Plant attributes explain the distribution of soil microbial communities in two contrasting regions of the globe.

DOI:
10.1111/nph.15161

Document Version
Accepted author manuscript

New Phytologist

Plant attributes explain the distribution of soil microbial communities in two contrasting regions of the globe.

Journal:	*New Phytologist*
Manuscript ID	NPH-MS-2018-26416.R1
Manuscript Type:	MS - Regular Manuscript
Date Submitted by the Author:	12-Mar-2018

Complete List of Authors:	Delgado-Baquerizo, Manuel; University of Colorado Boulder, Cooperative Institute for Research in Environmental Sciences. University of Colorado, Boulder Fry, Ellen ; The University of Manchester, School of Earth and Environmental Sciences Eldridge, David; University of New South Wales, School of Biological, Earth and Environmental Sciences de Vries, Franciska; The University of Manchester, Faculty of Life Sciences Manning, Peter; Biodiversity and Climate Research Centre, Biodiversity and Climate Research Centre Hamonts, Kelly ; University of Western Sydney, Hawkesbury Institute for the Environment Kattge, Jens; Max-Planck Institute, Biogeochemistry Bönisch, Gerhard; Max-Planck-Institute, Biogeochemistry Singh, Brajesh; University of Western Sydney, Hawkesbury Institute for the Environment Bardgett, Richard; The University of Manchester, Faculty of Life Sciences

Key Words:	Plant functional traits, Bacteria, Fungi, Biodiversity, Terrestrial ecosystems

Plant attributes explain the distribution of soil microbial communities in two contrasting regions of the globe.

Manuel Delgado-Baquerizo[1,2]*, Ellen L. Fry[3], David J. Eldridge[4], Franciska T. de Vries[3], Pete Manning[5], Kelly Hamonts[6], Jens Kattge[7], Gerhard Boenisch[7], Brajesh K. Singh[6,8], Richard D. Bardgett[3]*

[1]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO 80309.

[2] Departamento de Biología, Geología, Física y Química Inorgánica, Escuela Superior de Ciencias Experimentales y Tecnología, Universidad Rey Juan Carlos, c/ Tulipán s/n, 28933 Móstoles, Spain.

[3]School of Earth and Environmental Sciences, Michael Smith Building, The University of Manchester, Oxford Road, Manchester M13 9PT, UK.

4Centre for Ecosystem Science, School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, New South Wales 2052, Australia.

[5]Senckenberg Biodiversity and Climate Research Centre, Senckenberganlage 25, Frankfurt, Germany.

[6]Hawkesbury Institute for the Environment, Western Sydney University, Penrith, 2751, New South Wales, Australia.

[7]Max Planck Institute for Biogeochemistry, P.O. Box 10 01 64, 07701 Jena, Germany.

[8]Global Centre for Land-Based Innovation, Western Sydney University, Penrith South DC, NSW 2751, Australia.

***Authors for correspondence:**

Manuel Delgado-Baquerizo. E-mail: M.DelgadoBaquerizo@gmail.com

Richard D. Bardgett. E-mail: richard.bardgett@manchester.ac.uk

Summary

- We lack strong empirical evidence for links between plant attributes (plant community attributes and functional traits) and the distribution of soil microbial communities at large spatial scales.
- Using datasets from two contrasting regions and ecosystem types in Australia and England, we report that aboveground plant community attributes such as diversity (species richness), cover, and functional traits can predict a unique portion of the variation in the diversity (number of phylotypes) and community composition of soil bacteria and fungi, that cannot be explained by soil abiotic properties and climate. We further identify the relative importance and evaluate the potential direct and indirect effects of climate, soil properties and plant attributes in regulating the diversity and community composition of soil microbial communities.
- Finally, we deliver a list of examples including common taxa from Australia and England that are strongly related to specific plant traits, such as specific leaf area index, leaf N, and N fixation.
- Together, our work provides new evidence that plant attributes, especially plant functional traits, can predict the distribution soil microbial communities at the regional scale and across two hemispheres.

Key words: Plant functional traits; Bacteria; Fungi; Biodiversity; Terrestrial ecosystems.

Introduction

Soil microbial communities play important roles in driving multiple ecosystem functions and services including climate regulation, nutrient cycling, water regulation, and food and fibre production (Bardgett & van der Putten 2014; Delgado-Baquerizo *et al.* 2017). Previous studies have provided evidence that abiotic factors such as climate (Maestre *et al.* 2015; Zhou *et al.* 2016) and soil chemical properties (pH, soil carbon and nutrients; Lauber *et al.* 2009; Maestre *et al.* 2015; Tedersoo *et al.* 2014) are the main predictors of the distribution of soil microbial communities across the globe. Much less is known, however, about the role of plant attributes including community-level attributes, such as diversity and cover, and functional traits in regulating the distribution of soil microbial communities at regional scales (i.e., hundreds of kilometers). Identifying the relative importance of plant attributes in predicting the distribution of soil microbial communities is of paramount importance, as plant communities are highly sensitive to climate, N enrichment, and land use intensification (Allan *et al.* 2015; Le Bagousse-Pinguet *et al.* 2017), and resulting shifts in vegetation might have cascading effects on the diversity and functioning of soil microbial communities (García-Palacios *et al.* 2016; Deraison *et al.* 2015; Le Bagousse-Pinguet *et al.* 2017).

The identity of plant genotypes or lichen species, major biological components of cold and warm deserts, has recently been highlighted as a major predictor of the distribution of soil bacteria at the local scale (Leff *et al.* 2017; Liu *et al.* 2017). Much less is known, however, about the role of other plant attributes such as plant diversity (number of species), plant cover, and plant functional traits as predictors of the diversity (number of phylotypes) and community composition of soil bacterial and fungal communities. While empirical evidence is lacking, the conceptual links among plant attributes and microbial community composition are reasonably well established (Hooper *et al.* 2000; Wardle *et al.* 2004; Lavorel 2013; Bardgett 2017). Plant community attributes and functional traits can directly affect soil microbes by altering the quality (which can be represented by measures such as specific leaf area –SLA– and tissue nutrient content; Cornelissen *et al.* 2003) and quantity of resource inputs *via* litter and detritus (which can be represented via measures such as plant biomass and canopy cover). Both the quantity and quality of resources have been demonstrated to regulate the diversity and community composition of soil microbial communities (Hooper *et al.* 2000; Wardle *et al.* 2004; Schneider *et al.* 2012; Zhou *et al.* 2015). Moreover, microcosm studies have demonstrated that changes in litter quality during decomposition strongly influence the composition and diversity of soil microbial communities (Schneider *et al.* 2012; Zhou *et al.* 2015). Plant diversity

could also alter the distribution of microbial communities by promoting a greater diversity of litter types, promoting niche differentiation and resource partitioning (Wardle *et al.* 2004; Gould *et al.* 2016), and facilitating the existence of multiple mutualism (e.g., mycorrhizae and rhizobia) and antagonistic (e.g., plant-pathogen) interactions with soil microbes. Other effects on plant community attributes and functional traits of soil microbes include changes in habitat conditions (e.g., soil structure, shading, water regulation) and soil chemistry (e.g., root exudation and nutrient uptake), which are both known to strongly affect the structure and function of microbial communities (Bardgett 2017; Le Bagousse-Pinguet *et al.* 2017).

Plant traits have been used to predict broad-scale shifts in the biomass of fungi and bacteria at the individual plant (Orwin 2010), community (Legay *et al.* 2014) and regional scale (hundreds of kms; de Vries *et al.* 2012; Grigulis *et al.* 2013). Further, plant functional traits are also known to influence the abundance of particular groups of soil microorganisms, such as mycorrhizal fungi (e.g., López-García et al. 2014; 2017), and specific groups involved in N cycling, such as archaeal ammonia oxidisers (Moreau *et al.* 2015; Thion *et al.* 2016). However, the role of plant functional traits in regulating the diversity (number of phylotypes; richness) and community composition (relative abundance of phylotypes) of soil bacteria and fungi remains relatively poorly understood. Recent studies that have evaluated the link between plant functional traits and the taxonomic diversity of soil microbial communities at a local scale have revealed no clear relationships, despite strong effects of plant species identity (Barberán *et al.* 2015; Fry *et al.* 2017; Leff *et al.* 2018). However, whether plant traits can explain variation in microbial diversity and composition at larger spatial scales, and across regions and ecosystem types at the global scale, remains largely unexplored. This is despite the suggestion that the relationship between plant traits and the diversity and community composition of soil microbial communities are likely to be strongest at regional scales (hundreds of kms) where taxonomic and trait diversity is considerable, and the effect of plant attributes on microbial communities could be statistically detected (Wardle 2005). We posit, therefore, that regional scale variation in plant traits will be strongly correlated with changes in diversity and community composition of bacterial and fungal communities.

Here, we evaluate the role of plant attributes, including (1) plant community attributes (plant diversity and cover) and (2) functional traits, in predicting the distribution of community composition and diversity of soil bacteria and fungi in two contrasting ecosystem types located in two different hemispheres. Given the strong theoretical link between plant attributes and soil

microbial communities, we hypothesized that plant attributes would explain additional variation in microbial community composition and diversity that is unaccounted for by key drivers such as climate or soil properties. Such hypothesis should be valid across regions differing markedly in climate, vegetation and soils. As such, we used two contrasting regional datasets (hundreds of kms) from Australia and England, which included natural forests and a range of grassland types (Fig. S1; de Vries *et al.* 2012; Delgado-Baquerizo *et al.* 2016c). The English dataset has previously been used to identify the role of plant traits in predicting the biomass of fungi and bacteria and their relative abundance (de Vries *et al.* 2012), but the role of plant attributes as predictors of microbial community composition and diversity remain unaddressed. Our intention was not to merge the two data sets, which differed in their sampling design, vegetation, soil and climatic conditions, and plant trait information, but to test our hypotheses across two regions with markedly different vegetation, climate and soils. In doing so, we provide a general and robust test of the importance of plant traits for explaining regional scale variation in the composition and diversity of soil microbial communities across a range of different ecosystems.

Material and Methods

Study sites

We used two separate regional datasets (Fig S1). The first included micro-habitat level information on three distinct vegetation classes micro-habitat (grasses, N-fixing shrubs and trees) across twenty natural forest locations from eastern Australia (Fig. S1) (Delgado-Baquerizo *et al.* 2016c). Locations in Australia are distributed across a >1000 km environmental transect (Fig. S1). These sites were originally chosen to represent a wide range of aridity conditions, from arid to humid forest communities, and with perennial vegetation cover ranging from 18 to 98%. These ecosystems consistently had independent patches of vegetation dominated by trees (*Eucalyptus* spp.), N-fixing shrubs (*Acacia* spp.), and perennial grasses (*Rhytidosperma* spp.). Total annual precipitation and mean temperature ranged from 280 mm to 1167 mm and from 12.8º C to 17.5ºC, respectively. The second dataset was from England and included plot-level information from 180 grasslands varying in management intensity (unimproved, semi-improved and improved grasslands) and covering the main acid, calcicolous, mesotrophic, and wet grassland types of the United Kingdom (see de Vries *et al.* 2012 and Manning et al. 2015). Locations in England spanned all major grassland regions of the country, distributed across a north to south transect of approximately 500km^2. Across all grasslands,

total annual precipitation and mean temperature ranged from 573 mm to 1355 mm and from 6.3º C to 10.2ºC, respectively.

Soil sampling

Soil samples from the top ~7cm were collected in Australia and England as explained in Appendix S1. In brief, in Australia, three soil cores were collected under the three most common plant functional groups micro-habitat: grasses (*Rhytidosperma* genus including species *R. caespitosum, R. pilosum* or *R. racemosum*), N-fixing shrubs (*Acacia* genus including species *A. dealbata, A. decora, A. genistifolia, A. implexa* or *A. wilhelmiana*) and trees (*Eucalyptus* genus including species *Eucalyptus largiflorens. E. microcarpa, E. populnea, E. rossii, E. socialis* or *E. tereticornis*). The same genus of these plant taxa was present across all plots. A total of 60 soil samples (20 sites x three micro-habitats) were collected. Sampling was conducted in March (2014). In England, soil samples were collected June-July 2005 from 180 sites covering the main grassland habitat classifications in the UK, namely acid, calcicolous, mesotrophic, and wet grasslands (De Vries et al. 2012; Manning et al. 2015). The survey covered a wide range of grassland communities within each habitat type and included a total of 256 grassland plant species, confirming the representative nature the national survey (Rodwell 1992). In terms of dominant graminoid species, unimproved acid grasslands were typically dominated by *Festuca ovina*, *Deschampsia flexuosa* and *Agrostis capillaris*, calcicolous grasslands were typically dominated by *Festuca rubra*, *Festuca ovina*, *Bromus erectus* and *Carex flacca*; mesotrophic grasslands were typically dominated by *Agrostis canina*, *Festuca rubra*, and *Poa trivialis*; and wet grasslands were dominated by *Carex distichia* and *Molinia caerulea*. Semi-improved grasslands in all four habitat types grasslands became increasingly dominated by *Lolium perenne*, and improved grasslands also strongly promoted *Holcus lanatus* in acid and mestrophic grasslands, *Poa trivialis* in calcicolous grasslands and mesotrophic grasslands, and *Agrostis stolonifera* in wet grasslands.

Climate and soil properties

In all cases, we obtained information on mean annual temperature and Aridity Index (positively related to precipitation)(1km) for the surveyed sites from the Worldclim database (www.worldclim.org). Moreover, we obtained information on total soil organic C, total N and P and pH as explained in Appendix S1.

Plant attributes

The Australian and English samples contain shared information on five plant attributes: diversity (species richness), percentage plant cover, Specific Leaf Area (SLA), leaf N content and N fixation (proportion of N fixing plants in England and presence of *Acacia* species –the only N-fixer micro-habitat– in the Australian dataset). In addition, the two datasets include a subset of distinct plant functional traits such as leaf C and P, plant height, canopy width and canopy height in the Australian dataset and leaf dry matter content (LDMC) and relative growth rate (RGR) in the English dataset. Both datasets were originally independently generated and with each study designed to explicitly include variables that were hypothesized to account for variation in soil properties and functions within their respective regions. For example, plant functional traits such as plant height, canopy width and canopy height may explain differences in microbial communities in forests from Australia, but not in English grasslands where they vary little. Detailed information on how plant traits were measured in these two datasets is available in Appendix S1.

Soil microbial community

Soil DNA was extracted from both sets of soil samples using the Powersoil® DNA Isolation Kit (Mo Bio Laboratories, Carlsbad, CA, USA). In England, 161 from 180 samples were included in further analyses due to DNA amplification problems. Amplicons targeting the bacterial 16S rRNA gene and fungal ITS2 region were sequenced using the Illumina MiSeq platform and the 341F/805R (bacteria) and FITS7/ITS4 (fungi) primer sets (Appendix S1). Bioinformatic analyses were conducted using UPARSE and MOTHUR (Appendix S1). Operational Taxonomic Units (OTU) were picked at 97% sequence similarity in both cases. The resulting OTU abundance tables were rarefied. As these analyses were done together for the Australian and English datasets, OTU identities are directly comparable between them.

