Soil Protists in Three Neotropical Rainforests are Hyperdiverse and Dominated by Parasites

Frédéric Mahé,[1] Colomban de Vargas,[2,3] David Bass,[4,5] Lucas Czech,[6] Alexandros Stamatakis,[6,7] Enrique Lara,[8] Jordan Mayor,[9] John Bunge,[10] Sarah Sernaker,[11] Tobias Siemensmeyer,[1] Isabelle Trautmann,[1] Sarah Romac,[2,3] Cédric Berney,[2,3] Alexey Kozlov,[6] Edward A. D. Mitchell,[8,12] Christophe V. W. Seppey,[8] David Singer,[8] Elianne Egge,[13] Rainer Wirth,[14] Gabriel Trueba,[15] and Micah Dunthorn[1]*

[1]Department of Ecology, University of Kaiserslautern, Erwin-Schrödinger-Straße, 67663 Kaiserslautern, Germany.

[2]CNRS, UMR 7144, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France.

[3]Sorbonne Universités, UPMC Univ Paris 06, UMR 7144, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff, France.

[4]Department of Life Sciences, The Natural History Museum London, Cromwell Road, London SW7 5BD, United Kingdom.

[5]Centre for Environment, Fisheries & Aquaculture Science (Cefas), Barrack Road, The Nothe, Weymouth, Dorset DT4 8UB, United Kingdom.

[6]Heidelberg Institute for Theoretical Studies, Schloß-Wolfsbrunnenweg, 69118 Heidelberg, Germany.

[7]Institute for Theoretical Informatics, Karlsruhe Institute of Technology, Am Fasanengarten, 76128 Karlsruhe, Germany.

[8]Laboratory of Soil Biodiversity, Université de Neuchâtel, Rue Emile-Argand, 2000 Neuchâtel, Switzerland.

[9]Department of Forest Ecology and Management, Swedish University of Agricultural Sciences, Skogsmarksgränd, 90183 Umeå, Sweden.

[10]Department of Computer Science, Cornell University, Gates Hall, Ithaca, New York 14853, USA.

[11]Department of Statistical Science, Cornell University, Malott Hall, Ithaca, New York 14853, USA.

[12]Jardin Botanique de Neuchâtel, Pertuis-du-Sault, 2000 Neuchâtel, Switzerland.

[13]Department of Biosciences, University of Oslo, Blindernveien, 0316 Oslo, Norway.

[14]Department of Plant Ecology and Systematics, University of Kaiserslautern, Erwin-Schrödinger-Straße, 67663 Kaiserslautern, Germany.

[15]Instituto de Microbiología, Universidad San Francisco de Quito, Diego de Robles, Cumbayá, Ecuador.

*Correspondence to: E-mail: dunthorn@rhrk.uni-kl.de.

Abstract:

Although animals and plants in tropical rainforests are known to be hyperdiverse within and between communities, it is unknown if similar patterns are reflected at the microbial scale with unicellular eukaryotes or protists. Using environmental DNA sequencing of soils sampled in rainforests from Costa Rica, Panama, and Ecuador, we found that the Apicomplexa dominated the protist communities; these parasites potentially promote animal diversity in the forests by reducing population growth rates in a density-dependent manner similar to the Janzen-Connell hypothesis for tropical trees. Extremely high OTU diversity and high OTU turnover between samples within the same forests suggest that protists, not arthropods, are the most diverse eukaryotes in tropical rainforests, and that rainforest soil protists are at least as diverse, and potentially more diverse, than those in the marine plankton.

Main Text:

Since the works of early naturalists such as von Humboldt and Bonpland (*1*), we have known that animals and plants in tropical rainforests have high species richness, and can exhibit high species turnover across relatively short distances. Subsequent work has strengthened this view, and further revealed that parasites contribute to high tropical tree diversity (*2*) and that arthropods are the most diverse eukaryotes in these forests (*3*). The focus on macro-organisms in these studies is primarily because they are familiar and readily observable to us. We do not know if the less familiar and less readily observable protists—microbial eukaryotes excluding animals, plants, and fungi (*4*)—inhabiting these same ecosystems exhibit similar diversity patterns.

