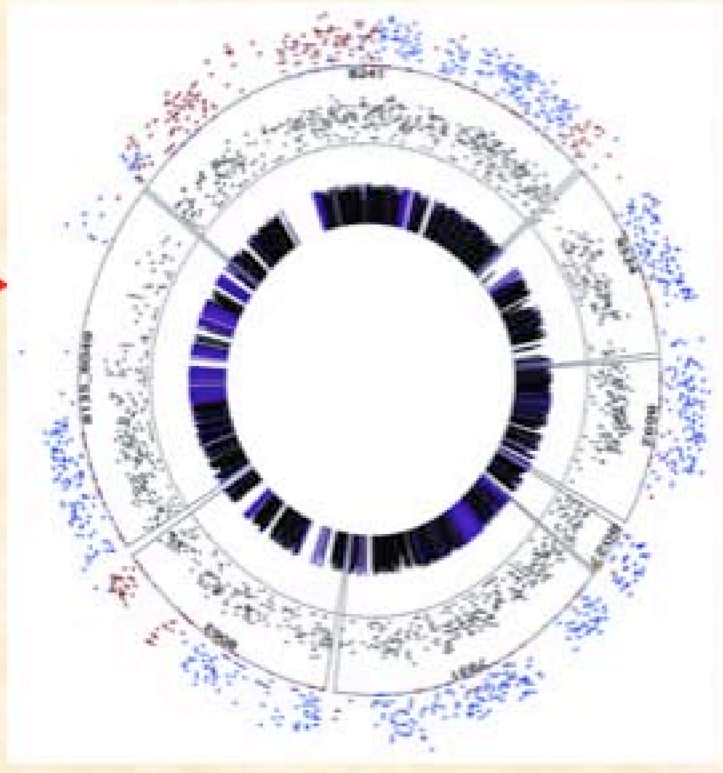
Novel experimental approaches for deep proteome characterization of soil microbial systems

Nathan VerBerkmoes

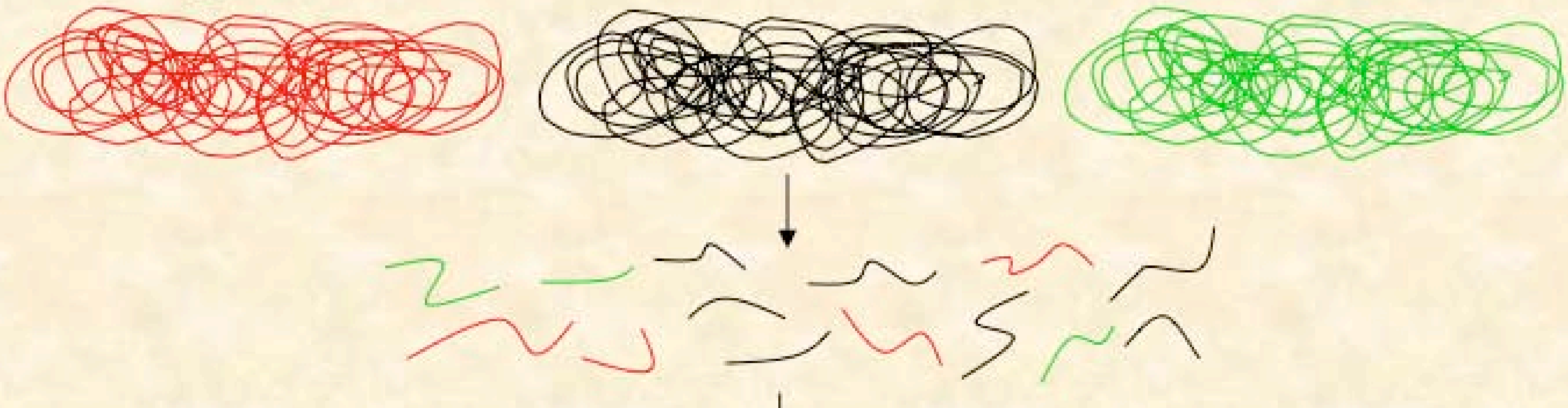
Organic and Biological Mass Spectrometry Group
Oak Ridge National Laboratory



Overview

- MS-based proteomics for natural microbial communities (metaproteomics).
- New methods for soil proteomics
- Integrating groundwater, filter and soil samples in the Rifle field site
- New methods for MS-based proteomics

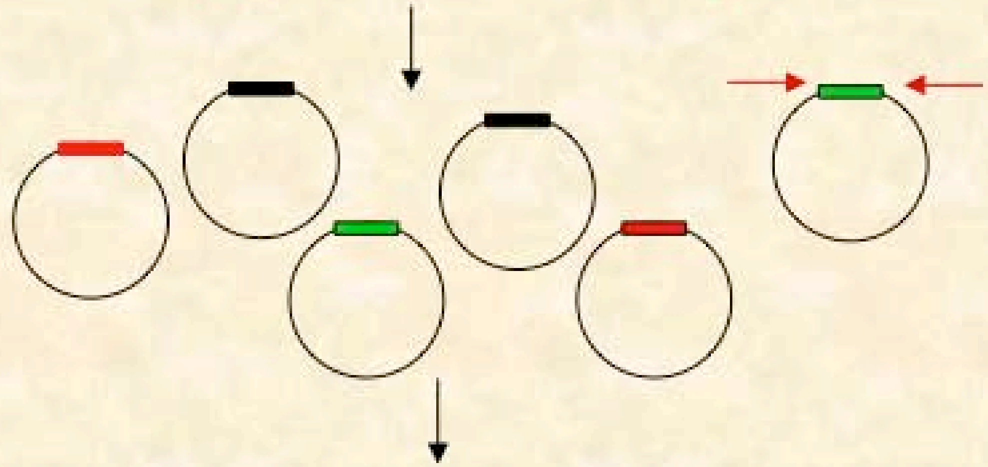
Environmental Microbial Community Genomics: Sanger, 454 or Illumina??



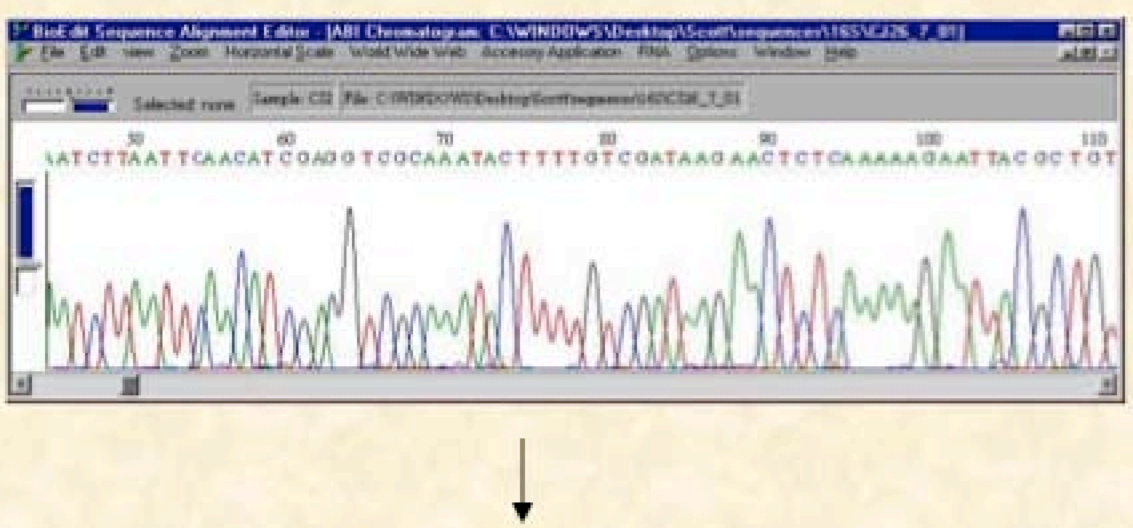
Extract DNA from natural sample

Shear

Tyson et al. Nature 2004



3 – 4 kb shotgun library



End sequence clones

Assemble reads by alignment identity

…ACGGCTGCGTTACATCGATCAT
ACATCGATCATTTACGATACCATTG…

ORNL Proteome Pipeline for Microbial Species

Community or Environmental Sample

↓

Sample Collection

↓

Protein extraction/Sample Preparation

↓

Liquid Chromatography

↓

Mass Spectrometry

↓

Proteome Bioinformatics

↓

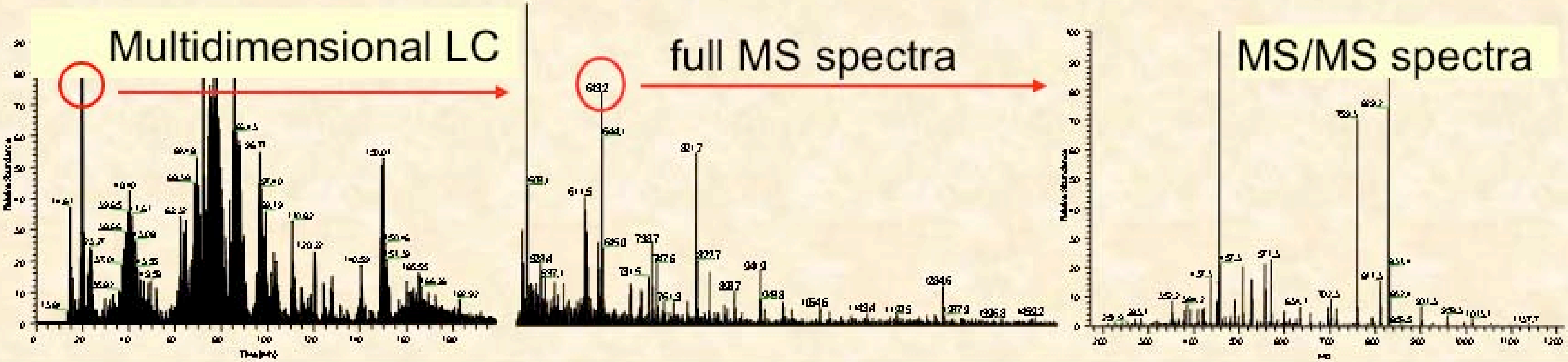
Biology

Protein extraction and Sample Prep
Will there be a one size fits all method??

- Chemical Lyses: Guanidine or Urea plus DTT
- Detergent Lyses: SDS, non-ionic detergent
- Sonication
- Bead Beating
- French Press
- TCA Prec.

