

ECOGRAPHY

Research

Locality or habitat? Exploring predictors of biodiversity in Amazonia

Camila D. Ritter, Alexander Zizka, Christopher Barnes, R. Henrik Nilsson, Fabian Roger and Alexandre Antonelli

C. D. Ritter (<http://orcid.org/0000-0002-3371-7425>) (camila.ritter@bioenv.gu.se), A. Zizka, R. H. Nilsson and A. Antonelli, Gothenburg Global Biodiversity Centre, Göteborg, Sweden, and Dept of Biological and Environmental Sciences, Univ. of Gothenburg, Göteborg, Sweden. AA also at: Gothenburg Botanical Garden, Göteborg, Sweden, and Dept of Organismic and Evolutionary Biology, Harvard Univ., Cambridge, MA, USA. – C. Barnes, Natural History Museum of Denmark, Univ. of Copenhagen, Copenhagen, Denmark. – F. Roger (<http://orcid.org/0000-0001-9865-7542>), Centre for Environmental and Climate Research, Lund Univ., Lund, Sweden.

Ecography

41: 1–13, 2018

doi: 10.1111/ecog.03833

Subject Editor: Miguel Araújo
Editor-in-Chief: Miguel Araújo
Accepted 16 May 2018



Amazonia is an environmentally heterogeneous and biologically megadiverse region, and its biodiversity varies considerably over space. However, existing knowledge on Amazonian biodiversity and its environmental determinants stems almost exclusively from studies of macroscopic above-ground organisms, notably vertebrates and trees. In contrast, diversity patterns of most other organisms remain elusive, although some of them, for instance microorganisms, constitute the overwhelming majority of taxa in any given location, both in terms of diversity and abundance. Here, we use DNA metabarcoding to estimate prokaryote and eukaryote diversity in environmental soil and litter samples from 39 survey plots in a longitudinal transect across Brazilian Amazonia using 16S and 18S gene sequences, respectively. We characterize richness and community composition based on operational taxonomic units (OTUs) and test their correlation with longitude and habitat. We find that prokaryote and eukaryote OTU richness and community composition differ significantly among localities and habitats, and that prokaryotes are more strongly structured by locality and habitat type than eukaryotes. Our results 1) provide a first large-scale mapping of Amazonian soil biodiversity, suggesting that OTU richness patterns might follow substantially different patterns from those observed for macro-organisms; and 2) indicate that locality and habitat factors interact in determining OTU richness patterns and community composition. This study shows the potential of DNA metabarcoding in unveiling Amazonia's outstanding diversity, despite the lack of complete reference sequence databases for the organisms sequenced.

Keywords: biodiversity gradient, metabarcoding, operational taxonomic units (OTUs)

Background

The Amazon drainage basin (Amazonia) has the highest biodiversity of all tropical rainforests and is a global biodiversity hotspot (Hansen et al. 2013). However, large areas of Amazonia are severely understudied, and a major proportion of its biodiversity remains



www.ecography.org

© 2018 The Authors. This is an Online Open article

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

poorly characterized. Despite the gaps in our understanding of Amazonian biodiversity, some broad-scale patterns have been identified.

On a large scale, a west-to-east diversity gradient from highly diverse areas on the Andean slopes in western Amazonia towards relatively less diverse areas on the Guiana shield and the eastern Amazonian lowlands has been observed in many animal and plant groups (ter Steege et al. 2003, Bass et al. 2010, Hoorn et al. 2010, Zizka et al. 2018). The drivers for this gradient remain elusive, but have been attributed to bedrock geology (Tuomisto et al. 2017), historical processes including mountain and basin formation (Hoorn et al. 2010), marine incursions (Bates 2001, Lovejoy et al. 2006, Antonelli et al. 2009), and soil fertility (Moran et al. 2000).

On a finer scale, distinct vegetation types play a major role in structuring the distribution of plant and animal species

throughout Amazonia. These vegetation (habitat) types are closely linked to soil characteristics, flooding regime, and nutrient availability. The four most widespread and important habitat types are: 1) unflooded terra-firme forests, considered the most biodiverse habitat type, are generally characterized by latosols (Falesi 1984, Fig. 1D); 2) várzeas, fertile forests that are seasonally flooded by nutrient-rich white-water rivers up to 240 d yr⁻¹ (Junk et al. 1989; Fig. 1C); 3) igapós, less fertile forests that are seasonally flooded by nutrient-poor black-water rivers (Junk et al. 2011; Fig. 1F); and 4) the naturally open areas, which are forest-free 'islands' in the 'sea' of forest, dominated by grasses and shrubs (campinas) or low canopy forest (Fine et al. 2005; Fig. 1E). Terra-firme forests cover the majority of Amazonia in terms of area, whereas várzeas and igapós jointly cover between 5–7% (Peres 1997), and naturally open areas around 1.6%. These four habitat types support distinct communities of plants and animals