To evaluate if macro-organismic diversity patterns are reflected at the microbial scale with protists, we conducted an environmental DNA metabarcoding study by sampling soils in 279 locations in a variety of lowland Neotropical forest types in: La Selva Biological Station, Costa Rica; Barro Colorado Island, Panama; and Tiputini Biodiversity Station, Ecuador (**Table S1**) (*5*). This metabarcoding approach had the power to uncover known and new taxa on a massive scale that can be subsequently linked to novel morphologies (*6, 7*). By amplifying DNA extracted from the soils with general primers targeting the V4 region of 18S rRNA and sequencing it with Illumina MiSeq (**Fig. S1**), we were able to unveil most eukaryotic groups and assess the diversity and the relative dominance of free-living and parasitic lineages. Sequence reads were clustered into operational taxonomic units (OTUs) with Swarm (*8*) (**Fig. S2**). To provide a comparative context for the rainforest soils, we applied the the same bioinformatics pipeline to re-analyze the *Tara*-Oceans consortium's V9 rRNA metabarcoding data from the world oceans (*6*), re-sequenced 40 European near-shore planktonic and benthic samples from the

BioMarKs consortium (*9*) with V4 and V9 primers (**Table S2**), and newly sequenced 29 soil samples from temperate forests in Switzerland with V9 primers (**Table S3**).

Sequencing of the rainforest soils resulted in ~132.3 million cleaned reads. Of the ~50.1 million reads assigned to the protists, 75.3% had a maximum similarity of <80% to references in the Protist Ribosomal Reference (PR^2) database (*10*) (**Fig. 1**, **Fig. S3**). By contrast, in our re-analysis of ~367.8 million cleaned protist reads from the world's open oceans only 3.1% were <80% similar to the PR^2 database (**Fig. 1**). Most of the protist reads from the near-shore sites and Swiss soils and most of the fungi in the rainforest soils were likewise highly similar to the PR^2 database (**Fig. S4**).

Reads <80% similar to known references are usually considered spurious and removed in environmental protist sequencing studies (*11*). Three quarters of our rainforest soil protist data would be discarded if we applied this conservative cleaning step; however, PR^2 and similar databases are biased towards marine and temperate terrestrial reference sequences. We therefore deployed a phylogeny-aware cleaning step that dereplicated the reads into ~10.6 million amplicons (= strictly identical reads to which an abundance value can be attached) and placed them into a phylogenetic context using the Evolutionary Placement Algorithm (EPA) (*12*) with a tree inferred from 512 full-length reference sequences from all major eukaryotic clades (**Table S4**). This approach found only 0.012% of the amplicons (representing 0.0056% of the reads) placed outside of any known clade, and 5.4% of the amplicons (representing 5.7% of the reads) placed both outside and inside known clades (**Fig. 2, Fig. S5**). Some amplicons may have been grouped incorrectly because they are short and highly divergent (*13*). We thus retained OTUs whose centroid (= the most abundant amplicon in the OTU) placed only within known clades

with a high likelihood score, resulting in the removal of only 6.8% of the reads and 7.7% of the OTUs.

The remaining protist reads from the rainforest soils clustered into 26,860 OTUs. As in the marine plankton (*6*), more protist OTUs were detected than animals, plants, and fungi combined (**Table S5**), and unlike the marine plankton, most of these OTUs had a low similarity to the PR^2 database (**Fig. 1, Fig. S6**). Taxonomic assignment of the protist OTUs showed that 84.4% of the reads (between 76.6% and 93.0% per forest) and 50.6% of the OTUs (between 43.1% and 61.5% per forest) were affiliated to the Apicomplexa (**Fig. 3, Fig. S7**). Apicomplexa are widespread and are parasites of animals (*14, 15*), although a few free-living species are known (*16*). The observed read- and OTU-abundances of the Apicomplexa in the marine and Swiss soil protist data were all ≤2.6% (**Fig. S8-S9**). A high abundance of Apicomplexa was also found in rainforest samples amplified with primers designed to target the closely related Ciliophora (**Fig. S10**).

