

ORIGINAL ARTICLE

Functional metaproteome analysis of protein extracts from contaminated soil and groundwater

Dirk Benndorf¹, Gerd U Balcke², Hauke Harms³ and Martin von Bergen¹¹Department of Proteomics, UFZ-Helmholtz Centre for Environmental Research, Leipzig, Germany;²Department of Hydrogeology, UFZ-Helmholtz Centre for Environmental Research, Leipzig, Germany and³Department of Environmental Microbiology, UFZ-Helmholtz Centre for Environmental Research, Leipzig, Germany

Using proteins from soil or groundwater as functional biomarkers requires efficient extraction. We developed an extraction method in which the separation of proteins from the inorganic and organic constituents of the soil matrix was achieved by a combination of 0.1 M NaOH treatment and phenol extraction. Incubation with NaOH released humic acids and proteins from soil minerals, and simultaneously, disrupted microorganisms. The subsequent phenol extraction separated the proteins from the humic organic matter. Protein extracts were applied to sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and 2D-electrophoresis (2-DE). Spots and bands were excised and individual proteins identified by liquid chromatography online linked to mass spectrometry (MS) via electrospray ionization source (LC-ESI-MS). To assess the suitability of the method for the functional analysis of environmental metaproteomes, it was applied to soil that had been enriched in chlorophenoxy acid-degrading bacteria by incubation with 2,4-dichlorophenoxy acetic acid (2,4-D) for 22 days. The method was also used to analyze groundwater from the aquifer of a chlorobenzene-contaminated site. The identification of enzymes such as chlorocatechol dioxygenases was consistent with bacterial metabolic pathways expected to be expressed in these samples. The protocol enabled thus the analysis of the metaproteome of soil and groundwater samples. It thereby provides a means to study the diversity of environmental microbial communities while addressing functional aspects more directly than metagenome or even metatranscriptome analysis.

The ISME Journal (2007) 1, 224–234; doi:10.1038/ismej.2007.39; published online 31 May 2007

Subject Category: integrated genomics and post-genomics approaches in microbial ecology

Keywords: environmental proteomics; 2D-electrophoresis (2-DE); mass spectrometry (MS); chlorophenoxy acids; chlorobenzene

Introduction

Various strategies were used to study the relationship between the ecosystem functioning and the structure of microbial communities. Major goals of these efforts are to attribute key functions to specific community members and, in view of the ecosystem stability, to reveal cooperation between community members and functional redundancies. Besides measuring enzyme activities, respiration rates and metabolite concentrations (Landi *et al.*, 2000; Moreno *et al.*, 2001), nucleic acids (Eilers *et al.*, 2000) and lipids (Fredrickson *et al.*, 1995) are often used as

markers for microbial identity, and to a certain extent, metabolic potential and its expression in environmental samples (Kanagawa, 2003; Fields *et al.*, 2006). Structural data obtained from lipids are based on the occurrence of fatty acids or quinones that are specific for certain taxa. However, there are no specific markers for all known bacteria. Using phospholipid fatty acids (PLFA), a metabolic function is only inferred from the presence of microorganisms known for that function. In comparison to lipids, the amplification of structural and functional genes can be used as a good indicator for the presence of microorganisms and their metabolic potential. Meanwhile, DNA and RNA from environmental samples can be quantified in one step using microarray with several thousands of probes for known genes and pathways involved in biodegradation and metal resistance (Zhou, 2003; Rhee *et al.*, 2004). Although the analysis of thousands of genes from environmental samples is already possible, the selection of probes in advance imposes a restriction

Correspondence: Dr D Benndorf, Department of Proteomics, UFZ-Helmholtz Centre for Environmental Research, Permoserstr. 15, 04318 Leipzig, Germany.

E-mail: dirk.benndorf@ufz.de

Received 15 February 2007; revised and accepted 26 April 2007; published online 31 May 2007

to genes already known. Gene probes are also used for a combination of fluorescence *in situ* hybridization (FISH) and microautoradiography (MAR) to assign structural and functional genes to single cells of microbial communities (Wagner *et al.*, 2006). In liquid samples, population dynamics can be followed by multiparametric flow cytometry and phylogenetic identification of subcommunities by 16S ribosomal RNA (16S rRNA) gene sequencing (Kleinstaub *et al.*, 2006).

Another strategy to assign activity to microorganisms is the use of substrates labeled with stable isotopes. The stable isotopes are predominantly incorporated into active biomass. Its enrichment in PLFA (Boschker *et al.*, 1998) and DNA (Radajewski *et al.*, 2000) of active microorganisms allows their differentiation from inactive community members. However, intermediates of the labeled substrate added can be also used by other species that are not involved in the pathway and may cause false results (Manefield *et al.*, 2006).

In comparison to lipids and nucleic acids, proteins are promising alternative markers since they reflect the actual functionality with respect to metabolic reactions and regulatory cascades, and give more direct information about microbial activity than functional genes and even the corresponding messenger RNAs (mRNAs; Wilmes and Bond, 2006). In other respect, the use of proteins also bears the potential to reveal the identity of the active microorganisms via database analysis using the level of homology to other species. Furthermore, PCR-based techniques in nucleic acid detection are often biased by their preference for some templates, and thus, not ideal for quantitative analysis (Kanagawa, 2003). Finally, regulation of metabolic processes can take place at the post-transcriptional level, hence uncoupling quantities of mRNA transcripts and actual activity. Thus, the presence of specific proteins in environmental samples is a potentially reliable indicator for microbial function.

Meanwhile, a variety of powerful methods enables to visualize and identify the proteome of pure microbial cultures and also increasing proteomes of microbial communities are analyzed. Proteomics has become a useful complement of functional genomics. The availability of numerous genomes triggered this development by facilitating the identification of proteins, although the available genome data are still likely to represent only a tiny fraction of the genetic information present in environmental samples. The most widely used approach in proteomics is the separation of proteins by 2D-electrophoresis (2-DE), which combines isoelectric focusing (IEF) in the first dimension and sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) in the second dimension (Klose, 1975; O'Farrell, 1975; Görg *et al.*, 1988). This separation is followed by in-gel digestion of selected spots and protein identification by mass spectrometry (MS). In the case of proteins from fully sequenced organisms,

the identification is achieved by data bank comparison of peptide mass fingerprint (matrix assisted laser desorption ionization time of flight MS (MALDI-ToF-MS)) or fragmentation data of the peptides (electrospray ionization source tandem mass spectrometry (ESI-MS/MS) or MALDI-ToF/ToF-MS). Identification of peptides from unsequenced species can be achieved in two ways; either the peptides and their MS/MS fragments resemble those of known proteins (in the case of high homology) or the peptides are sequenced by MS and exhibit sufficient homology to published sequences (in the case of lower homology). Owing to the ever increasing number of fully sequenced genomes, there is a fairly good chance that homologies can be found. A prerequisite for this procedure is high quality MS data requiring pure samples containing enough copies of every protein.

