

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

Guillaume Lentendu^{a,1}, Frédéric Mahé^{a,b}, David Bass^{c,d}, Sonja Rueckert^e, Thorsten Stoeck^a, Micah Dunthorn^a

^aDepartment of Ecology, University of Kaiserslautern, Erwin-Schrödinger-Straße, 67663 Kaiserslautern, Germany

^bCIRAD, UMR LSTM, 34398 Montpellier, France

^cDepartment of Life Sciences, The Natural History Museum London, Cromwell Road, London SW7 5BD, UK

^dCentre for Environment, Fisheries & Aquaculture Science (Cefas), Barrack Road, The Nothe, Weymouth, Dorset DT4 8UB, UK

^eSchool of Applied Sciences, Edinburgh Napier University, Edinburgh, EH11 4BN, Scotland, UK

¹Email: lentendu@rhrk.uni-kl.de

Abstract

Animal and plant communities in tropical rainforests are known to have high alpha diversity within forests, but low beta diversity between forests. By contrast, it is unknown if the microbial protists inhabiting the same ecosystems exhibit similar biogeographic patterns. To evaluate the biogeographies of soil protists in three lowland Neotropical rainforests using metabarcoding data, we estimated taxa-area and distance-decay relationships for three large protist taxa and their subtaxa, at both the OTU and phylogenetic levels, with presence-absence and abundance based measures, and compared the estimates to null models. High local alpha and low regional beta diversity patterns were persistently found for both the parasitic Apicomplexa and the free-living Cercozoa and Ciliophora, even though they have different ecological functions and different dispersal modes. In comparison to the null models, both OTU and phylogenetic diversities showed spatial structures between forests, but only phylogenetic diversity showed spatial patterns 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. As with arthropods, the protists' high alpha diversity within forests presents problems for estimating their local diversity, and shows that regional diversity cannot be easily estimated because of low turnover between forests.

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 and 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, supporting a finer scale biogeographic structure 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 biogeographic patterns were used to unravel spatial vs. non-spatial ecological and evolutionary processes (Hardy 2008, Kraft et al. 2011, Zhang et al. 2013).

Recently it was shown that similar to macro-organisms, soil-inhabiting microbial protists were exceedingly species rich with low similarity among samples within and between lowland rainforests in Costa Rica, Panama, and Ecuador (Mahé et al. 2017). Here we evaluated the taxa-area and distance-decay relationships of this protist diversity from those soil samples at both the OTU- and phylogenetic-diversity levels to ask if they exhibit biogeographical patterns, and if those patterns are similar for taxa with different ecological functions and different dispersal modes. Uncovering these answers allowed us to infer general eukaryotic biogeographic patterns in the Neotropical rainforests for both the macro- and micro-organisms, and to gain insights on how protist distributions affect our ability to estimate total local and regional diversity in the tropics.

Methods

All codes used here, and all numbers used for the figures, can be found in HTML format (**File S1**). Details of sampling and sequencing can be found in Mahé et al. (Mahé et al. 2017). Briefly, soil samples were taken from: La Selva Biological Station, Costa Rica; Barro Colorado Island (BCI), Panama; and Tiputini Biodiversity Station, Ecuador. DNA from every two consecutive samples was combined in equal concentration to reduce costs, resulting in 144 sample sites. DNA was amplified for the V4 region of 18S rRNA with general eukaryotic V4 primers (2010), and sequenced with Illumina MiSeq. Reads were clustered into OTUs with Swarm v2.1.5 (Mahé et al. 2015) using $d=1$ with the fastidious option on. Chimeras were identified and removed with VSEARCH v1.6.0 (Rognes et al. 2016). Reads and OTUs assigned by Mahé et al. (Mahé et al. 2017) using the Protist Ribosomal Reference (PR²) database v203 (Guillou et al. 2013) to the Apicomplexa, Cercozoa, and Ciliophora were used for this study (**Dataset S1**). The taxonomy of Apicomplexa was modified to follow Rueckert et al. (Rueckert et al. 2011): Colpodellidae not in the Apicomplexa, and *Cryptosporidium* in the Gregarinasina.

Using the OTUs, richness as well as Sørensen (based on presence-absence) and Bray-Curtis (based on relative abundance) dissimilarity indexes were estimated using Vegan v2.4-3 (Oksanen et al. 2013). Using OTU representatives, multiple sequence alignments were generated with MAFFT v7.221 (Kato and Standley 2013) using the FFT-NS-i strategy. Maximum Likelihood trees were inferred with RAxML v7.3.0 (Stamatakis 2006) using the GTR+CAT model and hill-climbing algorithm. The phylogenetic species variability index (Helmus et al. 2007) for alpha diversity was estimated using 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 and Knight 2005) as implemented in phyloseq v1.20.0 (McMurdie and Holmes 2013). Phylogenetic variability was preferred to Faith's PD index because of its independence from OTU richness, making it more suitable for testing for phylogenetic patterns independently from the amount of diversity covered (Helmus et al. 2007, Helmus and Ives 2012).

All diversity indexes were estimated based on 1000 randomly rarefied OTU tables normalized to the sample with the lowest number of reads. To allow for statistical robustness, samples were analyzed if they gather 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. Taxa-area relationships (TAR) were estimated using the double logarithmic generalization of Arrhenius' equation (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. Distance-decay relationships (DDR) were estimated using the double log transformation of the Nekola and White formula (Nekola and 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. Analyses of variance were used to test significance of linear regressions and significant difference between linear regression slopes (i.e. between observed and null model slopes, see below) were assessed using analysis of covariance. Regression statistics were calculated in each 1000 rarefied matrices and averaged, with p-value being corrected for multiple comparisons (Benjamini and Hochberg 1995).

Areas for the TAR were estimated in circle of increasing radius (Zinger et al. 2014) ranging from 0.5 to 5 km with a 0.5 km increment within forests with each sample being successively the circle center. 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 1152 and 1159 km for Ecuador samples, incrementing by 1 km inside each range. Only lowland Neotropical rainforests contained within each circle was 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 and Knowles 2002). All analyses were conducted in R (R Core Team 2016).

Two null models were constructed in which only the tested parameter was randomized while keeping 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 based on OTU richness and community similarity 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 based on phylogenetic variability and phylogenetic similarity indexes.

Results

Taxonomic assignments. Although many protist 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 (**Fig. 1**). These OTUs can be viewed as estimates of species diversity in molecular environmental studies of microbes (Bik et al. 2012, de Vargas et al. 2015, Mahé et al. 2017).

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 and Schrével 2013). Their oocysts and/or gamontocysts are expelled from the macro-organisms with the faeces and can be found in soils worldwide. In *Cryptosporidium*, for example, there are 26 valid species (Ryan et al. 2014), yet we found 4,500 OTUs. The blood parasites in the Haemospororida accounted for only 0.004% of the reads and 0.04% of the OTUs.

The dominant group within Cercozoa was the sarcomonads (44.9% of reads, 38.2% of OTUs), which is comprised of 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 and Melkonian 2013). Other detected algivorous cercozoans included the vampyrellid amoebae (Berney et al. 2013). Numerous testate filose amoebae were also found, most of which were the silica-shelled euglyphids and organic-shelled thecofiloseans, but 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). Although there are fewer than 250 described colpodean species, we found 1,008 OTUs assigned to this taxon. 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).

Null models. Two null models were computed in order to test for the non-randomness of the below OTU- and phylogenetic-diversity patterns. The null model 1 randomized the sample's geographic locations inside the considered scale (forest or regional) and was applied to OTU-based diversity measures; this null model allowed the OTU richness accumulation and the OTU composition turnover to change without any spatial constraint (Scheiner et al. 2011). The null model 2 randomized the OTUs along the tips of the phylogenetic trees at the considered scale and was applied to phylogenetic-based diversity measures; this null model allowed phylogenetic relatedness to vary independently from community composition in order to test for sample's geographic position on phylogenetic diversity only (Hardy 2008) .

