nature ecology & evolution

ARTICLES

PUBLISHED: 20 MARCH 2017 | VOLUME: 1 | ARTICLE NUMBER: 0091

Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests

Frédéric Mahé[1], Colomban de Vargas[2,3], David Bass[4,5], Lucas Czech[6], Alexandros Stamatakis[6,7], Enrique Lara[8], David Singer[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], Elianne Egge[13], Guillaume Lentendu[1], Rainer Wirth[14], Gabriel Trueba[15] and Micah Dunthorn[1]*

High animal and plant richness in tropical rainforest communities has long intrigued naturalists. It is unknown if similar hyperdiversity patterns are reflected at the microbial scale with unicellular eukaryotes (protists). Here we show, using environmental metabarcoding of soil samples and a phylogeny-aware cleaning step, that protist communities in Neotropical rainforests are hyperdiverse and dominated by the parasitic Apicomplexa, which infect arthropods and other animals. These host-specific parasites potentially contribute to the high animal diversity in the forests by reducing population growth in a density-dependent manner. By contrast, too few operational taxonomic units (OTUs) of Oomycota were found to broadly drive high tropical tree diversity in a host-specific manner under the Janzen-Connell model. Extremely high OTU diversity and high heterogeneity between samples within the same forests suggest that protists, not arthropods, are the most diverse eukaryotes in tropical rainforests. Our data show that protists play a large role in tropical terrestrial ecosystems long viewed as being dominated by macroorganisms.

Since the works of early naturalists such as von Humboldt and Bonpland[1], we have known that animal and plant communities in tropical rainforests are exceedingly species rich. For example, one hectare can contain more than 400 tree species[2] and one tree can harbour more than 40 ant species[3]. This hyperdiversity of trees has been partially explained by the Janzen-Connell model[4,5], which hypothesizes that host-specific predators and parasites reduce plant population growth in a density-dependent manner[6,7]. Sampling up in the tree canopies and below on the ground has further led to the view that arthropods are the most diverse eukaryotes in tropical rainforests[8,9].

The focus on eukaryotic macroorganisms in these studies is primarily because they are familiar and readily observable to us. We do not know whether the less familiar and less readily observable protists—microbial eukaryotes that are not animals, plants or fungi[10]—inhabiting these same ecosystems exhibit similar diversity patterns. To evaluate if macroorganismic 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. This metabarcoding approach has the power to uncover known and new taxa on a massive scale[11]. By amplifying DNA extracted from the soils with broadly targeted primers for the V4 region of 18S rRNA and sequencing it using the Illumina MiSeq platform, we were able to detect most eukaryotic lineages, and assess the diversity and relative dominance of free-living and parasitic lineages.

Results

Sequencing of the Neotropical rainforest soil samples resulted in 132.3 million cleaned V4 reads (Supplementary Table 1). 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 database (PR^2)[12] (Fig. 1 and Supplementary Fig. 1). By contrast, our re-analysis of 367.8 million reads from the hypervariable V9 region of the 18S rRNA from the open oceans[11] produced only 3.1% reads <80% similar to the PR^2 database, and most of our new V4 and V9 reads from European near-shore marine environments were likewise highly similar to the PR^2 database (Fig. 1 and Supplementary Fig. 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 Cedex, France. [3]Sorbonne Universités, UPMC Université de Paris 06, UMR 7144, Station Biologique de Roscoff, Place Georges Teissier, 29680 Roscoff Cedex, France. [4]Department of Life Sciences, The Natural History Museum London, Cromwell Road, London SW7 5BD, UK. [5]Centre for Environment, Fisheries and Aquaculture Science, Barrack Road, The Nothe, Weymouth, Dorset DT4 8UB, UK. [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, Quito, Ecuador. *e-mail: dunthorn@rhrk.uni-kl.de

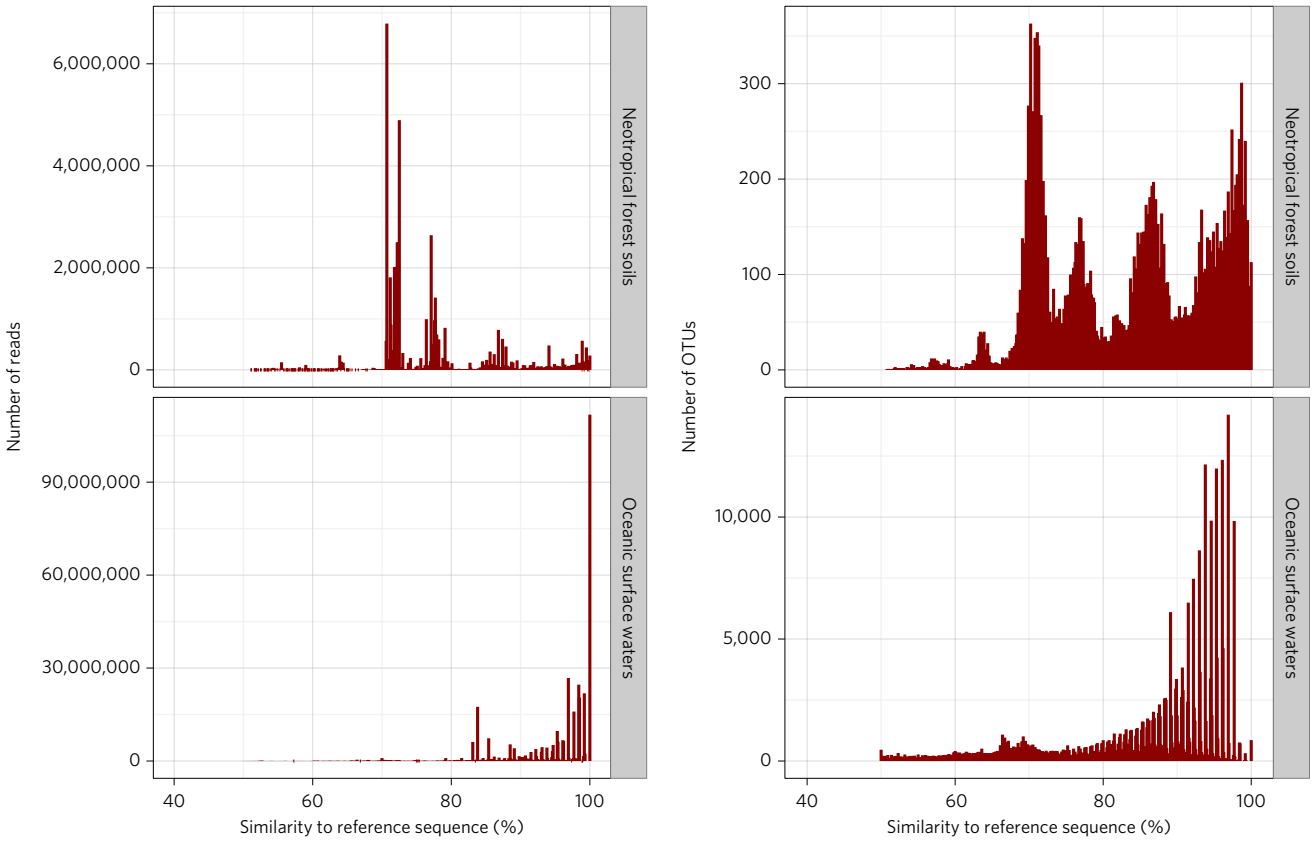

Figure 1 | Similarity of protists to the taxonomic reference database. In contrast to marine data, most of the reads and OTUs from the Neotropical rainforest soils were <80% similar to references in the PR^2 database. Only 8.1% of soil reads had a similarity ≥95%, whereas 68.1% of the marine reads from the *Tara* Oceans study of the world's open oceans had a similarity ≥95%.

