Development of a biomarker for *Geobacter* activity and strain composition; Proteogenomic analysis of the citrate synthase protein during bioremediation of U(VI).

Michael J. Wilkins[1], Stephen J. Callister[1], Marzia Miletto[2], Kenneth H. Williams[3], Carrie D. Nicora[1], Derek R. Lovley[2], Philip E. Long[4] and Mary S. Lipton[1]

[1]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington 99353; [2]Department of Microbiology, University of Massachusetts, Amherst, Massachusetts 01002; [3]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720; [4]Energy and Environment Directorate, Pacific Northwest National Laboratory, Richland, Washington 99353

February 2010

Keywords*: Geobacter; Proteogenomics; Rifle IFRC; Fe(III) reduction; U(VI) reduction; citrate synthase*

Abstract

Monitoring the activity of target microorganisms during stimulated bioremediation is a key problem for the development of effective remediation strategies. At the U.S. Department of Energy's Integrated Field Research Center (IFRC) site in Rifle, CO, the stimulation of *Geobacter* growth and activity via subsurface acetate addition leads to precipitation of U(VI) from groundwater as U(IV). Citrate synthase (*gltA*) is a key enzyme in *Geobacter* central metabolism that controls flux into the TCA cycle. Here, we utilize shotgun proteomic methods to demonstrate that the measurement of *gltA* peptides can be used to track *Geobacter* activity and strain evolution during *in situ* biostimulation. Abundances of conserved *gltA* peptides tracked Fe(III) reduction and changes in U(VI) concentrations during biostimulation, whereas changing patterns of unique peptide abundances between samples suggested sample-specific strain shifts within the *Geobacter* population. Abundances of unique peptides also indicated potential differences at the strain level between Fe(III)-reducing populations stimulated during *in situ* biostimulation experiments conducted a year apart at the Rifle IFRC. These results offer a novel technique for the rapid screening of large numbers of proteomic samples for *Geobacter* species and will aid monitoring of subsurface bioremediation efforts that rely on metal reduction for desired outcomes.

Introduction

Enzymatic reduction of soluble U(VI) to insoluble U(IV) by indigenous stimulated microorganisms has emerged as a promising remediation strategy in certain subsurface environments. At the Department of Energy (DOE) Integrated Field Research Challenge (IFRC) site in Rifle, CO, acetate addition to the subsurface results in a "bloom" of *Geobacter* microorganisms that is correlated with a decrease in U(VI) concentrations to below the MCL (Anderson *et al*., 2003; Vrionis *et al*., 2005). As biostimulation progresses, the subsurface microbiology shifts from an Fe(III)-reducing community (dominated by *Geobacter* species) to a sulfate reducing community, whose activity results in elevated sulfide production. In some field experiments, this transition is associated with a decrease in the efficiency of U(VI) removal from groundwater. Thus, to ensure the efficacy of this process and better understand how it can be improved across the treatment area, previous research has focused on detecting the activity of *Geobacter* species and the evolution of the subsurface community as biostimulation progresses (Holmes *et al*., 2004; Vrionis *et al*., 2005; Mouser *et al*., 2009; Wilkins *et al*., 2009).

Proteogenomic analysis to date at the Rifle site has studied a "global" view of the *Geobacter* community during the biostimulation campaigns conducted during both 2007 and 2008 (Wilkins *et al*., 2009). While these analyses revealed a significant amount of data about the community physiology and strain make-up during acetate addition, time-intensive computational searching of the data means that this technique is excessively detailed for rapid screening of *Geobacter* species in large numbers of samples. Instead, a biomarker that quickly allows the identification of *Geobacter* activity in a sample is needed. Biomarker studies of these populations at the mRNA level have focused on the potential for

a range of genes to track *Geobacter* activity in the subsurface. The nitrogen fixation gene, *nifD*, has shown promise in this role (Holmes *et al*., 2004), while Holmes et al. (Holmes *et al*., 2005) demonstrated that the citrate synthase gene, *gltA,* could be used to estimate the rates of *Geobacter* metabolism in a number of environments. However, the extraction of mRNA from large numbers of samples is a difficult and time-intensive process, and has associated problems of rapid mRNA degradation. In contrast, shotgun proteomic techniques have the potential to offer a faster method of analysis due to high-throughput pipelines in place at a number of institutions (Lipton *et al*., 2002; Callister *et al*., 2006; VerBerkmoes *et al*., 2008; Wilkins *et al*., 2009).

The citrate synthase protein (subsequently referred to as CS in the text), which is responsible for controlling flux into the TCA cycle by catalyzing the condensation of acetyl-CoA and oxaloacetate to produce citric acid, has a number of characteristics that make it a suitable candidate as a *Geobacter*-specific peptide-based biomarker. The amino acid sequence in members of the *Geobacteraceae* is more closely related to a eukaryotic CS than other prokaryotic sequences (Bond *et al*., 2005), limiting the potential for false positive identifications from other subsurface species. Within the *Geobacteraceae* however, certain regions of the protein are highly conserved and have the potential to act as biomarkers for general *Geobacter* abundance and activity (Fig. 1). While other proteins contain highly conserved regions that could potentially act as *Geobacter* biomarkers, these regions frequently match other closely related species, such as *Pelobacter* and *Desulfuromonas*. In addition, the analysis of more divergent regions of the CS protein sequence shows potential as a technique to track strain-level changes within the *Geobacter* community. Clear shifts in CS unique peptide abundances and diversity are observed over the duration of biostimulation, and therefore may be used to "fingerprint" the microbial community at a specific period during the biostimulation process. Given that these community

"fingerprints" may be characteristic of certain time points during the biostimulation process, they are potentially useful indicators of changes in the biogeochemistry of the system (Wilkins *et al*., 2009). Finally, given that the subsurface at Rifle is amended with acetate to stimulate microbial growth, proteins involved in the efficient utilization of this substrate (e.g. TCA cycle proteins) are abundant within proteomic samples. Citrate synthase peptides are therefore suitable for roles as biomarkers given that they are easily detected where *Geobacter* species are active. Given that the reduction of contaminants such as U(VI) is tightly linked to respiratory processes (Lovley *et al*., 1991; Gorby and Lovley, 1992), the fluctuating abundance of a protein involved in key metabolic processes such as citrate synthase will act as an effective proxy for *Geobacter* activity.

