

Global hotspots of mycorrhizal fungal richness are poorly protected

<https://doi.org/10.1038/s41586-025-09277-4>

Received: 13 February 2024

Accepted: 12 June 2025

Published online: 23 July 2025

Open access



Michael E. Van Nuland¹✉, Colin Averill^{2,3}, Justin D. Stewart^{1,4}, Oleh Prylutskyi⁵, Adriana Corrales¹, Laura G. van Galen^{1,2}, Bethan F. Manley¹, Clara Qin¹, Thomas Lauber², Vladimir Mikryukov⁶, Olesia Dulia⁶, Giuliana Furci⁷, César Marín^{4,8}, Merlin Sheldrake^{1,4}, James T. Weedon⁹, Kabir G. Peay^{10,11}, Charlie K. Cornwallis¹², Tomáš Větrovský¹³, Petr Kohout¹³, Petr Baldrian¹³, Leho Tedersoo^{6,14}, Stuart A. West¹⁵, Thomas W. Crowther², E. Toby Kiers^{1,4}, SPUN Mapping Consortium* & Johan van den Hoogen^{1,2}

Mycorrhizal fungi are ecosystem engineers that sustain plant life and help regulate Earth's biogeochemical cycles^{1–3}. However, in contrast to plants and animals, the global distribution of mycorrhizal fungal biodiversity is largely unknown, which limits our ability to monitor and protect key underground ecosystems^{4,5}. Here we trained machine-learning algorithms on a global dataset of 25,000 geolocated soil samples comprising >2.8 billion fungal DNA sequences. We predicted arbuscular mycorrhizal and ectomycorrhizal fungal richness and rarity across terrestrial ecosystems. On the basis of these predictions, we generated high-resolution, global-scale maps and identified key reservoirs of highly diverse and endemic mycorrhizal communities. Intersecting protected areas with mycorrhizal hotspots indicated that less than 10% of predicted mycorrhizal richness hotspots currently exist in protected areas. Our results describe a largely hidden component of Earth's underground ecosystems and can help identify conservation priorities, set monitoring benchmarks and create specific restoration plans and land-management strategies.

Mycorrhizal fungi form nutritional symbioses with >80% of plant species and build extensive underground hyphal networks that can constitute >30% of the living microbial biomass of soils¹. The fungi use their networks to forage in the soil for phosphorus, nitrogen and trace elements, and receive carbon from plant root systems in return. The diversity and activity of these fungal communities is fundamental to the functioning of Earth's terrestrial ecosystems, with an estimated 3.6 billion tons of carbon annually allocated through plants to mycorrhizal fungi^{2,3}. However, despite their importance, our understanding of the global distribution of mycorrhizal fungal biodiversity lags far behind plants and animals. As a result, it is difficult to develop appropriate measures to protect these organisms and the integrity of their symbiotic partnerships^{4,5}.

Over the past decade, the emergence of comprehensive, large-scale molecular datasets on belowground fungi have made it possible to establish global baselines of mycorrhizal fungal biogeography^{6–8}. These efforts have helped reveal the dominance of two types of mycorrhizal symbionts spread across Earth's terrestrial biomes: ectomycorrhizal (EcM) fungi and arbuscular mycorrhizal (AM) fungi. These fungal types differ in their physiology, symbiotic behaviour and contribution to carbon cycling^{9,10}. Although only about 2% of terrestrial plant species

depend on EcM associations, these host species constitute >25% of global vegetation area and dominate in high-latitude forests¹¹. AM fungi form associations with around 80% of plant species¹¹ and dominate in tropical forests, grasslands and croplands globally. A third type—ericoid mycorrhizal (ErM) fungi—are common in high-elevation and high-latitude ecosystems but are rarer globally¹¹. The scarcity of ErM fungi in large-scale fungal datasets has left a sizable gap in our understanding of their global diversity patterns or conservation opportunities.

Previous work has used plant genus as a proxy of mycorrhizal type to reveal fundamental insights into the distributions and functions of mycorrhizal symbioses^{12–14}. However, efforts to explore mycorrhizal fungal biogeography from direct microbial observations, with methods such as DNA-based surveys of soil samples^{15–19}, have shown that aboveground and belowground biodiversity patterns are often not correlated²⁰. This finding means that patterns of plant communities may not accurately reflect distributions of mycorrhizal diversity belowground. Moreover, our ability to identify the scale of this mismatch is limited by geographical biases in the locations of mycorrhizal samples used for predictive modelling. Such large data gaps inevitably result in extensive model extrapolation into untrained environmental space²¹, which is particularly unreliable in under-sampled tropical regions where AM fungi are

¹Society for the Protection of Underground Networks (SPUN), Dover, DE, USA. ²Institute of Integrative Biology, Department of Environmental Systems Science, ETH Zürich, Zürich, Switzerland.

³Funga Public Benefit Corporation, Austin, TX, USA. ⁴Amsterdam Institute for Life and Environment (A-LIFE), Section Ecology and Evolution, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. ⁵Falz-Fein Biosphere Reserve “Askania Nova”, Kherson Oblast, Ukraine. ⁶Mycology and Microbiology Center, University of Tartu, Tartu, Estonia. ⁷Fungi Foundation, Brooklyn, NY, USA. ⁸Centro de Investigación e Innovación para el Cambio Climático (CiiCC), Universidad Santo Tomás, Valdivia, Chile. ⁹Amsterdam Institute for Life and Environment (A-LIFE), Section Systems Ecology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. ¹⁰Department of Earth System Science, Stanford University, Stanford, CA, USA. ¹¹Department of Biology, Stanford University, Stanford, CA, USA. ¹²Department of Biology, Lund University, Lund, Sweden. ¹³Laboratory of Environmental Microbiology, Institute of Microbiology of the Czech Academy of Sciences, Prague, Czech Republic. ¹⁴College of Science, King Saud University, Riyadh, Saudi Arabia. ¹⁵Department of Biology, Oxford University, Oxford, UK. *A list of authors and their affiliations appears at the end of the paper. [✉]e-mail: michael@spun.earth

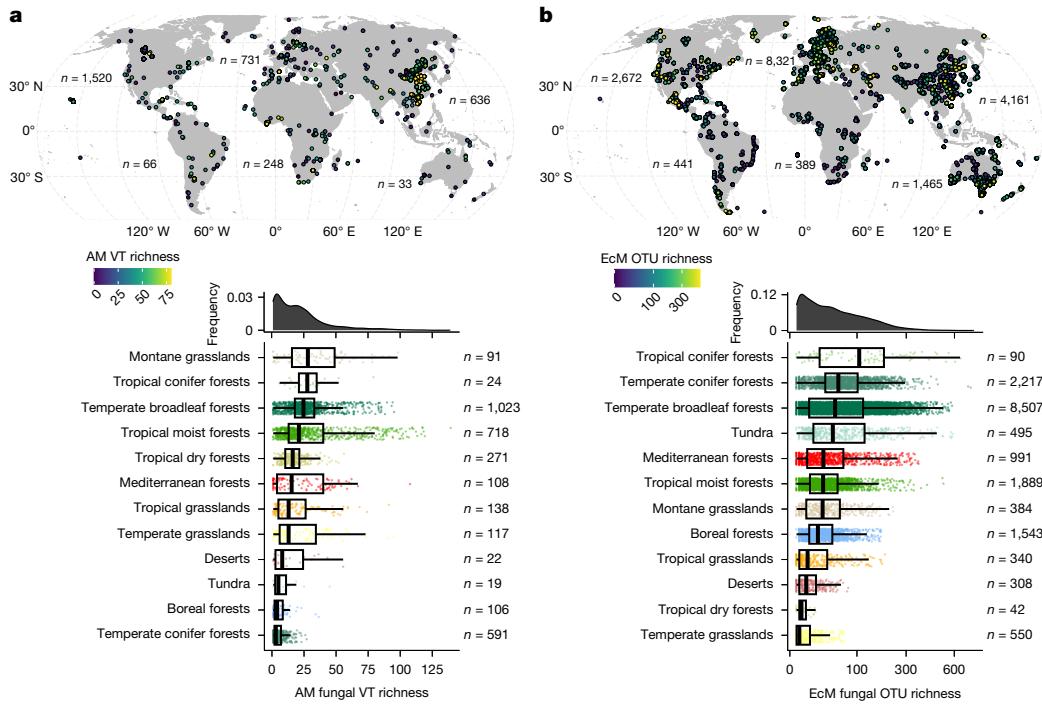


Fig. 1 | Sample locations and mycorrhizal richness trends by biome.

a,b, Distribution of sites and richness estimates for AM fungi (**a**) and EcM fungi (**b**). VT were created from SSU sequences for AM fungi, and 97% similar OTUs were created from ITS sequences and assigned to EcM fungi. Mycorrhizal richness patterns across terrestrial biomes are shown in boxplots (mangroves, flooded grasslands and rock and ice biomes not shown owing to low numbers of samples in both datasets). Global distributions of estimated richness are

shown in density curves above biome-level boxplots. Boxplots indicate the median (centre line), first and third quartiles (lower and upper box edges) and $1.5 \times$ the interquartile range (IQR) (box whiskers). EcM richness data are shown as square-root transformed for visualization. Richness estimates were calculated using a rarefaction and extrapolation approach that incorporates sequencing depth per sample. Points shown here are all samples that passed quality-control checks and used as training data.

likely to dominate. Generating high-resolution global distributions of mycorrhizal fungal diversity from direct observations—and quantifying the extent of model uncertainty²²—is essential to guide future conservation planning⁵. Protecting mycorrhizal fungal diversity will help avoid species loss across other taxonomic groups and help maintain critical ecosystem functions (for example, carbon sequestration) for realizing nature-based climate mitigation strategies^{4,5}. To address these challenges, we took the following steps: (1) built machine-learning models to create high-resolution (1 km^2) spatial predictions of mycorrhizal fungal richness and endemism (rarity-weighted richness) at the global scale; (2) identified mycorrhizal fungal richness and endemism hotspots and assessed the extent of their current environmental protections; and (3) characterized spatial uncertainty and data limitations.

Calculating global fungal richness

First, we explored the distribution and range of the training data (Fig. 1). This dataset consisted of a globally distributed collection of nearly 25,000 soil samples containing >2.8 billion fungal DNA sequences from 130 countries compiled in the GlobalFungi, GlobalAMFungi and Global Soil Mycobiome consortium databases^{5–8}. These are the largest datasets of fungal internal transcribed spacer (ITS) and small subunit (SSU) rRNA amplicon sequences assembled so far. These sequences were analysed using virtual taxa (VT) for AM fungi and 97% similar operational taxonomic units (OTUs) for EcM fungi. With these data, we estimated AM and EcM fungal richness using a rarefaction and extrapolation approach, and rarity-weighted richness (hereafter ‘rarity’; Extended Data Fig. 1), which is a metric of relative endemism used to guide conservation priorities²³.

Across both AM and EcM datasets, temperate forests were the most sampled biomes (32–61% of all samples), and mangroves and flooded grasslands were the least sampled biomes (0.1% of all samples). There

were only a few AM fungal samples from desert and tundra biomes, and few EcM fungal samples from tropical dry forests and tropical conifer forests. Compared with other biomes, soil samples from montane grasslands, tropical conifer forests and temperate broadleaf forests showed the highest AM fungal richness estimates (Fig. 1a). Similarly, montane grasslands had the greatest AM fungal rarity estimates, followed by Mediterranean forests and tropical grasslands with the next highest in AM fungal endemism levels (Extended Data Fig. 1a). For EcM fungi, samples from coniferous forests (tropical and temperate) and temperate broadleaf forests had the highest richness estimates (Fig. 1b). EcM fungal endemism was also highest in tropical conifer forests, whereas tundra and Mediterranean forests showed a greater concentration of rare EcM fungi than temperate forests (Extended Data Fig. 1b). We used these datasets to train random forest machine-learning models to predict global AM and EcM fungal richness and endemism patterns.

We wanted to know whether mycorrhizal fungi show clear global diversity patterns like plants and animals. To that end, we used maps of predicted richness to plot mean richness and rarity across latitudes. Predicted AM fungal richness was highest in ecosystems near the equator and gradually declined towards the poles. This result was in agreement with the classical latitude diversity gradient hypothesis observed across many taxa^{24,25} (Fig. 2a). By contrast, predicted EcM fungal richness was lowest near the equator and showed more species-rich communities across northern latitudes and in southern regions of South America and Australia, which created an inverse latitudinal diversity gradient (Fig. 2b and Extended Data Fig. 2). Latitudinal patterns of AM fungal endemism showed a similar pattern as AM richness, but EcM fungal endemism had a moderate increase near the equator compared with EcM richness (Fig. 3). These relationships probably emerged through a complex interplay of host vegetation^{14,26}, degree of host specificity^{11,27} and plant–soil nutrient economies^{28,29}. Notably, these analyses highlighted that plant diversity is not a straightforward indicator for

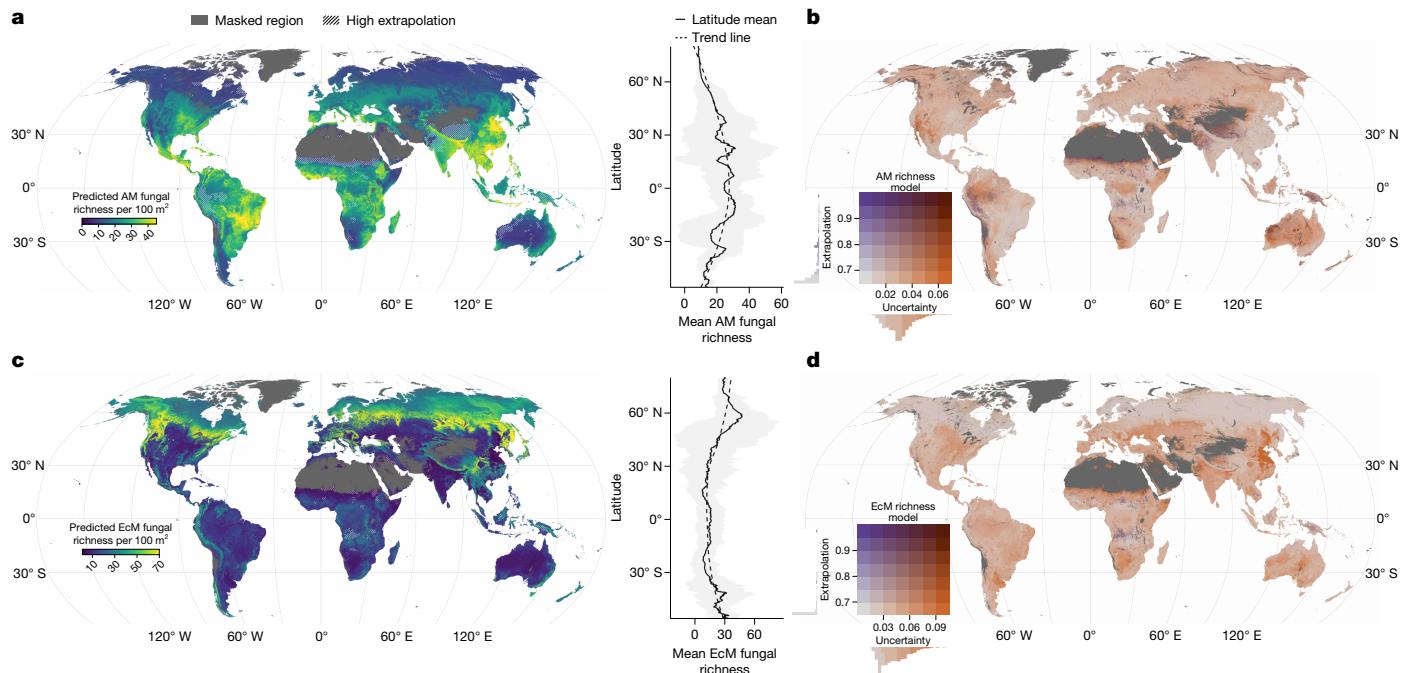


Fig. 2 | Global predictions and latitudinal trends of mycorrhizal fungal richness. **a,c**, Richness maps show the predicted number of AM fungal VT (**a**) and EcM fungal OTUs (**c**) per 100 m² (pixels approximately 1 km²). Predicted richness values are capped at 45 VT per 100 m² (AM) and 70 OTUs per 100 m² (EcM) for visualization. Crosshatches are superimposed over areas that are underrepresented by the training data (highly extrapolated) where model predictions should be interpreted with caution. Plots show mean richness trends across latitude (shaded area is ± 2 s.e.m.). **b,d**, Bivariate maps show the

combination of pixel-level uncertainty (orange gradient) and extrapolation (purple gradient) of AM (**b**) and EcM (**d**) fungi richness predictions. Uncertainty is measured as the coefficient of variation across $n = 100$ bootstrapped model predictions. Extrapolation reflects the degree of environmental difference and geographical distance from samples in the training dataset. Histograms on the charts show the frequency of pixels in different uncertainty and extrapolation levels visualized here. Masked areas (grey) are sparsely vegetated zones and dense urban areas based on global land-cover data.

mycorrhizal fungal biodiversity at global scales^{20,26}. This finding is a problem for the assumption that Earth's mycorrhizal fungi will be automatically protected under new biodiversity initiatives, as current conservation policy tends to rely on plant and animal diversity metrics for determining priority areas and ecosystem value^{5,30}.