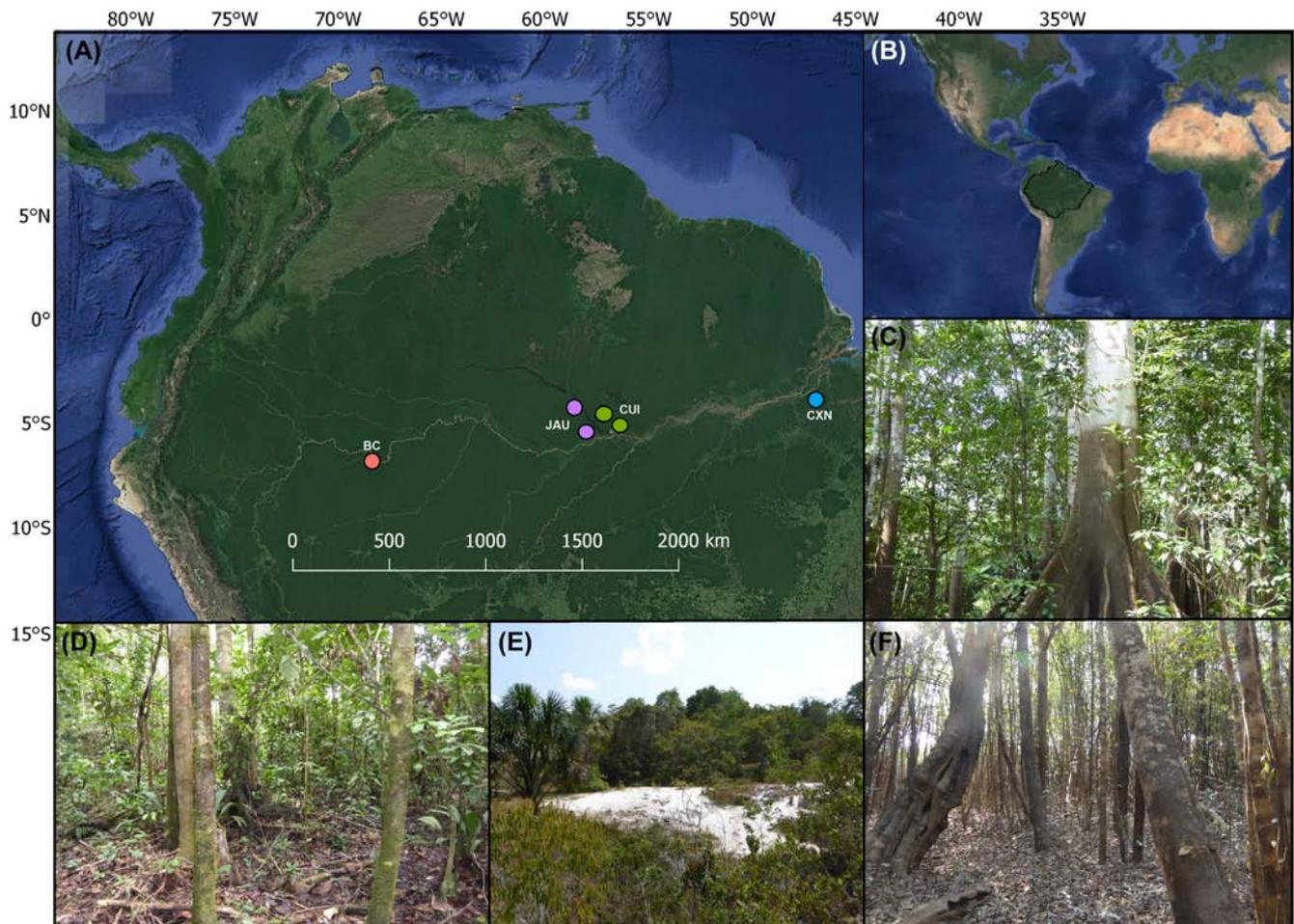


Figure 1. Sampling locations and major vegetation types in Amazonia surveyed in this study. Map constructed with QGIS (2012). The sampling was designed to cover a wide longitudinal range in Amazonia as well as all main habitat types. (A) Map with the main sampling localities (BC = Benjamin Constant; JAU = Jaú; CUI = Cuieras and; CXN = Caxaiuanã); (B) the Amazon drainage basin; (C) várzea (seasonally flooded forest) from Benjamin Constant, detail of water mark at ca 15 m height; (D) terra-firme forest from Benjamin Constant, detail of highly diverse understorey vegetation; (E) Campina (white-sand areas) from Campina's reserve, detail of exposed white sand soil and low vegetation cover and; (F) Igapó (seasonally dry water streams) from Cuieras reserve, detail showing less diverse understorey vegetation.

and are often associated with differences in species richness and community composition.

The biodiversity patterns described for Amazonia have been deduced from large organisms such as birds, mammals, and trees (Bass et al. 2010, Hoorn et al. 2010, Zizka et al. 2018). In contrast, we know very little about the ‘hidden biodiversity’ of inconspicuous but species-rich microorganisms such as fungi, nematodes, and bacteria (but see Bates et al. 2013, Peay et al. 2013, Lentendu et al. 2017, Mahé et al. 2017). This is problematic, because these groups likely comprise the vast majority of biodiversity in terrestrial ecosystems (Mora et al. 2011) and play pivotal roles in nutrient cycles and ecosystem functioning (Dominati et al. 2010). This lack of basic understanding of biodiversity pattern is not only problematic from a scientific perspective, but may also compromise the effectiveness of conservation strategies and hinder the sustainable development of Amazonia.

The scarce knowledge of the majority of Amazonian biodiversity is mostly due to difficulties in sampling of, and taxonomic identification in, these groups, notably insects, fungi, nematodes, and bacteria. High-throughput DNA sequencing methods overcome many of the present challenges to studying biodiversity (Biggs et al. 2015). In particular, metabarcoding (Taberlet et al. 2012) allows quantification of genetic, and, to a certain degree taxonomic, diversity of a locality without the need to spend years collecting and examining specimens (Gibson et al. 2014). Even in highly diverse and poorly sampled environments such as tropical rainforests – areas for which reference sequence databases are very thinly populated – the use of operational taxonomic units (OTUs, Blaxter et al. 2005) defined on the basis of sequence similarity makes the assessment and comparison of biodiversity across sites possible (Balmford and Whitten 2003, Giam et al. 2012, Stahlhut et al. 2013).

In this study, we investigate patterns in genetic soil and litter diversity on a west-to-east transect across Brazilian Amazonia along the Amazon River. We analyse environmental DNA through metabarcoding, using ribosomal 16S (prokaryote) and 18S (eukaryote) gene sequences as proxies to provide one of the first large-scale biodiversity assessments across Amazonia. Specifically, we test two hypotheses on correlations of OTU richness and community structure from 39 study plots.

Hypothesis 1 – microorganism OTU richness is similar to the patterns previously documented for vertebrates and plants, decreasing in richness from west-to-east throughout Amazonia. At the finer level, richness is determined by habitat type, with a gradient related to low stress level and high nutrient availability (terra-firme > várzea > igapó > campinas).

Hypothesis 2 – OTU community structure is linked to geographic vicinity, such that community similarity decreases with geographic distance as observed in vertebrates. OTU community structure reflects habitat type and is more similar within than among habitat types.

Material and methods

Sampling localities

We sampled 39 plots at four localities across a large longitudinal range in Brazilian Amazonia (Fig. 1). The fieldwork took place between 5th and 29th of November 2015. Localities were selected to maximize geographic distance and the number of habitat types (terra-firme, igapós, várzeas, and campinas). The locations are: A) Benjamin Constant (BC) – a municipality near the border of Brazil, Colombia, and Peru, situated approximately 1100 km west of Manaus at the upper Amazonas River (4.383°S, 70.017°W). This region is accessible by boat only and is characterized by a low human population density and relatively low rates of deforestation. The region is situated in the southern margin of the Amazon River and supports large areas of várzeas, terra-firme, and some igapó forests. B) Jaú (JAU) – a national park that encompasses an area of 2 272 000 ha. It is located on the lower Rio Negro (1.850°S, 61.616°W), 200 km northwest of Manaus. About 70% of the forested area is covered by terra-firme forest (Borges et al. 2001). There is considerable heterogeneity in local plant communities in the terra-firme forests due to soil mosaics in the region. This heterogeneity might also be related to human disturbance (Ferreira and Prance 1999). Approximately 12% is covered by igapó forests. Jaú also includes Novo Airão (2.620°S, 60.944°W), a site with campinas of low shrubby vegetation with only few trees reaching above 5 m in height. These open areas are close to the road AM 325 and are thus subject to relatively high anthropogenic influences. C) Reserva do Cuieras (CUI) and Reserva da Campina (CUI) – Reserva do Cuieras is a nature reserve covering 22.7 ha, and it is located about 70 km north of Manaus (2.609°S, 60.217°W). The vegetation is a mosaic of evergreen forest with a canopy height of about 35–40 m, with emergent trees over 45 m tall. Igapós and campinaranas cover 43% of the area, whereas terra-firme forests occupy 57% (Zanchi et al. 2011). Reserva da Campina is a nature reserve located 60 km north of Manaus (2.592°S, 60.030°W), on the right side of the Negro river. It covers approximately 900 ha, of which 6.5 ha is campinas and campinaranas. The campinas (2.6 ha) are mosaics of shrub islands surrounded by white bare sandy soil, with a canopy height of about 4–7 m. D) Caxiuanã (CXN) – a national forest of 371 000 ha of rainforest located 350 km west of Belém (1.735°S, 51.463°W) in the lower Amazon region of northern Brazil. About 85% of the area is covered by terra-firme forest and about 10% by várzea and igapó forests, but this reserve also has some campinas (Behling and Costa 2000).

Sampling design

We installed three temporary circular plots with a 28 m radius in each habitat type at each locality, totalling 39 plots (9 at BC, JAU, and CUI and 12 at CXN), following the soil sampling protocol of Tedersoo et al. (2014). Inside each plot,

we selected 20 trees at random and collected litter samples and two soil cores in opposite directions of each tree, summing to a total of 40 litter and 40 soil samples per plot. We pooled all samples to obtain one litter and one soil sample for each plot. Litter was defined as all organic material above the mineral soil and varied from 0 to ca 50 cm in thickness. We collected the soil samples from the top 5 cm of the mineral soil using a metal probe with a 2.5 cm diameter. We used gloves and masks and changed equipment in between each new plot to reduce the risk of cross-plot contamination. The samples were stored in sterilized white silica gel 1–4 mm (pre-treated by two minutes of microwave heating (800 W) and 15 min of UV light). All plots were provided with GPS coordinates. All dry samples were processed at the Univ. of Gothenburg, Sweden.

