Analysis of Biostimulated Microbial Communities from Two Field Experiments Reveals Temporal and Spatial Differences in Proteome Profiles

Stephen J. Callister[1*], Michael J. Wilkins[1*], Carrie D. Nicora[1], Kenneth H. Williams[2], Jillian F. Banfield[3], Nathan C. VerBerkmoes[4], Robert L. Hettich[4], Lucie N'Guessan[5], Paula J. Mouser[5], Hila Elifantz[5], Richard D. Smith[1], Derek R. Lovley[5], Mary S. Lipton[1] and Philip E. Long[1]

[1]Pacific Northwest National Laboratory, Richland WA, 99352; [2]Lawrence Berkeley National Laboratory, Berkeley CA, 94720; [3]University of California Berkeley, Berkeley CA, 94720; [4]Oak Ridge National Laboratory, Oak Ridge TN, 37831; [5]University of Massachusetts Amherst, Amherst MA, 01002

Brief: Comparison of proteomics measurements made during 2007 and 2008 field experiments reveal the failure of a microbial community to rebound following the cessation of biostimulation.

*Equal Co-authors

Abstract

Stimulated by an acetate-amendment field experiment conducted in 2007, anaerobic microbial populations in the aquifer at the Rifle Integrated Field Research Challenge site in Colorado reduced mobile U(VI) to insoluble U(IV). During this experiment, planktonic biomass was sampled at various time points to quantitatively evaluate proteomes. In 2008, an acetate-amended field experiment was again conducted in a similar manner to the 2007 experiment. As there was no comprehensive metagenome sequence available for use in proteomics analysis, we systematically evaluated 12 different organism genome sequences to generate sets of aggregate genomes, or "pseudo-metagenomes", for supplying relative quantitative peptide and protein identifications. Proteomics results support previous observations of the dominance of Geobacteraceae during biostimulation using acetate as sole electron donor, and revealed a shift from an early stage of iron reduction to a late stage of iron reduction. Additionally, a shift from iron reduction to sulfate reduction was indicated by changes in the contribution of proteome information contributed by different organism genome sequences within the aggregate set. In addition, the comparison of proteome measurements made between the 2007 field experiment and 2008 field experiment revealed differences in proteome profiles. These differences may be the result of alterations in abundance and population structure within the planktonic biomass samples collected for analysis.

1.0 Introduction

At the U. S. Department of Energy's Integrated Field Research Challenge (IFRC) site at Rifle, Colorado field research activities are advancing scientific understanding of biogeochemical reduction of soluble Uranium (VI) to insoluble Uranium (IV) (*1*). Biostimulation activities at the site include the addition of acetate to create redox conditions in the aquifer suitable for the enzymatic reduction of Fe(III) and U(VI) (*2*). During Fe(III) reduction, members of the *Geobacteraceae* family appear to dominate the aquifer microbial community relative to other phylogenetic groups (*1, 3, 4*). As the duration of biostimulation increases, subsurface geochemical conditions shift from an Fe(III)-reducing system to one dominated by microbial sulfate reduction (*5*). This transition is associated with a decrease in the rate of U(IV) removed from solution.

In 2007 and 2008, biostimulation field experiments were conducted to quantitatively compare spatially and temporally, the proteomes associated with different phases of Fe(III) and U(VI) reduction to look at the dynamics of the microbial community proteome in relationship to geochemical measurements. Three biomass samples were collected during the 2007 experiment; two from monitoring well D07 during early and late phases of Fe(III) reduction and an additional sample from well D05 when conditions for acetate and Fe(II) concentrations approximated those of early phase Fe(III) reduction (*6*). During the 2008 experiment, three biomass samples were recovered at the start of acetate addition from well D04.

Herein, we report the effects of the biostimulation experiments in terms of microbial community function and structure derived from liquid chromatography-mass spectrometry-based proteomics analyses utilizing isolate genomic data (*7*). Initially, in the absence of metagenomic data, community proteomics analysis of the samples collected in 2007 employed only genome sequences from *Geobacteraceae* to confirm the prominence of this family in the aquifer microbial community, infer shifts in metabolism, and changes in the community structure with increasing biostimulation (*6*). To investigate other bacterial species that may play a key role in biogeochemical processes at Rifle, a broader range of genome sequences from Fe(III) and sulfate-reducing bacteria from both the α- and δ-sub-divisions of the *Proteobacteria* were utilized. These data were used in subsequent analyses to determine the extent to

which communities rebound after cessation of biostimulation. We evaluated the contribution of proteins provided by each organism's genome sequence, as well as the relative abundance differences of identified proteins within and between monitoring wells. In addition to providing insight into the dynamics of a microbial community proteome, the resulting proteomics datasets may prove useful for discovering environmental markers to monitor the persistence (post biostimulation) and attenuation of biologically reduced U(IV).

2.0 Experimental

2.1 Field experiments and sample collection. The biostimulation experiment in 2008 followed the same protocols as previously published for the earlier field experiment in 2007 *(8)*. In brief, wells D04, D05 and D07 are part of a well gallery composed of 10 injection wells and 12 down-gradient monitoring wells arranged in three rows. Well D04 is present in the first row of monitoring wells, whereas D05 and D07 are present in the second row (5 meters down gradient from the injection wells). Acetate:bromide (50 mM:5 mM) amended groundwater was injected into the subsurface to provide a target acetate concentration of ~ 5 mM. The analysis of aquifer samples to determine Fe(II) and U(VI) concentrations is described elsewhere (*6*).