Here, we have applied these principles to field data obtained during *in situ* biostimulation experiments at Rifle during two successive field seasons (2007 and 2008). Results indicate that certain highly abundant conserved peptides can act as indicators of *Geobacter* activity, while abundance patterns for unique peptides matching CS can act as a community "fingerprint" that can then be linked to specific periods of the biostimulation process. The development of a mass-tag search database containing >200 *Geobacter* CS amino acid sequences will allow future proteomic samples to be rapidly screened for *Geobacter* activity and community composition, two factors that have a significant impact on U(VI) removal efficiency.

Materials and Methods

Sample Collection. Samples for groundwater geochemistry and planktonic biomass for proteomics was collected over two consecutive *in situ* biostimulation projects at the Rifle IFRC site located in Rifle, CO. The in situ biostimulation experiments were carried out in the same flow-cell during August and September 2007, and July, August and October 2008. During the 2008 biostimulation experiment, acetate addition was ceased for a 8 day period to allow a "groundwater flush" to take place within the aquifer. For other details on flow-cell size, operating conditions and biomass collection, see Wilkins et al. (Wilkins *et al*., 2009). Analysis focused on the three samples collected from wells D05 and D07 during the 2007 experiment, and nine planktonic biomass samples collected from well D04 during the 2008 experiment. DNA for metagenomic analysis was extracted from biomass collected from well D05 during a period of dominant Fe(III) reduction during the 2007 biostimulation experiment. This sequence was made available courtesy of B. A. Methé at the J. Craig Venter Institute, Rockville, MD. Groundwater samples for DNA analysis of *gltA* clones were collected during the 2008 biostimulation experiment in well D04. Per sample, 10 L of groundwater was concentrated on a Supor®-200 membrane filter (ø=293 mm, pore size = 0.2 μm; Pall Life Sciences, Ann Arbor, MI). Filters were quickly sealed into Mylar bags, flash frozen in an ethanol-dry ice bath, and stored at -80 °C until nucleic acids extraction. These DNA samples for *gltA* gene analysis were collected on the following dates: 07/19/08, 07/29/08, 08/02/08, 08/14/08 and 08/22/08. Geochemical measurements were carried out as previously described (Wilkins *et al*., 2009).

DNA extraction and *gltA* cloning and sequencing. DNA was extracted from a portion of the filter, crushed with liquid nitrogen, using the FastDNA® SPIN Kit for soil (MP Biomedicals, Solon, OH, USA). DNA was quantified using a NanoDrop spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA).

An approximately 803-bp DNA fragment was amplified using the primers CS18F (5'-CTCGCGACATCCGGAGTCT-3') and CS821R (5'-TGTCCGGCGTTCAGGGTAT-3') (Holmes, et al., 2005)

targeting the *Geobacteraceae* citrate synthase-encoding gene (*gltA*), and the following PCR protocol: initial denaturation at 95˚C for 5 min, 30 cycles of 1 min denaturation at 95˚C, 1.5 min annealing at 55˚C-50.5˚C (0.5˚C decrease per cycle during the first 10 cycles), 1.5 min elongation at 72˚C, and final elongation at 72˚C for 10 min. Positive controls, *i.e.* purified *gltA* PCR product from *Geobacter metallireducens*, and negative controls without DNA were always included in PCR amplification experiments. The reaction was carried in a PTC200 Peltier Thermal Cycler (MJ Research, Waltham, MA, USA). The 50 µL reaction mixture contained 100 ng of DNA, 1X Q-Solution (Qiagen), 1X PCR Buffer (Qiagen), 1.5 mM $MgCl_2$ (Qiagen), 200 µM concentrations of each deoxynucleotide (Sigma, Saint Louis, MO, USA), 0.5 µM concentrations of each primer, 0.5X BSA (New England Biolabs, Beverly, MA, USA), and 1.25 U of Taq DNA polymerase (Qiagen). The presence and size of the amplification products were determined by agarose (1% [w/v]) gel electrophoresis. Bands of the expected size were purified from the gel by excision with a sterile surgical blade and purified with the QIAquick Gel Extraction Kit as recommended by the manufacturer (Qiagen). Four µL of the agarose gel-purified DNA mixture were immediately ligated into the pCR2.1®-TOPO® vector (TOPO TA Cloning® Kit, Invitrogen, Carlsbad, CA, USA). The *gltA* sequence was determined for *E. coli* recombinant vector-containing colonies with the primers M13F and M13R, in an ABI 3730xl DNA Analyzer using the Sanger chain-terminator method with fluorescently labeled nucleotides. Chromatograms were visually inspected using the software 4Peaks v1.7 (www.mekentosj.com). Recovered *gltA* sequences were initially compared to GenBank database (Benson *et al*., 2005) for preliminary identification using the programs BLASTN and BLASTX (http://www.ncbi.nml.nih.gov/BLAST).

Proteomics via LC–MS-MS and peptide putative mass and elution time tag assignments. Global as well as soluble and insoluble protein fractions were extracted from cell pellets by using established protocols (Lipton *et al*., 2002; Adkins *et al*., 2006). Briefly, frozen cells were thawed on ice, washed using 100 mM NH4HCO3, pH 8.4, buffer, and then suspended in a new aliquot of this buffer. Cells were lysed via pressure cycling technology using a barocycler (Pressure BioSciences, Inc., South Easton, MA). The suspended cells were subjected to 20 s of high pressure at 35 kilopounds

per square inch, followed by 10 s of ambient pressure for 10 cycles. The protein concentration was determined by a Coomassie assay (Thermo Scientific, Rockford, IL). Protein extraction, digestion, and high-performance LC fractionation were performed as previously described (Callister *et al*., 2006). From each collected fraction, peptides were analyzed by reversed phase high-performance LC separation coupled with the use of an LTQ ion trap mass spectrometer (ThermoFisher Scientific Corp., San Jose, CA) operated in a data-dependent MS-MS mode. Spectra generated by LC-MS/MS were analyzed using the SEQUEST algorithm in conjunction with predicted protein annotations concatenated from CS sequences described below. CS sequences were obtained from seven isolate *Geobacter* genomes, metagenomic sequence, and translated *gltA* clones amplified from DNA from well D04 during the 2008 biostimulation experiment. SEQUEST results were preliminarily filtered (Xcorr values of ≥1.9, ≥2.2, or ≥3.5 for 1+, 2+, or ≥3+ if seen once; Xcorr values of ≥1.9 if seen two or more times; no cleavage rules; minimum length of 6 residues), extracted, and processed using the PRISM proteomics pipeline developed in-house (Kiebel *et al*., 2006).