Mycorrhizal fungal hotspots

If plant diversity is not a direct indicator of mycorrhizal fungal diversity, then protecting underground biodiversity requires explicit consideration that is based on the development of new high-resolution identification and monitoring approaches. For effective underground conservation, the identification of global hotspots of mycorrhizal biodiversity is particularly important. We therefore used our spatial predictions to locate areas of highly species-rich and endemic mycorrhizal fungal communities (hotspots). Hotspots were defined as pixels in the upper 95th percentile of predicted richness and rarity values globally, as per previous definitions^{25,31}. This analysis resulted in richness cut-off values of 39.9 VT for AM fungi and 60.0 OTUs for EcM fungi, and unitless rarity cut-off values of 0.24 for AM fungi and 1.27 for EcM fungi.

Sampling bias can affect the relative rarity of species, which means it is important to determine whether predicted endemism hotspots are biologically real or an artefact of uneven global inventories of mycorrhizal fungi. We trained the AM and EcM fungal rarity machine-learning models with additional covariate layers of global sampling intensity. These layers were created using kernel density interpolation from sample coordinates in the SSU and ITS training datasets (Methods). This approach enabled us to compare empirical rarity predictions under current sampling intensities against predictions that simulated universally 'high-sampling' scenarios, as previously reported³². For AM fungi,

there was substantial overlap between the empirical and high-sampling rarity hotspot predictions (Extended Data Fig. 3a), which indicated that current sampling efforts are generally capturing the distribution of endemic AM fungal taxa. However, there was large divergence in the empirical and high-sampling predictions of EcM rarity hotspots, particularly in tundra and tropical forest biomes (Extended Data Fig. 3b). This result suggests that increasing sampling efforts in these habitats will probably reveal new or overlooked hotspots of rare and endemic EcM fungi. For both mycorrhizal types, we used the high-sampling predictions for subsequent spatial analyses.

We predicted major hotspots of AM fungal richness and endemicity across savannas of the Brazilian Cerrado, tropical forests across Southeast Asia and Guinean forests in West Africa (Figs. 2a and 4). Our models showed that these areas may contain more than 45 AM fungal species (VT) per 100 m². Species-rich communities of AM fungi were likewise predicted in parts of East and North China in ecosystems that are rapidly undergoing significant anthropogenic land-use changes³³. Smaller regions, like the montane grasslands in Ethiopia, dry forests throughout Central America and Mediterranean woodland and scrubland in Europe, were also predicted to contain highly diverse collections of AM fungi (Fig. 2a). However, these areas were not predicted to be endemicity hotspots. Instead, AM fungal endemicity hotspots were predicted to occur in tropical and subtropical forests in the Congo basin and eastern Amazon basin (Figs. 3a and 4). Although montane grasslands had the most species-rich AM fungal samples in the model training data, this biome did not emerge as a substantial AM fungal richness or rarity hotspot at the global level. This result may be due to the low number of samples from these habitats or the extent and location of montane grasslands compared with other biomes.

We predicted EcM fungal richness hotspots throughout northern forest ecosystems (Figs. 2b and 5). These included much of the Siberian and

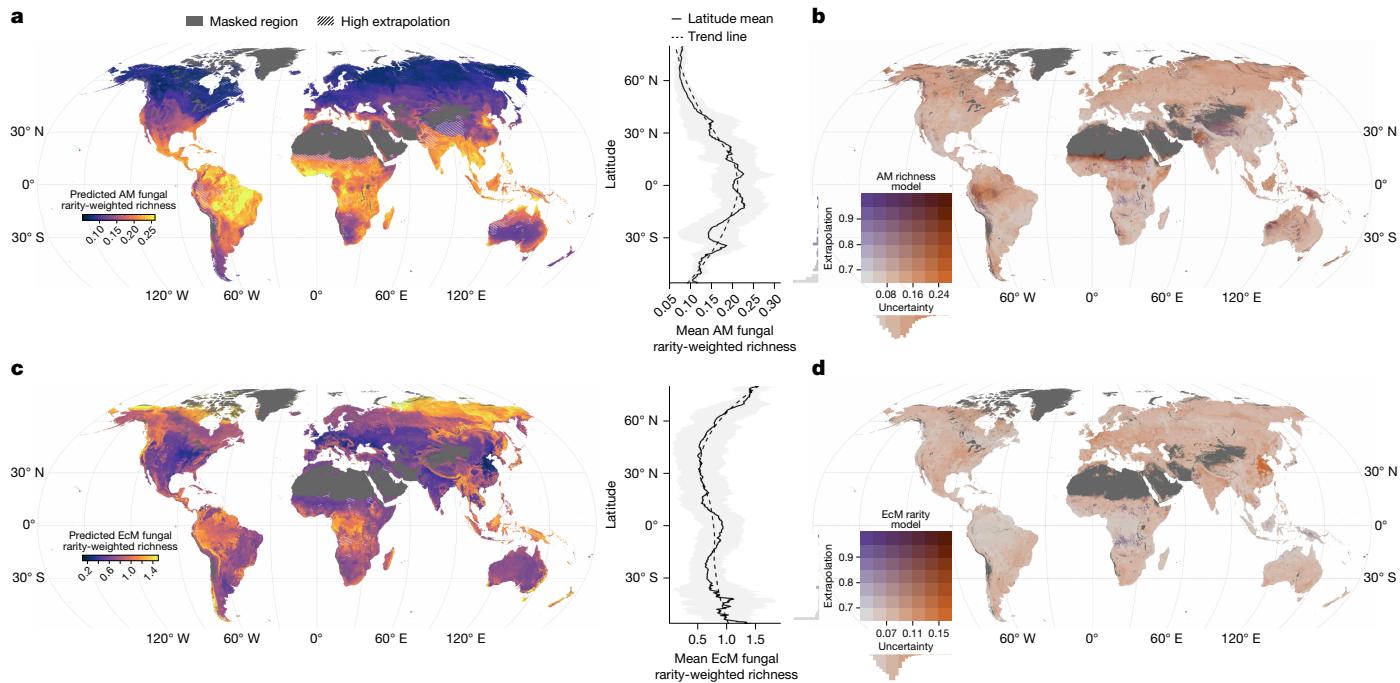


Fig. 3 | Global predictions and latitudinal trends of mycorrhizal fungal endemism. **a,c**, Endemism maps show the predicted rarity-weighted richness of AM fungal VT (**a**) and EcM fungal OTUs (**c**) per 100 m² (pixels approximately 1 km²). Rarity-weighted richness is a unitless metric, and predicted values are capped at 0.26 (AM) and 1.5 (EcM) for visualization. These spatial predictions reflect the simulated high-sampling scenarios to limit unequal sampling effects on rarity patterns. Crosshatches are superimposed over areas that are underrepresented by the training data (highly extrapolated) where model predictions should be interpreted with caution. Plots show mean rarity trends across latitude (shaded area is ± 2 s.e.m.). **b,d**, Bivariate maps show the

combination of pixel-level uncertainty (orange gradient) and extrapolation (purple gradient) of AM (**b**) and EcM (**d**) fungi rarity-weighted richness predictions. Uncertainty is measured as the coefficient of variation across $n = 100$ bootstrapped model predictions. Extrapolation reflects the degree of environmental difference and geographical distance from samples in the training dataset. Histograms on the charts show the frequency of pixels in different uncertainty and extrapolation levels visualized here. Masked areas (grey) are sparsely vegetated zones and dense urban areas based on global land-cover data.

Canadian boreal forest regions, temperate coniferous forests across Western United States and Canadian mountain ranges and temperate broadleaf and mixed forests in Central Europe and the North American Great Lakes region. We predicted that these places could have underground ecosystems averaging >100 EcM fungal species (OTUs) per 100 m². However, most of these northern forest biomes were not identified as EcM fungal endemism hotspots. Instead, we predicted that tundra ecosystems contained the greatest EcM rarity levels at high latitudes (Figs. 3b and 5). We also identified areas of global EcM diversity hotspots—especially rarity hotspots—in tropical biomes, such as tropical moist forests in China's Yunnan–Guizhou plateau and mainland Southeast Asia, montane rainforests across Indonesia and Guianan highland forests in Venezuela. Of particular note are the predicted EcM fungal endemism hotspots in tropical conifer forests in Central America that contain the highest richness of EcM host plant lineages³⁴. EcM fungal richness and rarity hotspots in the Southern hemisphere were predicted to be concentrated primarily in the Andes Mountain range and the temperate mixed forests in southern Chile, the east coast of Australia and New Zealand (that is, all habitats of the Nothofagaceae plant family).

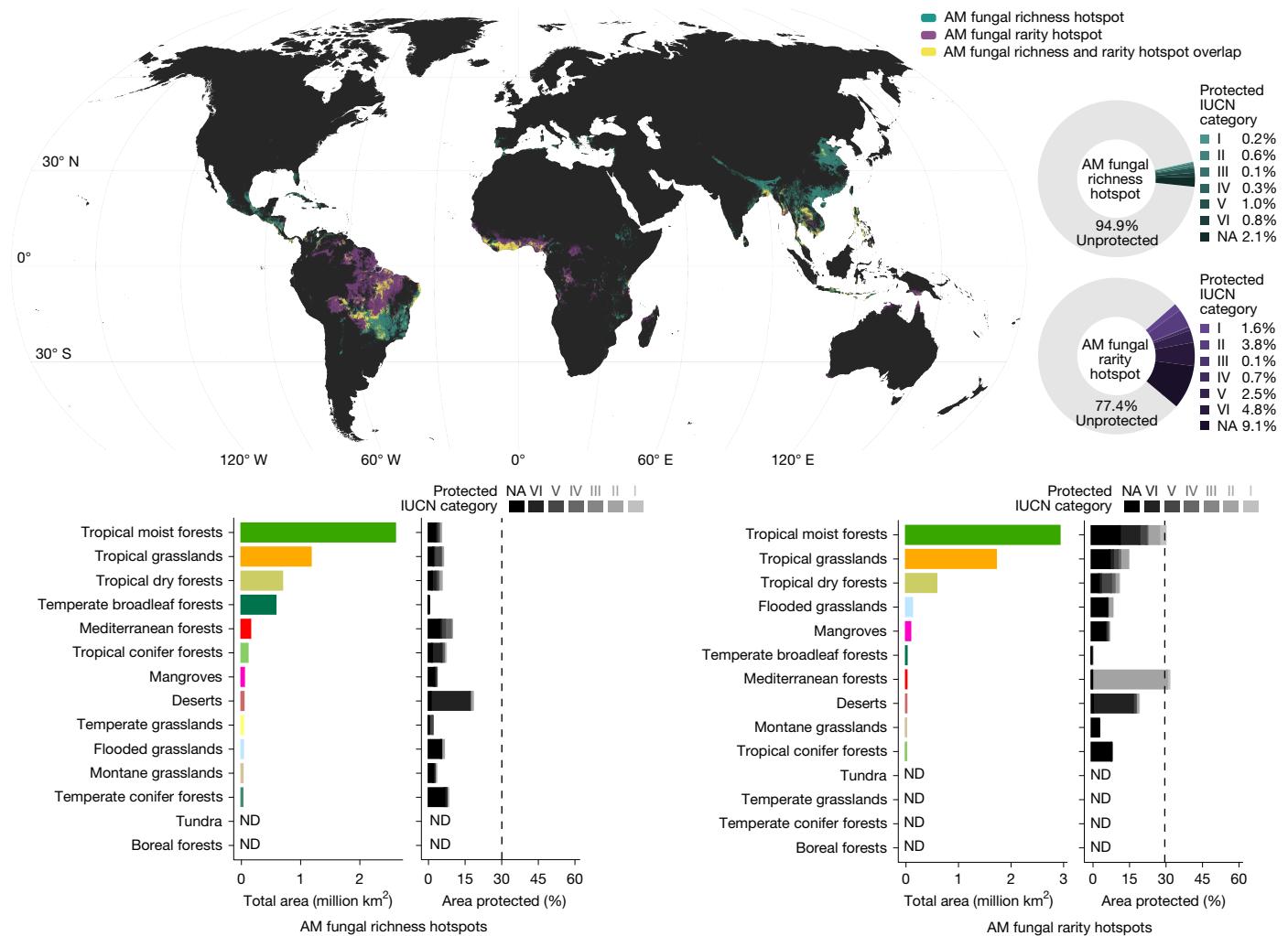
Mycorrhizal hotspot protections

To assess the extent to which mycorrhizal diversity is currently protected, we overlaid mycorrhizal hotspot pixels with protected areas defined by the World Database of Protected Areas³⁵. We then calculated the total extent to which the hotspots were within current protected area delineations. This analysis revealed that the centres of mycorrhizal biodiversity were unequally distributed across biomes; nearly all had

low protective coverage, and the management stringencies for protected hotspots differed among the mycorrhizal types (Figs. 4 and 5). Globally, we predicted that only 9.5% of mycorrhizal richness hotspots occur in protected habitats, including 5.1% (about 280,000 km²) of AM hotspots and 13.9% (around 756,000 km²) of EcM hotspots. Mycorrhizal rarity hotspots were protected at roughly twice the rate (22.9% overall), with 22.6% (about 1.2 million km²) and 23.2% (around 1.3 million km²) of AM and EcM fungal rarity hotspots overlapping with protected areas, respectively. This result is probably because high levels of endemism often result from geographical barriers that cause insularity. Moreover, protected areas are more frequently established in remote, hard-to-access areas where human pressures are anticipated to remain low, regardless of conservation action³⁶.

The protected areas considered here cover roughly 16 million km² and contain everything from strict nature reserves to managed resource areas (International Union for Conservation of Nature (IUCN) management categories I–VI and not applicable (NA) or undefined (NA/undefined)). Analyses of hotspot protections by management category revealed that most protected AM fungal hotspots are in the least-strictly preserved habitats (Fig. 4). Specifically, 76.4% (richness) and 72.8% (rarity) of protected AM hotspots are under categories V, VI or NA/undefined. By contrast, protected EcM fungal hotspots are mostly under the strictest preservation levels, including nearly half of protected EcM richness and rarity hotspot area in IUCN categories I–III (Fig. 5).

At the continent scale, predicted mycorrhizal fungal hotspots in Asia have the lowest protection for both AM fungi (richness = 2.2% and rarity = 7.5% protected) and EcM fungi (richness = 11.3% and rarity = 17.2% protected), even though these hotspots are concentrated in different biomes for each mycorrhizal type (tropical moist forests compared

**Fig. 4 | Arbuscular mycorrhizal fungal hotspots and global protected areas.**

Predicted richness and endemism hotspots (95th percentile of predictions) for AM fungi. Coloured areas on the map show richness hotspots (green), rarity hotspots (purple) and the overlap of richness and rarity hotspots (yellow). Black areas indicate non-hotspots. Bar graphs show the total hotspot size and

percentage overlap with protected areas by biome, and grey bars reflect different IUCN management categories (I, most strictly preserved habitats; NA, unassigned category). The dashed line at 30% hotspot area protected reflects the ambitions of 30 × 30 target goals under the Kunming–Montreal Global Biodiversity Framework. ND, no data.

with boreal and tundra systems; see the Zenodo archive (<https://doi.org/10.5281/zenodo.1487158>³⁷ for continent-level results). Europe has the highest protected AM fungal richness hotspots (19.6% protected), which we predicted to occur mainly in Mediterranean forests across the continent. Australia and Oceania showed the largest protected area overlap with AM fungal rarity hotspots (35.7% protected), primarily in tropical grasslands. Predicted EcM fungal richness hotspots are also the most well protected in Australia and Oceania (40.2% protected), which largely occur throughout temperate broadleaf forests. Both Australia and Oceania and South America have the largest protected area overlap with predicted EcM fungal endemism hotspots (46% protected).

Environmental predictors

Given the distinct hotspot locations and latitude–diversity relationships, we used shapley additive explanations (SHAP) to analyse which of the 24 environmental covariates were the most important predictors of mycorrhizal fungal richness and rarity. This method quantifies the additive contribution of each feature of a machine-learning model (for example, temperature and rainfall) on its predicted values, as opposed to measuring feature importance based on changes in model performance³⁸. Therefore, SHAP is a useful method for interpreting the

magnitude and directionality of predictors in our machine-learning models with respect to changes in mycorrhizal diversity. Moreover, in exploratory tests, the inclusion of spatial predictors (Moran Eigenvector layers) generally did not change the identity or order of important variables in either of the richness models. This result confirmed that the rankings of the most important predictors of mycorrhizal fungal richness are unlikely to be biased by spatially autocorrelated processes not included in the model (Supplementary Figs. 1 and 2).

