Global diversity and geography of soil fungi

Leho Tedersoo[1]*†, Mohammad Bahram[2]†, Sergei Põlme[1], Urmas Kõljalg[2], Nourou S. Yorou[3], Ravi Wijesundera[4], Luis Villarreal Ruiz[5], Aída M. Vasco-Palacios[6], Pham Quang Thu[7], Ave Suija[2], Matthew E. Smith[8], Cathy Sharp[9], Erki Saluveer[2], Alessandro Saitta[10], Miguel Rosas[11], Taavi Riit[2], David Ratkowsky[12], Karin Pritsch[13], Kadri Põldmaa[2], Meike Piepenbring[11], Cherdchai Phosri[14], Marko Peterson[2], Kaarin Parts[2], Kadri Pärtel[2], Eveli Otsing[2], Eduardo Nouhra[15], André L. Njouonkou[16], R. Henrik Nilsson[17], Luis N. Morgado[18], Jordan Mayor[19], Tom W. May[20], Luiza Majuakim[21], D. Jean Lodge[22], Su See Lee[23], Karl-Henrik Larsson[24], Petr Kohout[2], Kentaro Hosaka[25], Indrek Hiiesalu[2], Terry W. Henkel[26], Helery Harend[2], Liang-dong Guo[27], Alina Greslebin[28], Gwen Grelet[29], Jozsef Geml[18], Genevieve Gates[12], William Dunstan[30], Chris Dunk[19], Rein Drenkhan[31], John Dearnaley[32], André De Kesel[33], Tan Dang[7], Xin Chen[34], Franz Buegger[13], Francis Q. Brearley[35], Gregory Bonito[20], Sten Anslan[2], Sandra Abell[36], Kessy Abarenkov[2]

[1]Natural History Museum, University of Tartu, Tartu, Estonia.

[2]Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia.

[3]Faculté d´Agronomie, Université de Parakou, Parakou, Benin.

[4]Department of Plant Sciences, University of Colombo, Colombo 3, Sri Lanka.

[5]Postgrado en Recursos Genéticos y Productividad-Genética, LARGEMBIO, Colegio de Postgraduados-LPI 6, México City, Mexico.

[6]The Fungal Biodiversity Centre, CBS-KNAW, Utrecht, The Netherlands.

[7]Vietnamese Academy of Forest Sciences, Hanoi, Vietnam.

[8]Department of Plant Pathology, University of Florida, Gainesville, Florida, USA.

[9]Natural History Museum, Bulawayo, Zimbabwe.

[10]Department of Agricultural and Forest Sciences, Università di Palermo, Palermo, Italy.

[11]Department of Mycology, Goethe University Frankfurt, Frankfurt am Main, Germany.

[12]Tasmanian Institute of Agriculture, Hobart, Tasmania, Australia.

[13]Institute of Soil Ecology, Helmholtz Zentrum München, Neuherberg, Germany.

[14]Department of Biology, Nakhon Phanom University, Nakhon Phanom, Thailand.

[15]Instituto Multidisciplinario de Biología Vegetal, Córdoba, Argentina.

[16]Department of Biological Sciences, University of Bamenda, Bambili, Cameroon.

[17]Department of Biological and Environmental Sciences, University of Gothenburg, Göteborg, Sweden.

[18]Naturalis Biodiversity Center, Leiden, The Netherlands.

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

[20]Royal Botanic Gardens Melbourne, Melbourne, Victoria, Australia.

[21]Institute for Tropical Biology and Conservation, University Malaysia Sabah, Sabah, Malaysia.

[22]Center for Forest Mycology Research, USDA-Forest Service, Luquillo, Puerto Rico.

[23]Forest Research Institute Malaysia, Kepong, Selangor, Malaysia.

[24]Natural History Museum, University of Oslo, Oslo, Norway.

[25]Department of Botany, National Museum of Nature and Science, Tsukuba, Japan.

[26]Department of Biological Sciences, Humboldt State University, Arcata, California, USA.

[27]State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China.

[28]CONICET - Facultad de Cs. Naturales, Universidad Nacional de la Patagonia SJB, Esquel, Chubut, Argentina.

[29]Ecosystems and Global Change team, Landcare Research, Lincoln, New Zealand.

[30]School of Veterinary & Life Sciences, Murdoch University, Western Australia, Australia.

[31]Institute of Forestry and Rural Engineering, Estonian University of Life Sciences, Tartu, Estonia.

[32]Faculty of Health, Engineering and Sciences, University of Southern Queensland, Toowoomba, Queensland, Australia.

[33]Botanic Garden Meise, Meise, Belgium.

[34]College of Life Sciences, Zhejiag University, Hangzhou 310058, China.

[35]School of Science and the Environment, Manchester Metropolitan University, Manchester, United Kingdom.

[36]School of Marine and Tropical Biology, James Cook University, Cairns, Queensland, Australia.

†Equal contribution

*Corresponding author. E-mail: leho.tedersoo@ut.ee

Abstract

Fungi play major roles in ecosystem processes, but the determinants of fungal diversity and biogeographic patterns remain poorly understood. By using DNA metabarcoding data from hundreds of globally distributed soil samples, we demonstrate that fungal richness is decoupled from plant diversity. The plant-to-fungus richness ratio declines exponentially towards the poles, indicating strong biases in previous fungal diversity estimates. Climatic factors, followed by edaphic and spatial variables, constitute the best predictors of fungal richness and community composition at the global scale. Fungi follow general biogeographic patterns related to latitudinal diversity gradients but with several notable exceptions. These findings significantly advance our understanding of fungal diversity patterns at the global scale and permit integration of fungi into a general macro-ecological framework.

One-sentence summary

A massive, global-scale metagenomic study detects hotspots of fungal diversity and macroecological patterns, and indicates that plant and fungal diversity are uncoupled.

INTRODUCTION: The kingdom Fungi is one of the most diverse groups of organisms on Earth and they are integral ecosystem agents that govern soil carbon cycling, plant nutrition, and pathology. Fungi are widely distributed in all terrestrial ecosystems, but the distribution of species, phyla, and functional groups has been poorly documented. Based on 365 global soil samples from natural ecosystems, we determined the main drivers and biogeographic patterns of fungal diversity and community composition.

RATIONALE: We identified soil-inhabiting fungi using 454 pyrosequencing and comparison against taxonomically and functionally annotated sequence databases. Multiple regression models were used to disentangle the roles of climatic, spatial, edaphic, and floristic parameters on fungal diversity and community composition. Structural equation models were used to determine the direct and indirect effects of climate on fungal diversity, soil chemistry and vegetation. We also examined if fungal biogeographic patterns matched paradigms derived from plants and animals — namely, that species' latitudinal ranges increase towards the poles (Rapoport's rule) and diversity increases towards the equator. Finally, we sought group-specific global biogeographic links among major biogeographic regions and biomes using a network approach and area-based clustering.

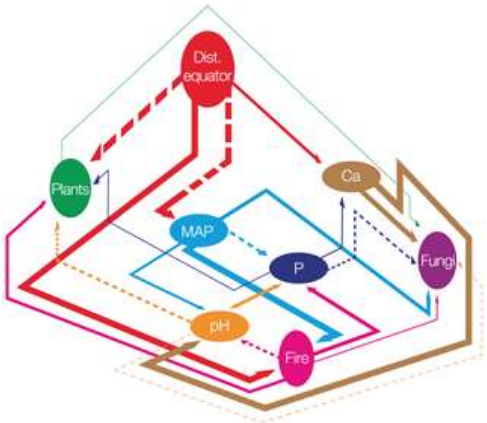
RESULTS: Metabarcoding analysis of global soils revealed fungal richness estimates approaching the number of species recorded to date. Distance from equator and mean annual precipitation had the strongest effects on richness of fungi including most fungal taxonomic and functional groups. Diversity of most fungal groups peaked in tropical ecosystems, but ectomycorrhizal fungi and several fungal classes were most diverse in temperate or boreal ecosystems and many fungal groups exhibited distinct preferences for specific edaphic conditions (e.g. pH, calcium, phosphorus). Consistent with Rapoport´s rule, the geographic

range of fungal taxa increased toward the poles. Fungal endemicity was particularly strong in tropical regions, but multiple fungal taxa had cosmopolitan distribution.

