**This is a pre- or post-print of an article published in
Becher, D., Bernhardt, J., Fuchs, S., Riedel, K.
Metaproteomics to unravel major microbial players in leaf
litter and soil environments: Challenges and perspectives
(2013) Proteomics, 13 (18-19), pp. 2895-2909.**

Review Article

Metaproteomics to Unravel Major Microbial Players in Leaf Litter & Soil Environments: Challenges & Perspectives

Dörte Becher¹, Jörg Bernhardt¹, Stephan Fuchs¹, and Katharina Riedel^{1,2*}

¹ Ernst-Moritz-Arndt-University of Greifswald, Institute of Microbiology, Friedrich-Ludwig-Jahnstrasse 15, D-17487 Greifswald

² Helmholtz-Centre for Infection Research, Inhoffenstrasse 7, D-38124 Braunschweig

* Correspondence: Katharina Riedel, Friedrich-Ludwig-Jahnstrasse 15, D-17487 Greifswald, Phone: +49 3834 86 4233, Fax: +49 3834 86 4202, E-mail: riedela@uni-greifswald.de

Abbreviations: 1-DE, one-dimensional gel electrophoresis; 2-DE, two-dimensional gel electrophoresis; TCA, trichloroacetic acid; PVPP; polyvinylpolypyrrolidone

Key words: metaproteomics, soil, rhizosphere, bioremediation, leaf litter, decomposition

Abstract

Soil and litter-born microorganisms vitally contribute to biogeochemical cycles. However, changes in environmental parameters but also human interferences may alter species composition and elicit alterations in microbial activities. Soil and litter metaproteomics, implying the assignment of soil and litter proteins to specific phylogenetic and functional groups, has a great potential to provide essential new insights into the impact of microbial diversity on soil ecosystem functioning. This article will illuminate challenges and perspectives of current soil and litter metaproteomics research, starting with an introduction into an appropriate experimental design and *state-of-the-art* proteomics methodologies. This will be followed by a summary of important studies aiming on (i) the discovery of the major biotic drivers of leaf litter decomposition, (ii) metaproteomics analyses of rhizosphere-inhabiting microbes, and (iii) global approaches to study bioremediation processes. The review will be closed by a brief outlook on future developments and some concluding remarks, which should assist the reader to develop successful concepts for soil and litter metaproteomics studies.

1 Introduction

Terrestrial natural environments such as soil but also leaf litter harbour highly complex living communities and are thus characterized by interactive processes such as nutrient competition, synergistic interdependencies, predation and intra- and interspecies signalling. Microbial residents of these habitats, i.e. bacteria and fungi, play a major role in the degradation of virtually all natural (but also man-made) compounds, thereby determining biogeochemical cycles and exerting a lasting effect on biosphere and climate. In particular, the activity of microbes involved in leaf litter decomposition has been recognized to be of utmost importance as a large pool of organic carbon, nitrogen and phosphorus, immobilized in 10^{11} tons of yearly fallen leaf litter, has to be reconverted to plant-accessible nutrients. The global mean steady state turnover time of this process has been estimated to last 1.4 to 3.4 years [1]. Also, the root system of most land plants is intimately associated with soil microbes, e.g. mycorrhizal arbuscular fungi providing an interface between plants and soil [2] or soil-borne bacteria supporting nutrient acquisition, protecting against plant pathogens, and enhancing plant growth (for review see [3]). Studies aiding the identification of central biotic drivers in soil and litter will therefore improve our understanding how biodiversity and functionality of these ecosystems are related [4] and may also facilitate human intervention in environmental key processes. However, it is well recognized that the microbiota of soil and to a lesser extent also leaf litter supports an extremely high biodiversity; a few grams of soil might contain thousands of microbial and hundreds of invertebrate species [5, 6]. Moreover, at least 90% of all microbial species present in soil (and litter) are “un-cultivable” [7]. Consequently, soil has to be regarded as one of the most challenging environments for microbial ecologists [8], which might explain why contributions of individual species to soil and litter environmental processes remained so far largely unknown.

In the last two decades, a multitude of novel molecular tools bypassing the need to isolate and cultivate individual microbial species have been developed, which have the potential to revolutionize our current ecological concepts including those related to soil environments [9]. A crucial step towards the culture-independent assessment of soil

community composition was the introduction of the 16S rRNA gene sequencing approach (for review see [10]). Today, the strategy to sequence single conserved genes is increasingly replaced by novel shotgun sequencing techniques enabling the mapping of entire soil metagenomes (for review see [8, 11]). However, these approaches do not allow linking microbial diversity and functionality. Moreover, most microbes overcome environmental changes by altering their gene/protein expression profiles and to a much lesser extend by genomic rearrangements such as horizontal gene transfer. Therefore, alternative techniques, including a combination of fluorescent in situ hybridization and microautoradiography (CARD-FISH; [12, 13]) or stable isotope probing (SIP) technologies [14, 15] have been developed to investigate identities and specific metabolic functions of individual microbes in complex communities. Most recently, gene expression [16, 17] and protein profiling (for review see [18]) of soil and litter environments (designated as metatranscriptomics and metaproteomics, respectively) have evolved as powerful tools for a comprehensive and simultaneous analysis of community structure and functionality.

Metaproteomics allows the qualitative and quantitative assessment of the protein complement of an environment (e.g. soil and litter) at a given time point (for review see [19-21]) ranging from simple protein cataloguing to comparative and quantitative proteomics analyses, and even the determination of amino acid sequences and genotypes, e.g. by strain-resolved proteogenomics. Besides its great potential, metaproteomics, especially when dealing with highly complex environments such as soil and litter, is facing numerous challenges. This is mirrored by the fact, that the actual output of soil metaproteomics studies appears still limited. Major obstacles for such studies are: (i) the tremendous heterogeneity of soil and litter [22], (ii) the comparably little protein yield which can be obtained from soil and litter matrices compared to other environmental samples [18, 23], (iii) the wide range of protein abundance levels (iv) the (up-to-now) limited amount of genomic sequence information of the mostly uncultivable soil biota, and (v) the unavailability of appropriate bioinformatic tools to analyse, integrate and visualize comprehensive global datasets. It is, however, encouraging to note that constantly improving extraction methods alongside

advances in down-stream mass spectrometry, a potentially growing pool of genomic data and the rapid development of sophisticated software, capable of dealing with extensive datasets, might help to overcome the current limitations of soil and litter metaproteomics approaches [21].

This review will address the great potential, but also challenges and pitfalls implicated in current soil metaproteomics analyses. The article starts with a general description of an appropriate experimental design suited for soil and litter metaproteomics studies together with an overview on *state-of-the-art* metaproteomics technologies. The following chapters will summarize important publications in the field aiming on (i) deciphering major players in leaf litter decomposition, (ii) investigations on structure and function of rhizosphere-inhabiting microbes, and (iii) metaproteome analyses of bioremediating microbial communities. Finally, an outlook on future developments and concluding remarks will be given.

2 Soil and litter Metaproteomics - Experimental Design and *State-of-the-Art* Technologies

A soil and litter metaproteomics experiment typically comprises three basic steps: (i) sample preparation including protein extraction, purification and concentration, (ii) protein or peptide separation and mass spectrometry (MS) analysis, and (iii) peptide identification based on the obtained tandem mass spectrometry (MS/MS) data, followed by an assignment of the peptides to proteins in an appropriate database [21]. **Figure 1** summarizes methodological strategies employed in soil and litter metaproteomics studies. The following paragraphs give an overview on *state-of-the-art* proteomics techniques focussing on the requirements of soil or litter environments and will discuss weaknesses and strengths of different experimental strategies.

2.1 Protein Extraction Strategies

Protein extraction from soil and litter, albeit its high heterogeneity (see above), can be considered as one of the crucial steps in obtaining reliable and meaningful proteomics datasets [9, 24]. Generally, two strategies can be distinguished: (i) an indirect protein extraction from previously isolated or enriched soil or litter microbial cells or (ii) a direct protein extraction from the soil or litter matrix [25]. Whilst the indirect approach might circumvent problems arising from the interfering effects of humic substances or minerals [26], the latter strategy might result in a more complete protein recovery, due to minimization of artificial influences on the proteome profile during sample preparation [27, 28].

One of the major challenges during sample preparation is the purification of proteins from humic acids, which are often present in soil or litter in high concentrations and tend to be co-extracted [20]. Humic compounds are known to adsorb proteins [29], and thus to interfere with the extraction, purification [30], and quantification of proteins by colorimetric assays [31]. Moreover, they may disturb the separation of proteins by SDS-PAGE and are known to hamper the chromatographic separation of tryptic peptides prior mass spectrometry analysis [18, 27].

