%) data. In the physical data PCA, large values are associated with coarse texture and small values with fine texture. The increasing values of PCA 1 are associated with silt (-0.45), clay (-0.38), fine sand fraction (0.28), coarse sand fraction (0.52), and total sand

Comentado [h15]: This value doesn't fit with the one showed in the Figure 2.

Comentado [h16]: It is not strong

Comentado [h17]: It could be stronger if you take into account the relative abundance of each OTU. Computing the analysis with

Comentado [h18]: It should be called S3

Comentado [h19]: prokaryotic

Comentado [h20]: eukaryotic

fraction (0.55; Table S2). The first axis of PC1 is separated by habitat type with some overlap of campinas (Fig. 5A). For the chemical data, large values are associated with poor soils. The largest negative association was -0.35, namely the sum of all exchangeable bases (SB). The largest positive association was 0.29, viz. the aluminium saturation index (Table S3). The habitat types are not well separated along the first axis of the chemical PCA, with highly scattered igapós and várzeas values. Campinas and terra firme, on the other hand, are associated with positive values (poor soils) (Fig. 5B). For all datasets, the best GLM models – considering the deviance information criterion (DIC) and Watanabe-Akaike information criterion (WAIC) – were those that included spatial correlation (Table S4). For prokaryotes, we identified organic carbon as an important predictor of OTU richness for both soil and litter. In addition, chemical PC1 was a significant predictor for soil OTU richness and physical PC1 for litter OTU richness. For eukaryote OTU richness, the most important predictors were pH and organic carbon, for both litter and soil (Table 1). Overall, soil organic carbon had the strongest effect on OTU richness for prokaryotes and eukaryotes in litter and soil, showing a negative correlation.

The Mantel test showed a significant association between environmental distance (soil properties) and community similarity (Jaccard dissimilarity index) for all datasets (prokaryote – soil [R = 0.52, p = 0.001], litter [R = 0.57, p = 0.001]; eukaryote – soil[R = 0.38, p = 0.001], litter [R = 0.54, p = 0.001]). Accounting for geographic distances using partial Mantel tests caused only a small decrease in the correlation coefficients, and all the correlations remained highly significant (prokaryote – soil [R = 0.48, p = 0.001], litter [R = 0.53, p = 0.001]; eukaryote – soil [R = 0.35, p = 0.001], litter [R = 0.51, p = 0.001]).

Comentado [h21]: Not True. There is a mix of samples in the negative part of PC1 with some outliers from campinas and terra-firme.

Comentado [h22]: The same is shown in the Physical plot with respect to campinas.

Comentado [h23]: Change in Table text OUT for OTU

Comentado [h24]: Add a space

The variation partitioning revealed that a moderate total percentage of the community variation was explained by soil physic-chemical data in the soil prokaryote model (33%) and in the litter prokaryote model (37%; Fig. 6). Soil physic-chemical data had a higher explanatory power for prokaryotes than for eukaryotes, with a total of 12% of community variation explained for soil and 18% for the litter communities for eukaryotes. Inside each dataset (prokaryote and eukaryote), the litter communities are more structured by soil characteristics than are the soil communities. All variables explained small but significant proportions of the variance in all communities, and showed some weak but significant interactions (Fig. 6).

Comentado [h25]: physico

4. Discussion

4.1. Contrasting litter and soil diversity

A regression between soil and litter richness of prokaryotes and a non-significant regression between soil and litter OTU richness of eukaryotes was observed. Contrary to our expectation of litter being more dominated by plants and nematodes and soil by microorganisms, we could not observe a difference in proportion of taxonomic groups between the soil and litter (Fig. 4). This means that on the Amazon basin scale, the taxonomic composition at higher levels (phylum and order) is consistent between litter and soil. Interestingly, the dominance order of the phyla in our samples was only partly congruent with the one recently reported in a large global dataset (Delgado-Baquerizo et al., 2018). While we, too, found Proteobacteria to be the most dominant phylum, the second most abundant phylum was Chloroflexi in our samples

Comentado [h26]: It's difficult to say when not considering relative abundance

Comentado [h27]: It's absolutely wrong!
Delgado-Baquerizo et al. took into account the relative abundance of each OTU before evaluating the dominance of the different phyla (Figure 1c of that study).

while this phylum was only the 5th most abundant in the global dataset. Moreover, the rank-abundance distribution of the most dominant phyla was remarkably more even in our tropical sample than in the global sample, with Proteobacteria accounting for just over 20% of all reads (versus almost 40% in the global dataset) and 8 phyla accounting for > 5% of relative frequency each versus only 4 in Delgado-Baquerizo et al., (2018). For eukaryotes, too, we found contrasting results. For instance, unlike Porazinska et al. (2012) who found a dominance of nematodes in the litter of tropical forest, we found very similar proportions of nematode OTUs in soil and litter, with the highest diversity in the soil (Fig. 4). These differences could be related to sampling or primer biases but also may be related to differences in the forest structure between studies. Taken together these differences highlight the need of more studies across Amazon basin.

