Global mycorrhizal plant distribution linked to terrestrial carbon stocks

Authors: Nadejda A. Soudzilovskaia[1]*, Peter M. van Bodegom[1], César Terrer[2,3], Maarten van't Zelfde[1], Ian McCallum[4], M. Luke McCormack[5], Joshua B. Fisher[6,7], Mark Brundrett[8], Nuno César de Sá[1], Leho Tedersoo[9]

Affiliations:

[1]Institute of Environmental Sciences, Leiden University, 2333 CC Leiden, the Netherlands.

[2]Institut de Ciència i Tecnologia Ambientals (ICTA) Universitat Autonoma de Barcelona, Barcelona, Spain.

[3]Department of Earth System Science, Stanford University, Stanford, CA, USA.

[4]Ecosystems Services and Management Program, International Institute for Applied Systems Analysis, Schlossplatz 1, A-2361, Laxenburg, Austria.

[5]Center for Tree Science, The Morton Arboretum, 4100 Illinois Route 53, Lisle, IL 60532, USA.

[6]Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, CA, 91109, USA

[7]Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles

[8]School of Biological Sciences, Faculty of Science, University of Western Australia, Crawley, 6009 WA, Australia

[9]Natural History Museum and Institute of Ecology and Earth Sciences; University of Tartu; 14a Ravila; 50411 Tartu; Estonia

*Correspondence to: n.a.soudzilovskaia@cml.leidenuniv.nl

Abstract

Vegetation impacts on ecosystem functioning are mediated by mycorrhiza, a plant-fungal association formed by most plant species. Ecosystems dominated by distinct mycorrhizal types differ strongly in their biogeochemistry. Quantitative analyses of mycorrhizal impacts on ecosystem functioning are hindered by the absence of information on mycorrhizal distribution. We present the first global high-resolution maps of vegetation biomass distribution among main types of mycorrhizal associations. Arbuscular, ecto-, ericoid and non-mycorrhizal vegetation store 241±15, 100±17, 7±1.8 and 29 ± 5.5 GT carbon in aboveground biomass, respectively. Soil carbon stocks in both topsoil and subsoil are positively related to the biomass fraction of ectomycorrhizal plants in the community, though the strength of this relationship varies across biomes. We show that human-induced transformations of Earth's ecosystems have reduced ectomycorrhizal vegetation, with potential knock-on effects on terrestrial carbon stocks. Our work provides a benchmark for spatially explicit global quantitative assessments of mycorrhizal impacts on ecosystem functioning and biogeochemical cycles.

Short title: Global distribution of mycorrhizal plants

One Sentence Summary: First maps of the global distribution of mycorrhizal plants reveal global losses of ectomycorrhizal vegetation, and quantitative links between mycorrhizal vegetation patterns and terrestrial carbon stocks.

Mycorrhizas are mutualistic relationships between plants and fungi, in which fungi supply plants with nutrients and plants provide carbon to fungi[1]. Among known mycorrhizal types, arbuscular mycorrhiza (AM), ectomycorrhiza (EcM), and ericoid mycorrhiza (ErM) are geographically the most widespread, colonizing over 85% of vascular plants across vegetated terrestrial biomes[1-4]. Due to the enhancement of plant nutrient acquisition[1] and the large biomass of fungal networks in soil[5], the presence and type of mycorrhiza is among the key determinants of ecosystem functioning[6-9] and biogeochemical cycling[10-13]. Thus, mycorrhizal types of plants likely also affect the global distribution of soil carbon stocks. There is growing

evidence that ecosystems dominated by EcM and ErM vegetation exhibit higher topsoil carbon to nitrogen ratios (C/N) compared to ecosystems dominated by AM plants[11,12,14,15], although in temperate forests the pattern may be reversed in deeper soil layers[17]. The mechanisms driving these differences are heavily debated in current literature, with distinct physiological traits of mycorrhizal fungi most likely playing a critical role[16-20].

Although it can be argued that high abundance of EcM plants is a consequence rather than a driver of high soil C stocks, a large body of recent findings provides evidence that EcM symbionts may be the key drivers of topsoil carbon accumulation through two interacting mechanisms. First, EcM fungi produce greater biomass of more recalcitrant mycelium compared to AM fungi[5]. Second, while EcM fungi are more efficient in taking up N in N-poor soils than AM fungi or roots[10,21], EcM fungi immobilize most of the N in their own biomass. This suppresses saprotrophic decomposition process[22] and reinforces the competitive advantage of EcM and ErM plants via enhanced N limitation[22,23].

A full understanding of the global carbon and nitrogen stocks requires quantitative models on the distribution of mycorrhizal types in ecosystems. Despite the existence of regional maps of current[24,25] and past[26] mycorrhizal vegetation, and on the distribution of mycorrhizal fungal species[27,28] we still lack global information on the distribution of biomass of mycorrhizal plants, which is a much better proxy for mycorrhizal impacts on ecosystem functioning than the biodiversity of mycorrhizal symbionts. While current terrestrial biosphere models simulate feedbacks between the carbon cycle and vegetation distribution[29], such models ignore mycorrhizal types and their effects on nutrient cycling. Integration of such information is a promising way towards a more realistic simulation of carbon and nutrient fluxes associated with plant nutrition[16,18,21] and soil carbon cycles[15,30]. Quantitative models of the distribution of

mycorrhizal vegetation constitute an important missing link between the known effects of mycorrhizas in biogeochemical cycles and their global impacts[31].

Human activities such as forest logging, urbanization, and agricultural practices have altered 50-75% of the Earth's terrestrial ecosystems[32], transforming areas with previously natural EcM and ErM vegetation into AM and non-mycorrhizal (NM) vegetation. So far, the impact of anthropogenic land use shifts on biogeochemical cycles associated with mycorrhiza have remained poorly known due to the lack of appropriate spatial information.

Based on a comprehensive quantitative evaluation of plant-mycorrhizal associations and the distribution of vascular plant species across biomes and continents, we assembled the first high-resolution digital maps of the global distribution of biomass fractions of AM, EcM, ErM, and NM plants. Building on these maps, we assessed (i) the amount of aboveground biomass carbon currently stored in each type of mycorrhizal vegetation, (ii) the impact of conversion of natural ecosystems to croplands on the distribution of mycorrhizal types globally, and (iii) the difference in relationships between relative abundance of AM and EcM plants in an ecosystem and ecosystem soil carbon content in topsoil (0-20cm), medium (20-60) and deep (60-100 cm) subsoil layers.

Assembly of mycorrhizal vegetation maps

To generate global maps of mycorrhizal vegetation, we estimated biomass fractions of AM, EcM, ErM and NM plants within each combination of continent x ecoregion x land cover type. Extended Data Figure 1 illustrates the data assembly processes for the maps. Ecoregions follow Bailey[33] (Supplementary table 1) and land cover types were retrieved from the ESA CCI land cover map[34], which specifies cover and biomass fractions of trees, shrubs and

herbaceous plants (Supplementary Table 2). For each combination, we determined the dominant species or group of species from 1,568 vegetation surveys (Supplementary Table 3). For these species, we determined mycorrhizal type using the recently published FungalRoot database[30] (Supplementary Table 4). Integrating these data, we obtained mycorrhizal plant biomass fractions of AM, EcM, ErM and NM plants for each combination of Bailey ecoregion, continent, and land cover type (Supplementary Tables 5 and 6). These fractions were overlain on a global grid.

Our maps (Figure 1) provide quantitative estimates of the distribution of aboveground biomass fractions among AM, EcM and ErM plants within areal units of 10 arcmin. The use of a detailed map of ecoregions[33] provides much greater resolution compared to the biome-based patterns of mycorrhizal distributions reported by Read[3] >25 years ago, whereas the land cover map[34] enabled us to provide accurate spatial positioning of ecosystem boundaries based on satellite-derived data, explicitly taking into account human-driven transformations of vegetation.

We validated the map data using four independent datasets: (i) forest biomass structure for Eurasia[35], (ii) a global dataset of forest biomass structure used for an analysis of mycorrhizal impacts on carbon *vs* nitrogen dynamics[19], (iii) estimates of mycorrhizal associations in the USA based on remote sensing[36], and (iv) West Australian map of mycorrhizal root abundance[24] (Extended Data Figure 2). This validation revealed that the vast majority of the data (87% of the AM data points and 89% of the EcM data points) deviate by <25% from the measurements[19, 35,36], when excluding ESA land use classes comprising poorly resolved combinations (i.e. mixed classes of land cover types[34], such as “Tree cover, broadleaved, evergreen, closed to open (>15%)”, see methods for details) that were difficult to couple to

our classification scheme. The relationship between the validation data and our estimates is shown in Extended Data Figure 3.

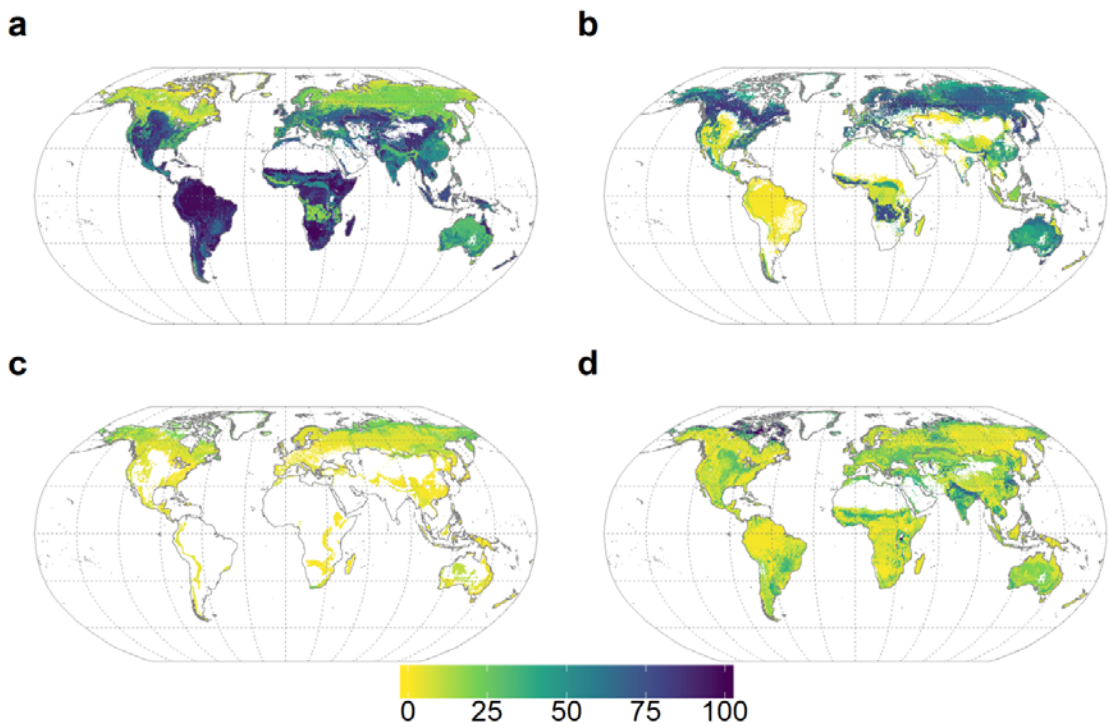


Figure 1. Percentage of aboveground plant biomass of a - arbuscular mycorrhizal plants, b - ectomycorrhizal plants, c - ericoid mycorrhizal plants, and d -non-mycorrhizal plants. The map resolution is 10 arcmin. See Extended Data Figure 4 for associated uncertainty values.

Our maps of mycorrhizal vegetation were assembled based on multiple published datasets, using a number of conversion factors to obtain per pixel values of mycorrhizal plants biomass fractions. These conversions as well as the fact that the plant species distribution data (Supplementary Table 4) originates from multiple sources constitute important uncertainty sources in our dataset. We examined the uncertainty of our maps based on uncertainties of tree, shrubs and herbaceous plant fractions within the land cover types[37], and the number of

data sources used to assess mycorrhizal fractions of plant biomass within each combination of Bailey ecoregion x continent; see methods for details. Extended Data Figure 4 shows spatial distribution of map uncertainties. The mean uncertainties of AM, EM, ER, NM maps are 19.6, 17.6, 14.6, and 15.0% at the 90% confidence interval. Overall, tropical areas have the highest uncertainties of the mycorrhizal fraction data, reaching 50% (AM). in the Amazon region. Therefore, our maps should be used with caution for these areas. Future sampling efforts of mycorrhizal vegetation distribution should be more focussed on tropical areas of Asia, Africa and South America.

Mycorrhizal vegetation and aboveground carbon stocks

By linking our maps of mycorrhizal vegetation to satellite observations of global aboveground biomass carbon[38], we estimated the amount of aboveground biomass carbon stored in arbuscular, ecto-, ericoid and non-mycorrhizal vegetation as 241±15, 100±17, 7±1.8 and 29 ± 5.5 GT (mean values ± uncertainty at 90% confidence interval; Figure 2). In this analysis, the data were scaled to a resolution of 15 arcmin to match biomass estimates[38]. Most of the aboveground carbon stock stored in arbuscular mycorrhizal vegetation is situated in tropical forests (Figure 2a). Extended Data Table 1 shows per biome distribution of the carbon stocks among mycorrhizal types.