Taxa-area relationships. Using OTUs, the TAR slopes (z) within forests varied from 0.36 to 0.57 (**Fig. 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 when using all samples (**Figs. S2-S3**).

Using phylogenies of the OTUs, the TAR slopes within forests varied from -0.00048 to 0.06 and were mostly significant (**Fig. 2**). The significant slopes had low R^2 values from 0.05 to 0.27, and they all differed from null model 2 except in Costa Rica. 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 significant and mostly differed significantly from null model 2 (**Figs. S3-S4**).

Distance-decay relationships. Using OTUs, most DDR slopes (β) of protist taxa and subtaxa within forests were not significant using both presence-absence and abundance-based measures,

and thus a null model was not evaluated at that scale. DDR slopes of protist taxa between forests varied from -0.31 to -0.22 using presence-absence data, and -0.49 to -0.36 using abundance data (**Fig. 3**). These slopes were all significant with low R^2 ranging 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 all significant and all differed from null model 1 (**Fig. S5-S6**).

Using phylogenies of the OTUs, most DDR slopes of protist taxa and subtaxa within forests were likewise not significant using both presence-absence and abundance-based measures, and thus a null model was also not evaluated at that scale. DDR slopes of protist taxa between forests varied from -0.08 to -0.03 for presence-absence, and -0.04 to -0.03 for abundance-based (**Fig. 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 (**Fig. S6**).

OTU- and phylogenetic-measures of similarity showed moderate to high correlations within forests (Spearman's rho from 0.41 to 0.89 for presence-absence, and 0.29 to 0.84 for abundance based measures, $p < 0.001$) as well as when using all samples (from 0.47 to 0.66 for presence-absence, and 0.5 to 0.59 for abundance based measures) (**Table S1**). These results explain the similar trends between DDR slopes of the two kinds of similarity measures.

Discussion

At the OTU level, 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 accumulation of area *per se*, but due to the accumulation of samples as there was no local difference with the null model; this pattern suggests that non-spatial ecological processes such as environmental filtering drive the within-forest protist diversity. When the area was expanded to include multiple forests the TAR slopes were less steep giving rise to low beta diversity. The steeper regional slopes compared to the null model suggests that spatial ecological processes such as dispersal limitations drive the between-forest protist diversity. The DDR slopes showed that as more samples were taken within a forest there was no significant decrease in community similarity related to distance, but there was a significant non-random decrease when samples were compared between forests. The distance pattern within forests means that no matter how far the samples are from each other they will have few OTUs shared between them, and demonstrates again that protist diversity at the forest scale is not spatially structured. The even lower community similarity between forests also supports the spatial structure of protist community at regional scale.

At the phylogenetic level, TAR and DDR slopes were similar to the OTU level within and between forests, although additional non-random phylogenetic patterns were detected within forests. The area patterns mean that the more area you sample the more new clades you encounter, highlighting a phylogenetic over-clustering of protists at local and regional scales. The absence of geographic structure in phylogenetic similarity within forests reflected the absence of geographic structure in OTU community similarity, as both type of measures were significantly correlated. The distance patterns between forests suggest that, like OTU diversity, phylogenetic diversity is geographically structured at the regional scale. This finding that additional forests contained new OTUs as well as new clades of protists is applicable for the conservation of microbial diversity (Cotterill et al. 2013, Gómez and Nichols 2013, Griffith 2012), as the destruction of unconserved tropical forests would lead not only to the loss of macro-organisms but also to the possible loss of protist clades.

These TAR and DDR patterns of the protist 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 (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, Martiny et al. 2011, Mazel et al. 2014, Novotny et al. 2007, Venter et al. 2017, Zinger et al. 2014). The molecular data were clustered into OTUs with Swarm, which can be more accurate than other clustering methods that rely on global clustering thresholds (Forster et al. 2016, Mahé et al. 2015). Like any clustering method, Swarm may nevertheless over-cluster environmental molecular data (i.e., produce too few OTUs) if there is no or low variability between species (Mahé et al. 2017). Swarm may also under-cluster (i.e., produce too many OTUs) if the sequencing depth was too shallow (Mahé et al. 2015), or if there is not concerted evolution among multiple genomic copies of the targeted gene region, but the extent of such intra-cellular polymorphisms is debated (Brown et al. 2015, Decelle et al. 2014).

Together our sampling and molecular data did show that protists in the lowland Neotropical rainforests exhibit biogeographical patterns. Here we persistently found—for the first time—these 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 protists inhabiting temperate soils only evaluated one or a few of these aspects (Bahram et al. 2016, Bates et al. 2013, Fernández et al. 2016, Lara et al. 2016, Venter et al. 2017). We additionally evaluated two null hypotheses in order to reject the random origin of the observed biogeographic patterns. 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 slopes were steeper than the slopes produced by the null models in which either the sample locations or the phylogenetic placements were randomly shuffled.

The biogeographical patterns of high alpha diversity within forests and low beta diversity between forests were persistent for both the parasitic Apicomplexa and the 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 Haemospororida, 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, while their phylogenetic diversity is structured by spatially-related processes, at least partially; similar patterns were also observed for Neotropical trees within forests (Swenson et al. 2013). 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 were likewise 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 eukaryotes in Neotropical lowland rainforests. Unlike the animals and plants, though, for the protists we did not observe distance-decay patterns within forests; 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 cm² or m² scale). 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), this 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 in just the soils of a forest, as well as those that are inhabiting the tree canopy and other microhabitats), as well as determining the most appropriate scale to describe local biogeographic structures.

ACKNOWLEDGMENTS

We thank Jordan Mayor and Tobias Siemensemeyer for help in collecting. Funding came from the Deutsche Forschungsgemeinschaft grants DU1319/1-1 and DU1319/1-2 to M.D.

REFERENCES

- Arrhenius (1921). Species and area. *J. Ecol.* 9:95-99.
- Azovsky (2002). Size-dependent species-area relationships in benthos: is the world more diverse for microbes? *Ecography* 25:273-282.
- Bahram, Kohout, Anslan, Harend, Abarenkov, Tedersoo (2016). Stochastic distribution of small soil eukaryotes resulting from high dispersal and drift in a local environment. *ISME J.* 10:885-896.
- Bass, Howe, Mylnikov, Vickerman, Chao, Edwards Smallbone, Snell, Cabral, Cavalier-Smith (2009). Phylogeny and classification of Cercomonadida (Protozoa, Cercozoa): *Cercomonas*, *Eocercomonas*, *Paracercomonas*, and *Cavernomonas* gen. nov. *Protist* 160:483-521.
- Basset, Cizek, Cuénoud, Didham, Guilhaumon, Missa, Novotny, Ødegaard, Roslin, Schmidl, Tishechkin, Winchester, Roubik, Aberlenc, Bail, Barrios, Bridle, Castaño-Meneses, Corbara, Curletti, Duarte da Rocha, De Bakker, Delabie, Dejean, Fagan, Floren, Kitching, Medianero, Miller, Gama de Oliveira, Orivel, Pollet, Rapp, Ribeiro, Roisin, Schmidt, Sørensen, Leponce (2012). Arthropod diversity in a tropical forest. *Science* 338:1481-1484.
- Basset, Cizek, Cuénoud, Didham, Novotny, Ødegaard, Roslin, Tishechkin, Schmidl, Winchester, Roubik, Aberlenc, Bail, Barrios, Bridle, Castaño-Meneses, Corbara, Curletti, Duarte da Rocha, De Bakker, Delabie, Dejean, Fagan, Floren, Kitching, Medianero, Gama de Oliveira, Orivel, Pollet, Rapp, Ribeiro, Roisin, Schmidt, Sørensen, Lewinsohn, Leponce (2015). Arthropod distribution in a tropical rainforest: tackling a four dimensional puzzle. *PLoS ONE* 10:e0144110.
- Bates, Clemente, Flores, Walters, Parfrey, Knight, Fierer (2013). Global biogeography of highly diverse protistan communities in soil. *ISME J.* 7:652-659.
- Benjamini, Hochberg (1995). Controlling false discovery rate - a practical and powerful approach. *J. R. Stat. Soc. Ser. B-Methodol.* 57:289-300.
- Berney, Romac, Mahé, Santini, Siano, Bass (2013). Vampires in the oceans: predatory cercozoan amoebae in marine habitats. *ISME J.* 7:2387-2399.
- Bik, Porazinska, Creer, Caporaso, Knight, Thomas (2012). Sequencing our way towards understanding global eukaryotic biodiversity. *Trends Ecol. Evol.* 27:233-243.

