



## Structure and function of bacterial metaproteomes across biomes

F. Bastida<sup>a,\*</sup>, N. Jehmlich<sup>b,1</sup>, R. Starke<sup>c</sup>, K. Schallert<sup>d</sup>, D. Benndorf<sup>e</sup>, R. López-Mondéjar<sup>c</sup>, C. Plaza<sup>f</sup>, Z. Freixino<sup>g</sup>, C. Ramírez-Ortuno<sup>a</sup>, A. Ruiz-Navarro<sup>a</sup>, M. Díaz-López<sup>a</sup>, A. Vera<sup>a</sup>, J. L. Moreno<sup>a</sup>, D.J. Eldridge<sup>h</sup>, C. García<sup>a</sup>, M. Delgado-Baquerizo<sup>i</sup>

<sup>a</sup> CEBAS-CSIC, Department of Soil and Water Conservation, Campus Universitario de Espinardo, 30100, Murcia, Spain

<sup>b</sup> Helmholtz-Centre for Environmental Research - UFZ, Department of Molecular Systems Biology, Permoserstr. 15, 04318, Leipzig, Germany

<sup>c</sup> Laboratory of Environmental Microbiology, Institute of Microbiology of the Czech Academy of Sciences, Vídeňská 1083, 14220, Praha 4, Czech Republic

<sup>d</sup> Otto von Guericke University, Bioprocess Engineering, Universitätsplatz 2, 39106, Magdeburg, Germany

<sup>e</sup> Max Planck Institute for Dynamics of Complex Technical Systems, Bioprocess Engineering, Sandtorstraße 1, 39106, Magdeburg, Germany

<sup>f</sup> Instituto de Ciencias Agrarias, Consejo Superior de Investigaciones Científicas, Serrano 115 Bis, 28006, Madrid, Spain

<sup>g</sup> Department of Ecology and Hydrology, University of Murcia, Campus de Espinardo, 30100, Murcia, Spain

<sup>h</sup> School of Biological, Earth and Environmental Sciences, University of New South Wales, Sydney, New South Wales, Australia

<sup>i</sup> Departamento de Sistemas Físicos, Químicos y Naturales, Universidad Pablo de Olavide, Sevilla, Spain

### ARTICLE INFO

#### Keywords:

Bacterial communities  
Biodiversity  
Metaproteomics  
Functionality  
Climate  
Vegetation

### ABSTRACT

Soil microbes, and the proteins they produce, are responsible for a myriad of soil processes which are integral to life on Earth, supporting soil fertility, nutrient fluxes, trace gas emissions, and plant production. However, how and why the composition of soil microbial proteins (the metaproteome) changes across wide gradients of vegetation, climatic and edaphic conditions remains largely undetermined. By applying high-resolution mass spectrometry to soil samples collected from four continents, we identified the most common proteins in soils, and investigated the primary environmental factors driving their distributions across climate and vegetation types. We found that soil proteins involved in carbohydrate metabolism, DNA repair, lipid metabolism, transcription regulation, tricarboxylic acid cycling, nitrogen (N) fixation and one-carbon metabolism dominate soils across a wide range of climates, vegetation types and edaphic conditions. Vegetation type and climate were important factors determining the community composition of the topsoil metaproteome. Moreover, we show that vegetation type, climate, and key edaphic properties (mainly soil C fractions, pH and texture) influenced the proportion of important proteins involved in biogeochemical cycles and cellular processes. We also found that protein-based taxonomic information based on proteins has a greater resolution than 16S rRNA gene sequencing with regards to the ability to detect significant correlations with environmental variables. Together, our work identifies the dominant proteins produced by microbes living in a wide range of soils, and advances our understanding of how environmental changes can influence the structure and function of the topsoil metaproteome and the soil processes that they support.

### 1. Introduction

Topsoil proteins are the ultimate catalyzers of multitude of biological functions (Hettich et al., 2013; Starke et al., 2019) which allow microbial communities to drive fundamental ecosystem services, including soil fertility, climate change regulation, pollutants degradation and waste decomposition (Bardgett and van der Putten, 2014; Crowther et al., 2019; Delgado-Baquerizo et al., 2020). Previous studies have

investigated the cross-biome drivers of taxonomic and functional diversity across contrasting vegetation and climatic types using DNA amplicon sequencing and metagenomics (Bahram et al., 2018; Delgado-Baquerizo et al., 2018; Fierer et al., 2012). However, DNA-based methods are limited by their reduced capacity to account for active microbial communities and processes, restricting the capacity of metagenomics to efficiently predict ecosystem functions (Carini et al., 2016).

Proteomics has been proposed as a suitable approach for assessing

\* Corresponding author.

E-mail address: [fbastida@cebas.csic.es](mailto:fbastida@cebas.csic.es) (F. Bastida).

<sup>1</sup> Authors contributed equally.

the functionality of soil microbial communities through identifying the real catalyzers of soil processes -the proteins-at the local scale (Bastida et al., 2016; Hultman et al., 2015; Liu et al., 2019; Starke et al., 2017). Yet, we know very little about how and why the structure and function of the topsoil metaproteome changes across contrasting vegetation types and climates which are globally distributed. Advancing our knowledge on the structure and function of the topsoil proteome is integral to potentially predict shifts in important microbial-driven process in a changing planet. Here, we conducted a survey of soils collected from 60 sites that span broad gradients in environmental context (edaphic conditions, climate, and vegetation types) (Fig. S1; Table S1) with the topsoil metaproteomes and the relative abundance of individual protein categories characterized using high resolution mass spectrometry. For instance, mean annual precipitation and temperature ranged from 81 to 2161 mm and  $-2.8$ – $21.0$  °C, respectively. Soil pH and carbon content ranged from 3.7 to 9.1 and 0.09–37.81%, respectively. We focused on bacterial proteins because these organisms are one of the most abundant and diverse on Earth (Bardgett and van der Putten, 2014; Delgado-Baquerizo et al., 2018).

Recent evidence suggests that microbial activity is particularly responsive to changes in plant communities and associated shifts in soil organic carbon (C) availability (Bastida et al., 2016; Lladó et al., 2019; Žifčáková et al., 2017). Here, we test the hypothesis that the metaproteome of bacterial communities differ across vegetation and climate types. This hypothesis is based on the assumption that the same factors that are known to be important in structuring bacterial communities, including soil pH, and the quantity and quality of soil organic C (Fierer, 2017; Bahram et al., 2018), also vary across vegetation and climate types (Jobbágy and Jackson, 2000). Because of this, we hypothesized that vegetation structure and contrasting C fractions (e.g., light vs. mineral-associated carbon) should be important drivers of the soil metaproteome. As a whole, our work aims to identify the most common protein groups in soils and advance our knowledge of the major environmental drivers regulating the distribution of the soil metaproteome across broad gradients of vegetation, climatic, and edaphic conditions.

## 2. Material and methods

### 2.1. Soil sampling

Soil and vegetation data were collected between 2016 and 2017 from 60 locations distributed across climate and vegetation types (SI Appendix, Fig. S1). These locations include a wide range of globally distributed soil, vegetation (including forest, prairies and shrublands) and climate based on aridity index (mesic and drylands) types. Sampling was designed to obtain wide gradients of edaphic characteristics (Bastida et al., 2019; Delgado-Baquerizo et al., 2019) (Table S1). For example, precipitation and temperature ranged from 81 to 2160 mm per year and  $-2.80$ – $21.0$  °C, respectively. A standardized protocol was used for soil survey (Maestre et al., 2012). In each location, we surveyed a 50 m  $\times$  50 m plot. Three parallel transects of the same length, spaced 25 m apart were added. The cover of perennial vegetation was measured in each transect using the line-intercept method (Maestre et al., 2012). Plant cover ranged between 6.7 and 100%. One composite topsoil (five 0–10 cm soil cores) sample was collected under the dominant ecosystem features across our plots. Following field sampling, soils were sieved ( $<2$  mm) and frozen at  $-20$  °C. Climatic information was extracted from WorldClim (Hijmans et al., 2005).