A Mass-Tag database was constructed from SEQUEST search results that allowed the rapid screening of proteomic data for CS peptide matches (Smith *et al*., 2002). Label-free arbitrary abundance measurements of peptides from sampling events D07(1), D05, and D07(2) were obtained from LC-MS measurements. Triplicate (9 total analyses) measurements were made on a custom-built HPLC system coupled via ESI to a 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 mass and elution time features prior to matching to the reference peptide database.

Data handling & Phylogenetic analysis. Peptide abundances were calculated by integrating the signal strength under each peak of the LC-MS spectra (Smith *et al*., 2004). Peptide abundances were subsequently log transformed and normalized to a common baseline using central tendency normalization (Callister *et al*., 2006). Manipulation of these data, including the generation of heat maps, was carried out using DANTE (Polpitiya *et al*., 2008), a software tool publically available at http://ncrr.pnl.gov/. Significance of changes in abundances between groups of peptides was

assessed using paired t-tests. Where necessary, significance cutoffs (0.05) are displayed in the text. Citrate synthase amino acid sequences were aligned using the ClustalW algorithm at the EBI website (http://www.ebi.ac.uk/Tools/clustalw2/index.html). Alignment results were imported into the MEGA software, which was used to construct Neighbor-Joining phylogenetic trees (Tamura *et al*., 2007). Non-metric multidimensional scaling, principal components analysis, and environmental vector plotting was carried out using the vegan package (Oksanen *et al*., 2009) in the statistical software R.

Results and Discussion

For this study, a proteomic search database consisting of CS sequences from metagenomic sequence, isolate *Geobacter* genomes and over 200 cloned *Geobacter* copies of the protein was generated. Within such a proteomic search database, unique peptides are present in only one species and only one protein. In contrast, non-unique peptides are present two times or greater, either in different proteins within the same organism, or within copies of a protein in multiple organisms. Alignment of the CS sequences from this search database and other bacterial genomes, together with *in silico* tryptic digests, identified multiple peptides that were highly conserved within *Geobacter* species but absent from other prokaryotic and eukaryotic *gltA* sequences (Fig. 1, A and B). These sequences were the candidates for general *Geobacter* biomarker peptides. Analysis of the "global" proteomic datasets from the 2007 and 2008 biostimulation campaigns revealed that a subset of these conserved peptides were the most abundantly detected where *Geobacter* species were present. Peptides with sequences TIPETFEALPK, SLVTDISYLDPQEGIR and QVVPEYVYTAVR were selected on this basis as the conserved "*Geobacter*" peptide markers. BlastP analysis of these sequences against the NCBI database revealed that the only significant matches to these sequences were peptides from *Geobacter* species. From a total

of 232 citrate synthase sequences in the mass tag database, the "TIPETFEALPK" peptide was present 171 times, the "SLVTDISYLDPQEGIR" peptide was present 166 times, and the "QVVPEYVYTAVR" peptide was present 124 times.

In silico tryptic digests also revealed a large number of unique peptides present within the mass-tag database constructed from CS sequences. Because these peptides are unique to one sequence in the database they can be used to constrain the genotype of the strains present in the subsurface at any time point. These peptides come from more divergent regions of the CS protein (Fig. 1).

Conserved biomarkers

Three proteogenomic datasets were collected during a 13-day period of the 2007 biostimulation field campaign. As explained in Wilkins et al. (Wilkins *et al*., 2009), all were taken during the period of dominant Fe(III) reduction in the subsurface. The heat map in figure 2 illustrates the presence of *Geobacter* species in all samples, as identified by high abundances of conserved CS peptides (TIPETFEALPK, SLVTDISYLDPQEGIR and QVVPEYVYTAVR). Evidence for enzymatic Fe(III) reduction is present in all three samples in the form of aqueous Fe(II), while a rapid decrease in U(VI) concentrations is observed in the later D07 sample (Fig. 2A). Acetate amendment was stopped in this biostimulation experiment before the Fe(III)-reducing system transitioned into sulfate-reducing conditions.

During the 2008 biostimulation experiment, acetate addition to the subsurface stimulated a period of dominant Fe(III) reduction, followed by a transition period where the subsurface geochemistry shifted towards dominant sulfate-reducing conditions. Nine planktonic biomass samples were collected from downgradient well D04 during the experiment and therefore represent the microbial community changes that accompany the

geochemical shifts. Previous research has shown that a microbial community shift from Fe(III)-reducing microorganisms to sulfate-reducing bacteria occurs over this transition period (Vrionis *et al*., 2005). Additionally, U(VI) removal from groundwater is typically most effective during this period of dominant Fe(III) reduction, with rates decreasing during the transition period (Anderson *et al*., 2003). However, geochemical measurements alone cannot accurately predict key process changes, due to complex cycling processes that occur between reduced Fe and S species. During periods of dominant Fe(III) reduction, the precipitation of FeS via the reaction of biogenic sulfide with Fe(II) can prohibit the detection of low-level sulfate reduction. Conversely, during periods of dominant sulfate reduction in the aquifer, low-level Fe(III) reduction is difficult to detect due to the titration of Fe(II) from solution by S^{2-}. It is therefore of key importance to be able to quickly screen microbial biomass for organisms of interest, which in this case are the dominant *Geobacter* communities that develop following acetate addition.

Conserved peptide abundances indicate planktonic *Geobacter* activity during the first half of the biostimulation experiment, when the system is apparently dominated by Fe(III) reduction and U(VI) concentrations are lowest (samples “Arthur” through “Omar”)(Fig 3A). Analysis of the sample “Rene”, taken shortly after the cessation of acetate addition (23 days from start date), suggests that planktonic *Geobacter* populations are starting to decrease, and evidence of the biomarker peptides has mainly disappeared in the following sample, “Teddy” (taken 27 days from start of experiment). As the system geochemistry begins to transition into sulfate reducing conditions (identified by rising S^{2-} concentrations and decreasing Fe(II) concentrations), the remaining samples (“Boris” and “Elida”) show no evidence for the presence of planktonic *Geobacter* activity (Fig. 3A). During the period where *Geobacter* abundances decrease (as inferred by conserved peptide abundances), U(VI) concentrations rebound to 3 mg/L. This is likely due to both the temporary cessation

of acetate addition and the beginning of dominance by sulfate reduction towards the end of the experiment (Fig. 3B).

