A peer-reviewed version of this preprint was published in PeerJ on 25 September 2018.

Ritter CD, Zizka A, Roger F, Tuomisto H, Barnes C, Nilsson RH, Antonelli A. 2018. High-throughput metabarcoding reveals the effect of physicochemical soil properties on soil and litter biodiversity and community turnover across Amazonia. PeerJ 6:e5661 https://doi.org/10.7717/peerj.5661

High-throughput metabarcoding reveals the effect of physicochemical soil properties on soil and litter biodiversity and community turnover across Amazonia

Camila D Ritter [Corresp., 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, Göteborg, Sweden
[2] Department of Biological and Environmental Sciences, University of Gothenburg, Gothenburg, Sweden
[3] Centre for Environmental and Climate Research, Lund University, Lund, Sweden
[4] Department of Biology, University of Turku, Turku, Finland
[5] Natural History Museum of Denmark, University of Copenhagen, Denmark, Copenhagen, Denmark
[6] Gothenburg Botanical Garden, Göteborg, Sweden, Gothenburg, Sweden
[7] Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, United States

Corresponding Author: Camila D Ritter
Email address: camila.ritter@gu.se

Background. Knowledge on the globally outstanding Amazonian biodiversity and its environmental determinants stems almost exclusively from aboveground organisms, notably plants. In contrast, the environmental factors and habitat preferences that drive diversity patterns for micro-organisms in the ground remain elusive, despite the fact that micro-organisms constitute the overwhelming majority of life forms in any given location, in terms of both diversity and abundance. Here we address how the diversity and community turnover of operational taxonomic units (OTU) of micro-organisms in soil and litter respond to soil physicochemical properties; whether OTU diversities and community composition in soil and litter are correlated with each other; and whether they respond in a similar way to soil properties. **Methods.** We used recently inferred OTUs from high-throughput metabarcoding of the 16S (prokaryotes) and 18S (eukaryotes) genes to estimate OTU diversity (OTU richness and effective number of OTUs) and community composition for prokaryotes and eukaryotes in soil and litter across four localities in Brazilian Amazonia. All analyses were run separately for prokaryote and eukaryote OTUs, and for each group using both presence-absence and abundance data. Combining these with novel data on soil chemical and physical properties, we identify abiotic correlates of soil and litter micro-organism diversity and community structure using regression, ordination, and variance partitioning analysis. **Results.** Soil organic carbon content was the strongest factor explaining OTU diversity (negative correlation) and pH was the strongest factor explaining turnover for prokaryotes and eukaryotes in both soil and litter. We found significant effects also for other soil variables, including both chemical and

physical properties. The correlation between OTU diversity in litter and in soil was non-significant for eukaryotes and weak for prokaryotes, suggesting that diversity of in one substrate should not be used as a proxy for diversity in the other. The community compositions of both prokaryotes and eukaryotes were more separated for habitat type than for substrate (soil and litter). **Discussion.** In spite of the limited sampling (four localities, 39 plots), our results provide a broad-scale view 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 properties, OTU diversity patterns, and community composition and turnover. The lack of strong correlation between OTU diversity in litter and in soil suggests independence of diversity drives of these substrates and highlights the importance of including both measures in biodiversity assessments. Massive sequencing of soil and litter samples holds the potential to complement traditional biological inventories in advancing our understanding of the factors affecting tropical diversity.

Author Cover Page

THIS IS A REVISION OF MANUSCRIPT: 26029

Article submission to PeerJ

Manuscript category: Research Articles

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

Title:

High-throughput metabarcoding reveals the effect of physicochemical soil properties on soil and litter biodiversity and community turnover across Amazonia.

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, Göteborg, Sweden.

[2] Department of Biological and Environmental Sciences, University of Gothenburg, Sweden.

[3] Centre for Environmental and Climate Research, Lund University, Sweden.

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

[5] Natural History Museum of Denmark, University of Copenhagen, Denmark.

[6] Gothenburg Botanical Garden, Göteborg, Sweden.

[7] Department of Organismic and Evolutionary Biology, Harvard University.

*Corresponding author: Camila Ritter, kmicaduarre@gmail.com. Phone: +46 723924447, Carl Skottsbergs gata 22B, Box 461, SE 405 30 Göteborg, Sweden

§Both are senior authors in this paper.

Abstract

Background. Knowledge on the globally outstanding Amazonian biodiversity and its environmental determinants stems almost exclusively from aboveground organisms, notably plants. In contrast, the environmental factors and habitat preferences that drive diversity patterns for micro-organisms in the ground remain elusive, despite the fact that micro-organisms constitute the overwhelming majority of life forms in any given location, in terms of both diversity and abundance. Here we address how the diversity and community turnover of operational taxonomic units (OTU) of micro-organisms in soil and litter respond to soil physicochemical properties; whether OTU diversities and community composition in soil and litter are correlated with each other; and whether they respond in a similar way to soil properties.

Methods. We used recently inferred OTUs from high-throughput metabarcoding of the 16S (prokaryotes) and 18S (eukaryotes) genes to estimate OTU diversity (OTU richness and effective number of OTUs) and community composition for prokaryotes and eukaryotes in soil and litter across four localities in Brazilian Amazonia. All analyses were run separately for prokaryote and eukaryote OTUs, and for each group using both presence-absence and abundance data. Combining these with novel data on soil chemical and physical properties, we identify abiotic correlates of soil and litter micro-organism diversity and community structure using regression, ordination, and variance partitioning analysis.

Results. Soil organic carbon content was the strongest factor explaining OTU diversity (negative correlation) and pH was the strongest factor explaining turnover for prokaryotes and eukaryotes in both soil and litter. We found significant effects also for other soil variables, including both chemical and physical properties. The correlation between OTU diversity in litter and in soil was

non-significant for eukaryotes and weak for prokaryotes, suggesting that diversity of in one substrate should not be used as a proxy for diversity in the other. The community compositions of both prokaryotes and eukaryotes were more separated for habitat type than for substrate (soil and litter).

Discussion. In spite of the limited sampling (four localities, 39 plots), our results provide a broad-scale view 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 properties, OTU diversity patterns, and community composition and turnover. The lack of strong correlation between OTU diversity in litter and in soil suggests independence of diversity drives of these substrates and highlights the importance of including both measures in biodiversity assessments. Massive sequencing of soil and litter samples holds the potential to complement traditional biological inventories in advancing our understanding of the factors affecting tropical diversity.

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 properties are often considered crucial in regulating the biotic dynamics, vegetation, and diversity patterns at local to regional scales (Vogel et al., 2009; Laurence et al., 2010; Higgins et al., 2011; Friesen et al., 2011; Tuomisto et al., 2016).

For instance, diversity and community composition of plants are influenced by geology and physicochemical soil properties (e.g. Vogel et al., 2009; Laurence et al., 2010; Higgins et al., 2011; Friesen et al., 2011; Tuomisto et al., 2014, 2016, Tedersoo et al., 2016). In particular, the availability of soil nutrients and soil cation concentration are important factors determining plant species composition and turnover (Tuomisto et al., 2003; Laurence et al., 2010; Higgins et al., 2011; Baldeck et al., 2016; Tuomisto et al., 2016; Cámara-Leret et al., 2017). Additionally, soil properties, in particular phosphorus, can 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; Bates et al., 2013; Barnes et al., 2016).

Different soil layers may show different patterns of biodiversity (Hinsinger et al., 2009). For instance, the taxonomic composition of nematode species clustered in six trophic guilds (bacterial feeders, fungal feeders, root associates, plant parasites, omnivores, and predators) has

been found to vary between the mineral soil layer and the organic matter layer (litter) above it (Porazinska et al., 2012). In Amazonia, litter layers vary with habitat type and the length of the inundation period. Unflooded forests (terra-firme) are particularly rich in litter. In addition, flooded forests are also rich in litter and the litter layer increases with increasing length of the inundation period (Myster, 2017). Besides inundation, several other factors influence litter accumulation and thereby decomposition rates and nutrient cycling. These include leaf abscission and species composition (e.g. Gregorich et al., 2016), which have implications for the diversity and community structure of soil and litter-inhabiting organisms. However, the strength of soil-litter interactions varies, and a study conducted in Canada reported no influence of soil physicochemical properties on litter decomposition (Gregorich et al., 2016). This may indicate a difference in biodiversity patterns (due to different drivers and biomass) between soil and litter layers.

The diversity and composition of Amazonian soil and litter communities remain poorly understood, despite recent studies on soil micro-organismic communities (e. g. Basset et al., 2012; Mahé et al., 2017; Ritter et al., 2018). This lack of knowledge, especially in taxonomic groups such as fungi, protists, nematodes, and bacteria, is problematic given the important roles of these groups in a wide range of biotic processes (Falkowski et al., 2008; Stajich et al., 2009; Friesen et al., 2011). To tackle his lack of knowledge, high-throughput amplicon-based sequencing analyses such as DNA metabarcoding (Taberlet et al., 2012) now allow examination of soil diversity patterns (Bardgett and van der Putten, 2014). However, most studies so far have been focused on one or a few taxonomic groups, which renders general conclusions on the effects of soil properties on biodiversity difficult (e.g. Faoro et al., 2010; Laurence et al., 2010;

Navarrete et al., 2013; Barnes et al., 2016). Understanding microbial diversity and communities and their relation to soil physicochemical properties on a broad taxonomic scale is therefore crucial in any location, but particularly so in mega-diverse regions such as Amazonia to access general conclusion about the abiotic drivers of diversity.

