Consistent predictors of microbial community composition across spatial scales in grasslands reveal low context-dependency

Dajana Radujković[1*], Sara Vicca[1], Margaretha van Rooyen[2], Peter Wilfahrt[3,4], Leslie Brown[5], Anke Jentsch[3], Kurt O. Reinhart[6], Charlotte Brown[7,8], Johan De Gruyter[1], Gerald Jurasinski[9,24], Diana Askarizadeh[10], Sandor Bartha[11], Ryan Beck[12], Theodore Blenkinsopp[7], James Cahill[7], Giandiego Campetella[13], Roberto Canullo[13], Stefano Chelli[13], Lucas Enrico[14], Lauchlan Fraser[15], Xiying Hao[12], Hugh A. L. Henry[16], Maria Hohn[17], Mohammad Hassan Jouri[18], Marian Koch[19], Rachael Lawrence Lodge[20], Frank Yonghong Li[21], Janice M. Lord[20], Patrick Milligan[22], Hugjiltu Minggagud[21], Todd Palmer[22], Birgit Schröder[9], Gábor Szabó[23], Tongrui Zhang[21], Zita Zimmermann[11], Erik Verbruggen[1]

[1] University of Antwerp, Department of Biology, Plants and Ecosystems (PLECO), Universiteitsplein 1, 2610 Wilrijk, Belgium
[2] University of Pretoria, Department of Plant and Soil Science, Pretoria, South Africa
[3] University of Bayreuth, Department of Disturbance Ecology, 95447 Bayreuth, Germany
[4] University Minnesota, Department of Ecology, Evolution, and Behavior, Saint Paul, MN USA
[5] Applied Behavioural Ecology & Ecosystem Research Unit, Dept. Environmental Sciences, University of South Africa, P/Bag X6, Florida, 1710, South Africa
[6] United States Department of Agriculture- Agricultural Research Service (or USDA-ARS), Fort Keogh Livestock& Range Research Laboratory, Miles City, Montana, USA
[7] University of Alberta, Department of Biological Sciences, Edmonton, Alberta, Canada
[8] Desert Laboratory on Tumamoc Hill, University of Arizona, Tucson Arizona
[9] University of Rostock, Landscape Ecology, Rostock, Germany
[10] University of Tehran, Department of Rehabilitation of Arid and Mountainous Regions, Faculty of Natural Resources, Iran.
[11] Centre for Ecological Research, Institute of Ecology and Botany, H-2163 Vácrátót, Alkotmány u. 2-4., Hungary
[12] Agriculture and Agri-Food Canada, Lethbridge Research and Development Centre, Canada
[13] Unit of Plant Diversity and Ecosystems Management, School of Biosciences and Veterinary Medicine, University of Camerino, Via Pontoni, 5, 62032, Camerino, Italy
[14] Instituto Multidisciplinario de Biología Vegetal (CONICET-UNC) and FCEFyN, Universidad Nacional de Córdoba, Velez Sarsfield 1611, Córdoba, Argentina
[15] Thompson Rivers University, Department of Natural Resource Science, Kamloops, British Columbia, 805 TRU Way, V2C 0C8, Canada
[16] University of Western Ontario, Department of Biology, London, Ontario, Canada
[17] Hungarian University of Agriculture and Life Sciences, Department of Botany, H-1118 Budapest, Menesi street 44, Hungary
[18] Department of Natural Resources, Islamic Azad University, Nour Branch, Iran
[19] University of Rostock, Soil Physics, Rostock, Germany. deceased
[20] Department of Botany - Te Tari Huaota, University of Otago, Dunedin, New Zealand
[21] School of Ecology and Environment, Inner Mongolia University, Hohhot, 010021 China
[22] University of Florida, Department of Biology, Gainesville, FL 32601 USA
[23] Hungarian University of Agriculture and Life Sciences, Environmental Sciences Doctoral School, H-2100 Gödöllő, Páter K. u. 1., Hungary
[24] Institute of Botany and Landscape Ecology, University of Greifswald, Germany

dajana.radujkovic@uantwerpen.be; sara.vicca@uantwerpen.be; gretel.vanrooyen@up.ac.za; wilf0020@umn.edu; lrbrown@unisa.ac.za; anke.jentsch@uni-bayreuth.de; kurt.reinhart@ars.usda.gov; cbrown2@ualberta.ca; Johan.DeGruyter@uantwerpen.be; gerald.jurasinski@uni-greifswald.de; d.askarizadeh@ut.ac.ir; bartha.sandor@ecolres.hu; ryan.beck@canada.ca; theodore.blenkinsopp@ualberta.ca; james.cahill@ualberta.ca; diego.campetella@unicam.it; roberto.canullo@unicam.it; stefano.chelli@unicam.it; lenrico@unc.edu.ar; lfraser@tru.ca xiying.hao@canada.ca; hhenry4@uwo.ca; hohnmariamargit@gmail.com; mjouri@gmail.com rachael.lawrencelodge@otago.ac.nz; lifyhong@126.com; janice.lord@otago.ac.nz ; pmilligan@ufl.edu; emh529@126.com; tmp@ufl.edu; birgit.schroeder@uni-rostock.de; kabazit@gmail.com; zh4ngtongrui@163.com; vadrezeda@gmail.com; Erik.Verbruggen@uantwerpen.be

Corresponding author: Dajana Radujković, dajana.radujkovic@uantwerpen.be, +32 471 125 466
University of Antwerp, Universiteitsplein 1, 2610 Wilrijk, Belgium.
https://orcid.org/0000-0003-4981-5879

Running title: Consistent predictors of microbial communities

Keywords: bacterial community composition, fungal community composition, spatial scales, plant community composition, plant productivity, soil properties

Abstract

Environmental circumstances shaping soil microbial communities have been studied extensively. However, due to disparate study designs, it has been difficult to resolve whether a globally consistent set of predictors exists, or context-dependency prevails. Here, we used a network of 18 grassland sites (11 of those containing regional plant productivity gradients) to examine i) if similar abiotic or biotic factors predict both large-scale (across sites) and regional-scale (within sites) patterns in bacterial and fungal community composition, and ii) if microbial community composition differs consistently at two levels of regional plant productivity (low vs high). Our results revealed that bacteria were associated with particular soil properties (such as base saturation) and both bacteria and fungi were associated with plant community composition across sites and within the majority of sites. Moreover, a discernible microbial community signal emerged, clearly distinguishing high and low-productivity soils across different grasslands independent of their location in the world. Hence, regional productivity differences may be typified by characteristic soil microbial communities across the grassland biome. These results could encourage future research aiming to predict the general effects of global changes on soil microbial community composition in grasslands and to discriminate fertile from infertile systems using generally applicable microbial indicators.

Introduction

Soil microbial communities, especially bacteria and fungi, perform fundamental ecological functions in terrestrial ecosystems, primarily through their role in biogeochemical processes (Falkowski et al., 2008; Tedersoo et al., 2014) and numerous direct and indirect interactions they form with plants (van der Heijden et al., 2008). The predictors of bacterial and fungal community composition have been thoroughly

studied in an attempt to understand soil microbial biogeography (de Vries et al., 2012; Fierer & Jackson, 2006; Tedersoo et al., 2014) and to be able to predict the effect of environmental changes on microbial communities and the functions they perform (Pold & DeAngelis, 2013). While global scale analyses strongly indicate that factors such as soil pH, climate, plant productivity (Delgado-Baquerizo et al., 2018; Fierer & Jackson, 2006) or climate and plant community composition (Prober et al., 2015; Větrovský et al., 2019) can universally predict some aspects of large-scale soil bacterial and fungal community composition, respectively, such relationships often cannot explain the variation in community composition at smaller scales. Indeed, several studies have indicated that the drivers of microbial community composition may strongly vary with spatial and/or environmental contexts (Chalmandrier et al., 2019; Martiny et al., 2011; Shi et al., 2018) and that predictability of the soil microbiome therefore depends on spatial scale (Averill et al., 2021). This disparity may simply be caused by the fact that gradients of explanatory variables are longer at global scales (ideally encompassing the full global range of e.g. climate, soil pH, or plant community composition), facilitating statistical power detecting them. However, it may also indicate that the factors driving soil communities are context-dependent. If this is the case, it strongly impedes our capacity to make general predictions about the effect of altered environmental conditions on soil communities and the processes they perform.

Variation in the strength and sign of ecological relationships under different environmental, spatial, or ecological settings (i.e. context-dependency) is common in nature (Chamberlain et al., 2014; Maestre et al., 2005; Tedersoo et al., 2015). Context-dependency in the processes that structure microbial communities may arise for several (non-mutually exclusive) reasons, including historical legacies (Fukami, 2015), stochastic events in community assembly processes (Beck et al., 2015), or dispersal limitation (Peay et al., 2010). The drivers of microbial community composition could thus strongly differ depending on region, presence of keystone taxa (Banerjee et al., 2018), or environmental conditions (Hendershot et al.,

2017). However, the apparent context-dependency could also arise as an artefact of a study design or due to methodological differences among studies (Catford et al., 2022).

Thus far, it has been challenging to examine the context-dependency vs. generality in predictors of bacterial and fungal community composition across sites because most studies have either been restricted in spatial extent or were otherwise not suitable to evaluate context-dependency. While global-scale studies can detect general patterns and predictors of microbial communities, the lack of local replication within these global studies complicates distinguishing between possible drivers that covary across locations. For instance, microbial and plant communities on the one hand, and soil properties on the other, both strongly covary with geographical distances and climate (Steidinger et al., 2019). Regional- and local-scale studies may be better suited to assess the effect of soil properties and plant communities along an environmental (e.g. productivity or fertility) gradient, but findings may not generalize across multiple individual gradients (Alzarhani et al., 2019). Hence, understanding context-dependency requires a simultaneous examination of large and regional (local) scale predictors of microbial community composition.

Here, we used a network of 18 grassland sites (each with two to six 64 m^2 plots; Fig. 1), 11 of which contained plots located along a regional gradient in plant productivity (Fraser et al., 2015), to examine the consistency of predictors of soil bacterial and fungal community composition under different spatial scales and environmental contexts. Given that grassland productivity is intrinsically related to biodiversity, soil fertility and plant-soil interactions (Delgado-Baquerizo et al., 2017; Guerrero-Ramírez et al., 2019), and therefore to the overall ecological functioning of the system, different regional productivity levels represent distinct underlying environmental contexts for the development of soil microbial communities. This thus provides an ideal setting to replicate a strong and contrasting ecological signal. For instance, at

high productivity, plant competition for light is strong, favouring acquisitive, fast-growing plant species (DeMalach et al., 2016). This could have add-on effects on soils: high input of easily decomposable plant litter selects for more acquisitive microbiota such as copiotrophic bacteria (Ling et al., 2017; Ramirez et al., 2010) which are selected over fungi, oligotrophic bacteria and microbes engaged in nutritional symbioses with plants (de Vries et al., 2007).

To examine whether similar predictors explain variation in microbial community composition across scales, we first analyse the importance of 17 environmental variables in explaining large-scale bacterial and fungal community dissimilarities and test if plant community composition can explain additional variation. We then examine whether important, regionally-varying, predictors identified at the large scale can likewise consistently predict within-site microbial community composition, thus truly ruling out any covariances between sites. Finally, we examine whether two contrasting grassland productivity levels (regionally low and high) have a consistent association with overall microbial community composition across different sites. If the drivers of microbial communities are strongly context-dependent, we hypothesize that the important environmental or plant community predictors identified at the large scale would be poor or inconsistent predictors of within-site variability. Likewise, if the effect of plant productivity on microbial community composition varies strongly across grassland sites (i.e. depending on climatic conditions, biogeography, soil type), we expect no significant differences between communities at high and low productivity levels as well as no common indicator taxa of particular productivity levels.