Among the climate features, temperature (annual mean temperature and maximum temperature of warmest month) was generally the most important climatic predictor across all AM and EcM models (Extended Data Figs. 4 and 5). By calculating mean absolute SHAP values, these variables were two to five times more influential than other predictors, contributing ±0.07 VT to each AM fungal richness prediction and ±0.11 OTUs to each EcM fungal richness prediction on average. Specifically, higher temperatures predicted greater AM fungal richness and lower EcM fungal richness. Potential evapotranspiration was also an influential climate predictor for rarity models, with drier areas associated with higher AM and EcM fungal endemism (Extended Data Figs. 4 and 5). However, the evapotranspiration trends were opposite for EcM diversity metrics, whereby drier areas were associated with lower EcM fungal richness. Using paleoclimate

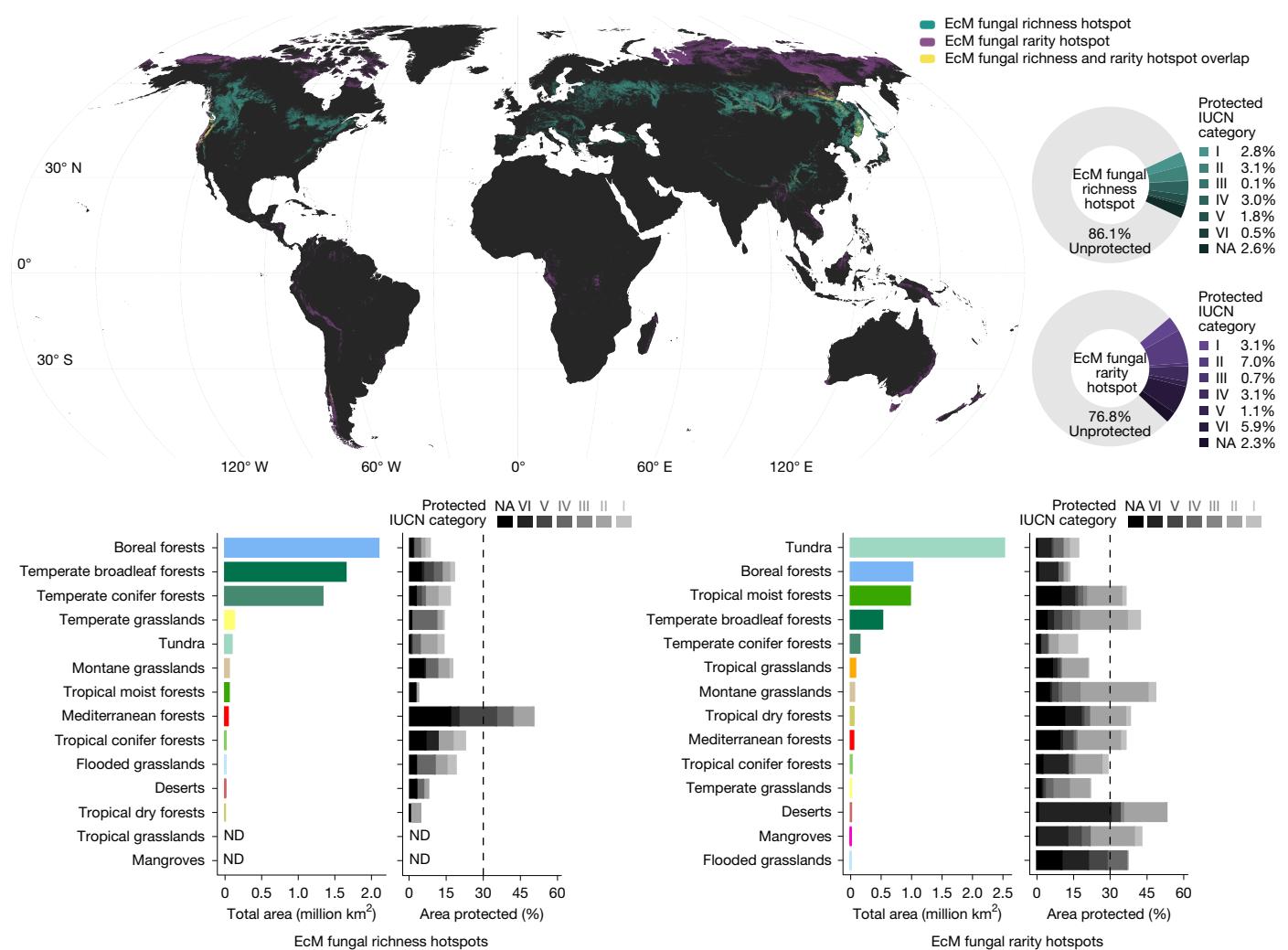


Fig. 5 | Ectomycorrhizal fungal hotspots and global protected areas.

Predicted richness and endemism hotspots (95th percentile of predictions) for EcM fungi. Coloured areas on the map show richness hotspots (green), rarity hotspots (purple) and the overlap of richness and rarity hotspots (yellow). Black areas indicate non-hotspots. Bar graphs show the total hotspot size and

percentage overlap with protected areas by biome, and grey bars reflect different IUCN management categories (I, most strictly preserved habitats; NA, unassigned category). The dashed line at 30% hotspot area protected reflects the ambitions of 30 × 30 target goals under the Kunming–Montreal Global Biodiversity Framework.

data, EcM fungal rarity was strongly and positively associated with climatic stability since the Pliocene (Extended Data Fig. 5; mean absolute SHAP value of ± 0.04 OTUs), which indicates that current endemism patterns reflect the evolutionary history of EcM fungal symbioses over millions of years. These findings add to recent work showing a link between climate and soil fungi at global scales^{17,39}. Moreover, the results probably reflect joint climate tolerances of mycorrhizal fungi and their plant hosts, as well as climatic controls on soil organic matter decomposition^{13,15,17,28,39}.

The importance of climatic controls in predicting mycorrhizal diversity was further supported by strong links between fungal richness and soil organic carbon. Greater soil organic carbon was associated with higher EcM richness and lower AM richness (Extended Data Fig. 4; mean absolute SHAP values of ± 0.11 OTUs and ± 0.08 VT, respectively). This result is consistent with past findings that soil organic carbon stocks are positively associated with EcM plant dominance^{14,40} and probably relates to metabolic differences between AM and EcM fungi in enzymatic capabilities and plant host litter chemistry^{9,41}. In contrast to EcM, AM fungal richness peaked in the tropics, where climate and other biotic conditions favour rapid mineralization of plant litter that may drive AM fungal dominance over EcM symbiosis. Soil organic carbon was not among the top 10 most important predictors of EcM fungal endemism

(mean absolute SHAP value of ± 0.01 OTUs; Extended Data Fig. 5), which helps explain why we predicted substantial EcM rarity hotspots across tropical forests with generally lower soil carbon stocks than temperate or boreal forests¹². Notably, soil pH (as a geospatial covariate layer) was not among the most important predictors identified in AM or EcM models, despite past work describing its importance in shaping overall soil fungal diversity patterns and AM fungal niche traits^{16,42}.

In addition to climate, we were interested in testing how human activities affect diversity patterns. Given that 95% of the Earth's terrestrial surface is affected by some level of human modification⁴³, we used the summed percentage of cultivated and managed areas and urban and built-up habitats to measure anthropogenic land-cover (dense urban centres were masked from predictions; Methods). Anthropogenic land-cover emerged as an important and positive predictor of AM fungal richness (mean absolute SHAP value of ± 0.07 VT; Extended Data Fig. 4). Although sampling bias in the SSU dataset from human-modified habitats may be partially responsible, disturbed ecosystems also tend to contain more ruderal, wind-dispersed AM fungal species that can increase local species richness at the expense of homogenizing regional-scale species composition^{44,45}. Human-mediated land modification was a less important feature in the AM fungal rarity model (mean absolute SHAP value of ± 0.004 VT;

Article

Extended Data Fig. 5) than in the AM fungal richness model, which could be interpreted as supporting the regional homogenization hypothesis. An open question is whether changes in AM fungal-associated vegetation, such as new crops and exotic plants, create more opportunities for diverse plant–AM fungal interactions to occur^{44,46}. Different types of human disturbance are likely to have different effects on AM fungal diversity, and future research on AM fungi in human-modified habitats would benefit from analyses with more specific spatial layers on anthropogenic landscapes.

Aboveground plant biomass was the top predictor for EcM richness and rarity models (richness mean absolute SHAP value of ± 0.13 OTUs; rarity mean absolute SHAP value of ± 0.05), with higher biomass positively predicting higher EcM fungal richness and rarity. In general, this result reflects the dominance of EcM fungi in forest ecosystems, including estimates that 60% of all tree stems on Earth form EcM symbioses¹³. Elevation and slope were also among the most important predictors of EcM rarity (mean absolute SHAP values of ± 0.04 and ± 0.03 , respectively; Extended Data Fig. 5), and we predicted hotspots of endemic EcM fungal communities across mountainous terrain in the Andes, Sierra Madre, Sierra Nevada and Cascade Mountain ranges (Fig. 5). For EcM fungi, aboveground biomass and topography may be more reliable indicators of richness and rarity than plant diversity. This finding is encouraging given that remote-sensing tools are increasingly able to detect complex forest structures of significant ecological value for protection⁴⁷. Our results further underscore the critical planetary benefits of high-biomass forests that are supported by and foster endemic, species-rich EcM fungal communities.

Uncertainty analyses

For spatial predictions of biodiversity to be useful in conservation policy, it is important to measure and communicate the uncertainty associated with each prediction²². This is especially true for microbial biodiversity predictions that are more challenging to verify compared to plant and animal distributions⁴⁸. To do so, we calculated two pixel-level metrics of uncertainty (Methods). In brief, we defined statistical uncertainty as the coefficient of variation in the predictions across $n = 100$ bootstrap samples of the training data to create a confidence interval around the prediction mean of each pixel. Next, we quantified extrapolation as the extent to which a given pixel departs from the sampling locations and the range of environmental conditions in the training data.

The uncertainty range of richness models were similar among mycorrhizal types. The coefficient of variation (the relative dispersion of pixel values around predicted means) ranged mostly between 0 and 0.06 for AM and EcM richness models (Fig. 2). AM fungal rarity models showed slightly higher uncertainties than for EcM fungal rarity interpolations (0.10–0.20 coefficient of variation; Fig. 3). This result is probably caused by differences in the characteristics of AM and EcM fungi that may affect local-scale sample variation or species detection (for example, host specificity and dispersal ability) and differences in the underlying datasets (for example, sample sizes and coverage across environmental gradients). The predictive accuracy of both fungal richness models was comparable (random cross-validation $R^2 = 0.61$ for AM and $R^2 = 0.63$ for EcM), which indicated that both spatial predictions are robust despite the differences in training dataset size. Spatial patterns of uncertainty were different for both mycorrhizal types and diversity metrics (Extended Data Fig. 6). For example, EcM fungal richness models showed approximately double the uncertainty in desert, grassland and tropical forest biomes compared with AM models. AM fungal model uncertainties were highest in areas with low predicted richness, such as desert and tundra regions (Fig. 2b).

Locations with high extrapolation reflected poor representation in the model predictors and were primarily in regions that are consistently underrepresented in global soil biodiversity data⁴⁹ (Extended Data Figs. 7 and 8). For AM fungi, areas that require greater extrapolation

included northern India, the Tibetan plateau, Indonesia and parts of the Amazon rainforest (Figs. 2 and 3). Therefore, extra caution is required when examining AM fungal predictions in these regions. Overall, the EcM predictions were less dependent on model extrapolation because they were built from a larger training dataset (over five times more ITS samples distributed across a wider range of environmental conditions than the SSU samples used for AM fungi). However, there were small regions with a high degree of EcM model extrapolation in Sub-Saharan Africa, northern Canada and Southeast Asia (Figs. 2 and 3).

Important critiques of machine-learning geospatial approaches have been recently noted²¹. We performed the following actions to address these concerns: (1) accounted for spatial autocorrelation variables in the model (Supplementary Figs. 1 and 2); (2) created spatial blocks to limit the distance between sample and prediction locations for additional cross-validation (Supplementary Fig. 3); and (3) provided clear data on prediction uncertainty (Figs. 2 and 3 and Extended Data Figs. 7 and 8). Moreover, we conducted multiple cross-validation procedures as there is currently no consensus on best practices for validating spatial models and measuring their inaccuracies (Methods). Overall, our models performed similar to (or better than) the accuracy of recent machine-learning predictive mapping approaches used for other soil organisms (Methods and Extended Data Fig. 9).

Conclusions

Our high-resolution, global-scale maps of AM and EcM fungal diversity suggest that <10% of mycorrhizal richness hotspots currently exist in protected areas. The same analysis for non-fungal taxa showed that both plant and animal biodiversity hotspots are much better protected. Replicating the analysis with spatial data of other taxonomic groups showed roughly 3-fold higher protection of richness hotspot areas for vascular plants (27% protected), trees (34% protected), ants (28% protected) and vertebrates (33–41% protected for amphibians, birds, mammals and reptiles), which is probably because threatened species in these groups drive most conservation interventions^{31,50}. Endemic mycorrhizal fungi may be faring better: overall 22.9% of predicted rarity hotspot areas overlapped with protected areas, similar to the protection coverage of vertebrate rarity centres (19–21% protected for amphibians, birds, mammals and reptiles). To fully realize the 30×30 target goals of protecting Earth's biodiversity under the Kunming–Montreal Global Biodiversity Framework, it is necessary to identify underground conservation priorities, set monitoring benchmarks and create specific restoration plans⁵¹. For example, our data can be used to help develop land-management strategies to protect and maintain mycorrhizal diversity reservoirs⁵², including identifying soil-management practices developed over millennia by local populations⁵³. We also recommend that any use of these predictive biodiversity maps in a specific locale should be cross-referenced with the model uncertainty and extrapolation layers in the same area to maximize positive conservation outcomes²².

Our results should be seen as a first step towards understanding the baseline protected levels of Earth's mycorrhizal fungal diversity. Overlaying protected areas on biodiversity hotspots does not tell us how successful these conservation strategies are at maintaining or improving mycorrhizal fungal biodiversity. Because IUCN management categories vary in strictness of habitat preservation, it is also important to analyse hotspot protections on the basis of these management categories, few of which are entirely free of human pressures. A critical next step will be to measure the effectiveness of protected areas to mitigate fungal biodiversity loss, as has been done using counterfactual analysis for other organisms⁵⁴. This strategy will help determine which protected area conditions and objectives best deliver conservation benefits to mycorrhizal fungi—such as focusing on rare or threatened species, unique ecosystems and their entire communities, vital ecosystem services and/or locations of cultural and social significance. Conservation efforts will also be more effective by continuously monitoring

mycorrhizal fungi, as they have known vulnerabilities to environmental change and disturbance^{19,52}, and there is a need to track their recovery following conservation and restoration actions.

Owing to insufficient data, we were unable to robustly model ErM or orchid mycorrhizal fungal richness patterns. For instance, the average ErM fungal richness per sample was 0.6 species, and <3% of samples contained more than 4 ErM fungal species. These are both important fungal symbioses to understand in a global context because of their unique contributions to plant ecology and ecosystem functions¹¹. Additional sequencing projects and data-mining efforts will facilitate large-scale analyses of these mycorrhizal types. A second issue is that using SSU data to understand the biogeography of AM fungal richness and rarity may be problematic owing to overly conservative species definitions that could underrepresent the true diversity of this mycorrhizal type (Methods). AM fungi have a unique genetic organization with high intragenomic variation that makes their sequencing and species definition more difficult than for EcM fungi, and only 332 AM fungal species have been formally described⁸. Alternative approaches, such as long-read sequencing of both environmental AM fungi and culture collections and the use of multiple metabarcoding target regions, may provide additional insights into the definition and distribution of AM fungal biodiversity⁵⁵.

Although our models performed similar to (or better than) those for other soil organisms (for example, nematodes, springtails and earthworms; Methods), they are still an imperfect representation of mycorrhizal patterns in nature. In particular, 1 km² is a relatively coarse spatial scale to quantify the average soil mycorrhizal richness expected per 100 m². There can be large landscape heterogeneity in a pixel of this size, and soil fungal sampling is usually restricted to a smaller collection grid and sequencing <1 g of soil. Currently, 1 km² is the highest spatial resolution possible for making reliable global predictions given the coverage of relevant covariate layers, but this technical limitation will be overcome as more global environmental datasets are created at 250-m and 30-m resolution, closer to the scale of field-sampling methods. We are also actively exploring how much within-pixel variation affects model accuracy as part of an ongoing effort to independently ground-truth these spatial predictions. Ground-truth validation is particularly important for the following regions: (1) under-sampled environments poorly represented in the model; (2) regions with high uncertainty despite good training data coverage; (3) areas with high landscape-related heterogeneity (for example, steep mountains); and (4) areas that are particularly vulnerable to rapid climate change and habitat loss from human activities.

Given the importance of these organisms for the productivity of ecosystems and the functioning of landscapes, it is surprising that mycorrhizal diversity remains highly underrepresented in conservation agendas. Explicit consideration of mycorrhizal biodiversity hotspots like those identified here can help direct and implement conservation strategies towards protecting the most diverse and endangered underground ecosystems.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-025-09277-4>.