CONCLUSIONS: Climatic factors, followed by edaphic and spatial patterning, are the best predictors of soil fungal richness and community composition at the global scale. Richness of all fungi and functional groups is causally unrelated to plant diversity with the exception of ectomycorrhizal root symbionts, suggesting that plant-soil feedbacks do not influence the diversity of soil fungi at the global scale. The plant-to-fungi richness ratio declined exponentially towards the poles, indicating that current predictions assuming globally constant ratios overestimate fungal richness by 1.5-2.5-fold. Fungi follow similar biogeographic patterns as plants and animals with the exception of several major taxonomic and functional groups that run counter to overall patterns. Strong biogeographic links among distant continents reflect relatively efficient long-distance dispersal compared with macro-organisms.

Figure caption

Direct and indirect effects of climatic and edaphic variables on plant and fungal richness. Line thickness corresponds to relative path coefficients. Dashed lines indicate negative relationships. Abbreviations: MAP, mean annual precipitation; Fire, time since last fire.



Introduction

Fungi are eukaryotic microorganisms that play fundamental ecological roles as decomposers, mutualists, or pathogens of plants and animals; they drive carbon cycling in forest soils, mediate mineral nutrition of plants, and alleviate carbon limitations of other soil organisms. Fungi comprise some 100,000 described species (accounting for synonyms), but the actual extent of global fungal diversity is estimated at 0.8 to 5.1 million species (*1*).

Globally, the biomass and relative proportions of microbial groups, including fungi, co-vary with the concentration of growth-limiting nutrients in soils and plant tissues. Such patterns suggest that the distribution of microbes reflects latitudinal variation in ecosystem nutrient dynamics (*2-4*). Richness of nearly all terrestrial and marine macro-organisms is negatively related to increasing latitude (*5*) — a pattern attributed to the combined effects of climate, niche conservatism, and rates of evolutionary radiation and extinction (*6*). Although morphological species of unicellular microbes are usually cosmopolitan (*7*), there is growing evidence that the distribution of micro-organisms is shaped by macro-ecological and community assembly processes (*8*). Only a few of these biogeographic processes have been demonstrated for fungi at the local scale (*9*). Despite their enormous diversity and importance in ecosystem function, little is known about general patterns of fungal diversity or functional roles over large geographic scales. Here we use a global dataset to disentangle the roles of climatic, edaphic, floristic, and spatial variables governing global-scale patterns of soil fungal diversity. We also address macro-ecological phenomena and show that fungi largely exhibit strong biogeographic patterns that appear to be driven by dispersal limitation and climate.

Materials and Methods

Sample preparation

We collected 40 soil cores from natural communities in each of 365 sites across the world using a uniform sampling protocol (Fig. 1A; Data S1). Most plots (2500 m^2) were circular, but in steep mountain regions and densely forested areas, some plots were oblong. We randomly selected twenty trees located at least 8 m apart. In two opposite directions, 1-1.5 m from each tree trunk, loose debris was removed from the forest floor. PVC tubes (5 cm diam.) were hammered into the soil down to 5 cm depth. These soil cores almost always included fine roots and comprised both the organic layer and top mineral soil. Although deep soil may contain some unique organisms adapted to anoxic conditions or low nutrient levels, our sampling was limited to topsoil for the following reasons. First, in the vast majority of soil types, >50% of microbial biomass and biological activity occur in the topmost organic soil layer. Second, deeper sampling was impossible in shallow, rocky soils or those with high clay concentrations and hardpans. Third, differences among soil horizons may be masked by other variables across large geographic scales (*10*). The 40 soil cores taken in each site were pooled, coarse roots and stones removed, and a subset of the soil was air-dried at <35 °C. Dried soil was stored in zip-lock plastic bags with silica gel to minimize humidity during transit. In the laboratory, dried soil was ground into fine powder using bead beating.

DNA was extracted from 2.0 g of soil using the PowerMax Soil DNA Isolation kit (MoBio, Carlsbad, CA USA) following manufacturer's instructions. PCR was performed using a mixture of six forward primers (in equimolar concentration) analogous to ITS3 and a

degenerate reverse primer analogous to ITS4 (hereafter referred to as ITS4ngs). Forward and reverse primers were shortened and modified to completely match >99.5% of all fungi (except ca. 60% of Tulasnellaceae that exhibit highly divergent 5.8S rDNA and Microsporidia that exhibit re-arrangements in ribosomal DNA; Table S1). The ITS4ngs primer was tagged with one of 110 identifiers (MIDs, 10-12 bases) that were modified from those recommended by Roche to differ by >3 bases, start only with adenosine, and consist of between 30-70% adenosine and thymidine in order to optimize the adapter ligation step. The PCR cocktail consisted of 0.6 µl DNA extract, 0.5 µl each of the primers (20 pmol), 5 µl 5xHOT FIREPol Blend Master Mix (Solis Biodyne, Tartu, Estonia), and 13.4 µl double-distilled water. PCR was carried out in four replicates using the following thermocycling conditions: an initial 15 min at 95 °C, followed by 30 cycles at 95 °C for 30 s, 55 °C for 30 s, 72 °C for 1 min, and a final cycle of 10 min at 72 °C. PCR products were pooled and their relative quantity was estimated by running 5 µl amplicon DNA on 1% agarose gel for 15 min. DNA samples yielding no visible band were re-amplified using 35 cycles in an effort to obtain sufficient PCR product, whereas samples with a very strong band were re-amplified with only 25 cycles. It is important to use as few cycles as possible to minimize chimera formation and to be able to interpret sequence abundance in a semiquantitative manner (*11*). We used negative (for DNA extraction and PCR) and positive controls throughout the experiment. Amplicons were purified with Exonuclease I and FastAP thermosensitive alkaline phosphatase enzymes (Thermo Scientific, Pittsburgh, PA USA). Purified amplicons were subjected to quantity normalization with a SequalPrep Normalization Plate Kit (Invitrogen, Carlsbad, CA, USA) following manufacturer's instructions. Normalized amplicons were divided into five pools that were subjected to 454 adaptor ligation, emulsion PCR, and 454 pyrosequencing using the GS-FLX+

technology and Titanium chemistry as implemented by Beckman Coulter Genomics (Danvers, MA, USA).

Bioinformatics

Pyrosequencing on five half-plates resulted in 2,512,068 reads with a median length of 409 bases. The sequences were re-assigned to samples in mothur 1.32.2 (www.mothur.org) based on the barcodes and then trimmed (parameters: minlength=300; maxambigs=1; maxhomop=12; qwindowaverage=35; qwindowsize=50; bdiffs=1) to exclude short and low-quality sequences, resulting in 2,231,188 high quality sequences. We used ITSx 1.0.7 (http://microbiology.se/software/itsx) to remove the flanking 5.8S and 28S rRNA genes for optimal resolution of ITS2 clustering and removal of compromised and non-target sequences. As a filter to remove most of the partial sequences we retained only sequences >99 bp in length. Chimera control was exercised through UCHIME 4.2 (www.drive5.com/uchime/). After these filtering steps, 1,397,679 sequences were retained and further clustered at 90.0% and 95.0-99.0% sequence similarity thresholds (*12*) as implemented in CD-Hit 4.6.1 (www.cd-hit.org). Clustering revealed 37,387, 59,556, 66,785, 77,448, 94,255, and 157,956 taxa based on 90.0%, 95.0%, 96.0%, 97.0%, 98.0%, and 99.0% sequence similarity thresholds, respectively. The longest sequence of each Operational Taxonomic Unit (OTU), based on clustering at 98.0% sequence similarity, was selected as the representative for BLASTn searches (word size=7; penalties: gap=-1; gap extension=-2; match=1) against the International Nucleotide Sequence Databases Collaboration (INSDC: www.insdc.org) and UNITE (unite.ut.ee) databases. In addition, we ran BLASTn searches against established reference sequences of all fungi in 99.0% similarity clusters that include third-party taxonomic and

metadata updates (*12*) as implemented in the PlutoF workbench (*13*). For each query, we considered the 10 best-matching references to annotate our global sequences as accurately as possible. If no reliable taxon name was available, we ran manual BLASTn searches against INSDC with 500 best matching sequences as output. We typically relied on 90%, 85%, 80%, and 75% sequence identity as a criterion for assigning OTUs with names of a genus, family, order, or class, respectively. Sequence identity levels were raised in subsets of Sordariomycetes, Leotiomycetes, and Eurotiomycetes, because these taxa contain multiple genera and families that have unusually conserved ITS sequences. As a rule, we considered e-values of BLASTn search results $<e^{-50}$ reliable to assign sequences to the fungal kingdom, whereas those $>e^{-20}$ were considered ´unknown´. E-values between e^{-20} and e^{-50} were manually checked against the 10 best matches for accurate assignment. We followed INSDC for higher-level taxonomy of eukaryotes (*14*) and the Index Fungorum (www.indexfungorum.org) for species through class-level taxonomy of fungi. Our group of taxonomic experts assigned each fungal genus, family, or order to functional categories (Data S2). If different functional categories were present within a specific genus, we chose the dominant group (>75% of species assigned to a specific category) or considered its ecology unknown (<75% of species assignable to a single category). All Glomeromycota were considered to be arbuscular mycorrhizal (AM). Taxa were considered to be ectomycorrhizal (EcM) if they best matched any sequences of known EcM lineages (*15*) and exhibited sequence length / BLASTn scores above lineage-specific thresholds. For several taxonomic groups, we constructed phylogenetic trees to assess the performance of clustering, sequence quality of singletons, accuracy of OTU separation, and taxonomic assignments (Fig. S1). In the course of this project, we provided

10,232 third-party taxonomic re-annotations to INSDC sequences to improve subsequent identification of fungal sequences and made these available through the UNITE database.

