Using proteins to study how microbes contribute to soil ecosystem services: The current state and future perspectives of soil metaproteomics

Robert Starke[1], Nico Jehmlich[2], Felipe Bastida[3]

[1]Laboratory of Environmental Microbiology, Institute of Microbiology of the CAS, Vídeňská 1083, 14220 Praha 4, Czech Republic; [2]Helmholtz-Centre for Environmental Research – UFZ, Department of Molecular Systems Biology, Permoserstraße 15, 04318 Leipzig, Germany; [3]CEBAS-CSIC, Campus Universitario de Espinardo, Murcia, E-30100, Spain

Abstract

Metaproteomics was established to analyse both the structure and the function of microbial communities and, particularly in soils, their contribution to ecosystem services. In this review, we provide an overview on how the study of the soil metaproteome can provide fundamental information on the role of microbial communities in soil ecosystem services. We further discuss the strengths and weaknesses of soil metaproteomics in comparison to other culture-independent OMIC techniques. We critically review its bottlenecks but also provide strategies to mitigate and possible directions for future research as the direct link of structure and function is advantageous and complementary to metagenomics, metatranscriptomics and metametabolomics.

Soil ecosystem services provided by microbes

Soil, a key component of the terrestrial ecosystem, operates at the interface of the atmosphere, biosphere, hydrosphere and lithosphere [1], and provides many ecosystem services - acting as a habitat for soil organisms, regulator of water quality, modifier of the atmospheric composition, medium for plant growth and recycling system for nutrients and organic wastes, etc. [2]. Even though soil is one of the most important carbon (C) reservoirs within the Earth's ecosystems [3,4], most of its microbes have not yet been described [5,6]. But, one gram of soil can contain billions of organisms belonging to thousands of different species [7]. In particular, the mean prokaryotic density of 10,000,000 organisms per gram of soil [8] is at least one magnitude of order higher than that found per millilitre of water in the ocean [9]. This implies the importance of understanding microbially-driven soil processes such as C-cycling and soil fertility that ultimately lead to human wellbeing. Surprisingly, microbial biomass typically accounts for only 1% of soil organic carbon (SOC) [10,11] but its contribution to C cycling is extensive, through its implication in the mineralisation and stabilisation of the soil organic matter (SOM) derived from plants [12–14]. In addition, soil microbial community has been shown to be responsive against climate change factors [15–21]. Hence, *understanding the structure and the function of the soil microbial*

community is crucial to the determination of its contribution to soil ecosystem services and how global change and land use intensification will affect this.

The standard metric of quantification in soil microbial ecology studies is the reading of nucleic acid counts. The data derived from genomic studies permit: i) the identification and quantification of the relative abundance of microbial populations through amplicon-based approaches (i.e., 16S rRNA genes; ITS fragments, etc.); ii) the quantification and the estimation of the diversity of functional genes (e.g., genes related to N cycling, such as amoA, nirS and nirK; and, more recently, genes involved in P cycling, such as PhoD and PhoC [22]); iii) the inference of their functional role through statistical methods which relate nucleic acids to functional parameters (i.e., extracellular enzyme activities, basal respiration, etc.); and iv) metagenomic approaches, which aim to decipher the whole repertoire of genes in a soil microbial community which attempts to potentially connect phylogeny and functionality.

Through sequencing approaches, several studies have revealed the core composition of soil microbial communities at the global scale. For instance, Delgado-Baquerizo and colleagues found that only 2% of bacterial phylotypes, which amounted to 511 individual phylotypes, consistently account for almost half of the soil bacterial communities worldwide [23]. This "most wanted" list included *Proteobacteria*, *Actinobacteria*, *Acidobacteria*, *Planctomycetes*, *Chloroflexi*, *Verrucomicrobia*, *Bacteroidetes*, *Gemmatimonadetes*, *Firmicutes* and *Armatimonadetes* in decreasing order. Complementary, Tedersoo and colleagues used DNA metabarcoding to describe that fungal richness is decoupled from plant diversity but fungi show similar latitudinal diversity gradients to other organisms [24]. Similar to bacteria, the abundant phylotypes globally, comprising of *Agaricomycetes* (50.1%), *Sordariomycetes* (8.0%), *Eurotiomycetes* (7.7%), *Leotiomycetes* (7.1%) and *Mortierellomycotina* (6.3%), describe more than half of the fungal communities worldwide. When comparing metaproteomics and 16S rRNA gene sequencing in samples from North-America and South-Europe, the patterns of bacterial composition at phylum level resembled similar patterns, but the relative abundance of *Proteobacteria* and *Planctomycetes* was greater in metaproteomics than metagenomics [25]. Nevertheless, although the composition of microbial communities at the phylum level can be partially predicted (either with metagenomics and metaproteomics), the composition of microbial communities at finer taxonomic levels could probably arise in greater variations across soil types and climates.

However, sequencing harbors limitations such as the quantitative accuracy [26,27] which is why the actual functionality of microbes should be deciphered by analysis of the catalysts of soil ecosystem services – the proteins, which provide both functional and phylogenetic information [28–30]. This is fundamental since it has been discovered recently that a large part (up to 40%) of the soil DNA can

belong to dead cells (relic DNA) which are not active anymore [31], whereas extracellular proteins in soil (e.g., extracellular enzymes) can persist and remain active through stabilisation by humic substances and clays [32].

The significance of soil metaproteomics

In 2009, Bastida and colleagues published a review that ~~focused~~evaluated the initial potentials of soil metaproteomics. This review was focused ~~–~~on the methodology of soil metaproteomics with the emphasis on the different protein extraction techniques and peptide identification by mass spectrometric analysis as the identification of soil proteins may provide information about the biogeochemical potential of soils and pollutant degradation and act as an indicator of soil quality, identifying which proteins and microorganisms are affected by a degradation process [33]. In their review from 2016, Keiblinger and colleagues discussed the opportunities and limitations of selected protein extraction techniques in soil and leaf litter metaproteomics, including a step-by-step guideline on application, sampling, sample preparation, extraction and data evaluation strategies, and used recent application to discuss how linking phylogeny and taxonomy can help to gain deeper insights in terrestrial microbial ecology [34]. In this review, we focus on recent applications of metaproteomics but also highlight the bottlenecks and offer a case-study where major limitations are explored (e.g. biomass and soil organic C content). Further, we establish a set of research questions that can be potentially investigated in the coming years with the use of metaproteomics.

Kommentar [FB1]: Split this sentence within 2 sentences

Among the culture-independent OMIC techniques deployed to gain deeper insights into the structure and function of microbial communities [35], metaproteomics has gained more and more interest from the scientific community as a central element in microbial ecology studies [30], since it deciphers the functional relationships between community members [36], particularly in soils [15,36–53]. Thus far, 49 articles on soil metaproteomics have been published, in comparison to 54 on soil metatranscriptomics and 749 on soil metagenomics (PubMed as of July 16th, 2018). Genomics is still the method of choice used by many soil scientists to evaluate the diversity and composition of the soil microbial community, due to easier handling, lower costs and commercially-available sequencing provided by companies (for instance, Agilent Technologies Inc. and Illumina Inc.) or institutes (for instance, the Joint Genome Institute or Argonne National Laboratory). In comparison, proteomics involves complex sample-extraction protocols (which still need to be improved) and cost-intense and sensitive equipment to which only a few laboratories in the world have access. However, the ability of metaproteomics to link the phylogenetic structure to the function is a clear advantage over nucleic acid and metabolomic approaches in soil microbial ecology.

The applications of soil metaproteomics

Soil metaproteomics has been used to describe the soil microbial community in relation to ecosystem services, such as agriculture [38,41,45,54], bioremediation [36,43,52] and C cycling [40,42,47–49,51], and climate change factors (i.e., temperature) [15,37,39,53]. The results obtained from nucleic acids and proteomics, at least at the phylogenetic level, have shown good correlation [15,48,53]. Hence, the use of proteins to describe ecosystem functioning in soils is of interest because it captures both phylogenetic and functional information [15,41,42,48].

Carbon (C) and nitrogen (N) cycles

Carbon cycling is fundamental to the Earth's ecosystem since C is the main component of all biological compounds. The greenhouse gases carbon dioxide (CO_2) and methane, two forms of C, absorb and retain heat in the Earth's atmosphere and are thus partially responsible for the greenhouse effect [3], which is why an understanding of C cycling is crucial to the comprehension of climate change. Indeed, the concentration of atmospheric CO_2 recorded in summer 2018 was the highest since records began, which underlines the need for a better understanding of the biotic factors (e.g., microbes) controlling the mineralisation of SOM, which is driven by heterotrophic soil microbes [55]. Within the C-cycle, the restoration of C-limited soils, such as those frequently found under arid and semi-arid climates, by the application of organic amendments (i.e., sludges, composts, etc.) is a major task that is necessary to maintain soil fertility. Through a metaproteomic approach, it has been found that organic amendments strongly impact soil ecosystem processes, at both the cellular and extracellular level, depending on the type of organic amendment [47]. Further, metaproteomics revealed that not only the nutrient concentrations but also the C:N:P stoichiometry influenced both the structure and the activity of the microbial community during beech litter decomposition [40], whereas deforestation induced a long-term loss of bacterial biomass and enzyme activity but increased bacterial diversity [49]. Moreover, the active diversity (measured by metaproteomics), but not the total microbial diversity (through amplicon-based genomics), was shown to be related to the availability of organic C [48] in a semi-arid ecosystem. Lastly, metaproteomic approaches have been utilised to track the N-cycle. The degradation of plant-derived polymeric organic C (i.e., lignin, cellulose, etc.) is dominated by fungi [56,57],but bacteria were recently shown to dominate the short-term assimilation of plant-derived N, presumably in easily-available sources, by tracing ^{15}N incorporation into proteins of the microbial community [41].

