

1 **Global Hotspots of Mycorrhizal Fungal Richness are Poorly Protected**

2

3 Michael E. Van Nuland^{1,*}, Colin Averill^{2,3}, Justin D. Stewart^{1,4}, Oleh Prylutskyi⁵, Adriana
4 Corrales¹, Laura G. van Galen^{1,2}, Bethan F. Manley¹, Clara Qin¹, Thomas Lauber², Vladimir
5 Mikryukov⁶, Olesia Dulia⁶, Giuliana Furci⁷, César Marín^{4,8}, Merlin Sheldrake^{1,4}, James Weedon⁹,
6 Kabir G. Peay^{10,11}, Charlie K. Cornwallis¹², Tomáš Větrovský¹³, Petr Kohout¹³, Petr Baldrian¹³,
7 Leho Tedersoo^{6,14}, Stuart A. West¹⁵, Thomas W. Crowther², E. Toby Kiers^{1,4}, SPUN Mapping
8 Consortium¹⁶, Johan van den Hoogen^{1,2}

9

10 ¹ Society for the Protection of Underground Networks (SPUN), Dover, DE, USA

11 ² Crowther Lab, Department of Environmental Systems Science, ETH Zürich, Zürich, Switzerland

12 ³ Funga Public Benefit Corporation, Austin TX USA

13 ⁴ Amsterdam Institute for Life and Environment (A-LIFE), Section Ecology & Evolution, Vrije
14 Universiteit Amsterdam, Amsterdam, the Netherlands

15 ⁵ Department of Mycology and Plant Resistance, V.N. Karazin Kharkiv National University,
16 Kharkiv, Ukraine

17 ⁶ Mycology and Microbiology Center, University of Tartu, Tartu, Estonia

18 ⁷ Fungi Foundation, Brooklyn, NY, USA

19 ⁸ Centro de Investigación e Innovación para el Cambio Climático (CiiCC), Universidad Santo
20 Tomás, Valdivia, Chile

21 ⁹ Amsterdam Institute for Life and Environment (A-LIFE), Section Systems Ecology, Vrije
22 Universiteit Amsterdam, Amsterdam, the Netherlands

23 ¹⁰ Department of Earth System Science, Stanford University, Stanford, CA, USA

24 ¹¹ Department of Biology, Stanford University, Stanford, CA, USA

25 ¹² Department of Biology, Lund University, Lund, Sweden

26 ¹³ Laboratory of Environmental Microbiology, Institute of Microbiology of the Czech Academy
27 of Sciences, Czech Republic

28 ¹⁴ College of Science, King Saud University, Riyadh, Saudi Arabia

29 ¹⁵ Department of Biology, Oxford University, Oxford, United Kingdom

30 ¹⁶ SPUN Mapping Consortium author details provided in Supplemental document

31 * Corresponding author (michael@spun.earth)

32 **Abstract**
33 Symbiotic mycorrhizal fungi help regulate the Earth's biogeochemical cycles and mediate
34 terrestrial vegetation patterns. However, the global distribution of mycorrhizal fungal biodiversity
35 is largely unknown. Unlike plants and animals, we do not know where mycorrhizal diversity
36 hotspots are located. This limits our ability to monitor and protect key underground ecosystems.
37 To resolve this, we trained machine learning algorithms on 25,000 geolocated soil samples
38 comprising >2.8 billion fungal DNA sequences. This allowed us to predict arbuscular mycorrhizal
39 and ectomycorrhizal fungal richness and rarity across Earth's terrestrial ecosystems. Incorporating
40 model uncertainty estimates, we identified regions with extraordinarily diverse and endemic
41 mycorrhizal communities. Arbuscular mycorrhizal fungal hotspots, with over 60 species per soil
42 sample, include Southeast Asian tropical forests, West African Guinean forests, and the Brazilian
43 Cerrado. Ectomycorrhizal fungal hotspots, with over 100 species per sample, include Siberian and
44 Canadian boreal forests, North American temperate conifer forests, and broadleaf forests in China.
45 Intersecting protected areas with mycorrhizal hotspots (95th percentile of diversity predictions)
46 shows that <10% of richness hotspots and <23% of rarity hotspots are currently protected. Our
47 results visualize a previously unidentified and hidden element of biodiversity, radically altering
48 how we view the Earth with the potential to transform international conservation priorities.

49

50

51 **Keywords:** Arbuscular mycorrhizal fungi, biodiversity hotspot, conservation, ectomycorrhizal
52 fungi, endemism, machine learning, protected area, soil, symbiosis, rarity

53 **Main**

54 Mycorrhizal fungi form nutritional symbioses with most plant species (>80%), building extensive
55 underground hyphal networks that can comprise over 30% of the living biomass of soils¹. The
56 fungi use their networks to forage in the soil for phosphorus, nitrogen, and trace elements, and
57 receive carbon from plant root systems in return. The diversity and activity of these fungal
58 communities is fundamental to the functioning of Earth's terrestrial ecosystems, with an estimated
59 3.6 billion tons of carbon allocated via plants to mycorrhizal fungi annually^{2,3}. Yet despite their
60 importance, our understanding of the global distribution of mycorrhizal fungal biodiversity lags
61 far behind plants and animals. As a result, it is difficult to develop appropriate measures to protect
62 the integrity of these ecosystems^{4,5}.

63

64 Over the last decade, the emergence of comprehensive, large-scale molecular datasets on
65 belowground fungi have unlocked our potential to establish global baselines of mycorrhizal fungal
66 biogeography⁶⁻⁸. This has revealed the dominance of two types of mycorrhizal symbionts spread
67 throughout Earth's terrestrial biomes: ectomycorrhizal (EcM) fungi and arbuscular mycorrhizal
68 (AM) fungi. These fungal types differ in their physiology, symbiotic behavior, and contribution to
69 carbon cycling^{9,10}. Although only ~2% of terrestrial plant species depend on EcM associations,
70 these host species comprise over 25% of global vegetation area, dominating in high-latitude
71 forests¹¹. AM fungi form associations with ~80% of plant species¹¹, dominating in tropical forests,
72 grasslands, and croplands globally. A third type—ericoid mycorrhizal (ErM) fungi—can be common
73 in high-elevation and high-latitude ecosystems but are much rarer globally¹¹. The scarcity of ErM
74 fungi in large-scale fungal datasets has left a sizeable gap in understanding their global diversity
75 patterns or conservation opportunities.

76

77 Previous work has used plant genus as a proxy of mycorrhizal type, revealing fundamental insights
78 into the distributions and functioning of mycorrhizal symbioses¹²⁻¹⁴. However, efforts to explore
79 mycorrhizal fungal biogeography from direct microbial observations, with methods such as DNA-
80 based surveys of soil samples¹⁵⁻¹⁹, have suggested that aboveground and belowground biodiversity
81 patterns are often uncorrelated²⁰. This means that patterns of plant communities may not accurately
82 reflect distributions of mycorrhizal diversity belowground. Yet, our ability to identify the scale of
83 this mismatch is limited by geographic bias in the locations of mycorrhizal samples for predictive

84 modeling, which results in extensive extrapolation into un-trained environmental space²¹. This is
85 particularly worrying in tropical regions where AM fungi are likely to dominate. Generating high-
86 resolution global distributions of mycorrhizal fungal diversity from direct observations—and
87 quantifying the extent of model uncertainty—is essential to guide future conservation planning²².
88 Protecting mycorrhizal fungal diversity will help avoid species loss across other taxonomic groups
89 and maintain critical ecosystem functions (e.g., carbon sequestration) for realizing nature-based
90 climate mitigation strategies^{4,5}.

91
92 To address these challenges, we built machine-learning spatial models of mycorrhizal diversity
93 from a globally distributed set of nearly 25,000 soil samples containing over 2.8 billion fungal
94 DNA sequences found in 130 countries compiled in the GlobalFungi, GlobalAMFungi, and Global
95 Soil Mycobiome consortium databases (Fig. 1)⁶⁻⁸. These are the largest datasets of fungal internal
96 transcribed spacers (ITS) and small subunit (SSU) rRNA amplicon sequences assembled to date,
97 which we analyzed using Virtual Taxa (VT) for AM fungi, and 97% similar operational taxonomic
98 units (OTU) for EcM fungi. With these data, we (i) created high-resolution (1-km²) spatial
99 predictions of mycorrhizal fungal richness and endemism (rarity-weighted richness) at the global
100 scale, (ii) identified mycorrhizal fungal diversity hotspots and assessed the extent of their current
101 environmental protections, and (iii) characterized spatial uncertainty and data limitations.

102
103 **Results & Discussion**

104
105 First, we explored the spread of the training data (Fig. 1). This consisted of AM and EcM fungal
106 richness estimates, calculated using a rarefaction/extrapolation approach, and rarity-weighted
107 richness (hereafter “rarity”; Supplemental Fig. 1), a useful metric of relative endemism to guide
108 conservation priorities²³. Across both AM and EcM datasets, we found temperate forests to be the
109 most sampled biomes (32-61% of all samples), and mangroves and flooded grasslands the least
110 sampled biomes (0.1% of all samples). There were only a few AM fungal samples from desert and
111 tundra biomes, and few EcM fungal samples from tropical dry forests and tropical conifer forests.
112 Compared to other biomes, we found that soil samples from montane grasslands, tropical conifer
113 forests, and temperate broadleaf forests show the highest AM fungal richness estimates (Fig. 1A).
114 Similarly, montane grasslands had the greatest AM fungal rarity estimates, followed by

115 Mediterranean forests and tropical grasslands with the next highest in AM fungal endemism levels
116 (Supplemental Fig. 1A). For EcM fungi, samples from coniferous forests (tropical and temperate)
117 and temperate broadleaf forests had the highest richness estimates (Fig. 1B). EcM fungal rarity
118 was also highest in tropical conifer forests, but tundra and Mediterranean forests showed a greater
119 concentration of rare EcM fungi than temperate forests (Supplemental Fig. 1B). We used these
120 datasets for training random forest machine learning models to predict global AM and EcM fungal
121 richness and endemism patterns.

122

123 Using maps of predicted richness, we asked whether mycorrhizal fungi show clear global diversity
124 patterns like plants and animals by plotting mean richness and rarity across latitude. We found that
125 predicted AM fungal richness was highest in ecosystems near the equator and gradually declined
126 towards the poles, in agreement with the classic latitude diversity gradient hypothesis observed
127 across many taxa (Fig. 2A)^{24,25}. In contrast, predicted EcM fungal richness was lowest near the
128 equator and showed more species-rich communities across northern latitudes and in southern
129 regions of South America and Australia, creating an inverse latitude diversity gradient (Fig. 2B;
130 Supplemental Fig. 2). Latitudinal patterns of AM fungal endemism showed a similar pattern as
131 AM richness, but EcM fungal rarity had a moderate increase near the equator compared to EcM
132 richness (Fig. 3). These relationships likely emerge based on a complex interplay of host
133 vegetation^{14,26}, degree of host specificity^{11,27}, and plant-soil nutrient economies^{28,29}. Most
134 importantly, these analyses stress that plant diversity is not a straightforward indicator for
135 mycorrhizal biodiversity at global scales^{20,26}. This is a problem for current conservation policy,
136 which tends to rely on plant and animal diversity metrics for determining priority areas and
137 ecosystem value^{5,30}.

138

139 (i) *Mycorrhizal richness and endemicity hotspots*

140

141 If plant diversity is not a clear indicator of mycorrhizal fungal diversity, then protecting
142 underground biodiversity requires explicit consideration, based on the development of new high-
143 resolution identification and monitoring approaches. For effective underground conservation,
144 locating global hotspots of mycorrhizal biodiversity is particularly important. We therefore used
145 our spatial predictions to identify areas of extremely species-rich and endemic mycorrhizal fungal

146 communities (“hotspots”). Hotspots were defined as pixels in the upper 95th percentile of predicted
147 richness and rarity values globally, following refs. 25 and 31. This resulted in richness cutoffs of
148 52.8 VTs (AM fungi) and 60.1 OTUs (EcM fungi), and unitless rarity cutoffs of 0.236 (AM fungi)
149 and 1.155 (EcM fungi).

150

151 We predict major hotspots of AM fungal richness and endemism across the savannas of the
152 Brazilian Cerrado, tropical forests across Southeast Asia, and Guinean forests in West Africa (Fig.
153 2A; Fig. 4). Our models show that these areas may contain more than 60 AM fungal species (VTs)
154 per soil sample. Extremely species-rich communities of AM fungi are likewise predicted in parts
155 of East and North China within ecosystems that are rapidly undergoing significant anthropogenic
156 land-use changes³². Smaller regions, like the Pampas grasslands in Argentina, dry forests
157 throughout Central America, and coastal plains in the Southeastern US, are also predicted to
158 contain highly diverse collections of AM fungi (Fig. 2A), but these areas are not predicted to be
159 rarity hotspots. Instead, AM fungal endemism hotspots are located in tropical and subtropical
160 forests in the eastern Amazon basin (Fig. 3A; Fig. 4). Although montane grasslands had the most
161 species-rich AM fungal samples in the model training data, this biome did not emerge as a
162 substantial AM fungal richness or rarity hotspot at the global level, potentially due to fewer
163 samples from these habitats relative to other biomes.

164

165 We predicted EcM fungal richness hotspots throughout northern forest ecosystems (Fig. 2B; Fig.
166 5). These include much of the Siberian and Canadian boreal forest regions; temperate coniferous
167 forests across Western US and Canadian mountain ranges; and temperate broadleaf and mixed
168 forests in Central Europe and the North American Great Lakes Region. We predict these places
169 can have soils averaging over 100 EcM fungal species (OTUs) per sample. However, the vast
170 majority of these northern forest biomes were not identified as EcM fungal endemism hotspots –
171 instead, we predict tundra ecosystems to contain the greatest EcM rarity levels at high latitudes
172 (Fig. 3B; Fig. 5). We also identified areas of global EcM diversity hotspots – especially rarity
173 hotspots – in tropical biomes, such as tropical moist forests in China’s Yunnan-Guizhou plateau
174 and mainland Southeast Asia, montane rainforests across Indonesia, and Guianan highland forests
175 in Venezuela. Of particular note are the predicted EcM fungal endemism hotspots in tropical
176 conifer forests in Mexico that contain the highest richness of EcM host plant lineages³³. EcM

177 fungal richness and rarity hotspots in the Southern hemisphere are predicted to be concentrated
178 primarily within the Andes Mountain range and the temperate mixed forests in southern Chile, the
179 east coast of Australia, and New Zealand.

