Body size determines soil community assembly in a tropical forest

Authors
Lucie Zinger*, Pierre Taberlet², Heidy Schimann³, Aurélie Bonin², Frédéric Boyer², Marta De Barba², Philippe Gaucher⁴, Ludovic Gielly², Charline Giguët-Covex²‡, Amaia Iribar¹, Maxime Réjou-Méchain¹, Gilles Rayé², Delphine Rioux², Vincent Schilling¹, Blaise Tymen¹, Jérôme Viers⁶, Cyril Zouiten⁶, Wilfried Thuiller², Eric Coissac², Jérôme Chave¹.

Affiliations
¹ Université Toulouse 3 Paul Sabatier, CNRS, IRD, UMR 5174 Evolution et Diversité Biologique (EDB), F-31062 Toulouse, France.
² Univ. Grenoble Alpes, CNRS, Laboratoire d’Ecologie Alpine (LECA), F-38000 Grenoble, France.
³ INRA, UMR Ecologie des Forets de Guyane (AgroParisTech, CIRAD, CNRS, Université des Antilles, Université de la Guyane), 97310 Kourou, France.
ABSTRACT

Tropical forests shelter an unparalleled biological diversity. The relative influence of environmental selection (i.e. abiotic conditions, biotic interactions) and stochastic-distance dependent neutral processes (i.e. demography, dispersal) in shaping the composition of tropical soil communities is largely understudied. Here, we provide a comprehensive assessment of the assembly processes underlying the diversity of soil macrofauna in French Guiana. We apply eDNA metabarcoding to explore the relative contribution of environmental selection and stochastic processes on community composition across a gradient of forest disturbance. Our results indicate that environmental selection plays a major role in shaping soil communities, while stochastic processes have only a minor impact on community structure. However, the balance between these two processes varies depending on the environmental gradients considered, highlighting the importance of considering multiple environmental factors to understand the assembly of tropical soil communities.
This article is protected by copyright. All rights reserved.
that their influence depends on the spatial scale of observation (Levin, 1992; Leibold et al., 2004; Ricklefs, 2004). For example, the influence of environmental selection has been shown to increase as the observational scale increases, due to the inclusion of greater environmental heterogeneity (De Cáceres et al., 2012; Barton et al., 2013; Chase, 2014; Dini-Andreote, Stegen, van Elsas, & Salles, 2015). However, much less is known on how community assembly processes vary across organisms, and especially when it comes to the whole soil biota.

Soils are structurally complex environments (Hinsinger, Bengough, Vetterlein, & Young, 2009) and harbour a large and elusive diversity of organisms (Bardgett & van der Putten, 2014). Amplicon-based DNA analysis of environmental samples, or metabarcoding (Taberlet, Coissac, Pompanon, Brochmann, & Willerslev, 2012) has recently enabled unravelling novel macroecological patterns for soil fauna, nematodes, bacteria, fungi, protists, and archaea across biomes and habitats (Fierer, Strickland, Liptzin, Bradford, & Cleveland, 2009; Wu, Ayres, Bardgett, Wall, & Garey, 2011; Bardgett & van der Putten, 2014; Tedersoo et al., 2014; Mahé et al., 2017). These patterns often co-vary with environmental conditions such as pH and nutrient quality/availability. They also depend on plant cover due to trophic and mutualistic/pathogenic interactions (Lavelle et al., 1997; Wardle, 2002; Dini-Andreote & van Elsas, 2013), but this relationship appears context-dependent (Peay, Baraloto, & Fine, 2013; Barberán et al., 2015; Tedersoo et al., 2016). More generally, soil communities are often reported to be driven by environmental selection (Hanson, Fuhrman, Horner-Devine, & Martiny, 2012; Bardgett & van der Putten, 2014), although certain studies also unveiled dispersal limitation and/or ecological drift in fungi, mesofauna, and bacteria to a lesser extent.
Much of the difficulty in drawing general conclusions about soil community assembly lies in that empirical studies often focus on single taxonomic groups (but see Ramirez et al., 2014; Powell et al., 2015; Schuldt et al., 2015; Tedersoo et al., 2016). Soil organisms exhibit large differences in life-history traits; in particular their body size, which spans six orders of magnitude (0.1 µm to 10 cm; Bardgett & van der Putten 2014). Organisms of different body size have contrasting metabolisms, ontogenies, population sizes, and diversities (Brown Gillooly, Allen, Savage, & West, 2004; Woodward et al., 2005). Body size is also indicative of organism trophic status in its broadest sense (Brose et al., 2012; Briones, 2014), and is a key trait in the modelling of food webs (Williams & Martinez, 2000; Petchey, Beckerman, Riede, & Warren, 2008). Finally, body size also influences the way organisms perceive – or move across – space (Hubbell, 2001; Ettema & Wardle, 2002; Berg, 2012). This suggests that community assembly processes not only depend on the spatial scale of the study, but also on the size of the focal organisms.

For example, dispersal of microorganisms is mediated by external agents and is assumed to occur across large spatial distances in most cases. This, together with their large population size and short generation time, contributes to creating a homogeneous regional microbial species pool. Local microbial communities are then selected from the regional species pool by local environmental conditions (Hanson, Fuhrman, Horner-
Devine, & Martiny, 2012; Barberán, Casamayor, & Fierer, 2014), although this selection can be offset by high immigration rates, a process previously coined “mass effect” or “homogenizing dispersal” (Leibold et al., 2004; Stegen et al., 2013). In contrast, larger organisms, such as trees or mesofauna, are thought to be limited in their ability to disperse, have smaller population sizes, and longer generation times. They hence tend to lag behind their optimal habitat conditions and are more prone to local extinction. In addition, they may not perceive small-scale environmental variations as compared to microorganisms (Ettema & Wardle, 2002; Woodward et al., 2005; Berg, 2012). As a consequence, their distribution patterns depend less, or not at all, on local environmental conditions (Bie et al., 2012 and references within). Hence, body size differences may have enormous consequences for the spatial scaling of soil biodiversity and its associated major biogeochemical cycles and ecosystem services (Lavelle et al., 1997; Wardle, 2002; Bardgett & van der Putten, 2014). However, existing knowledge on the scaling relationship between body sizes and soil communities’ assembly processes is often based on meta-analyses and is therefore indirect (Lavelle et al., 1997; Ettema & Wardle, 2002; Berg, 2012).

Here, we addressed the question of how body size determines community assembly in a tropical forest, by assessing soil community assembly across taxonomic groups spanning several body size classes. In these ecosystems, a large part of our knowledge on spatial patterns of biodiversity and their underlying processes comes from large permanent forest plots (Hubbell, 2001; Condit et al., 2002; Anderson-Teixeira et al., 2015). In that spirit, we assessed the processes governing soil community assemblages by combining an extensive characterization of abiotic and plant cover...
conditions with a comprehensive survey of soil biodiversity using DNA metabarcoding in a 12-ha plot, where samples were collected every ten meters. We sought to (i) determine which factors, both abiotic (e.g. soil chemistry) and biotic (plant diversity, identity), influence soil community composition, (ii) evaluate the relative importance of environmental selection versus ecological drift in shaping soil communities, and (iii) determine how these effects relate to taxon body size. We hypothesized that in communities of small-bodied organisms, environmental selection should be more important relative to larger organisms, which should instead display spatial patterns resulting from ecological drift.

