The biogeography of relative abundance of soil fungi and bacteria in top surface soil

Kailiang Yu1*, Johan van den Hoogen1, Zhiqiang Wang2, Colin Averill1, Devin Routh1, Gabriel R. Smith3,1, Rebecca E. Drenovsky4, Kate M. Scow5, Fei Mo6, Mark P Waldrop7, Yuanhe Yang8,9, Weize Tang9,10, Franciska T. De Vries11, Richard D. Bardgett12, Peter Manning13, Felipe Bastida14, Sara G. Baer15, Elizabeth M. Bach16, Carlos García14, Qingkui Wang17, Linna Ma8, Baodong Chen18,9, Xianjing He19, Sven Teurlincx20, Amber Heijboer21,22, James A. Bradley23,24, Thomas W Crowther1*

1Institute of Integrative Biology, ETH Zürich, Zürich, Switzerland
2Institute for Advanced Study, Chengdu University, Chengdu, China
3Department of Biology, Stanford University, California, USA
4Biology Department, John Carroll University, Ohio, USA
5Department of Land, Air and Water Resources, University of California, Davis, California, USA
6College of Agronomy, Northwest A&F University, Shaanxi, PR China
7U.S. Geological Survey, Geology, Minerals, Energy, and Geophysics Science Center. Menlo Park, California, USA
8State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing, China
9University of Chinese Academy of Sciences, Beijing, China
10Key Laboratory of Vegetation Restoration and Management of Degraded Ecosystems, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China
11Institute of Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, the Netherlands
Department of Earth and Environmental Sciences, University of Manchester, Oxford Road, Manchester, UK

Senckenberg Biodiversity and Climate Research Centre, Frankfurt, Germany

CEBAS-CSIC. Department of Soil and Water Conservation. Campus Universitario de Espinardo, Murcia, Spain.

Kansas Biological Survey and Department of Ecology & Evolutionary Biology, University of Kansas, Lawrence, Kansas, USA

The Nature Conservancy, Nachusa Grasslands, Franklin Grove, IL, USA

Huitong Experimental Station of Forest Ecology, CAS Key Laboratory of Forest Ecology and Management, Institute of Applied Ecology, Shenyang, PR China

State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing, China

Key Laboratory of the Three Gorges Reservoir Region’s Eco-Environment, Ministry of Education, Chongqing University, Chongqing, China

Department of Aquatic Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, Netherlands

Biometris, Wageningen University & Research, Wageningen, Netherlands

Ecology and Biodiversity Group, Department of Biology, Institute of Environmental Biology, Utrecht University, Padualaan, Netherlands

School of Geography, Queen Mary University of London, London, E1 4NS, UK

Interface Geochemistry, GFZ German Research Centre for Geosciences, Potsdam, Germany

Corresponding author: kai86liang@gmail.com or tom.crowther@usys.ethz.ch
Abstract. Fungi and bacteria are the two dominant groups of soil microbial communities worldwide. By controlling the turnover of soil organic matter, these organisms directly regulate the cycling of carbon between the soil and the atmosphere. Fundamental differences in the physiology and life history of bacteria and fungi suggest that variation in the biogeography of soil fungal and bacterial relative abundance could drive striking differences in carbon decomposition and soil organic matter formation across different biomes. However, a lack of global and predictive information on the distribution of these organisms in terrestrial ecosystems has prevented the inclusion of soil fungal and bacterial relative abundance and the associated processes into global biogeochemical models. Here, we used a global scale dataset (>3000 distinct observations of soil fungal and bacterial abundance) in the top soil surface (up to 15 cm) (>3000 distinct observations of soil fungal and bacterial abundance) to generate the first quantitative and spatially high resolution (1km²) explicit map of soil fungal proportion, defined as fungi/fungi + bacteria, across terrestrial ecosystems. We reveal striking latitudinal trends where fungal dominance increases in cold and high latitude environments with large soil carbon stocks. There was strong non-linear response of fungal dominance to environmental gradient, i.e., mean annual temperature (MAT) and net primary productivity (NPP). Fungi and bacteria dominated in regions with low and high MAT and NPP, respectively, thus representing slow vs. fast soil energy channels, a concept with a long history in soil ecology. These high-resolution models provide the first steps towards representing the major soil microbial groups and their functional differences in global biogeochemical models to improve predictions of soil organic matter turnover under current and future climate scenarios.