The preparation of protein extracts is thus of paramount importance. This is the main reason why proteomics was restricted mostly to microbial isolates in culture. In fact, its application to soil requires considerably improved protocols of protein extraction and sample preparation (Nannipieri, 2006; Ogunseitan, 2006) since these are very critical steps for gaining high resolution and sensitivity of 2-DE. The extraction of water-soluble proteins from soil has been proposed recently (Schulze *et al.*, 2005). In this approach, minerals were dissolved by hydrofluoric acid and the extracted proteins identified by MS coupled to liquid chromatography. Methods based on freeze-thawing or bead beating of soil for extracting intra- or extracellular proteins were also proposed (Singleton *et al.*, 2003). The greatest difficulty lies in the separation of proteins from humic acids. Another environmental medium of emerging importance is groundwater due to the increasing need for high quality drinking water. Here, the challenge consists in the low abundance of biomass. Until now, several authors presented metaproteome data from different environmental systems: soil particles (Schulze *et al.*, 2005), activated sludge (Wilmes and Bond, 2004), biofilms (Ram *et al.*, 2005) and seawater (Kan *et al.*, 2005). Additionally, metaproteomics was used to determine the stress response of mixed cultures after exposure to cadmium (Lacerda *et al.*, 2007).

The objectives of this paper were (i) to develop an extraction protocol to allow the analysis of the protein composition of environmental samples by a combination of SDS-PAGE or 2-DE and liquid chromatography online linked to MS/MS via electrospray ionization source (LC-ESI-MS/MS) and (ii) to use the metaproteomic information to attribute metabolic functions to community members, that is to link structural and functional aspects of the microbial diversity of these ecosystems. The method was applied to soil microcosms that had been enriched in inoculated chlorophenoxy acid-degrading bacteria by incubation with 2,4-dichlorophenoxy acetic acid (2,4-D) and to groundwater from the

Table 1 Experimental conditions used for soil percolation experiments and in the groundwater treatment reactor

	Soil column 1	Soil column 2	Groundwater treatment reactor
Column material	Columns filled with 220 g compost soil		Columns filled with lignite-containing glacial sandy material (porosity 0.2)
Treatment	Continuous percolation with mineral medium containing 2 mM 2,4-D (2,4-dichlorophenoxy acetic acid). Final flow rate: 6 ml h ⁻¹		Continuously recharged with anoxic, contaminated groundwater (15 to 20 mg l ⁻¹ chlorobenzene) from the site at an average linear velocity of 1.4 m day ⁻¹
Community	Autochthonous community	Inoculated with a mixture of 2,4-D-degrading bacteria	Autochthonous community
Sample	5 g soil per column		1 l water from effluent

chlorobenzene-contaminated aquifer of Bitterfeld (Saxony-Anhalt, Germany), where chlorobenzene is aerobically degraded by a microbial community of known composition but of unknown roles of its individual community members. Both environments are examples for ecosystems contaminated with organic compounds and they were chosen to demonstrate that metaproteomics may be used for indication of biodegradation processes.

Experimental procedures

Enrichment of chlorophenoxy acid-degrading bacteria in soil percolation experiments

Two columns (Table 1) of 250 ml volume were filled with 220 g compost soil from Leipzig (Germany), which is rich in humic organic matter. The first column was inoculated with 50 ml of a mixture of *Cupriavidus necator* JMP134, *Rhodoferrax* sp P230 and *Sphingomonas herbicidovorans* B488 (30 mg dry weight) previously grown on succinate as carbon and energy source. In the second column, no bacteria were added and only the indigenous microorganisms were present as starter culture. The columns were percolated with minimal medium (Muller and Babel, 1986), containing 2 mM 2,4-D at a flow rate of 3 ml h⁻¹. After 3 days, the flow was increased to 6 ml h⁻¹ and after 5 days to 12 ml h⁻¹. After 7 days, the flow rate was decreased to 6 ml h⁻¹ and remained the same until the end of the experiment after 22 days. The concentration of 2,4-D was determined by HPLC separation and UV-detection as described previously (Oh and Tuovinen, 1990). Five grams of soil were sampled after homogenization of the column content.

Treatment of groundwater samples

The groundwater (Table 1) originated from the effluent of a field reactor (0.6 m diameter × 12 m length) filled with lignite-containing glacial sandy material (porosity 0.2) of the quaternary aquifer at Bitterfeld, Germany. The reactor was continuously recharged with anoxic, contaminated groundwater (15 mg l⁻¹ to 20 mg l⁻¹ chlorobenzene) from the site at an average linear velocity of 1.4 m day⁻¹. During the sampling period, a short pulse of pure oxygen

gas was sparged daily into the interstitial voids of the reactor material. Upon gas injection, the oxygen gas dissolved creating intermittently aerobic conditions for several hours throughout the full length of the reactor. Cellular biomass (motile and sheared-off cells from biofilms covering the reactor filling) of 1 l reactor effluent was obtained by centrifugation (16 000g, 10 min, 4°C). The cells were collected on filters, stained with 4',6-diamidino-2-phenylindole (DAPI) and counted using fluorescence microscopy (Vogt *et al.*, 2004).

Protein extraction and purification

The original compost soil was used to develop the extraction and purification method applied later on also for groundwater samples with low protein contents. This led to the following protocol: 5 g of the original compost soil or the soils from percolation experiments were treated with 10 ml 0.1 M NaOH for 30 min (Figure 1). The suspension was centrifuged 10 min at 16 000g and 20°C. About 6 ml of supernatant were mixed with 16 ml liquid phenol (10 g phenol and 1 ml water) and 10 ml water and shaken for 1 h at 20°C. Afterwards, the phases were separated by centrifugation (10 min at 14 000g). About 15 ml of the lower phenol phase were collected and washed by mixing with 15 ml water, followed by 5 min shaking and subsequent centrifugation. The proteins in the phenol phase (15 ml) were precipitated with the 5-fold volume of 0.1 M ammonium acetate in methanol at -18°C, overnight. Then, the sample was centrifuged (10 min, 16 000g, 0°C), the pellet was suspended by ultrasonication in 10 ml 0.1 M ammonium acetate in methanol, incubated 15 min at -18°C and centrifuged again. The pellet was successively washed in 2 ml 0.1 M ammonium acetate in methanol, 2 ml 80% acetone, 2 ml 70% ethanol, each washing step including 15 min incubation and subsequent centrifugation.

The cell pellet from the groundwater was incubated with 400 µl 0.1 M NaOH (1400 min⁻¹, 20°C). The sample was centrifuged (10 000g, 10 min, 4°C) and 400 µl supernatant was mixed with 900 µl liquid phenol and 400 µl distilled water. The phases were separated by centrifugation (10 min at 16 000g) and protein in the phenol phase (1 ml) was precipitated

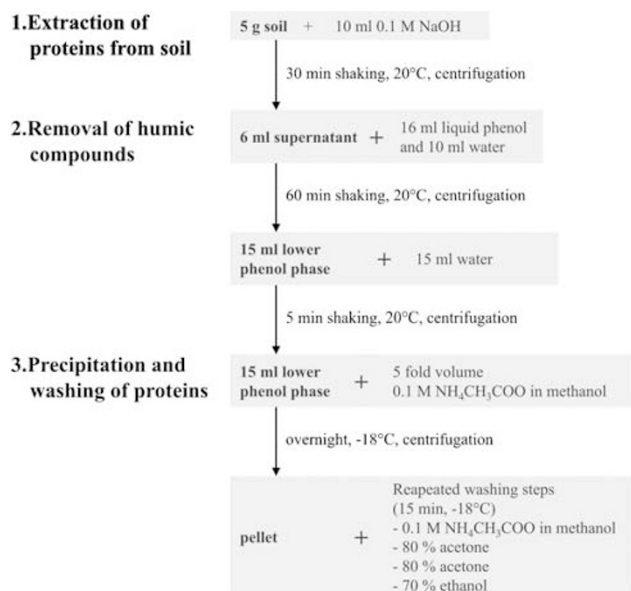


Figure 1 Scheme for extraction and purification of proteins from soil.

with the 5-fold volume 0.1 M ammonium acetate in methanol at -18°C , overnight. Afterwards, the pellet was resuspended and washed as described above.