Statistical analyses

Estimates of the mean annual temperature (MAT), mean annual precipitation (MAP), soil moisture, and soil carbon at 30 arc second resolution were obtained from the WorldClim database (www.worldclim.org). Estimates of potential evapotranspiration (PET) and net primary productivity (NPP) at 30 arc minute resolution were obtained from the Atlas of the Biosphere (www.sage.wisc.edu/atlas/maps.php). Variation coefficients for MAT and MAP were computed based on the average monthly values to represent seasonality of temperature and precipitation. We also calculated the difference of MAP to PET to evaluate the effect of rainfall surplus or deficit. Based on vegetation type and geographical distribution, sites were categorized into biogeographic regions and biomes following the classification of the World Wildlife Foundation (http://worldwildlife.org) with a few exceptions: i) temperate deciduous forests in the Northern and Southern hemispheres were treated separately; ii) tropical montane forests (>1500 m elevation) were separated from the tropical lowland moist forests; and, iii) grasslands and shrublands of all geographic origins were pooled. At each site, we also determined the age of vegetation, time since the last fire, and EcM plant species along with their relative contribution to stand basal area. EcM plants are usually conspicuous trees or prominent shrubs that are relatively easy to identify and their mycorrhizal status is verifiable in the field using root excavation and microscopy. Complete lists of tree species were available for <10% of the sites, so we did not directly include plant community composition parameters in our analyses (but see below).

263 Concentrations of N, C, $^{13}C/^{12}C$, and $^{15}N/^{14}N$ were determined from 1-20 mg of soil
264 using GC-combustion coupled to isotope-ratio mass spectrometry (*16*). Concentrations of soil
265 calcium, potassium, magnesium, and phosphorus were determined as in Tedersoo et al. (*16*).
266 Soil pH was measured in 1 N KCl solution.
267 For analyses of fungal richness, we calculated residuals of OUT richness in relation to
268 the square root of the number of obtained sequences to account for differences in sequencing
269 depth. This method outperformed the commonly used rarefaction to the lowest number of
270 sequences method, which removes most of the data (*17*). We also calculated the richness of
271 major class-level taxonomic and functional groups (comprising >100 OTUs). We excluded
272 outlying samples dominated by a few OTUs of molds, which are indicative of poor sample
273 preservation (relative abundance of sequences belonging to Trichocomaceae >5%,
274 Mortierellaceae >20%, or Mucoraceae >20%, that exceeded three times the mean + standard
275 deviation). Although these samples were fairly homogeneously distributed across the world,
276 they had conspicuously lower fungal richness. We also excluded samples that yielded less than
277 1200 sequences per sample.
278 To determine the relationship between plant and fungal richness, we relied on co-
279 kriging values from the global vascular plant species richness dataset (*18*), which covered
280 96.7% of our sites. These scale-free values of plant richness were then regressed with residuals
281 from the best fit models for fungal richness and fungal functional groups. We further calculated
282 the ratio of relative plant richness to fungal richness and fitted this ratio with latitude using
283 polynomial functions to test the assumed uniformity of plant-to-fungal richness ratios at the
284 global scale (*1, 19, 20*). To account for potential latitudinal biases in plant-to-fungal diversity
285 estimates, we took into account the non-uniform distribution of land surfaces by calculating an

286 Inverse Distance Weighting (IDW) spatial interpolation of standardized ratios of plant-to-
287 residual fungal diversity using the *gstat* package in R (*21*). We then used IDW to interpolate
288 total fungal diversity beyond sampling sites, by accounting for MAP as based on the best-
289 fitting multiple regression model.

290 Distance from the equator, altitude, age of vegetation, time since last fire, climatic
291 variables, and concentrations of nutrients were log-transformed prior to analyses to improve
292 the distribution of residuals and reduce non-linearity. To account for potential autocorrelation
293 effects, we calculated spatial eigenvectors using SAM ver. 4 (*22*). To determine the best
294 predictors of global fungal diversity, we included edaphic, climatic, floristic, and spatial
295 variables in multiple regression models. Due to the large number of predictors, we pre-selected
296 16 candidate predictors that were revealed by exploratory multiple linear and polynomial
297 regression analyses, based on coefficients of determination and forward selection criteria. The
298 most parsimonious models were determined based on the corrected Akaike information
299 criterion (AICc), which penalizes over-fitting. Finally, components of the best models were
300 forward-selected to determine their relative importance as implemented in the *packfor* package
301 in R.

302 To test the direct effects of climatic variables on richness of fungi and their functional
303 groups, and indirect climatic effects (via soil nutrients and vegetation), we used Structural
304 Equation Modeling (SEM) in Amos ver. 22 (SPSS Software, Chicago, IL, USA). Model fits
305 were explored based on both chi-square test and Root Mean Square Error of Approximation
306 (RMSEA). First, we included all potentially important variables (inferred from both the
307 multiple regression models and correlations for individual response variables to construct
308 separate SEM models. We tested all direct and indirect relations between exogenous and

endogenous variables including their error terms. Then, we used backward elimination to remove non-significant links to maximise whole model fit. Finally, we combined the obtained SEM models in a unified path model, following the same elimination procedure.

In addition to full models, we specifically tested the relationships between OTU richness and distance from the equator and soil pH, because these or closely related variables were usually among the most important predictors. For these analyses, we calculated residuals of richness that accounted for other significant variables of the best models. To address non-linear relationships, we fitted up to fifth order polynomial functions and selected best fit models based on AICc values.

The relative effects of climatic, edaphic, spatial, and floristic variables on the total fungal community composition and on particular functional groups were determined using Hellinger dissimilarity (calculated if >90% sites were represented by >1 shared OTUs), exclusion of all OTUs that occurred once, and a multi-stage model selection procedure as implemented in the DISTLM function of Permanova+ (www.primer-e.com/permanova.htm). Considering computational requirements, 15 candidate variables were pre-selected based on unifactorial (marginal test based on largest F_{pseudo} values) and multifactorial (forward selection) models. Spatial eigenvectors were not included in these analyses, because they were typically of minor importance in variation partitioning analyses (see below), and to avoid making the models computationally prohibitive. Optimal models were selected based on the AICc. To obtain coefficients of determination (cumulative $R^2_{adjusted}$) and statistics (F_{pseudo} and P-values) for each variable, components of the best models were forward selected. In parallel, we prepared Global Nonmetric Multidimensional Scaling (GNMDS) graphs using the same options. Significant variables were fitted into the GNMDS ordination space using the *envfit*

function in the *vegan* package of R. We also grouped all climatic, edaphic, spatial, and floristic variables into a variation partitioning analysis by integrating procedures in the *vegan* and *packfor* packages of R. Besides group effects, variation partitioning estimates the proportion of shared variation among these groups of predictors.

For global biogeographic analyses, we excluded OTUs from the order Hypocreales and family Trichocomaceae (both Ascomycota), because the ITS region provides insufficient taxonomic resolution and known biological species are grouped together within the same OTU (*23*). We tested the differences among fungal taxonomic and functional groups for the occurrence frequency (number of sites detected) and latitudinal range of OTUs using a non-parametric Kruskal-Wallis test and Bonferroni-adjusted multiple comparisons among mean ranks. To test the validity of Rapoport's rule in soil fungi, we calculated the average latitudinal range of OTUs for each site (*24*). The average latitudinal range was regressed with the latitude of study sites by polynomial model selection based on the AICc criterion. This analysis was run with and without OTUs only detected at a single site (range=0). Because the results were qualitatively similar, we report results including all OTUs. To construct biogeographic relationships among major regions and biomes, we generated cross-region and cross-biome networks based on the number of shared OTUs. We excluded occurrences represented by a single sequence per site. Ward clustering of biogeographic regions and biomes were constructed using the Morisita-Horn index of similarity, which is insensitive to differences in samples size, by use of the *pvclust* package of R. In this procedure, P-values are inferred for non-terminal branches based on multiscale bootstrap resampling with 1,000 replicates.