- Brodzick, Knowles 2002. EASE-Grid: A Versatile Set of Equal-Area Projections and Grids. In: Goodchild M, Kimerling AJ editors. *Discrete Global Grids*. Santa Barbara, CA, National Center for Geographic Information & Analysis.
- Brown, Chain, Crease, MacIsaac, Cristescu (2015). Divergence thresholds and divergent biodiversity estimates: can metabarcoding reliably describe zooplankton communities? *Ecol. Evol.* 5:2234-2251.
- Channan, Collins, Emanuel. 2014. Global mosaics of the standard MODIS land cover type data. College Park, Maryland, University of Maryland and the Pacific Northwest National Laboratory.
- Chave, Muller-Landau, Levin (2002). Comparing classical community models: theoretical consequences for patterns of diversity. *Amer. Nat.* 159:1-23.
- Condit, Hubbell, Lafrankie, Sukumar, Manokaran, Foster, Ashton (1996). Species-area and species-Individual relationships for tropical trees: a comparison of three 50-ha plots. *J. Ecol.* 84:549-562.
- Condit, Pitman, Leigh, Chave, Terborgh, Foster, Núñez V., Aguilar, Valencia, Villa, Muller-Landau, Losos, Hubbell (2002). Beta-diversity in tropical forest trees. *Science* 295:666-669.
- Connell (1978). Diversity in tropical rain forests and coral reefs. *Science* 199:1302-1310.
- Cotterill, Augustin, Medicus, Foissner (2013). Conservation of protists: the Krauthügel Pond in Austria. *Diversity* 5:374-392.
- de Vargas, Audic, Henry, Decelle, Mahé, Logares, Lara, Berney, Le Bescot, Probert, Carmichael, Poulain, Romac, Colin, Aury, Bittner, Chaffron, Dunthorn, Engelen, Flegontova, Guidi, Horák, Jaillon, Lima-Mendez, Lukeš, Malviya, Morard, Mulot, Scalco, Siano, Vincent, Zingone, Dimier, Picheral, Searson, Kandels-Lewis, Coordinators, Acinas, Bork, Bowler, Gorsky, Grimsley, Hingamp, Iudicone, Not, Ogata, Pesant, Raes, Sieracki, Speich, Stemann, Sunagawa, Weissenbach, Wincker, Karsenti (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science* 348:1261605.
- Decelle, Romac, Sasaki, Not, Mahé (2014). Intracellular diversity of the V4 and V9 regions of the 18S rRNA in marine protists (Radiolarians) assessed by high-throughput sequencing. *PLoS ONE* 9:e104297.
- Desportes, Schrével (2013). *Treatise on Zoology—Anatomy, Taxonomy, Biology: The Gregarines (2 vols)*. Leiden, Brill.
- Drakare, Lennon, Hillebrand (2006). The imprint of the geographical, evolutionary and ecological context on species-area relationships. *Ecol. Lett.* 9:215-227.
- Fecchio, Pinheiro, Felix, Faria, Pinho, Lacorte, Braga, Farias, Aleixo, Tkach, Collins, Bell, Weckstein (2017). Host community similarity and geography shape the diversity and

distribution of haemosporidian parasites in Amazonian birds. *Ecography*
doi:10.1111/ecog.03058.

- Fernández, Fournier, Rivera, Lara, Mitchell, Hernández (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 (2002). Global dispersal of free-living microbial eukaryote species. *Science* 296:1061-1063.
- Forster, Dunthorn, Stoeck, Mahé (2016). Comparison of three clustering approaches for detecting novel environmental microbial diversity. *PeerJ* 4:e1692.
- Friedl, Sulla-Menashe, Tan, Schneider, Ramankutty, Sibley, Huang (2010). MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sens. Environ.* 114:168-182.
- Gómez, Nichols (2013). Neglected wild life: parasitic biodiversity as a conservation target. *Int. J. Parasitol. Parasites Wildl.* 2:222-227.
- González-Caro, Umaña, Álvarez, Stevenson, Swenson (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.
- Griffith (2012). Do we need a global strategy for microbial conservation? *Trends Ecol. Evol.* 27:1-2.
- Guillou, Bachar, Audic, Bass, Berney, Bittner, Boutte, Burgaud, de Vargas, Decelle, Del Campo, Dolan, Dunthorn, Edvardsen, Holzmann, Kooistra, Lara, Le Bescot, Logares, Mahé, Massana, Montresor, Morard, Not, Pawlowski, Probert, Sauvadet, Siano, Stoeck, Vaultot, Zimmermann, Christen (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 (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.
- Helmus, Bland, Williams, Ives (2007). Phylogenetic measures of biodiversity. *Amer. Nat.* 169:E68-E83.
- Helmus, Ives (2012). Phylogenetic diversity–area curves. *Ecology* 93:S31-S43.
- Hess, Melkonian (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, Lage, Hughes, Bohannan (2004). A taxa-area relationship for bacteria. *Nature* 432:750-753.
- Hubbell (2001). *The unified neutral theory of biodiversity and biogeography*. Princeton, Princeton University Press.
- Hubbell, He, Condit, Borda-de-Água, Kellner, ter Steege (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, Standley (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30:772-780.
- Kembel (2009). Disentangling niche and neutral influences on community assembly: assessing the performance of community phylogenetic structure tests. *Ecol. Lett.* 12:949-960.
- Kembel, Cowan, Helmus, Cornwell, Morlon, Ackerly, Blomberg, Webb (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26:1463-1464.
- Kraft, Comita, Chase, Sanders, Swenson, Crist, Stegen, Vellend, Boyle, Anderson, Cornell, Davies, Freestone, Inouye, Harrison, Myers (2011). Disentangling the drivers of β diversity along latitudinal and elevational gradients. *Science* 333:1755-1758.
- Lara, Roussel-Delif, Fournier, Wilkinson, Mitchell (2016). Soil microorganisms behave like macroscopic organisms: patterns in the global distribution of soil euglyphid testate amoebae. *J. Biogeogr.* 43:520-532.
- Lozupone, Knight (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* 71:8228-8235.
- Lynn (2008). *The ciliated protozoa: characterization, classification, and guide to the literature, 3rd edition*. Dordrecht, Springer.
- Mahé, Rognes, Quince, de Vargas, Dunthorn (2015). Swarm v2: highly-scalable and high-resolution amplicon clustering. *PeerJ* 3:e1761.
- Mahé, de Vargas, Bass, Czech, Stamatakis, Lara, Singer, Mayor, Bunge, Sernaker, Siemensmeyer, Trautmann, Romac, Berney, Kozlov, Mitchell, Seppey, Egge, Lentendu, Wirth, Trueba, Dunthorn (2017). Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. *Nat. Ecol. Evol.* 1:0091.
- Martiny, Eisen, Penn, Allison, Horner-Devine (2011). Drivers of bacterial β -diversity depend on spatial scale. *Proc. Natl. Acad. Sci. U.S.A.* 108:7850-7854.
- Mazel, Guilhaumon, Mouquet, Devictor, Gravel, Renaud, Cianciaruso, Loyola, Diniz-Filho, Mouillot, Thuiller (2014). Multifaceted diversity–area relationships reveal global hotspots of mammalian species, trait and lineage diversity. *Glob. Ecol. Biogeogr.* 23:836-847.