As soil and (to a lesser extend) also litter are characterized by an enormous heterogeneity with respect to texture, pH, electrical conductivity, carbon/nitrogen/phosphorus stoichiometry and, most importantly, humic acids (e.g. polyphenolic compounds) content, the development of an “universal” extraction protocol suited for all the different matrices is certainly impractical. **Table 1** summarizes different extraction strategies for soil and leaf litter together with the number of resolved (in case no MS analysis was performed) protein spots or (if applicable) assigned non-redundant proteins. The protocols, aiming on an enrichment of proteins from soil and litter and on the removal of disturbing compounds, are either based on (i) aqueous buffers often combined with mechanophysical treatments (e.g. bead beaters or sonification) [31, 32], or (ii) organic solvents (e.g. phenol) [27, 33-36], (iii) strong acids (e.g. hydrofluoric acid) [28] or bases (e.g. NaOH) [27], and (iv) detergents (e.g. SDS) [23, 24, 34, 37-39].

One of the most recent studies dealing with soil metaproteomics [25] aimed on a comprehensive evaluation of well-established soil protein extraction protocols. To this end the authors extracted proteins from two types of soil (forest soil and potting soil) by (i) grinding the samples with liquid nitrogen in presence of 10% polyvinylpolypyrrolidone (PVPP) to clean samples from interfering substances, (ii) protein extraction with either SDS, NaOH (described in [27]), or SDS-phenol with and without pre-purification steps (described in [24]), and (iii) protein precipitation with either TCA or ammonium acetate in methanol. After washing and resuspension, proteins were trypsin-digested and subsequently analyzed by two-dimensional liquid chromatography coupled to tandem mass spectrometry (2D-LC-MS/MS). The SDS-phenol based protocol without additional washing steps resulted in the highest number of assigned proteins for forest soil extracts (494 non-redundant proteins), while the SDS protocol proved best for the potting soil (237 non-redundant proteins) [25]. Hence, the efficiency of the tested protocols seemed to be strongly dependent on soil composition. Notably, the study also demonstrated a surprisingly small overlap of the protocols with respect to the assignment of proteins to different phylogenetic and function groups. This finding suggests that a combination of different extraction protocols may result in significantly higher metaproteome coverages than applying only one specific protocol [25]. In conclusion, the diversity of the protocols listed in **Table 1** and in particular the study of Keiblinger *et al.* [25] emphasizes the importance of optimizing extraction protocols every time when new types of soil or litter should be analyzed and depending on the biological question that has been asked.

After the successful extraction of soil and litter proteins the researcher is faced with extremely complex samples containing proteins from hundreds of different (micro)organisms in concentrations varying over more than six orders of magnitude. A specific challenge in the analysis of such samples is therefore the efficient prefractionation in a first step and the unambiguous identification of the proteins in a second step. As prefractionation methods usually 2D gel-based, 1D gel-based or gel-free methods are used.

2.2 Gel-based Metaproteome Analyses

During two-dimensional gel electrophoresis (2-DE) [40] proteins are first separated along a pH gradient by isoelectric focussing and then by a second separation according to their molecular mass (for review see [41]). Subsequently, the separated proteins are subjected to *in-gel* tryptic digestion and mass spectrometry analysis for identification. For a long time 2-DE was regarded as the “gold standard” of proteomics research and has, albeit rarely, also been employed in community proteomics [42-44] and even for soil metaproteome analyses [34, 35, 45, 46]. However, 2-DE suffers from numerous limitations, which are especially critical for the analyses of soil and litter metaproteomes: (i) protein extracts from soil or leaf litter often contain substances (such as humic acids or salts), which interfere with the sensible first dimension (i.e. the isoelectric focussing), (ii) proteins with extreme molecular masses and isoelectric points as well as lipophilic proteins are mostly lost [47, 48], (iii) the co-migration of proteins, which occurs likely in highly complex protein mixtures, hampers accurate identification and quantification, and (iv) the relatively low sensitivity as compared to other (2D-gel-free) approaches, resulting in the loss of low-abundant proteins and generally lower level of identifications [20]. Very recent technological developments provide higher resolution and a better reproducibility of protein separation in 2D gels [49] that might circumvent at least some of the above mentioned drawbacks and thus the future application of 2-DE in soil and litter metaproteomics cannot be excluded.

2.3 2D-gel-free Metaproteome Analyses & Mass Spectrometric Analysis

Most of the current metaproteome studies are either based on a pre-separation of proteins by one-dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis (1-DE) followed by tryptic digestion and liquid chromatography coupled to tandem mass spectrometry (GeLC-MS/MS) or the tryptic digestion in a first step followed by two-dimensional or multidimensional liquid chromatography coupled to tandem mass spectrometry (2D-LC-MS/MS or MudPIT). Both methods have been shown to deliver reliable

results for the investigation of highly complex protein samples originating from terrestrial, e.g. [23, 38, 50], or aquatic habitats, e.g. [51-53], but also from microbial communities in humans (microbiome), e.g. [54-56].

In our hands and as also shown by others [57, 58], however, the GeLC-MS/MS approach outperforms the 2D-LC-MS/MS approach and is especially suited for soil or leaf litter, which often contains high concentrations of disturbing compounds. Firstly, the 1-DE represents an additional, very efficient purification step that separates proteins e.g. from humic acids. Secondly, a higher variety of extraction buffers are applicable, as the proteins are also separated from the extraction buffer including salt; a further protein precipitation step to avoid any interference between the extraction buffer and the LC-system is thus not mandatory. Finally, during 1-DE a separation on the protein and not peptide level is carried out and separates highly from less abundant proteins, thereby increasing the probability to identify lower abundant proteins.

The rapid developments on the technological side are crucial for the growing impact of environmental metaproteome analyses. The introduction of ultra-high pressure liquid chromatography (UPLC) as well as modified stationary phases provided e.g. a significant improvement in resolution and sensitivity on the chromatographic side [59, 60]. Even more important was the rapid development of high resolution mass spectrometers as essential base for successful metaproteomics analyses. The first step during a mass spectrometric analysis is the determination of the accurate mass of all eluting peptides at a given time point. Depending on the signal intensity of the peptides, so called MS/MS spectra are generated by collision induced dissociation (CID), which are then compared with a reference database for peptide identification. With the introduction of fast scanning instruments [61-63] a crucial step towards meaningful analyses of highly complex samples has been made as more MS/MS spectra can be acquired in a limited time period, which is especially important for samples of highest complexity such as soil and litter samples. Important developments were especially made for Orbitrap-based MS instruments: due to an improved design and technology the MS/MS scan speed of the Orbitrap Velos has been doubled compared to the

former Orbitrap generation, i.e. the LTQ-Orbitrap XL. Moreover, the increased sensitivity of the Orbitrap Velos leads to MS/MS spectra with more signals and improved quality. All together these improvements lead to significantly more identifications especially when highly complex peptide mixtures are analyzed as exemplarily shown for a marine metaproteomics dataset in **Figure 2**.

A similar impact can certainly be expected from mass spectrometric analyses relying on “data-independent” acquisition, during which complex peptide mixtures are analyzed in an untargeted and thus unbiased mode of acquisition. Basically, two approaches for this method exist. The first one, MS^E, is a vendor-specific acquisition mode, which is characterized by alternating scan modes: overview scans targeting the intact peptide masses eluting from a reversed-phase column are constantly alternated with collision-induced dissociation scans, fragmenting all precursor ions simultaneously [64]. In the following bioinformatic post-processing step a reassignment of the elution profiles of peptide fragment masses to their respective precursors in the overview scan is carried out [65]. The second approach relies on mass spectrometric equipment being capable of fast acquisitions of larger m/z windows for triggering tandem mass spectra bypassing the decision step of selecting precursor masses suitable for fragmentation [66]. A further development of this method provides an ion count independent acquisition [67]. A more recent developed method, called SWATH™ MS, systematically queries sample sets for the presence and quantity of essentially any protein of interest. This MS approach is using the information available in fragment ion spectral libraries to generate complete fragment ion maps [68]. Interested readers find a comprehensive overview on recent developments in mass spectrometry methods in a recent review written by Bensimon *et al.* [69].