### 2.2. Soil chemical, physical and microbial analyses

For all soil samples, we measured electrical conductivity, pH, texture, soil organic C (soil C) content and available P (Olsen P) content. Soil properties were determined using standardized protocols (Maestre et al., 2012). Soil pH was measured in all the soil samples with a pH meter, in a 1: 2.5 mass: volume soil and water suspension. Soil texture

(% of fine fractions: clay + silt) was determined according to Kettler et al. (2001). Soil organic C was determined colorimetrically after oxidation with  $K_2Cr_2O_7$  and  $H_2SO_4$  at 150 °C for 30 min (Anderson and Ingram, 1993). Total N was obtained using a CN analyzer (LECO CHN628 Series, LECO Corporation, St Joseph, MI USA). The content of Olsen P was analyzed following the method of Olsen and Sommers (1983) (Olsen and Sommers, 1983). Soil C content ranged between 0.1 and 38%, pH between 3.8 and 9.1, and percentage clay + silt varied between 0.3 and 62%, respectively. A soil sample representative of each site was fractionated to isolate soil organic C pools characterized by different mechanisms of protection from decomposition and stability. We used the density-based fractionation scheme developed by Golchin et al. (1994) (Golchin et al., 1994) with slight modifications (Plaza et al., 2019; Sohi et al., 2001) to isolate two different soil organic C fractions: a free organic C fraction (FR\_OC), located between soil aggregates and accessible to decomposers; and a mineral-associated organic C fraction, protected from microbial decomposition by sorption to mineral surfaces (MA\_OC). This light fraction (FR\_OC) separated here by density has been suggested to be similar to the particulate organic matter (POM) fraction separated by size (Lavalley et al., 2020; Mikutta et al., 2019). The isolated fractions were oven-dried at 60 °C, weighed, ground with a ball mill and analyzed for organic C concentration. The organic C concentration was determined by dry combustion using an elemental analyzer; before the analysis, the MA\_OC fraction was fumigated with HCl to remove carbonates (Harris et al., 2001).

Bacterial biomass was quantified through phospholipid fatty acids (PLFAs) extracted from 0.5 g freeze-dried subsample, by using the method described in Bligh and Dyer (1959) (Bligh and Dyer, 1959) and modified by Buyer et al. (2012) (Buyer and Sasser, 2012). The extracted fatty acid methyl esters were analyzed on an Agilent Technologies 7890B gas chromatograph with an Agilent DB-5 ms column (Agilent Technologies, CA, USA). The fatty acids selected to represent bacterial biomass are the PLFAs i15:0, a15:0, 15:0, i16:0, 16:1 $\omega$ 7, 17:0, i17:0, a17:0, cy17:0, 18:1 $\omega$ 7 and cy19:0, and the fatty acid representative of fungal biomass is the 18:2 $\omega$ 6 (Frostegård and Bååth, 1996; Rinnan and Bååth, 2009).

The composition of soil bacterial communities was analyzed through amplicon sequencing using the Illumina MiSeq platform. Ten grams of frozen soil samples (from composite soil samples as explained above) were cooled using liquid nitrogen and ground using a mortar and pestle. Soil DNA was extracted using the Powersoil® DNA Isolation Kit (MoBio Laboratories, Carlsbad, CA, USA). A portion of the bacterial 16S rRNA gene was sequenced using the 515F/806R primer set (Lauber et al., 2009; Ramirez et al., 2014). Bioinformatic analyses were carried out with QIIME (Caporaso et al., 2010), USEARCH (Edgar, 2010) and UNOISE3 (Edgar, 2016). Phylotypes (i.e. ASVs) were identified at the 100% identity level. A detailed description of DNA approaches can be found in Delgado-Baquerizo et al. (2019) (Delgado-Baquerizo et al., 2019).

### 2.3. Protein extraction from soil and mass spectrometry analysis

Protein extraction was performed according to the method described elsewhere (Bastida et al., 2014; Chourey et al., 2010). The cell lysis and disruption of soil aggregates were performed by boiling at 100 °C for 10 min in sodium dodecyl sulfate (SDS) buffer. The proteins were separated by 12% SDS-PAGE and, after electrophoresis, the gels were stained using colloidal Coomassie brilliant blue. The gel area containing the protein mixture of each sample was sliced into one piece. The samples were further processed by in-gel reduction and alkylation of cysteine residues, in-gel tryptic cleavage, and elution as well as desalting of tryptic peptides (Bastida et al., 2016). The peptide lysates were reconstituted in 0.1% formic acid prior to LC-MS measurement. Separation of peptide lysates was performed using an 85-min, non-linear gradient from 3.2% to 80% acetonitrile, in 0.1% formic acid, on a C18 analytical column (Acclaim PepMap100, 75  $\mu$ m inner diameter, 25 cm, C18, Thermo

Scientific) in a UHPLC system (Ultimate nanoRSLC 3000, Thermo Fisher Scientific, Idstein, Germany). Mass spectrometry was performed on a Q Exactive HF mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA) coupled with a TriVersa NanoMate (Advion, Ltd., Harlow, UK) source in LC chip coupling mode. The mass spectrometer full scans were measured in the Orbitrap mass analyzer within the mass range of 350–1550  $m/z$ , at 60,000 resolution, using an automatic gain control target of  $1 \times 10^6$  and a maximum fill time of 100 ms. An MS/MS isolation window for ions in the quadrupole was set to 1.4  $m/z$ . The MS/MS scans were acquired using the higher energy dissociation mode at a normalized collision-induced energy of 28%, within a scan range of 200–2000  $m/z$  and using a resolution of 15,000. The exclusion time to reject masses from repetitive MS/MS fragmentation was set to 30 s. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al., 2018) partner repository with the dataset identifier PXD018448.

For data analysis, the raw files from the mass spectrometer were converted into peaklists in the mascot generic file format (MGF) using MSConverter. Protein database searches and subsequent data functional and phylogenetic analyses were carried out using the Meta-ProteomeAnalyzer (Heyer et al., 2019) (MPA, version 2.14) using the OMSSA and X!Tandem search engines. Search parameters were set to 10 ppm peptide ion tolerance and 0.1 Da fragment ion tolerance, carbamidomethyl as fixed and methionine oxidation as variable modification, using UniProtKB/SwissProt as protein database (roughly 556,000 proteins, 11/2017). The database searches were performed in target-decoy mode and a false discovery rate of 10% was used. This high rate was used considering that this study is focused in broad protein categories and the wide variety and heterogeneity of soils. Proteins were grouped if they were sharing at least one peptide across all samples. A comparison matrix was created listing non-redundant spectrum count per protein group per sample. The comparison matrix was exported and used for subsequent data analysis. From each soil sample, we obtained the relative abundance of different protein groups, as well as their taxonomic origin at phylum and order levels (Supporting Information, Dataset).

## 2.4. Statistical analyses

We used permutational analysis of variance (PERMANOVA) to investigate the significant differences in the relative abundance of protein groups across climate (dryland vs mesic) and vegetation (forests, prairies and shrublands) types. In these analyses, each plot is considered a statistical replicate. Put simply, in our study we are using Earth as a grid across which we are collecting data from different plots or sites (replicates) from different ecosystem types. Having more than one sample within each plot would have been considered pseudo-replication as our question was related to comparing the abundance of protein groups across different vegetation and climate types globally rather than comparing them across plots within a given ecosystem type. Further, gradient designs, as we have used here, are powerful tools for detecting patterns in ecological responses to continuous and interacting environmental drivers as they generally outperform replicated designs in terms of prediction success of responses (Kreyling et al., 2018). Spearman correlation and Robust Linear Models (RLM) analyses were carried out to investigate the relationship between the relative abundance of protein groups or microbial populations, and environmental variables. We used Spearman correlations, as they provide useful information on the direction and strength of the associations between two variables, do not require normality of data, and linearity is not a strict assumption of these analyses. Non-metric multidimensional scaling (NMDS) was used for study the structure of bacterial metaproteomes across vegetation and climate types.