Comentado [h28]: These data are not shown neither in results nor in supplementary material.

The OTU community composition was weakly correlated among samples at the OTU level, but we could not observe it on the phylum or order levels (Fig. 4). This is expected since for instance fungi can be anticipated to be a dominant group in any soil environment, including tropical forests (Tedersoo et al. 2017), but the dominant fungal taxa (OTU) may vary considerably even on local and sub-local scales in these forests (Urbanová et al. 2015). Urbanová et al. (2015) found similar results for bacteria and fungi; the phylum level indicated the same magnitude of diversity in soils and litter, but there were striking differences on the OTU level. The habitat types explain a higher proportion of community variation than does the substrate type, which is expected since both substrates would share a large number of organisms.

Comentado [h29]: Not shown anywhere. All the sampling locations are mixed in the results presented in Figure 4.

4.2. Soil predictors of OTU composition

Soil proprieties are thought to be useful predictors of diversity and composition of microorganism communities, as inferred from several vegetation types globally (e.g. Lauber et al., 2009; Fierer et al., 2012; Navarrete et al., 2013; Barnes et al., 2016). In this study, we found that in Amazonia the soil properties we quantified have variable effects on OTU richness and community composition for litter and soil, and that they furthermore vary between prokaryotes and eukaryotes.

Considering the results from the linear models, pH was the second strongest factor in explaining both litter and soil eukaryote richness, although it had only a weak effect on prokaryote richness. Lauber et al. (2009) found pH to be the main factor in explaining bacterial phylogenetic diversity and phylogenetic composition, where soils with pH between 4.5 and 8 had the lowest micro-organismal diversity, even in tropical forests with high macro-organism diversity. In our samples, the variation in pH was moderate, from 3.65 to 5.14. The variation observed for the effect of pH (and other soil variables) in different sets of organisms (e.g. Acidobacteria and Actinobacteria) could help us understand the low percentage of explained variation for prokaryotes.

Biotic and abiotic conditions jointly determine soil proprieties, which in turn interact with biotic and abiotic factors in a feedback loop. It is therefore important to consider environmental and biological interactions between variables. Indeed, our variance analysis reveals several interactions between soil compounds, although these interactions are weak.

Comentado [h30]: properties

Comentado [h31]: eukaryotic

Comentado [h32]: prokaryotic

Comentado [h33]: Where are the results to support this assumption?

This analysis is important for providing a better understanding of the study system, but it is limited to the variables we are able to sample. The first axis of physical PCA was well separated by the habitat types, however the first axis of chemical PCA was less well separated for flooded areas (igapós and várzeas). The campinas, which are associated with soil of the white sand type, are grouped in the extreme of positive values followed by terra-firme. In our results the large values of first axis of chemical PCA are associated with poor soils (Table S3). These results agree with previous studies, which report both habitats as being related with poor soils (Falesi, 1984; Prance, 1996; Fine et al., 2005).

We found a negative correlation of soil organic carbon with OTU richness for all groups. This is a puzzling result, since soil organic carbon is often used as an indicator of soil biomass (Fierer et al., 2009), and soil biodiversity has previously been found to be correlated with carbon sequestration (Wagg et al., 2013). However, the relationship between soil biodiversity and carbon has varied across studies (Nielson et al., 2011). Furthermore, Fierer et al. (2012) and de Lima Brossi et al. (2014) found that soil organic matter was related to microbial community composition in several different vegetation types. The negative correlation between soil organic carbon content and OTU richness reported here might be related to high decomposition rates, keeping the carbon stock locked in aboveground biomass and low in the soil. Our results support the findings of Wall et al. (2008), who found a positive influence of richness of soil biota on decomposition rates in wet tropical environments. Along the same line, Wagg et al. (2013) found that soil diversity and soil community composition are related through nutrient cycling. Decreases in soil diversity and changes in soil communities alter the ability of soil organisms to break down organic matter and recycle nutrients, rendering the

Comentado [h34]: As I told you above, the separation is scattered in both plots for some samples.