Methods

Sampling sites and data collection

Data were collected from 18 Herbaceous Diversity Network (HerbDivNet) grassland sites (Fraser et al., 2015) located in 12 countries (Fig. 1). The sites include different types of grasslands (xeric, mesic and hydric) spanning a wide range of climatic conditions (mean annual temperature ranges from 1.5 °C to 20.1 °C and mean precipitation ranges from 294 mm to 1237 mm). Peak annual biomass values spanned a range from 13 g/m^2 to 1187 g/m^2. Each of the 18 sites contained between two and six 8 × 8 m plots: 11 sites contained six plots, one site contained four plots, one site three plots and five sites contained two plots (Table S1); giving a total of 83 plots. Most sites were chosen to represent a site-specific gradient in productivity based on their plant biomass production; with six plots (two replicates of low, medium and high productivity) located within the same region with little to no variation in climatic conditions (Fraser et al., 2015). However, seven sites contained fewer plots and did not contain a productivity gradient (Fig. 1).

Soil sampling and storage

Soil samples were taken in a single sampling event at the peak of the growing season in the period between 2017 and 2018 (Table S1). For each plot within a site, five soil subsamples to a depth of 10 cm were taken from four corners and the centre of the plot. Subsamples for microbial analyses were taken and stored in pure ethanol (a total of 415 samples) and the rest of the sample was pooled into one composite sample (a total of 83 samples), air-dried and passed through a 2 mm sieve. Soil microbial communities were analysed separately per plot to capture the heterogeneity of the plot, restrict

coincidental inhibitory compounds to interfere with PCR of entire plots, and enable future additional analysis on within-plot variation. All samples were further analysed at the University of Antwerp. Samples for microbial analyses stored in ethanol were kept cool until DNA extraction (see below). Storage in ethanol has been shown to yield similar DNA recovery as cold storage conservation with successful amplifications after one year of storage (Harry et al., 2000).

Plant sampling

Plant species presence and total aboveground biomass were measured from each m^2 of each 64 m^2 plot in a single event at the peak of the growing season. Most sites were surveyed in 2012, but some were surveyed in later years (2015 or 2017; Table S1). Litter was first excluded from the total biomass and live biomass was dried and weighed. Based on this, the average peak biomass production [g/m^2] was calculated for each plot.

The data on the presence of different plant species at each 1 m^2 subplot of the plot (Fraser et al., 2015) was used to derive the 'frequency' of different species per 64 m^2 plot (e.g. the species that is found only in one subplot got a value of 1 and the one found in every subplot, a value of 64), which was used as a measure of relative abundance. Further analyses of plant community composition distances were based on species aggregated to genera (as in Prober et al. (2015)) rather than to the species level because plant species turnover across different plots and sites would often be 100% and thus produce continuous data for highly similar communities only, reducing information content. For most sites, plant sampling was performed several years before soil sampling (Table S1), however, the overall composition of dominant plant genera remained relatively unchanged over the years.

Climatic, nitrogen (N) deposition and soil data

Mean annual precipitation (MAP) [mm/y] and temperature (MAT) [°C] were derived from the CHELSA database (Karger et al., 2017) based on the geographical position (latitude and longitude) of each plot, which was also used to calculate geographical distances [km] between the plots. Data on total inorganic N deposition [kg/ha/yr] were derived from Ackerman et al. (2018). We used the average values over the years available in the database to account for long-term fertilization by atmospheric N deposition.

We analysed 14 soil properties: soil organic matter (SOM), total N, total carbon (C), total phosphorus (P), available P, base saturation (BS), cation exchange capacity (CEC), pH, soil texture (sand, clay, silt), extractable Ca, Mg and K. These soil properties are related to soil fertility and plant productivity (Vicca et al., 2018), are known to affect soil microbial community composition (de Vries et al., 2012; Tedersoo et al., 2014; Zheng et al., 2019) and can be compared across different sites.

Analyses of soil properties

SOM [%] was calculated as the loss of dry matter at 550°C expressed as a percentage of dry matter (Heiri et al., 2001). Total soil N and total C [%] were determined on ground soil, dried 48h at 70°C, on a Flash 2000 CN analyser (ThermoFisher Scientific, Waltham, MA, USA). Total P [ppm] was determined using acid digestion with H_2SO_4, salicylic acid, H_2O_2 and selenium (Novozamsky et al., 1983). Available P [ppm] was analysed following the Olsen extraction method (Olsen 1954) using a Continuous Flow Analyser (CFA) SAN++ (Skalar, Breda, The Netherlands). CEC [meq/100g] and BS [%] were estimated based on the exchangeable H^+ and total exchangeable bases – TEB [meq/100g]. For this, cations in ammonium acetate

extract [ppm] were measured following Reeuwijk (2002) using an Inductively Coupled Plasma Spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA) and acidity was determined following Brown (1943). pH of air-dried soil was measured using a pH meter (Hanna HI 3222; Hanna Instruments, Woonsocket, RI, USA) in 1:2.5 w:v soil in 1M KCl suspension (Blakemore et al., 1987). Soil texture was analysed by determining the percentage of primary particles (sand: 2000-53 µm, silt: 53-2.0 µm, and clay: < 2.0 µm), following the method of Gee & Bauder (1986).

Analyses of microbial communities

Sample preparation, sequencing and bioinformatics analyses

DNA was isolated from 415 soil samples of 0.25-0.35 g using the DNeasy PowerSoil Kit according to the manufacturer's protocol (Qiagen, Venlo, the Netherlands). The bacterial 16S V4 region was amplified using the 515F-806R primer pair (Caporaso et al., 2011) and the fungal ITS1 region was amplified using general fungal primers ITS1 and ITS2, modified according to Smith & Peay, (2014). Each 25 µl reaction mixture contained 2 µl of the sample, 0.5 µM of each forward and reverse primer, 1X PCR buffer, 200 µM dNTPs and 1 U Phusion High-Fidelity DNA polymerase (New England Biolabs, Ipswich, MA, USA). PCR conditions were as follows: initial denaturation at 98 °C for 60 s, followed by 30 (35 for fungi) cycles of: denaturation at 98 °C for 30 s, annealing at 55 °C for 30 s, extension at 72 °C for 30 s; and an additional extension of 72 °C for 10 min. The success of amplification was tested on 1.5% agarose gel. For the samples that did not amplify successfully, amplification was attempted again with a modified mixture that contained 2 µl of the sample and 1 µM of forward and reverse primer. Successful PCR products were diluted 50-fold and a second PCR was performed using dual barcoded primers with Illumina adapters (2.5 µl of diluted PCR products and 0.1 µM of each primer). The conditions were: 98 °C for 60 s, 12 cycles: at

98 °C for 10 s, 63 °C for 30 s, 72 °C for 30 s; and 72 °C for 5 min. PCR products were run on an agarose gel and successful amplicons were purified and normalized using the SequalPrep Normalization Plate Kit (ThermoFisher Scientific) and pooled into a single library. The library was purified using QIAquick Gel Extraction Kit (Qiagen, Venlo, the Netherlands) and quantified using qPCR (KAPA Library Quantification Kits, Kapa Biosystems, Wilmington, MA, USA). The libraries were sequenced via 2 x 300 cycles using the Illumina MiSeq platform (Illumina Inc; San Diego, CA, USA).

The sequences were analysed using the USEARCH (v8.1.1861) and VSEARCH (Rognes et al., 2016) software following the UPARSE pipeline (Edgar, 2013). After trimming to 280 bp and 250 bp for bacteria and fungi respectively, the paired-end reads were merged and primers were removed. This trim length was chosen because it was the optimal length for merging paired reads by removing reduced-quality bases at the end. Merged sequences were quality filtered using the expected number of errors (E) as a measure of read quality, with a threshold of $E_{max} = 0.5$. This yielded 10.8 M and 4.02 M of good-quality reads, for bacteria and fungi, respectively. Following singleton removal, the sequences were clustered into OTUs based on 97% similarity using the UPARSE-OTU algorithm (Edgar, 2013) which automatically detects and filters out chimaeras. Filtered reads were then mapped to the OTUs with an identity threshold of 0.97, yielding an OTU table for bacteria and fungi. Representative OTUs were aligned to the SILVA database (bacteria) (Quast et al., 2013) (release 138) and UNITE database (fungi) (Kõljalg et al., 2005) (release date 2.2.2019), using the *sintax* command in USEARCH with a 0.8 cut-off, resulting in 19,248 and 13,967 OTUs for bacteria and fungi, respectively.

Further steps were performed using R software (R Core Team, 2015). The number of reads per subsample was rarefied using the *rrarefy* function in *vegan* (Oksanen et al., 2015) to 6,046 for bacteria and 1,231 reads for fungi as rarefaction curves showed that the number of taxa was levelling off for most subsamples at these depths (Fig. S1). Most bacterial subsamples had more sequences than the chosen rarefaction

depth but 60 samples had fewer sequences than this threshold. 13 samples had too few sequences or were clear outliers (as determined by comparing them to other samples from the same plot using non-metric multidimensional scaling ordinations), and they were therefore discarded. 47 were upsampled to contain 6,046 sequences while retaining the original proportions, leaving 402 bacterial samples. As we were interested in plot-level community composition, this procedure was done in cases when it could be verified that the upsampled bacterial communities do not notably deviate from those in other subsamples of their group (i.e. those sampled from the same plot) which demonstrated that their overall quality was not compromised (Fig. S2 and Fig. S3a). For fungi, 13 samples had too few sequences and were considered as failed, leaving 402 samples in total. Both for bacteria and fungi, at least 3 samples per plot were retained. Several technical replicates were sequenced (same sample but separate PCR procedure and sequencing) showing clustering with each other and other samples of the same plot (see example in Fig. S3). Given that we chose to sequence five subsamples and then aggregate them per plot (see below), rather than pool them and sequence one sample per plot, subsamples contained a relatively low number of sequences, but each plot was ultimately represented by approximately 30,000 and 6,000 sequences for bacteria and fungi, respectively.

To annotate *sintax*-assigned fungal sequences to known genera in the UNITE database, we used NCBI's BLAST algorithm with default settings. OTUs were then assigned to particular taxa if they had a maximum E-value of 10^{-36} and from this, the lowest E-value hit with a known genus was selected. If there were none, the genus level was left unassigned. OTUs were subsequently assigned to three major functional groups: saprotrophs, pathotrophs and symbiotrophs using FUNGuild (Nguyen et al., 2016).

Analysis of microbial abundance

DNA extracts of the five subsamples per plot were pooled to give 83 plot samples in total. The abundance of bacterial and fungal gene copies per sample was quantified using qPCR targeting the 16s V4 region (with the 515F–806R primer pair) for bacteria and the 18s region for fungi (primer set FR1 / FF390 (Chemidlin Prévost-Bouré et al., 2011)), which were chosen because high length variation of the ITS1 region precludes accurate quantification. Each 20 µl reaction mixture contained 4 µl of the sample, 0.5 µM of each forward and reverse primer, 1 x ROX high and 10 µl of KAPA SYBR FAST qPCR master mix (Kapa Biosystems, Wilmington, MA, USA). qPCR conditions were as follows: initial denaturation at 95°C for 3 min, followed by 40 cycles of: denaturation at 95 °C for 3 s, annealing at 57 °C (52 °C for fungi) for 20 s, extension at 72 °C for 12 s; finishing with 35 s at 50 °C. Prior gel-electrophoresis with these primers and reaction conditions showed the reactions were highly specific. Melting curve analysis of all amplicons was conducted to confirm that fluorescence signals originated from specific amplicons and not from primer-dimers or other artefacts. Standard curves were generated using duplicates of 10-fold dilutions of amplicons derived using the same primers, isolated from gel using QIAquick Gel Extraction Kit (Qiagen, Venlo, the Netherlands) and quantified using Qubit fluorometer (Invitrogen GmbH, Karlsruhe, Germany). Fungal and bacterial gene copy numbers were derived from a regression equation based on the standard curves (with minimal $R^2 > 0.99$) by relating the quantification cycle (Cq) value of each sample to the Cq values of standards with the known number of copies. All reactions were performed in duplicate and the number of bacterial and fungal copies was then averaged (a deviation of Cq between replicates < 1 was used as a passing criterion) and expressed per g of soil dry weight.