Statistical analyses

All statistical analyses were independently done for each dataset (Australia and England) and microbial group (bacteria or fungi). First, we evaluated the relationship between bacterial and fungal community dissimilarity with the dissimilarity of plant attributes (plant cover, diversity and functional traits) across plots. To do this, we calculated Bray–Curtis dissimilarities to generate independent community distance matrices at the OTU level for bacterial and fungal communities in the Australia and English datasets. Similarly, the Euclidean distance was used to independently create a matrix of distance for plant drivers for the Australia and English datasets. We then

correlated the matrix of plant community attributes and traits distances to the dissimilarity matrix of bacteria and fungi in Australia and England using Mantel test correlations.

Second, we used two independent approaches to assess whether plant attributes can predict a unique portion of the variation of soil microbial diversity and community composition. We first conducted Variation Partitioning (R package Vegan; Oksanen *et al.* 2015) as an exploratory analysis to identify whether plant attributes: (1) plant functional traits; and (2) plant diversity and cover, explain a unique portion of the variation in microbial diversity and composition, after accounting for key microbial drivers such as location (latitude and longitude), climate (aridity index and mean annual temperature) and soil properties (total C, N and P and pH; Table 1).

We then used a multi-model inference approach based on information theory and non-parametric distance-based linear regressions (DISTLM; McArdle & Anderson 2001) to evaluate whether plant attributes (plant cover, diversity and traits) explained a unique proportion of the variation in bacterial and fungal diversity (richness; number of phylotypes) and community composition (at the OTU level) after accounting for other important microbial drivers such as soil properties (total C, N and P and pH) and climate (aridity index and mean annual temperature). Location (latitude and longitude; Table 1) was included in all models to account for spatial autocorrelation. The Euclidean and Bray-Curtis distances were used for microbial diversity and composition, respectively in these analyses. We carried out these analyses using the PERMANOVA+ for PRIMER statistical package. We ranked all the models that could be generated with different combinations of our independent variables according to the second-order Akaike information criterion (AICc) and considered a ΔAICc>2 threshold to differentiate between two substantially different models (Burnham & Anderson 2002). Differences<2 in AICc between alternative models indicate that they do not differ significantly in their explanatory power. The full statistical reasoning for this approach can be found elsewhere (e.g., Zuur et al. 2009). We then selected the best of those models including all parameters in Table 1, and compared the AICc of the best model with competing models containing: (1) all parameters in model A, but plant functional traits (Model B); (2) included all parameters in model A, but plant community attributes (cover and PDiv) (Model C); or (3) all parameters in model A but plant functional traits and community attributes (Model D) (Table 2).

Third, we conducted two independent analyses to assess the importance of plant attributes, soil properties, and climate as predictors of soil microbial community composition and diversity. We

first used Random Forest analyses (Archer 2016), as explained in Delgado-Baquerizo et al. (2016a), to identify the most important predictors (Table 1) of bacterial and fungal diversity and community composition. For simplicity, and given that, at this point, we were interested in the responses of the entire microbial community composition rather than on single taxa, in the case of bacterial and fungal community composition, we conducted these analyses on the axes of a NMDS conducted on bacterial and fungal composition data at the lowest taxonomic rank (Fig. S3, stress = 0.08 and = 0.12, respectively). We then used Structural Equation Modelling (SEM) to build a system-level understanding of the major direct and indirect effects of climate, soil properties, and plant attributes on the composition and diversity of soil bacteria and fungi (*a priori* model available in Fig. S2 and Appendix S1 for details). For simplicity, and due to the data constraints of fitting SE models with many paths, we only included in these models those variables that were identified as major predictors of the diversity and composition of bacteria from the best models of our distance-based multi-model approach. Importantly, in general, similar variables were identified as important predictors in our Random Forest results (see below). Therefore, although we used the same *a priori* model in all cases (Fig. S2), SE models conducted for the different datasets contain different predictors and were constructed independently. The only exception to this was latitude and longitude, which were included in all the models to account for spatial autocorrelation in our models, and to represent other variables that might co-vary with latitude and longitude but which are not included in our analyses. Analyses were performed independently for each dataset. With a good model fit, we were then free to interpret the path coefficients of the model and their associated *P* values. In the case of England we accounted for any effect from management practices on our results, by repeating the SEM analyses using the residuals from a one-way ANOVA in which management practice (managed, intermediate intensity managed, and intensively managed) were treated as a fixed factor and bacterial diversity or composition as a response variable (i.e. residuals of bacterial diversity or composition). This results in a more conservative test of plant effects on microbial communities as functional traits are known to covary with management (see de Vries *et al.* 2012).

Finally, we used Random Forest analysis (Archer 2016) to identify the microbial phylotypes that were most strongly associated with a particular plant trait. We focused on shared dominant taxa (>50 reads across all samples) between Australia and England for these analyses. Moreover, we focused on shared plant community attributes (cover and diversity) and functional traits (SLA, leaf

N and N fixation), and microbial phylotypes for the Australia and English datasets. Analyses were conducted independently for the Australia and English datasets and for fungal and bacterial communities. For both datasets, we first identified the top unique and shared (significance; $P < 0.05$) microbial phylotypes accounting for the variation of particular plant traits (i.e., those microbial phylotypes that are selected from Random Forest model as important predictors of each plant trait). The reserved approach enabled us to identify particular phylotypes that consistently characterize particular plant attributes in both Australia and England. We then conducted Spearman correlations among shared phylotypes in Australia and England with particular plant traits for which these phylotypes are good predictors. The major goal for these analyses is to provide a list of examples that could make the basis of experimental studies to look at the links between particular microbial phylotypes and plant attributes in more detail.

Results

Microbial and plant attributes in Australia and England

The Australia and English datasets varied markedly in fungal and bacterial community composition (Figs. S3-S4). Proteobacteria and Acidobacteria were the dominant bacterial phyla in England, while Actinobacteria was the dominant phylum in Australia (Fig. S4). In both datasets, the fungal community was dominated by Ascomycota (Fig. S4), with Zygomycota and Basidiomycota being the second most abundant fungal phyla in England and Australia, respectively. Fungal diversity was greater in the Australian dataset, but bacterial diversity did not differ between datasets (Fig. S3). See Appendix S1 for details on the statistical analyses conducted to evaluate these general patterns in microbial diversity and composition. In both datasets, there was considerable heterogeneity in soil properties and microbial communities. For example, in Australia, pH and soil C ranged from 4.8-8.9 and 1.3-12.3%, respectively (Table 1). Similarly, bacterial and fungal diversity ranged from 955-2833 and 489-813 phylotypes, respectively. In England, soil pH and C ranged from 4.1-7.8 and 1.4-12.8%, respectively (Table 1), and bacterial and fungal diverity from 820-3329 and 243-763 phylotypes, respectively.

Plant attributes varied greatly among plots in both datasets. For example, plant cover ranged from 78.3 to 249.5% (i.e., due to multiple vegetation layers in grassland communities) in England and from 18.3 to 98.3% in Australia. Plant species diversity ranged from 2 to 36 species across grassland plots in England and from 11 to 41 species in forest plots in Australia. Values for CWM SLA ranged from 5.8 to 16.3 $cm^2\ g^{-1}$ in England and from 6.1 to 127.1 $cm^2\ g^{-1}$ in Australia, and

CWM Leaf N ranged from 1.7 to 3.5% in England and from 0.5 to 2.9% in Australia. The percentage of N fixers in England ranged from 0 to 42.4% of total cover (presence of *Acacia* spp. micro-habitats characterized the only N fixer in the Australian dataset; Table 1).

Linking plant attributes and microbial community composition

The Euclidean matrix of distance for plant attributes were positively and significantly related to Bray-Curtis matrix of distance including the community composition of soil bacteria and fungi in the Australia and English datasets (*via* Mantel test) (Fig. 1), indicating that certain plant community attributes/traits and microbial taxa tend to co-occur in nature. Variation partitioning modeling suggested that, in general, plant attributes explained unique portions of the variation in bacterial and fungal communities from both Australia and England (Fig. 2; Figs. S5 and S6; Table S1). Shared variation explaining microbial community composition and diversity among different predictors (e.g. climate and location, soil properties and plant attributes) cannot be attributed to any of those groups of predictors in particular. Because of this, we only compared the unique portion of the variation in microbial communities explained in a singular manner by either: climate and location, soil properties or plant attributes.

Moreover, using distance-based multi-model inference and variation partitioning modeling, we found that plant attributes explained a unique proportion of the variation in soil microbial communities that was unaccounted for by soil properties, climate or location (Table 2). Removal of all plant attributes from these models always resulted in poorer model fit in all cases (ΔAIC>2.00). In Australia, our best fitting models selected canopy height, plant cover and leaf P as the major predictors of bacterial community composition and diversity, respectively (Table 2). Plant cover, height and width were selected as major plant predictors of the diversity of soil fungi in Australia (Table 2). The only exception was the community composition of soil fungi in Australia which was best predicted by pH and Aridity Index, and for which models were not improved by the inclusion of plant attributes (Table 2). In England, plant diversity, leaf N and LDMC were selected as major predictors for bacterial composition. The same predictors, but also the cover of N fixers, were also the major drivers of bacterial diversity in this dataset (Table 2). Finally, plant diversity and leaf N were selected as the major predictors of fungal composition, whereas cover, diversity, RGR and SLA were the best predictors of fungal diversity in the English dataset.

We then used Random Forest analyses to identify the importance of plant attributes, soil properties and climate in predicting microbial community composition and diversity (Fig. 3). Plant

attributes were selected as significant predictors of the diversity and community composition of bacteria and fungi in Australia and England (Fig. 3). In addition, soil properties and climate were key significant predictors of bacterial and fungal attributes; although no soil property or climate variable was selected as a significant driver of the diversity of fungi in Australia. Most predictors in the best fitting models (Table 2) were also selected as significant drivers of bacterial and fungal diversity and community composition by our Random Forest analyses (Fig. 3), thus demonstrating that the identity of the main predictors was robust to the statistical method used.

We then used SEM to gain deeper insights on the role of plant attributes and functional traits in predicting the community diversity and composition of fungi and bacteria in two Hemispheres. Each SEM included the predictors of each microbial attribute selected in the best fitting (ΔAIC > 2) models described above and in Table 2. We detected multiple significant direct effects of plant attributes on soil microbial community composition and diversity after accounting for other key drivers such as climate and soil properties (Figs. 4 and 5). In both the Australia and English datasets, plant cover had a negative direct effect on the diversity of bacteria and/or fungi (Fig. 4). In Australia, canopy height was the major plant attribute explaining the composition of bacteria (Fig. 4). In England, plant diversity had a positive effect on the diversity of bacteria and fungi (Fig. 4). Also, plant diversity and leaf N showed direct effects on the composition of bacteria and fungi (Fig. 5).

We also identified some indirect effects of location and climate on the composition or diversity of soil bacteria and fungi via plant attributes (Figs. 4 and 5) in the Australian and English datasets. For example, plant width was indirectly related to the composition of fungi *via* changes in soil pH for the Australian dataset (Fig. 4). In addition, we also found direct effects of climate (mainly from Aridity Index) on the diversity of soil bacteria and fungi in England (Fig. 5). Aridity Index also operated via its effects on the plant cover, CWM SLA, and CWM leaf N of temperate grassland plant communities in England, but it did not affect these attributes in Australia (Figs. 4 and 5).

Further correlation analyses (Spearman) exploring links among plant attributes and microbial community diversity and composition for Australia and England are available in Fig. S7. Soil pH and C were the most consistent abiotic factors explaining the community composition and/or diversity of fungi and/or bacteria for the Australian and English datasets (Figs 4 and 5). Importantly, in the case of England, the direction and strength of the multiple direct and indirect effects in our SEM were mostly maintained after controlling for management practices by using the residuals of

bacterial diversity or composition from a one-way ANOVA, as explained in the Method section (Fig. S8).

Finally, we used Random Forest analyses to identify particular bacterial and fungal species that are associated with certain plant community attributes and plant traits in both the Australian and English datasets. A subset of phylotypes (total 57 OTUs) shared by the Australian and English datasets –bioinformatic analyses were done simultaneously for both datasets allowing direct comparison of OTUs– were significantly associated with particular plant traits (Fig. S9). For example, the relative abundance of OTU_1699 (unidentified species from family Ellin5301; phylum *Gemmatimonadetes*) was strongly and positively correlated to N fixation (% coverage of N fixing plants across English grasslands and presence of Acacia sp. in Australia) in both the Australian and English datasets ($P<0.01$). Similarly, the relative abundance of OTU_98 (Unidentified species from genus *Candidatus Solibacter*; phylum *Acidobacteria*) was strongly positively related to SLA, in the Australian and English datasets ($P<0.05$). Finally, the relative abundance of OTU_8 (Uncultured Mortierellaceae; division Zygomycota) and the relative abundance of OTU_43313 (Erythrobacteraceae; phylum proteobacteria) were found to be strongly negatively related to plant cover and plant diversity, respectively, in both Australia and England (Fig. S9 for complete list of taxa).

Discussion

Our study provides strong observational evidence, from two contrasting regions of the globe that aboveground plant attributes such as diversity, cover and functional traits, can help explain the diversity and community composition of soil bacterial and fungal communities, at a regional scale (hundreds of kilometers). We also provided examples for microbial phylotypes that are strongly related to particular plant traits such as SLA, leaf N, and N fixation across two very different regions of the world. We did this using two separate datasets from Australia and England, which differed markedly in climate (dryland vs. mesic), vegetation (forest vs. grasslands), and microbial community composition (Figs. S3 and S4). Our distance-based and variation partitioning models provided evidence that plant attributes explain a unique proportion of variation in the composition and diversity of microbial communities that is unaccounted for by other key microbial drivers such as climate and soil properties, which are routinely proposed to be the main determinants of microbial community structure and diversity at large spatial scales. Our SEMs provided an integrative understanding of the role of plant attributes in driving soil microbial communities once we

controlled for multiple environmental drivers. These results provide further evidence of strong, direct links between particular aboveground plant attributes and the diversity and composition of soil fungal and bacterial communities at regional scales.

Our findings accord with the results of microcosm experiments that demonstrate the importance of plant functional traits (e.g. litter chemistry) for soil microbial community composition (Schneider *et al.* 2012; Zhou *et al.* 2015). However, they are in contrast to recent studies that did not find significant relationships between the local distribution of plant traits and soil microbial community composition within Panamanian tropical forest (Barberán *et al.* 2015) or in grassland sites in England (Fry *et al.* 2017; Leff et al. 2018). This likely relates to the different spatial scale used in these studies; our study considers variation in microbial communities at a regional scale, whereas studies of Barberán *et al.* (2015), Fry al. (2017) and Leff et al. (2018) examined local scales where variation in both plant traits and microbial communities, and their drivers, is less and thus shows weaker patterns of association.