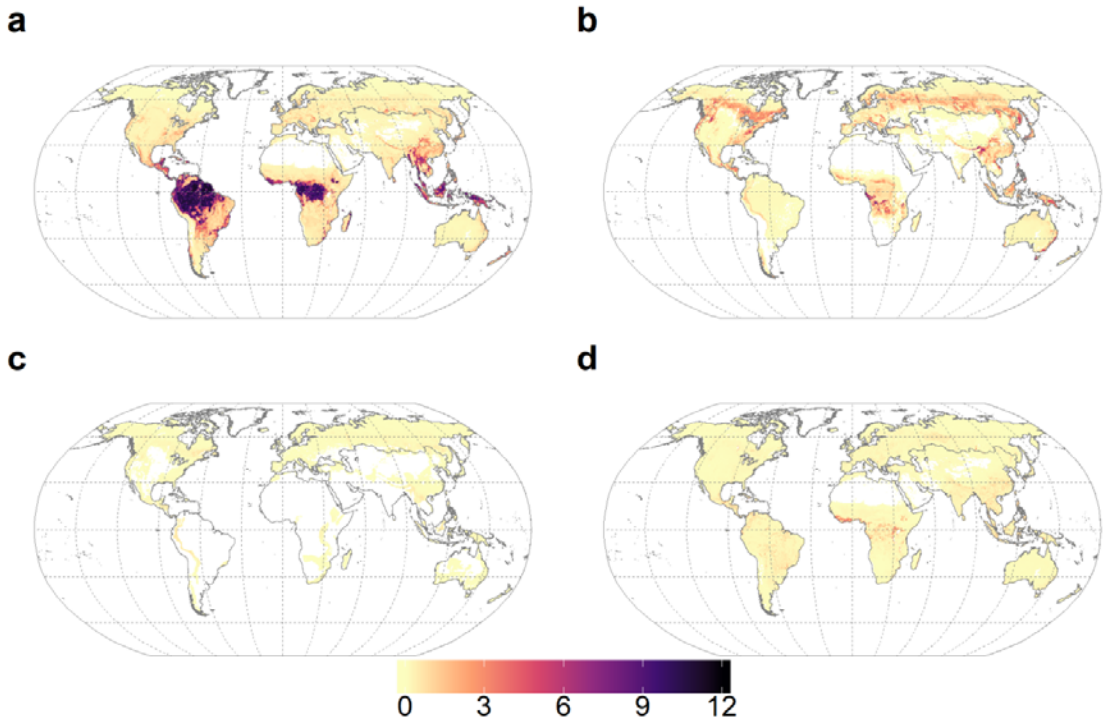


Figure 2. Amount of carbon (Mt per grid cell of 15 arcmin) stored in plant biomass in vegetation of different mycorrhizal types: (a) arbuscular mycorrhizal plants, (b) ectomycorrhizal plants, (c) ericoid mycorrhizal plants, (d) non-mycorrhizal plants. The amount of aboveground biomass carbon stored in arbuscular, ecto-, ericoid and non-mycorrhizal vegetation as 241±15, 100±17, 7±1.8 and 29 ± 5.5 GT (mean values ± uncertainty at 90% confidence interval)

Impacts of human-induced land transformations on distribution patterns of mycorrhizal vegetation

Agricultural practices drive the replacement of natural vegetation by facultatively AM crops[1,39], which could also be *de-facto* non-mycorrhizal due to destruction of hyphal networks by ploughing and excess fertilisation[40,41]. Using past vegetation estimates, Swaty and co-workers[26] showed that across conterminous USA, agriculture has reduced the relative abundance of ectomycorrhizal plants compared to other mycorrhizal types. However, global quantifications of agricultural impacts on distribution of mycorrhizas have not been possible

until now. Based on the current land use data underlying our maps (Supplementary Table 4), we assessed mycorrhizal distributions on Earth in the absence of croplands. For each ecoregion-continent-land cover combination that contained croplands, we replaced current biomass fractions by estimates of per-grid cell biomass fractions in AM, EcM, ErM and NM plants that would be *expected* at these locations based on natural vegetation types (see Methods for details, and Supplementary Tables 7-8 for data). Based on these data, we generated maps presenting *potential natural* distributions of biomass fraction of AM, EcM, ErM, and NM plants in a cropland-free world (Extended Data Figures 6 and 7). The current biomass fractions of AM plants have increased in Europe, parts of Asia, and North America, but declined in Africa, Asia (mostly India) and South America, coinciding with increase in non-mycorrhizal vegetation (Figure 3, Extended Data Figure 8). Our analysis suggests that EcM biomass has declined in all continents, primarily due to a replacement of natural forests by agricultural lands, whereas ErM biomass has remained unchanged.

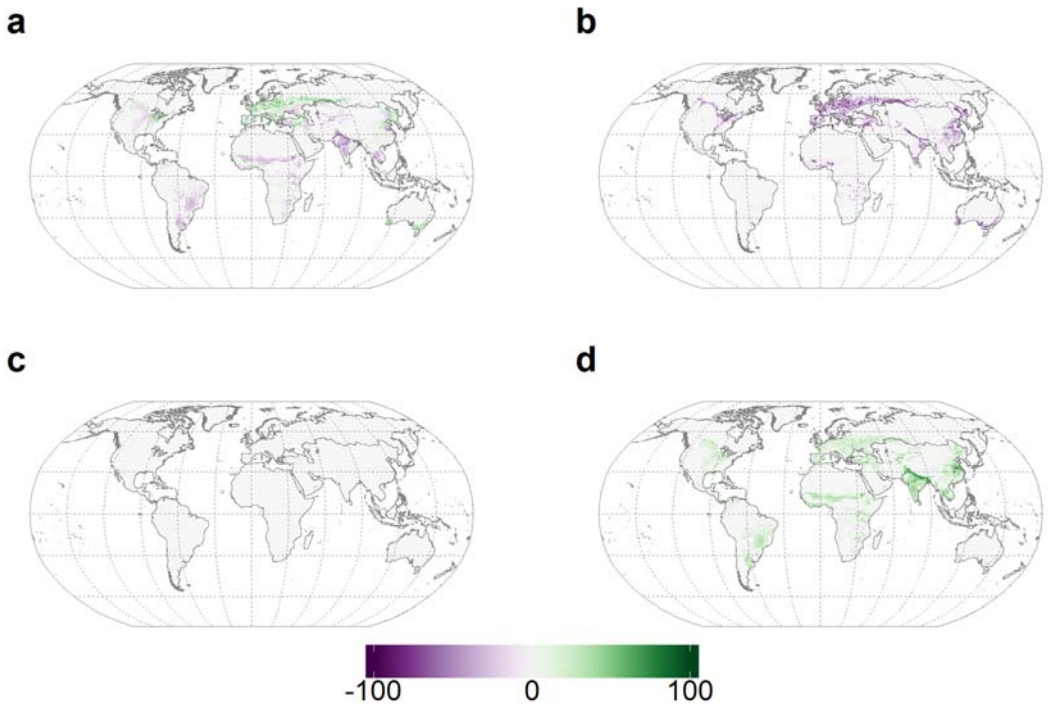


Figure 3. Changes in biomass fractions of mycorrhizal vegetation induced by crop cultivation and pastures. a - arbuscular mycorrhizal plants, b - ectomycorrhizal plants, c - ericoid mycorrhizal plants, d - non-mycorrhizal plants. Purple colors indicate losses, green colors indicate gains. Uncertainties are shown in Extended Data Figure 7.

Relationships between biomass fractions of mycorrhizal types and soil carbon stocks

Recent field research in the US temperate forests suggests that soil carbon content increases with increasing EcM abundance in top soil layers, but, depending on forest type, this relationship may be reversed in deeper soil[17,42]. This is in agreement with the Microbial Efficiency-Matrix Stabilization hypothesis, which predicts that ecosystems with rapid decomposition, such as most AM-dominated forests[19], enhance soil organic matter (SOM) stabilization by accelerating the production and deposition of microbial residues[43,44,45].

To separate the effects of biome and mycorrhizal type on soil C on a global scale, we modelled the relationships between soil carbon content, biome type[46], and biomass fractions of AM and EcM plants. We did not analyse the relationship between ErM plant cover and soil C content, due to a small proportion of the ErM plant biomass in the majority of ecosystems. We conducted separate analyses for the topsoil (uppermost 20 cm soil layer) and subsoil (20-60, and 60-100cm soil layers), as obtained from the ISRIC-WISE Soil Property Databases, at a resolution of 30 arcsec[47]. The data sources used for these analyses are independent: the ecoregion classification and hence mycorrhizal type distribution does not account for edaphic parameters, whereas our soil C data is unrelated to that of vegetation. Model comparisons were based on the Akaike information criterion (AIC). The relative importance of each predictor was examined using the Lindemann-Merenda-Gold (LMG) metric, providing the fraction of variance explained by each predictor, within the total variance explained by the model.

Our global analysis revealed a moderately strong positive relationship between the biomass fractions of EcM plants and both topsoil and subsoil carbon (Figure 5, Extended Data Figure 8). Consistent with the current paradigm[45], cf[11], our analysis revealed that biome is the main predictor of soil carbon stocks. Even so, the grid cell biomass fraction of EcM plants still accounted for one third of the explained variation in both topsoil carbon and in subsoil (Table 1). The interaction between biome and EcM biomass fractions was significant ($P<0.001$) but only marginally important (LMG=1%), suggesting that the increase in topsoil carbon along with an increase in EcM plant biomass is mostly independent from the environment (Table 1, Figure 5, Extended Data Figures 8,9, Extended Data Table 2). The total aboveground carbon stock stored in the EcM plants was a relatively worse predictor of soil carbon stocks compared to the EcM relative biomass fraction, showing in all cases higher AIC and lower R^2.

In contrast to EcM, AM biomass fractions per grid cell showed negative but inconsistent relationships to soil carbon, with a contrasting positive trend in tundra (Figure 5, Extended Data Figure 10). The latter relationship explained only a small amount of variance in tundra soil carbon stocks, (12% in top 0-20cm, 1.7% in 20-60 cm layer and 0.2% in 60-100 cm layer) and could arise from less accurate data of soil C in tundra[47] due to local landscape heterogeneity and prevalence of facultative AM plants with no or low mycorrhizal colonization[48].

Our analysis of the relationships between mycorrhizal type and soil carbon is based on maps, which feature large uncertainties. Analyses of relationships between ISRIC-WISE predicted soil carbon and the original data that were used to generate the ISRIC soil map yield R^2-values in the range of 0.4-0.6 [49,50]. This uncertainty adds ambiguity to our analysis, and reduces the reliability of quantitative estimates of the relationships between EcM plant biomass fractions and soil C. However these uncertainties equally apply to the analysis of AM vs soil C as well as to that of the EcM vs soil C, due to the fact that the analysis is based on the same geographical data points. Therefore, we consider that the high uncertainty of the ISRIC-WISE soil data is unlikely to affect the qualitative nature of our conclusion that AM and EcM vegetation differently relate to soil C.

