Consistent patterns of high alpha and low beta diversity in tropical parasitic and free-living protists

Guillaume Lentendu[1] | Frédéric Mahé[1,2] | David Bass[3,4] | Sonja Rueckert[5] | Thorsten Stoeck[1] | Micah Dunthorn[1]

[1]Department of Ecology, University of Kaiserslautern, Erwin-Schrödinger-Straße, 67663 Kaiserslautern, Germany
[2]CIRAD, UMR LSTM, 34398 Montpellier, France
[3]Department of Life Sciences, The Natural History Museum London, Cromwell Road, London SW7 5BD, UK
[4]Centre for Environment, Fisheries & Aquaculture Science (Cefas), Barrack Road, The Nothe, Weymouth, Dorset DT4 8UB, UK
[5]School of Applied Sciences, Edinburgh Napier University, Edinburgh, EH11 4BN, Scotland, UK

Correspondence: Guillaume Lentendu, Department of Ecology, University of Kaiserslautern, Erwin-Schrödinger-Straße, 67663 Kaiserslautern, Germany. Email: lentendu@rhrk.uni-kl.de

Running title: Tropical protist biogeography

Funding information: Deutsche Forschungsgemeinschaft grants DU1319/1-1 to M.D.

Abstract

Tropical animals and plants are known to have high alpha diversity within forests, but low beta diversity between forests. By contrast, it is unknown if microbes inhabiting the same ecosystems exhibit similar biogeographic patterns. To evaluate the biogeographies of tropical protists, we used metabarcoding data of species sampled in the soils of three lowland Neotropical rainforests. Taxa-area and distance-decay relationships for three of the dominant protist taxa and their subtaxa were estimated at both the OTU- and phylogenetic-levels, with presence-absence and abundance based measures. These estimates were compared to null models. High local alpha and low regional beta diversity patterns were consistently found for both the parasitic Apicomplexa and the largely free-living Cercozoa and Ciliophora. Similar to animals and plants, the protists showed spatial structures between forests at the OTU- and phylogenetic-levels, and only at the phylogenetic level within forests. These results suggest that the biogeographies of macro- and micro-organismal eukaryotes in lowland Neotropical rainforests are partially structured by the same general processes. However, and unlike the animals and plants, the protist OTUs did not exhibit spatial structures within forests, which hinders our ability to estimate local and regional diversity of protists in tropical forests.

KEYWORDS

Apicomplexa, biogeography, Cercozoa, Ciliophora, Neotropics, phylogeny

1 | INTRODUCTION

The distributions of animal and plant species within and between forests has been a central focus in tropical biology research (Connell, 1978; Hubbell, 2001). Understanding these macro-organismal biogeographic patterns has enabled biologists to establish conservation strategies (Hubbell et al., 2008; Moritz et al., 2001; Turner & Corlett, 1996), to infer how diversity arose through the combination of historical, evolutionary and ecological processes (Chave et al., 2002; Myers et al., 2013; Rosindell et al., 2011), and to estimate the total diversity of different taxa (Basset et al., 2012; Novotny et al., 2007; Slik et al., 2015). One way of looking at these patterns is to evaluate taxa-area relationships (TAR), which measures the increase in richness as more area is sampled (Arrhenius, 1921; Drakare et al., 2006). Another way is to evaluate distance-decay relationships (DDR), which measures the decrease of community similarity in relation to distance (Morlon et al., 2008; Soininen et al., 2007). Together these patterns allow for comparing the biogeography of taxa that have different ecological functions and different dispersal modes (Novotny et al., 2007; Saito et al., 2015; Wetzel et al., 2012; Zinger et al., 2014).

Most studies inferring TAR and DDR of tropical macro-organisms have focused on species diversity using morphological observations, but a few have used molecular approaches. At local scales, TAR and DDR slopes of arthropods and plants were estimated to be steep, reflecting high alpha diversity and high turnover within forests (Basset et al., 2012; Condit et al., 2002; Hubbell, 2001; Novotny et al., 2007). By contrast, TAR and DDR slopes at regional scales were low, reflecting the low beta diversity of animals and plants between forests (Condit et al., 1996; Condit et al., 2002; Novotny et al., 2007; Plotkin et al., 2000). Fewer TAR and DDR studies of tropical macro-organisms have analyzed phylogenetic diversity, in which evolutionary relationships are considered, but these studies have likewise found high local alpha and low

regional beta diversity patterns (González-Caro et al., 2014; Swenson et al., 2012; Zhang et al., 2013). The phylogenetic-based TAR slopes were observed to be steeper within forests compared to species-based TAR slopes for mammals and trees, reflecting finer-scale biogeographic structures at the phylogenetic level (Mazel et al., 2014; Swenson et al., 2013).

In both species- and phylogenetic-based studies of tropical macro-organisms, assessing the non-random origin of the TAR and DDR patterns was used to unravel spatial vs. non-spatial ecological and evolutionary processes (Hardy, 2008; Kraft et al., 2011; Zhang et al., 2013). The main spatial process driving biogeography is dispersal limitation, and it is associated with observed non-null TAR and DRR slopes along a spatial gradient (Helmus & Ives, 2012; Nekola & White, 1999). Dispersal limitations were identified as the main cause of the low similarity between rainforest tree communities at the regional scale in South American and Cameroon (Condit et al., 2002; Hardy & Sonké, 2004), while herbivorous arthropods did not appear to be dispersal limited in lowland rainforests in Papua New Guinea (Novotny et al., 2007). The main non-spatial process driving biogeography is environmental filtering, and is associated with observed non-null TAR and DRR slopes along an environmental gradient (Helmus & Ives, 2012; Nekola & White, 1999). Such environmental filtering at the regional scale was proposed as a cause for phylogenetic clustering of hummingbirds in Ecuador (Graham et al., 2009), and in trees in French Guyana (Baraloto et al., 2012). Environmental filtering was also proposed to explain the local to regional steep phylogeny-based TAR slopes of Eurasian mammals due to a mosaic of biomes (Mazel et al., 2015).

There has been a debate whether microbial protists exhibit fundamentally different patterns of biogeography than animals and plants (Adl & Gupta, 2006; Fenchel & Finlay, 2006; Finlay, 2002; Foissner, 2006; Geisen et al., 2017). But more recent studies of either TAR or DDR

have shown that protists do exhibit biogeographical patterns in some environments (Bates et al., 2013; Fernández et al., 2016; Geisen et al., 2018; Lara et al., 2016; Venter et al., 2017; Wetzel et al., 2012). For example, using TAR on soil protists analyzed with metabarcoding, Venter et al. (2017) found significantly increasing slopes (0.12-0.18) within and between German temperate grasslands and forests. Using DDR, Wetzel et al. (2012) found using morphology that planktonic river diatoms had flatter slopes than benthic species, suggesting that dispersal limitations in the benthic species increases biogeographical structure, and Boenigk et al. (2018) found using metabarcoding that protist communities in lakes had flatter slopes across the Alps than within local regions not separated by the mountains. And using phylogenetic-based DDR on soil protists analyzed with metabarcoding, Bates et al. (2013) found significant slopes at the global scale, suggesting limited dispersal of entire clades of species, although other environmental parameters could better explain the community composition changes (e.g., climate moisture index). The question remain, though, what protistan taxa-area and distance-decay patterns look like in most ecosystems (especially in relatively little sampled tropical forests), and if the microbial eukaryotic patterns are similar to those seen in the macro-organisms inhabiting those same ecosystems.

Mahé et al. (2017) recently looked at the biodiversity of soil-inhabiting protists in lowland rainforests in Costa Rica, Panama, and Ecuador using general primers amplifying the V4 hyper-variable region of the 18S rRNA. They found that similar to macro-organisms, the microbial eukaryotes were exceedingly species rich and that there was low similarity in community composition between samples. Here we evaluated the taxa-area and distance-decay relationships of the three most abundant protist taxa in Mahé et al.'s (2017) data: the Apicomplexa, which are parasites of animals that disperse by vectors or cysts; and the Cercozoa

and Ciliophora, which are largely free-living bacterivores that disperse by active movements or cysts (Bass & Cavalier-Smith, 2004; Grattepanche et al., 2018; Lynn, 2008; Votýpka et al., 2017). These taxa together represented 85.9% of the protist reads and 74.2% of the protist OTUs amplified with the V4 primers, and they allowed for robust statistical comparisons at both the OTU- and phylogenetic-diversity levels. With these Neotropical metabarcoding data, we asked three main questions: 1) Do the protists exhibit biogeographic patterns that are consistent for taxa with different ecological functions and different dispersal modes?, 2) Are the protistan biogeographical patterns similar to those seen in tropical animals and plants?, and 3) How do these patterns affect our ability to estimate local and regional diversity of tropical protists? Uncovering these answers allowed us to infer general eukaryotic biogeographic patterns in the Neotropical rainforests for both macro- and microbial-eukaryotic organisms.

2 | MATERIALS AND METHODS

2.1 | Sampling and sequencing

All codes used here, all numbers used for the figures, and all sample coordinates can be found in HTML format (**File S1**). Details of sampling and sequencing can be found in Mahé et al. (2017). Briefly, soil samples were collected from La Selva Biological Station (Costa Rica) and (Barro Colorado Island (Panama) in October 2012 and June 2013, and from Tiputini Biodiversity Station (Ecuador) in October 2013 (**Figure S1**). Here we designate within forest samples as originating from the same field station. The soils were collected at the surface layer (after removal of leaf litter and stones) throughout each station in a variety of site conditions (except in areas excluded to humans). DNA was extracted with the PowerSoil DNA Elution Accessory Kit (MO BIO), and amplified with general eukaryotic primers that targeted the V4 hyper-variable region of the 18S rRNA (Stoeck et al., 2010). These primers amplify many protist groups , but not all (Adl et al., 2014; Forster et al., 2016; Hu et al., 2015; Lentendu et al., 2014; Logares et al., 2014; Stoeck et al., 2010). DNA amplified from every two consecutive sampling sites was combined in equal concentration to reduce costs, resulting in 144 operational samples. Sequencing was performed with Illumina MiSeq sequencing with v3 chemistry and software MCS v2.3.0.3 and RTA v1.18.42.0.