SDS-PAGE and 2D-PAGE of extracted proteins

Extracted proteins were separated by SDS-PAGE (Laemmli, 1970) or 2D-polyacrylamide gel electrophoresis (2D-PAGE; Klose, 1975; O'Farrell, 1975; Görg *et al.*, 1988). Before electrophoresis, the protein pellets were dissolved in 10 μl deionized water by sonication (sonication bath, 5 min). For SDS-PAGE, 2 μl to 5 μl solubilized proteins were mixed with sample buffer (Laemmli, 1970), incubated 5 min at 90°C and loaded on SDS gels (4% stacking gel, 12% separating gel). For 2D-PAGE, 8 μl of solubilized protein were mixed with 50 μl IEF sample buffer (8 M urea, 2 triton X100, 2% CHAPS, 0.28 dithiothreitol (DTT), 0.5% IPG-buffer pH 3-10 NL and bromophenol blue) and treated as described previously (Benndorf *et al.*, 2006). After electrophoresis, gels were stained with colloidal Coomassie brilliant blue and dried in a stream of unheated air.

Identification of proteins

For identification of proteins from SDS-PAGE of soil or groundwater samples, the complete lanes were divided into bands and digested overnight with trypsin (Santos *et al.*, 2004). For identification of proteins from 2D gel of groundwater, the most abundant spots were excised and digested overnight. The extracted peptides were separated by reversed-phase nano-LC (LC1100 series, Agilent Technologies, Palo Alto, CA, USA; column: Zorbax 300SB-C18, 3.5 μm , 150 \times 0.075 mm; eluent: 0.1%

formic acid, 0% to 60% acetonitrile) and analyzed by MS/MS (LC/MSD TRAP XCT mass spectrometer, Agilent Technologies). Database searches were carried out with MS/MS ion search (MASCOT, <http://www.matrixscience.com>) against NCBIInr. Alternatively, the mass spectra were searched against the NCBIInr using the software Spectrum Mill (Agilent Technologies).

Results and discussion

Using proteins as biomarkers requires their extraction and purification from environmental samples. Our objective was thus to develop an extraction protocol for a wide range of samples from soil, rich in humic organic matter, to groundwater, comprising only low amounts of protein. Since organic soil is the most complex of these matrices, initially an extraction and purification protocol suitable for compost soil was developed and later adapted to aqueous samples. There are basically two different strategies for protein extraction from soil. Whereas some groups used buffers containing salt and detergents to release proteins from the organic and inorganic components of the soil matrix (Rigou *et al.*, 2006), others extracted proteins by dissolving the soil mineral matrix in concentrated HF (Schulze *et al.*, 2005). Under more rigorous conditions more proteins are extracted, but also more contaminating humic compounds, which make further purification necessary. Our protocol consists of unselective extraction, protein purification and precipitation (Figure 1). An initial extraction step with 0.1 M NaOH is widely used to extract humic compounds from soil (isolation of International Humic Substances Society (IHSS) soil fulvic and humic acids, <http://www.ihss.gatech.edu/soilhafa.html>). This method has the side effect (often in combination with SDS) to lyse efficiently bacteria for subsequent extraction of protein (Guerlava *et al.*, 1998). The dark-brown color of NaOH extracts from untreated compost soil indicated that besides intracellular proteins of soil microorganisms, high amounts of humic organic matter were eluted, which we attempted to remove by a water and liquid phenol two phase extraction that had been used to purify proteins from lysates of olive leaves (Wang *et al.*, 2003) and fungal spores (Benndorf, unpublished results). Phenol is also used to remove proteins in DNA and RNA purification protocols (Kirby, 1956), because its gentle hydrophobicity makes it a better solvent for proteins than water. After phase separation, most of the brown color was found in the water phase, which was separated from a clear light yellow phase by an opaque creamy inter-phase. After precipitation of the contents of all phases, nearly the entire pellet obtained from the phenol phase was soluble in sample buffer for SDS-PAGE, whereas the pellets from the water phase and the inter-phase contained much insoluble

dark-brown material. During SDS-PAGE of the soluble material of all phases, some material precipitated in the sample slots and produced brown smear over the entire length of the lanes (not shown). A subsequent Coomassie stain also detected the brown substances, thus confirming that this dye visualizes humic compounds (Aoyama, 2006). The phenol phase contained significantly less brown smear showing that most humic compounds were successfully removed. These findings correspond to the data of (Simonart *et al.*, 1967) who used a similar procedure for enrichment of bulk soil proteins. However, even in the lane of the phenol phase there was no defined band visible. A likely reason is the lack of dominant organisms and enzymes in the highly complex and heterogeneous control soil (Singleton *et al.*, 2003). Another possibility is that our procedure also extracted large amounts of extracellular protein from soil particles (Nannipieri, 2006) which is presumably also heterogeneous and prevented the detection of bands. To overcome this problem and to introduce target proteins for metaproteome analysis, we enriched the soil with autochthonous chlorophenoxy acid-degrading bacteria by percolating the soil with 2,4-D and inoculated another 2,4-D-percolated soil column with the 2,4-D-degrading bacteria *C. necator* JMP134, *Rhodoferrax* sp P230 and *S. herbicidovorans* B488. We opted against simply spiking soil with either protein or massive amounts of known bacteria to approach the natural situation, where a bacterial community has time to establish tight interactions with its habitat including physicochemical interactions with the soil matrix. After 22 days, 2,4-D (2 mM in the influent) was completely degraded during the passage through the inoculated column, whereas 33 μ M 2,4-D remained in the eluate of the column with autochthonous bacteria, showing that 2,4-D-degrading communities had established in both columns. Protein extracts from both columns were separated by SDS-PAGE and some bands were detected before blue background (Figure 2). The band patterns were similar except for bands at 29 and 12 kDa that were only present in the sample that had been inoculated with lab cultures (Figure 2, lane 2). The separation of both samples by 2-DE failed, although phenol extraction is one of the most powerful purification steps in sample preparation. Additionally, other purification strategies, for example trichloroacetic acid (TCA) precipitation and acetone washes, were tried. Mostly much brown material enriched in the acidic half of the IPG-strip. An alternative to remove that material would be liquid IEF. Finally, each lane of the SDS-PAGE from soil samples was cut into 23 pieces and submitted to protein identification by nanoLC-ESI-MS/MS. 2,4-Dichlorophenoxyacetate dioxygenase and outer membrane protein (porin) were identified in both samples, whereas catechol 1,2-dioxygenase, an enzyme found in the metabolism of chlorinated catechol intermediates, and molecular chaperone