Keywords:
1. Introduction

Fungi and bacteria are the dominant members of soil microbial communities worldwide in terms of diversity, abundance and biomass (Bahram et al., 2018). Representing distinct kingdoms of life, bacteria and fungi systematically differ in a multitude of physiological and life-history traits with direct implications for global soil biogeochemical cycles (Waring et al., 2013; Wieder et al., 2015) including the decomposition of organic matter, which contributes to one of the largest fluxes of CO$_2$ on Earth (Glassman et al., 2018). Compared to bacteria, fungi generally have slower growth and turnover rates (Rousk and Bååth, 2007), greater carbon (C) to nutrient stoichiometry (Waring et al., 2013), greater capacity to degrade a wider and more recalcitrant range of substrates (Strickland and Rousk, 2010) and potentially higher C use efficiency (Soares and Rousk, 2019).

For these reasons, a new generation of soil and ecosystem models have begun to explicitly represent these fundamentally distinct fast and slow cycling microbial groups, suggesting that spatially-explicit information about the relative abundance of fungal and bacteria in a region can dramatically improve the accuracy of global carbon cycling model predictions (Shi et al., 2018; Sulman et al., 2014; Wieder et al., 2013, 2015). Generating an understanding of the factors affecting the biogeography of the relative abundance of fungal and bacteria in soil, and its connection to global carbon cycle, would represent a breakthrough step forward in our general understanding of the natural history of soil microbial life.

Temperature, precipitation, soil pH and soil C:N have all been linked to the balance of fungi vs. bacteria within soil communities across different spatial scales (Bahram et al., 2018; Strickland and Rousk, 2010; Tedersoo et al., 2014). Relative to fungi, bacteria tend to dominate in locations with high soil nutrient contents or in frequently disturbed soils that limit the growth of fungal hyphae or make N more available (Fierer et al., 2009; Van Der Heijden et al., 2008;
Strickland and Rousk, 2010). However, until now, the relative importance of these different environmental drivers remains relatively unclear at global scale, and the biogeography of these major functional groups (fungi vs. bacteria) has only been demonstrated at local and regional scales. A recent analysis suggested that the relative soil bacterial abundance is high in tropical latitudes and decreases in abundance towards the high-latitude boreal regions, where fungi tend to dominate (Bahram et al., 2018). Translating these broad-scale trends into quantitative, spatially explicit information will be necessary if we intend to represent regional variations in soil community functioning (Wieder et al., 2013, 2015), or predict future changes in terrestrial carbon and nutrient cycling.

Some progress was made in the quantitative and spatially explicit understanding of global biogeographic patterns of fungal and bacterial biomass and their biomass ratio. By synthesizing phospholipid-derived fatty acids data from 1323 locations across the globe, and extrapolating linear relationships with environmental factors, a recent study generate the global maps of fungal and bacterial biomass and their biomass ratio at the resolution of 0.5 degree for top soil (0-30 cm) (He et al., 2020). This approach provided the support for the broad-scale latitudinal trends, with high fungal dominance in high-latitude regions. Yet to date, there are three crucial knowledge gaps to be addressed. First, we still lack a high resolution evaluation of the spatial patterns and regional contingencies in fungal:bacterial ratios, which would allow representation of microbial-mediated mechanisms that operate within and/or across ecosystems at fine scales (Frindte et al., 2019; Zhu et al., 2017). Second, the response of soil microbial community composition across environmental gradients are expected to be non-linear, with strong interactive effects of different environmental characteristics that give rise to thresholds that diverge from the global latitudinal trends (Sengupta et al., 2021; Wang et al., 2018; Waring et al., 2013). This
non-linear linkage of soil microbial communities with environmental resource gradient has not
been assessed, while it has fundamental implications on ecosystem functions and management
solutions (Sengupta et al., 2021; Wang et al., 2018). Third, there are distinct difference of soil
nutrients, soil microbial community and the associated biogeochemical processes across soil
depths, i.e., from top surface soil (i.e., 0-10 cm) to top subsurface soil (i.e., 0-30 cm) (Lavahun et
al., 1996; Yue et al., 2015). A continental-scale empirical study further showed that strong
positive associations among soil microbial community, fertility and plant productivity are limited
to the top surface soil (Delgado-Baquerizo et al., 2017), thus highlighting its potential dominant
role regulated by top surface soil microbial communities on ecosystem functions and the
research needs of generating a global spatially explicit understanding of soil fungi and bacteria in
top surface soil.