2. MATERIAL AND METHODS

2.1. Study site and sampling

The study site is located at the Nouragues Ecological Research Station, in the lowland rain forest of French Guiana (latitude: 4° 4’ 28” N, longitude: 52° 40’ 45” W, Fig. 1). Rainfall is 2861 mm.y⁻¹ (average 1992-2012), with a two-month dry season (< 100 mm.month⁻¹), from late August to early November, and a shorter dry season in March. Our sampling campaign was conducted November 7-20, 2012, towards the end of the dry season, which lasted from early September to late November. Cumulative rainfall during the 60 days preceding the sampling session was 134 mm, with 44 days without rain, and over 90% of the rainfall concentrated in seven days.
We surveyed a 12-ha (300 x 400 m) plot established in 1992. This plot extends on a gentle slope (7.5°) between a ridge and a small creek (Chave et al., 2008). The 5,640 trees occurring in the plot (diameter at breast height ≥ 10 cm) belong to over 600 species, with the two dominant species accounting each for only 2.3% of individuals (Baraloto et al., 2012). Sand and clay fractions are about 40% each in the soil top 10 cm. The parent material is Caribbean granite. Soil edaphic conditions are typical of tropical oxisols, with an acidic pH (pH = 5.0) and low exchangeable cation content (ECEC = 3.5 cmolc.kg⁻¹). The C:N and N:P ratios are of 13.4 and 40.5 respectively (median values), typical of tropical forests (Turner & Engelbrecht, 2010).

We sampled the plot following a regular grid scheme with a 10-m resolution grid, excluding bordering points, hence resulting in a total of 1,132 sampling points. At each point, 50-100g soil cores were collected with an auger at ~10 cm depth, excluding the O-horizon. We did so because the organic and mineral horizons harbour different arthropods and microbial communities (Baldrian et al., 2012; Yang et al., 2014). Lumping together these compartments may hence complicate the interpretation of the spatial distribution of certain soil clades. Consequently, we focused on the surface soil layer, which it is the most biogeochemically active in the mineral horizon (Silver et al., 2000). The soil cores were stored and sealed in sterile plastic bags after collection and transported to the field station laboratory. Extracellular DNA was extracted from 15 g of soil per soil core as described previously (Taberlet et al., 2012, Zinger et al., 2016) within 4 hours after sample collection to prevent microbial growth. This method was used due to its easiness of implementation for large-scale samplings. It also yields ecological signals similar to that obtained with total DNA, although it may under-sample...
rare taxa (Zinger et al., 2016). Its use for tropical soils is relevant because both micro and macroorganisms continuously DNA release in the environment that is most likely recycled rapidly in tropical soils, but persistent enough to buffer microbial diel-scale dynamics (Taberlet, Bonin, Zinger, & Coissac, 2018). Soil DNA was extracted twice for each soil core, and the remaining soil material was dried and stored for analytical chemistry analyses.

2.2. Molecular analyses

Soil biodiversity was surveyed through DNA metabarcoding using three DNA markers, with primers targeting three hypervariable regions of the ssu rRNA gene in Archaea, Bacteria and Eukaryota domains respectively (see Table 1). The corresponding primers have not been reported to present amplification biases (see Material S1 for a detailed description of priming sites conservatism across phyla). We conducted duplicated PCRs for each marker and each DNA extract, hence representing a total of 18,112 independent PCRs. To discriminate PCR products after sequencing, forward and reverse primers were tagged with a combination of two different 8-nucleotide labels. PCR and sequencing protocols are provided in Supplementary Methods. To control for potential contaminants (Salter et al., 2014) and false positive caused by tag-switching events (Esling, Lejzerowicz, & Pawlowski, 2015), the sequenced multiplexes comprised extractions/PCR blank controls, unused tag combinations, as well as positive controls for which specific analyses are provided in Material S2.
2.3. Sequence analyses and curation

About one billion sequencing reads were produced and curated mainly using the OBITools package (Boyer et al., 2016) and the R software v.3.3.2 (R Core Team, 2016) following the protocols provided in the Supplementary Methods (see also Material S2 for a validation of the data curation pipeline). Briefly, paired-end reads were assembled, assigned to their respective samples/marker and dereplicated. Low-quality sequences were excluded; the remaining ones were clustered into operational taxonomic units (OTUs) and assigned to taxonomic clade. We paid particular attention to minimize PCR/sequencing errors, contaminant and false positive sequences as well as potential non-functional PCRs by using conservative quality-check criteria. Raw and curated datasets statistics are reported in Table S1. Finally, the sequencing depth of each sample was standardized for each marker by randomly resampling a number of reads equal to the first quartile of read number across samples. Although such procedure has been recently questioned (McMurdie & Holmes, 2014), the data loss caused by standardization is here minimal and had no or weak effects on the dataset characteristics and the retrieved patterns of diversity compared to raw datasets (Fig. S1).

2.4. Focus taxonomic groups and body size

Within the Eukaryota and Bacteria domains, we distinguished groups on the basis of their taxonomic affiliation at the phylum or class level (Table 2). We did so because broadly defined functional traits such as body size and trophic categories are relatively well conserved within phyla (Briones, 2014; Martiny et al., 2015). This approach circumvents the difficulties in retrieving body size for each OTU, for which fine-level taxonomic assignation and/or morphological descriptions are seldom...
available. We restricted our analysis to the most abundant phyla (i.e. representing ≥ 1% of the total bacterial or eukaryotic OTU diversity). The OTUs unambiguously assigned to non-fungal unicellular eukaryotes clades (e.g. Stramenopiles, Rhizaria, Alveolata, Cercozoa, Amoebidae) constituted only a low fraction of the eukaryotic dataset (1.95%).

To include in our analysis this often-neglected, yet important part of the soil biota (Geisen et al. 2018), we therefore considered them as one single group, hereafter referred to as “protists”. Archaea were analysed as a single group due to imprecise taxonomic assignments.

Average body sizes per focus taxonomic group were inferred from Portillo, Leff, Lauber, & Fierer (2013) for bacterial cells, from Bécard & Pfeffer (1993) and Ingold (2001) for fungal propagules, and from Briones (2014) for the other groups (Table 2). Body size classes were defined as the size of the dispersal unit (propagule), rather than that of the mature individual. In fungi, spore size can be used as the operational definition of body size (Peay, Kennedy, & Bruns, 2008). Indeed, mature bodies may extend over large areas through mycelium growth (i.e. the vegetative part of fungi). The intraspecific variability of these mature forms span one to two orders of magnitude (Douhan, Vincenot, Gryta, & Selosse, 2011), and it is thus difficult to decide on an effective definition of a ‘body size’. In contrast, spore size is more conserved and was shown to be proportional to the reproduction rate and to the fructification size in certain fungal groups (Meerts, 1999; Douhan, Vincenot, Gryta, & Selosse, 2011). This definition also cuts across the domains of life as it corresponds to average cell size for unicellular organisms (i.e. archaea, bacteria and protists) and average body size of adults for the soil mesofauna.

This article is protected by copyright. All rights reserved.
2.5. Environmental parameters

We measure a total of 60 contextual parameters (Fig. S2) corresponding essentially to soil chemistry (including important nutrients such as Mg, S, B, Ca, Mn), but also to lidar and plant-DNA derived data.