Although peptide biomarker abundances track Fe(II) production during the initial stages of biostimulation during the 2008 experiment, later samples show a disconnect between elevated Fe(II) concentrations and low (or absent) biomarker abundances (e.g. "Teddy"). As biostimulation progresses towards the transition period, these data suggest that rapid *Geobacter* growth that characterizes some earlier time points during the most efficient U(VI) removal, has ceased. Less active cell growth results in fewer planktonic cells for biomass sampling, and we can infer that any low-level Fe(III) reduction that occurs during the remainder of biostimulation is likely carried out by mineral-attached. There are multiple possible factors responsible for the slowdown of *Geobacter* cell growth such as the decreasing concentrations of certain "bioavailable" Fe(III) oxides or other nutrient limitations. The clustering of 2008 samples via non-metric multi-dimensional scaling (NMDS) using both conserved "biomarker" and unique CS peptide abundances, and the projection of environmental variables onto these axes suggest that acetate concentrations best explain the shifts in *Geobacter* population numbers (Fig 5A). This further confirms the importance of ensuring that the key microbial communities during biostimulation receive excess carbon concentrations wherever possible. This plot additionally shows a negative correlation between U(VI) and the samples containing abundant *Geobacter*, and the visual correlation between low U(VI) concentrations and high biomarker abundance is clear in figure 3A. Thus, these biomarker abundances are useful as indicators of efficient U(VI) removal from groundwater. During the 2008 biostimulation experiment the highest abundances are correlated with the lowest U(VI) concentrations (≤1 mg/L) (Fig. 3A), and where peptide abundances begin to decrease (~ day 15, sample "Rene" onwards), U(VI) concentrations start to rebound within the sampling well. Within the 2007 data, this pattern

is repeated with the D07[2] sample, while the biomass for the D07[1] sample was sampled just prior to a rapid decrease in U(VI) concentrations (Fig. 2A).

Unique peptide abundance patterns

While conserved peptides can be used to assess *Geobacter* abundance within a community, fine-scale strain-level shifts between samples can be assessed using unique peptide abundances. Global analysis of unique peptide abundances in 2007 datasets indicated strain shifts and increasing community diversity within the *Geobacter* population over the duration of the biostimulation period. While this was characterized by an increase in peptides matching *G. lovleyi* and a decrease in peptides matching *G. bemidjiensis* and strain M21 as biostimulation progressed, total peptide data revealed that the community was still dominated by strains most closely related to *G. bemidjiensis* and strain M21, but with increased similarity to *G. lovleyi* at certain loci (Wilkins *et al*., 2009). While the abundance of unique peptides matching CS in these global datasets suggested that this protein would be a good candidate for this study, the addition of cloned *gltA* sequences to the search database for this study increased the strain-level resolution of the analysis (Fig. 4A and 4B).

Changes in the presence and abundance of unique CS peptides over the course of the 2007 experiment are clearly visible between the two D07 samples (Fig. 2B). Phylogenetic analysis of CS sequences matching these unique peptides reveals that the majority fall within clusters “1, 2 and 3” in figure 4A, and are most closely related to the *G. bemidjiensis* and strain M21 copies of CS. One exception is a cloned sequence (cluster “4”; C0814_40) that is more closely related to *G. uraniireducens* and *G. lovleyi*. Those sequences that match peptides unique to sample D07[1] are generally associated with the *G. bemidjiensis*/strain M21 region of the tree (red labels, grey shaded region, Fig. 4A), whereas increases in the

diversity of sequences matching CS peptides unique to D07[2] are apparent (blue labels, Fig. 4A). In this later sample, unique peptides matching five isolate copies of CS and five cloned copies of CS are detected, more than double the number of sequences matching unique peptides in D07[1]. The sample specific presence-or-absence nature of these peptides is illustrated in figure 5B, where peptides unique to one sample constrain the clustering distances between samples.

Shifts in the abundance of these peptides are also apparent between the 2007 D07 samples, as illustrated in figure 4B. Abundance increases in unique peptides matching *G. lovleyi, G. uraniireducens* and clone C0814_40 ("cluster 4") are coupled to decreases in *G. bemidjiensis*/metagenomic sequence peptides (Fig. 4B) between D07[1] and D07[2] (both shifts $p<0.05$). Less significant decreases are also observed within clusters "1, 2 and 3" between the same time points. These patterns amongst CS unique peptides are indicative of the *Geobacter* community shifting from a population dominated by a few strains closely-related to *G. bemidjiensis* and strain M21 during the early period of Fe(III) reduction, to increasing strain diversity later in Fe(III) reduction. Mapping these unique peptides on CS protein alignments, and identifying regions of multiple coexisting peptides previously demonstrated this effect (Wilkins *et al*., 2009). However, given that the majority of cloned sequences are located within clusters "1, 2 and 3" (Fig 4A), we can infer that despite fine-scale strain-level shifts occurring within the *Geobacter* community during this experiment, the dominant strains still have greatest similarity to *G. bemidjiensis* and strain M21.

During the 2008 biostimulation experiment, approximately five samples (Arthur thru Omar) were recovered during the period of dominant Fe(III) reduction. Phylogenetic analysis of sequences matching unique peptides detected in these samples reveals differences relative to both D07 samples from 2007. Whereas sequences from "cluster 1" are absent in 2008 samples, there are clear increases in the number of sequences from

"cluster 2" relative to the 2007 data (Fig. 4C). Unique peptide abundance patterns confirm these differences (Fig. 4D); abundances of unique peptides matching *G. bemidjiensis*/metagenomic sequence suggest the presence of strains that were dominant at the beginning of the 2007 experiment (sample D07[1]) when acetate was added to the pristine aquifer. However, the abundance of *G. lovleyi* unique peptides indicates some similarity to strains that were present towards the end of biostimulation in 2007 (sample D07[2]). In addition, relatively similar abundance patterns between these five samples (Fig. 4D) may suggest the presence of a more stable community of Fe(III)-reducing bacteria than was present during the 2007 biostimulation experiment. Indeed, the NMDS plot in figure 5A reveals the tight clustering of samples "Arthur" thru "Hanna", indicating significant similarity between these samples.