In this study, we test the effect of physicochemical soil properties on soil and litter biodiversity and community turnover at four localities along a west-to-east transect across Brazilian Amazonia. We base diversity estimates on operational taxonomic units (OTUs) from environmental DNA of the ribosomal 16S (prokaryote) and nuclear ribosomal 18S (eukaryote) genes. Specifically, we seek to answer the following questions: Are OTU diversity and turnover related to physical and chemical soil properties? If so, what are the most important soil properties? Are OTU diversity and community composition correlated when quantified for the litter layer vs. the underlying soil? All questions are addressed separately for eukaryotes (18S) and prokaryotes (16S) using both presence-absence and abundance data.

2. Materials and Methods

2.1. Sampling design and localities

We sampled four localities along the Amazon River (Fig. 1) following the sampling design of Tedersoo et al. (2014). Detailed locality descriptions are available in Ritter et al. (2018). Briefly, we sampled all depths of the litter layer above the mineral soil (all organic matter, including leaves, roots, and animal debris) and the top 5 cm of the mineral soil in a total of 39 circular

127 plots, each with a radius of 28 m. We chose 20 random trees inside each plot and collected litter
128 and soil on both sides of each tree. We then pooled the samples by substrate to obtain one soil
129 and one litter sample per plot. We sampled in different habitat types, which can be summarized
130 as terra-firme, várzeas, igapós, and campinas. These are four of the commonly recognised main
131 Amazonian environments. Terra-firme is characterised by not being inundated during the annual
132 flood season, and terra-firme forests generally have tall stature and complex canopy structure
133 (IBGE, 2004). In contrast, várzeas and igapós are seasonally flooded and remain submerged
134 during parts of the year. Várzeas are flooded by white-water rivers, which carry a large load of
135 suspended sediments, whereas igapós are flooded by black-water rivers, which bring a small load
136 of suspended sediments but a high concentration of organic acids (Junk et al., 2011). Finally,
137 campinas have nutrient-impoverished sandy soils and forests with a reduced stature and
138 relatively simple canopy structure (Prance, 1996; IBGE, 2004).

139

140 Our sampling was carried out in four areas: *Benjamin Constant* (9 plots covering terra-firme,
141 várzea and igapó), our westernmost locality, approximately 1,100 km west of Manaus in the
142 upper Amazonas River (4.383° S, 70.017° W; Fig 1A); *Jaú national park* (6 plots covering terra-
143 firme and igapó; 1.850° S, 61.616°W; Fig 1B) and *Novo Airão* (3 plots covering campinas;
144 2.620° S, 60.944°W; Fig 1C), on the west side of the Negro River; *Reserva do Cuieras* (6 plots
145 covering terra-firme and igapó; 2.609º S, 60.217º W; Fig 1D) and *Reserva da Campina* (3 plots
146 covering campinas; 2.592º S, 60.030º W; Fig 1E), on the east side of the Negro River; and
147 *Caxiaunã* (12 plots covering campinas, terra-firme, várzea, and igapó), a national forest located
148 350 km west of Belém (1.7352° S, 51,463°W; Fig 1F), which constitutes our easternmost

149 locality. The sample collection was authorized by Brazilian authorities: ICMBio (registration
150 number 48185-2) and IBAMA (registration number 127341).

151

152 2.2. PHYSICOCHEMICAL SOIL ANALYSES

153 We determined the physicochemical soil properties of each plot from three soil samples per plot,
154 totalling 117 samples. The pH was measured in water (soil: water ratio 1:2.5). The exchangeable
155 concentrations were measured for sodium (Na), potassium (K), and phosphorus (P) using
156 Mehlich-1 extraction (unit mg/dm^3) and for calcium and magnesium (Ca, Mg) using KCl (1
157 mol/L) extraction (unit $cmol_c/dm^3$). The sum of all exchangeable bases (SB; which comprises
158 K^+, Ca^{2+}, Mg^{2+}, and Na^+; unit $cmol_c/dm^3$) was then calculated. We also estimated exchangeable
159 aluminium (Al and H+Al unit $cmol_c/dm^3$) extracted with calcium acetate (0.5 mol/L at pH 7.0),
160 aluminium saturation index (m; unit %), and Base Saturation Index (V; unit %). The effective
161 cation exchange capacity (t) as well as the cation exchange capacity (T) were measured at pH 7.0
162 (unit $cmol_c/dm^3$). The organic matter (O.M) was quantified (unit g/kg) and the organic carbon
163 (C) was estimated from the organic matter as: C = MO / 1,724 - Walkley-Black (unit g/kg). Soil
164 texture was characterized by quantifying the fractions of clay (< 0.002 mm), silt (0.002 – 0.05
165 mm), fine sand (0.05 – 0.2 mm), coarse sand (0.2 – 2 mm), and total sand (0.05 – 2 mm) (unit %
166 of soil weight). We did not quantify nitrogen levels due to the highly volatile nature of nitrogen;
167 its concentration changes quickly during sample storage due to the activity of soil bacteria, and
168 freezing the samples in our remote sampling localities was not feasible. All analyses were
169 commissioned from EMBRAPA Ocidental (Brazil), following the protocol described in
170 Donagema et al. (2011). Afterwards, we used the mean of the three soil samples from the same

plot to obtain a representative value for the measurement of each variable for each plot.

2.3. DNA EXTRACTION, AMPLIFICATION, AND SEQUENCING

The detailed laboratory procedures are described in Ritter et al. (2018). Briefly, 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. (2018) 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, and remove singletons. We inferred operational taxonomic units (OTU) at the 97% sequence similarity level as usually used for OTU clustering (meaning that sequences differing by more than 3% are considered to belong to different OTUs; Stackebrandt and Goebel, 1994; Blaxter et al., 2005). We used the SINA v1.2.10 for ARB SVN (revision 21008; Pruesse, Peplies, and Glöckner, 2012) taxonomic reference dataset for both markers and used SILVAngs 1.3 for taxonomic assignments (Quast et al., 2013).

2.6. Construction of corrected OTU tables

Presence/absence analyses – PCR biases, variation in the copy number of 16S/18S genes per cell/genome, as well as differences in size and biomass across the targeted organisms can compromise a straightforward interpretation of OTU reads as an abundance measure (Elbrecht and Leese, 2015; Pawluczyk et al., 2015). Since the number of observed OTUs is dependent on the number of reads, we first rarefied all samples to the lowest number of reads obtained from any one plot (22,209 for 16S and 25,144 for 18S; Fig S1). One sample containing only 1,359 reads was excluded from the 18S data analysis prior to rarefaction to avoid having to downsize the other samples to such a low number of reads (McMurdie and Holmes, 2014). The OTU richness of each plot was computed after rarefaction 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 rarefied OTU tables to presence/absence for both prokaryote (16S) and eukaryote (18S) data.

Abundance analyses Despite known limitations of methods, read abundances can be meaningful, especially for 16S. Therefore, we carried out analyses also using abundance data. We calculated true OTU diversity of order q = 1, which is equivalent to the exponential of the Shannon entropy (Jost, 2006). It can be interpreted as the effective number of OTUs, i.e. the number of OTUs in an idealised community where the geometric mean of the proportional OTU abundances is the same as in the original sample, but all OTUs are equally abundant (Tuomisto, 2010). The effective number of OTUs is more robust against biases arising from uneven sampling depth than the simple number of OTUs, so for diversity we used the unrarefied read counts as OTU abundance. However, the results were virtually identical when we used the

rarefied OTU table (correlation = 1 for both 16S and 18S). For the remaining abundance-based analyses, we transformed read counts using the "varianceStabilizingTransformation" function in DESeq2 (Love et al., 2014) as suggested by McMurdie and Holmes (2014). This transformation normalizes the count data with respect to sample size (number of reads in each sample) and variances, based on fitted dispersion-mean relationships (Love et al., 2014).

2.7 Statistical analyses

Preparation of environmental data – We first normalised all soil variables to zero mean and unit variance using the "scale" function of vegan. We then performed two principal component analyses (PCAs) to reduce the number of variables. The first PCA used the chemical soil properties, i.e. all variables based on concentrations of elements. The second PCA used the physical soil properties, i.e. grain size classes. We input missing sand, silt, and clay information for three plots, based on regression weights from the observed data using the mice v. 2.30 R package (Buuren and Groothuis-Oudshoorn, 2011) before performing the PCAs. We used the first axis of each PCA (explaining 66% and 65% of the total variation, respectively) in the subsequent analyses. Given the expected importance of soil organic carbon content (Nielson et al., 2011) and pH (Lauber et al., 2009), we used these as independent variables.

Hypothesis testing – For all analyses we used pH, organic carbon, chemical PC1, and physical PC1 as explanatory variables. All analyses were carried out using presence-absence data and relative abundance data in parallel: in the case of diversity, richness corresponds to presence-absence data and effective number of OTUs to relative abundance data. Overall, each kind of

238 analysis was carried out eight times: one for each of the eight possible combinations of organism
239 group (prokaryote or 16S, eukaryote or 18S), substrate (soil, litter) and abundance measure
240 (presence-absence, proportional abundance).

241

242 *Do OTU diversities reflect physical or chemical soil properties?* To address the first
243 question, we performed Bayesian general linear models (GLM), as implemented in the R-INLA
244 v. 17.6.20 R package (Rue et al., 2009). The response variables were the eight different variants
245 of OTU diversity in turn, and in each case the soil properties were used as explanatory variables.
246 We tested the effect of spatial auto-correlation by comparing analyses of standard GLMs with
247 GLM analysis using stochastic partial differential equations (SPDE) that explicitly consider
248 spatial correlation.