We used EPA to place Apicomplexa OTU centroids into a more focused phylogeny inferred from 190 full-length reference sequences from all major Alveolate clades (**Table S6**). While the OTUs were generally distributed across the whole phylogeny (**Fig. S11**), 80.2% grouped with the gregarines, which predominantly infect invertebrates (*14*). About 23.8% of these gregarine OTUs placed within the lineage formed by the millipede parasite *Stenophora,* and the insect parasites *Amoebogregarina*, *Gregarina*, *Leidyana*, and *Protomagalhaensia*. An additional 13.5% of the OTUs clustered with two environmental gregarine sequences collected from brackish sediment. Many other OTUs placed within gregarine lineages thought to be primarily parasites of marine annelids and polychaetes. Non-gregarine Apicomplexa OTUs largely grouped with the blood parasites *Plasmodium* (which cause malaria) and close relatives,

some of which cycle through arthropods and vertebrates. It is unknown if this immense diversity of Apicomplexa is derived from animals in the soils alone. Some of this diversity could be a shadow of the tree-canopy communities that have rained down from above, and some may be unknown free-living forms. Some may have evolved to exploit the nutrient-rich environment of the numerous fungi found in the same samples (**Fig. S12**); while chitinases are unknown in the Apicomplexa, mycophagy has evolved in the closely related Cercozoa (*17*) and Ciliophora (*18*).

These dominating apicomplexan parasites potentially have a significant influence on the diversity of animals in the three rainforests. In the Janzen-Connell model (*19, 20*), host-specific parasites are hypothesized to contribute to high tropical tree diversity by reducing population growth rates in a density-dependent manner (*2*). The Apicomplexa, which are usually host-specific although switching is known (*14, 15, 21, 22*), may likewise contribute to high tropical animal diversity via infections (e.g., *23, 24*) that limit population growth rates of species that become locally abundant. In our model, we expect Apicomplexa to dominate both the below- and above-ground protist communities in tropical forests worldwide, and we expect most animal populations in those forests to have apicomplexan parasites. As with tropical trees (*25*), the nature of this density-dependent effect on the animals may vary by the mode of apicomplexan transmission.

The second through fifth most abundant protist microbes were the predominantly heterotrophic Cercozoa, Ciliophora, Conosa and Lobosa (**Fig. 3, Fig. S7**). Photo- or mixo-trophic Chlorophyta, Dinophyta, Haptophyta, and Rhodophyta were also uncovered; Haptophyta are mostly marine, but phylogenetic placement of these OTUs showed many were closely related to freshwater species (**Fig. S13**). Algae increase the net uptake of carbon in some terrestrial ecosystems (*26*), but the contribution of these algae to the carbon cycle in Neotropical rainforests

is unknown. Although we observed taxonomically diverse heterotrophic and phototrophic microbial eukaryotic groups, almost 70% of the reads were derived from 50 hyperdominant OTUs that were mostly the apicomplexan parasites (**Fig. S14, Table S7**); on the other hand, with macro-organisms, resistance to parasites was a predicted feature for the 227 hyperdominant tree species found in the Amazon basin (*27*).

Using Breakaway (*28*), we estimated 26,863 (S.E. 1) observed plus unobserved protist OTUs using the combined soil samples, suggesting that the sequencing depth enabled detection of a rich diversity that strongly represents the total diversity within each sample. Bray-Curtis dendrograms (**Fig. S15**) and non-metric multidimensional scaling (**Fig. S16**) showed that the OTU composition was slightly more similar among the samples from Costa Rica and Panama than those from Ecuador. However, the Jaccard similarity index (**Fig. 4, Fig. S17**) estimated high OTU turnover between samples even within the same forest (lower OTU turnover was found for the fungi in the same forests (**Fig. S18**) and for the marine and Swiss soil protists (**Fig. S19-S20**)). Within-forest OTU rarefaction curves (**Fig S21**) based on the sample accumulation further estimated logarithmic increases without plateaus. Overall, this indicates that our sampling effort has unveiled only a fraction of the full extent of protist hyperdiversity in the three studied rainforests.

These observations and analyses offer a number of unique insights. First, just as there is high richness and high species turnover in rainforests at the macro-organismic scale, there are similar patterns of hyperdiversity with protists at the microbial scale. Given the low similarity of protist community compositions between samples and the 6-fold ratio of protist to animal OTUs, a concerted and comparable count will likely show that protists, and not arthropods as traditionally thought (*3*), are the most diverse eukaryotes in Neotropical rainforests. Second,

scratching the soil surface in just three Neotropical rainforests revealed a comparable amount of protist diversity as sampled in the sunlit surface layer of the world's open oceans. While the *Tara*-Oceans consortium (*6*) had a 7.3-fold higher sequencing depth from 29% more size-fractionated circumglobal samples, our re-analysis of their data found only 11.3 times more protist OTUs than found in the rainforest soils. Third, a diversity survey combining terrestrial and marine protists may reveal that these microbes are the most diverse eukaryotes on Earth. If they are it would not be due to inordinate speciation in just a few clades since the mid-Phanerozoic, e.g., the beetles (*29*), but because of the accumulation of rich and functionally complex protist lineages beginning in the early Proterozoic (*30*), which built up the multifaceted invisible foundation of our now familiar macroscopic ecosystems.