Protein extraction and Sample Prep
What do we want in a method

- Simple and Straightforward
- Un-biased
- Reproducible
- Fast
- Sensitive
- Don’t add unwanted chemical modifications-oxidations, SDS adducts, carbamomethyations



- protein mixture —digest→ peptides (---Lys,---Arg) —LC→ ***measure*** **peptides and fragmentation**

Computational match

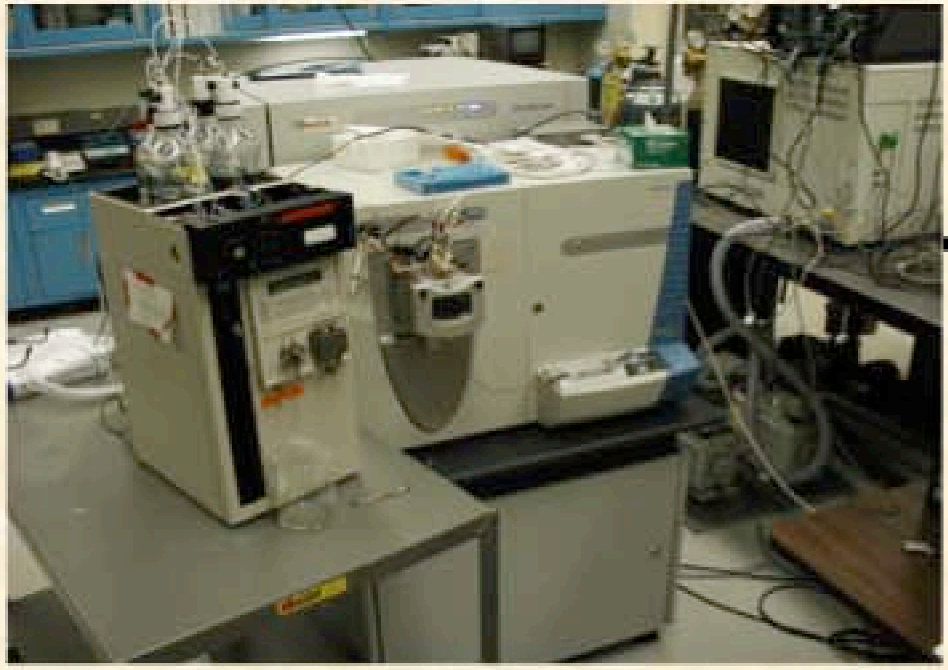
- genome sequence —computer→ predict proteins, fragment at Lys, Arg → ***predict*** **peptides and fragmentation**

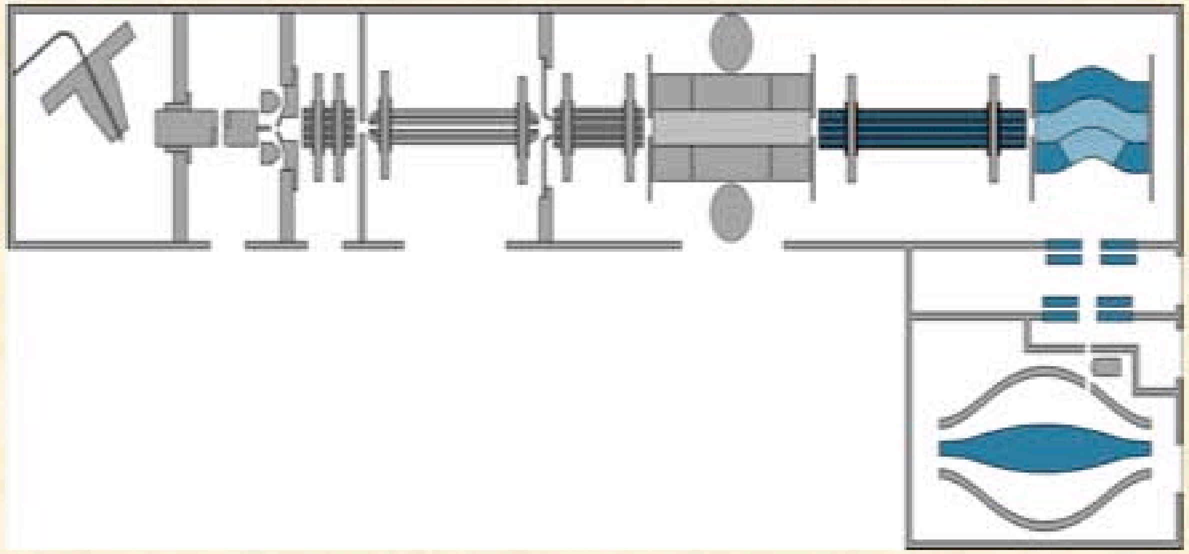
Leptospirillum **group II_scaffold_14_GENE_20**

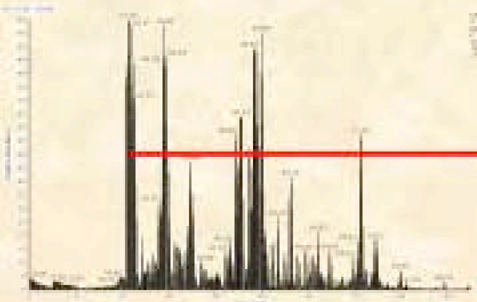
```
MNKWAGAVLGTVTLGLLSATAYSAELDILKPNRVPADQIAAAKAMKPPFPVTA
AVIAKGKEVFNGAGTCYTCHGVGGKGDGPGAAGMDPSPRFTNHQFDQVRTAGE
MVWVVSNGSPLQPAMVGFVSAGITDKQAWEAVMYERSLGCGGDMDC.......
.
```

High resolution and accurate mass measurements of both Full scan and MS/MS spectra on LC time scales

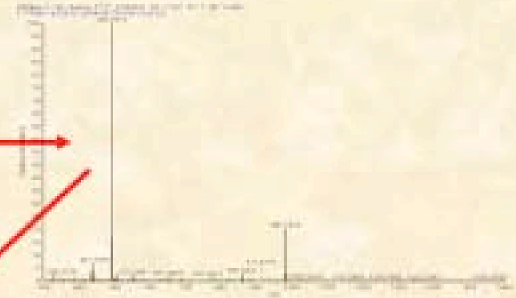
Solution: LTQ FT Orbitrap



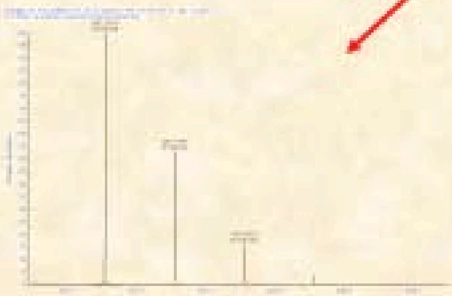




Base Peak Chromatogram

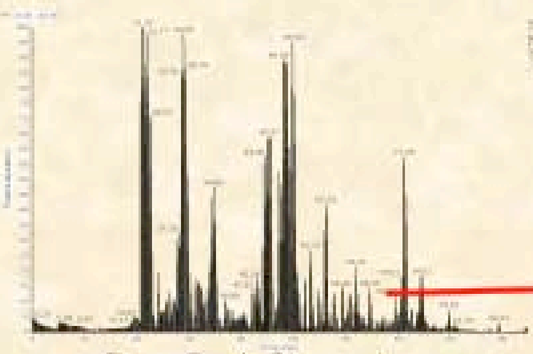


Full Scan MS

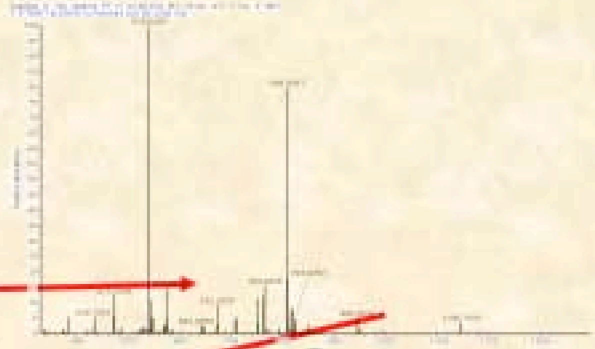


LeptoII_scaff_14_GENE_20_SNP1
VPADQIAAAK
Actual Monoisotopic Mass= 982.544
Observed Monoisotopic Mass= 982.546
PPM= 2.0 ppm

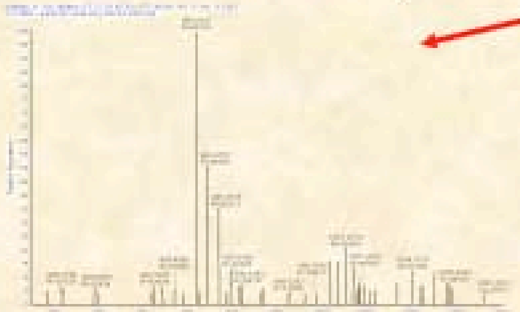
Highest point in Dynamic Range!
This is because of AGC.



Base Peak Chromatogram



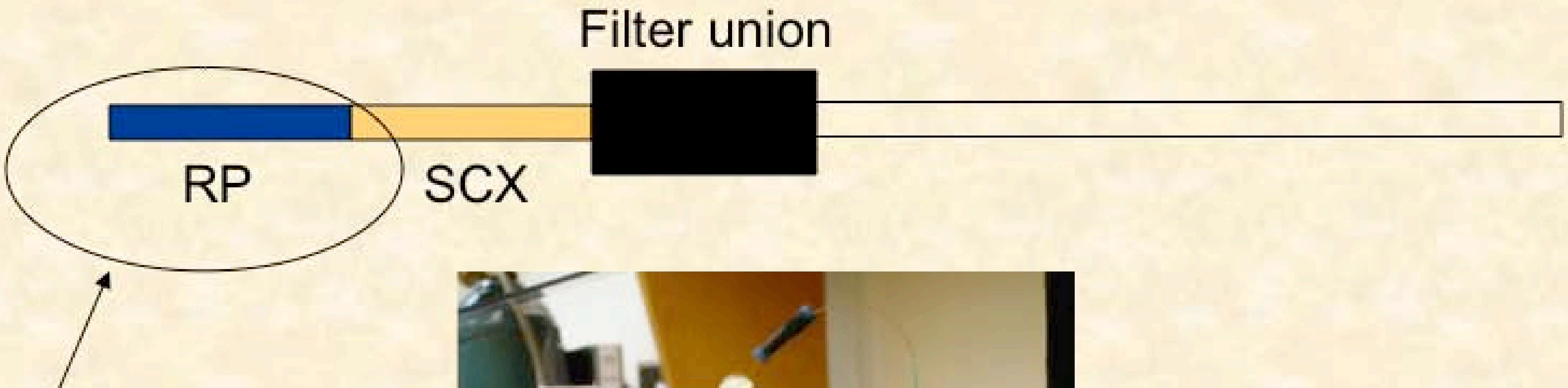
Full Scan MS



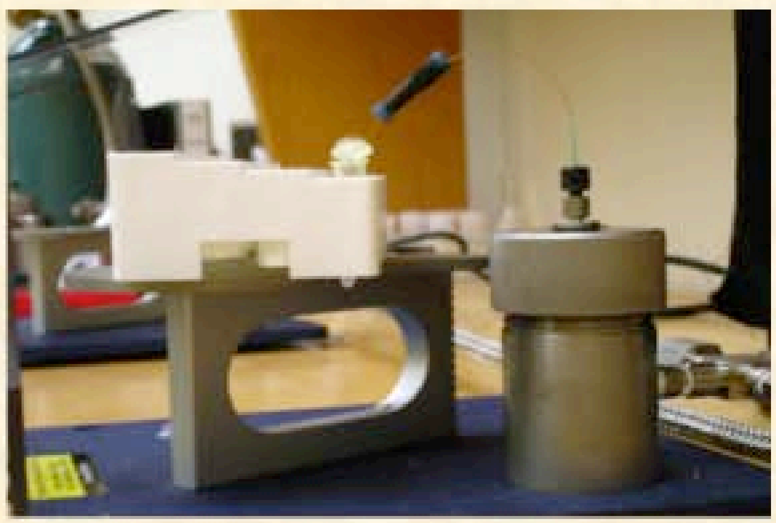
LeptoII_scaff_14_GENE_20_SNP1
SLGCGGDMDCVTGSADWVSK
Actual Monoisotopic Mass= 1986.811
Observed Monoisotopic Mass= 1986.813
PPM= 1.0 ppm

Lowest point in Dynamic Range!
This is because of AGC.