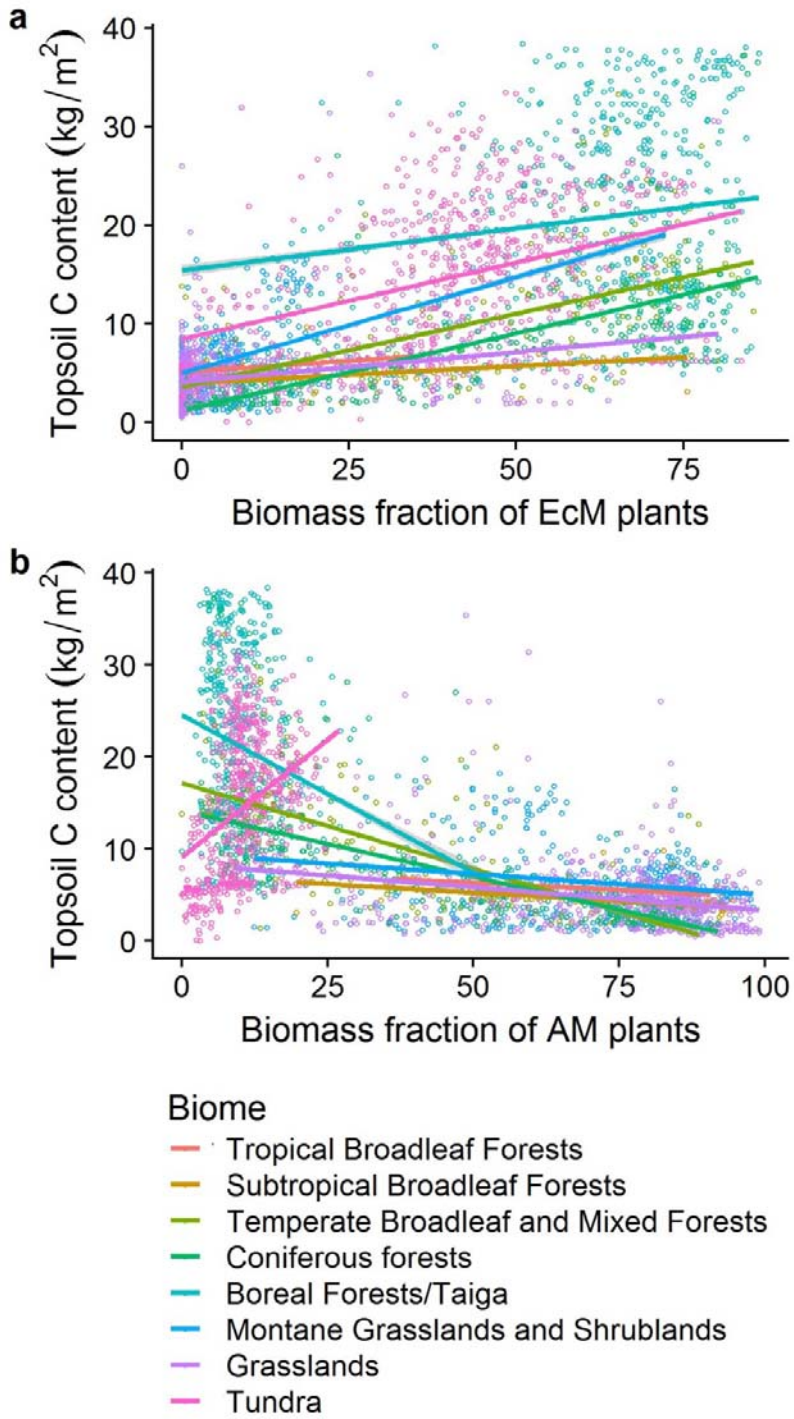


Figure 4. Quantitative relationships between topsoil (0-20cm) C and biomass fraction of a – EcM plants and b - AM plants in natural ecosystems. The outcomes of individual models are presented in the Supplementary Table 2. Croplands were excluded from the analysis. Per-biome predictions are shown in different colours.

Implications

Our work provides the first quantitative estimates of the global biomass distribution of arbuscular, ecto-, ericoid and non-mycorrhizal plants, accounting for human-induced transformation of habitats. Previous research has shed light onto the distribution patterns of mycorrhizal plant and fungal species richness [27,28,51], and onto low-resolution (1 arcdegree) distribution patterns of mycorrhizal trees. In contrast, our maps directly reflect the global distribution of biomass fractions of mycorrhizal plants across all biomes and all main vegetation types. Availability of such data at high resolution of 10 arcmin provides an opportunity for multiple potential analyses aimed at unraveling mycorrhizal impacts on ecosystem functioning at large geographical scales. To the best of our knowledge, this study for the first time provides a global quantification of mycorrhiza type-soil C relationships based on field (as opposed to modelled) data.

Our maps were derived at spatial resolutions allowing identification of the global patterns of mycorrhizal distributions and are most appropriate for global and large-geographical scale analyses of mycorrhizal impacts on ecosystem functioning and global drivers thereof. Recent estimates suggest that the total soil carbon loss due to agricultural practices accounts for 133 GT, with great acceleration of losses during the past 200 years[52]. Our analyses accounting for pre-agricultural patterns in EcM plant distribution point to large-scale losses of ectomycorrhizal vegetation, with potentially strong effect on the amount of C stored in soils. Analyses of agricultural impacts presented in this paper are based on the assumption that these impacts are limited to shifts in plant species composition, and do not encompass shifts in soil water and nutrient availability, which could affect activity of mycorrhizal fungi, depth distribution of mycorrhizas in the soil, and shifts among mycorrhizal fungal species composition due to an introduction of exotic species. While such simplifications are necessary

for the analyses reported in this paper, they should be considered when interpreting our results. Furthermore, our analyses do not address other human impacts that can lead to shifts among AM and EcM vegetation, such as climate change, introduction of invasive species and nitrogen deposition. The latter is known to be an especially important driver of mycorrhizal vegetation shifts[42,53], as it negatively affects abundance of ectomycorrhizal fungi in soil[54]. Given that nitrogen deposition leads to replacement of ectomycorrhizal plants by arbuscular and non-mycorrhizal vegetation, it further enhances agricultural impacts on soil carbon losses.

The question whether increased domination of EcM plants in an ecosystem is associated with higher soil carbon content across both top- and subsoil is heavily debated[11,12,17-20,31]. However the studies thereof are based on a limited number of observations[11] or regional-scale analyses[11,12,17,53,55]. Our analysis shows that across large geographical scales, higher cover of EcM vegetation is broadly associated with greater soil C stocks in both topsoil and subsoil, while AM vegetation has more variable, weaker and mostly negative relationships. This analysis does not provide evidence of causality of this relationship as multiple environmental variables such as climate, soil nutrients, especially nitrogen availability, and soil texture may affect both soil carbon and mycorrhizal plant distributions.

Nonetheless, our study has established a quantitative predictive framework relation between the dominance of mycorrhizal types and soil C stocks. Complete and directional understanding of the complex nature of the hierarchy of environmental drivers controlling soil carbon patterns requires further detailed (experimental) investigations of the hierarchy of different predictors and importance of local edaphic variables.

Our estimates of the carbon stocks in AM and EcM aboveground biomass together with the quantitative relationships between soil carbon stocks and AM or EcM plant dominance in ecosystems provide new qualitative insights into the global carbon cycle, highlighting the substantial role of mycorrhizas therein. Tropical forests, usually dominated by AM symbiosis (Figure 2), contain 162 GT (44%) of global aboveground biomass[38], whereas the predominately EcM temperate and boreal forests altogether store only 21% of global aboveground biomass carbon[38], indicating that contribution of EcM vegetation to the aboveground biomass carbon is relatively small. In contrast, belowground carbon stocks are positively correlated to the proportion of EcM plant biomass, suggesting that mycorrhizal contribution to large carbon stocks in these regions occurs primarily through the carbon supply to belowground organs and mycorrhizal fungi, which is further emphasized with slowed decomposition processes[22]. Furthermore, our analyses revealed a relatively stronger relationship between soil C and EcM plant biomass fraction (%) than between soil C and the amount of carbon stored in EcM plants. These findings suggest that belowground carbon allocation by plants through mycorrhizal pathways is not directly proportional to the aboveground plant biomass, supporting the view of the importance of mutualism-parasitism trade-off in ectomycorrhizal associations[56] across biomes.

Taken together, this study provides a benchmark for relating ecosystem processes to the functioning of distinct types of mycorrhizas on a global scale. So far, quantitative global information about mycorrhizal distribution was virtually absent in spite of high demand for such data [12,13,31]. In spite of some uncertainty, our mycorrhizal distribution maps provide an essential source for systematic analyses of mycorrhizal biogeography and the environmental drivers thereof. Because our maps are based on field data and not on modelling of a vegetation-to-environment relationships, they provide independent data for examining drivers

of mycorrhizal distributions and their impacts on ecosystem functioning, without introducing a circular reasoning caused by the use of environmental data. Inclusion of mycorrhizal distribution into vegetation models allows testing hypotheses about mycorrhizal impacts on ecosystem functioning and related ecosystem services. Our integrated models enable quantifying relationships between mycorrhizal abundances in ecosystems as well as soil and vegetation carbon content in global-scale analyses of biogeochemical cycles. In particular, the results of our study suggest that restoration of native vegetation especially in abandoned agricultural and barren land may help alleviate anthropogenic soil carbon losses and ameliorate increase in atmospheric greenhouse gases.

Methods

Construction of mycorrhizal vegetation maps

Data sources

To construct the maps of mycorrhizal biomass fraction distribution, we integrated (1) the data about dominant plant species and their growth forms within each continent x ecoregions x land cover type combination, (2) data about mycorrhizal type of these species, and (3) data about cover and biomass fractions of trees, shrubs and herbaceous plants within individual land cover types. Here we defined a species or a set of species with a similar mycorrhizal strategy and growth form as 'dominant' if it constituted >20% of vegetation biomass. Extended Data Figure 1 shows a flowchart of the map assembly process.

We selected the global ecoregion map of Bailey (2014) [33] with 98 ecoregions (Supplementary Table 1), provided by the Oak Ridge National Laboratory Distributed Active Archive Center[57] (spatial resolution 10 arcmin), as a basis for mapping of global-scale distribution of mycorrhizal types. This map was preferred over that of biomes provided by Olson and co-workers[46] and others because of higher level of detail for ecoregions and because the boundaries of ecoregions were more strongly related to the distribution of mycorrhizal types. Ecoregions spanning across multiple continents were considered separately for each continent. We used the continent division based upon the FAO Global Administrative Unit Layers (http://www.fao.org/geonetwork/srv/en/).

We determined the land cover types present in each ecoregion using a satellite observation-based map for the year 2015 generated by the European Space Agency [34]. This map includes 38 land cover categories such as croplands, urban areas, grasslands and forests of various types, with a spatial resolution of 300 m (Supplementary Table 2).

Based on vegetation surveys assigned to ecoregions (1,568 data sources; Supplementary Table 3) we determined the dominant plant species and their growth form for each continent x ecoregion x land cover type combination. We used relative abundances of plant species averaged across different data points, as available across vegetation surveys, with equal weight to all observations.

We assigned mycorrhizal type (AM, EcM, ErM, NM) to each dominant species using the FungalRoot database[30]. Species with dual EcM-AM and AM-NM categories were allocated to both types equally (50% weight). Mycorrhizal status of species with no empirical records were extrapolated from congeneric and confamilial species. Therefore, all Diapensiaceae and Ericaceae species were considered ErM[67], except for *Enkianthus* (AM), *Arbuteae*, *Pyroleae*, *Monotropeae* and *Pterosporeae* (all subtypes of EcM). Because of multiple incorrect reports and alternative definitions for EcM, we took a conservative approach by considering plants to be EcM only when this was supported by multiple independent studies and the proportion of conflicting reports was <50%[68]. Although most crop plants are able to form arbuscular mycorrhizas, intensive agricultural practices and breeding may lead to reduction or loss of mycorrhizal infection [40,41]. Therefore, rain fed and flooded croplands were considered to feature partly AM and partly NM vegetation (Supplementary Table 5, 7), unless data indicating dominance of NM vegetation was available. Crop species that belong to *Brassicaceae* family were considered NM.

<u>Biomass of mycorrhizal types</u>

Combining the prevailing dominant plant species, their growth form and mycorrhizal type, we estimated the biomass proportions of EcM, AM, ErM and NM vegetation in each ecoregion by continent by land cover combination (Supplementary Tables 5 and 6). We considered that

in forests with a sparse understorey, trees contribute 90-95% of the biomass and the understorey accounts for 5-10% of biomass[36,58-61]. In forests with a dense layer of shrubs, trees, shrubs and herbs/dwarf shrubs contribute 70±15%, 20±10% and 10±5% to plant biomass[36,58-61], respectively. In shrublands, we consider shrubs and herbs to account for 90±5% and 10±5% of biomass, respectively [35,58-61]. We considered that woodlands harbour 30±10% of biomass in trees, 30±10% of biomass in shrubs, and the remaining biomass in herbaceous vegetation[35, 58-61]. This resulted in biomass proportions of each mycorrhizal type in continent x ecoregion x land cover type (Tables S5, S6).

As we focused on the biomass of mycorrhizal plants and not on species diversity; we did not attempt to map the distribution of orchid mycorrhiza. Orchid species are never abundant in ecosystems in terms of biomass, and they are therefore unlikely to play an essential role in biogeochemical cycles at large regional scales.

Assembly of raster maps

We generated raster maps based on the proportional mycorrhizal type biomass data. We overlaid the raster map of Bailey ecoregions (10 arcmin resolution)[57] with the raster of ESA CCI land cover data (300 m resolution)[34], which we converted to 10 arcmin using a nearest neighbour approach. The resulting raster was overlain with the polygon map of continents, rasterized at 10 arcmin. To each pixel, we assigned the corresponding mycorrhizal type proportions, considering the prevailing combination of Bailey ecoregion x land cover in each continent. Because Some certain ecoregions covered multiple isolated parts of continents and proportions of mycorrhizal type distribution differed in these regional areas by >2-fold (Supplementary Tables 2,3), we split these ecoregions into two or more subregions using the ArcGIS 10.2.2 Raster Calculator.

Impact of croplands

We estimated the effect of agriculture on mycorrhizal type distribution by substituting the proportions of mycorrhizal types in croplands (landcover types 10,11,12,20 and 30; see Supplementary Table 3) by estimates of mycorrhizal type proportions from natural land cover types in these ecoregions. For ecoregions naturally harbouring more than one vegetation type (e.g., grasslands, shrublands and forests), we considered that the *expected* vegetation represents a mixture of these land cover types. This resulted in an additional dataset describing combinations of vegetation as defined by ecoregion x continent x land cover without croplands (Supplementary Table 7 for continent data and Supplementary Table 8 for island data). Maps of potential mycorrhizal type distribution in a cropland-free world (Extended Data Figure 4) were created based on this dataset, following the aforementioned procedures.