2.4 Data Analysis, Interpretation & Visualization

Identification: Tandem mass spectrometry determination of fragment ion masses, generated from specific parent ions, has emerged as most commonly used tool for protein assignments

of complex natural environments. The generated mass data are fed into Software packages such as MASCOT [70], SEQUEST [71], X-tandem [72], or MyriMatch [73], which employ sophisticated algorithms to assign them to one or several proteins in the database with a certain probability. Albeit highly suitable for metaproteome analyses of non-sequenced microbial communities for which no genomic data exist, *de-novo* sequencing of peptides with highly accurate mass spectrometers [74] allowing for protein identifications by BLAST searches, is still too expensive in time and labour to become routinely used.

Quantification. Despite its well-known drawbacks (see above) 2-DE has dominated quantitative proteomics for many years, but has to our knowledge never been used for quantitative soil and litter metaproteome analyses. Instead, 2D-gel free, label-free LC-MS-based technologies (for review see [75, 76]) have emerged as powerful tools for quantitative metaproteomics, which are either based on measuring signal intensities of peptide precursor ions or on counting fragment spectra of peptides. A crucial advantage of these approaches is their comparably high dynamic range, which is of particular importance when multifaceted and large protein changes within highly complex samples, such as soil or litter, have to be anticipated (for review see [77]). Although promising, these approaches still encounter certain limitations: (i) they are less accurate than 2D-PAGE or label-based quantitative technologies as samples have to be analysed separately and (ii) the analysis of complex datasets places enormous demands on computational power, but the development of suitable software tools still lags behind the advances in mass spectrometry [78, 79].

Interpretation. Metaproteomic data sets include hundreds of protein identifications and, thus, their interpretation is a challenging task. A powerful workflow should integrate (i) wise protein inference algorithms, (ii) high quality protein annotation data, and (iii) “functional binning”. Peptide to protein mapping can involve serious difficulties if there are several conserved proteins matching. This is a much discussed problem in Shotgun proteomics [80-82] but a universal solution has not been found yet. Thus, researchers have to decide

whether and how ambiguous identifications should be reduced. In general, MS data interpretation strongly depends on quality of protein annotation. If possible, information of expertly curated resources such as SwissProt or Refseq should be considered [83, 84]. Alternatively, functional data can be transferred from well-characterized sequence or Hidden Markov Model (HMM) profile collections using appropriate alignment tools (for an excellent review see [85]). These tools can also help to assign higher-level functional categories (such as *TIGR Roles* or *the SEED families*). Thus, the functional diversity can be reduced to an acceptable level. Tools supporting such integrative approaches such as *Prophane* or *Megan* in an automated manner can be highly beneficial for metaproteomic data interpretation [50, 54, 86].

Visualization. Briefly, two extremes on a scale of detailedness can be distinguished: (i) Intelligent condensation or dimensionality reduction for accessing the data as an entirety and (ii) detailed mapping of all or a subset of data for in depth analyses. Tools for (a) quality control and validation of the applied experimental techniques as well as (b) for interpreting the biologically relevant data are in use.

(a) The quality of a sample run is visualized on the separation level by using two- or three dimensional representations of (2D) gel [87] or LC-MS/MS data sets as shown in **Figure 2** and [88]. For benchmarking the reproducibility of a sample concerning proteome coverage, Euler or VENN diagrams [89], concerning measurement of protein quantities data distribution plots, e.g. histograms, density plots, box plots [90] and/or violin plots [91, 92] are suitable.

(b) For illustrating the compositions of environmental samples (available biota or functional gene/protein classes) on a qualitative level tables or trees are established, on a quantitative level traditional charts (bar charts, bubble charts, pie charts) are widely used. For considering also hierarchical information as known from gene functional classification schemes or phylogenetic trees so called squarified [93] or Voronoi treemaps [94, 95] have been established. Changes in protein quantities of a set of spatially or temporally related samples

are traditionally displayed in line/bar graphs [90] or parallel plots [96]. For the simultaneous display of large data sets heat maps should be applied [97]. More specialized tools considering hierarchical taxonomic as well as gene functional data or results from hierarchical clustering approaches (e. g. k-means) use clustered heat maps [98] or stream graphs [99].

3 Metaproteome Analyses of Leaf Litter & Soil

As outlined before, soil and litter environments are due to their high species diversity (most of which uncultivable) together with their high content of inhibitory compounds (i.e. humic acids) an extremely challenging “playground” for metaproteomics analyses [18, 20]. Consequently, the number of studies aiming on a comprehensive characterization of protein expression profiles obtained from soil and litter samples and tackling a specific biological question is still limited. Until recently, the majority of the soil metaproteomics studies have focused on the development or optimization of appropriate protein extraction protocols (see paragraph 2.1). A pioneering study was, however, published already 2005 by Schulze *et al.* [28] who investigated the phylogenetic origin and metabolic function of proteins isolated from dissolved organic matter (DOM) and soil particles derived from lake and forest ecosystems by a GeLC-MS/MS metaproteomics approach. Due to the poor database information at that time the number of protein identifications was somewhat moderate. Nevertheless, the study delivered some first insights into general principles: (i) the authors observed a negative correlation between protein abundance and increasing soil depth or soil organic carbon content, respectively, (ii) the proportion of bacterial proteins was significantly higher in lake compared to soil environments, and (iii) enzymes involved in the degradation of recalcitrant polymers, i.e. cellulose and lignin, were only identified from soil particles, but not from free soil DOM as they tend to bind to recalcitrant compounds within the soil matrix.

3.1 Unraveling Major Players in Leaf Litter Decomposition

Leaf litter decomposition, defined as breakdown of dead plant material, which is composed of cellulose, hemicellulose, lignin, pectin but also proteins, plays a central role in carbon cycling and may thus on a long term also affect our climate [100]. Whilst numerous studies have investigated the influence of biogeochemical parameters on the litter decomposition process itself, e.g. [101, 102], little has been known about structure and functionality of leaf litter decomposing microbial communities. This lack of knowledge motivated Schneider *et al.* [38] to employ a *state-of-the-art* metaproteomics approach for the investigation of two central questions: who are the major microbial players in leaf litter decomposition and how do season and leaf-litter stoichiometry (nutrient content) influence the leaf litter decomposing community? To this end, beech leaf litter from different samples sites in Austria was collected during winter and spring time and analyzed by GeLC-MS/MS-based proteomics and complementary molecular and biochemical approaches such as phospholipid fatty acid (PLFA) analyses and measurements of polymer degrading enzyme activities.

One of the main findings of Schneider *et al.* [38] was, that independent from sampling site or season only fungal, but no bacterial polymer-degrading enzymes could be detected; this was in good agreement with a previous study of the same authors investigating beech litter decomposition in a defined low-complexity model system [37]. In conclusion, fungi appear to be the major players at least at early stages of the decomposition process. Notably, the active part of the microbial community changed over time. Whilst *Ascomycota* dominated early stages of the process, the contribution of *Basidiomycota* to leaf litter decomposition increased at later time points; a result that confirmed earlier studies employing DNA-based sequencing [103]. Schneider *et al.* [38] noticed a positive correlation between leaf litter stoichiometry and the abundance of microbes and polymer degrading enzymes, especially cellulases. More specifically, it was shown that in leaf litter with high nitrogen, phosphorus and micronutrient (i.e. K, Fe, Mn) content the abundance and activity of microbial decomposers is significantly increased compared to low-nutrient containing leaf litters. Notably, as suggested earlier by Schimel *et al.* [104], fungal enzyme production was

defined as “rate-limiting” step in organic-matter decomposition as litter decomposition was fastest at sites with highest abundances of polymer-degrading exoenzymes. **Figure 3** summarizes important steps in early phases of leaf litter decomposition together with site dependent parameters affecting structure and functionality of the decomposing microbial community.

Most of the proteomics data generated by Schneider *et al.* [38] could be conclusively confirmed by the employed complementary approaches, i.e. the phylogenetic composition of the microbial leaf-litter communities predicted by spectral counting was almost identical to the one determined by PLFA analyses. However, it is important to note, that the determination of enzymatic activities in leaf litter point to a major drawback of the metaproteomics analyses, namely that protein abundances do not necessarily correlate with protein activities. This limitation has to be kept in mind whenever proteomics data should contribute to a functional concept of complex ecosystems.

3.2 Global Analyses of Rhizosphere Microbial Communities

The rhizosphere comprising plant roots and the surrounding area of soil is defined as a highly dynamic ecosystem, in which plants shape the microbial population density and activity by their root exudates (for review see [3, 105]). In return, rhizosphere microbes are providing a multitude of benefits to their host including (i) plant growth promotion, (ii) stimulation of pathogen-resistance or direct defense against pathogens, and (iii) nutrient supply e.g. via nitrogen fixation or phosphate solubilization [106]. Whilst numerous proteomics studies (for review see [107]) aimed on the identification of plant proteins specifically induced by the presence of rhizosphere model organisms such as nitrogen-fixing rhizobia and mycorrhizal fungi, much fewer efforts have been undertaken to study the rhizosphere *in situ*.

