Direct Cellular Lysis/Protein Extraction Protocol for Soil Metaproteomics

ARTICLE *in* JOURNAL OF PROTEOME RESEARCH · OCTOBER 2010

Impact Factor: 4.25 · DOI: 10.1021/pr100787q · Source: PubMed

research articles Journal of proteome research

Direct Cellular Lysis/Protein Extraction Protocol for Soil Metaproteomics

Karuna Chourey,[†] Janet Jansson,[‡] Nathan VerBerkmoes,[†] Manesh Shah,[†] Krystle L. Chavarria,[‡] Lauren M. Tom,[‡] Eoin L. Brodie,[‡] and Robert L. Hettich*[,†]

Oak Ridge National Laboratory, Tennessee, United States, and Lawrence Berkeley National Laboratory, California, United States

Received July 28, 2010

We present a novel direct protocol for deep proteome characterization of microorganisms in soil. The method employs thermally assisted detergent-based cellular lysis (SDS) of soil samples, followed by TCA precipitation for proteome extraction/cleanup prior to liquid chromatography−mass spectrometric characterization. This approach was developed and optimized using different soils inoculated with genome-sequenced bacteria (Gram-negative *Pseudomonas putida* or Gram-positive *Arthrobacter chlorophenolicus)*. Direct soil protein extraction was compared to protein extraction from cells isolated from the soil matrix prior to lysis (indirect method). Each approach resulted in identification of greater than 500 unique proteins, with a wide range in molecular mass and functional categories. To our knowledge, this SDS-TCA approach enables the deepest proteome characterizations of microbes in soil to date, without significant biases in protein size, localization, or functional category compared to pure cultures. This protocol should provide a powerful tool for ecological studies of soil microbial communities.

Keywords: soil metaproteomics • soil microbiome • *Arthrobacter* • *Pseudomonas* • microbial proteomics • mass spectrometry • proteomics

Introduction

The advent of high performance mass spectrometry-based technology platforms for deep proteomic characterization of microbial proteins in environmental samples has ushered in a new era of scientific discovery referred to as community proteomics, or metaproteomics.[1,2] This powerful approach integrates genomics with proteomics (termed proteogenomics) for exploration of environmental microbial communities, which are comprised mostly of uncultivated and uncharacterized populations[3−6] To date, this technology has been demonstrated primarily for communities with limited-moderate microbial diversity, such as acid mine drainage communities.[1,2] However, there is a driving interest in extension of metaproteomics to more complex environments with a high microbial diversity, such as soils and sediments. This interest is fueled by our need to better understand the functional roles of soil microorganisms in key processes such as pollutant bioremediation and biogeochemical cycling.[7−9] To date, metaproteomics technologies have had limited success for characterizing microbes in soil, due to the heterogeneity and hydrophobic nature of the soil matrix in addition to the high microbial diversity of soils.[10] Proteins released in soil can potentially be absorbed to soil particles and soil organic matter, thus hampering their recovery. Furthermore, humic materials that are normally found in soil have been shown to degrade proteins,[11] and may cause peptide signal suppression in LC−MS measurements. To achieve success in application of metaproteomics to soil microbes, it is imperative to have a method that can effectively lyse the microbial cells directly in the soil matrix and also quantitatively extract the proteins from the sample without significant losses or contaminants that interfere with downstream processing and analysis of the sample.[12]

Recent attempts to extract proteins from soil have shown only modest success and have only achieved low protein recoveries.[13] Only a few proteins have been successfully extracted from soil using different approaches, as evidenced by visualization of a few bands on SDS-PAGE gels.[14] For these studies, some approaches focus on direct soil protein extraction,[14,15] while others use an indirect approach that first extracts microbial cells from soils prior to protein extraction.[16,17] The direct approaches are particularly limited by low protein yields due to the electrostatic or hydrophobic interactions with soil particles resulting in protein sorption or due to coextraction of contaminants that interfere with downstream processing steps.[13,18,19]

One approach developed for direct soil protein extraction involved blocking hydrophobic sites in the soils using bovine serum albumin (BSA) and Triton-X prior to microbial lysis.[20] This method was developed for extraction of a single known enzyme from soil and is not applicable for extraction of whole cell proteomes. Another method for soil proteome extraction

* To whom correspondence should be addressed. Dr. Robert L. Hettich, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6131. E-mail: hettichrl@ornl.gov. Phone: 865-574-4968.
[†] Oak Ridge National Laboratory.
[‡] Lawrence Berkeley National Laboratory.

10.1021/pr100787q

Published on Web 10/18/2010

used sodium hydroxide and phenol to lyse the soil microorganisms but yielded only about 12 protein identifications.[14] Other protein extraction protocols (guanidine lysis, sonication, or bead beating) that have been successfully used in different environmental samples, including AMD biofilms[1] and mouse cecal material,[21] have, in our hands, proven to be ineffective for direct extraction of proteins from soil (unpublished results). Recently, a new method was described that involves a sequential extraction of proteins from soil using citrate, phenol, and detergent.[15] This method was successful for detection of glomalin-related soil proteins using 1-D and 2-D SDS-PAGE, but this approach was not demonstrated for deep proteome identification.

Our aim was to develop a robust experimental approach for the efficient extraction of proteins from soil microorganisms to obtain a high protein yield and a more extensive, unbiased proteome characterization than afforded by other methods available to date. In particular, we sought to develop a procedure that is suitable for *direct* extraction of proteins from soil microbes without prior separation from their soil matrix (direct extraction), which could be compared and contrasted with approaches that involve extraction of proteins from cells that were first physically enriched from soil (indirect extraction).

Experimental Section

Our experimental approach was designed to enhance direct proteome extraction from microbes in soil. Diverse soils from two different geographical locations (Hanford, WA and Hopland, CA) were used as test-beds to establish short-term microcosms containing different well-characterized microbial inoculants (one Gram negative and one Gram positive strain). A concurrent cellular lysis/proteome solubilization approach was developed and optimized for these soil microcosms, so that the resulting mass spectrometric data could be compared with results obtained using a commonly employed guanidine assisted cell lysis method used for isolates. To evaluate the efficacy of this direct approach, it was necessary to compare and contrast these results from proteome mass spectrometric data obtained from an indirect method in which the microbes are separated from the soil matrix prior to cellular lysis/proteome extraction, a general method that we previously have demonstrated for cecal samples.[21]

Bacterial Species Information, Incubation Conditions, and Microcosm Establishment

Pseudomonas putida strain F1 (ATCC 700007) was grown overnight aerobically at 30 °C with vigorous shaking (200 rpm) in 50 mL Luria−Bertani (LB) medium. The cells were pelleted by centrifugation at 2850× *g* for 10 min, and the supernatant was discarded. The cells were suspended in 10 mL of fresh LB medium and half of the culture (approximately 2×10^9 cells) was transferred to 5 g soil (Hanford, WA, sandy soil) in a 50 mL Falcon tube (VWR). The other half was transferred to a fresh Falcon tube. Both the inoculated soil and the liquid medium (LM) were incubated for 5 h at 30 °C, with no shaking. After 5 h, the microcosm and the liquid culture were flash frozen in liquid nitrogen and kept at −80 °C until further analysis. Since the purpose of this initial experiment was to evaluate the overall ability of the detergent method to extract proteomes directly from microbes in soil, no specific cell counting was done after microcosm incubation, since it was not deemed critical to precisely quantify the modest degree in cell population change expected after 5 h incubation in stationery mode.