In this analysis, we did not consider forest plantations to be croplands. Therefore, changes in land cover induced by forest restoration through, for instance pine plantations or eucalypt plantations, are not addressed as vegetation changes induced by cropland cultivation. The total area occupied by AM tree plantations exceeds that of EcM tree plantations [62,63] (Extended Data Table 3), suggesting that exclusion of tree plantations leads to conservative estimates of the global reduction of EcM vegetation.

Map validation

The maps of the current distributions of mycorrhizal biomass fractions were validated using the datasets of forest biomass structure for Eurasia[35], global analysis of impacts on mycorrhizas on carbon vs nitrogen dynamics[19], the USA-based analysis of mycorrhizal

associations conducted with remote sensing techniques[36], and the map of mycorrhizal root biomass in West Australia by Brundrett [24] (see Extended Data Figure 2).

The data of forest biomass structure for Eurasia[35] provide information on per-plot tree species abundances for a large number of European sites. As the data contain all records obtained since the 19th century, we used only the data recorded after 1999. Using our database of plant species and associated mycorrhizal types we assigned every tree species with its mycorrhizal type (1344 data points, Extended Data Figure 2). This provided us with a per-site data of the relative biomass of AM and EcM trees. We used these data as proxies for AM and EcM biomass fractions to compare with the data in our maps. We used the same approach for the data of Lin and co-workers[19], which represent plot-based records of vegetation structure for 100 sites across the globe accompanied with data about plant mycorrhizal associations.

The dataset of Fisher *et al* [36] provides the relative cover of AM and EcM plants from Landsat scenes centered on four sites in USA: Lilly-Dickey Woods (Indiana), long-term research site of Smithsonian Conservation Biology Institute (Virginia), Tyson Research Center Plot (Missouri), and a long-term research site of Wabikon Forest Dynamics (Wisconsin). Given that the dataset comprises forested areas only, we considered areal coverage of AM and EcM plants provides a good estimate of AM and EcM plant biomass. Using this dataset, we directly compared the AM and EcM coverage per pixel to the data of our maps.

In the datasets [19,35,36] biomass and areal fractions of AM and EcM plants are always considered to sum up to 100%. As these datasets do not provide information about non-mycorrhizal and ericoid plants, we estimated these as 5-10% and 0-20%, respectively, depending on the dominant tree association, and accordingly reduced the values of AM and

EM biomass fractions in the validation calculations. While we considered this approach to be acceptable for the validation of AM and EcM data, the data quality is not high enough to validate the NM and ErM maps.

As the thee datasets[19,35,36]represent forest data, we evaluated whether all data points or raster cells[36] were indeed located in forest areas. This was done using the ESA land cover categories data[34]. All data points that were located out of the current areas registers by ESA [34] as forests were excluded from the analysis.

The West Australian map of mycorrhizal root abundance[24] provides information about the percentage of total biomass of plant roots featuring AM and EM root colonization and about the percentage of non-mycorrhizal root biomass. We used this data as a proxy for biomass fractions of AM, EcM and NM plants. In order to quantify the differences between the validation datasets and our maps, we have calculated the Mean Averaged Error (MAE) of the difference between our maps and validation datasets. MAE expresses the deviation between two spatial data sets from the 1:1 line. For AM, EcM and NM vegetation fractions MAE is 18.7%, for EcM it is 13.6%, for NM it is 4.7%, respectively. Due to a virtual absence of information about distribution of solely ErM vegetation, direct validation of the ErM maps was impossible."

To further assess the uncertainties in the maps, we examined which land use classes represent the data points that deviate from the observed data by more than 25 percent units. Our analysis showed that the large proportion of those deviations (60% for EcM and 40% for AM) fall into those land use classes that represent a poorly described mixture of evergreen or mixed forests and grasslands, i.e. ESA classes described as various forms of "closed to open

(>15%) forest" (Supplementary Table 3). Further improvement of the ESA classification data will provide a possibility to improve precision of our maps.

Due to the rarity of datasets on field-examined mycorrhizal vegetation distributions at large special scale we had to validate our datasets used available data on plant species distribution and to accept a number of assumptions and/or recalculations in order to make the data comparable. These adjustments may have affected the quality of the validation dataset and therefore the validation.

Maps spatial uncertainty analysis

We quantified the uncertainty in our maps of mycorrhizal vegetation fractions by applying the error propagation rules to the formulas used to calculate the biomass fractions of mycorrhizal plants per gridcell. Hereto we used the data provided by 35, [REF CCI 1and 2, and 35,58-61] to estimate the uncertainty associated with relative biomass of trees, shrubs, and herbaceous vegetation in each CCI land cover class. The uncertainty in the proportion of each mycorrhizal type in a given Bailey x continent combination was set to $1/\sqrt{n}$, where $n=b+c/3+g/20$. In this formula b is a number of literature sources describing the vegetation composition in a given Bailey ecoregion x continent combination (Supplementary Table 3), while c and g are the numbers of literature sources describing vegetation composition at continent level and global levels, respectively. These latter sources were given less weight, because of their general orientation, though these sources were used only if they were providing information relevant for the combination of Bailey ecoregion x continent under consideration. This procedure was applied to the maps of the current distribution of mycorrhizal fractions (Figure 1) and for the distribution of mycorrhizal fractions in the cropland-free world (Extended Data Figure 5). The

resulting maps of uncertainties are shown in Extended Data Figures 4 and 6, respectively. We calculated the uncertainties in the estimations of changes in biomass fractions of mycorrhizal vegetation induced by crop cultivation and pastures (Figure 3) by applying the error propagation rule for the mathematical operation of subtraction. The resulting uncertainty map is shown in Extended Data Figure 7.

Estimations of carbon stocks in the aboveground biomass of AM, EcM, ErM and NM plants.

In order to estimate the amount of carbon stored globally in the biomass of plants that belong to different mycorrhizal types, we multiplied the mycorrhizal type biomass fractions by the data of the global distribution of carbon stored in aboveground plant biomass obtained from passive microwave-based satellite observations[38]. As these data have resolution of 15 arcmin, we converted our data of mycorrhizal biomass fractions to this resolutionTo calculate the total amount of carbon stored in AM, EcM, ErM and NM plants (illustrated in Figure 2, and reported per biome in the Extended Table 1), we summed the data on carbon stocks per unit area globally and per biome[46].

To assess uncertainty of estimations of carbon storage in mycorrhizal vegetation we used the rule of uncertainty propagation through the operation of multiplication. Thereto we used the uncertainty of mycorrhizal biomass fraction data and the per biome uncertainty of biomass carbon data, as provided by Liu et al[38] . Given that the data of carbon stored in vegetation of different mycorrhizal types is a product of two datasets, each featuring a high uncertainty[38], the values of aboveground biomass carbon stored in arbuscular, ecto-, ericoid and non-mycorrhizal vegetation are locally associated with large uncertainties (e.g. Supplementary Table 9). Therefore, we recommend to use these data exclusively for large-scale estimates of

biomass carbon storage. However, following the central limit theorem, the uncertainties of the total amount of carbon stored in vegetation of each mycorrhizal type are much lower: 15, 17, 1.8 and 5.5 GT for AM, EcM, ErM and NM vegetation, respectively.

Statistical analysis of the relationship between soil carbon content and biomass distribution of mycorrhizal types

We examined whether soil carbon content was related to the biomass fractions of AM and EcM plants, using generalized linear model regressions, with topsoil (0-20 cm) or subsoil (20-60 cm and 60-100 cm) C content per m^2 as a response variable, and biome and ecosystem biomass fraction of AM or EcM plants as predictors. The data for soil carbon content in the top 20 cm of soil were obtained from the ISRIC WISE Soil Property Databases at a resolution of 30 arcsec[47]. As we were interested in relationships between AM or EcM coverage and soil carbon in natural vegetated environments, we excluded urban and agricultural areas, lakes and "Rock and Ice" areas according to the ESA land cover categories[34], from the analysis. We also excluded wetlands, inundated areas, and extremely dry regions (deserts and Mediterranean areas), because we considered that harsh abiotic conditions instead of biotic interactions are likely to shape biogeochemical cycles in these areas; and we excluded five land cover categories for which our maps showed higher uncertainties, i.e. those where the extent of forest vs grassland cover was unclear. Those land cover categories included "Tree cover, broadleaved, evergreen, closed to open (>15%)", "Tree cover, broadleaved, deciduous, closed to open (>15%)", "Tree cover, needleleaved, evergreen, closed to open (>15%)", "Tree cover, needleleaved, deciduous, closed to open (>15%)" and "Sparse vegetation (tree, shrub, herbaceous cover)".

Data for biome types were obtained from the map of terrestrial biomes[46]. To create a more balanced dataset, we combined all natural grasslands into one single category and all coniferous forests into another single category. To minimize the impacts of unprecise estimations of biome borders we excluded from forest biomes areas that we recognized as grasslands according to the ESA land cover data[34]. Similarly, we excluded from grassland biome areas that were fully covered by forests and shrubs according to[34]. However, given that AM-EcM-NM interactions are likely to relate to C stocks in the areas featuring forest-grassland and forest-tundra mosaics, we kept such areas in the dataset. As soil carbon content data is known to feature high uncertainties[47], we opted to run the analyses at a resolution of 1 arcdegree, which allows grasping large-scale tendencies, while reducing the problem of P-value fallacy[69].

Our models comprised topsoil and subsoil carbon content as a response variable. As predictors, we have examined the following combinations of data: (i) biome, biomass fraction of EcM plants, and their interactions, (ii) biome, biomass fraction of AM plants, and their interactions, (iii) biome, amount of carbon stored in the aboveground biomass of EcM plants (product of our EcM map and[38]), and their interactions, and (iv) biome, amount of carbon stored in the aboveground biomass of AM plants (product of our AM map and [38]), and their interactions. All models were examined independently and evaluated based on the Akaike Information Criterion.

In all models, we assessed the total variance explained by the model by R-squared metric and the relative importance of each predictor using the Lindemann, Merenda and Gold (LMG) metric. This metric shows the proportion of variance explained by each of model predictors within the entire variance explained.

Given significant interactions between biome factor and biomass fractions of mycorrhizal plants (Table 1), we fitted linear models with EcM or AM biomass fractions as predictors and topsoil and subsoil C stocks as response variables to the data within individual biomes. The outcome of this analysis is reported in the Extended Data Table 2.

Additionally, we examined whether the amount of carbon stored in AM and EcM plants per unit area was a better predictor of soil C stocks than the fractions of AM or EcM biomass per unit area. This analysis was performed in exactly the same manner as the analysis of the impacts of AM and EcM biomass. We assessed the resulting models using the Akaike Information Criterion and R^2, and detected that these models were worse than the those based on the fractions of AM or EcM biomass per unit area. We also examined how fractions of AM or EcM biomass per unit area, were related to soil C-to-N ratio, and detected that these relationships had a pattern very similar to that of soil C.

Dealing with spatial autocorrelations

The above-described analyses were checked for spatial autocorrelation using a Moran's I metric. These tests revealed a spatial autocorrelation of residuals, which is expected given the nature of our soil and vegetation data with spatial points nested within biomes. We examined how incorporation of a spatial dependence structure affected our models. Hereto we ran generalized least square (GLS) models of the same sets of predictors as described in the section "Statistical analysis of the relationship between soil carbon content and biomass distribution of mycorrhizal types" but included spatial correlation structures. Because calculation of global spatial autocorrelation matix is computationally intensive, we chose a Monte-Carlo type approach for this analysis. For each analysis we drew 100 times a random

sample of 2% datapoints (1624 points), taking care that all biomes are included into each sample, and for each sample we ran a gls accounting for autocorrelation structure of the data. Hereto we used the algorithms proposed by Zuur et al.[70]. The averaged outcomes of these analyses for the relationship between topsoil C, AM, and EcM fractions of vegetation biomass are shown in the Supplementary Table 11. The analyses accounting for autocorrelations yielded the same conclusions as the analyses that did not account for autocorrelations: there is a positive relationship between biomass fraction of EcM plants in vegetation and soil C, and this relationship is biome-independent (non-significant Biome * EcM interaction); in contrast, the relationship between biomass fraction of AM plants in vegetation and soil C was idiosyncratic and biome-dependent.

Including spatial coordinates explicitly in the model, or explicit accounting for autocorrelation matrix, can be problematic to interpret, because covariation between spatial coordinates and environmental variables can obscure the interpretation of the relative importance of the predictors[71]. Given that the main goal of our analysis was to detect the global links between AM and EcM dominance and soil C, but not to predict this relationship in new areas or under future climatic scenarios we prioritized the interpretation of the models rather than their predictive power, and therfore report the outcomes of the models that do not account for autocorrelations (Table 1).

Materials & Correspondence

Correspondence and material requests should be addressed to Nadejda Soudzilovskaia (n.a.soudzilovskaia@cml.leidenuniv.nl).