- Högberg, M. N. & Högberg, P. Extramatrical ectomycorrhizal mycelium contributes one-third of microbial biomass and produces, together with associated roots, half the dissolved organic carbon in a forest soil. *New Phytol.* **154**, 791–795 (2002).
- van Der Heijden, M. G., Martin, F. M., Selosse, M. A. & Sanders, I. R. Mycorrhizal ecology and evolution: the past, the present, and the future. *New Phytol.* **205**, 1406–1423 (2015).
- Hawkins, H. J. et al. Mycorrhizal mycelium as a global carbon pool. *Curr. Biol.* **33**, R560–R573 (2023).
- Cavicchioli, R. et al. Scientists' warning to humanity: microorganisms and climate change. *Nat. Rev. Microbiol.* **17**, 569–586 (2019).
- Guerra, C. A. et al. Tracking, targeting, and conserving soil biodiversity. *Science* **371**, 239–241 (2021).
- Větrovský, T. et al. GlobalFungi, a global database of fungal occurrences from high-throughput-sequencing metabarcoding studies. *Sci. Data* **7**, 228 (2020).
- Tedersoo, L. et al. The Global Soil MycoBiome consortium dataset for boosting fungal diversity research. *Fungal Divers.* **111**, 573–588 (2021).
- Větrovský, T. et al. GlobalAMFungi: a global database of arbuscular mycorrhizal fungal occurrences from high-throughput sequencing metabarcoding studies. *New Phytol.* **240**, 2151–2163 (2023).
- Frey, S. D. Mycorrhizal fungi as mediators of soil organic matter dynamics. *Annu. Rev. Ecol. Evol. Syst.* **50**, 237–259 (2019).
- Tedersoo, L. & Bahram, M. Mycorrhizal types differ in ecophysiology and alter plant nutrition and soil processes. *Biol. Rev.* **94**, 1857–1880 (2019).
- Brundrett, M. C. & Tedersoo, L. Evolutionary history of mycorrhizal symbioses and global host plant diversity. *New Phytol.* **220**, 1108–1115 (2018).
- Crowther, T. W. et al. The global soil community and its influence on biogeochemistry. *Science* **365**, eaav0550 (2019).
- Steidinger, B. S. et al. Climatic controls of decomposition drive the global biogeography of forest-tree symbioses. *Nature* **569**, 404–408 (2019).
- Soudzilovskia, N. A. et al. Global mycorrhizal plant distribution linked to terrestrial carbon stocks. *Nat. Commun.* **10**, 5077 (2019).
- Kivlin, S. N., Hawkes, C. V. & Treseder, K. K. Global diversity and distribution of arbuscular mycorrhizal fungi. *Soil Biol. Biochem.* **43**, 2294–2303 (2011).
- Tedersoo, L. et al. Global diversity and geography of soil fungi. *Science* **346**, 1256688 (2014).
- Větrovský, T. et al. A meta-analysis of global fungal distribution reveals climate-driven patterns. *Nat. Commun.* **10**, 5142 (2019).
- Guerra, C. A. et al. Global hotspots for soil nature conservation. *Nature* **610**, 693–698 (2022).
- Tedersoo, L. et al. Global patterns in endemism and vulnerability of soil fungi. *Global Change Biol.* **28**, 6696–6710 (2022).
- Cameron, E. K. et al. Global mismatches in aboveground and belowground biodiversity. *Conserv. Biol.* **33**, 1187–1192 (2019).
- Meyer, H. & Peblesma, E. Machine learning-based global maps of ecological variables and the challenge of assessing them. *Nat. Commun.* **13**, 2208 (2022).
- Jansen, J. et al. Stop ignoring map uncertainty in biodiversity science and conservation policy. *Nat. Ecol. Evol.* **6**, 828–829 (2022).
- Albuquerque, F., Astudillo-Scalia, Y., Loyola, R. & Beier, P. Towards an understanding of the drivers of broad-scale patterns of rarity-weighted richness for vertebrates. *Biodivers. Conserv.* **28**, 3733–3747 (2019).
- Kinlock, N. L. et al. Explaining global variation in the latitudinal diversity gradient: meta-analysis confirms known patterns and uncovers new ones. *Global Ecol. Biogeogr.* **27**, 125–141 (2018).
- Sabatini, F. M. et al. Global patterns of vascular plant alpha diversity. *Nat. Commun.* **13**, 4683 (2022).
- Toussaint, A. et al. Asymmetric patterns of global diversity among plants and mycorrhizal fungi. *J. Veg. Sci.* **31**, 355–366 (2020).
- Kokkoris, V. et al. Codependency between plant and arbuscular mycorrhizal fungal communities: what is the evidence? *New Phytol.* **228**, 828–838 (2020).
- Read, D. J. Mycorrhizas in ecosystems. *Experientia* **47**, 376–391 (1991).
- Phillips, R. P., Brzostek, E. & Midgley, M. G. The mycorrhizal-associated nutrient economy: a new framework for predicting carbon–nutrient couplings in temperate forests. *New Phytol.* **199**, 41–51 (2013).
- Jung, M. et al. Areas of global importance for conserving terrestrial biodiversity, carbon and water. *Nat. Ecol. Evol.* **5**, 1499–1509 (2021).
- Jenkins, C. N., Pimm, S. L. & Joppa, L. N. Global patterns of terrestrial vertebrate diversity and conservation. *Proc. Natl. Acad. Sci. USA* **110**, E2602–E2610 (2013).
- Kass, J. M. et al. The global distribution of known and undiscovered ant biodiversity. *Sci. Adv.* **8**, eabp9908 (2022).
- Chen, C. et al. China and India lead in greening of the world through land-use management. *Nat. Sustain.* **2**, 122–129 (2019).
- Tedersoo, L. in *Biogeography of Mycorrhizal Symbiosis* Vol. 230 (ed. Tedersoo, L.) 469–531 (Springer, 2017).
- Bingham, H. C. et al. Sixty years of tracking conservation progress using the World Database on Protected Areas. *Nat. Ecol. Evol.* **3**, 737–743 (2019).
- Joppa, L. N. & Pfaff, A. High and far: biases in the location of protected areas. *PLoS ONE* **4**, e8273 (2009).
- van den Hoogen, J., van Nuland, M. & Kumar, S. Data and code for: Global Hotspots of Mycorrhizal Fungal Richness are Poorly Protected. Zenodo <https://doi.org/10.5281/zenodo.14871588> (2025).
- Lundberg, S. M. et al. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2**, 56–67 (2020).
- Mikryukov, V. et al. Connecting the multiple dimensions of global soil fungal diversity. *Sci. Adv.* **9**, eadj8016 (2023).
- Averill, C., Turner, B. L. & Finzi, A. C. Mycorrhiza-mediated competition between plants and decomposers drives soil carbon storage. *Nature* **505**, 543–545 (2014).
- Hicks Pries, C. E. et al. Differences in soil organic matter between EcM- and AM-dominated forests depend on tree and fungal identity. *Ecology* **104**, e3929 (2023).
- Davison, J. et al. Temperature and pH define the realised niche space of arbuscular mycorrhizal fungi. *New Phytol.* **231**, 763–776 (2021).
- Kennedy, C. M., Oakleaf, J. R., Theobald, D. M., Baruch-Mordo, S. & Kiesecker, J. Managing the middle: a shift in conservation priorities based on the global human modification gradient. *Glob. Change Biol.* **25**, 811–826 (2019).
- Chaudhary, V. B., Nolimal, S., Sosa-Hernández, M. A., Egan, C. & Kastens, J. Trait-based aerial dispersal of arbuscular mycorrhizal fungi. *New Phytol.* **228**, 238–252 (2020).

45. Guerra, C. A. et al. Global projections of the soil microbiome in the Anthropocene. *Global Ecol. Biogeogr.* **30**, 987–999 (2021b).
46. Guzman, A. et al. Crop diversity enriches arbuscular mycorrhizal fungal communities in an intensive agricultural landscape. *New Phytol.* **231**, 447–459 (2021).
47. Lang, N., Jetz, W., Schindler, K. & Wegner, J. D. A high-resolution canopy height model of the Earth. *Nat. Ecol. Evol.* **7**, 1778–1789 (2023).
48. Barron, E. Conservation of abundance: How fungi can contribute to rethinking conservation. *Conserv. Soc.* **21**, 99–109 (2023).
49. Guerra, C. A. et al. Blind spots in global soil biodiversity and ecosystem function research. *Nat. Commun.* **11**, 3870 (2020).
50. Senior, R. A. et al. Global shortfalls in documented actions to conserve biodiversity. *Nature* **630**, 387–391 (2024).
51. Labouyrie, M. et al. Patterns in soil microbial diversity across Europe. *Nat. Commun.* **14**, 3311 (2023).
52. Averill, C. et al. Defending Earth's terrestrial microbiome. *Nat. Microbiol.* **7**, 1717–1725 (2022).
53. Fleischman, F. et al. Restoration prioritization must be informed by marginalized people. *Nature* **607**, E5–E6 (2022).
54. Langhammer, P. F. et al. The positive impact of conservation action. *Science* **384**, 453–458 (2024).
55. Lutz, S. et al. Global richness of arbuscular mycorrhizal fungi. *Fungal Ecol.* **74**, 101407 (2025).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

SPUN Mapping Consortium

Noelia Barriga-Medina¹⁶, Paola Bonfante¹⁷, Alper Cevirgel¹⁸, Peter Chatanga¹⁹, Bala Chaudhary²⁰, Matteo Chialva¹⁷, S. Caroline Daws^{11,21}, Mark Day²², Aurélie Deveaux²³, Vincent Diringer¹, Katie Franklin²⁴, Nicole Hynson²⁵, Alyona Koshkina²⁶, Luisa Lanfranco¹⁷, Antonio Leon-Reyes¹⁶, Sol Llerena¹⁶, Liteboho Maduna¹⁹, Francis Martin²³, Jean-Paul Maurice²⁷, Sebolele Molete¹⁹, Andrés Avella Muñoz²⁸, Liam F. Nokes²⁰, Cesar A. Parra Aldana^{29,30}, Rachel Pringle¹, Dario X. Ramirez-Villacis^{16,31,32}, Juan David Rosales¹, Cosmo Sheldrake¹, Aigerim Soltabayeva²⁴, Genevieve Stephens²³, J. Benjamin Stielow³⁵, Nicolas Suberbielle³⁶, Matsepo M. Taole¹⁹, Lorenzo Tolari²⁴, Cristian Moreno Tormo¹, Jacob Ulzen³⁷, Rocío Urrutia-Jalabert^{38,39}, Zander S. Venter⁴⁰, Andressa M. Venturini¹¹ & Alex S. Wegmann⁴¹

¹⁶Laboratorio de Biotecnología Agrícola y de Alimentos, Universidad San Francisco de Quito, Quito, Ecuador. ¹⁷Department of Life Sciences and Systems Biology, University of Torino, Torino, Italy. ¹⁸Department of Medical Microbiology and Infection Prevention, Virology and Immunology Research Group, University Medical Center Groningen, Groningen, The Netherlands. ¹⁹Department of Biology, Faculty of Science and Technology, National University of Lesotho, Roma, Lesotho. ²⁰Department of Environmental Studies, Dartmouth College, Hanover, NH, USA. ²¹Outer Coast, Sitka, AK, USA. ²²Royal Society for the Protection of Birds, Cambridge, UK. ²³Université de Lorraine, INRAE, UMR Interactions Arbres/Microorganismes, INRAE-Nancy, Champenoux, France. ²⁴The Nature Conservancy, Honolulu, HI, USA. ²⁵Pacific Biosciences Research Center, University of Hawai'i at Mānoa, Honolulu, HI, USA. ²⁶Association for the Conservation of Biodiversity of Kazakhstan, Astana, Kazakhstan. ²⁷Mycological Society of Lorraine, Neufchâteau, France. ²⁸Facultad del Medio Ambiente y Recursos Naturales, Universidad Distrital Francisco José de Caldas, Bogotá, Colombia. ²⁹Ingeniero Forestal, Universidad Distrital, Bogota, Colombia. ³⁰Magister en Bosques y Conservación Ambiental, Universidad Nacional de Colombia, Bogota, Colombia. ³¹Department of Microbial Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, The Netherlands. ³²Plant-Microbe Interactions, Department of Biology, Science for Life, Utrecht University, Utrecht, The Netherlands. ³³Faculty of Exact and Natural Sciences, Taxonomy and Ecology of Fungi Group, University of Antioquia, Medellín, Colombia. ³⁴Biology Department, School of Science and Humanities, Nazarbayev University, Astana, Kazakhstan. ³⁵Helmholtz Institute for One Health, Bioinformatics Group, Greifswald, Germany. ³⁶Conservatoire Botanique National de Corse, Office de l'Environnement de la Corse, Corte, France. ³⁷Forest and Horticultural Crops Research Centre, University of Ghana, Accra, Ghana. ³⁸Departamento de Ciencias Forestales, Facultad de Ciencias Agropecuarias y Medioambiente, Universidad de la Frontera, Temuco, Chile. ³⁹Center of Climate and Resilience Research, CR2, Santiago, Chile. ⁴⁰Terrestrial Ecology Section, Norwegian Institute for Nature Research, Oslo, Norway. ⁴¹The Nature Conservancy, Sacramento, CA, USA.

Methods

Fungal data

Fungal occurrence records were generated from the data mining of published ITS and SSU sequencing studies collected for the GlobalFungi and GlobalAMFungi databases (see previous studies^{6,8} for details) and the ITS region from full-length sequences in the Global Soil Mycobiome consortium database⁷. For EcM occurrences used in this study, we considered only samples of the ITS2 barcode deposited in the GlobalFungi dataset as this marker is less biased by length variability compared with ITS1 (refs. 56,57). Moreover, ITS2 is better represented in the source database and ITS1 samples failed technical validation (see below). In brief, raw sequences and metadata from 255 ITS studies (representing the fourth release of the GlobalFungi database) were processed using an established bioinformatic pipeline that incorporates sequence quality checks, extraction of full ITS2 fungal regions using ITSx (v.1.1.2)⁵⁸ and clustering into OTUs at 97% similarity level with subsequent exclusion of global singletons and chimeric sequences using USEARCH (v.11.0.0667)⁵⁹. We used BLASTN searches against UNITE (v.8.3, released 10.5.2021) to assign putative taxonomy to non-singleton OTUs. Default BLAST parameters were used, and representative sequences were considered to belong to the closest BLAST hit genera in the case of >92% similarity and >95% coverage. Considering that the majority of EcM fungal genera do not include species belonging to other functional guilds, we found that these BLAST results criteria were reasonable. We excluded OTUs represented by sequences with e value > 10⁻⁵⁰. The resulting OTU table with taxonomy assignments was compared against the Fungal Traits database (v.1.1) to subset EcM fungi⁶⁰.

In early tests, we used the ITS dataset to model AM fungi, but their low abundance detected from using this target region was leading to clearly erroneous spatial predictions (for example, zero AM species predicted in regions known to be dominated by AM vegetation). As a result, SSU data from the GlobalAMFungi database were used to estimate AM fungal species richness⁸. In brief, raw sequences and metadata from 45 SSU studies (representing the first release of the GlobalAMFungi database) were subjected to sequence quality checks, trimming of the sequences to the V4 region of the SSU and assignment of the sequences to VT from the MaarjAM database (v.5.6.2019 release)⁶¹. We also used BLASTN with a required sequence coverage of ≥98% and sequence similarity ≥97%, based on the most recent release of type sequences of VT. VT were defined on the basis of phylogenetic grouping of SSU sequences at roughly species-level diversity (which may result in conservative taxonomic designations of AM fungal taxa^{55,62}). Molecular identification based on VT is typically used for AM fungal metabarcoding studies using the SSU region, and OTUs are more traditionally used for general fungal community sequencing (including EcM fungi) based on the ITS region. Both VT and OTUs are commonly used to approximate species-level assignments in sequencing studies of mycorrhizal fungi. The EcM OTU data used for this analysis is the result of clustering of GlobalFungi sequences followed by taxonomic assignment. By contrast, the AMF pipeline results in more conservative estimates of diversity, as SSU sequences are assigned to representative VT using stringent BLAST parameters. This results in the discarding of sequences without a good match to these known VT, which is necessary to filter non-fungal sequences (for example, Annelids) but also probably removes undescribed AM fungal taxa. In total, 2.8 billion fungal sequences were processed across 24,982 samples, which identified 332 VT of AM fungi (encompassing 86% of defined VT in the MaarjAM database) and 164,439 EcM fungal OTUs.

Richness estimates

We used analytical estimators for the rarefaction and extrapolation of species richness to measure the number of mycorrhizal species (OTUs or VT) in each sample⁶³. This approach creates a sequencing depth-based rarefaction and extrapolation sampling curve (that is, a

sample-specific species accumulation curve), with diversity estimates and 95% confidence intervals calculated at curve asymptotes (using the R package iNEXT⁶⁴). Extrapolation end points were computed as twice the sequencing depth for each sample (default setting). Use of the rarefaction and extrapolation approach to estimate mycorrhizal richness enables a robust comparison of mycorrhizal patterns across studies of multiple sequencing technologies (although variation in error rates may persist), primer sets and sequencing intensities. The extrapolated richness estimates here are within similar ranges to past studies that estimated global diversity of mycorrhizal fungi^{26,39}. Moreover, sample richness values were highly consistent among 96%, 97% and 98% clustering similarity thresholds (Supplementary Fig. 4).

We removed outliers in mycorrhizal richness estimates before spatial modelling. Two Australian studies with ITS samples from desert and Mediterranean biomes had unusually high EcM OTU richness estimates and standard deviations (two orders of magnitude higher than other Australian ecoregions^{65,66}). Moreover, these studies have previously been identified as potentially inaccurate based on a recent database comparison⁶. All samples from these two studies were removed ($n = 2,070$). We then filtered samples by biome by removing estimated richness values that were more than five times the interquartile range higher than the biome-level median estimate. We only filtered values at the highest end of the distribution to avoid potentially removing ‘true’ estimates of low or zero mycorrhizal richness at a given location. See Supplementary Table 1 for a summary of outliers removed per biome.

We also calculated rarity-weighted richness to estimate the relative endemism of mycorrhizal fungal communities. Using samples from the outlier filtered dataset (see above), we created a species-level prevalence score as the number of samples for which each OTU or VT occurs and then summed the inverse prevalence score of all species present in each sample⁶⁷. This rarity metric is useful for identifying site endemism or areas with a high concentration of rare taxa²³. Rarity-weighted richness was originally used with a geographically unbiased dataset for which each pixel contained information on bird observations⁶⁷, but has since proven useful in other contexts to predict patterns of relative species endemism^{32,68}. To account for the possibility that geographically uneven sampling efforts affect fungal species prevalence scores and biases predictions of fungal endemism, we simulated a global high-sampling scenario for rarity spatial models (see below).