Pollution, bioremediation and restoration

The growth of the human population is increasing the pollution of the environment and hence a detailed understanding of the microbial populations associated with the biodegradation of pollutants is required, as well as of the main biochemical pathways of biodegradation, which have biotechnological applications. In fact, metaproteomics revealed that petroleum pollution induced an increase in the relative abundance of *Proteobacteria* but a decrease in *Rhizobiales* - which ultimately led to promotion of the removal of polycyclic aromatic hydrocarbons and alkanes [36]; this work also provided parallel information on the biochemical pathways of the biodegradation. Further, the rhizosphere community of constructed wetlands, generally used to treat contaminated water, was shown to have diurnal polyhydroxyalkanoate metabolism, which is aligned with the diurnal cycle in plants [58]. Similarly, naphthalene, an aromatic contaminant, resulted in the separation of a bio-stimulated community - able to utilise naphthalene - from a non-stimulated community exhibiting a greater metabolic window than previously predicted [52], suggesting a core community for specific functionality. Such an active community was found in permafrost-affected cryosols, for atmospheric-methane-oxidising bacteria [37].

Climate change

Climate change is expected to induce a loss of SOC by increasing soil respiration rates [55,59], which will ultimately exacerbate the greenhouse effect [60]. Recent evidence indicates that warming is accelerating the activity of soil microorganisms at the global scale, with the subsequent increase in CO_2 release from soil to atmosphere [61]. As soil respiration is mainly performed by heterotrophic soil microorganisms [55], understanding the structure and function of the microbial community is vital to comprehend responses to climate change. Changes in the phylogenetic composition of the microbial community and in its functional potential and activity were shown to represent different states of thaw in soils [53]. In contrast, warming resulted only in subtle differences at the phylum and class levels of the community whereas functionality increased, particularly energy production and conversion [39]. Similarly, the structure and diversity of the total bacterial community were unaffected by drought even though microbial biomass and ecosystem multifunctionality decreased, in a semi-arid ecosystem [15]. In fact, climate was reported to be a fundamental driver of the protein abundance of various phyla in a wide range of soils from different continents [25].

Soil fertility and agriculture

Agricultural productivity does not depend only on the physical and chemical quality of soil (e.g., texture, density, nutrients and water availability, etc.) or the absence of plant pathogens. The soil microbial community plays a critical role in agricultural productivity through its cycling of nutrients, making them available for plant growth. Indeed, the soil microbial diversity is fundamental to the maintenance of soil productivity and the resulting economic benefits [62]. Soil metaproteomics could help to dig deeper into the functions and role of microbial populations in plant-soil interactions in an agricultural context. In this sense, crop yields and the optimal growth conditions have been evaluated through metaproteomic approaches. For instance, the mere presence of plants influences the metaproteome and soil bacterial functions [63]. Further, a yield decline of crops (i.e., sugarcane) was shown to be related to the soil metaproteome, as well as to enzyme activities and the catabolic diversity of the microbial community [38]. Moreover, changes in microbial abundance in the rhizosphere have been related to the protection of tomato plants from wilt disease, presumably by beneficial microbial proteins in healthy soil [64], and to ecological and functional adaptations to varying water management in semi-arid ecosystems [41].

The bottlenecks of soil metaproteomics

Protein extraction, sample processing and measurement

Protein extraction from soil is time-consuming as it comprises protein extraction (2 days) followed by gel electrophoresis (1 day), tryptic digestion for at least 16 hours or overnight and desalting (2 days) (Figure 1). The commonly-used extractants are sodium dodecyl sulphate and trichloroacetic acid (SDS-TCA) [50] and SDS-phenol [65]. The extracted proteins are then analysed by 2D-SDS-PAGE gel electrophoresis for imaging and MALDI-TOF analysis [66] or, more commonly, by 1D-SDS-PAGE for LC-MS/MS analysis. The 1D approach comprises tryptic digestion of the proteins in-solution or in-gel, desalting via ZipTip columns or SCX spin columns, measurement with a coupled system of liquid chromatography and tandem mass spectrometry (LC-MS/MS), database searches with Proteome Discoverer (using Mascot or Sequest as the algorithm) against every known protein-coding sequence or a metagenome, the use of Percolator to calculate false discovery rates (FDR) [67], phylogenetic and functional assignment of protein groups with the "Proteomics results Pruning & Homology group Annotation engine" (PROPHANE) [68], calculation of protein abundances - as either the normalised spectral abundance factor (NSAF) [69] or the area under the curve (AUC) - and, finally, data upload to ProteomeXchange via PRIDE [70] (Please refer to the reviews by Muth and colleagues [71], and Heyer and colleagues [72] for further details on the bioinformatic data analysis of metaproteomes). However, slight differences in the

extraction method, sample preparation, instrumentation and data analysis were found among soil metaproteomic studies. Each of these steps has been optimised but requires further validation to obtain high-quality proteomics data. However, among these steps, protein extraction and data analysis (see next section) can influence the outcome the most.

The extraction of extracellular proteins remains a challenge due to their physicochemical interactions with SOM and soil mineral particles [33], which were reported to lower the extraction efficiencies and final peptide yields [46,73]. For example, an increasing content of montmorillonite, an expansible 2:1 clay, in soil was found to decrease the amount of extracted proteins and to hinder protein identification, likely due to conformational changes or degradation [74]. In addition to the interaction with clay, it was demonstrated that the contact between proteins and soil-borne humic substances, potentially due to hydrophobic interactions, led to protein modifications that affected the identification by MS; this was especially prominent in proteins of lower molecular weight and of less-complex structure [75]. Previously hypothesised were hydrophobic interactions caused by dehydration, which could destabilise and partially disassemble humic substances in contact with the proteins [76] or supramolecular aggregation of low molecular weight organic molecules held together by weak molecular surface forces [77]. However, it remains unknown which modifications specifically lead to lower identification rates.

The relationship between the concentration of SOC or microbial biomass (both as *S*) and the number of identified peptides (*v*) using equation 1 as previously described [78] are illustrated in Figure 2, with V_{max} as the maximum number of identified peptides and K_m the concentration of SOC or microbial biomass at which half of the maximum number of peptides are identified.

Eq. 1: $$v = \frac{V_{max}[S]}{K_m + [S]}$$

These results were obtained from a gradient of semi-arid soils that covered a range of SOC concentrations. The SOC concentration has a negative impact on the number of spectra and, consequently, on the number of identified peptides (Figure 2a). The relationship between the SOC concentration and the number of identified peptides reveals a plateau of extracted proteins at a breaking point of 1.5% SOC, despite the greater microbial biomass in soils richer in SOC that should logically lead to higher protein content. In fact, the efficiency of protein extraction is lowered not only by increasing SOC but also by increasing microbial biomass content (Figure 2b). The number of peptide identifications stagnated when the microbial biomass reached a breaking point of 16.8 nmol fatty acids per gram of soil, which follows previously-reported shielding from disruption treatments at high cell densities, due to aggregation [79]. Hence, (i) more-efficient cell lysis techniques and (ii) more-efficient protein extraction techniques must be investigated to assure the quality and quantity of proteomics

data. Further, the extraction techniques and buffers commonly used for the measurement of extracellular enzyme activities were shown to have low extraction efficiencies, which points to strong adsorption of proteins on humic substances and clays [80]. In addition, a definitive quantification of the extracted proteins was still lacking in many studies [46,80] as the scientific community has commonly relied on colorimetric methods - for instance, the Bradford assay - that are biased by the interference of humic compounds [81,82]. However, correct quantification is crucial, to evaluate both the efficiency of protein extraction and the full contribution of the protein repertoire to soil ecosystem services. In this regard, the quantification of the amino acids from the extracted proteins [46] or the amido black assay [83] could provide the most-straightforward approaches to estimating the amount of extracted proteins. Otherwise, total ion chromatogram counts from liquid chromatography could be used but accurate quantification requires internal standards and calibration for each protein of interest. However, the intensity cannot be directly related to abundance due to differences in ionisation efficiencies between peptides and discrepancies between ionisation and detection since only one out of every 100 to 10,000 analyte ions generated is detected [84–86].

Data analysis and functional assignments

The backbone of proteomic analysis is the database, as the measured peptides from the sample are matched, using data analysis programs such as MASCOT [87,88] or SEQUEST [89], with the database sequences. Hence, knowledge of what is in the sample is as important as the quantity and quality of the proteins within the database. Unsurprisingly, since most organisms in soil are unknown [5,6], the database searches usually involve the utilisation of all known protein sequences - which comprise 79,565,724 TrEMBL and 335,933 Swiss-Prot entries for bacteria, 8,114,426 TrEMBL and 33,876 Swiss-Prot entries for fungi, 3,591,467 TrEMBL and 20,986 Swiss-Prot entries for archaea and 629,133 TrEMBL and 4,636 Swiss-Prot entries for nematodes. These are weighted in favor of intensively-investigated organisms such as the bacterial *E. coli* K12 (4,497), the fungal *S. cerevisiae* (6,729) or the nematode *C. elegans* (26,846) (Uniprot as of July 7th, 2018 [90]). Instead of using all known sequences, one could provide a database by using shotgun sequencing of DNA [91,92], which is both costly and time-consuming and requires additional programs and knowledge. A possible solution could be to make sequences publicly available as a community effort like the Human Genome Project [93], which has already been started for proteomic datasets with PRIDE [70,94–98] but, hereafter, soil studies focusing on all clades of the ecosystem should be emphasised.