DNA extraction

For total DNA extraction, we used the PowerMax[®] Soil DNA Isolation Kit (MO BIO Laboratories, USA), according to the manufacturer's instructions. We use 10 g (dry weight) from all soil samples and 15 ml of the litter samples (corresponding to 3–10 g of dry weight litter, depending on texture and composition in that we standardized the samples by volume). Each DNA sample was concentrated and cleaned following the PowerMax[®] Soil DNA Isolation's instructions (MO BIO Laboratories, USA).

PCR amplification

16S: we targeted the V3-V4 region (~460 bases) of the 16S rRNA gene using the forward primer (5'-CCTACGGGN GGCWGCAG-3') and reverse primer (5'-GACTACH VGGGTATCTAATCC-3') from Klindworth et al. (2013). The amplification and sequencing was done by Macrogen (Republic of Korea) following standard protocols using the Illumina MiSeq 2×300 platform.

18S: we targeted the V7 region of the 18S rRNA gene using the forward and reverse primers (5'-TTTGTCTG STTAATTSCG-3') and (5'-TCACAGACCTGTATTGC-3') designed by Guardiola et al. (2015) to yield 100–110 bases long fragments. Amplification was performed in a total volume of 25 µl and consisted of: 0.25 µl of AmpliTaq1 Gold DNA polymerase, 5U µl⁻¹, 2.5 µl Pfu polymerase buffer 10×, 0.5 µl dNTP (final concentration of each dNTP 200 µmol; all above mentioned reagents are from Promega[®], Sweden), 0.25 µl of 50 mol of forward and reverse primers, 20.25 µl of nuclease free water, and 1 µl of DNA template. The PCR conditions were an initial denaturation step of 2 min at 95°C and then 30 cycles of denaturation at 95°C for 1 min, hybridization at 50°C for 45 s, and elongation at 72°C for 1 min, followed by a final elongation at 72°C for 10 min and finishing at 4°C. Each sample was amplified three times and pooled to reduce biases of amplification efficiency variation on different species and stochastic effects of amplification (Carew et al. 2013, Edgar 2013, Piñol et al. 2015). The quality of the amplification was checked in

UV light using GelRed[™] stain (1%; Biotium, USA) on a 2% agarose gel. All samples were purified using the QIAquick[®] PCR purification kit. Dual PCR amplifications were performed for Illumina MiSeq sequencing (Illumina, USA), using fusion primers as described in Bourlat et al. (2016). For indexing, we used the Nextera XT DNA index kit (Illumina, USA) according to the manufacturer's instructions. We checked the quality of the PCR products on a 2% agarose gel. We then made size selection using magnetic beads and a magnetic stand, using the ratio 0.9:1 beads/PCR product. We checked the DNA concentration in a Qubit 30[®] fluorimeter (Invitrogen, Sweden), and we assessed the quality and size selection of the PCR products with a 2200 Agilent TapeStation[®] (Agilent, USA). We normalized and pooled the PCR product (with the same concentration) following the Illumina protocol. The samples were sequenced at SciLifeLab (Stockholm, Sweden) using an Illumina MiSeq 2×250 machine.

Sequence analyses and taxonomic assessment

We used the USEARCH/UPARSE ver. 9.0.2132 Illumina paired reads pipeline (Edgar 2013) to quality filtering, de-replicate and sort reads by abundance, to infer OTUs, and to remove singletons. We filtered the sequences to discard chimeras and clustered sequences into OTUs at a minimum similarity of 97% using a 'greedy' algorithm that performs chimera filtering and OTU clustering simultaneously (Edgar 2013). We used SILVAngs 1.3 (Quast et al. 2012) for assessment of the taxonomic composition of the OTUs, using a representative sequence from each OTU as query sequence. We used the SINA ver. 1.2.10 reference data for ARB SVN (revision 21008, Pruesse et al. 2012) for both markers.

Statistical analyses

We performed all statistical analyses in R ver. 3.4.2 (R Development Core Team). We pooled the litter and soil OTUs from each plot for all analyses as our goal was to address general diversity patterns and not to compare the OTU composition from different substrates. We used the stringr ver. 1.0.0 (Wickham 2015) and tidyverse ver. 1.1.1 (Wickham 2017) packages for data curation and the ggplot2 ver. 2.2.1 (Wickham 2009), ggfortify ver. 0.1.0 (Tang and Li 2016), gplots ver. 3.0.1 (Warnes et al. 2015), gridExtra ver. 2.2.1 (Auguie and Antonov 2016), qgraph ver. 1.4.4 (Epskamp et al. 2012), and viridis ver. 0.4.0 (Garnier 2016) packages for data visualization. Scripts for all analyses are provided in Supplementary material Appendix 1.

We rarefied all samples to equal depth, where the depth was determined by the lowest number of reads obtained from a single plot (55 111 for 16S and 10 919 for 18S; Supplementary material Appendix 1 Fig A1). We subsequently transformed the OTU tables to presence/absence for both prokaryote (16S) and eukaryote (18S) data. Read abundances have been shown to be largely unreliable, especially for the 18S marker due to the large variance in biomass of the

study organisms (e.g. protozoa vs trees) (Carew et al. 2013, Deagle et al. 2013). Additionally, both the 16S and the 18S genes are multicopy genes, and hundreds of 18S copies per cell are known from some eukaryotes (Lindner et al. 2013).

To test hypothesis 1, we fitted a two-way ANOVA model with longitude and habitat as predictors and prokaryote (16S) and eukaryote (18S) OTU richness as response variables, respectively. We fitted a two-way ANOVA model with locality and habitat as predictors and performed a Tukey Honest significance test to evaluate the significance of among-group differences. We furthermore performed a general linear model considering habitat interaction with longitude and the effect on taxonomic groups ($\text{glmm} = \text{richness} \sim \text{habitat} \times \text{longitude} + (1/\text{taxa})$).

To test hypothesis 2, we performed a permutational multivariate analysis of variance (PERMANOVA) test using habitat and longitude as predictors and dissimilarity matrices using Jaccard index of prokaryote (16S) and eukaryote (18S) OTUs as response variables, respectively, using the ‘vegan’ package ver. 2.4-3 (Oksanen et al. 2007) in R. Additionally, we constructed two-dimensional non-metric multidimensional scaling (NMDS) ordinations of the presence/absence matrices of prokaryote (16S) and eukaryote (18S) data using the Jaccard dissimilarity index, as implemented in the metaMDS function in the vegan package to analyse community dissimilarity among all samples. We then used the ‘envfit’ method implemented in vegan to fit locality and environmental type onto the NMDS ordination as a measure of the correlation of these factors with the NMDS axes. We tested the isolating effect of distance with a Mantel test, and visualized the OTU community similarity among plots with the ‘qgraph’ function in R using a similarity index ($1/\text{Jaccard dissimilarity}$). Furthermore we performed variation partitioning on the prokaryote (16S) and eukaryote (18S) communities to investigate the compositional effects of changing habitat and locality. Variation partitioning resolves the contribution of habitat and locality to the total community variation (Legendre and Legendre 1998), but also resolves the variation shared between factors (i.e. shared between both habitat and locality). We used the ‘varpart’ function of the vegan package and assessed the significance for each section of the variation partitioning approach using redundancy analysis. We analysed the community matrix in models against habitat, locality, or both together as explanatory variables. We finally constructed Venn diagrams with the ‘gplots’ package in R to check the number of exclusive and shared OTU as a complement of community structure.