Geospatial modelling, validation and analysis

Spatial predictions of mycorrhizal fungal richness and rarity were created using a random forest modelling approach (modified from ref. 69). We first sampled a collection of >72 global environmental covariates at each of the locations in the dataset. To reduce overfitting and variance inflation, we removed highly correlated variables such that the final set of predictors comprised 24 environmental variables. These contained macroclimatic, soil texture and physicochemical information, vegetation, radiation and topographical indices and anthropogenic variables. Details of all predictors, including descriptions, units and source information, are previously described⁷⁰ and provided in Supplementary Table 2. Variables describing soil structure and physicochemical properties were obtained from SoilGrids⁷¹, limited to the upper 5 cm of soil. Climate information (that is, mean annual temperature, annual precipitation, monthly maximum temperature and precipitation seasonality) was obtained from CHELSA⁷², and climate stability over palaeoecological timescales was obtained from the Climate Stability Index⁷³. Spectral vegetation indices (that is, MODIS NPP product MOD17A3HGF V6.1, averaged annually) were obtained from the Google Earth Engine Data Catalogue. We used aboveground biomass data from CDIAC⁷⁴. We used EarthEnv^{75,76} to collect consensus land cover information (including human development percentage) and eight topographical layers that capture a broad range of complex terrain features (for example, geographical isolation). The predicted plant diversity layer was obtained from a previous study⁷⁷. The potential

Article

evapotranspiration layer was obtained from CGIAR⁷⁸. Resolve Ecoregion classifications were used to categorize sampling locations into biomes⁷⁹. All spatial covariate layers were reprojected and resampled to a unified pixel grid in EPSG:4326 (WGS84) at 30-arcsec resolution (approximately 1 km² at the equator). Areas covered by permanent snow or ice (for example, the Greenland ice cap and glaciated mountain ranges, which were identified using SoilGrids⁷¹), barren land cover (sparsely vegetated regions defined in EarthEnv^{75,76}) and highly urban and built-up areas were excluded from the analyses. Antarctic areas were excluded from analyses owing to limited coverage of covariate layers in the region.

To harmonize the data across the different experimental approaches of the original studies, we included six types of project-specific variables. This included two continuous variables (sample area and soil mass for DNA extraction) and the following four categorical variables: sequencing platform (Illumina, IonTorrent, PacBio, 454Roche and DNBSEQ-G400); target gene region (ITS1 and ITS2); sample type (soil, topsoil and rhizosphere); and primer set (37 different kinds). As an intermediate step, we created random forest spatial models (described below) to derive predicted values of mycorrhizal richness for comparison to observed data at the same geolocation. We filtered samples from the categorical project-specific variables that failed technical validation by comparing observed with predicted richness trends (failures were defined by zero correlation). This process removed all DNBSEQ-G400 samples ($n = 3,738$), all ITS1 samples ($n = 33$; these samples also had corresponding ITS2 sequence data that were retained) and 14 different primer sets (total $n = 809$), which were filtered before running the final models. To create spatial predictions, we harmonized the project-specific variables to the following most common levels: sample area = 100 m²; soil mass for DNA extract = 0.5 g; sequencing platform = Illumina; target gene = ITS2 (implicit because ITS1 samples were removed); sample type = soil; and primer set = ITS3–ITS4 for EcM. Before modelling, all categorical project-specific variables were transformed to binary variables (that is, one-hot encoding or dummy variables). See Supplementary Figs. 5–10 for more details on technical validation and data filtering.

After outlier removal and validation, the SSU dataset contained 3,234 samples (332 total VT) for AM fungal analysis, and the ITS dataset had 17,519 samples (41,086 total OTUs) for EcM fungal analysis. To create the training datasets and to reduce potential overfitting, we used only distinct observations of mycorrhizal fungal richness (that is, samples with non-identical richness or rarity values). This meant that when multiple samples fell within the same 1-km² pixel, we removed duplicate richness estimates to ensure that each observation contributed equally to the learning process of the model (multiple identical observations caused the random forest models to overfit towards those combinations of richness levels and environmental variables). Using only distinct observations enabled us to retain important within-pixel variation rather than aggregating across multiple richness estimates per pixel. To deal with the zero values in the dataset, rarefied richness values were $\log_{10}[x + 1]$ transformed before modelling. In the random forest models, we fixed the number of trees at 250 and explored different combinations of hyperparameters by varying the number of variables considered at each split (4, 6, 8, 10 or 12) and the minimum number of samples per leaf (2, 4, 6, 8, 10 or 12). In total, 30 hyperparameter combinations were tested for each AM and EcM response variable.

There is ongoing debate about the best methodology for validating spatial models^{80–83}. As a result, we tested each model using both random cross-validation and k -fold nearest neighbour distance matching cross-validation (kNNMD)⁸⁴. Random folds were assigned stratified per biome to ensure equal coverage of each fold. For both AM and EcM, the final predictions are an ensemble (mean) of the top ten best performing models based on coefficient of determination (R^2) with kNNMD ten-fold cross validation. This approach resulted in the following predictive accuracy metrics for richness models: AM random

cross-validated $R^2 = 0.61$; AM kNNMD cross-validated $R^2 = 0.20$; EcM random cross-validated $R^2 = 0.63$; and EcM kNNMD cross-validated $R^2 = 0.28$ (Extended Data Fig. 9a). The predictive accuracy metrics for rarity models were as follows: AM random cross-validated $R^2 = 0.56$; AM kNNMD cross-validated $R^2 = -2.55$; EcM random cross-validated $R^2 = 0.61$; and EcM kNNMD cross-validated $R^2 = 0.19$ (Extended Data Fig. 9b). Overall, these performance estimates were similar to (or exceeded) the predictive accuracy of global machine-learning models for other soil organisms^{69,85,86}. The poor performance of the AM fungal rarity model based on kNNMD cross-validation suggests there may be spatial or environmental structures in the dataset that the model is unable to detect, which is possible if the VT dataset fails to sufficiently capture AM fungal endemism patterns because of conservative taxonomic assignments from the AM fungal reference database⁶¹. We used the SHAP algorithm to interpret the importance and directionality of environmental features³⁸. Standard feature importance algorithms, like the ones implemented in random forest or boosted regression tree models, measure feature importance as a change in model performance and are more susceptible to feature collinearity. By contrast, SHAP handles feature collinearity relatively well and is based on the magnitude of feature contribution. We selected SHAP as the most suitable method here because of the possibility of feature collinearity among the environmental predictors and because we were interested in the effect of the features on the predicted value (as opposed to the effect on model error) to interpret the importance and directionality of the features.

As we expected the model performance to decrease with distance to training locations, we also performed spatially buffered leave-one-out cross-validation (SLOO-cv)⁸⁰, a computationally intensive approach in which a separate model is trained for every unique location in the dataset and leaving out locations in a range of buffer sizes. At larger buffer sizes (that is, >500 km), SLOO-cv R^2 values were in line with kNNMD cross-validation R^2 values. To transform these findings into a visual representation, we plotted the R^2 values against the distance to the nearest sampling location to create a spatial product (Supplementary Fig. 3). The resulting map was consistent with our previously described extrapolation map.

To generate a spatial understanding of our predictive accuracy, we created 100 bootstrap samples by resampling the training datasets with replacement using biome-based stratification. Leveraging the hyperparameter settings of the best performing random forest model, we generated 100 global prediction images. These were subsequently used to calculate a bootstrap coefficient of variation (derived by dividing the standard deviation by the mean) and 95% confidence intervals for each pixel (Extended Data Figs. 7 and 8). To locate environmental conditions and corresponding geographical regions that are under-represented in the training data, we first converted the data into the principal component space. We then selected the first 13 and 14 axes that cumulatively accounted for 90% of the total variance for the models for AM and EcM, respectively. We evaluated whether raster pixels were associated with environmental conditions internal or external to the convex hull circumscribing the data points along each pairwise combination of principal component axes. We defined the degree of model extrapolation as the proportion of total pairwise combinations for which a pixel lies outside the corresponding convex hull. Finally, we combined this environmental extrapolation data with a map of geographical distance between sampling locations (distance and environmental extrapolation maps were combined in a 2:1 ratio, after scaling each from 0 to 1) to produce a global spatial assessment of the representativeness of our datasets (Extended Data Figs. 7 and 8). The extrapolation map was then used to mark pixels that were most under-represented by the training data using a 5% extrapolation cut-off value (that is, areas with <95% coverage in training data space).

To explore possible spatial dependency in the data, we fit semi-variogram models per mycorrhizal type on richness estimates and model residuals using the R package automap (v.1.1-9)⁸⁷. For AM fungi,

we observed autocorrelation in the extrapolated richness data up to about 900 km; for EcM fungi this range reached about 550 km. This level of spatial autocorrelation in response variables is unsurprising given the scale of our georeferenced datasets, similar to a recent analysis showing that spatial weights of around 800 km were best at handling spatial autocorrelation in a global plant diversity dataset⁷⁷. For model residuals, semi-variograms showed weak, although significant, spatial dependency for both AM and EcM fungal data (Supplementary Fig. 2). We then performed a Global Moran's *I*-test with permutations calculated using the R package sfdep⁸⁸, including five nearest points as neighbours and a Gaussian kernel function for building spatial weight matrices. For the richness estimates, spatial autocorrelation was positive and highly significant (Moran's *I* of 0.71 for AM and 0.64 for EcM), whereas spatial autocorrelation was low in model residuals for both fungal types (Moran's *I* of 0.11 for AM and 0.23 for EcM; Supplementary Fig. 1), similar to a previous study that reported limited effects of geographical distance on fungal alpha diversity³⁹. To incorporate spatial processes into the prediction model, we applied spatial eigenvector mapping using distance-based Moran's eigenvector maps (dbMEMs)⁸⁹. We calculated dbMEMs using the R package adespatial (v.0.3-21)⁹⁰, with the custom function quickMEM⁹¹. The inclusion of dbMEMs in the modelling approach showed mycorrhizal hotspots predicted in the same areas but reduced some finer-scale variation in richness predictions in the non-spatial models (Supplementary Fig. 10). These results led us to conclude that the models without dbMEMs capture the most variation; however, there might be some fine-scale spatial processes that are not explained for AM fungi here.

We evaluated how uneven sampling density might change rarity patterns by comparing model predictions that used current sampling intensity with those that used a simulated globally homogenous and high-sampling intensity, as previously described³². We created 1-km² resolution sampling intensity layers for the SSU and ITS training data-sets using kernel density interpolation from sample coordinates with a 5° radius and uniform decay rate in QGIS (v.3.40). These layers were rescaled 0–1 and included as covariates in the rarity random forest models to make empirical predictions of AM and EcM fungal rarity on the basis of the current distribution of sampling efforts. We then extrapolated rarity models under a global and equally high-sampling scenario by setting all pixels in the sampling intensity covariate layer to the maximum value of 1. Comparing patterns from the empirical and high-sampling rarity models showed where future sampling may reveal overlooked and under-realized hotspots of mycorrhizal fungal rarity³². For all analyses (for example, latitude trends and hotspot protections), we used the high-sampling rarity models to avoid biases in endemism predictions that were based on sampling intensity.

To measure how mycorrhizal richness and rarity varies with latitude, predicted richness and rarity pixel values were averaged at every 0.2 latitude degrees within 1°-wide longitude bands⁹² using functions in the R package Tidyverse (v.2.0.0)⁹³. Correlations between latitude and estimated mycorrhizal richness and rarity of samples are also provided in Extended Data Fig. 2. Mycorrhizal hotspots were defined by setting a cut-off at the 95th percentile of predicted richness values. These hotspot pixels were then overlaid with the World Database of Protected Areas cropped to spatial predictions here (WDPA; www.protectedplanet.net)³⁵ before calculating the total extent to which mycorrhizal hotspots in different biomes fall within current protected area delineations. Hotspot overlap with protected areas was calculated for each IUCN management category (terrestrial WDPA categories I–VI and unassigned). To understand mycorrhizal hotspot protections relative to other organisms, we compiled recently published spatial data on the species diversity of vascular plants²⁵, trees⁹⁴, ants³² and four groups of vertebrates (amphibians, birds, mammals and reptiles; IUCN⁹⁵). The same hotspot analysis described above was applied to these spatial layers, and the total amount of richness and rarity centre overlap in protected areas was calculated for

comparisons. All spatial analyses were based on raster data in Equal Earth projection at 1-km² pixel resolution, and maps were visualized in Robinson projection.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Data are archived on Zenodo (<https://doi.org/10.5281/zenodo.14871588>)³⁷. Full map predictions (1-km² global GeoTIFFs) are available after submitting a data request (<https://spun.earth/data-request>). Links to spatial data layers used for modelling are listed in Supplementary Table 2. The FungalTraits database (v.1.2; <https://doi.org/10.1007/s13225-020-00466-2>) and the MaarjAM database (v.5.6.2019; <https://maarjam.ut.ee/>) were used for fungal bioinformatic analyses. Fungal data used in this study are available from previous publications^{6–8,60}.

Code availability

Code is available from GitHub (https://github.com/SocietyProtectionUndergroundNetworks/richness_maps/) and permanently archived in the public Zenodo repository (<https://doi.org/10.5281/zenodo.14871588>)³⁷.

56. Tedersoo, L. & Lindahl, B. Fungal identification biases in microbiome projects. *Environ. Microbiol. Rep.* **8**, 774–779 (2016).
57. Yang, R. H. et al. Evaluation of the ribosomal DNA internal transcribed spacer (ITS), specifically ITS1 and ITS2, for the analysis of fungal diversity by deep sequencing. *PLoS ONE* **13**, 206428 (2018).
58. Bengtsson-Palme, J. et al. Improved software detection and extraction of ITS1 and ITS 2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods Ecol. Evol.* **4**, 914–919 (2013).
59. Edgar, R. C. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996–998 (2013).
60. Pöhlme, S. et al. FungalTraits: a user-friendly traits database of fungi and fungus-like stramenopiles. *Fungal Divers.* **105**, 1–16 (2020).
61. Öpik, M. et al. The online database MaarjAM reveals global and ecosystemic distribution patterns in arbuscular mycorrhizal fungi (Glomeromycota). *New Phytol.* **188**, 223–241 (2010).
62. Bruns, T. D. & Taylor, J. W. Comment on “Global assessment of arbuscular mycorrhizal fungus diversity reveals very low endemism”. *Science* **351**, 826–826 (2016).
63. Chao, A. et al. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecol. Monogr.* **84**, 45–67 (2014).
64. Hsieh, T. C., Ma, K. H. & Chao, A. iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods Ecol. Evol.* **7**, 1451–1456 (2016).
65. Bissett, A. et al. Introducing BASE: the Biomes of Australian Soil Environments soil microbial diversity database. *GigaScience* **5**, s13742–016 (2016).
66. Yan, D. et al. High-throughput eDNA monitoring of fungi to track functional recovery in ecological restoration. *Biol. Conserv.* **217**, 113–120 (2018).
67. Usher, M. B. in *Wildlife Conservation Evaluation* (ed. Usher, M. B.) 3–44 (Chapman & Hall, 1986).
68. Albuquerque, F. & Beier, P. Predicted rarity-weighted richness, a new tool to prioritize sites for species representation. *Ecol. Evol.* **6**, 8107–8114 (2016).
69. van den Hoogen, J. et al. Soil nematode abundance and functional group composition at a global scale. *Nature* **572**, 194–198 (2019).
70. van den Hoogen, J. et al. A global database of soil nematode abundance and functional group composition. *Sci. Data* **7**, 103 (2020).
71. Hengl, T. et al. SoilGrids250m: global gridded soil information based on machine learning. *PLoS ONE* **12**, e0169748 (2017).
72. Karger, D. N. et al. Climatologies at high resolution for the earth’s land surface areas. *Sci. Data* **4**, 170122 (2017).
73. Herrando-Moraira, S. et al. Climate Stability Index maps, a global high resolution cartography of climate stability from Pliocene to 2100. *Sci. Data* **9**, 48 (2022).
74. Ruesch, A. & Gibbs, H. K. New IPCC Tier-1 global biomass carbon map for the year 2000. *ESS-DIVE* <https://doi.org/10.15485/1463800> (2008).
75. Tuuvalu, M. N. & Jetz, W. A global 1-km consensus land-cover product for biodiversity and ecosystem modelling. *Global Ecol. Biogeogr.* **23**, 1031–1045 (2014).
76. Tuuvalu, M. N. & Jetz, W. A global, remote sensing-based characterization of terrestrial habitat heterogeneity for biodiversity and ecosystem modelling. *Global Ecol. Biogeogr.* **24**, 1329–1339 (2015).
77. Cai, L. et al. Global models and predictions of plant diversity based on advanced machine learning techniques. *New Phytol.* **237**, 1432–1445 (2023).
78. Trabucco, A. & Zomer, R. J. Global aridity index and potential evapo-transpiration (ETO) climate database v2. *figshare* <https://doi.org/10.6084/m9.figshare.7504448.v3> (2019).
79. Dinerstein, E. et al. An ecoregion-based approach to protecting half the terrestrial realm. *BioScience* **67**, 534–545 (2017).