Our SEM results indicate that plant cover had a strong negative effect on the diversity of bacteria and/or fungi in both Australia and England. More specifically, our results suggest that increases in percentage plant cover might lead to the exclusion of microbial species via the competition-to-exclusion principle (Eldridge *et al.* 2017). In addition, unlike Australia, in England, plant leaf N content (e.g. positive for bacterial diversity), SLA (e.g. negative for fungal diversity), and species diversity (e.g. positive for fungi and bacteria) were also important drivers of the distribution of the diversity and community composition of fungi and bacteria. All these plant attributes are considered key functional markers which relate to soil fertility and the quantity and quality of plant inputs (Garnier *et al.* 2004). This finding suggests that in temperate grasslands, the community composition and diversity of soil microbial communities may be strongly affected by both the range and quality of the resources entering soil from plant communities, in the form of litter (note that we used leaf nutrients in our study), but they may also be related to an effect of root turnover and exudation (de Vries *et al.* 2012; Grigulis *et al.* 2013). For example, highly diverse plant communities can influence the community composition and diversity of soil microbial communities via greater variability in litter quality (niche partitioning), but also by promoting a higher diversity of resources (e.g. via rhizodeposition; Paterson *et al.* 2007). Plant leaf N and diversity were also major drivers of microbial community composition in the studied grasslands, suggesting that these plant attributes can promote/inhibit the relative abundance of particular microbial taxa. Conversely, other

plant traits not measured in England, such as canopy height (likely to be relatively constant in temperate grasslands, and therefore uninformative in England), regulated the community composition of bacteria in Australian forests. Together, the above discussed results suggest that litter quality might be the major plant driver of microbial community composition in temperate grasslands, where plant inputs to soil are relatively large. Further, in the English dataset, there was almost complete vegetation coverage across grassland sites. Conversely in Australia, where plant cover was always less than 100% (18-98%), litter quantity rather than quality likely plays a more important role in influencing the composition of soil microbial communities. We would like to highlight that our study focused on aboveground plant attributes, which were found to account for a unique portion of the variation in the distribution of soil microbial community composition and diversity. However, we did not have available information on belowground attributes for our study sites. As such, we can only guess that including belowground plant attributes would have increased the explanatory power of our models, however, further research need to be done to support this assumption.

In addition to demonstrating that plant attributes can explain regional scale variation in bacterial and fungal community composition, our study provides a unique inventory of phylotypes (i.e., species equivalent) that are strongly associated with particular plant traits, such as SLA index, leaf N content and/or N fixation, in two markedly different regions of the globe. This information and approach could be used to: (1) predict the distribution of particular microbial taxa using plant functional traits, with potential implications for the understanding of ecosystem functioning; and (2) help to identify the potential role of certain microbial species, with as yet unestablished functional roles, in driving particular ecosystem functions (e.g. decomposition rates). Some of these phylotypes responded in a similar manner to increases in the values for particular plant traits. For example, the relative abundance of OTU_1699 (family Ellin5301) was strongly positively related to N fixation (% coverage of N fixers across English grasslands and presence of *Acacia* sp. in Australia) in both Australian and English datasets. Regrettably, little is currently known about the ecology of these bacterial taxa. Furthermore, the relative abundance of OTU_98 (*Candidatus Solibacter* sp.) was strongly positively correlated to SLA in both datasets ($\rho>0.164$, $P<0.05$). Species from the genus *Candidatus Solibacter* are known to be chemoorganotrophic organisms that use organic C for growth and energy (Ward *et al.* 2009); as such, they might gain resources from litter inputs, especially those of high decomposability (i.e., often characterized by a higher SLA). In the same

vein, OTU_8 (*Mortierellaceae* sp), a saprophyte that can act as a facultative parasite (Fitzpatrick and Morton 1930), was negatively related to plant cover in England and Australia (see extended discussion on phylotypes showing opposite patterns in both datasets in Appendix S2).

Plant attributes such as diversity, vegetation cover, and plant traits are highly sensitive to climate change and land use intensification (Allan *et al.* 2015; García-Palacios *et al.* 2016; Deraison *et al.* 2015; Le Bagousse-Pinguet *et al.* 2017). Supporting this notion, our SEM identified multiple indirect effects of climate on fungal diversity and community composition, driven indirectly by changes in plant attributes. For example, increases in aridity related to changes in plant cover, SLA, and leaf N of temperate grasslands in England, which could be taken to suggest that predicted increases in aridity resulting from climate change (Huang *et al.* 2016) might indirectly alter the diversity and composition of grassland soil fungal communities. In this respect, our SEM results could be used to generate new hypotheses that could potentially lead to management strategies. For instance, our approach could help identify how plant traits mediate climate effects, and lately provide strategies for the management of these traits, that mitigate climate impacts on soil microbial communities and soil processes. This is especially significant given the known importance of changes in fungal communities for biogeochemical cycles and plant community dynamics in grasslands (van der Heijden *et al.* 2008), and hence the potential for this to alter the capacity of these ecosystems to provide essential goods and services, such as food production and climate regulation (Bardgett & van der Putten 2014; Delgado-Baquerizo *et al.* 2016a).

Together, our work provides new evidence, from an observational study, for the important role of plant attributes in explaining variation in soil microbial communities across two markedly different mesic and dryland ecosystem types of the world. More precisely, in both forested ecosystems and temperate grasslands, plant attributes explained a unique proportion of the variation in soil microbial communities that could not be explained by factors such as soil abiotic properties and climate. Our findings also advance understanding of the links between plant traits and soil microbial communities by identifying a suite of phylotypes strongly associated with particular plant traits such as SLA, leaf N and N fixation across a broad range of ecosystem types. Such information suggests that it might be possible to predict the distribution of certain microbial taxa at large spatial scales using plant functional traits. Given the importance soil microbial communities for ecosystem functioning, such knowledge is critical to improve our ability to predict likely changes in ecosystem function under global change and to manage terrestrial ecosystems sustainably.

Statement of authorship

M.D-B., R.D.B. and B.K.S. designed this study. The Australia dataset was compiled by D.J.E. and M.D-B. The England dataset was compiled by R.D.B., F.T.V. and P.M. Lab analyses were conducted by E.L.F. and M.D-B. J.K., G.B., D.J.E. and M.D-B. provided plant trait data. B.K.S. provided Miseq Illumina data. K.H. performed bioinformatic analyses. M.D-B. conducted statistical modeling. The manuscript was written by M.D-B with contributions from all co-authors.

Data accessibility

Data associated with this paper has been deposited in figshare: https://figshare.com/s/f61108b8c4fa89074296 (10.6084/m9.figshare.5861061).

Acknowledgments

M.D-B. acknowledge support from the Marie Sklodowska-Curie Actions of the Horizon 2020 Framework Programme H2020-MSCA-IF-2016 under REA grant agreement n° 702057. E.L.F. was supported by the NERC Biodiversity and Ecosystem Services and Sustainability programme (Wessex BESS, ref. NE/J014680/1). R.D.B. acknowledges the UK Department for Environment, Food and Rural Affairs (DEFRA) (Grant BD1451) and Biotechnology and Biological Sciences Research Council (BBSRC) (Grant BB/I009000/2) for supporting the generation of the English dataset, and D. Millward, R.S. Smith, and S. Barlow who performed the English vegetation surveys. B.K.S., M.D-B are supported by the Australian Research Council project (DP170104634). D.J.E. was supported by the Hermon Slade Foundation. We also thank MPI-BGC Jena, who host TRY, and the international funding networks supporting TRY (IGBP, DIVERSITAS, GLP, NERC, QUEST, FRB and GIS Climate).

References

Allan E, Manning P, Alt F, Binkenstein J, Blaser S, Blüthgen N, Böhm S, Grassein F, Hölzel N, Klaus VH, et al. 2015. Land use intensification alters ecosystem multifunctionality via loss of biodiversity and changes to functional composition. *Ecology Letters* **18**: 834–843.

Archer E. 2016. *rfPermute: Estimate Permutation p-Values for Random Forest Importance Metrics*. R package version 1.5.2.

Barberán A, McGuire KL, Wolf JA, Jones FA, Wright SJ, Turner BL, Essene A, Hubbell SP, Faircloth BC, Fierer N. 2015. Relating belowground microbial composition to the taxonomic, phylogenetic, and functional trait distributions of trees in a tropical forest. *Ecology Letters* **18**: 1397-1405.

Bardgett RD, van der Putten WH. 2014. Belowground biodiversity and ecosystem functioning. *Nature* **515**: 505-11.

Bardgett RD, Wardle DA. 2010. *Aboveground – belowground linkages: biotic interactions, ecosystem processes and global change*. Oxford, UK: Oxford University Press

Bardgett RD. 2017. Plant trait-based approaches for interrogating belowground function. *Biology and Environment: Proceedings of the Royal Irish Academy* **117**: 1-13.

Burnham KP, Anderson DR. 2002. *Model Selection and Multimodel Inference. A Practical Information-Theoretical Approach*. Springer, Heidelberg.

Cornelissen JHC, Lavorel S, Garnier E, Díaz S, Buchmann N, Gurvich DE, Reich PB, ter Steege H, Morgen HD, van der Heijden MGA et al. 2003. A handbook of protocols for standardised and easy measurement of plant functional traits. *Australian Journal of Botany* **51**: 335-380.

De Deyn GB, Cornelissen JHC, Bardgett RD. 2008. Plant functional traits and soil carbon sequestration in contrasting biomes. *Ecology Letters* **11**: 516–531.

de Vries FT, Manning P, Tallowin JRB, Mortimer SR, Pilgrim ES, Harrison KA, Hobbs PH, Quirk H, Shipley B, Cornelissen JHC et al. 2012. Abiotic drivers and plant traits explain landscape-scale patterns in soil microbial communities. *Ecology Letters*, **15**: 1230-1239.

Delgado-Baquerizo M, Maestre FT, Reich PB, Trivedi P, Osanai Y, Liu Y-R, Hamonts K, Jeffries TC, Singh BK. 2016a. Carbon content and climate variability drive global soil bacterial diversity patterns. *Ecological Monographs* **86**: 373–390.

Delgado-Baquerizo M, Maestre FT, Eldridge DJ, Singh BK. 2016b. Microsite differentiation drives the abundance of soil ammonia oxidizing bacteria along aridity gradients. *Frontiers in Microbiology* **7**: 505.

Delgado-Baquerizo M, Eldridge DJ, Ochoa V, Gozalo B, Singh, BK, Maestre FT. 2017. Soil microbial communities drive the resistance of ecosystem multifunctionality to global change in drylands across the globe. *Ecology Letters* **20**: 1295–1305.

Deraison H, Badenhausser I, Loeuille N, Scherber C, Gross N. 2015 Functional trait diversity across trophic levels determines herbivore impact on plant community biomass. *Ecology Letters* **18**: 1346-1355.

Eldridge DJ, Delgado-Baquerizo M, Travers SK, Val J, Oliver I, Hamonts K, Singh BK. 2017. Competition drives the response of soil microbial diversity to increased grazing by vertebrate herbivores. *Ecology* DOI: 10.1002/ecy.1879.

546 Fry EL, Pilgrim ES, Tallowin JRB, Smith RS, Mortimer SR, Beaumont DA, Simkin J, Harris SJ,
547 Shiel RS, Quirk H. et al. 2017. Plant, soil and microbial controls on grassland diversity
548 restoration: a long-term, multi-site mesocosm experiment. *Journal of Applied Ecology* **54:**
549 1320-1330.
550 García-Palacios P, Maestre FT, Kattge J, Wall DH. 2013. Climate and litter quality differently
551 modulate the effects of soil fauna on litter decomposition across biomes. *Ecology Letters* **16**:
552 1045-1053.
553 García-Palacios P, Shaw EA, Wall DH, Hättenschwiler S. 2016. Temporal dynamics of biotic and
554 abiotic drivers of litter decomposition. *Ecology Letters* **19**: 554-563.
555 Garnier E, Cortez J, Billès J, Navas M-L, Roumet C, Debussche M, Laurent G, Blanchard A, Aubry
556 D, Bellmann A. et al. (2004) Plant functional markers capture ecosystem properties during
557 secondary succession. *Ecology* **85:** 2630-2637.
558 Gould IJ, Quinton JN, Weigelt A, De Deyn GB, Bardgett RD. 2016. Plant diversity and root traits
559 benefit physical properties key to soil function in grasslands. *Ecology Letters* **19**: 1140-1149.
560 Grigulis K, Lavorel S, Krainer U, Legay N, Baxendale C, Dumont M, Kastl E, Arnoldi C, Bardgett
561 RD, Poly F et al. 2013. Relative contributions of plant traits and soil microbial properties to
562 mountain grassland ecosystem services. *Journal of Ecology* **101**: 47-57.
563 Hooper DU, Bignell DE, Brown VK, Brussard L, Dangerfield JM, Wall DH, Wardle DA, Coleman
564 DC, Giller KE, Lavelle P. et al. 2000. Interactions between aboveground and belowground
565 biodiversity in terrestrial ecosystems: patterns, mechanisms, and feedbacks. *BioScience* **50**:
566 1049-1061.
567 Huang J, Yu H, Guan X, Wang G, Guo R. 2016. Accelerated dryland expansion under climate
568 change. *Nature Climate Change* **6**: 166–171.
569 Laliberté E. 2017. Belowground frontiers in trait-based plant ecology. *New Phytologist* **213**: 1597-
570 1603.
571 Lauber CL, Hamady M, Knight R, Fierer N. 2009. Soil pH as a predictor of soil bacterial community
572 structure at the continental scale: a pyrosequencing-based assessment. *Applied*
573 *Environmental Microbiology* **75**: 5111-5120.
574 Lavorel S. 2013. Plant functional effects on ecosystem services. *Journal of Ecology* **101**: 4-8.

Le Bagousse-Pinguet Y, Gross N, Maestre FT, Maire V, de Bello F, Fonseca CR, Kattge J, Valencia E, Leps J, Liancourt P. 2017. Testing the environmental filtering concept in global drylands. *Journal of Ecology* doi: 10.1111/1365-2745.12735.

Leff JW, Lynch RC, Kane NC, Fierer N. 2017. Plant domestication and the assembly of bacterial and fungal communities associated with strains of the common sunflower, *Helianthus annuus*. New Phytologist 214, 412–423.

Leff, JW, Bardgett, R, Wilkinson, A, Jackson, BG, Pritchard, W, De Long, J, Oakley, S, Mason, KE, Ostle, NJ, Johnson, D, et al. 2018. Predicting the structure of soil communities from plant community taxonomy, phylogeny, and traits.*The ISME Journal* (in press).

Legay N, Baxendale C, Grigulis K, Krainer U, Kastl E, Schloter M, Bardgett RD, Arnoldi C, Bahn M, Dumont M. et al. 2014 Contribution of above- and below-ground plant traits to the structure and function of grassland soil microbial communities. *Annals of Botany* **114**: 1011–1021.

Liu Y-R, Delgado-Baquerizo M, Trivedi P, He Y-Z, Wang J-T, Singh BK. 2017. Identity of biocrust species and microbial communities drive the response of soil multifunctionality to simulated global change. *Soil Biology and Biochemistry* **107**: 208-217.

López-García Á, Azcón-Aguilar C, Barea JM. 2014. The interactions between plant life form and fungal traits of arbuscular mycorrhizal fungi determine the symbiotic community. *Oecologia* **176**: 1075-86.

López-García Á, Varela-Cervero S, Vasar M, Öpik M, Barea JM, Azcón-Aguilar C. 2017. Plant traits determine the phylogenetic structure of arbuscular mycorrhizal fungal communities. *Molecular Ecology* **26**: 6948-6959.