2.2 | Bioinformatics and statistics

Details of bioinformatic processing can be found in Mahé et al. (2017). Briefly, reads were pair-ended with PEAR v0.9.8 (Zhang et al., 2014) and filtered with filtered with Cutadapt v1.9 (Martin, 2011). Reads were then clustered into OTUs with Swarm v2.1.5 (Mahé et al., 2015b) using d=1 with the fastidious option on, which clusters amplicons by using a local clustering

threshold rather than relying on an arbitrary global threshold such as 97% similarity. Chimeras were identified and removed with VSEARCH v1.6.0 (Rognes et al., 2016). Low abundant OTUs were removed from the combined dataset only if they included ≤2 reads, and were found in only one sample, and were <99% similar to an accession in the taxonomically curated Protist Ribosomal Reference (PR^2) database v203 (Guillou et al., 2013). OTUs were also removed if they did not phylogenetically place within known eukaryotic clades using the Evolutionary Placement Algorithm (Berger et al., 2011) as implemented in RAxML v8.1.15 (Stamatakis, 2014). Reads and OTUs assigned by Mahé et al. (2017) to the main taxa of Apicomplexa (81.2% reads, 59.2% OTUs), Cercozoa (5.6% reads, 24.9% OTUs), and Ciliophora (3.4% reads, 9.2% OTUs) were used for this study (**Dataset S1**), using the PR^2 database, were used for this study. These three groups comprised the majority of the protist reads and OTUs in this dataset (Mahé et al., 2017). The taxonomy of Apicomplexa was modified to follow Rueckert et al. (2011): Colpodellidae not in the Apicomplexa, and *Cryptosporidium* in the Gregarinasina.

Statistical analyses here were conducted in R (R Core Team, 2016). To allow for statistical robustness, samples were analyzed if they gathered at least 80,000 Apicomplexa reads, or 1,000 Cercozoa or Ciliophora reads (100, 108 and 110 samples, respectively). Subtaxa within the Apicomplexa, Cercozoa, and Ciliophora were also analyzed if they had at least ten OTUs, and were found in at least five samples within each forest, and each sample had at least 100 reads.

To assess the main taxa and subtaxa biogeographies, OTU- and phylogeny-based alpha and beta diversity indexes were first calculated. Using the OTUs, richness for alpha diversity, as well as the Sørensen (based on presence-absence) and Bray-Curtis (based on relative abundance) dissimilarity indexes for beta diversity were estimated using the R package Vegan v2.4-3

(Oksanen et al., 2013). Using the phylogenies, phylogenetic species variability index (Helmus et al., 2007) for alpha diversity was estimated using the R package picante v1.6-2 (Kembel et al., 2010) and for beta diversity using the unweighted (based on presence-absence) and weighted (based on relative abundance) UniFrac distances (Lozupone & Knight, 2005) as implemented in phyloseq v1.20.0 (McMurdie & Holmes, 2013). Phylogenetic trees were inferred using multiple sequence alignments of OTU representatives generated with MAFFT v7.221 (Katoh & Standley, 2013) using the FFT-NS-i strategy and picking the highest likelihood tree inferred with RAxML v7.3.0 (Stamatakis, 2006) using the GTR+CAT model and hill-climbing algorithm. Phylogenetic variability was preferred to Faith's PD index because of its independence from OTU richness, making it more suitable for testing phylogenetic patterns independently from the amount of diversity covered (Helmus et al., 2007; Helmus & Ives, 2012).

In order to avoid any bias in alpha and beta diversity indexes due to unequal sequencing depth of the main taxa and subtaxa in the different samples, read counts in each sample were normalized to the sample with the lowest number of reads in each main taxa and subtaxa OTU matrix. The normalization was bootstrapped 1,000 times to produce robust statistical estimates of the diversity indexes and the subsequent taxa-area and distance-decay patterns.

To estimate taxa-area relationships (TAR), the double logarithmic generalization of Arrhenius' equation was applied on alpha diversity indexes (Arrhenius, 1921): $log(S_{obs}) = log(c) + z \times log(A)$; where S_{obs} is the species richness or phylogenetic variability, A the sampled area, c a constant and z the slope. Areas for the TAR were estimated in circle of increasing radius ranging from 0.5 to 5 km with a 0.5 km increment within forests with each sample being successively the circle center following Zinger et al. (2014). For area between forests, the same principle was applied using only the Panama samples as circle center, with additional radius

ranges between 475 and 484 km to cover Costa Rica samples and between 1,152 and 1,159 km for Ecuador samples, incrementing by 1 km inside each range. Only lowland Neotropical rainforests contained within each circle were considered to calculate the areas. For Costa Rica and Panama locations, tropical rainforest coverage was recovered from the 2012 MODIS Land Cover in native resolution (Channan et al., 2014; Friedl et al., 2010), with Panama areas being overlaid with the Barro Colorado Nature Monument Land boundaries. The regional lowland Neotropical rainforest areas were estimated by overlaying the "Evergreen Broadleaf forest" class of the 2012 MODIS Land Cover in 5' x 5' resolution with the "Tropical rain forest" class of the 2010 FAO's Global Ecological Zones (FAO 2015), keeping only the union of both layers clipped to the Neotropics. Areas were determined using the EASE-Grid equal area projection (Brodzik & Knowles, 2002), with the rgeos package (Bivand & Rundel, 2017).

To estimate distance-decay relationships (DDR), the double log transformation of the Nekola and White formula was applied on beta diversity indexes (Nekola & White, 1999): $log(S_{com}) = log(a) + \beta \times log(D)$; where S_{com} was the community or phylogenetic similarity (1 – dissimilarity index), D the geographic distance, a a constant and β the slope. Distances between each pair of sample were calculated using the great circle distance on WGS84 ellipsoid as implemented in the sp package v1.2-4 (Bivand et al., 2013).

Analyses of variance were used to test for significance of TAR and DDR linear regressions, and analyses of covariance were used to assess significant differences between linear regression slopes (i.e. between observed and null model slopes, see below). Regression statistics, analyses of variances, and covariances were calculated in each of the 1000 bootstraps, and were averaged, with p-value being corrected for multiple comparisons (Benjamini & Hochberg, 1995). In order to control for a potential temporal effect due to repeated sampling of same location at

two time points in Costa Rica and Panama, TAR and DDR were additionally calculated for each sampling campaign separately. This analysis was only conducted on Costa Rica samples which contain enough samples and reads (see previous thresholds) for each three phyla investigated at both dates to allow robust comparisons.

2.3 | Null models

Two null models were computed in order to test for the non-randomness of the OTU- and phylogenetic-diversity patterns, in which only the tested parameter was randomized while not changing all other structural features unrelated to the relevant null hypotheses tested (Hardy, 2008). Null model 1: the geographic positions of the samples were randomly shuffled inside each forest or among all three forests. This null model randomized each OTU's geographic range without modifying the OTU richness and community composition of single samples. This procedure erased potential OTU spatial structures and resulted in non-spatial TAR and DDR (Scheiner et al., 2011). The null hypothesis tested was that the sample's geographic position was independent from the community composition. Null model 1 TAR and DDR regressions were computed for OTU-based indexes. Null model 2: the OTUs were randomly shuffled along the tip of the phylogenetic tree within the entire regional OTU pool or for each forest OTU pools separately (Hardy, 2008; Kembel, 2009). This null model allowed the phylogenetic relatedness to vary independently from the OTU spatial range, OTU richness and community composition and is suitable to test for phylogenetic structuration in a geographical context (Hardy, 2008). Null model 2 TAR and DDR regressions were computed for phylogeny-based indexes.

3 | Results

3.1 | Taxonomic assignments

Although many protistan taxa were identified in the soils that were sampled throughout the three Neotropical rainforests and sequenced with a metabarcoding approach (Mahé et al., 2017), here we only evaluated the three most abundant groups: Apicomplexa, 35,134,460 reads and 12,247 OTUs; Cercozoa, 2,404,404 reads and 5,158 OTUs; and Ciliophora, 1,479,001 reads and 1,903 OTUs (**Figure 1**).

Within Apicomplexa, the Gregarinasina, including *Cryptosporidium*, was the dominant taxon (65.9% of the reads, 61.7% of the OTUs). Gregarines are obligate parasites infecting the intestines, coeloms and reproductive vesicles of invertebrates, while *Cryptosporidium* infects the intestines of vertebrates (Desportes & Schrével, 2013). Their oocysts and/or gamontocysts are expelled from the macro-organisms with the faeces and can be found in soils worldwide. Only 0.004% of the reads and 0.04% of the OTUs were assigned to the blood parasites in the Haemospororida.

The dominant group within Cercozoa was the sarcomonads (44.9% of reads, 38.2% of OTUs), which comprises the Cercomonadida and Glissomonadida. These two groups are generally bacterivorous or fungivorous flagellates (Bass et al., 2009), although one percent of the glissomonad reads were assigned to an algivorous viridiraptorid genus (Hess & Melkonian, 2013). Other detected algivorous cercozoans included the vampyrellid amoebae (Berney et al., 2013). Numerous testate filose amoebae were also found, most of which where the silica-shelled euglyphids and organic-shelled thecofiloseans, while only a few Endomyxa were detected despite including the plant pathogenic/symbiotic Plasmodiophorida which are often found in other soils (Neuhauser et al., 2014).

Within Ciliophora, Colpodea was the dominant taxon (67.2% of reads, 53.0% of the OTUs). Colpodeans are primarily bacterivores, but some are fungivorous, and have r-selected growth strategies that can quickly dominate the ciliate communities by ex-cysting from their resting stages (Lynn, 2008). The second and third most abundant ciliates groups were the largely bacterivorous Phyllopharyngea and Spirotrichea. Only one OTU found in Panama was assigned to Armophorea, which thrive in anoxic environments such as insect guts (Lynn, 2008).