Arthrobacter chlorophenolicus A6G[22] was previously genetically modified with the *gfp* gene encoding the green fluorescent protein (GFP) under control of a constitutive promoter.[22] The A6G strain was cultivated in 10% LB + 50 μg/mL kanamycin at 28 °C. To prepare the A6G cells for soil inoculation, 2 L of culture were grown overnight to mid log phase and collected by centrifugation at 10 000× *g* for 10 min. The cells were resuspended in 40 mL phosphate buffered saline (PBS) and preincubated overnight at 4 °C to starve and preadapt the cells to soil conditions and cold temperature, as previously described.[22]

Soil was collected at the University of California Hopland Research and Extension Center, from permanent managed mixed grassland (Hopland, CA). This soil is a medium-texture loam derived from hard sandstone and shale, classified as an ultic haploxeralf.[23] Soil was collected as previously described,[24] 4-chlorophenol was added as a specific growth substrate to 20 g portions of soil in 50 mL Falcon tubes to a concentration of 150 μg/g and the resulting mixture was preincubated at 4 °C overnight. The preadapted A6G culture was diluted and 2 mL portions were added to the soil to result in final cell concentrations of approximately 10^9, and 10^8 cells/g soil and mixed with a sterile metal spatula for 1 min. An additional 1 mL of cold sterile PBS was then added to the soil samples and they were mixed for an additional minute to disperse the cells. Uninoculated control samples were prepared in the same way by adding 3 mL cold, sterile PBS buffer and mixing. Samples were incubated for 24 h at 4 °C and either frozen in liquid nitrogen and stored at −80 °C prior to direct extraction using the SDS-TCA approach, or processed immediately by differential centrifugation (DC) using the protocols described below. Cell counts (CFU/g) at the time of sampling verified the cell number per g soil to be 1.43×10^9 and 1×10^8 for the two inoculation doses, respectively.

Direct Soil Protein Extraction Method (SDS-TCA)

Frozen soil was thawed partially and 5 g soil measured out in a 50 mL Pyrex glass bottle to which 10 mL of Alkaline-SDS buffer was added (5% SDS, 50 mM Tris-HCl, pH 8.5; 0.15 M NaCl; 0.1 mM EDTA; 1 mM $MgCl_2$; 50 mM Dithiothretiol (DTT)). Soil was dispersed in the buffer by vigorous vortexing for 2−3 min and then immersed in a boiling water bath for 10 min for effective cellular lysis and inactivation of the proteases that are released due to lysis. Bottle lids were capped loosely to provide adequate venting. The resulting slurry was cooled for 5 min and then transferred to a 50 mL Falcon tube followed by vigorous vortex mixing for 3 min. The solution was centrifuged at 2095× *g* for 10 min and the clear supernatant containing the whole cell lysate was transferred to fresh tubes. To each tube, chilled 100% tricholoroacetic acid (TCA) was added to achieve ∼25% TCA as final concentration. Tubes were inverted gently and kept at −10° overnight to precipitate proteins present in the solution. Samples were centrifuged at 20 800× *g* for 20 min to obtain a concentrated protein pellet and the supernatant (containing detergent and other contaminants) was discarded. Protein pellets were washed with 1 mL chilled (∼−10°) acetone followed by centrifugation at 20 800× *g* for 10 min. This step helped wash off excess TCA and lingering SDS. Protein pellets were gently disaggregated in acetone by vortexing during the first acetone addition. After centrifugation, acetone was gently removed and discarded, and a fresh 1 mL

of acetone was added. This acetone washing step was repeated three times. Protein pellets were air-dried until all acetone evaporated, solubilized in 2 mL of guanidine buffer (6 M Guanidine HCl, 10 mM DTT in Tris $CaCl_2$ buffer (50 mM Tris, 10 mM $CaCl_2$, pH 7.6) and incubated at 60° for 1 h with intermittent vortexing to disrupt and dissolve the protein pellet. Protein solutions were diluted 6-fold using Tris $CaCl_2$ buffer, pH 7.6 and subjected to proteolytic digestion by addition of 40 μg trypsin[25] and overnight incubation with gentle mixing at 37°. Following proteolytic digestion, samples were desalted using C_{18} Sep-Pak solid phase extraction cartridges (Waters, Milford, MA), and concentrated to ~5 mL. The sample was solvent exchanged with 0.1% formic acid in HPLC grade water and concentrated to 1/10th starting volume using a Savant SpeedVac (ThermoFisher Scientific, Waltham, MA). Samples were filtered using Durapore PVDF filters, 0.45 μm (Millipore) and stored at −80 °C prior to two-dimensional (2-D) LC−MS/MS analysis.

Direct Soil Protein Extraction using Guanidine HCl. Five grams of soil, thawed to 4°, were transferred to a 50 mL Falcon tube and suspended in 10 mL Guanidine solution (6 M guanidine HCl;10 mM DTT dissolved in Tris buffer (50 mM Tris/10 mM $CaCl_2$ pH 7.6). Slurries were incubated overnight at 37 °C. Following cell lysis and denaturation, samples were centrifuged at 2095× *g* for 10 min (Allegra X-15R centrifuge; Beckman Coulter) to pellet soil debris. Supernatants were transferred to a fresh tube and diluted 6-fold with the Tris buffer. This was followed by addition of trypsin (20 μg) for proteolytic digestion for 6 h at 37 °C with gentle rocking on a nutating mixer (VWR International), followed by a second trypsin addition (20 μg) and overnight incubation at 37 °C. Following the proteolytic digestion, another 20 mM DTT reduction step was performed for 1 h at 37 °C with gentle rocking. Samples were then centrifuged 2095× *g* for 10 min (Allegra X-15R centrifuge; Beckman Coulter) to pellet cellular debris. Samples were then desalted with a Sep-Pak C_{18} Lite cartridge, solvent was exchanged into 100% H_2O, 0.1% formic acid and followed by filtration using an Ultrafree-MC centrifugal filter device (Millipore, MA). All prepared samples were stored at −80 °C until LC−MS/MS analysis.