Maestre FT, Delgado-Baquerizo M, Jeffries TC, Eldridge DJ, Ochoa V, Gozalo B, Quero JL, García-Gómez M, Gallardo A, Ulrich W. et al. (2015) Increasing aridity reduces soil microbial diversity and abundance in global drylands. *Proceedings of the National Academy of Sciences of the United States of America* **112**: 15684–15689.

Manning P, de Vries FT, Tallowin JRB, Smith R, Mortimer SR, Pilgrim ES, Harrison KA, Wright, DG, Quirk H, Benson J. et al. 2015. Simple measures of climate, soil properties and plant traits predict national-scale grassland soil carbon stocks. *Journal of Applied Ecology* **52**: 1188–1196.

McArdle BH, Anderson MJ. 2001. Fitting multivariate models to community data, a comment on distance-based redundancy analysis. *Ecology* **82**: 290–297.

Moreau D, Pivato B, Bru D, Busset H, Deau F, Faivre C, Matejicek A, Strbik F, Philippot L, Mougel C. 2015. Plant traits related to nitrogen uptake influence plant-microbe competition. *Ecology* **96**, 2300-2310.

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P. et al. 2015. *vegan: Community Ecology Package.* R package version 2.3-0.

Orwin KH. 2010 Linkages of plant traits to soil properties and the functioning of temperate grassland. *Journal of Ecology* **98:** 1074-1083.

Paterson E, Gebbing T, Abel C, Sim A, Telfer G. 2007. Rhizodeposition shapes rhizosphere microbial community structure in organic soil. *New Phytologist* **173**: 600-10.

Powell JR. 2015. Deterministic processes vary during community assembly for ecologically dissimilar taxa. *Nature Communications* **6**: 8444.

Prober SM, Leff JW, Bates ST, Borer ET, Firn J, Harpole WS, Lind EM, Seabloom EW, Adler PB, Bakker JD. et al. 2015. Plant diversity predicts beta but not alpha diversity of soil microbes across grasslands worldwide. *Ecology Letters* **18**: 85–95.

Ramirez KS, Leff JW, Barberán A, Bates ST, Betley J, Crowther TW, Kelly EF, Oldfield EE, Shaw A, Steenbock C. et al. 2014. Biogeographic patterns in belowground diversity in New York City's Central Park are similar to those observed globally. *Proceedings of the Royal Society of London [Biology]* **281**: 20141988.

Rodwell JS. (Ed.). 1992. *British Plant Communities, Volume 3. Grasslands and montane communities*. Cambridge University Press, Cambridge.

Schneider T, Keiblinger KM, Schmid E, Sterflinger-Gleixner K, Ellersdorfer G, Roschitzki B, et al. 2012. Who is who in litter decomposition? Metaproteomics reveals major microbial players and their biogeochemical functions. *The ISME Journal* **6**: 1749–1762.

Tedersoo L, Bahram M, Põlme S, Kõljalg U, Yorou NS, Wilesundra R, Villarreal Ruiz L, Vasco-Palacios AM, Thu PQ, Suija A. et al. 2014. Fungal biogeography. Global diversity and geography of soil fungi. *Science* **346**: 1256688.

Thion CE, Poirel JD, Cornulier T, de Vries FT, Bardgett RD, Prosser JI. 2016. Plant nitrogen-use strategy as a driver of rhizosphere archaeal and bacterial ammonia oxidiser abundance. *FEMS Microbiology Ecology* DOI: 10.1093/femsec/fiw091.

van der Heijden MGA, Bardgett RD, van Straalen NM. 2008. The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecology Letters* **11**: 296-310.

Wardle D. 2005. *How plant communities influence decomposer communities*. Biological Diversity and Function in Soils, Cambridge University Press, Cambridge, UK.

Wardle DA, Bardgett RD, Klironomos JN, Setälä H, van der Putten WH, Wall DH. 2004. Ecological linkages between aboveground and belowground biota. *Science* **304:** 1629.

Wehner J, Antunes PM, Powell JR, Caruso T, Rillig MC. 2011. Indigenous arbuscular mycorrhizal fungal assemblages protect grassland host plants from pathogens. *PLoS ONE* **6**: e27381.

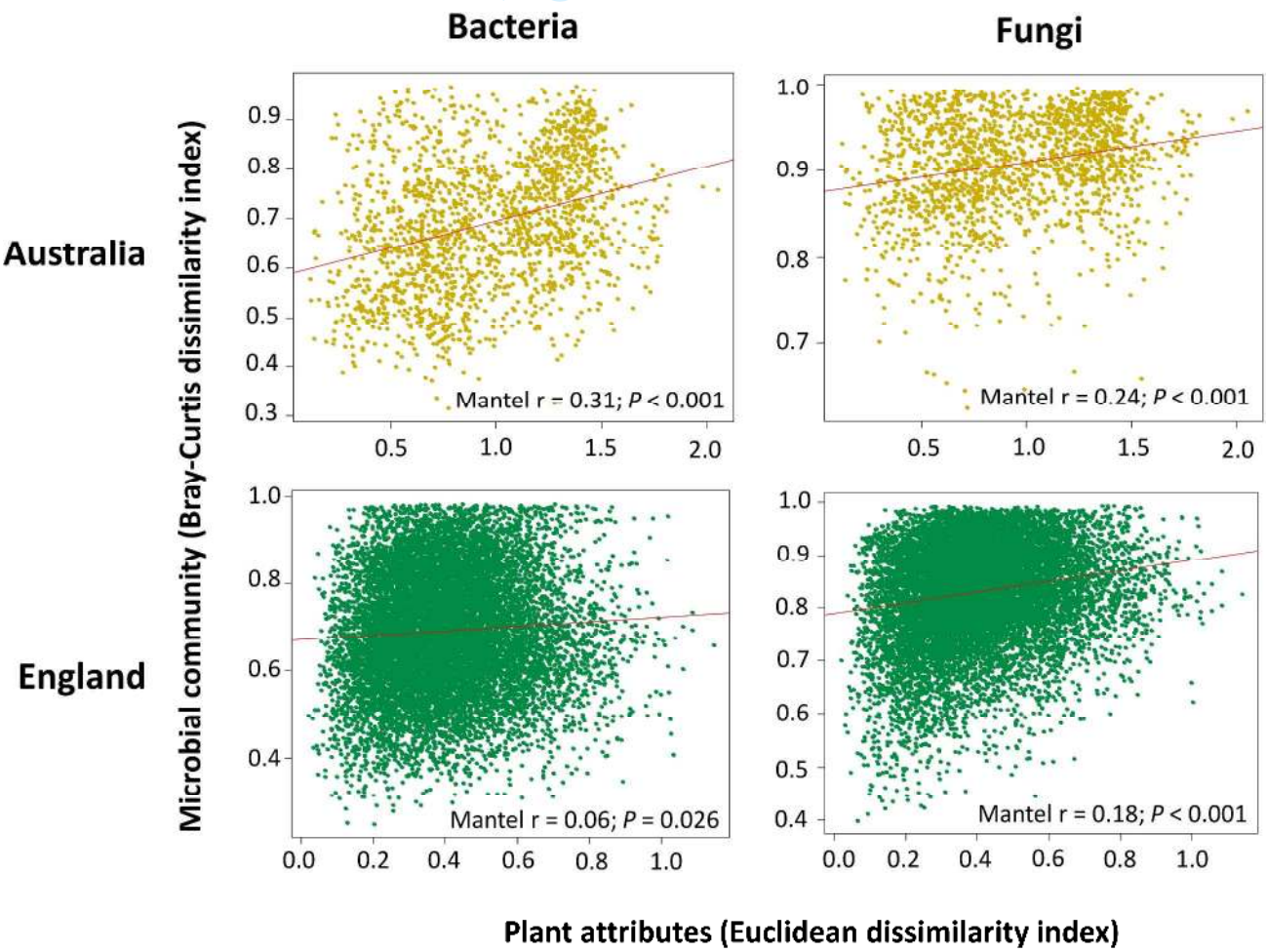
Zhou J, Deng Y, Shen L, Wen C, Yan Q, Ning D, Qin Y, Xue K, He Z, Voordeckers JW. et al. 2016. Temperature mediates continental-scale diversity of microbes in forest soils. *Nature Communications* **7**: 12083.

Zhou Y, Clark M, Su J, Xiao C. 2015. Litter decomposition and soil microbial community composition in three Korean pine (*Pinus koraiensis*) forests along an altitudinal gradient. *Plant and Soil* **386**, 171–183.

Zuur AF, Ieno EN, Walker NJ, Savaliev A, Smith GM. 2009. *Mixed effects models and extensions in Ecology with R.* Springer.

665
666
667
668
669
670
671
672
673
674
675
676

677 **Figure Legends**



678

Figure 1. Relationship between the matrix of dissimilarity from multiple plant traits, cover and diversity (Euclidean distance) and the beta diversity of bacteria and fungi (community composition dissimilarity based on Bray-Curtis distance) for the Australia (n = 60) and England (n = ~160) datasets. The solid lines represent the fitted linear regressions.

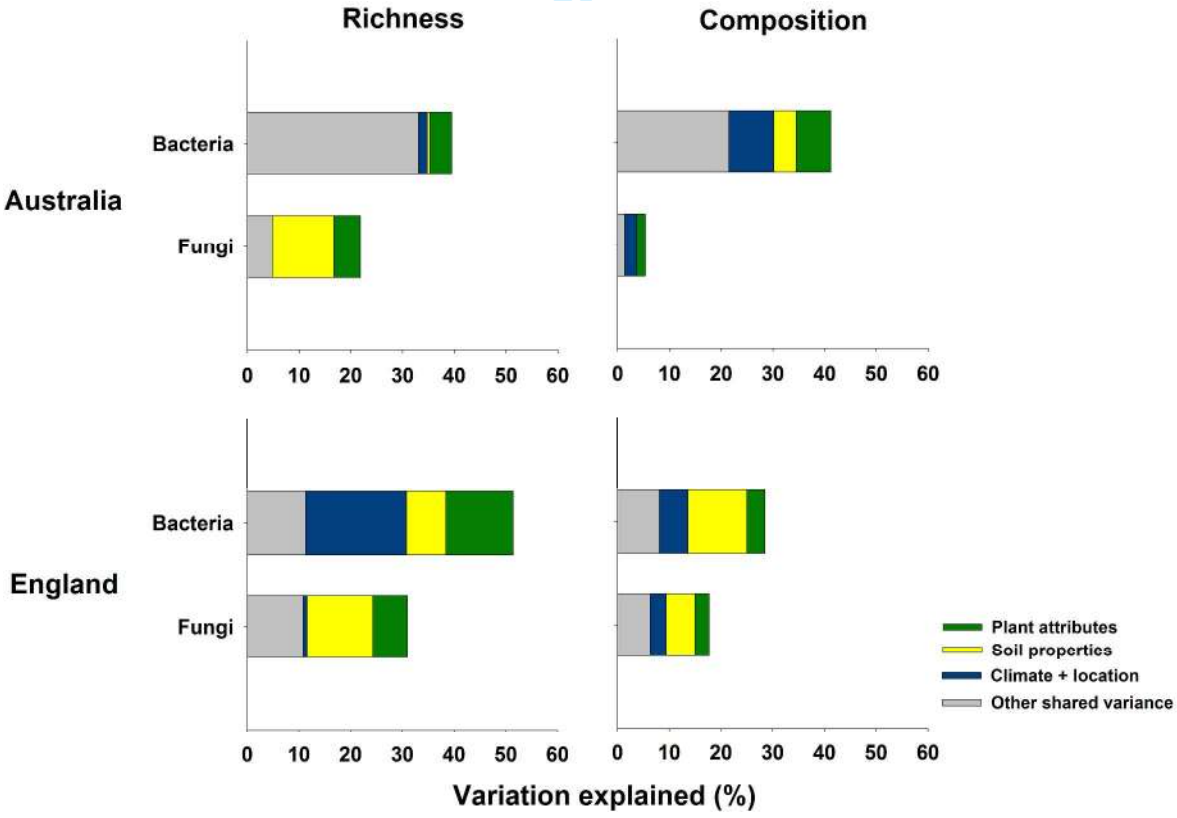


Figure 2. Relative contribution of the different predictors used to model bacterial and fungal composition and diversity. Panels represent results from Variation Partitioning modelling aiming to identify the percentage variance of bacterial and fungal community composition and diversity explained by plant attributes (cover, diversity and functional traits), soil properties and climate in

Australia and England. Unique and shared variance from plant cover, diversity and functional traits in predicting microbial community composition and diversity were merged in this figure for simplicity. An alternative version of this figure showing the unique and shared variance of each group of predictors can be found in Supplementary Figs. 5 and 6. P-values associated with the relative contribution of the different predictors are available in Table S1.

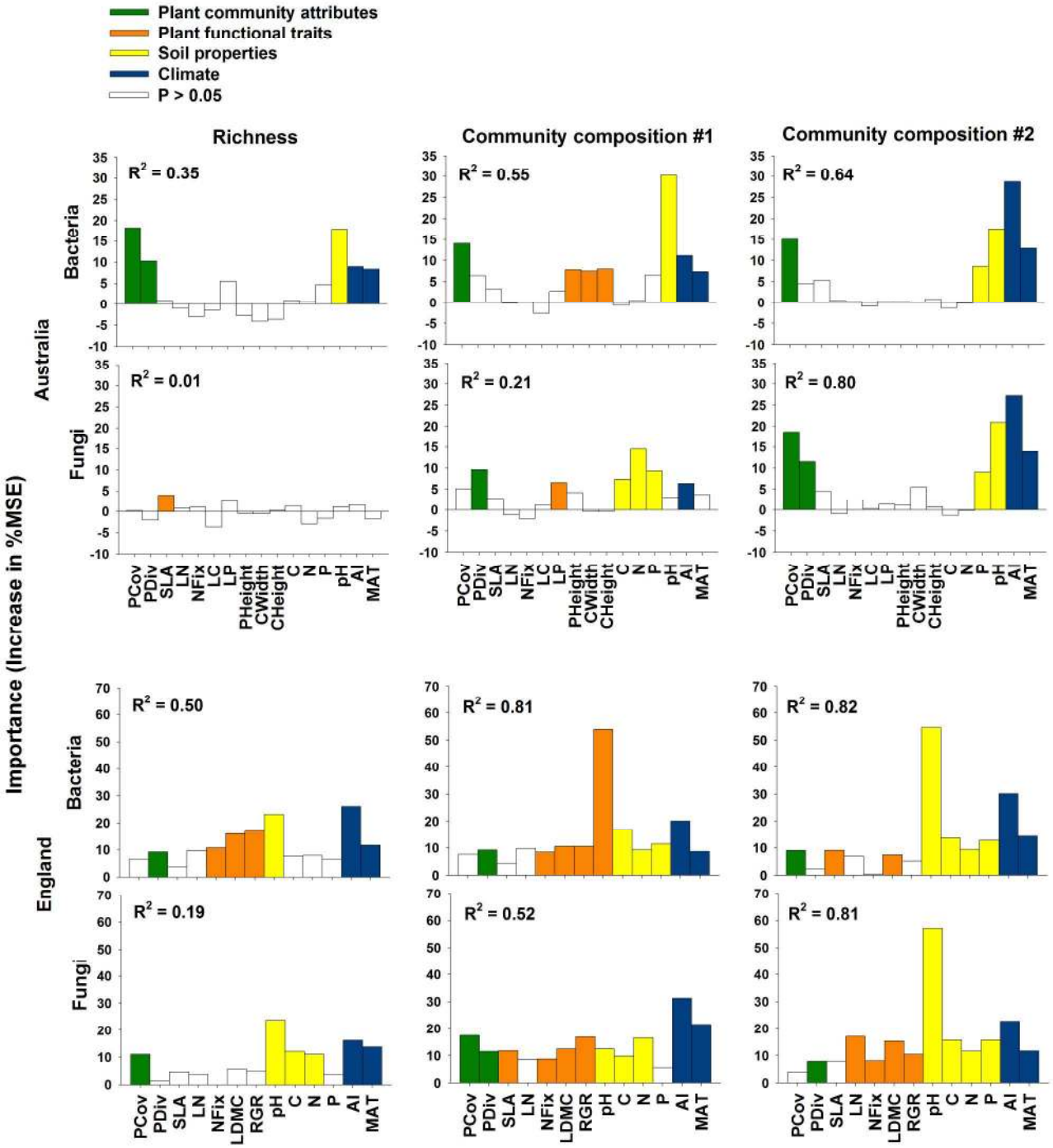


Figure 3. Random Forest analysis aiming to identify the best individual predictors of the diversity and community composition of bacteria and fungi in Australia and England. Predictors include plant attributes, soil properties and climate (Table 1). MSE = Mean Square Error. Community composition #1 and #2 represent the first and second axis of a NMDS including the community composition of bacteria or fungi (See Fig. S3).