References and Notes:

1. A. von Humboldt, A. Bonpland, *Personal narrative of travels to the equinoctial regions of the new continent, during the years 1799-1804*. (Longman, Hurst, Rees, Orme & Brown, London, 1814-1829).

2. R. Bagchi *et al.*, Pathogens and insect herbivores drive rainforest plant diversity and composition. *Nature* **506**, 85-88 (2014).

3. Y. Basset *et al.*, Arthropod diversity in a tropical forest. *Science* **338**, 1481-1484 (2012).

4. S. M. Adl *et al.*, The revised classification of eukaryotes. *J. Eukaryot. Microbiol.* **59**, 429-493 (2012).

5. Materials and methods are available as Supporting Material.

6. C. de Vargas *et al.*, Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**, 1261605 (2015).

7. S.-E. Tarnawski, E. Lara, From environmental sequences to morphology: observation and characterisation of a paulinellid testate amoeba (*Micropyxidiella edaphonis* gen. nov. sp. nov. Euglyphida, Paulinellidae) from soil using fluorescent in situ hybridization. *Protist* **166**, 264-270 (2015).

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

9. R. Logares *et al.*, Patterns of rare and abundant marine microbial eukaryotes. *Curr. Biol.* **24**, 813-821 (2014).

10. L. Guillou *et al.*, 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 (2013).

11. T. Stoeck *et al.*, Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol. Ecol.* **19**, 21-31 (2010).

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

13. M. Dunthorn *et al.*, Placing environmental next-generation sequencing amplicons from microbial eukaryotes into a phylogenetic context. *Mol. Biol. Evol.* **31**, 993-1009 (2014).

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

15. F. O. Perkins, J. R. Barta, R. E. Clopton, M. A. Peirce, S. J. Upton, in *An Illustrated Giude to the Protozoa, 2nd ed.,* J. J. Lee, G. F. Leedale, P. Bradbury, Eds. (Society of Protozoologists, Lawrence, KS, 2002).

16. J. Janouškovec *et al.*, Factors mediating plastid dependency and the origins of parasitism in apicomplexans and their close relatives. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 10200-10207 (2015).

17. K. Dumack, M. E. H. Müller, M. Bonkowski, Description of *Lecythium terrestris* sp. nov. (Chlamydophryidae, Cercozoa), a soil dwelling protist feeding on fungi and algae. *Protist* **167**, 93-105 (2016).

18. W. Foissner, Colpodea (Ciliophora). *Protozoenfauna* **4/1**, i-x, 1-798 (1993).

19. D. H. Janzen, Herbivores and the number of tree species in tropical forests. *Am. Nat.* **104**, 501-528 (1970).

20. J. H. Connell, in *Dynamics of population,* P. J. Den Boer, G. R. Gradwell, Eds. (Wageningen, Pudoc, 1970).

21. V. A. Ellis *et al.*, Local host specialization, host-switching, and dispersal shape the regional distributions of avian haemosporidian parasites. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 11294-11299 (2015).

22. S. Rueckert, P. M. Villette, B. S. Leander, Species boundaries in gregarine apicomplexan parasites: a case study-comparison of morphometric and molecular variability in *Lecudina* cf. *tuzetae* (Eugregarinorida, Lecudinidae). *J. Eukaryot. Microbiol.* **58**, 275-283 (2011).

23. M. Asghar *et al.*, Hidden costs of infection: Chronic malaria accelerates telomere degradation and senescence in wild birds. *Science* **347**, 436-438 (2015).

24. A. M. Bouwma, K. J. Howard, R. L. Jeanne, Parasitism in a social wasp: effect of gregarines on foraging behavior, colony productivity, and adult mortality. *Behav. Ecol. Sociobiol.* **59**, 222-233 (2005).

25. R. P. Freckleton, O. T. Lewis, Pathogens, density dependence and the coexistence of tropical trees. *Proc. R. Soc. B* **273**, 2909-2916 (2006).

26. W. Elbert *et al.*, Contribution of cryptogamic covers to the global cycles of carbon and nitrogen. *Nat. Geosci.* **5**, 459-462 (2012).