Data deposition

DNA sequences: BioSample accessions SAMN09081679-SAMN09081756; NCBI SRA: PRJNA464362, Bioproject 410658. Final DNA sequence assembly uploaded as supplementary material. Sampling locations, soil physical-chemical data, OTU tables, and R-scripts are available from the Dryad Digital Repository: <<http://dx.doi.org/10.5061/dryad.4hv2b38>> (Ritter et al. 2018).

Results

We obtained a total of 2 984 233 reads and 6625 OTUs for prokaryotes (16S) and 9 149 502 reads and 15 840 OTUs for eukaryotes (18S). See Supplementary material Appendix 1 Table A1 for the number of sequences and OTUs for each plot (see also Supplementary material Appendix 1 Fig. A2 for rarefaction OTU number by plot) after rarefaction. The correlation between prokaryote (16S) and eukaryote (18S) OTU richness was weak overall and absent if considering each locality separately, with just the CUI location correlating significantly (Supplementary material Appendix 1 Fig. A5 and A4).

Taxonomic composition

The taxonomic composition of the prokaryote component shows that the groups with the highest number of OTUs were Proteobacteria (22% of the taxa identified in our samples, equivalent to about 1000 OTUs per habitat and locality, most of which belonged to Alphaproteobacteria: Fig. 2A and 2B) and Chloroflexi (15%, average ~600 OTUs; Fig. 2A and 2B). For eukaryotes, the group with the highest number of OTUs was Fungi (30%, ~2400 OTUs, mainly Ascomycota and Basidiomycota: Fig. 2C and 2D) followed by Cercozoa (15%, ~1100 OTUs: Fig. 2C and 2D) and Alveolata (10%, ~750 OTUs; Fig. 2C and 2D). Most of the eukaryotic OTUs for all taxa were relatively ‘range-restricted’, occurring in less than five plots. In contrast, prokaryote OTUs were generally widespread across many plots (Supplementary material Appendix 1 Fig. A6 and A7). Due to the small replicate number (39 plots) in relation to the number of taxonomic groups, we did not explicitly test the patterns of richness or the community composition for habitat types or locality for each taxonomic group on the OTU level. However, based on a qualitative (visual) inspection of the results, we did not observe any difference in taxonomic composition by locality and habitat for neither prokaryotes (16S) nor eukaryotes (18S). The general linear model to assess the effect of habitat and longitude by taxonomic groups was non-significant ($p > 0.05$ for both prokaryotes and eukaryotes).

Hypothesis 1 – OTU richness by locality and habitat

We found a significant effect of longitude on OTU richness for prokaryotes ($F_{[1,34]} = 5.36$, $p < 0.05$). The highest mean OTU richness for prokaryotes (16S) was found in BC (2076 OTUs), followed by CXN (1756) and JAU (1587), and the lowest was found in CUI (1462). The analysis of variance showed a significant effect of locality ($F_{[3,32]} = 26.63$, $p < 0.001$). The Tukey HSD test showed a significant pairwise difference between BC and all other locations, as well as between CXN and CUI. Concerning habitat type, the mean OTU number was the highest in várzeas (1948) followed by campinas (1827), terra-firmes (1686), and finally igapós (1570). The analysis of variance showed a significant effect

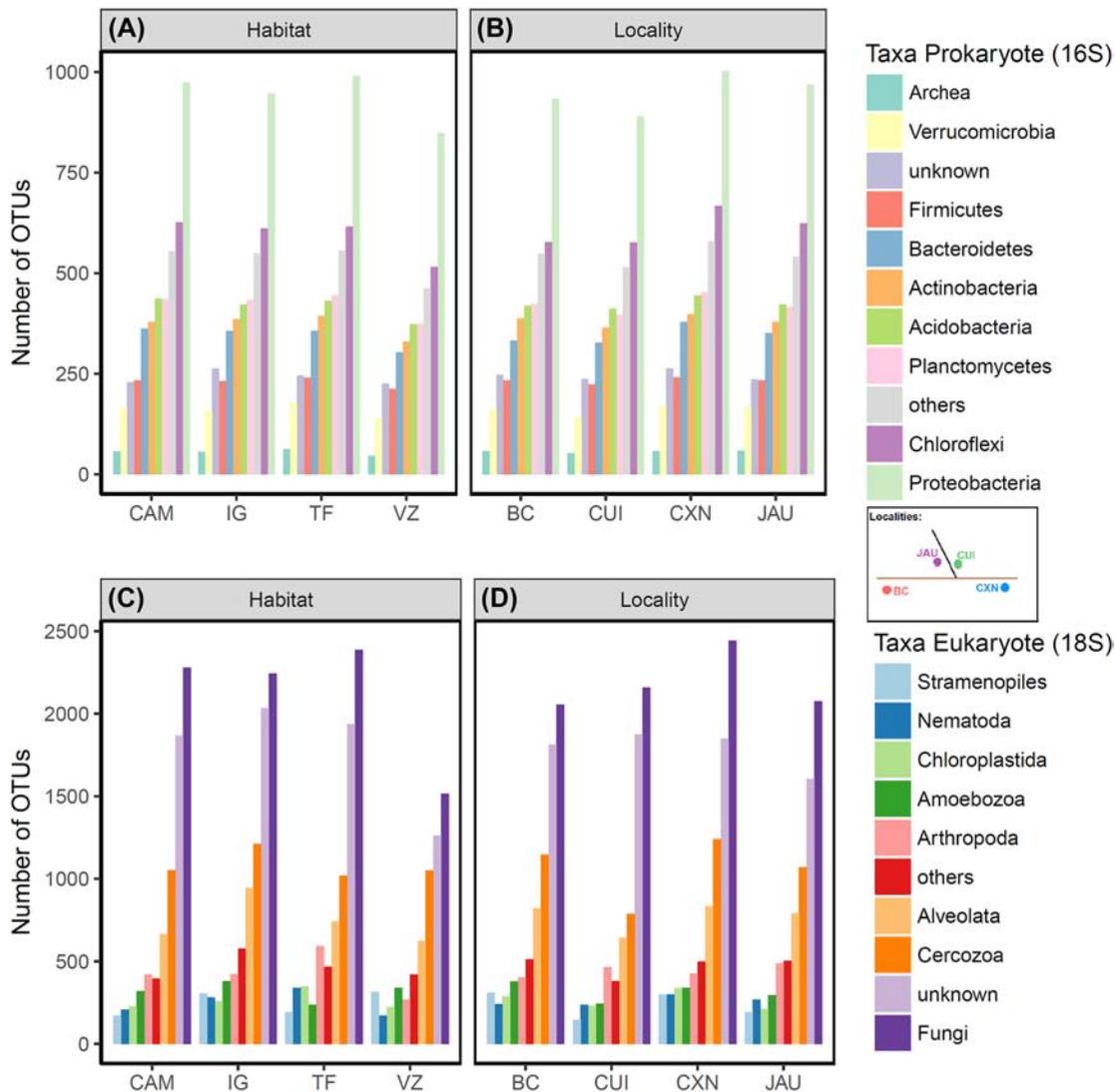


Figure 2. Taxonomic composition of OTU communities. The total number of OTUs by taxonomic groups for (A) habitat and (B) locality for prokaryotes; and (C) habitat and (D) locality for eukaryotes. Localities are BC: Benjamin Constant; JAU: Jaú; CUI: Cuieras and; CXN: Caxiuanã. Habitats are CAM = campinas; IG = igapós; TF = terra-firmes; and VZ = várzeas. Inset: schematic position of each locality, with the brown line representing the Amazon River and the black line the Negro River. There is no clear taxonomic variation among groups by locality or habitat type, in either the prokaryote or the eukaryote data.

of habitat type ($F_{[3,32]} = 10.87$, $p < 0.001$) on the prokaryote (16S) OTU richness (Fig. 3A). The Tukey HSD test showed a significant difference in OTU richness between campinas and both igapós and terra-firme.