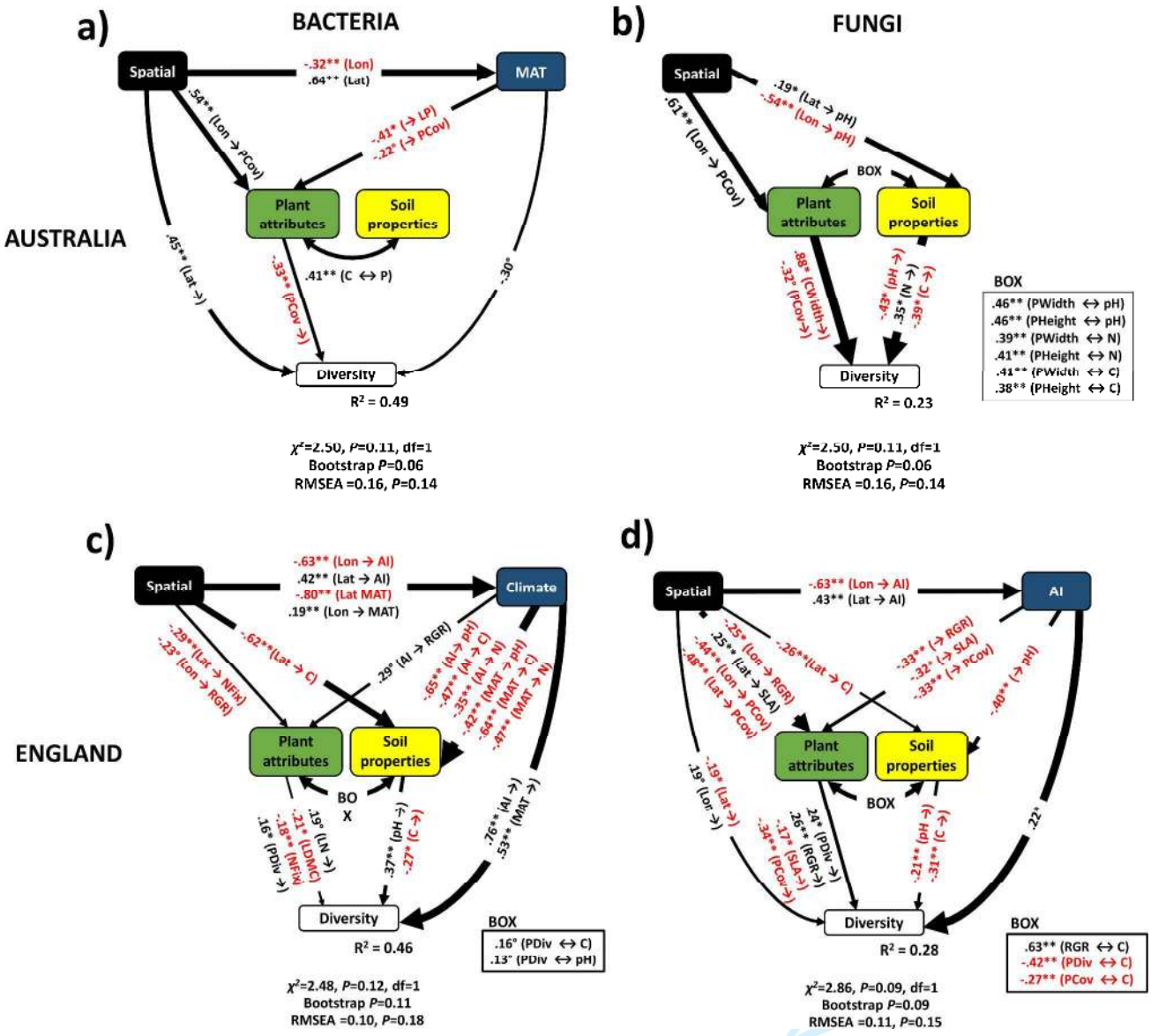


Figure 4. Structural equation model describing the effects of multiple drivers (selected from Table 1) on the diversity of bacteria (a and c) and fungi (b and d) for the Australia (n = 60) and England (n = ~160) datasets. Numbers adjacent to arrows are indicative of the effect size of the relationship. R^2 denotes the proportion of variance explained. Climate, soil properties and plant predictors are included in our models as independent observable variables, however we group them in the same box in the model for graphical simplicity. All predictors within each both are allowed to co-vary. This does not apply to model in which only one predictor for a given group is included. In this case, the name of the predictor stand alone (e.g. soil pH). Significance levels of each predictor are °P<0.10, *P<0.05, **P<0.01. Negative effects in red.

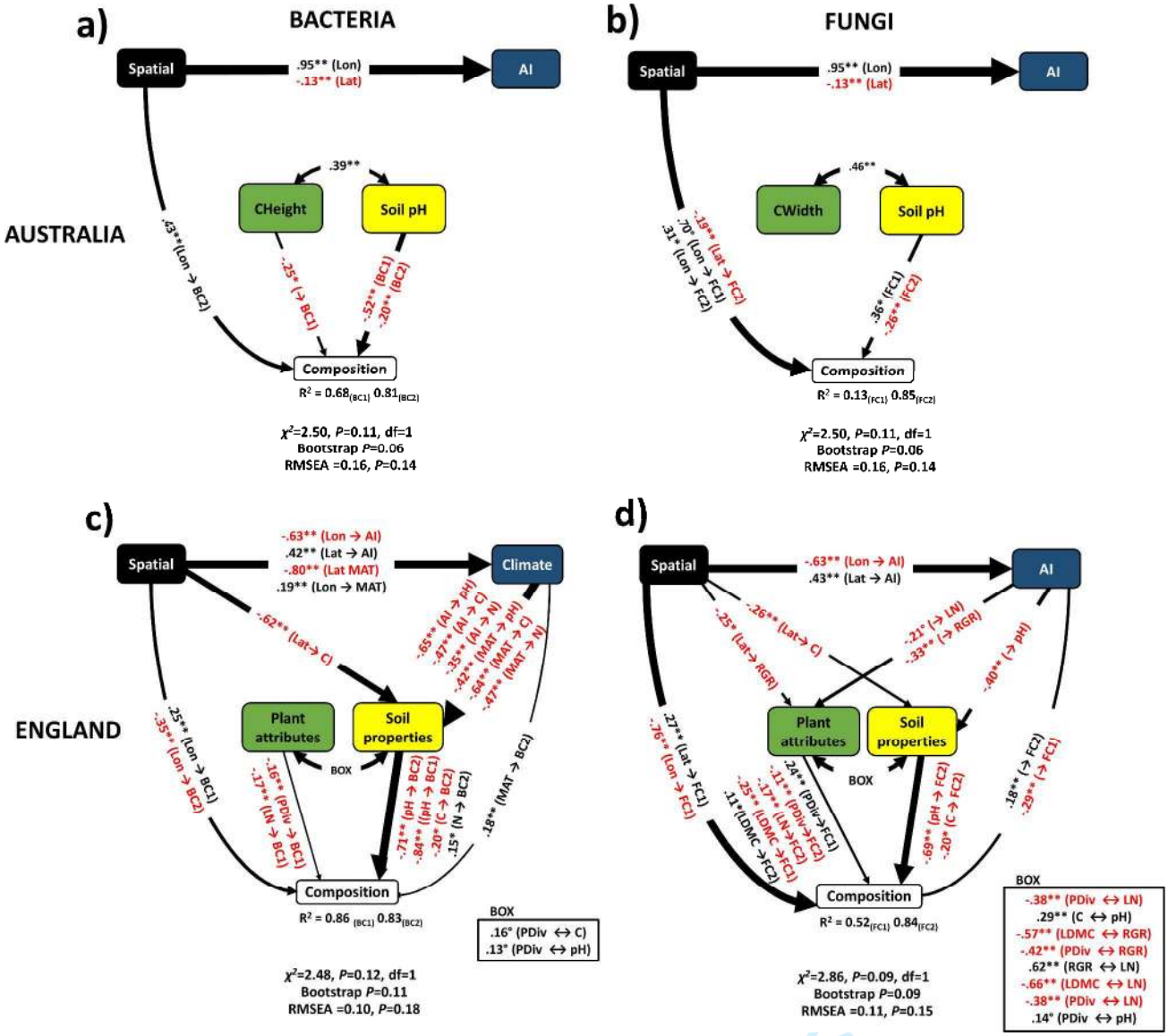


Figure 5. Structural equation model describing the effects of multiple drivers (selected from Table 1) on the composition (two axes from a NMDS) of (a and c) and fungi (b and d) for the Australia (n = 60) and England (n = ~160) datasets. Numbers adjacent to arrows are indicative of the effect size of the relationship. R^2 denotes the proportion of variance explained. Climate, soil properties and plant predictors are included in our models as independent observable variables, however we grouped them in the same box in the model for graphical simplicity. Significance levels of each predictor are °P<0.10, *P<0.05, **P<0.01. Negative effects in red.

Table 1. Complete list of predictors used in this study.

Group of predictors	Variables	Acronym	Value range (Australia)	Value range (England)	Units
Location	**Latitude**	Lat	-34.7 to -33.3	50.7 to 54.6	Decimal degrees
	Longitude	Lon	145.7 to 151.1	-4.4 to 0.9	Decimal degrees
Climate	**Aridity Index**	AI	0.3 to 0.9	0.9 to 2.4	Unitless
	Mean annual temperature	MAT	12.8 to 17.5	6.3 to 10.2	ºC
Soil properties	**Soil C**	C	1.3 to 12.3	1.4 to 12.8	%
	Soil N	N	0.1 to 0.6	0.2 to 1.1	%
	Soil P	P	$3.1 \cdot 10^{-3}$ to $6.0 \cdot 10^{-2}$	$1.6 \cdot 10^{-2}$ to 0.2	%
	pH	pH	4.8 to 8.9	4.1 to 7.8	Unitless
Plant community-level traits	**Plant richness**	PDiv	11 to 41	2 to 36	Number of species
	Plant cover	PCov	18.3 to 98.3	78.3 to 249.5	%
Plant functional traits	**Specific Leaf Area**	SLA	6.1 to 127.1	5.8 to 16.3	$cm^2\ g^{-1}$
	Leaf N	LN	0.5 to 2.9	1.7 to 3.5	%
	N fixation	NFix	0 to 1	0 to 0.42	Australia: Presence/Absence N fixers England: Proportion of N fixers (0-1)
	Leaf C	LC	0.5 to 2.9	-	%
	Leaf P	LP	$2.1 \cdot 10^{-2}$ to 0.2	-	%
	Plant height	PHeight	0.2 to 22.0	-	m
	Canopy width	CWidth	0.1 to 21.0	-	m
	Canopy height	CHeight	$6.0 \cdot 10^{-2}$ to 7.0	-	m
	Leaf dry matter content	LDMC	-	14.9 to 34.8	$g^{-1}g^{-1}$
	Relative growth rate	RGR	-	0.1 to 0.3	$g^{-1}g^{-1}d^{-1}$

Table 2. Best-fitting model predicting the distribution of microbial PDiv and composition (bacteria and fungi). Model A include all parameters in Table 1. Model B included all parameters in model A, but plant functional traits. Model C included all parameters in model A, but plant community attributes (cover and PDiv). Model D included all parameters in model A but plant functional traits and community attributes. Location (latitude and longitude) inclusion was forced in all models to account for spatial autocorrelation. Models are ranked by AIC. AIC measures the relative goodness of fit of a given model; the lower its value, the more likely the model to be correct. ΔAIC are difference between the AIC of each model and that of the best model. See Table 1, for the acronyms of the variables included in this table.

Database	Microbial	Models	Climate	Soil	Plant predictors	R^2	AIC	ΔAIC
Australia	**Bacterial composition**	**A**	AI	pH	CHeight	0.380	451.20	0
		B	AI	pH		0.334	453.44	2.24
		C	AI	pH	CHeight	0.380	451.20	0.00
		D	AI	pH		0.334	453.44	2.24
	Bacterial richness	**A**	MAT	C + P	PCov + LP	0.462	283.44	0
		B	MAT	C + P	PCov	0.441	283.73	0.29
		C	AI			0.357	286.09	2.65
		D	AI			0.357	286.09	2.65
	Fungal composition	**A**	AI		CWidth	0.172	497.37	
		B	AI	pH		0.170	497.55	0.18
		C	AI		CWidth	0.172	497.37	0.00
		D	AI	pH		0.170	497.55	0.18
	Fungal richness	**A**		C + N + pH	PCov + PHeight + CWidth	0.222	218.51	
		B		C + N + pH	PCov	0.159	219.13	0.62
		C		pH		0.049	220.82	2.31
		D		pH		0.049	220.82	2.31
England	**Bacterial composition**	**A**	AI + MAT	C + N + pH	PDiv + LN + LDMC	0.412	1150.00	
		B	AI + MAT	C + N + P + pH	PDiv	0.394	1152.60	2.60
		C	AI + MAT	C + N + pH	RGR + LN + LDMC	0.406	1151.70	1.70
		D	AI + MAT	C + N + P + pH		0.377	1155.10	5.10
	Bacterial richness	**A**	AI + MAT	C + N + pH	PDiv + RGR + LN + LDMC + NFix	0.485	647.88	

	B	AI + MAT	C + N + P + pH		0.360	673.52	25.64
	C	AI + MAT	C + N + pH	LDMC + NFix	0.459	649.27	1.39
	D	AI + MAT	C + N + P + pH		0.360	673.52	25.64
Fungal composition	**A**	AI	C + pH	PDiv + LN	0.233	1269.70	
	B	AI	C + pH	PDiv	0.237	1270.90	1.20
	C	AI	C + pH	RGR + LN + LDMC	0.237	1271.10	1.40
	D	AI	C + pH		0.215	1273.50	3.80
Fungal richness	**A**	AI	C + pH	PCov + PDiv + RGR + SLA	0.282	802.24	
	B	AI	C + pH	PCov + PDiv	0.250	805.05	2.81
	C	AI	C + pH	RGR + SLA	0.240	807.25	5.01
	D	AI	C + pH		0.210	809.37	7.13

Supplementary Information

Plant attributes explain the distribution of soil microbial communities in two contrasting regions of the globe.