Recent studies have focused on the community composition to assess ecosystem functioning but lack direct functional classification even though the peptides identified with MS were assigned to proteins

with both taxonomic and functional information. Further, in complex communities containing thousands of different species per gram of soil [7], not many proteins will be assigned to individual species of the community as one measurement generally yields a few thousand proteins. Moreover, these microbial communities are not homogenous, generally having a few highly-abundant species and many species of low abundance. Worse still, proteins with ubiquitous functions such as ATP synthases or chaperones - which every cell contains, regardless of its type - are more abundant than specialised enzymes such as benzoyl-CoA reductases and the many proteins whose functions are still unknown. These functional proteins are generally less abundant than "housekeeping" proteins (e.g., proteins involved in the Krebs cycle, cellular respiration, lipid metabolism, DNA replication, etc.), which makes them hard to detect among the most-abundant proteins. Hence, while the taxonomic affiliation generates profound data on what is there, the functional information is biased by ubiquitous enzymes and, depending on the degree of heterogeneity, by specialised enzymes of the most-abundant organisms. Furthermore, the identification of extracellular soil enzymes (i.e., glucosidases, phosphatases, cellulases, etc.), which are deeply involved in nutrient cycles [32], is hampered by the inefficient extraction of the extracellular metaproteome of soil because proteins are frequently linked to humic substances or clays [80]; hence, the genomic database is still poor in this respect. To circumvent these issues and directly assess functionality, targeted approaches, instead of the commonly-used untargeted approaches [99,100], have to be considered; this implies the selection of specific proteins for the measurements, by targeting the masses of peptides from proteins. Briefly, to achieve consistent quantification of sets of proteins across a variety of different samples and replicates, targeted MS - such as selected reaction monitoring [101,102] - could provide specific assays for the detection and quantification of proteins over the whole range of cellular concentrations [103].

Stable isotope probing (SIP) of proteins

Metaproteomics can provide information on both phylogeny and functionality [36] in microbial ecology studies [30], but only the use of SIP can directly link microbial populations within the whole community to a specific function [104]. Possible labelling targets in proteins are C, hydrogen, N, oxygen and sulphur [105]. All of which can be now analysed by the tool MetaProSIP [106] that uses the identification of the peptide on MS^2-level together with detecting and describing the incorporation pattern on MS^1-level by relative isotope abundance (RIA) and labelling ratio (LR). Of these, C is the one used most frequently in environmental SIP studies [107,108] due to its high abundance in amino acids and since it is a main component of the bacterial biomass [109]. Otherwise, hydrogen has the second-highest abundance in proteins, but its use is limited by the rapid HD-exchange [110] which will occur in living cells and during

sample preparation. In addition, the toxicity of highly-deuterated water has been described for eukaryotes; in particular, cell division is hampered during the formation of the mitotic spindle [111]. By contrast, prokaryotes were reported to grow in artificial conditions with a high deuterium content [112], up to 98% D_2O [113]. However, the chromatographic properties of deuterated compounds - originating from the higher hydrophilicity of CD bonds, compared to CH bonds [114] - result in significantly-different retention times [115]. Nitrogen, on the other hand, has a lower abundance in amino acids, resulting in smaller mass shifts and reduced sensitivity [116]. Oxygen has not been used in protein-SIP experiments, perhaps because of the hampering of direct incorporation from the substrate into biomass due to the exchange of 18-oxygen for 16-oxygen water through enzymatic reactions [117]. Lastly, sulphur is only present in cysteine and methionine, which is beneficial for the tracking of general metabolic activity [118] but, due to the low amount of sulphur in proteins, according to the average model, the utilisation of labelled sulphur in protein-SIP appears inauspicious. As of now, the only reported isotopic labelling of soils has been done with 15-nitrogen, using labelled plant material (tobacco) to investigate its assimilation by the soil microbial community, through protein-SIP [42]. In this study, between 11 and 26 labelled peptides were identified per time point over the course of 14 days; this showed that bacteria dominated the short-term assimilation of plant N. However, if one relates this number of labelled peptides to the existence of 10,000,000 organisms per gram of soil [8], then the validity of the data and their value regarding the determination of ecosystem functioning are challenged. Metabolic labelling of energy or nutrient sources leads to dynamic and unpredictable mass shifts [109] and, hence, small amounts of label indicate a low number of active key players within the microbial community with respect to specific substrates such as N. This is further complicated by the fact that microbial biomass, even though its derived products significantly contribute to SOM [12–14], typically accounts for only 1% of SOC [10,11] and many of its species are thought to be dormant [16,31,119,120]. To achieve higher labelling efficiency, the utilisation of heavy water, labelled with either deuterium or 18-oxygen, is promising as it assesses the global activity regardless of the specific substrate. It was successfully demonstrated in acid-mine-drainage biofilms using D_2O [121] and in soil communities using $H_2{}^{18}O$ in DNA [122–124] and RNA [125,126]. In fact, SIP of fully-18-oxygen-labelled water in RNA was used recently to elucidate that 94% of soil taxa produced new rRNA and therefore were metabolically active [127] - which is the opposite of the common-acceptedly state of high dormancy of soil microbes. However, this approach has yet to be demonstrated for proteins and in soils. In particular, 18-oxygen seems more promising than deuterium as (i) the abiotic HD-exchange of acidic hydrogens once the proteins are in contact with unlabelled water [110] could cause depletion of the label and (ii) the cell

division of eukaryotes is hampered due to the negative effect of deuterium on the mitotic spindle formation [111], even though it has been reported that bacteria survive in 98% D_2O [113]. Hence, deuterated water could negatively impact the activity of eukaryotes such as fungi and nematodes, which would falsely indicate high bacterial activity.

Research questions to be answered by soil metaproteomics

To date, the knowledge derived from soil metaproteomics has been rather descriptive, mainly providing information on the community composition and cellular functional attributes. However, the advent of new ecological concepts in the field of soil metaproteomics [128] will provide a functional basis for the development of microbial-ecological theories in soil science. In this respect, we propose several fundamental research questions that could be investigated with the use of metaproteomics (of course, in combination with classical and other OMIC approaches in soil science). For instance, *how do land use and associated changes in soil nutrients determine the protein repertoire and its functional repercussion within soil ecosystem services?* Further, environments with higher phylogenetic diversity are supposed to be more resilient/resistant against harmful stressors such as drought [15–21,129]. Hence, *does the diversity of proteins reflect the resilience/resistance of a soil microbial community?* This would be critical with respect to gaining a mechanistic understanding how microbial communities can resist climate change factors such as global warming or aridity. If so, *can we predict "functional markers" of climate-change resistance in the soil microbial community?* Finally, it has been shown recently how microbes interact with each other at the global scale [130] and how climate can affect not only soil microbial communities but also the co-occurrence networks [131]. Consequently, *is metaproteomics able to provide functional insights into these interactions?* While proteomics has proven to be powerful, with advantages over other culture-independent OMIC tools and culture-dependent techniques, the bottlenecks must be reviewed and actively targeted to sustain high-quality research using proteins from environmental samples. In fact, better integration of techniques and research groups can guide the scientific community to a more-comprehensive understanding of soil ecosystem services since the data are complementary and not exclusive. Besides, more effort must be made to improve methods of protein extraction from soil as well as the identification of low-abundance proteins with functional importance. In conclusion, despite the current limitations, soil metaproteomics has rapidly developed since 2007 [132] and there is room for terrific improvements in the near future. Soil metaproteomics is contributing to our understanding of fundamental and global aspects of soil microbial ecology (and will continue to do so), which will ultimately help us to understand the role of microbes in ecosystem

services in the light of the land-use intensification and global changes forecasted for the coming decades.

References

[1] I. Szabolcs, The concept of soil resilience, CAB Int. (1994).

[2] E. Dominati, M. Patterson, A. Mackay, A framework for classifying and quantifying the natural capital and ecosystem services of soils, Ecol. Econ. (2010). doi:10.1016/j.ecolecon.2010.05.002.

[3] P. Falkowski, R.J. Scholes, E. Boyle, J. Canadell, D. Canfield, J. Elser, N. Gruber, K. Hibbard, P. Hogberg, S. Linder, F.T. Mackenzie, B. Moore, T. Pedersen, Y. Rosental, S. Seitzinger, V. Smetacek, W. Steffen, The global carbon cycle: A test of our knowledge of earth as a system, Science (80-.). (2000). doi:10.1126/science.290.5490.291.

[4] D. Schimel, Terrestrial ecosystems and the carbon cycle, Glob. Chang. Biol. (1995). doi:10.1111/j.1365-2486.1995.tb00008.x.

[5] J. Gans, M. Wolinsky, J. Dunbar, Microbiology: Computational improvements reveal great bacterial diversity and high toxicity in soil, Science (80-.). (2005). doi:10.1126/science.1112665.

[6] A. Dance, Soil ecology: What lies beneath, Nature. (2008). doi:10.1038/455724a.

[7] V. Torsvik, L. Øvreås, Microbial diversity and function in soil: From genes to ecosystems, Curr. Opin. Microbiol. (2002). doi:10.1016/S1369-5274(02)00324-7.

[8] X. Raynaud, N. Nunan, Spatial ecology of bacteria at the microscale in soil, PLoS One. (2014). doi:10.1371/journal.pone.0087217.

[9] W.B. Whitman, D.C. Coleman, W.J. Wiebe, Prokaryotes: The unseen majority, Proc. Natl. Acad. Sci. (1998). doi:10.1073/pnas.95.12.6578.