Here, we present a global analysis of total and relative abundance of soil fungi and bacteria
in soil surfaces (defined as top 10-15 cm) informed from over 3000 spatially distinct surface soil
observations from phospholipid-derived fatty acids (PLFA) (Fig. 1a). The use of PLFA data
provides an opportunity to provide quantitative insights into the abundances of major
functional groups. We conducted the analysis on the abundances in view of the uncertainty in
conversion factors used to convert abundance derived from PLFA to biomass (Frostegård et al.,
2011; Klamer and Bååth, 2004). We used machine learning to link the variation in soil fungi and
bacteria abundances to global variation in 95 climate, vegetation, and soil variables. This allowed
us to 1) explore the environmental drivers of fungal and bacterial dominance, defined as fungal
proportion - fungi/(fungi + bacteria), where values closer to 1 indicate a higher fungal dominance
and values closer to zero indicate a greater bacterial dominance (see Methods); 2) examine the
non-linear response or pattern of fungal proportion across environmental gradients, i.e., mean
annual temperature-MAT and net primary productivity-NPP. Based on the observed relationships (by accounting for the non-linearity), we generated a quantitative spatially-explicit global map (1 km$^2$) of fungal proportion, and assessed how soil fungal and bacterial dominance varies with key climate, soil, vegetation and geographic drivers.

2. Material and methods

2.1. Data acquisition of soil microbe composition

We compiled data of abundance of soil fungi and bacteria and fungal proportion, defined as fungi/(fungi + bacteria). We focused on phospholipid-derived fatty acids (PLFA) and the data derived from PLFA reported the balance between fungal and bacterial PLFAs (Frostegård et al., 2011) can provide a valuable estimation of the comparative dominance of both functional groups. The data based on qPCR was not included because of difference in units with PLFA. With non-significant difference using data of fungal proportion and fungi : bacteria ratio, we focused on and reported the results on the fungal proportion rather than fungi : bacteria ratio because 1) The fungal proportion is insensitive to whether fungi or bacteria are the numerator (i.e. bacterial proportion = 1 – fungal proportion), and 2) fungal proportion had more spread frequency distribution and thus led to better machine learning predictions (Fig. S1). The data was compiled by a primary literature review through Google Scholar, Web of Science (http://apps.webofknowledge.com) and China National Knowledge Infrastructure Database (http://cnki.net) till 30 June, 2020 using the keywords “fungi”, “bacteria”, “abundance”, “PLFA”. To be included in our data analysis, the study had to at least have the following metadata: longitude and latitude, sampling date, sampling depth, information on land use (agriculture, tree plantations, or natural sites), units and the methods used. In total, this led to 319 references. We further used the following criteria to select eligible references and datasets: (1) when the studies were manipulative experiments, we only included the
data from “control” plots (Chen et al., 2016). (2) we standardized our efforts by focusing on all samples that were collected from the top surface soils (= 0-10/15 cm) because this layer contains greater biomass and has the majority of sample size. (3) we used the datasets based on reporting abundance with units of nmol, umol, or mol% since the majority (>90%) of datasets reported abundance. Thus, we exclude all datasets reporting biomass instead of abundance. (4) we excluded observations located in sea since our study focuses on terrestrial ecosystems. (5) we only included the datasets on soil samples derived from field experiments and thus excluded the datasets from incubation experiments. (6) some datasets reported in original references as average across sampling sites or sampling dates were included.

The criteria were carefully scrutinized by three independent researchers and this ultimately led to 179 eligible references (see Supplementary references for PLFA) used for this study. In total, we compiled a dataset of fungal proportion (n = 3224) at a global scale. The subset of data (n = 1795) with only natural ecosystems (Fig. S2a) were used to examine the potential role of land use change (see Supplementary Methods). The results showed minimal difference of the two scenarios of including all data and natural ecosystems (Fig. S2a) were used to examine its environmental controls and geospatial modelling in making the global map (Fig. 1a; Fig. S2a).