- McMurdie, Holmes (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* 8:e61217.
- Mora, Tittensor, Adl, Simpson, Worm (2011). How many species are there on earth and in the ocean? *PLoS Biology* 9:e1001127.
- Moritz, Richardson, Ferrier, Monteith, Stanisci, Williams, Whiffin (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, Chuyong, Condit, Hubbell, Kenfack, Thomas, Valencia, Green (2008). A general framework for the distance-decay of similarity in ecological communities. *Ecol. Lett.* 11:904-917.
- Myers, Chase, Jimnez, Jørgensen, Araujo- Murakami, Paniagua-Zambrana, Seidel (2013). Beta-diversity in temperate and tropical forests reflects dissimilar mechanisms of community assembly. *Ecol. Lett.* 16:151-157.
- Nekola, White (1999). The distance decay of similarity in biogeography and ecology. *J. Biogeog.* 26:867-878.
- Neuhauser, Kirchmair, Bulman, Bass (2014). Cross-kingdom host shifts of phytomyxid parasites. *BMC Evol. Biol.* 14:33.
- Novotny, Basset, Miller, Weiblen, Bremer, Cizek, Drozd (2002). Low host specificity of herbivorous insects in a tropical forest. *Nature* 416:841-844.
- Novotny, Miller, Hulcr, Drew, Basset, Janda, Setliff, Darrow, Stewart, Auga, Isua, Molem, Manumbor, Tamtia, Mogia, Weiblen (2007). Low beta diversity of herbivorous insects in tropical forests. *Nature* 448:692-695.
- Oksanen, Blanchet, Kindt, Legendre, Minchin, O'Hara, Simpson, Solymos, Stevens, Wagner. 2013. Vegan: community ecology package. R package version 2.0-7.
- Pitman, Terborgh, Silman, Nuñez V (1999). Tree species distributions in an upper amazonian forest. *Ecology* 80:2651-2661.
- Plotkin, Potts, Yu, Bunyavejchewin, Condit, Foster, Hubbell, LaFrankie, Manokaran, Seng, Sukumar, Nowak, Ashton (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. Vienna, Austria, R Foundation for Statistical Computing.
- Rognes, Flouri, Nichols, Quince, Mahé (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4: e2584

- Rosindell, Hubbell, Etienne (2011). The Unified Neutral Theory of Biodiversity and Biogeography at age ten. *Trends Ecol. Evol.* 26:340-348.
- Rueckert, Simdyanov, Aleoshin, Leander (2011). Identification of a divergent environmental DNA sequence clade using the phylogeny of gregarine parasites (Apicomplexa) from crustacean hosts. *PLoS ONE* 6:e18163.
- Ryan, Fayer, Xiao (2014). *Cryptosporidium* species in humans and animals: current understanding and research needs. *Parasitology* 141:1667-1685.
- Saito, Soininen, Fonseca-Gessner, Siqueira (2015). Dispersal traits drive the phylogenetic distance decay of similarity in Neotropical stream metacommunities. *J. Biogeogr.* 42:2101-2111.
- Scheiner (2003). Six types of species-area curves. *Glob. Ecol. Biogeogr.* 12:441-447.
- Scheiner, Chiarucci, Fox, Helmus, McGlenn, Willig (2011). The underpinnings of the relationship of species richness with space and time. *Ecol. Monogr.* 81:195-213.
- Slik, Arroyo-Rodríguez, Aiba, Alvarez-Loayza, Alves, Ashton, Balvanera, Bastian, Bellingham, van den Berg, Bernacci, da Conceição Bispo, Blanc, Böhning-Gaese, Boeckx, Bongers, Boyle, Bradford, Brearley, Breuer-Ndoundou Hockemba, Bunyavejchewin, Calderado Leal Matos, Castillo-Santiago, Catharino, Chai, Chen, Colwell, Chazdon, Clark, Clark, Clark, Culmsee, Damas, Dattaraja, Dauby, Davidar, DeWalt, Doucet, Duque, Durigan, Eichhorn, Eisenlohr, Eler, Ewango, Farwig, Feeley, Ferreira, Field, de Oliveira Filho, Fletcher, Forshed, Franco, Fredriksson, Gillespie, Gillet, Amarnath, Griffith, Grogan, Gunatilleke, Harris, Harrison, Hector, Homeier, Imai, Itoh, Jansen, Joly, de Jong, Kartawinata, Kearsley, Kelly, Kenfack, Kessler, Kitayama, Kooyman, Larney, Laumonier, Laurance, Laurance, Lawes, Amaral, Letcher, Lindsell, Lu, Mansor, Marjokorpi, Martin, Meilby, Melo, Metcalfe, Medjibe, Metzger, Millet, Mohandass, Montero, de Morisson Valeriano, Mugerwa, Nagamasu, Nilus, Ochoa-Gaona, Onrizal, Page, Parolin, Parren, Parthasarathy, Paudel, Permana, Piedade, Pitman, Poorter, Poulsen, Poulsen, Powers, Prasad, Puyravaud, Razafimahaimodison, Reitsma, dos Santos, Roberto Spironello, Romero-Saltos, Rovero, Rozak, Ruokolainen, Rutishauser, Saiter, Saner, Santos, Santos, Sarker, Satdichanh, Schmitt, Schöngart, Schulze, Suganuma, Sheil, da Silva Pinheiro, Sist, Stevart, Sukumar, Sun, Sunderland, Suresh, Suzuki, Tabarelli, Tang, Targhetta, Theilade, Thomas, Tchouto, Hurtado, Valencia, van Valkenburg, Van Do, Vasquez, Verbeeck, Adekunle, Vieira, Webb, Whitfeld, Wich, Williams, Wittmann, Wöll, Yang, Adou Yao, Yap, Yoneda, Zahawi, Zakaria, Zang, de Assis, Garcia Luize, Venticinque (2015). An estimate of the number of tropical tree species. *Proc. Natl. Acad. Sci. U.S.A.* 112:7472-7477.
- Soininen, McDonald, Hillebrand (2007). The distance decay of similarity in ecological communities. *Ecography* 30:3-12.
- Stamatakis (2006). RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688-2690.

- Stoeck, Bass, Nebel, Christen, Jones, Breiner, Richards (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, Loisel, Blake, Ricklefs (2016). Resource predictability and specialization in avian malaria parasites. *Mol. Ecol.* 25:4377-4391.
- Swenson, Erickson, Mi, Bourg, Forero-Montaña, Ge, Howe, Lake, Liu, Ma, Pei, Thompson, Uriarte, Wolf, Wright, Ye, Zhang, Zimmerman, Kress (2012). Phylogenetic and functional alpha and beta diversity in temperate and tropical tree communities. *Ecology* 93:S112-S125.
- Swenson, Mi, Kress, Thompson, Uriarte, Zimmerman (2013). Species-time-area and phylogenetic-time-area relationships in tropical tree communities. *Ecol. Evol.* 3:1173-1183.
- Turner, Corlett (1996). The conservation value of small, isolated fragments of lowland tropical rain forest. *Trends Ecol. Evol.* 11:330-333.
- Venter, Nitsche, Domonell, Heger, Arndt (2017). The protistan microbiome of grassland soil: diversity in the mesoscale. *Protist* doi:10.1016/j.protis.2017.03.005.
- Wetzel, Bicudo, Ector, Lobo, Soininen, Landeiro, Bini (2012). Distance decay of similarity in Neotropical diatom communities. *PLoS ONE* 7:e45071.
- Zhang, Swenson, Chen, Liu, Li, Huang, Mi, Ma (2013). Phylogenetic beta diversity in tropical forests: implications for the roles of geographical and environmental distance. *J. Syst. Evol.* 51:71-85.
- Zinger, Boetius, Ramette (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.