Comentado [h35]: Both too scattered to assume this.

Comentado [h36]: largest

Comentado [h37]: Not in the bibliography

Comentado [h38]: The reference is incomplete

return of nutrients to the above-ground community difficult (Wardle et al., 2004). These findings stress the complex nature of carbon-diversity dynamics and the plant–soil feedback mediated by soil biota (Mangan et al. 2010).

5. Conclusions

In this study we found significant correlations between physio-chemical soil properties and genetic diversity in Amazonia. Across the study area, we found that OTU richness and community composition are in part explained by different soil compounds. These compounds interact in a complex way, which stresses the importance of considering multiple factors and their interactions in the characterization of biodiversity patterns. Our most striking result was the negative correlation between organic carbon and OTU richness, and the effect of organic carbon on community composition. Soils are crucial for carbon cycling in terrestrial ecosystems, and our results suggest that a better understanding of the relationship between diversity (above and belowground) and carbon cycles might be essential for modelling carbon deposition and diversity patterns in the world's largest and most biodiverse rainforest.

Comentado [h39]: physico

Acknowledgements

We thank the Brazilian authorities ICMBio (registration number 48185-2) and IBAMA (registration number 127341), for the permits granted for this research; Anna Ansebo, Sven Toresson, and Ylva Heed for laboratory and administrative assistance; Mats Töpel for help with bioinformatics; Hans ter Steege for advice on sampling localities and experimental design; and members of our research group for discussions and suggestions. The authors also

397 acknowledge support from Science for Life Laboratory, the Knut and Alice Wallenberg
398 Foundation, the National Genomics Infrastructure funded by the Swedish Research Council,
399 and Uppsala Multidisciplinary Center for Advanced Computational Science for assistance
400 with massively parallel sequencing and access to the UPPMAX computational infrastructure.
401 Additional computational analyses were run at the University of Gothenburg bioinformatics
402 cluster at the Department of Biological and Environmental Sciences
403 (http://albiorix.bioenv.gu.se/).

404

405 **Funding** – This study had primary financial support from CNPq (Conselho Nacional de
406 Desenvolvimento Científico e Tecnológico - Brazil: 249064/2013-8) and the Swedish
407 Research Council (B0569601). AA is further supported by the European Research Council
408 under the European Union's Seventh Framework Programme (FP/2007-2013, ERC Grant
409 Agreement n. 331024), the Swedish Foundation for Strategic Research, a Wallenberg
410 Academy Fellowship, the Faculty of Sciences at the University of Gothenburg, and the David
411 Rockefeller Center for Latin American Studies at Harvard University.

412

413 **References**

414 Auguie, B., Antonov, A., and Auguie, M. B. 2016. Package 'gridExtra'.

415 Baldeck, C.A., Tupayachi, R., Sinca, F., Jaramillo, N. and Asner, G.P., 2016. Environmental
416 drivers of tree community turnover in western Amazonian forests. *Ecography*, *39*(11),
417 pp.1089-1099. DOI: 10.1111/ecog.01575.

418 Bardgett, R.D. and Van Der Putten, W.H., 2014. Belowground biodiversity and ecosystem
419 functioning. *Nature*, *515*(7528), pp.505-511. DOI: 10.1038/nature13855.

420 Barnes, C.J., Maldonado, C., Frøslev, T.G., Antonelli, A. and Rønsted, N., 2016.
421 Unexpectedly high beta-diversity of root-associated fungal communities in the Bolivian
422 Andes. *Frontiers in Microbiology*, *7*. DOI: 10.3389/fmicb.2016.01377.

423 Basset, Y., Cizek, L., Cuénoud, P., Didham, R.K., Guilhaumon, F., Missa, O., Novotny, V.,
424 Ødegaard, F., Roslin, T., Schmidl, J. and Tishechkin, A.K., 2012. Arthropod diversity in a
425 tropical forest. *Science*, *338*(6113), pp.1481-1484. DOI: 10.1126/science.1226727.
426 Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Floyd, R. and Abebe, E., 2005.
427 Defining operational taxonomic units using DNA barcode data. *Philosophical Transactions of*
428 *the Royal Society B: Biological Sciences*, *360*(1462), pp.1935-1943. DOI:
429 10.1098/rstb.2005.1725

430

431 Buckley, D.H. and Schmidt, T.M., 2001. Environmental factors influencing the distribution of
432 rRNA from Verrucomicrobia in soil. *FEMS Microbiology Ecology*, *35*(1), pp.105-112.
433 DOI: 10.1111/j.1574-6941.2001.tb00793.x.