The spatial variations of fungi and bacteria ratio or fungal proportion across latitude could be influenced by either changes (increases or decreases) in abundance of fungi or bacteria or both. Thus, to better understand the biogeographic pattern of fungal and bacterial composition, we also analysed the spatial patterns of abundance of fungi and bacteria by using the abundance data with the same unit (nmol g⁻¹ PLFA). In total, our data compiling led to a final subset of 2753, and 2759
samples which were used for further analyses of abundance of fungi and bacteria, respectively (Fig. S3). As compared to the larger sample size of fungal proportion (n = 946 for all data), the data of abundance of fungi (n = 646 for all data) and bacteria (n = 647 for all data) aggregated within the 30 arc-seconds (~1-km²) pixel via an average were used for the analysis of their spatial trends across vegetation biome, vegetation type and latitude (see Supplementary Methods).

2.2. Geospatial modelling

A stack (n = 95) of ecologically relevant, global map layers including soil physical, chemical and nutrient properties, climate conditions, vegetative indices, radiation and topographic variables and anthropogenic covariates (Supplementary Table 1) were used to derive the environmental factors which could affect fungal proportion. All of these covariate map layers were standardized at 30 arc-seconds resolution (≈1km² at the equator) (van den Hoogen et al., 2019). These covariates were then derived based on the georeferenced coordinates of the soil samples aggregated at 30 arc-seconds resolution.

We used the Random Forest machine learning algorithm (see Supplementary Methods) with the derived 95 covariates to extrapolate these relationships between fungal proportion and environmental conditions across the globe and generate the first spatially-explicit, quantitative map of fungal proportion at a global scale. The strength of prediction was evaluated using k-fold cross validation (with k = 10) and the best model having high coefficient of determination and low standard deviation in the mean cross-validation were used to generate the global map of fungal proportion. The standard error sharply decreased with increasing sample size across all vegetation biomes and the analysis showed that an efficient prediction required a large sample size (n > 500) (Fig. S4). To evaluate the sensitivity, we also generate the uncertainty (standard deviation as a fraction of the mean predicted value) map of fungal proportion by using a stratified bootstrapping
procedure (van den Hoogen et al., 2019). The stratification category was the sampled biomes of each point feature (fungal proportion) with the total number collection of fungal proportion points to avoid biases. In total, 100 bootstrap iterations were used, thus generating 100 global maps of fungal proportion used to quantify statistically robust 95% confidence intervals per pixel.

2.3. Environmental drivers and statistic analysis

To examine the environmental controls of soil microbial composition at a global scale, we chose the top drivers (Chen et al., 2016; Drenovsky et al., 2010a; de Vries et al., 2012) which include soil properties, climate conditions, vegetation index and human activities (see Supplementary Methods). These variables were examined to avoid multicollinearity using a matrix of pairwise correlations to remove any variable with high correlations ($R>0.7$) with other predictor variables (Anderegg et al., 2013). Random Forest machine learning algorithm was then used to determine variable importance for each variable (Breiman, 2001). Mean decrease in accuracy (%IncMSE) and mean decrease gini (IncNodePurity) were reported and the variables with greater values of %IncMSE and IncNodePurity are more important in influencing fungal proportion. Partial functions of most important variables (MAT and NPP) were plotted using forestFloor package to examine their influences on fungal proportion.

3. Code and data availability of machine learning

The code and data of machine learning is available at https://github.com/KailiangYu/Biogeography-of-soil-microbes.git.

4. Results and discussion

4.1. Raw data patterns of fungal proportion

Globally, we observed greater than 10-fold variation in soil fungal proportion across all sites, ranging from 0.01 to 0.6 (Fig. 1b). At a global scale, we found clear latitudinal trends, with the
abundance of both fungi and bacteria increasing in high-latitude regions. Yet, the abundance of fungi increased with latitude at a greater rate than the abundance of bacteria (Fig. S5), resulting in a higher proportion of fungi in the cold, high-latitude regions. These latitudinal trends lend support to the general global patterns detected in a previous broad-scale analysis (Bahram et al., 2018) and observed in tundra and boreal forest ecosystems (mean ± 1SE: 0.23 ± 0.02; Fig. 1b). As such, the highest fungal dominance was high elevation and cold grasslands (i.e., Montane grasslands) with large soil organic C (SOC) content generally harbor higher proportion of fungi, relative to bacteria (Fig. 1b).