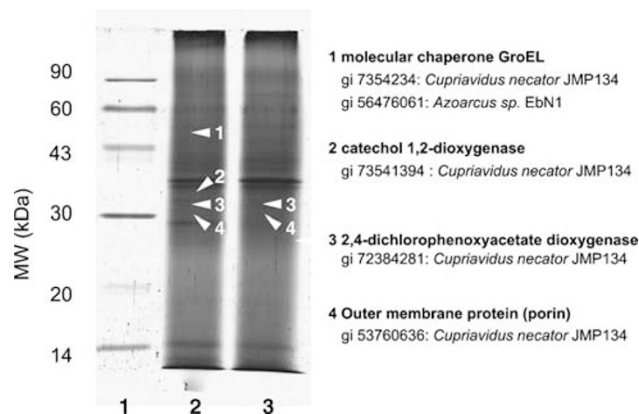


Figure 2 Proteins extracted from soil columns percolated with 2,4-dichlorophenoxy acetic acid (2,4-D). Lane 1: standard proteins; lane 2: soil from column augmented with *C. necator* JMP134, *Rhodoferrax* sp P230 and *S. herbicidovorans* B488; and lane 3: soil from column without inoculation. Phenol phases were precipitated and washed with organic solvents before sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and staining with Coomassie brilliant blue G250.

GroEL were only identified in the sample from the column inoculated with lab cultures. The presence of both metabolic enzymes known to be involved in the catabolism of 2,4-D corresponds well with the observed degradation of 2,4-D in both columns. The failure to identify catechol 1,2-dioxygenase in the column with autochthonous bacteria may be due to low gene expression or insufficient sensitivity of the MS analysis, as it corresponds with the lower degradation rate observed in that column. Unfortunately, the database search failed to identify the proteins at 29 and 12 kDa that were only present in the extract from the inoculated column. The appearance of distinguishable amounts of an outer membrane porin and the molecular chaperone GroEL indicates that percolation with 2,4-D increased the microbial biomass, presumably of species involved in degradation of 2,4-D. Porins are involved in the transport of small molecules and chaperones in maintaining the integrity of proteins. All four proteins, regardless of the origin of the column matched best (hits with highest Mowse score in MASCOT) with entries of *C. necator* JMP134, a well described 2,4-D-degrading species. For the column without specific inoculation, it appears that an organism identical or related with *C. necator* JMP134 that was present in compost soil was enriched upon growth on 2,4-D. For the inoculated column, it may indicate that this autochthonous strain and/or its inoculated relative outcompeted *Rhodoferrax* sp P230 and *S. herbicidovorans* B488 under the applied conditions.

Assuming that the entire 1.285 g of 2,4-D supplied to the column during the 22 days was metabolized and finally converted into biomass (assumed yield: 0.2 g dry mass/g of 2,4-D; protein content: 0.5 g protein/g dry of mass) and remained as such, 1 g soil

should contain at least 0.58 mg protein of 2,4-D-degraders, which is 2 to 10 times more than the typical protein content of natural soils (Vance *et al.*, 1987). It appears thus a relatively high amount of protein enabled the identification of at least some proteins. For calculating the efficiency of extraction, a protein determination after the phenol extraction and subsequent precipitation with organic solvent was not possible since the precipitates were only soluble in SDS-sample buffer or IEF sample buffer-containing high concentrations of detergents and reducing reagents. On the basis of a comparison of the intensity of the complete lane on SDS gels with a bovine serum albumin (BSA) in standard of the same gel, a rough estimation of the extraction yield gave that only 31 μ g protein were extracted from 5 g soil that, in theory, could have contained 2.5 mg protein. This corresponds to an extraction efficiency of about 1% and illustrates the difficulties of quantitative protein extraction from soil. There may be some potential for improvement by starting extractions with higher amounts of soil, increasing the ionic strength during the elution step or adding metal ion-complexing compounds to minimize the interactions between proteins and humic acids (Criquet *et al.*, 2002). However, elution of soil organic matter with NaOH is a very strong and effective procedure (80% of SOM; Stevenson, 1982) that probably also minimizes the interactions between proteins and humic compounds. Therefore, the potential effect of increasing the ionic strength or adding complexing compounds appears to be limited.

The protocol was also applied to a chlorobenzene-contaminated groundwater sample containing about 10^7 – 10^8 cells l^{-1} . The centrifugation pellet of 1 l groundwater contained, besides cells, also particles from the reactor material consisting of sand and lignite microparticles as could be seen from its dark color. Proteins were extracted with 0.1 M NaOH and purified by phenol extraction as described above. One-third of the final precipitate was separated by SDS-PAGE (Figure 3). About 20 Coomassie blue-stained bands were distinguished. The remaining two-thirds of the sample were separated by 2-DE in which about 100 spots were detected (Figure 4). The results show that the extraction protocol worked well with the groundwater samples that were low in humic organic matter. Nineteen bands from 1D gels and 50 most abundant spots from 2D gels were excised and used for protein identification. Table 2 summarizes 29 proteins that were identified in bands from the 1D gel and Table 3 summarizes 26 proteins identified in the 2D gel. The proteins belong to several functional groups and include enzymes as well as chaperones. In both gels together, for instance all enzymes of the chlorobenzene degradation pathway via 3-chlorocatechol to 3-oxoadipate (Supplementary Figure 1) were detected (van der Meer *et al.*, 1998). The best matches of all catabolic enzymes were achieved with *Proteobac-*

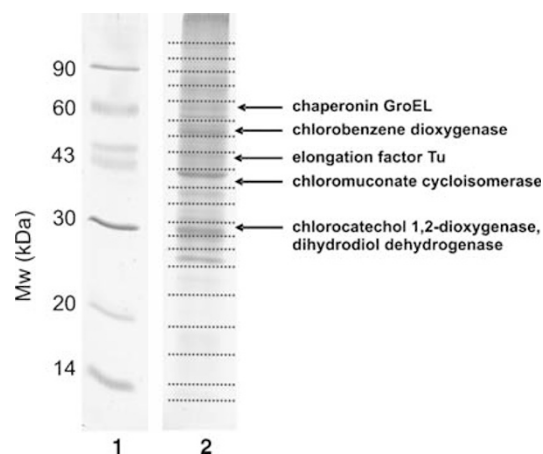


Figure 3 Proteins extracted from groundwater separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Lane 1: standard proteins; and lane 2: proteins extracted from ground water. Phenol phase was precipitated and washed with organic solvents before SDS-PAGE and staining with Coomassie brilliant blue G250. Arrows point towards gel slices cut from gel for protein identification.