Reads <80% similar to known references are often considered spurious and removed in environmental protist sequencing studies[13]. Three-quarters of our rainforest soil protist data would be discarded if we applied this conservative cleaning step. Furthermore, PR^2 and similar databases are biased towards marine and temperate references. To solve these problems, reads were dereplicated into 10.6 million amplicons (strictly identical reads to which an abundance value can be attached), and subsequently placed using the Evolutionary Placement Algorithm (EPA)[14], as implemented in RAxML[15], onto a phylogenetic tree inferred from 512 full-length references from all major eukaryotic clades (Fig. 2 and Supplementary Fig. 2). We conservatively retained operational taxonomic units (OTUs) constructed with Swarm[16] (Supplementary Fig. 3) whose most abundant amplicon fell only within known clades with a high likelihood-weight score. This novel phylogeny-aware cleaning step effectively discarded highly divergent amplicons[17], resulting in the removal of only 6.8% of the cleaned reads and 7.7% of the OTUs.

The retained protist reads from the Neotropical rainforest soils clustered into 26,860 OTUs. As in a recent sampling of the sunlit surface layer of the world's open oceans[11], more protist OTUs were detected than animals (4,374, of which 39% were assigned to the Arthropoda), plants (3,089) and fungi (17,849) combined (Supplementary Table 1). The OTUs found in the samples may not all correspond to soil-dwelling species: some of these hyperdiverse protists and other eukaryotes could be a shadow of the tree-canopy communities from cells that have rained down from above. Taxonomic assignment of the protists showed that 84.4% of the reads and 50.6% of the OTUs were affiliated to the Apicomplexa (Fig. 3 and Supplementary Fig. 4). Apicomplexa are widespread parasites of animals[18,19]. As an independent line of evidence for their dominance in the Neotropical rainforest soils, ten samples from both Costa Rica and Panama were amplified with primers designed to specifically target the closely related Ciliophora (both the Ciliophora and the Apicomplexa are in the Alveolata[20]) and sequenced using a Roche/454 pyrosequencer. From these ciliate-specific amplifications, 47.8% of the 297,892 reads and 28.1% of the 1,082 OTUs were assigned to the Apicomplexa (Supplementary Fig. 5). In contrast to the results presented here, read- and OTU-abundances of the Apicomplexa were observed to be substantially lower in marine and other terrestrial environments (Supplementary Fig. 4).

Using EPA we placed the Apicomplexa OTUs into a more focused phylogeny inferred from 190 full-length references from all major Alveolata clades. While the OTUs were generally distributed across the whole tree (Supplementary Fig. 6), 80.2% of them were grouped with the gregarines. Gregarines predominantly infect arthropods and other invertebrates[18]. About 23.8% of these gregarine OTUs were 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 were placed with two environmental gregarine sequences collected from brackish sediment[21]. Many other OTUs were grouped within gregarine lineages thought to be primarily parasites of marine annelids and polychaetes. Non-gregarine Apicomplexa OTUs were largely grouped with the blood parasites *Plasmodium* (some of which cause malaria) and close relatives, including those that can cycle through arthropods and vertebrates such as birds.

The second to fifth most diverse protist taxa were the predominantly predatory Cercozoa, Ciliophora, Conosa and Lobosa (Fig. 3), accounting for a combined total of 5,572,490 reads and 10,338 OTUs. Fewer photo- or mixo-trophic Chlorophyta, Dinophyta, Haptophyta, and Rhodophyta were also found, accounting for a combined total of 699,187 reads and 1,096 OTUs. Haptophyta, for

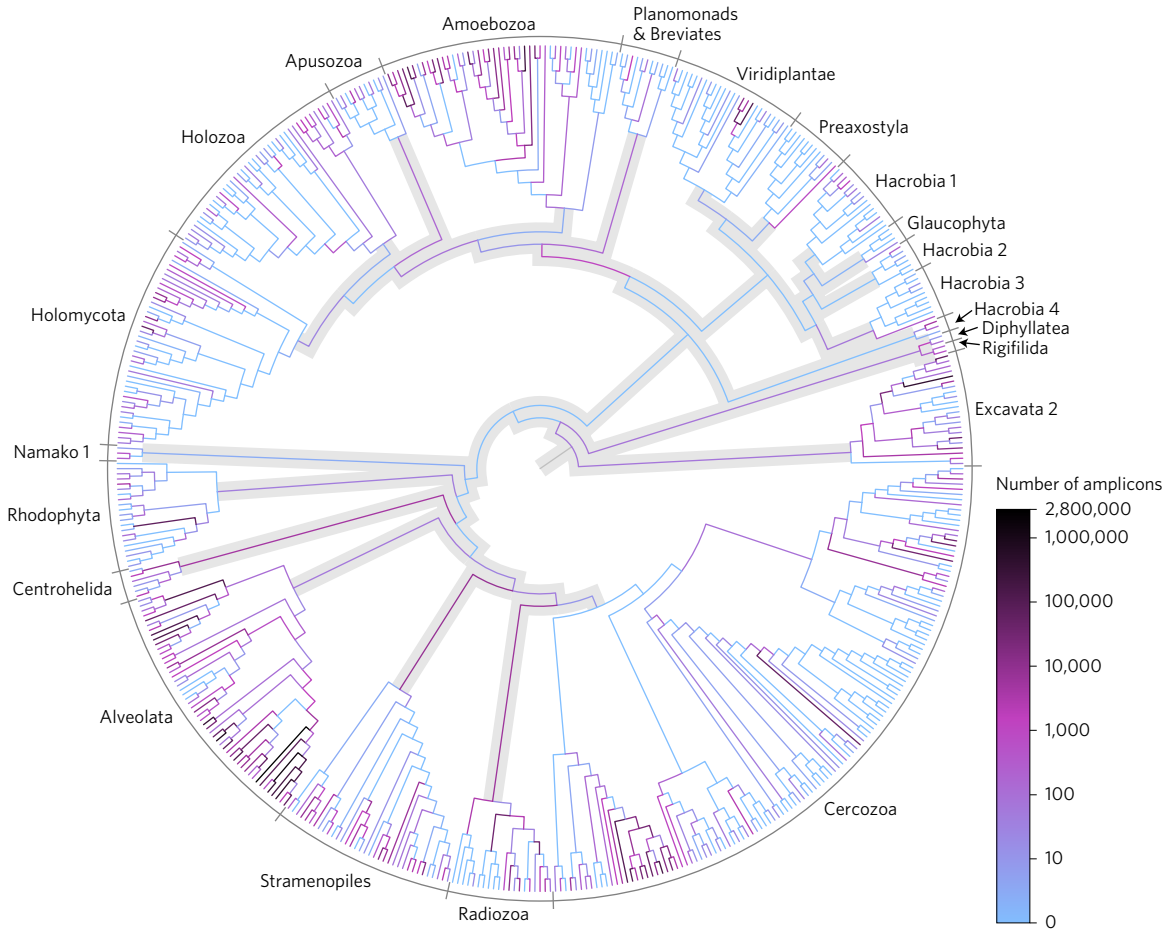

Figure 2 | Phylogenetic placement of Neotropical soil protist reads on a taxonomically unconstrained global eukaryotic tree. Reads were dereplicated into strictly identical amplicons. Inferred relationships between these major taxa may differ from those obtained with phylogenomic data. Alveolata includes Apicomplexa and Ciliophora; Holozoa includes animals; Holomycota includes fungi. Branches and nodes outside of known clades are shaded grey. In our conservative approach only OTUs that placed within known clades with high likelihood-weight scores were retained.

example, are mostly marine[22] but a phylogenetic inference of these OTUs showed that most of them had close relationships with freshwater species (Supplementary Fig. 7). Although algae are thought to affect the net uptake of carbon in some terrestrial ecosystems[23], the contribution of these soil-inhabiting algae to the carbon cycle in Neotropical rainforests is unknown. Only 131 OTUs were assigned to the Oomycota. Oomycota are efficient parasites with flagellated stages that disperse through soils. Using the best BLAST hits to GenBank references, we inferred close relationships between many of these Oomycota OTUs and taxa that infect either animals or plants (Supplementary Fig. 8).