## 3. Results and discussion

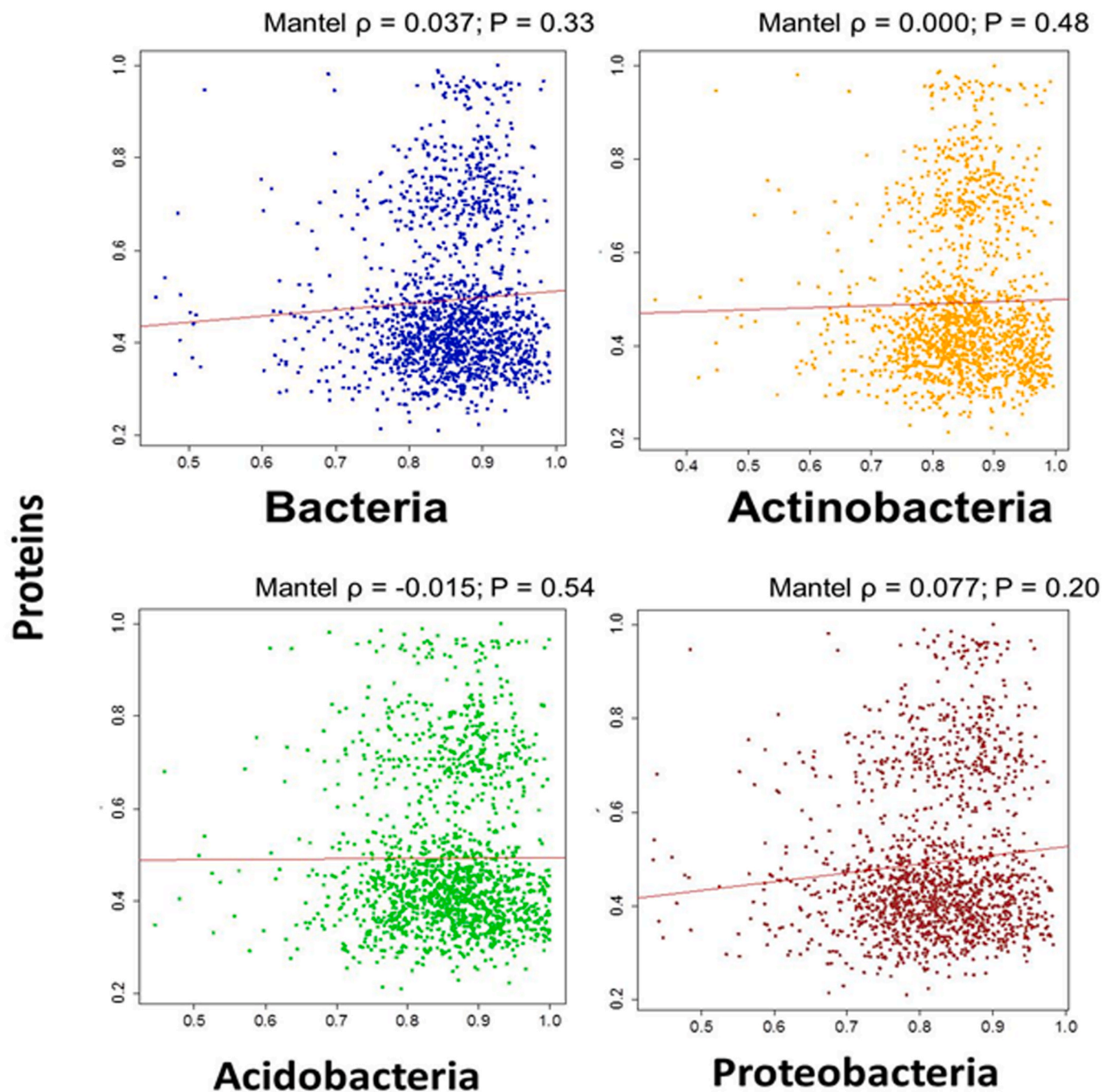
We quantified the proportion (relative abundance; %), based on their

spectral abundances, of a total of 5286 proteins that were assigned to 193 different functional groups (dataset, Supporting Information). Among them, the most abundant proteins were those involved in protein biosynthesis, transcription, amino acid biosynthesis, stress response, transcription regulation, glycolysis, tricarboxylic acid cycle, ion transport, DNA repair, lipid metabolism, N fixation, antibiotic resistance, one-carbon (C1) metabolism, cell cycle and cell division, lipid biosynthesis, carbohydrate metabolism, fatty acid metabolism, and DNA replication (Fig. S2). Some continental and global metagenome soil surveys have also reported a high abundance of genes involved in protein metabolism, amino acid and carbohydrates transport and metabolism, replication, DNA metabolism, energy production, transcription, cell division, etc. (Fierer et al., 2012; Bahram et al., 2018). Our study reflects the dominance of these protein groups (Fig. S2), but also the high proportion of proteins involved in nitrogen-fixation, C1-metabolism and antibiotic resistance. These results may indicate that proteins involved in rhizosphere colonization, where C1 metabolism (Knief et al., 2012) and microbial competition (i.e. antibiotic resistance mechanisms; Hou and Kolodkin-Gal, 2020) are critical traits, are among the most important proteins in a wide variety of soils across biomes.

Our results indicate that metaproteomic analyses complement, and provide a unique perspective on soil bacterial communities compared with DNA-based analyses of taxonomic composition (proportion of amplicon sequence variants – ASVs- or phyla determined from 16S rRNA gene amplicon sequencing; relative abundance, %). In particular, Mantel tests provided evidence that there is no significant correlation between the matrix of distance (Bray-Curtis) based on the taxonomic composition of bacterial communities (proportion of ASVs) and that based on the community composition of the soil metaproteome (proportion of 193 groups of proteins). This absence of correlation was also confirmed when we repeated Mantel analyses based on the proportion of taxa with the most abundant bacteria phyla (Fig. 1 and Fig. S2). These analyses suggest that our results are consistent across contrasting phylogenies and unlikely to be biased toward the preferential annotation of proteins from any given dominant phylum (Fig. 1). This lack of correlation between 16S rRNA gene amplicon analyses of DNA and metaproteomics could derive from the large abundance of inactive DNA in soil (Carini et al., 2016) while some proteins can remain active through their stabilization in organic matter and clay (Burns et al., 2013), from different methodological biases or from different turnover times of the biomolecules (Starke et al., 2019). Another explanation is that protein abundances may simply be poorly correlated with taxon abundance, either because of the large variation in protein production or turnover rates across different taxa, or because similar proteins may be produced by very different taxa. Together, these findings suggest that metaproteomics provide complementary information to that of DNA-based marker gene analyses when evaluating the links between environmental factors and soil microbial communities. Our results are in agreement with earlier studies showing that DNA-based surveys likely miss considerable portions of active microbial populations when compared to RNA approaches (Baldrian et al., 2012), and with local studies which have highlighted distinct microbial community composition when comparing DNA- and protein-based approaches (Bastida et al., 2017; Thorn et al., 2019).

Overall, proteins produced by Proteobacteria, and Alphaproteobacteria in particular, dominated the metaproteome profiles in terms of their relative abundances, followed by Actinobacteria and Firmicutes (Fig. 2A). Other relatively abundant phyla represented in the topsoil metaproteome included Bacteroidetes, Tenericutes, Cyanobacteria, Acidobacteria, Chlorobi and Spirochaetes (Fig. 2A). These results provide a complementary picture of the dominant phyla in soils across the globe, compared with observations of bacterial community composition inferred from 16S rRNA gene amplicon sequencing. For example, in terms of protein abundance, the relative abundance of Proteobacteria (ranging between 60 and 70% of the identified proteins) was greater than that observed from sequencing-based taxonomic approaches in our



