Biogeosciences Discussions, 1, 195–218, 2004
www.biogeosciences.net/bgd/1/195/
SRef-ID: 1810-6285/bgd/2004-1-195

Biogeosciences Discussions

Biogeosciences Discussions is the access reviewed discussion forum of *Biogeosciences*

Environmental proteomics – what proteins from soil and surface water can tell us: a perspective

W. Schulze

Center for Experimental Bioinformatics, University of Southern Denmark, Campusvej 55, 5230 Odense, Denmark

Received: 28 June 2004 – Accepted: 12 July 2004 – Published: 22 July 2004

Correspondence to: W. Schulze (waltraud@bmb.sdu.dk)

Abstract

Mass spectrometry based proteomics is widely used to study cellular processes in model organisms. However, it has not much been applied in environmental research because it was thought that free proteins would not be sufficiently stable in the environments. Based on recent observations that protein can readily be detected as a component of dissolve organic carbon, this article gives an overview about the possible use of proteomic methods in ecology and environmental sciences. At this stage, there are two areas of interest: (1) the identification of phylogenetic groups contributing to the DOC pool, and (2) identification of the origin of specific enzymes that are important for ecosystem processes. In this paper methods of mass spectrometry based proteomics were applied to identify proteins from DOC and water samples from different environments. It is demonstrated, that environmental proteomics is capable to distinguish the active set of organisms of different horizons of soils, and from various sources of surface water. Currently the limitation is given by the present knowledge of the genome of soil organisms. In addition, environmental proteomics allows to relate protein presence to biogeochemical processes, and to identify the source organisms for specific enzymes. Taking laccases as an example, it is shown that this enzyme is excreted into soils by a whole range of organisms from different phylogenetic groups. Further applications, such as in pollution reseach are conceivable. In summary, environmental proteomcis opens a new area of research between the fields of microbiology and biogeochemistry.

1. Introduction

Dissolved organic matter (DOM) plays an important role in the carbon biogeochemistry coupling terrestrial and aquatic carbon pools. Dissolved organic nitrogen and carbon are significant for C and N cycles of terrestrial ecosystems and undergo variations in season and depth profile (Michalzik and Matzner, 1999; Kaiser et al., 2001), and

BGD

1, 195–218, 2004

the composition and origin of DOM may depend on the organisms living in a given environment. Dissolved organic carbon and dissolved organic nitrogen have so far been well studied with respect to their $\delta^{13}C$ and $\delta^{15}N$ origin and their basic chemical structure (Michalzik and Matzner, 1999; Gleixner et al., 2001; Kaiser et al., 2001). ^{15}N NMR analysis has demonstrated that a significant amount of nitrogen is present in amide form (Almendros et al., 1991), and thus possibly as protein. However, not much is actually known about these proteins, their origin or enzymatic functions within the pool of dissolved organic matter. The identity of the proteins, their phylogenetic origin, their functions and spatial distribution could shed an entirely new light on ecosystem biology and at the same time link to biogeochemical processes.

Proteomics is one of the fastest developing research areas, and contributes substantially to our understanding of organisms at the cellular level (Aebersold and Mann, 2003; Tyers and Mann, 2003). There are now ongoing approaches to isolate and sequence random DNA samples extracted from environments, such as oceans and soil (Tyson et al., 2004; Venter et al., 2004). However, in contrast to DNA and RNA, proteins mirror the taxonomy of the active pool of organisms and in addition can have an active biological function. By directly analyzing the proteome of the environment, we can conclude about the identity of the organisms contributing to the DOM pool, and at the same time also understand the functional contribution of certain proteins to biogeochemical processes. However, the proteomic methods have not yet been widely applied to study proteins derived from environmental samples. In a pioneer study, the protein composition of water samples collected from different soil layers and soil particles was analyzed (Schulze et al., 2004). There, it was shown that a large number of proteins is present in soil water, and that they can mirror the organismic groups being present in ecosystems.

By further persuing the analysis of proteomic fingerprints of different environments, in this paper mass spectrometry based proteomics was applied to exploit taxonomic and functional information of the protein component of DOM. Proteomic analysis of environmental samples was applied to (i) analysis of organisms contributing along a

BGD
1, 195–218, 2004

decomposition line of plant material, (ii) a comparison of taxonomic units present in different surface waters, and (iii) comparison taxonomic units of DOC from different soil types. Furthermore, functional information was exploited by classifying identified proteins according to (iv) size and (v) cellular function. It is shown that environmental proteomics has the potential to develop into a novel field of ecological and environmental research.

2. Methods

2.1. Soil and surface water collection

Surface waters were collected directly, filtered, and freeze-dried. Percolating soil water was obtained using glass ceramic suction plates. Water was filtered through a 0.45 μm acetate filter membranes prior to freeze-drying.

2.2. Sample preparation and analysis by LC-MS

Water samples were desalted and purified proteins were digested in-solution by trypsin prior to analysis by LC-MS/MS following the procedures described by (Schulze et al., 2004). Acquired spectra were searched against the NCBI protein database (http://www.ncbi.nlm.nih.gov/) using the Mascot algorithm (Perkins et al., 1999). The following search parameters were applied: maximum of one missed trypsin cleavage, cysteine carbamidomethylation, methionine oxidation, and a maximum 0.2 Da error tolerance in both the MS and MS/MS data (40 ppm after dynamic recalibration). Only fully tryptic peptides were allowed and all hits were manually verified against the raw mass spectrometric data using accepted rules for peptide fragmentation in a quadrupole-TOF hybrid mass spectrometer. On average, 30% of the proteins were identified by a single tryptic peptide.