These apparent differences between *Geobacter* populations recovered from the subsurface over two years may potentially be linked to the use of the same flow cell for both experiments. A range of analyses have demonstrated that following acetate addition to a pristine aquifer system, a bloom of *Geobacter* growth is stimulated (Vrionis *et al*., 2005, Mouser *et al*., 2009, Wilkins *et al*., 2009). This initial enrichment of *Geobacter* may be dominated by only a few strains that couple the highest growth rates to the most efficient utilization of acetate and Fe(III) oxides. As the duration of biostimulation progresses however, strain diversity within the Fe(III)-reducing microbial community has been seen to increase (Wilkins *et al*., 2009) and this is reflected in the increased diversity of unique CS peptides observed in sample D07[2] (Fig. 4 A and B). Whether this is due to the emergence of slower growing bacteria, the initial effects of sulfate reducers, or changes in the availability of different Fe(III) oxides as terminal electron acceptors, is currently unknown. The 2007 biostimulation experiment was ceased at this point, towards the end of Fe(III) reduction but prior to the transition into sulfate reduction. Potentially, carbon

concentrations arising from residual acetate and breakdown products from the initial biomass "bloom" may have allowed a relatively stable and diverse community of Fe(III)-reducing microorganisms to persist at low levels in the subsurface between experiments. Initial Fe(II) concentrations in well D04 prior to acetate amendment in 2007 (1.52 mg L^{-1}) are significantly lower than those seen the following year in 2008 (3.13 mg L^{-1}), suggesting Fe(III)-reducing activity during the period in-between biostimulation experiments. Following the subsequent addition of acetate at the start of the 2008 campaign, some members of this pre-existing community may be enriched together with the fast-growing *G. bemidjiensis* and M21-like strains, resulting in a community structure that is different from that seen at the start of acetate enrichment in a pristine aquifer.

This demonstrated ability to detect strain-level differences between *Geobacter*-dominated communities using just CS unique peptides will enable the analysis of biostimulated *Geobacter* communities during future experiments. Comparisons between new data and patterns observed during the 2007 and 2008 experiments may allow us to infer biogeochemical processes occurring in the subsurface, providing a useful tool to complement geochemical measurements. This rapid screening will be carried out in conjunction with the analysis of conserved CS peptides, which we have demonstrated may be used as indicators of planktonic *Geobacter* abundance and activity during biostimulation. Given the strong correlation between high "biomarker" abundance and low U(VI) concentrations, these abundances will also be used in conjunction with geochemical measurements for improved prediction of U(VI) removal rates in the subsurface. The availability of a dedicated *Geobacter* CS search database will further ensure that these analyses can be carried out in an efficient manner.

Acknowledgements

We thank the city of Rifle, CO, the Colorado Department of Public Health and Environment, and the U.S. Environmental Protection Agency, Region 8, for their cooperation in this study. Pacific Northwest National Laboratory is managed under contract DE-AC05-76RL01830 with Battelle Memorial Institute. Lawrence Berkeley National Laboratory is operated for the U.S. Department of Energy by the University of California under contract DE-AC02-05CH11231. Portions of this work were performed at the Environmental Molecular Sciences Laboratory, a DOE national scientific user facility located at the Pacific Northwest National Laboratory. This research was sponsored by the Environmental and Remediation Sciences Program, Biological and Environmental Research, Office of Science, U.S. Department of Energy.

References

Adkins, J.N., Mottaz, H.M., Norbeck, A.D., Gustin, J.K., Rue, J., Clauss, T.R.W., *et al*. (2006) Analysis of the *Salmonella typhimurium* proteome through environmental response toward infectious conditions. *Mol Cell Proteomics* **5**: 1450-1461.

Anderson, R.T., Vrionis, H.A., Ortiz-Bernad, I., Resch, C.T., Long, P.E., Dayvault, R., *et al*. (2003) Stimulating the in situ activity of *Geobacter* species to remove uranium from the groundwater of a uranium-contaminated aquifer. *Appl Environ Microbiol* **69**: 5884-5891.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Wheeler, D.L. (2005) GenBank, *Nucl. Acids Res.* **33**: D34-38.

Bond, D.R., Mester, T., Nesbo, C.L., Izquierdo-Lopez, A.V., Collart, F.L., and Lovley, D.R. (2005) Characterization of citrate synthase from *Geobacter sulfurreducens* and evidence for a family of citrate synthases similar to those of eukaryotes throughout the *Geobacteraceae*. *Appl Environ Microbiol* **71**: 3858-3865.

Callister, S.J., Dominguez, M.A., Nicora, C.D., Zeng, X., Tavano, C.L., Kaplan, S., *et al*. (2006) Application of the accurate mass and time tag approach to the proteome analysis of sub-cellular fractions obtained from *Rhodobacter sphaeroides* 2.4.1. Aerobic and photosynthetic cell cultures. *J Proteome Res* **5**: 1940-1947.

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

Gorby, Y.A., and Lovley, D.R. (1992) Enzymatic uranium precipitation. *Environ Sci Technol* **26**: 205-207.

Holmes, D.E., Nevin, K.P., and Lovley, D.R. (2004) In situ expression of *nifD* in *Geobacteraceae* in subsurface sediments. *Appl Environ Microbiol* **70**: 7251-7259.

Holmes, D.E., Nevin, K.P., O'Neil, R.A., Ward, J.E., Adams, L.A., Woodard, T.L., *et al*. (2005) Potential for quantifying expression of the *Geobacteraceae* citrate synthase gene to assess the activity of *Geobacteraceae* in the subsurface and on current-harvesting electrodes. *Appl Environ Microbiol* **71**: 6870-6877.

Kiebel, G.R., Auberry, K.J., Jaitly, N., Clark, D.A., Monroe, M.E., Peterson, E.S., *et al*. (2006) PRISM: A data management system for high-throughput proteomics. *Proteomics* **6**: 1783-1790.

Lipton, M.S., Pasa-Tolic, L., Anderson, G.A., Anderson, D.J., Auberry, D.L., Battista, K.R., *et al*. (2002) Global analysis of the *Deinococcus radiodurans* proteome by using accurate mass tags. *Proc Natl Acad Sci USA* **99**: 11049-11054.

Lovley, D.R., Phillips, E.J.P., Gorby, Y.A., and Landa, E. (1991) Microbial reduction of uranium. *Nature* **350**: 413-416.