Results and Discussion

Taxonomic and functional diversity

Pyrosequencing analysis of global soil samples revealed 1,019,514 quality-filtered sequences that were separated into 94,255 species-level OTUs (see supplementary information). Altogether 963,458 (94.5%) sequences and 80,486 (85.4%) OTUs were classified as Fungi. Most other taxa belonged to animals (Metazoa, 3.3%), plants (Viridiplantae, 3.1%), alveolates (Alveolata, 2.8%), and amoebae (mostly Rhizaria, 1.3%). Kingdom-level assignment of 3.8% OTUs remained elusive. The fungal subset included 35,923 (44.6%) OTUs that were represented by a single sequence; these were removed from further analyses to avoid overestimating richness based on these potentially erroneous sequences (*25*). The remaining 44,563 non-singleton fungal OTUs in our data set numerically correspond to approximately half of the described fungal species on Earth (*1*). For comparison, there are currently 52,481 OTUs based on 98.0% similarity clustering of all fungal ITS sequences in publicly available databases (*12*). Global soil sampling revealed representatives of all major phyla and classes of Fungi. Of fungal taxa, Basidiomycota (55.7%), Ascomycota (31.3%), Mortierellomycotina (6.3%) and Mucoromycotina (4.4%) encompassed the largest proportion of sequences (Fig. 2), whereas the most OTU-rich phyla were the Ascomycota (48.7%), Basidiomycota (41.8%), Chytridiomycota (2.3%), and Cryptomycota (syn. Rozellida; 2.1%) (Fig. S2; Data S1). Except for the recently described phylum Cryptomycota (*26*), the relative proportions of major phyla correspond to the proportional distribution of taxa described and sequenced to date (*12*, www.indexfungorum.org). Below the phylum level, approximately 6% of all fungal OTUs could not be assigned to any known class of fungi. Further clustering of unidentified fungal

sequences at 70% sequence similarity revealed 14 distinct taxonomic groups comprising >7 OTUs, suggesting that there are several deeply divergent class-level fungal lineages that have not yet been described or previously sequenced.

Our classification revealed that 10,801 (24.2%) fungal OTUs exhibited >98% sequence similarity, and 33.8% exhibited >97% similarity, to pre-existing ITS sequences in public databases. This is consistent with Taylor et al. (*19*), reporting 48% of OTUs amplified from Alaskan soils with >97% similarity to any database sequences. In our study, only 4353 fungal OTUs (9.8%) were matched to sequences from herbarium specimens or fully described culture collections at >98.0% sequence similarity. Although many type collections are yet to be sequenced, the paucity of matches to database entries indicates that a majority of soil-inhabiting fungal taxa remain undescribed (*19*-*20*). These results highlight the current lack of data from understudied tropical and subtropical ecosystems. The phenomenon of high cryptic diversity and low success in naming OTUs at the genus or species level have been found in other groups of soil microbes and invertebrates, emphasizing our poor overall knowledge of global soil biodiversity (*27*-*28*).

The main fungal phylogenetic and functional groups were present in all ecosystems, but their relative proportions varied several-fold across biomes (Figs. 2, S2-S4). The ratio of Ascomycota to Basidiomycota OTUs was highest in grasslands and shrublands (1.86) and tropical dry forests (1.64) but lowest in the temperate deciduous forests (0.88). Chytridiomycota, Cryptomycota, and Glomeromycota were relatively more diverse in the grasslands and shrublands, accounting for 4.6%, 3.6%, and 1.4% of OTU richness, respectively. The relative OTU richness of Mortierellomycotina and Mucoromycotina (including most fast-growing molds but also some plant symbionts) peaked in the tundra biome

(4.8% and 2.7%, respectively), but their abundance was lowest in tropical dry forests (1.0% and 0.6%, respectively). Archaeorhizomycetes, a recently described class of Ascomycetes from a boreal forest (*29*), was most diverse in tropical moist and montane forests, particularly in northern South America and New Guinea.

Among all fungal taxa, OTUs assigned to saprotrophs, EcM mutualists, and plant pathogens comprised 19,540 (43.8%), 10,334 (23.2%), and 1770 (4.0%), respectively (Fig. S4). Other trophic categories were contained <1% of remaining OTUs. EcM fungi contributed 34.1% of all taxa in the northern temperate deciduous forests, but accounted for a relatively low proportion (11.9%) in grasslands and shrublands, reflecting the paucity of host plants in these ecosystems. Similarly, the proportion of EcM fungal taxa was lowest in northern South America (8.0%), where AM trees often dominate. Plant pathogens were relatively more abundant and diverse in lowland tropical moist (6.2%) and dry (6.3%) forests.

Predictors of global richness

Structural equation models revealed that climate has both a strong direct effect on plant and fungal richness and functional groups, but it also indirectly affects these metrics by altering edaphic conditions (Main text; Fig. S5). Both SEM and regression models suggest that the best predictors of diversity differed among phylogenetic and functional groups of fungi. Positive effects of mean annual precipitation (MAP) and soil Ca concentration were the strongest predictors of total fungal diversity, explaining 7.2% and 8.9% of residual richness, respectively (Table S2). Richness of EcM fungi responded positively to the relative proportion and species richness of EcM plants (explaining 18.3% and 8.5% of variance, respectively), as well as soil pH (13.0%). EcM host species richness (5.9%) and soil pH (20.4%) remained the strongest

423 predictors in the best model for sites with EcM vegetation accounting for >60% of basal area, a
424 critical point above which the proportion of EcM plants had no further effect on EcM fungal
425 richness. MAP had a strong positive effect (14.8%) on richness of saprotrophs. Diversity of
426 plant pathogens declined with increasing distance from the equator (17.8%) and soil C/N ratio
427 (11.6%). Animal parasites responded positively to MAP (20.3%), whereas monthly variation of
428 precipitation (MAP CV) had a negative impact on richness of mycoparasites (fungus-parasitic
429 fungi; 8.2%). Richness of the AM Glomeromycota was negatively related to the age of
430 vegetation (7.3%) but positively related to potential evapotranspiration (PET, 3.5%) and soil
431 pH (4.3%). Of the major taxonomic groups, the richness of Ascomycota in general (18.5%)
432 and that of Archaeorhizomycetes (21.7%) were negatively related to distance from the equator
433 in best-fit models. Climatic variables were the best predictors for richness of
434 Mortierellomycotina (MAT: negative effect, 26.1%) and the ascomycete classes
435 Dothideomycetes (MAT: positive effect, 20.9%), Lecanoromycetes (MAT: negative effect,
436 26.7%), Leotiomycetes (MAT: negative effect, 30.1%), Orbiliomycetes (MAT: positive effect,
437 12.8%), and Sordariomycetes (MAP: positive effect, 33.4%). The richness of Chytridiomycota
438 and the ascomycete class Pezizomycetes was best explained by a positive response to soil pH
439 (8.6% and 40.5%, respectively). Concentration of soil nutrients or their ratio to other nutrients
440 were the strongest predictors for OTU richness of Cryptomycota (N concentration: positive
441 effect, 10.1%), Geoglossomycetes (N/P ratio: positive effect, 3.7%), Mucoromycotina (C/N
442 ratio: positive effect, 19.0%), and Wallemiomycetes (P concentration: negative effect, 14.9%).
443 The richness of Basidiomycota and its class Agaricomycetes were best explained by a positive
444 response to soil Ca concentration (13.5% and 12.8%, respectively).