2.5.1. Soil chemistry and airborne lidar. Total content in soil chemical elements was assessed every 20 m, corresponding to a quarter of the collected soil samples. Soil samples were ground and 100 mg of powder were digested in Teflon beakers in a clean room following this protocol: 1 ml H₂O₂ + 1 ml HNO₃ during 24 hours at room temperature and during 24 hours at 80°C; evaporation to dryness at 80°C; 1 ml HNO₃ + 1 mL HF during 24 hours at 80°C; evaporation to dryness; 20 drops of HCl + 10 drops of HNO₃ (aqua regia) during 24 hours at 115°C; evaporation to dryness. The residue was then removed in a matrix HNO₃ 2 % for analysis by ICPMS (Agilent 7500ce). Indium and rhenium were used as internal standards to correct for instrumental drift and eventual matrix effects during ICPMS analyses. The quality of the analysis was checked by analyzing international certified reference waters (CNRC SLRS-5, NIST SRM 1643e). The accuracy was better than 5% relative to the certified values and the analytical error (relative standard deviation) was better than 5% for concentrations ten times higher than the detection limits. Carbon and nitrogen concentration was measured by a CHN elemental analyzer (NA 2100 Protein, CE Instruments). In total, our dataset included 55 elemental concentrations per sample. These were krigged using an exponential variogram model so as to obtain values for all the points of the initial sampling design. This analysis was conducted with the sp (http://rspatial.r-forge.r-project.org/) and gstat (http://gstat.r-forge.r-project.org/) R packages. In our study, we did not assess
soil pH due to strong methodological constraints, although it is a common predictor of soil microbial diversity. However, previous studies in French Guiana revealed limited spatial variation compared to what could be expected in oxisols in general, and negligible variation in our study plot (pH=4.2±0.1, Fig. S3). In a previous study at a site encompassing greater soil pH variation than at our site (Barberán et al. 2015), soil pH was found to have a limited effect on tropical soil microbial communities. We hence expect pH to have a small explanatory power of community composition at our site.

A 1-m² digital elevation model (DEM) was obtained at our site using airborne lidar scanning (Réjou-Méchain et al., 2015; Tymen et al., 2016). From the DEM we derived slope, and topographic wetness indices (see Supplementary Methods). We also used the lidar data to define a canopy closure index, which quantifies how much light reaches the ground level at each point.

2.5.2. Plant sequencing data. Plants can influence soil communities through direct biotic interactions or local modification of the soil chemistry. To quantify this effect, we characterized the plant composition at the scale of the sampling point (i.e. root or litter presence) by amplifying a plastid DNA marker (P6 loop of the trnL intron) from soil DNA extracts, following the same analytical protocol as explained above (see also Table 1). We preferred the DNA metabarcoding approach over traditional plant census because it is a direct local measure, and because plant censuses usually include only trees ≥ 10 cm dbh, and thus exclude understory plants and lianas. Moreover, inferring the influence of plants on a local soil sample based on the stem locations is highly
challenging. Since PCR failed for 30% of the sampling points (330 of the 1,132), we decided to infer plant composition by a multinomial resampling of reads from neighbouring points. This approach is reasonable because tree roots influence is detectable up to 20 m from their corresponding stems in tropical forests (Barberán et al., 2015). We then inferred several plant-related characteristics to explain soil community assembly. These corresponded to plant diversity (Shannon index) and the identity of the three most dominant plant genera in each soil sample. In the molecular dataset, plant dominant genera corresponded to 66 different plant taxa, 42 of which were unambiguously identified. The three dominant plant genera represented on average 70 ± 15% of the plant reads in each sample, hence providing a good description of the local dominant plants.

2.6. Statistical analyses

We assessed the spatial autocorrelation of each focus taxonomic group using Mantel correlograms (999 permutations with FDR correction of p-values). This was done on both OTU diversity and community compositional dissimilarities. OTU diversity was calculated as the exponential of the Shannon entropy, an effective number of species that is less sensitive to rare OTUs (Jost, 2006) and hence more robust to describe microbial diversity (Haegeman et al., 2013). Community dissimilarities were obtained with the Bray-Curtis index. Correlation of spatial patterns between focus taxonomic groups and with plants was assessed with Pearson product-moment correlation or Mantel tests for OTU diversity and community dissimilarity, respectively.
To disentangle the relative importance of niche-based and neutral processes in the assembly of soil communities, we did not test neutral models (e.g. Hubbell, 2001) against our data because they assume an unbiased resampling of individuals and species. This assumption is not necessarily met with metabarcoding data (e.g. molecular biases, sampling of DNA fragments instead of individuals), which can bias estimates of Hubbell’s neutral theory parameters (Sommeria-Klein, Zinger, Taberlet, Coissac, & Chave, 2016). Instead, we used two complementary methods (as reviewed in Vellend et al., 2014). First we used a null model approach based on community pairwise dissimilarities, as proposed by Chase & Myers (2011) and extended by Stegen et al. (2013) to incorporate species abundances. This approach consists in determining whether community dissimilarity patterns deviate from a null scenario. Consequently, it does not require a priori knowledge of the local environmental conditions. For each sample, the null scenario was generated by random resampling of OTUs and reads in the total community matrix, while preserving the sample richness and number of reads in the sample. OTUs total occurrences and abundances were used as probabilities of selecting an OTU and its associated number of reads respectively. Dissimilarities across null communities were then computed with the Bray-Curtis index. This process was repeated 1000 times and the resulting dissimilarities were averaged out to obtain null expectations of community dissimilarities for each pair of samples. We then defined the null deviation as the difference between observed and averaged null dissimilarities (Tucker, Shoemaker, Davies, Nemergut, & Melbourne, 2016), which returns values between -1 and 1 and deviating from 0 when the observed dissimilarities differ from the null model. Values departing from 0 usually indicate environmental selection in well-sampled communities (Fig. S1, Tucker, Shoemaker, Davies, Nemergut, & Melbourne, 2016), but can also correspond to high dispersal limitation (values > 0) or
homogenizing dispersal (values < 0; Stegen et al., 2013). To better differentiate these processes, null-model inferences can be complemented with analyses of the community phylogenetic structure (Stegen et al., 2013). We did not conduct such an analysis here because obtaining phylogenetic trees for the 19 taxonomic groups studied proved an insurmountable challenge, due to the unavailability of backbone phylogenies for most groups, and the uncertainty associated to phylogenetic placement methods for short sequences.

The null model approach cannot identify potential biotic or abiotic drivers of community assembly. To complement it, we applied in parallel another approach based on variation partitioning and redundancy analysis (RDA; Legendre, 2008). This analysis consists in decomposing the variance of community composition explained by environmental or spatial variables alone as well their combined effects. Initially, the contribution of pure environmental vs. spatial component was considered to be indicative of the relative importance of niche-base processes vs. dispersal limitation (Cottenie, 2005; Legendre, 2008). This interpretation has been questioned because it applies only when the environmental context is well characterized, spatially not structured (Smith & Lundholm, 2010, Vellend et al., 2014), and when neutral processes are not correlated with the environment. We will consider these limits in our interpretations.