Non-metric multidimensional scaling (Supplementary Fig. 9) and Bray-Curtis dendrograms (Supplementary Fig. 10) showed that the protist community composition was slightly more similar among the samples from Costa Rica and Panama than those from Ecuador, reflecting similar diversity patterns seen in animals and plants in Central and South American forests[24]. Within each forest, we estimated fewer than 1,732 unobserved OTUs exclusively from the samples taken (Supplementary Table 2), suggesting that our sequencing depth detected a high fraction of the total diversity within each of the samples. However, within-forest OTU rarefaction curves, based on the sample accumulation, estimated logarithmic increases without plateaus (Supplementary Fig. 11), and the Jaccard similarity index estimated high OTU heterogeneity between samples even within the same forest (Supplementary Fig. 12). This high heterogeneity between samples indicates that our sampling effort unveiled only a fraction of the protist hyperdiversity in the three Neotropical rainforests.

Discussion

The broad sampling and deep sequencing of soils from three Neotropical rainforests revealed numerous protist taxa. Much of this unravelled diversity was detected by our phylogeny-aware cleaning step that retained OTUs with low similarity to existing 18S rRNA references (Figs 1 and 2). Using EPA phylogenetic placements, rather than relying solely on pairwise sequence similarities, is a powerful approach to metabarcoding studies aimed at discovering novel taxa in environments where little is known and few references are available—such as tropical soils. Even if almost all the tips are missing in the tree—that is, there are no references closely related to the environmental data— a tree that comprises at least the major eukaryotic lineages allows phylogenetic placements to retain most of the novel diversity that would have otherwise been discarded. While the placements can be evaluated at the amplicon or OTU levels, placing only the OTU representatives produced by Swarm drastically reduces computation time without significant information loss.

Even with the phylogeny-aware cleaning step, our molecular approach probably underestimated diversity across all microbial eukaryotic and macroorganismic taxa. For example, while the broadly targeted V4 primers used here amplify a wide variety of eukaryotic lineages[13,25–29], other primers could have amplified additional taxa[30–32] (Supplementary Table 1). The resolution power

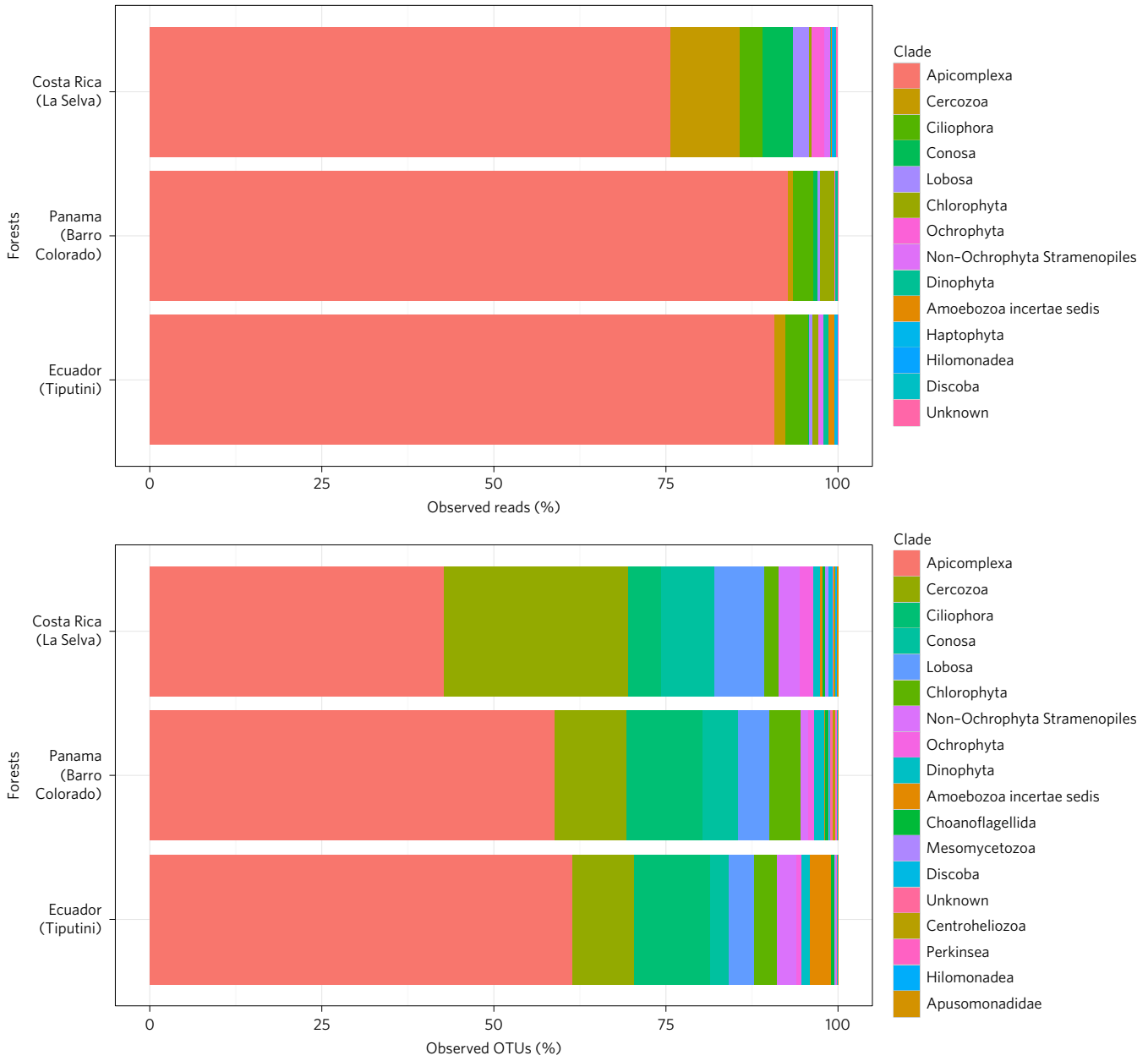

Figure 3 | Taxonomic identity and relative abundances of soil protist reads and OTUs in three Neotropical rainforests. Each taxon shown represents at least 0.1% of the total data. Across the three forests, 76.6–93.0% of the reads and 43.1–61.5% of the OTUs were assigned to the Apicomplexa.

of metabarcoding also has its limits. The Swarm clustering method, which relies on iterative local clustering thresholds, often results in fine-grained OTUs that can reveal additional diversity compared with traditional clustering methods that rely on global clustering thresholds[16,33] (Supplementary Fig. 3). Species with identical barcode sequences (that is, no variation between taxa) can nevertheless be indistinguishable by any clustering method, which may mask potentially different ecological roles and functions.