[10] H.M. Serna-Chavez, N. Fierer, P.M. Van Bodegom, Global drivers and patterns of microbial abundance in soil, Glob. Ecol. Biogeogr. (2013). doi:10.1111/geb.12070.

[11] N. Fierer, M.S. Strickland, D. Liptzin, M.A. Bradford, C.C. Cleveland, Global patterns in belowground communities, Ecol. Lett. (2009). doi:10.1111/j.1461-0248.2009.01360.x.

[12] H.M. Throckmorton, J.A. Bird, L. Dane, M.K. Firestone, W.R. Horwath, The source of microbial C has little impact on soil organic matter stabilisation in forest ecosystems, Ecol. Lett. (2012). doi:10.1111/j.1461-0248.2012.01848.x.

[13] M. Schweigert, S. Herrmann, A. Miltner, T. Fester, M. Kästner, Fate of ectomycorrhizal fungal biomass in a soil bioreactor system and its contribution to soil organic matter formation, Soil Biol. Biochem. (2015). doi:10.1016/j.soilbio.2015.05.012.

[14] R. Kindler, A. Miltner, M. Thullner, H.H. Richnow, M. Kästner, Fate of bacterial biomass derived fatty acids in soil and their contribution to soil organic matter, Org. Geochem. (2009). doi:10.1016/j.orggeochem.2008.09.005.

[15] F. Bastida, I.F. Torres, M. Andrés-Abellán, P. Baldrian, R. López-Mondéjar, T. Větrovský, H.H. Richnow, R. Starke, S. Ondoño, C. García, F.R. López-Serrano, N. Jehmlich, Differential sensitivity of total and active soil microbial communities to drought and forest management, Glob. Chang.

Biol. (2017). doi:10.1111/gcb.13790.

[16] E. Blagodatskaya, Y. Kuzyakov, Active microorganisms in soil: Critical review of estimation criteria and approaches, Soil Biol. Biochem. (2013). doi:10.1016/j.soilbio.2013.08.024.

[17] R.L. Barnard, C.A. Osborne, M.K. Firestone, Responses of soil bacterial and fungal communities to extreme desiccation and rewetting, ISME J. (2013). doi:10.1038/ismej.2013.104.

[18] C. V. Hawkes, S.N. Kivlin, J.D. Rocca, V. Huguet, M.A. Thomsen, K.B. Suttle, Fungal community responses to precipitation, Glob. Chang. Biol. (2011). doi:10.1111/j.1365-2486.2010.02327.x.

[19] A. Kaisermann, P.A. Maron, L. Beaumelle, J.C. Lata, Fungal communities are more sensitive indicators to non-extreme soil moisture variations than bacterial communities, Appl. Soil Ecol. (2015). doi:10.1016/j.apsoil.2014.10.009.

[20] K. Felsmann, M. Baudis, K. Gimbel, Z.E. Kayler, R. Ellerbrock, H. Bruehlheide, J. Bruckhoff, E. Welk, H. Puhlmann, M. Weiler, A. Gessler, A. Ulrich, Soil bacterial community structure responses to precipitation reduction and forest management in forest ecosystems across Germany, PLoS One. (2015). doi:10.1371/journal.pone.0122539.

[21] K.J. Romanowicz, Z.B. Freedman, R.A. Upchurch, W.A. Argiroff, D.R. Zak, Active microorganisms in forest soils differ from the total community yet are shaped by the same environmental factors: The influence of pH and soil moisture, FEMS Microbiol. Ecol. (2016). doi:10.1093/femsec/fiw149.

[22] T.D. Fraser, D.H. Lynch, J. Gaiero, K. Khosla, K.E. Dunfield, Quantification of bacterial non-specific acid (phoC) and alkaline (phoD) phosphatase genes in bulk and rhizosphere soil from organically managed soybean fields, Appl. Soil Ecol. (2017). doi:10.1016/j.apsoil.2016.11.013.

[23] M. Delgado-Baquerizo, A.M. Oliverio, T.E. Brewer, A. Benavent-González, D.J. Eldridge, R.D. Bardgett, F.T. Maestre, B.K. Singh, N. Fierer, A global atlas of the dominant bacteria found in soil, Science (80-.). (2018). doi:10.1126/science.aap9516.

[24] L. Tedersoo, M. Bahram, S. Põlme, U. Kõljalg, N.S. Yorou, R. Wijesundera, L.V. Ruiz, A.M. Vasco-Palacios, P.Q. Thu, A. Suija, M.E. Smith, C. Sharp, E. Saluveer, A. Saitta, M. Rosas, T. Riit, D. Ratkowsky, K. Pritsch, K. Põldmaa, M. Piepenbring, C. Phosri, M. Peterson, K. Parts, K. Pärtel, E. Otsing, E. Nouhra, A.L. Njouonkou, R.H. Nilsson, L.N. Morgado, J. Mayor, T.W. May, L. Majuakim, D.J. Lodge, S. Lee, K.H. Larsson, P. Kohout, K. Hosaka, I. Hiiesalu, T.W. Henkel, H. Harend, L.D. Guo, A. Greslebin, G. Grelet, J. Geml, G. Gates, W. Dunstan, C. Dunk, R. Drenkhan, J. Dearnaley, A. De Kesel, T. Dang, X. Chen, F. Buegger, F.Q. Brearley, G. Bonito, S. Anslan, S. Abell, K. Abarenkov, Global diversity and geography of soil fungi, Science (80-.). (2014). doi:10.1126/science.1256688.

[25] F. Bastida, T.W. Crowther, I. Prieto, D. Routh, C. García, N. Jehmlich, Climate shapes the protein abundance of dominant soil bacteria, Sci. Total Environ. (2018). doi:10.1016/j.scitotenv.2018.05.288.

[26] B.A. Hungate, R.L. Mau, E. Schwartz, J. Gregory Caporaso, P. Dijkstra, N. van Gestel, B.J. Koch, C.M. Liu, T.A. McHugh, J.C. Marks, E.M. Morrissey, L.B. Price, Quantitative microbial ecology through stable isotope probing, Appl. Environ. Microbiol. (2015). doi:10.1128/AEM.02280-15.

[27] L.M. Feinstein, J.S. Woo, C.B. Blackwood, Assessment of bias associated with incomplete extraction of microbial DNA from soil, Appl. Environ. Microbiol. (2009). doi:10.1128/AEM.00120-09.

[28] R.L. Hettich, C. Pan, K. Chourey, R.J. Giannone, Metaproteomics: Harnessing the power of high performance mass spectrometry to identify the suite of proteins that control metabolic activities in microbial communities, Anal. Chem. (2013). doi:10.1021/ac303053e.

[29] P. Wilmes, A. Heintz-Buschart, P.L. Bond, A decade of metaproteomics: Where we stand and what the future holds, Proteomics. (2015). doi:10.1002/pmic.201500183.

[30] M. Von Bergen, N. Jehmlich, M. Taubert, C. Vogt, F. Bastida, F.A. Herbst, F. Schmidt, H.H. Richnow, J. Seifert, Insights from quantitative metaproteomics and protein-stable isotope probing into microbial ecology, ISME J. (2013). doi:10.1038/ismej.2013.78.

[31] P. Carini, P.J. Marsden, J.W. Leff, E.E. Morgan, M.S. Strickland, N. Fierer, Relic DNA is abundant in soil and obscures estimates of soil microbial diversity, Nat. Microbiol. (2016). doi:10.1038/nmicrobiol.2016.242.

[32] R.G. Burns, J.L. DeForest, J. Marxsen, R.L. Sinsabaugh, M.E. Stromberger, M.D. Wallenstein, M.N. Weintraub, A. Zoppini, Soil enzymes in a changing environment: Current knowledge and future directions, Soil Biol. Biochem. (2013). doi:10.1016/j.soilbio.2012.11.009.

[33] F. Bastida, J.L. Moreno, C. Nicolás, T. Hernández, C. García, Soil metaproteomics: A review of an emerging environmental science. Significance, methodology and perspectives, Eur. J. Soil Sci. (2009). doi:10.1111/j.1365-2389.2009.01184.x.

[34] K.M. Keiblinger, S. Fuchs, S. Zechmeister-Boltenstern, K. Riedel, Soil and leaf litter metaproteomics-A brief guideline from sampling to understanding, FEMS Microbiol. Ecol. (2016). doi:10.1093/femsec/iw180.

[35] F.A. Herbst, V. Lünsmann, H. Kjeldal, N. Jehmlich, A. Tholey, M. von Bergen, J.L. Nielsen, R.L. Hettich, J. Seifert, P.H. Nielsen, Enhancing metaproteomics-The value of models and defined environmental microbial systems, Proteomics. (2016). doi:10.1002/pmic.201500305.

[36] F. Bastida, N. Jehmlich, K. Lima, B.E.L. Morris, H.H. Richnow, T. Hernández, M. von Bergen, C. García, The ecological and physiological responses of the microbial community from a semiarid soil to hydrocarbon contamination and its bioremediation using compost amendment, J. Proteomics. (2016). doi:10.1016/j.jprot.2015.07.023.

[37] M.C.Y. Lau, B.T. Stackhouse, A.C. Layton, A. Chauhan, T.A. Vishnivetskaya, K. Chourey, J. Ronholm, N.C.S. Mykytczuk, P.C. Bennett, G. Lamarche-Gagnon, N. Burton, W.H. Pollard, C.R. Omelon, D.M. Medvigy, R.L. Hettich, S.M. Pfiffner, L.G. Whyte, T.C. Onstott, An active atmospheric methane sink in high Arctic mineral cryosols, ISME J. (2015). doi:10.1038/ismej.2015.13.