2.3. Taxonomic and functional classification

The protein sequence derived from MS/MS spectra of tryptic peptides bears taxonomic information of the origin of the protein. In most cases, the sequences obtained from tryptic peptides were unique to a specific group of organisms or even single species. Since full proteomic information in the database is available only for a limited number of organisms, the identified proteins were grouped according to their taxonomic origin on broader taxonomic levels following the nomenclature of the NCBI taxonomy browser (http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/). There is no reason for this methodology not to identify single species in future. Especially for bacteria, a vast genomic information is available which readily allows to distinguish finer categories. For this task it is not necessary to recover a full set of proteins for any organism, but it is sufficient to recover one protein per organism that contains species specific information. In this study, proteins originating from bacteria, archaea, and viruses were not separated into further subgroups. Proteins from eucaryotes were sorted by their origin from green plants, metazoa, fungi, and "unicellular eucaryota" containing all those taxa that do not belong to the three major groups of eucaryota. Proteins from plants were further grouped into algae and vascular plants. Proteins from metazoa were classified into platyhelminthes, protostomia (annelid worms, insects, and mollusks), nematoda, and vertebrata (mammals, birds, reptiles amphibia, fish). In some cases, sequenced peptides identified highly conserved regions of proteins common to different taxonomic groups. These proteins were designated as 'not classified'. Protein sequences were tested for redundancy by alignment of all identified sequences. Sequences with identity greater than 98% were considered identical. The functional attributes of the identified proteins were classified following the Enzyme Commission Classification Scheme.

2.4. Future perspectives

The methods described here are certainly preliminary and need to be improved. Currently, we are developing strategies to purify and concentrate proteins directly from

BGD
1, 195–218, 2004

the DOC and water sources, in order to avoild freeze-drying. Also, improved sensitivity of MS instruments in combination with more thorough fragmentation (i.e. MS^3) will improve data interpretation.

3. Results and discussion

3.1. Analysis of organisms contributing to decomposition of plant material

Proteins were extracted from three independent samples collected from four different stages of decomposition in the organic layer of a permafrost soil in Central Siberia (Fig. 1). Plant proteins were indeed found to be the largest fraction in extracts from green plant material (a layer of green moss), but also in this layer the second largest fraction is consisting of bacterial proteins. The fraction of plant proteins decreased along the decomposition line, while the fraction of fungal proteins increased in deeper litter layers. In the L (litter) and FH ("foerna" and humus) horizons, fungi and bacteria are the dominating groups of protein origin. With increasing decomposition, the total number of identified proteins decreases and drops to 30% in the FH horizon compared to the total number of proteins identified in the green moss layer.

Although the result is not unexpected, the analysis of proteins at different stages of decomposing plant material is well suited to validate the methodology of protein detection and classification of organisms. Also the high fraction of bacterial proteins in protein extracts of green plant material is not surprising as bacteria are expected to be present in all environmental samples, but have not been recognized as such by ecologists. A thorough analysis of microbial biomass (g C/100 g soil) suggested a ratio of bacteria to fungi of 35/65 the FH layer in a natural beech forest (Ellenberg et al., 1986). The protein analysis revealed equal amounts of bacterial and fungal proteins in the FH layer. However, protein analysis is not a direct measure of biomass but rather emphasizes the actual turnover and metabolic activity of the respective phylogenetic group. The proteomic method applied here thus results in a fingerprint of a majority of

BGD
1, 195–218, 2004

phylogenetic groups present and active in the sample. This is different from most other approaches focusing on specific groups of organisms but overseeeing the diversity of the ecosystem compartment as such.

3.2. Protein origin of surface waters

Proteins analyzed from filtered freeze-dried samples of different surface water sources reveals a vast diversity of protein origin (Fig. 2). A peat bog lake was dominated by bacterial proteins, with only 22% of proteins originating from other organisms, mainly viruses, vertebrates, and protostomia. In total, proteins from seven phylogenetic groups were distinguished. An acidic creek, flowing through a bog region also showed a high fraction of bacterial proteins. The pattern of protein origin and protein amount of a small stream in the forest tundra was dependent on water flow levels. At low baseflow, only low amounts of protein was detected, and these originated only from four different phylogenetic groups. In contrast, after a summer stormflow, the amount of proteins increased by factor 7. The composition of proteins in the DOC of the stream water at stormflow was dominated by plant proteins, but proteins from a total of eight other phylogenetic groups were identified as well. The observed increase in protein abundance was accompanied by an increase in DOC from 16.9 $\mathrm{mgC\,L^{-1}}$ at low flow levels to 23.3 $\mathrm{mgC\,L^{-1}}$ at the intense stormflow. Finally, rain water contained only very few proteins originating from bacteria and viruses. These examples show that there is a strong variation in the phylogenetic origin of proteins as well as total number of proteins contributing to DOC of different surface waters at different seasonal conditions. Although a detailed interpretation of every difference between the various water sources is not possible at this stage, the analysis clearly shows that each environment has its proteomic 'fingerprint' which is important on a broader scale. For example the increase in total protein and especially of plant proteins after a stormflow can be interpreted by the accumulation of fresh plant material in the flood, and the absence of proteins from eukaryotic organisms in rain water is not surprising either. The decrease in protein amount in the stream at low water levels could indicate that protein turnover

is important. In this case, the water percolates from the organic layer (shown in Fig. 1) through the mineral soil until it appears in the stream. Along this path, proteins from the organic layer disappear and new proteins emerge.