27. H. ter Steege *et al.*, Hyperdominance in the Amazonian tree flora. *Science* **342**, 1243092 (2013).

28. A. Willis, J. Bunge, Estimating diversity via frequency ratios. *Biometrics* **71**, 1042-1049 (2015).

29. B. D. Farrell, "Inordinate fondness" explained: why are there so many beetles? *Science* **281**, 555-559 (1998).

30. A. H. Knoll, Paleobiological perspectives on early eukaryotic evolution. *Cold Spring Harb. Perspect. Biol.* **6**, a016121 (2014).

Acknowledgments:

We thank the Gauss Centre for Supercomputing with SuperMUC access, and Thorsten Stoeck, Laura Katz, Håvard Kauserud, David Montagnes, and Owen T. Lewis. Funding primarily came from the Deutsche Forschungsgemeinschaft grant DU1319/1-1 to M.D.; the EU ERANet program BiodivERsA (BioMarKs 2008-6530) and the French Government "Investissements d'Avenir" programmes OCEANOMICS (ANR-11-BTBR-0008) to C.dV.; the Klaus Tschira Foundation to L.C., A.S., and A.K.; and the Swiss National Science Foundation to E.A.D.M. and E.L. The data reported here are available in the GenBank SRA bioproject # SUB582348.

Figure Captions.

Fig. 1: Similarity of protists to the taxonomic reference database PR^2 (*10*). Only 8.1% of the protist V4 reads from the three Neotropical forest soils had a similarity ≥95% with PR^2 reference sequences. By contrast, 68.1% of the protist V9 reads from the *Tara*-Oceans open oceans had a similarity ≥95% with PR^2 reference sequences; plotting this marine diversity shows that the approximately 112 million reads with 100% similarity to PR^2 references correspond to only 857 OTUs.