180

181 Sampling bias can affect the relative rarity of OTUs or VTs, which means it is important to
182 determine if predicted rarity-weighted richness hotspots are biologically real or an artifact of
183 uneven global inventories of mycorrhizal fungi. We trained the AM and EcM fungal rarity
184 machine-learning models with additional covariate layers of global sampling intensity, which we
185 created using kernel density interpolation from sample coordinates in the SSU and ITS training
186 datasets (see Methods). This allowed us to compare empirical rarity predictions under current
187 sampling intensities against predictions that simulate universally ‘high-sampling’ scenarios, as in
188 ref. 34. For AM fungi, there is substantial overlap between the empirical and ‘high-sampling’ rarity
189 hotspot predictions (Supplemental Fig. 3A), indicating that current sampling efforts are generally
190 capturing the distribution of endemic AM fungal taxa. However, we find large divergence in the
191 empirical and ‘high-sampling’ predictions of EcM rarity hotspots, particularly in tundra and
192 tropical forest biomes (Supplemental Fig. 3B). This means that increasing sampling efforts in these
193 habitats is likely to reveal underrealized hotspots of rare and endemic EcM fungi.

194

195 (ii) *Protection of mycorrhizal hotspots*

196

197 To assess the extent to which mycorrhizal diversity is currently protected, we next overlaid
198 mycorrhizal hotspot pixels with protected areas defined by the World Database of Protected
199 Areas³⁵. We then calculated the total extent to which the hotspots fall within current protected area
200 delineations. This revealed that the centers of mycorrhizal biodiversity are unequally distributed
201 across biomes, nearly all have low protective coverage, and the management stringencies for
202 protected hotspots differ between mycorrhizal types (Figs. 4-5). Globally, we predict just 9.9% of
203 mycorrhizal richness hotspots occur in protected habitats, including 4.8% (~267,000 km²) of AM
204 hotspots and 15% (~821,000 km²) of EcM hotspots. Mycorrhizal rarity hotspots are protected at
205 roughly twice the rate (22.5% overall), with 20.3% (~1.1 million km²) and 24.5% (~1.3 million
206 km²) of AM and EcM fungal rarity hotspots overlapping with protected areas, respectively. This

207 is likely because protected areas are more frequently established in remote areas where human
208 pressures are expected to remain low regardless of conservation action³⁶.

209

210 The protected areas considered here cover roughly 16 million km² and contain everything from
211 strict nature reserves to managed resource areas (IUCN management categories I-VI and
212 NA/undefined). Analyzing hotspot protections by management category reveals most protected
213 AM fungal hotspots are in the least-strictly preserved habitats (Fig. 4). Specifically, 81.6%
214 (richness) and 71.9% (rarity) of protected AM hotspots are under categories V, VI, or
215 NA/undefined. In contrast, protected EcM fungal hotspots are mostly under the strictest
216 preservation levels, including nearly half of protected EcM richness and rarity hotspot area in
217 IUCN categories I-III (Fig. 5).

218

219 In tropical biomes, we found that only 4-7% of the total size of AM richness hotspots currently
220 overlap with designated protected areas, while between 9-29% of tropical AM rarity hotspots are
221 under some form of protection. This reflects the spatial discrepancy in AM richness vs. rarity
222 hotspots (Fig. 4), particularly comparing species-rich Southeast Asian tropical biomes and the
223 Amazon basin—the latter having larger protected areas where we predict greater AM fungal
224 endemism to occur. We found a similar pattern for EcM fungi in tropical moist forests: a relatively
225 small (9,700 km²) EcM richness hotspot showed 11% overlap with protected areas, but more than
226 33% of a relatively large (660,000 km²) EcM rarity hotspot is under protection within this same
227 biome (Fig. 5). However, the major EcM hotspots in northern latitude systems show just 10-16%
228 overlap with boreal forest protected areas, and 17-27% overlap with tundra protected areas.

229

230 At the continent scale, we found predicted mycorrhizal fungal hotspots in Asia had the lowest
231 protection for both AM fungi (richness = 2.0% and rarity = 6.3% protected) and EcM fungi
232 (richness = 11.9% and rarity = 17.3% protected), even though these hotspots were concentrated in
233 different biomes for each mycorrhizal type (tropical moist forests vs. boreal and tundra systems)
234 (Supplemental Figs. 4-15). Europe has the highest protected AM fungal richness hotspots (19.2%
235 protected) which we predict to occur mainly in Mediterranean forests across the continent.
236 Australia/Oceania showed the largest protected area overlap with AM fungal rarity hotspots
237 (32.4% protected) primarily in tropical grasslands. We find that predicted EcM fungal richness

238 hotspots are also most well-protected in Australia/Oceania (40.3% protected) which largely occur
239 throughout temperate broadleaf forests. In contrast, we find South America had the largest
240 protected area overlap with predicted EcM fungal endemism hotspots (48% protected).

241

242 (iii) *Drivers of mycorrhizal hotspots and biogeography*

243

244 Given the distinct hotspot locations and latitude-diversity relationships, we analyzed which
245 environmental covariates were the most important predictors of mycorrhizal fungal richness and
246 rarity using SHapley Additive exPlanations (SHAP) analysis. This method quantifies the additive
247 contribution of each feature of a machine learning model (e.g. temperature, rainfall) on its
248 predicted values, as opposed to measuring feature importance based on changes in model
249 performance³⁷. Therefore, it is a useful method for interpreting the magnitude and directionality
250 of predictors in our machine-learning models related to changes in mycorrhizal diversity.
251 Additionally, we found that including spatial predictors (Moran Eigenvector layers) generally did
252 not change the identity or order of important variables in either richness models. This confirms
253 that the rankings of the most important predictors of mycorrhizal fungal richness are unlikely to
254 be biased by spatially autocorrelated processes not included in the model (Supplemental Fig. 16-
255 17).

256

257 In testing 22 factors, we found that mean annual temperature and potential evapotranspiration were
258 two of the most important climate covariates across all AM and EcM models (Supplemental Fig.
259 18-19). By calculating mean absolute SHAP values, we found that these variables are two to five
260 times more influential than other predictors, contributing ± 1.7 VTs to each AM fungal richness
261 prediction, and ± 0.1 OTUs to each EcM fungal richness prediction on average. Specifically, higher
262 temperature and evapotranspiration values predicted greater AM fungal richness and lower EcM
263 fungal richness. However, the evapotranspiration trend reversed for EcM endemicity – higher EcM
264 fungal rarity was associated with increasingly dry areas. This supports recent work showing a link
265 between climate and soil fungi at global scales and likely reflects joint climate tolerances of
266 mycorrhizal fungi and their hosts, as well as climatic controls on soil organic matter
267 decomposition^{13,15,17,28,38}.

268

269 The importance of climatic controls in predicting mycorrhizal diversity is further supported by
270 strong links between fungal richness and soil organic carbon. We found greater soil organic carbon
271 was associated with higher EcM richness and lower AM richness (Supplemental Fig. 18; mean
272 absolute SHAP value = ± 2.9 VTs and ± 0.1 OTUs). This is consistent with past findings that soil
273 organic carbon stocks are positively associated with EcM plant dominance^{14,39}, and likely relates
274 to metabolic differences between AM and EcM fungi in enzymatic capabilities and plant host litter
275 chemistry^{9,40}. In contrast to EcM, AM fungal richness peaks in tropics, where climate and other
276 biotic conditions favors rapid mineralization of plant litter that may drive AM fungal dominance
277 over EcM symbiosis. Soil organic carbon was not among the top 10 most important predictors of
278 EcM fungal endemism (mean absolute SHAP value = ± 0.01 ; Supplemental Fig. 19), which helps
279 explain why we predict substantial EcM rarity hotspots across tropical forests with generally lower
280 soil carbon stocks than temperate or boreal forests¹². Interestingly, soil pH (as a geospatial
281 covariate layer) was not among the most important predictors identified in either AM or EcM
282 models, despite past work describing its importance in shaping overall soil fungal diversity
283 patterns and AM fungal niche traits^{16,41}.

284

285 In addition to climate, we were also interested in testing how human activities affect diversity
286 patterns. Given that 95% of the Earth's terrestrial surface is affected by some level of human
287 modification, we used the summed percentage of urban/built-up and cultivated/managed areas to
288 measure anthropogenic land-cover⁴² (dense urban centers were masked from predictions; see
289 Methods). Anthropogenic land-cover emerged as an important and positive driver of AM fungal
290 richness (mean absolute SHAP value = ± 2.9 VTs; Supplemental Fig. 18). While sampling bias in
291 the SSU dataset from human-modified habitats may be partially responsible, disturbed ecosystems
292 also tend to contain more ruderal, wind-dispersed AM fungal species that increase local species
293 richness at the expense of homogenizing regional scale species composition^{43,44}. We found that
294 human land modification was a less important feature in the AM fungal rarity model (mean
295 absolute SHAP value ± 0.006 VTs; Supplemental Fig. 19) than in the AM fungal richness model,
296 which could be interpreted as supporting the regional homogenization hypothesis. An open
297 question is whether changes in AM fungal associated vegetation, such as novel crops and exotic
298 plants, create more opportunities for diverse plant-AM fungal interactions to occur^{14,45}.

299

300 Aboveground plant biomass was the top predictor for EcM richness and rarity models (richness
301 mean absolute SHAP value = ± 0.14 OTUs; rarity mean absolute SHAP value = ± 0.06), with higher
302 biomass positively predicting higher EcM fungal richness and rarity. In general, this result reflects
303 the dominance of EcM fungi in forest ecosystems, including estimates that 60% of all tree stems
304 on Earth form EcM symbioses¹³. Elevation and slope were also among the most important
305 predictors of EcM rarity (mean absolute SHAP value ± 0.04 and ± 0.03 , respectively; Supplemental
306 Fig. 19), and we predict hotspots of endemic EcM fungal communities across mountainous terrain
307 in the Andes, Sierra Madre, Sierra Nevada, and Cascade Mountain ranges (Fig. 5). For EcM fungi,
308 aboveground biomass and topography may be more reliable indicators of richness and rarity than
309 plant diversity. This is encouraging given that remote sensing tools are increasingly able to detect
310 complex forest structures of significant ecological value for protection⁴⁶. Our results further
311 underscore the critical planetary benefits of high-biomass forests supported by and fostering
312 endemic, species-rich EcM fungal communities.

313

314 (iv) *Uncertainty analyses*

315

316 For spatial predictions of biodiversity to be useful in conservation policy, it is important to quantify
317 the confidence associated with each prediction²². This is especially true for microbial biodiversity
318 predictions that are more challenging to verify compared to plant and animal distributions⁴⁷. To
319 quantify uncertainty, we calculated two pixel-level metrics of model variance (see Methods).
320 Briefly, we define statistical uncertainty as the coefficient of variation in the predictions across
321 n=100 bootstrap samples of the training data, and we define extrapolation as the extent to which a
322 given pixel departs from the sampling locations and the range of environmental conditions in the
323 training data.

324

325 We found global AM fungal richness predictions have higher uncertainties compared to EcM
326 fungi, though the uncertainty range of rarity models was similar among mycorrhizal types (Fig. 6).
327 This is likely caused by differences in the characteristics of AM and EcM fungi that may affect
328 local-scale sample variation (e.g., host specificity, dispersal ability) and differences in the
329 underlying datasets (sample sizes and coverage across environmental gradients). The predictive
330 accuracy of the AM fungal richness model was higher than the EcM model (random cross-

331 validation $R^2 = 0.71$ vs 0.59), indicating that AM spatial predictions are robust despite the smaller
332 training dataset. Moreover, biomes with high uncertainty for one mycorrhizal type were not
333 consistently uncertain for the other type (Supplemental Fig. 20). For instance, AM fungal models
334 showed approximately double the uncertainty in desert, grassland, and coniferous forest biomes
335 compared to ECM models. ECM uncertainties were highest in areas with low predicted richness,
336 such as steppe and prairie regions (Fig. 6B).

337

338 Locations with high extrapolation reflect poor representation in the model predictors and are
339 primarily in regions that are consistently underrepresented in global soil biodiversity data⁴⁸
340 (Supplemental Fig. 21-22). For AM fungi, areas requiring greater extrapolation included northern
341 India, the Tibetan plateau, Indonesia, neotropical forests east of the Andes Mountains, and parts
342 of the Amazon basin (Fig. 6A). Therefore, extra caution is required in examining AM fungal
343 predictions in these regions. Overall, the ECM predictions are less dependent on model
344 extrapolation because they are built from a larger training dataset (over five times more ITS
345 samples distributed across a wider range of environmental conditions than the SSU samples used
346 for AM fungi). However, there are small regions with a high degree of ECM model extrapolation
347 in Sub-Saharan Africa, northern Canada, and Southeast Asia (Fig. 6B).

348

349 Important critiques of machine-learning geospatial approaches have been recently noted²¹. We
350 addressed these concerns by: (i) accounting for spatial auto-correlation variables in the model
351 (Supplemental Fig. 16-17), (ii) creating spatial blocks to limit the distance between sample and
352 prediction locations for additional cross-validation (Supplemental Fig. 23), and (iii) providing
353 clear data on prediction uncertainty (Fig. 4; Supplemental Figs. 21-22). Additionally, we
354 conducted multiple cross-validation procedures as there is currently no consensus on best-practices
355 for validating spatial models and measuring their inaccuracies (Methods). Overall, our models
356 perform similarly (or better) to the accuracy of recent machine learning predictive mapping
357 approaches for other soil organisms (Methods; Supplemental Fig. 24).