[38] W. Lin, L. Wu, S. Lin, A. Zhang, M. Zhou, R. Lin, H. Wang, J. Chen, Z. Zhang, R. Lin, Metaproteomic analysis of ratoon sugarcane rhizospheric soil, BMC Microbiol. (2013). doi:10.1186/1471-2180-13-135.

[39] D. Liu, K.M. Keiblinger, A. Schindlbacher, U. Wegner, H. Sun, S. Fuchs, C. Lassek, K. Riedel, S. Zechmeister-Boltenstern, Microbial functionality as affected by experimental warming of a temperate mountain forest soil—A metaproteomics survey, Appl. Soil Ecol. (2017). doi:10.1016/j.apsoil.2017.04.021.

[40] T. Schneider, K.M. Keiblinger, E. Schmid, K. Sterflinger-Gleixner, G. Ellersdorfer, B. Roschitzki, A. Richter, L. Eberl, S. Zechmeister-Boltenstern, K. Riedel, Who is who in litter decomposition

Metaproteomics reveals major microbial players and their biogeochemical functions, ISME J. (2012). doi:10.1038/ismej.2012.11.

[41] R. Starke, F. Bastida, J. Abadía, C. García, E. Nicolás, N. Jehmlich, Ecological and functional adaptations to water management in a semiarid agroecosystem: A soil metaproteomics approach, Sci. Rep. 7 (2017). doi:10.1038/s41598-017-09973-w.

[42] R. Starke, R. Kermer, L. Ullmann-Zeunert, I.T. Baldwin, J. Seifert, F. Bastida, M. von Bergen, N. Jehmlich, Bacteria dominate the short-term assimilation of plant-derived N in soil, Soil Biol. Biochem. (2016). doi:10.1016/j.soilbio.2016.01.009.

[43] M.A. Williams, E.B. Taylor, H.P. Mula, Metaproteomic characterization of a soil microbial community following carbon amendment, Soil Biol. Biochem. (2010). doi:10.1016/j.soilbio.2010.03.021.

[44] L. Wu, H. Wang, Z. Zhang, R. Lin, Z. Zhang, W. Lin, Comparative metaproteomic analysis on consecutively Rehmannia glutinosa-monocultured rhizosphere soil, PLoS One. (2011). doi:10.1371/journal.pone.0020611.

[45] E. Zampieri, M. Chiapello, S. Daghino, P. Bonfante, A. Mello, Soil metaproteomics reveals an inter-kingdom stress response to the presence of black truffles, Sci. Rep. (2016). doi:10.1038/srep25773.

[46] F. Bastida, T. Hernández, C. García, Metaproteomics of soils from semiarid environment: Functional and phylogenetic information obtained with different protein extraction methods, J. Proteomics. (2014). doi:10.1016/j.jprot.2014.02.006.

[47] F. Bastida, N. Selevsek, I.F. Torres, T. Hernández, C. García, Soil restoration with organic amendments: Linking cellular functionality and ecosystem processes, Sci. Rep. (2015). doi:10.1038/srep15550.

[48] F. Bastida, I.F. Torres, J.L. Moreno, P. Baldrian, S. Ondoño, A. Ruiz-Navarro, T. Hernández, H.H. Richnow, R. Starke, C. García, N. Jehmlich, The active microbial diversity drives ecosystem multifunctionality and is physiologically related to carbon availability in Mediterranean semi-arid soils, Mol. Ecol. (2016). doi:10.1111/mec.13783.

[49] F. Bastida, C. García, M. von Bergen, J.L. Moreno, H.H. Richnow, N. Jehmlich, Deforestation fosters bacterial diversity and the cyanobacterial community responsible for carbon fixation processes under semiarid climate: A metaproteomics study, Appl. Soil Ecol. (2015). doi:10.1016/j.apsoil.2015.04.006.

[50] K. Chourey, J. Jansson, N. Verberkmoes, M. Shah, K.L. Chavarria, L.M. Tom, E.L. Brodie, R.L. Hettich, Direct cellular lysis/protein extraction protocol for soil metaproteomics, J. Proteome Res. (2010). doi:10.1021/pr100787q.

[51] K. Chourey, S. Nissen, T. Vishnivetskaya, M. Shah, S. Pfiffner, R.L. Hettich, F.E. Löffler, Environmental proteomics reveals early microbial community responses to biostimulation at a uranium- and nitrate-contaminated site, Proteomics. (2013). doi:10.1002/pmic.201300155.

[52] M.E. Guazzaroni, F.A. Herbst, I. Lores, J. Tamames, A.I. Peláez, N. López-Cortés, M. Alcaide, M. V. Del Pozo, J.M. Vieites, M. Von Bergen, J.L.R. Gallego, R. Bargiela, A. López-López, D.H. Pieper, R. Rosselló-Móra, J. Sánchez, J. Seifert, M. Ferrer, Metaproteogenomic insights beyond bacterial response to naphthalene exposure and bio-stimulation, ISME J. (2013).

502 doi:10.1038/ismej.2012.82.

503 [53] J. Hultman, M.P. Waldrop, R. Mackelprang, M.M. David, J. McFarland, S.J. Blazewicz, J. Harden,
504 M.R. Turetsky, A.D. McGuire, M.B. Shah, N.C. VerBerkmoes, L.H. Lee, K. Mavrommatis, J.K.
505 Jansson, Multi-omics of permafrost, active layer and thermokarst bog soil microbiomes, Nature.
506 (2015). doi:10.1038/nature14238.

507 [54] F. Bastida, I.F. Torres, M. Andrés-Abellán, P. Baldrian, R. López-Mondéjar, T. Větrovský, H.H.
508 Richnow, R. Starke, S. Ondoño, C. García, F.R. López-Serrano, N. Jehmlich, Differential sensitivity
509 of total and active soil microbial communities to drought and forest management, Glob. Chang.
510 Biol. 23 (2017). doi:10.1111/gcb.13790.

511 [55] M. Tuomi, P. Vanhala, K. Karhu, H. Fritze, J. Liski, Heterotrophic soil respiration-Comparison of
512 different models describing its temperature dependence, Ecol. Modell. (2008).
513 doi:10.1016/j.ecolmodel.2007.09.003.

514 [56] S.E. Hannula, H.T.S. Boschker, W. de Boer, J.A. van Veen, 13C pulse-labeling assessment of the
515 community structure of active fungi in the rhizosphere of a genetically starch-modified potato
516 (Solanum tuberosum) cultivar and its parental isoline, New Phytol. (2012). doi:10.1111/j.1469-
517 8137.2012.04089.x.

518 [57] I.F. Torres, F. Bastida, T. Hernández, P. Bombach, H.H. Richnow, C. García, The role of lignin and
519 cellulose in the carbon-cycling of degraded soils under semiarid climate and their relation to
520 microbial biomass, Soil Biol. Biochem. (2014). doi:10.1016/j.soilbio.2014.04.007.

521 [58] V. Lünsmann, U. Kappelmeyer, A. Taubert, I. Nijenhuis, M. Von Bergen, H.J. Heipieper, J.A.
522 Müller, N. Jehmlich, Aerobic toluene degraders in the rhizosphere of a constructed wetland
523 model show diurnal polyhydroxyalkanoate metabolism, Appl. Environ. Microbiol. (2016).
524 doi:10.1128/AEM.00493-16.

525 [59] J. Lloyd, J.A. Taylor, On the Temperature Dependence of Soil Respiration, Funct. Ecol. (1994).
526 doi:10.2307/2389824.

527 [60] M.A. Bradford, W.R. Wieder, G.B. Bonan, N. Fierer, P.A. Raymond, T.W. Crowther, Managing
528 uncertainty in soil carbon feedbacks to climate change, Nat. Clim. Chang. (2016).
529 doi:10.1038/nclimate3071.

530 [61] B. Bond-Lamberty, V.L. Bailey, M. Chen, C.M. Gough, R. Vargas, Globally rising soil heterotrophic
531 respiration over recent decades, Nature. (2018). doi:10.1038/s41586-018-0358-x.

532 [62] L. Brussaard, P.C. de Ruiter, G.G. Brown, Soil biodiversity for agricultural sustainability, Agric.
533 Ecosyst. Environ. (2007). doi:10.1016/j.agee.2006.12.013.

534 [63] M. Mattarozzi, M. Manfredi, B. Montanini, F. Gosetti, A.M. Sanangelantoni, E. Marengo, M.
535 Careri, G. Visioli, A metaproteomic approach dissecting major bacterial functions in the
536 rhizosphere of plants living in serpentine soil, Anal. Bioanal. Chem. (2017). doi:10.1007/s00216-
537 016-0175-8.

538 [64] R. Manikandan, G. Karthikeyan, T. Raguchander, Soil proteomics for exploitation of microbial
539 diversity in Fusarium wilt infected and healthy rhizosphere soils of tomato, Physiol. Mol. Plant
540 Pathol. (2017). doi:10.1016/j.pmpp.2017.10.001.

541 [65] K.M. Keiblinger, I.C. Wilhartitz, T. Schneider, B. Roschitzki, E. Schmid, L. Eberl, K. Riedel, S.

Zechmeister-Boltenstern, Soil metaproteomics - Comparative evaluation of protein extraction protocols, Soil Biol. Biochem. (2012). doi:10.1016/j.soilbio.2012.05.014.

[66] O. V. Singh, N.S. Nagaraj, Transcriptomics, proteomics and interactomics: Unique approaches to track the insights of bioremediation, Briefings Funct. Genomics Proteomics. (2006). doi:10.1093/bfgp/eli006.