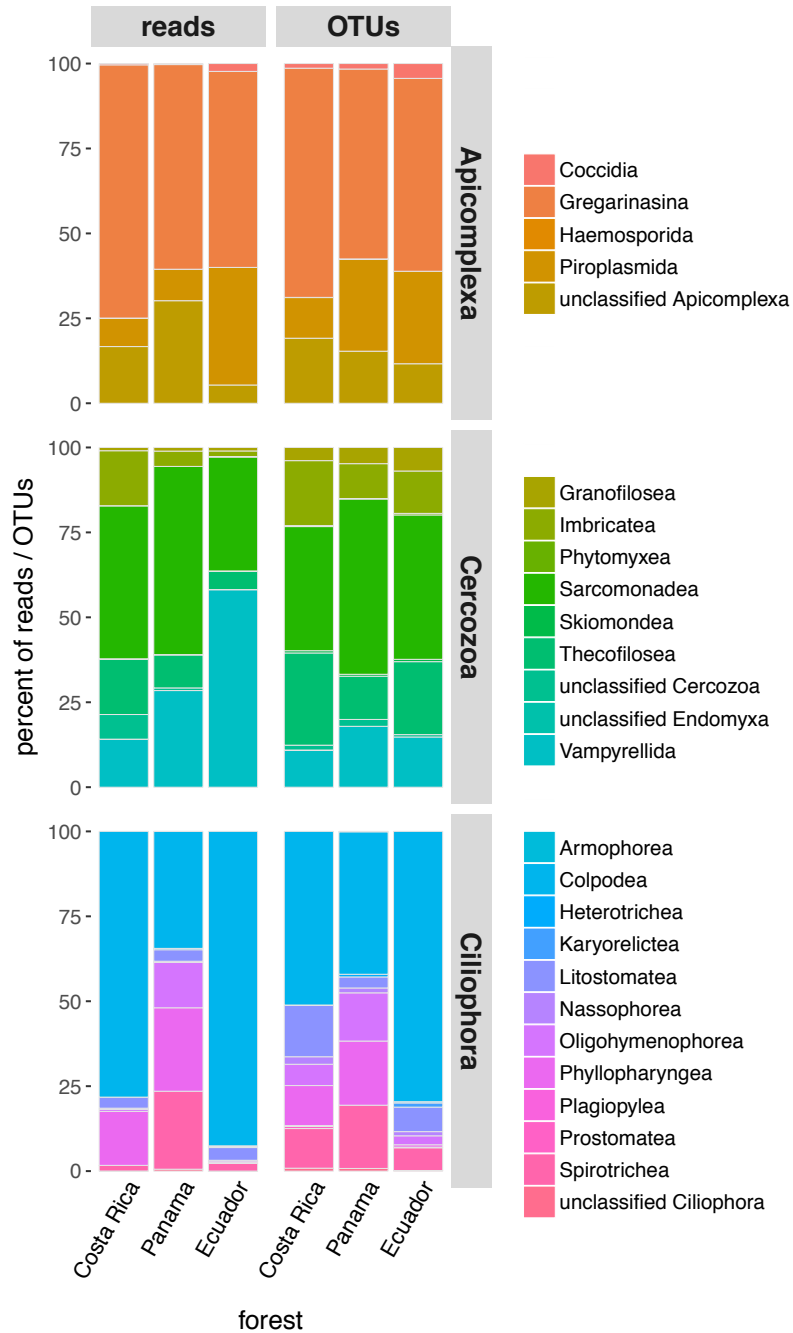


Figure 2. Taxa-area relationships (TAR) within the three forests, and between all sample sites. Shaded areas are 95 % confidence intervals. Slopes were significant for both OTU richness and phylogenetic variability. Slopes were steeper within forest than among forests. Stars denote significant steeper 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.

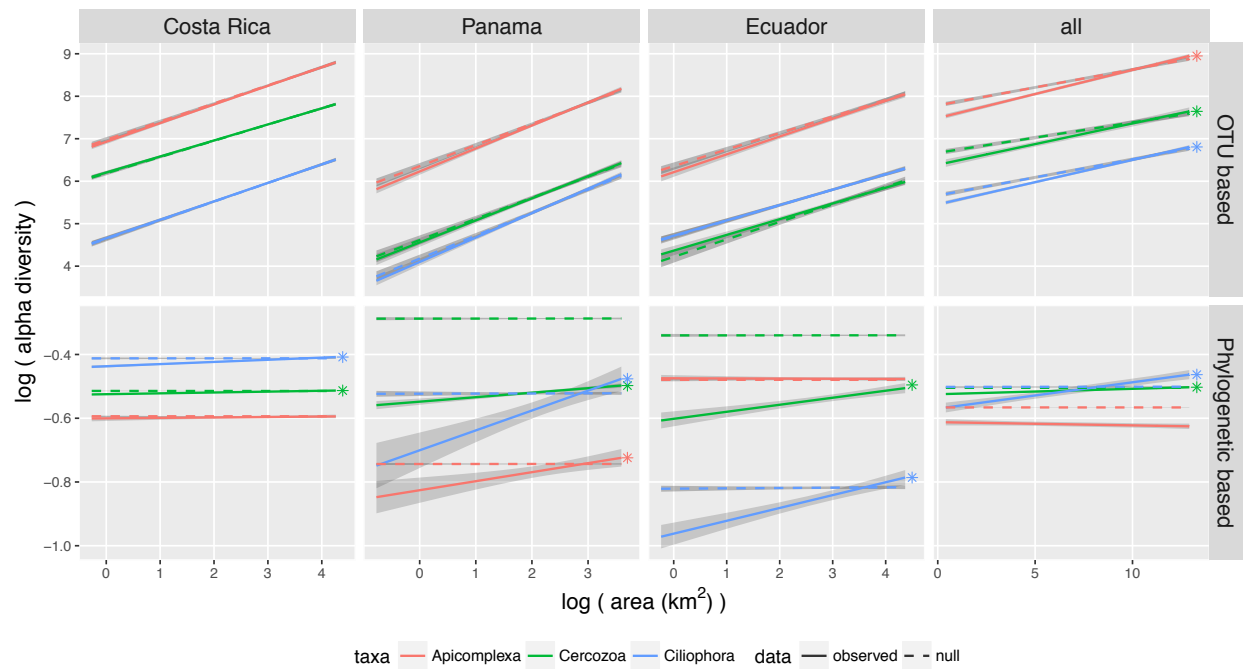


Figure 3. Distance-decay relationships (DDR) for three protist groups using all samples of the three forests. Shaded areas are 95 % confidence intervals. Slopes were not significant within forests for both OTU richness and phylogenetic variability, and therefore are not shown. Stars denote significant 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.