The parasitic Apicomplexa dominated the Neotropical soil protist communities by accounting for the majority of both reads and OTUs (Fig. 3). This pattern of dominating parasites contrasts with the prevailing view that soil protist communities are dominated by predators of bacteria[34], although a considerable presence of protist predators of fungi and animals, as well as protist parasites, has been observed elsewhere[35–41]. These dominating protist parasites also potentially contribute to the high animal diversity in the rainforests by the same mechanisms that other parasites contribute to high tree diversity as hypothesized in the Janzen-Connell model[4,5]. Apicomplexa infect animals (for example, refs [42,43]) and are usually host-specific, although host switching is known[18,19,42,43]. They therefore potentially limit population growth of species that become locally abundant. In this tentative model we expect Apicomplexa to dominate both the below- and aboveground protist communities in tropical forests worldwide, most animal populations in these forests will have apicomplexan parasites, and animals that become locally dominant will have escaped from their apicomplexan enemies or at least evolved to be able to cope with this burden. This relationship between animals and the Apicomplexa could also be reciprocal, with each group contributing to the diversity of the other. This top-down force by parasitic protists on tropical animal diversity may complement the bottom-up response that was proposed for plants on herbivorous insects[8].

In contrast to the Apicomplexa parasites, few parasitic Oomycota were detected in the Neotropical soil protist communities

(Supplementary Fig. 8). Along with fungi and insects, the Oomycota have long been thought to be one of the drivers of the Janzen-Connell model. However, the degree of host-specificity is unknown for most Oomycota species[46], and a fungicide study in Belize documented a non-significant effect by these protists[6]. If the Oomycota broadly drive the Janzen-Connell model, then we can expect their diversity to be very high, mirroring tree diversity. However, we did not find enough Oomycota OTUs for them to be host-specific plant parasites under this model; although, as mentioned above, OTUs can mask functional diversity.

Just as there is high species richness in Neotropical rainforests at the macroorganismic scale, the high OTU numbers of both free-living and parasitic protists (Supplementary Table 1) and the high heterogeneity between samples (Supplementary Fig. 12) show that such hyperdiversity also exists at the microbial eukaryotic scale, where its degree is even greater. Given the low similarity of protist compositions between samples, and the six-to-one ratio for protist to animal OTUs, a concerted and comparable count will probably show that protists are more diverse than arthropods in tropical rainforests. This would certainly be the case if, as we suggest, every arthropod species has at least one apicomplexan parasite, and the Apicomplexa are only one component of the total protist hyperdiversity (with fungi being the second most diverse group). If protists are the most diverse eukaryotes in tropical rainforests, it would not be due to inordinate speciation in just a few clades since the mid-Phanerozoic (for example, the beetles[47]), it would be because of the diversification of rich and functionally complex protist lineages, beginning in the early Proterozoic[48], that built up the multifaceted and interactive unseen foundation of these now familiar macroscopic terrestrial ecosystems.

Methods

Neotropical soil samples. A total of 279 soil samples were taken from the following locations: La Selva Biological Station, Costa Rica in October 2012 and June 2013; Barro Colorado Island, Panama in October 2012 and June 2013; and Tiputini Biodiversity Station, Ecuador in October 2013 (see Supplementary Code 1 for sample coordinates). The aim was to collect soils from around each field station (excluding areas where humans were not permitted and areas of secondary growth) without targeting any specific sub-environments (samples were collected from a variety of soil types, including swamp mud, sandy soils and hill-top dry soils). Fieldwork and government-issued sampling permits were made possible by the Organization for Tropical Studies and the Ministerio del Ambiente y Energía and Sistema Nacional de Areas de Conservación in Costa Rica, the Smithsonian Tropical Research Institute in Panama, and David Romo Vallejo at the Universidad San Francisco de Quito in Ecuador.

At the site of collection, the loose-leaf layer and stones were removed. A 5 ml sample of the top-most layer of surface soil (with visible plant roots and visible animals removed) was immediately placed into 8 ml of LifeGuard Soil Preservation Solution (Mo Bio). DNA was isolated using the PowerSoil DNA Elution Accessory Kit (Mo Bio). Like other large-scale sequencing studies[11,49], only DNA was used in this study, not RNA, which can also be retained in sediments, even ancient ones[50,51]. See Supplementary Fig. 13 to see that RNA can be found in cysts, and for a comparison of DNA and RNA abundances; in addition, the low similarity of OTUs among samples within each forest, even those that were collected near each other (Supplementary Fig. 12), point to the soils not having the taxa, that would be expected if most of the DNA was from a large seed bank of dead or inactive species that collected overtime. The samples were mostly taken at least 100 m apart. DNA from every two consecutive samples was combined in equal concentration to reduce costs; many of these samples failed to amplify and only those that worked are shown here. The combined samples were amplified with a high fidelity polymerase for the V4 region of 18S rRNA following ref. [52] with universal eukaryotic V4 primers (TAReuk454FWD1 and TAReukREV3)[13]. This primer pair has not shown a preference for amplifying Apicomplexa, and the V4 primers were able to amplify all Oomycota in GenBank in an *in silico* analysis.

After PCR cleanup, a second PCR with primers containing sample-specific tags and the Illumina MiSeq sequencing adapters was performed following the manufacturer's instructions. Amplicons were gel size-selected for appropriate length, and library quality was assessed using a Bioanalyzer 2100 (Agilent) and Qubit 2.0 (Life Technologies). Illumina MiSeq sequencing was performed at GATC Biotech AG (Konstanz, Germany), using v3 chemistry and the software MCS v2.3.0.3 and RTA v1.18.42 (Illumina). PhiX DNA (20%) was spiked into the MiSeq flowcell to increase base diversity. Twenty amplified products (each from a combined two samples) were loaded onto one MiSeq flowcell.

Ten samples from La Selva, Costa Rica, and ten samples from Barro Colorado Island, Panama, were also amplified with primers designed to specifically target the closely related Ciliophora[53]. After PCR cleanup, a second, nested PCR was performed with the above V4 primers, and then sequenced using a Roche/454 GS FLX+ Titanium system and v2.6 of the associated software. The 20 amplified products were loaded onto one-quarter of a plate.

Marine samples. Reads from the open oceans came from the *Tara* Oceans consortium[11]. Starting with the *Tara* Oceans raw reads (Sequence Read Archive accession numbers: PRJEB6610 and PRJEB7315) allowed us to treat the sequencing data with the same bioinformatic pipeline. Our pipeline's goal was to improve the sensitivity of OTU detection by applying most quality- and frequency-based filtering steps after clustering instead of applying them before. Removal of low-abundance OTUs was performed on the combined data and not at the sample level. We also replaced the paired-read merger FLASH (Fast Length Adjustment of SHort reads)[54] with the newer tool PEAR (Paired-End reAd merger)[55], which is advertised as yielding higher-quality results. After a careful investigation, it turned out that the large increase in OTU number we observed compared with de Vargas *et al.*[11] was mainly due to our use of PEAR rather than FLASH.