[67] L. Käll, J.D. Canterbury, J. Weston, W.S. Noble, M.J. MacCoss, Semi-supervised learning for peptide identification from shotgun proteomics datasets, Nat. Methods. (2007). doi:10.1038/nmeth1113.

[68] T. Schneider, E. Schmid, J. V. de Castro, M. Cardinale, L. Eberl, M. Grube, G. Berg, K. Riedel, Structure and function of the symbiosis partners of the lung lichen (Lobaria pulmonaria L. Hoffm.) analyzed by metaproteomics, Proteomics. (2011). doi:10.1002/pmic.201000679.

[69] B. Zybailov, A.L. Mosley, M.E. Sardiu, M.K. Coleman, L. Florens, M.P. Washburn, Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae, J. Proteome Res. (2006). doi:10.1021/pr060161n.

[70] J.A. Vizcaíno, E.W. Deutsch, R. Wang, A. Csordas, F. Reisinger, D. Ríos, J.A. Dianes, Z. Sun, T. Farrah, N. Bandeira, P.A. Binz, I. Xenarios, M. Eisenacher, G. Mayer, L. Gatto, A. Campos, R.J. Chalkley, H.J. Kraus, J.P. Albar, S. Martinez-Bartolomé, R. Apweiler, G.S. Omenn, L. Martens, A.R. Jones, H. Hermjakob, ProteomeXchange provides globally coordinated proteomics data submission and dissemination, Nat. Biotechnol. (2014). doi:10.1038/nbt.2839.

[71] T. Muth, B.Y. Renard, L. Martens, Metaproteomic data analysis at a glance: advances in computational microbial community proteomics, Expert Rev. Proteomics. (2016). doi:10.1080/14789450.2016.1209418.

[72] R. Heyer, K. Schallert, R. Zoun, B. Becher, G. Saake, D. Benndorf, Challenges and perspectives of metaproteomic data analysis, J. Biotechnol. (2017). doi:10.1016/j.jbiotec.2017.06.1201.

[73] L. Giagnoni, F. Magherini, L. Landi, S. Taghavi, A. Modesti, L. Bini, P. Nannipieri, D. Van Der Lelie, G. Renella, Extraction of microbial proteome from soil: Potential and limitations assessed through a model study, Eur. J. Soil Sci. (2011). doi:10.1111/j.1365-2389.2010.01322.x.

[74] L. Giagnoni, A. Migliaccio, P. Nannipieri, G. Renella, High montmorillonite content may affect soil microbial proteomic analysis, Appl. Soil Ecol. (2013). doi:10.1016/j.apsoil.2013.07.010.

[75] M. Arenella, L. Giagnoni, G. Masciandaro, B. Ceccanti, P. Nannipieri, G. Renella, Interactions between proteins and humic substances affect protein identification by mass spectrometry, Biol. Fertil. Soils. (2014). doi:10.1007/s00374-013-0860-0.

[76] J.E. Tomaszewski, R.P. Schwarzenbach, M. Sander, Protein encapsulation by humic substances, Environ. Sci. Technol. (2011). doi:10.1021/es200663h.

[77] A. Piccolo, The supramolecular structure of humic substances, Soil Sci. (2001). doi:10.1097/00010694-200111000-00007.

[78] P. Čapek, P. Kotas, S. Manzoni, H. Šantrůčková, Drivers of phosphorus limitation across soil microbial communities, Funct. Ecol. (2016). doi:10.1111/1365-2435.12650.

[79] A. Beck, K. Hunt, R. Carlson, Measuring Cellular Biomass Composition for Computational Biology

Applications, Processes. (2018). doi:10.3390/pr6050038.

[80] F. Bastida, N. Jehmlich, I.F. Torres, C. García, The extracellular metaproteome of soils under semiarid climate: A methodological comparison of extraction buffers, Sci. Total Environ. (2018). doi:10.1016/j.scitotenv.2017.11.134.

[81] M.M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, Anal. Biochem. (1976). doi:10.1016/0003-2697(76)90527-3.

[82] P.K. Smith, R.I. Krohn, G.T. Hermanson, A.K. Mallia, F.H. Gartner, M.D. Provenzano, E.K. Fujimoto, N.M. Goeke, B.J. Olson, D.C. Klenk, Measurement of protein using bicinchoninic acid, Anal. Biochem. (1985). doi:10.1016/0003-2697(85)90442-7.

[83] J. Schulz, S. Dettlaff, U. Fritzsche, U. Harms, H. Schiebel, W. Derer, N.E. Fusenig, A. Hülsen, M. Böhm, The amido black assay: a simple and quantitative multipurpose test of adhesion, proliferation, and cytotoxicity in microplate cultures of keratinocytes (HaCaT) and other cell types growing adherently or in suspension, J. Immunol. Methods. (1994). doi:10.1016/0022-1759(94)90069-8.

[84] P. Kebarle, L. Tang, From Ions in Solution to Ions in the Gas Phase: The Mechanism of Electrospray Mass Spectrometry, Anal. Chem. (1993). doi:10.1021/ac00070a001.

[85] N.B. Cech, C.G. Enke, Practical implications of some recent studies in electrospray ionization fundamentals, Mass Spectrom. Rev. (2001). doi:10.1002/mas.10008.

[86] R.D. Smith, J.A. Loo, C.G. Edmonds, C.J. Barinaga, H.R. Udseth, New Developments in Biochemical Mass Spectrometry: Electrospray Ionization, Anal. Chem. (1990). doi:10.1021/ac00208a002.

[87] D.N. Perkins, D.J.C. Pappin, D.M. Creasy, J.S. Cottrell, Probability-based protein identification by searching sequence databases using mass spectrometry data, Electrophoresis. (1999). doi:10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2.

[88] T. Koenig, B.H. Menze, M. Kirchner, F. Monigatti, K.C. Parker, T. Patterson, J.J. Steen, F.A. Hamprecht, H. Steen, Robust prediction of the MASCOT score for an improved quality assessment in mass spectrometric proteomics, J. Proteome Res. (2008). doi:10.1021/pr700859x.

[89] J.K. Eng, A.L. McCormack, J.R. Yates, An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database, J. Am. Soc. Mass Spectrom. (1994). doi:10.1016/1044-0305(94)80016-2.

[90] A. Bateman, M.J. Martin, C. O'Donovan, M. Magrane, E. Alpi, R. Antunes, B. Bely, M. Bingley, C. Bonilla, R. Britto, B. Bursteinas, H. Bye-AJee, A. Cowley, A. Da Silva, M. De Giorgi, T. Dogan, F. Fazzini, L.G. Castro, L. Figueira, P. Garmiri, G. Georghiou, D. Gonzalez, E. Hatton-Ellis, W. Li, W. Liu, R. Lopez, J. Luo, Y. Lussi, A. MacDougall, A. Nightingale, B. Palka, K. Pichler, D. Poggioli, S. Pundir, L. Pureza, G. Qi, S. Rosanoff, R. Saidi, T. Sawford, A. Shypitsyna, E. Speretta, E. Turner, N. Tyagi, V. Volynkin, T. Wardell, K. Warner, X. Watkins, R. Zaru, H. Zellner, I. Xenarios, L. Bougueleret, A. Bridge, S. Poux, N. Redaschi, L. Aimo, G. ArgoudPuy, A. Auchincloss, K. Axelsen, P. Bansal, D. Baratin, M.C. Blatter, B. Boeckmann, J. Bolleman, E. Boutet, L. Breuza, C. Casal-Casas, E. De Castro, E. Coudert, B. Cuche, M. Doche, D. Dornevil, S. Duvaud, A. Estreicher, L. Famiglietti, M. Feuermann, E. Gasteiger, S. Gehant, V. Gerritsen, A. Gos, N. Gruaz-Gumowski, U. Hinz, C. Hulo, F. Jungo, G. Keller, V. Lara, P. Lemercier, D. Lieberherr, T. Lombardot, X. Martin, P. Masson,

A. Morgat, T. Neto, N. Nouspikel, S. Paesano, I. Pedruzzi, S. Pilbout, M. Pozzato, M. Pruess, C. Rivoire, B. Roechert, M. Schneider, C. Sigrist, K. Sonesson, S. Staehli, A. Stutz, S. Sundaram, M. Tognolli, L. Verbregue, A.L. Veuthey, C.H. Wu, C.N. Arighi, L. Arminski, C. Chen, Y. Chen, J.S. Garavelli, H. Huang, K. Laiho, P. McGarvey, D.A. Natale, K. Ross, C.R. Vinayaka, Q. Wang, Y. Wang, L.S. Yeh, J. Zhang, UniProt: The universal protein knowledgebase, Nucleic Acids Res. (2017). doi:10.1093/nar/gkw1099.

[91] S. Anderson, Shotgun DNA sequencing using cloned DNase I-generated fragments, Nucleic Acids Res. (1981). doi:10.1093/nar/9.13.3015.

[92] R. Staden, A strategy of DNA sequencing employing computer programs, Nucleic Acids Res. (1979). doi:10.1093/nar/6.7.2601.

[93] T. Human, G. Project, Human Genome Project, Genome. (2001). doi:10.1080/19371918.2011.579488.

[94] E.W. Deutsch, A. Csordas, Z. Sun, A. Jarnuczak, Y. Perez-Riverol, T. Ternent, D.S. Campbell, M. Bernal-Llinares, S. Okuda, S. Kawano, R.L. Moritz, J.J. Carver, M. Wang, Y. Ishihama, N. Bandeira, H. Hermjakob, J.A. Vizcaíno, The ProteomeXchange consortium in 2017: Supporting the cultural change in proteomics public data deposition, Nucleic Acids Res. (2017). doi:10.1093/nar/gkw936.