Manuel Delgado-Baquerizo, Ellen L. Fry, David J. Eldridge, Franciska T. de Vries, Pete Manning, Kelly Hamonts, Jens Kattge, Gerhard Boenisch, Brajesh K. Singh, Richard D. Bardgett.

Correspondence to:

Manuel Delgado-Baquerizo. E-mail: M.DelgadoBaquerizo@gmail.com

Richard D. Bardgett. E-mail: richard.bardgett@manchester.ac.uk

This PDF file includes:

Appendices S1-S2

Table S1

Figures S1-S9

Appendix S1. Supplementary Methods.

Soil sampling

In Australia, soil samples were collected in March 2014. A 30 m x 30 m site representative of the dominant vegetation was established at 20 locations across New South Wales. At each site, three soil cores (0-5 cm depth) were collected under the three most common plant functional groupsmicro-habitat: grasses (*Rhytidosperma* spp.), N-fixing shrubs (*Acacia* spp.) and trees (*Eucalyptus* spp.). The same genus of these plant taxa was present across all plots. A total of 60 soil samples (20 sites x three functional group micro-habitats) were collected in this study. Soil cores were then mixed to obtain a composite soil sample per micro-habitat for each of the sites. Following field sampling, the soil was separated into two fractions. A fraction of the soil was immediately frozen at -20 ºC for molecular analyses.

In England, sampling was conducted in June/July 2005. The sites consisted of triplets of extensively managed, intermediate intensively managed, and intensively managed grasslands at 60 locations giving 180 sites. At each of these a 25 m x 25 m plot of homogenous vegetation was established and five soil cores (0-7 cm depth) were taken at random and pooled to produce a composite sample for microbial and chemical analysis. Following field sampling, the soil was sieved (2 mm mesh) and separated into two fractions, one of which was immediately frozen at -80 ºC for molecular analyses.

Climate

In all cases, we obtained information on mean annual temperature and precipitation (1 km) for the surveyed sites from the Worldclim database (*www.worldclim.org*). In addition, for each site we obtained the Aridity Index (Precipitation/evapotranspiration) from the Global Potential Evapotranspiration database (Zomer *et al.* 2008), which is based on interpolations provided by WorldClim. We used Aridity Index (which is positively related to precipitation) rather than mean annual precipitation because aridity includes both mean annual precipitation and potential evapotranspiration, and is therefore a more accurate metric of the water availability at each site.

Soil properties

For the Australian samples, concentration of soil total organic carbon (C) was determined as described in Delgado-Baquerizo *et al.* (2016c). Soil total N was measured with a CN analyzer (Leco CHN628 Series, LECO Corporation, St Joseph, MI, USA) and total phosphorus (P) was measured using a SKALAR San++ Analyzer (Skalar, Breda, The Netherlands) after digestion with sulphuric

acid. For the English samples, total soil C and N were measured on an Elementar Vario EL elemental analyzer (Hanau, Germany), and soil P was measured by combustion and digestion in sulfuric acid, followed by quantification of orthophosphate by automated colorimetry. In all cases, soil pH was measured in a soil and water suspension with a pH meter.

Plant attributes

The Australia dataset includes *de novo* information on eight plant traits for multiple genus corresponding with the three sampled micro-habitatmicro-habitat in each plot: grasses (*Rhytidosperma* spp.), N-fixing shrubs (*Acacia* spp.) and trees (*Eucalyptus* spp.). These plant traits include leaf C, N and P, SLA index, plant height, canopy width, canopy height (distance from canopy to ground), and ability to fix N (hereafter N fixation). The concentrations of leaf C, N and P for *Rhytidosperma* spp., *Acacia* spp., and *Eucalyptus* spp. were determined using the same methods explained above for soil. In all cases, a composite sample from ten individuals was collected per plot. Average plant height, width and canopy height and ability to fix N were determined in the field for each plot. We used a clinometer to measure the height of all large trees (> 2m) and a graduated pole to measure trees and shrubs less than 2 m tall. SLA was measured in the lab using a standardized protocol (Cornelissen *et al.* 2003) and N fixation was measured as presence of *Acacia* species (the only N-fixer micro-habitat in this dataset). Total plant cover and diversity (number of species) were recorded at each site as explained in Maestre *et al.* (2015). Note that in this dataset, sampling effort was focused on dominant plants –which are expected to affect microbial communities via their plant attributes and functional traits–, however, other less dominant species were also present in these plots allowing us to obtain a metric of plant diversity per plot.

In England, we used the plant functional trait dataset of de Vries et al. (2012), which included community weighted mean (CWM) values for five plant traits which were assigned to all plant species occurring in the 180 plots: leaf dry matter content (LDMC), relative growth rate (RGR), leaf N content (LNC), SLA, and proportion of N-fixer plants (hereafter N fixation). In each site, the plant cover of all vascular plant species, total plant cover and species richness were recorded in five $1m^2$ quadrates and averaged (de Vries *et al.* 2012). Cover data were also combined with trait data obtained from the TRY database (Kattge *et al.* 2008) to determine community abundance (plant cover) weighted means of each trait (CWMs), following de Vries *et al.* (2012); information on the cover of N-fixing plant species was also gathered for each plot.

Sequence data processing

After visual assessment of the quality of all Illumina R1 and R2 reads using FastQC (Andrews 2010), low quality regions (Q<20) were trimmed from the 5' end of the sequences (20 bp from R1 and 82 bp from R2 for primer set 341F/805R; 5 bp from R1 and 35 bp from R2 for primer set FITS7-ITS4R) using SEQTK (https://github.com/lh3/seqtk). The paired ends were subsequently joined using FLASH (Magoc & Salzberg 2011). Remaining primer sequences were removed from the resulting reads using SEQTK and a further round of quality control was conducted in mother (Ward *et al.* 2009) to discard short sequences (<380 bp for 16S and <150 bp for ITS), as well as sequences with ambiguous characters or more than 8 homopolymers. Operational Taxonomic Units (OTUs) were built at 97% sequence similarity using UPARSE (Edgar *et al.* 2013). Singletons were discarded, as well as chimeric sequences identified by the UCHIME algorithm using the recommended SILVA gold 16S rRNA gene or UNITE reference databases for bacteria and fungi, respectively. OTU abundance (Edgar *et al.* 2011) tables were constructed by running the usearch_global command (http://www.drive5.com/). Taxonomy was assigned to OTUs in mothur using the naïve Bayesian classifier with a minimum bootstrap support of 60% and the Greengenes database version 13_8 (DeSantis *et al.* 2006; McDonald *et al.* 2012) for bacteria or the dynamic UNITE version 6 dataset (Koljalg *et al.* 2013) for fungi. The resulting OTU abundance tables were rarefied to an even number of sequences per sample, corresponding to the minimum number of sequences for a single soil sample (13225 sequences/sample for bacteria and 13433 sequences/sample for fungi), using mother (Schloss *et al.* 2009). We further removed phylotypes that only had one read per OTU across all samples.

General patterns in microbial diversity and composition

We first examined the community composition and diversity (number of species) of bacteria and fungi in the Australian and English datasets. To obtain a metric of microbial community composition at the OTU level, we used a non-metric multidimensional ordination (NMDS). Fungal and bacterial community compositions were analysed separately, but in both cases, simultaneously included data from the Australia and English datasets. We retained the first two axes from a 2D solution (stress ~ 0.1 in all cases). We conducted NMDS ordinations with the package Vegan from R (Oksanen *et al.* 2015) using the Bray-Curtis dissimilarity index. We evaluated overall differences in microbial diversity and composition between the Australia and English dataset by conducting one-way PERMANOVA with dataset (Australia/England) as a fixed factor. The PERMANOVA aiming to assess overall differences in microbial community composition between datasets was carried out

including information at the OTU level and not the axes of the NMDS. These analyses were done using the PERMANOVA+ for PRIMER statistical package (PRIMER-E Ltd., Plymouth Marine Laboratory, UK).

Structural Equation Modeling.

Unlike regression or ANOVA, SEM offers the ability to separate multiple pathways of influence and view them as parts of a system, and thus is useful for investigating the complex relationships among predictors commonly found in natural ecosystems (Grace 2006). The probability that a path coefficient differs from zero was tested using bootstrap resampling. Bootstrapping is preferred to the classical maximum-likelihood estimation in these cases because in bootstrapping probability assessments are not based on the assumption that the data match a particular theoretical distribution.

The goodness of fit of SEM models was checked using the following: the Chi-square test, the root mean square error of approximation (RMSEA) and the Bollen-Stine bootstrap test (Schermelleh-Engel *et al.* 2003). Our *a priori* models attained an acceptable/good fit by all criteria in all cases, and thus no post hoc alterations were made. SEM models were conducted with the software AMOS 20 (IBM SPSS Inc, Chicago, IL, USA).

Appendix S2. *Extended discussion on particular microbial phylotypes associated with particular plant functional traits.*

Other phylotypes showed opposite patterns in both datasets, but were still characteristic of particular plant traits (see the complete list of examples in Table S1). Examples include OTU_10654 (family Rhodospirillaceae) or OTU_4 (*Arthrobacter oxydans*) as bacteria associated with SLA or OTU_3517 (*Catenulostroma hermanusense*; plant pathogen) and OTU_3470 (Acremonium R8_9; saprophyte) as fungi linked to leaf N. These inconsistencies may be related to the strong differences between the Australian and English datasets in terms of vegetation, or to the capturing of two sides of a unimodal, or other non-linear relationship, as CWM values of some plant traits (e.g. SLA index and leaf N) were very different between the two regions (García-Palacios *et al.* 2013).

Table S1. P-values associated to the relative contribution of the different predictors used to model the richness and community composition of bacteria and fungi in Australia and England.

Dataset	Microbial attributes	Plant traits	Plant diversity + cover	Soil properties	Location + climate
Australia	Bacterial richness	0.627	0.002	0.003	< 0.001
	Bacterial composition	< 0.001	< 0.001	< 0.001	< 0.001
	Fungal richness	0.991	0.230	0.308	0.789
	Fungal composition	0.002	< 0.001	< 0.001	< 0.001
England	Bacterial richness	0.001	0.770	0.005	0.053
	Bacterial composition	< 0.001	< 0.001	< 0.001	< 0.001
	Fungal richness	0.831	0.003	< 0.001	< 0.001
	Fungal composition	< 0.001	< 0.001	< 0.001	< 0.001

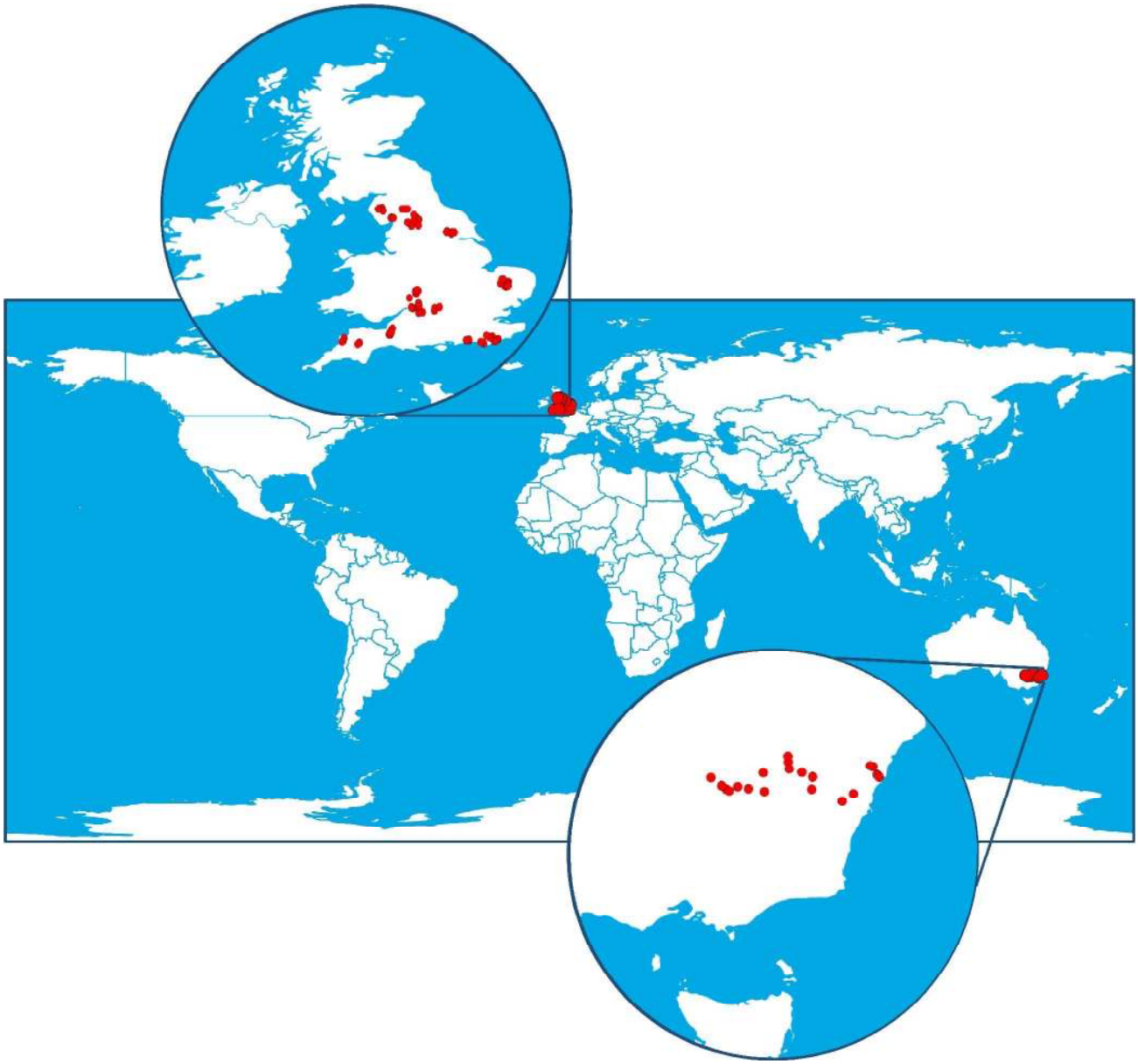


Figure S1. Locations of the sites included in this study for the Australia (n = 60) and England (n ~160) datasets.