In 2011, a Chinese research group published two studies reporting metaproteome analyses of different rhizosphere soil samples derived from rice, *Pseudostellariae heterophylla*, *Rehmanniae* sp. and tobacco fields in different cropping areas by employing a

combination of 2-DE and MALDI-Tof/Tof MS analyses employing the NCBI database [45, 46]. In the first study that mainly focused on rice rhizosphere soil samples, about 900 spots were resolved on 2D-gels, 189 of which were successfully identified corresponding to 122 different protein species of rice, bacterial, fungal and metazoan origin [45]. The assigned plant and microbial proteins were mostly involved in protein, energy, nucleotide, and secondary metabolism or signal transduction and resistance mechanisms suggesting that a multitude of metabolic pathways and intra- and interspecies signaling define the complex interactions in the rhizosphere soil biotic community [45]. In the second study the metaproteomes of crop rhizosphere soils subjected to increasing periods of *Rehmannia glutinosa* (a flowering herb used in traditional Chinese medicine) monoculturing were compared by the above mentioned methodological approach [46]. In total, 33 protein spots were identified abundance of which was significantly changed when protein profiles derived from control soil and soils from one-year and two-year *R. glutinosa* monocultured fields were compared. With increasing monoculturing up-regulated plant proteins were found to be involved in carbon and nitrogen metabolism, while most of the up-regulated microbial proteins were shown to participate in protein metabolism and cell-wall biosynthesis. The authors conclude that the accumulation of *R. glutinosa*-specific exudates during monoculturing modifies the structure and functionality the rhizosphere community, which in turn might affect nutrient cycling resulting in plant growth and development retardation [46].

3.3 Metaproteomics of Soil Microbes Involved in Bioremediation

Man-made pollutants, especially industrial wastes, lead to a rapid deterioration of our environment (for review see [108]). Bioremediation as an important branch of environmental biotechnology exploits (mostly) soil-borne fungi (for review see [109]), bacteria (e.g. [110]), or archaea (e.g. [111]) operating under oxic and anoxic conditions to remove harmful pollutants such as polycyclic aromatic hydrocarbons or heavy metals. Modern 'omics' approaches facilitate a comprehensive exploration of the catabolic potential of these beneficial microbes,

thereby providing a solid basis for synthetic microbiology constructing novel catabolic pathways to remove persisting, recalcitrant pollutants from the environment [109].

An early metaproteomics study published in 2003 by Singleton *et al.* investigated the effects of cadmium treatments on total soil protein, protein expression and microbial biomass [112]. Proteins from control and contaminated soils were extracted by a simple snap-freeze protocol, followed by acetone precipitation; protein profiles were then compared by 1-DE. Whilst cadmium-contamination resulted in a significant decrease of microbial biomass and total soil protein, no obvious changes in the protein expression profiles of control and Cd-contaminated soils could be detected [112]. However, due to the lack of any further analysis, an impact of cadmium on soil metaproteome cannot be excluded.

In 2007, Benndorf *et al.* [27] employed an environmental proteomics approach to analyze soil that had been enriched in 2,4-dichlorophenoxy acetic acid (2,4-D)-degrading bacteria. Proteins were extracted from 2,4-D percolated and un-treated soil microcosms by a sequential protocol based on NaOH and phenol (see also **Table 1**) and the resulting protein mixtures were analyzed by a GeLC-MS/MS approach, followed by a search of the obtained mass spectra against the NCBI database. Albeit only four proteins have been identified, one of the central enzymes involved in 2,4-D degradation, the 2,4-dichlorophenoxyacetate dioxygenase, was found in both (pre-treated and un-treated) soils. This finding suggests a comparable bioremediating capacity of the autochthonous and 2,4-D-degrader-enriched communities [27].

More recently, Bastida *et al.* [113] investigated the impact of hydrocarbon (i.e. diesel) contamination on structure and functionality of a semiarid soil microbial community. To this end, the authors were growing microbes enriched from diesel-polluted and control soils in rich medium and analyzed the protein profiles of these mixed cultures by either an immediate protein digestion followed by LC-MS/MS or a pre-fraction of proteins by 1-DE followed by LC-MS/MS (GeLC-MS/MS). Interestingly, the metaproteome profiles indicate a relative increase in the complexity of the soil microbial community with hydrocarbon contamination. Whilst *Bacillus* sp. dominated the communities of un-polluted and diesel-contaminated soils, species such as *Ralstonia solanacearum*, *Synechococcus elongatus* and *Clostridium* sp.

were only found in the polluted soils [113]. The study nicely demonstrates that culture-dependent metaproteomics can actually contribute to a better understanding of how hydrocarbon contamination affects the structure of soil microbial communities.

A similar strategy was employed by Williams *et al.* [114], who investigated the metaproteome of microorganisms directly enriched from toluene-contaminated and glucose-amended soils or derived from enrichment cultures containing toluene and soil extracts. To this end, the authors were combining 1-DE and MALDI-Tof/Tof-MS analyses, followed by searching the MS-data against the NCBI database resulting in the confidential identification of 47 proteins. Proteins identified from microbes enriched from toluene-amended soils as well as toluene-containing cultures included (amongst others) ABC transporters, extracellular solute-binding and outer membrane proteins possibly involved in removing toluene from the bacterial cells. Notably, in microbes enriched from contaminated soils stress-related proteins, i.e. the superoxide dismutase SodB and the chaperonin GroEL appeared to be highly abundant when compared to the un-treated controls. Even though no proteins involved in the toluene degradation pathway could be identified by Williams *et al.* [114], the presented results indicate a strong adaptation of the soil microbial community to the presence of toluene.

4 Remaining Challenges & Future Perspectives

Although being a rapidly developing and highly promising research field, soil and litter metaproteomics is still in its infancy due to the numerous technical drawbacks typical for the soil and litter matrix (outlined in detail before) but also generally connected to environmental proteomics analyses.

An overall problem of metaproteomics studies dealing with highly complex habitats, for which no metagenome sequence information exists, are certainly the often occurring “ambiguous protein identifications” that implement the risk to “over-interpret” the generated data. Simply spoken, a metaproteomics dataset can only be as good as the database used for spectrum to peptide matching in the mass spectrometric analysis. Another problem that should not be underestimated is the exponentially growing size of the available protein

sequences, which results in extremely time-consuming and hardware-demanding database searches. It is thus recommendable to either design habitat-specific databases by restricting the general database to organisms or phylogenetic groups expected to be present in the respective environment or (certainly better) to combine metaproteomics with metagenome sequencing. As shown for habitats such as the plant phyllosphere [115], the application of a database that is based on a matching metagenome significantly enhances the proportion of assignable spectra within a metaproteomics dataset. In case no matching genomic database is available the use of unmatched soil or litter metagenomes would be a valuable alternative. Unfortunately, until today only a limited number of soil metagenomes have been made available to the public, e.g. by the DOE Joint Genome Institute (<http://genome.jgi.doe.gov/>); however, the initiative of the TerraGenome Project [116], which bundles the international efforts to sequence soil metagenomes will hopefully rapidly increase the number of appropriate soil and litter genomic datasets.

Other obstacles, which are more related to the specific nature of the soil and litter environment, include the heterogeneous species distribution, the wide dynamic range of protein abundance levels, and the tendency of secreted and hydrophobic proteins to bind to the soil matrix. These limitations are probably responsible for soil and litter metaproteomic studies having to date only identified “the tip of the iceberg”. Albeit we have still to be patient until the sensitivity of mass spectrometry will allow “single cell proteomics”, constant improvements in soil extraction protocols, mass spectrometry instrumentation (see also paragraph 2.3) and computational analyses (see also paragraph 2.4) summarized as “next generation proteomics” [117] guarantees for a much better soil and litter metaproteome coverage (including low-abundant proteins) together with a more accurate quantification of soil or litter proteins in the future.

In contrast to metagenomics, metaproteomics can not only define the theoretical genetic inventory of a given habitat, but has also the potential to link structure and functionality of the inhabiting biota. However, questions such as which species or enzymes are responsible for the degradation of a specific substrate within soil or litter remain still

unanswered. Recently, a promising new strategy enabling functional insights into metaproteomics analyses has been introduced by Martin von Bergen and colleagues (for review see [118]). The so-called protein-based stable-isotope probing (protein-SIP) approach is based on a metabolic labeling by stable isotopes of carbon, nitrogen, or sulfur. The incorporation of stable isotopes into proteins can subsequently be used as a surrogate marker for metabolic activity (details can be found elsewhere in this special issue) and thus allows the identification of major players responsible for specific biochemical processes in soil and litter habitats.