Planktonic biomass for proteomics analysis was harvested from the subsurface during the course of *in-situ* Fe(III) and U(VI) bioreduction. During the 2007 experiment, two samples were collected from D07, 9-days and 21-days after the start of acetate injection (D07(1) and D07(2), respectively). One sample was collected 15-days after acetate injection from well D05 (Fig. 2). During the 2008 experiment, biomass samples were recovered from well D04. In this study, we utilized samples collected 5, 7, and 10 days after start of biostimulation (Fig 3A). Groundwater was prefiltered (20 μm) and biomass harvested using a Pelicon tangential flow filtration (0.2 μm) system (Millipore, MA, USA) to filter 500 L at a rate of ~2 L min^{-1}. To preserve biomass, the extracted groundwater was passed through chilling baths that contained an ice-rock salt mixture. Groundwater temperature was maintained at ~1 ^{o}C. Biomass was

concentrated to ~ 200 mL following filtration and then centrifuged at 4000 rpm for 40 min at 4 ^{o}C to form a cell pellet. The pellet was resuspended in ~ 5 mL of groundwater and immediately frozen in an ethanol-dry ice mix.

2.2 Protein extraction from planktonic biomass samples. Harvested cell pellets were washed and suspended in 100 mM NH_4HCO_3 (pH 8.4) and then lysed via Pressure Cycling Technology (PCT) using a barocylcer (Pressure BioSciences Inc., South Easton, MA). The barocycler was operated for 20 s at 35 kpsi, followed by 10 s at ambient pressure. These conditions were repeated for 10 cycles. Following cell lysis, global, soluble, and insoluble protein fractions were extracted from the cell lysates, using established protocols (*7*).

2.3 Liquid chromatography-tandem mass spectrometry (LC-MS/MS) and reference peptide database generation. Peptides from the protein digests were fractionated to reduce sample complexity, using high-resolution reversed-phase HPLC according to established protocols (*7*). From each fraction, 10 µg of peptides were analyzed on a quad column HPLC system (built in-house) coupled to an LTQ mass spectrometer (ThermoFisher Scientific, San Jose, CA). Reverse-phase separation of peptide digests occurred by way of in-house manufactured columns (60 cm x 360 µm o.d. × 75 µm i.d. fused silica capillary tubing) packed with 3 µm Jupiter C_{18} stationary phase (Phenomenex, Torrence, CA). Prior to sample injection, the HPLC system was equilibrated with 100% mobile phase A (0.2% acetic acid and 0.05% TFA in water), and 50 min post injection, mobile phase B (0.1% TFA in 90% acetonitrile/10% water) was introduced to displace mobile phase A, which generated an approximate exponential gradient. Split flow controlled the gradient speed operating under constant pressure (10 kpsi). Separated peptides were ionized (positive) using an electrospray ionization interface (manufactured in-house) that consisted of chemically etched electrospray emitters (150 mm o.d. × 20 mm i.d).

The SEQUEST algorithm was employed to search MS/MS spectra against theoretical spectra to assign peptide sequences (*9*). Theoretical spectra were derived from the genome sequences of 12 bacteria

(Table 1) obtained from the Joint Genome Institute's database of Integrated Microbial Genomes (http://www.jgi.doe.gov/). These genome sequences were systematically concatenated by the sequential addition of sequences into 12 pseudo-metagenomes to generate the reference peptide databases shown in Table 1. The order of sequence addition was based upon decreased relative distance (calculated from aligned 16S rDNA sequences) between *Geobacter* strain M21 and the remaining 11 organisms. Because *Geobacter* strain M21 was isolated from the Rifle site, this organism appears first in the database (P430).

Each database contained full-tryptic peptides of at least six amino acids in length, having 1+, 2+, and 3+ charge states and SEQUEST cross-correlation (Xcorr; an indicator of the statistical match between the observed and theoretically generated MS/MS spectra) scores of ≥ 1.9, 2.2, or 3.5, respectively. Each database was further filtered to achieve less than a 10% false discovery rate (FDR) for peptide identifications. To estimate the FDR, a mixture of two multivariate Gaussian distributions, one depicting random identifications and the other depicting true identifications was fitted to SEQUEST scores Xcorr, and ΔCn (the difference between the Xcorr value of the current candidate sequence and the value of the top candidate sequence). A p-value was calculated for each peptide sequence as the probability that the identification belonged to the random distribution, and from that value, a q-value (the expected FDR) was derived (*10, 11*).

Initially, peptides were identified from 25 LC-MS/MS analyses of the 2007 samples using each of the 12 reference databases. This allowed us to evaluate how the addition of an organism's genome sequence affected the number of identified peptides and proteins. Based upon this evaluation, the reference peptide database that provided the most proteome information was expanded to include peptides identified from an additional 132 LC-MS/MS analyses. This expanded reference database was used to assign peptide and protein information to quantitative LC-MS measurements of samples collected in both years.