249

250 *Do OTU community turnovers reflect differences in physical and chemical soil properties?* To
251 address the second question, we performed multiple regressions on dissimilarity matrices
252 (MRM), as implemented in the function "MRM" of the R package *ecodist* v.2.0.1 (Goslee and
253 Urban, 2007). The response variables were dissimilarity matrices based on the eight different
254 variants of OTU turnover (as calculated using the Jaccard dissimilarity) in turn. In each case, the
255 explanatory variables were four distance matrices based on soil properties and one geographical
256 distance matrix (all calculated using Euclidean distances). Statistical significance of the
257 regression coefficients was determined with 10,000 permutations. Additionally, we used
258 variance partitioning analysis based in dissimilarity matrixes to quantify the unique and shared

contributions of each of the explanatory variables to explaining variance in the response distance matrix (Tuomisto and Ruokolainen 2006).

Does OTU diversity of the litter layer predict the OTU diversity of the underlying soil? To address the third question, we analysed the relationship between litter and soil OTU richness and diversity for prokaryotes (16S) and eukaryotes (18S) using a linear regression model (the *lm* function in R).

Are OTU community turnover patterns in the litter layer similar to those in the underlying soil? To address the fourth question, we first performed non-metric multidimensional scaling (NMDS) ordinations as implemented in the metaMDS function in the R package vegan. Compositional dissimilarity was quantified with the Jaccard dissimilarity index. Ordinations based on the same organism group and abundance data type but different substrates were then compared. Next, we used the Permutational Analysis of Variance (PERMANOVA) to assess whether substrate type has an effect on community composition. Finally, we illustrated which were the dominant taxonomic groups (orders or phyla) with bar-plots. As with the other questions, all analyses were repeated for all possible combinations of organism group (prokaryotes or 16S and eukaryotes or 18S) and abundance data type (presence-absence and proportional abundance, in the bar-plots we used the rarefied abundance data).

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), entropart

(Marcon and Hérault, 2015), broom v.0.4.4 (Robinson, 2018), and viridis v. 0.4.0 (Garnier, 2016). Scripts for all analyses are provided in the supplementary material.

3. Results

3.1. OTU diversity and turnover in relation to soil properties (research questions 1 and 2)

In the physical soil data PCA, large values on the first PC were associated with coarse texture (coarse sand fraction loading 0.52, total sand fraction loading 0.55) and small values with fine texture (silt loading -0.45, clay loading -0.38; Table S2). The flooded forests (igapós and várzeas) generally had fine-textured soils (negative values of PC1). The unflooded forests (terra-firme and campinas) were more widely distributed along PC1, with some plots having similar values with várzeas and igapós (Fig. 2A). In the chemical soil data PCA, large values on the first PC were associated with poor soils. The most negative loading was -0.35 for the sum of exchangeable bases (SB), and the largest positive loading was 0.29 for aluminium saturation index (Table S3). The habitat types were not well separated along the first axis of the chemical PCA, as most plots of all habitat types had poor soils (large values of PC1) and just a few scattered várzea and igapó plots had more cation-rich soils (small values of PC1; Fig. 2B).

In general, organic carbon and pH had the strongest effects on OTU diversity. This was the case both for prokaryotes and eukaryotes, for richness and effective number of OTUs) and for soil and litter (Table 1). In addition, PC1 of chemical soil properties was an important predictor for prokaryotic OTU richness in the soil and for prokaryotic effective number of OTUs in both soil

303 and litter, with stronger effect in soil than in litter. For eukaryotes, soil texture had an important
304 effect on OTU diversity, albeit different on each substrate: positive for soil and negative for litter
305 (Table 1). Overall, soil properties had strong effects on OTU diversity in litter.

306

307 OTU community turnover was significantly associated with differences in soil properties,
308 especially those in organic carbon and pH, which were significant for all communities. The pH
309 effect was strong for all prokaryote (16S) datasets and for eukaryotes (18S) in soil when relative
310 abundance data were used (Table 2). Organic carbon had the strongest effect for eukaryotes in
311 soil when presence/absence data were used and for eukaryotes in litter with both
312 presence/absence and relative abundance data. Chemical PC1 was significant for prokaryotes in
313 soil (both presence/absence and relative abundance) and for eukaryotes in litter when
314 presence/absence data were used. Texture PC1 was significant only for eukaryotes in litter (both
315 presence/absence and relative abundance; Table 2). Geographical distance was a significant
316 explanatory factor for all datasets, but as closer places usually are more environmentally similar,
317 we cannot separate the effect of spatial correlation from soil property effects.

318

319 A moderate percentage of the variation in Jaccard dissimilarities was explained by soil
320 physicochemical properties in the presence/absence data, for prokaryotes (31% in soil and 35%
321 in litter; Fig. 3). For eukaryotes, the total explanatory power of soil physicochemical properties
322 was smaller (12% in soil and 16% in litter). For prokaryotes and eukaryotes, the litter
323 communities were more structured by soil characteristics than were the soil communities. All
324 variables explained small but significant proportions of the variance in all communities and

showed some weak but significant interactions considering presence/absence matrices (Fig. 3) and a similar, strong proportion of the variance in abundance data (Fig. S2). Organic carbon had the strongest effect in all substrates and for both organism groups (ranging from 0.03 for eukaryotes in soil through 0.05 for eukaryotes in litter to 0.08 for prokaryotes in both soil and litter).

3.2. Similarities in OTU diversity and turnover patterns between litter and soil (research questions 3-4)

We found a weak positive regression between OTU richness of prokaryotes in litter and in soil (adj. $R^2 = 0.25$, $p < 0.001$; Fig. 4A) and between the effective number of prokaryote OTUs in litter and in soil (adj. $R^2 = 0.1$, $p = 0.03$; Fig. 4B). For eukaryotes, the corresponding correlations were not significant (Fig. 4C and 4D). The plot "CXNCAMP3" had very low soil OTU richness, and excluding this data point strengthened the correlation of OTU richness between soil and litter for prokaryotes (to adj. $R^2 = 0.46$, $p < 0.001$; Fig. S3A), but not for eukaryotes (Fig. S3B).

The OTU communities in litter and in soil tended to be separated in the NMDS ordination space, although there was some overlap especially for the igapó plots (Fig. 5). The PERMANOVA test indicated weak but significant effects (all $p < 0.001$) of substrate type on compositional dissimilarities of both prokaryotes ($R^2 = 0.06$, $F = 5.83$, for presence/absence data and $R^2 = 0.07$, $F = 6.7$, for abundance data) and eukaryotes ($R^2 = 0.03$, $F = 2.8$ for presence/absence data and $R^2 = 0.04$, $F = 3.74$, for abundance data). Habitat type had an even stronger effect on the compositional dissimilarities of both prokaryotes ($R^2 = 0.17$, $F = 5.48$, for presence/absence data

and R^2 = 0.18, F = 5.68, for abundance data) and eukaryotes (R^2 = 0.1, F = 2.8, for presence/absence data and R^2 = 0.1, F = 2.98, for abundance data). Taxonomic composition at the order and phylum level was similar in litter and in soil both for prokaryotes and for eukaryotes (Fig. 6).

4. Discussion

4.1. Soil predictors of OTU diversity and turnover

In this study, we tested the impact of physicochemical soil properties on the OTU diversity and community turnover of prokaryotes and eukaryotes in soil and litter across Brazilian Amazonia. We found that that the soil properties we quantified had variable effects on OTU diversity (richness and effective number of OTUs) and community turnover for litter and soil, and that the effect varied between prokaryotic and eukaryotic organisms. The variable with the highest explanatory power was overall organic carbon for both prokaryotes and eukaryotes. OTU diversity and community turnover were better explained by soil properties in litter than in soil.

Considering the results from the linear models, in general organic carbon and pH were the strongest factors in explaining eukaryotic litter and soil prokaryotic and eukaryotic diversity. Our results show a positive correlation between soil pH and OTU diversity, which is expected since much of the soils in Amazonia are acidic. For instance, for soil communities, 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 highest bacterial diversity. Tropical

368 forests with high macro-organismic diversity had soil with pH < 4.5 and had the lowest bacterial
369 diversity (Lauber, 2009). In our samples, pH was overall low and its variation was moderate,
370 from 3.65 to 5.14, thereby in less acid soil we found highest OTU diversity.

371

372 We also found that pH was significant for all community turnovers, and that it was the strongest
373 variable in explaining community turnover for prokaryotes (in both soil and litter) and for
374 eukaryotes in soil. However, we found that pH had no effect on prokaryote diversity in litter. The
375 consistent effect of pH for prokaryotes and eukaryotes in the soil but inconsistent effect in litter
376 support the findings of Gregorich et al. (2016), who found no correlation of soil properties and
377 litter decomposition. This point to the independence of the environmental factors regulating each
378 substrate.

379

380 We found significant effects of organic carbon on all community turnovers and the strongest
381 effect for eukaryotes in litter. Additionally, we found a negative correlation of soil organic
382 carbon with OTU diversity for all groups. Soil biodiversity has previously been found to have an
383 effect on carbon sequestration (Wagg et al., 2013). However, the relationship between soil
384 biodiversity and carbon has varied across studies (Nielson et al., 2011). Furthermore, Fierer et al.
385 (2012) and de Lima Brossi et al. (2014) found that soil organic matter was related to microbial
386 community composition in several different vegetation types. The negative correlation between
387 soil organic carbon content and OTU diversity reported here might be related to high nutrient
388 turnover in high-diversity soil/litter, keeping the carbon stock locked in aboveground biomass.
389 Our results support the findings of Wall et al. (2008), who found a positive influence of the

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 the related changes in community composition alter the communities´ capacity to break down organic matter and recycle nutrients, slowing down the return of nutrients to the above-ground communities (Wardle et al., 2004). These findings stress the complex nature of carbon-diversity dynamics and the plant–soil feedback loop mediated by soil biota (Mangan et al., 2010). They furthermore highlight a connection between decomposition rates and biodiversity in Amazonia that should be better explored.