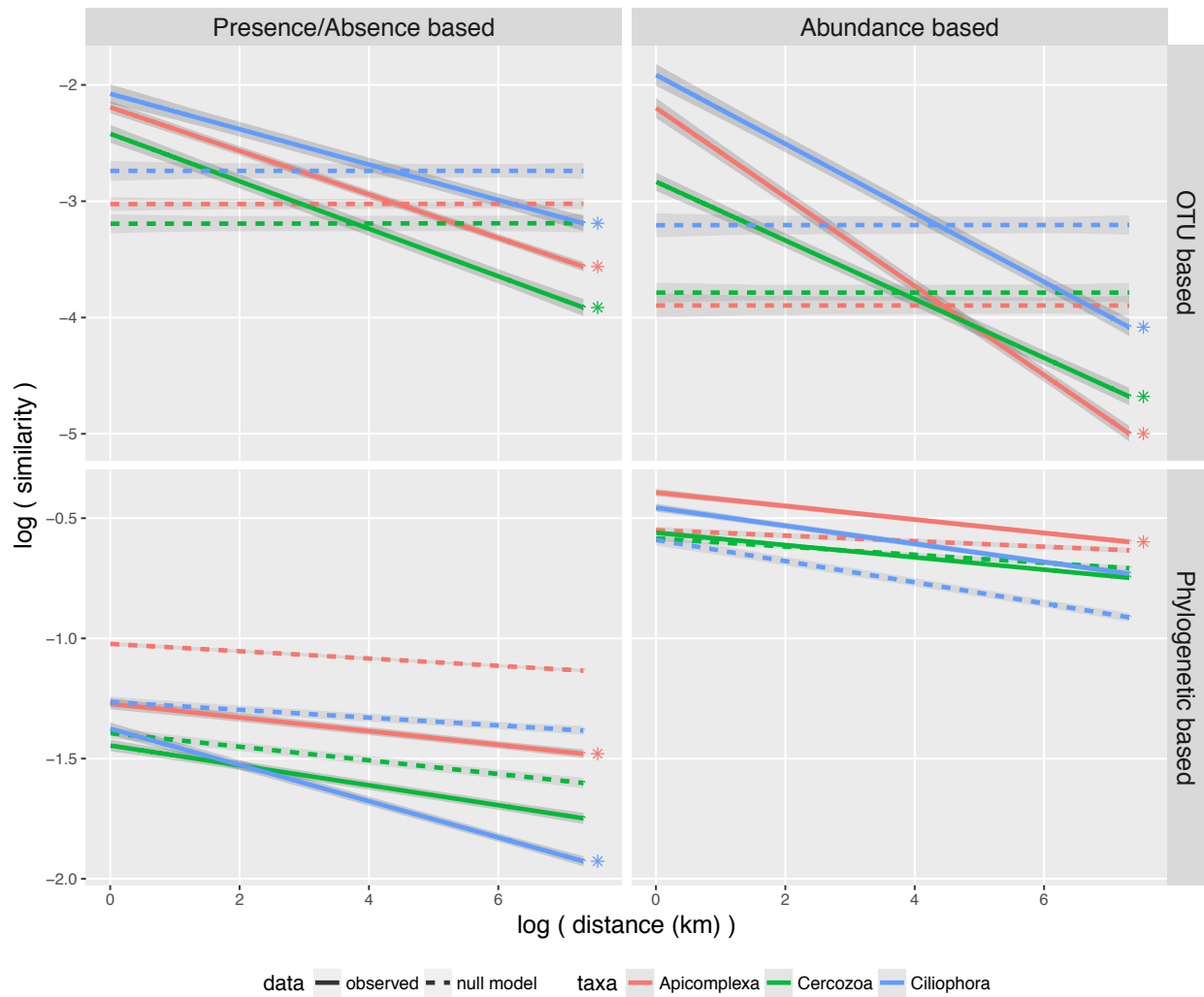


Figure S1. Map of the sampled Neotropical region with inset into sampled forests in Costa Rica (La Selva Biological Station), Panama (Barro Colorado Island) and Ecuador (Tiputini Biodiversity Station). Red crosses are individual sampling points.

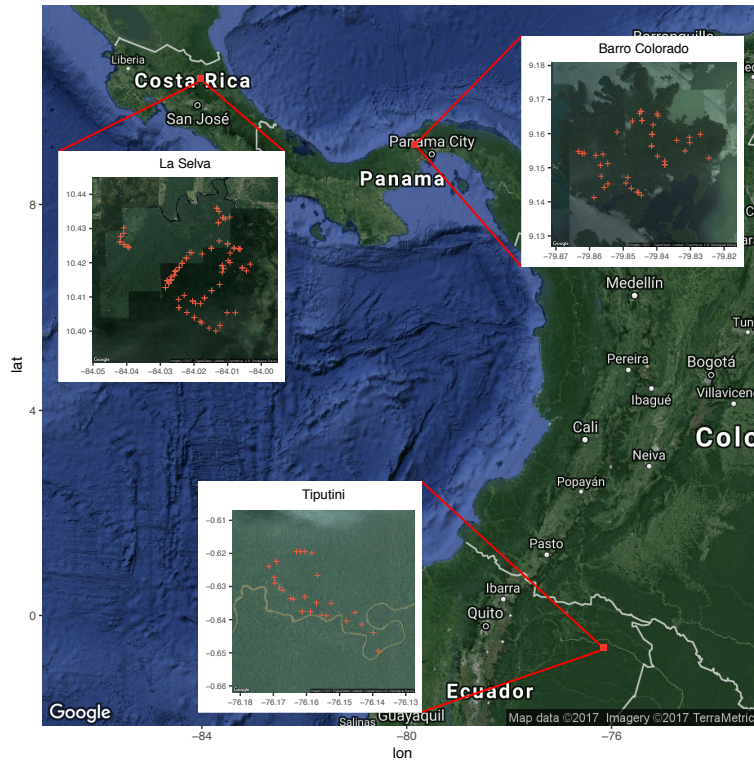


Figure S2. OTU richness based Taxa-Area Relationship (TAR) within forests for each sub-taxa. None of the slopes significantly differed from null model 1.

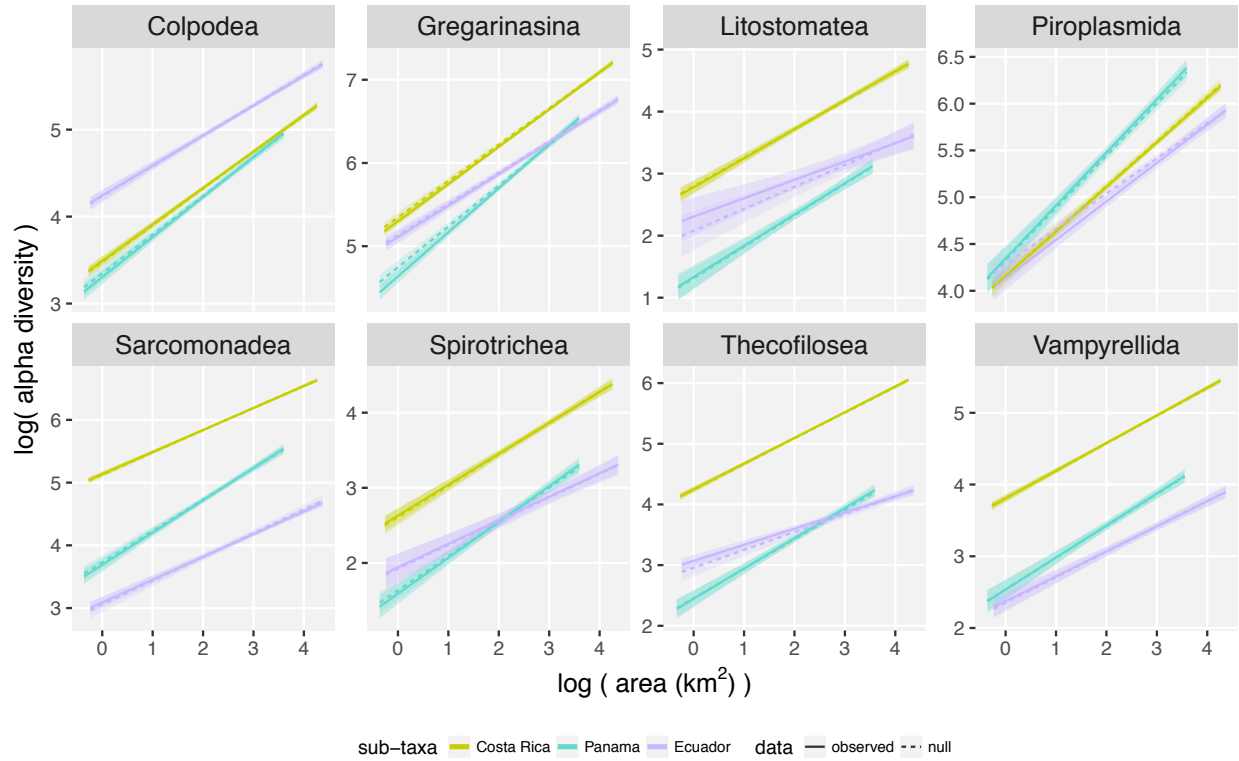


Figure S3. Taxa-Area Relationship (TAR) of sub-taxa across forests. Stars denoted significant steepest slopes than the null model. Null model 1 (samples random placement) was used for OTU based analyses and null model 2 (phylogenetic tree tips shuffling) was used for phylogenetic based analyses.