2.4 High resolution, high mass accuracy LC-MS and label free quantitation. The accurate mass and time tag proteomics approach *(7, 12)* afforded label-free arbitrary abundance estimates of peptides from LC-MS measurements of biomass samples collect during the 2007 experiment (D07(1), D07(2), D05) and 2008 experiment (D04; collected 5, 7, and 10 days following the start of biostimulation). Triplicate measurements were performed on a custom-built HPLC system coupled via electrospray ionization to an LTQ-Orbitrap™ mass spectrometer (ThermoFisher Scientific, San Jose, CA). HPLC conditions were the same as reported above. Mass measurement accuracy and elution time accuracy cut-offs of 5 ppm and 1%, respectively, were applied to observed mass and elution time features prior to matching these features to information in the reference peptide database to identify peptides.

Peptide abundances were calculated by integrating the signal strength under each peak of the LC-MS spectra (*13*). These abundances were normalized using central tendency normalization (*14*) to a common baseline, and then "rolled up" using the Z-score rollup algorithm available in DANTE (*15*) to a protein abundance estimate. Protein abundances were clustered using a Euclidean distance metric and hierarchical algorithm (average linkage) available in the data analysis software package OmniViz™ (version 6.0). Non-metric multidimensional scaling of rolled-up protein abundances was accomplished using the VEGAN package in R (*16*).

3.0 Results and Discussion

The construction of a database of aggregate sets of genome sequences, or "pseudo-metagenome" provided proteomics data for comparing 2007 and 2008 experiments. While an optimal strategy for proteomics analysis of the field experiment samples would have been to use metagenomes representative of the monitoring wells to identify peptide sequences, the absence of such information mandated an alternative approach. As a result, we constructed "pseudo-metagenomes" from multiple moderately close and closely related organisms that allowed us to identify peptides. With each genome addition to the pseudo-metagenome, the number of peptides identified from the 25 LC-MS/MS datasets increased, from

~12,300 identified from the *Geobacter* sp. M21 genome to ~32,900 identified using an extended pseudo-metagenome of 12 organisms (Fig. S1 of the Supporting Information, database P430 to P450). Expectedly, a greater degree of peptide ambiguity (peptides shared between organisms) was observed with each additional genome sequence added (Fig. 1A). The measured abundances of these ambiguous peptides cannot be definitively assigned to a single protein or organism. Much of the observed ambiguity is attributed to using more than one *Geobacter* sequence. For example, addition of the *G. metallireducens* sequence approximately doubled the number of ambiguous peptides (Fig. 1A, P434 to P435), and addition of the *G. bemidjiensis* sequence caused the number of ambiguous peptides to exceed the number of unique peptides (Fig. 1B, P438 to P440). At the protein level (Fig. 1C), the largest percentage of *Geobacter* strain M21 proteins, identified by ≥2 unique peptides drops considerably with the addition of the *G. bemidjiensis* genome sequence. A balance in the "trade-off" between new proteins identified from unique peptides verses the loss of proteins resulting from ambiguous peptides was observed when the genomes of *G. uraniireducens* and *G. bemidjiensis* were excluded from the "pseudo-metagenome (Fig. 1C, P437). As a result, this latter "pseudo-metagenome" was used to construct the final proteomics database for quantitative proteome comparison of the 2007 and 2008 samples.

2008 global protein profiles do not approximate those of 2007. To investigate how peptide and protein abundances shifted over the course of the two biostimulation experiments, Z-score values (*18*) from normalized abundance data were calculated. Non-metric multi-dimensional scaling clustering of the 18 datasets (three technical replicates for each of six samples) revealed tight clustering between technical replicates, but significant differences between the three samples from 2007 (Fig. 2A). Conversely, all of the samples from well D04 collected in 2008 clustered closely together, and nearer to D07(2) than to D07(1) and D05. The cluster separation of D07(1), D07(2) and D05 highlights the influence of temporal and spatial variability on the observed global protein complement during the 2007 experiment. In a larger context, we hypothesize that the cluster separation in the observed proteomes of the 2007 and 2008

experiments is a result of a "legacy" effect; that is, biostimulation performed in 2007 impacted the proteome of the microbial community present at the start of the 2008 experiment.

Proteome differences among the samples collected in 2007 were assessed. The clustering of Z-score values (Fig. 2B) indicates the temporal impact on proteomes, as nearly half of the proteins in common (~750) between D07(1) and D07(2) showed a significant difference between the "early" phase of Fe(III)-reduction (D07(1) and D05) and the "end" phase of Fe(III)-reduction (D07(2)). Although these few data points are not directly linked to geochemical variables, the temporal effect here correlates best with Fe(II) concentrations (Fig. 2). Note, measurements for sulfate concentration in wells D07 and D05 are shown in Supporting Information Figure S2. While correlations suggest that other geochemical variables play a role in the differences between proteome measurements, these data suggest the importance of "bioavailable" Fe(III) availability on the proteome. Additionally, a greater proteome similarity between D07(1) and D05 matches can be linked to similar Fe(II) concentrations at the time of sampling. Note that the significant differences in the abundances of peptides common to all samples also were observed at the protein level (Fig. S3 of the Supporting Information).