5 Concluding Remarks

Albeit soil ecologists have recognized long ago that there is an urgent need to understand the complex mechanisms by which soil biodiversity and fertility are maintained, it was only recently that researchers have started to employ *state-of-the-art* ‘omics’ technologies to answer the multitude of questions related to the functional significance of the soil-associated microbiota [4]. Metagenomics, metatranscriptomics, and metaproteomics hold a great potential to contribute to a better understanding of how genetic diversity and community structure and function in soil are related [119]. However, metagenomics provides just a “blueprint of life” and the detection of a gene transcript by metatranscriptomics analyses only “suggests” the expression of a certain gene. Consequently, the direct detection of proteins by metaproteomics is the ultimate tool to link specific taxa to metabolic processes in soil and litter environments even though we have still to face numerous limitations (see above).

What can we learn from challenges and pitfalls reported by the up-to-now relatively small number of soil and litter metaproteomics studies? Firstly, we should be aware that metaproteomics for several reasons (e.g. lab- and time consuming and costly) can still not be regarded as a “high-through-put” technology, which certainly limits its usefulness and implements an intelligent and somehow “modest” experimental design. Secondly, it is essential to test and optimize soil and litter protein extraction protocols whenever a new study should be initiated, as the efficiency of the protocols strongly depends on the physical

and biogeochemical properties of the respective soil or litter sample. Thirdly, in case quantitative information should be generated, an appropriate number of biological but also technical replicates are “sine qua non”, which in turn limits the number of “original samples” that can be investigated. Last but not least it is certainly highly advisable to evaluate these data by complementary approaches such as metatranscriptomics, FISH analyses or (whenever applicable) phenotypical assays, e.g. measurement of enzyme activities.

Keeping the potential but also the drawbacks of soil and litter metaproteomics in mind, we are sure that this rapidly developing research field will enable us to re-visit and answer old questions in soil ecology, e.g. how biotic and abiotic factors influence soil biodiversity and fertility or how anthropogenic interferences and climate changes affect structure and functionality of soil microbial communities.

Acknowledgements

We thank Christian Lassek & Uwe Wegner for critical reading, Giovanni & Peppone R. for encouragement and Florian Winter from the group of Sophie Zechmeister-Boltenstern for graphic design. The research of K. Riedel and S. Zechmeister-Boltenstern reported in this review was funded by a grant from the Austrian Science Foundation (FWF).

Conflict of interest statement

The authors declare to have no conflicts of interest.

References

- [1] Matthews, E., Global litter production, pools, and turnover times: Estimates from measurement data and regression models. *J. Geophys. Res.* 1997, *102*, 18771–18800.
- [2] Smith, S. E., Smith, F. A., Roles of arbuscular mycorrhizas in plant nutrition and growth: new paradigms from cellular to ecosystem scales. *Annu. Rev. Plant Biol.* 2011, *62*, 227-250.
- [3] Berendsen, R. L., Pieterse, C. M., Bakker, P. A., The rhizosphere microbiome and plant health. *Trends Plant Sci.* 2012, *17*, 478-86.
- [4] Wardle, D. A., The influence of biotic interactions on soil biodiversity. *Ecology Lett.* 2006, *9*, 870-886.
- [5] Torsvik, V., Goksøyr, J., Daae, F. L., Sørheim, R. et al., In: Beyond the Biomass: Compositional and Functional Analysis of Soil Microbial Communities. Ritz, K., Dighton, J. Giller, K. E. (Eds.). *Use of DNA to determine the diversity of microbial communities*. Wiley Chichester, 1994, pp. 39-48.
- [6] Wardle, D. A., Communities and Ecosystems. Linking the Above-ground and Below-ground Components. Princeton University Press, Princeton, NJ, 2002.
- [7] Amann, R. I., Ludwig, W., Schleifer, K. H., Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *FEMS Microbiol. Rev.* 1995, *59*, 143-169.
- [8] Mocali, S., Benedetti, A., Exploring research frontiers in microbiology: the challenge of metagenomics in soil microbiology. *Res. Microbiol.* 2010, *161*, 497-505.
- [9] Nannipieri, P., In: Nannipieri, P., Smalla, K. (Eds.). *Role of Stabilised Enzymes in Microbial Ecology and Enzyme Extraction from Soil with Potential Applications in Soil Proteomics*. Series: Soil Biology, Springer Verlag, Heidelberg, 2006, *8*, pp. 75-94.
- [10] Schloss, P. D., Handelsman, J., Toward a census of bacteria in soil. *PLoS Comput. Biol.* 2006, *2*, e92.
- [11] Simon, C., Daniel, R., Metagenomic analyses: past and future trends. *Appl. Environ. Microbiol.* 2011, *77*, 1153-1161.

- [12] Lee, N., Nielsen, P. H., Andreasen, K. H., Juretschko, S. et al., Combination of fluorescent in situ hybridization and microautoradiography – a new tool for structure-function analyses in microbial ecology. *Appl. Environm. Microbiol.* 1999, 65, 1289-1297.
- [13] Kubota, K., CARD-FISH for Environmental Microorganisms: Technical Advancement and Future Applications. *Microbes Environ.* 2012, Oct 31 [Epub ahead of print].
- [14] Wellington, E. M., Berry, A., Krsek, M., Resolving functional diversity in relation to microbial community structure in soil: exploiting genomics and stable isotope probing. *Curr. Opin. Microbiol.* 2003, 6, 295-301.
- [15] Saleh-Lakha, S., Miller, M., Campbell, R. G., Schneider, K., et al., Microbial gene expression in soil: methods, applications and challenges. *J. Microbiol. Methods* 2005, 63, 1-19.
- [16] Carvalhais, L. C., Dennis, P. G., Tyson, G. W., Schenk, P. M., Application of metatranscriptomics to soil environments. *J. Microbiol. Methods* 2012, 91, 246-251.
- [17] Su, C., Lei, L., Duan, Y., Zhang, K. Q. et al., Culture-independent methods for studying environmental microorganisms: methods, application, and perspective. *Appl. Microbiol. Biotechnol.* 2012, 93, 993-1003.
- [18] Bastida, F., Moreno, J. L., Nicolas, C., Hernandez, T. et al., Soil metaproteomics: a review of an emerging environmental science. Significance, methodology and perspectives. *Eur. J. Soil Science* 2009, 60, 845-859.
- [19] Keller, M., Hettich, R., Environmental proteomics: a paradigm shift in characterizing microbial activities at the molecular level. *Microbiol. Mol. Biol. Rev.* 2009, 73, 62-70.
- [20] Siggins, A., Gunnigle, E., Abram, F., Exploring mixed microbial community functioning: recent advances in metaproteomics. *FEMS Microbiol Ecol.* 2012, 80, 265-280.
- [21] Schneider, T., Riedel, K., Environmental proteomics: analysis of structure and function of microbial communities. *Proteomics* 2010, 10, 785-798.
- [22] Wilmes, P., Bond, P. L., Microbial community proteomics: elucidating the catalysts and metabolic mechanisms that drive the Earth's biogeochemical cycles. *Curr. Opin. Microbiol.* 2009, 12, 310-317.