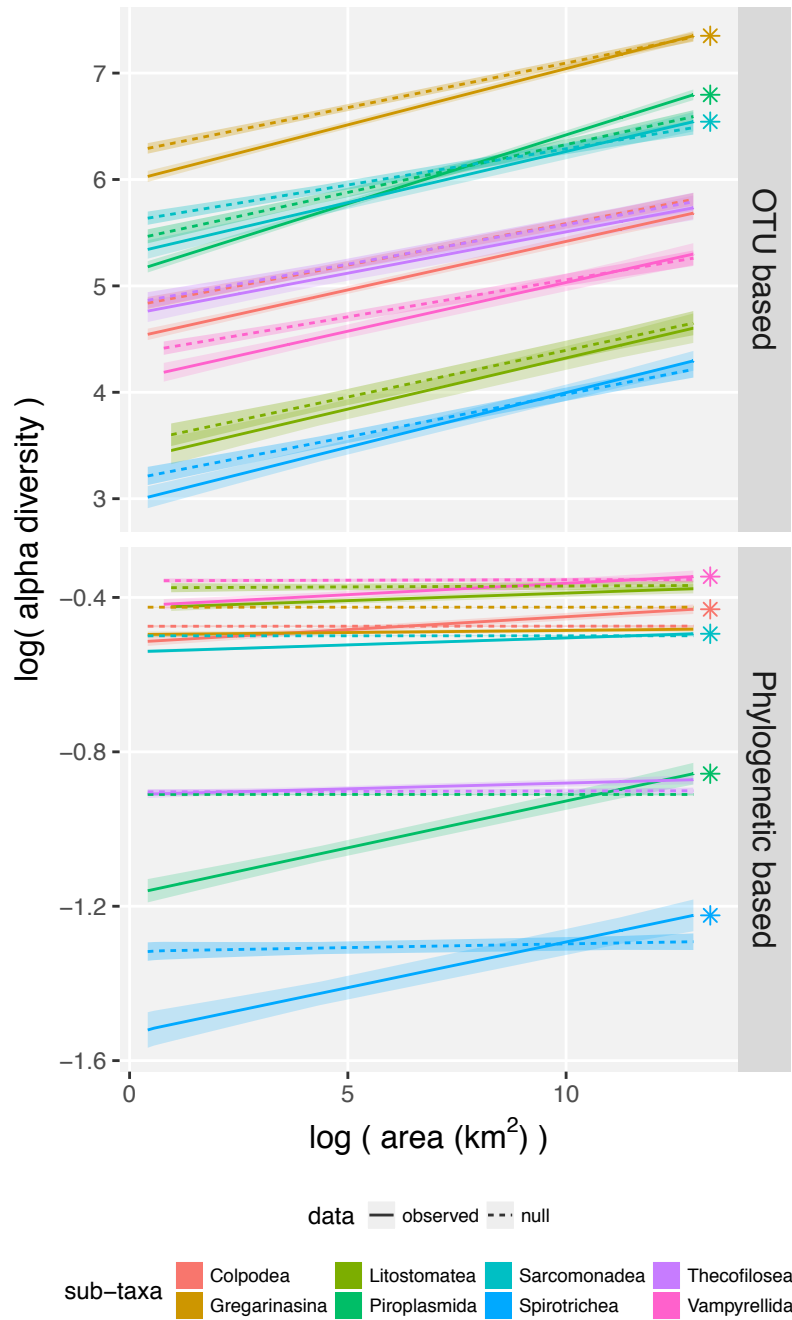


Figure S4. Phylogenetic variability based Taxa-Area Relationship (TAR) within forests for each sub-taxa. Stars denoted significant steepest slopes compared to null model 2.

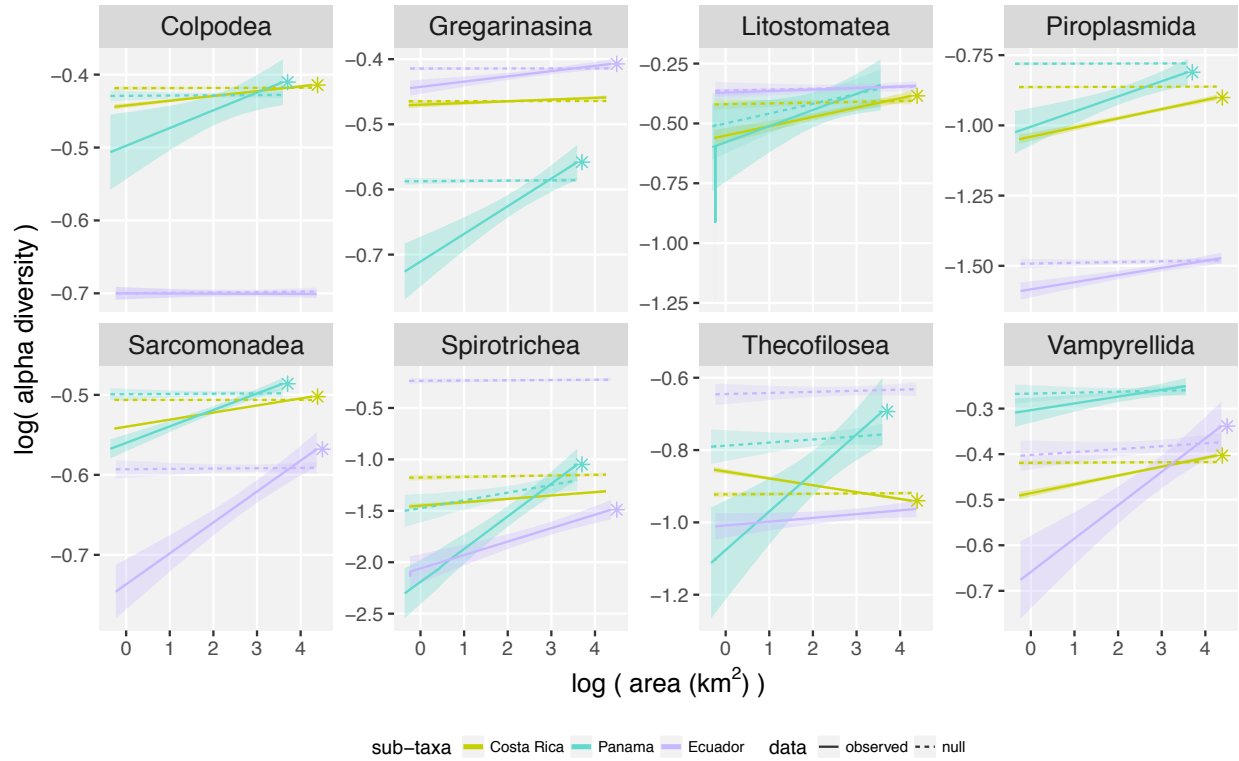


Figure S5. OTU relative abundance based Distance-Decay Relationship (DDR) of sub-taxa within forests. All fitted linear regression lines were not significant.