Shifts in community structure and function explain differences between 2007 to 2008 experiments. Microbial community functional and structural changes are intricately linked across all samples. Therefore, a Z-score significance cutoff-value of 0.5 was selected to identify proteins exhibiting significant abundance changes among samples D07(1), D07(2), and D04(day 5). These proteins were classified according to their general function, and shifts in protein group categories were identified (Fig. 3). Additionally, shifts in microbial populations were inferred from percent contributions of proteome information provided by each genome sequence that underwent a significant abundance change between two samples (Table 2). Changes in protein numbers within each functional cluster between early and late phases of Fe(III) reduction (i.e., D07(1) and D07(2)) agreed with trends reported by Wilkins et al. (*6*); that is, the abundance of proteins that matched *Geobacter* strain M21 indicated a strain(s) closely matching this Fe(III)-reducing species predominates during biostimulation. In his report, Wilkins et al.

used a "pseudo-metagenome" constructed from *Geobacter* genome sequences only. However, with additional organism sequences, further shifts in community function and structure could be inferred from the proteomics data obtained for the 2007 experiment samples. For example, comparing D07(1) and D07(2), we observed a clear decrease in *Geobacter* strain M21 proteins coupled to proportional increases in proteins that matched *A. dehalogens*, *P. carbinolicus*, *G. lovleyi*, *G. sulfurreducens* and *D. desulfuricans*. A proportion of these increased proteins is associated with energy production such as ATP synthase subunits and TCA cycle enzymes, which helps to explain the relatively small decrease in the "energy production and conversion" category between the two time points (Fig. 3A) that is likely due to increased energy requirements during growth. However, the large decrease observed within the "translation, ribosomal structure and biogenesis" category over the same time period suggests that while some strains are slowly increasing in abundance, the growth rate for the dominant strains (e.g., *Geobacter* strain M21) has greatly decreased in D07(2). The presence of co-existing unique peptides in orthologous proteins confirms the presence of multiple strains rather than a single organism that has a "composite" protein phenotype (see Fig. S4 of the Supporting Information).

Changes in community structure during the 2007 experiment are coupled to proteomics-inferred shifts in microbial respiration. While Fe(III) reduction is the dominant respiratory pathway in D07(2), the detection of proteins associated with dissimilatory sulfate reduction (ATP-sulfurylase, sulfite reductase) suggest the beginning of a shift from Fe(III) to sulfate as the primary electron acceptor for microbial growth. Increases in detected peptides that match proteins from the sulfate-reducing species *D. desulfuricans* further supports this result (Table 2). Although this transition has been observed during previous biostimulation experiments at the site (*1*), acetate addition to the subsurface during the 2007 experiment was stopped intentionally before sulfate reduction became the dominant biogeochemical process (Fig. 2). The detection of proteins responsible for sulfate reduction during this "late" stage of Fe(III) reduction indicates that we have captured the beginning of the transition period.

Low-level sulfate reduction at the Rifle site is typically hard to detect using traditional aqueous geochemical measurements alone, because of abiotic reactions between the products of Fe(III)- and

sulfate reduction that result in the formation of FeS precipitates. These precipitates preclude in-field chemical assay detection of low concentrations of aqueous S^{2-}. The ability of the proteomics methods used in this study to detect proteins involved in sulfate-reduction suggests that these techniques are useful for assessing low-level sulfate reduction in these environments, particularly when combined with geochemical indicators sensitive to the onset of dissimilatory sulfate reduction, such as the isotopic fraction of groundwater sulfate and increases in $\delta^{34}S(SO_4^{2-})$ (*19*).

More detailed comparisons between D04(day 5), i.e., a 2008 sample taken at the beginning of acetate amendment and the 2007 samples from well D07 were used to examine any potential legacy effects that may exist within the microbial populations at the beginning of the 2008 acetate amendment. Relative to D07(1), decreases in *Geobacter* strain M21 proteins are coupled to increases in the abundances of the seven other species in D04(day 5) (Table 2). Three TCA cycle proteins from *G. lovleyi*, citrate synthase, malate dehydrogenase, and isocitrate dehydrogenase are all up-regulated, while multiple RNA polymerases and methyltransferases are detected that match *R. ferrireducens*. This pattern of greater species diversity at the start of the 2008 experiment relative to 2007 is the likely reason that the D04 samples collected over three time points all cluster more closely to the proteomic signature of biomass recovered at the end of the 2007 experiment (D07(2)) (Fig. 2A). However, D04(day 5) still contains a greater number of up-regulated proteins from *Geobacter* strain M21 relative to D07(2), with significant decreases in proteins matching *R. ferrireducens* and *G. lovleyi* (Table 2). These abundance shifts indicate that at the start of the 2008 experiment, fast-growing strain M21-type *Geobacter* species are rapidly enriched upon acetate addition in a similar manner to that seen at the beginning of the 2007 experiment (Table 2). These strains have been shown to predominate in a range of environments and thus it is not surprising that they quickly dominate the microbial community, despite the increased diversity inferred from the proteomic data (*20*). The rapid growth of these strains following the addition of acetate to the aquifer at the beginning of the 2008 experiment is reflected in changes in up-regulated protein categories.

Relative to both D07 samples, there were increased abundances of proteins within the "translation, ribosomal structure and biogenesis" category in D04(day 5), suggestive of a higher growth

rate during the initial stages of the 2008 experiment (Fig. 3B and C). Approximately 55-65 % of these proteins (mainly ribosomal proteins and tRNA synthases) were associated with *Geobacter* strain M21, which further suggests that this was the dominant strain accounting for the majority of the growth. A similar pattern is observed in the "energy production" category; of the proteins up-regulated in D04(day 5) relative to D07(2), almost 60% can be assigned to *Geobacter* strain M21. While the presence of up-regulated nitrogen fixation proteins within this category demonstrates the importance of this process in supporting high growth rates under relatively nutrient poor conditions, the increased abundance of proteins associated with dissimilatory sulfate reduction in D04(day 5) relative to D07(2) suggests that the activity of sulfate-reducing bacteria is rapidly induced following the addition of acetate during the second year of biostimulation.