Data availability statement

All the codes and author's data used to create the maps of mycorrhizal vegetation biomass, and all the data underlying Figures 1-4, Extended Data Figs 3-6, Table 1 and Supplementary Tables 9-11 are available at a GitHub repository https://github.com/nasoudzilovskaia/Soudzilovskaia_NatureComm_MycoMaps and at the Netherlands governmental data repository DANS (Data Archiving and Networked Services (https://dans.knaw.nl/en) [available after paper acceptance]

All the links to the publically available datasets used within the process of assembly of the maps of mycorrhizal vegetation, in addition to the author's data, are provided in the reference list of the manuscript.

Code availability statement

All the codes used to the generate maps of mycorrhizal vegetation biomass are available at a GitHub repository https://github.com/nasoudzilovskaia/Soudzilovskaia_NatureComm_MycoMaps

Table 1. Summary of statistical linear models (lm) predicting soil carbon stocks at 0-20, 20-60, and 60-100cm depth based on biome and fraction of EcM or AM plants in vegetation biomass. R2 – ordinary least square goodness of fit measure. LMG - relative importance of individual predictors in a model examined through the Lindemann-Merenda-Gold (LMG) metric. The LMG shows the percentage of variance explained by each of model predictors within the entire variance explained. The P values show the outcome of ANOVA type I models (n= 78883 in all models).

Predicted Variable	Model	R^2	Predictor	P-value	LMG (%)
Topsoil C 0-20cm	**EcM + Biome + EcM x Biome**	**0.53**	**EcM**	**<0.001**	**42**
			Biome	<0.001	57
			EcM x Biome	<0.001	1
Subsoil C 20-60cm	**EcM + Biome + EcM x Biome**	**0.38**	**EcM**	**<0.001**	**39**
			Biome	<0.001	60
			EcM x Biome	<0.001	1
Subsoil C 60-100cm	**EcM + Biome + EcM x Biome**	**0.33**	**EcM**	**<0.001**	**35**
			Biome	<0.001	64
			EcM x Biome	<0.001	1
Topsoil C 0-20cm	**AM + Biome + AM x Biome**	**0.54**	**AM**	**<0.001**	**38**
			Biome	<0.001	56
			AM x Biome	<0.001	6
Subsoil C 20-60cm	**AM + Biome + AM x Biome**	**0.33**	**AM**	**<0.001**	**29**
			Biome	<0.001	67
			AM x Biome	<0.001	2

Subsoil C 60-100cm	**AM + Biome + AM x Biome**	**0.32**	**AM**	**<0.001**	**31**
			Biome	<0.001	67
			AM x Biome	<0.001	2

References

1 Smith, S. E. & Read, D. J. *Mycorrhizal Symbiosis*. (Academic Press, 2008).

2 Brundrett, M. C. Mycorrhizal associations and other means of nutrition of vascular plants: understanding the global diversity of host plants by resolving conflicting information and developing reliable means of diagnosis. *Plant and Soil* **320**, 37-77, doi:10.1007/s11104-008-9877-9 (2009).

3 Read, D. J. Mycorrhizas in ecosystems. *Experientia* **47**, 376-391, doi:10.1007/bf01972080 (1991).

4 Brundrett, M. C. Mycorrhizas in natural ecosystems. *Advances in Ecological Research* **21**, 171-313 (1991).

5 Leake, J. R. *et al.* Networks of power and influence: the role of mycorrhizal mycelium in controlling plant communities and agroecosystem functioning. *Canadian Journal of Botany* **82**, 1016-1045, doi:10.1139/b04-060 (2004).

6 Bennett, J. A. *et al.* Plant-soil feedbacks and mycorrhizal type influence temperate forest population dynamics. *Science* **355**, 181-184 (2017).

7 Teste, F. P. *et al.* Plant-soil feedback and the maintenance of diversity in Mediterranean-climate shrublands. *Science* **355**, 173-176 (2017).

8 Shi, Z., Wang, F. & Liu, Y. Response of soil respiration under different mycorrhizal strategies to precipitation and temperature. *Journal of Soil Science and Plant Nutrition* **12**, 411-420 (2012).

9 Wurzburger, N., Brookshire, E. N. J., McCormack, M. & Lankau, R. Mycorrhizal fungi as drivers and modulators of terrestrial ecosystem processes *New Phytologist* **213** 996-999 (2017).

10 Tedersoo, L. & Bahram, M. Mycorrhizal types differ in ecophysiology and alter plant nutrition and soil processes. *Biological Reviews*, in press, doi:10.1111/brv.12538 (2019).

11 Averill, C., Turner, B. L. & Finzi, A. C. Mycorrhiza-mediated competition between plants and decomposers drives soil carbon storage. *Nature* **505**, 543–545, doi:10.1038/nature12901 (2014).

12 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 Phytologist* **199**, 41-51, doi:10.1111/nph.12221 (2013).

13 Terrer, C., Vicca, S., Hungate, B. A., Phillips, R. P. & Prentice, I. C. Mycorrhizal association as a primary control of the CO2 fertilization effect. *Science* **353**, 72-74 (2016).

14 Read, D. J., Leake, J. R. & Perez-Moreno, J. Mycorrhizal fungi as drivers of ecosystem processes in heathland and boreal forest biomes. *Canadian Journal of Botany-Revue Canadienne De Botanique* **82**, 1243-1263, doi:10.1139/b04-123 (2004).

15 Sulman, B. N. *et al.* Feedbacks between plant N demand and rhizosphere priming depend on type of mycorrhizal association. *Ecology letters* **20**, 1043-1053 (2017).

16 Brzostek, E. R., Fisher, J. B. & Phillips, R. P. Modeling the carbon cost of plant nitrogen acquisition: Mycorrhizal trade-offs and multipath resistance uptake improve

predictions of retranslocation. *Journal of Geophysical Research: Biogeosciences* **119**, 1684-1697 (2014).

17 Craig, M. E. *et al.* Tree mycorrhizal type predicts within-site variability in the storage and distribution of soil organic matter. *Global change biology* (2018).

18 Shi, M., Fisher, J. B., Brzostek, E. R. & Phillips, R. P. Carbon cost of plant nitrogen acquisition: global carbon cycle impact from an improved plant nitrogen cycle in the Community Land Model. *Global change biology* (2015).

19 Lin, G., McCormack, M. L., Ma, C. & Guo, D. Similar below-ground carbon cycling dynamics but contrasting modes of nitrogen cycling between arbuscular mycorrhizal and ectomycorrhizal forests. *New Phytologist* **213**, 1440-1451 (2017).

20 Averill, C. & Hawkes, C. V. Ectomycorrhizal fungi slow soil carbon cycling. *Ecology Letters* **19**, 937-947, doi:10.1111/ele.12631 (2016).

21 Terrer, C. *et al.* Ecosystem responses to elevated CO2 governed by plant–soil interactions and the cost of nitrogen acquisition. *New Phytologist* **217**, 507-522 (2018).

22 Fernandez, C. W. & Kennedy, P. G. Revisiting the 'Gadgil effect': do interguild fungal interactions control carbon cycling in forest soils? *New phytologist* **209**, 1382-1394 (2016).

23 Näsholm, T. *et al.* Are ectomycorrhizal fungi alleviating or aggravating nitrogen limitation of tree growth in boreal forests? *New Phytologist* **198**, 214-221, doi:10.1111/nph.12139 (2013).

24 Brundrett, M. C. Distribution and evolution of mycorrhizal types and other specialised roots in Australia. *Ecological Studies* **230**, 361-394 (2017).

25 Menzel, A.. *et al.* Distribution patterns of arbuscular mycorrhizal and non-mycorrhizalplant species in Germany. *Perspectives in Plant Ecology, Evolution and Systematics* **21**, 78–88 (2016).

26 Swaty, R., Michael, H. M., Deckert, R. & Gehring, C. A. Mapping the potential mycorrhizal associations of the conterminous United States of America. *Fungal Ecology* (2016).

27 Tedersoo, L. *et al.* Global diversity and geography of soil fungi. *Science* **346**, 1078-1088, doi:10.1126/science.1256688 (2014).

28 Davison, J. *et al.* Global assessment of arbuscular mycorrhizal fungus diversity reveals very low endemism. *Science* **349**, 970-973 (2015).

29 Zaehle, S. *et al.* Evaluation of 11 terrestrial carbon–nitrogen cycle models against observations from two temperate F ree-A ir CO 2 E nrichment studies. *New Phytologist* **202**, 803-822 (2014).

30 Soudzilovskaia, N. *et al.* FungalRoot: Global online database of plant mycorrhizal associations. *Biorxiv pre-print repository*. doi: 10.1101/717488

31 Soudzilovskaia, N. A. *et al.* Quantitative assessment of the differential impacts of arbuscular and ectomycorrhiza on soil carbon cycling. *New Phytologist* **208**, 280-293 (2015).

32 Ellis, E. C., Goldewijk, K. K., Siebert, S., Lightman, D. & Ramankutty, N. Anthropogenic transformation of the biomes, 1700 to 2000. *Global Ecology and Biogeography* **19**, 589-606, doi:10.1111/j.1466-8238.2010.00540.x (2010).

33 Bailey, R. G. *Ecoregions. The Ecosystem Geography of the Oceans and Continents*. (Springer New York, 2014).

34 ESA. CCI Land cover map 2015 (https://www.esa-landcover-cci.org/). (2017).

35 Schepaschenko, D. *et al.* A dataset of forest biomass structure for Eurasia. *Scientific data* **4**, 170070 (2017).

36 Fisher, J. B. *et al.* Tree–mycorrhizal associations detected remotely from canopy spectral properties. *Global Change Biology* (2016).
37 Hartley, A. J. *et al.* Uncertainty in plant functional type distributions and its impact on land surface models." *Remote Sensing of Environment* **203**, 71-89 (2017).
38 Liu, Y. Y. *et al.* Recent reversal in loss of global terrestrial biomass. Nature Climate change. . *Nature Climate change* **5**, 470-474 (2015).
39 Wang, B. & Qiu, Y. L. Phylogenetic distribution and evolution of mycorrhizas in land plants. *Mycorrhiza* **16**, 299-363, doi:10.1007/s00572-005-0033-6 (2006).
40 Thompson, J. Decline of vesicular-arbuscular mycorrhizae in long fallow disorder of field crops and its expression in phosphorus deficiency of sunflower. *Australian Journal of Agricultural Research* **38**, 847-867 (1987).
41 Mbuthia, L. W. *et al.* Long term tillage, cover crop, and fertilization effects on microbial community structure, activity: Implications for soil quality. *Soil Biology and Biochemistry* **89**, 24-34 (2015).
42 Jo I. et al. Neglecting plant–microbe symbioses leads to underestimation of modeled climate impacts. *Science Advances, in press.* (2019).
43 Cotrufo, M. F., Wallenstein, M. D., Boot, C. M., Denef, K. & Paul, E. The M icrobial E fficiency-M atrix S tabilization (MEMS) framework integrates plant litter decomposition with soil organic matter stabilization: do labile plant inputs form stable soil organic matter? *Global Change Biology* **19**, 988-995 (2013).
44 Cotrufo, M. F. *et al.* Formation of soil organic matter via biochemical and physical pathways of litter mass loss. *Nature Geoscience* **8**, ngeo2520 (2015).
45 Poirier, V., Roumet, C. & Munson, A.D. The root of the matter: Linking root traits and soil organic matter stabilization processes. *Soil Biology and Biochemistry* **120**, 246-259 (2018).
46 Olson, D. M. *et al.* Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *Bioscience* **51**, 933-938 (2001).
47 Batjes, N. Harmonized soil property values for broad-scale modelling (WISE30sec) with estimates of global soil carbon stocks. *Geoderma* **269**, 61-68 (2016).
48 Newsham, K., Upson, R. & Read, D. Mycorrhizas and dark septate root endophytes in polar regions. *Fungal Ecology* **2**, 10-20 (2009).
49 Hengl, T. *et al.* SoilGrids250m: Global gridded soil information based on machine learning. *PLoS one* **12**, e0169748 (2017).
50 Batjes, N. H. *et al.* WoSIS: providing standardised soil profile data for the world. *Earth System Science Data* **9**, 1-14 (2017).
51 Steidinger, B.S. *et al.* Climatic controls of decomposition drive the global biogeography of forest-tree symbioses. Nature **569**, 404–408 (2019).
52 Sanderman, J., Hengl, T. & Fiske, G. J. Soil carbon debt of 12,000 years of human land use. *Proceedings of the National Academy of Sciences*, 201706103 (2017).
53 Averill, C., Dietze, M. C. & Bhatnagar, J. M. Continental-scale nitrogen pollution is shifting forest mycorrhizal associations and soil carbon stocks. *Global change biology* (2018).
54 Ekblad, A. *et al.* Production and turnover of ectomycorrhizal extramatrical mycelial biomass and necromass under elevated CO2 and nitrogen fertilization. *New Phytologist* **211**, 874-885, doi:10.1111/nph.13961 (2016).
55 Zhu, K., McCormack, M. L., Lankau, R. A., Egan, J. F. & Wurzburger, N. Association of ectomycorrhizal trees with high carbon-to-nitrogen ratio soils across temperate forests is driven by smaller nitrogen not larger carbon stocks. *Journal of Ecology* **106**, 524-535 (2018).