3.2 | Taxa-area relationships

Using OTUs, the TAR slopes (*z*) of the main taxa within forests varied from 0.36 to 0.57 (**Figure 2**). These slopes were all significant ($p < 0.001$) with high R^2 values between 0.71 and 0.84, and none differed significantly from the slopes produced by the null model 1 (ANCOVA $p > 0.05$). TARs between forests (i.e., using all samples from all three forests) were much flatter with slopes from 0.1 to 0.11. These slopes were all significant with low to moderate R^2 values from 0.25 to 0.60, but all were significantly steeper than null model 1. Subtaxa (i.e., smaller taxa within Apicomplexa, Cercozoa, or Ciliophora) slopes were likewise all significant, with most slopes differing from null model 1 between forests (**Figures S2-S3**). OTU-based TAR slopes of the main taxa for the two sampling years in Costa Rica vary in the same range as the Costa Rica slopes of the full dataset (**Figure S5**). The observed slopes at the two sampling dates were consistently significant ($p < 0.001$) and did not differ from the null model (ANCOVA $p > 0.05$). The steep TAR slopes showed that as more area was sampled within a forest there was an increase in the number of observed protists giving rise to high alpha diversity. This increase in OTUs was not due to the increase in sampled area, but due to the accumulation of samples as there was no within-forest difference with the null model. When the area was expanded to

include multiple forests, the TAR slopes were less steep giving rise to low beta diversity. However, the steeper regional slopes compared to the null model showed that the increase in OTUs over larger area at the regional scale was not simply due to the accumulation of samples.

Using phylogenies of the OTUs, the TAR slopes of the main taxa within forests varied from -0.00048 to 0.06 and were mostly significant (**Figure 2**). The significant slopes had low R^2 values from 0.05 to 0.27, and they all differed from null model 2 except for Apicomplexa in Costa Rica and Ecuador. TARs between forests were much flatter with slopes from -0.001 to 0.0083. Only Cercozoa and Ciliophora had significant slopes that had low R^2 values from 0.07 to 0.10, but these were significantly steeper than null model 2. Subtaxa slopes were also mostly significantly and mostly differed significantly from null model 2 (**Figures S3-S4**). Phylogeny-based TAR slopes of the two sampling years in Costa Rica were significant and differed from null model only for Cercozoa and Ciliophora, just as for the full dataset (**Figure S5**). Although the June 2013 Ciliophora slope was the steepest (z = 0.08), it was not significant due to high variation among the 1000 bootstraps. Phylogeny-based TAR slopes had similar trends to the OTU-based TAR slopes, with most slope being stepper within than between forests, although additional non-random phylogenetic patterns were detected within forests. These patterns mean that as more area is sampled the more new clades will be encountered, highlighting a phylogenetic clustering in most of the main taxa and subtaxa within and between forests.

3.3 | Distance-decay relationships

Using OTUs, most DDR slopes (β) of the main taxa and subtaxa within forests were not significant using both presence-absence and abundance-based measures, and thus null models were not evaluated at that scale. DDR slopes of the main taxa between forests varied from -0.31

to -0.22 using presence-absence data, and -0.49 to -0.36 using abundance data (**Figure 3**). These slopes were all significant with low R^2 from 0.05 to 0.28, and all were significantly steeper than null model 1, as the latter resulted in non-significant (i.e., flat) slopes. Subtaxa slopes between forests were also all significant and all differed from null model 1 (**Figures S6-S7**). Most OTU-based DDR slopes of the main taxa at the two sampling dates in Costa Rica were not significant (**Figure S8**), similar to what was found with the full dataset (**Figure 3**). However, the October 2012 presence-absence based Apicomplexa and abundance-based Ciliophora have significant steepest negative slopes compared to the null model. This discrepancy with the full dataset patterns could have arose from the low number of samples covered at that date for the respective taxa (12 and 13). These OTU-based DDR slopes patterns showed that protist communities do have biogeographic patterns between forests but not within forests.

Using phylogenies of the OTUs, most DDR slopes of the main taxa and subtaxa within forests were likewise not significant using both presence-absence and abundance-based measures, and thus null models were also not evaluated at that scale. DDR slopes of the main taxa between forests varied from -0.08 to -0.03 for presence-absence, and -0.04 to -0.03 for abundance-based (**Figure 3**). These slopes were all significant with low R^2 from 0.04 to 0.16. Only the Apicomplexa for both measures and Ciliophora for presence-absence have significantly steeper DDR slopes than the null model 2. Subtaxa slopes between all samples were mostly significant, but only half differed from null model 2 (**Figure S7**). All phylogeny-based DDR slopes of the main taxa at the two sampling dates in Costa Rica were not significant (**Figure S8**), in congruence with the patterns observed for the full dataset (**Figure 3**). At the phylogenetic level, DDR slopes showed as well that protist communities do have biogeographic patterns between forests but not within forests.

To further explain the similarity between OTU- and phylogeny-based DDR patterns, we used Spearman's rank correlation coefficient (ρ) between both type of distance matrices. OTU- and phylogenetic-based measures of similarity showed moderate to high correlations within forests ($0.41 < \rho < 0.89$ for presence-absence, $0.29 < \rho < 0.84$ for abundance based measures, $p < 0.001$) as well as between forests ($0.47 < \rho < 0.66$ for presence-absence, and $0.5 < \rho < 0.59$ for abundance based measures) (**Table S1**). These results explain the similar trends between DDR slopes of the two kinds of similarity measures.

4 | Discussion

Together our analyses showed that protists in the three lowland Neotropical rainforests exhibit biogeographical patterns. Within forests, the steep TAR slopes of the OTUs reflected the high alpha diversity and high turnover locally. The absence of difference with the null model in TAR and the non-significant slopes in DDR analyses rejected the presence of spatially structured OTUs. The significant TAR of the phylogenies and their differences to the null model within forest, however, demonstrated that a spatial phylogenetic clustering occurred locally. Both of these results support environmental filtering as the dominant process and that microhabitats were non-randomly distributed at the local scale. Between forests, the low OTU- and phylogeny-based TAR slopes reflected the low beta diversity regionally. The non-random biogeographic structure of OTUs between forests and the non-random phylogenetic clustering of OTU belonging to the same forest as observed in both TAR and DDR analyses reveal that dispersal limitation is structuring protist communities at the regional scale.

Here we consistently found—for the first time—these biogeographic patterns at the OTU- and phylogenetic-diversity levels, using taxa-area and distance-decay relationships, with presence-absence and abundance-based measures, on three large protist taxa as well as their subtaxa. Previous studies aimed at analyzing the biogeographies of soil-inhabiting protists in other biomes only evaluated one or a few of these aspects (Bates et al., 2013; Boenigk et al., 2018; Fernández et al., 2016; Lara et al., 2016; Venter et al., 2017). Recent studies not based on taxa-area and distance-decay relationships have also identified environmental filtering as the main process affecting the composition of protist communities, with dispersal limitation affecting protists within similar habitats (Dupont et al., 2016; Fiore-Donno et al., 2016; Grossmann et al., 2016).

We evaluated two null hypotheses in order to reject the random origin of the observed biogeographic patterns of protists in the Neotropical rainforests. Low taxa-area slopes were previously used to argue that protists and other microbes lacked biogeographies (Azovsky, 2002; Finlay, 2002; Horner-Devine et al., 2004). We showed here, however, that there could be significant biogeographic structures even with low taxa-area and distance-decay slopes, as long as the observed slopes were steeper than the slopes produced by the null models in which either the sample locations or the phylogenetic placements were randomly shuffled. In particular, the null model that was applied to TAR in which the sample locations were shuffled, differentiated between a simple OTU accumulation with an increase in the number of samples added in a non-geographically meaningful order (i.e. non-spatial curve), with an OTU accumulation due to both increase in number of samples and increase in size of area covered. However, as the TAR slope is just a particular case of an accumulation curve, the steepness of the slope is not enough to assess the magnitude or significance of a geographic pattern.

These TAR and DDR patterns of the Neotropical rainforest protists were inferred with our sampling and molecular approaches. The soils were sampled in noncontiguous forests in three countries, but the areas were calculated following a nested sampling scheme of Scheiner (2003). Similar strategies of estimating biogeographic patterns from noncontiguous sampling were used in previous studies of aquatic and terrestrial macro- and micro-organisms (Bates et al., 2013; Boenigk et al., 2018; Martiny et al., 2011; Mazel et al., 2014; Novotny et al., 2007; Venter et al., 2017; Zinger et al., 2014). The phylogenetic-based indexes were inferred from the molecular data composed of sequences of the V4 hyper-variable region. While this short region has less phylogenetic signal than the full 18S rRNA locus, the V4 can indeed be used to infer protistan clades and has more phylogenetic signal than other short 18S regions used in

metabarcoding studies (Dunthorn et al., 2014; Mahé et al., 2015a; Pernice et al., 2013; Tragin et al., 2017). Similar use of V4 reads to infer phylogenetic patterns occurred in previous studies of protistan environmental diversity (del Campo et al., 2015; Egge et al., 2015; Heger et al., 2018; Logares et al., 2014; Mahé et al., 2017; Ward et al., 2018).

These biogeographical patterns of high alpha diversity within forests and low beta diversity between forests were consistent for both the parasitic Apicomplexa and the largely free-living Cercozoa and Ciliophora, even though they have different ecological functions and different dispersal modes. The apicomplexan patterns can be explained by them mirroring the patterns of their animal hosts: arthropod biogeographic patterns, at least for the herbivores, are themselves partially driven by their host plants (Basset et al., 2015; Novotny et al., 2007). Apicomplexan patterns should follow the animals if they are largely host-specific, although little is known of their biology in the Neotropics except for some Haemosporida, such as *Plasmodium* (Fecchio et al., 2017; Svensson-Coelho et al., 2016), which accounted for only a small fraction of Apicomplexa here. The cercozoan and ciliophoran patterns within forests suggest that their OTU diversity is driven by environmental filtering. The cercozoan and ciliophoran patterns can be explained by them reflecting the patchiness of their potential bacterial prey (Livermore & Jones, 2015; Ushio et al., 2010). However, cercozoan and ciliophoran phylogenetic diversities within forests were structured by spatially-related processes that are congruent with a decoupling of prey-predator phylogenetic similarity, as observed experimentally for selected cercozoan strains (Glücksman et al., 2010). Between forests, both of these free-living taxa showed spatial patterns of OTU diversity suggesting that their biogeographies are partially structured by dispersal limitations, while only Ciliophora exhibited

spatial phylogenetic pattern highlighting the influence of evolutionary history processes on their biogeographic structure.