Within similar climates, soil fungal and bacterial abundance as well as fungal proportion was greatest in ecosystems harboring woody vegetation compared to grasslands and managed (agricultural) ecosystems (Fig. S6). This finding is consistent with the idea that ecosystems dominated by woody plants generate lignified, more recalcitrant and nutrient poor soil C inputs that characteristically favor fungal dominance (Fierer et al., 2009; Strickland and Rousk, 2010), and have a biomass stoichiometry better suited to low nutrient environments (Waring et al., 2013). But we stress that this link of belowground soil microbial composition (fungi vs bacteria) with aboveground plant community composition (woody plants vs grasses) can be complex, non-linear and even divergent, as demonstrated by the non-existence of woody plants in grasslands and scarcity of grasses in forests but with well mixed fungi vs bacteria abundances. This raises the curiosity whether the interactions, associations or couplings of belowground soil microbial composition vs aboveground plant community composition are stronger in ecosystems where woody plants and grasses interact or coexist (i.e., savannas) (Yu and D’Odorico, 2015). It also remains unclear how this coupling could improve our understanding of ecosystem carbon cycling and other services.
Management of agricultural ecosystems often disrupts soil fungal networks (i.e. tillage, frequent dry/wet cycles due to irrigation, machine operations, etc.), which decreases the abundance of fungi relative to bacteria in agricultural soils (Fig. S6) (Drenovsky et al., 2010b; Jangid et al., 2011; Waldrop et al., 2017). A central concern in agricultural ecosystems is the tradeoff of increased food production to feed the increasing population vs the decreased soil carbon storage to accelerate the global climate change (Sanderman et al., 2017). This study showed the higher bacterial abundance relative to fungal abundance in soils of agricultural lands where soil carbon storage is low; this corresponds with the global trends of bacterial dominance in low latitude where soil carbon storage is low. These results suggest the potential strong but complex interactions and feedbacks of soil microbial composition and soil functions (i.e., soil carbon storage) (Bardgett et al., 2008), while the mechanistic links need further studies.

4.2. Drivers of fungal proportion

Globally, the fungal proportion in soil can be predicted by few primary environmental drivers (Fig. 2; Fig. S7). Specifically, mean annual temperature (MAT) and primary productivity (NPP) were strong determinants of fungal dominance. The responses of fungal proportion to both MAT and NPP were strongly non-linear, with warmer, more productive regions of the world (i.e. tropical forest biomes) showing lower dominance of fungi as compared to colder, less productive ecosystems (i.e. boreal forest and tundra biomes, Fig. 3; Fig. S8). This pattern is consistent with the idea that fungi and bacteria represent slow vs. fast soil energy channels, respectively (Crowther et al., 2019; Malik et al., 2016), a concept with a long history in soil ecology (Moore et al., 2003; Moore and William Hunt, 1988). This finding is important because it could potentially link the belowground slow – fungi vs fast – bacteria energy channels with aboveground plant slow growth rates – woody plants vs fast growth rates – grasses, while the linkage could be complex, non-linear
or even divergent. The fast vs slow concept or spectrum have fundamentally improved the understandings and predictions of land carbon storage across resource gradient or under global change. The faster growth could be typically trade off with higher mortality or heterotrophic respiration with resource (i.e., CO$_2$) enriched conditions (Jiang et al., 2020; Terrer et al., 2021; Yu et al., 2019), thus constraining land carbon storage. This raises the question of how the belowground fast vs slow energy channels and the aboveground fast vs slow growth spectrum could be potentially linked or integrated to assess land carbon storage.

Temperature can affect soil microbial composition in complex ways, via directly physiology or via indirectly soil substrate (Romero-Olivares et al., 2017). Previous studies have shown the non-linear response of soil fungal and bacterial ratio to soil substrate (Waring et al., 2013). The non-linear trends of the temperature sensitivity (Q10) of soil organic C decomposition, as regulated by soil fungal and bacterial ratio, were also found along latitude (Wang et al., 2018). Other environmental variables such as soil C to nitrogen ratio (C:N) have previously been found to be important drivers in influencing fungal proportion within local and regional scale analyses (Fierer et al., 2009; Waring et al., 2013). Our results suggest a more complicated relationship between fungal proportion and the soil C:N. In the low range of soil C:N values, fungal proportion decreased with soil C:N (Fig. S9a), suggesting the likely role of site-specific differences (i.e., climate or plant community) in out-weighting the influence of N availability (Soares and Rousk, 2019). Aside from these ecosystems, we observed a positive relationship between fungal proportion and soil C:N at a global scale, consistent with previous work at local and regional scales (Strickland and Rousk, 2010; Waring et al., 2013). Additionally, pH has been thought as a critical driver of microbial diversity and biomass in soils. At local scales, previous studies reported either no relationship, a negative correlation or convex curve between fungal and bacterial ratio and soil
pH (Rousk et al., 2009, 2010; de Vries et al., 2012). Our global scale analysis suggests a convex relationship between fungal proportion and soil pH, with fungi dominating only within a narrow pH range (<5-6) (Fig. S9b).