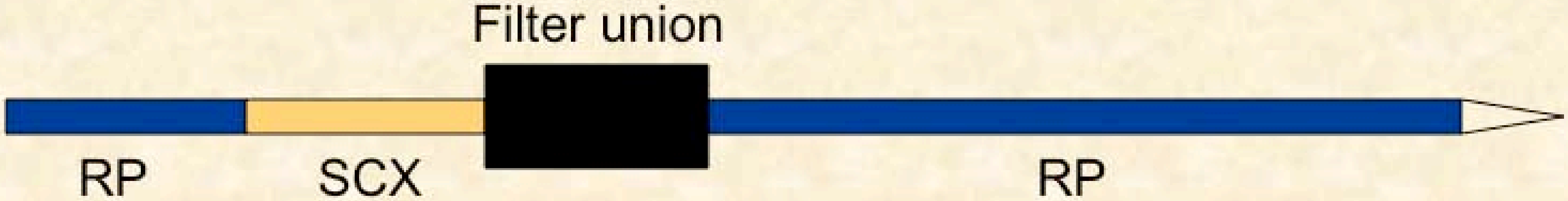
Split-phase MudPIT (pressure cell-packing and loading) described in McDonald et al. IJMS 2002



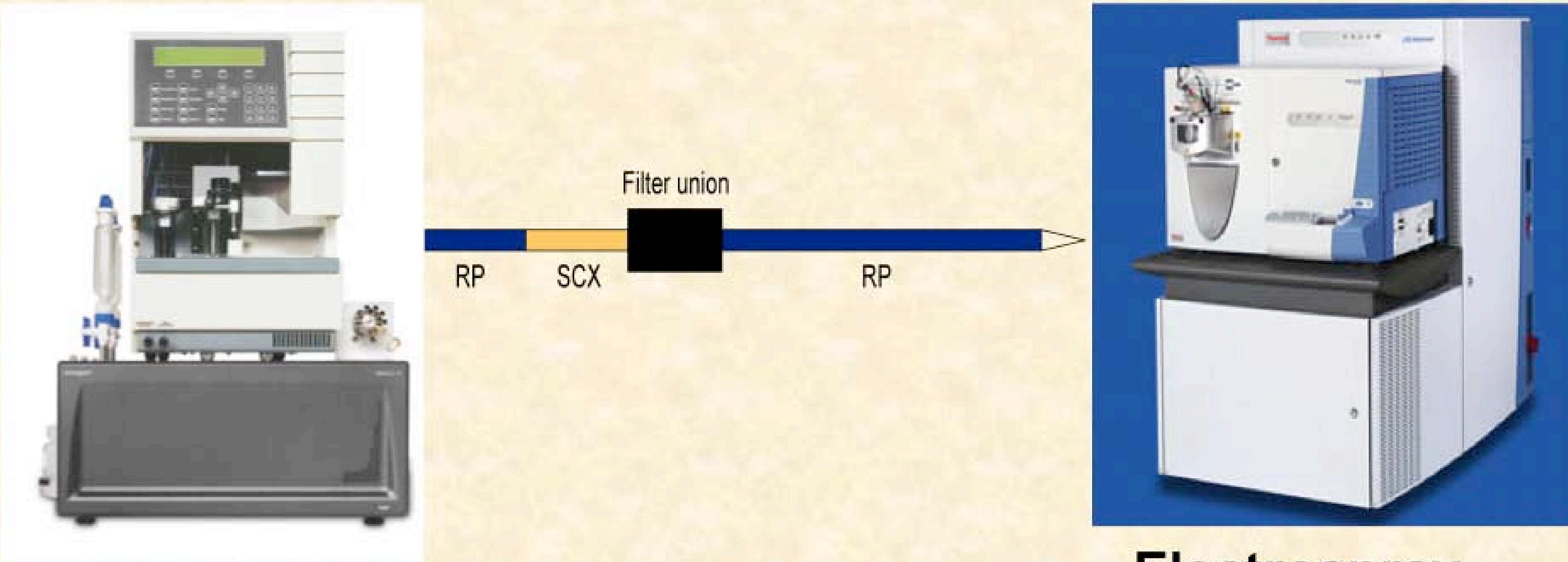
For "dirty" samples Omit this phase



Pressure cell packing and loading



Experimental Approach: Digest the proteins into peptides and examine by LC/LC-MS/MS