Samples from European near-shore marine sites (see Supplementary Code 1 for coordinates of each sample) came from the BioMarKs consortium[28,56]. We re-sequenced the samples here to obtain deeper sequencing. Total nucleic acids from sediment samples were extracted using a RNA Power Soil Total Isolation Kit combined with a DNA Elution Accessory Kit (Mo Bio) according to the manufacturer's instructions. To remove contaminating DNA in the RNA extracts, the samples were treated twice with DNase for 25 min at 37 °C with 2 U of Turbo DNA-free Kit (Ambion) following the manufacturer's instructions. RNA was reverse-transcribed into first-strand cDNA using Invitrogen's Superscript III (Thermo Fisher Scientific) following the manufacturer's instructions. The V4 region was amplified with the above primers, and the V9 region was amplified with general eukaryotic primers (1389F and 1510R)[57]. Amplified products were sequenced on an Illumina Genome Analyser IIx sequencer at CEA Genoscope (Évry, France).

Read cleaning and clustering. Fastq files were assembled via PEAR v0.9.8 using default parameters and converted to fasta format. Following ref. [52], assembled paired-end reads were filtered using Cutadapt[58] v1.9 and retained if they had both primers and no ambiguously named nucleotides (Ns). Reads were dereplicated into strictly identical amplicons with VSEARCH[59] v1.6.0, then clustered using Swarm[16] v2.1.5, with the parameter $d=1$ and the fastidious option on. The most abundant amplicon in each OTU was searched for chimeric sequences with VSEARCH, and their OTUs were removed even if they occurred in multiple samples.

Taxonomic assignment. Taxonomic assignment used VSEARCH's global pairwise alignments with the PR^2 database[12] v203 (based on GenBank release v203). Using Cutadapt, the PR^2 database was extracted for just the specific regions that were amplified and sequenced to allow for comparisons using a global pairwise alignment. Amplicons were assigned to their best hit, or co-best hits, in the reference database as reported by VSEARCH. The different steps of that taxonomic assignment strategy are grouped in a pipeline called Stampa[60]. Low-abundance OTUs were removed from the combined dataset only if they included ≤2 reads, were found in only one sample and were <99% similar to an accession in PR^2.

Stampa plots. To globally assess the results of Stampa, our taxonomic assignment pipeline, we produced 'Stampa plots'[60]. Stampa plots are simple distribution plots showing the number of reads per similarity value, where the similarity value is the best match between environmental and reference sequences. Stampa plots are a visual assessment of the taxonomic coverage of our reference database sequences: if most environmental reads or OTUs have high similarity values with references, then the coverage is good. Low similarity values indicate a lack of coverage, which can be a sign of amplification-sequencing artefacts (chimeras for instance), the presence of known taxa without reference sequences, the presence of new taxa, or if the coverage deficit is very large, a previously unexplored environment (such as Neotropical rainforest soils).

Phylogenetic placement pipeline. To put the 50,118,536 soil protist reads into a phylogenetic context, they were dereplicated into 10,567,804 strictly identical amplicons and placed onto a comprehensive eukaryotic reference tree. The corresponding multiple sequence alignment used to build this tree contained 512 full-length sequences from all major eukaryotic clades (see Supplementary Data 1 and 2 for GenBank accession numbers, sequences and clade assignments), based on the taxon sampling as summarized in a range of pan-eukaryotic phylogenomic studies reviewed in refs [61,62], with a bias towards lineages known to occur in soils from environmental sequencing studies; for example, refs [35,37,38]. To reduce phylogenetic artefacts, only high quality, full- or near-full length 18S rRNA reads were selected, and reads that have previously

been observed to form long branches were omitted. It is acknowledged that single gene trees are unable to resolve the backbone of the eukaryote phylogeny; however, this was not the intended purpose of this taxon selection or the phylogenetic analyses. The intent was to produce a comprehensive eukaryote sample tree on which to place the V4 reads generated by this study. Phylogenetic placements of query OTU representatives that were taxonomically assigned to the Apicomplexa were conducted using a full-length reference multiple sequence alignment and corresponding phylogeny comprising 190 taxa from all major Alveolate clades (see Supplementary Data 3 and 4 for GenBank accession numbers, sequences and clade assignments). The taxon sampling was collated from a range of publications relating to eukaryotic diversity in general and Alveolata diversity and phylogeny in particular; for example, refs [20,63–69]. It is also acknowledged that single gene trees are unable to resolve the backbone of the Alveolate phylogeny.

For a full explanation of the phylogenetic placement, see Supplementary Methods. In brief, we used RAxML[15] v8.1.15 to infer reference trees from the reference alignments, then used PaPaRa[70] v2.4 to align the query sequences to those reference alignments, and finally used the EPA[14] as implemented in RAxML to place the queries onto the trees. Placement results were visualized as 'heat trees' using Genesis[71] v0.2.0. The trees were inferred with and without taxonomic constraints, although the results were similar. We also assessed the quality of the phylogenetic placement positions to determine how confident the algorithm is when placing a query sequence on the branches of the reference tree by analysing the distribution of the likelihood weights for the placements, and by analysing the locality of placement distributions for each amplicon over the tree (Supplementary Fig. 14).

Phylogenetic placement of Haptophyta. For the 28 Haptophyta OTUs, we first downloaded 106 GenBank accessions following a previously described taxon-sampling method[22,72,73], and aligned them in MAFFT[74] v7. OTU representatives were then added to this alignment using the MAFFT 'add-in' function, and the tree was inferred with RAxML, using the substitution model GTR-G-I, the new rapid hill-climbing algorithm and 100 bootstrap runs.

Oomycete analyses. To tentatively assign functional roles of the Oomycota OTUs, we first compared each individual OTU with related sequences in GenBank using BLAST[75] for a taxonomic affiliation. In cases of identical BLAST hits, we proceeded to a quick phylogenetic neighbour joining analysis to determine as closely as possible its affiliation. Functional affiliation of the OTUs was then determined mostly based on Lara *et al.*[76].

Statistical analyses. We analysed frequency count data derived from OTU clustering to estimate the total (observed + unobserved) OTU richness in the population from which the observed samples were drawn. These data consisted of the number of OTUs with one representative in the sample (the 'singletons'), and then the number with two representatives, the number with three, and so on. We used two sets of statistical procedures. The first was implemented in the software package, Breakaway[77] v2.0. This method estimates a family of statistical models, known as Kemp-type distributions, from the frequency count data, by fitting ratios of successive frequency counts via nonlinear regression; an optimal selected model is returned, along with error and goodness-of-fit assessments. In the case of the combined (global) OTU data the lowest-order model was selected, which is equivalent to fitting a weighted linear regression to the ratios of successive frequency counts. The second was implemented in the software package CatchAll[78]. This method fits a family of mixed-Poisson models to the frequency data, as well as calculating a suite of nonparametric procedures. It returns an optimal selected model with errors and goodness-of-fit assessments, and comparisons of parametric and nonparametric analyses. CatchAll selected a mixture of three exponential-mixed Poisson models. In all cases, the results were virtually identical, showing almost no diversity extant in the population beyond that observed in the sample. The R package vegan[79,] was used to analyse frequency count data derived from OTU clustering. Different functions of Vegan were called to randomly subsample our samples (rrarefy function) and to estimate and compare species compositions, using Bray-Curtis distance and NMDS ordination (monoMDS function). Figures were made with R[80] and ggplot2[81].