Mouser, P.J., N'Guessan, A.L., Elifantz, H., Holmes, D.E., Williams, K.H., Wilkins, M.J., *et al*. (2009) 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* **43**: 4386-4392.

Oksanen, J., Roeland, K., Legendre, P., O"Hara, B., Simpson, G.L., Solymos, P., *et al*. (2009) vegan: Community Ecology Package. R package version 1.15-4

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

Smith, R.D., Anderson, G.A., Lipton, M.S., Pasa-Tolic, L., Shen, Y.F., Conrads, T.P., *et al*. (2002) An accurate mass tag strategy for quantitative and high-throughput proteome measurements. *Proteomics* **2**: 513-523.

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

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596-1599.

VerBerkmoes, N.C., Russell, A.L., Shah, M., Godzik, A., Rosenquist, M., Halfvarson, J., *et al*. (2008) Shotgun metaproteomics of the human distal gut microbiota. *ISME J* **3**: 179-189

Vrionis, H.A., Anderson, R.T., Ortiz-Bernad, I., O'Neill, K.R., Resch, C.T., Peacock, A.D., *et al*. (2005) Microbiological and geochemical heterogeneity in an in situ uranium bioremediation field site. *Appl Environ Microbiol* **71**: 6308-6318.

Wilkins, M.J., VerBerkmoes, N.C., Williams, K.H., Callister, S.J., Mouser, P.J., Elifantz, H., *et al*. (2009) Proteogenomic monitoring of *Geobacter* physiology during stimulated uranium bioremediation. *Appl Environ Microbiol* **75**: 6591-6599.

Figure Legends.

Figure 1. Alignment of eukaryotic and prokaryotic citrate synthase proteins illustrating certain conserved and more divergent regions associated with *Geobacter* copies of this protein. Red highlighted text in regions **A** and **B** correspond to two of the conserved peptides used as indicators of overall *Geobacter* activity. Blue highlighted amino acids in area **C** illustrate more divergent regions where differences occur among *Geobacter* sequences (e.g. 1 or 2 base pair differences).

Figure 2. Geochemical profiles for U(VI) and Fe(II) in well D07 during the 2007 biostimulation experiment. Heat-map abundances are shown for both the **(A)** conserved "biomarker" peptides and **(B)** unique peptides matching CS proteins in all three samples.

Figure 3. Geochemical profiles for U(VI), Fe(II) and S^{2-} in well D04 during the 2008 biostimulation experiment. Heat-map abundances are shown for both the **(A)** conserved "biomarker" peptides and **(B)** unique peptides matching CS proteins in all nine samples. Light blue zones show time periods of acetate injection. Uranium rebounds starts at ~day 20

Figure 4. (A) Neighbor-joining tree constructed from aligned citrate synthase sequences for which unique peptides were detected in the 2007 data. Isolate sequences are included for reference. Peptides were detected for those isolate sequences marked by a "*". Red and blue labels indicate sequences for which unique peptides were detected in only one of the samples (Red = D07[1], Blue = D07[2]). Isolate CS identifications are as follows; *G. bemidjiensis*(1), gi145617433; *G. bemidjiensis*(2), gi145620657; Strain M21(1), gm829064;

Strain M21(2), gm826531; Strain FRC32(1), gi110599265; Strain FRC32(2), gi110600278; *G. lovleyi*, gi118746732; *G. sulfurreducens*, gi39996208; *G. uraniireducens*, gi148263639; *G. metallireducens*(1), gi78222340; *G. metallireducens*(2), gi78223885. **(B)** Peptide % abundances for all unique peptides matched against the citrate synthase database in samples D07[1] and D07[2] from the 2007 "Winchester" experiment. **(C)** Neighbor-joining tree constructed from aligned citrate synthase sequences detected in the 2008 data. Isolate sequences are included for reference. Peptides were detected for those isolate sequences marked by a "*". Geobacter CS identifications are as above. **(D)** Peptide % abundances for all unique peptides matched against the citrate synthase database in samples Arthur thru Omar from the 2008 "Big Rusty" experiment.

Figure 5. (A) Non-metric multidimensional scaling diagram showing clustering of 2008 samples using both conserved "biomarker" and unique peptide abundances. Environmental vectors have been added illustrating the correlations between the clustering patterns of the biomass samples and geochemical gradients. **(B)** Principal components analysis bi-plot using 2007 unique peptides, illustrating the relative clustering of three biomass samples recovered during the 2007 biostimulation experiment. Four technical replicates are shown per sample. The points are shown as arrows, illustrating the sample-specific nature of the unique peptide abundances. Peptides only detected in sample D07[1] are defined by red arrows, those only detected in D05 by green arrows, and those only detected in D07[2] by blue arrows.

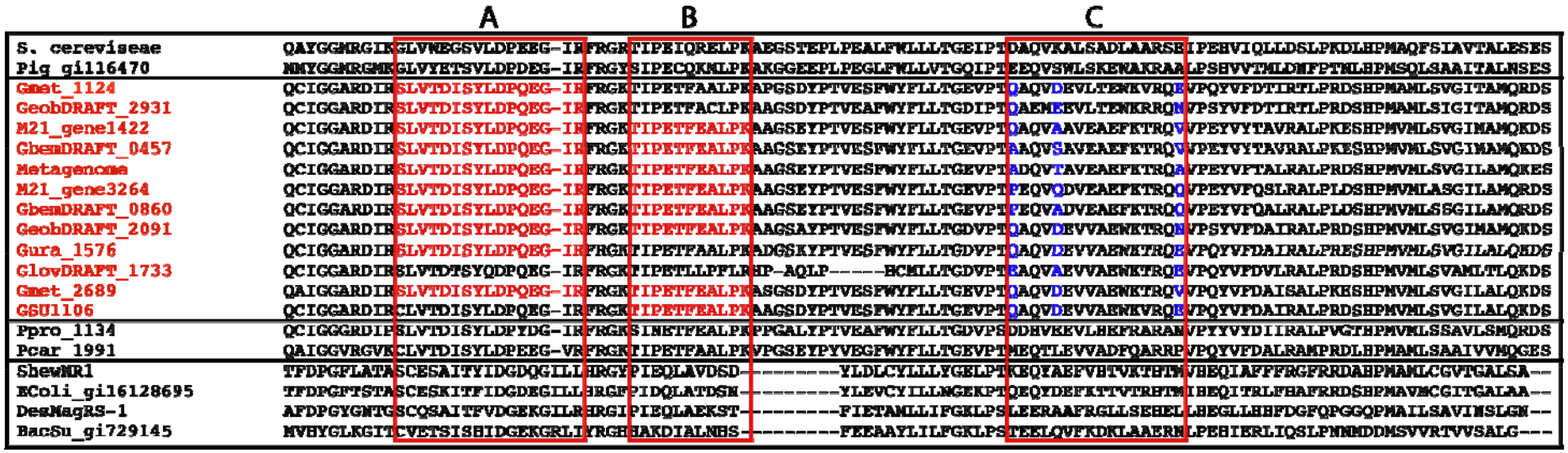
A
B
C
S. cerevisiae
Pig gi116470
Gmet_1124
GeobDRAFT_2931
M21_gene1422
GbemDRAFT_0457
Metagenome
M21_gene3264
GbemDRAFT_0860
GeobDRAFT_2091
Gura_1576
GlovDRAFT_1733
Gmet_2689
GSU1106
Ppro_1134
Pcar_1991
ShewMR1
EColi_gi16128695
DesMagRS-1
BacSu_gi729145