434 Buuren, S. and Groothuis-Oudshoorn, K., 2011. mice: Multivariate imputation by chained
435 equations in R. *Journal of Statistical Software*, *45*(3).

436 Cámara-Leret, R., Tuomisto, H., Ruokolainen, K., Balslev, H. and Munch Kristiansen, S.,
437 2017. Modelling responses of western Amazonian palms to soil nutrients. *Journal of*
438 *Ecology*, *105*(2), pp.367-381.

De Lima Brossi, M.J., Mendes, L.W., Germano, M.G., Lima, A.B. and Tsai, S.M., 2014. Assessment of bacterial bph gene in Amazonian Dark Earth and their adjacent soils. *PLOS ONE*, *9*(6), p.e99597. DOI: 10.1371/journal.pone.0099597.

Delgado-Baquerizo, M., Oliverio, A.M., Brewer, T.E., Benavent-González, A., Eldridge, D.J., Bardgett, R.D., Maestre, F.T., Singh, B.K. and Fierer, N., 2018. A global atlas of the dominant bacteria found in soil. *Science*, *359*(6373), pp.320-325. DOI: 10.1126/science.aap9516.

Donagema, G.K., De Campos, D.B., Calderano, S.B., Teixeira, W.G. and Viana, J.M., 2011. Manual de métodos de análise de solo. *Embrapa Solos-Documentos (INFOTECA-E)*.

Edgar, R.C., 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, *10*(10), pp.996-998. DOI: 10.1038/nmeth.2604.

Epskamp, S., Cramer, A.O., Waldorp, L.J., Schmittmann, V.D. and Borsboom, D., 2012. qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, *48*(4), pp.1-18. DOI: 10.1016/j.ecolecon.2010.05.002.

Falesi, I. C. 1984. Estado atual de conhecimento dos solos da Amazônia Brasileira. – *In*: Simpósio do Trópico Umido, Belém, 1984. *Embrapa- CPATU, Documentos*, 36: pp. 168-191.

Falkowski, P.G., Fenchel, T. and Delong, E.F., 2008. The microbial engines that drive Earth's biogeochemical cycles. *Science*, *320*(5879), pp.1034-1039. DOI: 10.1126/science.1153213.

Faoro, H., Alves, A.C., Souza, E.M., Rigo, L.U., Cruz, L.M., Al-Janabi, S.M., Monteiro, R.A., Baura, V.A. and Pedrosa, F.O., 2010. Influence of soil compounds on the diversity of bacteria in the Southern Brazilian Atlantic Forest. *Applied and Environmental*

Microbiology, *76*(14), pp.4744-4749. DOI: 10.1128/AEM.03025-09.

Fierer, N., Strickland, M.S., Liptzin, D., Bradford, M.A. and Cleveland, C.C., 2009. Global patterns in belowground communities. *Ecology Letters*, *12*(11), pp.1238-1249. DOI: 10.1111/j.1461-0248.2009.01360.x.

Fierer, N., Leff, J.W., Adams, B.J., Nielsen, U.N., Bates, S.T., Lauber, C.L., Owens, S., Gilbert, J.A., Wall, D.H. and Caporaso, J.G., 2012. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proceedings of the National Academy of Sciences*, *109*(52), pp.21390-21395. DOI: 10.1073/pnas.1215210110.

Fine, P.V., Daly, D.C., Muñoz, G.V., Mesones, I. and Cameron, K.M., 2005. The contribution of edaphic heterogeneity to the evolution and diversity of Burseraceae trees in the western Amazon. *Evolution*, *59*(7), pp.1464-1478. DOI: https://doi.org/10.1554/04-745

Friesen, M.L., Porter, S.S., Stark, S.C., Von Wettberg, E.J., Sachs, J.L. and Martinez-Romero, E., 2011. Microbially mediated plant functional traits. *Annual Review of Ecology, Evolution, and Systematics*, *42*, pp.23-46. DOI: 10.1146/annurev-ecolsys-102710-145039.

Garnier, S., 2016. viridis: Default color maps from 'matplotlib'[Software].

Harrell Jr, F.E. and Dupont, C., 2008. Hmisc: harrell miscellaneous. *R package version*, *3*(2).