3.3. Protomic fingerprint of different soil leachates

DOC of leachates from different soil types from temperate and arctic regions were investigated by proteomic analysis. Leachates from 5 cm depth of minera soil were analyzed from central European cambisol (beech forest on lime stone), dystric cambisol (acidic spruce forest on granitic soil), arenosol (acidic pine forest on sandstone), and histosol (sphagnum bog with Betula), and from gelic podzol originating from Siberian larch forest (Fig. 3). Total protein content of the different soil types was highest in the arenosol, whereas lowest protein amounts were found in the leachates of dystric cambisol of sub-monatane central European spruce forest and the Betula forest on Sphagnum bog. From all the distinguished taxonomic units, bacterial proteins were the highest fraction in all investigated soils and reached 80% in leachates of the arenosol. In other soil types, the fraction of bacterial proteins ranged from 30% to 45%. In soil leachates of cambisol and dystric cambisol a significant fraction of proteins originated from fungi and plants. Nematodal proteins were only found in the temperate forests but not the Siberian forest on permafrost.

Thus, the total protein content of DOC of soil leachates varies with climatic region and soil type. In addition, seasonal variations have been observed (Schulze et al., 2004). There are surprises, especially the strong dominance of bacterial proteins in leachates from arenosol of a pine forest needs to be analyzed in more detail.

3.4. Reproducibility of environmental proteomics

The analysis of environmental samples will be subject to variations between different sites and sampling times. In order to be able to meaningfully interpret the distribution of phylogenetic origin of proteins from DOC, the sample-to-sample variability was an-

alyzed. Samples from a central European beech forest (Hainich, Thuringia, Germany) were taken at the same time from three different sampling sites in a line 25 m apart from each other. All three samples were treated the same way for the protein analysis, and the same mass of DOC was used as of starting material. Although 30% more proteins were detected at site 2 compared to the other two sites, the pattern of protein origin was very similar between the three sites, indicating that the result of the proteomic analysis, the protein indentification and classification does not occur at random (Fig. 4), but the protein amounts reflect the small scale variability of soils in forests. Also the proteomic fingerprint of decomposing material (Fig. 1) was reproduced in three independent samples with the same distribution of taxonomic protein origin (data not shown). With improving sensitivity of protein mass spectrometry, and increasing efficiency of unambiguous protein indentification, a more detailed picture will emerge in future (Aebersold and Mann, 2003; Olsen et al., 2004).

3.5. Phylogenetic groups and protein number

The proteomic fingerprint of environmental samples is based on the classification of proteins into their taxonomic groups of origin. In a pioneer studied, only 12 different taxonomic categories were defined (archaea, bacteria, fungi, algae, plantae, platyhelminthes, nematoda, protostomia, vertebrata, viruses, "other eukaryotes", and unassigned proteins). Especially in samples with less than 20 identified proteins, the chance for proteins falling into different taxonomic groups increases with increasing protein number. However, in samples with 20 or more proteins the number of distinguished taxonomic groups ranged from 6 to 10 independent of protein number (Fig. 5). Only in one case, a sample from soil leachate of a beech forest (S2 in Fig. 4), all 12 distinguished taxonomic units were recognized. The taxonomic classification in these first studies was rather borad, as the genomic sequence information available in databases is still limiting. However, it has been shown that the diversity of larger taxonomic units correlates with species diversity (Báldi, 2003). Thus, the broad taxonmic classification applied here might be used as indication for broad taxonomic organization of an

ecosystem.

3.6. Protein size distribution and protein function

The size distribution of proteins identified from DOC peaks at 30 to 50 kDa (Fig. 6). This is in agreement of calculated average protein sizes for E.coli of 35 kDa and 51.8 kDa for human (Cagney et al., 2003). Since the protein size distribution in DOC is not changed from that of the living organisms, we believe that degradation and adsorption of proteins is not dependent on protein size, and that no bias of sample preparation and mass spectrometric analysis has been introduced.

The bacterial proteins identified were classified according to their cellular function. Most proteins identified were ribosomal proteins, followed by an additional large group of metabolic enzymes (Fig. 7). Protein synthesis and energy metabolism are also the two categories containing most proteins in a functional classification of proteomes of different organisms (The Arabidopsis Genome Initiative, 2000), indicating that the proteomic DOC analysis indeed reflects the the original protein composition of organisms. The apparent changes in the abundance of ribosomal proteins that can be observed between summer and winter (Fig. 7) cannot be readily explained and needs further investigation. It may indicate a lower growth activity of bacteria in winter and thus mirror ecosystem activity.