Figure 1.

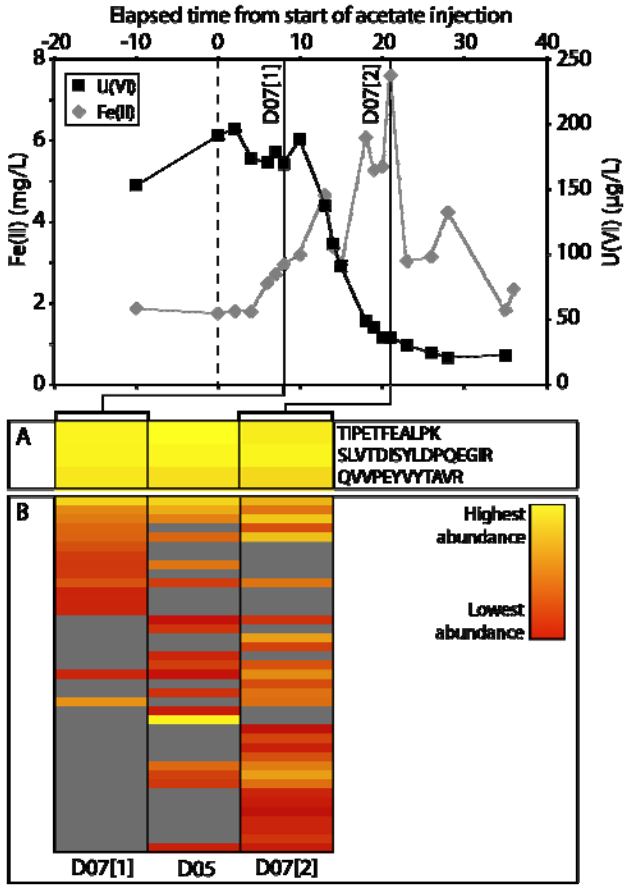
Elapsed time from start of acetate injection
-20
-10
0
10
20
30
40
U(VI)
Fe(II)
D07[1]
D07[2]
Fe(II) (mg/L)
U(VI) (µg/L)
0
2
4
6
8
50
100
150
200
250
A
TIPETFEALPK
SLVTDISYLDPQEGIR
QVVPEYVYTAVR
B
Highest abundance
Lowest abundance
D07[1]
D05
D07[2]

Figure 2.

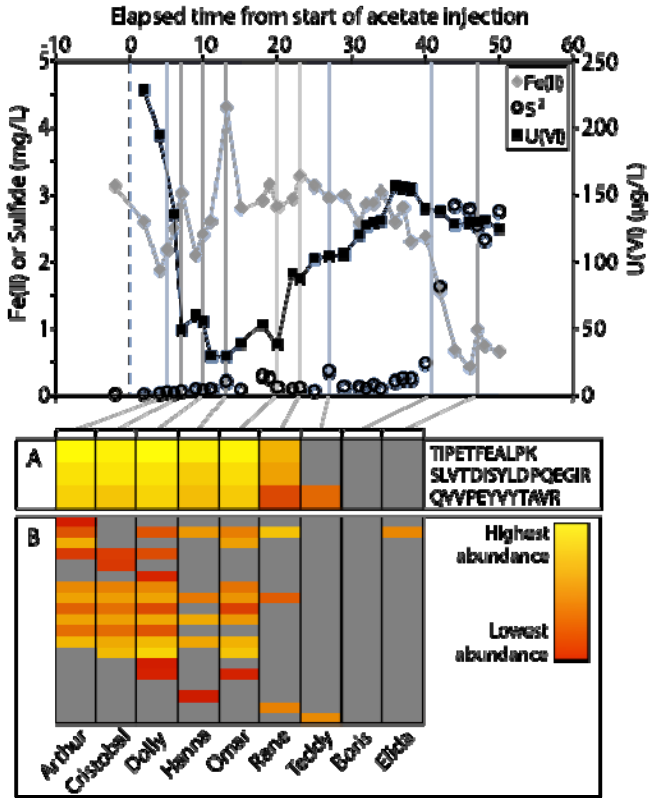
Elapsed time from start of acetate injection
-10
0
10
20
30
40
50
60
Fe(II) or Sulfide (mg/L)
U(VI) (µg/L)
Fe(II)
S2-
U(VI)
A
TIPETFEALPK
SLVTDISYLDPQEGIR
QYVPEYYYTAVR
B
Highest abundance
Lowest abundance
Arthur
Cristobal
Dolly
Hanna
Omar
Rene
Teddy
Boris
Elida

Figure 3.