These patterns of taxa-area and distance-decay relationships for the microbial protists are mostly similar to those found in tropical macro-organismic animals and plants (Basset et al., 2012; Condit et al., 2002; Novotny et al., 2007; Pitman et al., 1999): there was high alpha diversity within forests, but low beta diversity between forests. Our study therefore demonstrates that many of the same general processes partially structure the biogeography of macro- and micro-organismal eukaroytes in Neotropical lowland rainforests. Unlike for the animals and plants, we did not observe distance-decay patterns within forests at both the OTU and phylogenetic levels for the protists; this difference could be due to the extremely high diversity of protists at each sample location, or due to the sampling scale (such patterns could possibly be observed for protists at the cm^2 or m^2 scale). Similar results to our findings in Neotropical soil microbial protist communities were found in tropical plant communities inhabiting inselbergs, which exhibit significant phylogenetic clustering between micro-habitats, supporting environmental filtering at the local scale, while community similarity between similar microhabitats decreased with geographic distance, supporting dispersal limitation at the regional scale (Parmentier & Hardy, 2009).

These biogeographical observations at the OTU levels are pertinent to the problem of estimating the local and regional diversity of eukaryotes (Mora et al., 2011). As with estimating the diversity of tropical arthropods (Novotny et al., 2002; Novotny et al., 2007), the low regional turnover means that it is not simply a matter of discovering all protist species within a forest and then extrapolating that number by the total area of the Neotropics. And as with arthropods (Basset et al., 2012; Basset et al., 2015), the problem remaining for the protists is how to

effectively estimate their total diversity within a small area because of the low local similarity in OTUs between samples (i.e., how many samples must be taken to estimate the total number of protists inhabiting just the soils of a forest, as well as those that are inhabiting the tree canopies and other microhabitats), and how to determine the most appropriate scale to describe local biogeographic structures (i.e., how far apart should the samples be taken).

ACKNOWLEDGEMENTS

We thank Jordan Mayor and Tobias Siemensmeyer for help in collecting, and the editor and three anonymous reviewers for their constructive comments. Funding came from the Deutsche Forschungsgemeinschaft grants DU1319/1-1 to M.D.

DATA ACCESSIBILITY

All scripts and numbers to create the figure are provided in HTML format in File S1. Raw nucleotide sequences are available at the ENA Sequence Read Archive, BioProject PRJNA317860.

AUTHOR CONTRIBUTIONS

G.L. and M.D. designed the project. G.L., F.M., D.B., S.R., and M.D. performed the analyzes. G.L., F.M., D.B., S.R., T.S., and M.D. wrote the manuscript.

REFERENCES

Adl S., & Gupta V. V. S. R. (2006). Protists in soil ecology and forest nutrient cycling. *Can. J. For. Res.* 36, 1805-1817.

Adl S. M., Habura A., & Eglit Y. (2014). Amplification primers of SSU rDNA for soil protists. *Soil Biol. Biochem.* 69, 328-342.

Arrhenius O. (1921). Species and area. *J. Ecol.* 9, 95-99.

Azovsky A. I. (2002). Size-dependent species-area relationships in benthos: is the world more diverse for microbes? *Ecography* 25, 273-282.

Baraloto C., Hardy O. J., Paine C. E. T., Dexter K. G., Cruaud C., Dunning L. T., Gonzalez M.-A., Molino J.-F., Sabatier D., Savolainen V., & Chave J. (2012). Using functional traits and phylogenetic trees to examine the assembly of tropical tree communities. *J. Ecol.* 100, 690-701.

Bass D., & Cavalier-Smith T. (2004). Phylum-specific environmental DNA analysis reveals remarkably high global biodiversity of Cercozoa (Protozoa). *Int J Syst Evol Microbiol* 54, 2393-2404.

Bass D., Howe A. T., Mylnikov A. P., Vickerman K., Chao E. E., Edwards Smallbone J., Snell J., Cabral C., Jr., & Cavalier-Smith T. (2009). Phylogeny and classification of Cercomonadida (Protozoa, Cercozoa): *Cercomonas*, *Eocercomonas*, *Paracercomonas*, and *Cavernomonas* gen. nov. *Protist* 160, 483-521.

Basset Y., Cizek L., Cuénoud P., Didham R. K., Guilhaumon F., Missa O., Novotny V., Ødegaard F., Roslin T., Schmidl J., Tishechkin A. K., Winchester N. N., Roubik D. W., Aberlenc H.-P., Bail J., Barrios H., Bridle J. R., Castaño-Meneses G., Corbara B., Curletti G., Duarte da Rocha W., De Bakker D., Delabie J. H. C., Dejean A., Fagan L. L., Floren A., Kitching R. L., Medianero E., Miller S. E., Gama de Oliveira E., Orivel J., Pollet M., Rapp M., Ribeiro S. P., Roisin Y., Schmidt J. B., Sørensen L., & Leponce M. (2012). Arthropod diversity in a tropical forest. *Science* 338, 1481-1484.

Basset Y., Cizek L., Cuénoud P., Didham R. K., Novotny V., Ødegaard F., Roslin T., Tishechkin A. K., Schmidl J., Winchester N. N., Roubik D. W., Aberlenc H.-P., Bail J., Barrios H., Bridle J. R., Castaño-Meneses G., Corbara B., Curletti G., Duarte da Rocha W., De Bakker D., Delabie J. H. C., Dejean A., Fagan L. L., Floren A., Kitching R. L., Medianero E., Gama de Oliveira E., Orivel J., Pollet M., Rapp M., Ribeiro S. P., Roisin Y., Schmidt J. B., Sørensen L., Lewinsohn T. M., & Leponce M. (2015). Arthropod distribution in a tropical rainforest: tackling a four dimensional puzzle. *PLoS ONE* 10, e0144110.

Bates S. T., Clemente J. C., Flores G. E., Walters W. A., Parfrey L. W., Knight R., & Fierer N. (2013). Global biogeography of highly diverse protistan communities in soil. *ISME J.* 7, 652-659.

Benjamini Y., & Hochberg Y. (1995). Controlling false discovery rate - a practical and powerful approach. *J. R. Stat. Soc. Ser. B-Methodol.* 57, 289-300.

Berger S. A., Krompass D., & Stamatakis A. (2011). Performance, accuracy, and web server for evolutionary placement of short sequence reads under Maximum Likelihood. *Syst. Biol.* 60, 291-302.

Berney C., Romac S., Mahé F., Santini S., Siano R., & Bass D. (2013). Vampires in the oceans: predatory cercozoan amoebae in marine habitats. *ISME J.* 7, 2387-2399.

Bivand R. S., Pebesma E., & Gómez-Rubio V. (2013) *Applied spatial data analysis with R, 2nd ed.* Springer-Verlag, New York.

Bivand R. S., & Rundel C. (2017) rgeos: Interface to Geometry Engine - Open Source (GEOS), R package version 0.3-23, https://CRAN.R-project.org/package=rgeos.

Boenigk J., Wodniok S., Bock C., Beisser D., Hempel C., Grossmann L., Lange A., & Jensen M. (2018). Geographic distance and mountain ranges structure freshwater protist communities on a European scale. *Metabarcoding and Metagenomics* 2, e21519.

Brodzik M. J., & Knowles K. W. (2002) EASE-Grid: A Versatile Set of Equal-Area Projections and Grids. In: *Discrete Global Grids* (eds. Goodchild M, Kimerling AJ). National Center for Geographic Information & Analysis, Santa Barbara, CA.

Channan S., Collins K., & Emanuel W. R. (2014) Global mosaics of the standard MODIS land cover type data. University of Maryland and the Pacific Northwest National Laboratory, College Park. Maryland.

Chave J., Muller-Landau H. C., & Levin S. A. (2002). Comparing classical community models: theoretical consequences for patterns of diversity. *Amer. Nat.* 159, 1-23.

Condit R., Hubbell S. P., Lafrankie J. V., Sukumar R., Manokaran N., Foster R. B., & Ashton P. S. (1996). Species-area and species-Individual relationships for tropical trees: a comparison of three 50-ha plots. *J. Ecol.* 84, 549-562.

Condit R., Pitman N., Leigh E. G., Chave J., Terborgh J., Foster R. B., Núñez V. P., Aguilar S., Valencia R., Villa G., Muller-Landau H. C., Losos E., & Hubbell S. (2002). Beta-diversity in tropical forest trees. *Science* 295, 666-669.

Connell J. H. (1978). Diversity in tropical rain forests and coral reefs. *Science* 199, 1302-1310.

del Campo J., Mallo D., Massana R., de Vargas C., Richards T. A., & Ruiz-Trillo I. (2015). Diversity and distribution of unicellular opisthokonts along the European coast analyzed using high-throughput sequencing. *Environ. Microbiol.* 17, 3195-3207.

Desportes I., & Schrével J. (2013) *Treatise on Zoology—Anatomy, Taxonomy, Biology: The Gregarines (2 vols)* Brill, Leiden.

Drakare S., Lennon J. J., & Hillebrand H. (2006). The imprint of the geographical, evolutionary and ecological context on species-area relationships. *Ecol. Lett.* 9, 215-227.

Dunthorn M., Otto J., Berger S. A., Stamatakis A., Mahé F., Romac S., de Vargas C., Audic S., BioMarKs Consortium, Stock A., Kauff F., & Stoeck T. (2014). Placing environmental next-generation sequencing amplicons from microbial eukaryotes into a phylogenetic context. *Mol. Biol. Evol.* 31, 993-1009.