4.3. Biogeographic pattern from the machine learning model

Across all samples, the machine learning model was able to predict the variation in fungal and bacterial dominance with high predictive accuracy ($R^2 = 0.43/0.35$ in 10-fold cross validation; $R^2 = 0.92/0.91$ in final model; Fig. S10a-b). By extrapolating these relationships across terrestrial ecosystems, we could identify clear global trends in fungal dominance. Despite these general global scale patterns of increase in fungi dominance with latitude, our models also revealed regional contingencies that diverge from the global trends (Fig. 4a; Fig. S11a). For instance, Northeastern Europe is dominated by woody vegetation and exhibits high fungal proportion, while the United Kingdom and northern Kazakhstan have much lower fungal proportion despite being at comparable latitude, likely because these areas are dominated by herbaceous vegetation with lower lignin content than in woody tissues. Tibetan alpine grasslands are at comparatively much lower latitude but have high values of fungal proportion in part due to very high SOC stocks and cold temperatures. Model predictions of fungal proportion had high uncertainty in dry regions (i.e., Northern and Southern Africa, Australia, Western USA, eastern Mongolia) (Fig. 4b; Fig. S11b), presumably because of the low sample size in drylands and/or complex response of fungi and bacteria to water availability (Fierer et al., 2009; Strickland and Rousk, 2010). Indeed, our datasets are mostly concentrated to US, Europe and East Asia, thus highlighting the data gaps at tropical and boreal biomes. Even for the temperate biome, there were data gaps in west Australia and central Asia. Because of the unbalanced sample distribution, we also used a bootstrapping strategy (100 iterations) by randomly sampling 90% data with replacement. The results showed
the similar spatial patterns of fungal proportion (Fig. S12a) and uncertainty (Fig. S12b) as the
scenario of using full dataset without bootstrapping.

Our study differs from a previous study (He et al., 2020) in several aspects including
sample size (n > 3000), spatial resolution (1km$^2$), consideration of non-linearity (through random
forest analysis), soil depth (soil surface 0-10/15 cm). We also note that our analysis sticks to the
original data of abundance derived from PLFA instead of converting abundance to biomass.
Conversion of abundance to biomass needs the conversion factor, which has large uncertainty
(Frostegård et al., 2011; Klamer and Bååth, 2004). Our high resolution map would allow
representation of microbial-mediated mechanisms at fine scales to link with ecosystem functions.
For instance, the significant functional differences between fungi and bacterial mean that the
relative dominance of fungi vs. bacteria is likely to influence a wide range of ecosystem
functions such as C use efficiency (CUE) of the decomposer community (Six et al., 2006; Soares
and Rousk, 2019) and enzymatic activity in soil N vs P acquisition (Caldwell, 2005; Crowther et
al., 2019). At fine, local or even regional scales, these relationships between soil microbial
composition and ecosystem functions could only be well identified using fine scale maps of soil
microbial composition.

4.4. Implications and limitations of this study
It is generally accepted that the soil microbiome exerts major control over soil processes, and in
turn ecosystem functioning, and by extension the global biogeochemical cycles (Bahram et al.,
2018; Crowther et al., 2019; Van Der Heijden et al., 2008; Jenny, 1941). Fungi and bacteria
represent most of the diversity of life on Earth (Bardgett and van der Putten, 2014; Locey and
Lennon, 2016). Yet, inclusion of fungal and bacterial abundance into quantitative ecosystem and
Earth system models has been hindered by the paucity of information about organisms at
appropriate spatial scales. Here, we impose a global top-down constraint on the broad composition
of soil microbial life. By doing so, we hope to empower microbial, ecosystem and Earth-system
scientists to consider how this broad constraint on the soil biodiversity may inform and transform
how we understand terrestrial ecosystem functioning. As we develop a spatially-explicit
understanding of the global soil community, we will be able to better parameterize and benchmark
our predictions about the rate and efficiency of carbon turnover in soil and the feedbacks to
ongoing climate change.