Differences within the microbial community structure at the beginning of the 2007 and 2008 experiments suggest that the effects of acetate addition may persist over a larger time scale than was previously thought. While acetate amendment is occurring, clear microbial community shifts correlate with the availability of acetate (*3, 6*). These data suggest that once amendment has stopped, residual carbon concentrations and/or biomass breakdown products may continue to act as a stimulant for low-level microbial growth, which may explain the more diverse community structure detected at the start of the 2008 experiment. This "legacy" effect has implications for bioremediation projects intending to stimulate microbial activity. For both *in situ* field experiments and modeling tools employed to predict biogeochemical processes, the effects of these shifts on the ability to achieve desired outcomes (e.g. uranium immobilization) during subsequent field experiments needs to be examined closely. The inferred increase in abundance of enzymes associated with dissimilatory sulfate reduction in D04(day5) collected in 2008 relative to D07(2) collected in 2007 is an example of how an altered community may impact the efficacy of U(VI) reduction. The increased activity of sulfate-reducing enzymes at the beginning of the 2008 experiment may have played a key role in shortening the period of dominant Fe(III) reduction during which the rate of U(VI) reduction is typically the fastest (*8*).

This study demonstrated the strength of proteomics techniques for analyzing microbial populations in biostimulated environments. By systematically constructing a reference database encompassing a range of microorganisms expected to predominate under stimulated conditions, shifts in both microbial physiology and community structure can be inferred. Using this approach, we identified multiple strain variants within biomass samples and illustrated clear differences that occur between microbial communities, both on a spatial and temporal scale. These results have important implications for carrying out multiple biostimulation experiments within the same flow-cell, and can complement geochemical analyses to provide a greater understanding of the processes occurring in the subsurface at Rifle during biostimulation.

Acknowledgements

We thank the city of Rifle, Colorado, the Colorado Department of Public Health and Environment, and the U.S. Environmental Protection Agency Region 8 for their co-operation in this study. Research was supported by the U.S. Department of Energy Office of Biological and Environmental Research (DOE/BER) through the Environmental and Remediation Sciences Program. Portions of this work were performed in the Environmental Molecular Sciences Laboratory a DOE/BER national scientific user facility at the Pacific Northwest National Laboratory. PNNL and Lawrence Berkeley National Laboratory are managed under contracts DE-AC05-76RL01830 and DE-AC02-05CH11231 (DOE-LBNL), with Battelle Memorial Institute.

References

(1) Anderson, R. T.; Vrionis, H. A.; Ortiz-Bernad, I.; Resch, C. T.; Long, P. E.; Dayvault, R.; Karp, K.; Marutzky, S.; Metzler, D. R.; Peacock, A.; White, D. C.; Lowe, M.; Lovley, D. R. Stimulating the in situ activity of *Geobacter* species To remove uranium from the groundwater of a uranium-contaminated aquifer. *Appl. Environ. Microbiol.* **2003,** *69*, 5884-5891.

(2) N'Guessan, A. L.; Vrionis, H. A.; Resch, C. T.; Long, P. E.; Lovley, D. R. Sustained removal of uranium from contaminated groundwater following stimulation of dissimilatory metal reduction. *Environ. Sci. Technol.* **2008,** *42*, 2999-3004.

(3) Vrionis, H. A.; Anderson, R. T.; Ortiz-Bernad, I.; O'Neill, K. R.; Resch, C. T.; Peacock, A. D.; Dayvault, R.; White, D. C.; Long, P. E.; Lovley, D. R. Microbiological and geochemical heterogeneity in an in situ uranium bioremediation field site. *Appl. Environ. Microbiol.* **2005,** *71*, 6308-6318.

(4) Mouser, P. J.; N'Guessan, A. L.; Elifantz, H.; Holmes, D. E.; Williams, K. H.; Wilkins, M. J.; Long, P. E.; Lovley, D. R. Influence of heterogeneous ammonium availability on bacterial community structure and the expression of nitrogen fixation and ammonium transporter genes during in situ bioremediation of uranium-contaminated groundwater. *Environ. Sci. Technol.* **2009,** *43*, 4386-4392.

(5) Williams, K. H.; Kemna, A.; Wilkins, M. J.; Druhan, J.; Arntzen, E.; N'Guessan, A. L.; Long, P. E.; Hubbard, S. S.; Banfield, J. F. Geophysical monitoring of coupled microbial and geochemical processes during stimulated subsurface bioremediation. *Environ. Sci. Technol.* **2009,** *43*, 6717-6723.

(6)Wilkins, M. J.; VerBerkmoes, N. C.; Williams, K. H.; Callister, S. J.; Mouser, P. J.; Elifantz, H.; N'Guessan A, L.; Thomas, B. C.; Nicora, C. D.; Shah, M. B.; Abraham, P.; Lipton, M. S.; Lovley, D. R.; Hettich, R. L.; Long, P. E.; Banfield, J. F. Proteogenomic monitoring of *Geobacter* physiology during stimulated uranium bioremediation. *Appl. Environ. Microbiol.* **2009,** *75*, 6591-6599.