Dupont A., Griffiths R. I., Bell T., & Bass D. (2016). Differences in soil micro-eukaryotic communities over soil pH gradients are strongly driven by parasites and saprotrophs. *Environ. Microbiol.* 18, 2010-2024.

Egge E., Eikrem W., & Edvardsen B. (2015). Deep-branching novel lineages and high diversity of haptophytes in the Skagerrak (Norway) uncovered by 454 pyrosequencing. *J. Eukaryot. Microbiol.* 62, 121-140.

Fecchio A., Pinheiro R., Felix G., Faria I. P., Pinho J. B., Lacorte G. A., Braga E. M., Farias I. P., Aleixo A., Tkach V. V., Collins M. D., Bell J. A., & Weckstein J. D. (2017). Host community similarity and geography shape the diversity and distribution of haemosporidian parasites in Amazonian birds. *Ecography* doi:10.1111/ecog.03058.

Fenchel T., & Finlay B. J. (2006). The diversity of microbes: resurgence of the phenotype. *Phil. Trans. R. Soc. B* 361, 1965-1973.

Fernández L. D., Fournier B., Rivera R., Lara E., Mitchell E. A. D., & Hernández C. E. (2016). Water-energy balance, past ecological perturbations and evolutionary constraints shape the latitudinal diversity gradient of soil testate amoebae in south-western South America. *Glob. Ecol. Biogeogr.* 25, 1216-1227.

Finlay B. J. (2002). Global dispersal of free-living microbial eukaryote species. *Science* 296, 1061-1063.

Fiore-Donno A. M., Weinert J., Wubet T., & Bonkowski M. (2016). Metacommunity analysis of amoeboid protists in grassland soils. *Sci. Rep.* 6, 19068.

Foissner W. (2006). Biogeography and dispersal of micro-organisms: a review emphasizing protists. *Acta Protozool.* 45, 111-136.

Forster D., Dunthorn M., Mahé F., Dolan J. R., Audic S., Bass D., Bittner L., Boutte C., Christen R., Claverie J.-M., Decelle J., Edvardsen B., Egge E., Eikrem W., Gobet A., Kooistra W. H. C. F., Logares R., Massana R., Montresor M., Not F., Ogata H., Pawlowski J., Pernice M. C., Romac S., Shalchian-Tabrizi K., Simon N., Richards T. A., Santini S., Sarno D., Siano R., Vaulot D., Wincker P., Zingone A., de Vargas C., & Stoeck T. (2016). Benthic protists: the under-charted majority. *FEMS Microbiol. Ecol.* 92, fiw120.

Friedl M. A., Sulla-Menashe D., Tan B., Schneider A., Ramankutty N., Sibley A., & Huang X. (2010). MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sens. Environ.* 114, 168-182.

Geisen S., Mitchell E. A. D., Adl S., Bonkowski M., Dunthorn M., Ekelund F., Fernández L. D., Jousset A., Krashevska V., Singer D., Spiegel F. W., Walochnik J., & Lara E. (2018). Soil protists: a fertile frontier in soil biology research. *FEMS Microbiology Reviews* doi:10.1093/femsre/fuy006.

Geisen S., Mitchell E. A. D., Wilkinson D. M., Adl S., Bonkowski M., Brown M. W., Fiore-Donno A. M., Heger T. J., Jassey V. E. J., Krashevska V., Lahr D. J. G., Marcisz K., Mulot M., Payne R., Singer D., Anderson O. R., Charman D. J., Ekelund F., Griffiths B. S., Rønn R., Smirnov A., Bass D., Belbahri L., Berney C., Blandenier Q., Chatzinotas A., Clarholm M., Dunthorn M., Feest A., Fernández L. D., Foissner W., Fournier B., Gentekaki E., Hájek M., Helder J., Jousset A., Koller R., Kumar S., La Terza A., Lamentowicz M., Mazei Y., Santos S. S., Seppey C. V. W., Spiegel F. W., Walochnik J., Winding A., & Lara E. (2017). Soil protistology rebooted: 30 fundamental questions to start with. *Soil Biol. Biochem.* 111, 94-103.

Glücksman E., Bell T., Griffiths R. I., & Bass D. (2010). Closely related protist strains have different grazing impacts on natural bacterial communities. *Environ. Microbiol.* 12, 3105-3113.

González-Caro S., Umaña M. N., Álvarez E., Stevenson P. R., & Swenson N. G. (2014). Phylogenetic alpha and beta diversity in tropical tree assemblages along regional-scale environmental gradients in northwest South America. *J. Plant Ecol.* 7, 145-153.

Graham C. H., Parra J. L., Rahbek C., & McGuire J. A. (2009). Phylogenetic structure in tropical hummingbird communities. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19673-19678.

Grattepanche J.-D., Walker L. M., Ott B. M., Paim Pinto D. L., Delwiche C. F., Lane C. E., & Katz L. A. (2018). Microbial diversity in the eukaryotic SAR clade: illuminating the darkness between morphology and molecular data. *BioEssays* doi:10.1002/bies.201700198.

Grossmann L., Jensen M., Heider D., Jost S., Glucksman E., Hartikainen H., Mahamdallie S. S., Gardner M., Hoffmann D., Bass D., & Boenigk J. (2016). Protistan community analysis: key findings of a large-scale molecular sampling. *ISME J.* 10, 2269-2279.

Guillou L., Bachar D., Audic S., Bass D., Berney C., Bittner L., Boutte C., Burgaud G., de Vargas C., Decelle J., Del Campo J., Dolan J. R., Dunthorn M., Edvardsen B., Holzmann M., Kooistra W. H., Lara E., Le Bescot N., Logares R., Mahé F., Massana R., Montresor M., Morard R., Not F., Pawlowski J., Probert I., Sauvadet A. L., Siano R., Stoeck T., Vaulot D., Zimmermann P., & Christen R. (2013). The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res* 41, D597-604.

Hardy O. J. (2008). Testing the spatial phylogenetic structure of local communities: statistical performances of different null models and test statistics on a locally neutral community. *J. Ecol.* 96, 914-926.

Hardy O. J., & Sonké B. (2004). Spatial pattern analysis of tree species distribution in a tropical rain forest of Cameroon: assessing the role of limited dispersal and niche differentiation. *Forest Eco.l Manag.* 197, 191-202.

Heger T. J., Giesbrecht I. J. W., Gustavsen J., del Campo J., Kellogg C. T. E., Hoffman K. M., Lertzman K., Mohn W. W., & Keeling P. J. (2018). High-throughput environmental sequencing reveals high diversity of litter and moss associated protist communities along a gradient of drainage and tree productivity. *Environ. Microbiol.* doi:10.1111/1462-2920.14061.

Helmus Matthew R., Bland Thomas J., Williams Christopher K., & Ives Anthony R. (2007). Phylogenetic measures of biodiversity. *Amer. Nat.* 169, E68-E83.

Helmus M. R., & Ives A. R. (2012). Phylogenetic diversity–area curves. *Ecology* 93, S31-S43.

Hess S., & Melkonian M. (2013). The mystery of clade X: *Orciraptor* gen. nov. and *Viridiraptor* gen. nov. are highly specialised, algivorous amoeboflagellates (Glissomonadida, Cercozoa). *Protist* 164, 706-747.

Horner-Devine M. C., Lage M., Hughes J. B., & Bohannan B. J. M. (2004). A taxa-area relationship for bacteria. *Nature* 432, 750-753.

Hu S. K., Liu Z., Lie A. A. Y., Countway P. D., Kim D. Y., Jones A. C., Gast R. J., Cary S. C., Sherr E. B., Sherr B. F., & Caron D. A. (2015). Estimating protistan diversity using high-throughput sequencing. *J. Eukaryot. Microbiol.* 62, 688-693.

Hubbell S. P. (2001) *The unified neutral theory of biodiversity and biogeography* Princeton University Press, Princeton.

Hubbell S. P., He F., Condit R., Borda-de-Água L., Kellner J., & ter Steege H. (2008). How many tree species are there in the Amazon and how many of them will go extinct? *Proc. Natl. Acad. Sci. U.S.A.* 105, 11498-11504.

Katoh K., & Standley D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772-780.

Kembel S. W. (2009). Disentangling niche and neutral influences on community assembly: assessing the performance of community phylogenetic structure tests. *Ecol. Lett.* 12, 949-960.

Kembel S. W., Cowan P. D., Helmus M. R., Cornwell W. K., Morlon H., Ackerly D. D., Blomberg S. P., & Webb C. O. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26, 1463-1464.

Kraft N. J. B., Comita L. S., Chase J. M., Sanders N. J., Swenson N. G., Crist T. O., Stegen J. C., Vellend M., Boyle B., Anderson M. J., Cornell H. V., Davies K. F., Freestone A. L., Inouye B. D., Harrison S. P., & Myers J. A. (2011). Disentangling the drivers of β diversity along latitudinal and elevational gradients. *Science* 333, 1755-1758.

Lara E., Roussel-Delif L., Fournier B., Wilkinson D. M., & Mitchell E. A. D. (2016). Soil microorganisms behave like macroscopic organisms: patterns in the global distribution of soil euglyphid testate amoebae. *J. Biogeogr.* 43, 520-532.

Lentendu G., Wubet T., Chatzinotas A., Wilhelm C., Buscot F., & Schlegel M. (2014). Effects of long-term differential fertilization on eukaryotic microbial communities in an arable soil: a multiple barcoding approach. *Molec. Ecol.* 23, 3341-3355.

Livermore J. A., & Jones S. E. (2015). Local-global overlap in diversity informs mechanisms of bacterial biogeography. *ISME J.* 9, 2413.

Logares R., Audic S., Bass D., Bittner L., Boutte C., Christen R., Claverie J.-M., Decelle J., Dolan J., Dunthorn M., Edvardsen B., Gobet A., Kooistra W., Mahé F., Not F., Ogata H., Pawlowski J., Pernice M. C., Romac S., Shalchian-Tabrizi K., Simon N., Stoeck T., Santini S., Siano R., Wincker P., Zingone A., Richards T. A., de Vargas C., & Massana R. (2014). Patterns of rare and abundant marine microbial eukaryotes. *Curr. Biol.* 24, 813-821.