We found the mean OTU richness per locality for eukaryotes (18S) to be comparable with that of the prokaryotes (Fig. 3B). The largest OTU richness was found in BC (2051) followed by CXN (1808), JAU (1726), and CUI (1625). We found no significant effect of either longitude ($F_{[1,34]} = 1.67$, $p = 0.2$) or locality ($F_{[3,32]} = 2.6$, $p = 0.07$). The habitat type with the highest mean number of OTUs was campinas (1907), followed by várzeas (1836), terra-firmes (1783), and finally igapós (1728). We found no significant effect of habitat type ($F_{[3,32]} = 1.47$, $p = 0.24$).

Hypothesis 2 – OTU community structure by location and habitat

The PERMANOVA results showed a significant effect for habitat (16S [$R^2 = 0.35$, $p < 0.001$]; 18S [$R^2 = 0.12$, $p < 0.001$]) and longitude (18S [$R^2 = 0.22$, $p < 0.005$]; 16S [$R^2 = 0.07$, $p < 0.001$]) for both the prokaryote (16S) and the eukaryote (18S) communities. Variation partitioning also identified significant proportions of both prokaryote (16S) and eukaryote (18S) communities varying with habitat and locality, but with no shared variation between them. About 50% of the full prokaryote (16S) community variation was explained by the analysis, with 29% contributed by habitat and 21% by locality. A lower percentage of the total

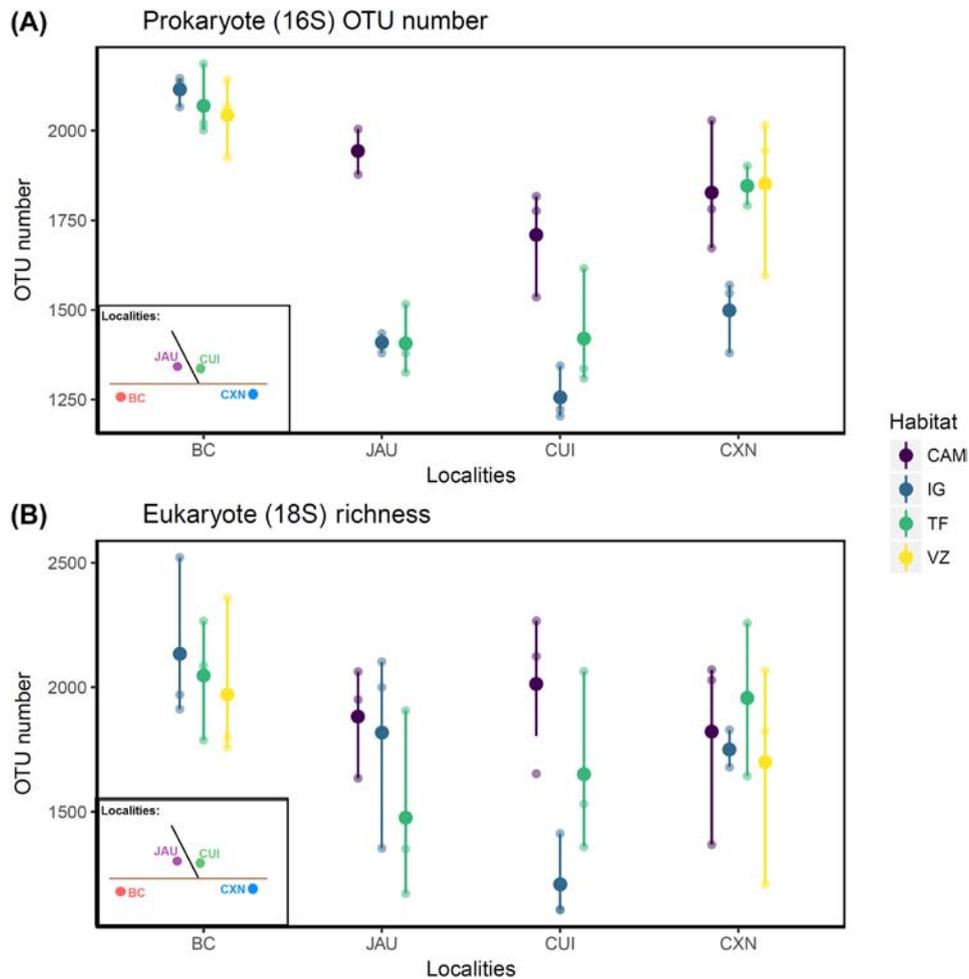


Figure 3. OTU richness per locality and habitat type. Colours indicate habitat type (CAM: campinas, IG: igapó, TF: terra firme, and VZ: várzea). Locations are ordered from west to east (BC: Benjamin Constant; JAU: Jaú; CUI: Cuieras; and CXN: Caxiuanã). Large circles are the median and small circles are the number of OTUs per plot. (A) prokaryotes (16S); (B) eukaryotes (18S). Note the difference in the range of values on the y-axis. Inset: schematic position of each locality, with the brown line representing the Amazon River and the black line the Negro River. These results indicate that OTU richness varies significantly with location and habitat type in 16S, with higher richness in BC and in campinas. This effect is not observed for the 18S data.

community variation was explained in the eukaryote (18S) analysis (27%), with habitat still explaining a larger proportion of the community variation (16%) than did locality (11%).

The similarity network analysis (Fig. 4) reveals a stronger (more aggregated) community structure for prokaryotes (Fig. 4A coloured by locality and Fig. 4B coloured by habitat type) than for eukaryotes (Fig. 4C coloured by locality and Fig. 4D coloured by habitat type). Plots with highest similarity occurred inside the same locality and habitat. In both prokaryote and eukaryote communities, BC clustered tightly together and distinctly from the other localities. The results of the NMDS show similarity in community composition among the plots and the influence of locality and environmental type. The envfit test indicated significant effects of locality on both the prokaryote ($R^2=0.38$; $p < 0.001$) and eukaryote ($R^2=0.33$; $p < 0.001$) communities. The envfit test also indicated a significant effect for habitat type on the

prokaryote ($R^2=0.54$; $p < 0.001$) and eukaryote ($R^2=0.54$; $p < 0.001$) communities.

The Venn diagrams show the number of unique and shared OTUs for each locality (Fig. 5A for 16S and Fig. 5C for 18S). The BC location had the highest number of unique OTUs (16S=591; 18S=2595), followed by CXN (16S=473; 18S=2239), JAU (16S=389; 18S=1562), and CUI (16S=376; 18S=1485) for both eukaryotes (18S) and prokaryotes (16S). For both these groups, the localities with the highest number of shared OTUs were BC and CXN (16S=335; 18S=786) followed by JAU and CUI (16S=201; 18S=654). The Venn diagrams also show the number of unique and shared OTUs for each habitat (Fig. 5B for 16S and Fig. 5D for 18S). The lowest number of unique OTUs was found in várzeas (16S=218; 18S=1164) followed by igapós (16S=308; 18S=2181). The habitat with the highest number of unique OTUs was campinas for prokaryotes (805) and terra-firme for eukaryotes