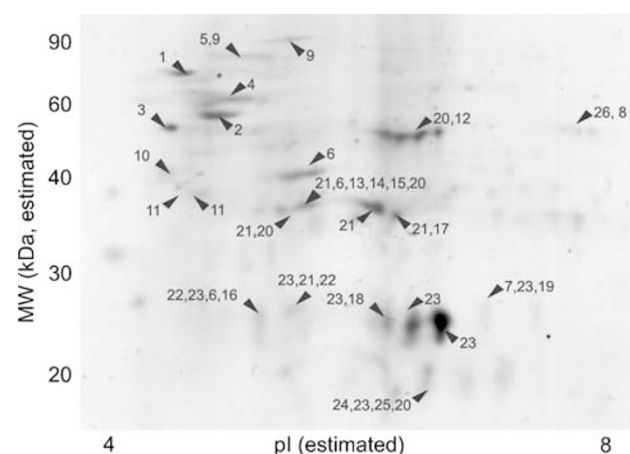


Figure 4 Proteins extracted from ground water separated by 2D-polyacrylamide gel electrophoresis (2D-PAGE). The phenol phase was precipitated and washed with organic solvents before sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and staining with Coomassie brilliant blue G250. Numbers mark identified proteins (refer to Table 2 for identifications). The isoelectric point and molecular weight scales were estimated by the values of the identified proteins.

teria sequences including the taxa *Acidovorax* and *Pseudomonas* that had been previously identified in that site (Alfreider *et al.*, 2002; Balcke *et al.*, 2004) of which members of the genus *Acidovorax* play a key role in the catabolism of chlorobenzene under oxygen limitation (Nestler *et al.*, 2007). This hypothesis has to be proved by 16S rRNA FISH. Some of the catabolic enzymes, for example, chlorobenzene dioxygenase, chloromuconate cycloisomerase, dihydrodiol dehydrogenase and chlorocatechol 1,2-dioxygenase appeared in more than one spot in the 2D gel. This could be due to some degree of proteolysis during extraction, posttranslational

Table 2 Identification of proteins extracted from groundwater and separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Number	Protein	Gene ID of best hit	Best matching species
C1	Chaperonin GroEL (HSP60 family)	22988122	<i>Burkholderia xenovorans</i>
C2	Chaperonin GroEL (HSP60 family)	22121848	<i>Comamonas testosteroni</i>
C3	Molecular chaperone (HSP70 family)	22985542	<i>Burkholderia xenovorans</i>
D1	Biphenyl dioxygenase (EC 1.14.-.-) terminal oxygenase component small chain	22981976	<i>Burkholderia xenovorans</i>
D2	Mutant biphenyl dioxygenase large subunit	6652685	<i>Pseudomonas pseudoalcaligenes</i>
D3	<i>cis</i> -2,3-Dihydroxy-2,3-dihydroisopropylbenzene dehydrogenase	2822269	<i>Pseudomonas putida</i>
D4	Chlorocatechol 1,2-dioxygenase II	5701903	<i>Pseudomonas chlororaphis</i>
D5	tetD chloromuconate cycloisomerase	7688269	<i>Pseudomonas chlororaphis</i>
D6	Maleylacetate reductase	7688272	<i>Pseudomonas chlororaphis</i>
M01	ATP synthase- β chain	15793517	<i>Neisseria meningitidis</i>
M02	ATP synthase F1, α -subunit	15642759	<i>Vibrio cholerae</i>
M03	Malate dehydrogenase	33593357	<i>Bordetella pertussis</i>
M04	Isocitrate dehydrogenase	15597820	<i>Pseudomonas aeruginosa</i>
M05	Citrate synthase	33597750	<i>Bordetella parapertussis</i>
M06	Succinyl-CoA synthetase, α -chain	33593523	<i>Bordetella pertussis</i>
M07	Succinyl-CoA synthetase, β -chain	33593524	<i>Bordetella pertussis</i>
M09	Dihydrolipoamide dehydrogenase-like protein	32455858	<i>Pseudomonas sp ADP</i>
M10	Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide acyltransferase	22983323	<i>Burkholderia xenovorans</i>
M11	Glutamine synthetase	17545977	<i>Ralstonia solanacearum</i>
M12	Fructose bisphosphate aldolase	17545292	<i>Ralstonia solanacearum</i>
N1	DNA-directed RNA polymerase, β -subunit	23467580	<i>Haemophilus somnus</i>
N2	DNA-directed RNA polymerase, β -subunit	15676060	<i>Neisseria meningitidis</i>
N3	Polyribonucleotide nucleotidyltransferase	30248196	<i>Nitrosomonas europaea</i>
O1	Alkyl hydroperoxide reductase subunit C	15595337	<i>Pseudomonas aeruginosa</i>
P1	Ribosomal protein L5	22982186	<i>Burkholderia xenovorans</i>
P2	Ribosomal protein S2	22985221	<i>Burkholderia xenovorans</i>
P3	50S ribosomal subunit L14	17547728	<i>Ralstonia solanacearum</i>
P4	Ribosomal protein S10	22982020	<i>Burkholderia xenovorans</i>
P5	GTPases – translation elongation factors	22982685	<i>Burkholderia xenovorans</i>

Protein number codes: C, chaperone; D, degradation; M, metabolism; N, nucleic acid synthesis; P, protein synthesis; O, other proteins. Proteins were identified using the software Spectrum Mill for database search against NCBI nr. Proteins with scores higher than 13 and with at least two matching peptides were considered as identified.

modification or isoenzymes originating from different genes as also described for the metaproteome of a natural acidophilic biofilm (Lo *et al.*, 2007). Proteolysis is not likely because of the extraction procedure, where the treatment with NaOH and the following treatment with phenol and organic solvents are mostly inactivating proteases. A clear distinction between posttranslational modification of isoenzymes of different genetic origin cannot be drawn yet, although significant differences in the mobility on the gel may indicate distinct isoenzymes. If this was true, this would indicate that we resolved a metaproteome comprising different species. Indeed two unique peptides with amino acid substitutions matching a homologue region confirmed that at least two chlorocatechol 1,2-dioxygenases of different genetic origins were present. One reason for the occurrence of several chlorocatechol 1,2-dioxygenases could be that specific enzymes dominate in a bacterial community that correspond best to the differential availability of substrates such as 3-chlorocatechol or oxygen in the aquifer (Kukor and Olsen, 1996; Krooneman *et al.*, 1998; Nestler *et al.*, 2007). The alignment of

partial sequences of MASCOT-identified elongation factors Tu (Figure 5) confirmed the presence of isoproteins of different genetic origins by unique peptides (grey background) with amino acid substitutions (black background) matching the homologue regions. Both results show that our protein extract represented a part of the metaproteome of the Bitterfeld aquifer.