Detecting RNA in cysts. To assess if viable and sequenceable RNA can be detected in protist cysts, we designed fluorescent *in situ* hybridization (FISH) probes specific to the kinetoplastids with 18S rRNA probes coupled to the Cy3 fluorochrome and designed to hybridize to ribosomes: KIN516, a wide-spectrum kinetoplastid probe (ACCAGACTTGTCCTCC); and Bodo1757, a probe designed on the V8 region and specific to the clade *Bodo* (CGAGCAAGTGAAACACTCGCC). Probes were checked for specificity using pure cultures of the *Neobodo designis* strain AND31 (AY965872) and the *Bodo saltans* strain SCCAP BS364. Cultures were fixed with three volumes of 4% paraformaldehyde for 2 h. Fixed cell suspensions were collected on a 0.2 µm nitrocellulose ISOPORE filter (Millipore) and rinsed with demineralized water. Hybridizations were performed at 45 °C for 90 min in a buffer containing 0.9 M NaCl, 20 mM Tris HCl, and 0.01% SDS, pH 7.2. Probes were used at a concentration of 50 ng µl^{-1}. The optimal amount of formamide in the hybridization buffer to reduce non-specific binding of the probe was set by performing hybridizations with increasing formamide concentrations for the different strains investigated. In general, hybridization with 35% formamide did not result in any unspecific binding of the probes to non-target organisms. For non-specific staining of all cells, DAPI (4′,6-diamidino-2-phenylindole) was added at a concentration of 1 µg ml^{-1} to the hybridization buffer. After hybridization, filters were washed for 15 min at 48 °C in a buffer containing 20 mM Tris-HCl (pH 7.2), 10 mM EDTA, 0.01% SDS and 80 mM NaCl, subsequently rinsed with distilled water and air dried. Prior to microscopic observation, filters were mounted with Citifluor solution. Probes were applied to soil taken from a garden compost in a garden in Switzerland. Soil was mixed and sieved through a 2 mm grid to remove large debris. The soil was then spread as a 1 cm layer on a 16 cm × 23 cm plastic tray, and wetted with autoclaved demineralized water (without flooding). Two layers of lens-cleaning tissues were spread on the surface, and four microscope cover slips measuring 24 mm × 50 mm each (Menzel Gläser) were laid on the wet tissues. The tray was left in the dark at 18 °C for one week. Protists attached to the coverslips were removed by rinsing with Neff's Amoeba Saline buffer and collected in a 10 ml tube. The cells were directly fixed with paraformaldehyde on a nitrocellulose filter (0.2 µm) and hybridized using the specifically designed probes along with DAPI as described above.

Measuring DNA and RNA. The DNA and RNA abundances of the 18S rRNA locus were measured in the following ciliates: *Colpoda magna*, from the American Type Culture Collection number 50128; *Dileptus* sp., from Carolina Biological Supply Company; *Halteria grandinella*, collected on October 2013 at a pond at the University of Kaiserslautern; *Paramecium tetraurelia* strain D4-2, provided by Martin Simon, Saarland University; and *Tetrahymena thermophila* wildtype B, provided by Josef Loidl, University of Vienna. Clonal populations were cultured with wheat grass powder medium inoculated with the bacterium *Klebsiella minuta*. Standard curves were made with cells from each species and were amplified for the 3′ end of 18S rRNA with the primers 1391-F (ref. [82]) and Euk-B (ref. [83]). PCR products were cleaned via a MinElute PCR Purification Kit from Qiagen and ligated to a pGEM®-T vector (Promega) and cloned. Plasmids were isolated with the FastPlasmid Mini Kit (5 Prime), and checked with an amplification using vector primers and Sanger sequencing. Plasmids underwent six tenfold dilutions (in series) to provide quantitative PCR (qPCR) standards ranging from 10^{-1} to 10^{-6} ng µl^{-1}. For qPCR of RNA, total RNA was extracted 12 separate times from single cells, first washed three times in Volvic water, then lysed using the Ambion Single Cell Lysis Kit (Thermo Fisher Scientific) with an integrated DNase step. cDNA was generated using the QuantiTect Reverse Transcription Kit (Qiagen) with another integrated DNase step. For qPCR of DNA, measurements were taken directly from 12 separate single cells. Amplifications were performed in a final volume of 20 µl, containing 10 µl of iQ SYBR Green Supermix (Bio-Rad), 1 µl of template DNA (cDNA or a single cell), 10 pg of each primer and 7 µl of RNase-free water. Cycling conditions were: 95 °C for 2 min; and 40 cycles of 95 °C for 15 s, 57 °C for 25 s, and 72 °C for 40 s. Melt curve data were collected after cycle 2 between 57 and 95 °C with the temperature increasing at a rate of 0.5 °C per 10 s. qPCR reactions were performed using an iQ Single Color Real-Time PCR Detection System (Bio-Rad), with a negative control and six tenfold serial dilutions in triplicates. The amplification efficiency E was estimated using the equation $E = (10^{-1/\text{slope}}) - 1$ with the calculated slope of the standard curves. We obtained efficiencies of between 90.1 and 105.5% and a coefficient of determination R^2 of between 0.987 and 1.000. Copy numbers were calculated as follows: number of copies = (amount × 6.022 × 10^{23}) / (length × 10^9 × 660), where amount is the mass of DNA or cDNA (in ng), length is the length of the linear fragment plus the plasmid, 660 is the average molecular weight (in daltons) of one base pair for double stranded DNA and 6.022 × 10^{23} is Avogadro's constant[84].

Code availability. All codes used in this study can be found in Supplementary Code 1 and 2 (HTML format).

Data availability. The data analysed in this study are available at GenBank's Sequence Read Archive under BioProject number PRJNA317860.

Received 7 July 2016; accepted 18 January 2017; published 20 March 2017

References

1. von Humboldt, A. & Bonpland, A. *Personal Narrative of Travels to the Equinoctial Regions of the New Continent, During the Years 1799–1804* (Longman, Hurst, Rees, Orme & Brown, 1814–1829).
2. Valencia, R., Balslev, H. & Paz Y Miño, G. High tree alpha-diversity in Amazonian Ecuador. *Biodivers. Conserv.* **3,** 21–28 (1994).
3. Wilson, E. O. The arboreal ant fauna of Peruvian Amazon forests: a first assessment. *Biotropica* **19,** 245–251 (1987).
4. Janzen, D. H. Herbivores and the number of tree species in tropical forests. *Am. Nat.* **104,** 501–528 (1970).