(7) Lipton, M. S.; Pasa-Tolic, L.; Anderson, G. A.; Anderson, D. J.; Auberry, D. L.; Battista, K. R.; Daly, M. J.; Fredrickson, J.; Hixson, K. K.; Kostandarithes, H.; Masselon, C.; Markillie, L. M.; Moore, R. J.; Romine, M. F.; Shen, Y. F.; Stritmatter, E.; Tolic, N.; Udseth, H. R.; Venkateswaran, A.; Wong, L. K.; Zhao, R.; Smith, R. D. Global analysis of the *Deinococcus radiodurans* proteome by using accurate mass tags. *Proc. Natl. Acad. Sci. U.S.A.* **2002,** *99*, 11049-11054.

(8) Williams, K. H.; Long, P. E.; N'Guessan, A. L.; Wilkins, M. J.; Davis, J. A.; Kerkhof, L. J.; McGuinness, L. M.; Yang, L.; Newcomer, D. R.; Spane, F. A.; Dayvault, R.; Lovley, D. R. Acetate availability and its influence on sustainable bioremediation of uranium-contaminated groundwater. *Geomicro. J.* **2010,** *in press*

(9) Eng, J. K.; McCormack, A. L.; Yates III, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **1994,** *5*, 976-989.
(10) Storey, J. D.; Tibshirani, R. Statistical significance for genome wide studies. *Proc. Natl. Acad. Sci. U.S.A.* **2003,** *100*, 9440-9445.

(11) Lopez-Ferrer, D.; Martinez-Bartolome, S.; Villar, M.; Campillos, M.; Martin-Maroto, F.; Vazquez, J. Statistical model for large-scale peptide identification in databases from tandem mass spectra using SEQUEST. *Anal. Chem.* **2004,** *76*, 6853-6860.

(12) Smith, R. D.; Anderson, G. A.; Lipton, M. S.; Pasa-Tolic, L.; Shen, Y. F.; Conrads, T. P.; Veenstra, T. D.; Udseth, H. R. An accurate mass tag strategy for quantitative and high-throughput proteome

measurements. *Proteomics* **2002,** *2*, 513-523.

(13) Smith, R. D.; Shen, Y. F.; Tang, K. Q. Ultrasensitive and quantitative analyses from combined separations-mass spectrometry for the characterization of proteomes. *Accounts Chem. Res.* **2004,** *37*, 269-278.

(14) Callister, S. J.; Barry, R. C.; Adkins, J. N.; Johnson, E. T.; Qian, W. J.; Webb-Robertson, B. J. M.; Smith, R. D.; Lipton, M. S. Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. *J. Proteome Res.* **2006,** *5*, 277-286.

(15) Polpitiya, A. D.; Qian, W. J.; Jaitly, N.; Petyuk, V. A.; Adkins, J. N.; Camp, D. G.; Anderson, G. A.; Smith, R. D. DAnTE: a statistical tool for quantitative analysis of -omics data. *Bioinformatics* **2008,** *24*, 1556-1558.

(16)Oksanen, J.; Roeland, K.; Legendre, P.; O"Hara, B.; Simpson, G. L.; Solymos, P.; Henry, M.; Stevens, H.; Wagner, H. vegan: Community Ecology Package. R package version 1.15-4. **2009** http://CRAN.R-project.org/package=vegan

(17) Gomez-Alvarez, V.; Teal, T. K.; Schmidt, T. M. Systematic artifacts in metagenomes from complex microbial communities. *ISME J*. **2009**, *3*, 1314-1317

(18) Callister, S. J.; Nicora, C. D.; Zeng, X. H.; Roh, J. H.; Dominguez, M. A.; Tavano, C. L.; Monroe, M. E.; Kaplan, S.; Donohue, T. J.; Smith, R. D.; Lipton, M. S. Comparison of aerobic and photosynthetic *Rhodobacter sphaeroides* 2.4.1 proteomes. *J. Microbiol. Meth.* **2006,** *67*, 424-436.

(19 Druhan, J. L.; Conrad, M. E.; Williams, K. H.; N'Guessan, L.; Long, P. E.; Hubbard, S. S. Sulfur isotopes as indicators of amended bacterial sulfate reduction processes influencing field scale uranium bioremediation. *Environ. Sci. Technol.* **2008,** *42*, 7842-7849.

(20) Holmes, D. E.; O'Neil, R. A.; Vrionis, H. A.; N'Guessan, L. A.; Ortiz-Bernad, I.; Larrahondo, M. J.; Adams, L. A.; Ward, J. A.; Nicoll, J. S.; Nevin, K. P.; Chavan, M. A.; Johnson, J. P.; Long, P. E.; Lovley, D. R. Subsurface clade of *Geobacteraceae* that predominates in a diversity of Fe(III)-reducing subsurface environments. *ISME J.* **2007,** *1*, 663.

Table legends

Table 1. Genome sequences of organisms used to construct a "pseudo-metagenome" for the identification of peptide sequences and proteins from Winchester and Big Rusty field experiments.

Table 2. Shifts in microbial community composition as illustrated by changes in percent contributions of proteome information provided by the genome sequences selected for this quantitative analysis

Figure legends

Figure 1. Counts for (A) ambiguous peptides, (B) unique peptides and (C) proteins inferred from unique peptides, as genome sequences are systematically added to the proteomic search database.