2D-Nano HPLC

Electrospray Linear Ion Trap Orbitrap

11 Ammonium Acetate salt pulses followed by 2 hour Reverse Phase Gradients

ORNL Proteome Informatics Pipeline

Extraction

↓

z Assignment

↓

Peptide Identification

↓

Protein Assembly & Filtering

↓

ppm Computation

↓

Sample Differentiation & Statistics

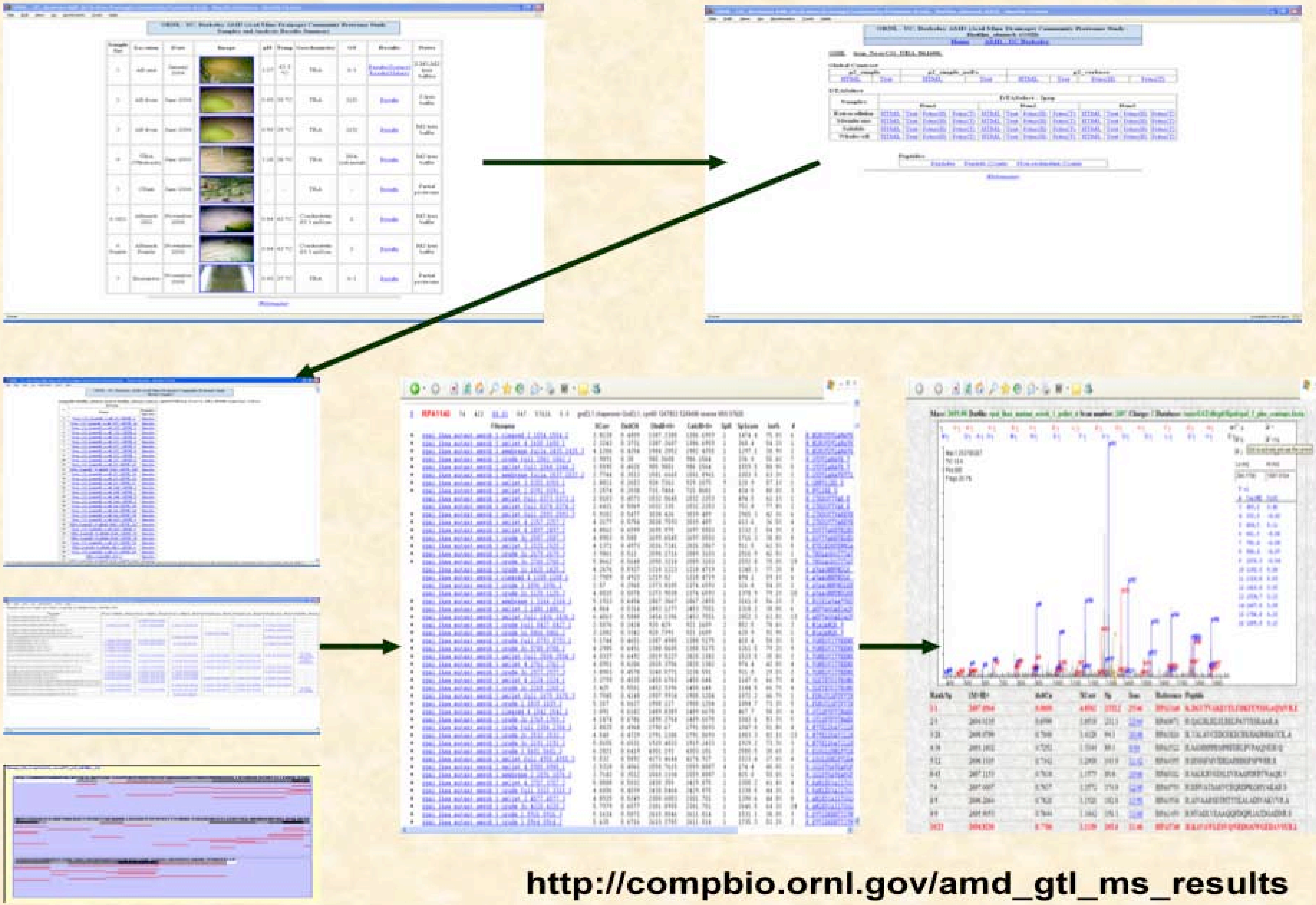
↓

Web-Based Publishing

↓

mySQL Database

ORNL Proteome Informatics Pipeline

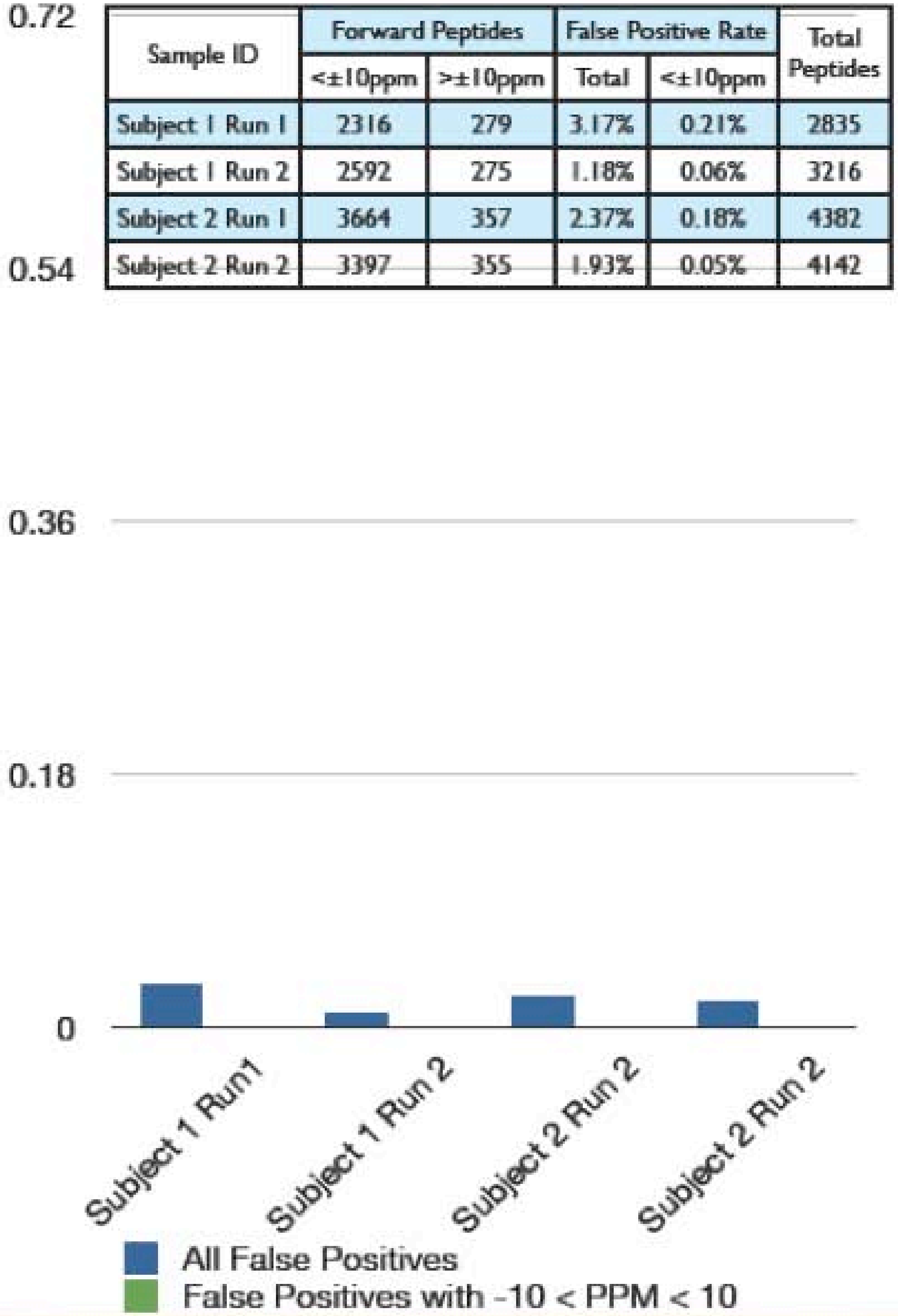


http://compbio.ornl.gov/amd_gtl_ms_results

False Positive Rates

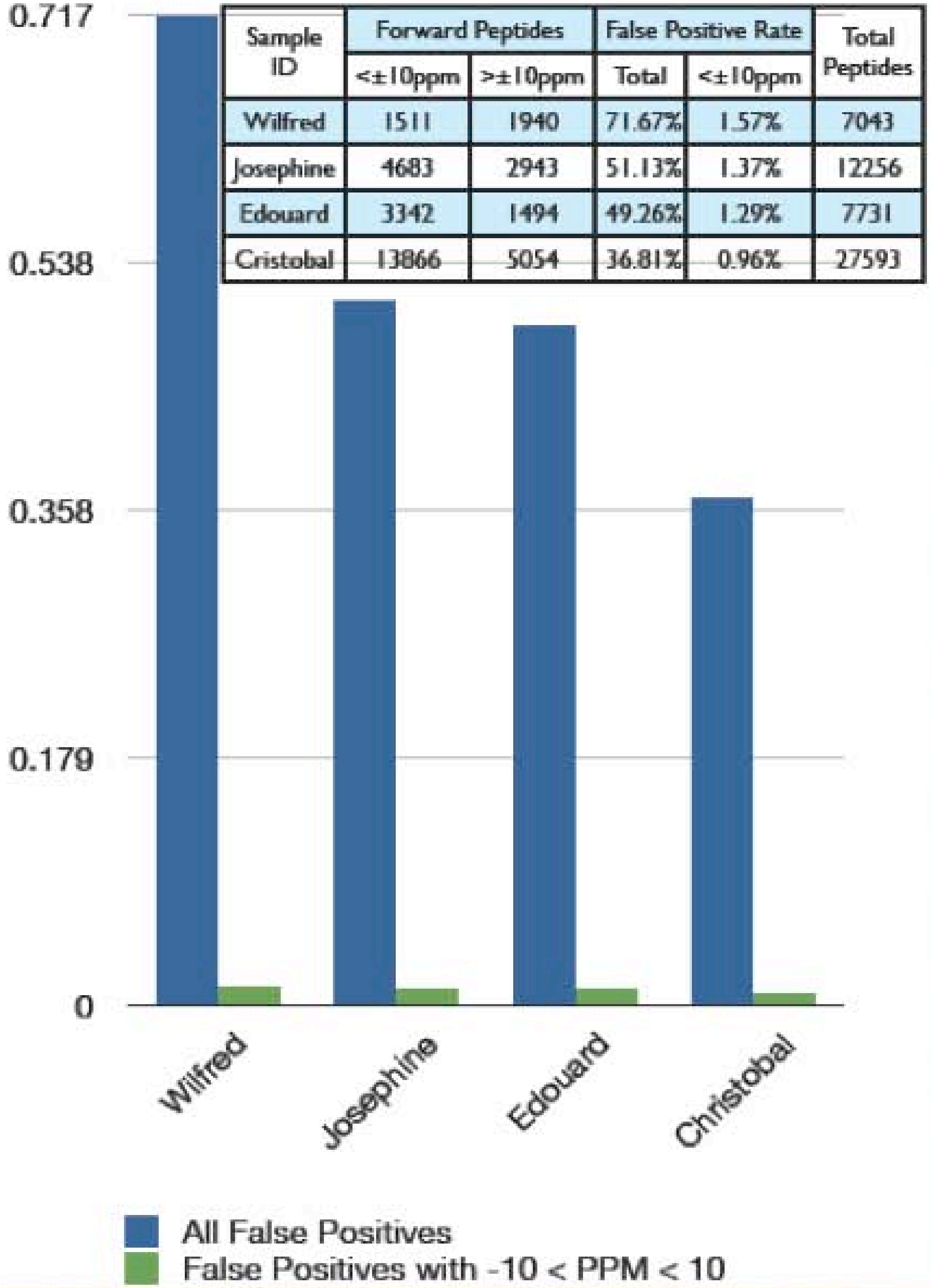
Human Gut Metaproteome

Sample ID	Forward Peptides		False Positive Rate		Total Peptides
	<±10ppm	>±10ppm	Total	<±10ppm	
Subject 1 Run 1	2316	279	3.17%	0.21%	2835
Subject 1 Run 2	2592	275	1.18%	0.06%	3216
Subject 2 Run 1	3664	357	2.37%	0.18%	4382
Subject 2 Run 2	3397	355	1.93%	0.05%	4142



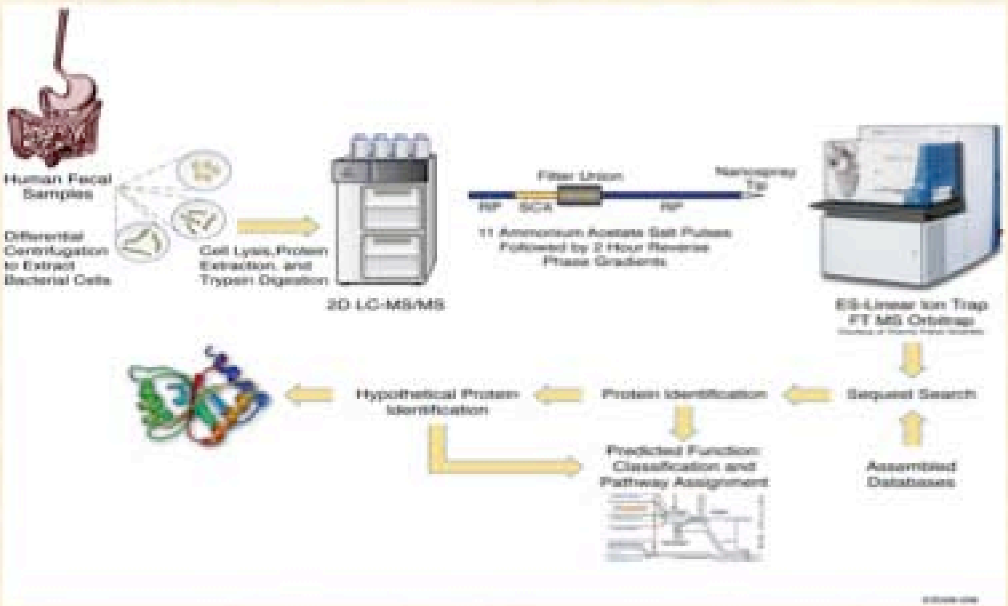
Rifle Gound Soil Sample

Sample ID	Forward Peptides		False Positive Rate		Total Peptides
	<±10ppm	>±10ppm	Total	<±10ppm	
Wilfred	1511	1940	71.67%	1.57%	7043
Josephine	4683	2943	51.13%	1.37%	12256
Edouard	3342	1494	49.26%	1.29%	7731
Cristobal	13866	5054	36.81%	0.96%	27593