Article

80. Ploton, P. et al. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nat. Commun.* **11**, 4540 (2020).
81. Wadoux, A. M. C., Heuvelink, G. B., De Bruin, S. & Brus, D. J. Spatial cross-validation is not the right way to evaluate map accuracy. *Ecol. Model.* **457**, 109692 (2021).
82. Meyer, H. & Pebesma, E. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods Ecol. Evol.* **12**, 1620–1633 (2021).
83. Milà, C., Mateu, J., Pebesma, E. & Meyer, H. Nearest neighbour distance matching leave-one-out cross-validation for map validation. *Methods Ecol. Evol.* **13**, 1304–1316 (2022).
84. Linnenbrink, J., Milà, C., Ludwig, M. & Meyer, H. KNNDM CV: k-fold nearest-neighbour distance matching cross-validation for map accuracy estimation. *Geosci. Model Dev.* **17**, 5897–5912 (2024).
85. Phillips, H. R. et al. Global distribution of earthworm diversity. *Science* **366**, 480–485 (2019).
86. Potapov, A. M. et al. Globally invariant metabolism but density–diversity mismatch in springtails. *Nat. Commun.* **14**, 674 (2023).
87. Hiemstra, P. H., Pebesma, E. J., Twenhöfel, C. J. & Heuvelink, G. B. Real-time automatic interpolation of ambient gamma dose rates from the Dutch radioactivity monitoring network. *Comput. Geosci.* **35**, 1711–1721 (2009).
88. Parry J. sfdep: Spatial Dependence for Simple Features. R package version 0.2.3 <https://CRAN.R-project.org/package=sfdep> (2023).
89. Dormann, C. F. et al. Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography* **30**, 609–628 (2007).
90. Dray S. et al. adespatial: Multivariate Multiscale Spatial Analysis. R package version 0.3-21 <https://CRAN.R-project.org/package=adespatial> (2023).
91. Borcard, D., Gillet, F. & Legendre, P. *Numerical Ecology with R* (Springer, 2018).
92. Zhu, K., Woodall, C. W. & Clark, J. S. Failure to migrate: lack of tree range expansion in response to climate change. *Glob. Change Biol.* **18**, 1042–1052 (2012).
93. Wickham, H. et al. Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
94. Liang, J. et al. Co-limitation towards lower latitudes shapes global forest diversity gradients. *Nat. Ecol. Evol.* **6**, 1423–1437 (2022).
95. *The IUCN Red List of Threatened Species Version 2022-2* (IUCN, 2022); <https://www.iucnredlist.org>.
- supported by grants from DOB Ecology and the Bernina Foundation. T.V. and P.K. acknowledge support from the Czech Science Foundation (21-17749S). P.B. acknowledges support from the MEYS (LC23152, LM2023055). L.T. acknowledges support from the Estonian Science Foundation (grant PRG632). K.G.P. is a fellow of the CIFAR program Fungal Kingdom: Threats and Opportunities, and acknowledges support from the US NSF (DEB-1845544) and DOE (DE-SC0023661). C.M. thanks the ANID-Chile projects SIA no. SA77210019 (2021) and Fondecyt Regular Project no. 1240186 (2024).

Author contributions J.v.d.H., C.A. and M.E.V.N. performed the initial processing and modelling of fungal datasets provided by T.V., P.K. and P.B. M.E.V.N., C.A., J.D.S., O.P., A.C., L.G.v.G. and E.T.K. provided feedback on the modelling process and interpreted results. C.Q. led the calculations of rarity-weighted richness. M.E.V.N. performed technical validation and additional spatial analyses. M.E.V.N. wrote the first draft of the manuscript, and C.A., J.D.S., O.P., A.C., L.G.v.G., B.F.M., C.Q., T.L., V.M., O.D., G.F., C.M., M.S., J.W., K.G.P., C.K.C., T.V., P.K., P.B., L.T., S.A.W., T.W.C., E.T.K. and J.v.d.H. provided comments and significant edits to the manuscript. Authors listed in the SPUN mapping consortium are included on the basis of their involvement with a co-led SPUN fungal sampling trip. In addition to providing feedback on the manuscript, these authors contributed in the following ways: (1) helped establish a global network of mycorrhizal researchers; (2) refined our sampling protocols and eDNA processing techniques; and (3) made substantial contributions to the conception and design of our future research plans involving mycorrhizal fungal data and modelling.

Competing interests E.T.K. and C.A. are the founders of the Society for the Protection of Underground Networks (SPUN), a non-governmental organization (NGO) that conducts research on mycorrhizal fungi for conservation and restoration. T.W.C. is the founder of Restor, an NGO that facilitates the global restoration movement. C.A. is the founder of Funga, an organization that facilitates the restoration of belowground fungal biodiversity. G.F. is the founding director of Fungi Foundation, an NGO that explores and educates about fungal diversity for applications and conservation. M.E.V.N., J.D.S., O.P., A.C., L.G.v.G., B.F.M., C.Q., T.L., V.M., O.D., C.M., M.S., J.W., K.G.P., C.K.C., T.V., P.K., P.B., L.T., S.A.W. and J.v.d.H. declare no competing interests.

Additional information

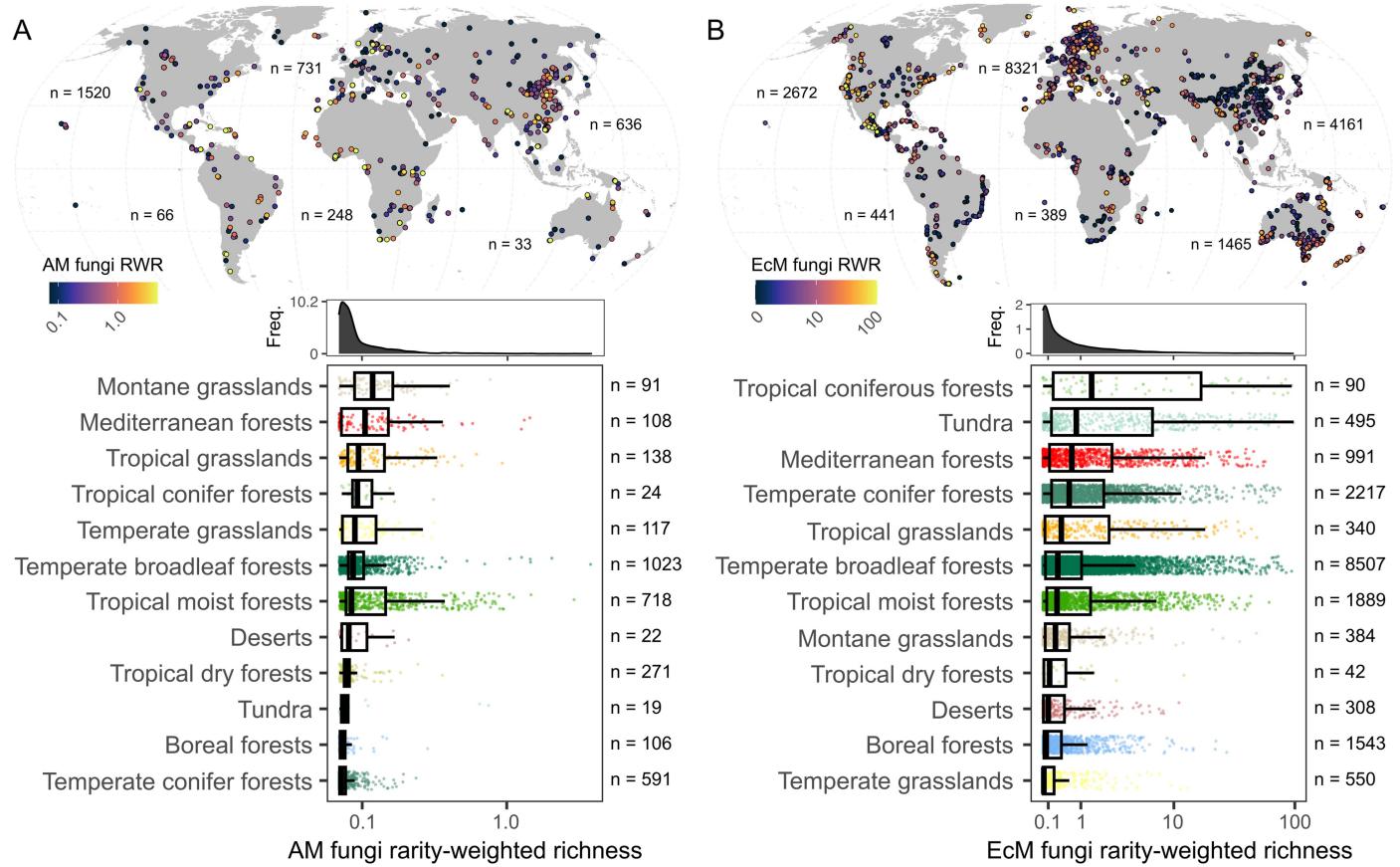
Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-025-09277-4>.

Correspondence and requests for materials should be addressed to Michael E. Van Nuland.

Peer review information *Nature* thanks the anonymous reviewers for their contribution to the peer review of this work.

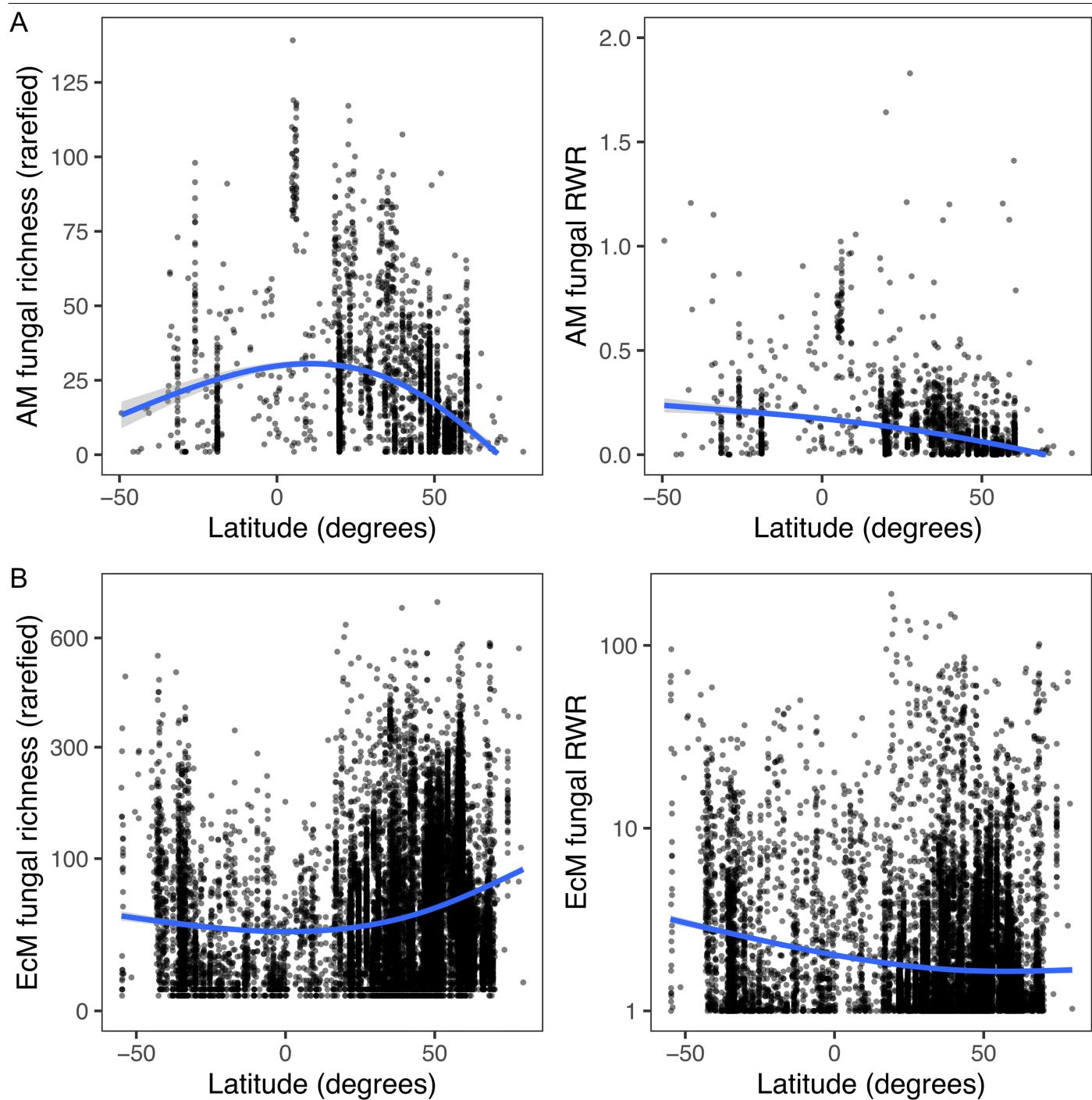
Reprints and permissions information is available at <http://www.nature.com/reprints>.

Acknowledgements SPUN acknowledges the support of the Jeremy and Hannelore Grantham Environmental Trust, the Paul Allen Family Foundation, the Schmidt Family Foundation, the Quadrature Climate Foundation and the Bezos Earth Fund. E.T.K. acknowledges support from NWO-VICI (202.012), NWO-Spinova (SPI.2023.2) and HFSP (RGP 0029). C.A. was supported by Ambizione grant no. PZOOP3_17990 from the Swiss National Science Foundation. J.D.S. was supported by the NWO-Gravity Grant Microp (024.004.014). T.W.C., J.v.d.H. and L.G.v.G. were



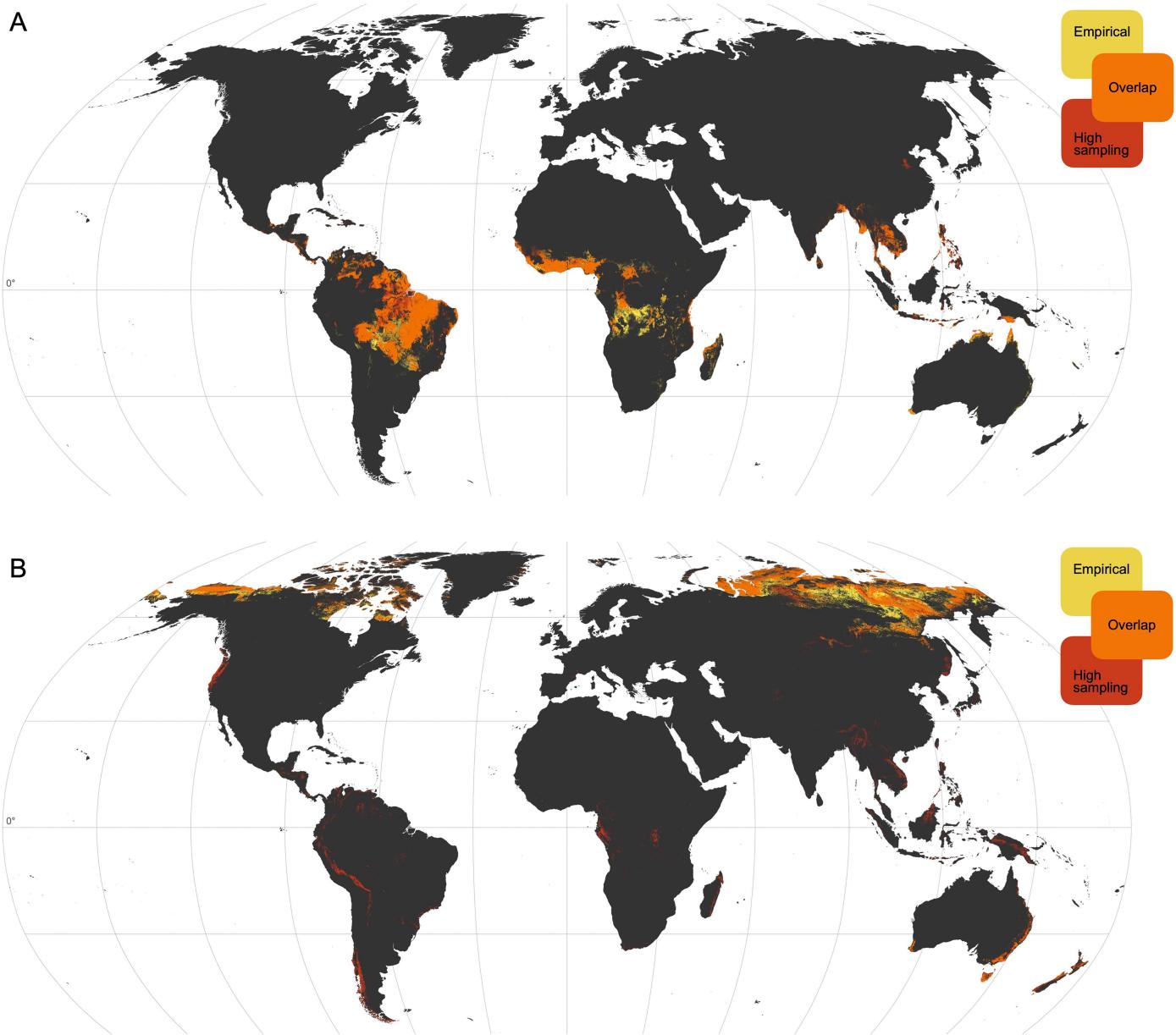
Extended Data Fig. 1 | Sample locations and mycorrhizal rarity trends by biome. Distribution of sites and rarity-weighted richness (RWR) estimates for **A**) arbuscular mycorrhizal (AM) fungi and **B**) ectomycorrhizal (EcM) fungi. Mycorrhizal rarity patterns across terrestrial biomes are shown in boxplots (Mangroves, Flooded grasslands, and Rock/Ice biomes not shown due to few samples in both datasets). Global distributions of relative rarity are shown in

density curves above biome-level boxplots. Boxplots indicate the median (center line), first and third quartiles (lower and upper box edges), and $1.5 \times \text{IQR}$ (box whiskers). Relative rarity data is unitless and shown as log-transformed for visualization. See Main Text for details on RWR calculation. Points shown here are all samples that passed quality control checks and used as training data for richness models.