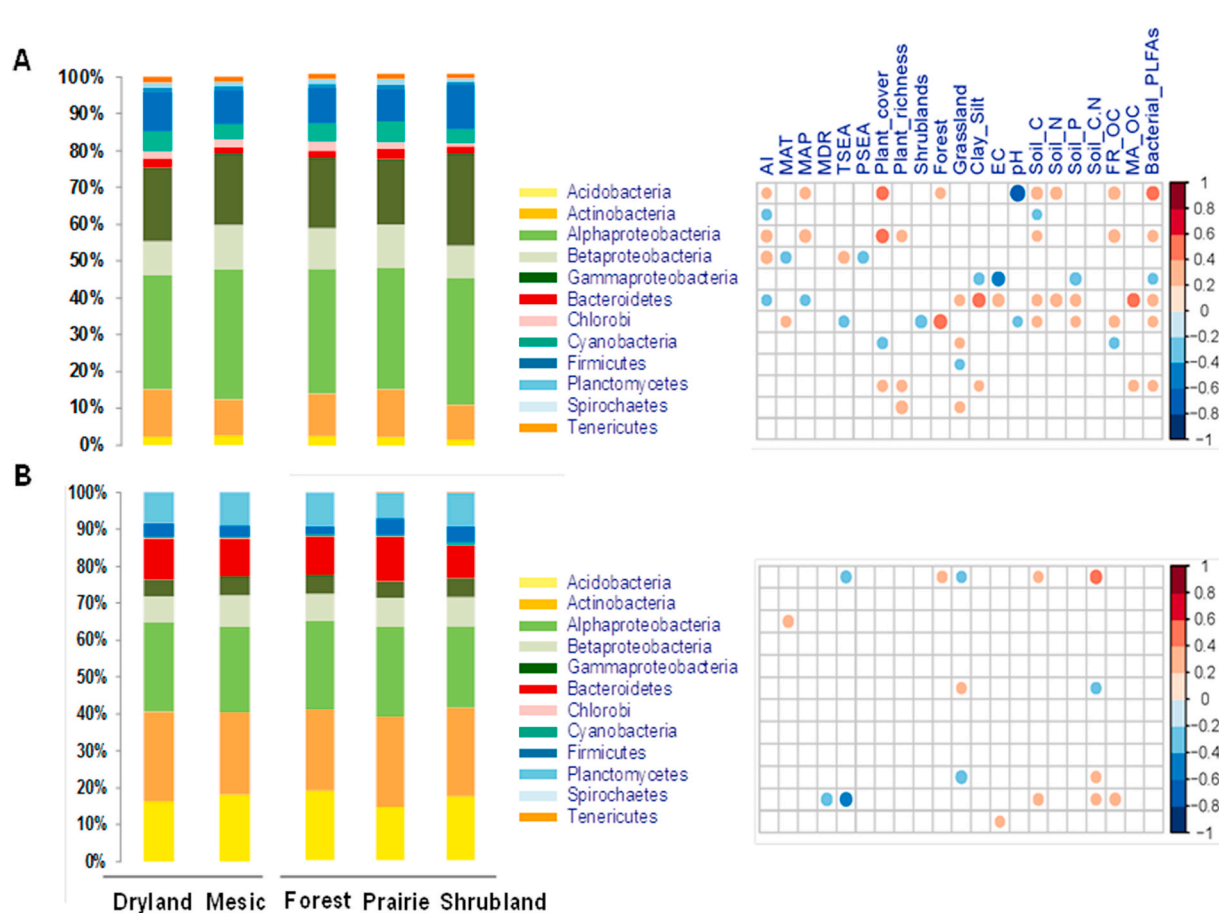


**Fig. 1. Relationship between the topsoil microbiome and metaproteome across locations.** Correlations between community composition (via 16S rRNA gene sequencing) and functional community composition (via metaproteomics). X axes represents community composition of bacteria (Bray-Curtis dissimilarity) and Y axes represents community composition of soil proteins (Bray-Curtis dissimilarity).

soils (30–40%, Fig. 2B), and in other large-scale genomic soil surveys (Bahram et al., 2018; Delgado-Baquerizo et al., 2018). Interestingly, the greater proportion of Proteobacteria proteins in comparison to the abundance of this group using 16S rRNA gene sequencing was also noticed in other metaproteome studies using different protein extraction methods and downstream bioinformatics pipelines (Thorn et al., 2019), thus indicating that the contribution of this phylum to the overall soil metaproteome is consistent among studies. This difference between genomic and proteomic approaches could be a product of the reference protein databases being biased towards well-characterized taxa (such as Proteobacteria) in contrast to other taxa, such as Acidobacteria, that are underrepresented in the reference protein databases. Consequently,

further investigations are needed to provide a more comprehensive picture of the taxonomy of bacterial communities based on soil metaproteomes, in particular for uncharacterized taxa. For this purpose, it would be important to improve the annotation of proteins that are likely originated from poorly characterized taxa.

We also found that protein-based taxonomic information (proportion of phyla/classes; relative abundance, %) based on proteins has a greater resolution than 16S rRNA gene sequencing with regards to the ability to detect significant correlations with environmental variables (Fig. 2A and B). The number of significant correlations between the information obtained from the proportion of protein-derived from phyla (proportion of phylum from protein data) and environmental variables were higher



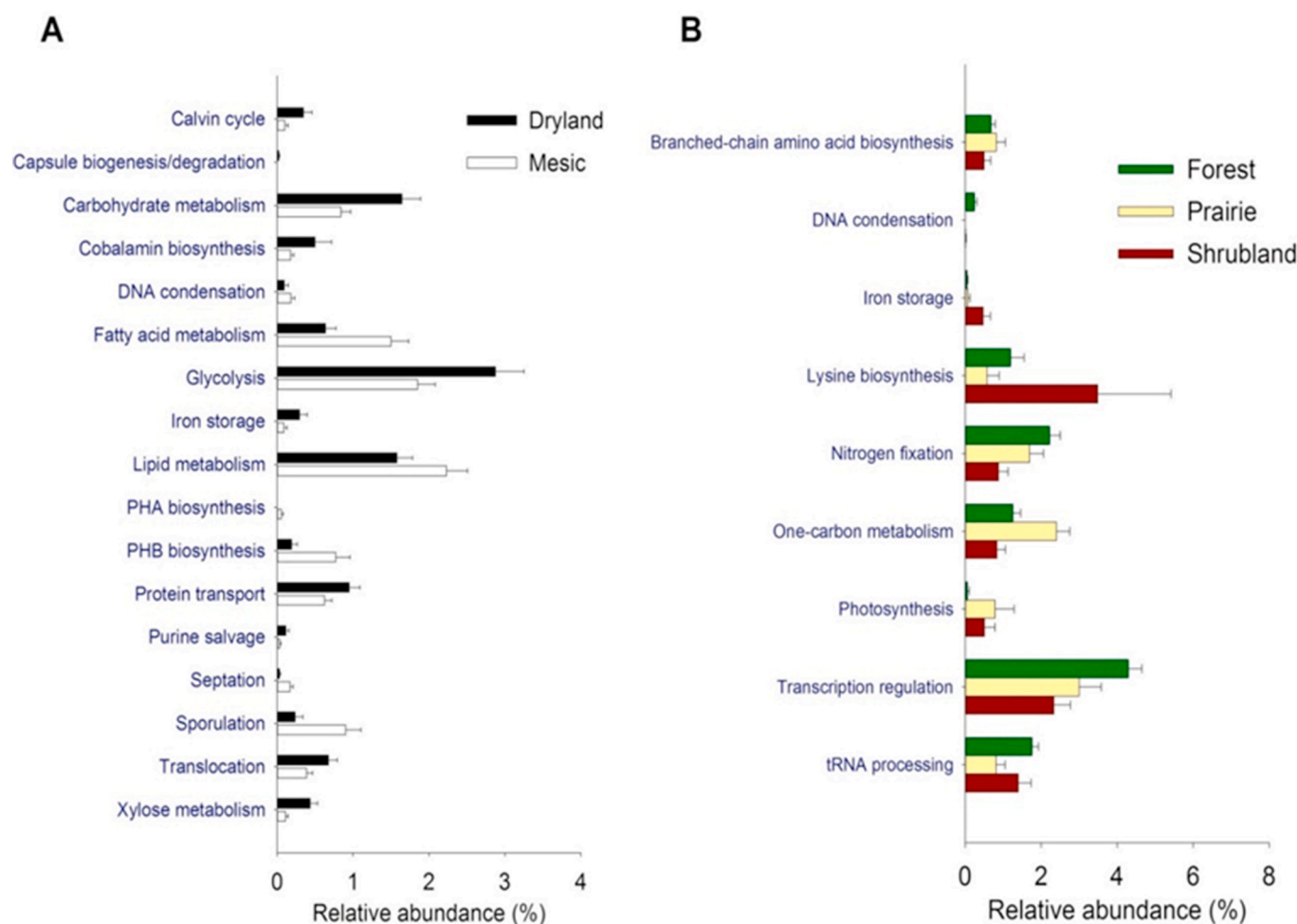
**Fig. 2. Taxonomic information based on the topsoil metaproteome.** Relative abundances of microbial taxa through proteomics (A) and 16S rRNA gene sequencing (B) across climate and vegetation types. The right part of each panel includes Spearman correlations ( $P < 0.05$ ) between microbial abundances and environmental variables. Only significant correlations are included ( $P < 0.05$ ) and color scale represents the correlation coefficient. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

than those found for taxonomic information from amplicon sequencing at a comparable phylum level (proportion of phylum from amplicon sequencing data) (Fig. 2). For example, we found that the proportion of Alphaproteobacteria proteins (including Rhizobiales, Fig. S3) correlated positively with plant cover, mean annual precipitation, soil C content and the light mineral-free C fraction (FR-OC; the fraction of organic matter not associated with minerals and thus more likely to be accessible to decomposers), while these correlations were not apparent from comparable analyses of alphaproteobacterial proportions as determined from 16S rRNA gene sequencing (Fig. 2; Fig. S4). Similarly, previous studies have revealed a high relative abundance of proteins from Rhizobiales in soils with greater organic matter availability (Bastida et al., 2015; Starke et al., 2016). Likewise, the relative abundance of cyanobacterial proteins correlated negatively with plant cover and accessible C fraction (FR-OC), while this correlation was not evident from the 16S rRNA gene analyses. Cyanobacteria have been shown to be proportionally important in soils with low carbon content (Kuske et al., 2002; Belnap and Lange, 2003). Moreover, in agreement with DNA-based studies highlighting an important role of texture on the composition of the soil microbial community (Kallenbach et al., 2016), our metaproteome study also captured some significant effects of soil physical parameters (i.e. clay plus silt content) on the proportion of microbial proteins from Bacteroidetes, Planctomycetes and Gammaproteobacteria (Fig. 2; Fig. S4).