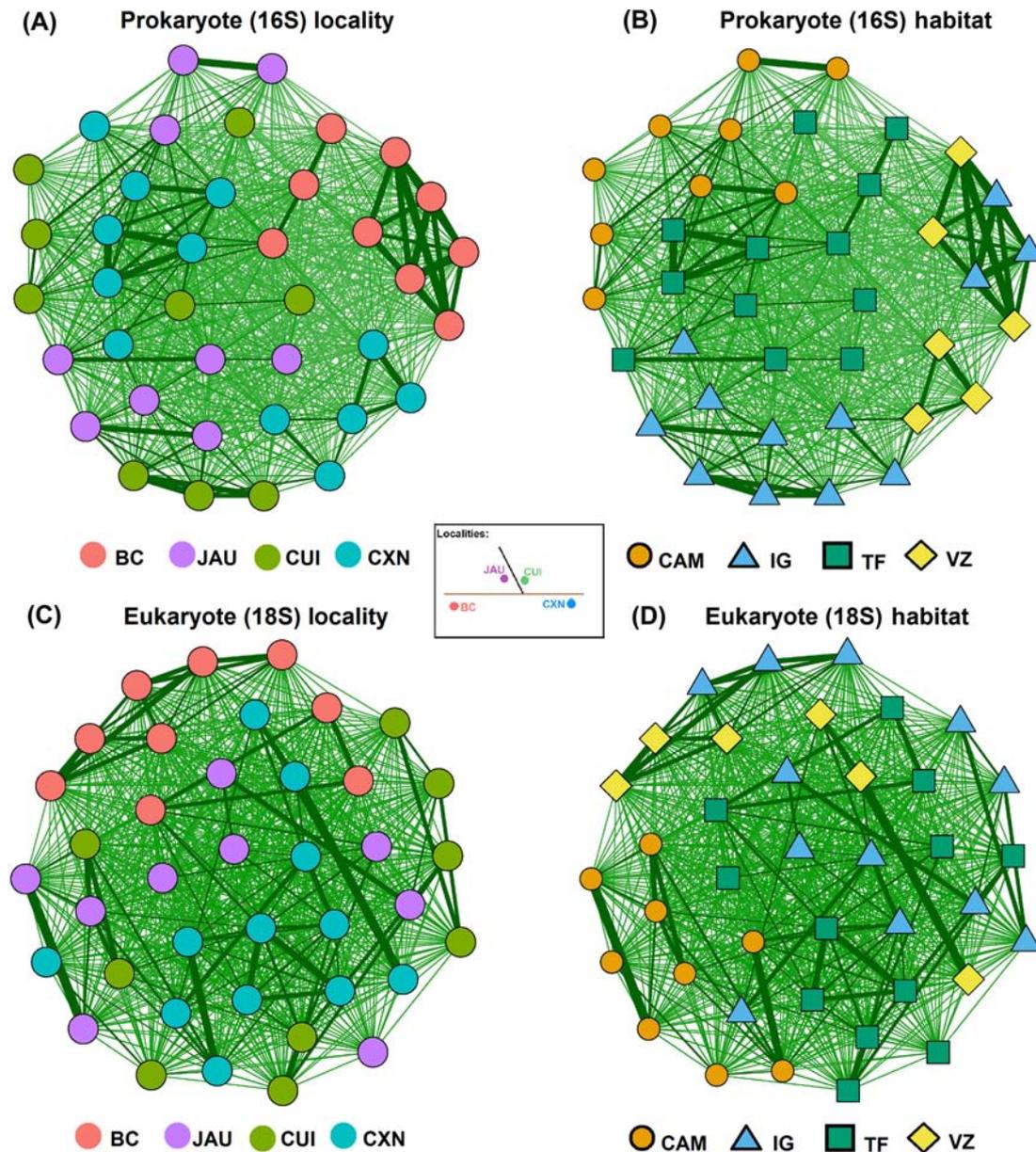


Figure 4. OTU community similarity network. The sizes of the green lines are proportional to the similarity index. The nodes are coloured by (A) localities for prokaryotes (16S); and (C) localities for eukaryotes (18S). The formats of nodes show different habitat types in (B) prokaryotes (16S); and (D) eukaryotes (18S); orange circles = campinas, blue triangles = igapós, green squares = terra-firme forest, and yellow lozenges = várzeas. Inset: schematic position of each locality, with the brown line representing the Amazon River and the black line the Negro River. The prokaryotes (16S) are more structured by locality and habitat than are the eukaryotes (18S).

(2378). The várzeas and igapós shared the largest number of OTUs (16S = 512; 18S = 1021) followed by terra-firmes and campinas (16S = 380; 18S = 785).

Discussion

Here we provide a first mapping of Amazonian biodiversity that considers not only macroscopic organisms but also the microscopic, microbial component across a large geographic scale. We found that prokaryote (16S) and eukaryote (18S)

OTU richness and community composition differ significantly among localities and habitats, with habitat type being a stronger predictor of diversity and community composition than locality.

Contrasting prokaryote and eukaryote diversity

The weak correlation found between prokaryotes (16S) and eukaryotes (18S) OTUs indicates that richness patterns may be different between the two groups. Previous reports found that localities with high bacterial diversity can have

(Wittmann et al. 2010), with white-water basins being generally more diverse than black-water basins. Furthermore, BC and CXN are the only localities with várzea, which is the environment with the highest OTU richness. All these factors are likely responsible for the higher OTU richness and the composition similarity between BC and CXN. BC and CXN are also the localities with the highest number of unique OTUs, which could emphasize their importance under conservation strategies. In contrast, both JAU and CUI are bathed by acidic, sediment-poor rivers, which is characteristic of both the Negro (JAU) and Cuieras (CUI) rivers.

We found different richness patterns across the environmental types surveyed. Based on previously documented patterns for macro-organisms, we expected richness to decrease in the order terra-firme > várzea > igapó > campina. Therefore, it was surprising to observe várzea > campina > terra-firme > igapó for prokaryotes and campina > várzea > terra-firme > igapó for our eukaryote dataset. Várzeas are considered to be a stressful environment for many organisms, with long periods of flooding, but they have fertile soils that could explain the more substantial OTU richness. Our finding that campinas was the richest habitat for eukaryotes and the second richest for prokaryotes is puzzling. Campinas are nutrient-poor (Prance 1996, Fine et al. 2005), with scleromorphic physiognomy (Anderson 1981) and are relatively low in diversity of macro-organisms (Wüster et al. 2005, Smith et al. 2012). The third place in richness was held by terra-firme. We found this to be at least as unexpected, given that terra-firmes are by far the most predominant vegetation type in the Amazon basin since the beginning of the Miocene, and contains a very high macro-organismal diversity (Jaramillo et al. 2006, Irion and Kalliola 2010). However, such high documented richness could in part be an effect of the disproportionately larger area of terra-firmes as compared to the other habitat types – an effect that should not be evident in our sampling design. Interestingly, the várzeas and igapós are more similar to each other in terms of community composition despite their differences in OTU richness, which could be related to similar environmental filters linked to stress by flooding, potentially favouring a shared set of specialized organisms. The community similarity between campinas and terra-firmes might be linked to the fact that campinas often are small ‘islands’ within large ‘seas’ of terra-firme forest, and in this way these campinas may receive DNA from the surrounding forests (from, e.g. leaves, insects, and fungal spores) in addition to their specific ‘specialized’ OTU community.

Most eukaryotic (18S) OTUs were restricted to few plots (Supplementary material Appendix 1 Fig. A8), with the majority restricted to one habitat or one locality. We found the number of unique (site-specific) prokaryote OTUs to decrease in the sequence campinas > terra-firmes > igapós > várzeas, whereas for eukaryotes the pattern was terra-firmes > campinas > igapós > várzeas, these patterns differs markedly from the results of our richness analyses. If we were to consider only richness as a conservation priority, várzeas would be the most important habitat type; but if we were to

preserve the most unique environment instead, várzeas would have been the least important habitat. Although these results should be viewed with some caution, not least for the tiny sample they represent out of the enormous region covered, they showcase the difficulties in prioritizing conservation areas based on single metrics (Orme et al. 2005).