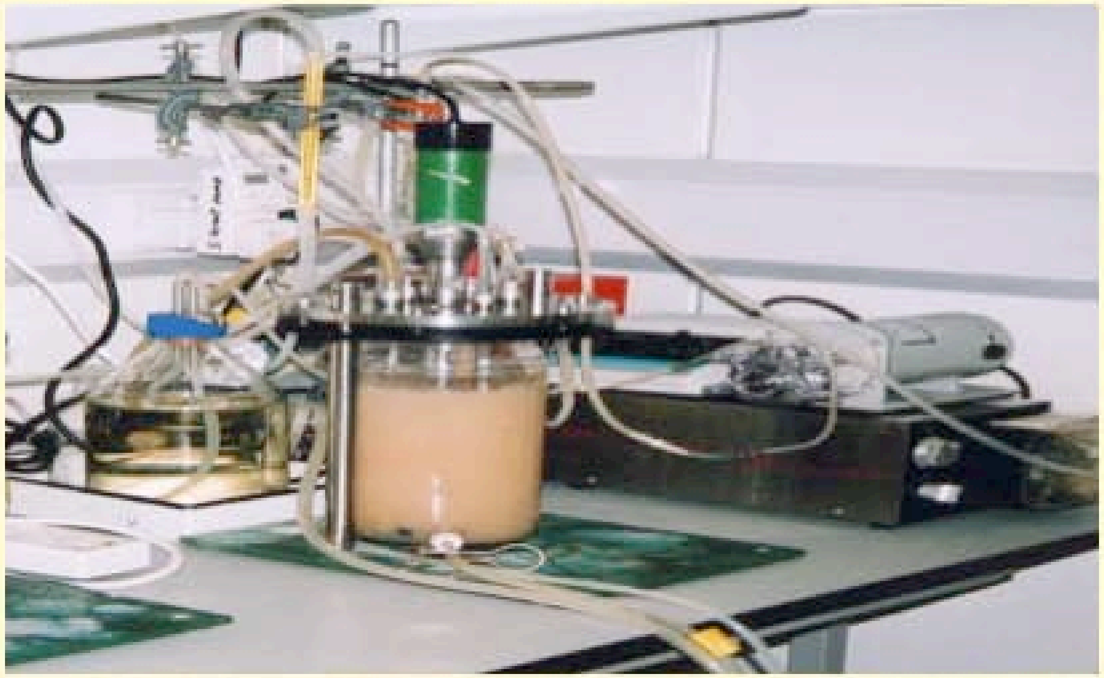
Released Published Datasets

Human Gut Metaproteome



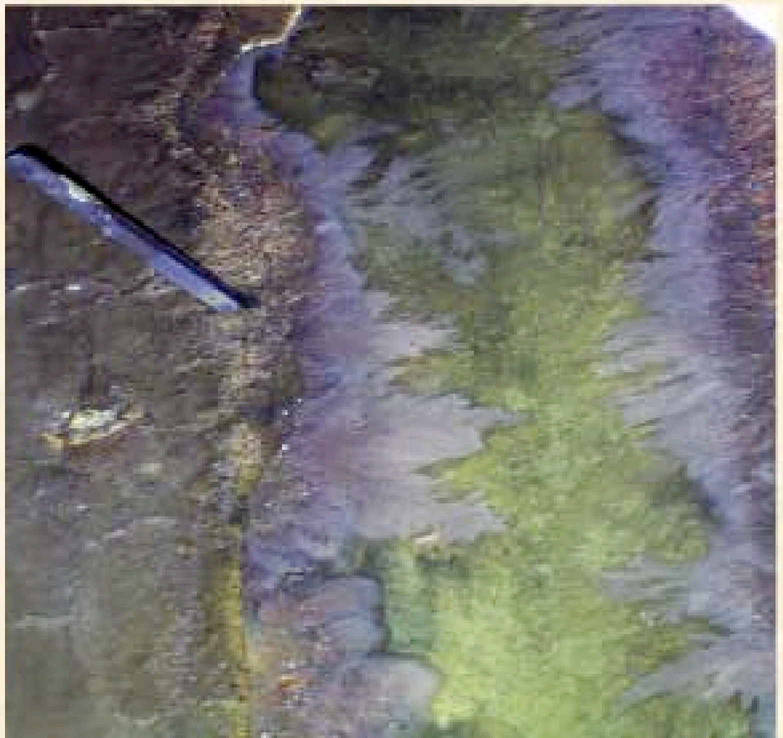
http://compbio.ornl.gov/
human_gut_microbial_metaproteome/

EBPR Sludge



http://compbio.ornl.gov/ebpr_sludge/

Acid Mine Drainage



http://compbio.ornl.gov/biofilm_amd/
http://compbio.ornl.gov/biofilm_amd_recombination/
http://compbio.ornl.gov/biofilm_amd_PIGT/

A golden soil sample prep method
What have we tried

- Chaotropes-Guanidine
- Basic extraction-NaOH
- Sonication
- Bead Beating
- Excess trypsin
- Blocking with ubiquitin

All of these failed in the end…

In Situ Soil Protein Extraction Methodology (SDS-TCA)

Soil samples dispersed in a detergent based lysis buffer and boiled for 20 min

↓ brief centrifugation at 2500 rpm, discard soil

Supernatant + chilled TCA

↓ Overnight incubation at 4^0C followed by centrifugation at 14,000 rpm, discard supernatant

Wash protein pellet with acetone, dry and dissolve in Guanidine-DTT solution.

↓

Overnight trypsin digestion followed by desalting via SPE and solvent exchange

↓

Peptides interrogated via 24h, 12-step, 2d-LC MS/MS

↓

Datasets analyzed using SEQUEST

Method Developed by Dr. K. Chourey and Dr. R. Hettich assistance from Dr. Jansson

Testing the efficiency of the *in-situ* lysing protocol

Live Cell spiking experiment

Grow *Pseudomonas putida F1* cells o/n in 10 ml LB medium

5 ml
Added to 5 gm sterilized soil (Hanford, WA)
(incubated for 5 h)

5 ml
Taken up directly for protein extraction

Sample subjected to SDS-TCA lysis protocol, followed by trypsin digestion of proteins, desalting the peptides and analyzing them via 24h 2d-LC-MS/MS on a LTQ MS

Datasets analyzed using SEQUEST software with *P. putida* F1 as the reference database

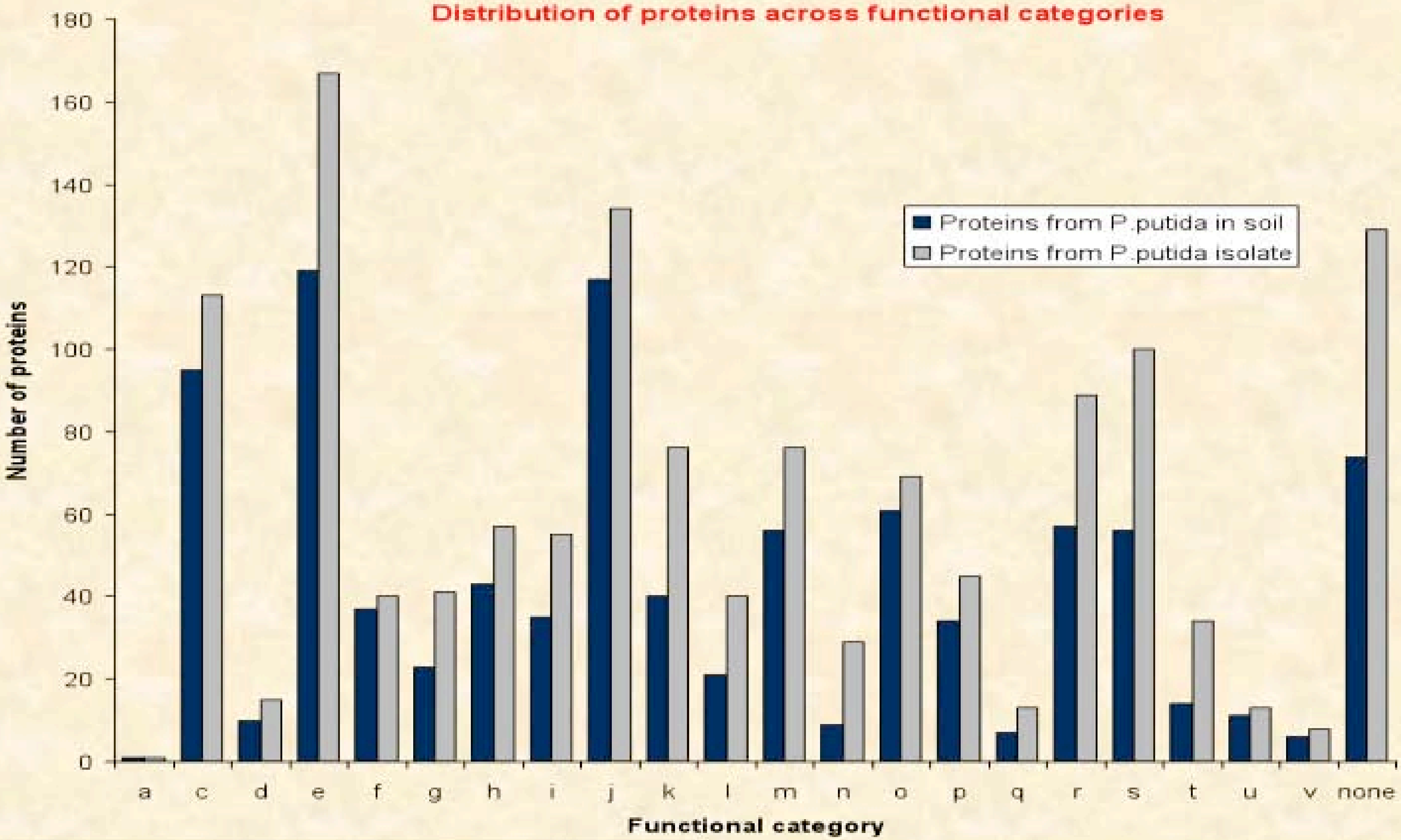
925 proteins identified

MW of proteins ranged from ~ 6 KDa to 181 KDa, with a single 269 KDa protein

1343 proteins identified

MW of proteins ranged from ~ 6 KDa to 181 KDa, with a single 475 KDa protein

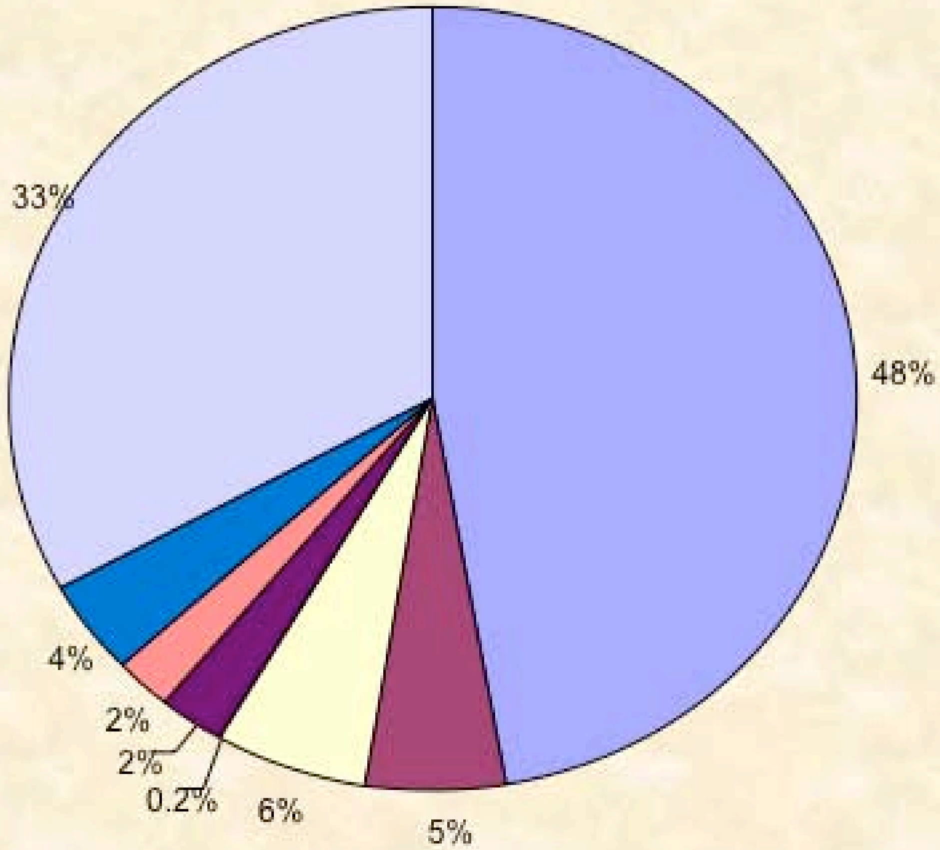
Results: The protocol works well for lysing cells within soil matrix.