358

359 (v) *Conclusions*

360

361 Our high-resolution, global scale maps of AM and EcM fungal diversity suggest that less than 10%
362 of predicted mycorrhizal richness hotspots currently exist within protected areas. The same
363 analysis for non-fungal taxa shows that both plant and animal biodiversity hotspots are much better
364 protected. Replicating the analysis with spatial data of other taxonomic groups shows roughly 3-
365 fold higher protection of richness hotspot areas for vascular plants (27% protected), trees (34%
366 protected), ants (28% protected), and vertebrates (33-41% protected for amphibians, birds,
367 mammals, and reptiles), likely because threatened species within these groups drive most
368 conservation interventions^{31,49}. Endemic mycorrhizal fungi may be faring better: overall 22.5% of
369 predicted rarity hotspot areas overlap with protected areas, similar to the protection coverage of
370 vertebrate rarity centers (19-21% protected for amphibians, birds, mammals, and reptiles). To fully
371 realize the 30x30 target goals of protecting Earth's biodiversity under the Kunming-Montreal
372 Global Biodiversity Framework, it is necessary to identify underground conservation priorities, set
373 monitoring benchmarks, and create specific restoration plans⁵⁰. For instance, our data can be used
374 to help develop land management strategies to protect and maintain mycorrhizal diversity
375 reservoirs⁵¹, including identifying soil management practices developed over millennia by local
376 populations⁵². Importantly, we recommend that any use of these predictive biodiversity maps in a
377 specific locale be cross-referenced with the model uncertainty and extrapolation layers in the same
378 area to maximize positive conservation outcomes²².

379

380 Our results should be seen as a first step towards understanding the baseline protected levels of
381 Earth's mycorrhizal fungal diversity. Overlaying protected areas on biodiversity hotspots does not
382 tell us how successful these conservation strategies are at maintaining or improving mycorrhizal
383 fungal biodiversity. Because IUCN management categories vary in strictness of habitat
384 preservation, it is also important to analyze hotspot protections based on these management
385 categories, few of which are entirely free of human pressures. A critical next step will be measuring
386 the effectiveness of protected areas for mitigating fungal biodiversity loss, as has been done using
387 counterfactual analysis for other organisms⁵³. This will help determine which protected area
388 conditions and objectives best deliver conservation benefits to mycorrhizal fungi—such as
389 focusing on rare or threatened species, unique ecosystems and their entire communities, vital
390 ecosystem services, and/or locations of cultural/social significance. Conservation efforts will also
391 be more effective by continuously monitoring mycorrhizal fungi, as they have known

392 vulnerabilities to environmental change and disturbance, and there is a need to track their recovery
393 following conservation and restoration actions.

394

395 Due to insufficient data, we were unable to robustly model ErM or orchid mycorrhizal fungal
396 richness patterns. For instance, the average ErM fungal richness per sample was 0.6 species, and
397 less than 3% of samples contained more than four ErM fungal species. These are both important
398 fungal symbioses to understand in a global context with their unique contributions to plant ecology
399 and ecosystem functioning¹¹. Further sequencing projects and data-mining efforts will allow for
400 large-scale analyses soon. A second issue is that using SSU data to understand the biogeography
401 of AM fungal richness and rarity may be problematic due to overly conservative species definitions
402 (see Methods). AM fungi have a unique genetic organization with high intragenomic variation that
403 makes their sequencing and species definition more difficult than for EcM fungi, and only 332
404 AM fungal species have been formally described⁸. Alternative approaches, such as long-read
405 sequencing of both environmental AM fungi and culture collections and the use of multiple
406 metabarcoding target regions, may provide additional insights to the definition and distribution of
407 AM fungal biodiversity.

408

409 While our models perform similarly (or better) to those for other soil organisms (e.g., nematodes,
410 springtails, and earthworms), they are still an imperfect representation of mycorrhizal patterns in
411 nature. In particular, 1-km² is a relatively coarse spatial scale to quantify the average mycorrhizal
412 richness expected per soil sample in that area. There can be large landscape heterogeneity within
413 a pixel of this size and soil fungal sampling is usually restricted to a smaller collection grid and
414 sequencing less than 1 gram of soil. At present, 1-km² is the highest spatial resolution possible for
415 making reliable global predictions given the coverage of relevant covariate layers, but this
416 technical limitation will be overcome as more global environmental datasets are created at 30-m
417 resolution, closer to the scale of field sampling methods. We are also actively exploring how much
418 within-pixel variation affects model accuracy as part of an ongoing effort to independently ground-
419 truth these spatial predictions. Ground-truth validation is especially critical for: (i) under-sampled
420 environments poorly represented in the model; (ii) regions with high uncertainty despite good
421 training data coverage; (iii) areas with high landscape-related heterogeneity (e.g., steep

422 mountains); and (iv) areas that are particularly vulnerable to rapid climate change and habitat loss
423 from anthropogenic pressure.

424

425 Given the importance of these organisms for the productivity of ecosystems, and the functioning
426 of landscapes, it is surprising that mycorrhizal diversity remains vastly underrepresented in
427 conservation agendas. Explicit consideration of microbial hotspots like those identified here can
428 help direct and implement conservation strategies towards protecting the most diverse and
429 endangered belowground ecosystems.

430

431 **Methods**

432 *Fungal data*

433 Fungal occurrence records were generated from data-mining of published ITS and SSU sequencing
434 studies collected in the GlobalFungi and GlobalAMFungi databases (see refs. 6 and 8 for details),
435 as well as the ITS region from full-length sequences in the Global Soil Mycobiome consortium
436 database⁷. For EcM occurrences used in this study, we considered only samples of the ITS2
437 barcode deposited in the GlobalFungi dataset, since this marker is less biased by length variability
438 compared to ITS1^{54,55}, is better represented in the source database, and ITS1 samples failed
439 technical validation (see below). Briefly, raw sequences and metadata from 255 ITS studies
440 (representing the 4th release of the GlobalFungi database) were processed through an established
441 bioinformatic pipeline that incorporates sequence quality check, extraction of full ITS2 fungal
442 region using ITSx v1.1.2⁵⁶, clustering into operational taxonomic units (OTU) at 97% similarity
443 level with subsequent exclusion of chimeric sequences using USEARCH v11.0.0667⁵⁷. We used
444 BLASTN search against UNITE version 8.3, released 10.5.2021 to assign putative taxonomy to
445 non-singleton OTU. Default BLAST parameters were used, and representative sequences were
446 considered to belong to the closest BLAST hit genera in the case of >92% similarity and >95%
447 coverage. Considering that the majority of EcM fungal genera do not include species belonging to
448 other functional guilds, we find these BLAST results criteria reasonable. We excluded OTUs
449 represented by sequences with e-value > e-50. The resulting OTU table with taxonomy
450 assignments was compared against the FungalTraits database v.1.2 to subset EcM fungi⁵⁸.

451

452 In early tests, we used the ITS dataset to model AM fungi, but their low abundance detected from
453 using this target region was leading to clearly erroneous spatial predictions (e.g., zero AM species
454 predicted in regions known to be dominated by AM vegetation). As a result, SSU data from the
455 GlobalAMFungi database was used to estimate AM fungal species richness⁸. Briefly, raw
456 sequences and metadata from 45 SSU studies (representing the 1st release of the GlobalAMFungi
457 database) were subjected to sequence quality check, trimming of the sequences to the V4 region
458 of the SSU and assignment of the sequences to Virtual Taxa (VT) from the MaarjAM database
459 5.6.2019 release⁵⁹, using BLASTN with a required sequence coverage of $\geq 98\%$ and sequence
460 similarity $\geq 97\%$, based on the most recent release of type sequences of VT. VT are defined based
461 on phylogenetic grouping of SSU sequences at roughly species level diversity (which may result
462 in conservative taxonomic designations of AM fungal taxa⁶⁰). Molecular identification based on
463 VTs is typically used for AM fungal metabarcoding studies using the SSU region, and OTUs are
464 more traditionally used for general fungal community sequencing (including EcM fungi) based on
465 the ITS region. Both VTs and OTUs are commonly used to approximate species-level assignments
466 in mycorrhizal fungal sequencing studies. Whereas the EcM OTU data used for this analysis is the
467 result of clustering of GlobalFungi sequences followed by taxonomic assignment, the AMF
468 pipeline results in more conservative estimates of diversity since SSU sequences are assigned to
469 representative VTs using stringent BLAST parameters. This results in the discarding of sequences
470 without a good match to these known VTs, which is necessary to filters non-fungal sequences
471 (e.g., Annelids) but also likely removes undescribed AM fungal taxa. In total, 2.8 billion fungal
472 sequences were processed across 24,982 samples which identified 332 VTs of AM fungi
473 (encompassing 86% of defined VT in the MaarjAM database), and 164,439 EcM fungal OTUs.
474

475 *Richness estimates*

476 We used Hill numbers ($q=0$) to measure the number of mycorrhizal species (OTUs or VTs) in each
477 sample⁶¹. This approach creates a sequencing depth-based rarefaction and extrapolation sampling
478 curve (i.e., a sample-specific species accumulation curve), with diversity estimates and 95%
479 confidence intervals calculated at curve asymptotes (iNEXT R package⁶²). Extrapolation endpoints
480 were computed as twice the sequencing depth for each sample. The rarefaction/extrapolation
481 approach to estimate mycorrhizal richness allows for a robust comparison of mycorrhizal patterns

482 across studies of multiple sequencing technologies (although variation in error rates may persist),
483 primer sets, and sequencing intensities.

484

485 We removed outliers in mycorrhizal richness estimates prior to spatial modeling. Two Australian
486 studies with ITS samples from Desert and Mediterranean biomes had unusually high EcM OTU
487 richness estimates and standard deviations (two orders of magnitude higher than other Australian
488 ecoregions; refs. 63, 64), and have been previously marked as potentially inaccurate based on a
489 recent database comparison⁶. All samples from these two studies were removed (N=2,070). We
490 then filtered samples by biome by removing estimated richness values that were more than five
491 times the interquartile range higher than the biome-level median estimate. We only filtered values
492 at the highest end of the distribution to avoid potentially removing ‘true’ estimates of low or zero
493 mycorrhizal richness at a given location. See Supplemental Table 1 for a summary of outliers
494 removed per biome.

495

496 We also calculated rarity-weighted richness to estimate the relative endemism of mycorrhizal
497 fungal communities. Using samples from the outlier filtered dataset (above), we created a species-
498 level prevalence score as the number of samples where each OTU or VT occurs, and then summed
499 the inverse prevalence score of all species present in each sample⁶⁵. This rarity metric is useful for
500 identifying site endemism, or areas with a high concentration of rare taxa²³. Rarity-weighted
501 richness was originally used with geographically un-biased dataset where each pixel contained
502 information on bird observations⁶⁵, but has since proven useful in other contexts to predict patterns
503 of relative species endemism^{34,66}. To account for the possibility that geographically uneven
504 sampling effort affects fungal species prevalence scores and biases predictions of fungal
505 endemism, we simulated a global ‘high-sampling’ scenario for rarity spatial models (see below).

506

507 *Geospatial modeling, validation, and analysis*

508 Spatial predictions of mycorrhizal fungal richness and rarity were created with a random forest
509 modeling approach (modified from ref. 67). We first sampled a collection of >70 global
510 environmental covariate layers at each of the locations within the dataset. These layers contain
511 macroclimatic, soil texture and physicochemical information, vegetation, radiation and
512 topographic indices and anthropogenic variables. To reduce overfitting and variance inflation, we

removed highly correlated variables, with the final set of predictors comprising 22 environmental variables. Details of all layers, including descriptions, units, and source information, are described in ref. 68. Variables describing soil structure and physicochemical properties were obtained from SoilGrids⁶⁹, limited to the upper 5 cm of soil. Climate information (i.e., mean annual temperature, annual precipitation, monthly maximum temperature, precipitation seasonality) was obtained from CHELSA⁷⁰. Spectral vegetation indices (i.e., MODIS NPP product MOD17A3HGF V6.1, averaged annually) were obtained from the Google Earth Engine Data Catalog. We used aboveground biomass data from CDIAC⁷¹. We obtained topographic and consensus landcover information (including human development percentage) from EarthEnv^{72,73}. The Potential Evapotranspiration layer was obtained from CGIAR⁷⁴. Resolve Ecoregion classifications were used to categorize sampling locations into biome⁷⁵. All spatial covariate layers were reprojected and resampled to a unified pixel grid in EPSG:4326 (WGS84) at 30 arc-sec resolution (approximately 1-km² at the equator). Areas covered by permanent snow or ice (e.g. the Greenland ice cap, glaciated mountain ranges; identified using SoilGrids), barren landcover (sparsely vegetated regions defined in EarthEnv), and highly urban/built-up areas were excluded from the analyses. Antarctic areas were excluded from analysis due to limited coverage of covariate layers in the region.

To harmonize the data across the different experimental approaches of the original studies, we included four types of project-specific variables: sequencing platform (Illumina, IonTorrent, PacBio, 454Roche, and DNBSEQ-G400), target gene region (ITS1, ITS2), sample type (soil, topsoil, 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 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. To create spatial predictions, we harmonized the project-specific variables to the most common levels: sequencing platform = Illumina; target_gene = ITS2 (implicit because ITS1 samples were removed); sample type = soil; primer set = ITS3 – ITS4 for EcM. Prior

544 to modeling, all project-specific variables were transformed from categorical to binary variables
545 (i.e., one-hot encoding or ‘dummy variables’). See Supplemental Figs. 25-29 for more details on
546 technical validation and data filtering.

547

548 After outlier removal and validation, the SSU dataset contained 3,234 samples (332 total VTs) for
549 AM fungal analysis, and the ITS dataset had 17,519 samples (41,086 total OTUs) for EcM fungal
550 analysis. To create the training datasets and to reduce potential overfitting, we used only distinct
551 observations of mycorrhizal fungal richness (i.e., samples with non-identical richness or rarity
552 values). This meant that when multiple samples fell within the same 1-km² pixel, we removed
553 duplicate richness estimates because multiple identical observations caused the random forest
554 models to overfit towards those combinations of richness levels and environmental variables.
555 Using only distinct observations allowed us to retain important within-pixel variation rather than
556 aggregating across multiple richness estimates per pixel. To deal with the zero-inflated data
557 structure of the EcM dataset, we adopted a two-step approach. First, a binary random forest
558 classification model was created to separate positive occurrence data (richness > 0) from samples
559 with undetected EcM communities (richness = 0). We then trained a regression random forest
560 model on the subset of positive occurrence samples with a log(x+1) transformation of richness.
561 Results of the regression model were multiplied by the binary classification model to create a
562 combined prediction. For AM we only trained a regression model as there were no zeros in this
563 dataset. We iteratively varied the model hyperparameters (ie., variables per split [4, 6, 8, 10, 12]
564 and minimum leaf population [2, 4, 6, 8, 10, 12], while keeping the number of trees constant at
565 250). In total, hyperparameters were iteratively changed across 30 models for each AM and EcM
566 response variable.