[95] R.G. Côté, J. Griss, J.A. Dianes, R. Wang, J.C. Wright, H.W.P. van den Toorn, B. van Breukelen, A.J.R. Heck, N. Hulstaert, L. Martens, F. Reisinger, A. Csordas, D. Ovelleiro, Y. Perez-Rivevol, H. Barsnes, H. Hermjakob, J.A. Vizcaíno, The PRoteomics IDEntification (PRIDE) Converter 2 Framework: An Improved Suite of Tools to Facilitate Data Submission to the PRIDE Database and the ProteomeXchange Consortium, Mol. Cell. Proteomics. (2012). doi:10.1074/mcp.O112.021543.

[96] H. Hermjakob, R. Apweiler, The proteomics identifications database (PRIDE) and the proteomExchange consortium: Making proteomics data accessible, in: Expert Rev. Proteomics, 2006. doi:10.1586/14789450.3.1.1.

[97] Y. Perez-Riverol, Q.-W. Xu, R. Wang, J. Uszkoreit, J. Griss, A. Sanchez, F. Reisinger, A. Csordas, T. Ternent, N. del-Toro, J.A. Dianes, M. Eisenacher, H. Hermjakob, J.A. Vizcaíno, PRIDE Inspector Toolsuite: Moving Toward a Universal Visualization Tool for Proteomics Data Standard Formats and Quality Assessment of ProteomeXchange Datasets, Mol. Cell. Proteomics. (2016). doi:10.1074/mcp.O115.050229.

[98] T. Ternent, A. Csordas, D. Qi, G. Gómez-Baena, R.J. Beynon, A.R. Jones, H. Hermjakob, J.A. Vizcaíno, How to submit MS proteomics data to ProteomeXchange via the PRIDE database, Proteomics. (2014). doi:10.1002/pmic.201400120.

[99] P. Picotti, R. Aebersold, B. Domon, The Implications of Proteolytic Background for Shotgun Proteomics, Mol. Cell. Proteomics. (2007). doi:10.1074/mcp.M700029-MCP200.

[100] B. Domon, S. Broder, Implications of New Proteomics Strategies for Biology and Medicine, J. Proteome Res. (2004). doi:10.1021/pr034082c.

[101] L. Anderson, C.L. Hunter, Quantitative Mass Spectrometric Multiple Reaction Monitoring Assays for Major Plasma Proteins, Mol. Cell. Proteomics. (2006). doi:10.1074/mcp.M500331-MCP200.

[102] V. Lange, J. a Malmström, J. Didion, N.L. King, B.P. Johansson, J. Schäfer, J. Rameseder, C.-H. Wong, E.W. Deutsch, M.-Y. Brusniak, P. Bühlmann, L. Björck, B. Domon, R. Aebersold, Targeted

quantitative analysis of Streptococcus pyogenes virulence factors by multiple reaction monitoring., Mol. Cell. Proteomics. (2008). doi:10.1074/mcp.M800032-MCP200.

[103] P. Picotti, B. Bodenmiller, L.N. Mueller, B. Domon, R. Aebersold, Full Dynamic Range Proteome Analysis of S. cerevisiae by Targeted Proteomics, Cell. (2009). doi:10.1016/j.cell.2009.05.051.

[104] N. Jehmlich, F. Schmidt, M. Von Bergen, H.H. Richnow, C. Vogt, Protein-based stable isotope probing (Protein-SIP) reveals active species within anoxic mixed cultures, ISME J. (2008). doi:10.1038/ismej.2008.64.

[105] M.W. Senko, S.C. Beu, F.W. McLaffertycor, Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions, J. Am. Soc. Mass Spectrom. (1995). doi:10.1016/1044-0305(95)00017-8.

[106] T. Sachsenberg, F.A. Herbst, M. Taubert, R. Kermer, N. Jehmlich, M. Von Bergen, J. Seifert, O. Kohlbacher, MetaProSIP: Automated inference of stable isotope incorporation rates in proteins for functional metaproteomics, J. Proteome Res. (2015). doi:10.1021/pr500245w.

[107] M. Taubert, C. Vogt, T. Wubet, S. Kleinsteuber, M.T. Tarkka, H. Harms, F. Buscot, H.H. Richnow, M. Von Bergen, J. Seifert, Protein-SIP enables time-resolved analysis of the carbon flux in a sulfate-reducing, benzene-degrading microbial consortium, ISME J. (2012). doi:10.1038/ismej.2012.68.

[108] N. Jehmlich, F. Schmidt, M. Hartwich, M. Von Bergen, H.H. Richnow, C. Vogt, Incorporation of carbon and nitrogen atoms into proteins measured by protein-based stable isotope probing (Protein-SIP), Rapid Commun. Mass Spectrom. (2008). doi:10.1002/rcm.3684.

[109] N. Jehmlich, C. Vogt, V. Lünsmann, H.H. Richnow, M. von Bergen, Protein-SIP in environmental studies, Curr. Opin. Biotechnol. (2016). doi:10.1016/j.copbio.2016.04.010.

[110] S.W. Englander, N.R. Kallenbach, Hydrogen exchange and structural dynamics of proteins and nucleic acids, Q. Rev. Biophys. (1983). doi:10.1017/S0033583500005217.

[111] W. Bild, V. Nastasa, I. Haulica, In vivo and in vitro research on the biological effects of deuterium-depleted water: 1. Influence of deuterium-depleted water on cultured cell growth, Rom. J. Physiol. Physiol. Sci. (2004) 53–67.

[112] O.B. Mosin, D.A. Skladnev, T.A. Egorova, V.I. Shvets, Biological Effects of Heavy Water, Bioorg. Chem. (1996) 861–874.

[113] D.A. Skladnev, O. V Mosin, T.A. Egorova, S. V Eremin, V.I. Shvets, Methylotrophic Bacteria as Sources of 2H-and 13C-amino Acids, Biotechnology. (1996) 14–22.

[114] P.J. Boersema, R. Raijmakers, S. Lemeer, S. Mohammed, A.J.R. Heck, Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics, Nat. Protoc. (2009). doi:10.1038/nprot.2009.21.

[115] R. Zhang, C.S. Sioma, R.A. Thompson, L. Xiong, F.E. Regnier, Controlling deuterium isotope effects in comparative proteomics, Anal. Chem. (2002). doi:10.1021/ac025614w.

[116] C. Pan, C.R. Fischer, D. Hyatt, B.P. Bowen, R.L. Hettich, J.F. Banfield, Quantitative Tracking of Isotope Flows in Proteomes of Microbial Communities, Mol. Cell. Proteomics. (2011). doi:10.1074/mcp.M110.006049.

[117] X. Ye, B. Luke, T. Andresson, J. Blonder, 18O stable isotope labeling in MS-based proteomics, Briefings Funct. Genomics Proteomics. (2009). doi:10.1093/bfgp/eln055.

[118] N. Jehmlich, F.-D. Kopinke, S. Lenhard, C. Vogt, F.-A. Herbst, J. Seifert, U. Lissner, U. Völker, F. Schmidt, M. von Bergen, Sulfur-36S stable isotope labeling of amino acids for quantification (SULAQ), Proteomics. (2012). doi:10.1002/pmic.201100057.

[119] N. Fierer, Embracing the unknown: Disentangling the complexities of the soil microbiome, Nat. Rev. Microbiol. (2017). doi:10.1038/nrmicro.2017.87.

[120] Y. Kuzyakov, E. Blagodatskaya, Microbial hotspots and hot moments in soil: Concept & review, Soil Biol. Biochem. (2015). doi:10.1016/j.soilbio.2015.01.025.

[121] N.B. Justice, Z. Li, Y. Wang, S.E. Spaudling, A.C. Mosier, R.L. Hettich, C. Pan, J.F. Banfield, 15N-and2H proteomic stable isotope probing links nitrogen flow to archaeal heterotrophic activity, Environ. Microbiol. (2014). doi:10.1111/1462-2920.12488.

[122] S.J. Blazewicz, E. Schwartz, Dynamics of18O Incorporation from H218O into Soil Microbial DNA, Microb. Ecol. (2011). doi:10.1007/s00248-011-9826-7.

[123] E. Schwartz, Analyzing microorganisms in environmental samples using stable isotope probing with H218O, Cold Spring Harb. Protoc. 4 (2009).

[124] E. Schwartz, Characterization of growing microorganisms in soil by stable isotope probing with H218O, Appl. Environ. Microbiol. (2007). doi:10.1128/AEM.02021-06.

[125] E.A. Rettedal, V.S. Brözel, Characterizing the diversity of active bacteria in soil by comprehensive stable isotope probing of DNA and RNA with H218O, Microbiologyopen. (2015). doi:10.1002/mbo3.230.

[126] R. Angel, R. Conrad, Elucidating the microbial resuscitation cascade in biological soil crusts following a simulated rain event, Environ. Microbiol. (2013). doi:10.1111/1462-2920.12140.

[127] K. Papp, R.L. Mau, M. Hayer, B.J. Koch, B.A. Hungate, E. Schwartz, Quantitative stable isotope probing with H218O reveals that most bacterial taxa in soil synthesize new ribosomal RNA, ISME J. (2018). doi:10.1038/s41396-018-0233-7.

[128] F. Bastida, N. Jehmlich, It's all about functionality: How can metaproteomics help us to discuss the attributes of ecological relevance in soil?, J. Proteomics. (2016). doi:10.1016/j.jprot.2016.06.002.

[129] B.S. Griffiths, L. Philippot, Insights into the resistance and resilience of the soil microbial community, FEMS Microbiol. Rev. (2013). doi:10.1111/j.1574-6976.2012.00343.x.