Lozupone C., & Knight R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* 71, 8228-8235.

Lynn D. H. (2008) *The ciliated protozoa: characterization, classification, and guide to the literature, 3rd edition* Springer, Dordrecht.

Mahé F., de Vargas C., Bass D., Czech L., Stamatakis A., Lara E., Singer D., Mayor J., Bunge J., Sernaker S., Siemensmeyer T., Trautmann I., Romac S., Berney C., Kozlov A., Mitchell E. A. D., Seppey C. V. W., Egge E., Lentendu G., Wirth R., Trueba G., & Dunthorn M. (2017). Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. *Nat. Ecol. Evol.* 1, 0091.

Mahé F., Mayor J., Bunge J., Chi J., Siemensmeyer T., Stoeck T., Wahl B., Paprotka T., Filker S., & Dunthorn M. (2015a). Comparing high-throughput platforms for sequencing the hyper-variable V4 region in environmental eukaryotic diversity surveys. *J. Eukaryot. Microbiol.* 62, 338-345.

Mahé F., Rognes T., Quince C., de Vargas C., & Dunthorn M. (2015b). Swarm v2: highly-scalable and high-resolution amplicon clustering. *PeerJ* 3, e1761.

Martin M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 1012.

Martiny J. B. H., Eisen J. A., Penn K., Allison S. D., & Horner-Devine M. C. (2011). Drivers of bacterial β-diversity depend on spatial scale. *Proc. Natl. Acad. Sci. U.S.A.* 108, 7850-7854.

Mazel F., Guilhaumon F., Mouquet N., Devictor V., Gravel D., Renaud J., Cianciaruso M. V., Loyola R., Diniz-Filho J. A. F., Mouillot D., & Thuiller W. (2014). Multifaceted

diversity–area relationships reveal global hotspots of mammalian species, trait and lineage diversity. *Glob. Ecol. Biogeogr.* 23, 836-847.

Mazel F., Guilhaumon F., Mouquet N., Devictor V., Gravel D., Renaud J., Cianciaruso M. V., Loyola R., Diniz-Filho J. A. F., Mouillot D., & Thuiller W. (2015). Mammalian phylogenetic diversity-area relationships at a continental scale. *Ecology* 96, 2814-2822.

McMurdie P. J., & Holmes S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* 8, e61217.

Mora C., Tittensor D. P., Adl S., Simpson A. G. B., & Worm B. (2011). How many species are there on earth and in the ocean? *Plos Biology* 9, e1001127.

Moritz C., Richardson K. S., Ferrier S., Monteith G. B., Stanisic J., Williams S. E., & Whiffin T. (2001). Biogeographical concordance and efficiency of taxon indicators for establishing conservation priority in a tropical rainforest biota. *Proc. R. Soc. B* 268, 1875-1881.

Morlon H., Chuyong G., Condit R., Hubbell S., Kenfack D., Thomas D., Valencia R., & Green J. C. (2008). A general framework for the distance-decay of similarity in ecological comunities. *Ecol. Lett.* 11, 904-917.

Myers J. A., Chase J. M., Jimnéz I., Jørgensen P. M., Araujo- Murakami A., Paniagua-Zambrana N., & Seidel R. (2013). Beta-diversity in temperate and tropical forests reflects dissimilar mechanisms of community assembly. *Ecol. Lett.* 16, 151-157.

Nekola J. C., & White P. S. (1999). The distance decay of similarity in biogeography and ecology. *J. Biogeog.* 26, 867-878.

Neuhauser S., Kirchmair M., Bulman S., & Bass D. (2014). Cross-kingdom host shifts of phytomyxid parasites. *BMC Evol. Biol.* 14, 33.

Novotny V., Basset Y., Miller S. E., Weiblen G. D., Bremer B., Cizek L., & Drozd P. (2002). Low host specificity of herbivorous insects in a tropical forest. *Nature* 416, 841-844.

Novotny V., Miller S. E., Hulcr J., Drew R. A. I., Basset Y., Janda M., Setliff G. P., Darrow K., Stewart A. J. A., Auga J., Isua B., Molem K., Manumbor M., Tamtiai E., Mogia M., & Weiblen G. D. (2007). Low beta diversity of herbivorous insects in tropical forests. *Nature* 448, 692-695.

Oksanen J., Blanchet F. G., Kindt R., Legendre P., Minchin P. R., O'Hara R. B., Simpson G. L., Solymos P., Stevens M. H. H., & Wagner H. (2013) Vegan: community ecology package. R package version 2.0-7.

Parmentier I., & Hardy O. J. (2009). The impact of ecological differentiation and dispersal limitation on species turnover and phylogenetic structure of inselberg’s plant communities. *Ecography* 32.

Pernice M. C., Logares R., Guillou L., & Massana R. (2013). General patterns of diversity in major marine microeukaryote lineages. *PLoS ONE* 8, e57170.

Pitman N. C. A., Terborgh J., Silman M. R., & Nuñez V P. (1999). Tree species distributions in an upper amazonian forest. *Ecology* 80, 2651-2661.

Plotkin J. B., Potts M. D., Yu D. W., Bunyavejchewin S., Condit R., Foster R., Hubbell S., LaFrankie J., Manokaran N., Seng L. H., Sukumar R., Nowak M. A., & Ashton P. S. (2000). Predicting species diversity in tropical forests. *Proc. Natl. Acad. Sci. U.S.A.* 97, 10850-10854.

R Core Team (2016) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rognes T., Flouri T., Nichols B., Quince C., & Mahé F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4, e2584

Rosindell J., Hubbell S. P., & Etienne R. S. (2011). The Unified Neutral Theory of Biodiversity and Biogeography at age ten. *Trends Ecol. Evol.* 26, 340-348.

Rueckert S., Simdyanov T. G., Aleoshin V. V., & Leander B. S. (2011). Identification of a divergent environmental DNA sequence clade using the phylogeny of gregarine parasites (Apicomplexa) from crustacean hosts. *PLoS ONE* 6, e18163.

Saito V. S., Soininen J., Fonseca-Gessner A. A., & Siqueira T. (2015). Dispersal traits drive the phylogenetic distance decay of similarity in Neotropical stream metacommunities. *J. Biogeogr.* 42, 2101-2111.

Scheiner S. M. (2003). Six types of species-area curves. *Glob. Ecol. Biogeogr.* 12, 441-447.

Scheiner S. M., Chiarucci A., Fox G. A., Helmus M. R., McGlinn D. J., & Willig M. R. (2011). The underpinnings of the relationship of species richness with space and time. *Ecol. Monogr.* 81, 195-213.

Slik J. W. F., Arroyo-Rodríguez V., Aiba S.-I., Alvarez-Loayza P., Alves L. F., Ashton P., Balvanera P., Bastian M. L., Bellingham P. J., van den Berg E., Bernacci L., da Conceição Bispo P., Blanc L., Böhning-Gaese K., Boeckx P., Bongers F., Boyle B., Bradford M., Brearley F. Q., Breuer-Ndoundou Hockemba M., Bunyavejchewin S., Calderado Leal Matos D., Castillo-Santiago M., Catharino E. L. M., Chai S.-L., Chen Y., Colwell R. K., Chazdon R. L., Clark C., Clark D. B., Clark D. A., Culmsee H., Damas K., Dattaraja H. S., Dauby G., Davidar P., DeWalt S. J., Doucet J.-L., Duque A., Durigan G., Eichhorn K. A. O., Eisenlohr P. V., Eler E., Ewango C., Farwig N., Feeley K. J., Ferreira L., Field R., de Oliveira Filho A. T., Fletcher C., Forshed O., Franco G., Fredriksson G., Gillespie T., Gillet J.-F., Amarnath G., Griffith D. M., Grogan J., Gunatilleke N., Harris D., Harrison R., Hector A., Homeier J., Imai N., Itoh A., Jansen P. A., Joly C. A., de Jong B. H. J., Kartawinata K., Kearsley E., Kelly D. L., Kenfack D., Kessler M., Kitayama K., Kooyman R., Larney E., Laumonier Y., Laurance S., Laurance W. F., Lawes M. J., Amaral I. L. d., Letcher S. G., Lindsell J., Lu X., Mansor A., Marjokorpi A., Martin E.

H., Meilby H., Melo F. P. L., Metcalfe D. J., Medjibe V. P., Metzger J. P., Millet J., Mohandass D., Montero J. C., de Morisson Valeriano M., Mugerwa B., Nagamasu H., Nilus R., Ochoa-Gaona S., Onrizal, Page N., Parolin P., Parren M., Parthasarathy N., Paudel E., Permana A., Piedade M. T. F., Pitman N. C. A., Poorter L., Poulsen A. D., Poulsen J., Powers J., Prasad R. C., Puyravaud J.-P., Razafimahaimodison J.-C., Reitsma J., dos Santos J. R., Roberto Spironello W., Romero-Saltos H., Rovero F., Rozak A. H., Ruokolainen K., Rutishauser E., Saiter F., Saner P., Santos B. A., Santos F., Sarker S. K., Satdichanh M., Schmitt C. B., Schöngart J., Schulze M., Suganuma M. S., Sheil D., da Silva Pinheiro E., Sist P., Stevart T., Sukumar R., Sun I.-F., Sunderland T., Suresh H. S., Suzuki E., Tabarelli M., Tang J., Targhetta N., Theilade I., Thomas D. W., Tchouto P., Hurtado J., Valencia R., van Valkenburg J. L. C. H., Van Do T., Vasquez R., Verbeeck H., Adekunle V., Vieira S. A., Webb C. O., Whitfeld T., Wich S. A., Williams J., Wittmann F., Wöll H., Yang X., Adou Yao C. Y., Yap S. L., Yoneda T., Zahawi R. A., Zakaria R., Zang R., de Assis R. L., Garcia Luize B., & Venticinque E. M. (2015). An estimate of the number of tropical tree species. *Proc. Natl. Acad. Sci. U.S.A.* 112, 7472-7477.