567

568 There is ongoing debate about the best methodology for validating spatial models⁷⁶⁻⁷⁹. As a result,
569 we tested each model using both random cross-validation and k-fold Nearest Neighbor Distance
570 matching cross-validation (kNNDM)⁷⁹. Random folds were assigned stratified per biome, to
571 ensure equal coverage of each fold. For both AM and EcM, the final predictions are an ensemble
572 (mean) of the top 10 best performing models based on coefficient of determination R² with
573 kNNDM 10-fold cross validation. This approach resulted in the following predictive accuracy
574 metrics for richness models: AM random cross-validated R² = 0.71, AM kNNDM cross-validated

575 $R^2 = 0.14$; EcM random cross-validated $R^2 = 0.59$, EcM kNNDM cross-validated $R^2 = 0.20$
576 (Supplemental Fig. 24A). The predictive accuracy metrics for rarity models are as follows: AM
577 random cross-validated $R^2 = 0.56$, AM kNNDM cross-validated $R^2 = -3.75$; EcM random cross-
578 validated $R^2 = 0.62$, EcM kNNDM cross-validated $R^2 = 0.27$ (Supplemental Fig. 24B). Overall,
579 these performance estimates are similar to (or exceed) the predictive accuracy of global machine
580 learning models for other soil organisms^{67,81,82}. The poor performance of the AM fungal rarity
581 model based on kNNDM cross-validation suggests that there are spatial/environmental structures
582 in the dataset that our model is unable to detect, which is possible if the VT dataset fails to
583 sufficiently capture AM fungal endemism patterns based on conservative taxonomic assignments
584 from the AM fungal reference database⁶⁰. We used SHapley Additive exPlanations (SHAP) to
585 interpret the importance and directionality of environmental features³⁷. Standard feature
586 importance algorithms, like the ones implemented in Random Forest or Boosted Regression Tree
587 models, measure feature importance as a change in model performance. In contrast, SHAP is based
588 on the magnitude of feature contribution. We selected SHAP as the most suitable method here
589 because we are interested in the effect of the features on the predicted value (as opposed to the
590 effect on model error) to interpret the importance and directionality of the features.

591
592 As we expect the model performance to decrease with distance to training locations, we also
593 performed spatially buffered leave-one-out cross-validation (SLOO-cv)⁸³, a computationally
594 intensive approach where a separate model is trained for every unique location in the dataset,
595 leaving out locations within a range of buffer sizes. At larger buffer sizes (i.e., >500 km), SLOO-
596 cv R^2 values are in line with kNNDM cross-validation R^2 values. To transform these findings into
597 a visual representation, we plotted the coefficient of determination (R^2), against the distance to the
598 nearest sampling location, creating a spatial product (Supplemental Fig. 23). The resulting map
599 was consistent with our previously described extrapolation map.

600
601 To generate a spatial understanding of our predictive accuracy, we created 100 bootstrap samples
602 by resampling the training datasets with replacement, using biome-based stratification. Leveraging
603 the hyperparameter settings of the best performing random forest model, we generated 100 global
604 prediction images. These were subsequently used to calculate a bootstrap coefficient of variation
605 (derived by dividing the standard deviation by the mean), and 95% confidence intervals for each

606 pixel (Supplemental Fig. 21-22). To locate environmental conditions and corresponding
607 geographic regions that are underrepresented in the training data, we first converted the data into
608 the principal component (PC) space. We then selected the first 13 and 14 axes that cumulatively
609 accounted for 90% of the total variance, for the models for AM and EcM, respectively. We
610 evaluated whether raster pixels were associated with environmental conditions internal or external
611 to the convex hull circumscribing each pairwise combination of PC axes. We defined the degree
612 of model extrapolation as the proportion of total pairwise combinations for which a pixel lies
613 outside the corresponding convex hull. Finally, we combined this environmental extrapolation data
614 with a map of geographic distance between sampling locations (distance and environmental
615 extrapolation maps were combined in a 2:1 ratio, after scaling each from 0 to 1), thereby producing
616 a global spatial assessment of the representativeness of our datasets (Supplemental Fig. 21-22).
617 The extrapolation map was then used to mark pixels that are most underrepresented by the training
618 data using a 5% extrapolation cutoff (i.e., areas with <95% coverage in training data space).

619

620 To explore possible spatial dependency in the data, we fit semivariogram models per mycorrhizal
621 type on richness estimates and model residuals using the R package *automap* v. 1.1-9⁸⁴. For AM
622 fungi, we observe autocorrelation in the extrapolated richness data up to ca. 900-km; for EcM
623 fungi this range reached ca. 550-km. This level of spatial autocorrelation in response variables is
624 unsurprising given the scale of our georeferenced datasets, similar to a recent analysis showed that
625 spatial weights of ~800-km were best at handling spatial autocorrelation in a global plant diversity
626 dataset⁸⁵. For model residuals, semi-variograms showed weak, although significant, spatial
627 dependency for both AM and EcM fungal data (Supplemental Fig. 17). We then performed a
628 Global Moran's *I* test with permutations calculated using the *sfdep* R package⁸⁶, including five
629 nearest points as neighbors and Gaussian kernel function for building spatial weight matrices. For
630 the raw data, spatial autocorrelation in estimated richness was positive and highly significant
631 (Moran's *I*: AM = 0.71; EcM = 0.64), whereas low in model residuals for both fungal types
632 (Moran's *I*: AM = 0.11; EcM = 0.23; Supplemental Fig. 16) similar to a previous study finding
633 limited effects of geographic distance on alpha diversity³⁸. To incorporate spatial processes into
634 the prediction model, we applied spatial eigenvector mapping using distance-based Moran's
635 eigenvector maps (dbMEM)⁸⁷. We calculated dbMEM using the R package *adespatial* v. 0.3-21⁸⁸,
636 with the custom function *quickMEM*⁸⁹. Including dbMEM in the modeling approach showed

637 mycorrhizal hotspots predicted in the same areas but reduced some finer-scale variation in richness
638 predictions in the non-spatial models (Supplemental Fig. 30). This led us to conclude that the
639 models without dbMEM capture most variation, yet there might be some fine-scale spatial
640 processes that are not explained for AM fungi here.

641

642 We evaluated how uneven sampling density might change rarity patterns by comparing model
643 predictions using current sampling intensity vs. simulating globally homogenous and high-
644 sampling intensity, as in ref. 34. We created 1-km² resolution sampling intensity layers for the
645 SSU and ITS training datasets using kernel density interpolation from sample coordinates with 5-
646 degree radius and uniform decay rate. These layers were rescaled 0-1 and included as covariates
647 in the rarity random forest models to make empirical predictions of AM and EcM fungal rarity
648 based on the current distribution of sampling efforts. We then extrapolated rarity models under a
649 global and equally ‘high-sampling’ scenario by setting all pixels in the sampling intensity covariate
650 layer to the maximum value of 1. Comparing patterns from the empirical vs. ‘high-sampling’ rarity
651 models shows where future sampling may reveal underrealized hotspots of mycorrhizal fungal
652 rarity³⁴. For all analyses (e.g., latitude trends and hotspot protections), we use the ‘high-sampling’
653 rarity models to avoid biases in endemism predictions based on sampling intensity.

654

655 To measure how mycorrhizal richness and rarity varies with latitude, predicted richness and rarity
656 pixel values were averaged at every 0.2 latitude degrees within one degree-wide longitude bands⁹⁰.
657 Correlations between latitude and estimated mycorrhizal richness and rarity of samples are also
658 provided in Supplemental Fig. 2. Mycorrhizal hotspots were defined by setting a cutoff at the 95th
659 percentile of predicted richness values. These hotspot pixels were then overlaid with the World
660 Database of Protected Areas cropped to spatial predictions here (WDPA;
661 www.protectedplanet.net, ref. 35) before calculating the total extent to which mycorrhizal hotspots
662 in different biomes fall within current protected area delineations. Hotspot overlap with protected
663 areas was calculated for each IUCN management category (terrestrial WDPA categories I-VI and
664 unassigned). To understand mycorrhizal hotspot protections relative to other organisms, we
665 compiled recently published spatial data on the species diversity of vascular plants²⁵, trees⁹¹, ants³⁴,
666 and four groups of vertebrates (amphibians, birds, mammals, and reptiles; IUCN⁹²). The same
667 hotspot analysis described above was applied to these spatial layers, and the total amount of

richness and rarity center 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 are visualized in Robinson projection.

671

672 **Acknowledgments**

SPUN is supported by grants from the Jeremy and Hannelore Grantham Environmental Trust, Paul Allen Family Foundation, the Schmidt Family Foundation. ETK acknowledges support from NWO-VICI (202.012) and HFSP (RGP 0029). CA was supported by Ambizione grant no. PZ00P3_17990 from the Swiss National Science Foundation. JDS was supported by the NOW-Gravity Grant Microp (024.004.014). TWC was supported by grants from DOB Ecology and the Bernina Foundation. TV and PK acknowledge support from the Czech Science Foundation (21-17749S), PB acknowledges the support from the MEYS (LC23152, LM2023055). LT acknowledges support from the Estonian Science Foundation (grant PRG632). KP 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).

683

684 **Author Contributions**

JvdH, CA, and MVN performed the initial processing and modeling of fungal datasets provided by TV, PK, and PB. MVN, CA, JS, OP, AC, LvG, and TK provided feedback on the modeling process and interpreted results. CQ led the calculations of rarity-weighted richness. MVN performed technical validation and additional spatial analyses. MVN wrote the first manuscript draft, and all authors provided comments and significant edits to the manuscript.

690

691 **Conflict of Interest Statement**

TK and CA are the founders of the Society for the Protection of Underground Networks (SPUN), an organization that conducts scientific research and advocates for the protection of belowground network forming fungi. TWC is the founder of Restor, a non-governmental organization (NGO) that facilitates the global restoration movement. CA is the founder of Funga, an organization that facilitates the restoration of belowground fungal biodiversity. GF is the founding director of Fungi Foundation, an NGO that explores and educates about fungal diversity for applications and conservation.

699

700 **Data Availability Statement**

701 Data and maps will be archived and made available on Zenodo prior to publication. Interactive
702 maps will be available on Google Earth Engine. Fungal data used in this study is available from
703 the following previous publications: <https://doi.org/10.1038/s41597-020-0567-7>;
704 <https://doi.org/10.1111/nph.19283>; <https://doi.org/10.1007/s13225-021-00493-7>;
705 <https://doi.org/10.1007/s13225-020-00466-2>.

706

707 **Code Availability Statement**

708 Code will be made available on Github prior to publication.

709

710 **References (Main Text)**

- 711 1. Höglberg, M.N. and Höglberg, P. (2002). Extramatrical ectomycorrhizal mycelium
712 contributes one-third of microbial biomass and produces, together with associated roots,
713 half the dissolved organic carbon in a forest soil. *New Phytologist*, 154, 791-795.
- 714 2. van Der Heijden, M.G., Martin, F.M., Selosse, M.A., & Sanders, I.R. (2015).
715 Mycorrhizal ecology and evolution: the past, the present, and the future. *New
716 Phytologist*, 205, 1406-1423.
- 717 3. Hawkins, H.J., Cargill, R.I., Van Nuland, M.E., Hagen, S.C., Field, K.J., Sheldrake, M.,
718 et al. (2023). Mycorrhizal mycelium as a global carbon pool. *Current Biology*, 33, R560-
719 R573.
- 720 4. Cavicchioli, R., Ripple, W.J., Timmis, K.N., Azam, F., Bakken, L.R., Baylis, M., et al.
721 (2019). Scientists' warning to humanity: microorganisms and climate change. *Nature
722 Reviews Microbiology*, 17, 569-586.
- 723 5. Guerra, C.A., Bardgett, R.D., Caon, L., Crowther, T.W., Delgado-Baquerizo, M.,
724 Montanarella, L., et al. (2021a). Tracking, targeting, and conserving soil
725 biodiversity. *Science*, 371, 239-241.
- 726 6. Větrovský, T., Morais, D., Kohout, P., Lepinay, C., Algora, C., Awokunle Hollá, S., et al.
727 (2020). GlobalFungi, a global database of fungal occurrences from high-throughput-
728 sequencing metabarcoding studies. *Scientific Data*, 7, 228.

- 729 7. Tedersoo, L., Mikryukov, V., Anslan, S., Bahram, M., Khalid, A.N., Corrales, A., et al.
730 (2021). The Global Soil Mycobiome consortium dataset for boosting fungal diversity
731 research. *Fungal Diversity*, 111, 573-588.
- 732 8. Větrovský, T., Kolaříková, Z., Lepinay, C., Awokunle Hollá, S., Davison, J.,
733 Fleyberková, A., et al. (2023). GlobalAMFungi: a global database of arbuscular
734 mycorrhizal fungal occurrences from high-throughput sequencing metabarcoding studies.
735 *New Phytologist*, 240, 2151-2163.
- 736 9. Frey, S.D. (2019). Mycorrhizal fungi as mediators of soil organic matter
737 dynamics. *Annual Review of Ecology, Evolution, and Systematics*, 50, 237-259.
- 738 10. Tedersoo, L., & Bahram, M. (2019). Mycorrhizal types differ in ecophysiology and alter
739 plant nutrition and soil processes. *Biological Reviews*, 94, 1857-1880.
- 740 11. Brundrett, M.C., & Tedersoo, L. (2018). Evolutionary history of mycorrhizal symbioses
741 and global host plant diversity. *New Phytologist*, 220, 1108-1115.
- 742 12. Crowther, T.W., van den Hoogen, J., Wan, J., Mayes, M.A., Keiser, A.D., Mo, L., et al.
743 (2019). The global soil community and its influence on biogeochemistry. *Science*, 365,
744 eaav0550.
- 745 13. Steidinger, B.S., Crowther, T.W., Liang, J., Van Nuland, M.E., Werner, G.D., Reich,
746 P.B., et al. (2019). Climatic controls of decomposition drive the global biogeography of
747 forest-tree symbioses. *Nature*, 569, 404-408.
- 748 14. Soudzilovskaia, N.A., van Bodegom, P.M., Terrer, C., Zelfde, M.V.T., McCallum, I., et
749 al. (2019). Global mycorrhizal plant distribution linked to terrestrial carbon
750 stocks. *Nature Communications*, 10, 5077.
- 751 15. Kivlin, S. N., Hawkes, C. V., & Treseder, K. K. (2011). Global diversity and distribution
752 of arbuscular mycorrhizal fungi. *Soil Biology and Biochemistry*, 43, 2294-2303.
- 753 16. Tedersoo, L. Bahram, M., Põlme, S., Kõljalg, U., Yorou, N. S., Wijesundera, R., et al.
754 (2014). Global diversity and geography of soil fungi. *Science*, 346, 1256688.
- 755 17. Větrovský, T., Kohout, P., Kopecký, M., Machac, A., Man, M., Bahnmann, B.D., et al.
756 (2019). A meta-analysis of global fungal distribution reveals climate-driven
757 patterns. *Nature Communications*, 10, 5142.
- 758 18. Guerra, C.A., Berdugo, M., Eldridge, D.J., Eisenhauer, N., Singh, B.K., Cui, H., et al.
759 (2022). Global hotspots for soil nature conservation. *Nature*, 610, 693-698.