We then conducted two-way PERMANOVA to investigate the role of climate (dryland and mesic) and vegetation (forest, prairies and shrublands) in driving the distribution of the metaproteome composition

(proportion of 193 proteins) across biomes. We found that both vegetation ( $P = 0.002$ ) and climate ( $P = 0.034$ ) influenced the community composition of the soil metaproteome, and further identified an interaction between both factors (vegetation  $\times$  climate interaction  $P = 0.016$ ), suggesting that the effects of climate on soil metaproteomes might be vegetation-type dependent (Fig. S5). These results suggest that shifts in vegetation, for example land use, encroachment, or deforestation associated with desertification (Berdugo et al., 2020) and climate change might have noticeable impacts on the overall community composition of the topsoil bacterial metaproteome (Bastida et al., 2018). Some of the effects of vegetation on the topsoil metaproteome might be indirectly driven by soil edaphic conditions. For example, vegetation types could have important effects on the soil metaproteome by regulating the amount of inputs and the quality of the organic matter to the soil system (Jobbágy and Jackson, 2000; Quideau et al., 2001). Indeed, soil C content and the most accessible C fraction (FR-OC) were lower in shrublands than in prairies and forests (Table S1), and the composition of the soil metaproteome of shrublands differed from that found in prairies and forests, particularly those in mesic climates (Fig. S5).

We next sought to determine those specific groups of soil proteins that were significantly associated with contrasting climates or vegetation types (Fig. 3). For this purpose, we performed a two-way PERMANOVA, and selected those protein groups which were significantly influenced ( $P < 0.05$ ) by either climate or vegetation type, but lacked significant interaction between both factors ( $P > 0.05$ ) (Fig. 3; Table S2). By doing so, we are free to independently interpret the effects of climate and vegetation on the soil metaproteome while ruling out complex



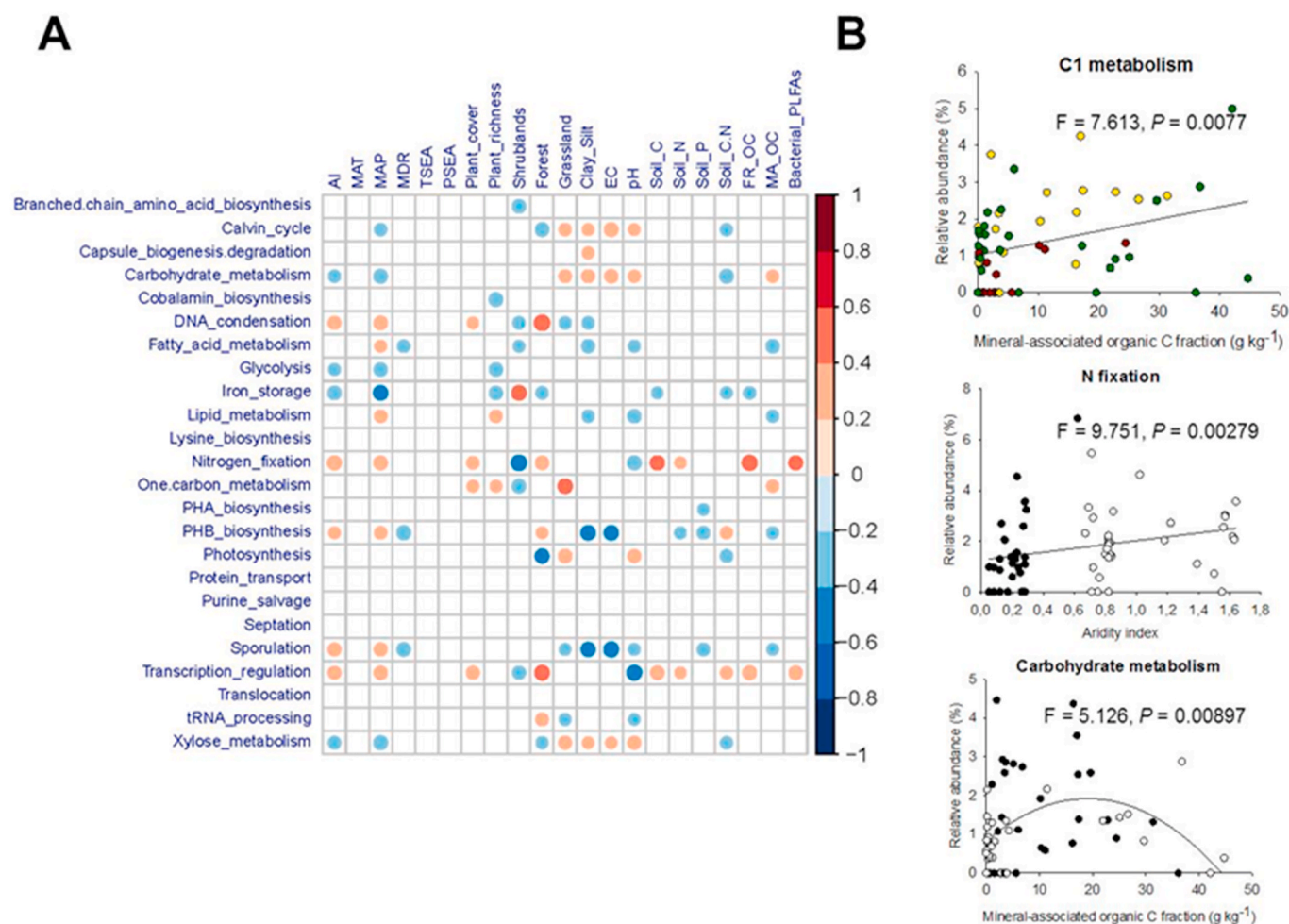
**Fig. 3. Key functional proteins across climates and vegetation types.** Relative abundance (%) (mean  $\pm$  SE) of functional groups of proteins across climate (A) and vegetation (B) types. For each factor, only biological functions with significant differences (PERMANOVA,  $P < 0.05$ ) and lack of interactions (vegetation  $\times$  climate  $P > 0.05$ ) are shown.

interactions between climate and vegetation types. We conducted further Spearman correlation analyses to identify the main climatic and soil properties associated with these protein groups (Fig. 4). Our results provide solid evidence that the proportion of important groups of proteins involved in metabolism and/or biosynthesis of carbohydrates, fatty acids, lipids, amino acids, and transcription regulation significantly differ across vegetation types and/or climate (Fig. 3). Many of these protein groups have been found in high relative abundances in continental-scale metagenomic surveys (Bahram et al., 2018; Fierer et al., 2012; Noronha et al., 2017). Further, our topsoil metaproteome survey allowed us to investigate important groups of proteins associated with particular biogeochemical processes. For example, the proportion of proteins involved in C1 metabolism (i.e. formaldehyde-activating enzyme, methanol dehydrogenase, methylamine metabolism) was influenced by vegetation, but not by climate type, and was more abundant in prairies than in forests or shrublands worldwide (Fig. 3). The metabolism of plant cell wall compounds, such as lignin, can produce methanol (Galbally and Kristine, 2002), which might be metabolized as source of carbon and energy by some microbes producing these enzymes (Sy et al., 2005). Further, C1 metabolism has been suggested to be important during bacterial colonization of phyllosphere, plant rests and rhizosphere (Knief et al., 2012). Our results also indicate that the proportion of C1 metabolism proteins correlated significantly and positively with the fraction of organic C associated with minerals, which is more protected from microbial decomposition (Lavalley et al., 2020), but did not correlate with total soil C content (Fig. 4).

Furthermore, the proportion of proteins associated with carbohydrate metabolism, including xylose metabolism and glycolysis, was higher in drylands than in more mesic ecosystems (Fig. 3). The relative abundance of this protein group correlated positively with the fraction of mineral-associated C and in fine-textured soils with high pH, and negatively with soil C/N ratio. However, there was no correlation with soil C content (Fig. 4). These results suggest that the quality—rather than quantity— and microbial availability of C which is regulated by organo-mineral interactions (Lavalley et al., 2020; Six et al., 2002), may be important factors shaping the relative abundance of proteins involved in carbohydrate metabolism. Under such conditions of intermediate or low C availability, such as those usually found in drylands (Maestre et al., 2012) (Table S1), a higher proportion of carbohydrate metabolising proteins might help microbes to access this source of mineral-protected C (MA-OC) (Fig. 4). In contrast, we found that the relative abundance of fatty acids and lipid metabolism proteins was greater in more mesic soils (Fig. 3). Further, our metaproteomic analyses also suggest that the microbial communities from more mesic soils with greater organic C accessibility tend to have higher relative abundance of proteins associated with the biosynthesis of polyhydroxybutyrate (PHB) (Figs. 3–4), a C storage polymer which is used as an energetic reserve by some soil microbes (Wang et al., 2006).

Environmental context also had an important role in regulating the relative abundance of N-fixation proteins across contrasting soil conditions. Our study revealed that vegetation type is an important factor shaping the proportion of this fundamental group of proteins (Fig. 3).