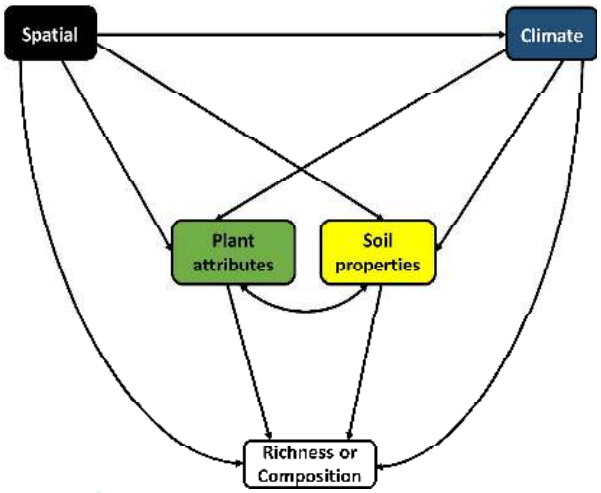


Figure S2. A priori structural equation model including direct and indirect effects of geographical location, climate, soil properties and plant attributes on the community composition or richness of soil bacteria and fungi.

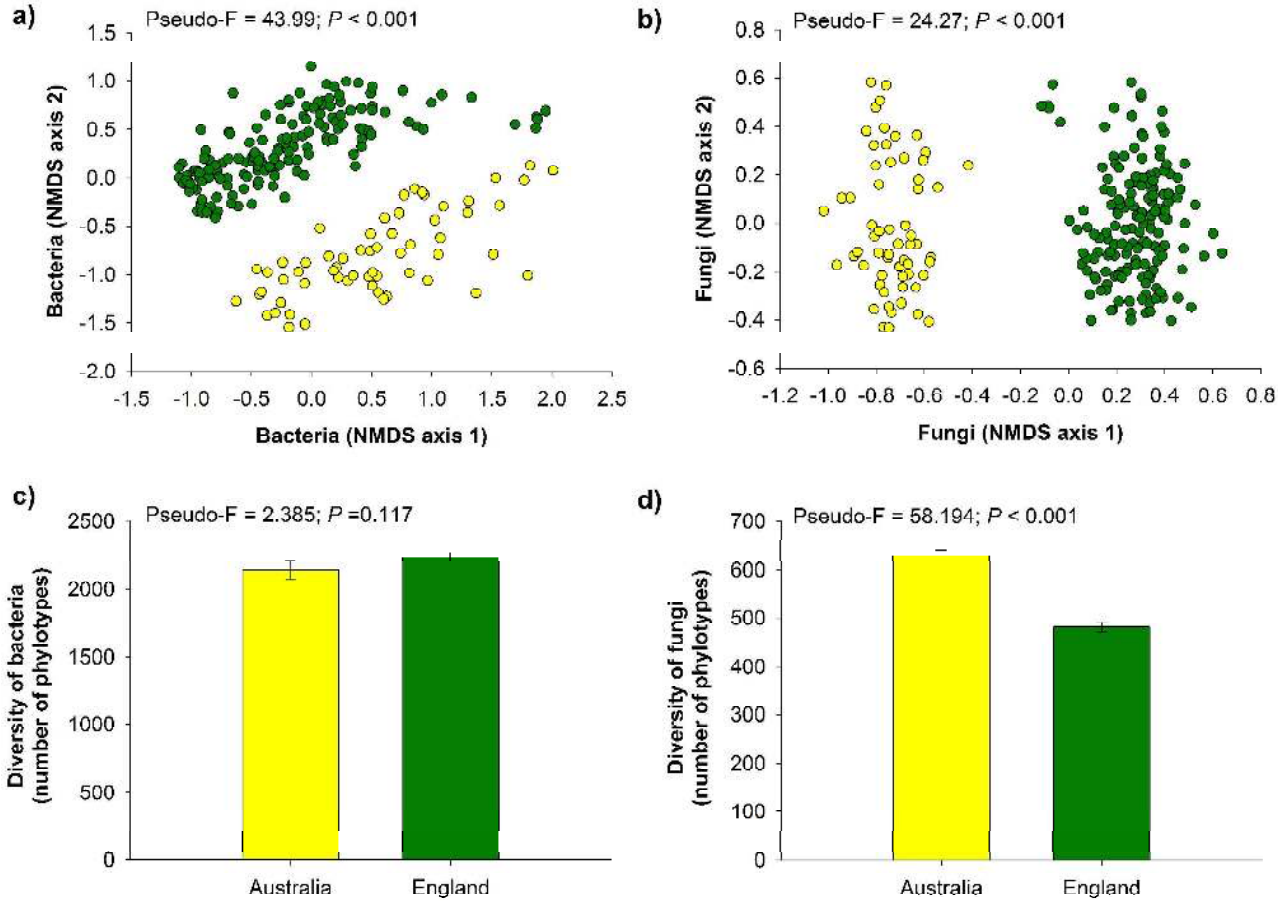


Figure S3. Community composition (a-b) and richness (c-d) of bacteria and fungi for the Australia (n = 60) and England (n ~160) datasets.

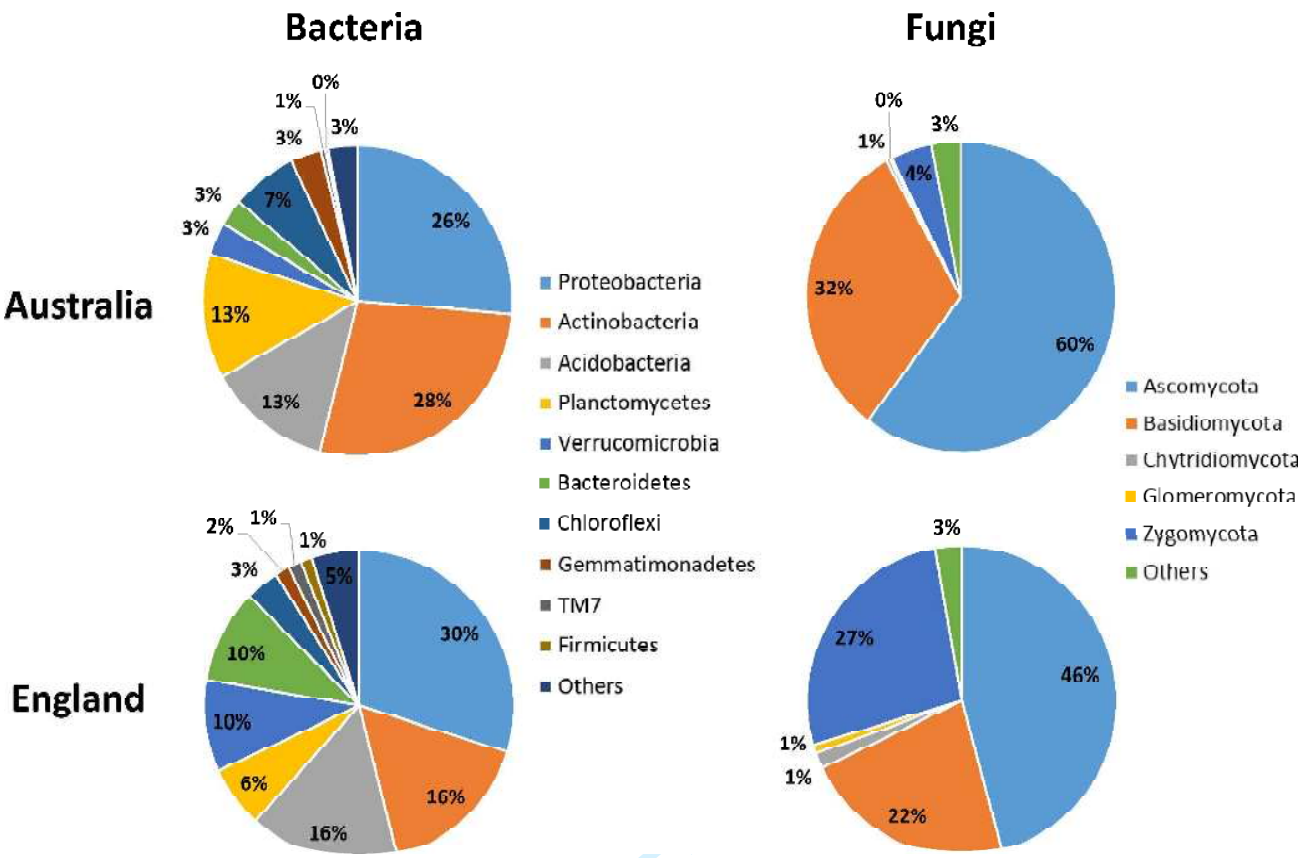


Figure S4. Composition of bacteria and fungi at the phyla level for the Australia (n = 60) and England (n ~160) datasets.

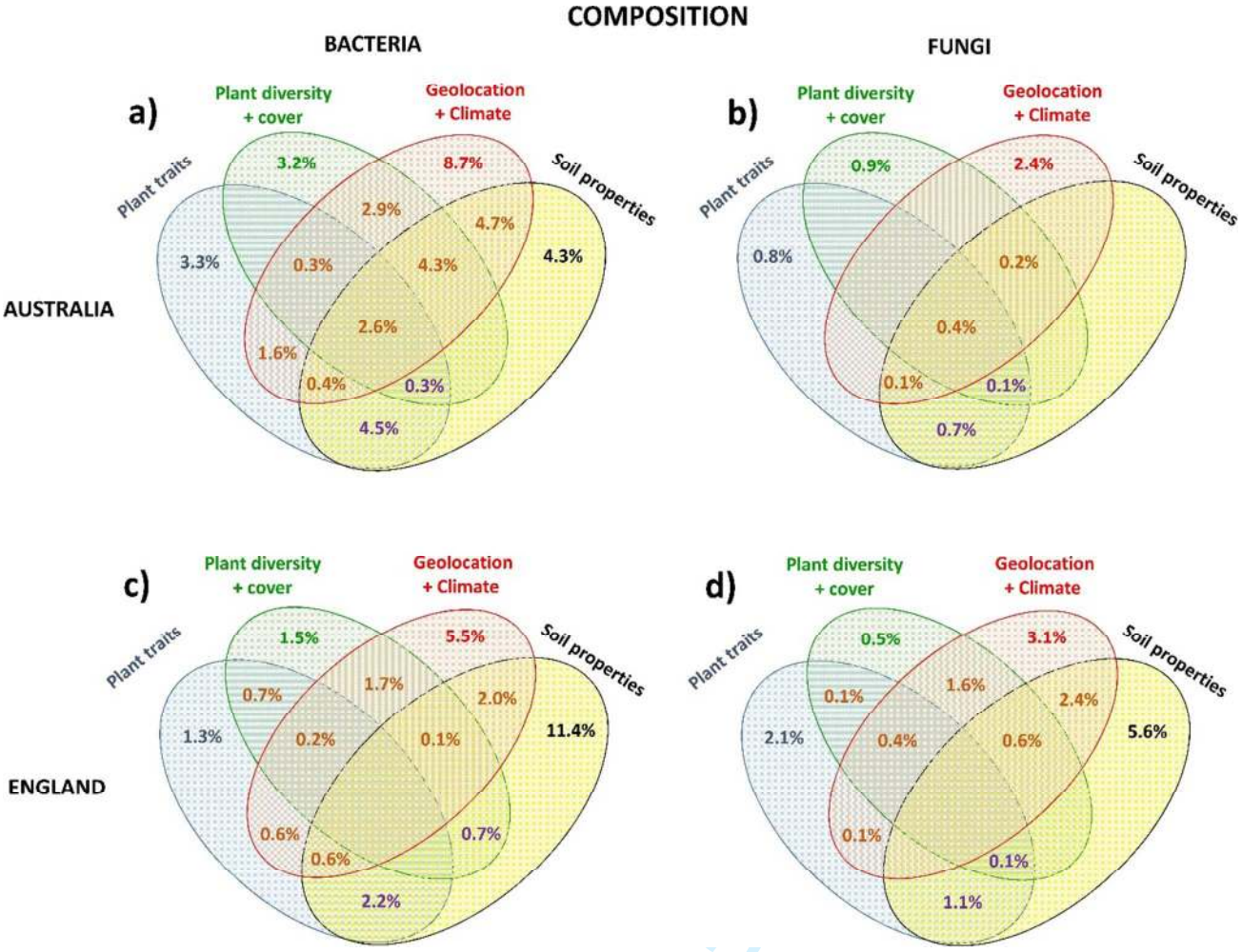


Figure S5. Variation partitioning modeling aiming to identify the relative contribution of (1) plant traits, (2) plant diversity and cover, (3) location and climate and (4) soil properties as predictors of the composition of bacteria and fungi at the OTU level. Shared effects of these variable groups are indicated by the overlap of circles. Only >0% portions of explained variation are plotted.

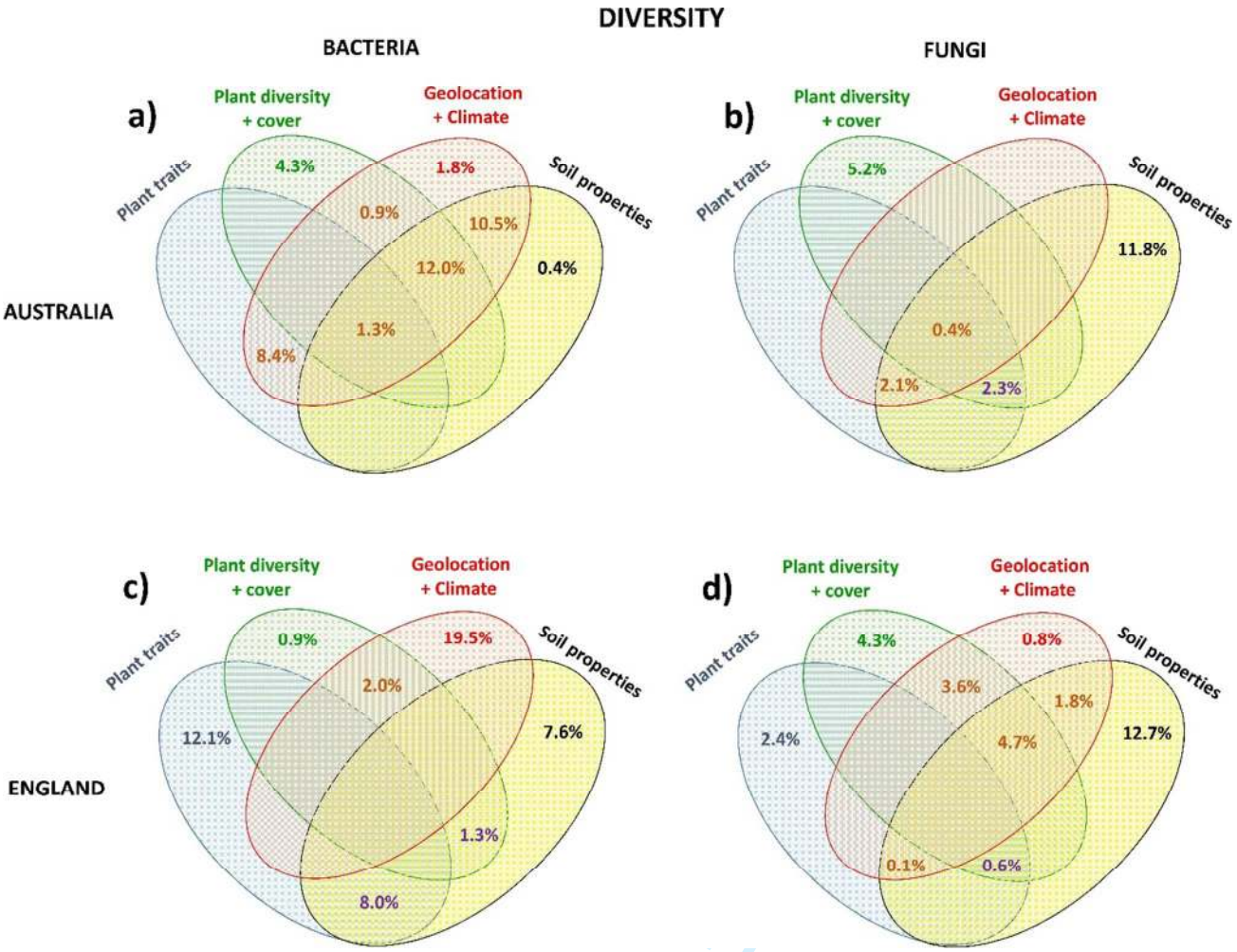


Figure S6. Variation partitioning modeling aiming to identify the relative contribution of (1) plant traits, (2) plant diversity and cover, (3) location and climate and (4) soil properties as predictors of the diversity of bacteria and fungi at the OTU level. Shared effects of these variable groups are indicated by the overlap of circles. Only >0% portions of explained variation are plotted.