- 760 19. Tedersoo, L., Mikryukov, V., Zizka, A., Bahram, M., Hagh-Doust, N., Anslan, S., et al.
761 (2022). Global patterns in endemicity and vulnerability of soil fungi. *Global Change
762 Biology*, 28, 6696-6710.
- 763 20. Cameron, E.K., Martins, I.S., Lavelle, P., Mathieu, J., Tedersoo, L., Bahram, M., et al.
764 (2019). Global mismatches in aboveground and belowground biodiversity. *Conservation
765 Biology*, 33, 1187-1192.
- 766 21. Meyer, H., & Pebesma, E. (2022). Machine learning-based global maps of ecological
767 variables and the challenge of assessing them. *Nature Communications*, 13, 2208.
- 768 22. Jansen, J., Woolley, S.N., Dunstan, P.K., Foster, S.D., Hill, N.A., Haward, M., &
769 Johnson, C. R. (2022). Stop ignoring map uncertainty in biodiversity science and
770 conservation policy. *Nature Ecology & Evolution*, 6, 828-829.
- 771 23. Albuquerque, F., Astudillo-Scalia, Y., Loyola, R., & Beier, P. (2019). Towards an
772 understanding of the drivers of broad-scale patterns of rarity-weighted richness for
773 vertebrates. *Biodiversity and Conservation*, 28, 3733-3747.
- 774 24. Kinlock, N.L., Prowant, L., Herstoff, E.M., Foley, C.M., Akin-Fajiye, M., Bender, N., et
775 al. (2018). Explaining global variation in the latitudinal diversity gradient: Meta-analysis
776 confirms known patterns and uncovers new ones. *Global Ecology and Biogeography*, 27,
777 125-141.
- 778 25. Sabatini, F.M., Jiménez-Alfaro, B., Jandt, U. et al. (2022). Global patterns of vascular
779 plant alpha diversity. *Nature Communications*, 13, 4683.
- 780 26. Toussaint, A., Bueno, G., Davison, J., Moora, M., Tedersoo, L., Zobel, M., et al. (2020).
781 Asymmetric patterns of global diversity among plants and mycorrhizal fungi. *Journal of
782 Vegetation Science*, 31, 355-366.
- 783 27. Kokkoris, V., Lekberg, Y., Antunes, P.M., Fahey, C., Fordyce, J.A., Kivlin, S.N., & Hart,
784 M.M. (2020). Codependency between plant and arbuscular mycorrhizal fungal
785 communities: what is the evidence?. *New Phytologist*, 228, 828-838.
- 786 28. Read, D. J. (1991). Mycorrhizas in ecosystems. *Experientia*, 47, 376-391.
- 787 29. Phillips, R.P., Brzostek, E., & Midgley, M. G. (2013). The mycorrhizal-associated
788 nutrient economy: a new framework for predicting carbon–nutrient couplings in
789 temperate forests. *New Phytologist*, 199, 41-51.

- 790 30. Jung, M., Arnell, A., De Lamo, X., García-Rangel, S., Lewis, M., Mark, J., et al. (2021).
791 Areas of global importance for conserving terrestrial biodiversity, carbon and
792 water. *Nature Ecology & Evolution*, 5, 1499-1509.
- 793 31. Jenkins, C. N., Pimm, S. L., & Joppa, L. N. (2013). Global patterns of terrestrial
794 vertebrate diversity and conservation. *Proceedings of the National Academy of
795 Sciences*, 110, E2602-E2610.
- 796 32. Chen, C., Park, T., Wang, X., Piao, S., Xu, B., Chaturvedi, R.K., et al. (2019). China and
797 India lead in greening of the world through land-use management. *Nature
798 Sustainability*, 2, 122-129.
- 799 33. Tedersoo, L. (2017). Global biogeography and invasions of ectomycorrhizal plants: past,
800 present and future. In *Biogeography of mycorrhizal symbiosis* (Vol. 230), Tedersoo, L.
801 (Ed.), pp 469-531. New York, NY, USA. Springer.
- 802 34. Kass, J.M., Guénard, B., Dudley, K.L., Jenkins, C.N., Azuma, F., Fisher, B.L., et al.
803 (2022). The global distribution of known and undiscovered ant biodiversity. *Science
804 Advances*, 8, eabp9908.
- 805 35. Bingham, H.C., Juffe Bignoli, D., Lewis, E., MacSharry, B., Burgess, N.D., Visconti, P.,
806 et al. (2019). Sixty years of tracking conservation progress using the World Database on
807 Protected Areas. *Nature Ecology & Evolution*, 3, 737-743.
- 808 36. Joppa, L. N., & Pfaff, A. (2009). High and far: biases in the location of protected
809 areas. *PloS one*, 4, e8273.
- 810 37. Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., et al. (2020).
811 From local explanations to global understanding with explainable AI for trees. *Nature
812 Machine Intelligence*, 2, 56-67.
- 813 38. Mikryukov, V., Dulya, O., Zizka, A., Bahram, M., Hagh-Doust, N., Anslan, S., et al.
814 (2023). Connecting the multiple dimensions of global soil fungal diversity. *Science
815 Advances*, 9, eadj8016.
- 816 39. Averill, C., Turner, B.L., & Finzi, A.C. (2014). Mycorrhiza-mediated competition
817 between plants and decomposers drives soil carbon storage. *Nature*, 505, 543-545.
- 818 40. Hicks Pries, C.E., Lankau, R., Ingham, G.A., Legge, E., Krol, O., Forrester, J., et al.
819 (2023). Differences in soil organic matter between EcM-and AM-dominated forests
820 depend on tree and fungal identity. *Ecology*, 104, e3929.

- 821 41. Davison, J., Moora, M., Semchenko, M., Adenan, S. B., Ahmed, T., Akhmetzhanova,
822 A.A., et al. (2021). Temperature and pH define the realised niche space of arbuscular
823 mycorrhizal fungi. *New Phytologist*, 231, 763-776.
- 824 42. Kennedy, C.M., Oakleaf, J.R., Theobald, D.M., Baruch-Mordo, S., Kiesecker, J. (2019).
825 Managing the middle: A shift in conservation priorities based on the global human
826 modification gradient. *Global Change Biology*, 25, 811-826.
- 827 43. Chaudhary, V.B., Nolimal, S., Sosa-Hernández, M.A., Egan, C., & Kastens, J. (2020).
828 Trait-based aerial dispersal of arbuscular mycorrhizal fungi. *New Phytologist*, 228, 238-
829 252.
- 830 44. Guerra, C.A., Delgado-Baquerizo, M., Duarte, E., Marigliano, O., Görgen, C., Maestre,
831 F.T., & Eisenhauer, N. (2021b). Global projections of the soil microbiome in the
832 Anthropocene. *Global Ecology and Biogeography*, 30, 987-999.
- 833 45. Guzman, A., Montes, M., Hutchins, L., DeLaCerda, G., Yang, P., Kakouridis, A., et al.
834 (2021). Crop diversity enriches arbuscular mycorrhizal fungal communities in an
835 intensive agricultural landscape. *New Phytologist*, 231, 447-459.
- 836 46. Lang, N., Jetz, W., Schindler, K., & Wegner, J. D. (2023). A high-resolution canopy
837 height model of the Earth. *Nature Ecology & Evolution*, 7, 1778-1789.
- 838 47. Barron, E. (2023) Conservation of abundance: How fungi can contribute to rethinking
839 conservation. *Conservation and Society*, 21, 99-109.
- 840 48. Guerra, C. A., Heintz-Buschart, A., Sikorski, J., Chatzinotas, A., Guerrero-Ramírez, N.,
841 Cesarz, S., (2020). Blind spots in global soil biodiversity and ecosystem function
842 research. *Nature Communications*, 11, 3870.
- 843 49. Senior, R. A., Bagwyn, R., Leng, D., Killion, A. K., Jetz, W., & Wilcove, D. S. (2024).
844 Global shortfalls in documented actions to conserve biodiversity. *Nature*, 1-5.
- 845 50. Labouyrie, M., Ballabio, C., Romero, F., Panagos, P., Jones, A., Schmid, M. W., et al.
846 (2023). Patterns in soil microbial diversity across Europe. *Nature Communications*, 14,
847 3311.
- 848 51. Averill, C., Anthony, M. A., Baldrian, P., Finkbeiner, F., Van den Hoogen, J., Kiers, T.,
849 et al. (2022). Defending Earth's terrestrial microbiome. *Nature Microbiology*, 7, 1717-
850 1725.

- 851 52. Fleischman, F., Coleman, E., Fischer, H., Kashwan, P., Pfeifer, M., Ramprasad, V., et al.
852 (2022). Restoration prioritization must be informed by marginalized people. *Nature*, 607,
853 E5-E6.
- 854 53. Langhammer, P. F., Bull, J. W., Bicknell, J. E., Oakley, J. L., Brown, M. H., Bruford, M.
855 W., et al. (2024). The positive impact of conservation action. *Science*, 384, 453-458.
- 856

857 **References (Methods)**

- 859 54. Tedersoo, L., & Lindahl, B. (2016). Fungal identification biases in microbiome
860 projects. *Environmental Microbiology Reports*, 8, 774-779.
- 861 55. Yang, R.H., Su, J.H., Shang, J.J., Wu, Y.Y., Li, Y., Bao, D.P., et al. (2018) Evaluation of
862 the ribosomal DNA internal transcribed spacer (ITS), specifically ITS1 and ITS2, for the
863 analysis of fungal diversity by deep sequencing. *PLoS ONE*, 13, 206428.
- 864 56. Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., et al.
865 (2013). Improved software detection and extraction of ITS1 and ITS 2 from ribosomal
866 ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing
867 data. *Methods in Ecology and Evolution*, 4, 914-919.
- 868 57. Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon
869 reads. *Nature Methods*, 10, 996-998.
- 870 58. Põlme, S., Abarenkov, K., Henrik Nilsson, R., Lindahl, B.D., Clemmensen, K.E.,
871 Kauserud, H., et al. (2020). FungalTraits: a user-friendly traits database of fungi and
872 fungus-like stramenopiles. *Fungal Diversity*, 105, 1-16.
- 873 59. Öpik, M., Vanatoa, A., Vanatoa, E., Moora, M., Davison, J., Kalwij, J. M., et al. (2010).
874 The online database MaarjAM reveals global and ecosystemic distribution patterns in
875 arbuscular mycorrhizal fungi (Glomeromycota). *New Phytologist*, 188, 223-241.
- 876 60. Bruns, T. D., & Taylor, J. W. (2016). Comment on “Global assessment of arbuscular
877 mycorrhizal fungus diversity reveals very low endemism”. *Science*, 351, 826-826.
- 878 61. Chao, A., Gotelli, N.J., Hsieh, T.C., Sander, E.L., Ma, K.H., Colwell, R.K., & Ellison,
879 A.M. (2014). Rarefaction and extrapolation with Hill numbers: a framework for sampling
880 and estimation in species diversity studies. *Ecological Monographs*, 84, 45-67.

62. Hsieh, T.C., Ma, K.H., & Chao, A. (2016). iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution*, 7, 1451-1456.
63. Bissett, A., Fitzgerald, A., Meintjes, T., Mele, P. M., Reith, F., Dennis, P. G., et al. (2016). Introducing BASE: the Biomes of Australian Soil Environments soil microbial diversity database. *GigaScience*, 5, s13742-016.
64. Yan, D., Mills, J.G., Gellie, N.J., Bissett, A., Lowe, A.J., & Breed, M.F. (2018). High-throughput eDNA monitoring of fungi to track functional recovery in ecological restoration. *Biological Conservation*, 217, 113-120.
65. Usher, M. B. (1986). Wildlife conservation evaluation: Attributes criteria and values. In M. Usher (Ed.). *Wildlife conservation evaluation*. (pp. 3–44). London: Chapman & Hall.
66. Albuquerque, F., & Beier, P. (2016). Predicted rarity-weighted richness, a new tool to prioritize sites for species representation. *Ecology and Evolution*, 6, 8107-8114.
67. van den Hoogen, J., Geisen, S., Routh, D., Ferris, H., Traunspurger, W., Wardle, D.A., et al. (2019). Soil nematode abundance and functional group composition at a global scale. *Nature*, 572, 194-198.
68. van den Hoogen, J., Geisen, S., Wall, D.H., Wardle, D.A., Traunspurger, W., de Goede, R.G., et al. (2020). A global database of soil nematode abundance and functional group composition. *Scientific Data*, 7, 103.
69. Hengl, T., Mendes de Jesus, J., Heuvelink, G.B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., et al. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE*, 12, e0169748.
70. Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., et al. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, 4, 1-20.
71. Ruesch, A. & Gibbs, H. K. (2008). New IPCC Tier-1 global biomass carbon map for the year 2000. <https://cdiac.ess-dive.lbl.gov>.
72. Tuanmu, M.N., & Jetz, W. (2014). A global 1-km consensus land-cover product for biodiversity and ecosystem modelling. *Global Ecology and Biogeography*, 23, 1031-1045.