A key question remains as to whether any of the enzymes identified in DOC may still be functional and be involved in geochemical processes. Just from the annotation of the identified proteins, none of the enzymes derived from DOC are true extracellular enzymes with possible biodegradative activities. However, in a more thorough analysis of proteins bound to soil particles, celllulases, laccases, and collagenases could be identified (see Table 1; Schulze, 2004, #1588). This indicates that there may well be 'functional' compartments in the soil with biodegradative activities attached to organic films of mineral particles. This is in agreement with enzyme activity tests indicating biodegradative activity in association with soil particles (Stemmer et al., 1998; Kjöller et al., 2000).

3.7. Interpretation of the data with respect to biodiversity

The results show that proteomic analysis of DOM in soil water is possible, and that DOM from different ecosystems clearly differ in their proteomic composition. In addition, it is shown that the size distribution of analyzed proteins reflects the size distribution of proteins in organisms, and also the functional classification of proteins is in agreement with protein classification within an organism. These are indications that the proteomic analysis of DOM can also be interpreted as a measure of relative abundance of proteins originating from phylogenetic groups. This interpretation is highly supported by the analysis of protein extracts from different layers of decomposing plant material, with a gradual increase in bacterial and fungal proteins, while plant proteins decrease. As the contribution of different phylogenetic groups to the protein pool varies between ecosystems and with soil depth, it emerges that each ecosystem has its characteristic 'proteomic fingerprint' reflecting the relative abundance or activity of different groups.

The analysis of species distribution and abundance in ecosystems has been of interest for ecologists for a long time, and remains important even today when characterizing different communities (Ellenberg et al., 1986; Ellwood and Foster, 2004). In a thorough survey of animal and plant species distribution, their interactions, and dynamics, a massive 'inventory' of a beech forest on sandstone in the Solling, Germany, was carried out over 20 years of investigation (Ellenberg et al., 1986). Displaying the data of this long-term thorough analysis in a similar way as the 'proteomic fingerprint' shows that each counting method emphasizes different groups of organisms. Displaying cell numbers of soil microorganisms gives a high number for bacteria, although they contribute only 30% of the biomass of soil microorganisms (Fig. 8). Including trees into a biomass assay will result in a different dominance than focusing on soil organisms and herbaceous vegetation only. The proteomic fingerprint of ecosystems is an important additional way of displaying contribution of different organisms to an ecosystem, because it has the chance to contain the entire organismic set of a system, which is usually not the case in diversity related ecosystem research. Just as counting or weigh-

ing has a its specific focus and bias, the proteomic fingerprint has a focus to picture the complete organismic food web, it may be biased towards emphasis of metabolically active organisms with a high cellular turnover rate (i.e. bacteria).

One limiting factor of the proteomic fingerprint currently is its rather low resolution of distinction of phylogenetic groups. Current efforts of sequencing DNA samples extracted from environments (Venter et al., 2004) are encouraging and will provide a basis for more accurate protein identifications and possibly will allow finer distinction of organisms. It could be demonstrated using an experimental dataset that cross-species protein identification by mass spectrometry (e.g. MS-BLAST) successfully identifies over 80% of the proteins by sequence similarity searches, because orthologue proteins share substantial sequence identity (Habermann et al., 2004). With improving sensitivity of protein mass spectrometry, and improved methods to purify proteins from environmental sources a more detailed picture may emerge. Peptide mass fingerprinting of tryptic digests of bacterial spores (Dickinson et al., 2004a, b) or mass spectrometric analysis of whole cells (Arnold and Reilly, 1998) emerge as a novel and more rapid tool to specifically distinguish microorganisms at the sub-species level. The advantage of protein analysis is in the rapid identification of taxonomic units over a broad range of the phylogenetic tree. In fact it may be the only method available to identify the whole organismic set-up in a very small sample, and in a very short time. Thus, we believe that the method of proteomic fingerprinting of environmental samples can become an important approach to compare environments and their seasonal changes. Most importantly, proteins mirror the active component of an ecosystem that results in the presence of DOM. Thus protein identification may improve our understanding of soil organic chemistry. Although in the examples analyzed here, finding of biodegradative enzymes seems like searching a needle in the haystack, methods of immunprecipitation or affinity purifications can serve as tools to study specific enzymes of interest.

BGD

1, 195–218, 2004

4. Conclusions

In conclusion, mass spectrometric analysis of proteins in DOC or from soil particles opens a new way of describing the biology of environments. With the possibility to also obtain information about potentially active components of DOC, environmental proteomics may become a powerful tool to biogeochemical processes in future. Although the approach has been used here for natural or semi-natural ecosystems, it is obvious, that the methodology may be very powerful to characterize effects of management on biodiversity. It may also have a potential in identifying organisms which are important in bio-degradation of environmental spills.

Acknowledgements. I would like to thank G. Gleixner, K. Kaiser, A. Prokushkin, and K. Kalbitz for donating DOC samples, and I am very grateful to E.-D. Schulze for critical comments on the manuscript.

References

Aebersold, R. and Mann, M.: Mass spectrometry-based proteomics, Nature, 422, 198–207, 2003.

Almendros, G., Frund, R., Gonzalez-Vila, F. J., Haider, K. M., Knicker, H., and Ludemann, H. D.: Analysis of 13C and 15N CPMAS NMR-spectra of soil organic matter and composts, FEBS Lett., 282, 119–121, 1991.