5. Connell, J. H. in *Dynamics of Population* (eds Den Boer, P. J. & Gradwell, G. R.) 298–392 (Wageningen, 1971).
6. Bagchi, R. *et al.* Pathogens and insect herbivores drive rainforest plant diversity and composition. *Nature* **506,** 85–88 (2014).
7. Terborgh, J. Enemies maintain hyperdiverse tropical forests. *Am. Nat.* **107,** 481–501 (2012).
8. Basset, Y. *et al.* Arthropod diversity in a tropical forest. *Science* **338,** 1481–1484 (2012).
9. Erwin, T. L. Tropical forests, their richness in Coleoptera and other arthropod species. *Coleopts. Bull.* **36,** 74–75 (1982).
10. Pawlowski, J. *et al.* CBOL Protist Working Group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS Biol.* **10,** e1001419 (2012).
11. de Vargas, C. *et al.* Eukaryotic plankton diversity in the sunlit ocean. *Science* **348,** 1261605 (2015).
12. Guillou, L. *et al.* The Protist Ribosomal Reference database (PR^2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.* **41,** D597–D604 (2013).
13. Stoeck, T. *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).
14. Berger, S. A., Krompass, D. & Stamatakis, A. Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood. *Syst. Biol.* **60,** 291–302 (2011).
15. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30,** 1312–1313 (2014).
16. Mahé, F., Rognes, T., Quince, C., de Vargas, C. & Dunthorn, M. Swarm v2: highly-scalable and high-resolution amplicon clustering. *PeerJ* **3,** e1761 (2015).
17. Dunthorn, M. *et al.* Placing environmental next-generation sequencing amplicons from microbial eukaryotes into a phylogenetic context. *Mol. Biol. Evol.* **31,** 993–1009 (2014).
18. Desportes, I. & Schrével, J. *Treatise on Zoology—Anatomy, Taxonomy, Biology: The Gregarines* Vols 1 & 2 (Brill, 2013).
19. Perkins, F. O., Barta, J. R., Clopton, R. E., Peirce, M. A. & Upton, S. J. in *An Illustrated Giude to the Protozoa* 2nd edn (eds Lee, J. J., Leedale, G. F. & Bradbury, P.) 190–369 (Society of Protozoologists, 2002).
20. Adl, S. M. *et al.* The revised classification of eukaryotes. *J. Eukaryot. Microbiol.* **59,** 429–493 (2012).
21. Dawson, S. C. & Pace, N. R. Novel kingdom-level eukaryotic diversity in anoxic environments. *Proc. Natl Acad. Sci. USA* **99,** 8324–8329 (2002).
22. Edvardsen, B., Egge, E. & Vaulot, D. Diversity and distribution of haptophytes revealed by environmental sequencing and metabarcoding — a review. *Perspec. Phycol.* **3,** 77–91 (2016).
23. Elbert, W. *et al.* Contribution of cryptogamic covers to the global cycles of carbon and nitrogen. *Nat. Geosci.* **5,** 459–462 (2012).
24. Gentry, A. H. *Four Neotropical Rainforests* (Yale Univ. Press, 1991).
25. Hu, S. K. *et al.* Protistan diversity and activity inferred from RNA and DNA at a coastal ocean site in the eastern North Pacific. *FEMS Microbiol. Ecol.* **92,** fiw050 (2016).
26. Filker, S., Gimmler, A., Dunthorn, M., Mahé, F. & Stoeck, T. Deep sequencing uncovers high and substantial novel diversity in protistan plankton of solar saltern ponds. *Extremophiles* **19,** 283–295 (2015).
27. Forster, D. *et al.* Benthic protists: the under-charted majority. *FEMS Microbiol. Ecol.* **92,** fiw120 (2016).
28. Logares, R. *et al.* Patterns of rare and abundant marine microbial eukaryotes. *Curr. Biol.* **24,** 813–821 (2014).
29. Richards, T. A. *et al.* Molecular diversity and distribution of marine fungi across 130 European environmental samples. *Proc. R. Soc. B* **282,** 20152243 (2015).
30. Adl, S. M., Habura, A. & Eglit, Y. Amplification primers of SSU rDNA for soil protists. *Soil Biol. Biochem.* **69,** 328–342 (2014).
31. Hu, S. K. *et al.* Estimating protistan diversity using high-throughput sequencing. *J. Eukaryot. Microbiol.* **62,** 688–693 (2015).
32. Lentendu, G. *et al.* Effects of long-term differential fertilization on eukaryotic microbial communities in an arable soil: a multiple barcoding approach. *Mol. Ecol.* **23,** 3341–3355 (2014).
33. Forster, D., Dunthorn, M., Stoeck, T. & Mahé, F. Comparison of three clustering approaches for detecting novel environmental microbial diversity. *PeerJ* **4,** e1692 (2016).
34. Adl, S. & Gupta, V. V. S. R. Protists in soil ecology and forest nutrient cycling. *Can. J. For. Res.* **36,** 1805–1817 (2006).
35. Bates, S. T. *et al.* Global biogeography of highly diverse protistan communities in soil. *ISME J.* **7,** 652–659 (2013).
36. Dumack, K., Müller, M. E. H. & Bonkowski, M. Description of *Lecythium terrestris* sp. nov. (Chlamydophryidae, Cercozoa), a soil dwelling protist feeding on fungi and algae. *Protist* **167,** 93–105 (2016).
37. Dupont, A., Griffiths, R. I., Bell, T. & Bass, D. Differences in soil micro-eukaryotic communities over soil pH gradients are strongly driven by parasites and saprotrophs. *Environ. Microbiol.* **18,** 2010–2024 (2016).
38. Geisen, S. *et al.* Metatranscriptomic census of active protists in soils. *ISME J.* **9,** 2178–2190 (2015).
39. Grossmann, L. *et al.* Protistan community analysis: key findings of a large-scale molecular sampling. *ISME J.* **10,** 2269–2279 (2016).
40. Foissner, W. *Colpodea (Cillophora)* (Protozoenfauna Vol. 4/1, Gustav Fischer Verlag, 1993).
41. Mitchell, E. A. D. Pack hunting by minute soil testate amoebae: nematode hell is a naturalist's paradise. *Environ. Microbiol.* **17,** 4145–4147 (2015).
42. Ellis, V. A. *et al.* Local host specialization, host-switching, and dispersal shape the regional distributions of avian haemosporidian parasites. *Proc. Natl Acad. Sci. USA* **112,** 11294–11299 (2015).
43. Rueckert, S., Villette, P. M. & Leander, B. S. 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).
44. Asghar, M. *et al.* Hidden costs of infection: chronic malaria accelerates telomere degradation and senescence in wild birds. *Science* **347,** 436–438 (2015).
45. Bouwma, A. M., Howard, K. J. & Jeanne, R. L. Parasitism in a social wasp: effect of gregarines on foraging behavior, colony productivity, and adult mortality. *Behav. Ecol. Sociobiol.* **59,** 222–233 (2005).
46. Freckleton, R. P. & Lewis, O. T. Pathogens, density dependence and the coexistence of tropical trees. *Proc. R. Soc. B* **273,** 2909–2916 (2006).
47. Farrell, B. D. "Inordinate fondness" explained: Why are there so many beetles? *Science* **281,** 555–559 (1998).
48. Knoll, A. H. Paleobiological perspectives on early eukaryotic evolution. *Cold Spring Harb. Perspect. Biol.* **6,** a016121 (2014).
49. Tedersoo, L. *et al.* Global diversity and geography of soil fungi. *Science* **346,** 1256688 (2014).
50. Orsi, W., Biddle, J. F. & Edgcomb, V. Deep sequencing of subseafloor eukaryotic rRNA reveals active fungi across marine subsurface provinces. *PLoS ONE* **8,** e56335 (2013).
51. Pawlowski, J., Esling, P., Lejzerowicz, F., Cedhagen, T. & Wilding, T. Environmental monitoring through protist next-generation sequencing metabarcoding: assessing the impact of fish farming on benthic foraminifera communities. *Mol. Ecol. Res.* **14,** 1129–1140 (2014).
52. Mahé, F. *et al.* Comparing high-throughput platforms for sequencing the hyper-variable V4 region in environmental eukaryotic diversity surveys. *J. Eukaryot. Microbiol.* **62,** 338–345 (2015).
53. Lara, E., Berney, C., Harms, H. & Chatzinotas, A. Cultivation-independent analysis reveals a shift in ciliate 18S rRNA gene diversity in a polycyclic aromatic hydrocarbon-polluted soil. *FEMS Microbiol. Ecol.* **62,** 365–373 (2007).
54. Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27,** 2957–2963 (2011).
55. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30,** 614–620 (2014).
56. Massana, R. *et al.* Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing. *Environ. Microbiol.* **17,** 4035–4049 (2015).
57. Amaral-Zettler, L. A., McCliment, E. A., Ducklow, H. W. & Huse, S. M. A method for studying protistian diversity using massively parallel sequencing of V9 hypervariable region of small-subunit ribosomal RNA genes. *PLoS ONE* **4,** 7 (2009).
58. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10–12 (2011).
59. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4,** e2584 (2016).
60. Mahé, F. Stampa: sequence taxonomic assignment by massive pairwise alignments. *GitHub* https://github.com/frederic-mahe/stampa (2016).
61. Burki, F. The eukaryotic tree of life from a global phylogenomic perspective. *Cold Spring Harb. Perspect Biol.* **6,** a016147 (2014).
62. del Campo, J. *et al.* The others: our biased perspective of eukaryotic genomes. *Trends Ecol. Evol.* **29,** 252–259 (2014).
63. Barta, J. R., Ogedengbe, J. D., Martin, D. S. & Smith, T. G. Phylogenetic position of the adeleorinid coccidia (Myzozoa, Apicomplexa, Coccidia, Eucoccidiorida, Adeleorina) inferred using 18S rDNA sequences. *J. Eukaryot. Microbiol.* **59,** 171–180 (2012).
64. Bass, D. & Cavalier-Smith, T. Phylum-specific environmental DNA analysis reveals remarkably high global biodiversity of Cercozoa (Protozoa). *Int. J. Syst. Evol. Microbiol.* **54,** 2393–2404 (2004).
65. Gomez, F., López-Garcia, P., Nowaczyk, A. & Moreira, D. The crustacean parasites *Ellobiopsis* Caullery, 1910 and *Thalassomyces* Niezabitowski, 1913 form a monophyletic divergent clade within the Alveolata. *Syst. Parasitol.* **74,** 65–74 (2009).