**Fig. 4. Environmental drivers of the topsoil metaproteome.** Spearman correlations between functional groups of proteins and environmental variables. Only significant correlations are included ( $P < 0.05$ ) and color scale represents the correlation coefficient (A). Relationships between environmental variables and the relative abundances of different protein groups (B). Circle color: green, forests; yellow, prairies; brown, shrublands; black, dryland; white, mesic. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

The proportion of proteins involved in N-fixation was higher in forest soils, followed by prairies, with shrubland soils having the lower relative abundance of N-fixation proteins. These results are in agreement with the greater estimated biological N fixation in forests than in grasslands (Yu and Zhuang, 2020). The relative abundance of this protein group correlated positively with high organic C accessibility (FR-OC fraction), soil C content, and aridity index, and negatively with soil pH (Fig. 4). A national-scale metagenome study in the United Kingdom highlighted that soils with low pH and high organic matter content supported microbial communities with more abundant genes for N fixation (Malik et al., 2017). Moreover, our metaproteome survey shows some additional functional adaptations of microbial communities to particular environmental conditions. For example, the relative abundance of proteins involved in iron storage was higher in drylands than in mesic soils, and particularly in shrublands (Fig. 3). We found negative correlations between the relative abundance of proteins involved in iron storage and aridity index, and with edaphic variables such C/N ratio and FR-OC (Fig. 4). The greater relative abundance of proteins involved in iron storage in drylands can be seen as a bacterial adaptation against the low iron availability associated to high soil pH in these environments (Moreno-Jiménez et al., 2019).

Soil metaproteomics is a promising approach to study microbial functionality. However, it is still in its relative infancy, and such analyses are associated with some important caveats that need to be considered. First, the absolute amount (mass) of extracted proteins was

not quantified, and thus there may be differential extraction efficiencies across soils. Second, we acknowledge that higher protein identification rates would likely be obtained if we had paired the metaproteomic data with metagenomic information (Starke et al., 2019), and that general databases (as the one used here, UniProtKB) can be biased to dominant cultured and sequenced bacteria. Moreover, current metaproteome assessments do not properly identify soil extracellular proteins, including extracellular hydrolases which are crucial for element cycles and soil fertility. This can be due to the fact that current methods do not properly address the extraction and identification of extracellular protein likely due to their stabilization in organic matter and soil particles (Bastida et al., 2018; Starke et al., 2019). Conversely, the preferential extraction of intracellular and membrane proteins can favour the identification of active proteins but not relic ones which are released and stabilized in soil after cell death and lysis. Further, our metaproteomic approach might be selective for certain microbial groups, such Proteobacteria. This can bias the results towards functional traits of this phylum, including a high proportion of proteins involved in N-fixation by Rhizobiales and C1-metabolism. However, our results focus on broad protein categories, and despite these limitations, we were able to identify the most dominant proteins in soils with contrasting properties across different environmental contexts. Future work should, however, seek to improve this methodology in order to advance our understanding of both dominant and rare proteins in soils globally.

Together, our topsoil metaproteome study helps advancing our

knowledge on the environmental drivers of the structure and function of the topsoil metaproteome, and provides critical insights on the adaptation of soil bacterial communities to their environment across climate and vegetation types. Importantly, we highlight the absence of correlation between the community composition based on DNA approach (16S rRNA gene sequencing) and protein-based analyses, highlighting that both approaches provide different information on the soil microbiome. Although metaproteomics approaches are still evolving to solve technical and methodological limitations, our study provides an important step toward building a predictive understanding on the environmental factors controlling the topsoil metaproteomes. Vegetation structure, climate, and key soil properties played fundamental roles in shaping the distributions of proteins involved in biogeochemical cycles. This information is integral to manage and support the functioning of sustainable terrestrial ecosystems in a world subject to important environmental changes.

## Funding

Data deposition: the primary data used in this paper will be deposited in Figshare and provide as Supporting Information dataset. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE repository with the dataset identifier PXD018448.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

Authors thank the Spanish Ministry and FEDER funds for the project AGL2017-85755-R (AEI/FEDER, UE), the i-LINK+ 2018 (LINKA20069) from CSIC, and funds from “Fundación Séneca” from Murcia Province (19896/GERM/15). We would like to thank the researchers involved in the CLIMIFUN project for the help with soil sampling. This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 702057. M.D.B. is supported by a Ramón y Cajal grant from the Spanish Government (agreement no. RYC2018-025483-I). C.P. acknowledges support from the Spanish State Plan for Scientific and Technical Research and Innovation (2013–2016), award ref. AGL201675762-R (AEI/FEDER, UE). K.S. was supported by the German Federal Ministry of Education and Research (de.NBI network, (project MetaProtServ, grant no. de-NBI-039031L0103). R.S. was supported by the Czech Science Foundation (20-02022Y).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.soilbio.2021.108331>.

## References

- Anderson, J.M., Ingram, J.S.I., 1993. *Tropical Soil Biology and Fertility: a Handbook of Methods*. CAB International, Wallingford.
- Bahrian, M., Hildebrand, F., Forslund, S.K., Anderson, J.L., Soudzilovskaia, N.A., Bodegom, P.M., Bengtsson-Palme, J., Anslan, S., Coelho, L.P., Harend, H., Huerta-Cepas, J., Medema, M.H., Maltz, M.R., Munda, S., Olsson, P.A., Pent, M., Pölme, S., Sunagawa, S., Ryberg, M., Tedersoo, L., Bork, P., 2018. Structure and function of the global topsoil microbiome. *Nature* 560, 233–237. <https://doi.org/10.1038/s41586-018-0386-6>.
- Baldrian, P., Kolarik, M., Stursova, M., Kopecky, J., Valaskova, V., Vetrovsky, T., Zifcakova, L., Snajdr, J., Ridl, J., Vlcek, C., Voriskova, J., 2012. Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. *The ISME Journal* 6, 248–258. <https://doi.org/10.1038/ismej.2011.95>.
- Bardgett, R.D., van der Putten, W.H., 2014. Belowground biodiversity and ecosystem functioning. *Nature* 515, 505–511. <https://doi.org/10.1038/nature13855>.
- Bastida, F., Hernández, T., García, C., 2014. Metaproteomics of soils from semiarid environment: functional and phylogenetic information obtained with different protein extraction methods. *Journal of Proteomics* 101, 31–42. <https://doi.org/10.1016/j.jprot.2014.02.006>.
- Bastida, F., Selevsek, N., Torres, I.F., Hernández, T., García, C., 2015. Soil restoration with organic amendments: linking cellular functionality and ecosystem processes. *Scientific Reports* 5. <https://doi.org/10.1038/srep15550>, 15550.
- Bastida, F., Torres, I.F., Moreno, J.L., Baldrian, P., Ondono, S., Ruiz-Navarro, A., Hernandez, T., Richnow, H.H., Starke, R., García, C., Jehmlich, N., 2016. The active microbial diversity drives ecosystem multifunctionality and is physiologically related to carbon availability in Mediterranean semi-arid soils. *Molecular Ecology* 25, 4660–4673. <https://doi.org/10.1111/mec.13783>.
- Bastida, F., Torres, I.F., Andrés-Abellán, M., Baldrian, P., López-Mondéjar, R., Vetrovsky, T., Richnow, H.H., Starke, R., Ondono, S., García, C., López-Serrano, F.R., Jehmlich, N., 2017. Differential sensitivity of total and active soil microbial communities to drought and forest management. *Global Change Biology* 23, 4185–4203. <https://doi.org/10.1111/gcb.13790>.
- Bastida, F., Jehmlich, N., Torres, I.F., García, C., 2018. The extracellular metaproteome of soils under semiarid climate: a methodological comparison of extraction buffers. *The Science of the Total Environment* 619–620, 707–711. <https://doi.org/10.1016/j.scitotenv.2017.11.134>.
- Bastida, F., Bastidal, F., García, C., Fierer, N., Eldridge, D.J., Bowker, M.A., Abades, S., Alfaro, F.D., Berhe, A.A., Cutler, N.A., Gallardo, A., García-Velázquez, L., Hart, S.C., Hayese, P.E., Hernandez, T., Hseu, Z.-Y., Jehmlich, N., Kirchmair, M., Lambers, H., Neuhauser, S., Pena-Ramirez, V.M., Perez, C.A., Reed, S.C., Santos, F., Siebe, C., Sullivan, B.W., Trivedi, P., Vera, A., Williams, M.A., Moreno, J.L., Delgado-Baquerizo, M., 2019. Global ecological predictors of the soil priming effect. *Nature Communications* 10. <https://doi.org/10.1038/s41467-019-11472-7>.
- Benlap, J., Lange, O.L., 2003. *Biological Soil Crusts: Structure, Function and Management*. Springer, Berlin, p. 503.
- Berdugo, M., Delgado-Baquerizo, M., Soliveres, S., Hernández-Clemente, R., Zhao, Y., Gaitán, J.J., Gross, N., Saiz, H., Maire, V., Lehmann, A., Rillig, M.C., Solé, R.V., Maestre, F.T., 2020. Global ecosystem thresholds driven by aridity. *Science* 367, 787–790. <https://doi.org/10.1126/science.aay5958>.
- Bligh, E.G., Dyer, W.J., 1959. A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology* 37, 911–917. <https://doi.org/10.1139/o59-099>.
- Burns, R.G., DeForest, J.L., Marxsen, J., Sinsabaugh, R.L., Stromberger, M.E., Wallenstein, M.D., Weintraub, M.N., Zoppini, A., 2013. Soil enzymes in a changing environment: current knowledge and future directions. *Soil Biology and Biochemistry* 58, 216–234. <https://doi.org/10.1016/j.soilbio.2012.11.009>.
- Buyer, J.S., Sasser, M., 2012. High throughput phospholipid fatty acid analysis of soils. *Applied Soil Ecology* 61, 127–130. <https://doi.org/10.1016/j.apsoil.2012.06.005>.
- Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittiger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* 7, 335–336. <https://doi.org/10.1038/nmeth.f.303>.
- Carini, P., Marsden, P.J., Leff, J.W., Morgan, E.E., Strickland, M.S., Fierer, N., 2016. Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. *Nature Microbiology* 2. <https://doi.org/10.1038/nmicrbiol.2016.242>, 16242.
- Chourey, K., Jansson, J., VerBerkmoes, N., Shah, M., Chavarria, K.L., Tom, L.M., Brodie, E.L., Hettich, R.L., 2010. Direct cellular lysis/protein extraction protocol for soil metaproteomics. *Journal of Proteome Research* 9, 6615–6622. <https://doi.org/10.1021/pr100787q>.
- Crowther, T.W., van den Hoogen, J., Wan, J., Mayes, M.A., Keiser, A.D., Mo, L., Averill, C., Maynard, D.S., 2019. The global soil community and its influence on biogeochemistry. *Science* 365. <https://doi.org/10.1126/science.aav0550>, eaav0550.
- 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, 320–325. <https://doi.org/10.1126/science.aap9516>.
- Delgado-Baquerizo, M., Bardgett, R.D., Vitousek, P.M., Maestre, F.T., Williams, M.A., Eldridge, D.J., Lambers, H., Neuhauser, S., Gallardo, A., García-Velázquez, L., Sala, O.E., Abades, S.R., Alfaro, F.D., Berhe, A.A., Bowker, M.A., Currier, C.M., Cutler, N.A., Hart, S.C., Hayes, P.E., Hseu, Z.-Y., Kirchmair, M., Peña-Ramírez, V.M., Pérez, C.A., Reed, S.C., Santos, F., Siebe, C., Sullivan, B.W., Weber-Grullon, L., Fierer, N., 2019. Changes in belowground biodiversity during ecosystem development. *Proceedings of the National Academy of Sciences* 116, 6891–6896. <https://doi.org/10.1073/pnas.1818400116>.
- Delgado-Baquerizo, M., Reich, P.B., Trivedi, C., Eldridge, D.J., Abades, S., Alfaro, F.D., Bastida, F., Berhe, A.A., Cutler, N.A., Gallardo, A., García-Velázquez, L., Hart, S.C., Hayes, P.E., He, J.-Z., Hseu, Z.-Y., Hu, H.-W., Kirchmair, M., Neuhauser, S., Perez, C.A., Reed, S.C., Santos, F., Sullivan, B.W., Trivedi, P., Wang, J.-T., Weber-Grullon, L., Williams, M.A., Singh, B.K., 2020. Multiple elements of soil biodiversity drive ecosystem functions across biomes. *NATURE ECOLOGY & EVOLUTION* 4, 210–220. <https://doi.org/10.1038/s41559-019-1084-y>.
- Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>.
- Edgar, R.C., 2016. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *BioRxiv*. <https://doi.org/10.1101/081257>, 81257.