- 912 73. Tuanmu, M.N., & Jetz, W. (2015). A global, remote sensing-based characterization of
913 terrestrial habitat heterogeneity for biodiversity and ecosystem modelling. *Global*
914 *Ecology and Biogeography*, 24, 1329-1339.
- 915 74. Trabucco, A., and Zomer, R.J. (2018). Global Aridity Index and Potential Evapo-
916 Transpiration (ET0) Climate Database v2. CGIAR Consortium for Spatial Information
917 (CGIAR-CSI). <https://cgiarcsi.community>.
- 918 75. Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N.D., Wikramanayake, E., et al.
919 (2017). An ecoregion-based approach to protecting half the terrestrial realm. *BioScience*,
920 67, 534-545.
- 921 76. Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., et al.
922 (2020). Spatial validation reveals poor predictive performance of large-scale ecological
923 mapping models. *Nature Communications*, 11, 4540.
- 924 77. Wadoux, A. M. C., Heuvelink, G. B., De Bruin, S., & Brus, D. J. (2021). Spatial cross-
925 validation is not the right way to evaluate map accuracy. *Ecological Modelling*, 457,
926 109692.
- 927 78. Meyer, H., & Pebesma, E. (2021). Predicting into unknown space? Estimating the area of
928 applicability of spatial prediction models. *Methods in Ecology and Evolution*, 12, 1620-
929 1633.
- 930 79. Milà, C., Mateu, J., Pebesma, E., & Meyer, H. (2022). Nearest neighbour distance
931 matching Leave-One-Out Cross-Validation for map validation. *Methods in Ecology and*
932 *Evolution*, 13, 1304-1316.
- 933 80. Linnenbrink, J., Milà, C., Ludwig, M., & Meyer, H. (2024). kNNDM CV: k-fold nearest-
934 neighbour distance matching cross-validation for map accuracy estimation. *Geoscientific*
935 *Model Development*, 17, 5897-5912.
- 936 81. Phillips, H.R., Guerra, C.A., Bartz, M.L., Briones, M.J., Brown, G., Crowther, T.W., et
937 al. (2019). Global distribution of earthworm diversity. *Science*, 366, 480-485.
- 938 82. Potapov, A.M., Guerra, C.A., van den Hoogen, J., Babenko, A., Bellini, B.C., Berg, M.
939 P., et al. (2023). Globally invariant metabolism but density-diversity mismatch in
940 springtails. *Nature Communications*, 14, 674.

- 941 83. Ploton, P., Mortier, F., Réjou-Méchain, M. et al. (2020). Spatial validation reveals poor
942 predictive performance of large-scale ecological mapping models. *Nature Communications*, 11, 4540.
- 943 84. Hiemstra, P.H., Pebesma, E.J., Twenhöfel, C.J., & Heuvelink, G.B. (2009). Real-time
944 automatic interpolation of ambient gamma dose rates from the Dutch radioactivity
945 monitoring network. *Computers & Geosciences*, 35, 1711-1721.
- 946 85. Cai, L., Kreft, H., Taylor, A., Denelle, P., Schrader, J., Essl, F., et al. (2023). Global
947 models and predictions of plant diversity based on advanced machine learning
948 techniques. *New Phytologist*, 237, 1432-1445.
- 949 86. Parry J (2023). sfdep: Spatial Dependence for Simple Features. R package version 0.2.3,
950 <<https://CRAN.R-project.org/package=sfdep>>.
- 951 87. Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., et al.
952 (2007). Methods to account for spatial autocorrelation in the analysis of species
953 distributional data: A review. *Ecography*, 30, 609–628.
- 954 88. Dray S., Bauman D., Blanchet G., Borcard D., Clappe S., Guénard G., et al. (2023).
955 adespatial: Multivariate Multiscale Spatial Analysis. R package version 0.3-21,
956 <<https://CRAN.R-project.org/package=adespatial>>.
- 957 89. Borcard, D., Gillet, F., & Legendre, P. (2018). Numerical Ecology with R. Springer
958 International Publishing.
- 959 90. Zhu, K., Woodall, C.W., & Clark, J.S. (2012). Failure to migrate: lack of tree range
960 expansion in response to climate change. *Global Change Biology*, 18, 1042-1052.
- 961 91. Liang, J., Gamarra, J.G., Picard, N., Zhou, M., Pijanowski, B., Jacobs, D.F., et al. (2022).
962 Co-limitation towards lower latitudes shapes global forest diversity gradients. *Nature
963 Ecology & Evolution*, 6, 1423-1437.
- 964 92. The IUCN Red List of Threatened Species Version 2022-2 (IUCN, 2022);
965 <https://www.iucnredlist.org>.
- 966

967 **Figure 1. Sample locations and mycorrhizal richness trends by biome.** Distribution of sites
968 and richness estimates for **A**) arbuscular mycorrhizal (AM) fungi and **B**) ectomycorrhizal (EcM)
969 fungi. Virtual Taxa (VT) were created from SSU sequences for AM fungi; 97% similar operational
970 taxon units (OTU) were created from ITS sequences and assigned to EcM fungi. Mycorrhizal
971 richness patterns across terrestrial biomes are shown in boxplots (Mangroves, Flooded grasslands,
972 and Rock/Ice biomes not shown due to few samples in both datasets). Global distributions of
973 estimated richness are shown in density curves above biome-level boxplots. Boxplots indicate the
974 median (center line), first and third quartiles (lower and upper box edges), and 1.5x IQR (box
975 whiskers). EcM richness data is shown as square-root transformed for visualization. Richness
976 estimates were calculated using a rarefaction/extrapolation approach incorporating sequencing
977 depth per sample. Points shown here are all samples that passed quality control checks and used
978 as training data.

979

980 **Figure 2. Global predictions and latitudinal trends of mycorrhizal fungal richness.** Maps
981 show the predicted richness of **A**) AM fungal VTs and **B**) EcM fungal OTUs per soil sample within
982 pixels (pixels approximately 1-km²). Predicted richness values are capped at 60 VTs (AM) and 70
983 OTUs (EcM) for visualization. Masked locations (grey) are sparsely vegetated zones and dense
984 urban areas based on global land-cover data. Cross-hatches are superimposed over areas that are
985 underrepresented by the training data (highly extrapolated) where model predictions should be
986 interpreted with caution. Plots on the right show mean richness trends across latitude (shaded area
987 is \pm 2 SE).

988

989 **Figure 3. Global predictions and latitudinal trends of mycorrhizal fungal endemism.** Maps
990 show the predicted rarity-weighted richness of **A**) AM fungal VTs and **B**) EcM fungal OTUs per
991 soil sample within pixels (pixels approximately 1-km²). Rarity-weighted richness is a unitless
992 metric, and predicted values are capped at 0.25 (AM) and 1.3 (EcM) for visualization. Masked
993 locations (grey) are sparsely vegetated zones and dense urban areas based on global land-cover
994 data. Cross-hatches are superimposed over areas that are underrepresented by the training data
995 (highly extrapolated) where model predictions should be interpreted with caution. Plots on the
996 right show mean rarity trends across latitude (shaded area is \pm 2 SE).

997

998

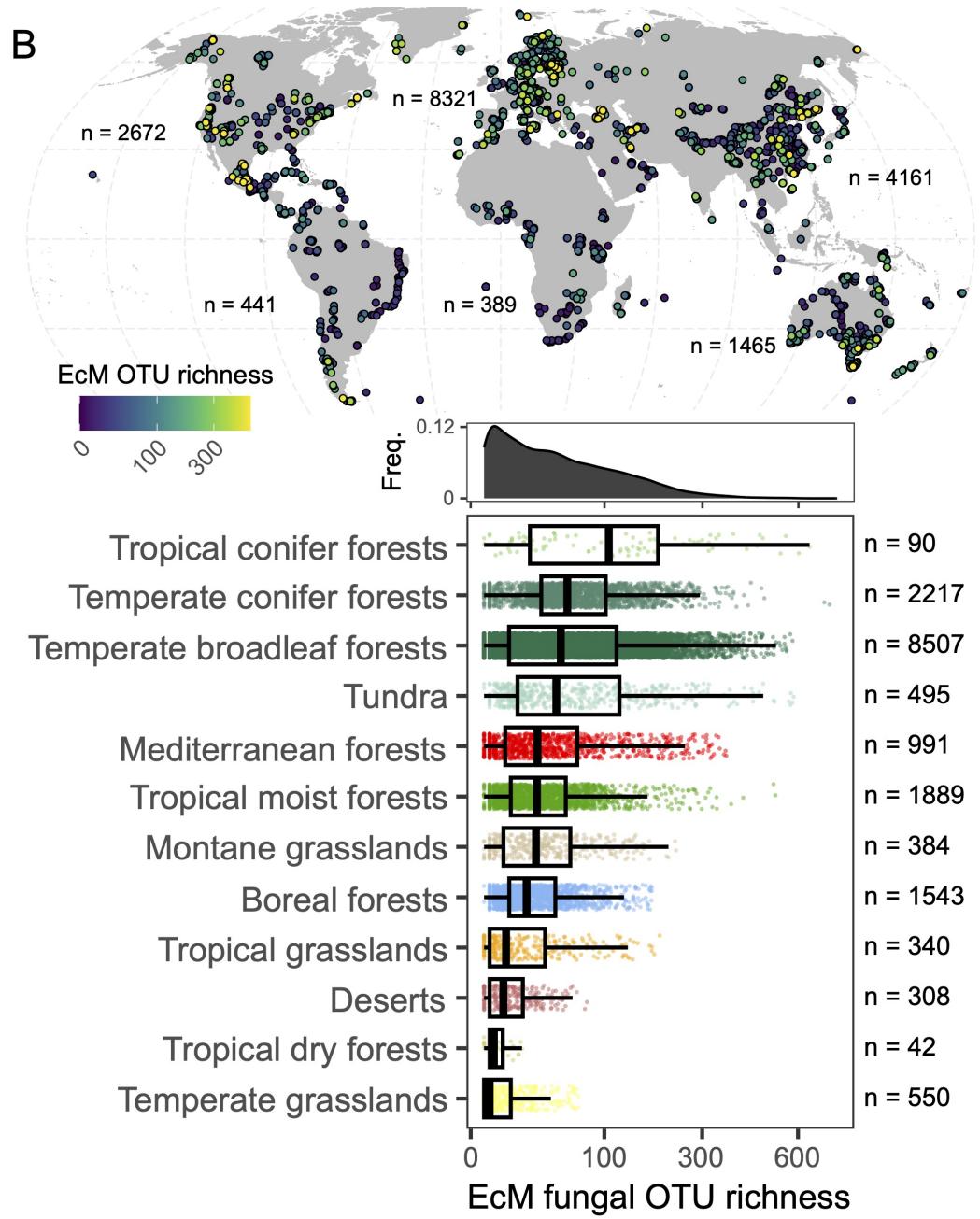
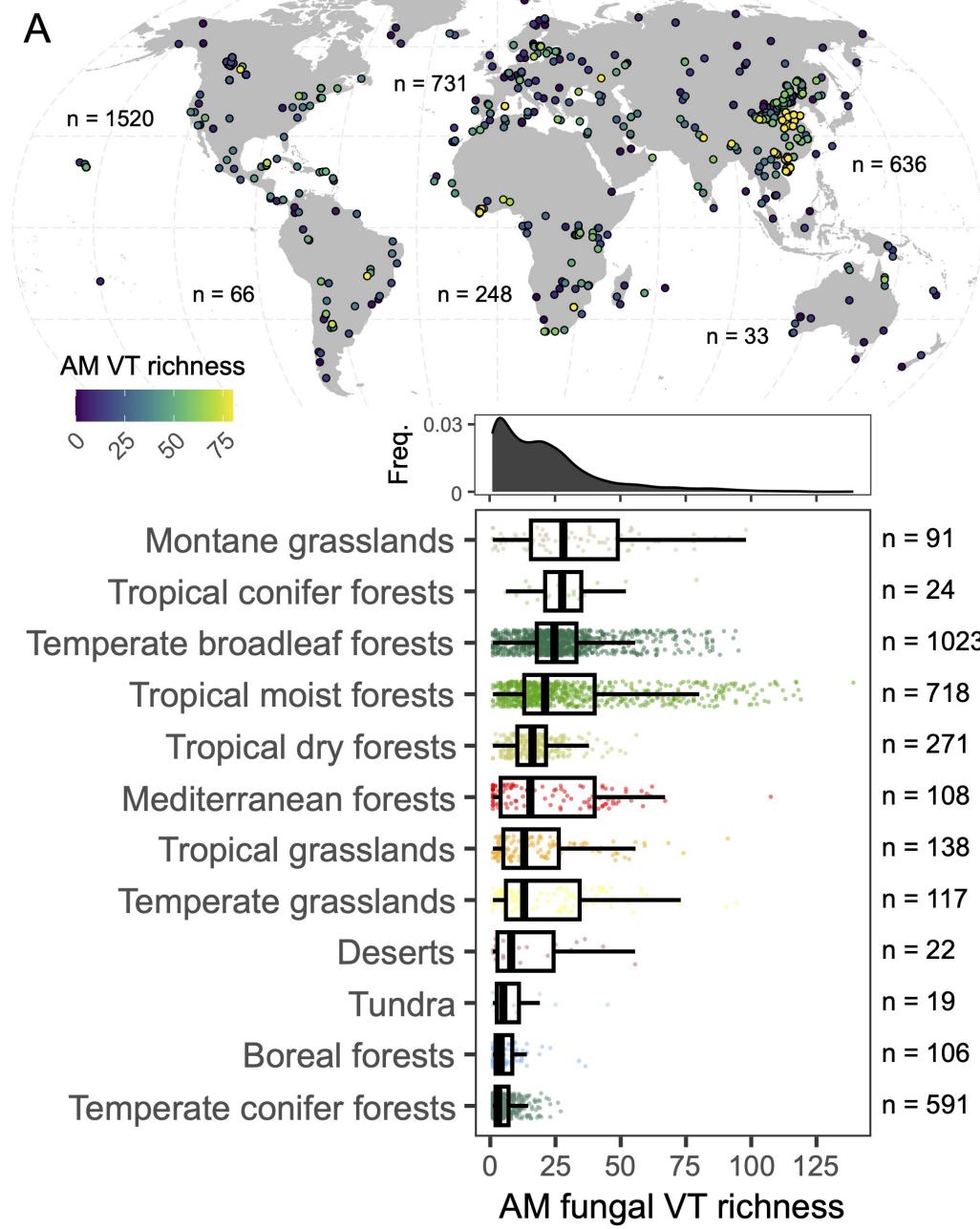
999 **Figure 4. Arbuscular mycorrhizal fungi hotspots and global protected areas.** Predicted
1000 richness and endemism hotspots (95th percentile of predictions) for AM fungi. Colored areas on
1001 map show richness hotspots (green), rarity hotspots (purple), and the overlap of richness and rarity
1002 hotspots (yellow); dark grey = non-hotspot. Bar-graphs show the total hotspot size and percent
1003 overlap with protected areas by biome, and grey bars reflect different IUCN management
1004 categories (I = most strictly preserved habitats; na = unassigned category). The dashed line at 30%
1005 hotspot area protected reflects the ambitions of 30x30 target goals under the Kunming-Montreal
1006 Global Biodiversity Framework.