Extended Data Fig. 2 | Observed mycorrhizal richness and rarity variation with latitude. Plots show A) AM fungi and B) EcM fungi. Points are the rarefied mycorrhizal richness values (estimated from samples) or rarity-weighted

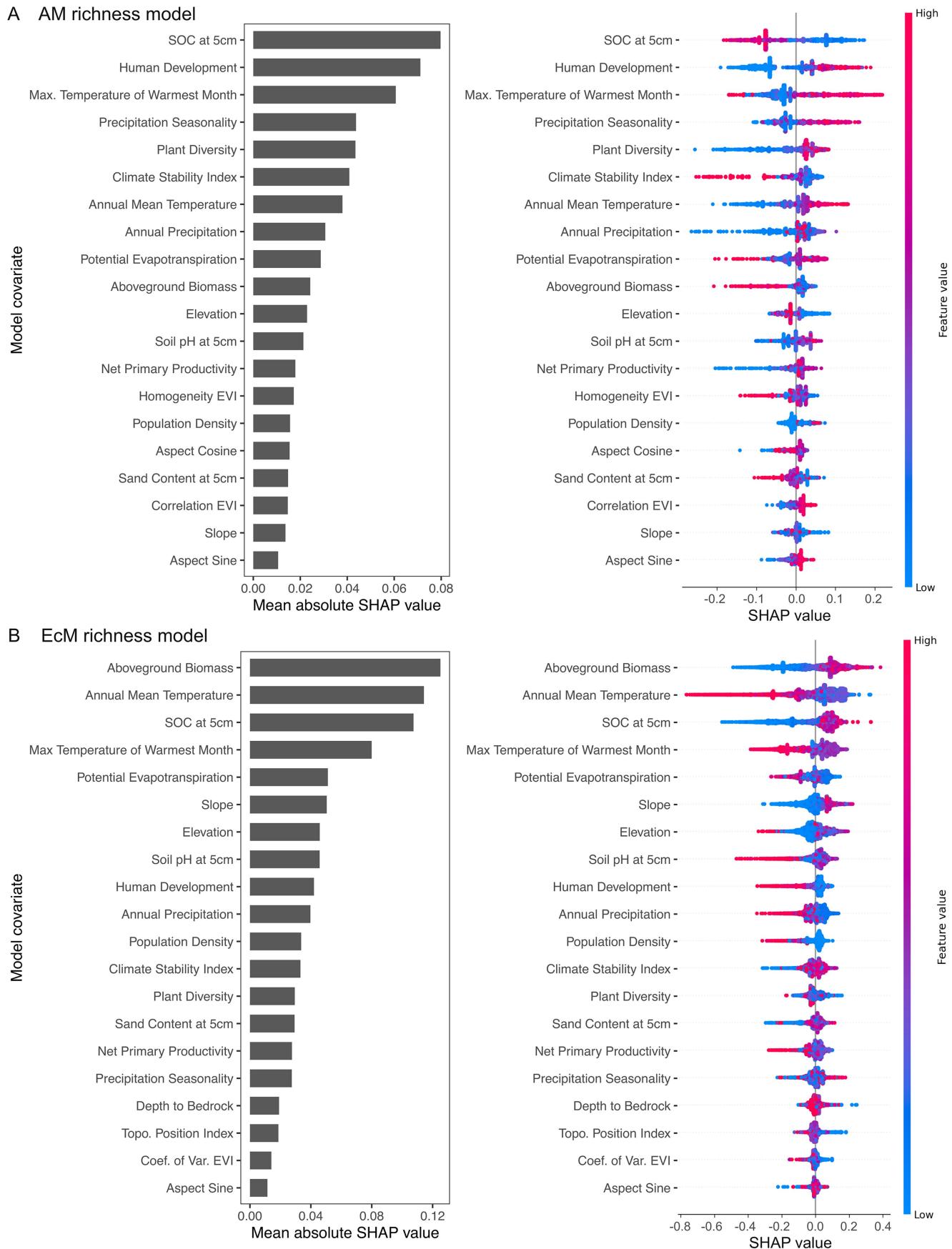
richness in relation to the latitude position where samples were collected. Curves show model fit based on a quadratic linear regression (second order polynomial).



Extended Data Fig. 3 | Spatial predictions of how increased sampling could change the global distribution of mycorrhizal fungal rarity hotspots. We built machine learning models of **A**) AM and **B**) EcM fungal rarity under two sampling scenarios. We first created empirical predictions (yellow) based on sample rarity values and the current geographic distribution of sample density (i.e., more concentrated sampling in North America, Europe, and Asia). We then built a second predictive model that simulated ‘high-sampling’ efforts (red) by

setting the sampling density covariate layer to the global max value in all pixels. Mapping predicted rarity hotspots under these different scenarios shows areas that may fall out of the top 5% of global endemism centers under increased sampling (yellow), hotspot predictions that are robust to future sampling efforts (orange), and places where future sampling is likely to reveal undiscovered rarity hotspots (red).

Article



Extended Data Fig. 4 | See next page for caption.

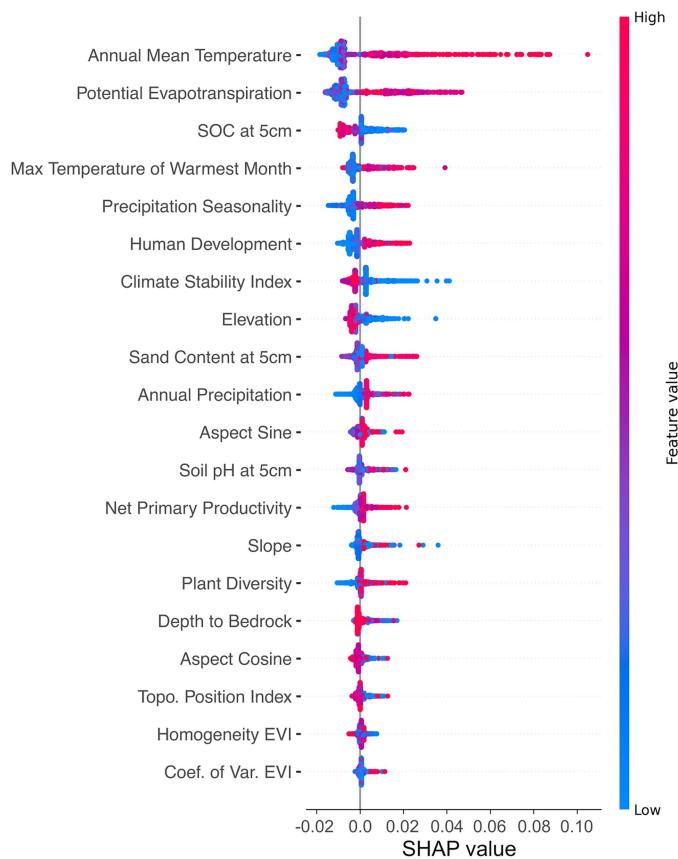
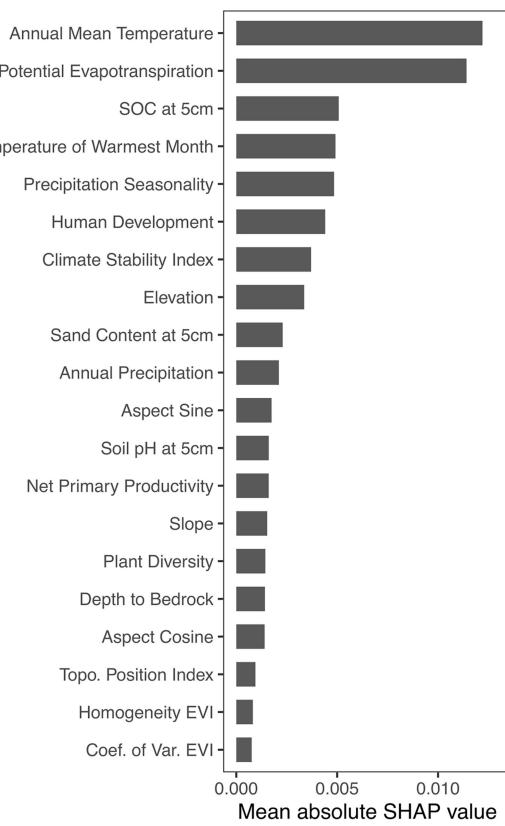
Extended Data Fig. 4 | SHapley Additive exPlanations (SHAP) value plots showing the magnitude and direction of relationships between environmental predictors and mycorrhizal richness. For **A**) AM fungi and **B**) EcM fungi, variables at the top of the graph are the most important predictors with decreasing importance down the figure. Bar graphs (left) show mean absolute SHAP values that measure the overall magnitude of each covariate

on model predictions in units of fungal richness. Beeswarm plots (right) show the directionality of SHAP values, with points left of the zero-line indicating a negative relationship between mycorrhizal richness and a given predictor. The red/blue color gradient represents fungal richness values associated with a predictor.

Article

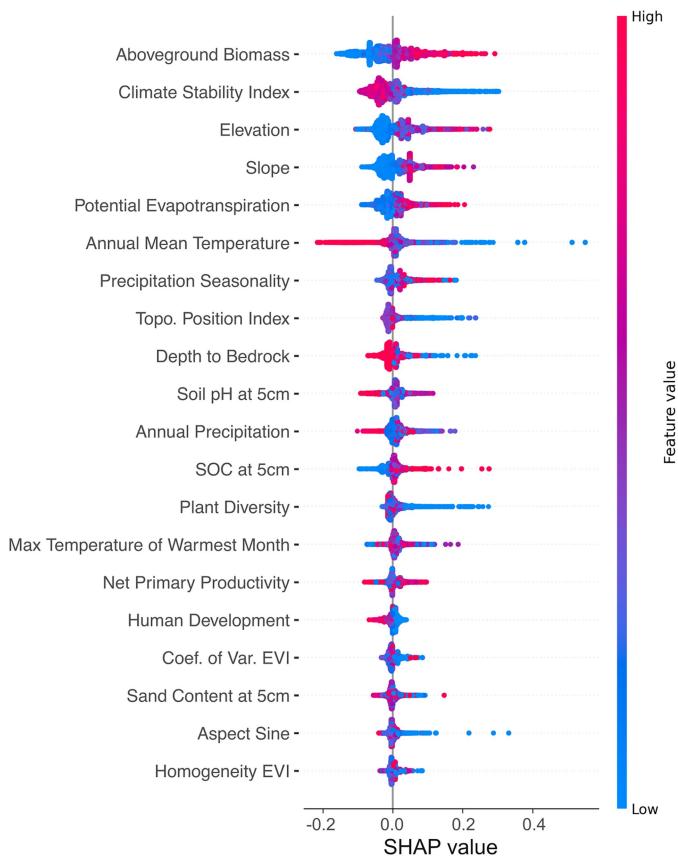
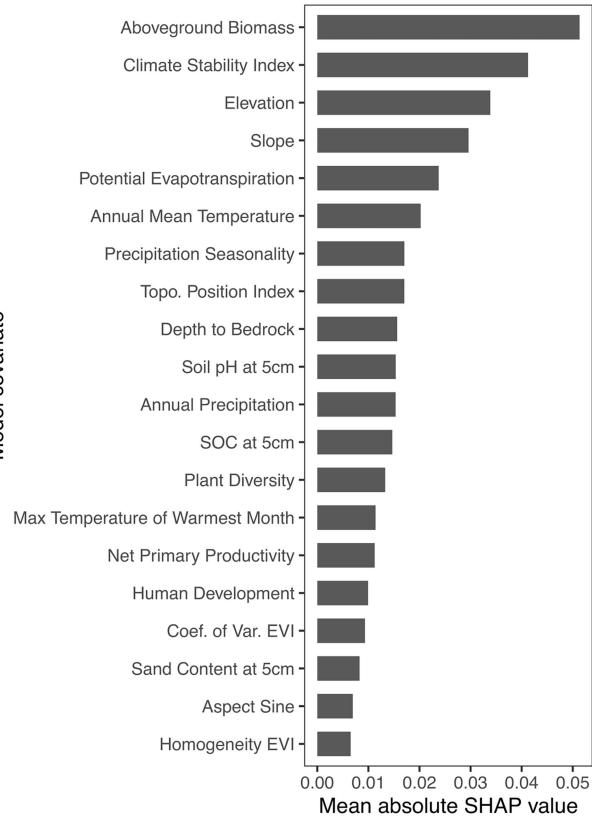
A AM rarity-weighted richness model