Higgins, M.A., Ruokolainen, K., Tuomisto, H., Llerena, N., Cardenas, G., Phillips, O.L., Vásquez, R. and Räsänen, M., 2011. Geological control of floristic composition in Amazonian forests. *Journal of Biogeography*, *38*(11), pp.2136-2149. DOI: 10.1111/j.1365-2699.2011.02585.x.

Hinsinger, P., Bengough, A.G., Vetterlein, D. and Young, I.M., 2009. Rhizosphere:

biophysics, biogeochemistry and ecological relevance. *Plant and Soil*, *321*(1-2), pp.117-152.

Junk, W.J., Piedade, M.T.F., Schöngart, J., Cohn-Haft, M., Adeney, J.M. and Wittmann, F., 2011. A classification of major naturally-occurring Amazonian lowland wetlands. *Wetlands*, *31*(4), pp.623-640. DOI: 10.1007/s13157-011-0190-7.

Kuramae, E.E., Yergeau, E., Wong, L.C., Pijl, A.S., Van Veen, J.A. and Kowalchuk, G.A., 2012. Soil compounds more strongly influence soil bacterial communities than land-use type. *FEMS Microbiology Ecology*, *79*(1), pp.12-24. DOI: 10.1111/j.1574-6941.2011.01192.x.

Lauber, C.L., Hamady, M., Knight, R. and Fierer, N., 2009. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Applied and Environmental Microbiology*, *75*(15), pp.5111-5120. DOI: 10.1128/AEM.00335-09.

Legendre, P. and Legendre, L., 1998. Numerical ecology: second English edition. *Developments in environmental modelling*, *20*.

Mangan, S.A., Schnitzer, S.A., Herre, E.A., Mack, K.M., Valencia, M.C., Sanchez, E.I. and Bever, J.D., 2010. Negative plant-soil feedback predicts tree-species relative abundance in a tropical forest. *Nature*, *466*(7307), pp.752-755. DOI: 10.1038/nature09273.

Navarrete, A.A., Kuramae, E.E., De Hollander, M., Pijl, A.S., Van Veen, J.A. and Tsai, S.M., 2013. Acidobacterial community responses to agricultural management of soybean in Amazon forest soils. *FEMS Microbiology Ecology*, *83*(3), pp.607-621. DOI: 10.1111/1574-6941.12018.

Oksanen, J., Kindt, R., Legendre, P., O'hara, B., Stevens, M.H.H. 2007. The vegan

package. *Community ecology package*, *10*, pp.631-637.

Osborne, C.A., Zwart, A.B., Broadhurst, L.M., Young, A.G. and Richardson, A.E., 2011. The influence of sampling strategies and spatial variation on the detected soil bacterial communities under three different land-use types. *FEMS Microbiology Ecology*, *78*(1), pp.70-79. DOI: 10.1111/j.1574-6941.2011.01105.x.

Porazinska, D.L., Giblin-Davis, R.M., Powers, T.O. and Thomas, W.K., 2012. Nematode spatial and ecological patterns from tropical and temperate rainforests. *PLOS ONE*, *7*(9), p.e44641. DOI: 10.1371/journal.pone.0044641.

Prance GT., 1996. Islands in Amazonia. *Philosophical Transactions of the Royal Society of London. B*, *29*(1341), pp. 823-33. DOI: 10.1098/rstb.1996.0077

Pruesse, E., Peplies, J. and Glöckner, F.O., 2012. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics*, *28*(14), pp.1823-1829. DOI: 10.1093/bioinformatics/bts252.

Prober, S.M., Leff, J.W., Bates, S.T., Borer, E.T., Firn, J., Harpole, W.S., Lind, E.M., Seabloom, E.W., Adler, P.B., Bakker, J.D. and Cleland, E.E., 2015. Plant diversity predicts beta but not alpha diversity of soil microbes across grasslands worldwide. *Ecology Letters*, *18*(1), pp.85-95. DOI: 10.1111/ele.12381.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. and Glöckner, F.O., 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, *41*(D1), pp.D590-D596. DOI: 10.1093/nar/gks1219.

R Development Core Team. 2017. R: a language environment for statistical computing. *Vienna: R Foundation for Statistical Computing. (http://www.r- project.org/).* Accessed 15/02/2017.

Ramirez, K.S., Leff, J.W., Barberán, A., Bates, S.T., Betley, J., Crowther, T.W., Kelly, E.F., Oldfield, E.E., Shaw, E.A., Steenbock, C. and Bradford, M.A., 2014. Biogeographic patterns in below-ground diversity in New York City's Central Park are similar to those observed globally. *Proceedings of the Royal Society B, 281*(1795). DOI: 10.1098/rspb.2014.1988.