Figure 2. (A) Non-metric multidimensional scaling plot illustrating the relative distribution of the datasets for 2007 planktonic biomass samples from wells D07(1) – 9-days following acetate injection, D07(2) – 21 days after acetate injection, and D05 – 15-days after acetate injection. 2008 planktonic samples were collected from well D04 following acetate injection at the time points indicated. (B) Geochemical data (previously published; Ref. 6) for Fe(II), acetate and U(VI) over the duration of biostimulation during the Winchester experiment at Rifle. Heat map demonstrating clustering patterns for unique peptides detected over the 3 samples.

Figure 3. Shifts in proteins that significantly increase in abundance in one sample relative to another. COG clusters; (A) Translation, ribosomal structure and biogenesis; (B) RNA processing and modification; (C) Transcription; (D) Replication, recombination and repair; (E) Chromatin structure and dynamics; (F) Cell cycle control, cell division and chromosome partitioning; (G) Nuclear structure; (H) Defense mechanisms; (I) Signal transduction mechanisms; (J) Cell wall, membrane and envelope biogenesis; (K) Cell motility; (L) Cytoskeleton; (M) Extracellular structures; (N) Intracellular trafficking, secretion, and vesicular transport; (O) Post-translational modification, protein turnover, and chaperones; (P) Energy production and conversion; (Q) Carbohydrate transport and metabolism; (R) Amino acid transport and metabolism; (S) Nucleotide transport and metabolism; (T) Coenzyme transport and metabolism; (U) Lipid transport and metabolism; (V) Inorganic ion transport and metabolism; (W) Secondary metabolites biosynthesis, transport and catabolism; (X) General function prediction only; (Y) Function unknown.

Table 1.

Organism	Reference Peptide Database											
	P430	P431	P432	P433	P434	P435	P436	P439	P437	P438	P440	P450
Geobacter strain M21	X	X	X	X	X	X	X	X	X	X	X	X
Rhodoferax ferrireducens T118		X	X	X	X	X	X	X	X	X	X	X
Desulfovibrio desulfuricans G20			X	X	X	X	X	X	X	X	X	X
Anaeromyxobacter dehalogens				X	X	X	X	X	X	X	X	X
Pelobacter carbinolicus DSM2380					X	X	X	X	X	X	X	X
Geobacter metallireducens GS-15						X	X	X	X	X	X	X
Geobacter sulfurreducens PCA							X	X	X	X	X	X
Geobacter lovleyi SZ								X	X	X	X	X
Geobacter strain FRC32									X	X	X	X
Geobacter uraniireducens Rf4										X	X	X
Geobacter bemidjiensis Bem											X	X
Arthrobacter sp. FB24												X

Table 2.

	2007-early D07[1] *vs* 2007-late D07[2]		**2007-early D07[1] *vs* 2008-early D04 day5**		**2007-late D07[2] *vs* 2008-early D04 day 5**	
Species/Strain	**D07[1]**	**D07[2]**	**D07[1]**	**D04**	**D07[2]**	**D04**
Geobacter strain M21	62	50	66	52	53	65
Rhodoferax ferrireducens T118	3.5	3.5	2.5	4	5	2.25
Desulfovibrio desulfuricans G20	3	4.5	2.5	3.25	2.75	2.25
Anaeromyxobacter dehalogens	5	9.5	4	5.5	4.5	4.5
Pelobacter carbinolicus DSM2380	3	4	1.5	5.5	2.75	2.75
Geobacter metallireducens GS-15	6	4	4.25	6	7	4.25
Geobacter sulfurreducens PCA	4	6.5	4.25	5.5	4.5	4
Geobacter lovleyi SZ	4	10	5	7	10.5	4.5
Geobacter strain FRC32	9.5	8	10	11.25	10	10.5

Figure 1

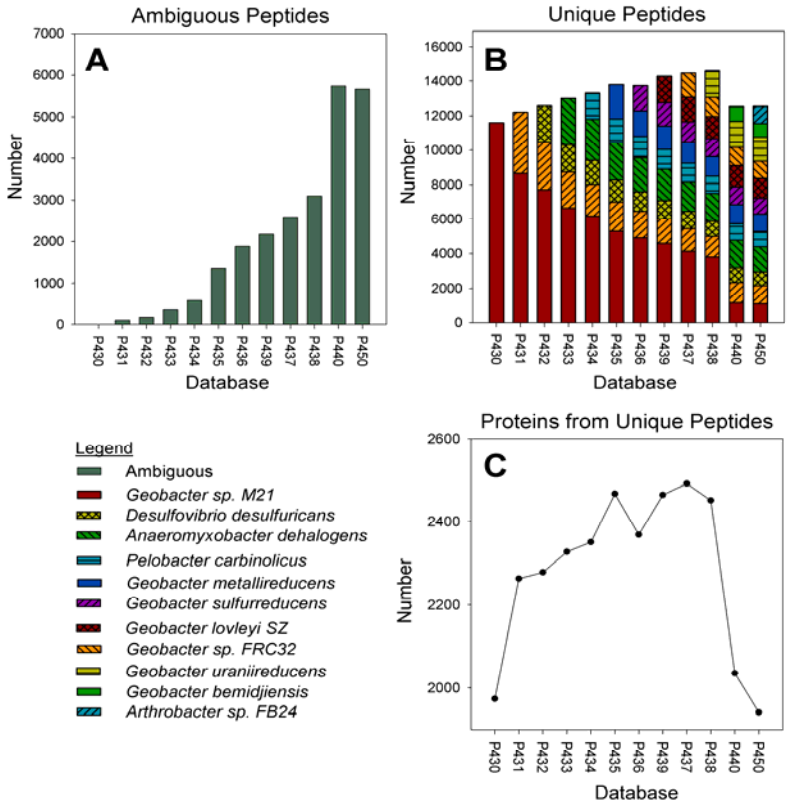


Figure 2.