Arnold, R. J. and Reilly, J. P.: Fingerprint mathcing of E.coli strains with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry of whole cells using a modified correlation approach, Rapid Commun. Mass Spectrom., 12, 630–636, 1998.

Báldi, A.: Using higher taxa as surrogates of species richness: a study based on 3700 Coleoptera, Diptera and Acri in Central-Hungarian reserves, Basic and applied Ecology, 4, 589–593, 2003.

Cagney, G., Amiri, S., Premawaraadena, T., Lindo, M., and Emili, M.: In silico proteome analysis to facilitate proteomics experiments using mass spectrometry, Proteome Science, 1, 1–15, 2003.

Dickinson, D. N., La Duc, M. T., Satomi, M., Wienfordner, J. D., Powell, D. H., and Venkateswaran, K.: MALDI-TOFMS compared with other polyphasic taxonomy approaches for the identification and classification of Bacillus pumilus spores, J. Microbiol. Meth., 58, 1–12, 2004a.

Dickinson, D. N., La Duc, M. T., Haskins, W. E., Gornushkin, I., Winefordner, J. D., Powell, D. H., and Venkateswaran, K.: Species differentiation of a diverse suite of bacillus spores by mass spectrometry-based protein profiling, Apppl. Environ. Microbiol., 70, 472–482, 2004b.

Ellenberg, H., Mayer, R., and Schauermann, J.: Ökosystemforschung – Ergebnisse des Sollingprojekts, Stuttgart: Ulmer Verlag, 1986.

Ellwood, M. D. F. and Foster, W. A.: Doubling the estimate of invertebrate biomass in a rainforest canopy, Nature 429, 459–551, 2004.

Gleixner, G., Czimczik, C., Kramer, C., Lühker, B., and Schmidt, M. W. I.: Plant compounds and their turnover and stability as soil orgainc matter, In Global Biogeochemical cycles in the climate system, edited by Schulze, E. D., Heimann, M., Harrison, S., Holland, E., Lloyd, L., Prentice, I. C., and Schimel, D., San Diego: Academic Press, 201–216, 2001.

Habermann, B., Oegerma, J., Sunyaev, S., and Shevchenko, A.: (The power and the limitations of cross-species protein identification by mass spectrometry-driven sequence similarity searches, Mol. Cell. Proteomics, 3, 238–249, 2004.

Kaiser, K., Guggenberger, G., Haumaer, L., and Zech, W.: Seasonal variations in the chemical composition of dissolved organic matter in organic forest floor layer leachates of old-growth Scots pine (Pinus sylvestris L.) and European beech (Fags sylvatica L.) stands in northeastern Bavaria, Germany, Biogeochem., 55, 103–143, 2001.

Kjöller, A., Miller, M., Struwe, S., Wolters, V., Pflug, A.: Diversity and role of microorganisms, Ecological Studies, 382–404, 2000.

Michalzik, B. and Matzner, E.: Dynamics of dissolved organic nitrogen and carbon in a Central European Norway spruce ecosystem, Eur. J. Soil Sci., 50, 579–590, 1999.

Olsen, J. V., Ong, S.-E., and Mann, M.: Trypsin cleaves exclusively C-terminal to arginine and lysine residues, Mol. Cell. Proteomics, 3, 608–614, 2004.

Perkins, D. N., Pappin, D. J. C., Creasy, D. M., and Cottrell, J. S.: Probability-based protein identification by searching sequence databases using mass spectrometry data, Electrophoresis, 20, 3551–3567, 1999.

Schulze, W., Gleixner, G., Kaiser, K., Guggenberger, G., Mann, M., and Schulze, E. D.: A proteomic fingerprint of dissolved organic carbon and soil particles, Oecologia, in press,

2004.
Stemmer, M., Gerzabek, M. H., and Kandeler, E.: Organc matter and enzyme activity in particle-size fractions of soils obtained after low-energy sonication, Soil Biol. Biochem., 30, 9–17, 1998.
The Arabidopsis Genome Initiative: Analysis of the genome sequence of the flowering plant Arabidopsis thaliana, Nature, 408, 796–816, 2000.
Tyers, M. and Mann, M.: From genomics to proteomics, Nature, 422, 193–197, 2003.
Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M., Solovyev, V. V., Rubin, E. M., Rokhsar, D. S., and Banfield, J. F.: Community structure and metabolism through reconstruction of microbial genomes from the environment, Nature, 428, 37–43, 2004.
Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D. B., Eisen, J. A., Wu, D., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parson, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y.-H., and Smith, H. O.: Environmental genome shotgun sequencing of the Sargasso Sea, Science, 306, 66–74, 2004.

Table 1. Enzymes involved in degradation of soil organic matter. The proteins were identified after elution of organic material from soil micro-particles.