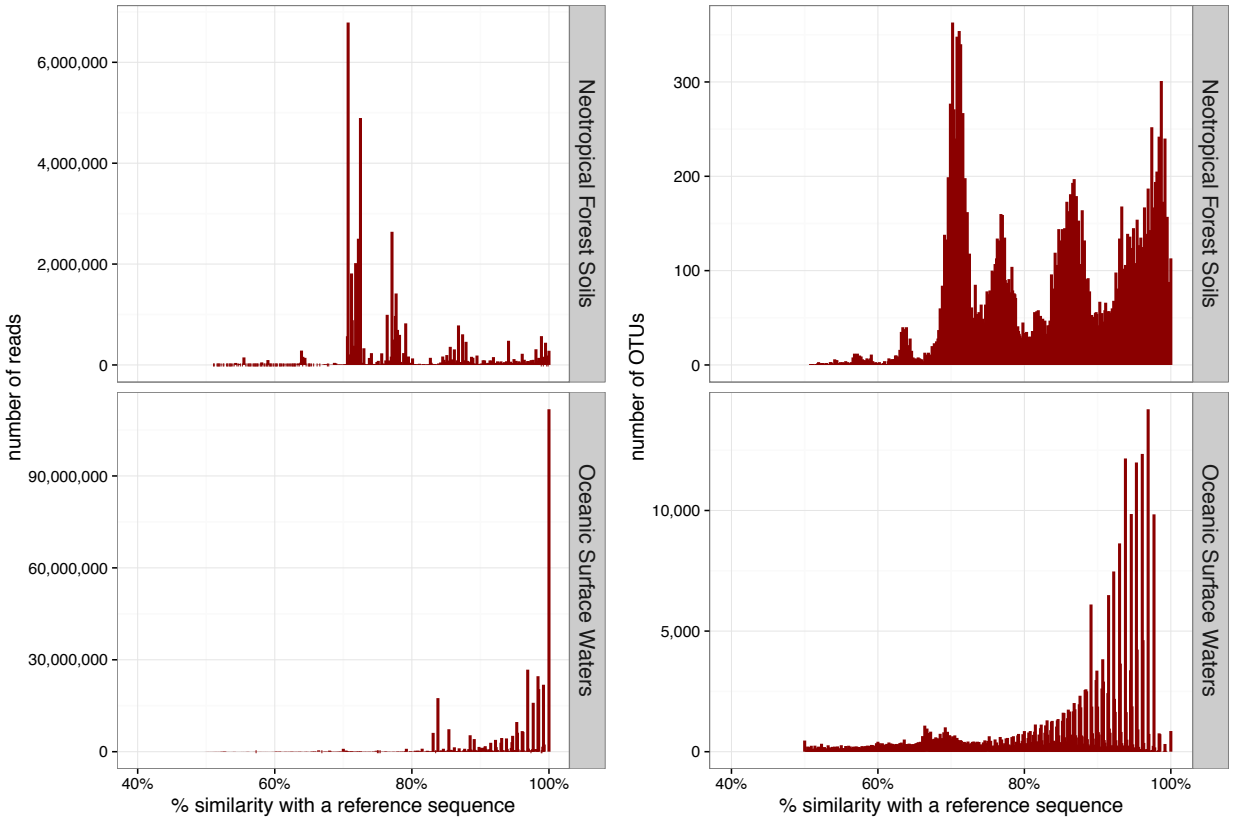

Fig. 2: Phylogenetic placement of Neotropical soil protist reads on a taxonomically unconstrained global eukaryotic tree. Reads were dereplicated into strictly identical amplicons. Inferred relationships between these major taxa may differ from those obtained with phylogenomic data. Alveolate includes Apicomplexa and Ciliophora; Holozoa includes animals; Holomycota includes fungi. Branches outside of known clades are shaded in gray.