56 Karst, J., Marczak, L., Jones, M. D. & Turkington, R. The mutualism-parasitism continuum in ectomycorrhizas: A quantitative assessment using meta-analysis. *Ecology* **89**, 1032-1042, doi:10.1890/07-0823.1 (2008).

57 ORNL DAAC. 2017. https://doi.org/10.3334/ORNLDAAC/1388; dataset "Bailey Ecoregions of the Continents (Province)"

58 Pan, Y., Birdsey, R. A., Phillips, O. L. & Jackson, R. B. The structure, distribution, and biomass of the world's forests. *Annual Review of Ecology, Evolution, and Systematics* **44**, 593-622 (2013).

59 Lamotte, M. in *Tropical ecological systems* 179-222 (Springer, 1975).

60 Ruiz-Benito, P. *et al.* Stand structure and recent climate change constrain stand basal area change in European forests: a comparison across boreal, temperate, and Mediterranean biomes. *Ecosystems* **17**, 1439-1454 (2014).

61 Poulter B. *et al* Plant functional type classification for earth system models: results from the European Space Agency's Land Cover Climate Change Initiative. *Geosci. Model Dev.*, **8**, 2315–2328 (2015).

62 FAO. Global forest resource assessement. (2000).

63 FAO, O. *Oilseeds and oilseed products. OECD-FAO Agricultural Outlook.*, (OECD Publishing, Paris. (2013).

64 Harley, J. L. & Harley, E. A check-list of mycorrhiza in the British flora. *New Phytologist* **105**, 1-102 (1987).

65 Tedersoo, L (ed). *Biogeography of mycorrhizal symbiosis* (Springer, 2017).

66 Brundrett, M. C. Global diversity and importance of mycorrhizal and nonmycorrhizal plants. *Ecological Studies* **230**, 533-556 (2017).

67 Kohout, P. Biogeography of ericoid mycorrhiza. *Ecological Studies* **230**, 179-193 (2017)..

68 Tedersoo, L. & Brundrett, M. C. Evolution of ectomycorrhizal symbiosis in plants. *Ecological Studies* **230**, 407-467 (2017).

69 Hoenig, J. M. & Heisey, D. M. The abuse of power: the pervasive fallacy of power calculations for data analysis. *The American Statistician* **55**, 19-24 (2001).

70 Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A. & Smith, G. M. *Mixed Effects Models and Extentsions in Ecology with R.* (Springer, 2009).

71 Dormann, C.F. (2007) Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography*, **16**, 129-138 (2007).

Acknowledgments

This study was supported by the vidi grant 016.161.318 issued to NAS by The Netherlands Organization for Scientific research and grants EcolChange, MOBERC and 1399PUT issued to LT by the Estonian Research Council. We thank Daisy Brickhill for editing the manuscript. We also thank Milagros Barcelo, Stijn Vaessen and Jinhong He who helped compiling data on plant mycorrhizal status. Michala Phillips assisted in the JBF data transfer. JBF contributed to this research, in part, from the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. California Institute of Technology. Government sponsorship acknowledged. Copyright 2018. All rights reserved. JBF was supported by funding from DOE BER TES and NASA IDS.

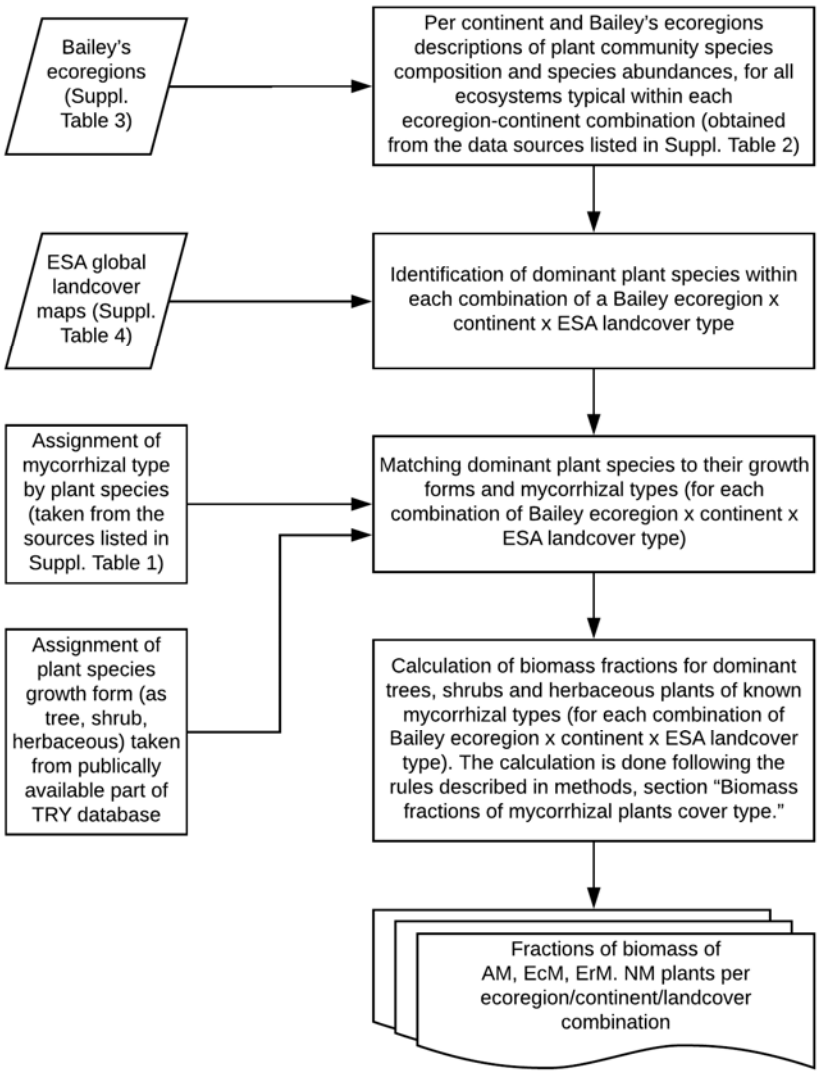
Author contributions

NAS and CT generated the idea. NAS, PMvB, LT and CT designed the study, and analyzed the data. NAS wrote the manuscript. NAS and IMC have prepared the figures, with the leading role of IMC. LT provided mycorrhizal type assignments to vegetation. CT, IMC, NCdS and MvhZ performed geoinformatics data processing. MLM, JBF, and MB provided data for map validation. PMvB, CT, LT and IMC have equally contributed to this work. All authors have contributed to preparation of manuscript drafts.

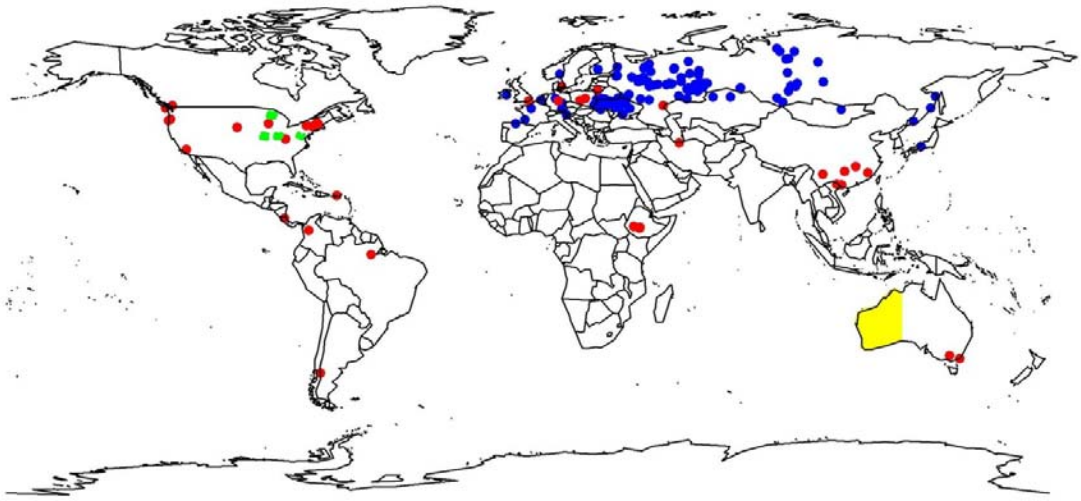
Competing Financial Interests statement

There are no competing financial interests associated with this work.

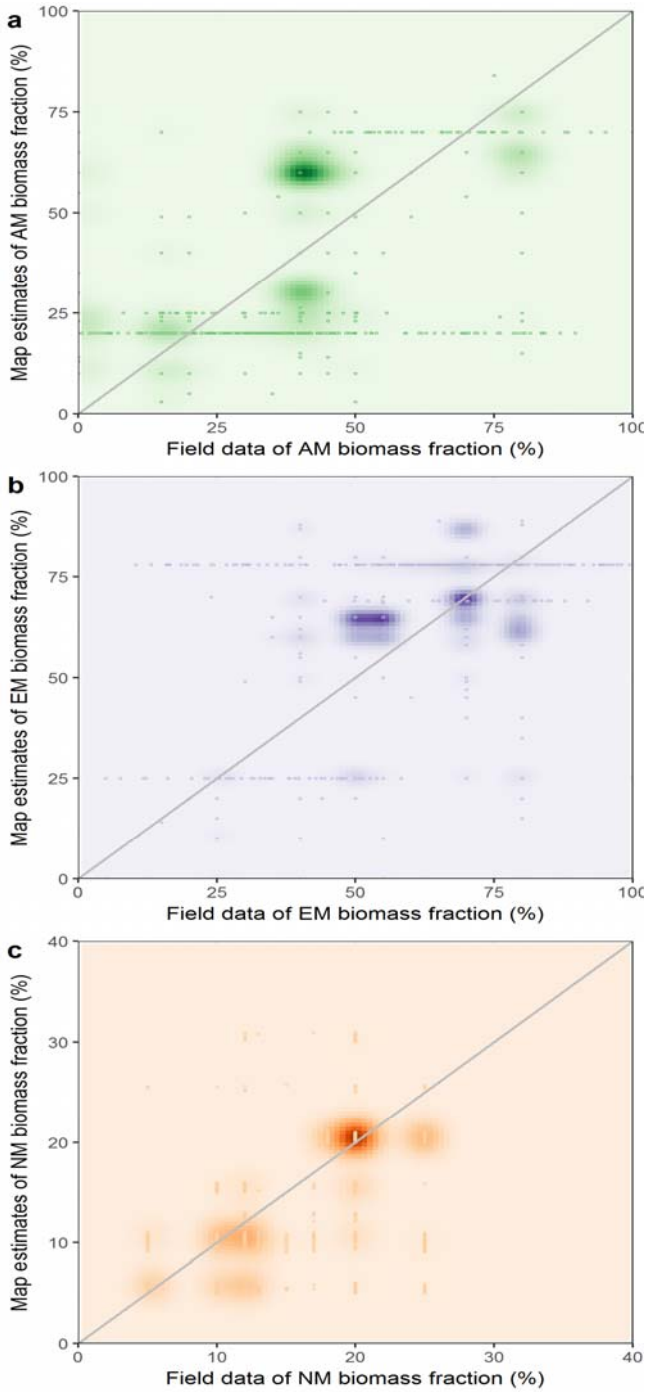
Supplementary data to the manuscript "Global mycorrhizal plants distribution linked to terrestrial carbon stocks"



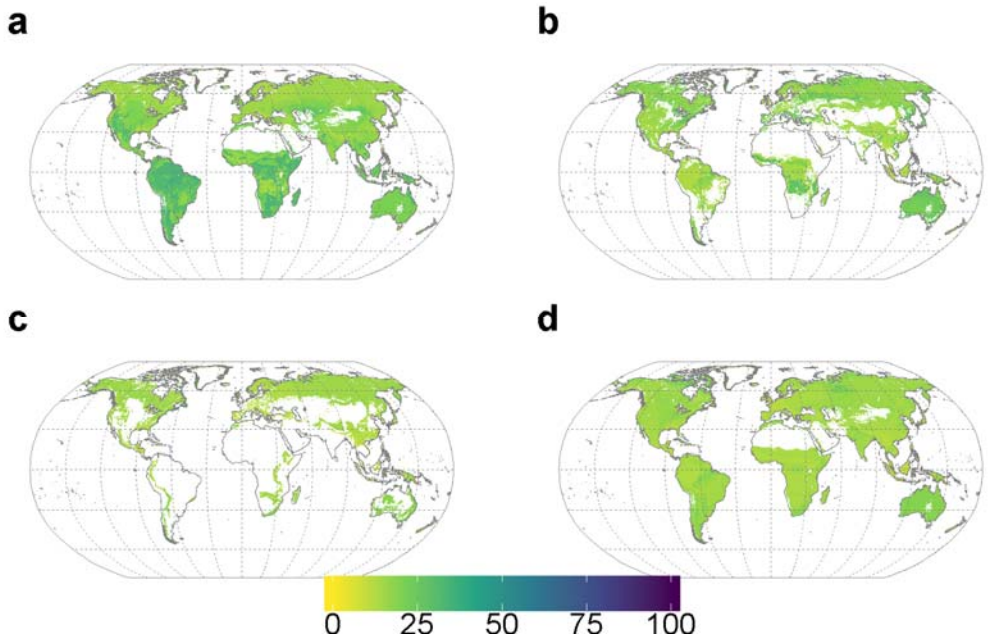
Extended Data Figure 1. Flowchart of the data assembly processes employed to construct maps of mycorrhizal types plant biomass fractions.