Statistical analyses

Do large-scale predictors consistently explain regional-scale variation in microbial community composition?

In this analysis, we first examined which environmental variables were the best predictors of bacterial and fungal community composition at large scale, across all plots. Then, we used the selected predictors to create an environmental distance matrix per region and examined how well it predicted regional (within-site) community composition for those regions that contained more than three plots. Moreover, we examined to what extent plant community distances can contribute to explaining large-scale variation when the environment is taken into account and how well plant distances explain regional variation compared to the environment.

Prior to analyses, we averaged the OTU relative abundances of five subsamples per plot (83 plots in total) to obtain one community measure per plot. All environmental variables (except pH and BS) were transformed using square root transformation, centred and scaled to reduce positive skewness and to allow for the comparison of effect sizes. Community data (fungi, bacteria, plants) were transformed with Hellinger transformation using the *decostand* function in the *vegan* package in R.

Climate, N deposition, geographical distances, soil variables and plant biomass were used as potential predictors of large-scale variation in microbial community composition (Table S2). The influence of different factors on the dissimilarity in bacterial and fungal communities across all the plots was analysed using multiple regression on distance matrices (MRM) in the *ecodist* package (Goslee & Urban, 2007) adopting the 'rank' method. A large-scale MRM model was first fitted using Bray-Curtis (BC)

bacterial/fungal distances as response variables and Euclidean distances in individual environmental variables and as predictors. In the case of two highly correlated variables (Pearson $r > 0.8$), only one of them was included in the model to avoid multicollinearity (SOM was omitted due to a high correlation with N and Ca due to a high correlation with CEC). Based on backward selection, the variables that did not significantly contribute to the model were sequentially removed leaving only those with $P < 0.05$. This was done to comprehensively capture the effect of the environment while accounting for the effect of geographical distances. To test if plant community dissimilarities can explain any unique (non-shared) variation in microbial community composition, we added it to the MRM model with the selected environmental variables after confirming that plant community distances are not highly correlated with the distances of environmental variables (selected environmental variables for bacterial and fungal models together explained 18% and 47% of the variation in plant community composition, respectively). Finally, we partitioned the variation explained by three major groups of variables in the MRM model: i) broad-scale predictors that vary strongly across large spatial scales – atmospheric variables and geographic distances, ii) regionally varying predictors – soil properties and plant biomass and iii) community predictor – plant community composition (Table S2). This was done by examining the variance explained (R^2 values) by each of the three groups separately in the MRM model and then calculating the percentage of shared variation between the groups. The variance partitioning results were visualised using the *eulerr* package in R.

To examine if the observed large-scale relationships persist at the regional scale, we created two environmental distance matrices (one for bacteria and one for fungi) using significant regionally varying predictors from the large-scale MRM model. The values for different variables were scaled and multiplied by the corresponding regression coefficient in the large-scale environmental MRM model (the coefficients represent the 'strength' of each variable as a predictor of the bacterial or fungal community) after which

the Euclidean distance matrices were created using the '*dist*' function. For 12 sites containing more than three plots, we used the MRM function, with the same settings as for the large-scale model, to examine the relationship between bacterial or fungal community distances and 1) environmental (Euclidean) and 2) plant (Bray-Curtis) distances. To assess the consistency of these relationships (environment – bacteria, plants – bacteria, environment – fungi, plants – fungi) across different sites, we calculated the variance of regression coefficients and reported, mean R^2 values and their standard deviations (sd). Hence, the higher and more consistent R^2 values (lower standard deviation) and the lower variance of regression coefficients for different sites, the higher the consistency of relationships.

Microbial community composition at different regional relative productivity levels

Our within-site productivity gradients allowed us to test whether there are consistent differences between relatively low-productivity and relatively high-productivity grasslands regardless of environmental context. For this analysis, the dataset was divided into two subsets: two replicate plots from each site with relatively high biomass production (i.e. high productivity plots) and two from each site with relatively low biomass production (i.e. low productivity plots) so that plots with high productivity within a site had at least 100% higher biomass than low productivity plots. Eleven sites met this relatively stringent productivity criterium, yielding two datasets each containing 22 plots (Fig. S4).

The differences in bacterial and fungal abundances (number of gene copies), bacterial to fungal ratio and three main fungal functional groups at low compared to high productivity levels were analysed using the *lme* function in *nlme* package with 'site' as a random effect.

To test if bacterial and fungal communities differed significantly between the two productivity levels with a consistent pattern across different sites, we performed PERMANOVA analysis using the *adonis* function in *vegan* adding 'site' as *strata* to control for inherent community differences between sites. We used multidimensional scaling (MDS) ordination to visualise the BC distance in bacterial and fungal communities (based on OTUs) at different productivity levels after controlling for the effect of site (i.e. that communities come from 11 different sites) using the *dbrda* function in *vegan*. To examine if the best predictors of bacterial and fungal community composition differed at different productivity levels, we created MRM models as described above for each of the productivity levels. Furthermore, we examined the OTUs that are significantly related to either low or high productivity levels using the *signassoc* function (with 999 permutations) from the *indicspecies* package. To avoid spurious correlation and increase the power of the analyses, we included only bacterial and fungal OTUs appearing in more than 50% and 25% of plots, respectively (the threshold for bacteria is higher as many more bacterial OTUs were shared across the plots). Bacterial and fungal OTUs which were significantly associated with one of the productivity levels (Sidak-corrected P-values < 0.05) were considered as indicator taxa. Fungal indicator OTUs were assigned to putative functional groups according to FUNGuild.

We selected the 20 most abundant bacterial and fungal families and tested if there was a significant difference in their abundance at different productivity levels using the *lme* function in *nlme* package with 'site' as a random intercept term. The P values for the test were corrected using Benjamini-Hochberg correction and the regression coefficients and standard deviations were visualised along with the total relative abundances of families per productivity level.

Results

Predictors of microbial community composition at large vs. regional scale

Our results revealed that the most important predictors of large-scale variation in bacterial community composition were soil properties (particularly soil base saturation and pH; Fig. 2a; Table S3). Plant community composition shared much of the variation with soil properties and atmospheric factors, but it still explained 5% of additional variation not accounted for by environmental predictors (Fig. 2a). At the regional scale, environmental distances (composed of the same soil predictors identified in the large-scale model) were a fairly good and consistent predictor of the regional-scale variation across sites (regression coefficient variance = 0.07; mean R^2 = 0.52, sd R^2= 0.28) (Fig. 3a; Fig. S5). Plant community composition was equally strongly associated with regional variation in bacterial community composition for most sites (regression coefficient variance = 0.06; mean R^2 = 0.52, sd R^2 = 0.29) (Fig. 3b; Fig S5).

The consistency between large- and regional-scale predictors was found for fungi as well, where the best large-scale predictor – plant community composition (Fig. 2b) – was also associated with regional variation in fungal community composition for most sites (regression coefficient variance = 0.08; mean R^2 = 0.47, sd R^2 = 0.28) (Fig. 3d; Fig. S5). Plant community composition alone was a better predictor at the large scale than all environmental (atmospheric and soil) variables combined (R^2 = 0.42 and R^2 = 0.38, respectively) (Fig. 2b, Table S4). Accordingly, the relationship between fungal community composition and the environmental distances was more variable from site to site (regression coefficient variance = 0.10; mean R^2 = 0.40, sd R^2 = 0.30) (Fig. 3c; Fig. S5). The relationships between regional geographical distances and bacterial/fungal distances per site are shown in Fig. S6.

Microbial community composition at different plant productivity levels

Bacterial and fungal community composition differed significantly between the two productivity levels (Fig. 4) when site differences were accounted for ($P < 0.001$, F.model = 2.2 and 1.2, $R^2 = 0.05$ and 0.03, respectively; Table S5). This indicates that there is a common community, shared across the widely distributed sites, which can separate regionally more and less productive grasslands. Despite these compositional differences, the predictors of microbial community composition at low and high productivity levels were similar and in line with the results in the previous section: base saturation was the most important predictor of bacterial community composition, whereas fungal community composition was most strongly associated with plant community composition (Table S6).

The linear mixed-effect model with 'site' as a random effect showed that total bacterial and fungal abundance (number of gene copies) was significantly higher in low than in high productivity plots ($P < 0.05$ and $P < 0.01$, respectively; Fig. 4b), while fungal: bacterial ratio did not differ significantly (Table S7). The relative abundance of the main fungal functional groups (saprotrophs, pathotrophs, symbiotrophs), did not differ significantly between the two productivity levels (Table S7).

Indicator species analysis demonstrated that there were 62 and 108 bacterial OTUs significantly associated with high and low productivity plots, respectively (Table S8). The most abundant indicators ($P<0.05$) of high-productivity grasslands belonged to the Bacilliaceae family along with the Clostridiaceae and family ELEV-16s-1332 from the Solirubrobacterales order (Fig. 5a). Interestingly, mixed effect models showed that the relative abundance of Bacilliaceae (and the genus *Bacillus*), which is by far the most dominant family at both productivity levels, was significantly higher at high productivity levels (Padj. < 0.05), and the relative abundance of family ELEV-16s-1332 was near-significantly (Padj < 0.1) increased at

high productivity levels (Fig. 5b). At low productivity, many of the highly abundant indicator OTUs came from the families Blastocatellaceae, Pseudonocardiaceae, Methylobacteriaceae and Rubrobacteraceae. Precisely the family Rubrobacteraceae (and its genus *Rubrobacter*) was significantly enriched at low compared to high productivity levels along with the family 319-6m6 from the Solirubrobacterales order whose members also appear as common indicators of low productivity grasslands (Fig. 5; Table S8).

For fungi, there were 8 indicators of high and 3 indicators of low productivity grasslands (Table S9). The indicators of high-productivity grasslands came from various families, four of them belonging to the order Hypocreales and two to the order Pleosporales (Fig. 6a). Moreover, most of them were assigned as putative pathotrophs (Table S9). Although none of the 20 most abundant fungal families was significantly different at different productivity levels, the families Pleosporaceae and Herpotrichiellaceae tended to be enriched at low productivity levels (Padj < 0.1; Fig. 6b).

Discussion

Despite considerable literature describing the most important predictors of soil microbial community composition in the grassland biome, until now it has been unclear whether these relationships are strongly context-dependent or whether they truly persist across different spatial scales and environmental contexts. In this study, we show that there may be generality in the way bacterial and fungal communities are shaped across two different spatial scales and productivity levels in grasslands.

Generality in the predictors of microbial community composition

Our results showed that soil abiotic factors (primarily base saturation and pH) are among the strongest predictors of bacterial community composition both across and within different grassland sites and at contrasting plant productivity levels; a finding that is consistent with previous research emphasising the importance of soil chemical properties (particularly pH) for continental-scale bacterial community turnover (Fierer & Jackson 2006; Lauber et al., 2009; Delgado-Baquerizo et al., 2018). However, within-site bacterial community composition was at the same time consistently associated with plant community composition for most sites. Due to the strong covariation of soil properties and plant communities within regions, it was difficult to separate their potential effects on bacterial communities. Moreover, a validation with data not included in the study could strengthen our general findings of the congruence between the large-scale and regional-scale predictors. Fungal communities were fairly strongly correlated with climatic variables such as mean annual temperature and atmospheric N deposition at the global scale as demonstrated in other large-scale studies (Tedersoo et al., 2014, Prober et al., 2015). Yet, contrary to the findings of Tedersoo et al. (2014), that climate and edaphic factors are the strongest predictors of global fungal community composition across different biomes, we demonstrate that plant community

composition was by far the best predictor of large-scale as well as the regional-scale patterns in grassland fungal communities. The relationship with soil abiotic factors was weaker at the large scale and agreed less between different sites. Plant community composition can thus predict both bacterial and fungal community composition within and across different grassland regions, whereas soil chemical properties might be stronger predictors of large-scale variation in bacterial communities. A clear link between plant and microbial community composition was found even though plants and soil were sampled in different years, implying that the strength of this association might even be somewhat underestimated in our study.