Biotic and abiotic interactions jointly determine soil properties, making it important to consider environmental and biological interactions between variables. Indeed, our variance analysis reveals several co-variances between soil properties, such as pH and organic carbon and physical and chemical properties. Although these interactions were weak, this analysis is important for providing a better understanding of the study system. Considering physicochemical soil properties, we had a partial separation of the major environmental types by the properties of their soils. It is in agreement with previous studies, which report an association of soil types and habitats in Amazonia (e.g. Falesi, 1984; Prance, 1996). The soil texture (first axis of the physical PCA) was well separated by the habitat types of flooded forests (igapós and várzeas), whereas the terra-firme and campinas were more spread in physical properties. On the other hand, the first axis of chemical PCA was less well separated for flooded areas (igapós and várzeas). This result was expected since there is a variation within habitat types, especially flooded forests (Kalliola et al., 1993;Tuomisto, Ruokolainen and Yli-Halla, 2003; Tuomisto et al., 2014, 2016). Furthermore, terra-firme forests have been reported to vary in soil nutrients (e.g. Tuomisto,

Ruokolainen and Yli-Halla, 2003; Tuomisto et al., 2014, 2016; Fine et al., 2005), consistent with the variation observed in our plots. However, due to the limited sampling in our studies, the variation we detected was small (Fig. 2B). Our finding that soil texture is similar among the flooded environments (várzeas and igapós) and that soil texture was an important factor for eukaryote diversity (both soil and litter) is consistent with the previously reported community similarity among these environmental types based on the data from the same samples (Ritter et al., 2018).

4.2. Contrasting litter and soil diversity

The correlations between soil and litter OTU richness and the effective number of OTUs were significant for prokaryotes but not for eukaryotes. This is congruent with previous reports that showed the independence of litter accumulation from properties of the underlying soils (Gregorich et al., 2016).

We expected a difference in taxonomic composition between litter and soil communities, with microbes dominating the soil as previously reported (e.g. Bates et al., 2013; Mahé et al. 2017) and plant and nematode OTUs dominating the litter due to its mainly being composed of leaves and roots. However, we found highest plant (Chloroplastida) richness and abundance in soil samples. Furthermore, unlike Porazinska et al. (2012) who found a dominance of nematodes in the litter of tropical forests, we found very similar proportions of nematode OTUs in soil and litter, with the highest richness and abundance in the soil (Fig. 6C and 6D). This pattern of no

clear taxonomic differences between soil and litter layers is consistent with respect to all dominant groups on the order and phylum levels (Fig. 6). This suggests that, on the Amazon basin scale, the taxonomic composition at higher levels (phylum and order) is consistent between litter and soil with some differences in abundance for some groups, such as Arthropoda and Chloroplastida for eukaryotes and Actinobacteri, Bacterioides and Chloroflexi for prokaryotes (Fig. 6).

Interestingly, the phyla that dominated our samples were only partly the same as those found dominant in a large global dataset (Fig. 6A and 6B; Delgado-Baquerizo et al., 2018). While we also found Proteobacteria to be the most frequent phylum considering both presence/absence and relative abundance, the second most frequent phylum was Chloroflexi in presence/absence data for soil and litter and abundance for soil in our samples, while this Chloroflexi only being the 5th most abundant in the global dataset. Actinobacteria, the second more abundant phyla in the global database, was the second most frequent phyla just for abundance in litter samples in our data. Moreover, the rank-abundance distribution of the most dominant phyla was 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 eight phyla representing more than 5% of relative frequency each (> 70% of relative frequency) versus only four phyla in Delgado-Baquerizo et al. (2018). Taken together, these differences highlight the need for more studies across the Amazon basin to better characterize the taxonomic composition.

The OTU community compositions of both prokaryotes and eukaryotes were better explained by

habitat type (terra firme, várzea, igapó, campina) than they were by substrate type (soil, litter), which was expected since both substrates should share a large number of organisms. The substrate types were weakly differentiated at the OTU level, but we could not observe any difference at the phylum or order levels for presence/absence and a small difference in for abundance data (Fig. 6). For instance, fungi usually dominate eukaryotic soil communities in any environment, including tropical forests (Tedersoo et al., 2017), but the dominant fungal taxa (OTU) may vary considerably even on local and sub-local scales (Urbanová et al., 2015). In a study conducted in the western parts of the Czech Republic, similar results for bacteria and fungi were found: the phylum level indicated the same taxonomic groups as dominant in soils and litter, but there were striking differences on the OTU level in these substrates (Urbanová et al., 2015).

5. Conclusions

In this study, we found OTU diversities to be related between soil and litter in prokaryotes, but not in eukaryotes. We also found that physicochemical soil properties can predict soil and litter diversity in Amazonia to some extent. In particular, we found a positive correlation for pH and a negative correlation for soil organic carbon content with respect to prokaryotic and eukaryotic OTU diversity. Furthermore, we found a significant effect of soil organic carbon content on community turnover. In general, our results stress the complexity of soil-biodiversity relationships, and hence the importance of considering multiple factors and their interactions in the characterization of biodiversity patterns. Soil biodiversity is crucial for carbon cycling in terrestrial ecosystems, and our results suggest that additional studies to better understand the

478 relationship between diversity (above and belowground) and carbon cycles may help modelling
479 carbon deposition and biodiversity patterns.

480

481 **Acknowledgements**

482 We thank Anna Ansebo, Sven Toresson, and Ylva Heed for laboratory and administrative
483 assistance; Mats Töpel for help with bioinformatics; and Quiterie Haenel for help with the DNA
484 extractions and fruitful discussions. We thank Hans ter Steege for advice on sampling localities
485 and experimental design. We are grateful to members of our research group for discussions and
486 suggestions. The authors also acknowledge support from the Science for Life Laboratory, the
487 Knut and Alice Wallenberg Foundation, the National Genomics Infrastructure funded by the
488 Swedish Research Council, and the Uppsala Multidisciplinary Center for Advanced
489 Computational Science for funding and/or assistance with massively parallel sequencing and
490 access to the UPPMAX computational infrastructure. Additional computational analyses were
491 run at the University of Gothenburg bioinformatics cluster at the Department of Biological and
492 Environmental Sciences (http://albiorix.bioenv.gu.se/).

493

494 **References**

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

496 Baldeck, C.A., Tupayachi, R., Sinca, F., Jaramillo, N. and Asner, G.P., 2016. Environmental
497 drivers of tree community turnover in western Amazonian forests. *Ecography*, *39*(11), pp.1089-

1099. DOI: 10.1111/ecog.01575.

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

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

Basset, Y., Cizek, L., Cuénoud, P., Didham, R.K., Guilhaumon, F., Missa, O., Novotny, V., Ødegaard, F., Roslin, T., Schmidl, J. and Tishechkin, A.K., 2012. Arthropod diversity in a tropical forest. *Science*, *338*(6113), pp.1481-1484. DOI: 10.1126/science.1226727.

Bates, S.T., Clemente, J.C., Flores, G.E., Walters, W.A., Parfrey, L.W., Knight, R. and Fierer, N., 2013. Global biogeography of highly diverse protistan communities in soil. *The ISME Journal*, *7*(3), p.652. DOI: 10.1038/ismej.2012.147.

Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Floyd, R. and Abebe, E., 2005. Defining operational taxonomic units using DNA barcode data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1462), pp.1935-1943. DOI: 10.1098/rstb.2005.1725

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

Buuren, S. and Groothuis-Oudshoorn, K., 2011. mice: Multivariate imputation by chained

equations in R. *Journal of Statistical Software*, *45*(3).

Cámara-Leret, R., Tuomisto, H., Ruokolainen, K., Balslev, H. and Munch Kristiansen, S., 2017. Modelling responses of western Amazonian palms to soil nutrients. *Journal of 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.

Elbrecht, V. and Leese, F., 2015. Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass—sequence relationships with an innovative metabarcoding protocol. *PLOS ONE*, *10*(7), p.e0130324. DOI: 10.1371/journal.pone.0130324.

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

538 biogeochemical cycles. *Science*, *320*(5879), pp.1034-1039. DOI: 10.1126/science.1153213.

539 Faoro, H., Alves, A.C., Souza, E.M., Rigo, L.U., Cruz, L.M., Al-Janabi, S.M., Monteiro, R.A.,
540 Baura, V.A. and Pedrosa, F.O., 2010. Influence of soil properties on the diversity of bacteria in
541 the Southern Brazilian Atlantic Forest. *Applied and Environmental Microbiology*, *76*(14),
542 pp.4744-4749. DOI: 10.1128/AEM.03025-09.

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

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

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

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

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

557 Goslee, S.C. and Urban, D.L., 2007. The ecodist package for dissimilarity-based analysis of

558 ecological data. *Journal of Statistical Software,* 22(7), pp.1-19.

559 Gregorich, E.G., Janzen, H., Ellert, B.H., Helgason, B.L., Qian, B., Zebarth, B.J., Angers, D.A.,
560 Beyaert, R.P., Drury, C.F., Duguid, S.D. and May, W.E., 2017. Litter decay controlled by
561 temperature, not soil properties, affecting future soil carbon. *Global Change Biology*, *23*(4),
562 pp.1725-1734. DOI: doi.org/10.1111/gcb.13502.