Our preliminary analyses indicated a linear relationship between dependent variables (soil diversity or community distribution) and explanatory variables, a prerequisite for using RDA. We Hellinger-transformed the OTU tables of each taxonomic group to downweigh rare OTUs, as well as to preserve the Euclidean distance among
samples (Legendre & Gallagher, 2001). We then constructed three parsimonious models corresponding to three non-exclusive hypotheses: (1) community composition is driven by the measured abiotic parameters, (2) community composition is driven by direct or indirect interactions with plants, or (3) community composition is driven by spatial processes, linked or not to environmental factors. Model 1 included soil chemistry and lidar-derived data as predictive factors. Model 2 included plant explanatory variables (Fig. 1, Table S2). For these two models, explanatory variables were preselected to reduce multicollinearity ($|\text{Pearson's } r| \geq 0.7$, Fig. S2), and normalized with a Box-Cox transformation to meet normality assumptions. Finally, model 3 included spatial eigenvectors as explanatory variables, derived from a Principal Components of Neighbours Matrices (PCNM) analysis (Dray, Legendre, & Peres-Neto, 2006). PCNMs allow modelling the spatial structure of community composition variation. To avoid inflation of $R^2$ statistics caused by the large number of PCNM eigenvectors (Gilbert & Bennett, 2010), we preselected them prior the forward selection procedure. We retained only PCNMs explaining a significant fraction of the variation in the biological response (at a significance level $p \leq 0.02$), as assessed through partial canonical redundancy analysis (pRDA). This pre-selection was independently performed for each study clade. Geographic coordinates were also included in the model to account for possible linear trends along the study area (Dray et al., 2006). Finally, we applied a classical forward selection procedure on model 1, 2 and 3 independently for each taxonomic group to limit inflation of the amount of explained variance and Type I error (Blanchet, Legendre, & Borcard, 2008). The resulting abiotic, plant and spatial matrices were used to conduct a variation-partitioning analysis. Significance of the total RDA models and of variation partitioning pure components were determined with 1,000 Monte-Carlo permutations. We here report adjusted R-squared statistics ($R^2_{\text{adj}}$), which This article is protected by copyright. All rights reserved.
are less inflated when the numbers of explanatory variables is high (Peres-Neto, Legendre, Dray, & Borcard, 2006). To better interpret the combined effects of environmental/spatial parameters, we identified the plant and abiotic parameters that correlated with PCNM eigenvector using pRDA.

To further gain insights into the spatial scales at which the community composition may co-vary with environmental conditions, we decomposed the community variation with a second “full” RDA model that included four categories of explanatory variables. The first category comprised environmental features, i.e. abiotic and plant variables. The three remaining categories corresponded to PCNMs eigenvectors with wavelengths ≥ 100 m, between 100 and 25 m, and < 25 m (Fig. S6). The $R^2_{adj}$ statistics obtained for pure or combined effects of each category of variable were used to calculate their relative contribution to the total explained variance. Variables used in these models corresponded to those preselected with the initial forward selection procedure, and tests for significance were conducted as described above. All variation-partitioning analyses were repeated by considering OTU diversity as a response variable to identify predictors of soil biodiversity.

Finally, we tested the effect of body size on community assembly by comparing taxon’s propagule sizes with the various statistics obtained above using Pearson product-moment correlation tests. All analyses were conducted with the vegan R package (http://vegan.r-forge.r-project.org/).
3. RESULTS

We found a total of 2,502 archaeal, 19,101 bacterial, and 11,470 eukaryotic OTUs within the 12-ha plot (Table 2). Many of these OTUs were rare (relative abundance < 0.1%): they accounted for 36 ± 10 % (mean ± standard deviation values hereafter) of archaea OTUs, 86 ± 2 % of bacterial OTUs, and 80 ± 6 % of eukaryotic OTUs. Bacterial OTUs identified at the phylum level (88% of OTUs, 94% of reads) corresponded primarily to Actinobacteria, Alphaproteobacteria, and Acidobacteria. Eukaryotic OTUs identified at least at the phylum level (51% of OTUs, 70% of reads) belonged to fungi (mainly Agaricomycetes, Glomeromycetes and Orbiliomycetes; Table S3), arthropods (mainly termites, mites and springtails) and annelids (Oligochaeta). For Archaea, the OTUs being assigned at least at the phylum level (3% of OTUs, 64% of reads) mainly corresponded to Nitrososphaeria and Methanomicrobia.

Figure 2 illustrates the spatial variation in local diversity for the focal taxonomic groups. Both positive and negative spatial autocorrelation were observed at various distances, depending on the taxon (Fig. 2, Fig. S4). Dissimilarities of soil community increased with body size (Fig. 3a, Pearson’s $r = 0.52$, $p = 0.02$). Compositional turnover also displayed positive spatial autocorrelation (Fig. S5) for distance classes that decreased with taxon body size (Pearson’s $r = -0.59$, $p = 0.007$). Negative spatial autocorrelation was observed at larger distances (> 100 m) mostly in microorganisms (Fig. S5). The average OTU distribution range of each taxonomic group was > 200 m in general, and showed no trend with body size (Table 2). Pairwise correlation of OTU diversity or compositional dissimilarity between focal taxonomic groups was moderate to low (|Pearson’s $r| < 0.7$), except for several bacterial clades. They showed low or no
correlation with plant diversity or community turnover (Table S4-5). Finally, the observed community dissimilarity values did not deviate from null expectations for all taxonomic groups, Firmicutes excepted (Fig. 3b). Null deviation values did not show any particular trend with respect to body size (Pearson’s $r = 0.18, p = 0.45$).

Abiotic and biotic conditions were noticeably heterogeneous, with abiotic gradients detectable at broad and intermediate spatial scales (Fig. 1,S6). However, the full RDA models, i.e. including abiotic, plant and spatial descriptors, explained only 3-13% of the variation in OTU composition across the 19 focal taxonomic groups (Fig. 4a, Table S6). Only $3 \pm 1\%$ of the variation was due to total abiotic effects, mainly through abiotic parameters that were spatially structured at broad to intermediate spatial scales (Fig. 1,S5, Table S6), such as soil aluminium and topographic variables (i.e. elevation, convexity and wetness). These abiotic parameters correlated negatively with the OTU diversity of most unicellular taxonomic groups, and worms for topographic variables (Fig. S7-8). When exploring plants’ influence on species composition across groups, we found that plant variables explained $2 \pm 1\%$ of the variation. The identity of the dominant plant genera in soil samples as detected with the plant DNA marker was the best plant-related predicting factor. Finally, spatial total effects explained $7 \pm 2\%$ of the variation, half of which corresponded to pure spatial effects.

When comparing these results across the focus taxonomic groups, we found that the unexplained component of RDA models correlated positively with dispersal unit size (Fig. 4b), and was strongly related to the averaged dissimilarities expected under the null model (Pearson’s $r = 0.82, p < 0.001$). We also found that the explained components corresponding to both pure environmental (i.e. abiotic and plant variables)
and broad/intermediate scale effects (combined with environmental effects or not) all negatively correlated with dispersal unit size (Fig. S9, Table S7). The fraction of community variation explained by fine-scale spatial variable was unrelated to environmental ones, and tended to increase with organism propagule size. Using OTU diversity instead of OTU composition as the dependent variable resulted in greater amount of variation explained by the tested factors (4 to 46%) and their relative importance for the two “full” RDA models, as well as their variation across body size was similar to what is reported above (Table S8-9, Fig. S10-11).