445 Although geographical distance *per se* had negligible effects on richness (Moran's
446 I=0.267), spatial predictors were included in the best richness models of nearly all functional
447 and phylogenetic groups (except Glomeromycota), indicating regional- or continental-scale
448 differences in OTU richness (Fig. 1B). Compared to other tropical regions, richness of fungi
449 was conspicuously lower in Africa, independent of biome type. These results might reflect the
450 relatively lower MAP in much of Africa compared with other tropical continents.
451 Alternatively, lower fungal richness could be related to the disproportionately strong shifts in
452 biomes during the Pleistocene, which impoverished the African flora (*18*).
453 Among edaphic variables, soil pH and Ca concentration were typically the most
454 important predictors of fungal OTU richness. These variables positively correlated with fungal
455 richness at the global scale ($F_{1,335}$=290.7; $R_{Pearson}$=0.682; P<0.001). The strong positive
456 influence of soil Ca concentration on richness of fungi, in particular Basidiomycota, is
457 congruent with a similar positive relationship found for Ca and EcM fungal richness associated
458 with Northern Hemisphere *Alnus* spp. (*30*). Exchangeable Ca is important for many
459 physiological processes in plants and microorganisms and it influences the turnover rate of soil
460 organic matter (*31*). In soil geochemical processes, pH and Ca concentration affect each other
461 and thus may have both direct and indirect effects on soil biota. Fungal functional groups were
462 differentially affected by pH. Richness of EcM fungi was greatest in slightly acidic to neutral
463 soils (Fig. S6), whereas saprotrophs, especially white rot decomposers, were more diverse in
464 moderately to strongly acidic soils. Richness of Pezizomycetes peaked distinctly in neutral
465 soils.
466
467 Macroecological patterns

In general agreement with biogeographic patterns of plants, animals, and foliar endophytic fungi (*5,32*), the overall richness of soil fungi increased towards the equator (Fig. 3A). However, major functional and taxonomic groups showed dramatic departures from the general latitudinal richness patterns (Figs. 3, S7). Namely, diversity of saprotrophic fungi, parasites, and pathogens increased at low latitudes, whereas richness of EcM fungi peaked at mid-latitudes, especially in temperate forests and Mediterranean biomes of the Northern Hemisphere (40-60 °N; Fig. S8). By contrast, saprotrophic fungi had a broad richness peak spanning from ca. 45 °S to 25 °N. Richness of Ascomycota, in particular that of Archaeorhizomycetes, Dothideomycetes, Eurotiomycetes, Orbiliomycetes, and Sordariomycetes, peaked in tropical ecosystems (Fig. S7). Conversely, the ascomycete classes Lecanoromycetes and Leotiomycetes as well as Microbotryomycetes (basidiomycete yeasts), Mortierellomycotina, and Mucoromycotina increased in diversity towards the poles, with no noticeable decline in boreal forests and tundra biomes. Agaricomycetes, Pezizomycetes, and Tremellomycetes exhibited distinct richness peaks at mid-latitudes. Richness of Agaricomycetes was greater in the Northern Hemisphere, whereas that of Microbotryomycetes, Tremellomycetes, and Wallemiomycetes peaked in the Southern Hemisphere temperate ecosystems (Fig. S8).

All of these phylogenetic groups originated >150 million years ago on the supercontinent Pangaea (*33*) and have had sufficient time for long-distance dispersal. However, our data suggest that particular regional biotic or abiotic conditions (e.g., soil pH and favorable climatic conditions) have likely stimulated evolutionary radiations in certain geographic areas and not in others. Adaptation to cold climate in younger fungal phyla has been suggested to explain differential latitudinal preferences among fungal groups (*34*). However, our global

analysis provided no support for this hypothesis (Fig. S9). Instead, it revealed that ancient lineages are relatively more common in non-wooded ecosystems.

Relation of plant and fungal richness

Plant and fungal richness were positively correlated (Fig. S10), but plant richness explained no residual richness of fungi based on the best regression model ($R^2_{adj}<0.01$; $P>0.05$). These results and SEM path diagrams suggest that correlations between plant and fungal richness are best explained by their similar response to climatic and edaphic variables (i.e., covariance) rather than by direct effects of plants on fungi. However, when separating functional categories, trophic groups of fungi exhibited differential response to plant diversity and relative proportion of potential hosts.

Plant pathogens usually attack a phylogenetically limited set of host plants (*35*), suggesting that that plant pathogens have at least partly co-evolved with their hosts and may have radiated more intensively in the tropics where high plant diversification and richness permit greater diversification. Strong phylogenetic signals in soil feedbacks, adaptive radiation, and negative density dependence (the Janzen-Connell hypothesis) have probably contributed to the pronounced richness of both plants and their pathogens at low latitudes (*36, 37*). However, our analyses revealed no significant effects of plant richness *per se* on residual richness of pathogens in soil. Similarly to pathogens, richness of AM fungi was unrelated to the proportion of AM host trees or interpolated host richness, which may result from non-specific associations with tree and understory species. Hence both AM and soil pathogen richness were unaffected by plant richness. By contrast, host richness explained 6% of variation in EcM fungal richness, indicating either niche differentiation of fungi in forests of mixed hosts or sampling effects

(i.e., forests with higher host diversity are more likely to include plant species that harbor high fungal diversity). With a few notable exceptions, most studies have found low levels of host preference or host specificity among EcM fungi (*38*). We found that relative EcM host density had a strong influence on EcM fungal richness, suggesting that greater availability of colonizable roots in soil provides more carbon for EcM fungi and thereby yields greater species density and local-scale richness regardless of latitude. The peak of EcM fungal taxonomic and phylogenetic richness in northern temperate biomes coincides with the geographical distribution and dominance of Pinaceae, which is the oldest extant EcM plant family (*15, 39*).

The ratio of plant-to-fungal richness decreased exponentially with increasing latitude, because plant diversity dropped precipitously toward the poles relative to fungal diversity (Fig. 4). This finding calls into question present global fungal richness estimates. These estimates assume similar spatial turnover of plant and fungal species and a constant plant-to-fungus ratio, and have been formulated based mostly on data from temperate and boreal ecosystems (*1, 19, 20*). Yet local-scale beta diversity of both plants and fungi differ among temperate and tropical sites (*40, 41*) and there are profound differences in plant species turnover depending on propagule size (*42*). Natural distribution of very few vascular plant species encompass several continents, but there are multiple fungal species with circumpolar or cosmopolitan distribution (*43, 44*; see Biogeography section below). While we cannot directly compare plant and fungal beta diversity, spatial turnover of plant species is inarguably greater (*42*). Based on the function of fungi-to-plant richness ratio to latitude and latitudinal distribution of land, we calculated that fungal richness is overestimated by 1.5- and 2.5-fold based on constant temperate (45° latitude) and boreal (65° latitude) richness ratios, respectively.

Since richness estimates are calculated based on the frequency of the rarest species, the reliability of singleton data call into question biologically meaningful extrapolations (*11*). In metabarcoding studies such as ours, sequencing errors tend to give rise to singleton sequences, and the number of rare artificial taxa grows rapidly with increasing sequencing depth (*25*). Therefore, despite the size of our dataset, it cannot readily be used to produce reliable taxonomic richness extrapolations.

Community ecology

Variation partitioning analysis revealed that climatic, edaphic, and floristic variables (and their shared effects) are the strongest predictors for community composition of all fungi and most of their functional groups (Fig. S11). However, the saprotroph community composition was most strongly explained by purely spatial variables. More specifically, PET and soil pH explained 2.4% and 1.5%, respectively, of the variation in total fungal community composition (Table S3; Fig. S12). PET contributed 3.8%, 2.8%, and 11.7% to community structure of saprotrophs, plant pathogens, and yeasts, respectively. Distance from the equator (1.3%) and soil pH (0.7%) were the strongest predictors of EcM fungal community composition, whereas mean annual temperature (4.0%) was the strongest predictor for animal parasites, and distance from the equator (3.5%) was the best predictor for mycoparasites (Table S3; Fig. S12).

These results indicate that both environmental and spatial predictors generally have a minor influence on species-level composition of fungi at the global scale. Nonetheless, the significant global-scale pH effect in several groups of fungi is consistent with the substantial influence of pH on the phylogenetic structure of soil fungal and bacterial communities in both local and continental scales (*27, 45*). The relatively stronger climatic and edaphic drivers of

richness at the class and phylum level suggest that phylogenetic niche conservatism in fungal lineages is similar to cross-biome distribution patterns in vascular plants (*46*) and protists (*47*).

Global biogeography

Consistent with Rapoport's rule formulated for macro-organisms (*24*) and later applied to marine bacteria (*48*), the mean latitudinal range of fungi strongly increased towards the poles (Fig. S13). These results also suggest that a greater proportion of fungi are endemic within tropical rather than extra-tropical ecosystems.

Major taxonomic and functional groups of fungi differed markedly in their distribution range (Figs. S14, S15). Animal parasites were more widely distributed compared with all other groups, suggesting that there are many generalist OTUs with global distribution. Saprotrophs and plant pathogens had broader distribution ranges than EcM and AM root symbionts. Taxa belonging to Mortierellomycotina, Mucoromycotina, Tremellomycetes, and Wallemiomycetes – groups that include a large proportion of saprotrophs and parasites that produce exceptionally large quantities of aerially dispersed mitospores – were generally most widely distributed. Besides the AM Glomeromycota, OTUs belonging to the ascomycete classes Archaeorhizomycetes, Geoglossomycetes, and Orbiliomycetes were detected from the fewest sites.