While the strong relationships between grassland plant and microbial, particularly fungal, community composition have previously been documented (Chalmandrier et al., 2019; Prober et al., 2015), the consistency of these relationships across different grasslands found in our study provides evidence that they are not just a matter of coincident spatial community turnover between microbes and plants. Instead, they are likely a result of similar environmental factors structuring plant and microbial communities, plant-microbe interactions, or both. Local experiments have shown that plant-microbe feedbacks play a central role both in microbial and plant community assembly processes (Radujković et al., 2020; Wubs et al., 2019). In such feedbacks, plant communities affect soil microorganisms both directly by providing a diverse set of hosts for mutualistic and antagonistic microorganisms and indirectly by altering edaphic factors and providing different quantity and quality of root exudates and litter (van der Heijden et al., 2008; Berg & Smalla 2009), where changes in microbial community composition in turn feedback to influence plant communities (van der Putten et al., 2016). While the data for the large-scale and within-site plant-microbe relationships in this study would ideally be independent, our research conclusively demonstrates that plant community composition 1) explained additional variation on top of all environmental factors across sites and 2) was a rather consistent predictor of both bacterial and fungal communities within most sites. These findings provide the support that plant-microbe feedbacks are

indeed important factors in shaping bacterial and fungal community composition in grasslands across large and regional scales. It is possible, however, that these plant-microbial relationships are much weaker at small scales, e.g. within the plots as shown by Chalmandrier et al. (2019) where predictability of microbial communities might decrease (Averill et al., 2021).

Universal link between plant productivity and soil microbial community composition

Bacterial and fungal community compositions were found to be more similar within low and high-productivity grasslands than between them when site-specific differences were accounted for. This suggests that plant productivity, and/or the myriad of factors related to it (including soil properties, soil fertility, plant diversity, and plant-soil interactions (Delgado-Baquerizo et al., 2017; Guerrero-Ramírez et al., 2019)), select for some of the same microbial taxa regardless of differences in climate and grassland type. Alternatively, these microbial taxa might be in part responsible for modulating plant productivity across different grasslands (van der Heijden et al., 2008). A link between bacterial taxa and plant productivity across contrasting biomes worldwide (forests, shrublands, grasslands) has previously been reported (Delgado-Baquerizo et al., 2018), where particular groups of globally dominant soil bacteria with a preference for low-productive sites were identified. Here, we show that similar conclusions hold for bacterial and fungal taxa even within the grassland biome, where differences in plant productivity are much smaller than across contrasting biomes.

The relative abundances of the three dominant fungal functional groups (saprotrophs, symbiotrophs and pathotrophs) did not differ significantly between productivity levels but total fungal abundance was significantly higher at low compared to high-productivity soils. Higher fungal abundance is common in less fertile soils (Bardgett & McAlister, 1999; Innes et al., 2004) where fungi are favoured over bacteria as the

predominant decomposers due to the higher recalcitrance of plant litter and their generally more resource-conservative lifestyles (Marschner et al., 2011). Moreover, plant reliance upon, and allocation to AMF is often higher in less fertile soils to secure P, N and other nutrients (Ven et al., 2019; Verbruggen et al., 2013). Surprisingly, in our study, bacterial abundance also increased slightly but significantly in low-productivity grasslands leading to no significant difference in bacterial to fungal ratio. It is possible that in low-productive grasslands there is an increase in oligotrophic bacteria capable of thriving under low resource conditions, compensating for the supposed decrease in copiotrophs relative to fungi.

There were 170 bacterial taxa identified as indicators of high and low-productivity soils. Specifically, the family Bacilliaceae and genus *Bacillus*, which were the most dominant family and genus in our grassland soils, 1) were significantly enriched at high productivity levels and 2) contained the most abundant indicator OTU of high-productivity grasslands. These findings demonstrate that members of the *Bacillus* genus could be reliable indicators of high-productivity soils across different grassland types. Members of this genus are well known as plant growth-promoting bacteria that can stimulate plant productivity by: releasing different volatile organic compounds that upregulate photosynthetic activity, reducing abiotic and biotic stress (biocontrol of plant pathogens), improving carbon sequestration processes, helping plants to acquire nutrients, e.g. thanks to their capacity to fixate N from the air (Gurikar et al., 2022; Mahapatra et al., 2022; Saxena et al., 2020). Clostridiaceae, whose members were also found to be abundant indicators of high productivity, is another family that contains taxa known to promote plant growth and perform N fixation under anaerobic conditions (Kennedy et al., 2004). Besides them, the members of family Elev-16s-1332 from Solirubrobacterlaes order were notably enriched at high productivity soils which is in line with the previous findings showing that this family is related to the high-yield production system in a maize field (Yu et al., 2022). On the other hand, the family Rubrobacteraceae and its genus *Rubrobacter* were significantly enriched in low-productive soil with several members that

were among the most abundant indicators of this productivity level. Rubrobacteraceae are highly desiccation-tolerant bacteria that have been shown to maintain a high intracellular concentration of osmoprotectants (Meier et al., 2021) and they could help plants to survive during low water levels in the soil. Such dry conditions are relatively common in our set of low-productive grasslands, which predominantly have sandy soils with poor water-holding capacity. Another family with increased abundance in low-productive soils and that contained multiple indicator taxa of such soils is Blastocatellaceae, whose members were typically discovered in dry, sandy soils depleted in nutrients (Ivanova 2020; Huber 2017). Family 0319.6M6 belonging to the order Solirubrobacterales is another potentially important indicator of low-productive grasslands, but currently not much is known about the ecology of members of this family.

Compared to bacteria, only a few fungal indicator taxa were identified, particularly in low-productivity grasslands. Therefore although we observed a clear response of the entire fungal community composition to different productivity levels, these effects may not be strong or consistent enough to be detected with indicator species analysis. The indicators of high-productive soils belonged to several different families but almost all of them were classified as pathotrophs including putative plant pathogens from the Hypocreales order. Plant pathogens often thrive under conditions of high productivity (Reynolds et al., 2003), and our result suggests that some of them (such as members of the Hypocreales order) are shared across different high-productive grasslands. Among significant fungal indicators of low-productive grasslands, we found the arbuscular mycorrhizal genus Glomus. AMF are known to be fairly cosmopolitan (Davison et al., 2015), and to improve plant growth in mycorrhizal symbiosis. They often decline with fertilization, particularly P (e.g. Lekberg et al. (2021)), which is especially true for those that predominate in grassland ecosystems (Verbruggen et al., 2015).

Conclusion

Several studies have argued that there are few general drivers of microbial community composition due to high levels of context dependency. If estimates derived from one system or spatial scale cannot be extrapolated to another, it is challenging to predict the effects of altered environmental conditions on soil microbial communities and the functions they perform. Our findings provide among the strongest evidence to date that the main factors that predict overall microbial (bacterial and fungal) community composition in grasslands are rather consistent, regardless of spatial scale, productivity, or climatic conditions. Moreover, particular regional plant productivity levels (low vs. high) are typified by relatively similar soil microbial communities across the grassland biome and are distinguishable by that characteristic. Particularly important bacterial indicators of high- and low- and low-productivity grasslands were members of the genera *Bacillus* and *Rubrobacter*, respectively. These results suggest that modelling soil microbial community composition under environmental changes or using generally applicable microbial indicators to discriminate fertile from infertile systems, are feasible tasks.

Acknowledgements

This research was supported by the Research Foundation—Flanders (FWO), the European Research Council grant ERC-SyG-610028 IMBALANCE-P and Methusalem funding of the Research Council UA. BS, ZZ and GS were supported by the GINOP-2.3.2-15-2016-00019 project. We thank Annelies Devos, Tom Van der Spiet, Kevin Van Sundert and Irene Ramirez Rojas for their help with the laboratory analyses of soil samples and three anonymous reviewers for providing suggestions that helped us to improve the manuscript. GJ wants to particularly mention the contribution of Marian Koch, who died suddenly and

unexpectedly in April 2020. Marian was a wonderful person and a dedicated vegetation ecologist who organized and led the fieldwork at the site Ger.r. We miss him.

References

Ackerman, D., Chen, X., & Millet, D. (2018). *Global nitrogen deposition (2°×2.5° grid resolution) simulated with GEOS-Chem for 1984-1986, 1994-1996, 2004-2006, and 2014-2016.* Retrieved from the Data Repository for the University of Minnesota.

Alzarhani, A. K., Clark, D. R., Underwood, G. J. C., Ford, H., Cotton, T. E. A., & Dumbrell, A. J. (2019). Are drivers of root-associated fungal community structure context specific? *ISME Journal*, *13*(5), 1330–1344.

Averill, C., Werbin, Z. R., Atherton, K. F., Bhatnagar, J. M., & Dietze, M. C. (2021). Soil microbiome predictability increases with spatial and taxonomic scale. *Nature Ecology & Evolution*, 1–10.

Banerjee, S., Schlaeppi, K., & van der Heijden, M. G. A. (2018). Keystone taxa as drivers of microbiome structure and functioning. *Nature Reviews Microbiology*, *16*(9), 567–576.

Bardgett, R. D., & McAlister, E. (1999). The measurement of soil fungal:bacterial biomass ratios as an indicator of ecosystem self-regulation in temperate meadow grasslands. *Biology and Fertility of Soils*, *29*(3), 282–290.

Beck, S., Powell, J. R., Drigo, B., Cairney, J. W. G., & Anderson, I. C. (2015). The role of stochasticity differs in the assembly of soil- and root-associated fungal communities. *Soil Biology and Biochemistry*, *80*, 18–25.

Berg, G., & Smalla, K. (2009). Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. *FEMS Microbiology Ecology*, *68*(1), 1–13.

Blakemore, L. C., Searle, P. L., & Daly B. K. (1987). Methods for chemical analysis of soils. *NewZealand Soil Bureau Scientific Report*, *80*.

Brown, I. C. (1943). Rapid method of determining exchangeable hydrogen and total exchangeable bases of soils. *Soil Science*, *56*, 53–357.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Fierer, N., & Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 4516–4522.

Catford, J. A., Wilson, J. R. U., Pyšek, P., Hulme, P. E., & Duncan, R. P. (2022). Addressing context dependence in ecology. *Trends in Ecology and Evolution*, *37*(2), 158–170.

Chalmandrier, L., Pansu, J., Zinger, L., Boyer, F., Coissac, E., Génin, A., ... & Thuiller, W. (2019). Environmental and biotic drivers of soil microbial β-diversity across spatial and phylogenetic scales. *Ecography*, *42*(12), 2144-2156.

Chamberlain, S. A., Bronstein, J. L., & Rudgers, J. A. (2014). How context dependent are species interactions? In *Ecology Letters* (Vol. 17, Issue 7, pp. 881–890). Blackwell Publishing Ltd.

Chemidlin Prévost-Bouré, N., Christen, R., Dequiedt, S., Mougel, C., Lelièvre, M., Jolivet, C., Shahbazkia, H. R., Guillou, L., Arrouays, D., & Ranjard, L. (2011). Validation and application of a PCR primer set to quantify fungal communities in the soil environment by real-time quantitative PCR. *PLoS ONE*, *6*(9), e24166.