Differential Centrifugation (Indirect Extraction)

The differential centrifugation protocol was modified from a previous method used for extraction of DNA from soil.[26] Soil samples (20 g) in 50 mL Falcon tubes were mixed with 20 mL cold PBS on ice using a VWR hand-held homogenizer at full speed (speed 6) for 2 × 30 s with a 30 s break in between homogenization steps to prevent the soil from overheating. Samples were centrifuged at a low speed (2500 g) for 5 min to gently pellet soil particles and the supernatant was collected in new 50 mL Falcon tubes. Soil pellets were re-extracted with an additional 20 mL cold PBS and the combined supernatants were centrifuged at high speed (10 000× *g* for 10 min, 4 °C). Supernatants were discarded and the pellet was resuspended in 250 μL cold PBS, briefly vortexed and transferred to a 2 mL microcentrifuge tube. An additional 250 μL PBS was added to the first tube to collect any remaining cells and combined with the first supernatant. Cell pellets were collected by centrifugation at 10 000× *g* for 10 min and immediately frozen in liquid nitrogen. Samples were stored frozen at −20 °C prior to cell lysis using the SDS-TCA or guanidine lysis methods described above.

2D-LC−MS/MS Measurements. Digested proteins were analyzed in duplicate using a 2-Dimensional (2D) nano-LC−MS/MS system with a split-phase column (RP-SCX) in a 24-h, 12-step setup[27,28] on a linear trapping quadrupole (LTQ XL) mass spectrometer (ThermoFisher Scientific, San Jose, CA) interfaced with an Ultimate HPLC system (LC Packings, a division of Dionex, San Francisco, CA). The LTQ was operated in a data-dependent mode with one full scan and 5 dependent MS/MS scans. The m/z isolation width was set to 3.00, the dynamic exclusion repeat count was set to 1, and the exclusion list size was set at 100 with exclusion duration of 60. Further details on the instrumentation setup have been described previously.[29] The peptide sample was loaded onto an SCX column followed by an offline column wash with a gradient from 100% aqueous solvent [95% H_2O/5% acetonitrile (ACN)/ 0.1% formic acid] to 100% organic solvent (30% H_2O/70% ACN/0.1% formic acid). This step was included to wash away any residual SDS or interfering substances. Omission of this step did occasionally cause problems, such as clogging of the front column and signal suppression. After the offline sample column wash, it was connected to a reverse-phase (C18) packed front column (New Objective) and subjected to separation analysis described in detail elsewhere.[29,30]

Proteome Bioinformatics. The *Arthrobacter chlorophenolicus* A6 predicted proteome database was downloaded from JGI's Integrated Microbial Genomes server (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi, September 2009 draft version). The protein sequence for Green Fluorescent Protein was also included in the proteome database since the strain used in this study carried a GFP tag. A total of 4641 proteins were predicted in this database. The predicted protein database for *Pseudomonas putida* F1 was also downloaded from IMG (2008 version) and contained a total of 5252 predicted proteins. An artificial metaproteome database consisting of all sequenced microbes (1606 microbes as of November 2009), including the Minnesota Farm soil metagenome,[31] was also generated using data from IMG and IMG/M to enable database searching for indigenous soil microbes. Common contaminants such as trypsin, keratin, etc. were also included in the databases. For all database searches, the MS/MS spectra were searched via SEQUEST,[32] using parameters described elsewhere.[29] The output data files were then filtered and sorted using the DTASelect algorithm[33] using the following parameters: fully tryptic peptides only with ΔCN of at least 0.08 and cross-correlation scores (Xcorr) of at least 1.8 (+1), 2.5 (+2), and 3.5 (+3). Post-translational modifications and other fixed modifications were not included in the search parameters. Slight variations in protein counts between duplicate runs was typical, but within acceptable limits. For discussion of specific proteins and functional categories, a representative single data set for each sample was used due to space constraints.

Results and Discussion

Here we describe a novel protocol for soil metaproteomics that exploits the lysing/solubilizing power of Sodium Dodecyl Sulfate (SDS) under elevated temperature conditions, followed by trichloroacetic acid (TCA) precipitation of proteins. Following removal of the detergent, the protein pellet is subjected to proteolytic digestion and mass spectrometric characterization (Figure 1). This technique provides efficient cell lysis and protein extraction, while filtering out humic materials and other interfering substances in soil. The use of detergents for cellular lysis is well-known, and has also been shown to be compatible

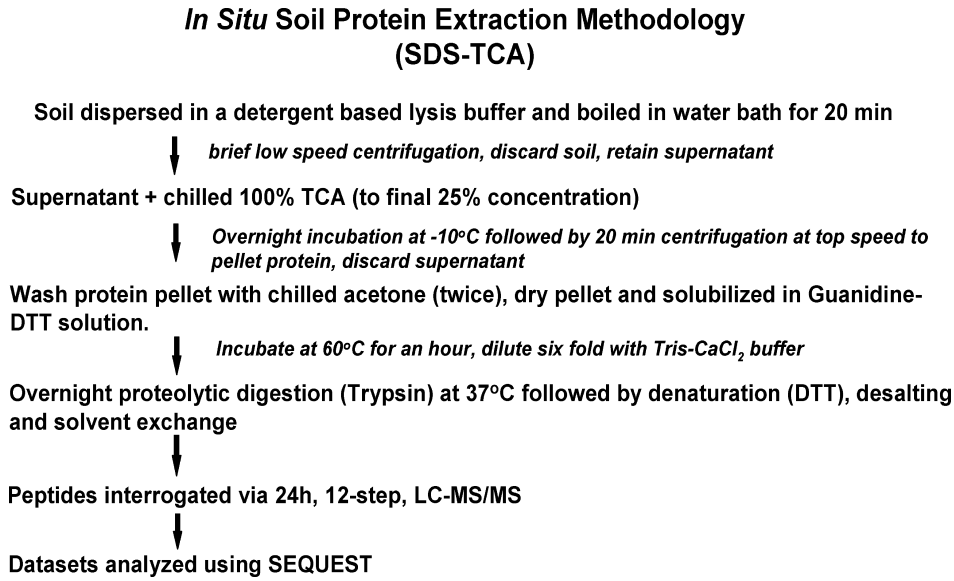

Figure 1. Flowchart illustrating the basic steps for direct soil protein extraction, highlighting detergent lysis, protein precipitation, proteolytic digestion, and MS characterization.