- [23] Chourey, K., Jansson, J., VerBerkmoes, N., Shah, M. et al., Direct cellular lysis/protein extraction protocol for soil metaproteomics. *J. Proteome Res.* 2010, 9, 6615-6622.
- [24] Wang, W., Vignani, R., Scali, M., Cresti, M., A universal and rapid protocol for protein extraction from recalcitrant plant tissues for proteomic analysis. *Electrophoresis* 2006, 27, 2782-2786.
- [25] Keiblinger, K. M., Wilhartitz, I. C., Schneider, T., Roschitzki, B., et al., Soil metaproteomics - Comparative evaluation of protein extraction protocols. *Soil Biol. Biochem.* 2012, 54, 14-24.
- [26] Taylor, E. B., Williams, M. A., Microbial Protein in Soil: Influence of Extraction Method and C Amendment on Extraction and Recovery. *Microbial Ecology* 2010, 59, 390-399.
- [27] Benndorf, D., Balcke, G. U., Harms, H., von Bergen, M., Functional metaproteome analysis of protein extracts from contaminated soil and groundwater. *ISME J.* 2007, 1, 224-234.
- [28] Schulze, W. X., Gleixner, G., Kaiser, K., Guggenberger, G. et al., A proteomic fingerprint of dissolved organic carbon and of soil particles. *Oecologia* 2005, 142, 335-343.
- [29] Nielsen, K. M., Calamai, L., Pietramellara, G., In: Nannipieri, P., Smalla, K. (Eds.). *Stabilization of Extracellular DNA and Proteins by Transient Binding to Various Soil Components*. Series: Soil Biology, Springer Verlag, Heidelberg, 2006, 8, pp. 141-157.
- [30] Giagnoni, L., Magherini, F., Landi, L. Taghavi, S. et al., Soil solid phases effects on the proteomic analysis of *Cupriavidus metallidurans* CH34. *Biol. Fert. Soils* 2012, 48, 425-433.
- [31] Criquet, S., Farnet, A. M., Ferre, E., Protein measurement in forest litter. *Biol. Fert. Soils* 2002, 35, 307-313.
- [32] Murase, A., Yoneda, M., Ueno, R., Yonebayashi, K., Isolation of extracellular protein from greenhouse soil. *Soil Biol. Biochem.* 2003, 35, 733-736.
- [33] Carpentier, S. C., Witters, E., Laukens, K., Deckers, P. et al., Preparation of protein extracts from recalcitrant plant tissues: an evaluation of different methods for two-dimensional gel electrophoresis analysis. *Proteomics* 2005, 5, 2497-2507.

- [34] Chen, S., Rillig, M. C., Wang, W., Improving soil protein extraction for metaproteome analysis and glomalin-related soil protein detection. *Proteomics* 2009, 9, 4970-4973.
- [35] Benndorf, D., Vogt, C., Jehmlich, N., Schmidt, Y., et al., Improving protein extraction and separation methods for investigating the metaproteome of anaerobic benzene communities within sediments. *Biodegradation* 2009, 20, 737-750.
- [36] Bastida, F., Algora, C., Hernández, T., García, C., Feasibility of a cell separation-proteomic based method for soils with different edaphic properties and microbial biomass. *Soil Biol. Biochem.* 2012, 45, 136-138.
- [37] Schneider, T., Gerrits, B., Gassmann, R., Schmid, E. et al., Proteome analysis of fungal and bacterial involvement in leaf litter decomposition. *Proteomics* 2010, 10, 1819-1830.
- [38] Schneider, T., Keiblinger, K. M., Schmid, E., Sterflinger-Gleixner, K. et al., Who is who in litter decomposition? Metaproteomics reveals major microbial players and their biogeochemical functions. *ISME J.* 2012, 6, 1749-1762.
- [39] Giagnoni, L., Magherini, F., Landi, L., Taghavi, S. et al., Extraction of microbial proteome from soil: potential and limitations assessed through a model study. *Eur. J. Soil Sci.* 2011, 62, 74–81.
- [40] O'Farrel, P. H., High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* 1975, 250, 4007-4021.
- [41] Weiss, W., Görg, A., High-resolution two-dimensional electrophoresis. *Methods Mol. Biol.* 2009, 564, 13-32.
- [42] Kan, J., Hanson, T. E., Ginter, J.M., Wang, K., et al., Metaproteomic analysis of Chesapeake Bay microbial communities. *Saline Systems* 2005, 1, 7-20.
- [43] Klaassens, E. S., de Vos, W. M., Vaughan, E. E., Metaproteomics approach to study the functionality of the microbiota in the human infant gastrointestinal tract. *Appl. Environ. Microbiol.* 2007, 73, 1388-1392.
- [44] Wilmes, P., Wexler, M., Bond, P. L., Metaproteomics provides functional insight into activated sludge wastewater treatment. *PLoS One* 2008, 3:e1778.

- [45] Wang, H. B., Zhang, Z. X., Li, H., He, H. B. et al., Characterization of metaproteomics in crop rhizospheric soil. *J. Proteome Res.* 2011, 10, 932-940.
- [46] Wu, L., Wang, H., Zhang, Z., Lin, R. et al., Comparative metaproteomic analysis on consecutively *Rehmannia glutinosa*-monocultured rhizosphere soil. *PLoS One* 2011, 6:e20611.
- [47] Gygi, S. P., Rist, B., Aebersold, R., Measuring gene expression by quantitative proteome analysis. *Curr. Opin. Biotechnol.* 2000, 11, 396-401.
- [48] Ong, S. E., Pandey, A., An evaluation of the use of two-dimensional gel electrophoresis in proteomics. *Biomol. Eng.* 2001, 18, 195-205.
- [49] Moche, M., Albrecht, D., Maaß, S., Hecker, M. et al., The new horizon in 2D electrophoresis - new technology to increase resolution and sensitivity. *Electrophoresis* 2013, Mar 13. doi: 10.1002/elps.201200618. [Epub ahead of print].
- [50] Schneider, T., Schmid, E., de Castro, J.V. Jr., Cardinale, M. et al., Structure and function of the symbiosis partners of the lung lichen (*Lobaria pulmonaria* L. Hoffm.) analyzed by metaproteomics. *Proteomics* 2011, 11, 2752-2756.
- [51] Morris, R. M., Nunn, B. L., Frazar, C., Goodlett, D. R. et al., Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. *ISME J.* 2010, 4, 673-685.
- [52] Leary, D. H., Hervey, W. J. 4th, Li, R. W., Deschamps, J. R. et al., Method development for metaproteomic analyses of marine biofilms. *Anal. Chem.* 2012, 84, 4006-4013.
- [53] Teeling, H., Fuchs, B. M., Becher, D., Klockow, C. et al., Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science* 2012, 336, 608-611.
- [54] Rudney, J. D., Xie, H., Rhodus, N. L., Ondrey, F. G. et al., A metaproteomic analysis of the human salivary microbiota by three-dimensional peptide fractionation and tandem mass spectrometry. *Mol. Oral Microbiol.* 2010, 25, 38-49.
- [55] Verberkmoes, N. C., Russell, A. L., Shah, M., Godzik, A. et al., Shotgun metaproteomics of the human distal gut microbiota. *ISME J.* 2009, 3, 179-189.

- [56] Kolmeder, C. A., de Been, M., Nikkilä, J., Ritamo, I. et al., Comparative metaproteomics and diversity analysis of human intestinal microbiota testifies for its temporal stability and expression of core functions. *PLoS One* 2012, 7, e29913.
- [57] Fang, Y., Robinson, D.P., Foster, L.J., Quantitative analysis of proteome coverage and recovery rates for upstream fractionation methods in proteomics. *J. Proteome Res.* 2010, 9, 1902-1912.
- [58] Piersma, S.R., Fiedler, U., Span, S., Lingnau, A., et al., Workflow comparison for label-free, quantitative secretome proteomics for cancer biomarker discovery: method evaluation, differential analysis, and verification in serum. *J. Proteome Res.* 2010, 9, 1913-1922.
- [59] Premstaller, A., Oberacher, H., Walcher, W., Timperio, A. M. et al., High-performance liquid chromatography-electrospray ionization mass spectrometry using monolithic capillary columns for proteomic studies. *Anal. Chem.* 2001, 73, 2390-2396.
- [60] Plumb, R., Castro-Perez, J., Granger, J., Beattie, I. et al., Ultra-performance liquid chromatography coupled to quadrupole-orthogonal time-of-flight mass spectrometry. *Rapid Commun. Mass. Spectrom.* 2004, 18, 2331-2337.
- [61] Olsen, J. V., de Godoy, L. M., Li, G., Macek, B. et al., Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* 2005, 4, 2010-2201.
- [62] Hu, Q., Noll, R. J., Li, H., Makarov, A. et al., The Orbitrap: a new mass spectrometer. *J. Mass Spectrom.* 2005, 40, 430-443.
- [63] Olsen, J. V., Schwartz, J. C., Griep-Raming, J., Nielsen, M. L. et al., A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Mol. Cell. Proteomics* 2009, 8, 2759-2769.
- [64] Bateman, R. H., Carruthers, R., Hoyes, J. B., Jones, C. et al., A novel precursor ion discovery method on a hybrid quadrupole orthogonal acceleration time-of-flight (Q-TOF) mass spectrometer for studying protein phosphorylation. *J. Am. Soc. Mass. Spectrom.* 2002, 13, 792-803.