Despite of the progress made in this study, here we clarify two limitations on this study.
First, our study highlights the data gaps in fungal proportion prediction in low latitude – tropical
biome and high latitude – boreal biome (i.e., boreal forests and tundra). Tropical vs boreal
biomes are hotspots or debated regions with their relative capacity and capability to sequestrate
atmospheric CO₂ and mitigate climate change in an increasingly changing climate (Schimel et
al., 2015; Tagesson et al., 2020). They are also regions with striking differences of soil
microbial composition (fungal proportion), plant communities and soil carbon storage, thus
suggesting their potentially strong interactions and feedbacks in these regions (Bardgett et al.,
2008). Boreal biome contains large amount of soil organic carbon which could be sensitive to
global change (i.e., warming), whereby soil microbial community (i.e., total biomass or the
relative abundance of soil fungi and bacteria) could play an essential role. Second, microbial
biomass (C) is more relevant to be linked with soil carbon cycling and carbon stock in term of
their own contribution by living carbon pools and the impacts of its microbial necromass (Liang
et al., 2019), while the conversion factor of converting abundance into biomass across space is
currently not available. To mechanistically and explicitly incorporate soil microbial composition
into biogeochemical models, the biogeographic patterns of abundance or biomass of each major
group (fungi vs bacteria), the relative ratio within fungi (i.e., saprotrophic fungi, arbuscular

mycorrhiza fungi vs ectomycorrhizal fungi) and/or bacteria (i.e., gram positive bacteria vs gram

negative bacteria) would also be critical in view of their striking functional difference (Averill et

al., 2014; Crowther et al., 2019). These knowledge gaps highlight the urgent research needs in

these new research endeavors with the increasing availability of datasets.

5. Conclusions

This study used a global scale dataset (>3000 distinct observations of soil fungal and bacterial

abundance) in the top soil surface (up to 15 cm) (>3000 distinct observations of soil fungal and

bacterial abundance) to generate the first quantitative and spatially high resolution (1 km²) explicit

maps of soil fungal and bacterial relative abundance across global terrestrial ecosystems. Our

machine learning approach (random forest) enabled us to link the variation in fungal proportion to

global variation in climate, soil, vegetation and other environmental drivers, whilst accounting for

the interactions and non-linearities among them. We found the striking latitudinal trends where

fungal dominance increases in cold and high latitude environments with large soil carbon stocks.

The fungal proportion in soil can be predicted by few primary environmental drivers – temperature

and NPP with strong non-linear effects of temperature and NPP. We demonstrated that fungi and

bacteria represent slow vs fast energy channels, whereby they dominate in regions of low MAT

and NPP vs high MAT and NPP, respectively. Overall, our spatially-explicit model would enable

us to explicitly represent the different contributions of fast - bacterial vs. slow – fungal energy

channels in spatially-explicit biogeochemical models, with the potential to enhance the accuracy

of soil carbon turnover and carbon storage predictions. We further highlight the data gaps in tropical

and boreal regions and needs of future research endeavors in generating high resolution
biogeographic patterns of biomass of each major microbial group, the relative biomass ratios across and within major microbial groups.
References