Table 1. Summary of Soil Microbe Proteome Study on Hanford and Hopland Soil Microcosm

sample matrix	inoculum	microbial biomass (cell number)	extraction protocol	cell lysis method	total proteins
LB liquid culture	*P. putida*	2×10^9	Direct	SDS-TCA	1343
Hanford soil	*P. putida*	2×10^9	Direct	SDS-TCA	925 (854)[a]
Hopland soil	*A. chlorophenolicus*	2×10^{10}	Indirect	SDS-TCA	600
Hopland soil	*A. chlorophenolicus*	5×10^9	Indirect	SDS-TCA	490
Hopland soil	*A. chlorophenolicus*	2×10^9	Indirect	SDS-TCA	582
Hopland soil	*A. chlorophenolicus*	2×10^{10}	Indirect	Guanidine lysis	732
Hopland soil	*A. chlorophenolicus*	5×10^9	Direct	SDS-TCA	816 (506)[a]
Hopland soil	*A. chlorophenolicus*	5×10^9	Direct	Guanidine lysis	145
Hopland soil	*A. chlorophenolicus*	5×10^8	Direct	SDS-TCA	555 (538)[a]
Hopland soil	None	NA[b]	Direct	SDS-TCA	716[c]/333[d]

[a] Technical replicate in parentheses. Data shown in Supplementary Tables 1B, 3D and 3E (Supporting Information) in descending order as shown in Table 1. [b] NA: Not applicable. [c] Redundant proteins. [d] Non redundant proteins.

for mass spectrometric (MS) analysis of eukaryotic cells,[34] although this particular approach is not amenable for the biomass concentrations and matrix complexity of soil microbial communities. For comparison, we also tested various standard cellular lysis methods (sonication, NaOH extraction, bead beating, and guanidine denaturation) for their ability to directly lyse cells in soil, but few proteins were detected, if any (data not shown).

Our novel direct SDS-TCA approach was first tested on sandy soil (Hanford, WA) spiked with the Gram-negative bacterium, *Pseudomonas putida* F1. For comparison, the proteome was also extracted from the same bacterium that was cultivated in liquid medium (LM). Our results showed good representation of the proteome under both conditions, with a total of 1343 and 925 proteins identified in LM and soil, respectively (Table 1). Identified proteins were sorted according to their relative abundances (spectral counts), functional categories, and localization (Supplementary Tables 1A, 1B, 2A, Supporting Information). In general, the distribution of proteins across Cluster of Orthologous Groups (COG) categories was similar in both samples, albeit in slightly reduced numbers for soil (Figure 2A). Proteins were grouped into COG categories based on their function and thus represent proteins from different cellular locations. Some COG categories, such as those for energy production and conversion, translation and biogenesis, amino acid transport and metabolism, transcription, and cellular motility were relatively less represented in the soil samples even after short-term (5 h) incubation in soil, and likely reflect an initial response of the bacterium to soil conditions, rather than extraction bias toward a particular COG category. The supposition is further supported by the protein localization data, discussed below, which reveals proteins with similar cellular localization profiles between the soil-extracted sample and the LM grown isolate, which would not be observed if there were extraction biases.

A scatter plot of peptide spectral counts between the LM and soil proteomes (Figure 2B) showed a modest correlation (although there were outliers primarily at the low end), suggesting minimal biases in protein recovery between these samples from very different matrices. A broad range of protein molecular masses (6−181 KDa) was detected in both LM and soil proteomes (Supplementary Tables 2B and 2C, Supporting Information). Both proteomes were evaluated for their protein localization distributions using PSORTdb,[35] and revealed representative proteins from both membrane and cytoplasmic locations present in each case (Supplementary Table 2A, Supporting Information, and Figure 3). This is the first demonstration of a direct proteome extraction method for microbes in soil that approaches protein yields and depth of proteome measurement comparable to established pure culture cellular lysis protocols.[36]

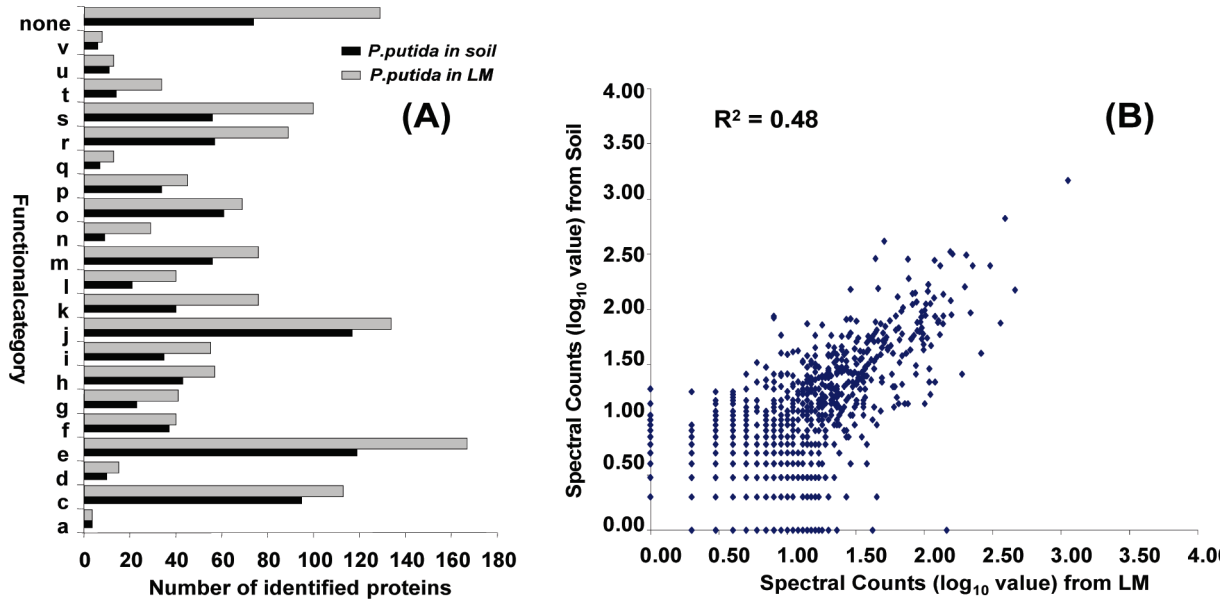

Figure 2. *P. putida* proteomes extracted using the direct SDS-TCA method. (A) *P. putida* proteome in soil and liquid medium (LM), sorted and compared by functional category (COG). Each letter corresponds to the functional category in the footnote. (B) Scatter plot of peptide spectral counts (log_{10} values) of soil proteins vs proteins identified in LM. The proteins on each axis represent those that were unique to LM (*x*-axis) or Soil (*y*-axis).