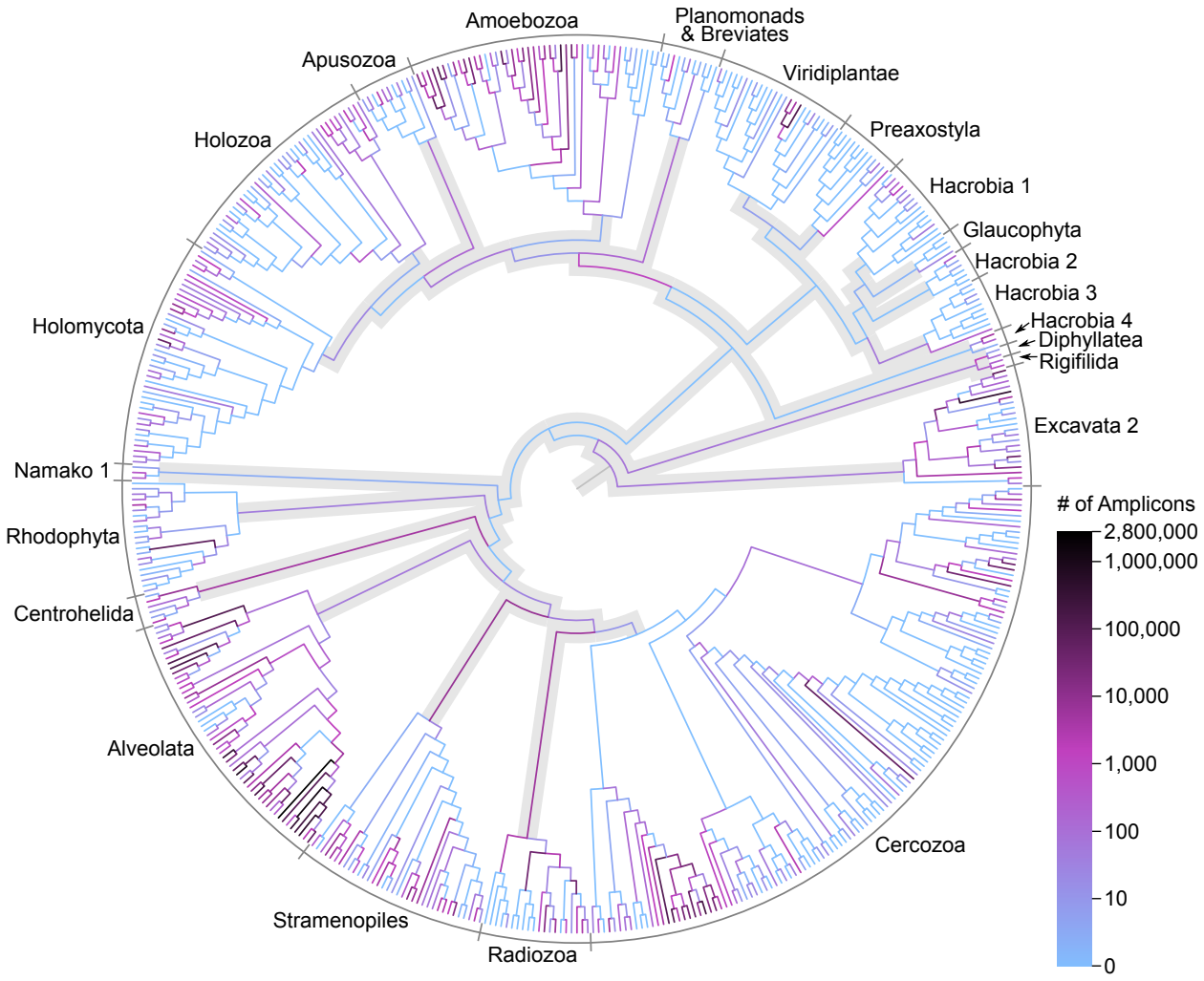

Fig. 3: Taxonomic identity and relative abundances of soil protist reads and OTUs in three Neotropical rainforests. Taxa shown each represent >0.1% of the data. Taxonomy is at the 3rd level of the PR^2 database.