Protein name	Species	Accession no. Enzyme no.
Cellulase	Clostridium acetobutylicum	gi|15893851
	Clostridium cellulovorans	gi|584895
	Myxobacter sp.	gi|6606317
	Irpex lacteus	gi|4586347
Collagenase	Desulfitobacterium hafniense	gi|23112072
	Porphyromonas gingivalis	gi|464477
Laccase	Trametes versicolor	gi|101946
	Thanatephorus cucumeris	gi|2147619

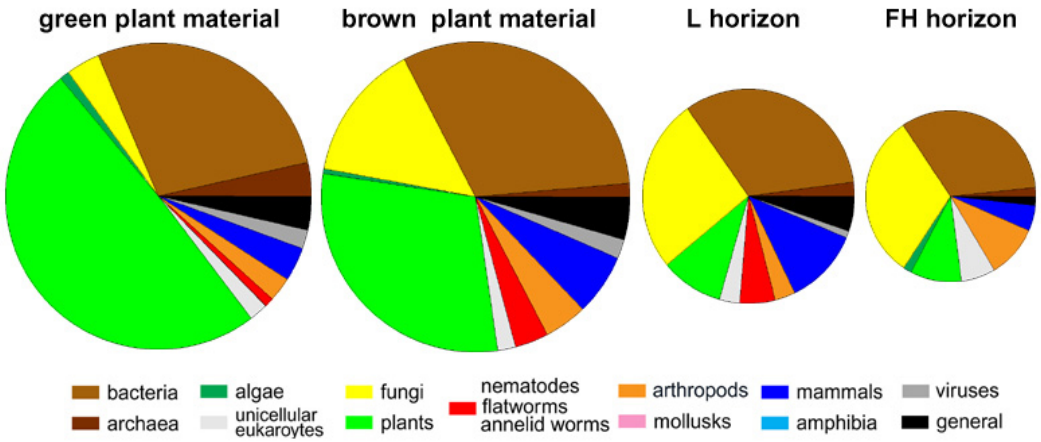

Fig. 1. Phylogenetic distribution of proteins extracted from decomposing plant material at different layers ranging from a green moss layer to the FH horizon. Areas of the pie charts represent the number of proteins identified.

BGD
1, 195–218, 2004

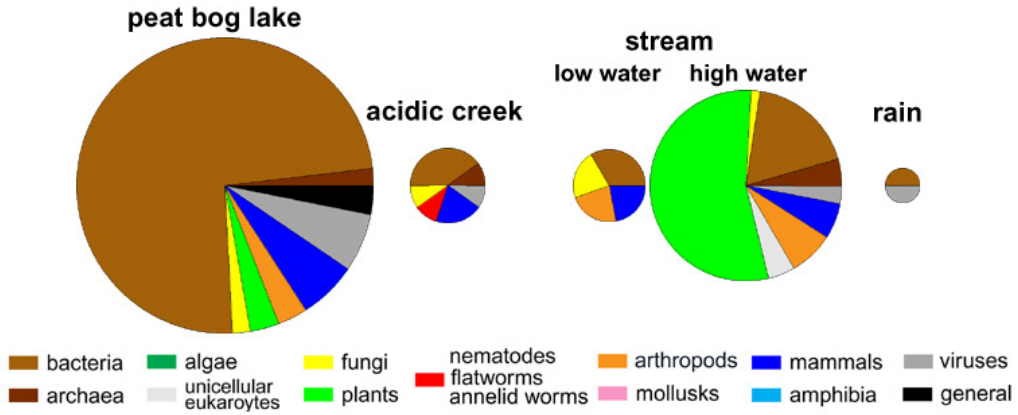

Fig. 2. Phylogenetic distribution of proteins identified in different sources of surface water, such as a peat bog lake, a creek in a bog area, a stream at low and high water and rain collected in a rain collector. Areas of the pie charts represent the number of proteins identified.

BGD
1, 195–218, 2004

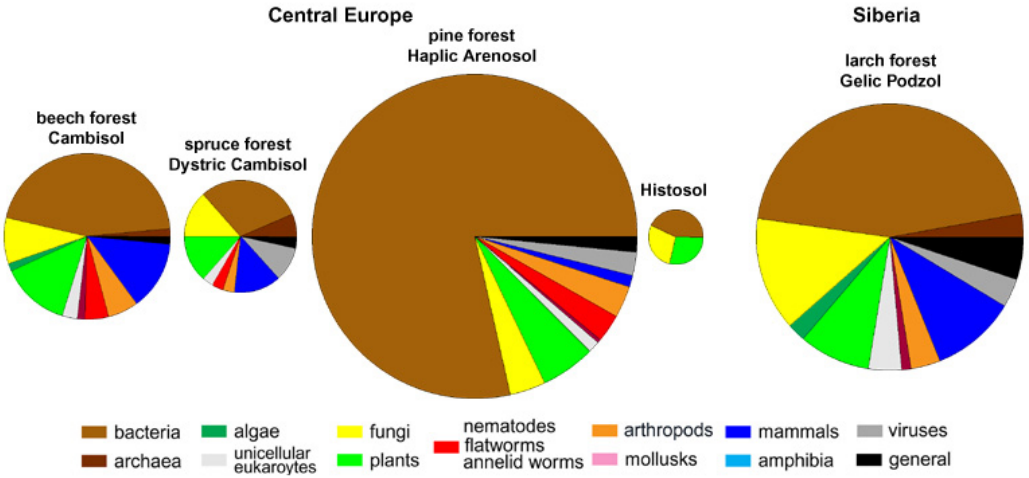

Fig. 3. Phylogenetic distribution of proteins identified from DOM leachates of different soil types. Areas of the pie charts represent the number of proteins identified.