4. DISCUSSION

4.1. Environmental drivers of soil communities

In this study, we assessed the environmental predictors that significantly explained soil community composition. Amongst them we identified the locally dominant plant genera, an effect previously termed the “tree identity effect” (Bardgett & van der Putten, 2014; Tedersoo et al., 2016). Neotropical trees species can be indeed involved in specific mutualistic/trophic interactions, as they greatly differ in the quantity, quality and chemistry of their litter (Hättenschwiler, Aeschlimann, Coûteaux, Roy, & Bonal, 2008), and most likely of their root exudates and physical structure. We found that soil total aluminium content was also a predictor of microbial community composition and correlated negatively with microbial diversity. Aluminium concentration strongly correlated with that of many other metals (e.g. Ti, Cu, Zn), all well known for their toxic properties (Piña & Cervantes, 1996; Lemire, Harrison, & Turner, 2013) and their capacity to immobilize soil nutrients and organic matter (Fujii,
Soils with high aluminium content also tend to be more acidic (Driscoll & Schecher, 1990; Abreu, Muraoka, & Lavorante, 2003), conditions under which microbial diversity is usually lower (Ramirez et al., 2014; Barberán et al., 2015). Finally, topography, slope, and the topographic wetness index, all closely related to soil moisture, also predicted the community composition for many taxonomic groups. They were the most important predictor of OTU diversity and composition of flat/earthworms, which are well known to have particular soil moisture optima (Fragoso & Lavelle, 1992; Jiménez, Decaëns, Lavelle, & Rossi, 2014). These variables also correlated with the spatial diversity and composition patterns of Deltaproteobacteria, an anaerobic clade favoured in wet conditions (Martiny et al., 2015). We hence interpret the decline of diversity in metal-rich and/or drier sites as a response to stressful conditions.

However, environmental parameters explained only a minor fraction of the variation in soil community composition. This result contrasts with previous studies, where soil organisms are often reported to co-vary relatively strongly with e.g., soil pH, moisture, nutrient or organic matter availability/quality (Dumbrell, Nelson, Helgason, Dytham, & Fitter, 2010; Wu, Ayres, Bardgett, Wall, & Garey, 2011; Ramirez et al., 2014; Tedersoo et al., 2016) or plant diversity (Basset et al., 2012; Peay, Baraloto, & Fine, 2013). Although we have characterized the environmental conditions to the best of our ability, we inevitably did not quantify directly all environmental predictors or biotic interactions. For example, lidar-derived measures, including soil wetness index, may lack precision for our sampling grain. Also we did not measure soil pH, but expect it to be homogeneous spatially and to have a negligible effect (see Methods). More generally,
many edaphic parameters are often spatially structured and their effects should be indirectly detected with PCNM variables (Smith & Lundholm, 2010; De Cáceres et al., 2012; Baldeck et al. 2013). Here, these spatial variables also explained a minor part of soil community variation, hence suggesting that environmental gradients have only a weak influence on our study site. Yet, other parameters inaccessible with our sampling grain such as soil microstructure or the geometry of roots may be also important in explaining the small-scale distribution of soil organisms (Berg, 2012; Baveye, Berthelin, & Munch, 2016).

The unexplained component of community variation, which was prominent in our study, has often been interpreted as resulting from local stochastic demographic processes (Legendre, 2008; De Cáceres et al. 2012). We suggest that in our study, this interpretation is valid. Indeed, we also found that the observed community dissimilarity patterns did not depart markedly from null expectations in any of the focal taxonomic groups. Also, the amount of unexplained variation across the focus taxonomic groups correlated remarkably well with corresponding averaged dissimilarities expected under the null model, a feature already observed (De Cáceres et al. 2012). Consequently, although the environmental context was not fully characterised in our study, it is unlikely that we underestimated the importance of environmental selection. Rather, all our analyses point towards the prominence of stochastic processes in the assembly of soil communities at the scale of our study plot, hence differing from most soil diversity studies (reviewed in Berg, 2012; Hanson et al., 2012; Barberán, Casamayor, & Fierer, 2014). On the other hand, environmental predictors (in particular soil Aluminium content and soil moisture) better explained soil diversity patterns. This suggests that
low moisture and high Aluminium sites do not only recruit particular taxa, as suggested by their weak, yet significant effects on community composition. These conditions also impose energetic constraints on the system carrying capacity that would limit local community size, causing a random species loss. Further analyses of the community phylogenetic structure could help confirming this hypothesis (Hurlbert & Stegen, 2014).

4.2. Underlying causes of community stochasticity

A major problem in attempts to explain patterns of biodiversity is the difficulty of observing ecological processes at the precise scale at which they manifest themselves (Levin, 1992). Soils are complex systems where ecological processes operate at a hierarchy of scales, and any attempt to interpret them should carefully examine the spatial scale of study and the grain of the sampling unit (Ettema & Wardle, 2002; Hinsinger, Bengough, Vetterlein, & Young, 2009; Berg, 2012; Baveye, Berthelin, & Munch, 2016). We suggest that at least part of the discrepancies between our findings and previous studies are related to a problem of size and sampling grain. Previous analyses were based on sampling units of typically e.g. 20 x 30 m distributed across landscapes or regions with potentially steep environmental gradients. For example, previously observed relationships between the species diversity/composition of soil organisms and local plant community composition (Basset et al., 2012; Peay, Baraloto, & Fine, 2013) are observable at spatial scales at which local plant community composition correlates with environmental conditions. However, the strength of this relationship decreases with decreasing sizes of sampling units and vanishes at the scale of the sampling point (Barberán et al., 2015). In our study, we did not observe such a relationship because we sampled local soil cores of ca. 15 g from which we inferred...
both local soil richness and plant community composition. Thus, we were able to relate
the co-occurrence of microbial cells and plant cells at the centimetric scale (Dini-
Andreote & van Elsas, 2013), and concluded there was a weak, yet significant, “tree
identity effect” as reported above.

Although our study plot was large and displayed gradients of abiotic conditions
and trees species distribution, these gradients are most likely not steep enough to drive
soil communities assembly as in other forests. Accordingly, previous studies have
reported weak effects of environmental selection at smaller spatial scales (0.04 to 0.2
ha) for soil microeukaryotes (Bahram et al., 2016), earthworms (Jiménez, Decaëns,
Lavelle, & Rossi, 2014), and bacteria (Barberán et al., 2015). This implies that soil
communities appear to be primarily driven by ecological drift and/or dispersal
processes. In our study plot, dispersal is unlikely to be limited. At this scale, dispersal of
microbes occurs through the movement of soil particles caused by e.g. raindrops and
water flows, above and belowground animals activity, and filamentous fungi or plant
root growth (Rillig et al. 2015, Joung et al., 2017). Similarly, passive dispersal of soil
mesofauna can spread through the entire plot and beyond, through transport of
cocoons by water flow, or phoresy on the macrofauna (Decaëns et al., 2016).
Accordingly, we found that OTUs had relatively large distribution ranges (> 200 m on
average) and that communities were not more dissimilar than null ones, contrary to the
dispersal limitation scenario (Stegen et al., 2013). In contrast, high levels of dispersal
(i.e. homogenizing dispersal) should offset both environmental effects and ecological
drift by increasing the chance for a soil-inhabiting species to occupy any location
through a source-sink dynamics (Vellend, 2010; Stegen et al., 2013). This should result
in communities that are more similar than expected by chance, which was not the case here. Priority effects, i.e. the fact that early colonizing species determine the local assemblage, may also cause stochastic patterns. However, this process is unlikely in our study because it is linked to environmental selection and would lead to communities more dissimilar than under null expectations. Hence, the stochasticity in the community assembly in our study plot is most likely caused by ecological drift without dispersal limitation.