The existing doubts about the metaproteomic nature of our results are a major challenge for the further development of the approach. The peptides of proteins identified by database search with MASCOT very often cover only small portions of proteins sequences (Supplementary Tables 1–3) and complicate the discrimination between isoenzymes and their taxonomic classification. However, the indirect, homology-based approach is necessary in cases where metagenomic information is not available. This is the typical situation in samples from soil and contaminated sites. Increasingly, examples are reported, where the availability of metagenomic data facilitated the metaproteomic analysis of acidic mine drainage by LC-MS/MS (Ram *et al.*, 2005). MS-based *de novo* sequencing would represent an

Table 3 Identification of proteins extracted from groundwater and separated by 2D-polyacrylamide gel electrophoresis (2D-PAGE)

Spot	Protein name	Gene ID of best hit	Species	Functional group	Genera of first five homologous hits
1	Heat shock protein 70	3095055	<i>Burkholderia pseudomallei</i>	Chaperones	<i>Burkholderia</i> , uncultured bacterium, <i>Ralstonia</i> , <i>Rubrivivax</i> , <i>Azoarcus</i>
2	Chaperonin GroEL (HSP60 family)	47574431	<i>Rubrivivax gelatinosus</i>	Chaperone	<i>Rubrivivax</i> , <i>Ralstonia</i> , <i>Rhodofera</i> , <i>Polaromonas</i> , <i>Burkholderia</i>
3	Trigger factor	78485520	<i>Thiomicrospira crunogena</i>	Chaperone	<i>Thiomicrospira</i>
4	Ribosomal protein S1	67668695	<i>Burkholderia cenocepacia</i>	Protein synthesis	<i>Burkholderia</i> , <i>Rhodofera</i> , <i>Rubrivivax</i> , <i>Polaromonas</i>
5	Translation elongation factor G	84711037	<i>Polaromonas naphthaleni-vorans</i>	Protein synthesis	<i>Polaromonas</i> , <i>Yersinia</i> , <i>Helicobacter</i> , <i>Rubrivivax</i> , <i>Thiomicrospira</i>
6	Elongation factor Tu	2654449	<i>Thiomonas cuprina</i>	Protein synthesis	<i>Thiomonas</i> , <i>Ralstonia</i> , <i>Rubrivivax</i> , <i>Bordetella</i> , <i>Methylococcus</i>
7	Ribosomal protein S2	74023265	<i>Rhodofera ferrireducens</i>	Protein synthesis	<i>Rhodofera</i> , <i>Polaromonas</i> , <i>Burkholderia</i> , <i>Nitrospira</i>
8	Transcription termination factor	47575036	<i>Rubrivivax gelatinosus</i>	Protein synthesis	<i>Rubrivivax</i> , <i>Francisella</i> , <i>Thermoanaerobacter</i> , <i>Rhodopirellula</i> , <i>Thermoanaerobacter</i>
9	Aconitate hydratase 2	4757495	<i>Rubrivivax gelatinosus</i>	Metabolism	<i>Rubrivivax</i> , <i>Xanthomonas</i> , <i>Magnetospirillum</i> , <i>Polaromonas</i> , <i>Pasteurella</i>
10	Enolase	68555733	<i>Ralstonia metallidurans</i>	Metabolism	<i>Ralstonia</i> , <i>Rubrivivax</i> , <i>Polaromonas</i> , <i>Dechloromonas</i> , <i>Geobacter</i>
11	Succinyl-CoA synthetase, β -subunit	67847535	<i>Polaromonas</i> sp	Metabolism	<i>Polaromonas</i> , <i>Rhodofera</i>
12	Citrate synthase	33593350	<i>Bordetella pertussis</i>	Metabolism	<i>Bordetella</i>
13	Isocitrate/isopropylmalate dehydro-genase	47575233	<i>Rubrivivax gelatinosus</i>	Metabolism	<i>Rubrivivax</i> , <i>Nitrosomonas</i>
14	Ketol-acid reductoisomerase	34496043	<i>Chromobacterium violaceum</i>	Metabolism	<i>Chromobacterium</i> , <i>Neisseria</i> , <i>Ralstonia</i> , <i>Microbulbifer</i> , <i>Burkholderia</i>
15	Aromatic-amino-acid aminotransferase	5565921	<i>Pseudomonas aeruginosa</i>	Metabolism	<i>Pseudomonas</i> , <i>Azotobacter</i> , <i>Ralstonia</i> , <i>Rhodofera</i>
16	Adenylate kinase, subfamily	74020663	<i>Rhodofera ferrireducens</i>	Metabolism	<i>Rhodofera</i>
17	Probable DNA-directed RNA polymerase (α -chain)	17430017	<i>Ralstonia solanacearum</i>	Metabolism	<i>Ralstonia</i> , <i>Bordetella</i> , <i>Rubrivivax</i> , <i>Burkholderia</i> , <i>Cupriavidus</i>
18	Glucose-1-phosphate thymidyltransferase	14717004	<i>Pseudomonas stutzeri</i>	Metabolism	<i>Pseudomonas</i> , <i>Photobacterium</i> , <i>Psychrobacter</i>
19	3-Oxoacyl-(acyl-carrier-protein) reductase	74021512	<i>Rhodofera ferrireducens</i>	Metabolism	<i>Rhodofera</i> , <i>Ralstonia</i>
20	Chlorobenzene dioxygenase (fragment)	45385177	<i>Acidovorax avenae</i>	Degradation	<i>Acidovorax</i> , <i>Burkholderia</i> , <i>Pseudomonas</i> , <i>Comamonas</i>
21	Chloromuconate cycloisomerase	26185980	<i>Achromobacter xylosoxidans</i>	Degradation	<i>Achromobacter</i> , <i>Delftia</i> , <i>Cupriavidus</i>
22	Dihydrodiol dehydrogenase	1256708	<i>Pseudomonas fluorescens</i>	Degradation	<i>Pseudomonas</i>
23	Chlorocatechol 1,2-dioxygenase	19386504	<i>Delftia acidovorans</i>	Degradation	<i>Delftia</i> , <i>Achromobacter</i> , <i>Pseudomonas</i>
24	Dienelactone hydrolase	151580	<i>Pseudomonas</i> sp	Degradation	<i>Pseudomonas</i> , <i>Achromobacter</i> , <i>Delftia</i> , <i>Cupriavidus</i>
25	Alkyl hydroperoxide reductase	71909129	<i>Dechloromonas aromatica</i>	Other	<i>Dechloromonas</i>
26	Putative multicopper oxidase	47574502	<i>Rubrivivax gelatinosus</i>	Other	<i>Rubrivivax</i>

Database search was conducted using the tandem mass spectrometry (MS/MS) ion search (MASCOT, <http://www.matrixscience.com>) against all bacterial entries of NCBI Inr (GenBank; <http://www.ncbi.nlm.nih.gov/index.html>) with subsequent parameters: trypsin digestion, up to one missed cleavage site, fixed modifications: carbamidomethyl (C), variable modifications: oxidation (M), peptide tol.: ± 1.2 Da, MS/MS tol.: ± 0.6 Da, peptide charge: +1, +2 and +3.

improvement of metaproteome analyses and was demonstrated as a possibility to extract protein sequence data from proteins of different sources (Wilmes and Bond, 2004; Schulze et al., 2005; Lacerda et al., 2007). Recently, there have been

substantial developments of instrumentation (CID-MALDI-ToF/ToF, (Samyn et al., 2006) and sample preparation methodology (for example by derivatization, Chen et al., 2004). Sequences obtained in this way can be compared with known genomes and