- [65] Silva, J. C., Gorenstein, M. V., Li, G. Z., Vissers, J. P. et al., Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol. Cell. Proteomics* 2006, 5, 144-156.
- [66] Panchaud, A., Scherl, A., Shaffer, S. A., von Haller, P. D. et al., Precursor acquisition independent from ion count: how to dive deeper into the proteomics ocean. *Anal. Chem.* 2009, 81, 6481-6488.
- [67] Panchaud, A., Jung, S., Shaffer, S. A., Aitchison, J.D. et al., Faster, quantitative, and accurate precursor acquisition independent from ion count. *Anal. Chem.* 2011, 83, 2250-2257.
- [68] Gillet, L. C., Navarro, P., Tate, S., Röst, H. et al., Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* 2012, 11, O111.016717.
- [69] Bensimon, A., Heck, A. J., Aebersold, R., Mass spectrometry-based proteomics and network biology. *Annu. Rev. Biochem.* 2012, 81, 379-405.
- [70] Perkins, D. N., Pappin, D. J., Creasy, D. M., Cottrell, J. S., Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 1999, 20, 3551-3567.
- [71] Eng, J. K., McCormack, A. L., Yates, III J. R., An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* 1994, 5, 976-989.
- [72] Craig, R., Beavis, R. C., TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* 2004, 20, 1466-1467.
- [73] Tabb, D. L., Fernando, C. G., Chambers, M. C., MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. *Proteome Res.* 2007, 6, 654-661.
- [74] Nesvizhskii, A. I., Vitek, O., Aebersold, R., Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat. Methods* 2007, 4, 787-797.

- [75] Neilson, K. A., Ali, N. A., Muralidharan, S., Mirzaei, M. et al., Less label, more free: approaches in label-free quantitative mass spectrometry. *Proteomics* 2011, 11, 535-553.
- [76] Graham, C., McMullan, G., Graham, R.L., Proteomics in the microbial sciences. *Bioeng. Bugs* 2011, 2, 17-30.
- [77] Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. et al., Quantitative mass spectrometry in proteomics: a critical review. *Anal. Bioanal. Chem.* 2007, 389, 1017-1031.
- [78] Ono, M., Shitashige, M., Honda, K., Isobe, T. et al., Label-free quantitative proteomics using large peptide data sets generated by nanoflow liquid chromatography and mass spectrometry. *Mol. Cell. Proteomics* 2006, 5, 1338-1347.
- [79] Choi, H., Fermin, D., Nesvizhskii, A. I., Significance analysis of spectral count data in label-free shotgun proteomics. *Mol. Cell. Proteomics* 2008, 7, 2373-2385.
- [80] Nesvizhskii, A. I., Aebersold, R., Interpretation of shotgun proteomic data: the protein inference problem. *Mol. Cell. Proteomics* 2005, 4, 1419-1440.
- [81] Webb-Robertson, B. J., Cannon, W. R., Current trends in computational inference from mass spectrometry-based proteomics. *Brief Bioinform.* 2007, 8, 304-17.
- [82] Huang, T., Wang, J., Yu, W., He, Z., Protein inference: a review. *Brief Bioinform.* 2012, 13, 586-614.
- [83] Boeckmann, B., Blatter, M. C., Famiglietti, L., Hinz, U. et al., Protein variety and functional diversity: Swiss-Prot annotation in its biological context. *C. R. Biol.*, 2005, 328, 882-899.
- [84] Pruitt, K. D., Tatusova, T., Brown, G. R., Maglott, D. R., NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res.* 2012, 40 (Database issue), D130-135.
- [85] Rentzsch, R., Orengo, C. A., Protein function prediction - the power of multiplicity. *Trends Biotechnol.* 2009, 27, 210-219.
- [86] Moog, G., Schmid, E., Becher, D., Bernhardt, J. et al., Challenges and Pitfalls of Metaproteomic Data Analysis Addressed by the Automated Workflow Prophan. Submitted to this issue.

- [87] Berth, M., Moser, F. M., Kolbe, M., Bernhardt, J., The state of the art in the analysis of two-dimensional gel electrophoresis images. *Appl. Microbiol. Biotechnol.* 2007, 76, 1223-43.
- [88] Linsen, L., Löcherbach, J., Berth, M., Becher, D. et al., Visual analysis of gel-free proteome data. *IEEE Trans. Vis. Comput. Graph.* 2006, 12, 497-508.
- [89] Kestler, H. A., Müller, A., Gress, T. M., Buchholz, M.. Generalized Venn diagrams: a new method of visualizing complex genetic set relations. *Bioinformatics* 2005, 21, 1592-1595.
- [90] R Development Core Team, The. "R: A language and environment for statistical computing." Vienna, Austria: R Foundation for Statistical Computing. 2012, <http://www.r-project.org/>
- [91] Jerry, L., Hintze, R. D. N., Violin plots: a box plot-density trace synergism. *Am. Stat.* 1998, 52, 181-184.
- [92] Potter, K., Kniss, J., Riesenfeld, R., Johhson, C., In: G, Munzner (Eds.). Visualizing summary statistics and uncertainty. *Proceedings of EuroVis* 2010, 823-831.
- [93] Shneiderman, B., Wattenberg, M., Ordered treemap layouts. In Proceedings of IEEE Symposium on Information Visualization 2001, 2001, 73-78.
- [94] Balzer, M., Deussen, O., Voronoi treemaps. IEEE Symposium on Information Visualization, 2005 (INFOVIS 2005).
- [95] Bernhardt, J., Funke, S., Hecker, M., Siebourg, J., Visualizing gene expression data via Voronoi treemaps. 6th International Symposium on Voronoi Diagrams in Science and Engineering, 2009 (ISVD'09). Copenhagen, Denmark: Computer Society Digital Library (CSDL); 2009.
- [96] McDonnell K. T., Mueller, K., In: Vilanova, A., Telea, A., Scheuermann, G., Möller, T. (Eds.). *Illustrative Parallel Coordinates*. Eurographics / IEEE-VGTC Symposium on Visualization 2008, 27, Nr. 3.
- [97] Wen, X., Fuhrman, S., Michaels, G. S., Carr, D. B. et al., Large-scale temporal gene expression mapping of central nervous system development. *Proc. Natl Acad. Sci. USA* 1998, 95, 334–339.

- [98] Eisen, M. B., Spellman, P. T., Brown, P. O., Botstein, D., Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* 1998, 95, 14863–14868.
- [99] Byron, L., Wattenberg, M., Stacked Graphs – Geometry & Aesthetics. *IEEE Trans. Vis. Comput. Graph.* 2008, 14, 1245 - 1252
- [100] Chapin, F. S., Matson, P. A., Mooney, H. A., Principles of terrestrial ecosystem ecology. New York: Springer, 2002.
- [101] Zhang, D., Hui, D., Luo, Y., Zhou, G., Rates of litter decomposition in terrestrial ecosystems: global patterns and controlling factors. *J. Plant Ecol.* 2008, 1, 85–93.
- [102] Prescott, C. E., Litter decomposition: what controls it and how can we alter it to sequester more carbon in forest soils? *Biogeochemistry* 2010, 101, 133–149.
- [103] Baldrian, P., Kolarik, M., Stursova, M., Kopecky, J. et al., Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. *ISME J.* 2012, 6, 248–258.
- [104] Schimel, J. P., Weintraub, M. N., The implications of exoenzyme activity on microbial carbon and nitrogen limitation in soil: a theoretical model. *Soil Biol. Biochem.* 2003, 35, 549–563.
- [105] Badri, D. V., Weir, T. L., van der Lelie, D., Vivanco, J. M., Rhizosphere chemical dialogues: plant-microbe interactions. *Curr. Opin. Biotechnol.* 2009, 20, 642-650.
- [106] Lugtenberg, B., Kamilova, F., Plant-growth-promoting rhizobacteria. *Annu. Rev. Microbiol.* 2009, 63, 541-56.
- [107] Mathesius, U., Comparative proteomic studies of root-microbe interactions. *J. Proteomics* 2009, 72, 353-366.
- [108] Harms, H., Schlosser, D., Wick, L. Y., Untapped potential: exploiting fungi in bioremediation of hazardous chemicals. *Nat. Rev. Microbiol.* 2011, 9, 177-92.
- [109] Ramos, J. L., Marqués, S., van Dillewijn, P., Espinosa-Urgel, M. et al., Laboratory research aimed at closing the gaps in microbial bioremediation. *Trends Biotechnol.* 2011, 29, 641-647.
- [110] Rojo, F., Degradation of alkanes by bacteria. *Environ. Microbiol.* 2009, 11, 2477-2490.