4.3. Organism body size determines assembly processes

Despite this apparent stochasticity, we still found some coherence in the spatial patterns across taxonomic groups. An important result from our analysis was the significant and linear increase of both the amount of unexplained variation in diversity or composition, and null community dissimilarities with increasing propagule size. We did not retrieve this trend for null deviation values because the body-size dependence of both observed and null community dissimilarities cancels out using this metric. One criticism of null models is that they preserve species abundances and local diversity distributions of the data, which can result from assembly mechanisms. Besides, null deviation values are difficult to compare across ecosystems or taxonomic groups due to their sensitivity to regional diversity and sampling effort (Bennett & Gilbert, 2015; Tucker, Shoemaker, Davies, Nemergut, & Melbourne, 2015). Finally, we also observed a negative correlation of body size with the spatial autocorrelation extent in community turnover, as well as with observed community dissimilarity. Our result implies that community stochasticity increases with body size.
The smallest body-size category includes bacteria, whose distribution showed
greater spatial autocorrelation at broad-to-intermediate scales. The variation-
partitioning approach including explicitly spatial scales as explanatory variables
suggests that this distribution pattern can be explained by two non-exclusive processes:
(i) a more pronounced environmental selection by the broad to intermediate-scale
environmental gradients at our site, (ii) a higher level of homogenizing dispersal at
intermediate scales, as suggested by the negative relationship between body size and
the pure intermediate-scale spatial component. Micro-eukaryote (i.e. fungi and protists)
distributions were less well explained by environmental gradients than bacteria. Similar
differences have been observed between bacterial and fungal community assembly
(Powell et al., 2015), which were explained by a greater dispersal limitation and
stronger priority effects in fungi. That this interpretation holds true for our study site
remains to be determined. Also, in fungi, the mycelium can spread across
heterogeneous environments and the apparent distribution of fungi may be less related
to environmental conditions at the scale of observation. Protists, archaea and certain
bacterial groups (i.e. Firmicutes and Bacteroidetes) displayed similar features. These
groups harbour low abundances in soils relative to other clades, they do not necessarily
exhibit dormant stages, and some of them can be associated with soil invertebrates
(Fierer, Strickland, Liptzin, Bradford, & Cleveland, 2009; Engel & Moran, 2013; Ramirez
et al., 2014; Barberán et al., 2015; Geisen et al., 2018). This suggests that these microbial
groups have smaller population sizes in soils and/or lower dispersal rates. They could
hence experience more extinction and be more prone to ecological drift compared to
most bacterial groups.

This article is protected by copyright. All rights reserved.
The distribution of OTUs of large-bodied organisms, i.e. arthropods, annelids and flatworms were even less related to environmental parameters and did not exhibit broad to intermediate spatial structures. Soil mesofauna communities were hence the most stochastic of the focus taxonomic groups at the scale of our 12-ha plot and grain studied. We also found that the contribution of the pure fine-scale spatial effects to the community variation tended to increase with increasing propagule size. This result is in agreement with the clumped spatial distribution of soil mesofauna, which usually displays spatial aggregation below 10 m (Berg, 2012; Jiménez, Decaëns, Lavelle, & Rossi, 2014; Bahram et al., 2016). This roughly corresponds to the horizontal distance earthworms can actively disperse per year, and suggests that although dispersal is not limited in the study plot, its rates may be lower for mesofauna than for microbes. In addition, studies using a finer sampling grain found that these aggregates co-vary poorly with environmental factors (Jiménez, Decaëns, Lavelle, & Rossi, 2014; Bahram et al., 2016). Indeed, at such small scales relative to mesofauna body size, population densities are low and hence more driven by drift (Chase, 2014). On the opposite end of the range, environmental selection of large-bodied organisms emerges at scales larger than 12-ha when soils properties are highly contrasted, as reported for tropical trees (Baldeck et al., 2013) or mesofauna (Basset et al., 2015; Decaëns et al., 2016).

The body-size dependence of drift, dispersal, and environmental selection is in line with Levin’s argument on the problem of scales (Levin, 1992), and confirms our initial hypothesis. For the studied area and sampling grain, we considered patterns of diversity at a range of spatial scales owing to differences in body size of the studied taxa relative to the studied area. We find that microbial groups, for which 12-ha represent a huge area, are better predicted by the environment than mesofauna ones, for which 12-
ha represent a smaller spatial extent. This mirrors the increases of environmental selection with the observational spatial scale reported for single taxonomic groups (De Cáceres et al., 2012; Barton et al., 2013; Chase, 2014; Dini-Andreote, Stegen, van Elsas, & Salles, 2015). This result points to an interesting juncture between two fields of ecological theory that have heretofore been loosely connected, namely the role of body size in explaining the scaling rules of metabolism (Brown, Gillooly, Allen, Savage, & West, 2004) and that of environment and dispersal in explaining the assembly of ecological communities (Ricklefs, 2004). In that light, the fact that large-bodied organisms are more stochastic than small-bodied organisms assembly is consistent with the main predictions of macroecology (Ricklefs, 2004) and soil science (Ettema & Wardle, 2002; Berg, 2012). We are aware that the taxonomic groups studied here harbour a large variation in propagule size, but this variation does not exceed the between-group body size variation (Briones 2014, Portillo, Leff, Lauber, & Fierer, 2013). Consequently, our finding of a log-linear relationship between the contribution of environmental and/or pure spatial effects in community assembly and propagule size is robust, and generalizes previous empirical observations in freshwater ecosystems (Bie et al., 2012 and references within). It provides an empirical evidence of spatial scaling rules across the soil food web. Although other important biological features also likely explain the differences of spatial distribution among soil taxonomic groups (e.g. clonal vs. sexual reproduction, dormancy, mutualistic/pathogenic interactions, active vs. passive transport), body size constitutes an operational parameter that cuts across the tree of life. Therefore, and perhaps unsurprisingly, body size has been considered as a key trait in the modelling of species interaction networks, as it constrains several essential species interaction variables such as handling times, densities, or attack rates (Petchey, Beckerman, Riede, & Warren, 2008; Brose, 2010).

This article is protected by copyright. All rights reserved.
The recent explosion of DNA-based studies has considerably increased our knowledge on the taxonomic, genetic and functional diversity of soil organisms (Bardgett & van der Putten, 2014), but has yet to provide a comprehensive understanding of the mechanisms shaping soil biodiversity (Baveye, Berthelin, & Munch, 2016). Here, we have evidence that this tropical soil community at the scale of a 12-ha plot is prominently neutral, mirroring tree communities in similar environments (Hubbell, 2001). However, it is likely that soil community assembly in other forests with more heterogeneous soil abiotic condition may be more prone to environmental selection than here. Still, one direct and important consequence of our result confirms that soil diversity, distribution, and extinction rates cannot be inferred from the patterns of plant or any other surrogate group at the field scale (Schuldt et al., 2015; Veresoglou, Halley, & Rillig et al., 2015). Yet, our results also suggest that accounting for the scaling relationship between body size and community assembly processes will help unravel spatial patterns in such complex, multitrophic communities and contribute to the development of currently emerging predictive spatial models of soil food webs (Galiana et al., 2018).