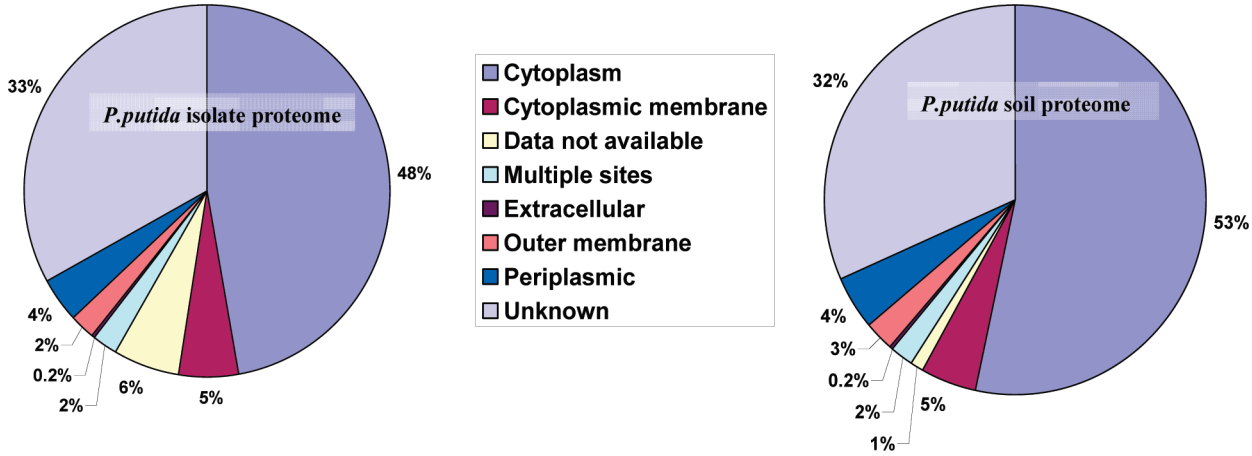

Figure 3. Identified proteome of *P. putida,* from soil or isolate (Liquid Medium: LM) compartmentalized as per their cellular localization using PSORTdb. The color-coded legend describes the cellular localization.

We extended evaluation of this protocol to include a different soil type (Hopland medium texture loam) inoculated with Gram-positive *Arthrobacter chlorophenolicus* for a longer incubation time (24 h). The cells were chromosomally tagged with the *gfp* gene, encoding green fluorescent protein (GFP) to differentiate the inoculum from possible indigenous *Arthrobacter* in the soil. The GFP protein also served as an internal protein standard control to help evaluate protein yields. The soil was amended with 150 ppm 4-chlorophenol, a specific substrate for *A. chlorophenolicus*, and the cells were inoculated into soil at two concentrations: 10^9 or 10^8 cells/gram soil. The goals of these incubation experiments were: (i) to evaluate and compare *direct* and *indirect* protein extraction protocols, and (ii) to test our protocol on a different soil type and with a more difficult to lyse Gram-positive bacterium.

Direct proteome extraction from 5 g soil samples using the SDS-TCA method yielded 816 *Arthrobacter* proteins (5×10^9 cells) or 555 proteins (5×10^8 cells), whereas the standard guanidine lysis protocol yielded only 145 proteins for the higher cell density sample (Table 1, Supplementary Table 3A, 3B, Supporting Information). In addition, the peptide spectral count and protein sequence coverage was much higher for the SDS-TCA extracted samples compared to those using the guanidine lysis approach (Supplementary Table 3A, Supporting Information). The direct SDS-TCA method again resulted in a broad distribution of protein masses (6−180 KDa), verifying the absence of bias on the basis of protein mass (Supplementary Table 3A, Supporting Information). A large variety of functional categories (COGs) were represented in the proteome profiles of the SDS lysed cells, but they were proportionately less represented in the guanidine lysed cells due to the lower number of proteins identified (Supplementary Table 3C, Supporting Information, and Figure 4A).

The SDS-TCA method was also used to lyse cells that were first extracted from the soil matrix by differential centrifugation (DC). This indirect protein extraction method resulted in a total of 490 proteins from 5 g soil containing 5×10^9 cells using SDS-TCA lysis and 732 proteins using guanidine lysis (Supplemen-

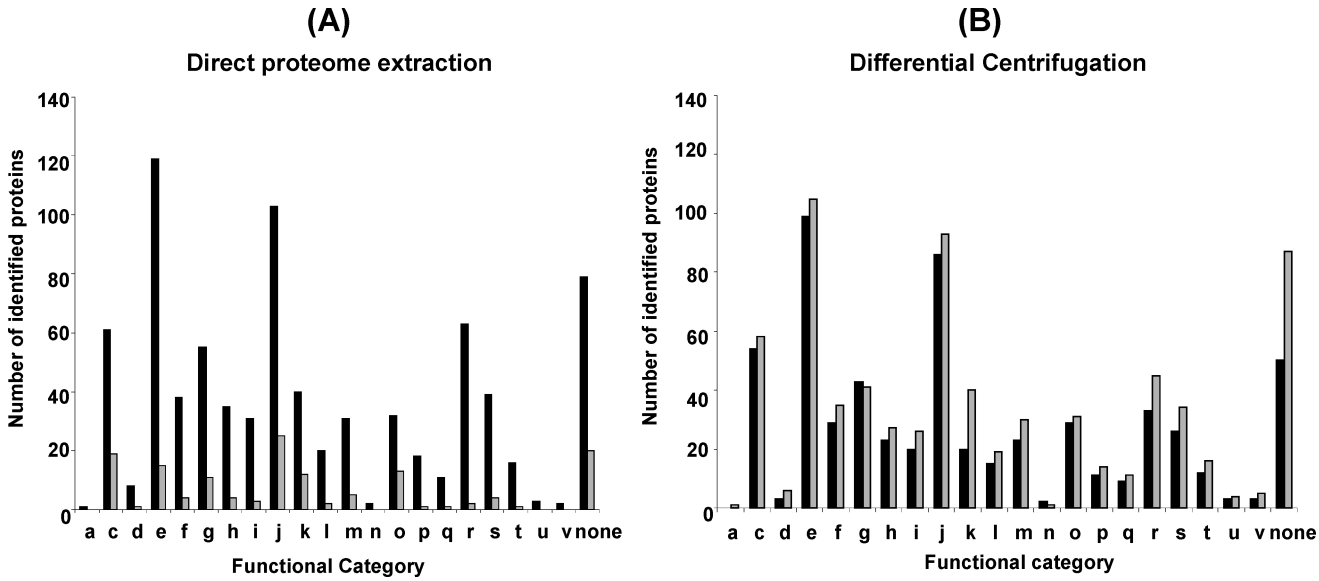

Figure 4. Functional category profile of *Arthrobacter* proteins extracted from soil via direct or indirect methods. (A) *A. chlorophenolicus* proteins directly extracted from soil microcosm (10^9 cells/g soil) using SDS (Black bars) or guanidine cell lysis (Gray bars). (B) *A. chlorophenolicus* proteins extracted from soil by an indirect extraction of cells from soil using differential centrifugation (DC) prior to cell lysis. The resulting microbial pellet was lysed using SDS-TCA (Black bars) or guanidine (Gray bars).