66. Rueckert, S. & Leander, B. S. Molecular phylogeny and surface morphology of marine archigregarines (Apicomplexa), *Selenidium* spp., *Filipodium phascolosomae* n. sp., and *Platyproteum* n. g. and comb. from North-Eastern Pacific peanut worms (Sipuncula). *J. Eukaryot. Microbiol.* **56,** 428–439 (2009).
67. Rueckert, S., Chantangsi, C. & Leander, B. S. Molecular systematics of marine gregarines (Apicomplexa) from North-eastern Pacific polychaetes and nemerteans, with descriptions of three novel species: *Lecudina phyllochaetopteri* sp. nov., *Difficilina tubulani* sp. nov. and *Difficilina paranemertis* sp. nov. *Int. J. Syst. Evol. Microbiol.* **60,** 2681–2690 (2010).
68. Slapeta, J. & Linares, M. C. Combined amplicon pyrosequencing assays reveal presence of the apicomplexan "type-N" (cf. *Gemmocystis cylindrus*) and *Chromera velia* on the Great Barrier Reef, Australia. *PLoS ONE* **8,** e76095 (2013).
69. Wakeman, K. C. & Leander, B. S. Molecular phylogeny of Pacific Archigregarines (Apicomplexa), including descriptions of *Veloxidium leptosynaptae* n. gen., n. sp., from the sea cucumber *Leptosynapta clarki* (Echinodermata), and two new species of *Selenidium*. *J. Eukaryot. Microbiol.* **59,** 232–245 (2012).
70. Berger, S. A. & Stamatakis, A. Aligning short reads to reference alignments and trees. *Bioinformatics* **27,** 2068–2075 (2011).
71. Czech, L. & Stamatakis, A. Genesis. A toolkit for working with phylogenetic data. *GitHub* https://github.com/lczech/genesis (2016).
72. Egge, E., Eikrem, W. & Edvardsen, B. Deep-branching novel lineages and high diversity of haptophytes in the Skagerrak (Norway) uncovered by 454 pyrosequencing. *J. Eukaryot. Microbiol.* **62,** 121–140 (2015).
73. Shalchian-Tabrizi, K., Reier-Røberg, K., Ree, D. K., Klaveness, D. & Bråte, J. Marine-freshwater colonizations of haptophytes inferred from phylogeny of environmental 18S rDNA sequences. *J. Eukaryot. Microbiol.* **58,** 315–318 (2011).
74. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30,** 3059–3066 (2002).
75. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).
76. Lara, E. & Belbahri, L. SSU rRNA reveals major trends in oomycete evolution. *Fungal Divers.* **49,** 93–100 (2011).
77. Willis, A. & Bunge, J. Estimating diversity via frequency ratios. *Biometrics* **71,** 1042–1049 (2015).
78. Bunge, J. *et al.* Estimating population diversity with CatchAll. *Bioinformatics* **28,** 1045–1047 (2012).
79. Oksanen, J. *et al.* vegan: Community Ecology Package. R package version 2.0-7 (2013).
80. R Core Team *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2014).
81. Wickham, H. *gplot2: Elegant Graphics for Data Analysis* (Springer, 2009).
82. Lane, D. J. in *Nucleic Acids Techniques in Bacterial Systematics* (eds Stackebrandt, E. & Goodfellow, M.) 115–147 (John Wiley & Sons, 1991).
83. Medlin, L., Elwood, H. J., Stickel, S. & Sogin, M. L. The characterization of enzymatically amplified eukaryotes 16S-like ribosomal RNA coding regions. *Gene* **71,** 491–500 (1988).
84. Zhu, F., Massana, R., Not, F., Marie, D. & Vaulot, D. Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiol. Ecol.* **52,** 79–92 (2005).

Acknowledgements

We thank T. Stoeck, L. Katz, H. Kauserud, O. T. Lewis and D. Montagnes for helpful comments. Funding primarily came from: a Deutsche Forschungsgemeinschaft grant DU1319/1-1 to M.D.; the Klaus Tschira Foundation to L.C., A.S and A.K; and the French Government's 'Investissements d'Avenir' program OCEANOMICS (ANR-11-BTBR-0008) to C.d.V., C.B. and S.R. Funding also came from: the Natural Environment Research Council grants NE/H000887/1 and NE/H009426/1 to D.B.; the National Science Foundation's International Research Fellowship Program (OISE-1012703) and the Smithsonian Tropical Research Institute's Fellowship program to J.M.; the Swiss National Science Foundation (310003A 143960); and the Swiss Federal Office for the Environment and Swiss National Science Foundation to E.L. and E.A.D.M. The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time on the GCS Supercomputer SuperMUC at the Leibniz Supercomputing Centre (www.lrz.de).

Author contributions

F.M. and M.D. conceived the project. F.M., C.d.V., J.M., T.S., S.R. and M.D. collected the samples. Sequencing was carried out by F.M., T.S., I.T., S.R and M.D. Data analysis was done by F.M., D.B., L.C., A.S., E.L., D.S., J.B., S.S., I.T., C.B., A.K., E.E. and M.D. The first draft of the manuscript was written by F.M., C.d.V., D.B., L.C., A.S. and M.D., and all authors contributed to discussing the results and editing the manuscript.

Additional information

Supplementary information is available for this paper.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to M.D.

How to cite this article: Mahé, F. *et al.* Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. *Nat. Ecol. Evol.* **1,** 0091 (2017).

Competing interests

The authors declare no competing financial interests.