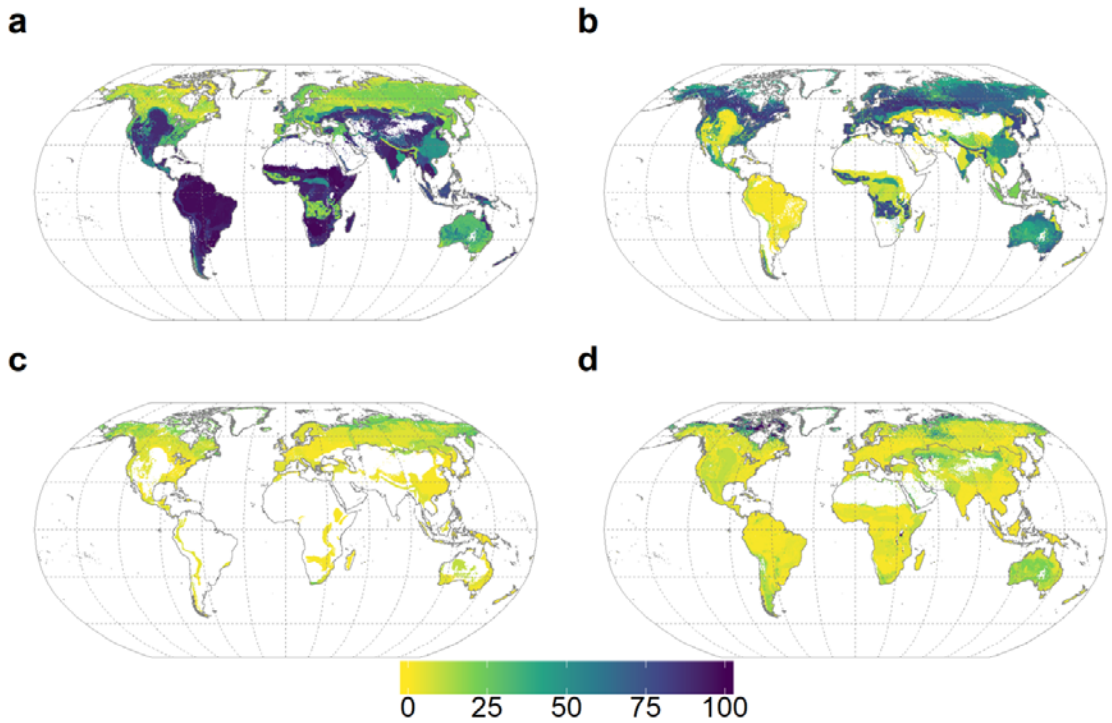
Extended Data Figure 2. Data used for map validation with the data of Schepachenko and co-workers [1] (blue points), Lin and co-workers [2] (red points), Fisher and co-workers [3] (four areas in USA, green) and Brundrett [4] (West Australia, yellow).



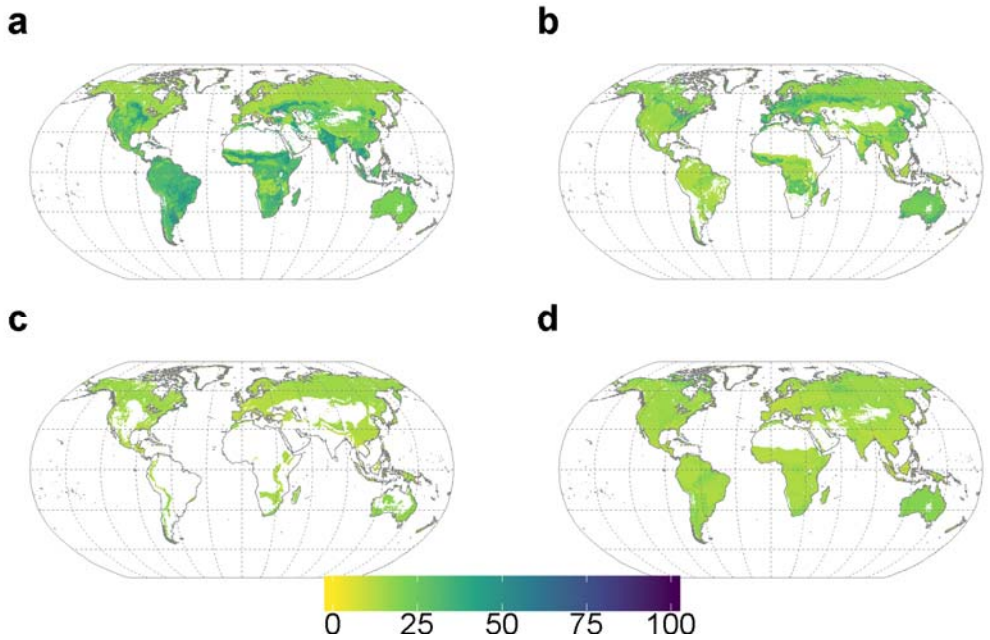
Extended Data Figure 3. Validation of the current distribution of (a) AM, (b) EcM, and (c) NM plants. The scatter plots show the relationships between the estimates of the mycorrhizal vegetation maps and the field measured data reported in the works of Schepachenko and co-workers [1] (Eurasia), Lin and co-workers [2] (global), Fisher an co-workers [3] (USA) and Brundrett [4] (West Australia). See Extended Data Figure 2 for the location of the validation sites. Due to overlap among multiple points, a 2d kernel density of the probability distribution is shown with increasing color intensity. The deviation from the 1:1 line is estimated as mean averaged error (MAE), being 18.7% for AM biomass vegetation fractions, 13.6% for ECM and 4.8% for NM.



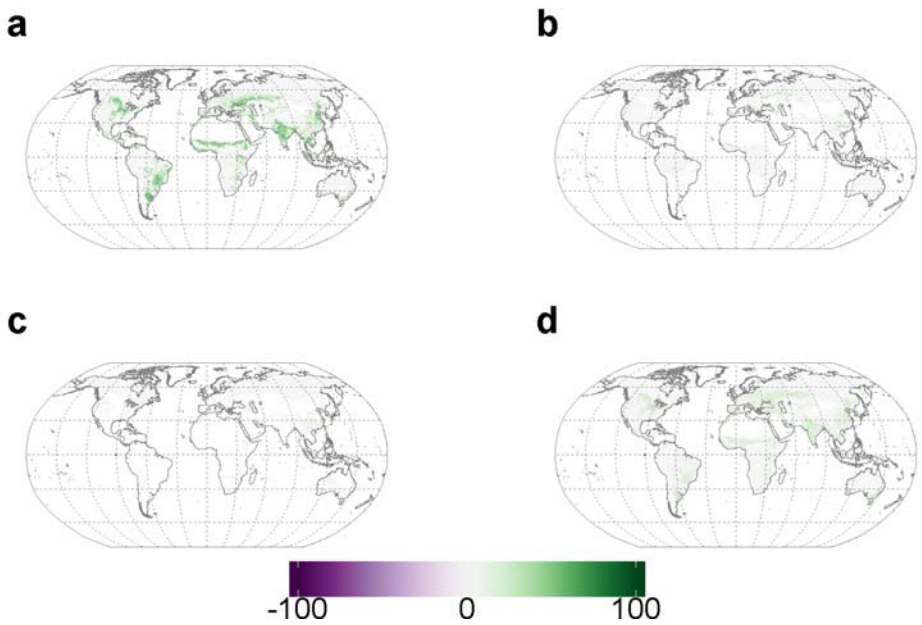
Extended Data Figure 4. Uncertainty in the percentage of aboveground plant biomass of a - arbuscular mycorrhizal plants, b - ectomycorrhizal plants, c - ericoid mycorrhizal plants, and d -non-mycorrhizal plants. Color intensity reflects uncertainty at 90% CI of the biomass fraction of mycorrhizal plants of a given type in a grid cell of 10 arcmin.



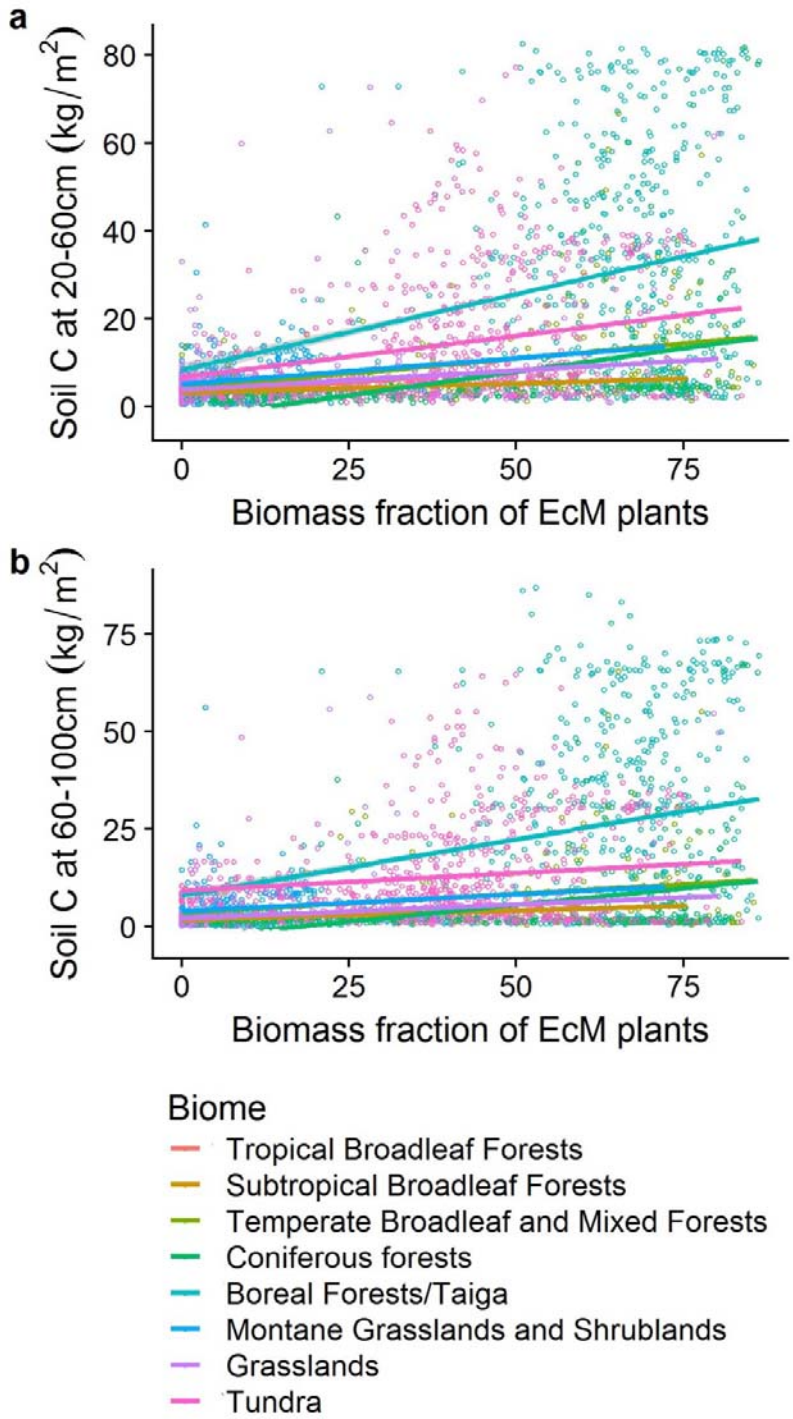
Extended Data Figure 5. Potential distribution of mycorrhizal vegetation in a cropland-free world. a - arbuscular mycorrhizal plants, b - ectomycorrhizal plants, c - ericoid mycorrhizal plants, and –d - non-mycorrhizal plants. Color intensity reflects biomass fraction of mycorrhizal plants of a given type in a grid cell of 10 arcmin.