Tedersoo, L., Bahram, M., Põlme, S., Koljalg, U., Yorou, N. S., Wijesundera, R., Ruiz, L. V.,
Vasco-Palacios, A. M., Thu, P. Q., Suija, A., Smith, M. E., Sharp, C., Saluveer, E., Saitta,
A., Rosas, M., Riit, T., Ratkowsky, D., Pöldmaa, K., Piepenbring, M., Phosri,
C., Peterson, M., Parts, K., Pärtel, K., Otsing, E., Nouhra, E., Njouonkou, A. L., Nilsson, R.
H., Morgado, L. N., Mayor, J., May, T. W., Majuakim, L., Lodge, D. J., Lee, S., Larsson, K.
H., Kohout, P., Hosaka, K., Hiesalu, I., Henkel, T. W., Harend, H., Guo, L. D., Greslebin,
A., Grelet, G., Geml, J., Gates, G., Dunstan, W., Dunk, C., Drenkhan, R., De Kesel, A.,
Dang, T., Chen, X., Buegger, F., Brearley, F. Q., Bonito, G., Anslan, S., Abell, S.
and Abarenkov, K.: Global diversity and geography of soil fungi, Science (80- ).,
Groenigen, K. J., Keenan, T. F., Sulman, B. N., Stocker, B. D., Reich, P. B., Pellegrini, A. F.
A., Pendall, E., Zhang, H., Evans, R. D., Carrillo, Y., Fisher, J. B., Van Sundert, K., Vicca,
S. and Jackson, R. B.: A trade-off between plant and soil carbon storage under elevated
de Vries, F. T., Manning, P., Tallowin, J. R. B., Mortimer, S. R., Pilgrim, E. S., Harrison, K. A.,
Hobbs, P. J., Quirk, H., Shipley, B., Cornelissen, J. H. C., Katge, J. and Bardgett, R. D.:
Abiotic drivers and plant traits explain landscape-scale patterns in soil microbial
M., Dick, R., Howard, D., Wylie, B. and Grace, J. B.: The interacting roles of climate, soils,
and plant production on soil microbial communities at a continental scale, Ecology,
Wang, Q., Liu, S. and Tian, P.: Carbon quality and soil microbial property control the latitudinal
pattern in temperature sensitivity of soil microbial respiration across Chinese forest
Waring, B. G., Averill, C. and Hawkes, C. V.: Differences in fungal and bacterial physiology
alter soil carbon and nitrogen cycling: Insights from meta-analysis and theoretical models,
Wieder, W. R., Bonan, G. B. and Allison, S. D.: Global soil carbon projections are improved by
modelling microbial processes, Nat. Clim. Chang., 3(10), 909–912,
Wieder, W. R., Allison, S. D., Davidson, E. A., Georgiou, K., Hararuk, O., He, Y., Hopkins, F.,
Luo, Y., Smith, M. J., Sulman, B., Todd-Brown, K., Wang, Y. P., Xia, J. and Xu, X.:
Explicitly representing soil microbial processes in Earth system models, Global
Yu, K. and D’Odorico, P.: Hydraulic lift as a determinant of tree-grass coexistence on savannas,
Peñuelas, J., Cailleret, M., Levanic, T., Gessler, A., Schaub, M., Ferretti, M. and Anderegg,
W. R. L.: Pervasive decreases in living vegetation carbon turnover time across forest climate
Zhou, J. and Yang, Y.: The microbe-mediated mechanisms affecting topsoil carbon stock in
Figure 1. Map of sample locations and fungal and bacterial abundance and fungal proportion data.

a. Sampling sites. A total of 3224 samples were collected and aggregated into 943 1-km² pixels that were used for geospatial modelling.

b. The median and interquartile range of abundance of fungi and bacteria and fungal proportion across vegetation biomes. Tundra and boreal forest, Mediterranean and desert have low sample sizes (<25) and thus were combined.
Figure 2. Mean decrease in accuracy (%IncMSE, mean and SD, a) and mean decrease gini (IncNodePurity, mean and SD, b) estimated from 1000 simulations of random forests. This is used to evaluate the importance of top environmental drivers on proportion of fungi derived from ‘all’ dataset.
Figure 3. Fungal proportion is primarily associated with climate-mean annual temperature (MAT) and net primary productivity (NPP). a–b. Partial feature contributions of primary environmental variables (a, MAT; b, NPP) to fungal proportion. c, Partial feature contributions of primary environmental variable interactions (MAT vs NPP) to fungal proportion.
Figure 4. Global map of fungal proportion (a) and bootstrapped (100 iterations) coefficient of variation (b) at the 30 arcsec (approximately 1 km²) pixel scale. Bootstrapped coefficient of variation is standard deviation divided by the mean predicted value as a measure of prediction accuracy. Sampling was stratified by biome.
Author contributions

KLY and TWC designed the project. KLY built the PLFA datasets with help from JVDH and ZQW. KLY performed the analysis with inputs from DR and CA. KLY, CA, and TWC wrote the paper with revisions from all other coauthors. GRS, RED, KMS, FM, MPW, YHY, FTDV, RDB, PM, FB, SGB, EMB, CG, QKW, LM, BD C, XJH, WZT, ST, AH, JAB contributed to PLFA datasets.