Craven, D., Isbell, F., Manning, P., Connolly, J., Bruelheide, H., Ebeling, A., ... & Eisenhauer, N. (2016). Plant diversity effects on grassland productivity are robust to both nutrient

enrichment and drought. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1694), 20150277.
Davison, J., Moora, M., Öpik, M., Adholeya, A., Ainsaar, L., Bâ, A., ... & Zobel, M. (2015). Global assessment of arbuscular mycorrhizal fungus diversity reveals very low endemism. *Science*, *349*(6251), 970-973.
de Vries, F. T., Bloem, J., van Eekeren, N., Brusaard, L., & Hoffland, E. (2007). Fungal biomass in pastures increases with age and reduced N input. *Soil Biology and Biochemistry*, *39*(7), 1620–1630.
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., Kattge, J., & Bardgett, R. D. (2012). Abiotic drivers and plant traits explain landscape-scale patterns in soil microbial communities. *Ecology Letters*, *15*(11), 1230–1239.
Delgado-Baquerizo, M., Oliverio, A. M., Brewer, T. E., Benavent-González, A., Eldridge, D. J., Bardgett, R. D., Maestre, F. T., Singh, B. K., & Fierer, N. (2018). A global atlas of the dominant bacteria found in soil. *Science*, *359*(6373), 320–325.
Delgado-Baquerizo, M., Powell, J. R., Hamonts, K., Reith, F., Mele, P., Brown, M. V., Dennis, P. G., Ferrari, B. C., Fitzgerald, A., Young, A., Singh, B. K., & Bissett, A. (2017). Circular linkages between soil biodiversity, fertility and plant productivity are limited to topsoil at the continental scale. *New Phytologist*, *215*(3), 1186–1196.
DeMalach, N., Zaady, E., & Kadmon, R. (2016). Light asymmetry explains the effect of nutrient enrichment on grassland diversity. *Ecology Letters*, 1–10. https://doi.org/10.1111/ele.12706
Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, *10*(10).
Falkowski, P. G., Fenchel, T., & Delong, E. F. (2008). The microbial engines that drive Earth 's biogeochemical cycles. *Science*, *320*(5879), 1034–1039.
Fierer, N., & Jackson, R. B. (2006). The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(3), 626–631.
Fraser, L. H., Pither, J., Jentsch, A., Sternberg, M., Zobel, M., Askarizadeh, D., Bartha, S., Beierkuhnlein, C., & Bennett, J. A. (2015). Worldwide evidence of a unimodal relationship between productivity and plant species richness. *Science*, *349*(6245), 302–306.
Fukami, T. (2015). Historical contingency in community assembly: integrating niches, species pools, and priority effects. *Annual Review of Ecology, Evolution, and Systematics*, *46*(1), 1–23.
Gee, G. W., & Bauder, J. W. (1986). Particle size analysis. In 'Methods of soil analysis. Part I. Physical and mineralogical methods'. *Agronomy Monograph*, *9*.
Goslee, S. C., & Urban, D. L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, *22*(7), 1–19.
Guerrero-Ramírez, N. R., Reich, P. B., Wagg, C., Ciobanu, M., & Eisenhauer, N. (2019). Diversity-dependent plant–soil feedbacks underlie long-term plant diversity effects on primary productivity. *Ecosphere*, *10*(4), e02704.
Gurikar, C., Gowda, N. A. N., Hanumantharaju, K. N., & Netravati, B. P. (2022). Role of Bacillus species in soil fertility with reference to rhizosphere engineering. *Rhizosphere Engineering*, 65–76.
Harry, M., Gambier, B., & Garnier-Sillam, E. (2000). Soil conservation for DNA preservation for bacterial molecular studies. *European Journal of Soil Biology*, *36*(1), 51–55.
Heiri, O., Lotter, A. F., & Lemcke, G. (2001). Loss on ignition as a method for estimating organic and carbonate content in sediments: Reproducibility and comparability of results. *Journal of Paleolimnology*, *25*(1), 101–110.

Hendershot, J. N., Read, Q. D., Henning, J. A., Sanders, N. J., & Classen, A. T. (2017). Consistently inconsistent drivers of microbial diversity and abundance at macroecological scales. *Ecology*, *98*(7), 1757–1763.

Huber, K. J., Geppert, A. M., Groß, U., Luckner, M., Wanner, G., Cooper, P., ... & Overmann, J. (2017). Aridibacter nitratireducens sp. nov., a member of the family Blastocatellaceae, class Blastocatellia, isolated from an African soil. *International journal of systematic and evolutionary microbiology*, *67*(11), 4487-4493.

Innes, L., Hobbs, P. J., & Bardgett, R. D. (2004). The impacts of individual plant species on rhizosphere microbial communities in soils of different fertility. *Biology and Fertility of Soils*, *40*(1), 7–13.

Ivanova, A. A., Zhelezova, A. D., Chernov, T. I., & Dedysh, S. N. (2020). Linking ecology and systematics of acidobacteria: Distinct habitat preferences of the Acidobacteriia and Blastocatellia in tundra soils. *PloS one*, *15*(3), e0230157.

Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., & Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, *4*. 1-20.

Kennedy, I. R., Choudhury, A. T. M. A., & Kecskés, M. L. (2004). Non-symbiotic bacterial diazotrophs in crop-farming systems: can their potential for plant growth promotion be better exploited? *Soil Biology and Biochemistry*, *36*(8), 1229–1244.

Kõljalg, U., Larsson, K.-H., Abarenkov, K., Nilsson, R. H., Alexander, I. J., Eberhardt, U., Erland, S., Høiland, K., Kjøller, R., Larsson, E., Pennanen, T., Sen, R., Taylor, A. F. S., Tedersoo, L., & Vrålstad, T. (2005). UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. *New Phytologist*, *166*(3), 1063–1068.

Lauber, C. L., Hamady, M., Knight, R., & Fierer, N. (2009). Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Applied and Environmental Microbiology*, *75*(15), 5111–5120.

Lekberg, Y., Arnillas, C. A., Borer, E. T., Bullington, L. S., Fierer, N., Kennedy, P. G., Leff, J. W., Luis, A. D., Seabloom, E. W., & Henning, J. A. (2021). Nitrogen and phosphorus fertilization consistently favor pathogenic over mutualistic fungi in grassland soils. *Nature Communications*, *12*(1), 3484.

Ling, N., Chen, D., Guo, H., Wei, J., Bai, Y., Shen, Q., & Hu, S. (2017). Differential responses of soil bacterial communities to long-term N and P inputs in a semi-arid steppe. *Geoderma*, *292*, 25–33.

Maestre, F. T., Valladares, F., & Reynolds, J. F. (2005). Is the change of plant-plant interactions with abiotic stress predictable? A meta-analysis of field results in arid environments. *Journal of Ecology*, *93*(4), 748–757.

Mahapatra, S., Yadav, R., & Ramakrishna, W. (2022). Bacillus subtilis impact on plant growth, soil health and environment: Dr. Jekyll and Mr. Hyde. *Journal of Applied Microbiology*, *132*(5), 3543–3562.

Marschner, P., Umar, S., & Baumann, K. (2011). The microbial community composition changes rapidly in the early stages of decomposition of wheat residue. *Soil Biology and Biochemistry*, *43*(2), 445–451.

Martiny, J. B. H., Eisen, J. A., Penn, K., Allison, S. D., & Horner-Devine, M. C. (2011). Drivers of bacterial β-diversity depend on spatial scale. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(19), 7850–7854.

Meier, D. V., Imminger, S., Gillor, O., & Woebken, D. (2021). Distribution of Mixotrophy and Desiccation Survival Mechanisms across Microbial Genomes in an Arid Biological Soil Crust Community. *MSystems*, *6*(1).

Nguyen, N. H., Song, Z., Bates, S. T., Branco, S., Tedersoo, L., Menke, J., Schilling, J. S., & Kennedy, P. G. (2016). FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecology*, *20*, 241–248.

Novozamsky, I., Houba, V. J. G., van Eck, R., & van Vark, W. (1983). A novel digestion technique for multi-element plant analysis. *Communications in Soil Science and Plant Analysis*, *14*(3), 239–248.

Oksanen, J., et al. (2015). *Vegan: community ecology package. R package version 2.4-0.* https://github.com/vegandevs/vegan.

Olsen S, Cole C, Watanabe F, D. L. (1954). Estimation of available phosphorus in soils by extraction with sodium bicarbonate. *USDA Circular Nr 939, US Gov. Print. Office, Washington, D.C*.

Peay, K. G., Bidartondo, M. I., & Elizabeth Arnold, A. (2010). Not every fungus is everywhere: scaling to the biogeography of fungal-plant interactions across roots, shoots and ecosystems. *New Phytologist*, *185*(4), 878–882.

Pold, G., & DeAngelis, K. M. (2013). Up against the wall: the effects of climate warming on soil microbial diversity and the potential for feedbacks to the carbon cycle. *Diversity*, *5*(2), 409-425.

Prober, S. M., Leff, J. W., Bates, S. T., Borer, E. T., Firn, J., Harpole, W. S., ... & Fierer, N. (2015). Plant diversity predicts beta but not alpha diversity of soil microbes across grasslands worldwide. *Ecology letters*, *18*(1), 85-95.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glockner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, *41*(D1), D590–D596.

R Core Team. (2015). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Vienna, Austria, URL https://www.R-project.org/.

Radujković, D., Diggelen, R., Bobbink, R., Weijters, M., Harris, J., Pawlett, M., Vicca, S., & Verbruggen, E. (2020). Initial soil community drives heathland fungal community trajectory over multiple years through altered plant–soil interactions. *New Phytologist*, *225*(5), 2140–2151.

Ramirez, K., Lauber, C., Knight, R., Bradford, M., & Fierer, N. (2010). Consistent effects of nitrogen fertilization on soil bacterial communities in contrasting systems. *Ecology*, *91*(12), 3463–3470.

Reeuwijk, L. (2002). *Technical Paper 09: Procedures for Soil Analysis (6th Edition)*. International Soil Reference and Information Centre,Wageningen, The Netherlands.

Reynolds, H. L., Packer, A., Bever, J. D., & Clay, K. (2003). Grassroots ecology: plant–microbe–soil interactions as drivers of plant community structure and dynamics. *Ecology*, *84*(9), 2281–2291.

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ*, *4*, e2584.

Saxena, A. K., Kumar, M., Chakdar, H., Anuroopa, N., & Bagyaraj, D. J. (2020). Bacillus species in soil as a natural resource for plant health and nutrition. *Journal of Applied Microbiology*, *128*(6), 1583–1594.

Shi, Y., Li, Y., Xiang, X., Sun, R., Yang, T., He, D., Zhang, K., Ni, Y., Zhu, Y. G., Adams, J. M., & Chu, H. (2018). Spatial scale affects the relative role of stochasticity versus determinism in soil bacterial communities in wheat fields across the North China Plain. *Microbiome*, *6*(1), 27.

Smith, D. P., & Peay, K. G. (2014). Sequence depth, not PCR replication, improves ecological inference from next generation DNA sequencing. *PLoS ONE*, *9*(2), e90234.

Steidinger, B. S., Crowther, T. W., Liang, J., Van Nuland, M. E., Werner, G. D., Reich, P. B., ... & Peay, K. G. (2019). Climatic controls of decomposition drive the global biogeography of forest-tree symbioses. *Nature*, *569*(7756), 404-408.

Tedersoo, L., Bahram, M., Cajthaml, T., Põlme, S., Hiiesalu, I., Anslan, S., Harend, H., Buegger, F., Pritsch, K., Koricheva, J., & Abarenkov, K. (2015). Tree diversity and species identity effects on soil fungi, protists and animals are context dependent. *The ISME Journal*, *10*(2), 346–362.

Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N. S., Wijesundera, R., ... & Abarenkov, K. (2014). Global diversity and geography of soil fungi. *science*, *346*(6213), 1256688.

van der Heijden, M. G., Bardgett, R. D., & van Straalen, N. M. (2008). The unseen majority: Soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecology Letters*, *11*(3), 269–310.

van der Putten, W. H., Bradford, M. A., Pernilla Brinkman, E., van de Voorde, T. F. J., & Veen, G. F. (2016). Where, when and how plant–soil feedback matters in a changing world. *Functional Ecology*, *30*(7), 1109–1121.