Rue, H., Martino, S., Lindgren, F., Simpson, D., Riebler, A. and Krainski, E.T., 2009. INLA: functions which allow to perform a full Bayesian analysis of structured additive models using Integrated Nested Laplace Approximation. *R package version 0.0*.

Stajich, J.E., Berbee, M.L., Blackwell, M., Hibbett, D.S., James, T.Y., Spatafora, J.W. and Taylor, J.W., 2009. Primer--The Fungi. *Current biology: CB*, *19*(18), p.R840. DOI: 10.1016/j.cub.2009.07.004

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. and Willerslev, E., 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, *21*(8), pp.2045-2050. DOI: 10.1111/j.1365-294X.2012.05470.x.

Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N.S., Wijesundera, R., Ruiz, L.V., Vasco-Palacios, A.M., Thu, P.Q., Suija, A. and Smith, M.E., 2014. Global diversity and geography of soil fungi. *Science*, *346*(6213), p.1256688. DOI: 10.1126/science.1256688.

Tedersoo, L., Bahram, M., Cajthaml, T., Põlme, S., Hiiesalu, I., Anslan, S., Harend, H., Buegger, F., Pritsch, K., Koricheva, J. and Abarenkov, K., 2016. Tree diversity and species identity effects on soil fungi, protists and animals are context dependent. *The ISME*

Journal, *10*(2), pp.346-362. DOI: 10.1038/ismej.2015.116.

Tedersoo, L., Bahram, M., Puusepp, R., Nilsson, R.H. and James, T.Y., 2017. Novel soil-inhabiting clades fill gaps in the fungal tree of life. *Microbiome*, *5*(1), p.42. DOI: 10.1186/s40168-017-0259-5.

Tuomisto, H., Zuquim, G. and Cárdenas, G., 2014. Species richness and diversity along edaphic and climatic gradients in Amazonia. Ecography, 37(11), pp.1034-1046.

Tuomisto, H., Moulatlet, G.M., Balslev, H., Emilio, T., Figueiredo, F.O., Pedersen, D. and Ruokolainen, K., 2016. A compositional turnover zone of biogeographical magnitude within lowland Amazonia. Journal of Biogeography, 43(12), pp.2400-2411.

Urbanová, M., Šnajdr, J. and Baldrian, P., 2015. Composition of fungal and bacterial communities in forest litter and soil is largely determined by dominant trees. *Soil Biology and Biochemistry*, *84*, pp.53-64. DOI: 10.1016/j.soilbio.2015.02.011.

Vogel, T.M., Simonet, P., Jansson, J.K., Hirsch, P.R., Tiedje, J.M., Van Elsas, J.D., Bailey, M.J., Nalin, R. and Philippot, L., 2009. TerraGenome: a consortium for the sequencing of a soil metagenome. *Nature Reviews Microbiology*, 7, 252. DOI: 10.1038/nrmicro2119

Wagg, C., Bender, S.F., Widmer, F. and van der Heijden, M.G., 2014. Soil biodiversity and soil community composition determine ecosystem multifunctionality. *Proceedings of the National Academy of Sciences*, *111*(14), pp.5266-5270. DOI: 10.1073/pnas.1320054111.

Wall, D. H., Bradford, M. A., S T, J., Trofymow, J. A., Behan-Pelletier, V., Bignell, D. E.,

Wardle, D.A., Bardgett, R.D., Klironomos, J.N., Setälä, H., Van Der Putten, W.H. and Wall, D.H., 2004. Ecological linkages between aboveground and belowground

biota. *Science*, *304*(5677), pp.1629-1633. DOI: 10.1126/science.1094875.

Wickham, H., 2016. *ggplot2: elegant graphics for data analysis*. Springer.

Wickham, H., 2017. tidyverse. *R package version, 1.1.1*.

Data Availability and Accessibility:

- DNA sequences: GenBank accessions XXXXXXX-XXXXXX; NCBI SRA: XXXXXXXXX
- Final DNA sequence assembly uploaded as online supplementary items
- Sampling locations, soil physical-chemical data, OTU tables, and R-scripts: Dryad doi:XXXXX

Comentado [h40]: You have to submit your fastq files to NCBI SRA and provide the BioProject number or the BioSample numers.

Authors' contributions

CDR, AA, and RHN conceived this study; CDR collected the data; CDR, AZ, and RHN performed the analyses; CDR wrote the manuscript with contributions from all authors. All authors read and approved the final version of the manuscript.