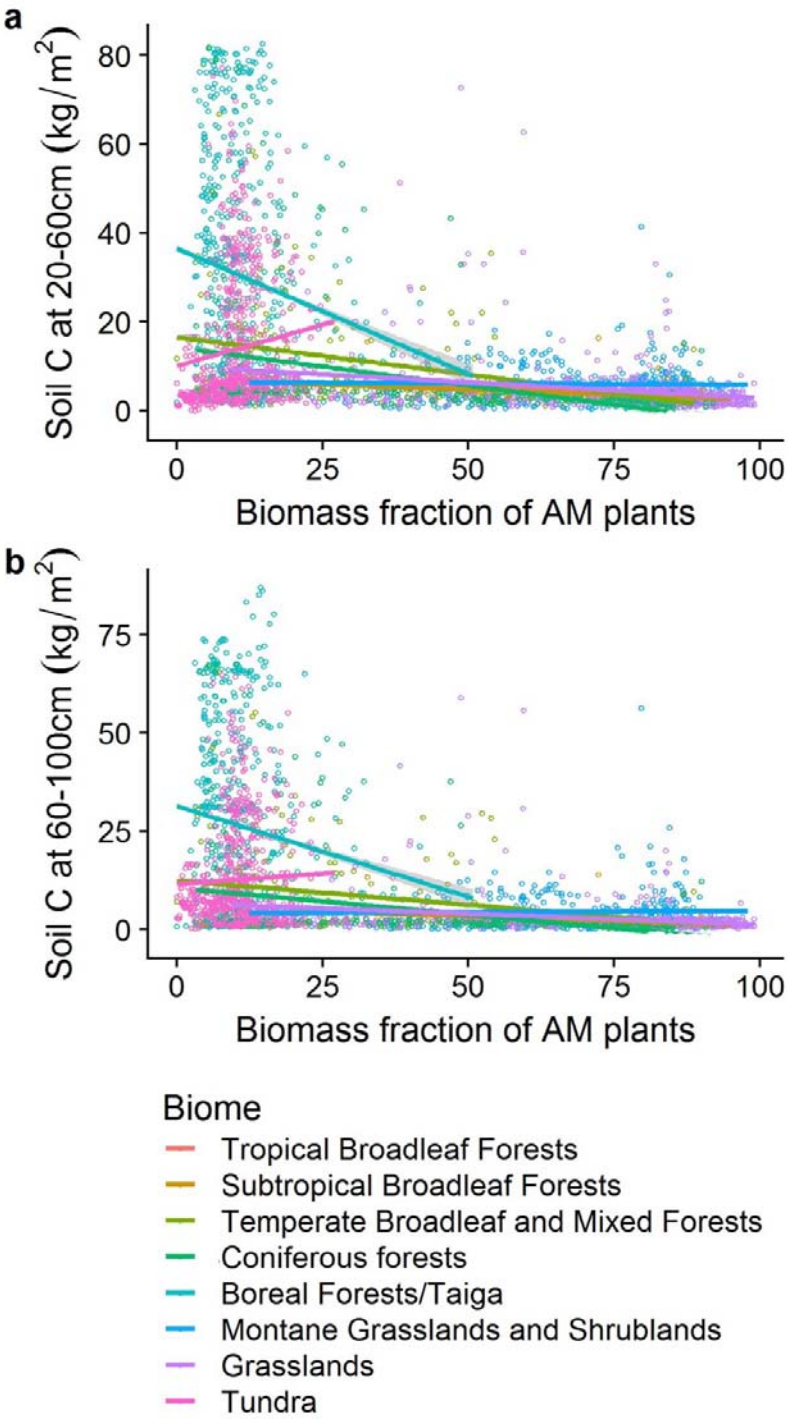
Extended Data Figure 6. Uncertainty in potential distribution of mycorrhizal vegetation in a cropland-free world of a - arbuscular mycorrhizal plants, b - ectomycorrhizal plants, c - ericoid mycorrhizal plants, and d - non-mycorrhizal plants. Color intensity reflects uncertainty at 90% confidence interval of the biomass fraction of mycorrhizal plants of a given type in a grid cell of 10 arcmin.



Extended Data Figure 7. Uncertainty in changes in biomass fractions of mycorrhizal vegetation induced by crop cultivation and pastures for a - arbuscular mycorrhizal plants, b - ectomycorrhizal plants, c - ericoid mycorrhizal plants, d - non-mycorrhizal plants. Purple colors indicate uncertainty in losses of biomass, green colors indicate uncertainty in gains of biomass. Color intensity reflects uncertainty at 90% CI) of biomass fraction of mycorrhizal plants of a given type in a grid cell of 10 arcmin.



Extended Data Figure 8. Quantitative relationships between subsoil carbon at 20-60 cm, and 60-100 cm layers, and biomass fraction of EcM plants in an ecosystem. Croplands were excluded from the analysis. Per biome predictions are shown.



Extended Data Figure 9. Quantitative relationships between subsoil carbon at 20-60 cm, and 60-100 cm layers, and biomass fraction of AM plants in an ecosystem. Croplands were excluded from the analysis. Per biome predictions are shown.

Supplementary Tables 1-8 are provided as separate Excel files.

Supplementary Table 9. Per biome distribution ± uncertainty at 90% CI of the vegetation carbon stocks among mycorrhizal types. The data presents the sum of carbon stored in vegetation of each mycorrhizal type, across all land cover categories present in each biome, including agricultural, urban and other human-transformed areas.

	Absolute values of C stock (GT)				**Uncertainty at 90% CI of the absolute values of C stock (GT)**			
Biome	**AM**	**EcM**	**ErM**	**NM**	**AM**	**EcM**	**ErM**	**NM**
Tropical and Subtropical Moist Broadleaf Forests	142.8	14.2	1.12	8.35	8.95	3.99	0.18	1.93
Tropical and Subtropical Dry Broadleaf Forests	4.10	0.51	0.02	0.86	2.61	0.86	0.03	1.21
Temperate Broadleaf and Mixed Forests	13.5	18.2	0.68	3.48	6.85	7.85	0.35	2.97
Boreal Forests/Taiga	7.30	33.5	2.97	3.47	3.28	7.30	1.58	2.76
Montane Grasslands and Shrublands	2.32	0.24	0.04	0.28	2.36	0.48	0.06	0.38
Grasslands	33.1	10.5	0.19	5.10	2.36	4.71	0.17	1.79
Tundra	0.49	1.38	0.51	0.92	0.37	1.16	0.47	1.29

Supplementary Table 10. Summary of statistical models predicting soil carbon stocks at 0-20, 20-60, and 60-100 cm depth based on the fraction of EcM or AM plants in vegetation biomass within individual biomes, Coeff. – slope as revealed by a generalized linear model (glm), SE – standard error of the slope.

		0-20 cm soil				20-60 cm soil				60-100 cm soil			
Biome		R^2	P	Coeff.	SE	R^2	P	Coeff.	SE	R^2	P	Coeff.	SE
Tropical Broadleaf Forests	EcM	0.49	<0.001	40.2	3.9	0.77	<0.001	83.3	4.3	0.94	<0.001	78.0	1.9
Tropical Broadleaf Forests	AM	0.45	<0.001	--26.5	2.8	0.68	<0.001	-53.7	3.5	0.95	<0.001	-54.1	1.2
Subtropical Broadleaf Forests	EcM	0.14	<0.001	33.0	2.2	0.08	<0.001	38.4	3.6	0.12	<0.001	43.4	3.1
Subtropical Broadleaf Forests	AM	0.18	<0.001	-33.8	1.9	0.09	<0.001	-37.2	3.2	0.14	<0.001	--40.9	2.8
Temperate Broadleaf and Mixed Forests	EcM	0.26	<0.001	140.6	3.4	0.07	<0.001	131.2	7.1	0.04	<0.001	91.5	6.7
Temperate Broadleaf and Mixed Forests	AM	0.34	<0.001	-177.0	3.5	0.09	<0.001	-16.6	7.7	0.05	<0.001	-121.7	7.5
Coniferous forests	EcM	0.28	<0.001	140.4	3.0	0.09	<0.001	182.2	7.9	0.07	<0.001	144.0	6.9
Coniferous forests	AM	0.26	<0.001	-131.7	3.0	0.06	<0.001	-148.4	7.7	0.05	<0.001	-116.0	6.8
Boreal Forests/Taiga	EcM	0.01	<0.001	75.7	4.2	0.02	<0.001	164.4	11.3	0.02	<0.001	215.1	10.6
Boreal Forests/Taiga	AM	0.03	<0.001	-294.8	10.7	0.02	<0.001	-540.0	29.0	0.01	<0.001	-458.7	17.0
Montane Grasslands and Shrublands	EcM	0.18	<0.001	191.5	3.9	0.05	<0.001	111.2	5.1	0.02	<0.001	81.2	5.6
Montane Grasslands and Shrublands	AM	0.03	<0.001	-37.70	2.3		0.574			0.00	<0.001	15.9	3.0
Grasslands	EcM	0.06	<0.001	58.9	1.6	0.05	<0.001	82.9	2.5	0.06	<0.001	68.6	2.0
Grasslands	AM	0.06	<0.001	-47.5	1.3	0.06	<0.001	-67.9	2.0	0.06	<0.001	-52.1	1.6
Tundra	EcM	0.12	<0.001	132.9	2.9	0.05	<0.001	165.6	5.9	0.01	<0.001	80.1	5.3
Tundra	AM	0.13	<0.001	481.1	10.3	0.02	<0.001	372.0	20.8	0.00	<0.001	134.1	18.5

Supplementary Table 11. Summary of generalized least squares (gls) statistical models predicting soil carbon stocks at 0-20, 20-60, and 60-100 cm depth based on biome and fraction of EcM or AM plants in vegetation biomass, and accounting for autocorrelations. Hereto we used the gls formula "correlation = corExp(form = ~y + x)" Each model has been run 100 times on a selection of 2% of the data points of the entire dataset. The value shown in the table are the mean values of the 100 runs. Degrees of freedom: $DF_{mycorrhizal\ type}$=1, DF_{Biome}=7, $DF_{Biome\ x\ mycorrhizal\ type}$ =7, $DF_{residuals}$ = 1608

Predicted Variable	Model	Predictor	P value	F value
Topsoil C 0-20cm	**EcM + Biome + EcM x Biome**	**EcM**	**<0.001**	**96.3**
		Biome	0.011	5.3
		EcM x Biome	ns	2.7
Subsoil C 20-60cm	**EcM + Biome + EcM x Biome**	**EcM**	**<0.001**	**61.1**
		Biome	0.016	4.0
		EcM x Biome	ns	2.2
Subsoil C 60-100cm	**EcM + Biome + EcM x Biome**	**EcM**	**<0.001**	**51.1**
		Biome	0.006	4.8
		EcM x Biome	ns	1.7
Topsoil C 0-20cm	**AM + Biome + AM x Biome**	**AM**	**<0.001**	**37.1**
		Biome	0.001	7.2
		AM x Biome	0.001	5.8
Subsoil C 20-60cm	**AM + Biome + AM x Biome**	**AM**	**<0.001**	**19.5**
		Biome	0.001	5.7
		AM x Biome	0.034	3.1
Subsoil C 60-100cm	**AM + Biome + AM x Biome**	**AM**	**<0.001**	**24.2**
		Biome	0.001	5.9
		AM x Biome	ns	2.6

Supplementary Table 12. Areas occupied by tree plantations of distinct mycorrhizal types. For the entire cover of unspecified tree species the OECD FAO (Organization for Economic Co-operation and Development of Food and Agriculture Organization of the United Nations) report[6] provides an estimate 0.536 mln km^2 . In this analysis we have split this area between EcM tree plantations and AM tree plantations.

	Area mln km^2	Data source
Global area plantations EcM trees		
Pinus spp. (pines)	0.400	*6*
Other coniferous trees (mostly EcM)	0.274	*6*
Unspecified tree species	0.268	*6*
Total	**0.742**	
Global area plantations AM/EcM trees		
Eucalyptus spp. (eucalypts)	0.200	*6*
Acacia spp. (wattles)	0.083	*6*
Total	**0.283**	
Global area plantations AM trees		
Elaeis guineensis (oil palm)	0.170	*7*
Hevea brasilensis (rubber tree)	0.098	*6*
Tectona grandis (teak)	0.057	*6*
Other broadleaf tree species (mostly AM)	0.335	*6*
Unspecified tree species	0.268	*6*
Total	**0.928**	

References to the Supplementary Data

1 Schepaschenko, D. *et al.* A dataset of forest biomass structure for Eurasia. *Scientific data* **4**, 170070 (2017).

2 Lin, G., McCormack, M. L., Ma, C. & Guo, D. Similar below-ground carbon cycling dynamics but contrasting modes of nitrogen cycling between arbuscular mycorrhizal and ectomycorrhizal forests. *New Phytologist* **213**, 1440-1451 (2017).

3 Fisher, J. B. *et al.* Tree–mycorrhizal associations detected remotely from canopy spectral properties. *Global Change Biology* (2016).

4 Brundrett, M. C. in *Biogeography of Mycorrhizal Symbiosis Ecol. Stud.* (ed L. Tedersoo) Ch. 17, 361-394 (2017).

5 Pontius, R. G., Thontteh, O. & Chen, H. Components of information for multiple resolution comparison between maps that share a real variable. *Environmental and Ecological Statistics* **15**, 111-142 (2008).

6 FAO. Global forest resource assessement. (2000).

7. FAO. *Oilseeds and oilseed products. OECD-FAO Agricultural Outlook.* (2013).