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

564 Higgins, M.A., Ruokolainen, K., Tuomisto, H., Llerena, N., Cardenas, G., Phillips, O.L.,
565 Vásquez, R. and Räsänen, M., 2011. Geological control of floristic composition in Amazonian
566 forests. *Journal of Biogeography*, *38*(11), pp.2136-2149. DOI: 10.1111/j.1365-
567 2699.2011.02585.x.

568 Hinsinger, P., Bengough, A.G., Vetterlein, D. and Young, I.M., 2009. Rhizosphere: biophysics,
569 biogeochemistry and ecological relevance. *Plant and Soil*, *321*(1-2), pp.117-152.

570 IBGE 2004. Mapa de Vegetação do Brasil 1:5,000,000. - IBGE - Instituto Brasileiro de
571 Geografia e Estatística.

572 Jost, L., 2006. Entropy and diversity. *Oikos, 113*(2), pp.363-375. DOI: 10.1111/j.2006.0030-
573 1299.14714.x.

574 Junk, W.J., Piedade, M.T.F., Schöngart, J., Cohn-Haft, M., Adeney, J.M. and Wittmann, F.,
575 2011. A classification of major naturally-occurring Amazonian lowland
576 wetlands. *Wetlands, 31*(4), pp.623-640. DOI: 10.1007/s13157-011-0190-7.

577 Kalliola, R., Puhakka, M., Salo, J., Linna, A. and Räsänen, M., 1993. Mineral nutrients in fluvial

578 sediments from the Peruvian Amazon. *Catena*, *20*(3), pp.333-349. DOI: 10.1016/0341-
579 8162(93)90009-E.

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

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

586 Laurance, S.G., Laurance, W.F., Andrade, A., Fearnside, P.M., Harms, K.E., Vicentini, A. and
587 Luizão, R.C., 2010. Influence of soils and topography on Amazonian tree diversity: a
588 landscape-scale study. *Journal of Vegetation Science*, *21*(1), pp.96-106. DOI: 10.1111/j.1654-
589 1103.2009.01122.x.

590 Love, M.I., Huber, W. and Anders, S., 2014. Moderated estimation of fold change and dispersion
591 for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), p.550. DOI: 10.1186/s13059-014-
592 0550-8.

593 Mahé, F., de Vargas, C., Bass, D., Czech, L., Stamatakis, A., Lara, E., Singer, D., Mayor, J.,
594 Bunge, J., Sernaker, S. and Siemensmeyer, T., 2017. Parasites dominate hyperdiverse soil protist
595 communities in Neotropical rainforests. *Nature Ecology & Evolution,* 1(4), p.0091.

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

599 Marcon, E. and Hérault, B., 2015. entropart: An R package to measure and partition
600 diversity. *Journal of Statistical Software*, *67*(8). DOI: 10.18637/jss.v067.i08.

601 McMurdie, P.J. and Holmes, S., 2014. Waste not, want not: why rarefying microbiome data is
602 inadmissible. *PLoS Computational Biology*, *10*(4), p.e1003531. DOI: 10.1371/
603 journal.pcbi.1003531.

604 Myster, R.W., 2016. The physical structure of forests in the Amazon Basin: a review. *The*
605 *Botanical Review*, *82*(4), pp.407-427. DOI: 10.1007/s12229-016-9174-x.

606 Myster, R.W., 2017. A comparison of the forest soils in the Peruvian Amazon: Terra firme,
607 palm, white sand and igap. *Journal of Soil Science and Environmental Management*, *8*(7),
608 pp.130-134. DOI: 10.5897/JSSEM2017.0644.

609 Navarrete, A.A., Kuramae, E.E., De Hollander, M., Pijl, A.S., Van Veen, J.A. and Tsai, S.M.,
610 2013. Acidobacterial community responses to agricultural management of soybean in Amazon
611 forest soils. *FEMS Microbiology Ecology*, *83*(3), pp.607-621. DOI: 10.1111/1574-6941.12018.

612 Nielsen, U.N., Ayres, E., Wall, D.H. and Bardgett, R.D., 2011. Soil biodiversity and carbon
613 cycling: a review and synthesis of studies examining diversity–function relationships. *European*
614 *Journal of Soil Science*, *62*(1), pp.105-116. DOI: 10.1111/j.1365-2389.2010.01314.x.

615 Oksanen, J., Kindt, R., Legendre, P., O'hara, B., Stevens, M.H.H. 2007. The vegan
616 package. *Community ecology package*, *10*, pp.631-637.

617 Osborne, C.A., Zwart, A.B., Broadhurst, L.M., Young, A.G. and Richardson, A.E., 2011. The
618 influence of sampling strategies and spatial variation on the detected soil bacterial communities

619 under three different land-use types. *FEMS Microbiology Ecology*, *78*(1), pp.70-79. DOI:
620 10.1111/j.1574-6941.2011.01105.x.

621 Pawluczyk, M., Weiss, J., Links, M.G., Aranguren, M.E., Wilkinson, M.D. and Egea-Cortines,
622 M., 2015. Quantitative evaluation of bias in PCR amplification and next-generation sequencing
623 derived from metabarcoding samples. *Analytical and Bioanalytical Chemistry*, *407*(7), pp.1841-
624 1848. DOI: 10.1007/s00216-014-8435-y.

625 Peres-Neto, P.R., Legendre, P., Dray, S. and Borcard, D., 2006. Variation partitioning of species
626 data matrices: estimation and comparison of fractions. *Ecology,* 87(10), pp. 2614-2625.

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

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

632 Pruesse, E., Peplies, J. and Glöckner, F.O., 2012. SINA: accurate high-throughput multiple
633 sequence alignment of ribosomal RNA genes. *Bioinformatics*, *28*(14), pp.1823-1829. DOI:
634 10.1093/bioinformatics/bts252.

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

638 QGIS Development Team, 2012. QGIS Geographic Information System. Open Source

639 Geospatial Foundation Project.

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

642 Ritter, C.D., Zizka, A., Barnes, C., Nilsson, R.H., Roger, F., and Antonelli, A., (Accepeted).
643 Locality or habitat? Exploring predictors of biodiversity in Amazonia. *Ecography.* DOI:
644 10.1111/ecog.03833.

645 Robinson, D., 2017. broom: Convert statistical analysis objects into tidy data frames. *R package*
646 *version 0.4.4.*

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

650 Stackebrandt, E. and Goebel, B.M., 1994. Taxonomic note: a place for DNA-DNA reassociation
651 and 16S rRNA sequence analysis in the present species definition in bacteriology. *International*
652 *Journal of Systematic and Evolutionary Microbiology*, *44*(4), pp.846-849. DOI:
653 10.1099/00207713-44-4-846.

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

657 Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C. and Willerslev, E., 2012. Towards
658 next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, *21*(8),

pp.2045-2050. DOI: 10.1111/j.1365-294X.2012.05470.x.

Tang, Y., Horikoshi, M. and Li, W., 2016. ggfortify: Unified interface to visualize statistical results of popular R Packages. *The R Journal*, *8*(2), pp.478-489.

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. 2010. A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. *Ecography* 33: 2–22. DOI: 10.1111/j.1600-0587.2009.05880.x

Tuomisto, H., Ruokolainen, K. and Yli-Halla, M., 2003. Dispersal, environment, and floristic variation of western Amazonian forests. *Science, 299*(5604), pp.241-244. DOI: 10.1126/science.1078037.

Tuomisto, H., Poulsen, A.D., Ruokolainen, K., Moran, R.C., Quintana, C., Celi, J., and Cañas, G.

679 2003. Linking floristic patterns with soil heterogeneity and satellite imagery in Ecuadorian
680 Amazonia. *Ecological Applications* 13: 352–371. DOI: 10.1890/1051-
681 0761(2003)013[0352:LFPWSH]2.0.CO;2

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

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

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

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

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

696 Wall, D. H., Bradford, M. A., S T, J., Trofymow, J. A., Behan-Pelletier, V., Bignell, D. E.,
697 Wardle, D.A., Bardgett, R.D., Klironomos, J.N., Setälä, H., Van Der Putten, W.H. and Wall,
698 D.H., 2004. Ecological linkages between aboveground and belowground
699 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: BioSample accessions SAMN09081679-SAMN09081756; NCBI SRA: PRJNA464362, Bioproject 410658.
- Final DNA sequence assembly uploaded as online supplementary items
- Sampling locations, soil physical-chemical data, OTU tables, and R scripts: included as supplementary material.

Authors' contributions

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

Table 1(on next page)

Soil effects on OTU richness and Shannon diversity:

The importance of soil properties differed between taxon, substrate and diversity metrics. Carbon content and pH were important in most of the cases. The table shows the coefficients of each predictor in four Bayesian general multivariate regression model using Stochastic Partial Differential Equations (SPDE) that explicitly consider spatial correlation, modelling OTU richness and effective number of OTUs dependent on soil properties for eukaryotes and prokaryotes in litter and soil, respectively. As the organic carbon content and pH are important variables for soil biota, we use them as independent variables. Bold indicates important predictor variables (credible intervals not crossing zero).