BGD
1, 195–218, 2004

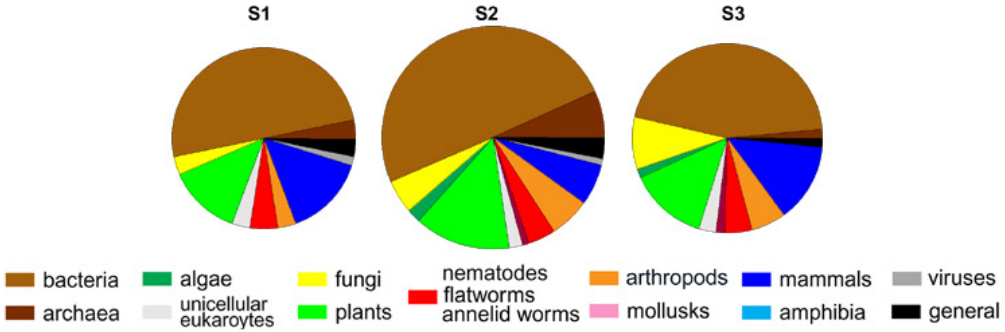

Fig. 4. Phylogenetic distribution of proteins identified from soil leachates along a transect in a beech forest on cambisolic soil. The distance between two samples was 25 m. Areas of the pie charts represent the number of proteins identified.

BGD
1, 195–218, 2004

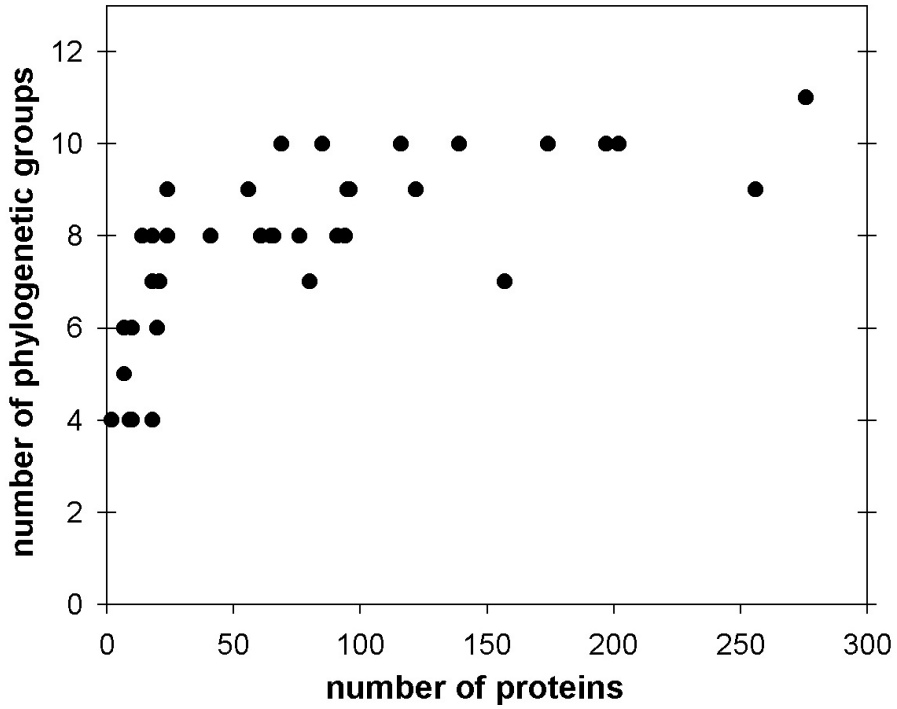

Fig. 5. The number of phylogenetic groups distinguished and the number of total proteins idenified for each sample analyzed. The number of phylogenetic groups distinguished does not depend on the number of proteins identified.

BGD
1, 195–218, 2004

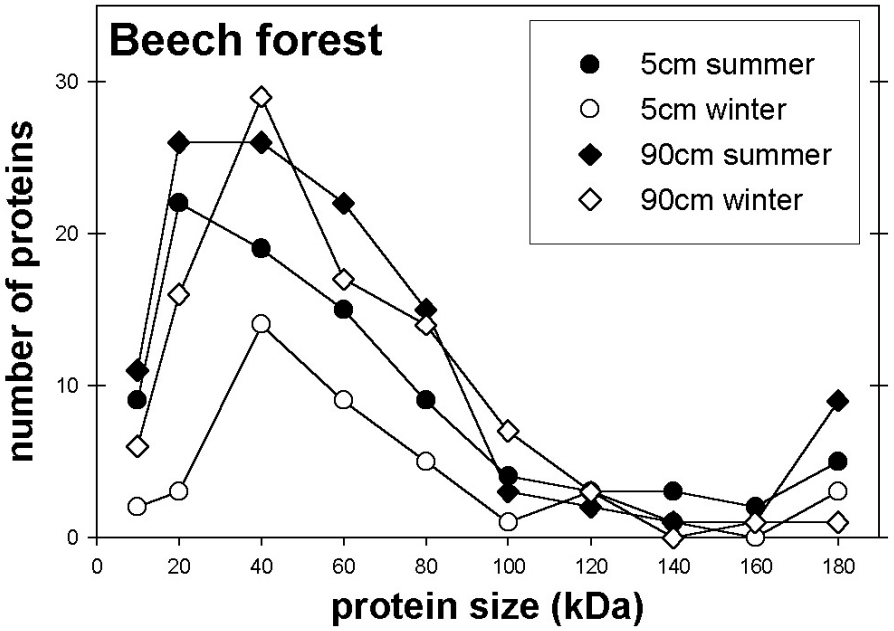

Fig. 6. Size distribution of identified proteins from DOM leachates of a beech forest on rendzic leptosol. Size distributions are shown for two different soil depths (5 cm and 90 cm) and seasons (summer and winter). Proteins from all organisms were included in the size distribution.