Ven, A., Verlinden, M. S., Verbruggen, E., & Vicca, S. (2019). Experimental evidence that phosphorus fertilization and arbuscular mycorrhizal symbiosis can reduce the carbon cost of phosphorus uptake. *Functional Ecology*, *33*(11), 2215–2225.

Verbruggen, E., van der Heijden, M. G. A., Rillig, M. C., & Kiers, E. T. (2013). Mycorrhizal fungal establishment in agricultural soils: Factors determining inoculation success. In *New Phytologist* (Vol. 197, Issue 4, pp. 1104–1109). John Wiley & Sons, Ltd.

Verbruggen, E., Xiang, D., Chen, B., Xu, T., & Rillig, M. C. (2015). Mycorrhizal fungi associated with high soil N:P ratios are more likely to be lost upon conversion from grasslands to arable agriculture. *Soil Biology and Biochemistry*, *86*, 1–4.

Větrovský, T., Kohout, P., Kopecký, M., Machac, A., Man, M., Bahnmann, B. D., ... & Baldrian, P. (2019). A meta-analysis of global fungal distribution reveals climate-driven patterns. *Nature communications*, *10*(1), 5142.

Vicca, S., Stocker, B. D., Reed, S., Wieder, W. R., Bahn, M., Fay, P. A., ... & Ciais, P. (2018). Using research networks to create the comprehensive datasets needed to assess nutrient availability as a key determinant of terrestrial carbon cycling. *Environmental Research Letters*, *13*(12), 125006.

Wubs, E. R. J., Putten, W. H., Mortimer, S. R., Korthals, G. W., Duyts, H., Wagenaar, R., & Bezemer, T. M. (2019). Single introductions of soil biota and plants generate long-term legacies in soil and plant community assembly. *Ecology Letters*, *22*(7), ele.13271.

Yu, N., Liu, J., Ren, B., Zhao, B., Liu, P., Gao, Z., & Zhang, J. (2022). Long-term integrated soil-crop management improves soil microbial community structure to reduce GHG emission and increase yield. *Frontiers in Microbiology*, *13*, 4035.

Zheng, Q., Hu, Y., Zhang, S., Noll, L., Böckle, T., Dietrich, M., Herbold, C. W., Eichorst, S. A., Woebken, D., Richter, A., & Wanek, W. (2019). Soil multifunctionality is affected by the soil environment and by microbial community composition and diversity. *Soil Biology and Biochemistry*, *136*, 107521.

Data accessibility: The data and the R script describing the steps of the analyses are deposited in the Dryad repository doi:10.5061/dryad.0k6djhb4z (if not yet active, the temporary link to the datasets is https://datadryad.org/stash/share/GW0t4-Cyhpf8FtTiB28oGrFek2n3qmVlSoOHCAT-CHY). Raw bacterial and fungal sequences are deposited in NCBI SRA database with the following accession numbers PRJNA966933 and PRJNA970213, respectively.

Benefit sharing: A research collaboration was developed with scientists from the countries providing soil samples. All collaborators are included as co-authors and the results of the research have been shared with them.

Author contributions: DR and EV conceived the study. DR performed lab work with the help of JDG and DR performed data analysis and interpretation together with EV. DR wrote the first draft of the paper with the help of EV and SV. Other co-authors performed soil sampling and provided information about sites and plant communities. All co-authors contributed significantly to the final version of the manuscript.

Figure legends

Figure 1 The location of 18 HerbDivNet sites in relation to global precipitation values. Red diamonds indicate 11 sites that contained a clear productivity gradient and yellow circles indicate other sites (containing from 2 to 6 plots but with no clear productivity gradient). All plots (n = 83) were used in the analyses of large-scale predictors of microbial community composition while 11 sites containing a productivity gradient (11 pairs of plots with relatively low and high productivity; a total of 44 plots) were used in the analyses of microbial community composition at high and low productivity levels.

Figure 2 Variance partitioning between selected variables in the large-scale model explaining a) bacterial and b) fungal community composition. The variance was partitioned between three groups of variables: i) broad-scale predictors – climate, N deposition and geographical distances; ii) regionally-varying environmental predictors – soil properties and biomass and iii) plant-community composition. The sizes of the bubbles correspond to the percentage of variance explained by each group (indicated by the numbers in the bubbles).

Figure 3 Relationships between regional (within-site) environmental/plant community distances and bacterial and fungal community distances **a)** bacterial distances vs environmental distances; **b**) bacterial distances vs plant distances **c**) fungal distances vs environmental distances; **d**) fungal distances vs plant distances for 12 sites that contained more than three plots. Environmental distances are created from a distance matrix of scaled soil variables that were identified as important predictors in the large-scale model and multiplied by the regression coefficient in that model. Colours of points and corresponding regression lines correspond to 12 different sites located in Argentina, Austria, Canada, Germany (2 sites), Hungary, Iran, Italy, Kenya, South Africa and USA. The relationships between regional geographical distances and bacterial/fungal distances per site are shown in Fig. S6. For complete site references, see Table S2.

Figure 4 Partial MDS ordination showing **a)** bacterial and fungal Bray-Curtis distances based on the relative abundances of OTUs (partialling out the effect of the site differences), coloured according to the productivity level of the sampling plots. b) Boxplots showing bacterial and fungal abundance (number of gene copies) at high and low productivity levels. Significant differences based on the mixed-effect model with 'site' as a random intercept are shown with * $P < 0.05$ and ** $P < 0.01$.

Figure 5 a) The coefficients and standard deviations of the mixed effect models (site as a random effect) examining the difference between the relative abundance of bacterial families at low and high productivity (left) and the overall abundance of these families at each productivity level (right). The 20 most abundant known bacterial families were examined. Correction of P values for multiple testing was

done using BH; * P < 0.05, **.** P < 0.1. b) Indicator bacterial taxa (orders and families of indicator OTUs) and their log-transformed mean abundances (along with standard errors) for high and low productivity levels. Only 50 most abundant indicator OTUs with known families were shown.

Figure 6 a) The coefficients and standard deviations of the mixed effect models (site as a random effect) examining the difference between the relative abundance of fungal families at low and high productivity (left) and the overall abundance of these families at each productivity level (right). 20 most abundant known fungal families were examined. Correction of P values for multiple testing was done using BH; **.** P < 0.1. b) Indicator fungal taxa (orders and families of indicator OTUs) and their log-transformed mean abundances (along with standard errors) for high and low productivity levels.

Figure 1.

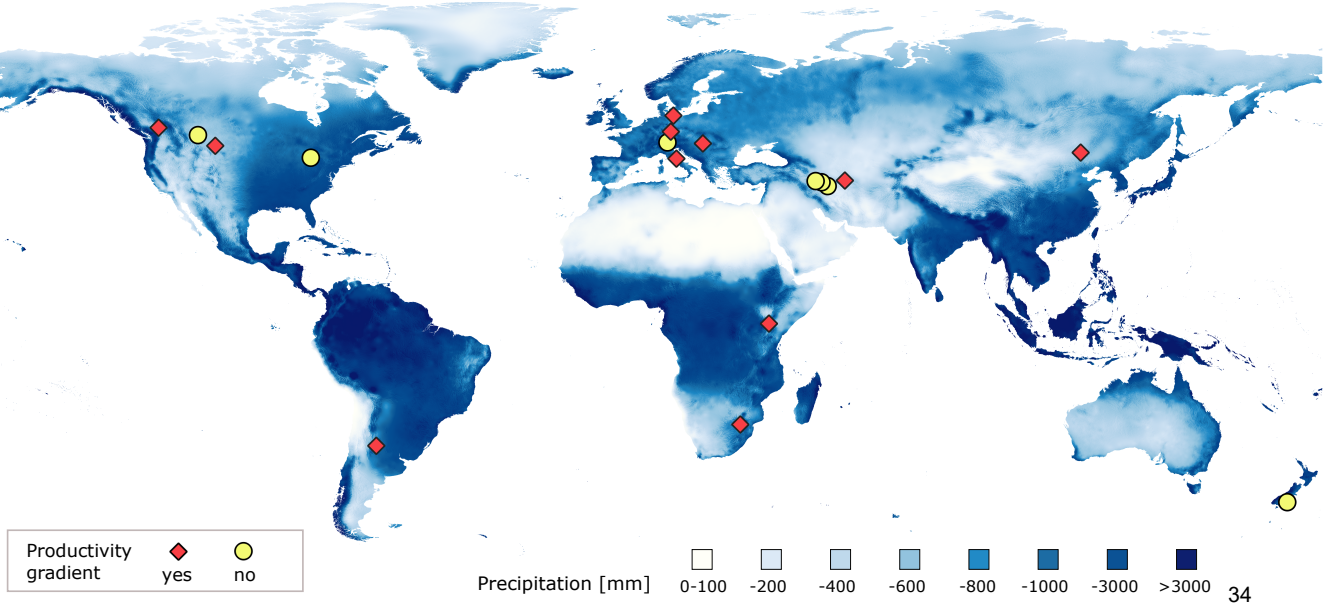


Figure 1

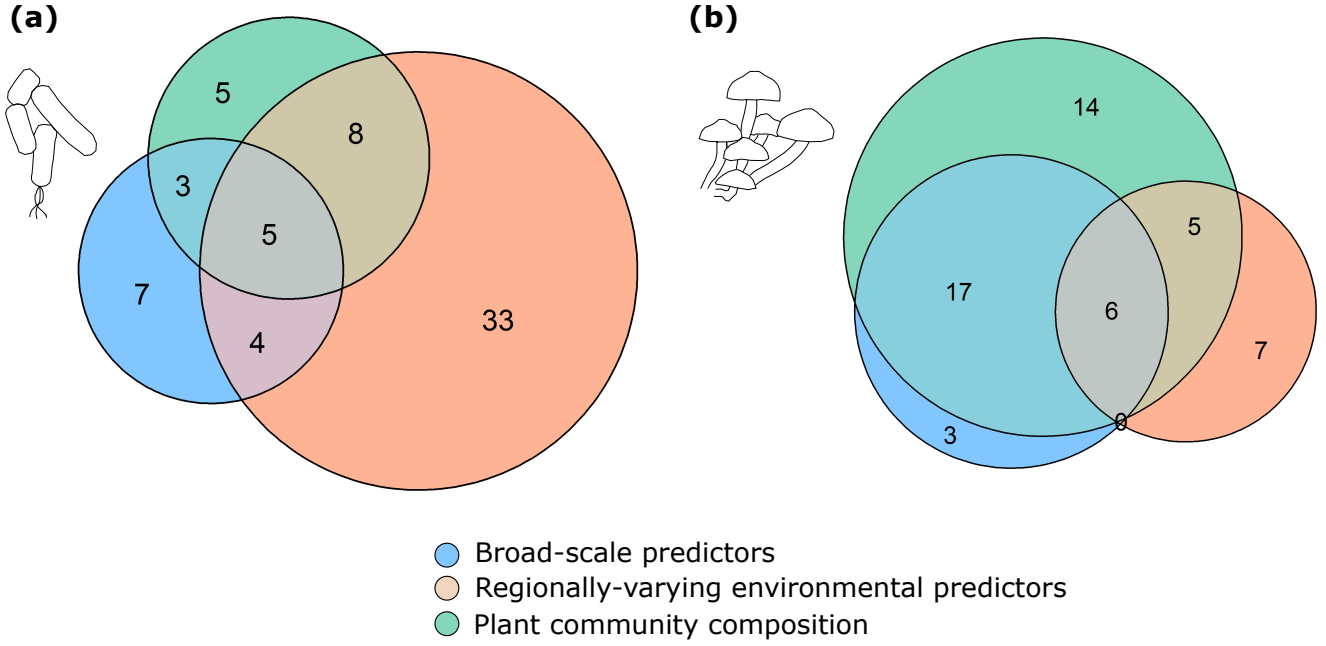

(a)
5
8
3
5
7
33
4
(b)
14
5
17
6
7
3
0
Broad-scale predictors
Regionally-varying environmental predictors
Plant community composition