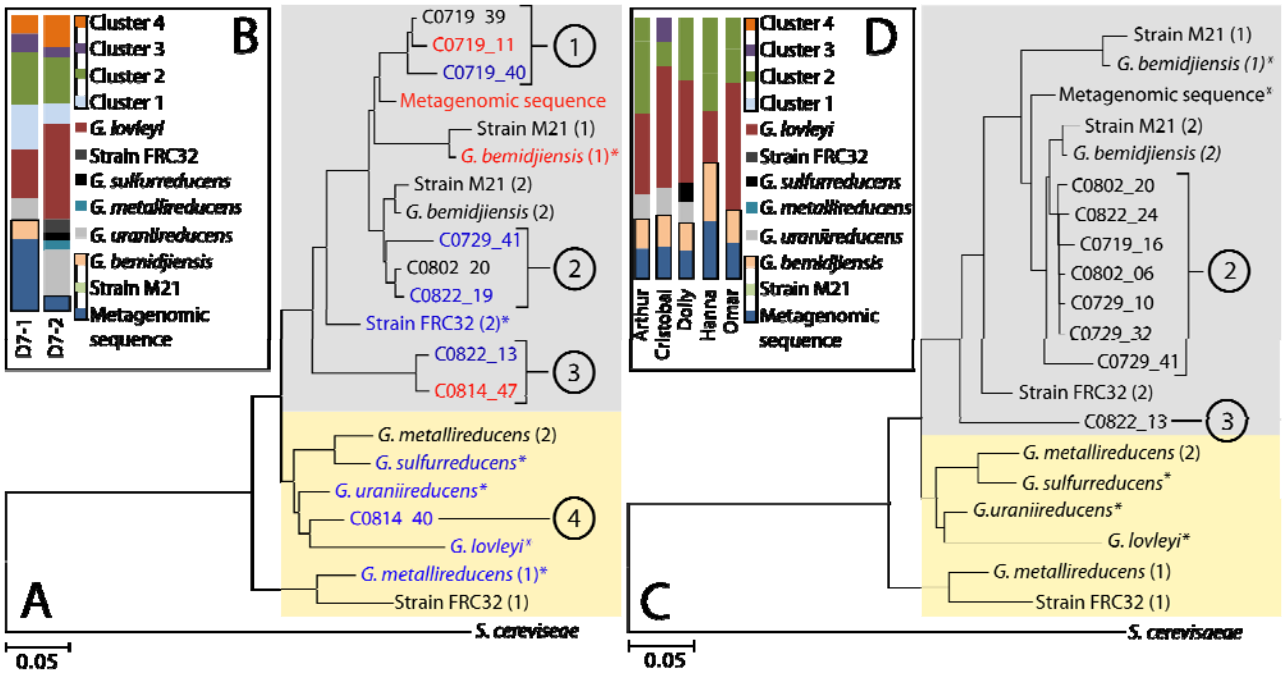
B
Cluster 4
Cluster 3
Cluster 2
Cluster 1
G. lovleyi
Strain FRC32
G. sulfurreducens
G. metallireducens
G. uraniireducens
G. bemidjiensis
Strain M21
Metagenomic sequence
D7-1
D7-2
C0719_39
C0719_11
C0719_40
Metagenomic sequence
Strain M21 (1)
G. bemidjiensis (1)*
Strain M21 (2)
G. bemidjiensis (2)
C0729_41
C0802_20
C0822_19
Strain FRC32 (2)*
C0822_13
C0814_47
G. metallireducens (2)
G. sulfurreducens*
G. uraniireducens*
C0814_40
G. lovleyi*
G. metallireducens (1)*
Strain FRC32 (1)
A
S. cerevisiae
0.05
D
Arthur
Cristobal
Dolly
Hanna
Omar
Metagenomic sequence
Strain M21 (1)
G. bemidjiensis (1)*
Metagenomic sequence*
Strain M21 (2)
G. bemidjiensis (2)
C0802_20
C0822_24
C0719_16
C0802_06
C0729_10
C0729_32
C0729_41
Strain FRC32 (2)
C0822_13
G. metallireducens (2)
G. sulfurreducens*
G.uraniireducens*
G. lovleyi*
G. metallireducens (1)
Strain FRC32 (1)
C
S. cerevisiae
0.05

Figure 4.

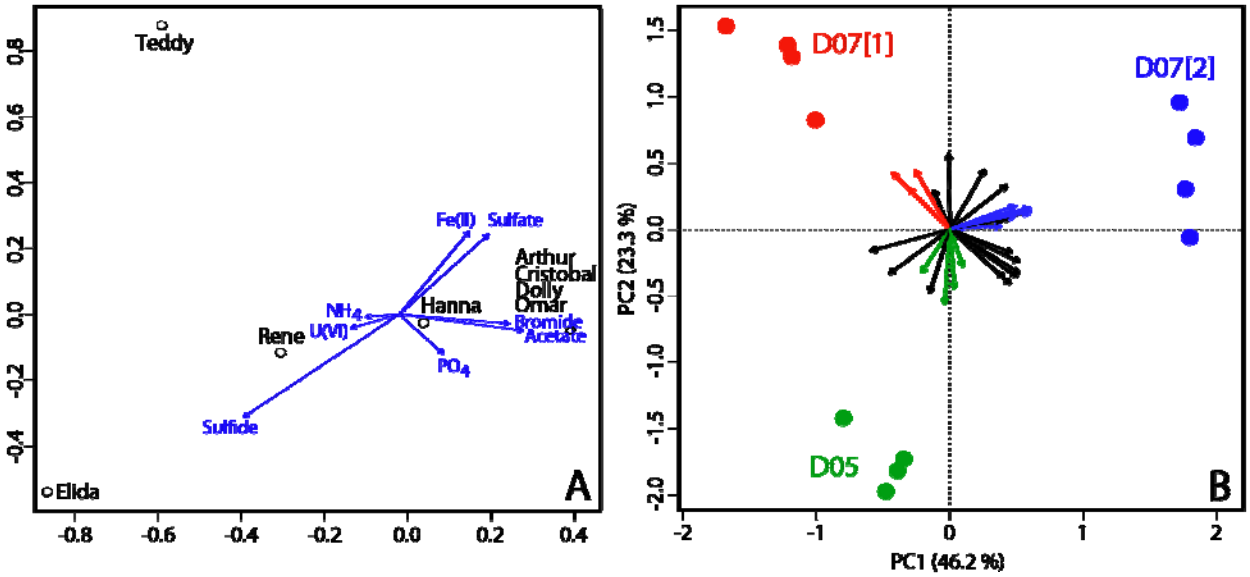
Teddy
Elida
Rene
Hanna
Arthur
Cristobal
Dolly
Omar
Fe(II)
Sulfate
NH4
U(VI)
Bromide
Acetate
PO4
Sulfide
A
D07[1]
D07[2]
D05
B
PC1 (46.2 %)
PC2 (23.3 %)

Figure 5.