BGD
1, 195–218, 2004

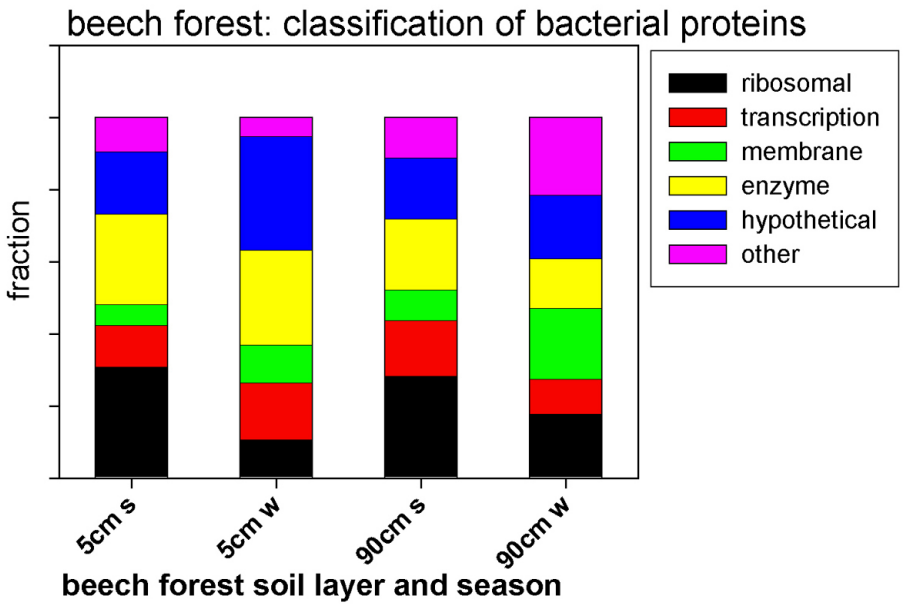

Fig. 7. Functional classification of the bacterial proteins identified from leachates of a beech forest on rendzic leptosol for two different soil depths (5 cm and 90 cm) and seasons (summer and winter).

BGD
1, 195–218, 2004

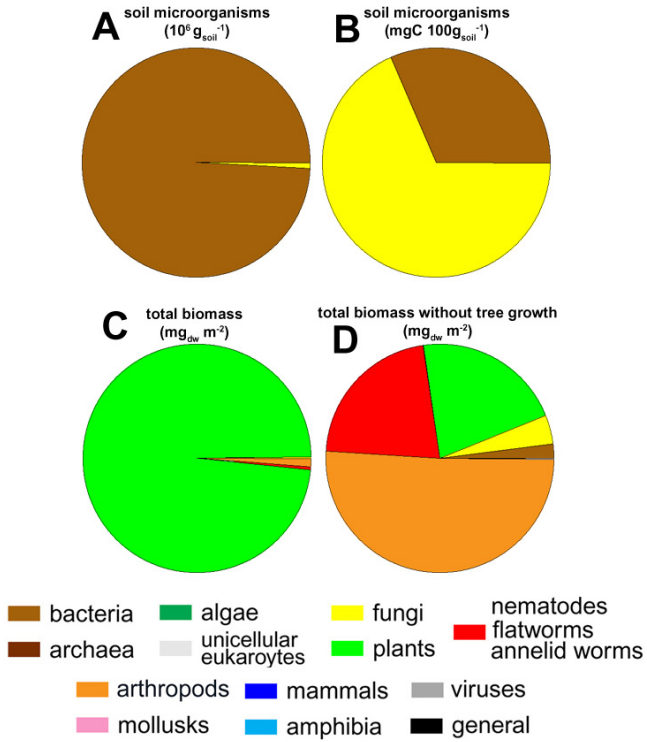

Fig. 8. Different ways of displaying distributions of phylogenetic groups assessed during the long-term study in the Solling, a beech forest on cambisolic soil (Ellenberg et al., 1986). **(A)**: distribution of soil microorganisms according to cell number. **(B)**: distribution of soil microorganisms according to biomass. **(C)**: Total biomass per square meter. The biomass of trees was included as annual growth rate. **(D)**: Total biomass per square meter, not taking into account the trees, but only herbaceous vegetation. The figures were drawn from numbers published in form of tables. Microorganism biomass per square meter was calculated from the published values in mgC per 100 g soil by using 34.82 g soil m^{-2} measured at a depth of 0–5 cm in cambisolic soil of the Hainich, Germany.

BGD
1, 195–218, 2004