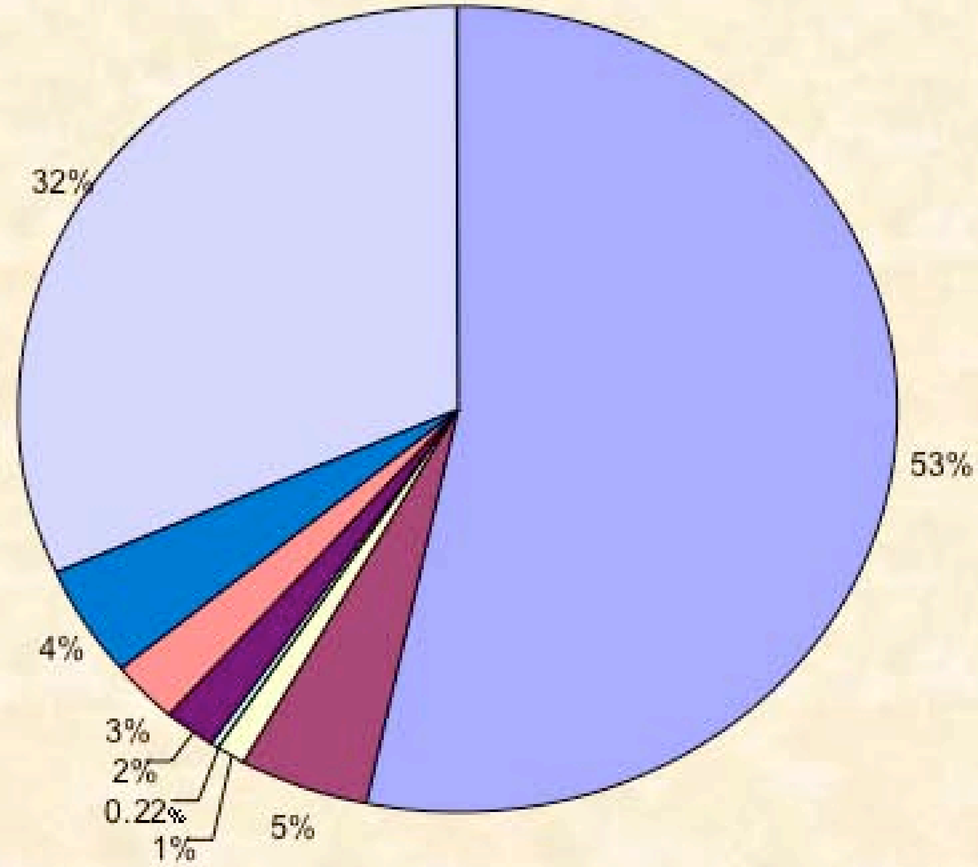
Key	Functional category (COG)
a	RNA processing and modification
c	Energy production and conversion
d	Cell cycle control, cell division and chromosome partitioning
e	Amino acid transport and metabolism
f	Nucleotide transport and metabolism
g	Carbohydrate transport and metabolism
h	Coenzyme transport and metabolism
i	Lipid transport and metabolism
j	Translation of ribosomal structure and biogenesis
k	Transcription

l	Replication, recombination and repair
m	Cell wall membrane and envelope biogenesis
n	Cell Motility
o	Post translational modification and protein turnover chaperones
p	Inorganic ion transport and metabolism
q	Secondary metabolites biosynthesis, transport and catabolism
r	General function prediction
s	Unknown Function
t	Signal transduction mechanisms
u	Intracellular trafficking, secretion and vesicular transport
v	Defense mechanisms

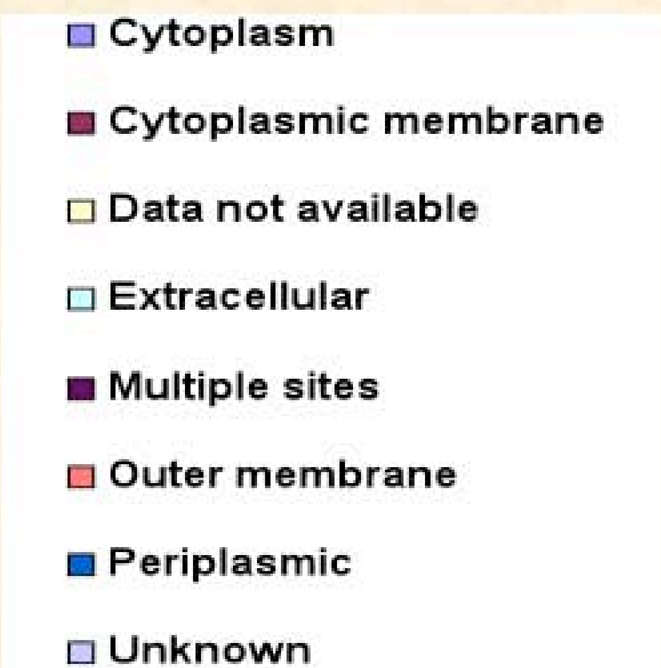
COMPARISON OF PROTEIN LOCALIZATION PROFILE OF IDENTIFIED PROTEINS



P.putida F1 proteome from liquid culture



P.putida F1 proteome from cells incubated in soil.



Indirect Soil Protein Extraction protocol

(Differential Centrifugation)

Grow *Arthrobacter chlorophenolicus* A6 cells in liquid media

↓

Culture suspended in PBS buffer and added to 20 gm Hopland soil to the final concentration of 10^9 or 10^8 cells (incubated at 4°C for 24 h)

↓

20 ml chilled PBS added to soil (kept on ice) and homogenized at top speed using hand held blender (30 sec on/off) twice.

↓

Slurry centrifuged at 5000 rpm x 5 min to sediment out soil particles. Supernatant containing cells reserved. Repeat the process with soil again. Pool all supernatant.

↓

Centrifuge supernatant at 10,000 rpm, 10 min at 4°C to precipitate cells. Wash cell pellet once with chilled PBS.

↓

Cell pellet subjected to SDS-TCA lysis protocol, followed by trypsin digestion of proteins, desalting the peptides and analyzing them via 24h 2d-LC-MS/MS on a LTQ MS

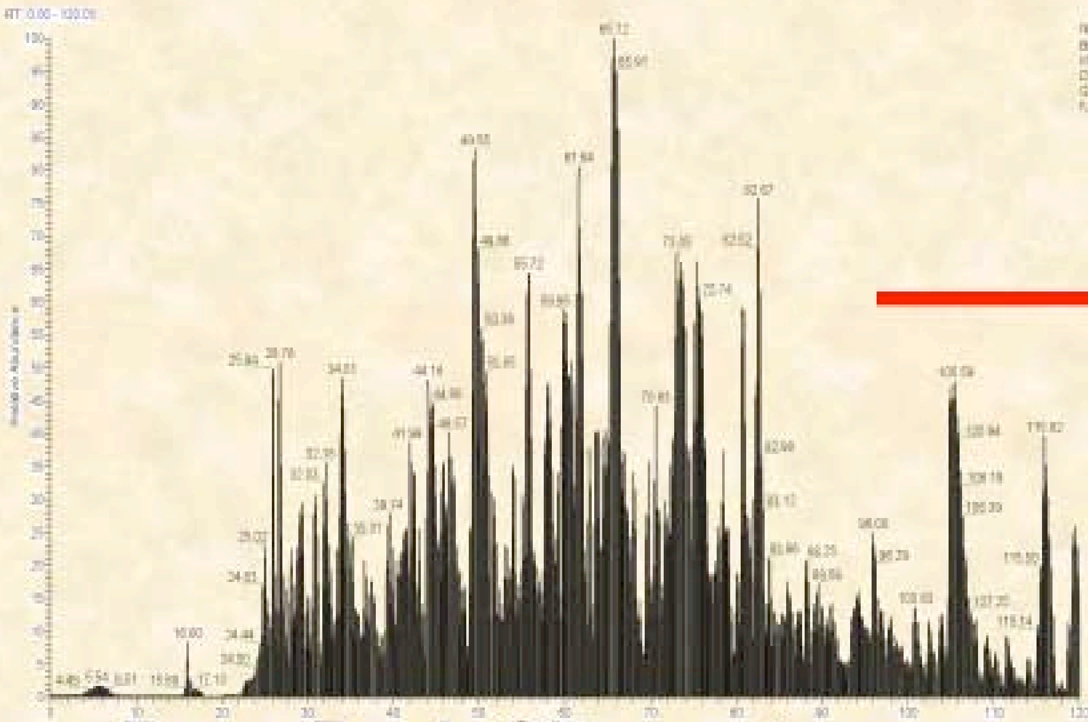
Results: The protocol isolates bacterial cells from soil particles effectively

Protocol Develop by Dr. Janet Jansson

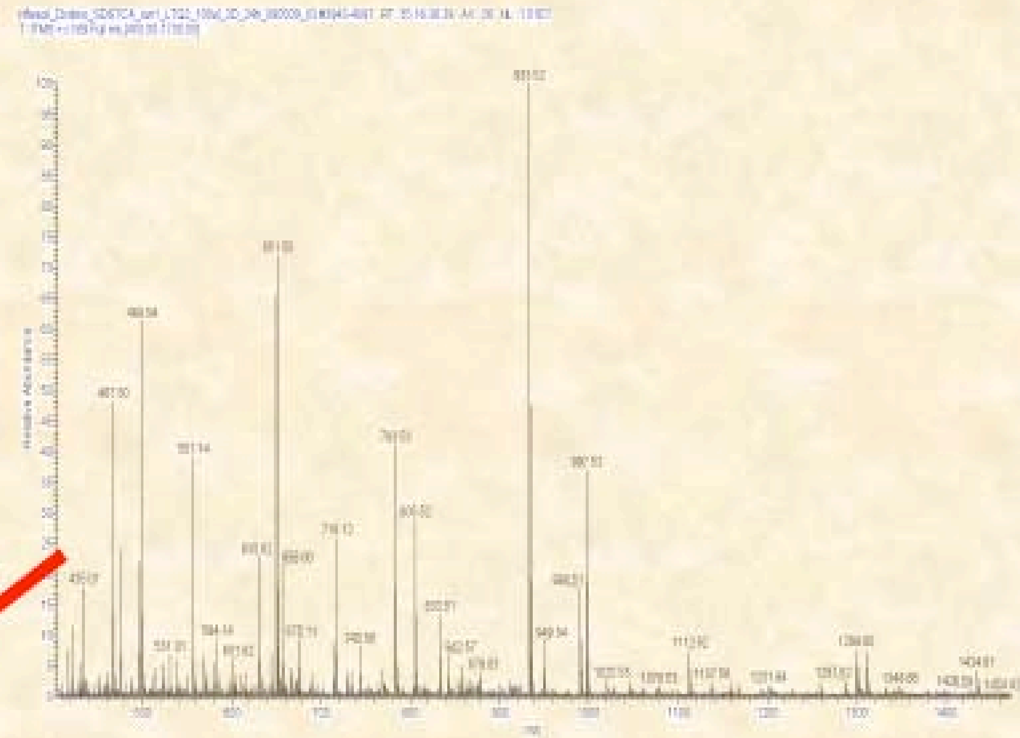
Results from Hopland soil spiked with *A. chlorophenolicus* A6 cells.