Taxonomic composition

The taxonomic resolution of metabarcoding data is limited by the availability of comprehensive reference databases (Coward et al. 2015). Such databases are generally meagre with respect to Amazonian biodiversity, even for well-studied (e.g. trees; Balmford and Whitten 2003, Giam et al. 2012). This reduces our ability to identify many of the OTUs to resolved taxonomic levels, in particular those from less studied group of organisms (e.g. platyhelminthes and Alveolata). A further complication with metabarcoding is the compromise between taxonomic coverage and taxonomic resolution. While the universal 18S primers can capture the majority of eukaryotic organisms, this gene is not variable enough to distinguish all eukaryotes at the species level (Hartmann et al. 2010, Lindahl et al. 2013). In addition, since most amplicon studies target only a part of the 18S (e.g. ~110 bases, as in this study) rather than its full length (~2000 bases), the available information and resolution are further decreased. This means that for most plants, family-level designations are usually the most resolved level of taxonomic composition possible using 18S fragments. For many insects and fungi, the precision may be at the order level (Lindahl et al. 2013, Liu et al. 2013).

The lower the taxonomic level considered, the more idiosyncratic will its distribution generally be, adding heterogeneity to general diversity patterns. For instance, Bass et al. (2010) found a west-to-east diversity gradient for Amazonian mammals, but this gradient was not observed by Maestri and Patterson (2016) using a lower taxonomic level (rodents). Looking at individual OTUs in our data, we found striking differences in richness and community composition across sites. When OTUs were collapsed into high taxonomic groups, such as orders and phyla, these differences were less pronounced (Supplementary material Appendix 1 Fig. A8 and A9).

Conclusions

The broad taxonomic coverage of 16S and 18S, together with the standardised sampling, extraction, and sequencing approaches applied here, allowed us to directly compare biodiversity patterns across a large spatial range in Amazonia. We stress the importance of considering the ‘hidden biodiversity’ (microscopic, subterranean, or otherwise inconspicuous species) to characterize a larger proportion of the total diversity patterns. We detected a different habitat gradient from what we expected initially, but as expected we found a longitudinal gradient for OTU richness (hypothesis 1) and community composition (hypothesis 2). Furthermore, we found habitat to be the strongest predictive factor of biodiversity, a pattern

that was particularly strong for the prokaryote communities. Our results show that the currently accepted diversity patterns in Amazonia do not hold for all organisms, which suggests that biodiversity patterns of different groups of organisms may be largely decoupled. We also found different patterns between richness and uniqueness of OTUs across sites and environmental types, showing the pitfalls in choosing single biodiversity metrics for prioritizing conservation areas.

Acknowledgements – We thank Anna Ansebo, Sven Toresson, and Ylva Heed for laboratory and administrative assistance; Mats Töpel for help with bioinformatics; Quiterie Haenel for help in the extractions and discussion on the begin of project; Hans ter Steege for advice on sampling localities and experimental design; and members of our research group for discussions and suggestions. The authors also acknowledge support from Science for Life Laboratory, the Knut and Alice Wallenberg Foundation, the National Genomics Infrastructure funded by the Swedish Research Council, and Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure. Additional computational analyses were run at the Univ. of Gothenburg bioinformatics cluster (<<http://albiorix.bioenv.gu.se/>>). We also thank Leho Tedersoo, Micah Dunthorn, and a third anonymous review for valuable suggestion on an earlier draft of this manuscript.

Funding – Funding for this study was provided by from CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico – Brazil: 249064/2013-8) and the Swedish Research Council (B0569601). AA is further supported by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013, ERC Grant Agreement no. 331024), the Swedish Foundation for Strategic Research, a Wallenberg Academy Fellowship, the Faculty of Sciences at the Univ. of Gothenburg, the Wenner-Gren Foundation, and the David Rockefeller Center for Latin American Studies at Harvard Univ.

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

Conflicts of interest – The authors declare have no conflicts of interest.

Permits – The permits were provided by Brazilian authorities, ICMBio (register number 48185-2) and IBAMA (register number 127341).

References

Anderson, A. B. 1981. White-sand vegetation of Brazilian Amazonia. – *Biotropica* 13: 199–210.

Antonelli, A. et al. 2009. Tracing the impact of the Andean uplift on Neotropical plant evolution. – *Proc. Natl Acad. Sci. USA* 106: 9749–9754.

Auguie, B. and Antonov, A. 2016. Package ‘gridExtra’. – <<https://cran.r-project.org/web/packages/gridExtra/>> accessed 15 December 2017.

Balmford, A. and Whitten, T. 2003. Who should pay for tropical conservation, and how could the costs be met? – *Oryx* 37: 238–250.

Barthlott, W. et al. 1999. Terminological and methodological aspects of the mapping and analysis of the global biodiversity. – *Acta Bot. Fenn.* 162: 103–110.

Bass, M. S. et al. 2010. Global conservation significance of Ecuador's Yasuní National Park. – *PLoS One* 5: e8767.

Bates, J. M. 2001. Avian diversification in Amazonia: evidence for historical complexity and a vicariance model for a basic diversification pattern. – In: Viera, I. et al. (eds), *Diversidade Biológica e Cultural da Amazônia*. Museu Paraense Emílio Goeldi, Belém, Pará, Brazil, pp. 119–137.

Bates, S. T. et al. 2013. Global biogeography of highly diverse protistan communities in soil. – *ISME J.* 7: 652–659.

Behling, H. and da Costa, M. L. 2000. Holocene environmental changes from the Rio Curuá record in the Caxiuana region, eastern Amazon Basin. – *Quat. Res.* 53: 369–377.

Biggs, J. et al. 2015. Using eDNA to develop a national citizen science-based monitoring programme for the great crested newt (*Triturus cristatus*). – *Biol. Conserv.* 183: 19–28.

Blaxter, M. et al. 2005. Defining operational taxonomic units using DNA barcode data. – *Phil. Trans. R. Soc. B* 360: 1935–1943.

Borges, S. H. et al. 2001. Birds of Jaú National Park, Brazilian Amazon: species check-list, biogeography and conservation. – *Ornitol. Neotrop.* 12: 109–140.

Bourlat, S. J. et al. 2016. Preparation of amplicon libraries for metabarcoding of marine eukaryotes using Illumina MiSeq: the dual-PCR method. – *Methods Mol. Biol.* 1452: 197–207.

Carew, M. E. et al. 2013. Environmental monitoring using next generation sequencing: rapid identification of macroinvertebrate bioindicator species. – *Front. Zool.* 10: 45.

Cowart, D. A. et al. 2015. Metabarcoding is powerful yet still blind: a comparative analysis of morphological and molecular surveys of seagrass communities. – *PLoS One* 10: e0117562.

Deagle, B. E. et al. 2013. Quantifying sequence proportions in a DNA-based diet study using Ion Torrent amplicon sequencing: which counts count? – *Mol. Ecol. Resour.* 13: 620–633.

Dominati, E. et al. 2010. A framework for classifying and quantifying the natural capital and ecosystem services of soils. – *Ecol. Econ.* 69: 1858–1868.

Edgar, R. C. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. – *Nat. Methods* 10: 996–998.

Epskamp, S. et al. 2012. qgraph: network visualizations of relationships in psychometric data. – *J. Stat. Softw.* 48: 1–18.

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.

Ferreira, L. V. and Prance, G. T. 1999. Ecosystem recovery in terra firme forests after cutting and burning: a comparison on species richness, floristic composition and forest structure in the Jaú National Park, Amazonia. – *Bot. J. Linn. Soc.* 130: 97–110.

Ferreira, L. V. et al. 2005. Riqueza e composição de espécies da floresta de igapó e várzea da Estação Científica Ferreira Penna: subsídios para o plano de manejo da Floresta Nacional de Caxiuana. – *Pesquisas Bot.* 46: 103–116.

Fine, P. V. et al. 2005. The contribution of edaphic heterogeneity to the evolution and diversity of Burseraceae trees in the western Amazon. – *Evolution* 59: 1464–1478.