Soininen J., McDonald R., & Hillebrand H. (2007). The distance decay of similarity in ecological communities. *Ecography* 30, 3-12.

Stamatakis A. (2006). RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688-2690.

Stamatakis A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312-1313.

Stoeck T., Bass D., Nebel M., Christen R., Jones M. D. M., Breiner H. W., & Richards T. A. (2010). Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol. Ecol.* 19, 21-31.

Svensson-Coelho M., Loiselle B. A., Blake J. G., & Ricklefs R. E. (2016). Resource predictability and specialization in avian malaria parasites. *Mol. Ecol.* 25, 4377-4391.

Swenson N. G., Erickson D. L., Mi X., Bourg N. A., Forero-Montaña J., Ge X., Howe R., Lake J. K., Liu X., Ma K., Pei N., Thompson J., Uriarte M., Wolf A., Wright S. J., Ye W., Zhang J., Zimmerman J. K., & Kress W. J. (2012). Phylogenetic and functional alpha and beta diversity in temperate and tropical tree communities. *Ecology* 93, S112-S125.

Swenson N. G., Mi X., Kress W. J., Thompson J., Uriarte M., & Zimmerman J. K. (2013). Species-time-area and phylogenetic-time-area relationships in tropical tree communities. *Ecol. Evol.* 3, 1173-1183.

Tragin M., Zingone A., & Vaulot D. (2017). Comparison of coastal phytoplankton composition estimated from the V4 and V9 regions of 18S rRNA gene with a focus on photosynthetic groups and especially Chlorophyta. *Environ. Microbiol.* doi:10.1111/1462-2920.13952.

Turner I. M., & Corlett R. T. (1996). The conservation value of small, isolated fragments of lowland tropical rain forest. *Trends Ecol. Evol.* 11, 330-333.

Ushio M., Kitayama K., & Balser T. C. (2010). Tree species-mediated spatial patchiness of the composition of microbial community and physicochemical properties in the topsoils of a tropical montane forest. *Soil Biol. Biochem.* 42, 1588-1595.

Venter P. C., Nitsche F., Domonell A., Heger P., & Arndt H. (2017). The protistan microbiome of grassland soil: diversity in the mesoscale. *Protist* 168, 546-564.

Votýpka J., Modrý D., Oborník M., Šlapeta J., & Lukeš J. (2017) Apicomplexa. In: *Handbook of the protists* (eds. Archibald JM, Simpson AGB, Slamovits CH, *et al.*), pp. 1-58. Springer, Cham.

Ward G. M., Neuhauser S., Groben R., Ciaghi S., Berney C., Romac S., & Bass D. (2018). Environmental sequencing fills the gap between parasitic haplosporidians and free-living giant amoebae. *J. Eukaryot. Microbiol.* doi:10.1111/jeu.12501.

Wetzel C. E., Bicudo D. d. C., Ector L., Lobo E. A., Soininen J., Landeiro V. L., & Bini L. M. (2012). Distance decay of similarity in Neotropical diatom communities. *PLoS ONE* 7, e45071.

Zhang J., Kobert K., Flouri T., & Stamatakis A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30, 614-620.

Zhang J.-L., Swenson N. G., Chen S.-B., Liu X.-J., Li Z.-S., Huang J.-H., Mi X.-C., & Ma K.-P. (2013). Phylogenetic beta diversity in tropical forests: implications for the roles of geographical and environmental distance. *J. Syst. Evol.* 51, 71-85.

Zinger L., Boetius A., & Ramette A. (2014). Bacterial taxa–area and distance–decay relationships in marine environments. *Mol. Ecol.* 23, 954-964.

Figure 1 Taxonomic assignment [%] of the reads and OTUs of the parasitic Apicomplexa, and the free-living Cercozoa and Ciliophora, collected in tropical rainforests in Costa Rica, Panama and Ecuador.

Figure 2 Taxa-area relationships (TAR) within the three forests, and between all sample sites. Shaded areas are 95 % confidence intervals. Slopes were significant for both OTU richness and phylogenetic variability. Slopes were steeper within forest than between forests. Stars denote significantly steeper slopes compared to the null model assessed by analyses of co-variance. Null model 1 (random placement of sample locations) was used for OTU based TARs and null model 2 (shuffling of phylogenetic tree tips) for phylogenetic based TARs. Some of the null regression lines (dotted) are not visible due to exact overlay with their respective observed regression lines (solid).

Figure 3 Distance-decay relationships (DDR) for three protist groups using all samples of the three forests. Shaded areas are 95 % confidence intervals. Slopes were not significant within forests for both OTU richness and phylogenetic variability, and therefore are not shown. Stars denote significant steeper slopes compared to null model assessed by analyses of co-variance. Null model 1 (random placement of sample locations) was used for OTU based DDRs and null model 2 (shuffling of phylogenetic tree tips) for phylogenetic based DDRs.

Figure S1 Map of the sampled Neotropical region with inset into sampled forests in Costa Rica (La Selva Biological Station), Panama (Barro Colorado Island) and Ecuador (Tiputini Biodiversity Station). Red crosses are individual sampling points.

Figure S2 OTU richness based Taxa-Area Relationship (TAR) within forests for each subtaxon. None of the slopes significantly differed from null model 1. Some of the null regression lines (dotted) are not visible, as they overlap exactly with their respective observed regression lines (solid).

Figure S3 Taxa-Area Relationship (TAR) of subtaxa across forests. Stars denoted significant steepest slopes than the null model. Null model 1 (samples random placement) was used for OTU-based analyses and null model 2 (phylogenetic tree tips shuffling) was used for phylogeny-based analyses.

Figure S4 Phylogenetic variability based Taxa-Area Relationship (TAR) within forests for each subtaxa. Stars denoted significant steepest slopes compared to null model 2.

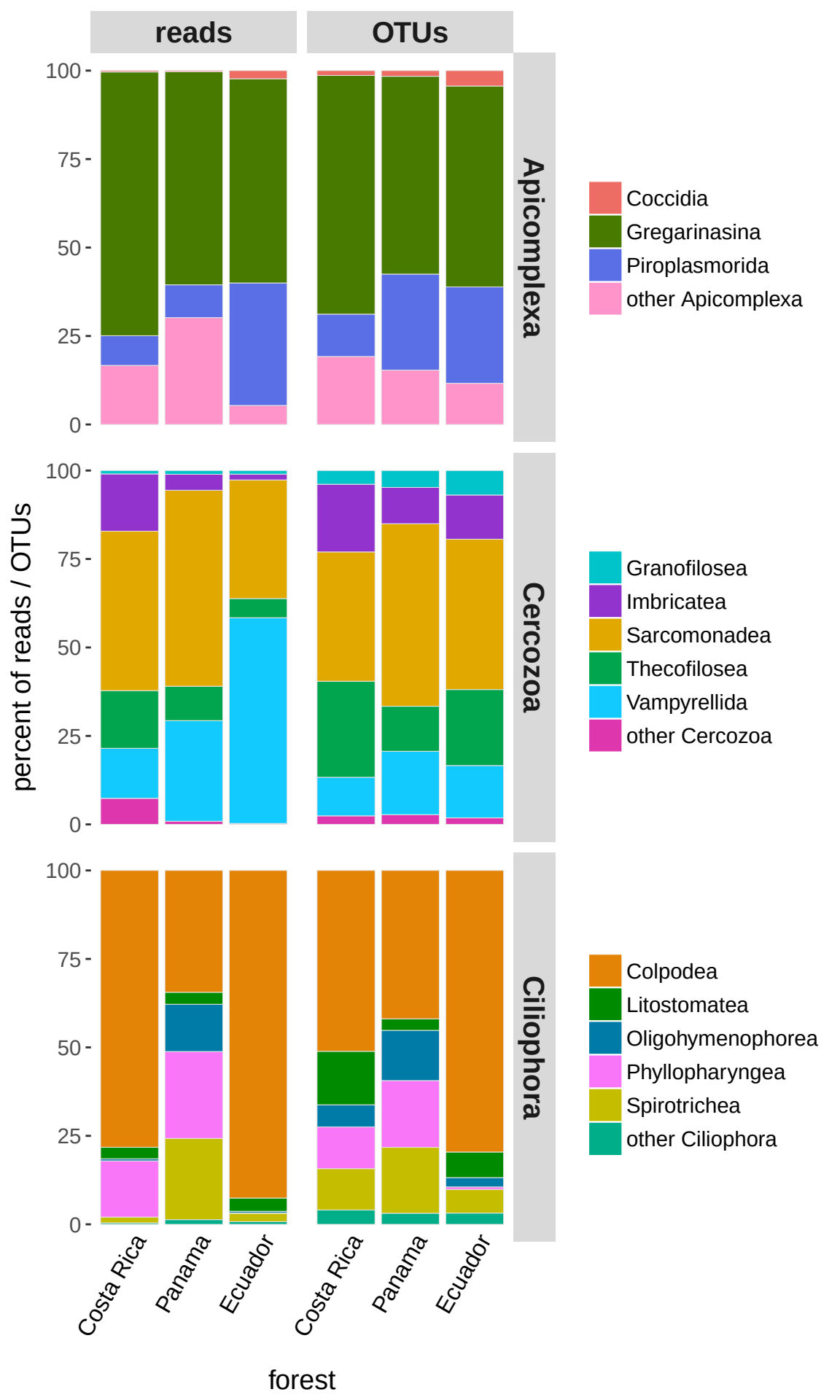
Figure S5 Taxa-Area Relationship (TAR) of each taxa in the Costa Rica forest (La Selva Biological Station) at the two sampling dates. Shaded areas are 95 % confidence intervals and stars denoted significant steepest slopes than the null model. Null model 1 (samples random placement) was used for OTU based analyses and null model 2 (phylogenetic tree tips shuffling) was used for phylogenetic based analyses.