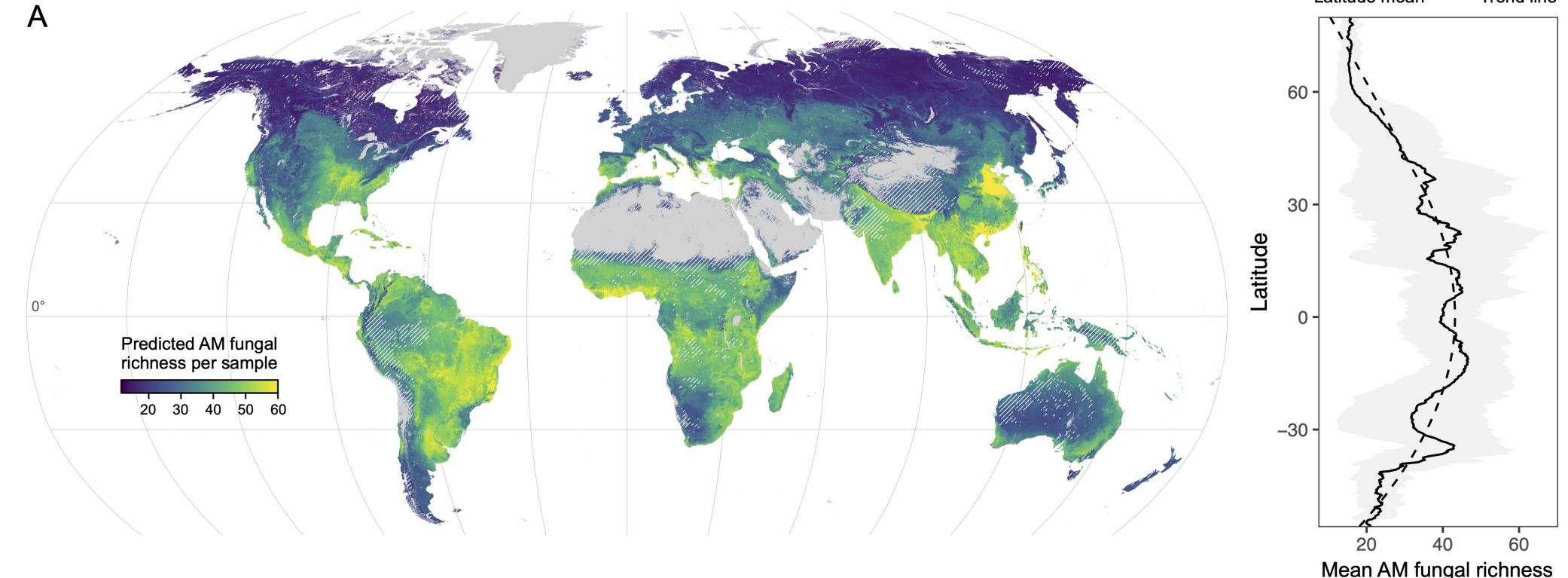
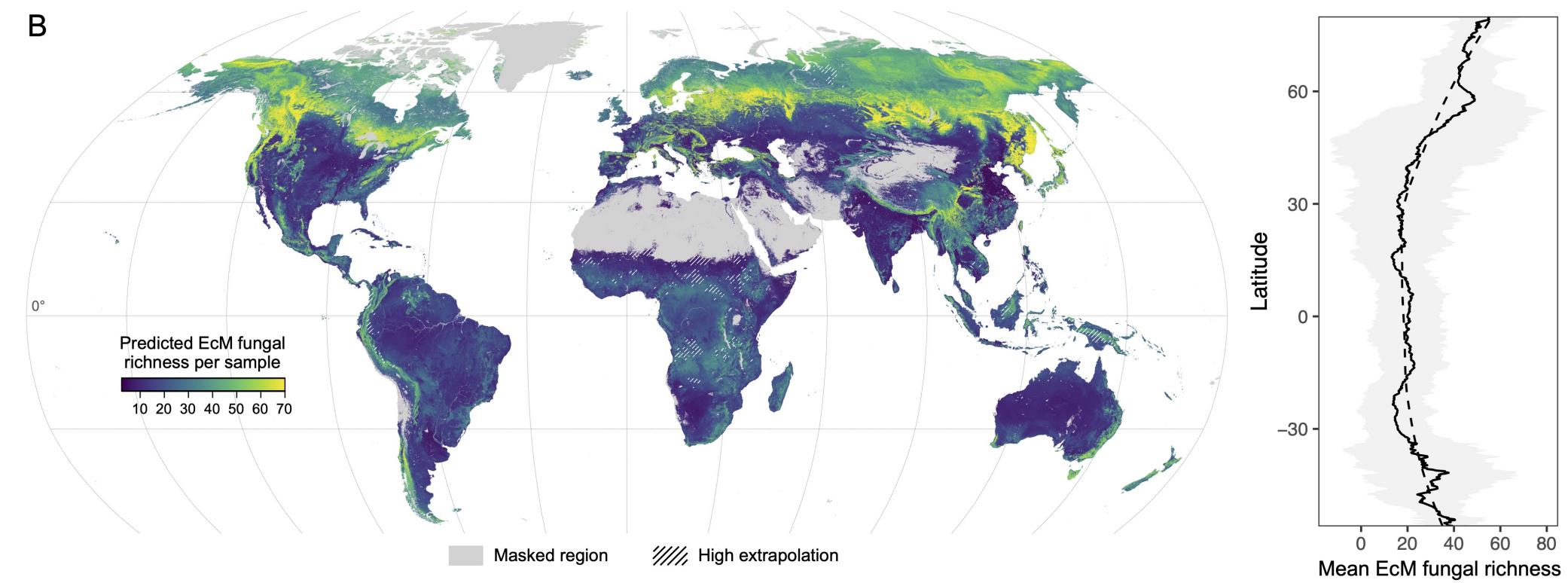
1007

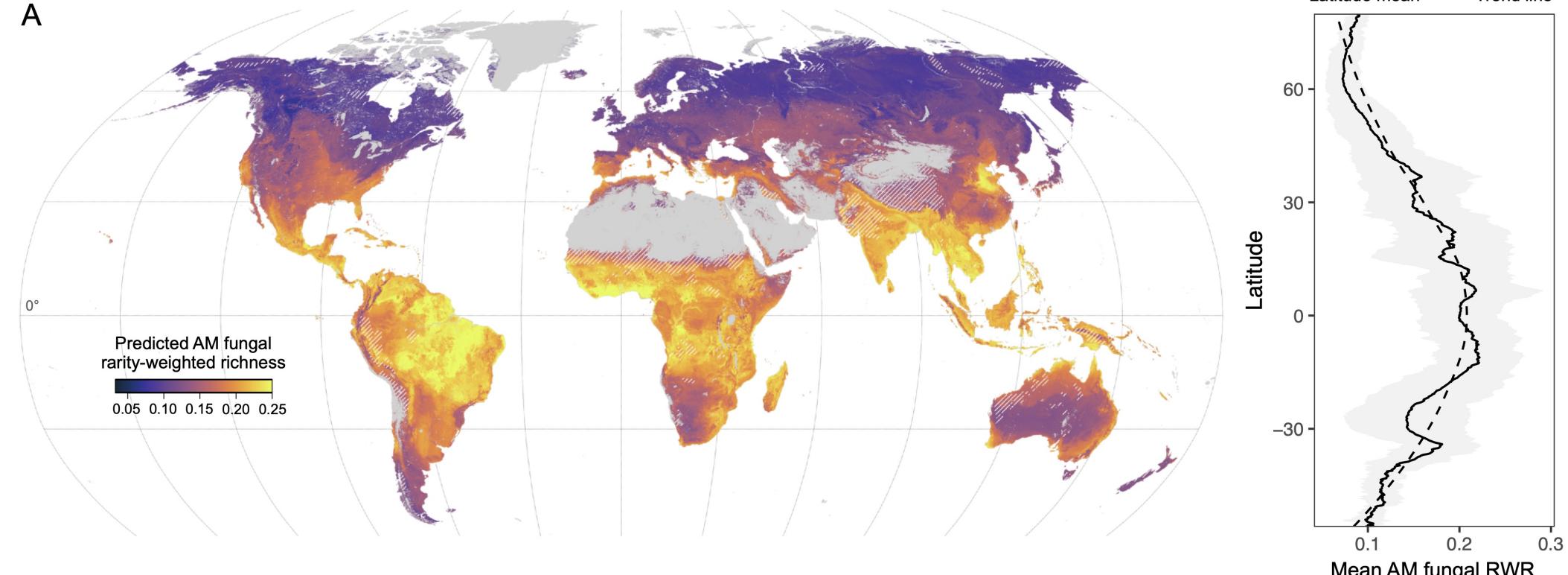
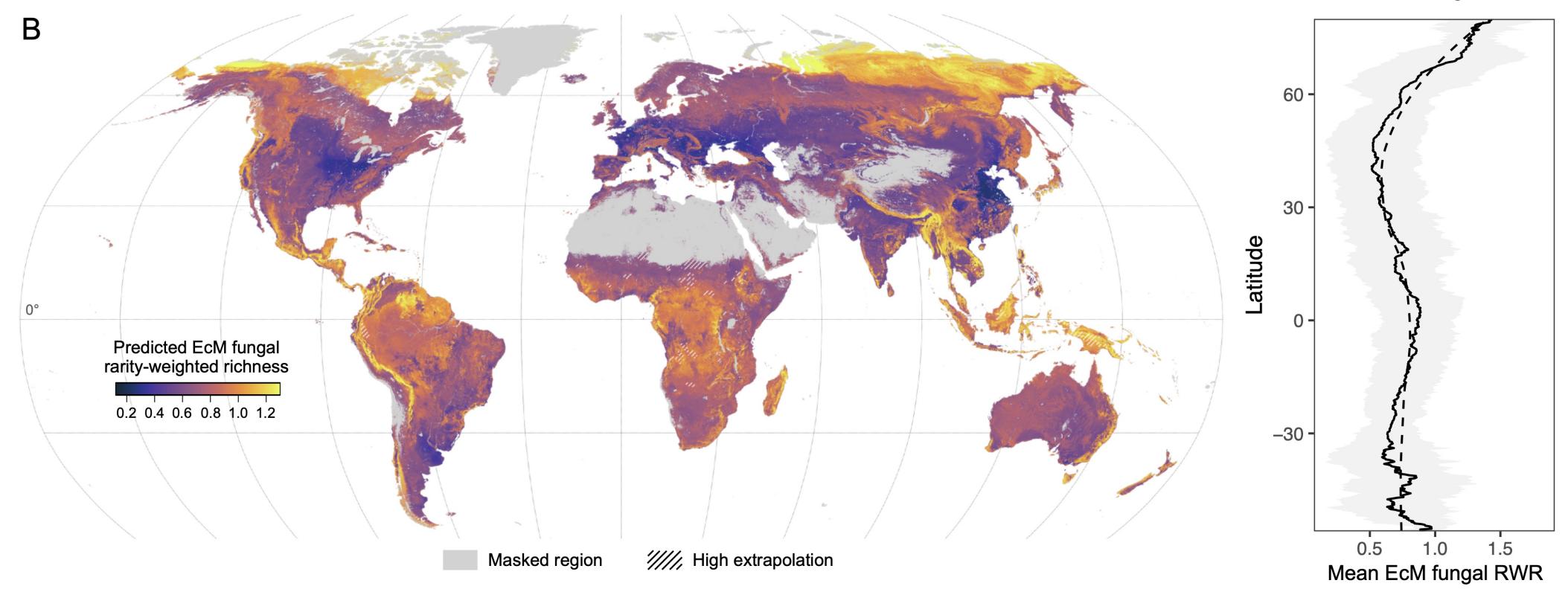
1008 **Figure 5. Ectomycorrhizal fungi hotspots and global protected areas.** Predicted richness and
1009 endemism hotspots (95th percentile of predictions) for EcM fungi. Colored areas on map show
1010 richness hotspots (green), rarity hotspots (purple), or the overlap of richness and rarity hotspots
1011 (yellow); dark grey = non-hotspot. Bar-graphs show the total hotspot size and percent overlap with
1012 protected areas by biome, and grey bars reflect different IUCN management categories (I = most
1013 strictly preserved habitats; na = unassigned category). The dashed line at 30% hotspot area
1014 protected reflects the ambitions of 30x30 target goals under the Kunming-Montreal Global
1015 Biodiversity Framework.