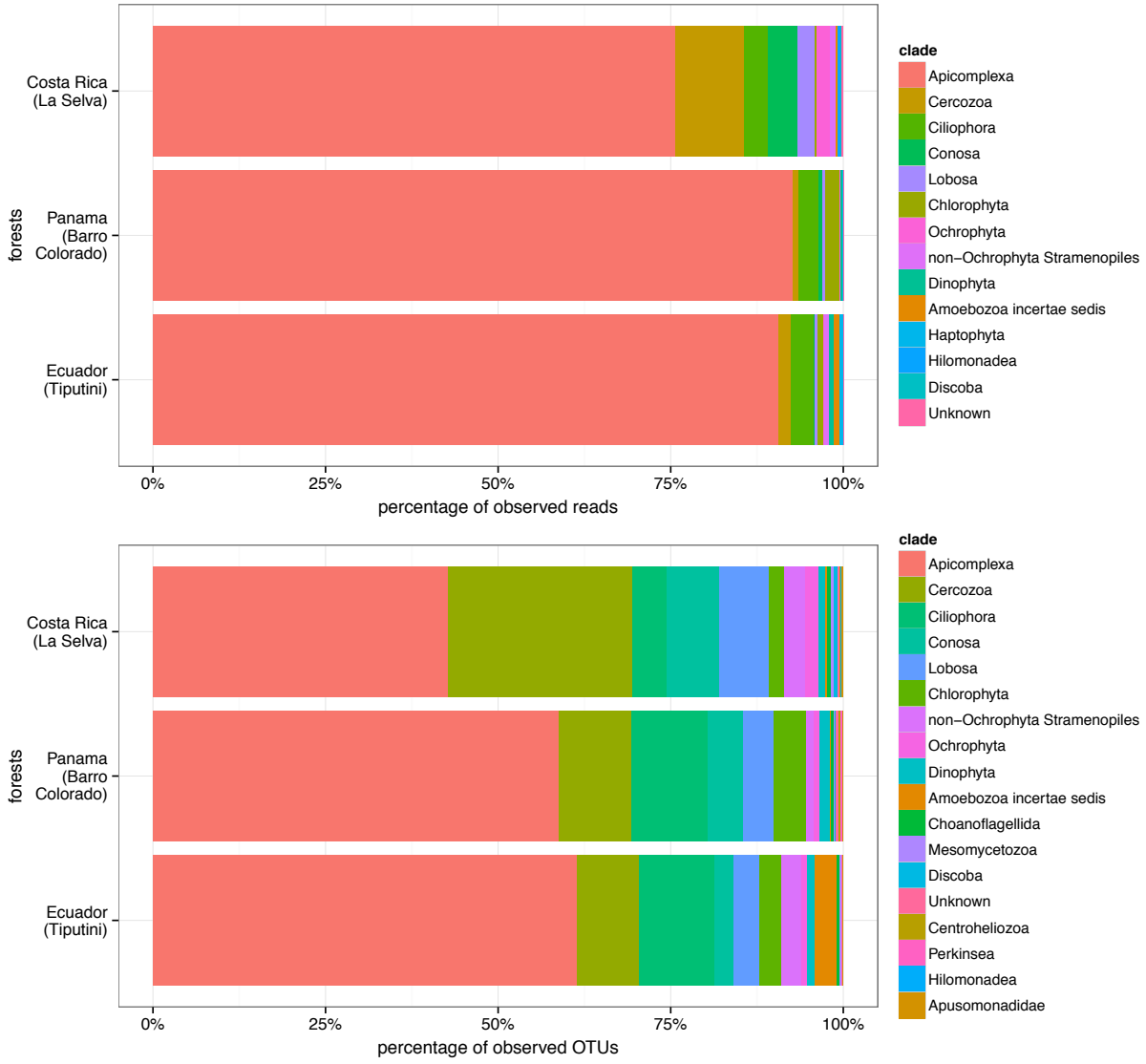
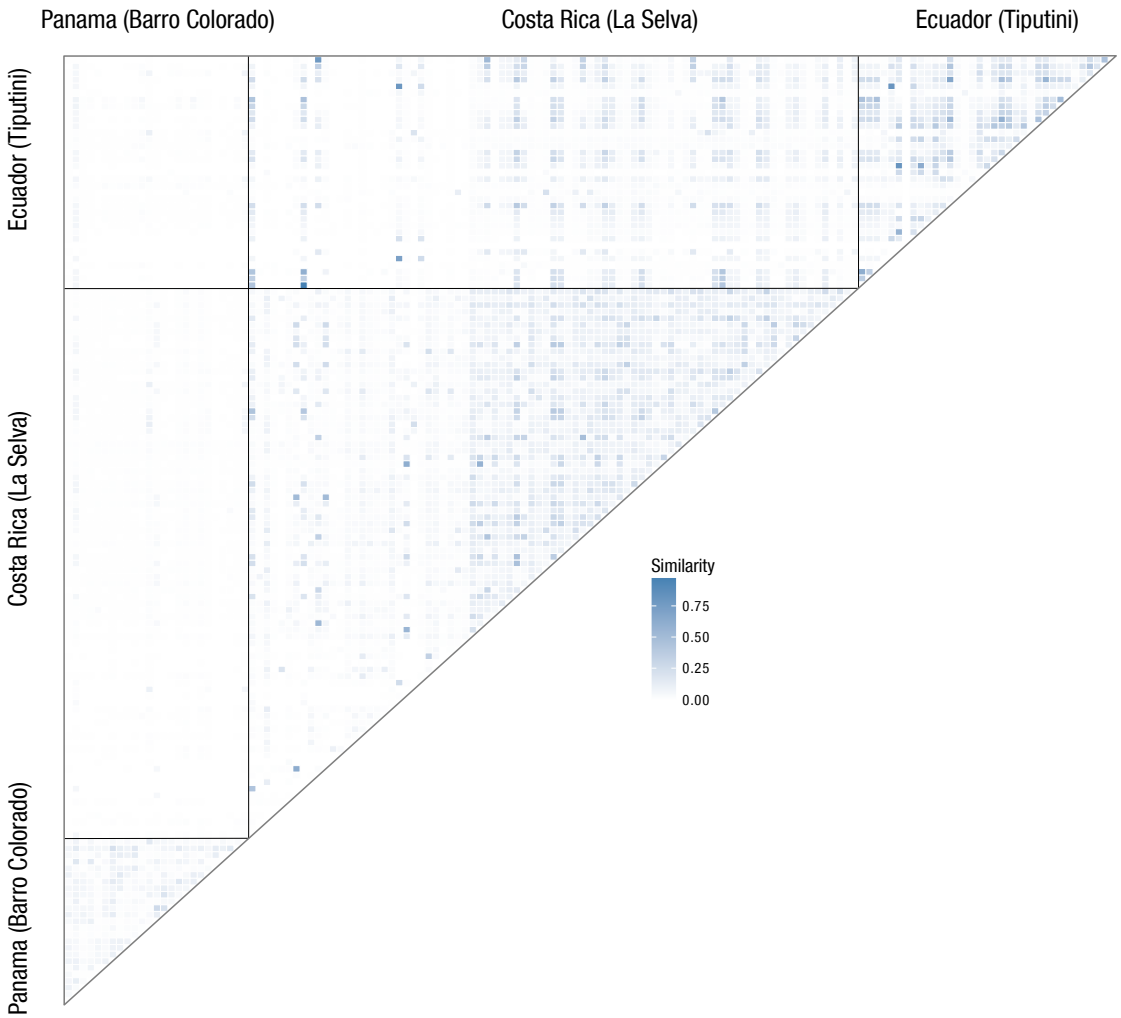
Fig. 4: Jaccard similarity index among Neotropical rainforest soil samples. Each square is a pairwise sample comparison. The lower the value, the higher the OTU turnover between samples. Sample names are in Figure S17.