The northernmost biogeographic regions (Europe, West Asia, East Asia, and North America) had the most similar fungal communities as revealed by shared fungal OTUs (Fig. 5). Based on the Morisita-Horn similarity index, the northern and southern temperate regions clustered together with marginally non-significant support (P=0.064; Fig. 6A). In spite of the large geographical distance separating them, paleo- and neotropical biogeographic regions

583 clustered together (P=0.059). However, biogeographic clustering of regions deviated markedly
584 in certain functional groups of fungi (Fig. 6). For instance, EcM fungi in the southern
585 temperate and tropical regions had greater similarity compared with northern temperate
586 ecosystems (P=0.001). Among biomes, boreal forests, temperate coniferous forests, and
587 temperate deciduous forests shared the largest numbers of fungal OTUs (Fig. S16). Fungal
588 OTUs in temperate deciduous forests were highly similar to Mediterranean and tropical
589 montane forests, whereas fungal OTUs in tropical montane forests were linked to tropical
590 moist forests, which in turn exhibited substantial connections with tropical dry forests and
591 savannas. As a result, cluster analysis supported separation of tropical and non-tropical biomes
592 (Fig. 6B). Consistent with biogeographic region-level analysis, lowland tropical biomes, arctic
593 tundra and boreal forests biomes, and temperate biomes formed three well-supported clusters.
594 Tropical montane forests and grasslands and shrublands were clustered with temperate biomes
595 based on distribution of all fungi and most functional groups. However in EcM fungi, taxa
596 from southern temperate forests, tropical montane forests, and grass/shrublands clustered with
597 tropical lowland and Mediterranean biomes. A relatively large proportion of EcM fungal taxa
598 were shared across various biomes in Australia and New Guinea, which explains these
599 deviating patterns. In contrast, plant pathogens from tropical montane forests clustered with
600 tropical lowland biomes rather than with temperate biomes.
601 Our biogeographic analyses complement the community-level results suggesting that
602 both climate and biogeographic history shape macro-ecological patterns of fungi. Co-migration
603 with hosts over Pleistocene land bridges (e.g., Beringia, Wallacea, Panamanian) and long-
604 distance dispersal by spores appear to have played important roles in shaping current fungal
605 distribution patterns (*30, 43*). The relative influence of climate and biotrophic associations with

606 host plants of varying extant distributions probably contribute to differences in the range and
607 biogeographic relationships among fungal functional groups (*49*). In addition, taxon-specific
608 constraints for dispersal, such as shape and size of propagules and sensitivity to UV light, may
609 differentially affect long-distance dispersal among taxa (*7*). For instance, Glomeromycota
610 OTUs, which form relatively large non-wind dispersed asexual spores, had the lowest average
611 geographical range. In general, region-based distribution patterns of fungi are somewhat
612 conflicting with clustering of plants and animals, where Holarctic lineages are deeply nested
613 within larger tropical groups (*50*). Consistent with macro-organisms, fungi from the Southern
614 Hemisphere temperate landmasses cluster together. Differences observed in macro-ecological
615 patterns among fungi, plants, and animals may originate from the relative strength of dispersal
616 limitation and phylogeographic history, but exaggeration by methodological differences among
617 studies cannot be discounted. The use of homogenous sampling and analytical methods, as
618 done in this study, are necessary to confidently compare macro-ecological patterns amongst
619 distinct life forms and to reliably test degrees of consistency among all kingdoms of life.
620
621 Conclusions and perspectives
622 Climatic variables explained the greatest proportion of richness and community composition in
623 fungal groups by exhibiting both direct and indirect effects through altered soil and floristic
624 variables. The strong driving climatic forces identified here open up concerns regarding the
625 impact of climate change on the spread of disease (*51*) and the functional consequences of
626 altered soil microorganism communities (*52*). The observed abrupt functional differences
627 between fungal communities in forested and treeless ecosystems, despite spatial juxtaposition,
628 suggests that plant life form and mycorrhizal associations determine soil biochemical processes

629 more than plant species *per se*. Loss of tree cover and shrub encroachment resulting from
630 drying and warming may thus have a marked impact on ecosystem functioning both above-
631 and belowground.

632 In addition to natural mechanisms, such as long-distance dispersal and migration over
633 past land bridges, global trade has enhanced the spread of some non-native soil organisms into
634 other ecosystems, where they sometimes become hazardous to native biota, economy, and
635 human health (*53*). Our results highlight how little insight we still have into natural microbial
636 distribution patterns, and this undermines our ability to appraise the actual role of humans in
637 shaping these biogeographic processes. Even larger-scale sampling campaigns are needed to
638 provide data for establishing natural distributions and building species distribution models
639 (*52*), which will enable us to predict the spread and habitat suitability of non-native
640 microorganisms.
641

642 **References and Notes**

643

644 1. M. Blackwell, *Am. J. Bot*. **98**, 426 (2011).

645 2. N. Fierer *et al*., *Ecol. Lett*. **12**, 1 (2009).

646 3. H. Serna-Chavez, N. Fierer, P. M. van Bodegom, *Glob. Ecol. Biogeogr*. **10**, 1162 (2013).

647 4. X. Xu, P. Thornton, W. M. Post, *Glob. Ecol. Biogeogr*. **22**, 737 (2013).

648 5. H. Hillebrand, *Am. Nat*. **163**, 192 (2004).

649 6. G. G. Mittelbach *et al*., *Ecol. Lett*. **10**, 315 (2007).

650 7. B. J. Finlay, *Science* **296**, 1061 (2002).

651 8. D. R. Nemergut *et al*., *Microbiol. Mol. Biol. Rev.* **77**, 342 (2013).

652 9. K. G. Peay, M. I. Bidartondo, A. E. Arnold, *New Phytol*. **185**, 878 (2010).

653 10. J. M. Talbot *et al*., *Proc. Natl. Acad. Sci. USA*, **111**, 6341 (2014)

654 11. B. D. Lindahl *et al*., *New Phytol*. **199**, 288 (2013).

655 12. U. Kõljalg *et al*., *Mol. Ecol.* **22**, 5271 (2013).

656 13. K. Abarenkov *et al., Evol. Bioinform*. **6,** 189 (2010).

657 14. S. M. Adl *et al., J. Eukaryot. Microbiol.,* **59**, 527 (2012).

658 15. L. Tedersoo, M. E. Smith*, Fung. Biol. Rev*. **27**, 83 (2013).

659 16. L. Tedersoo *et al*., *New Phytol*. **195**, 832 (2012).

660 17. I. Hiiesalu *et al., New Phytol.* **203**, 233 (2014).

661 18. H. Kreft, W. Jetz, *Proc. Natl. Acad. Sci. USA* **104**, 5925 (2007).

662 19. D. L. Taylor *et al*., *Ecol. Monogr*. **84**, 3 (2014).

663 20. H. E. O'Brien *et al*., *Appl. Environ. Microbiol*. **71**, 5544 (2005).