[130] M. Bahram, F. Hildebrand, S.K. Forslund, J.L. Anderson, N.A. Soudzilovskaia, P.M. Bodegom, J. Bengtsson-Palme, S. Anslan, L.P. Coelho, H. Harend, J. Huerta-Cepas, M.H. Medema, M.R. Maltz, S. Mundra, P.A. Olsson, M. Pent, S. Põlme, S. Sunagawa, M. Ryberg, L. Tedersoo, P. Bork, Structure and function of the global topsoil microbiome, Nature. (2018). doi:10.1038/s41586-018-0386-6.

[131] F.T. de Vries, R.I. Griffiths, M. Bailey, H. Craig, M. Girlanda, H.S. Gweon, S. Hallin, A. Kaisermann, A.M. Keith, M. Kretzschmar, P. Lemanceau, E. Lumini, K.E. Mason, A. Oliver, N. Ostle, J.I. Prosser, C. Thion, B. Thomson, R.D. Bardgett, Soil bacterial networks are less stable under drought than

741 fungal networks, Nat. Commun. (2018). doi:10.1038/s41467-018-05516-7.

742 [132] D. Benndorf, G.U. Balcke, H. Harms, M. von Bergen, Functional metaproteome analysis of protein
743 extracts from contaminated soil and groundwater, ISME J. (2007). doi:10.1038/ismej.2007.39.

744 [133] Y.Q. Yu, M. Gilar, P.J. Lee, E.S.P. Bouvier, J.C. Gebler, Enzyme-Friendly, Mass Spectrometry-
745 Compatible Surfactant for In-Solution Enzymatic Digestion of Proteins, Anal. Chem. (2003).
746 doi:10.1021/ac0346196.

747 [134] J. Rosenfeld, J. Capdevielle, J.C. Guillemot, P. Ferrara, In-gel digestion of proteins for internal
748 sequence analysis after one- or two-dimensional gel electrophoresis, Anal. Biochem. (1992).
749 doi:10.1016/0003-2697(92)90061-B.

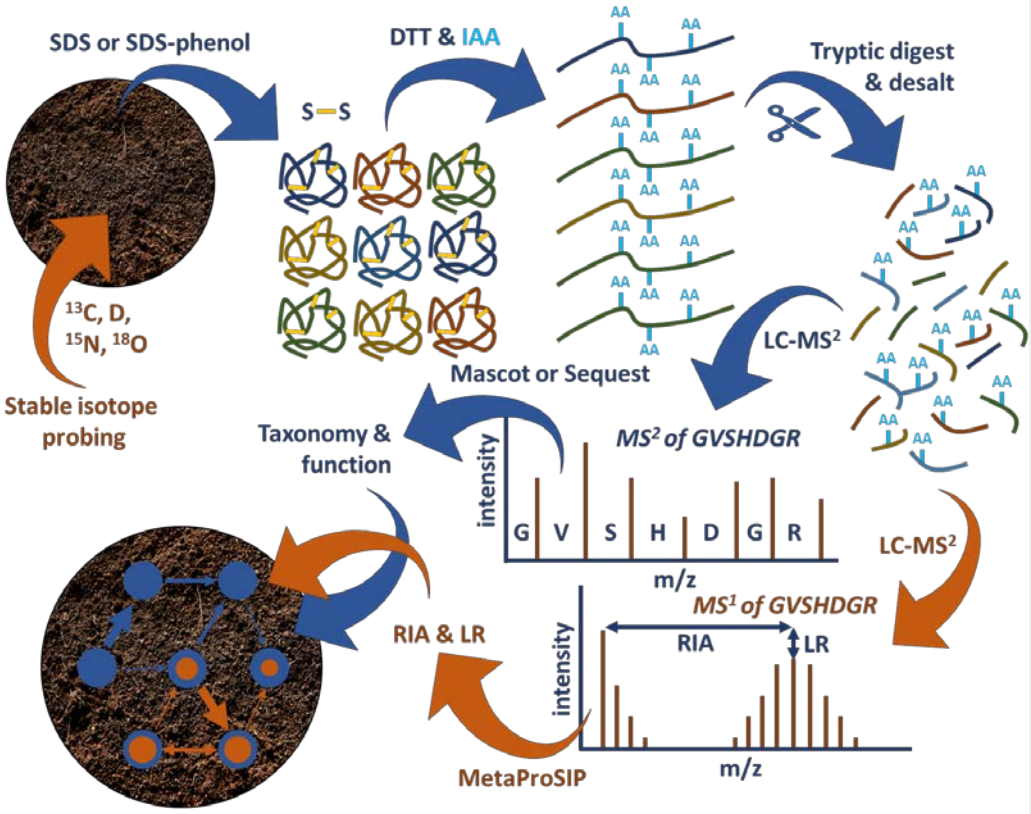
750 [135] B. Herbert, M. Galvani, M. Hamdan, E. Olivieri, J. MacCarthy, S. Pedersen, P.G. Righetti, Reduction
751 of alkylation of proteins in preparation of two-dimensional map analysis: Why, when, and how?,
752 Electrophoresis. (2001). doi:10.1002/1522-2683(200106)22:10<2046::AID-ELPS2046>3.0.CO;2-C.

753 [136] J. Rappsilber, M. Mann, Y. Ishihama, Protocol for micro-purification, enrichment, pre-
754 fractionation and storage of peptides for proteomics using StageTips, Nat. Protoc. (2007).
755 doi:10.1038/nprot.2007.261.

756 [137] J. Villén, S.P. Gygi, The SCX/IMAC enrichment approach for global phosphorylation analysis by
757 mass spectrometry, Nat. Protoc. (2008). doi:10.1038/nprot.2008.150.

758

759 **Figures**



760
761 **Figure 1:** The workflow of metaproteomics (blue) and protein stable isotope probing (orange) of soil samples.
762 Proteins are commonly extracted with sodium dodecyl sulphate (SDS) [65] or SDS-phenol [50], before in-solution
763 [133] or in-gel [134] tryptic digestion; sulphur bonds of proteins (S-S) are reduced with dithiothreitol (DTT) and
764 alkylated (shown as -AA) with iodoacetamide (IAA) [135]. Tryptic peptides are desalted using strong cation
765 exchange (SCX) columns [136,137] or ZipTip® pipette tips prior to mass spectrometric measurement (LC-MS^2). MS^2
766 is used to identify the sequence of the peptide for database matching, with MASCOT [87,88] or SEQUEST [89] as
767 the search algorithm. MS^1 is used for quantification as the normalised spectral abundance factor (NSAF) [69] and
768 stable isotope probing yields the relative isotope abundance (RIA) and labelling ratio (LR), estimated by
769 MetaProSIP [106]. Taxonomy and function are assigned via PROPHANE [68]. The raw data are finally uploaded to
770 ProteomeXchange via PRIDE [70].

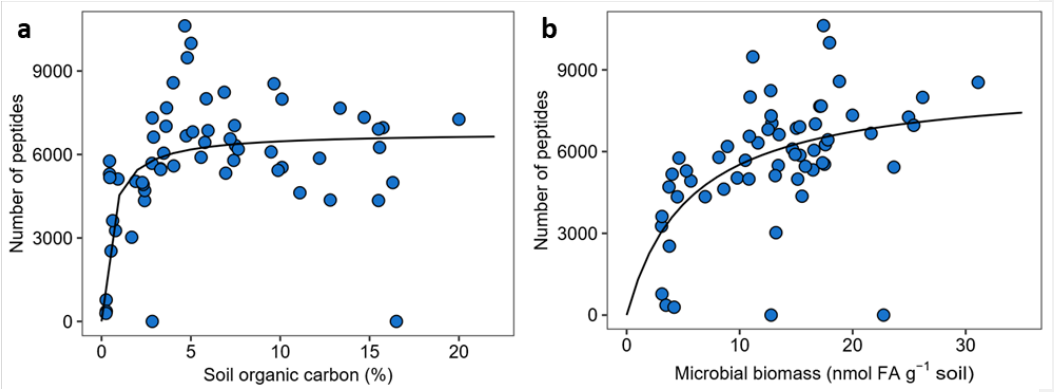


Figure 2: The relationship between soil organic carbon and the number of peptides. V_{max} was 6791.558±404.841 (P<2e-16) and K_m was 0.496±0.201 (P=0.016) (a). The breaking point was at a K_m of 1.48695, with 95% confidence intervals of 0.647 and 3.044. The relationship between microbial biomass and the number of peptides showed a V_{max} of 8620.149±1122.251 (P=2.1e-10) and a K_m of 5.609±2.372 (P=0.0214) (b). The breaking point was at a K_m of 16.827, with 95% confidence intervals of 7.087 and 37.799. The data were obtained from a gradient of semi-arid soils that covered a range of soil organic carbon concentration and microbial biomass estimated through the phospholipid fatty acids (PLFAs) content [48].

Acknowledgements

The authors are grateful to Dr. David Walker for English revision. RS thanks the Czech Science Foundation for the project 18-25706S. FB is grateful to the Spanish Ministry for the CICYT project AGL2017-85755, and the CSIC project 201740I008.

Congratulations to Journal of Proteomics' 10th Anniversary

The authors of this review would like to congratulate “Journal of Proteomics” on its 10th Anniversary. We are working in soil metaproteomics since a few years ago and it is always convenient to have a forum with recognized expertise where our research can be properly understood. This is particularly true when considering complex biological samples such as soils that require technical and bioinformatic developments. We are grateful to “Journal of Proteomics” for publishing some of our studies that are contributing to understand soil microbial ecology beyond genomics and keep us motivated to work further in the topic.