Soil	*Arthrobacter* (cells/g)	Extraction protocol	Lysis method	Total proteins	Comments
20 g Hopland soil	**10^9**	**Differential centrifugation**	**SDS-TCA**	490	1/4 cell lysate equivalent to 5 g soil
		"	**SDS-TCA**	600	**3/4 cell lysate equivalent to 15 g soil**
20 g Hopland soil	**10^8**	**Differential centrifugation**	**SDS-TCA**	582	1/4 cell lysate equivalent to 5 g soil
		"		457	**3/4 cell lysate equivalent to 15 g soil**
19 g Hopland soil	**10^9**	**Direct extraction (5 g soil)**	**SDS-TCA**	816	
19 g Hopland soil	**10^8**	**Direct extraction (5 g soil)**	**SDS-TCA**	555	
20 g Hopland soil	**No cells**	**Direct extraction (5 g soil)**	**SDS-TCA**	8	

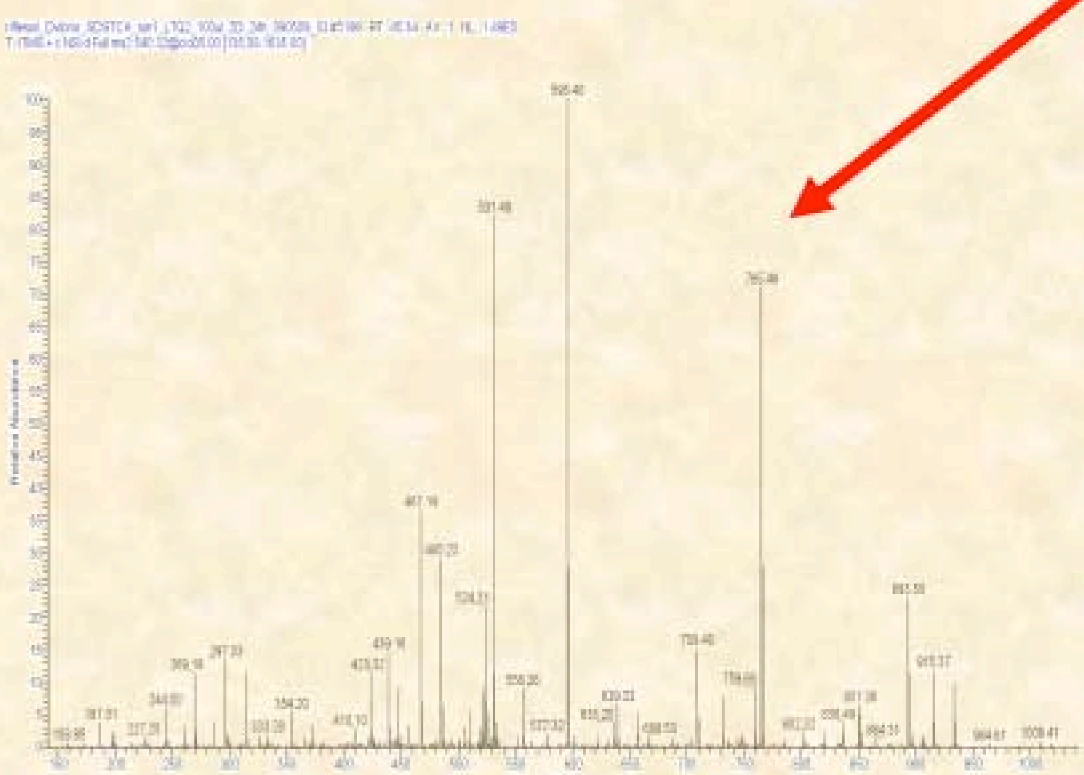
Desulfo isolate in Rifle Soil microcosms