- Fierer, N., 2017. Embracing the unknown: disentangling the complexities of the soil microbiome. *Nature Reviews Microbiology* 15, 579–590. <https://doi.org/10.1038/nrmicro.2017.87>.
- Fierer, N., Leff, J.W., Adams, B.J., Nielsen, U.N., Bates, S.T., Lauber, C.L., Owens, S., Gilbert, J.A., Wall, D.H., Caporaso, J.G., 2012. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proceedings of the National Academy of Sciences* 109, 21390–21395. <https://doi.org/10.1073/pnas.1215210110>.
- Frostegård, A., Bååth, E., 1996. The use of phospholipid fatty acid analysis to estimate bacterial and fungal biomass in soil. *Biology and Fertility of Soils* 22, 59–65. <https://doi.org/10.1007/BF00384433>.
- Galbally, I.E., Kristine, W., 2002. The production of methanol by flowering plants and the global cycle of methanol. *Journal of Atmospheric Chemistry* 43, 195–229. <https://doi.org/10.1023/A:1020684815474>.
- Golchin, A., Oades, J.M., Skjemstad, J.O., Clarke, P., 1994. Soil structure and carbon cycling. *Soil Research* 32, 1043–1068.
- Harris, D., Horváth, W.R., van Kessel, C., 2001. Acid fumigation of soils to remove carbonates prior to total organic carbon or CARBON-13 isotopic analysis. *Soil Science Society of America Journal* 65, 1853–1856. <https://doi.org/10.2136/sssaj2001.1853>.
- Hettich, R.L., Pan, C., Chourey, K., Giannone, R.J., 2013. Metaproteomics: harnessing the power of high performance mass spectrometry to identify the suite of proteins that control metabolic activities in microbial communities. *Analytical Chemistry* 85, 4203–4214. <https://doi.org/10.1021/ac303053e>.
- Heyer, R., Schallert, K., Büdel, A., Zoun, R., Dorl, S., Behne, A., Kohrs, F., Püttker, S., Siewert, C., Muth, T., Saake, G., Reichl, U., Benndorf, D., 2019. A robust and universal metaproteomics workflow for research studies and routine diagnostics within 24 h using phenol extraction, FASP digest, and the MetaProteomeAnalyzer. *Frontiers in Microbiology* 10, 1883. <https://doi.org/10.3389/fmicb.2019.01883>.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25, 1965–1978. <https://doi.org/10.1002/joc.1276>.
- Hou, Q., Kolodkin-Gal, I., 2020. Harvesting the complex pathways of antibiotic production and resistance of soil bacilli for optimizing plant microbiome. *FEMS Microbiology Ecology* 96. <https://doi.org/10.1093/femsec/iaaa142>.
- Hultman, J., Waldrop, M.P., Mackelprang, R., David, M.M., McFarland, J., Blazewicz, S. J., Harden, J., Turetsky, M.R., McGuire, A.D., Shah, M.B., VerBerkmoes, N.C., Lee, L. H., Mavrommatis, K., Jansson, J.K., 2015. Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes. *Nature* 521, 208–212. <https://doi.org/10.1038/nature14238>.
- Jobbágy, E.G., Jackson, R.B., 2000. The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecological Applications* 10, 423–436. [https://doi.org/10.1890/1051-0761\(2000\)010\[0423:TVDOSO\]2.0.CO;2](https://doi.org/10.1890/1051-0761(2000)010[0423:TVDOSO]2.0.CO;2).
- Kallenbach, C.M., Frey, S.D., Grandy, S., 2016. Direct evidence for microbial-derived soil organic matter formation and its ecophysiological controls. *Natural Computing* 7. <https://doi.org/10.1038/ncomms13630>.
- Knief, C., Delmotte, N., Chaffron, S., Stark, M., Innerebner, G., Wassmann, R., von Mering, C., Vorholt, J.A., 2012. Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere of rice. *The ISME Journal* 6, 1378–1390. <https://doi.org/10.1038/ismej.2011.192>.
- Kreyling, J., Schweiger, A.H., Bahn, M., Ineson, P., Migliavacca, M., Morel-Journel, T., Christiansen, J.R., Schtickzelle, N., Larsen, K.S., 2018. To replicate, or not to replicate – that is the question: how to tackle nonlinear responses in ecological experiments. *Ecology Letters* 21, 1629–1638. <https://doi.org/10.1111/ele.13134>.
- Kuske, C.R., Ticknor, L.O., Miller, M.E., Dunbar, J.M., Davis, J.A., Barns, S.M., Belpat, J., 2002. Comparison of soil bacterial communities in rhizospheres of three plant species and the interspaces in an arid grassland. *Applied and Environmental Microbiology* 68, 1854–1863. <https://doi.org/10.1128/AEM.68.4.1854-1863.2002>.
- 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, 5111–5120. <https://doi.org/10.1128/AEM.00335-09>.
- Lavallee, J.M., Soong, J.L., Cotrufo, M.F., 2020. Conceptualizing soil organic matter into particulate and mineral-associated forms to address global change in the 21st century. *Global Change Biology* 26, 261–273. <https://doi.org/10.1111/gcb.14859>.
- Liu, D., Keibler, K.M., Leitner, S., Wegner, U., Zimmermann, M., Fuchs, S., Lassek, C., Riedel, K., Zechmeister-Boltenstern, S., 2019. Response of microbial communities and their metabolic functions to drying–rewetting stress in a temperate forest soil. *Microorganisms*. <https://doi.org/10.3390/microorganisms7050129>.
- Lladó, S., Větrovský, T., Baldrian, P., 2019. Tracking the activity of individual bacteria in temperate forest soils shows guild-specific responses to seasonality. *Soil Biology and Biochemistry* 135, 275–282. <https://doi.org/10.1016/j.soilbio.2019.05.010>.
- Maestre, F.T., Quero, J.L., Gotelli, N.J., Escudero, A., Ochoa, V., Delgado-Baquerizo, M., García-Gómez, M., Bowker, M.A., Soliveres, S., Escolar, C., García-Palacios, P., Berdugo, M., Valencia, E., Gozalo, B., Gallardo, A., Aguilera, L., Arredondo, T., Blones, J., Boeken, B., Bran, D., Conceição, A.A., Cabrera, O., Chaiieb, M., Derak, M., Eldridge, D.J., Espinosa, C.I., Florentino, A., Gaitán, J., Gatica, M.G., Ghiloufi, W., Gómez-González, S., Gutiérrez, J.R., Hernández, R.M., Huang, X., Huber-Sannwald, E., Jankju, M., Miriti, M., Moneris, J., Mau, R.L., Morici, E., Naseri, K., Ospina, A., Polo, V., Prina, A., Pucheta, E., Ramírez-Collantes, D.A., Romão, R., Tighe, M., Torres-Díaz, C., Val, J., Veiga, J.P., Wang, D., Zaady, E., 2012. Plant species richness and ecosystem multifunctionality in global drylands. *Science* 335, 214–218. <https://doi.org/10.1126/science.1215442>.