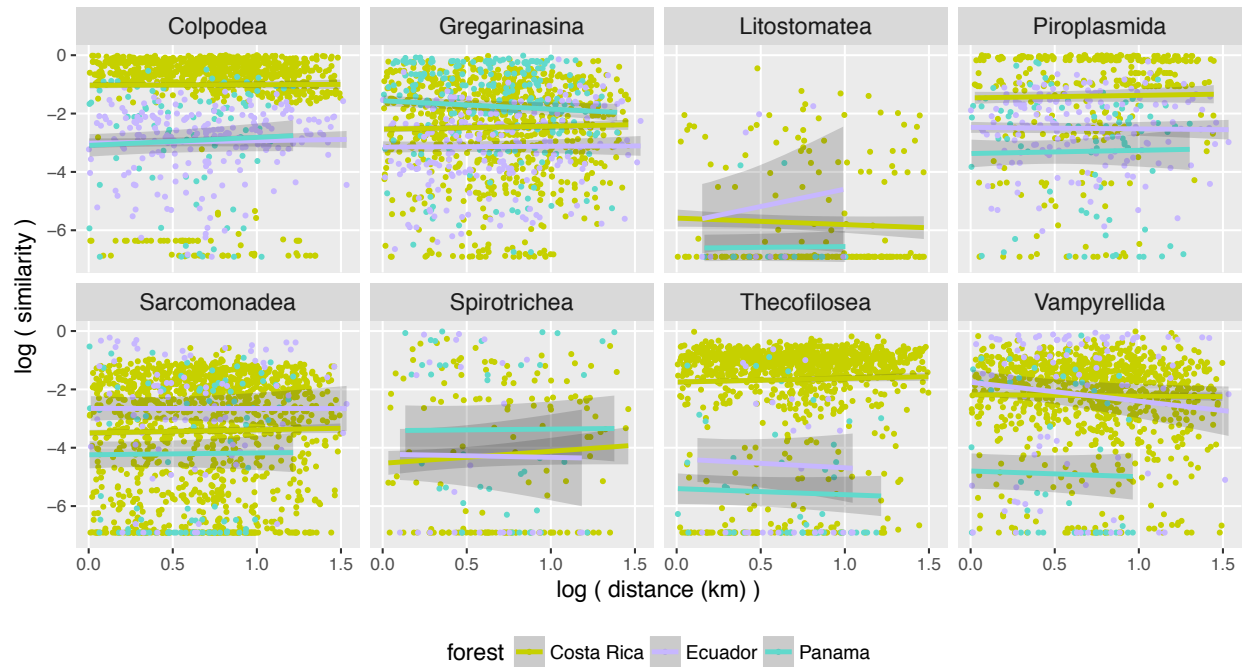


Figure S6. Distance-Decay Relationship (DDR) of sub-taxa across forests. A value of $1e-3$ was added to similarity indexes to avoid infinite negative values in the log space. Stars denoted significant steeper slopes than null model. Sub-taxa OTU based DDR slopes were all significantly negative and steepest than the null model 1. Only Thecofilosea and Vampyrellida have significant steeper phylogenetic based DDR slopes than in the null model 2 for both presence/absence and abundance based indexes. Piroplasmida have 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.

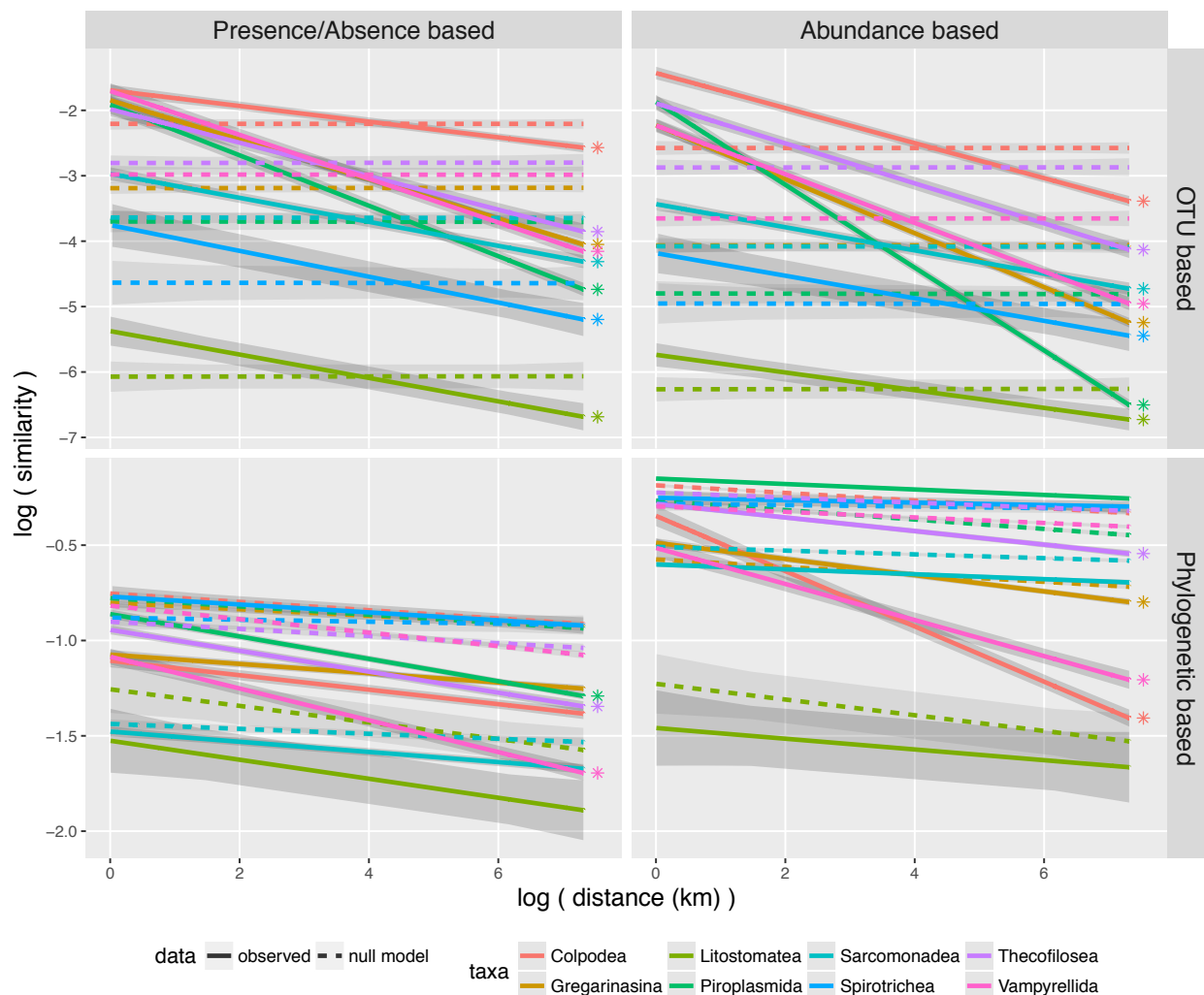


Table S1. Spearman's rho correlation coefficient calculated between OTU- and phylogenetic-based community similarity measures.

forest	taxa	Presence-Absence based	Abundance based
all	Apicomplexa	0.66*	0.59
	Cercozoa	0.6	0.54
	Ciliophora	0.47	0.5
Costa Rica	Apicomplexa	0.87	0.5
	Cercozoa	0.89	0.84
	Ciliophora	0.65	0.78
Panama	Apicomplexa	0.84	0.68
	Cercozoa	0.6	0.57
	Ciliophora	0.56	0.52
Ecuador	Apicomplexa	0.79	0.57
	Cercozoa	0.62	0.68
	Ciliophora	0.41	0.29

*All correlations significant with $p < 0.001$ after Benjamini and Hochberg (1995) correction.