Table 1: Soil effects on OTU richness and Shannon diversity: The importance of soil properties differed between taxon, substrate and diversity metrics. Carbon content and pH were important in most of the cases. The table shows the coefficients of each predictor in four Bayesian general multivariate regression model using Stochastic Partial Differential Equations (SPDE) that explicitly consider spatial correlation, modelling OTU richness and effective number of OTUs dependent on soil properties for eukaryotes and prokaryotes in litter and soil, respectively. As the organic carbon content and pH are important variables for soil biota, we use them as independent variables. Bold indicates important predictor variables (credible intervals not crossing zero).

			OTU richness			Effective number of OTUs		
Taxon	**Substrate**	**Predictor**	**Mean**	**0.025 quantile**	**0.975 quantile**	**Mean**	**0.025 quantile**	**0.975 quantile**
Prokaryotes	**Soil**	**Intercept**	6.14	2.31	9.95	6.07	2.03	10.43
		pH	**0.22**	**0.16**	**0.27**	**0.22**	**0.17**	**0.27**
		Carbon	**-0.13**	**-0.19**	**-0.08**	**-0.15**	**-0.20**	**-0.10**
		Chemical	**-0.08**	**-0.14**	**-0.02**	**-0.08**	**-0.14**	**-0.02**
		Physical	**0.02**	**0.00**	**0.04**	0.02	-0.01	0.04
	Litter	**Intercept**	4.14	-4.78	12.65	5.04	-3.97	13.71
		pH	0.03	-0.03	0.08	**0.05**	**0.00**	**0.10**
		Carbon	**-0.23**	**-0.28**	**-0.17**	**-0.21**	**-0.26**	**-0.16**
		Chemical	**-0.07**	**-0.12**	**-0.01**	-0.05	-0.11	0.01
		Physical	**0.13**	**0.10**	**0.15**	**0.12**	**0.10**	**0.15**
Eukaryotes	**Soil**	**Intercept**	3.28	-7.36	13.39	3.22	-6.49	13.22
		pH	**0.30**	**0.22**	**0.38**	**0.33**	**0.25**	**0.41**
		Carbon	**-0.36**	**-0.46**	**-0.26**	**-0.36**	**-0.46**	**-0.27**
		Chemical	-0.08	-0.20	0.04	-0.05	-0.17	0.06
		Physical	**0.11**	**0.07**	**0.15**	**0.10**	**0.07**	**0.14**
	Litter	**Intercept**	4.93	-10.81	19.82	6.07	-10.14	21.34
		pH	**0.24**	**0.15**	**0.34**	**0.26**	**0.16**	**0.35**
		Carbon	**-0.22**	**-0.32**	**-0.13**	**-0.23**	**-0.32**	**-0.14**
		Chemical	-0.01	-0.12	0.10	0.00	-0.10	0.11
		Physical	**-0.14**	**-0.19**	**-0.09**	**-0.15**	**-0.20**	**-0.10**

Table 2(on next page)

Association between environmental distance and community turnover:

Community dissimilarity is significantly associated with geographical and soil environmental distance for eukaryote and prokaryote communities in soil and litter. The Multiple Regressions were based on the geographical distance, Euclidean distance matrices of soil properties and community Jaccard dissimilarity indexes. Geographic distances were significant for all communities turnover; however, as geographical closest places are usually more environmental similar, we cannot separate the effect of soil properties from the spatial correlation. All community turnovers were significant with 10,000 permutations ($p < 0.001$) with the follow R^2: prokaryote soil ($R^2 = 0.36$for presence/absence and $R^2 = 0.36$ for relative abundance), prokaryote litter ($R^2 = 0.39$ for presence/absence and $R^2 = 0.35$ for relative abundance), eukaryote soil ($R^2 = 0.21$ for presence/absence and $R^2 = 0.20$ for relative abundance) and eukaryote litter ($R^2 = 0.32$ for presence/absence and $R^2 = 0.30$ for relative abundance).

Table 2: Association between environmental distance and community turnover: Community dissimilarity is significantly associated with geographical and soil environmental distance for eukaryote and prokaryote communities in soil and litter. The Multiple Regressions were based on the geographical distance, Euclidean distance matrices of soil properties and community Jaccard dissimilarity indexes. Geographic distances were significant for all communities turnover; however, as geographical closest places are usually more environmental similar, we cannot separate the effect of soil properties from the spatial correlation. All community turnovers were significant with 10,000 permutations ($p < 0.001$) with the follow R^2: prokaryote soil (R^2 = 0.36for presence/absence and $R^2 = 0.36$ for relative abundance), prokaryote litter ($R^2 = 0.39$ for presence/absence and $R^2 = 0.35$ for relative abundance), eukaryote soil ($R^2 = 0.21$ for presence/absence and $R^2 = 0.20$ for relative abundance) and eukaryote litter ($R^2 = 0.32$ for presence/absence and $R^2 = 0.30$ for relative abundance).

			Presence/besence		Relative abundance	
Taxon	**Substrate**	**Predictor**	**coefficients**	**p value**	**coefficients**	**p value**
Prokaryotes	**Soil**	**Intercept**	-24.28	1.00	-8.63	1.00
		Geo_dist	**0.15**	**0.00**	**0.15**	**0.00**
		Chemical	**0.23**	**0.00**	**0.21**	**0.01**
		Physical	**0.18**	**0.01**	**0.17**	**0.01**
		pH	**0.30**	**0.00**	**0.34**	**0.00**
		Carbon	**0.20**	**0.01**	**0.16**	**0.03**
	Litter	**Intercept**	-60.59	1.00	-37.45	1.00
		Geo_dist	**0.19**	**0.00**	**0.20**	**0.00**
		Chemical	**0.18**	**0.00**	0.11	0.09
		Physical	**0.16**	**0.00**	**0.19**	**0.00**
		pH	**0.32**	**0.00**	**0.33**	**0.00**
		Carbon	**0.32**	**0.00**	**0.28**	**0.00**
Eukaryotes	**Soil**	**Intercept**	43.54	1.00	59.63	1.00
		Geo_dist	0.09	0.06	**0.10**	**0.03**
		Chemical	0.12	0.16	0.11	0.21
		Physical	**0.15**	**0.04**	**0.16**	**0.03**

		pH	**0.22**	**0.00**	**0.24**	**0.00**
		Carbon	**0.29**	**0.00**	**0.23**	**0.01**
	Litter	**Intercept**	-8.86	1.00	10.65	1.00
		Geo_dist	**0.15**	**0.00**	**0.17**	**0.00**
		Chemical	**0.26**	**0.00**	**0.22**	**0.00**
		Physical	0.09	0.13	0.05	0.40
		pH	**0.19**	**0.00**	**0.24**	**0.00**
		Carbon	**0.33**	**0.00**	**0.29**	**0.00**

13

Figure 1

Study area and sampling locations.

Inset panels show details of each locality. A: Benjamin Constant; B: Jaú; C: Jaú, naturally open areas; D: Cuieras; E: Cuieras, naturally open areas; and F: Caxiuanã. The symbols in A–F represent different vegetation types that are linked with different soil properties: circles = open areas; triangles = forest seasonally flooded by black water rivers; squares = unflooded forest; and crosses = forest seasonally flooded by white water rivers. The sampling strategy was designed to cover a wide longitudinal range in Amazonia. The map was constructed using QGIS (2012).

**Note: Auto Gamma Correction was used for the image. This only affects the reviewing manuscript. See original source image if needed for review.*