		1					50
2654449	<i>Thiomonas cuprina</i>	MAKSKFERTK	PHVNVGTIGH	VDHGKTTTLTA	AITTVLSSKF	GG	EAKAYDQ
47574116	<i>Rubrivivax gelatinosus</i>	MAKSKFERTK	PHVNVGTIGH	VDHGKTTTLTA	AITTVLSSKF	GG	EAKAYDQ
73542880	<i>Ralstonia eutropha</i>	MAKEFERTK	PHVNVGTIGH	VDHGKTTTLTA	AIATVLAAKF	GG	AAKKYDE
33594477	<i>Bordetella pertussis</i>	MAKGFERTK	PHVNVGTIGH	VDHGKTTTLTA	AITTVLSNKF	GG	EARGYDQ
53804631	<i>Methylococcus capsulatus</i>	MSKEKFTRTK	PHVNVGTIGH	VDHGKTTTLTA	ALTRCMAAKF	GG	EFKAYDQ
22124392	<i>Yersinia pestis</i>	MSKEKFERTK	PHVNVGTIGH	VDHGKTTTLTA	AITTVLAKTY	GG	SARAFDQ
69938869	<i>Paracoccus denitrificans</i>	MAKAKFERTK	PHVNVGTIGH	VDHGKTTTLTA	AIT...KY	FG	EFKAYDQ
22087357	<i>Rickettsia bellii</i>	MAKAKFERTK	PHVNVGTIGH	VDHNETSLTA	AITMVLGKDN	KDI	KVKKYDE
		51					100
2654449	<i>Thiomonas cuprina</i>	IDAAP EE KAR	GITINTAHVE	YETANRHYAH	VDCPGHADV	KNMITGAAQM	
47574116	<i>Rubrivivax gelatinosus</i>	IDAAP EE KAR	GITINTAHVE	YETANRHYAH	VDCPGHADV	KNMITGAAQM	
73542880	<i>Ralstonia eutropha</i>	IDAAP EE KAR	GITINTAHVE	YETANRHYAH	VDCPGHADV	KNMITGAAQM	
33594477	<i>Bordetella pertussis</i>	IDAAP EE KAR	GITINTSHVE	YETETRHYAH	VDCPGHADV	KNMITGAAQM	
53804631	<i>Methylococcus capsulatus</i>	IDAAP EE KAR	GITIATAHVE	YESAARHYAH	VDCPGHADV	KNMITGAAQM	
22124392	<i>Yersinia pestis</i>	IDNAP EE KAR	GITINTSHVE	YDTPARHYAH	VDCPGHADV	KNMITGAAQM	
69938869	<i>Paracoccus denitrificans</i>	IDGAP EE KAR	GITISTAHVE	YESENRYAH	VDCPGHADV	KNMITGAAQM	
22087357	<i>Rickettsia bellii</i>	IDAAP EE KER	GITISTAHVE	YETMSRHYAH	VDCPGHADV	KNMITGAAQM	

Figure 5 Partial alignment of amino acid sequences of elongation factors Tu of selected species. From MASCOT result list, the first species and the corresponding sequence of all hits with at least non-redundant peptides were selected. Marked amino acids were covered by the peptides measured with electrospray ionization source tandem mass spectrometry (ESI-MS/MS). Bold letters: peptides that match to more than 50% of the species. Underlined letters: peptides that match to less than 50% of the species. Grey background: peptides that match to the same region of the proteins, but differ in at least one amino acid marked with black background. Numbers above the alignment correspond with amino acid residues of selected proteins.

this will reveal similarities to known functional proteins thus pointing at the function fulfilled by the protein. This approach will also yield data information about proteins that have not been observed before.

Conclusion

Alkaline extraction of protein from environmental samples followed by purification with liquid phenol extraction resulted in samples that could be separated on 1D or 2D gels. This is the first report on the MS identification of intracellular proteins from microbial communities inhabiting soil and groundwater.

The enzymes identified in a groundwater sample mirrored the observed metabolism of chlorobenzene and thus represent a part of the functional metaproteome of this environment. Isoforms of catechol 1,2-dioxygenase and of elongation factor Tu indicate that at least two species are involved in the biodegradation of chlorobenzene.

However, there is an urgent need for improved extraction and fractionation schemes for samples of high complexity, and in particular, high humic matter contents such as soil samples (Wilmes and Bond, 2006). On the basis of the proteins extracted from microcosms, the autochthonous community established on 2,4-D appears to be similar to the community established after inoculation with laboratory strains and shows the potential of autochthonous bacteria to compete with laboratory strains.

Finally, increasing availability of genome data from environmental samples and improved *de novo*

sequencing strategies will facilitate the analysis of environmental metaproteomes and their functional interpretation.

References

- Alfreider A, Vogt C, Babel W. (2002). Microbial diversity in an *in situ* reactor system treating monochlorobenzene contaminated groundwater as revealed by 16S ribosomal DNA analysis. *Sys Appl Microbiol* **25**: 232–240.
- Aoyama M. (2006). Properties of neutral phosphate buffer extractable organic matter in soils revealed using size exclusion chromatography and fractionation with polyvinylpyrrolidone. *Soil Sci Plant Nutr* **52**: 378–386.
- Balcke GU, Turunen LP, Geyer R, Wenderoth DF, Schlosser D. (2004). Chlorobenzene biodegradation under consecutive aerobic-anaerobic conditions. *FEMS Microbiol Ecol* **49**: 109–120.
- Benndorf D, Thiersch M, Loffhagen N, Kunath C, Harms H. (2006). *Pseudomonas putida* kt2440 responds specifically to chlorophenoxy herbicides and their initial metabolites. *Proteomics* **6**: 3319–3329.
- Boschker HTS, Nold SC, Wellsbury P, Bos D, de Graaf W, Pel R *et al.* (1998). Direct linking of microbial populations to specific biogeochemical processes by c-13-labelling of biomarkers. *Nature* **392**: 801–805.
- Chen P, Nie S, Mi W, Wang XC, Liang SP. (2004). *De novo* sequencing of tryptic peptides sulfonated by 4-sulfophenyl isothiocyanate for unambiguous protein identification using post-source decay matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Comm Mass Spectrom* **18**: 191–198.
- Criquet S, Farnet AM, Ferre E. (2002). Protein measurement in forest litter. *Biol Fertil Soils* **35**: 307–313.
- Eilers H, Pernthaler J, Glockner FO, Amann R. (2000). Culturability and *in situ* abundance of pelagic bacteria