Garnier, S. 2016. viridis: default color maps from ‘matplotlib’. – <<https://cran.r-project.org/web/packages/viridis/>> accessed 15 December 2017.

Giam, X. et al. 2012. Reservoirs of richness: least disturbed tropical forests are centres of undescribed species diversity. – *Proc. R. Soc. B* 279: 67–76.

- Gibson, J. et al. 2014. Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasytematics. – *Proc. Natl Acad. Sci. USA* 111: 8007–8012.
- Guardiola, M. et al. 2015. Deep-sea, deep-sequencing: metabarcoding extracellular DNA from sediments of marine canyons. – *PLoS One* 10: e0139633.
- Hansen, M. C. et al. 2013. High-resolution global maps of 21st-century forest cover change. – *Science* 342: 850–853.
- Hartmann, M. et al. 2010. V-Extractor: an open-source, high-throughput software tool to identify and extract hypervariable regions of small subunit (16S/18S) ribosomal RNA gene sequences. – *J. Microbiol. Methods* 83: 250–253.
- Hoorn, C. et al. 2010. Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. – *Science* 330: 927–931.
- Irion, G. and Kalliola, R. 2010. Long-term landscape development processes in Amazonia. – In: Hoorn, C. and Wesselingh, F. P. (eds), *Amazonia: landscape and species evolution a look into the past*. Blackwell Publishing, pp. 203–215.
- Jaramillo, C. et al. 2006. Cenozoic plant diversity in the Neotropics. – *Science* 311: 1893–1896.
- Junk, W. J. et al. 1989. The flood pulse concept in river-floodplain systems. – *Can. Spec. Publ. Fish. Aquat. Sci.* 106: 110–127.
- Junk, W. J. et al. 2011. A classification of major naturally-occurring Amazonian lowland wetlands. – *Wetlands* 31: 623–640.
- Klindworth, A. et al. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. – *Nucl. Acids Res.* 41: e1.
- Legendre, P. and Legendre, L. 1998. *Numerical ecology*, 2nd English ed. – Elsevier Science.
- Lentendu, G. et al. 2017. Persistent patterns of high alpha and low beta diversity in tropical parasitic and free-living protists. – *bioRxiv* 166892.
- Lindahl, B. D. et al. 2013. Fungal community analysis by high-throughput sequencing of amplified markers – a user’s guide. – *New Phytol.* 199: 288–299.
- Lindner, D. L. et al. 2013. Employing 454 amplicon pyrosequencing to reveal intragenomic divergence in the internal transcribed spacer rDNA region in fungi. – *Ecol. Evol.* 3: 1751–1764.
- Liu, S. et al. 2013. SOAPBarcode: revealing arthropod biodiversity through assembly of Illumina shotgun sequences of PCR amplicons. – *Methods Ecol. Evol.* 4: 1142–1150.
- Lovejoy, N. R. et al. 2006. Miocene marine incursions and marine/freshwater transitions: evidence from Neotropical fishes. – *J. South Am. Earth Sci.* 21: 5–13.
- Maestri, R. and Patterson, B. D. 2016. Patterns of species richness and turnover for the South American rodent fauna. – *PLoS One* 11: e0151895.
- Mahé, F. et al. 2017. Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. – *Nat. Ecol. Evol.* 1: 0091.
- Mendes, L. W. et al. 2015. Soil-borne microbiome: linking diversity to function. – *Microb. Ecol.* 70: 255–265.
- Mora, C. et al. 2011. How many species are there on Earth and in the ocean? – *PLoS Biol.* 9: e1001127.
- Moran, E. F. et al. 2000. Effects of soil fertility and land-use on forest succession in Amazonia. – *For. Ecol. Manage.* 139: 93–108.
- Oksanen, J. et al. 2007. The vegan package. – *Community Ecology Package* 10: 631–637.
- Orme, C. D. L. et al. 2005. Global hotspots of species richness are not congruent with endemism or threat. – *Nature* 436: 1016–1019.
- Peay, K. G. et al. 2013. Strong coupling of plant and fungal community structure across western Amazonian rainforests. – *ISME J.* 7: 1852–1861.
- Peres, C. A. 1997. Primate community structure at twenty western Amazonian flooded and unflooded forests. – *J. Trop. Ecol.* 13: 381–405.
- Piñol, J. et al. 2015. Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. – *Mol. Ecol. Resour.* 15: 819–830.
- Prance, G. T. 1996. Islands in Amazonia. – *Phil. Trans. R. Soc. B* 351: 823–833.
- Pruesse, E. et al. 2012. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. – *Bioinformatics* 28: 1823–1829.
- QGIS Development Team 2012. QGIS Geographic Information System. – Open Source Geospatial Foundation Project.
- Quast, C. et al. 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. – *Nucl. Acids Res.* 41: D590–D596.
- Ritter, C. D. et al. 2018. Data from: Locality or habitat? Exploring predictors of biodiversity in Amazonia. – *Dryad Digital Repository*, <<http://dx.doi.org/10.5061/dryad.4hv2b38>>.
- Smith, B. T. et al. 2012. Evaluating the role of contracting and expanding rainforest in initiating cycles of speciation across the Isthmus of Panama. – *Proc. R. Soc. B* 279: 3520–3526.
- Stahlhut, J. K. et al. 2013. DNA barcoding reveals diversity of Hymenoptera and the dominance of parasitoids in a sub-arctic environment. – *BMC Ecol.* 13: 2.
- Taberlet, P. et al. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. – *Mol. Ecol.* 21: 2045–2050.
- Tang, Y. and Li, W. 2016. lfd: an R package for local fisher discriminant analysis and visualization. – *arXiv preprint arXiv:1612.09219*.
- Tedersoo, L. et al. 2014. Global diversity and geography of soil fungi. – *Science* 346: 1256688.
- ter Steege, H. et al. 2003. A spatial model of tree α -diversity and tree density for the Amazon. – *Biodivers. Conserv.* 12: 2255–2277.
- Tuomisto, H. et al. 2017. Effect of sampling grain on patterns of species richness and turnover in Amazonian forests. – *Ecography* 40: 840–852.
- Warnes, G. R. et al. 2015. gplots: various R programming tools for plotting data. – R package ver. 2.17.0, <<http://CRAN.R-project.org/package=gplots>>.
- Wickham, H. 2009. ggplot2: elegant graphics for data analysis. – Springer.
- Wickham, H. 2015. stringr: simple, consistent wrappers for common string operations. – R package ver. 1.0.0, <<http://stringr.tidyverse.org>>, accessed 15 December 2017.
- Wickham, H. 2017. tidyverse. – R package ver. 1.1.1, <www.tidyverse.org/>, accessed 15 December 2017.
- Wittmann, F. et al. 2010. Phytogeography, species diversity, community structure and dynamics of central Amazonian floodplain forests. – In: Junk, W. J. et al. (eds), *Amazonian floodplain forests*. Springer, pp. 61–102.
- Wüster, W. et al. 2005. Tracing an invasion: landbridges, refugia, and the phylogeography of the Neotropical rattlesnake

- (Serpentes: Viperidae: *Crotalus durissus*). – Mol. Ecol. 14: 1095–1108.
- Zanchi, F. B. et al. 2011. Influence of drainage status on soil and water chemistry, litter decomposition and soil respiration in central Amazonian forests on sandy soils (Influência da drenagem na química do solo e da água, na decomposição da serapilheira e na respiração do solo em floresta de solo arenoso na Amazônia central). – Rev. Ambiente Água 6: 6.
- Zizka, A. et al. 2018. Finding needles in the haystack: where to look for rare species in the American tropics. – Ecography 41 doi:10.1111/ecog.02192

Supplementary material (Appendix ECOG-03833 at <www.ecography.org/appendix/ecog-03833>). Appendix 1.