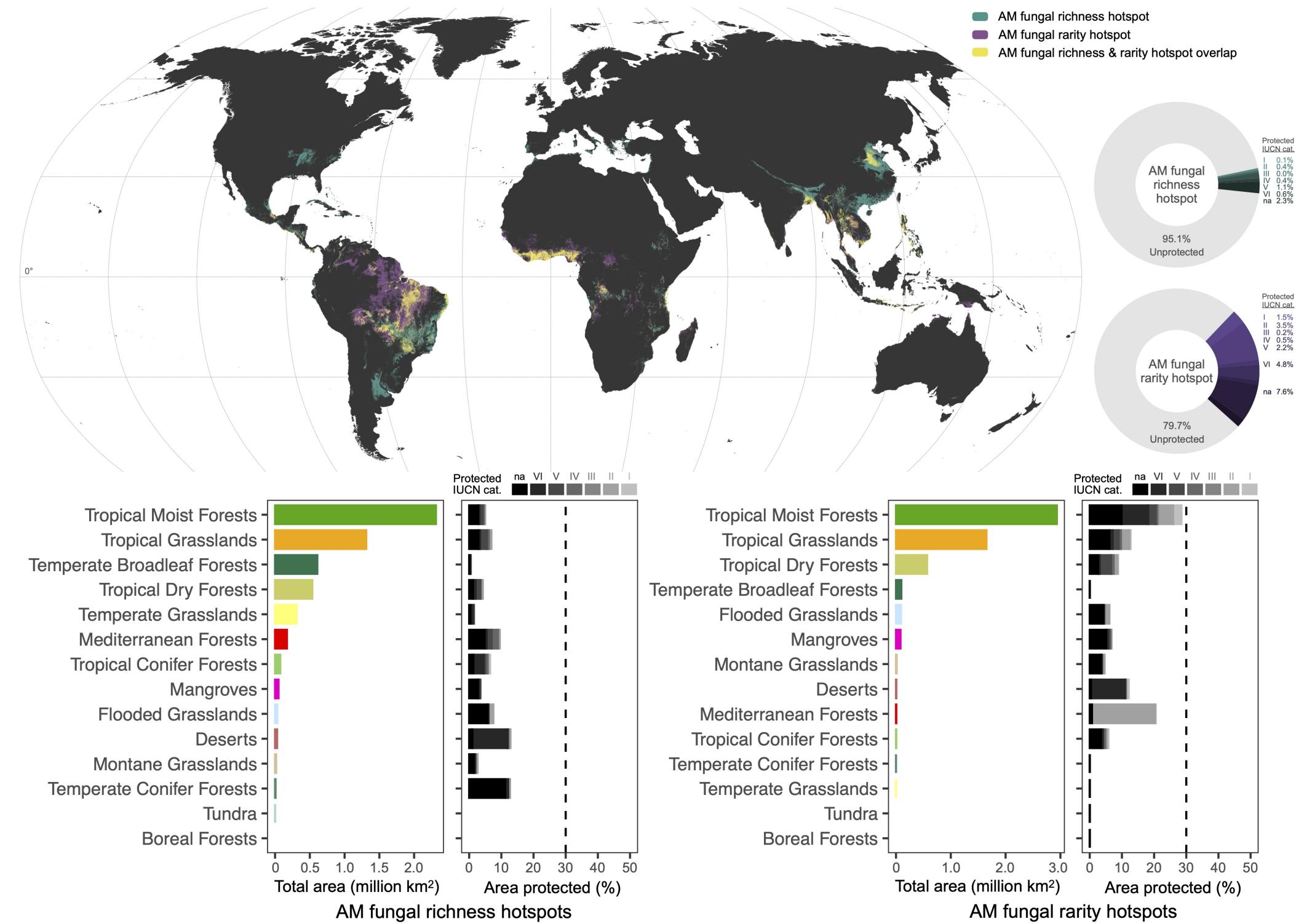
1016

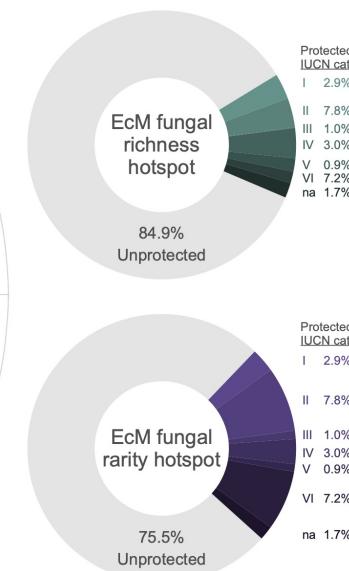
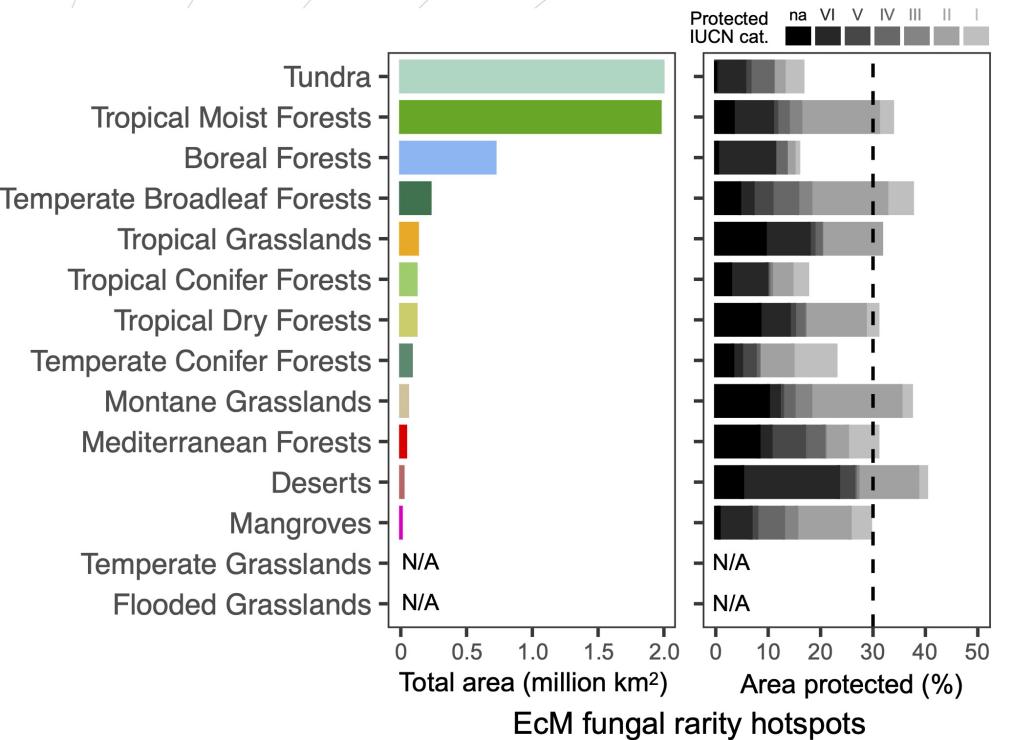
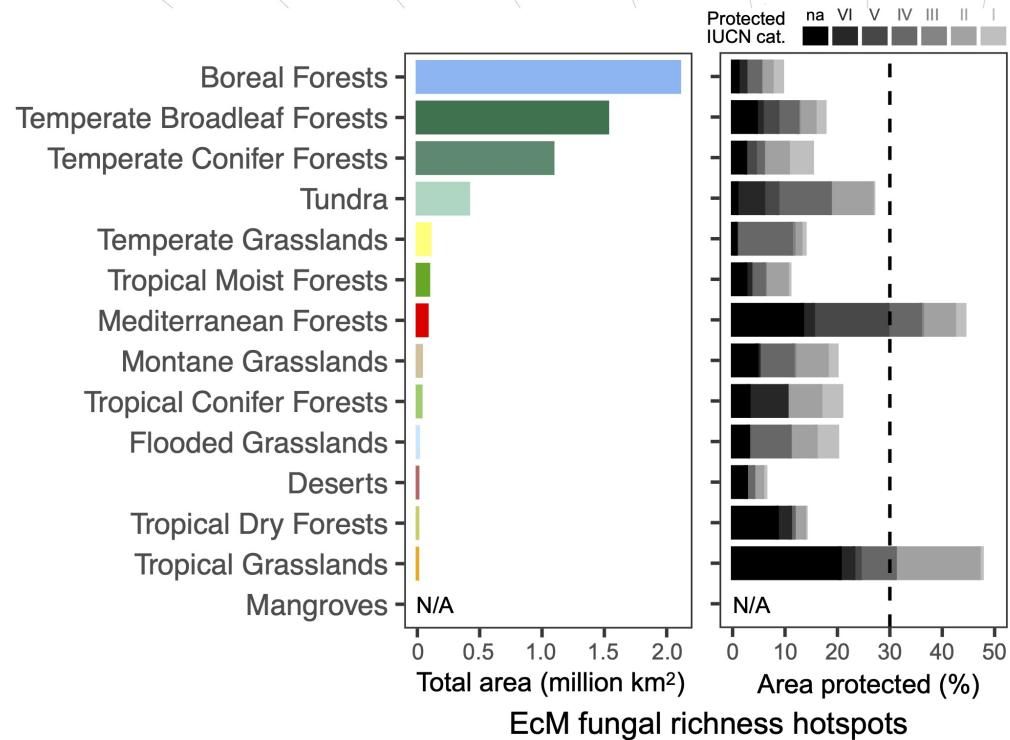
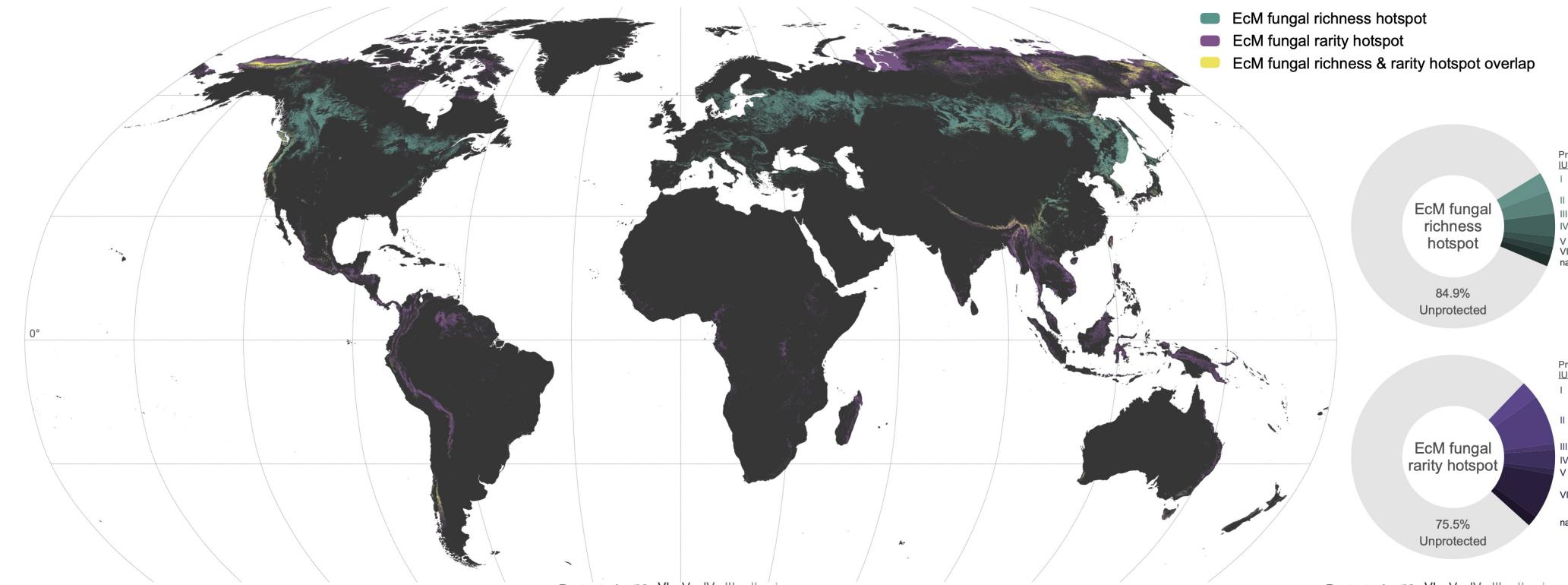
1017 **Figure 6. Visualizing uncertainty and extrapolation of spatial predictions.** Bivariate maps
1018 show the combination of pixel-level uncertainty (orange gradient) and extrapolation (purple
1019 gradient) of **A**) AM fungi predictions and **B**) EcM fungi predictions. Masked locations are shown
1020 in grey. Uncertainty is measured as the coefficient of variation across n=100 bootstrapped model
1021 predictions. Extrapolation reflects the degree of environmental difference and geographic distance
1022 from samples in the training dataset. Histograms on figure legends show the frequency of pixels
1023 in different uncertainty and extrapolation levels visualized here.

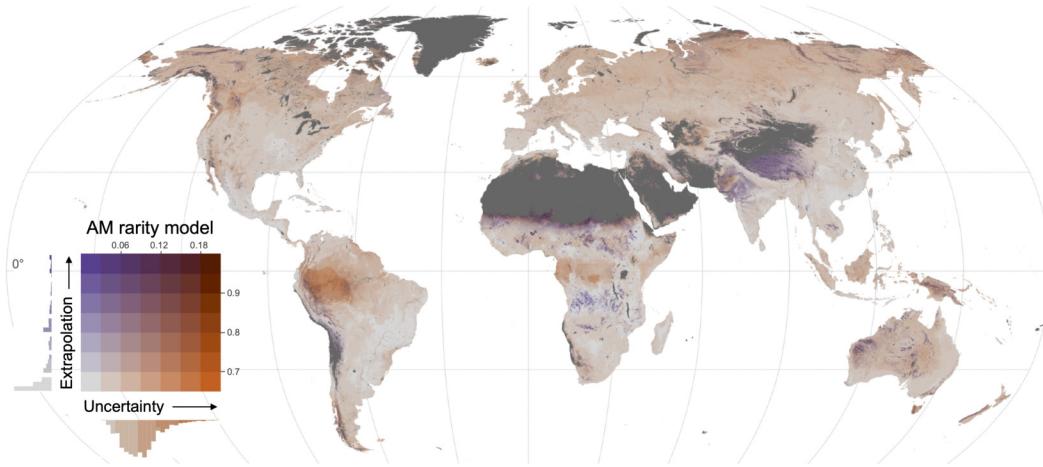
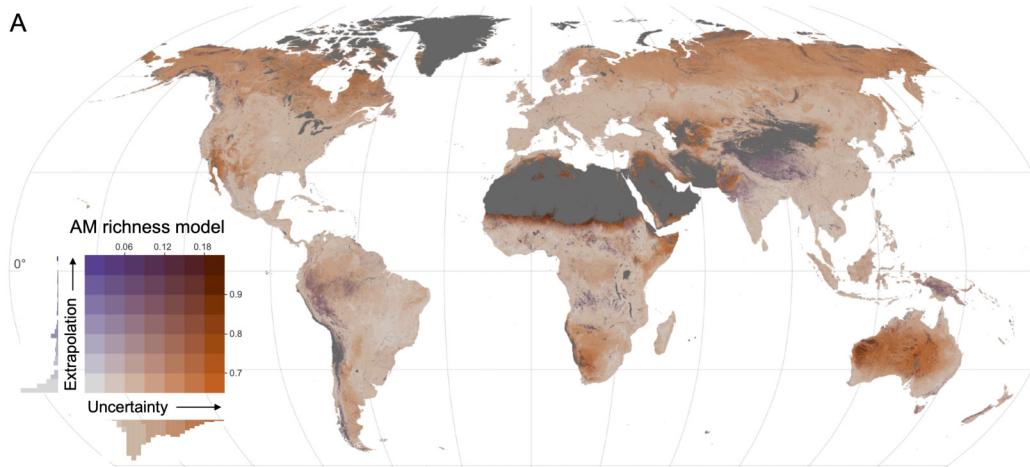


A**B**

A**B**





A**B**