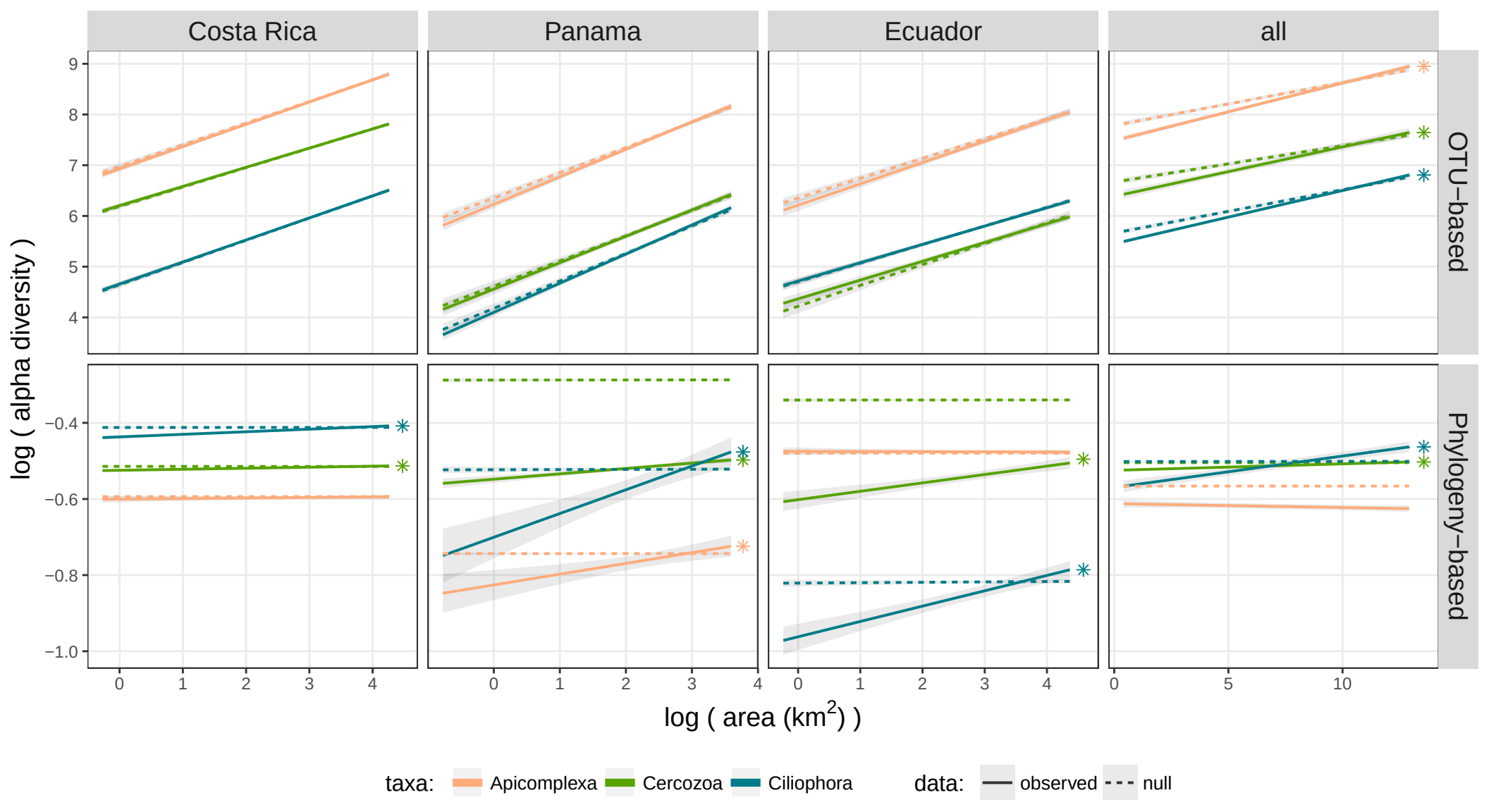
Figure S6 OTU relative abundance based Distance-Decay Relationship (DDR) of subtaxa within forests. All fitted linear regression lines were not significant.

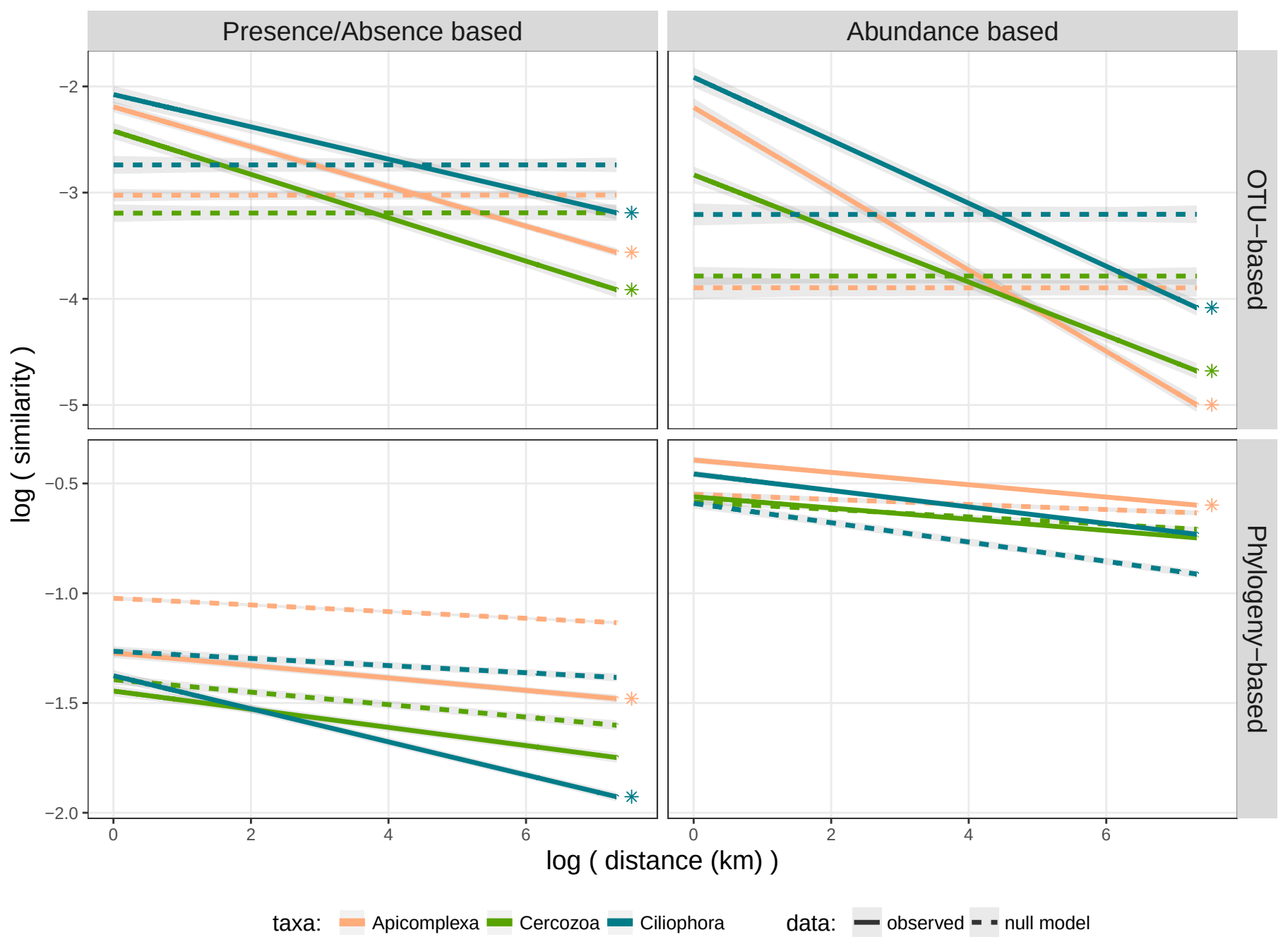
Figure S7 Distance-Decay Relationship (DDR) of subtaxa across forests. A value of 0.001 was added to similarity indexes to avoid infinite negative values in the log space. Stars denoted significant steeper slopes than null model. Subtaxa OTU based DDR slopes were all significantly negative and steeper than the null model 1. Only Thecofilosea and Vampyrellida had significant steeper phylogenetic based DDR slopes than in the null model 2 for both presence/absence and abundance based indexes. Piroplasmida had significant steeper phylogenetic based DDR slopes than in the null model 2 for the presence-absence based index while this was true for Colpodea and Gregarinasina only with the abundance based index.

Figure S8 Distance-Decay Relationship (DDR) of each taxa in the Costa Rica forest (La Selva Biological Station) at the two sampling dates. Shaded areas are 95 % confidence intervals and stars denoted significant steepest negative slopes than the null model. Null model 1 (samples random placement) was used for OTU based analyses and null model 2 (phylogenetic tree tips shuffling) was used for phylogenetic based analyses.

Table S1 Spearman's rho correlation coefficient calculated between OTU- and phylogeny-based community similarity measures. All correlations significant with $p < 0.001$ after Benjamini and Hochberg (1995) correction.


reads
OTUs
Apicomplexa
Cercozoa
Ciliophora
percent of reads / OTUs
100
75
50
25
0
Coccidia
Gregarinasina
Piroplasmorida
other Apicomplexa
Granofilosea
Imbricatea
Sarcomonadea
Thecofilosea
Vampyrellida
other Cercozoa
Colpodea
Litostomatea
Oligohymenophorea
Phyllopharyngea
Spirotrichea
other Ciliophora
Costa Rica
Panama
Ecuador
forest


Costa Rica
Panama
Ecuador
all
OTU−based
Phylogeny−based
log (alpha diversity)
log (area (km^2))
taxa:
Apicomplexa
Cercozoa
Ciliophora
data:
observed
null


Presence/Absence based
Abundance based
OTU−based
Phylogeny−based
log (similarity)
log (distance (km))
−2
−3
−4
−5
−0.5
−1.0
−1.5
−2.0
0
2
4
6
taxa:
Apicomplexa
Cercozoa
Ciliophora
data:
observed
null model