Model covariate



B EcM rarity-weighted richness model

Model covariate

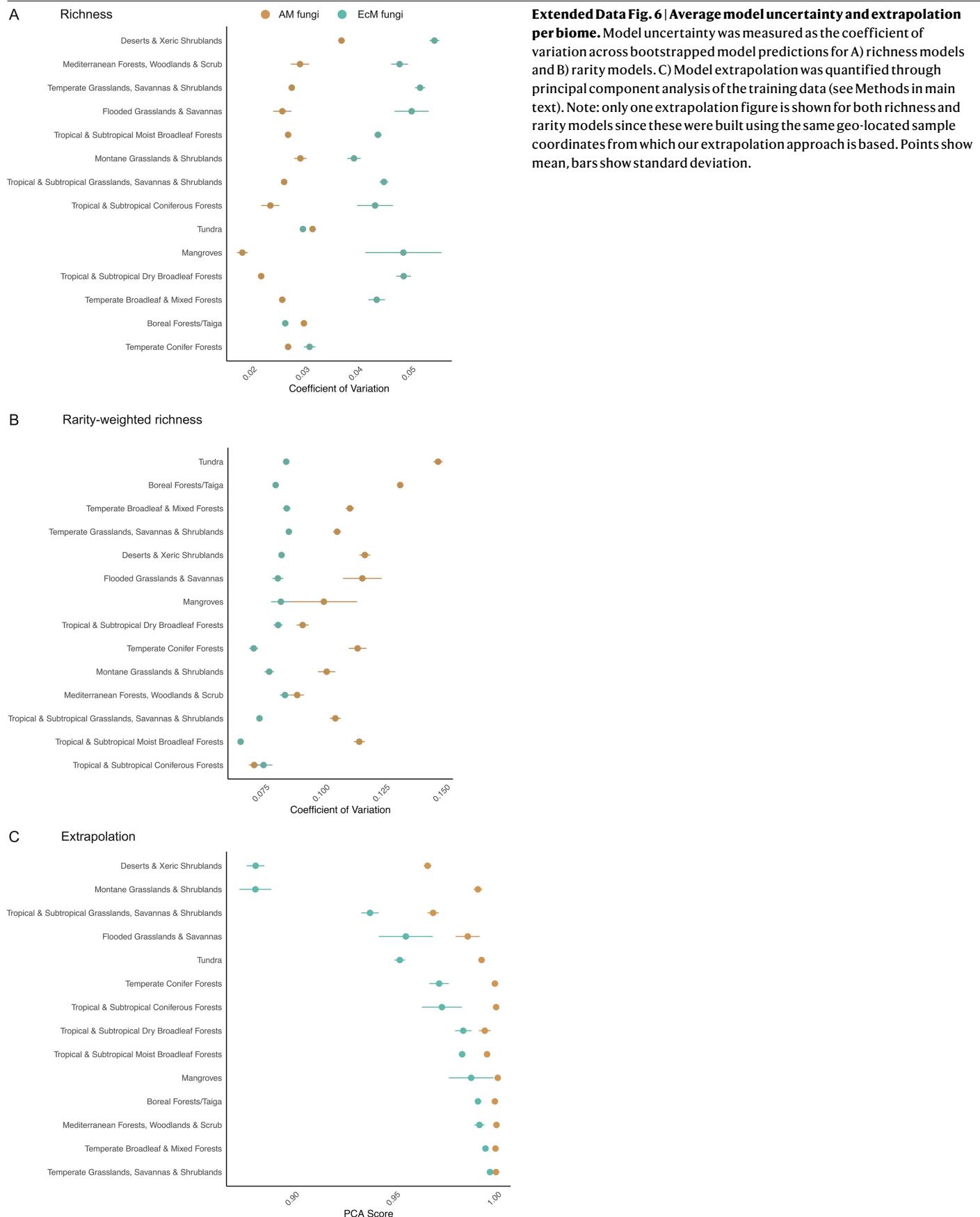


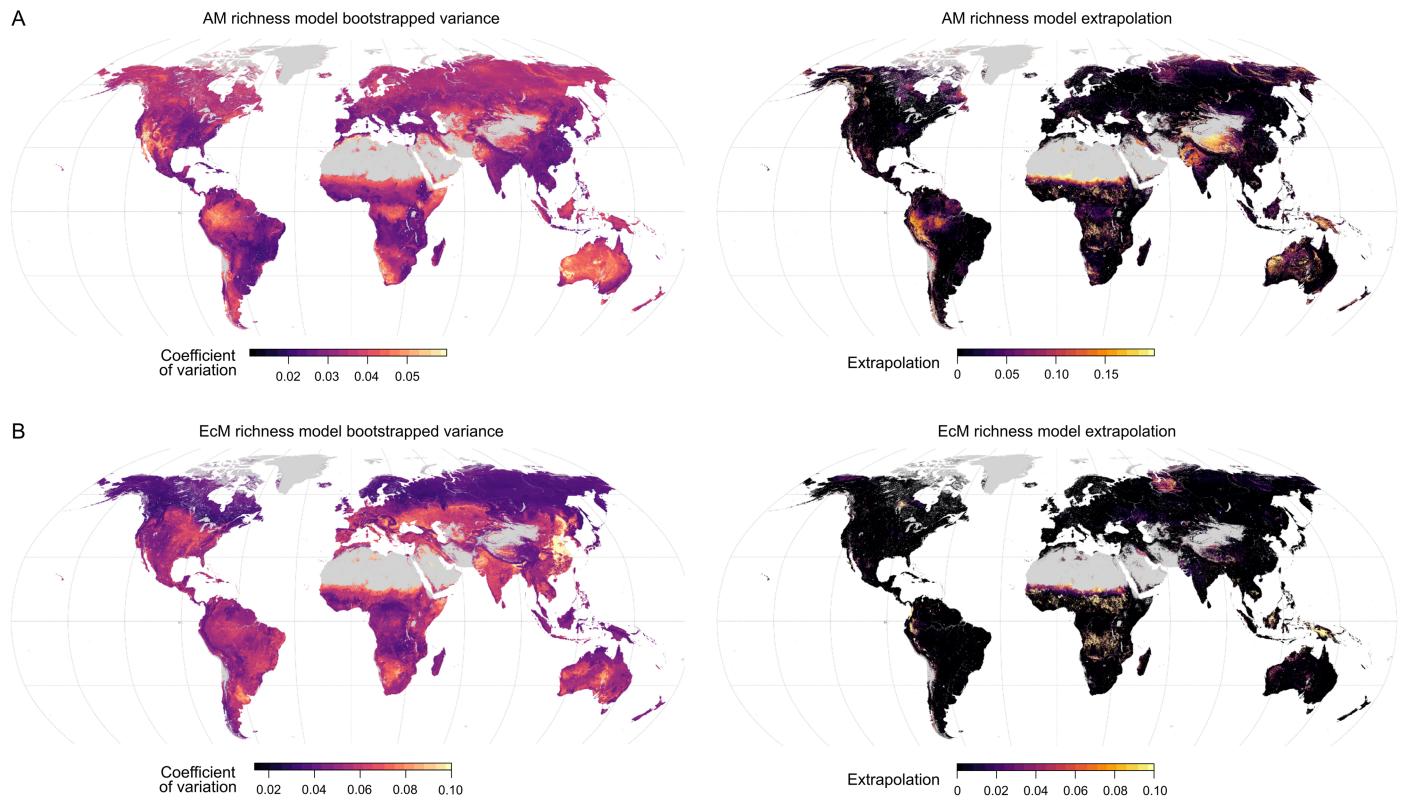
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | SHapley Additive exPlanations (SHAP) value plots showing the magnitude and direction of relationships between environmental predictors and mycorrhizal rarity. For A) AM fungi and B) EcM fungi, variables at the top of the graph are the most important predictors with decreasing importance down the figure. Bar graphs (left) show mean absolute SHAP values that measure the overall magnitude of each covariate on

model predictions in units of fungal rarity-weighted richness. Beeswarm plots (right) show the directionality of SHAP values, with points left of the zero-line indicating a negative relationship between mycorrhizal rarity and a given predictor. The red/blue color gradient represents fungal richness values associated with a predictor.

Article

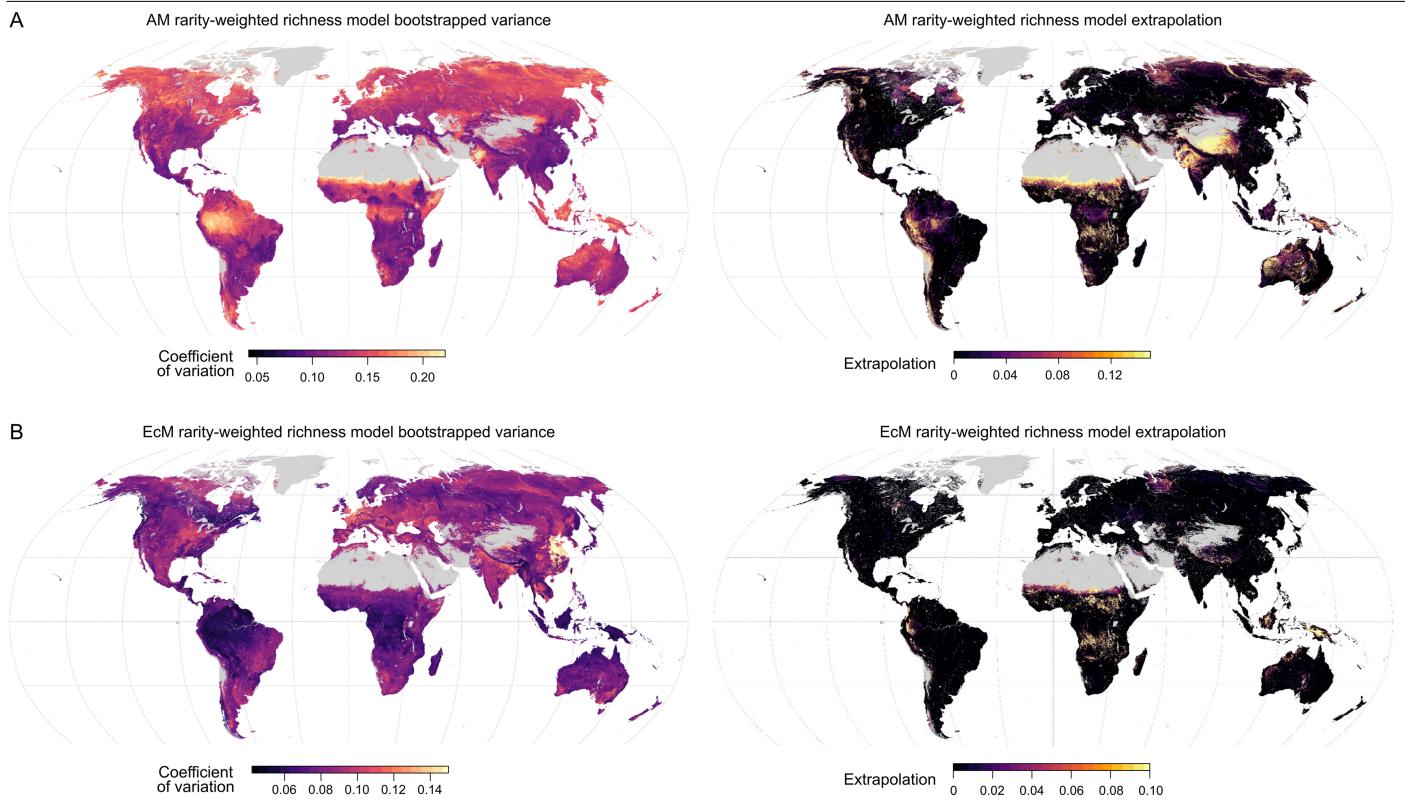




Extended Data Fig. 7 | Maps showing geographic distribution of model uncertainty and extrapolation for mycorrhizal fungal richness models. Model uncertainty and extrapolation shown for A) AM fungi richness and B) EcM fungi richness predictions. Brighter colors indicate areas with relatively

higher coefficient of variation across bootstrapped predictions (which we define as model uncertainty), or higher extrapolation due to poor environmental and spatial representation in the richness training data.

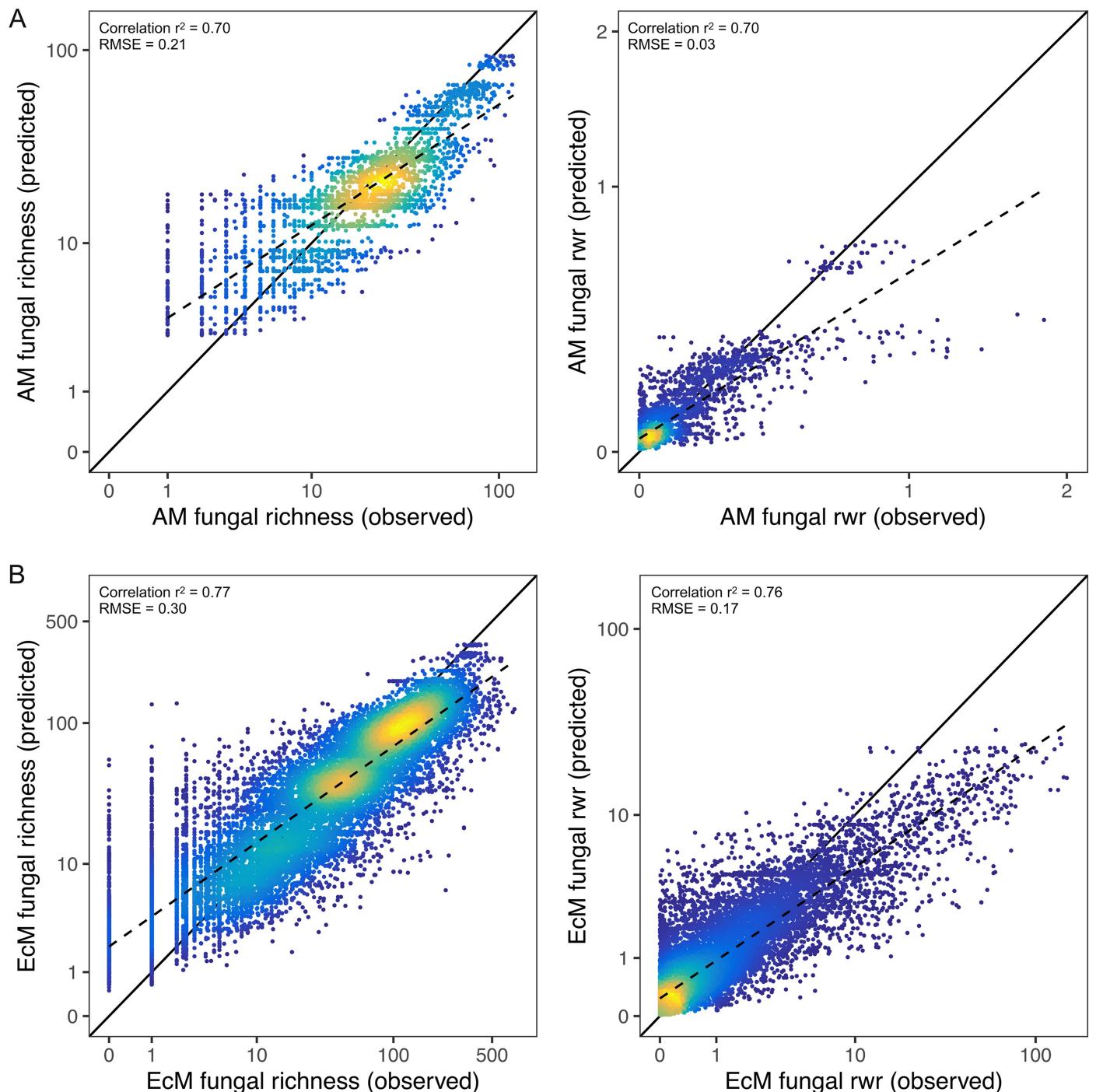
Article



Extended Data Fig. 8 | Maps showing geographic distribution of model uncertainty and extrapolation for mycorrhizal fungal endemism models.

Model uncertainty and extrapolation shown for A) AM fungi rarity-weighted richness and B) EcM fungi rarity-weighted richness predictions. Brighter colors

indicate areas with relatively higher coefficient of variation across bootstrapped predictions (which we define as model uncertainty), or higher extrapolation due to poor environmental and spatial representation in the rarity-weighted richness training data.



Extended Data Fig. 9 | Predicted versus observed plots of mycorrhizal richness and rarity-weighted richness (rwr) showing the model fit to the training data. For **A**) AM fungi and **B**) EcM fungi, the Xaxis represents observed data used to train the machine learning model and the Yaxis shows predicted values from the model. Colors indicate the density points in the

dataset. Solid lines show a 1:1 relationship between predicted versus observed values, indicating a perfect fit and 100% predictive accuracy. The dashed line shows the actual fit between the observed and predicted values. Axes scales are log-transformed.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
 - Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
 - Give P values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	No software was used for data collection.
Data analysis	<p>Coding scripts used for this manuscript are available here (Zenodo DOI:10.5281/zenodo.14871588) and here (Github (https://github.com/SocietyProtectionUndergroundNetworks/richness_maps/)).</p> <p>Software packages used for fungal bioinformatics: ITSx v1.1.2; USEARCH v11.0.0667; BLASTn search against UNITE version 8.3</p> <p>Python (v 3.12.2) packages for data processing and analysis: pandas; geopandas; numpy; subprocess; time; datetime; ee; shap; fasttreeshap; matplotlib; pathlib; multiprocessing; functools; scipy; sklearn; iterools</p> <p>R (v4.3.2) packages for data processing and analysis: data.table v.1.15.0; Tidyverse v 2.0.0; janitor v2.2.1; doParallel v1.0.17; caret v7.0.1; iNEXT v3.0.0; Terra v1.7-71; Raster v3.6-26; sf v1.0-20; automap v1.1-9; sfdep v0.2.4; adespatial v0.3-21;</p> <p>Spatial analysis software: QGIS (v3.40).</p>

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Data is archived on Zenodo (DOI:10.5281/zenodo.14871588). Full map predictions (1-km² global GeoTIFFs) are available after submitting a data request here: <https://spun.earth/data-request>. The FungalTraits database v.1.2 (<https://doi.org/10.1007/s13225-020-00466-2>) and MaarjAM database from 5.6.2019 (<https://maarjam.ut.ee/>) were used for fungal bioinformatic analysis. Links to spatial data layers used for modeling are listed in Supplementary Table 2.

Fungal databases used in this study is also available from the following previous publications:

GlobalFungi Database - <https://doi.org/10.1038/s41597-020-0567-7>;

GlobalAMFungi Database - <https://doi.org/10.1111/nph.19283>;

Global Soil Mycobiome Consortium Database - <https://doi.org/10.1007/s13225-021-00493-7>

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

NA

Reporting on race, ethnicity, or other socially relevant groupings

NA

Population characteristics

NA

Recruitment

NA

Ethics oversight

NA

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences

Behavioural & social sciences

Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

We combined data from three previously published fungal databases (refs. 6-8) to map mycorrhizal fungal diversity across Earth's terrestrial ecosystems. In total, we calculated fungal richness estimates for 25,000 geolocated soil samples comprising >2.8 billion fungal sequences. We used this data to train machine-learning models with 22 geospatial environmental layers for creating high-resolution (1km²) global spatial predictions of mycorrhizal richness. These spatial predictions were then further analyzed to assess the protected status of fungal biodiversity hotspots.

Research sample

Fungal occurrence records were generated from data-mining of published ITS and SSU soil sequencing studies collected in the GlobalFungi and GlobalAMFungi databases (see refs. 6 and 8 for details), as well as the ITS region from full-length sequences in the Global Soil Mycobiome consortium database (ref. 7).

Sampling strategy

See refs. 6-8 for details on sample processing and database development.

Data collection

NA

Timing and spatial scale

Samples are from a globally distributed set of published studies occurring from 2011-2022.

Data exclusions

Data outliers were identified and removed prior to modeling. Two Australian studies with ITS samples from Desert and Mediterranean biomes had unusually high EcM OTU richness estimates and standard deviations (two orders of magnitude higher than other Australian ecoregions; refs. 58, 59), and have been previously marked as potentially inaccurate based on a recent

database comparison⁶. All samples from these two studies were removed ($N=2,070$). We then filtered samples by biome by removing estimated richness values that were more than five times the interquartile range higher than the biome-level median estimate. We only filtered values at the highest end of the distribution to avoid potentially removing ‘true’ estimates of low or zero mycorrhizal richness at a given location. See Supplemental Table 1 for a summary of outliers removed per biome. We also filtered samples from project-specific variables that failed technical validation by comparing observed vs. predicted richness trends (failures defined by zero correlation). This process removed all DNBSEQ-G400 samples ($N=3,738$), all ITS1 samples ($N=33$; these samples also had corresponding ITS2 sequence data which were retained), and 14 different primer sets (total $N=809$), which were filtered prior to running the final models.

Reproducibility

All data and code to reproduce the global interpolations, map visualizations, and all spatial analyses are provided in the Zenodo archive (DOI:10.5281/zenodo.14871588).

Randomization

Randomization was not relevant to this study, samples retained their unique ID information for processing and analysis.

Blinding

Blinding was not relevant to this study, samples needed to retain georeferenced information to be usable for spatial analysis/inference.

Did the study involve field work? Yes No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- | | |
|-------------------------------------|--|
| n/a | Involved in the study |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Plants |

Methods

- | | |
|-------------------------------------|---|
| n/a | Involved in the study |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

Plants

Seed stocks

NA

Novel plant genotypes

NA

Authentication

NA