- from the north sea. *Appl Environ Microbiol* **66**: 3044–3051.
- Fields MW, Bagwell CE, Carroll SL, Yan T, Liu X, Watson DB et al. (2006). Phylogenetic and functional biomarkers as indicators of bacterial community responses to mixed-waste contamination. *Environ Sci Tech* **40**: 2601–2607.
- Fredrickson JK, Mckinley JP, Nierzwickibauer SA, White DC, Ringelberg DB, Rawson SA et al. (1995). Microbial community structure and biogeochemistry of miocene subsurface sediments – implications for long-term microbial survival. *Mol Ecol* **4**: 619–626.
- Görg A, Postel W, Gunther S. (1988). The current state of two-dimensional electrophoresis with immobilized ph gradients. *Electrophoresis* **9**: 531–546.
- Guerlava P, Izac V, Tholozan JL. (1998). Comparison of different methods of cell lysis and protein measurements in *Clostridium perfringens*: application to the cell volume determination. *Curr Microbiol* **36**: 131–135.
- Kan J, Hanson TE, Ginter JM, Wang K, Chen F. (2005). Metaproteomic analysis of chesapeake bay microbial communities. *Saline Systems* **1**: 7.
- Kanagawa T. (2003). Bias and artifacts in multitemplate polymerase chain reactions (pcr). *J Biosci Bioeng* **96**: 317–323.
- Kirby KS. (1956). New method for the isolation of ribonucleic acids from mammalian tissues. *Biochem J* **64**: 405–408.
- Kleinsteuber S, Riis V, Fetzer I, Harms H, Muller S. (2006). Population dynamics within a microbial consortium during growth on diesel fuel in saline environments. *Appl Environ Microbiol* **72**: 3531–3542.
- Klose J. (1975). Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues – novel approach to testing for induced point mutations in mammals. *Humangenetik* **26**: 231–243.
- Krooneman J, Moore ERB, van Velzen JCL, Prins RA, Forney LJ, Gottschal JC. (1998). Competition for oxygen and 3-chlorobenzoate between two aerobic bacteria using different degradation pathways. *FEMS Microbiol Ecol* **26**: 171–179.
- Kukor JJ, Olsen RH. (1996). Catechol 2,3-dioxygenases functional in oxygen-limited (hypoxic) environments. *Appl Environ Microbiol* **62**: 1728–1740.
- Lacerda CMR, Choe LH, Reardon KF. (2007). Metaproteomic analysis of a bacterial community response to cadmium exposure. *J Proteome Res* **6**: 1145–1152.
- Laemmli UK. (1970). Cleavage of structural proteins during assembly of head of bacteriophage-T4. *Nature* **227**: 680–685.
- Landi L, Renella G, Moreno JL, Falchini L, Nannipieri P. (2000). Influence of cadmium on the metabolic quotient, l-: D-glutamic acid respiration ratio and enzyme activity: microbial biomass ratio under laboratory conditions. *Biol Fertil Soils* **32**: 8–16.
- Lo I, Deneff VJ, VerBerkmoes NC, Shah MB, Goltzman D, DiBartolo G et al. (2007). Strain-resolved community proteomics reveals recombining genomes of acidophilic bacteria. *Nature* **446**: 537–541.
- Manefield M, Griffiths R, Bailey M, Whiteley A. (2006). *Stable Isotope Probing: A Critique of its Role in Linking Phylogeny and Function*, 1st edn. Springer: Berlin.
- Moreno JL, Garcia C, Landi L, Falchini L, Pietramellara G, Nannipieri P. (2001). The ecological dose value (ed50) for assessing cd toxicity on atp content and dehydrogenase and urease activities of soil. *Soil Biol Biochem* **33**: 483–489.
- Muller RH, Babel W. (1986). Glucose as an energy donor in acetate growing acinetobacter-calcoaceticus. *Arch Microbiol* **144**: 62–66.
- Nannipieri P. (2006). *Role of Stabilised Enzymes in Microbial Ecology and Enzyme Extraction from Soil with Potential Applications in Soil Proteomics*, 1st edn. Springer: Berlin.
- Nestler H, Kiesel B, Kaschabek SR, Mau M, Schlomann M, Balcke GU. (2007). Biodegradation of chlorobenzene under hypoxic and mixed hypoxic-denitrifying conditions. *Biodegrad* (published online doi:10.1007/s10532-007-9104-z).
- O'Farrell PH. (1975). High-resolution 2-dimensional electrophoresis of proteins. *J Biol Chem* **250**: 4007–4021.
- Ogunseitian OA. (2006). *Soil Proteomics: Extraction and Analysis of Proteins from Soils*, 1st edn. Springer: Berlin.
- Oh KH, Tuovinen OH. (1990). Degradation of 2,4-dichlorophenoxyacetic acid by mixed cultures of bacteria. *J Ind Microbiol* **6**: 275–278.
- Radajewski S, Ineson P, Parekh NR, Murrell JC. (2000). Stable-isotope probing as a tool in microbial ecology. *Nature* **403**: 646–649.
- Ram RJ, VerBerkmoes NC, Thelen MP, Tyson GW, Baker BJ, Blake RC et al. (2005). Community proteomics of a natural microbial biofilm. *Science* **308**: 1915–1920.
- Rhee SK, Liu XD, Wu LY, Chong SC, Wan XF, Zhou JZ. (2004). Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. *Appl Environ Microbiol* **70**: 4303–4317.
- Rigou P, Rezaei H, Grosclaude J, Staunton S, Quiquampoix H. (2006). Fate of prions in soil: adsorption and extraction by electroelution of recombinant ovine prion protein from montmorillonite and natural soils. *Environ Sci Technol* **40**: 1497–1503.
- Samyn B, Sergeant K, Memmi S, Debyser G, Devreese B, Van Beeumen J. (2006). Maldi-tof/tof de novo sequence analysis of 2-d page-separated proteins from *Halorhodospira halophila*, a bacterium with unsequenced genome. *Electrophoresis* **27**: 2702–2711.
- Santos PM, Benndorf D, Sa-Correia I. (2004). Insights into *Pseudomonas putida* KT2440 response to phenol-induced stress by quantitative proteomics. *Proteomics* **4**: 2640–2652.
- Schulze WX, Gleixner G, Kaiser K, Guggenberger G, Mann M, Schulze ED. (2005). A proteomic fingerprint of dissolved organic carbon and of soil particles. *Oecologia* **142**: 335–343.
- Simonart P, Batistic L, Mayaudon J. (1967). Isolation of protein from humic acid extracted from soil. *Plant Soil* **27**: 153–161.
- Singleton I, Merrington G, Colvan S, Delahunty JS. (2003). The potential of soil protein-based methods to indicate metal contamination. *Appl Soil Ecol* **23**: 25–32.
- Stevenson FJ. (1982). *Humus Chemistry – Genesis, Composition, Reactions*. Wiley: New York.
- van der Meer JR, Werlen C, Nishino SF, Spain JC. (1998). Evolution of a pathway for chlorobenzene metabolism leads to natural attenuation in contaminated groundwater. *Appl Environ Microbiol* **64**: 4185–4193.

- Vance ED, Brookes PC, Jenkinson DS. (1987). An extraction method for measuring soil microbial biomass-C. *Soil Biol Biochem* **19**: 703–707.
- Vogt C, Alfreider A, Lorbeer H, Hoffmann D, Wuensche L, Babel W. (2004). Bioremediation of chlorobenzene-contaminated ground water in an *in situ* reactor mediated by hydrogen peroxide. *J Contam Hydrol* **68**: 121–141.
- Wagner M, Nielsen PH, Loy A, Nielsen JL, Daims H. (2006). Linking microbial community structure with function: fluorescence *in situ* hybridization-microautoradiography and isotope arrays. *Curr Opin Biotechnol* **17**: 83–91.
- Wang W, Scali M, Vignani R, Spadafora A, Sensi E, Mazzuca S *et al*. (2003). Protein extraction for two-dimensional electrophoresis from olive leaf, a plant tissue containing high levels of interfering compounds. *Electrophoresis* **24**: 2369–2375.
- Wilmes P, Bond PL. (2004). The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. *Environ Microbiol* **6**: 911–920.
- Wilmes P, Bond PL. (2006). Metaproteomics: studying functional gene expression in microbial ecosystems. *Trends Microbiol* **14**: 92–97.
- Zhou JH. (2003). Microarrays for bacterial detection and microbial community analysis. *Curr Opin Microbiol* **6**: 288–294.

Supplementary Information accompanies the paper on The ISME Journal website (<http://www.nature.com/ismej>)