Figure 2

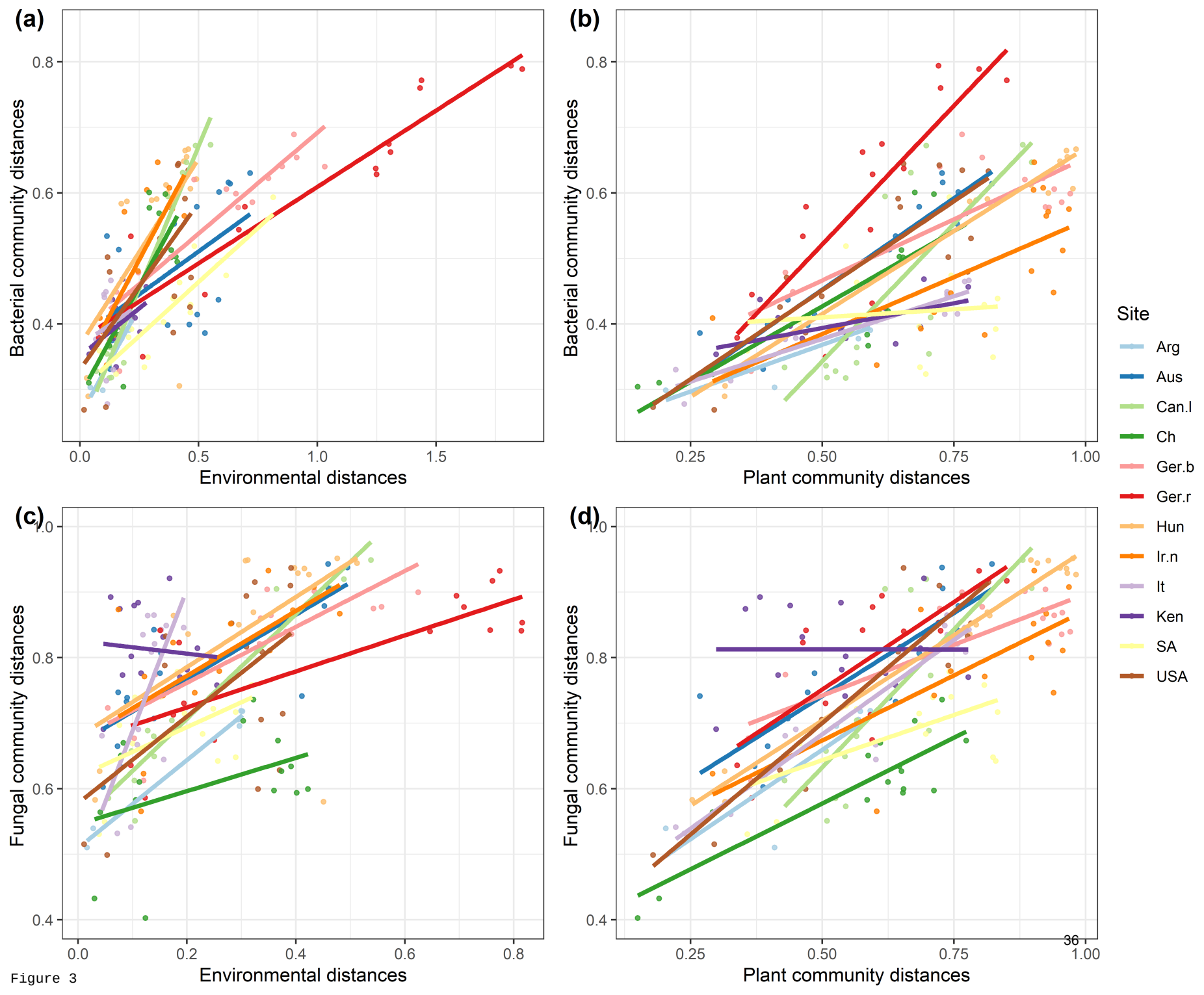

(a)
(b)
(c)
(d)
Bacterial community distances
Fungal community distances
Environmental distances
Plant community distances
0.0
0.5
1.0
1.5
0.25
0.50
0.75
1.00
0.2
0.4
0.6
0.8
Site
Arg
Aus
Can.I
Ch
Ger.b
Ger.r
Hun
Ir.n
It
Ken
SA
USA

Figure 3

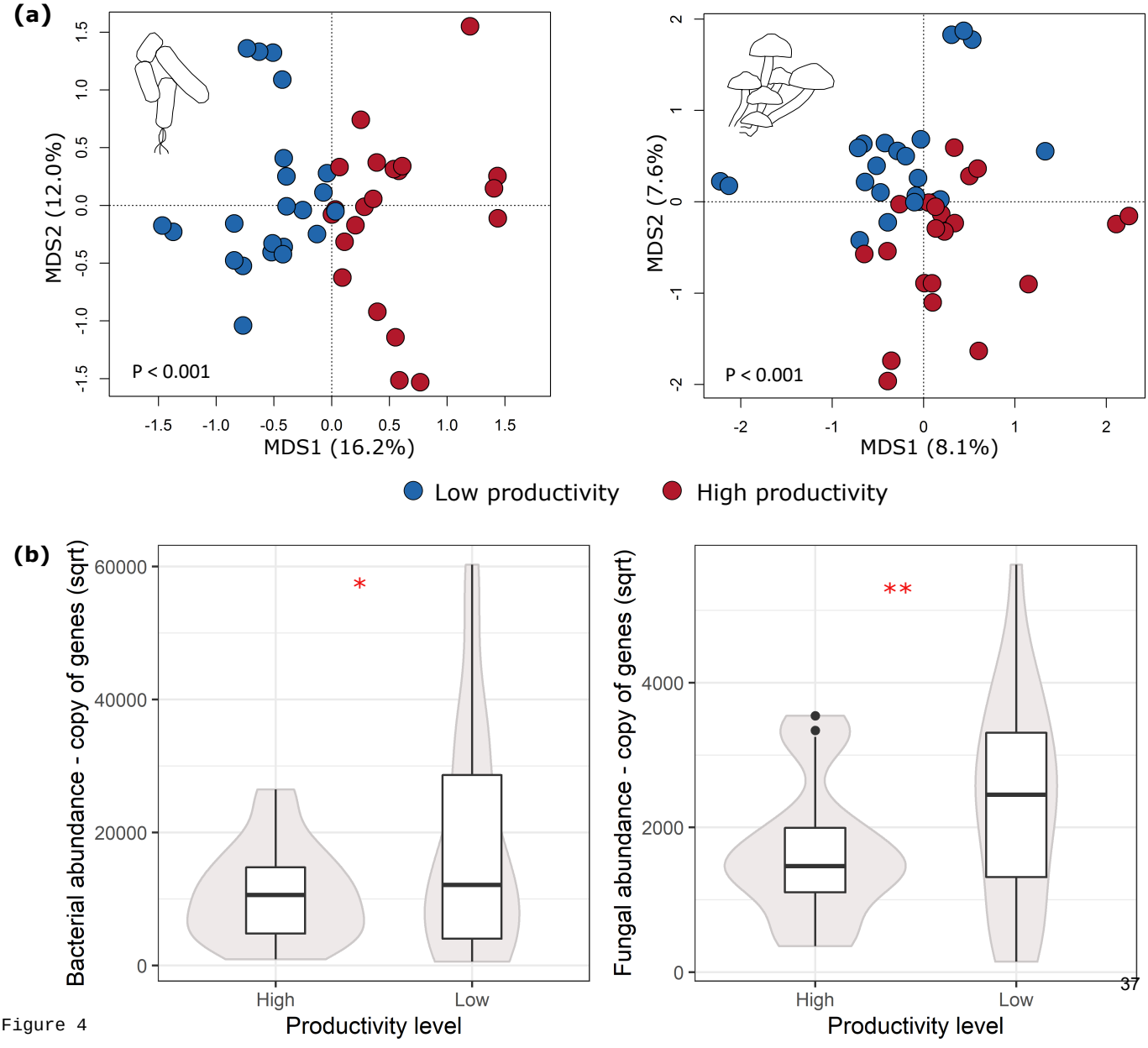
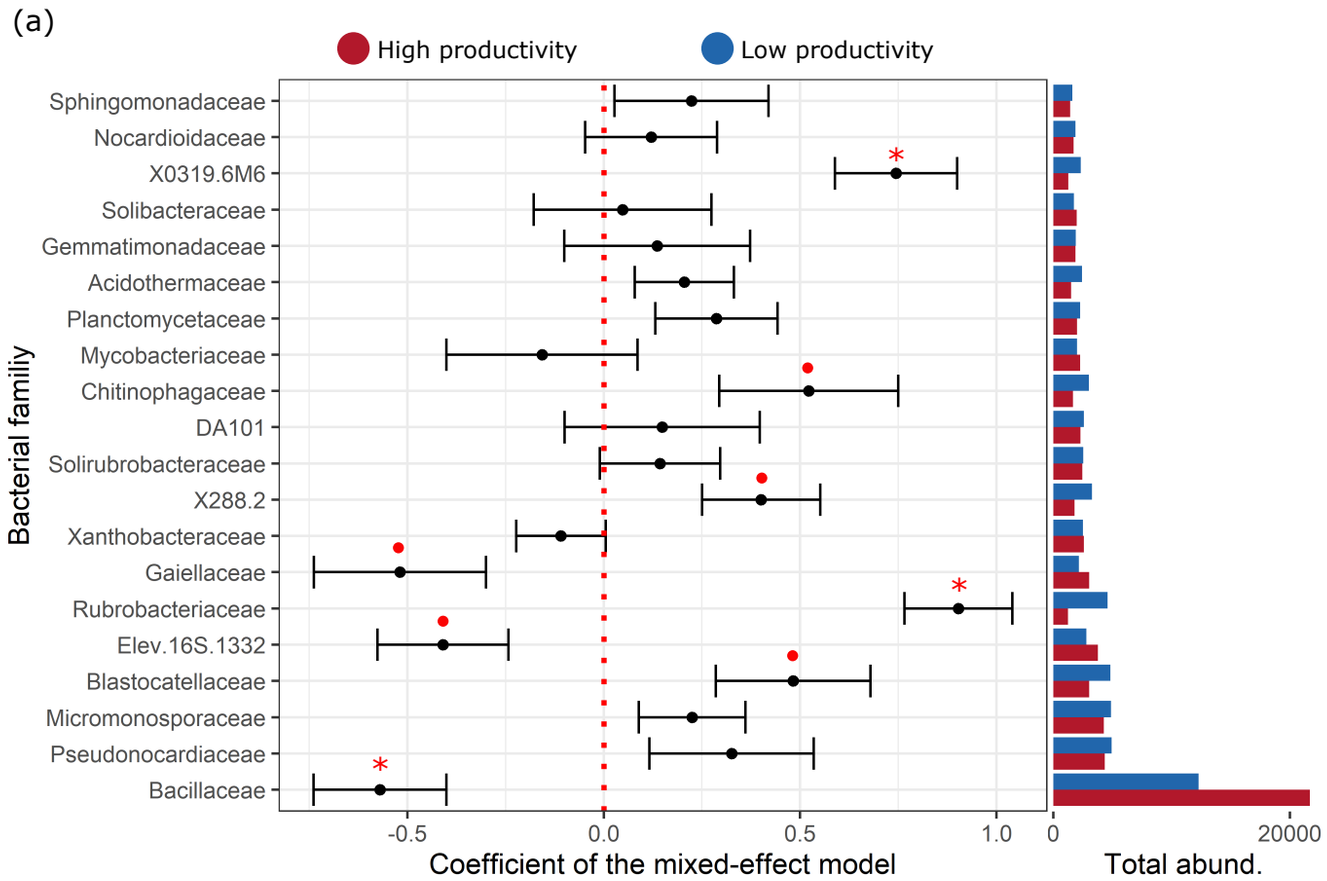

(a)
MDS2 (12.0%)
MDS1 (16.2%)
P < 0.001
MDS2 (7.6%)
MDS1 (8.1%)
P < 0.001
Low productivity
High productivity
(b)
Bacterial abundance - copy of genes (sqrt)
*
High
Low
Productivity level
Fungal abundance - copy of genes (sqrt)
**
High
Low
Productivity level

Figure 4


(a)
High productivity
Low productivity
Sphingomonadaceae
Nocardioidaceae
X0319.6M6
Solibacteraceae
Gemmatimonadaceae
Acidothermaceae
Planctomycetaceae
Mycobacteriaceae
Chitinophagaceae
DA101
Solirubrobacteraceae
X288.2
Xanthobacteraceae
Gaiellaceae
Rubrobacteriaceae
Elev.16S.1332
Blastocatellaceae
Micromonosporaceae
Pseudonocardiaceae
Bacillaceae
Bacterial familiy
-0.5
0.0
0.5
1.0
Coefficient of the mixed-effect model
0
20000
Total abund.