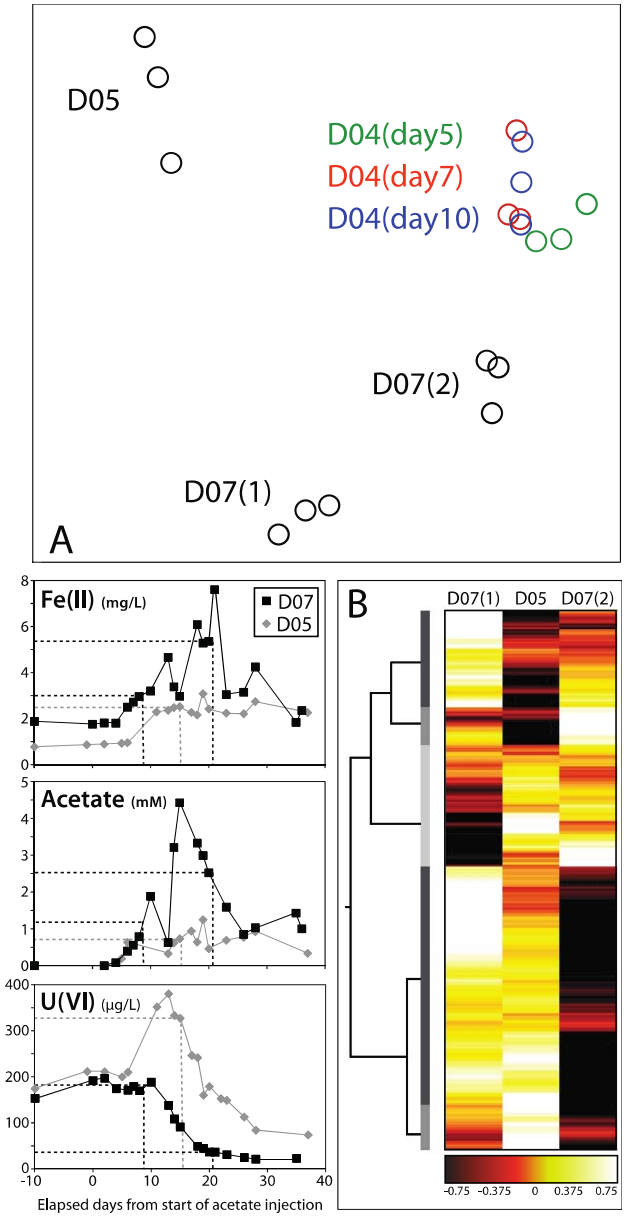
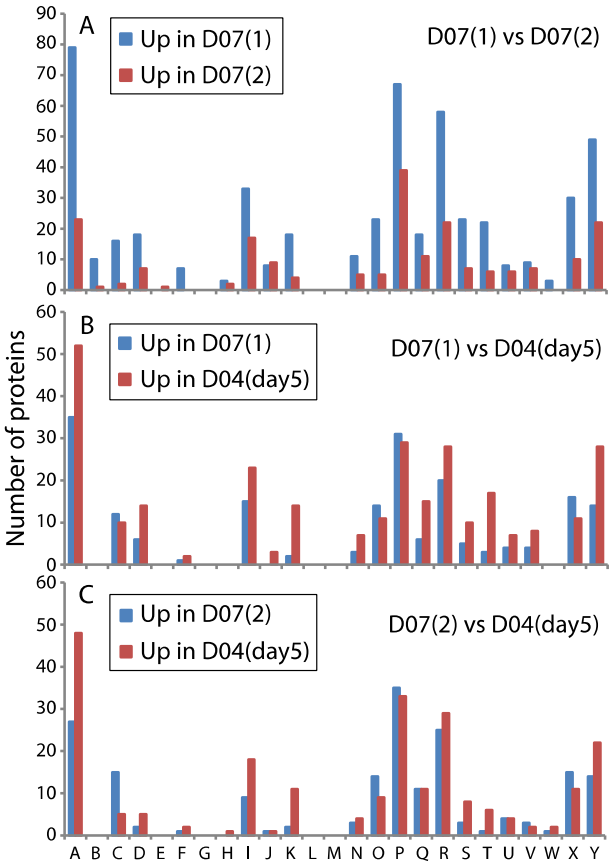

D05
D04(day5)
D04(day7)
D04(day10)
D07(2)
D07(1)
A
Fe(II) (mg/L)
D07
D05
Acetate (mM)
U(VI) (µg/L)
Elapsed days from start of acetate injection
B
D07(1) D05 D07(2)
-0.75 -0.375 0 0.375 0.75

Figure 3.


A
Up in D07(1)
Up in D07(2)
D07(1) vs D07(2)
B
Up in D07(1)
Up in D04(day5)
D07(1) vs D04(day5)
C
Up in D07(2)
Up in D04(day5)
D07(2) vs D04(day5)
Number of proteins
A B C D E F G H I J K L M N O P Q R S T U V W X Y

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.

Ernest Orlando Lawrence Berkeley National Laboratory is an equal opportunity employer.