21. R Core Team, R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing (2014).

22. T. F. Rangel *et al., Ecography* **33**, 46 (2010).

23. C. L. Schoch *et al*., *Proc. Natl. Acad. Sci. USA* **109**, 6241 (2012).

24. G. C. Stevens, *Am. Nat*. **133**, 240 (1989).

25. I. A. Dickie, *New Phytol*. **188**, 916 (2010).

26. M. D. M. Jones *et al*., *Nature* **474**, 200 (2011).

27. C. Lauber *et al*., *Appl. Environ. Microbiol*. **75**, 5111 (2009).

28. M. S. Robeson *et al*., *Proc. Natl. Acad. Sci. USA* **108**, 4406 (2011).

29. A. Rosling *et al*., *Science* **333**, 876 (2011).

30. S. Põlme *et al*., *New Phytol*. **198**, 1239 (2013).

31. P. B. Reich *et al*., *Ecol. Lett*. **8**, 811 (2005).

32. A. E. Arnold, *Fung. Biol. Rev*. **21**, 51 (2007).

33. M. L. Berbee, J. W. Taylor, *Fung. Biol. Rev.* **24**, 1 (2010).

34. K. K. Treseder *et al*., *Ecol. Lett.* **9**, 1086 (2014).

35. G. S. Gilbert, C. O. Webb, *Proc. Natl. Acad. Sci. USA* **104**, 4979 (2007).

36. X. Liu *et al*., *Ecol. Lett*. **15**, 111 (2012).

37. R. Bagchi *et al*., *Nature* **506**, 85 (2014).

38. M. Bahram *et al., Fung. Ecol.* **7**, 70 (2013).

39. L. Tedersoo *et al*., *Mol. Ecol*. **21**, 4160 (2012).

40. M. Bahram *et al., J. Ecol.* **101**, 1335 (2013).

41. H. Qian *et al., Glob. Ecol. Biogeogr.* **22**, 659 (2013).

42. H. Qian, *Glob. Ecol. Biogeogr.* **18**, 327 (2009).

687 43. J. Geml *et al*., *J. Biogeogr*. **34**, 74 (2012).
688 44. I. Timling *et al., Mol. Ecol.* **23**, 3258 (2014).
689 45. J. Rousk *et al*., *ISME J.* **4**, 1340 (2010).
690 46. M. D. Crisp *et al*., *Nature* **458**, 754 (2009).
691 47. S. T. Bates *et al*., *ISME J.* **7**, 652 (2013).
692 48. W. Jun Sul *et al*., *Proc. Natl. Acad. Sci. USA* **110**, 2342 (2013).
693 49. H. Sato *et al*., *Mol. Ecol.* **21,** 5599 (2012).
694 50. I. Sanmartin, F. Ronquist, *Syst. Biol.* **53**, 216 (2004).
695 51. S. Altizer *et al. Science* **341**, 514 (2013).
696 52. W. H. van der Putten *et al., Phil. Trans. R. Soc. B* **365**, 2025 (2010).
697 53. M.-L. *Desprez-Loustau et al., Trends Ecol. Evol.* **22**, 472 (2007).
698

699 **Acknowledgements**

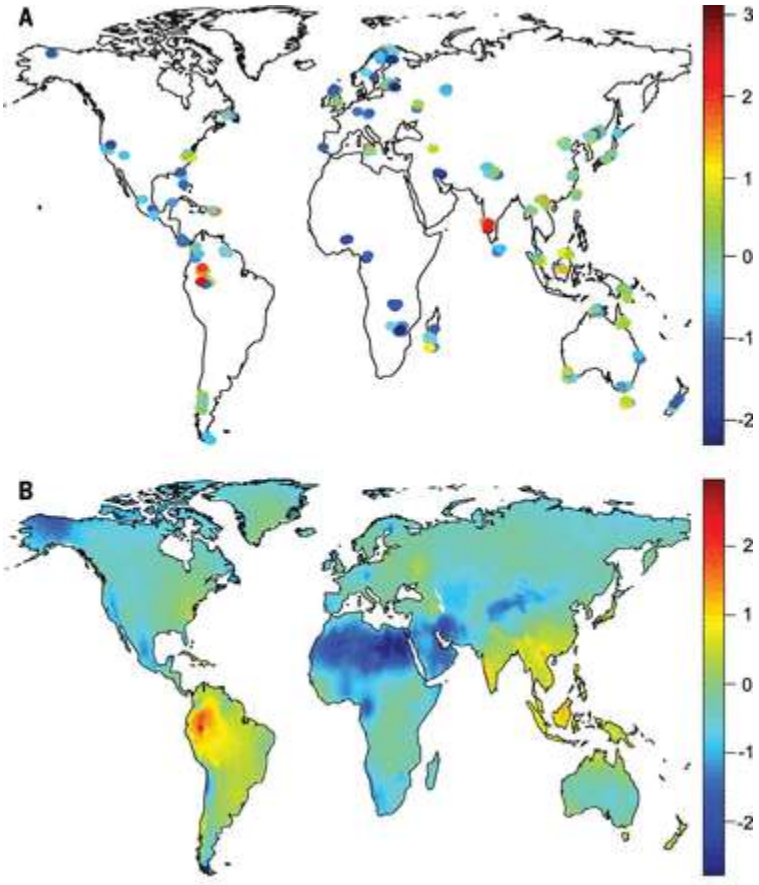
700

701 The sequence data and metadata are deposited in the Short Read Archive (accession
702 SRP043706) and UNITE databases. Data used for analyses are available as supplementary
703 online material Data S1 and S2. We thank H. Mann, D. Sveshnikov, F.O.P. Stefani, A. Voitk,
704 and Y. Wu for supplying single soil samples; R. Puusepp, M. Haugas, and M. Nõukas for
705 sample preparation; H. Kreft for providing interpolated plant diversity data; S. Jüris for
706 designing the printed figure; M.I. Bidartondo, K.G. Peay and three anonymous reviewers for
707 constructive comments on the manuscript; and relevant institutions of multiple countries for
708 issuing permissions for sampling and delivery. The bulk of this project was funded from
709 Estonian Science Foundation grants 9286, 171PUT, IUT20-30; EMP265; FIBIR; ERC; and in

part by numerous funding sources that facilitated co-author efforts in collecting and pre-processing samples.

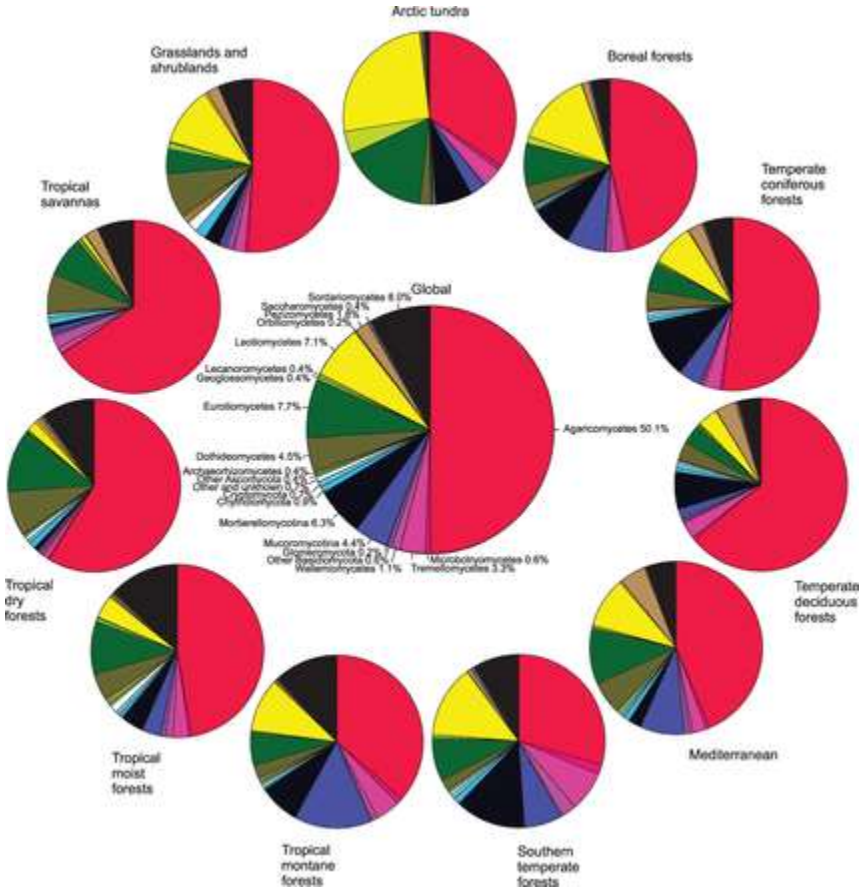
Figure legends

Fig. 1. Map of A) global sampling (circles as study sites); B) Interpolated taxonomic richness of all fungi using Inverse Distance Weighting (IDW) algorithm and accounting for the relationship with mean annual precipitation (based on the best multiple regression model). Different colors depict residual Operational Taxonomic Unit (OTU) richness of all fungi accounting for sequencing depth. Warm colors indicate OTU-rich sites, whereas cold colors indicate sites with fewer OTUs.



724 **Fig. 2.** Relative proportion of fungal sequences assigned to major taxonomic groups in

725 different biomes.



726

727

728

Fig. 3. Relationships between residual richness of fungal taxonomic or functional groups and distance from the equator. A, all fungi; B, ectomycorrhizal (EcM) fungi; C, saprotrophic fungi; D, plant pathogens; E, animal parasites; F, mycoparasites; G, white rot decomposers; and H, yeasts. Lines indicate best-fitting linear or polynomial functions.