**Acknowledgements**

We are indebted to the staff of the Nouragues Research Field Station (CNRS, French Guiana) and the Nouragues Natural Reserve, for commodities and technical help during the field session. We are grateful to Benjamin L. Turner for providing us soil pH data and advice on the manuscript. We also thank Johan Pansu, Guilhem Sommeria-Klein, Mélanie Roy and Renato Lima for helpful discussions, as well as the Genotoul bioinformatics platform Toulouse Midi-Pyrenees (Bioinfo Genotoul) and Pierre Solbes.
(EDB-Calc Cluster) for providing computing and storage resources. We are also grateful to anonymous referees for their constructive comments on an earlier version of the manuscript. The work was funded by the METABAR project (ANR-11-BSV7-0020) and has benefitted from "Investissement d'Avenir" grants managed by Agence Nationale de la Recherche (CEBA: ANR-10-LABX-25-01; TULIP: ANR-10-LABX-0041; OSUG@2020: ANR-10-LABX-56; ANAEE-France: ANR-11-INBS-0001). This work is dedicated to the memory of our colleague Serge Aubert.
References


Baldrian, P., Kolařík, M., Štursová, M., Kopecký, J., Valášková, V., Větrovský, T., ... & Vofíšková, J. (2012). Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. *The ISME journal, 6*(2), 248.


Baveye, P. C., Berthelin, J., & Munch, J. C. (2016). Too much or not enough: Reflection on


This article is protected by copyright. All rights reserved.


Biology and Biochemistry, 24(12), 1397-1408.


Hurlbert, A. H., & Stegen, J. C. (2014). When should species richness be energy limited, and how would we know?. Ecology letters, 17(4), 401-413.


This article is protected by copyright. All rights reserved.


This article is protected by copyright. All rights reserved.


Tedersoo, L., Bahram, M., Cajthaml, T., Põlme, S., Hiiesalu, I., Anslan, S., ... & Abarenkov, K. (2016). Tree diversity and species identity effects on soil fungi, protists and animals are context dependent. *The ISME journal, 10*(2), 346.


Vellend, M., Srivastava, D. S., Anderson, K. M., Brown, C. D., Jankowski, J. E., Kleynhans, E.

This article is protected by copyright. All rights reserved.


Zinger, L., Chave, J., Coissac, E., Iribar, A., Louisanna, E., Manzi, S., ... & Taberlet, P. (2016). Extracellular DNA extraction is a fast, cheap and reliable alternative for multi-taxa surveys based on soil DNA. Soil Biology and Biochemistry, 96, 16-19.
Data accessibility

Raw and curated sequencing data as well as associated metadata and bioinformatic codes are available on the Dryad Digital Repository (https://doi.org/10.5061/dryad.dc28m).

Author contributions

L.Z., P.T. and J.C. conceived the study. L.Z, P.T., H.S., A.B, M.D.B, P.G., L.G., C.G.C, A.I., M.R.M, G.R., E.C and J.C contributed to the fieldwork and/or to DNA extractions. P.T., A.B., A.I. and D.R. conducted the laboratory work to produce the metabarcoding data. J.V. and C.Z. performed the chemical analyses. B.T. provided the LiDAR data. L.Z. performed the bioinformatics and statistical analyses with the help of V.S., F.B., E.C, W.T. and J.C. The manuscript was written by L.Z. and J.C. with input from all co-authors.

Competing interests

L.G. and P.T. are co-inventors of patents related to the gh primers and the use of the P6 loop of the chloroplast trnL (UAA) intron for plant identification using degraded template DNA. These patents only restrict commercial applications and have no impact on the use of this locus by academic researchers.
Table 1: Barcodes characteristics and associated PCR conditions.

<table>
<thead>
<tr>
<th>Primers set</th>
<th>Archaea</th>
<th>Bacteria</th>
<th>Eukaryota</th>
<th>Viridiplantae</th>
</tr>
</thead>
<tbody>
<tr>
<td>Targeted region</td>
<td>V8-V9 regions, 16S rRNA gene</td>
<td>V5-V6 regions, 16S rRNA gene</td>
<td>V7 region, 18S rRNA gene</td>
<td>trnL-P6 loop, chloroplastic DNA</td>
</tr>
<tr>
<td>Forward primer (5’–3’)</td>
<td>CCTGCTCCTTGACACAC†</td>
<td>GGATTAGATACCTGGTAGT‡</td>
<td>TTTGTCTGTTAATTCG§</td>
<td>GGGCAATCCTGAGCCAA¶</td>
</tr>
<tr>
<td>Reverse primer (5’–3’)</td>
<td>CCTACGGCTACCTTGTAAC†</td>
<td>CACGACACGAGCTGAGC‡</td>
<td>TCACAGACCTGTATTGC§</td>
<td>CCATTGAGTCTCTGCACCTATC¶</td>
</tr>
<tr>
<td>PCR cycles characteristics†† (Denaturation, Annealing, Elongation – Final elongation)</td>
<td>95°C(30s), 58°C(30s), 72°C(90s) – 72°C(7min)</td>
<td>95°C(30s), 57°C(30s), 72°C(90s) – 72°C(7min)</td>
<td>95°C(30s), 45°C(30s) – 72°C(60s) – 72°C(7min)</td>
<td>95°C(30s), 50°C(30s), 72°C(60s) – 72°C(7min)</td>
</tr>
<tr>
<td>Illumina platform</td>
<td>HiSeq 2x100</td>
<td>MiSeq 2x250</td>
<td>HiSeq 2x150</td>
<td>HiSeq 2x100</td>
</tr>
</tbody>
</table>

† this study
‡ Fliegerova et al., 2014
§ Guardiola et al., 2015
¶ Taberlet, Gielly, Pautou, & Bouvet, 1991
†† All PCRs were preceded by an initial denaturation at 95°C (10 min)

This article is protected by copyright. All rights reserved.
Table 2. Soil biota characteristics in the 12-ha plot. Only the studied taxonomic groups (most abundant ones) are shown. Plant OTUs retrieved with the eukaryote marker were not included in our analysis. They represented only 2% of the eukaryotic OTUs (10% of reads). Relative abundances were calculated for each marker separately and correspond to the percent of reads of each clade. Richness values correspond to the number of OTUs (mean ± standard deviation [SD] for values per sample are reported in parentheses). Diversity values correspond to the effective number of OTUs (exponential Shannon diversity index; mean ± SD for values per sample). Note that the latter provides better estimates of species diversity than the number of OTUs per se for metabarcoding data (Haegeman et al. 2013, Material S2). OTUs distribution range corresponds to the maximum distance between samples where an OTU is detected (mean ± SD). n.a.: not applicable. The study plot is located in the Nouragues Reserve, French Guiana.