- [111] Bini, E., Archaeal transformation of metals in the environment. *FEMS Microbiol. Ecol.* 2010, *1*, 1-16.
- [112] Singleton, I., Merrington, G., Colvan, S., Delahunty, J. S., The potential of soil protein-based methods to indicate metal contamination. *Appl. Soil Ecol.* 2003, *23*, 25–32.
- [113] Bastida, F., Nicolas, C., Moreno, J. L., Hernandez, T. et al., Tracing Changes in the Microbial Community of a Hydrocarbon-Polluted Soil by Culture-Dependent. *Pedosphere* 2010, *20*, 479–485.
- [114] Williams, M. A., Taylor, E. B., Mula, H. P., Metaproteomic characterization of a soil microbial community following carbon amendment. *Soil Biol. Biochem.* 2010, *42*, 1148-1156.
- [115] Delmotte, N., Knief, C., Chaffron, S., Innerebner, G. et al., Community proteogenomics reveals insights into the physiology of phyllosphere bacteria. *Proc. Natl. Acad. Sci. U S A* 2009, *106*, 16428-16433.
- [116] Vogel, T. M., Simonet, P., Jansson, J. K., Hirsch, P. R. et al., TerraGenome: a consortium for the sequencing of a soil metagenome. *Nat. Rev. Microbiol.* 2009, *7*, 252.
- [117] Altelaar, A. F., Munoz, J. Heck, A. J., Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat. Rev. Genet.* 2013, *14*, 35-48.
- [118] Seifert, J., Taubert, M., Jehmlich, N., Schmidt, F. et al., Protein-based stable isotope probing (protein-SIP) in functional metaproteomics. *Mass Spectrom. Rev.* 2012, *31*, 683-697.
- [119] Burns, R. G., DeForest, J. L., Marxsen, J., Sinsabaugh, R. L. et al., Soil enzymes in a changing environment: Current knowledge and future directions. *Soil Biol. Biochem.* 2013, *58*, 216-234.

Figure legends:

Figure 1: General workflow summarizing *state-of-the-art* approaches employed in soil and litter metaproteomics analyses. *De novo sequencing is certainly highly useful for the identification of proteins, for which no sequence exists in the database, but becomes still not routinely employed. 2-DE, two-dimensional gel electrophoresis; 1-DE, one-dimensional SDS polyacrylamide gel electrophoresis; LC, liquid chromatography.

Figure 2: Comparison of identified peptides of the same (highly complex marine) sample measured at two different Orbitrap mass spectrometers: A) LTQ Orbitrap XL and B) Orbitrap Velos. The filter criteria for determination of peptide identification were: Xcorr 1.8 ($z = 1$); 2.2 ($z = 2$); 3.3 ($z = 3$); 95% peptide probability; max.10 ppm mass deviation. Grey dots represent the acquired MS spectra, which are the base for the data dependent acquisition of collision induced fragment spectra (MS/MS spectra). Blue dots represent the MS/MS spectra, which were acquired but not identified during the subsequent database search. All acquired MS/MS spectra, which could be successfully identified are shown in a scale from green to red dependent on the peptide identification probability score; highly reliable identifications of MS/MS spectra are represented by red dots. Whilst the LTQ Orbitrap XL only identified 215 out of 3416 spectra (6%, representing 106 different peptides), 793 out of 7643 spectra (10%, representing 516 different peptides) were identified by the Orbitrap Velos. These data nicely illustrate the impact of the higher scan rate (increased number of acquired spectra) and the higher sensitivity (increased percentage of identified spectra) on the performance of the Orbitrap Velos.

Figure 3: Schematic overview of early stages in litter decomposition together with site-dependent factors modulating structure and functionality of the leaf litter decomposing microbial community. Fe, iron; K, potassium; Mg, manganese; season, harvesting time point February or May. Figure designed by Florian Winter.

Table 1: Selected studies reporting different protocols for the extraction of soil and/or litter proteins (modified & updated from [18])

Matrix	Extraction & Purification Protocol	Proteome Analysis Methods	Numbers of detected Proteins	Reference
Evergreen oak litter	direct extraction powdering (mixer) 0.1 M CaCl ₂ , 0.1% Tween 80, 70 mg/ml PVPP filtering & dialysis	/	/	[31]
Recalcitrant plant tissue	direct extraction grinding with liquid nitrogen 1) precipitation with acetone/20% TCA/0.2% DTT, resuspension in 50 mM Tris/HCl, pH 8.5; 5 mM EDTA, 100 mM KCl, 1% DTT, 30% sucrose, protease inhibitor cocktail 2) phenol extraction, precipitation with ammonium acetate	2-DE ^a	depending on the extraction protocol 1) ~ 1000 protein spots 2) ~ 1150 protein spots	[33]
Aged leaves from different plants	direct extraction grinding with liquid nitrogen washing powder with TCA/acetone, methanol, and acetone 1:1 1% SDS, 50 mM Tris/HCl, pH 7.5 : phenol precipitation with 100 mM ammonium acetate	2-DE ^a	not specified	[24]
Beech leaf litter	direct extraction grinding with liquid nitrogen + 1% SDS, 50 mM Tris/KOH, pH 7.0 2 min sonication, 20 min boiling Speedvac concentration	GeLC-MS/MS ^b	assignment of 1724 non-redundant proteins	[37, 38]
Sediment (river gravel or lava granules)	direct extraction 5 min ultrasonication phenol extraction, precipitation with 100 mM ammonium acetate/methanol	2-DE ^a	~240 protein spots	[35]
Agricultural-abandoned soils	indirect extraction isolation of microbial cells by density-gradient centrifugation cell lysates were subjected to phenol extraction & precipitation with methanol/100 mM ammonium acetate	GeLC-MS/MS ^b	depending on the soil assignment of 11 to 71 non-redundant proteins	[36]
Greenhouse soil	direct extraction 67 mM phosphate buffer filtering & precipitation with TCA	1-DE ^c N-terminal sequencing	5 identified proteins	[32]
Soil microparticles	direct extraction Na polytungstate, 10% HF, neutralization gel filtration (sepharose 4B)	GeLC-MS/MS ^b	up to 78 identified proteins	[33]
Compost soil	100 mM NaOH vortexing, sonication phenol extraction, precipitation with methanol / 100 mM ammonium acetate	GeLC-MS/MS ^b	4 identified proteins	[27]
Agricultural soil	direct extraction optimized sequential extraction 1) 25 mM citrate, pH 8.0 2) 1% SDS, 20 mM DTT,	2-DE ^a	~250 protein spots	[34]

	50 mM Tris/HCl, pH 6.8, phenol extraction of combined extracts precipitation with methanol/100 mM ammonium acetate			
Grassland soil	direct extraction 5% SDS, 50 mM Tris/HCl, pH 8.5, 150 mM NaCl, 100 mM EDTA, 1 mM MgCl ₂ , 50 mM DTT 2-3 min vortexing, 20 min boiling precipitation with TCA resuspension in 6 mM guanidine HCl, 10 mM DTT in 50 mM Tris, 10 mM CaCl ₂ , pH 7.6	2D-LC- MS/MS ^d	assignment of 333 non-redundant proteins from non-spiked soil	[23]
Soil microcosms spiked with <i>Cupriavidus metallidurans</i> proteins	direct extraction 20 mM phosphate buffer saline (PBS) 1% SDS, 1 mg/ml DNase and RNase, protease inhibitor cocktail sonication, acetone precipitation resuspension in PBS/SDS and second precipitation with chloroform / methanol	2-DE ^a	~320 protein spots	[39]
Forest soil (FS) & potting soil (PS)	direct extraction grinding in liquid nitrogen with 10% PVVP 1) 1% SDS, 50 mM Tris/HCl, pH 7.5 sonication, 20 min boiling precipitation with TCA 2) 100 mM NaOH vortexing, sonication phenol extraction, precipitation with methanol/100 mM ammonium acetate 3) 1:1 1% SDS, 50 mM Tris/HCl, pH 7.5, phenol vortexing, sonication precipitation with methanol /100 mM ammonium acetate 4) same as 3) with additional washing steps prior extraction	2D-LC- MS/MS ^d	depending on the soil and protocol different numbers of non-redundant proteins 1) 226 FS / 237 PS 2) 293 FS / 124 PS 3) 494 FS / 198 PS 4) 372 FS / 80 PS	[25]

^a 2-DE = two-dimensional gel electrophoresis

^b GeLC-MS/MS = one-dimensional gel electrophoresis combined with liquid chromatography
and tandem mass spectrometry

^c 1-DE = one dimensional gel electrophoresis (SDS-PAGE)

^d 2D-LC-MS/S = two-dimensional liquid chromatography coupled to tandem mass spectrometry