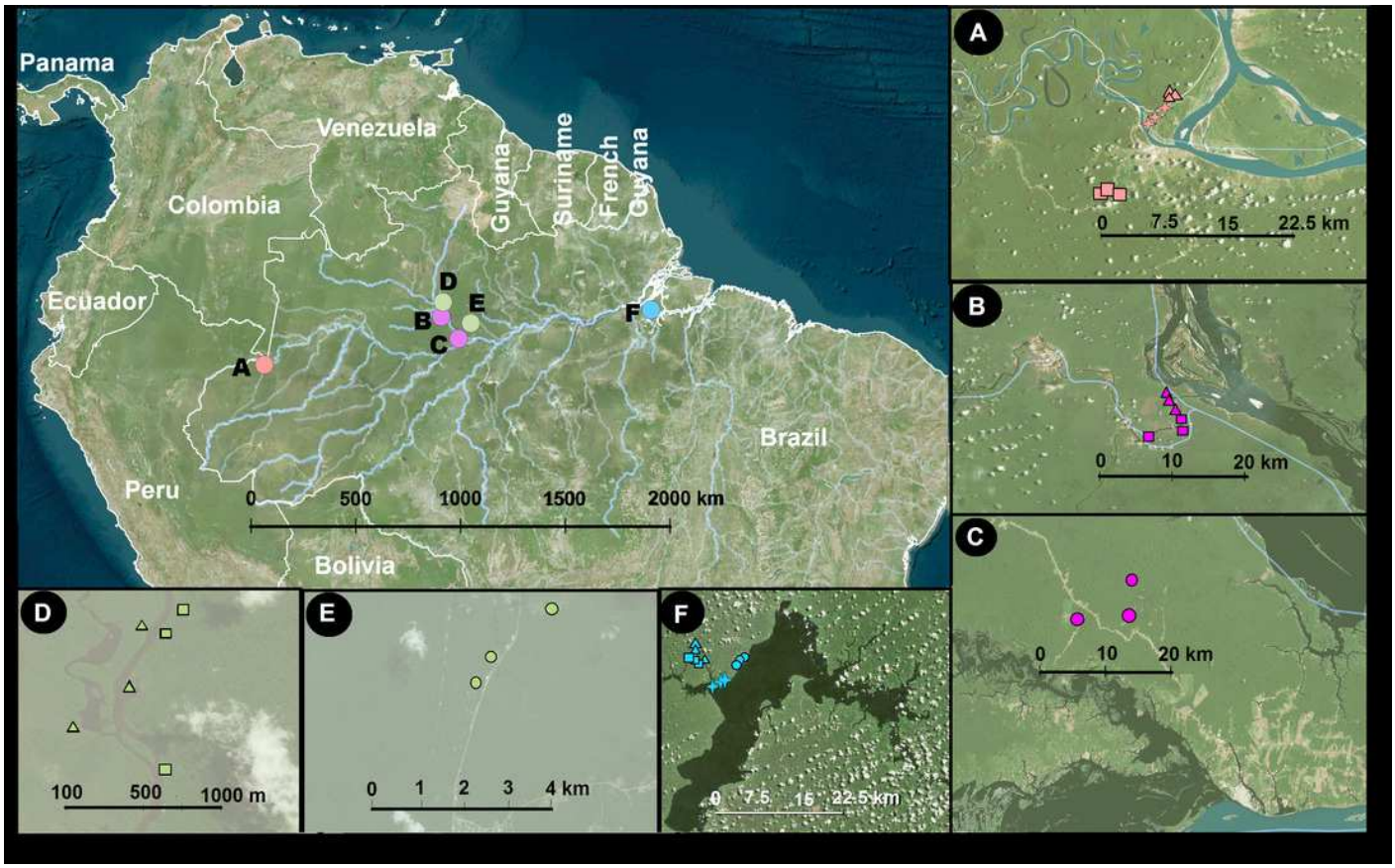

Figure 2

Physical and chemical soil similarity of sample sites across Amazonia.

The figure shows the study sites –coloured by habitat type– on the first two axes of a Principle Component Analysis for A) physical properties (silt, clay and fine, coarse, and total sand fraction) and B) chemical proprieties (phosphorus [P], exchangeable bases [Na, K, Ca, and Mg], the sum of all exchangeable bases [m], exchangeable aluminium [Al and H+Al], saturation index by aluminium [m], base saturation index [V], effective cation exchange capacity [t], and cation exchange capacity [T]). The blue rows show the values of each variable loadings in the two firs PCs. For physical PCA we can observe that flooded forest (igapós and várzeas) are associated with negative values in the first PC axis and a more spread distribution of terra-firme and campinas. For chemical PCA there is no separation of flooded forest, but campinas group in most positive values in the first PCA axis followed by terra-firmes.

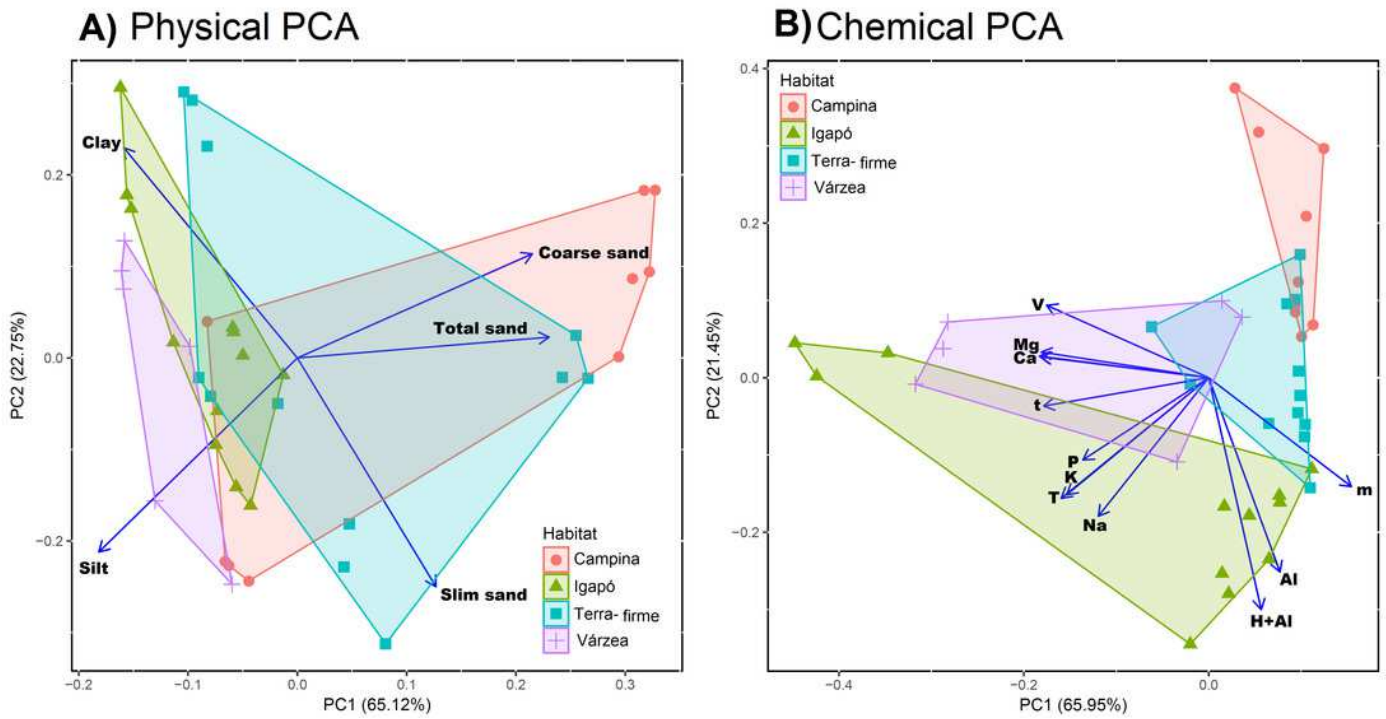

Figure 3

Variation in OTU community composition in Amazonian soil samples explained by soil characteristics.

Results of the variance partition analysis based on Jaccard dissimilarity distance-based analysis. Small but significant proportions of soil and litter communities vary with soil variables, and a small but significant proportion shows variation shared by soil variables. All values represent the proportion of variation explained by the factor/interaction. Chemical variables are shown in green (based on the first PCA of chemical variables, see Table S3 for details), physical variables in purple (based on first PCA axis of soil texture, see Table S2 for details), pH in blue, and carbon content in orange. The prokaryote communities are more structured by soil characteristics than are the eukaryote ones. Inside each taxonomic group, the litter communities are more structured by soil characteristics than are the soil communities.

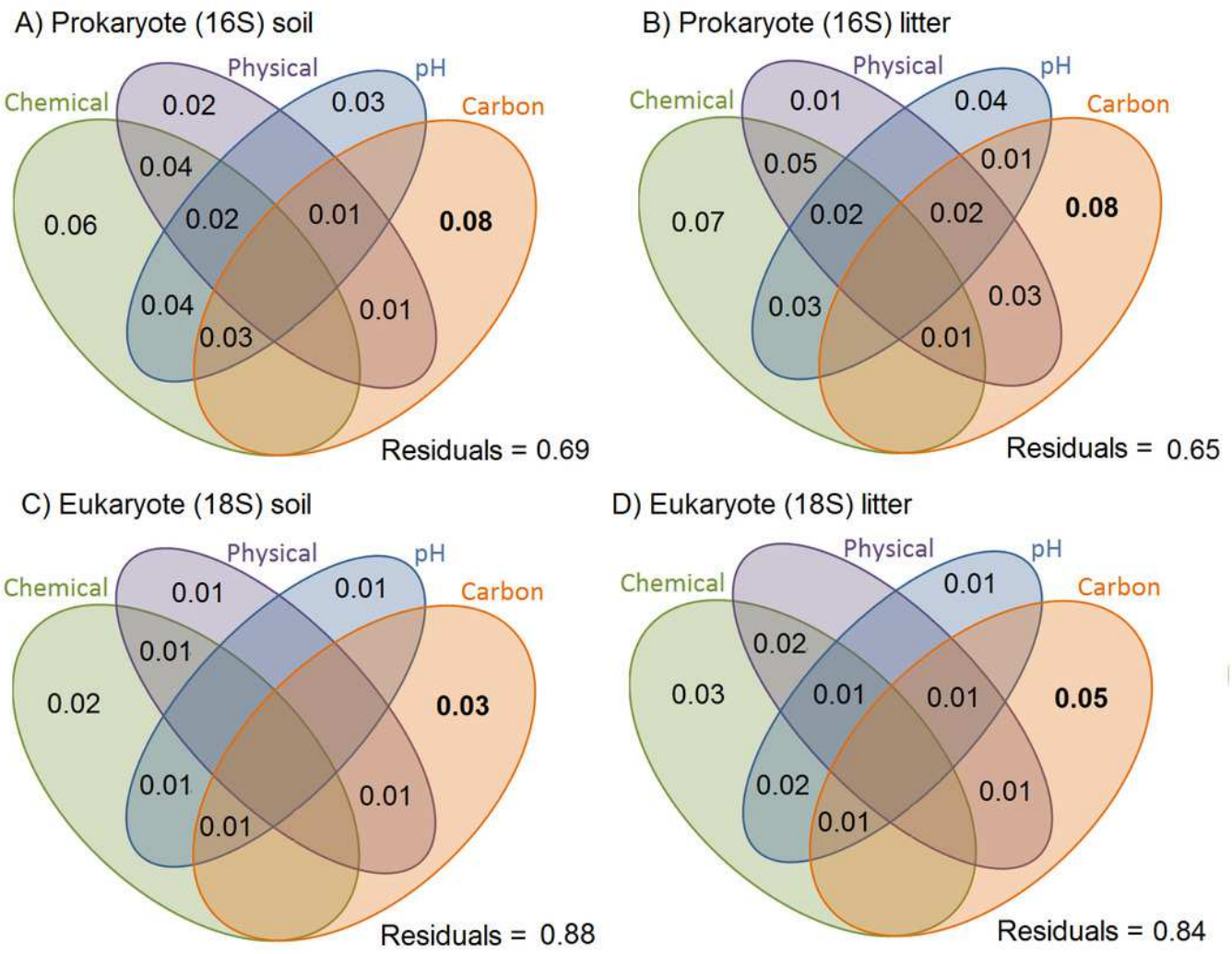

Figure 4

Relation between OTU diversity in soil and litter:

Prokaryotes (16S) are showed in A) OTU richness and B) effective number of OTUs; eukaryotes (18S) are showed in C) OTU richness and D) effective number of OTUs in the Amazonian soil samples**.** The blue line shows a linear regression with standard error indicated by the shaded area for significant correlations. The relationship between soil layers (litter *vs* soil) is weak and differs between taxa, with only prokaryotes showing a significant correlation for richness and effective number of OTUs. This result suggests that it OTU diversity in litter is unsuitable as proxy for the OTU diversity in the soil and *vice versa.*

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27012v1 | CC BY 4.0 Open Access | rec: 28 Jun 2018, publ: 28 Jun 2018

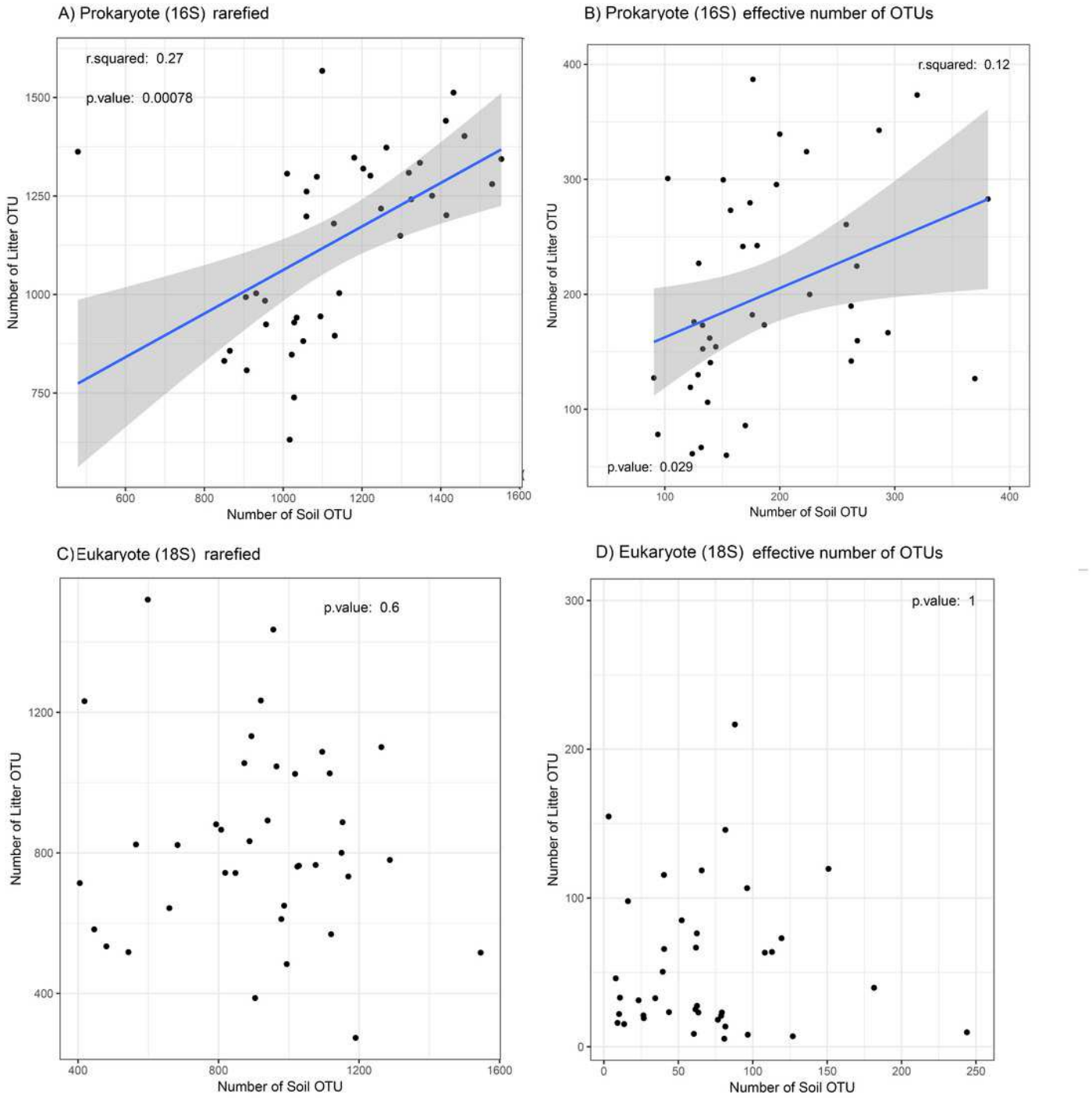

Figure 5

Community structure related to substrate type (litter and soil) and habitat types.

Visualization of differences in OTU composition (measured with the presence/absence matrixes using Jaccard dissimilarity index in A and B; and measured with relative abundance matrixes using Bray-Curtis dissimilarity index) using non-metric multidimensional scaling (NMDS) for A) and C) prokaryotes (16S) and B) and D) eukaryotes (18S). Symbols represent different habitats. Blue represent litter samples and green soil samples. A small but statistically significant (PERMANOVA test) separation between the substrates can be observed along the second ordination axis for both groups of organisms. The strongest and most significant separation is observed between habitat types.

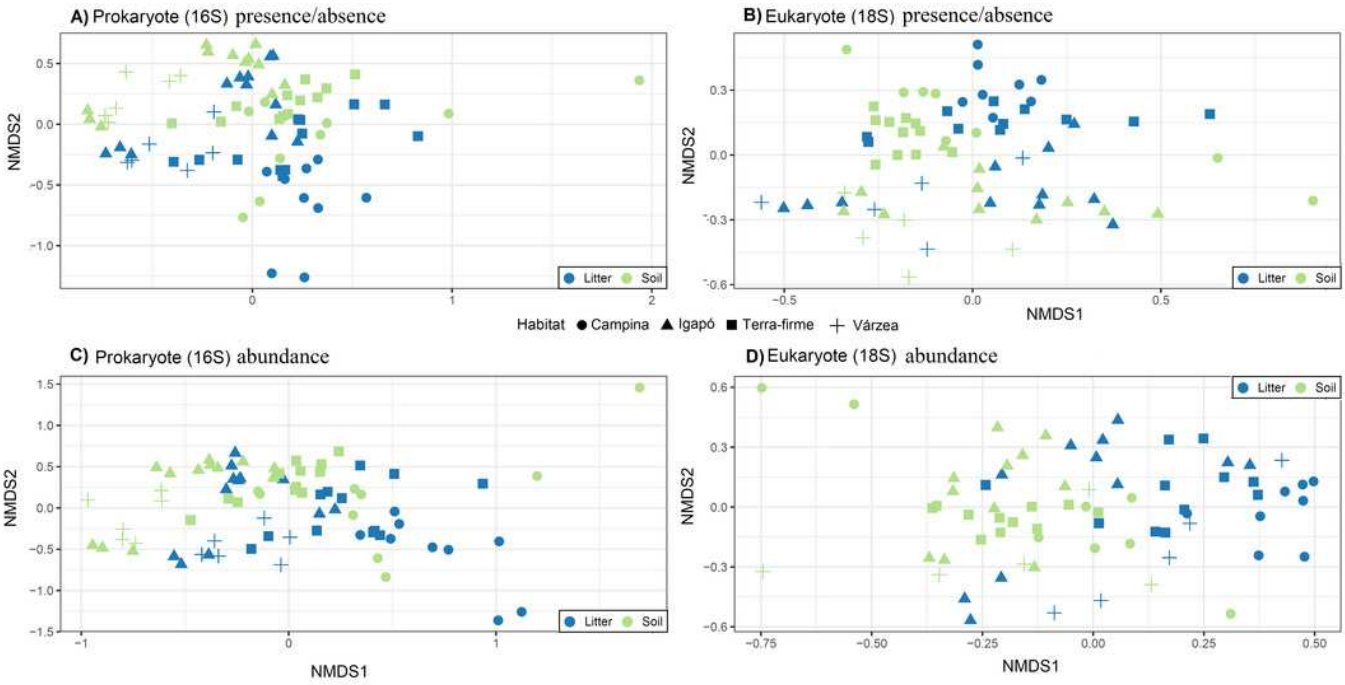

Figure 6

Taxonomic composition of Amazonian soil and litter micro-organismic communities.

The plots show the fraction of OTUs divided by taxonomic group for A) relative frequency of OTU presence for prokaryotes; B) relative abundance of OTU for prokaryotes; C) relative frequency of OTU presence for eukaryotes; D) relative abundance of OTU for eukaryotes. There is no clear taxonomic variation among groups in litter *vs* soil, in either the prokaryote or the eukaryote data for presence/absence. The relative abundance data shows a highest frequency of Actinobacteria in litter compared with soil and a highest abundance of Chloroflexi and Bacterioides in soil sample for prokaryotes. For eukaryotes is possible observe a highest relative abundance of Arthropoda and Chloroplastida in soil than litter samples.