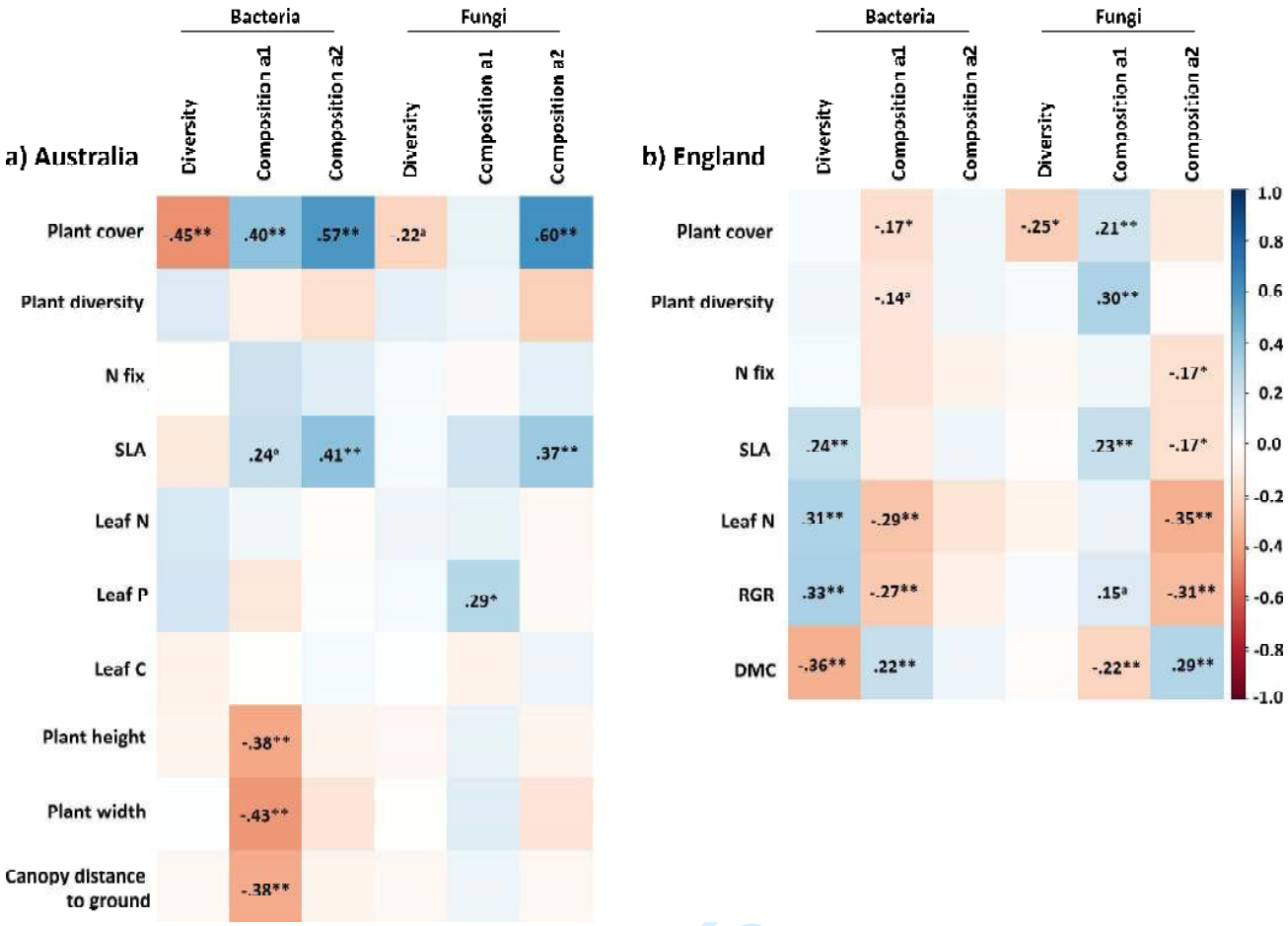


Figure S7. Correlations (Pearson) between plants traits, cover and diversity with the diversity and composition (two axes from a NMDS) of bacteria and fungi for the Australia (n = 60) and England (n = ~160) datasets. Significance levels of each predictor are *P < 0.05, **P < 0.01.

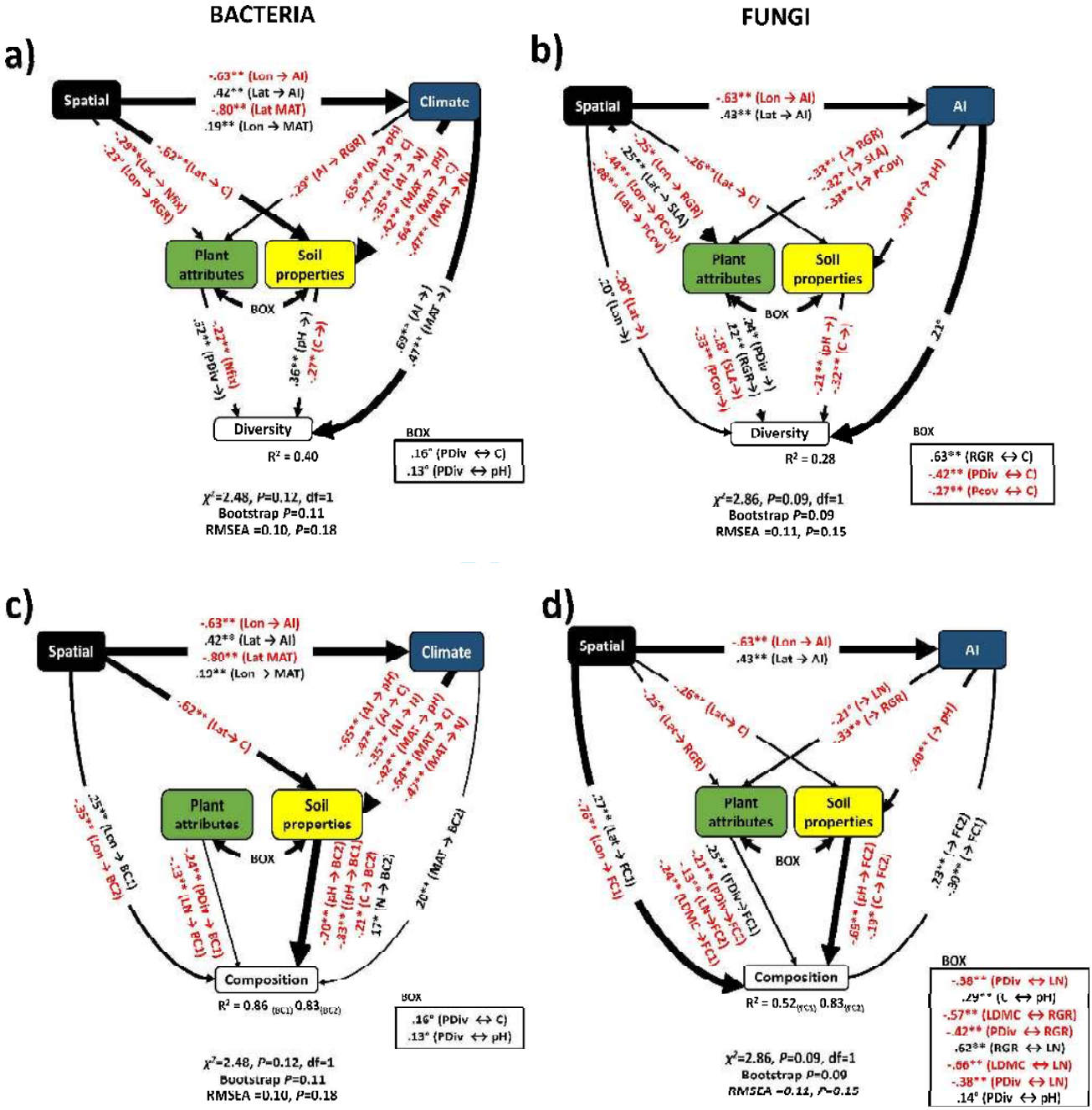


Figure S8. Structural equation model describing the effects of multiple drivers (selected from Table 1) on the residuals of richness and composition of bacteria and fungi for the England (n = ~160) dataset. Numbers adjacent to arrows are indicative of the effect size of the relationship. R^2 denotes the proportion of variance explained. Climate, soil properties and plant predictors are included in our models as independent observable variables, however we grouped them in the same box in the model for graphical simplicity. Significance levels of each predictor are °P < 0.10, *P < 0.05, **P < 0.01.

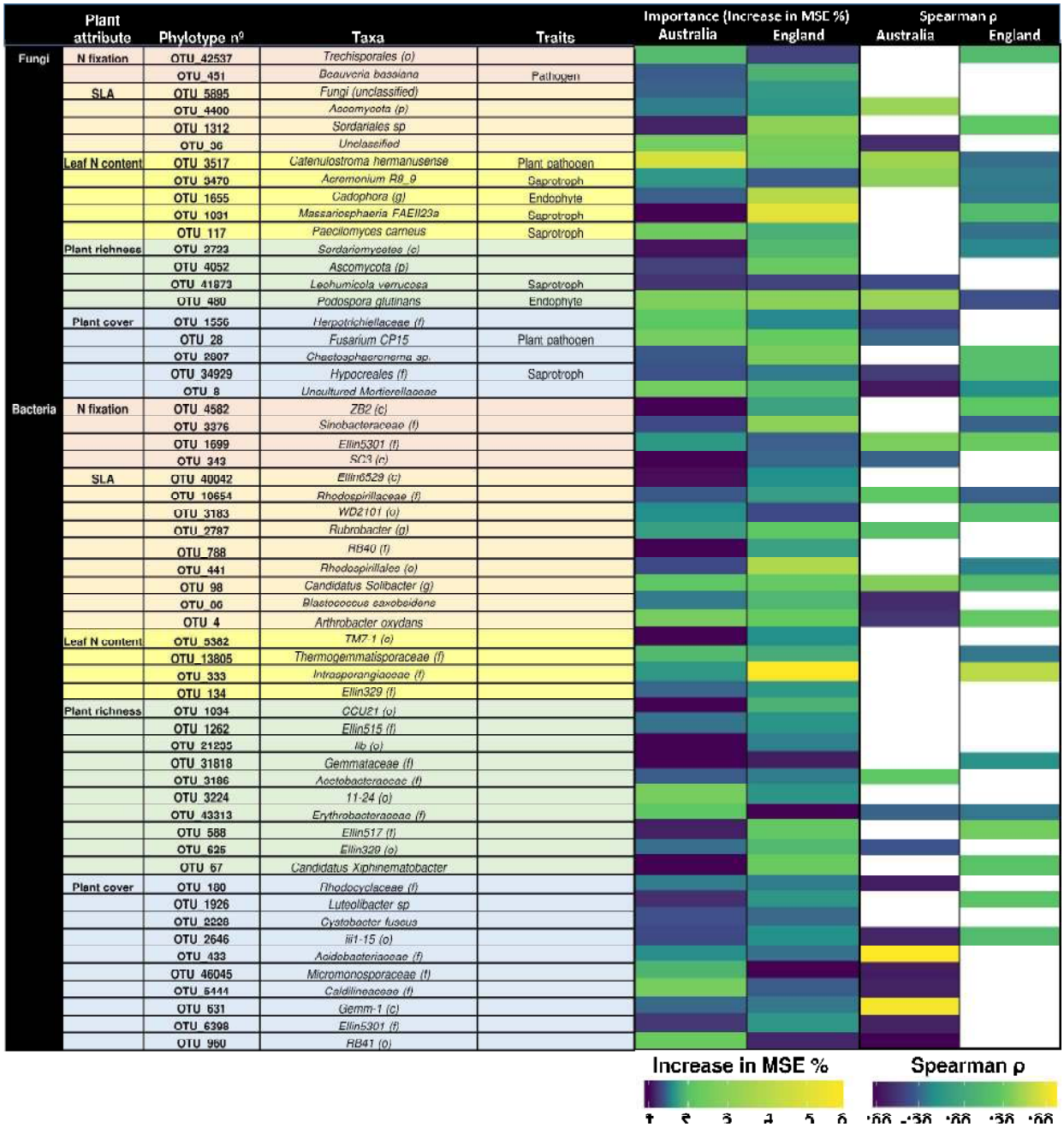


Figure S9. Shared phylotypes in the Australia and England dataset that were found to be universal predictors (via Random Forest analyses) of multiple plant attributes including plant community attributes (cover and richness) and traits (SLA index, N fixation and leaf N). Functional traits from fungal communities were identified using the FUNGUILD database (http://www.stbates.org/guilds/app.php). This Figure shows importance (MSE = Mean square error) for each microbial phylotype selected from Random Forest analyses as predictors of particular plant traits and correlation (Spearman P<0.05) between microbial phylotypes and plant traits.

References (not listed in the main text)

Andrews S. FastQCA (2010) Quality Control tool for High Throughput Sequence Data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

DeSantis T.Z. et al. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Applied and Environmental Microbiology, 72:5069-72.

Edgar R.C., Haas B.J., Clemente J.C., Quince C., Knight R. (2011) UCHIME improves sensitivity and speed of chimera detection. Bioinformatics, 27: 2194-2200.

Edgar R.G. (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nature Methods 10, 996-998.

Grace J.B. (2006) Structural Equation Modeling Natural Systems (Cambridge Univ. Press, Cambridge).

Kattge J. et al. (2008) TRY – a global database of plant traits. Global Change Biology. 17, 2905-2935.

Koljalg, U. et al. (2013) Towards a unified paradigm for sequence-based identification of fungi. Mol. Ecol. 22, 5271–5277.

McDonald, D. et al. (2012) An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J. 6, 610-618.

Magoc, T., Salzberg, S.L. (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics, 27: 2957-2963.

Schermelleh-Engel, K. et al. (2003) Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. Methods Psychol. Res. 8, 23–74.

Schloss, P.D. et al. (2009) Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. Appl. Environ. Microb. 75, 7537-7541.

Zomer, R.J., et al. (2008) Climate Change Mitigation, A Spatial Analysis of Global Land Suitability for Clean Development Mechanism Afforestation and Reforestation. Agric. Ecosyst. Envir. 126, 67-80.