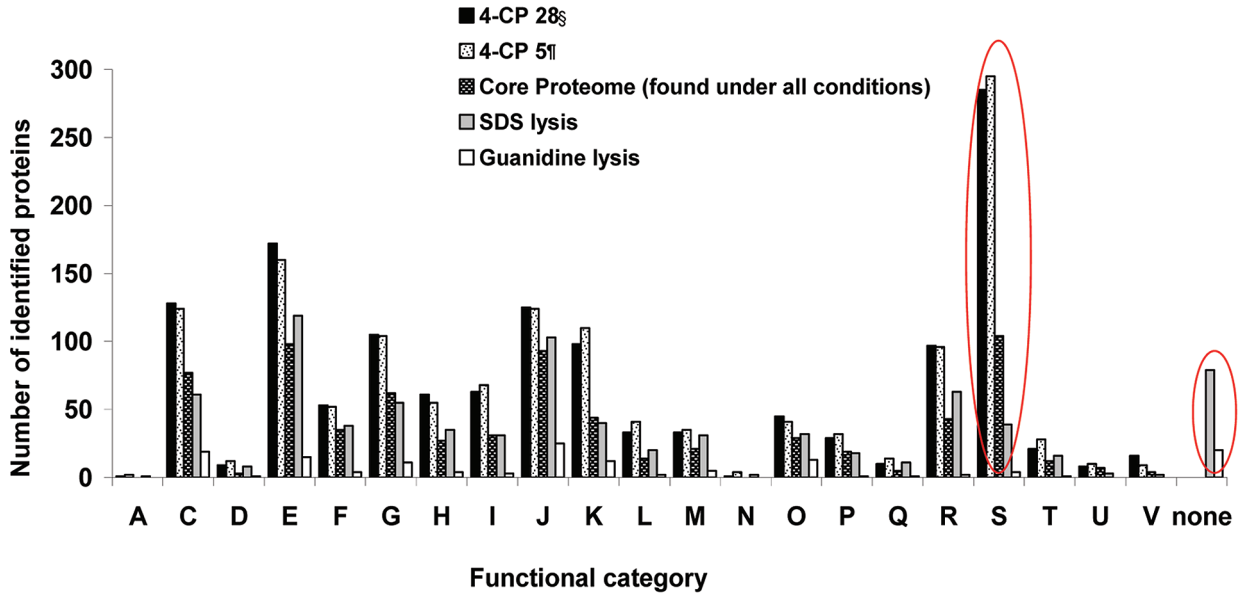

Figure 5. Distribution of proteins according to their functional category for proteomes derived from differently cultured and lysed *Arthrobacter* cells. Solid black bars, growth in liquid medium with 4-chlorophenol (4-CP) at 28°; white bars with black dots, growth in liquid medium with 4-CP at 5°; black bars with white dots, the core proteome found from grown using different growth substrates at different temperatures; gray bar, incubation in soil with 4-CP at 5° and lysed using SDS; white bar, incubation in soil with 4-CP at 5° and subjected to guanidine lysis. The red oval marks the category of proteins with unknown function, which is significantly higher using the draft protein database, compared to the finished database. The final database has an additional category called 'none' for proteins which have not yet been assigned a function. The description of abbreviations for COG categories can be found in the legend for Figure 4.

tary Table 4A and B, Supporting Information). The molecular weights of the proteins identified were similar in range for both SDS-TCA and guanidine lysis methods for this indirect DC approach (Supplementary Table 4B, Supporting Information). These results suggest that in the absence of the soil matrix, either lysis method can be employed effectively. The relative distribution of the proteins across functional categories was also similar for both SDS and guanidine lysis methods (Supplementary Table 4C, Supporting Information, and Figure 4B). Previous studies on *A. chlorophenolicus* have reported identification of ~1400 proteins (based on a draft genome) from a lab grown isolate,[37] providing a baseline metric for comparison with recovery of proteins from soil inoculated with *A. chlorophenolicus*. The previous study was focused on characterization of the *A. chlorophenolicus* proteome when cells are grown in liquid media under varying temperatures and amendments. Supplemental Table 7 (Supporting Information) and Figure 5 show how the current study compares with published work on this bacterium. One should keep in mind that previous work was done on liquid-media grown cells lysed in a different

manner and searched using a draft genome, whereas the current study is based on bacteria incubated in soil, and then detergent lysed and matched against finished version of database. Despite these differences, the proteome yield from soil-incubated *Arthrobacter* compares favorably for all COG categories to the earlier study. Figure 5 also includes a comparison with the 'core proteome' of *Arthrobacter*, as determined from various growth treatments. Note that the COG category "S", which are proteins with unknown function, is quite large in the earlier results since it was a draft genome, whereas in the final format, many proteins have been categorized into other groups with some still remaining in the functional category "None".

These results indicate that both direct and indirect extraction methods can be used to generate deep proteomic signatures (approaching a thousand proteins) from soil microorganisms (Supplementary Table 5A, B, Supporting Information). Note that the direct approach affords cellular lysis/proteome extraction in a single step (i.e., single processing tube) in which the boiling SDS condition ruptures the cells and readily denatures the sample, thereby reducing any undesirable proteome changes or losses that might be induced by more extensive or time-consuming sample handling. An additional consideration of the differential extraction procedure is unavoidable cellular loss, since normally only about one-third of the cell biomass is extracted while the remainder adheres to soil particles.[26] Thus, to bring the results from the indirect protocol on par with direct method, we either had to increase the cellular biomass in the same amount of soil, or use a slightly larger soil amount (which equates to more cells). There is also a potential for bias against secreted or extracellular proteins using the indirect approach (since they may be lost during cellular isolation), as compared to direct lysis of cells in the soil matrix.