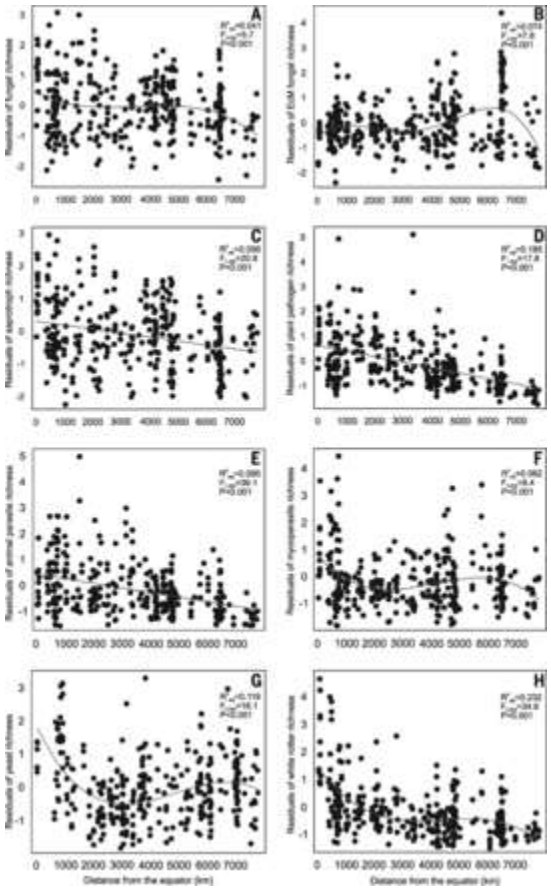


Fig. 4. Relationship between standardized plant richness to fungal richness ratio and distance from the equator based on (A) interpolated values and (B) polynomial regression. Residuals of fungal richness are taken from the best linear regression model accounting for other significant predictors. Warm colors indicate high plant-to-fungal richness ratio, whereas cold colors indicate low plant-to-fungal richness.

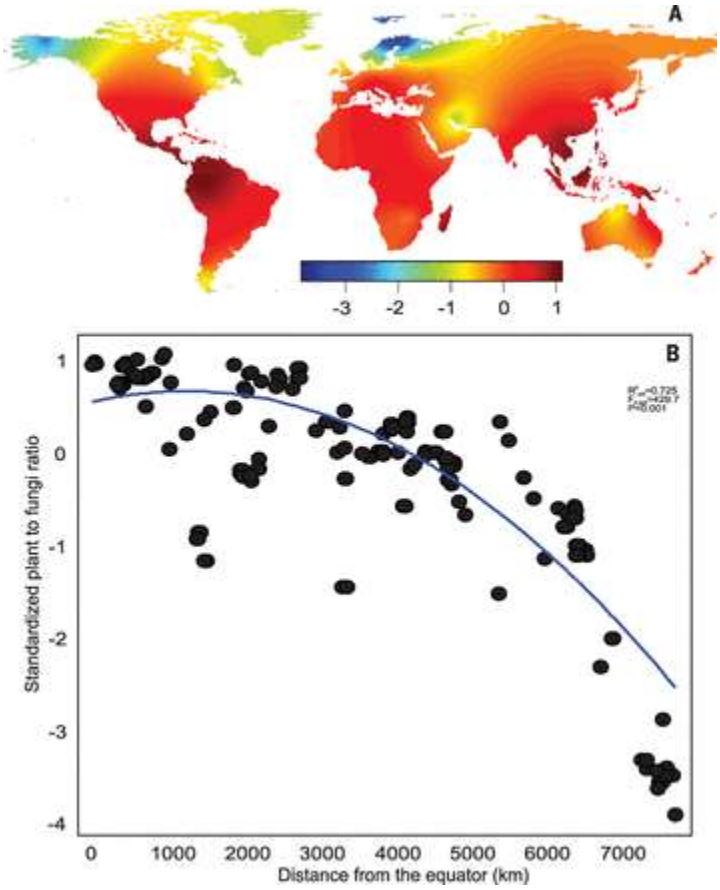


Fig. 5. Connectedness of biogeographic regions by shared Operational Taxonomic Units (OTUs) of ectomycorrhizal fungi (blue), saprotrophs (black), and plant pathogens (red). The width of lines and diameter of circles are proportional to the square root of the number of connections and sample size (number of sites), respectively. Numbers in circles indicate the number of OTUs found in each region. OTUs with a single sequence per site and OTUs belonging to Hypocreales and Trichocomaceae (in which the ITS region is too conservative for species-level discrimination) were excluded.

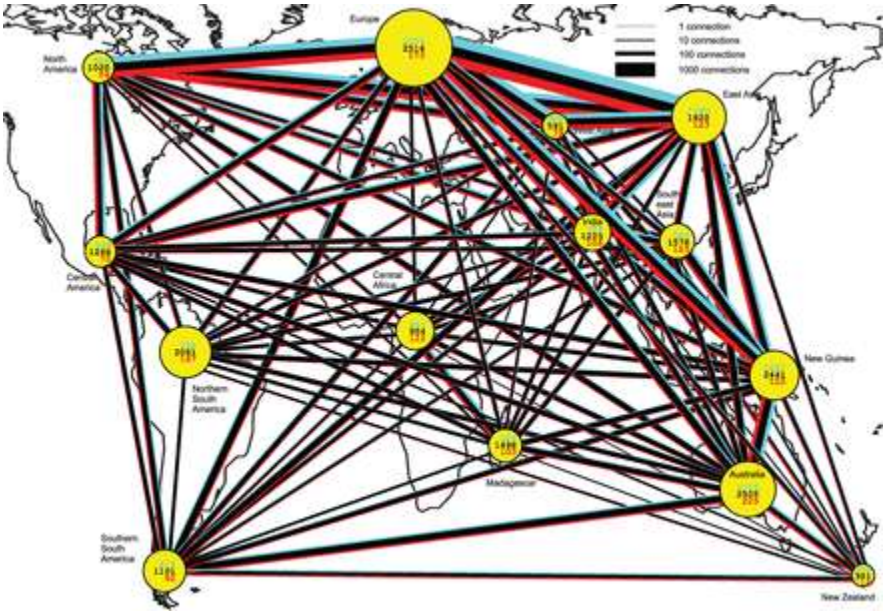
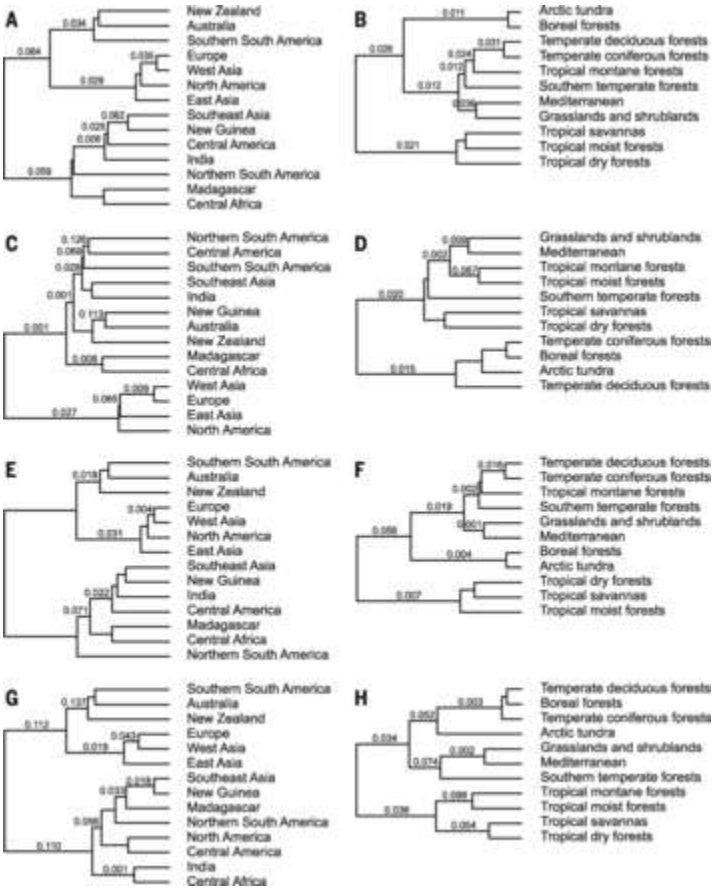


Fig. 6. Ward clustering of biogeographic regions (left panes) and biomes (right panes) based on the Morisita-Horn pairwise similarity index in A and B, all fungi; C and D, ectomycorrhizal fungi; E and F, saprotrophs; G and H, plant pathogens. Numbers above branches indicate P-values.



Supplementary Materials

Figs. S1-S16

Tables S1-S3

Data S1-S2