- Malik, A.A., Thomson, B.C., Whiteley, A.S., Bailey, M., Griffiths, R.I., 2017. Bacterial physiological adaptations to contrasting edaphic conditions identified using landscape scale metagenomics. *mBio* 8. <https://doi.org/10.1128/mBio.00799-17>.
- Mikutta, R., Turner, S., Schippers, A., Gentsch, N., Meyer-Stüve, S., Condron, L.M., Peltzer, D.A., Richardson, S.J., Eger, A., Hempel, G., Kaiser, K., Klotzbücher, T., Guggenberger, G., 2019. Microbial and abiotic controls on mineral-associated organic matter in soil profiles along an ecosystem gradient. *Scientific Reports* 9. <https://doi.org/10.1038/s41598-019-46501-4>.
- Moreno-Jiménez, E., Plaza, C., Saiz, H., Manzano, R., Flágr, M., Maestre, F.T., 2019. Aridity and reduced soil micronutrient availability in global drylands. *Nature Sustainability* 2, 371–377. <https://doi.org/10.1038/s41893-019-0262-x>.
- Noronha, M.F., Lacerda Júnior, G.V., Gilbert, J.A., de Oliveira, V.M., 2017. Taxonomic and functional patterns across soil microbial communities of global biomes. *The Science of the Total Environment* 609, 1064–1074. <https://doi.org/10.1016/j.scitotenv.2017.07.159>.
- Olsen, S.R., Sommers, L.E., 1983. Phosphorus. *Methods of Soil Analysis*. Agronomy Monographs. <https://doi.org/10.2134/agronmonogr9.2.2ed.c24>.
- Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., Pérez, E., Uszkoreit, J., Pfeuffer, J., Sachsenberg, T., Yilmaz, S., Tiwari, S., Cox, J., Audain, E., Walzer, M., Jarnuczak, A.F., Ternent, T., Brazma, A., Vizcaíno, J.A., 2018. The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Research* 47, D442–D450. <https://doi.org/10.1093/nar/gky1106>.
- Plaza, C., Giannetta, B., Benavente, I., Vischetti, C., Zaccone, C., 2019. Density-based fractionation of soil organic matter: effects of heavy liquid and heavy fraction washing. *Scientific Reports* 9. <https://doi.org/10.1038/s41598-019-46577-y>.
- Quideau, S.A., Chadwick, O.A., Benesi, A., Graham, R.C., Anderson, M.A., 2001. A direct link between forest vegetation type and soil organic matter composition. *Geoderma* 104, 41–60. [https://doi.org/10.1016/S0016-7061\(01\)00055-6](https://doi.org/10.1016/S0016-7061(01)00055-6).
- Ramírez, K.S., Leff, J.W., Barberán, A., Bates, S.T., Betley, J., Crowther, T.W., Kelly, E.F., Oldfield, E.E., Shaw, E.A., Steenbock, C., Bradford, M.A., Wall, D.H., Fierer, N., 2014. Biogeographic patterns in below-ground diversity in New York City's Central Park are similar to those observed globally. *Proceedings. Biological Sciences* 281. <https://doi.org/10.1098/rspb.2014.1988>.
- Rinnan, R., Bååth, E., 2009. Differential utilization of carbon substrates by bacteria and fungi in tundra soil. *Applied and Environmental Microbiology* 75, 3611–3620. <https://doi.org/10.1128/AEM.02865-08>.
- Six, J., Conant, R.T., Paul, E.A., Paustian, K., 2002. Stabilization mechanisms of soil organic matter: implications for C-saturation of soils. *Plant and Soil* 241, 155–176. <https://doi.org/10.1023/A:1016125726789>.
- Sohi, S.P., Mahieu, N., Arah, J.R.M., Powlson, D.S., Madari, B., Gaunt, J.L., 2001. A procedure for isolating soil organic matter fractions suitable for modeling. *Soil Science Society of America Journal* 65, 1121–1128. <https://doi.org/10.2136/sssaj2001.6541121x>.
- Starke, R., Kermer, R., Ullmann-Zeunert, L., Baldwin, I.T., Seifert, J., Bastida, F., von Bergen, M., Jehmlich, N., 2016. Bacteria dominate the short-term assimilation of plant-derived N in soil. *Soil Biology and Biochemistry* 96, 30–38. <https://doi.org/10.1016/j.soilbio.2016.01.009>.
- Starke, R., Bastida, F., Abadía, J., García, C., Nicolás, E., Jehmlich, N., 2017. Ecological and functional adaptations to water management in a semiarid agroecosystem: a soil metaproteomics approach. *Scientific Reports* 7. <https://doi.org/10.1038/s41598-017-09973-w>.
- Starke, R., Jehmlich, N., Bastida, F., 2019. Using proteins to study how microbes contribute to soil ecosystem services: the current state and future perspectives of soil metaproteomics. *JOURNAL OF PROTEOMICS* 198, 50–58. <https://doi.org/10.1016/j.jprot.2018.11.011>.
- Sy, A., Timmers, A.C.J., Knief, C., Vorholt, J.A., 2005. Methylobacterium metabolism is advantageous for Methylobacterium exortquens during colonization of Medicago truncatula under competitive conditions. *Applied and Environmental Microbiology* 71, 7245–7252. <https://doi.org/10.1128/AEM.71.11.7245-7252.2005>.
- Thorn, C.E., Bergesch, C., Joyce, A., Sambrano, G., McDonnell, K., Brennan, F., Heyer, R., Benndorf, D., Abram, F., 2019. A robust, cost-effective method for DNA, RNA and protein co-extraction from soil, other complex microbiomes and pure cultures. *Mol. Ecol. Resour.* 19, 439–455. <https://doi.org/10.1111/1755-0998.12979>.
- Wang, C., Meek, D.J., Panchal, P., Boruvka, N., Archibald, F.S., Driscoll, B.T., Charles, T. C., 2006. Isolation of poly-3-hydroxybutyrate metabolism genes from complex microbial communities by phenotypic complementation of bacterial mutants. *Applied and Environmental Microbiology* 72, 384–391. <https://doi.org/10.1128/AEM.72.1.384-391.2006>.
- Yu, Tong, Zhuang, Q., 2020. Modelling biological nitrogen fixation in global natural terrestrial ecosystems. *Biogeosciences* 17, 3643–3657. <https://doi.org/10.5194/bg-17-3643-2020>.
- Žifčáková, L., Větrovský, T., Lombard, V., Henrissat, B., Howe, A., Baldrian, P., 2017. Feed in summer, rest in winter: microbial carbon utilization in forest topsoil. *Microbiome* 5, 122. <https://doi.org/10.1186/s40168-017-0340-0>.