The constitutively expressed internal GFP protein standard could be detected in the higher biomass samples, where it represented 0.01−0.06% of the relative spectral counts corresponding to *Arthrobacter* proteins, in good accord with its expected representation in the predicted proteome of this bacterium (i.e., 0.02%). However, one should keep in mind that the absolute concentrations of proteins detected are dependent on several factors including protein stability, turnover rate, expression levels, and edaphic properties.

As a final test of our protocol, we used the SDS-TCA method to directly extract proteins from the indigenous microbial population in uninoculated Hopland soil for which we have no matching metagenome sequence data. Searches against all sequenced microbes (1606 as of November 2009) plus a soil metagenome in the JGI IMG database (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi) identified 716 redundant and 333 non-redundant proteins, which matched primarily to previously characterized soil microbes (Supplementary Table 6, Supporting Information). These results represent the deepest proteome coverage to date for indigenous microbes in natural soil. This soil typically experiences extended summer drought and as such is populated by bacteria with appropriate growth strategies.[38] One example of this in the proteome was the expression of small acid-soluble spore proteins, likely of *Bacillus* spp. origin that are present in *Bacillus* spores and protect the DNA backbone from chemical and enzymatic cleavage.[39] As they are rapidly degraded upon spore germination, their detection demonstrates the presence of dormant *Bacillus* spores. These identifications are based on protein/microbe similarities found using the nonredundant database and should be considered *tentative* predictions of expressed soil proteome, since we do not absolutely know the correspondence between genome sequences of microbial isolates and similar versions found in environmental samples. We anticipate that substantially more confirmed protein identifications would be possible if a matched metagenome of the soil microbial community were available.

Conclusions

Here we have demonstrated a novel, broadly applicable protocol for successful extraction of microbial proteomes from soil. In total, this SDS-TCA extraction method, when coupled with high performance mass spectrometry, currently provides the most extensive proteome coverage of microbes directly from soils. Extensive proteome information could be obtained from soils having different biomass and varying microbial compositions. The protocols worked effectively in complex soil matrices, with native soil organic compounds that have previously been found to inhibit protein yields from soil, and with added organic contaminants. While we have tested the protocol on two significantly different soil types with encouraging success, we recognize that it will be important to evaluate this approach for other diverse soil types, and thus we are currently extending this work to investigate its universality for proteome extraction from other types of soil. The soil proteome extraction is straightforward to perform in any standard laboratory and appears quite robust under a variety of conditions, and thus provides deep and unbiased proteome characterization which should be broadly applicable to understanding microbial processes in soil and for systems-biology research in other areas of microbial ecology.

Acknowledgment. This research was sponsored by the Environmental and Remediation Sciences Program (ERSP), Biological and Environmental Research (BER), Office of Science, U.S. Department of Energy. This work was supported in part by the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 with the Lawrence Berkeley National Laboratory and by Laboratory Directed Research and Development awards to JKJ and ELB. Oak Ridge National Laboratory is managed by University of Tennessee-Battelle LLC for the Department of Energy under contract DOE-AC05-00OR22725.

Supporting Information Available: Supplementary tables. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Ram, R. J.; VerBerkmoes, N. C.; Thelen, M. P.; Tyson, G. W.; Baker, B. J.; Blake, R. C.; Shah, M.; Hettich, R. L.; Banfield, J. F. Community proteomics of a natural microbial biofilm. *Science* **2005**, *308*, 1915–20.

(2) VerBerkmoes, N. C.; Denef, V. J.; Hettich, R. L.; Banfield, J. F. Systems biology: Functional analysis of natural microbial consortia using community proteomics. *Nat. Rev. Microbiol.* **2009**, *7* (3), 196–205.

(3) Wilmes, P.; Bond, P. Microbial community proteomics: elucidating the catalysts and metabolic mechanisms that drive the Earth's biogeochemical cycles. *Curr. Opin. Microbiol* **2009**, *12* (3), 310–7.

(4) DeLong, E. F. Microbial population genomics and ecology: the road ahead. *Environ. Microbiol.* **2004**, *6* (9), 875–8.

(5) Keller, M.; Hettich, R. Environmental Proteomics: a Paradigm Shift in Characterizing Microbial Activities at the Molecular Level. *Microbiol. Mol. Biol. Rev.* **2009**, *73* (1), 62–70.

(6) Cravatt, B. F.; Simon, G. M.; Yates, J. R. The biological impact of mass-spectrometry-based proteomics. *Nature* **2007**, *450*, 991–1000.