<table>
<thead>
<tr>
<th>Clade</th>
<th>Total rel. abundance (%)</th>
<th>Total Richness (per sample)</th>
<th>Total Diversity (per sample)</th>
<th>OTUs distribution range (m)</th>
<th>Propagule size (µm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archaea</td>
<td>n.a.</td>
<td>2,502 (43 ± 30)</td>
<td>30 (9 ± 8)</td>
<td>170 ± 163</td>
<td>0.5</td>
</tr>
<tr>
<td>Bacteria</td>
<td>n.a.</td>
<td>19,101 (1,391 ± 232)</td>
<td>885 (453 ± 102)</td>
<td>289 ± 158</td>
<td>n.a.</td>
</tr>
<tr>
<td>Acidobacteria</td>
<td>19.81</td>
<td>1,601 (231 ± 46)</td>
<td>132 (80 ± 32)</td>
<td>352 ± 136</td>
<td>0.5</td>
</tr>
<tr>
<td>Actinobacteria</td>
<td>20.58</td>
<td>1,696 (167 ± 37)</td>
<td>86 (54 ± 13)</td>
<td>313 ± 151</td>
<td>5</td>
</tr>
<tr>
<td>Bacteroidetes</td>
<td>1.02</td>
<td>767 (31 ± 12)</td>
<td>97 (22 ± 8)</td>
<td>242 ± 165</td>
<td>0.5</td>
</tr>
<tr>
<td>Chloroflexi</td>
<td>5.22</td>
<td>2,147 (136 ± 48)</td>
<td>268 (73 ± 29)</td>
<td>308 ± 148</td>
<td>1</td>
</tr>
<tr>
<td>Firmicutes</td>
<td>2.66</td>
<td>873 (19 ± 6)</td>
<td>4 (3 ± 2)</td>
<td>186 ± 164</td>
<td>1</td>
</tr>
<tr>
<td>Alphaproteobacteria</td>
<td>19.86</td>
<td>3,234 (303 ± 57)</td>
<td>242 (129 ± 25)</td>
<td>303 ± 155</td>
<td>0.5</td>
</tr>
<tr>
<td>Betaproteobacteria</td>
<td>3.47</td>
<td>482 (45 ± 13)</td>
<td>24 (14 ± 7)</td>
<td>297 ± 160</td>
<td>5</td>
</tr>
<tr>
<td>Deltaproteobacteria</td>
<td>4.36</td>
<td>1,357 (75 ± 31)</td>
<td>73 (34 ± 15)</td>
<td>279 ± 151</td>
<td>0.5</td>
</tr>
<tr>
<td>Gammaproteobacteria</td>
<td>8.59</td>
<td>888 (72 ± 11)</td>
<td>30 (21 ± 5)</td>
<td>260 ± 169</td>
<td>5</td>
</tr>
<tr>
<td>Verrucomicrobia</td>
<td>4.99</td>
<td>353 (51 ± 12)</td>
<td>33 (22 ± 5)</td>
<td>344 ± 139</td>
<td>0.5</td>
</tr>
<tr>
<td>Eukaryota</td>
<td>n.a.</td>
<td>11,470 (475 ± 137)</td>
<td>228 (51 ± 23)</td>
<td>254 ± 160</td>
<td>n.a.</td>
</tr>
<tr>
<td>Ascomycota</td>
<td>3.49</td>
<td>317 (25 ± 9)</td>
<td>22 (8 ± 3)</td>
<td>312 ± 150</td>
<td>100</td>
</tr>
<tr>
<td>Basidiomycota</td>
<td>16.94</td>
<td>387 (31 ± 8)</td>
<td>44 (8 ± 4)</td>
<td>305 ± 149</td>
<td>10</td>
</tr>
<tr>
<td>Glomeromycota</td>
<td>1.08</td>
<td>38 (7 ± 2)</td>
<td>8 (5 ± 1)</td>
<td>360 ± 141</td>
<td>200</td>
</tr>
<tr>
<td>Annelida</td>
<td>6.62</td>
<td>49 (8 ± 3)</td>
<td>6 (3 ± 1)</td>
<td>336 ± 155</td>
<td>20e³</td>
</tr>
<tr>
<td>Arthropoda</td>
<td>17.65</td>
<td>1,777 (63 ± 21)</td>
<td>91 (15 ± 7)</td>
<td>235 ± 168</td>
<td>10e³</td>
</tr>
</tbody>
</table>

This article is protected by copyright. All rights reserved.
<table>
<thead>
<tr>
<th>Kingdom</th>
<th>Value</th>
<th>Mean (± SE)</th>
<th>Median (± SE)</th>
<th>Mean (± SE)</th>
<th>Median (± SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nematoda</td>
<td>0.47</td>
<td>359 (9 ± 5)</td>
<td>35 (6 ± 2)</td>
<td>258 ± 153</td>
<td>100</td>
</tr>
<tr>
<td>Platyhelminthes</td>
<td>0.54</td>
<td>117 (4 ± 3)</td>
<td>15 (3 ± 1)</td>
<td>226 ± 169</td>
<td>20e³</td>
</tr>
<tr>
<td>Protists</td>
<td>1.95</td>
<td>1,610 (53 ± 40)</td>
<td>126 (27 ± 18)</td>
<td>265 ± 154</td>
<td>100</td>
</tr>
</tbody>
</table>
Figures captions

Figure 1. Study site description. a) Geographic location (red) of the study site in South America (top right), and in French Guiana. Elevation and hydrologic networks were retrieved from the Wordclim dataset (www.worldclim.org) and GIS Guiana layers (gisguyane.brgm.fr). b) Photography of the studied plot (12 ha, 300 x 400 m). Sampling points (every 10 meters) are marked with red sticks. Photo credit: P. Taberlet. c-f) Soil content in Carbon, Aluminium, Phosphorous and Potassium. g) site topography, h) soil wetness (TWI index, unitless) and i) canopy height as inferred from Lidar data. j) Distribution of the most dominant plant taxon in each soil sample as inferred from the plant molecular dataset. Only the six most frequent dominant taxon are shown: red: Apocynoideae; orange: Andira; yellow: Ingeae; green: Brosimum; lightblue: Sapotaceae; blue: Drypetes. x/y axes are not shown for representation purposes and range between 10-290 and 10-390 meters respectively.

Figure 2. Spatial distribution of OTU diversity per focal clade. The colour scale is expressed as effective numbers of OTUs (exponential Shannon diversity index) for archaea (a), bacterial clades (b-k) and eukaryotic clades (l-s), ranked by increasing body size (Table 2). Maps were obtained by ordinary kriging. Grey dots represent the samples available for each clade. The study plot is located in the Nouragues Reserve, French Guiana.

Figure 3. Community dissimilarities in the study plot (a), and their deviation from null model expectations (b) for each focus taxonomic group. Values of both observed dissimilarity and null deviation are summarized by their average (red dots) and 95% confidence intervals. In (b), the horizontal black dashed line indicates neutral assembly. Focal taxonomic groups on the x axis are ranked by increasing body size (Table 2). The study plot is located in the Nouragues Reserve, French Guiana.

Figure 4. Variation partitioning of OTU composition in each focus taxonomic group and across propagule size. (a) Variations in OTU composition are partitioned into pure (i.e. abiotic (A), plant (P) or spatial (S)) and shared components (A\cap P, P\cap S, A\cap S and A\cap P\cap S). See Table S6 for corresponding $R^2_{adj}$ values and their significance (for pure components and full models only, shared components not testable). (b) Relationship between propagule size and the amount of unexplained variance (i.e. residual $R^2_{adj}$, left y axis) by the RDA model (Pearson’s $r = 0.58, p = 0.009$) and Bray-Curtis dissimilarities obtained with the null model (right y axis; Pearson’s $r = 0.57, p = 0.01$). The study plot is located in the Nouragues Reserve, French Guiana.

This article is protected by copyright. All rights reserved.
This article is protected by copyright. All rights reserved.