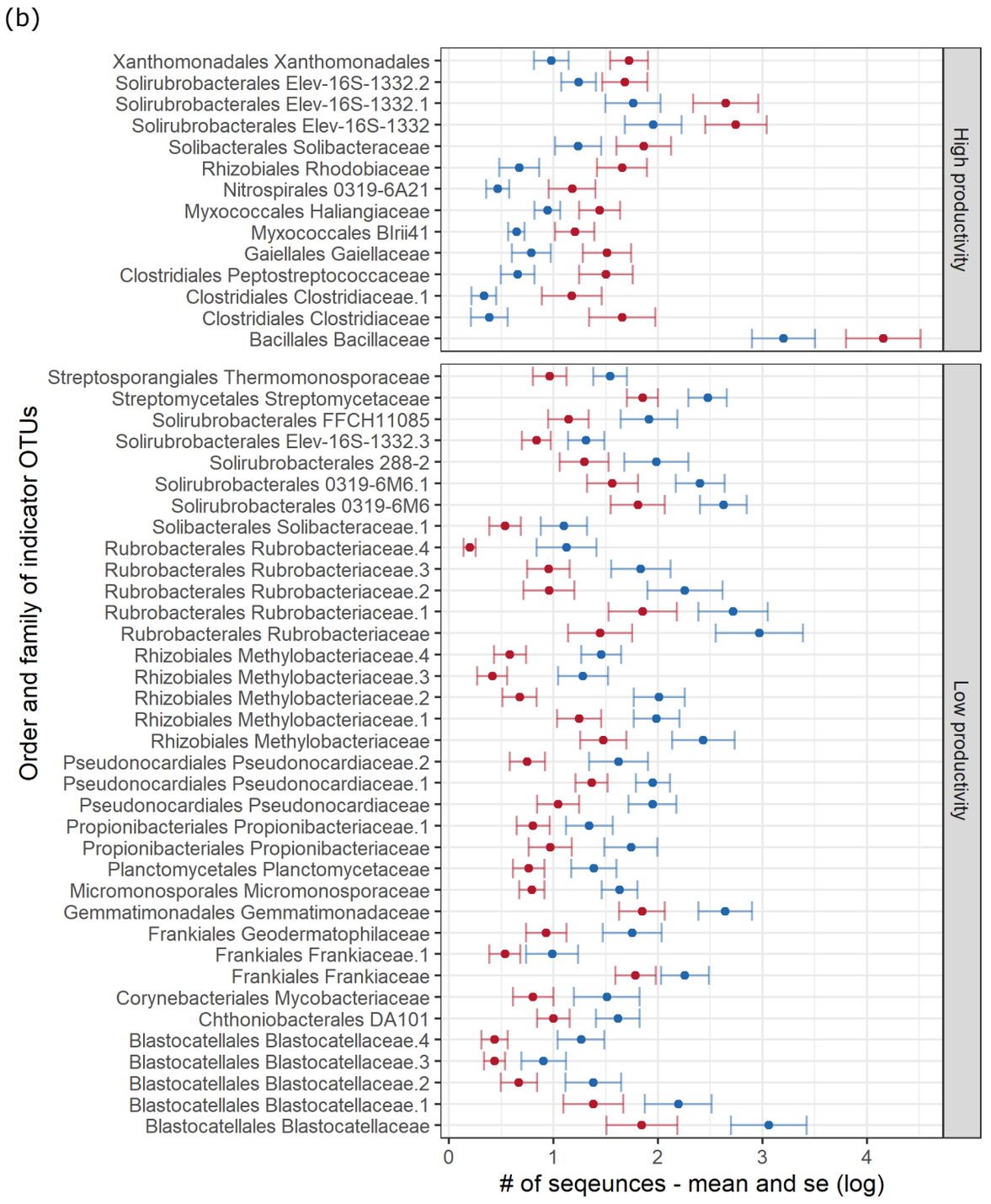

(b)
Xanthomonadales Xanthomonadales
Solirubrobacterales Elev-16S-1332.2
Solirubrobacterales Elev-16S-1332.1
Solirubrobacterales Elev-16S-1332
Solibacterales Solibacteraceae
Rhizobiales Rhodobiaceae
Nitrospirales 0319-6A21
Myxococcales Haliangiaceae
Myxococcales Blrii41
Gaiellales Gaiellaceae
Clostridiales Peptostreptococcaceae
Clostridiales Clostridiaceae.1
Clostridiales Clostridiaceae
Bacillales Bacillaceae
High productivity
Streptosporangiales Thermomonosporaceae
Streptomycetales Streptomycetaceae
Solirubrobacterales FFCH11085
Solirubrobacterales Elev-16S-1332.3
Solirubrobacterales 288-2
Solirubrobacterales 0319-6M6.1
Solirubrobacterales 0319-6M6
Solibacterales Solibacteraceae.1
Rubrobacterales Rubrobacteriaceae.4
Rubrobacterales Rubrobacteriaceae.3
Rubrobacterales Rubrobacteriaceae.2
Rubrobacterales Rubrobacteriaceae.1
Rubrobacterales Rubrobacteriaceae
Rhizobiales Methylobacteriaceae.4
Rhizobiales Methylobacteriaceae.3
Rhizobiales Methylobacteriaceae.2
Rhizobiales Methylobacteriaceae.1
Rhizobiales Methylobacteriaceae
Pseudonocardiales Pseudonocardiaceae.2
Pseudonocardiales Pseudonocardiaceae.1
Pseudonocardiales Pseudonocardiaceae
Propionibacteriales Propionibacteriaceae.1
Propionibacteriales Propionibacteriaceae
Planctomycetales Planctomycetaceae
Micromonosporales Micromonosporaceae
Gemmatimonadales Gemmatimonadaceae
Frankiales Geodermatophilaceae
Frankiales Frankiaceae.1
Frankiales Frankiaceae
Corynebacteriales Mycobacteriaceae
Chthoniobacterales DA101
Blastocatellales Blastocatellaceae.4
Blastocatellales Blastocatellaceae.3
Blastocatellales Blastocatellaceae.2
Blastocatellales Blastocatellaceae.1
Blastocatellales Blastocatellaceae
Order and family of indicator OTUs
Low productivity
0
1
2
3
4
# of seqeunces - mean and se (log)

Figure 5

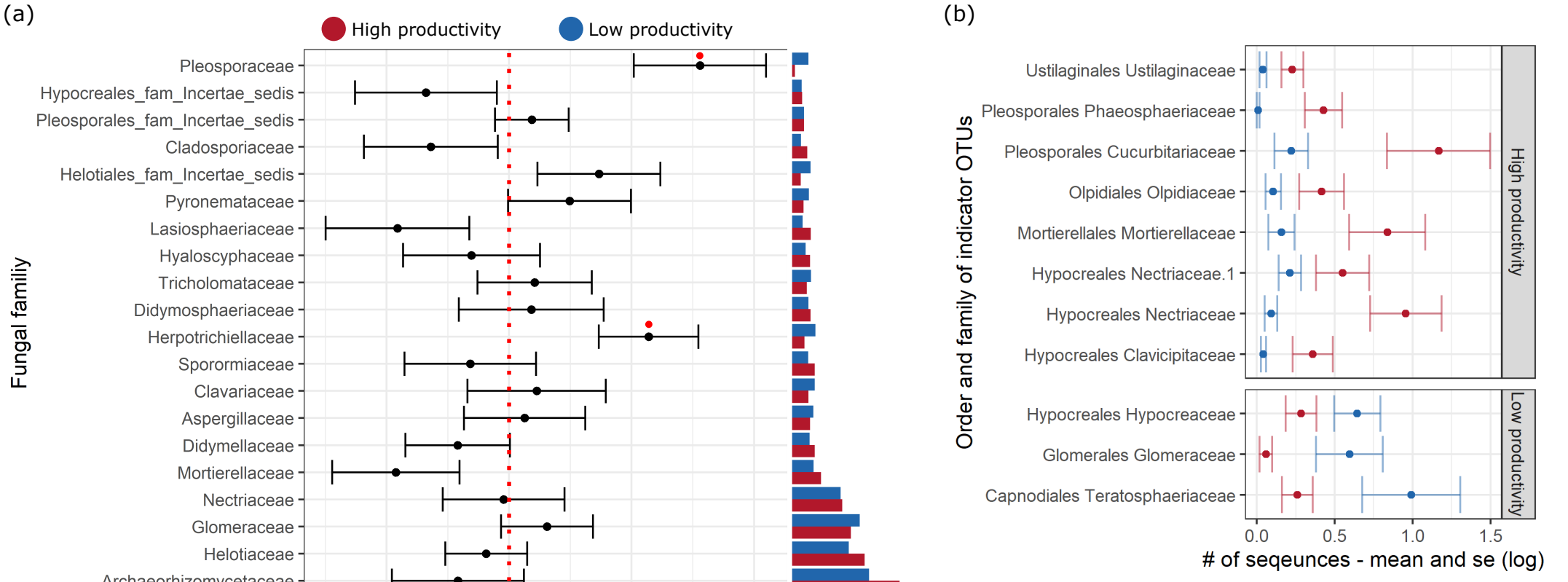

(a)
High productivity
Low productivity
Fungal familiy
Pleosporaceae
Hypocreales_fam_Incertae_sedis
Pleosporales_fam_Incertae_sedis
Cladosporiaceae
Helotiales_fam_Incertae_sedis
Pyronemataceae
Lasiosphaeriaceae
Hyaloscyphaceae
Tricholomataceae
Didymosphaeriaceae
Herpotrichiellaceae
Sporormiaceae
Clavariaceae
Aspergillaceae
Didymellaceae
Mortierellaceae
Nectriaceae
Glomeraceae
Helotiaceae
Archaeorhizomycetaceae
-0.5
0.0
0.5
1.0
Coefficient of the mixed-effect model
0
3000
Total abund.
(b)
Order and family of indicator OTUs
Ustilaginales Ustilaginaceae
Pleosporales Phaeosphaeriaceae
Pleosporales Cucurbitariaceae
Olpidiales Olpidiaceae
Mortierellales Mortierellaceae
Hypocreales Nectriaceae.1
Hypocreales Nectriaceae
Hypocreales Clavicipitaceae
High productivity
Hypocreales Hypocreaceae
Glomerales Glomeraceae
Capnodiales Teratosphaeriaceae
Low productivity
0.0
0.5
1.0
1.5
# of seqeunces - mean and se (log)

Figure 6