(7) Atlas, R. M.; Unterman, R. *Bioremediation*; ASM Press: Washington, DC., 1999; pp 666−81.
(8) Falkowski, P.; Fenchel, T.; Delong, E. The Microbial Engines That Drive Earth's Biogeochemical Cycles. *Science* **2008**, *320*, 1034–8.
(9) Wingate, L.; Ogee, J.; Cuntz, M.; Genty, B.; Reiter, I.; Seibt, U.; Yakir, D.; Maseyk, K.; Pendall, E. G.; Barbour, M. M.; Mortazavi, B.; Burlett, R.; Peylin, P.; Miller, J.; Mencuccini, M.; Shim, J. H.; Hunt, J.; Grace, J. The impact of soil microorganisms on the global budget of _18O in atmospheric CO2. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106* (52), 22411–5.
(10) Elsas, J. D.; Jansson, J. K.; Trevors, J. T. *Modern soil microbiology*, 2nd ed.; CRC Press, Taylor and Francis group: Boca Raton, FL, 2006.
(11) Solaiman, Z.; Kashem, M. A.; Matsumoto, I. *Environmental Proteomics: Extraction and Identification of Protein in Soil*; Springer-Verlag: Berlin Heidelberg, 2007; Vol. 2.
(12) Ogunseitan, O. A. *Soil Proteomics: Extraction and analysis of proteins from soils*; Springer-Verlag: Berlin Heidelberg, 2006; Vol. 8.
(13) Bastida, F.; Moreno, J. L.; Nicolas, C.; Hernandez, T.; Garcia, C. Soil metaproteomics: a review of an emerging environmental science. Significance, methodology and perspectives. *Eur. J. Soil Sci.* **2009**, *60*, 845–59.
(14) Benndorf, D.; Balcke, G. U.; Harms, H.; Bergen, M. V. Functional metaproteome analysis of protein extracts from contaminated soil and groundwater. *ISME J.* **2007**, *1*, 224–34.
(15) Chen, S.; Rillig, M. C.; Wang, W. Improving soil protein extraction for metaproteome analysis and glomalin-related soil protein detection. *Proteomics* **2009**, *9* (21), 4970–3.
(16) Taylor, E. B.; Williams, M. A. Microbial Protein in Soil: Influence of Extraction Method and C Amendment on Extraction and Recovery. *Microb. Ecol.* **2010**, *59*, 390–9.
(17) Benndorf, D.; Vogt, C.; Jehmlich, N.; Schmidt, Y.; Thomas, H.; Woffendin, G.; Shevchenko, A.; Richnow, H. H.; Bergen, M. V. Improving protein extraction and separation methods for investigating the metaproteome of anaerobic benzene communities within sediments. *Biodegradation* **2009**, *20*, 737–50.
(18) Nannipieri, P. *Role of Stabilised Enzymes in Microbial Ecology and Enzyme Extraction from Soil with Potential Applications in Soil Proteomics*; Springer-Verlag: Berlin Heidelberg, 2006; Vol. 8.
(19) Schulze, W. X.; Gleixner, G.; Kaiser, K.; Guggenberger, G.; Mann, M.; Schulze, E. D. A proteomic fingerprint of dissolved organic carbon and of soil particles. *Oecologia* **2005**, *142*, 335–43.
(20) Fornasier, F.; Margon, A. Bovine serum albumin and Triton X-100 greatly increase phosphomonoesterases and arylsulphatase extraction yield from soil. *Soil Biol. Biochem.* **2007**, *39* (10), 2682–4.
(21) Mahowald, M. A.; Rey, F. E.; Seedorf, H.; Turnbaugh, P. J.; Fulton, R. S.; Wollam, A.; Shah, M.; Wang, C.; Magrini, V.; Wilson, R. K.; Cantarel, B. L.; Coutinho, P. M.; Henrissat, B.; Crock, L. W.; Russell, A.; Verberkmoes, N. C.; Hettich, R. L.; Gordon, J. I. Characterizing a model human gut microbiota composed of members of its two dominant bacterial phyla. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106* (14), 5859–64.
(22) Elväng, A. M.; Jernberg, C.; Westerberg, K.; Jansson, J. K. Monitoring bioremediation of 4-chlorophenol-contaminated soil using gfp or luc-tagged Arthrobacter chlorophenolicus A6. *Environ. Microbiol.* **2001**, *3*, 1–12.
(23) Waldrop, M. P.; Firestone, M. K. Seasonal dynamics of microbial community composition and function in oak canopy and open grassland soils. *Microb Ecol* **2006**, *52* (3), 470–9.
(24) Clair, S. B. S.; Sudderth, E. A.; Fischer, M. L.; Torn, M. S.; Stuart, S. A.; Salve, R.; Eggett, D. L.; Ackerly, D. D. Soil drying and nitrogen availability modulate carbon and water exchange over a range of annual precipitation totals and grassland vegetation types. *Global Change Biol.* **2009**, *15* (12), 3018–30.
(25) Thompson, M. R.; VerBerkmoes, N. C.; Chourey, K.; Shah, M.; Thompson, D. K.; Hettich, R. L. Dosage-Dependent Proteome Response of Shewanella oneidensis MR-1 to Acute Chromate Challenge. *J. Proteome Res.* **2007**, *6* (5), 1745–57.
(26) Holben, W. E.; Jansson, J. K.; Chelm, B. K.; Tiedje, J. M. DNA probe method for the detection of specific microorganisms in the soil bacterial community. *Appl. Environ. Microbiol.* **1988**, *54*, 703–11.
(27) Washburn, M. P.; Wolters, D.; Yates, J. R. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **2001**, *19*, 242–7.
(28) McDonald, W. H.; Ohi, R.; Miyamoto, D. T.; Mitchison, T. J.; Yates, J. R. Comparison of three directly coupled HPLC MS/MS strategies for identification of proteins from complex mixtures: single-dimension LC-MS/MS, 2-phase MudPIT, and 3-phase MudPIT. *Int. J. Mass Spectrom.* **2002**, *219*, 245–51.
(29) Brown, S. D.; Thompson, M. R.; Verberkmoes, N. C.; Chourey, K.; Shah, M.; Zhou, J.; Hettich, R. L.; Thompson, D. K. Molecular dynamics of the Shewanella oneidensis response to chromate stress. *Mol. Cell. Proteomics* **2006**, *5* (6), 1054–71.
(30) Thompson, M. R.; Chourey, K.; Froelich, J. M.; Erickson, B. K.; VerBerkmoes, N. C.; Hettich, R. L. Experimental approach for deep proteome measurements from small-scale microbial biomass samples. *Anal. Chem.* **2008**, *80* (24), 9517–25.
(31) Tringe, S. G.; von Mering, C.; Kobayashi, A.; Salamov, A. A.; Chen, K.; Chang, H. W.; Podar, M.; Short, J. M.; Mathur, E. J.; Detter, J. C.; Bork, P.; Hugenholtz, P.; Rubin, E. M. Comparative metagenomics of microbial communities. *Science* **2005**, *308* (5721), 554–7.
(32) Eng, J. K.; McCormack, A. L.; Yates, 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*, 976–89.
(33) Tabb, D. L.; McDonald, W. H.; Yates, J. R. DTASelect and Contrast: Tools for assembling and comparing protein identifications from shotgun proteomics. *J. Proteome Res.* **2002**, *1*, 21–6.
(34) Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **2009**, *6* (5), 359–62.
(35) Gardy, J. L.; Laird, M. R.; Chen, F.; Rey, S.; Walsh, C. J.; Ester, M.; Brinkman, F. S. L. PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics* **2005**, *21* (5), 617–23.
(36) Thompson, D. K.; Chourey, K.; Wickham, G. S.; Thieman, S. B.; VerBerkmoes, N. C.; Zhang, B.; McCarthy, A. T.; Rudisill, M. A.; Shah, M.; Hettich, R. L. Proteomics reveals a core molecular response of Pseudomonas putida F1 to acute chromate challenge. *BMC Genomics* **2010**, *11*, 311.
(37) Unell, M.; Abraham, P. E.; Shah, M.; Zhang, B.; Rü ckert, C.; VerBerkmoes, N. C.; Jansson, J. K. Impact of Phenolic Substrate and Growth Temperature on the Arthrobacter chlorophenolicus Proteome. *J. Proteome Res.* **2009**, *8*, 1953–64.
(38) DeAngelis, K. M.; Brodie, E. L.; DeSantis, T. Z.; Andersen, G. L.; Lindow, S. E.; Firestone, M. K. Selective progressive response of soil microbial community to wild oat roots. *ISME J.* **2009**, *3* (2), 168–78.
(39) Setlow, P. Small, Acid-Soluble Spore Proteins of Bacillus Species: Structure, Synthesis, Genetics, Function, and Degradation. *Annu. Rev. Microbiol.* **1988**, *42* (1), 319–38.

PR100787Q