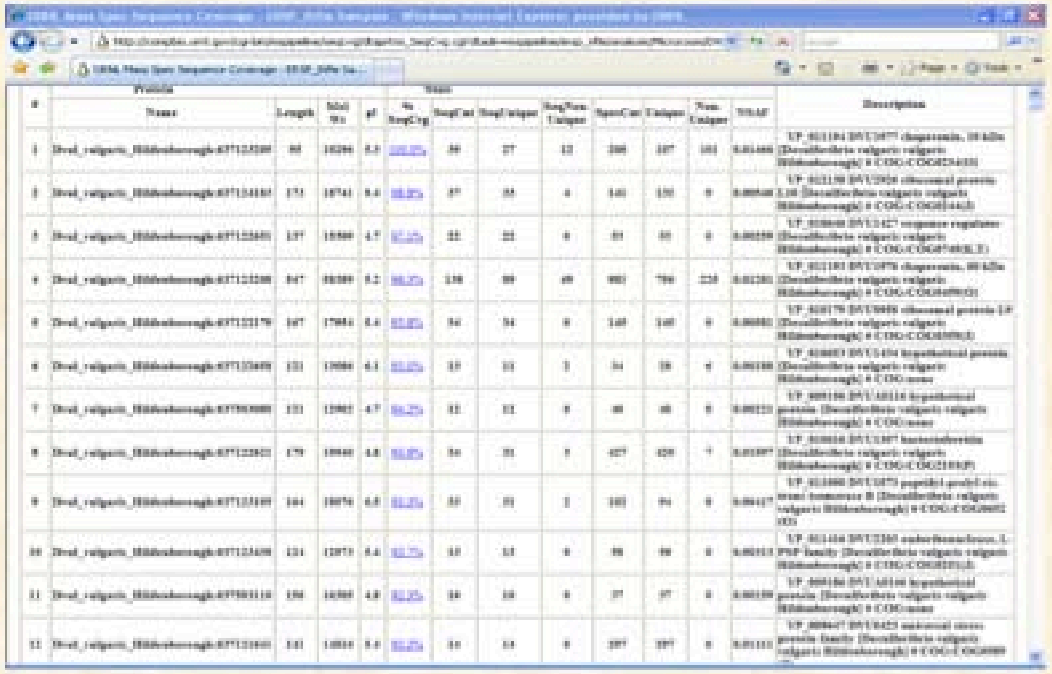
Base Peak Chromatogram



Full Scan Peptides

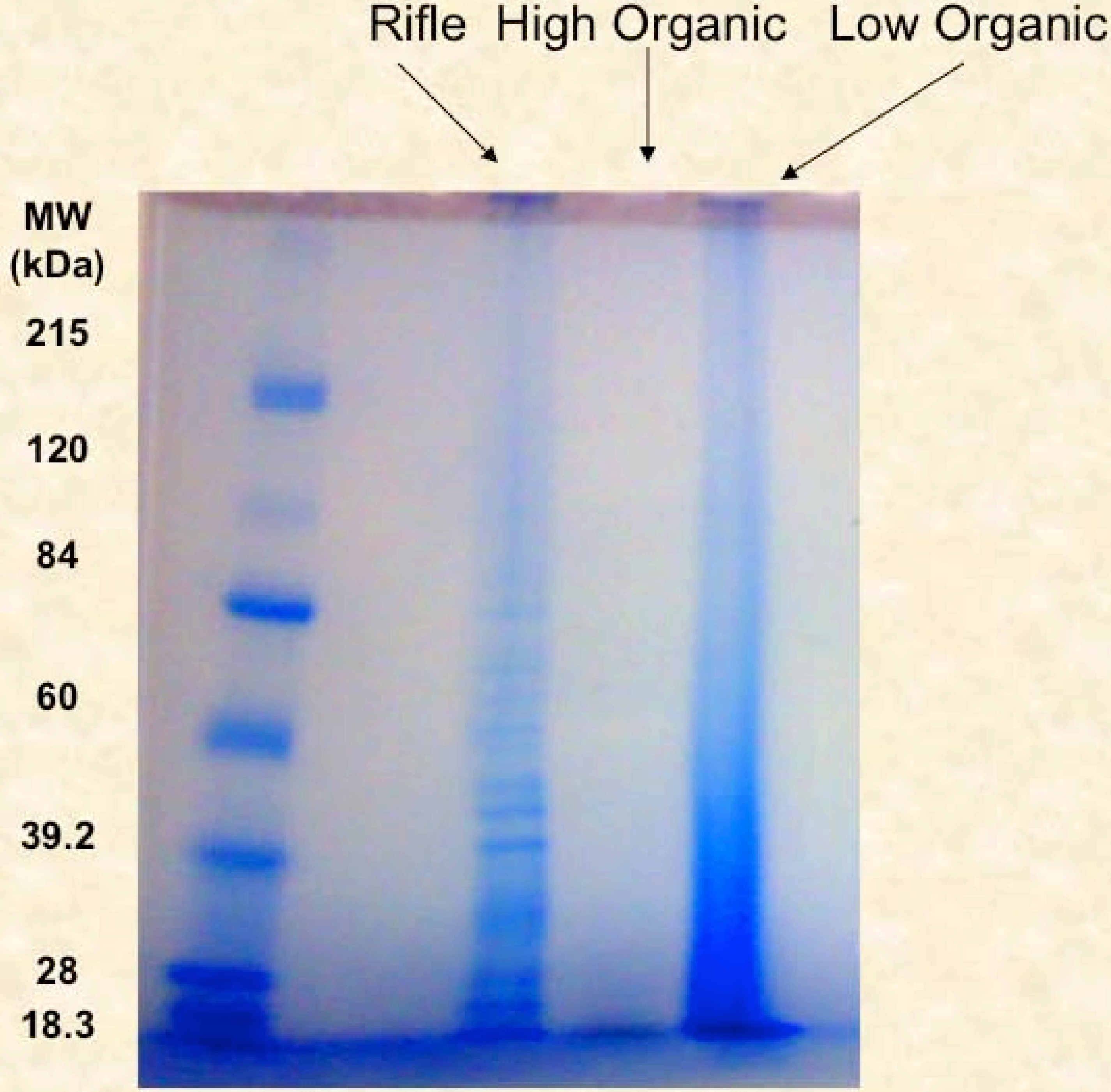


MS/MS spectra



~1300 Proteins Identified in 24h run

Sample provided by Kim Handley and Jill Banfield

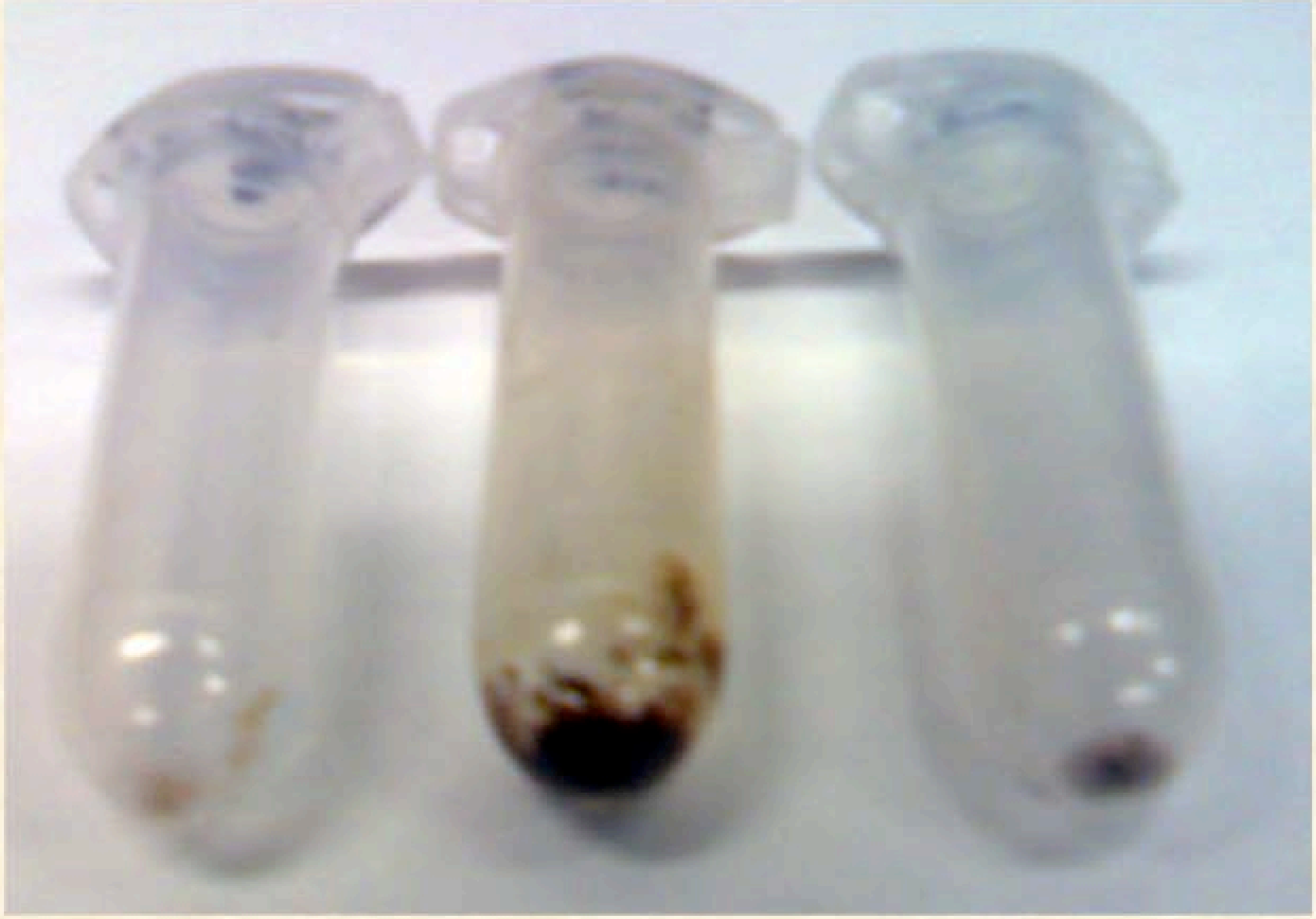


All Photos and work by Dr. Brian Dill

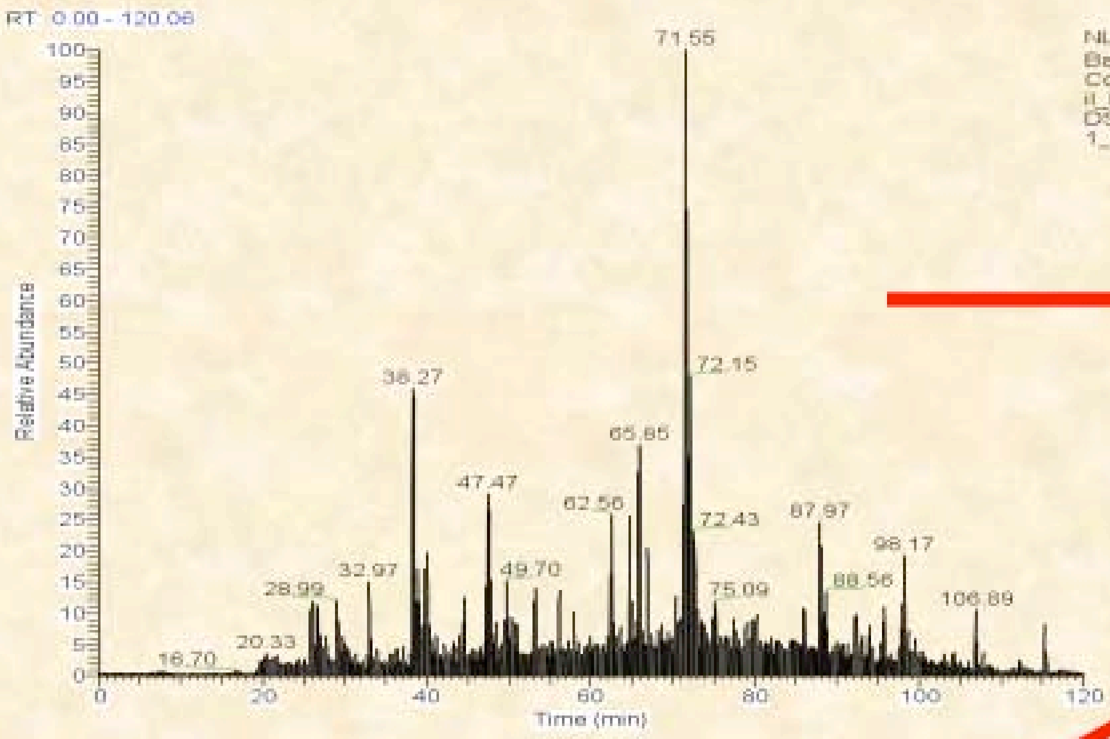
SDS-soil extract

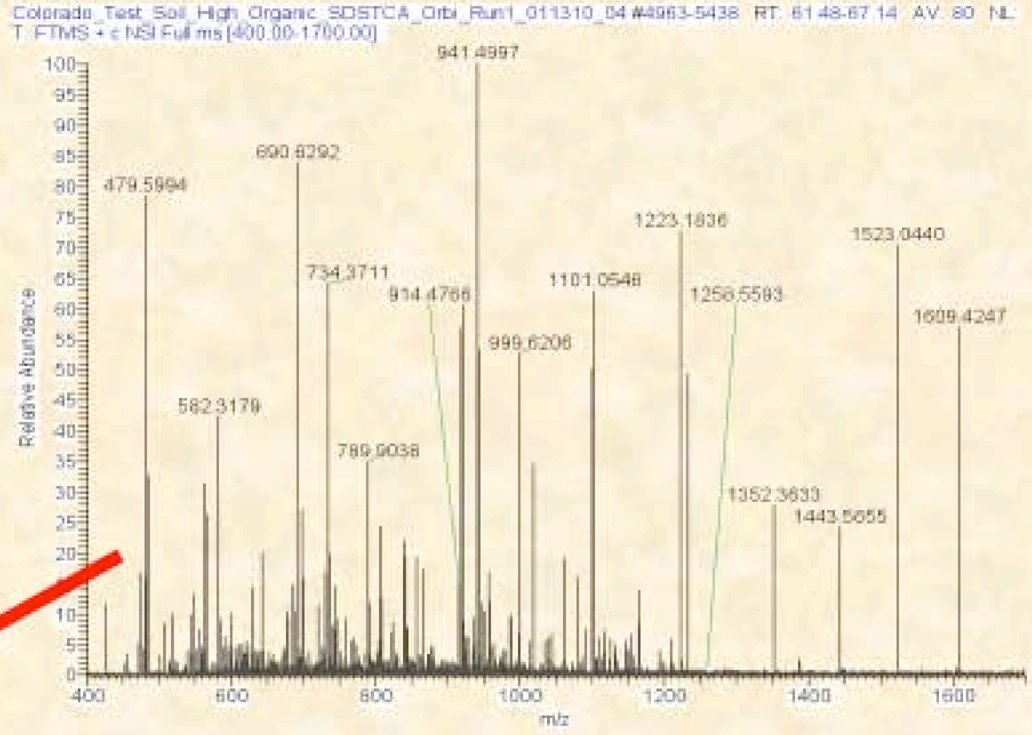
Aliquoted extract before TCA

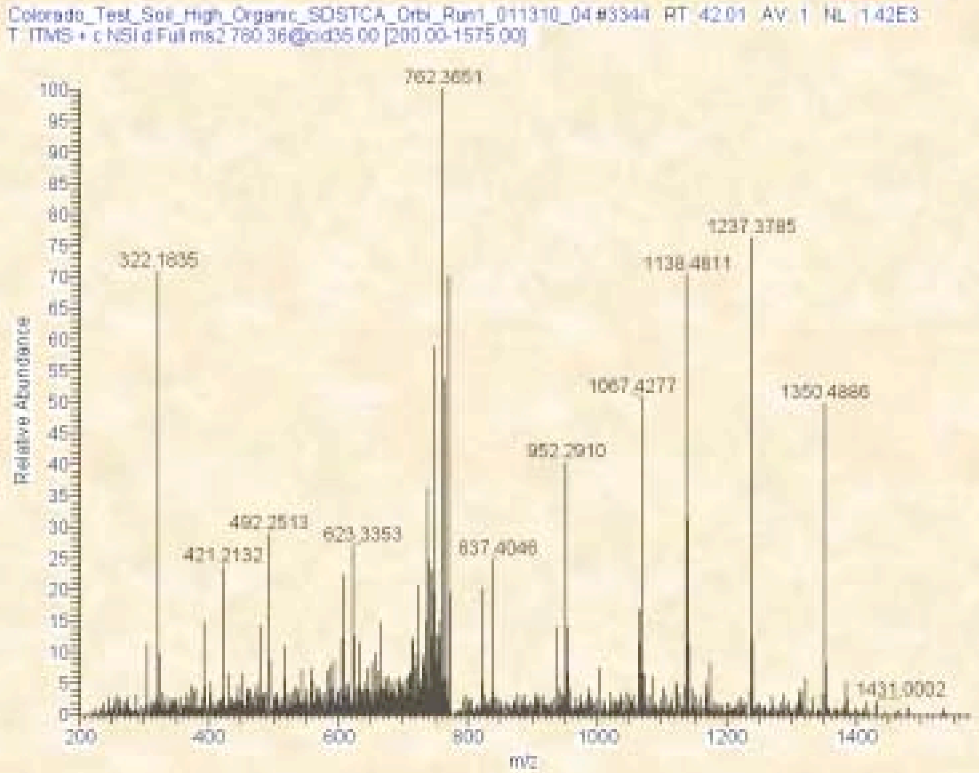
TCA pellets

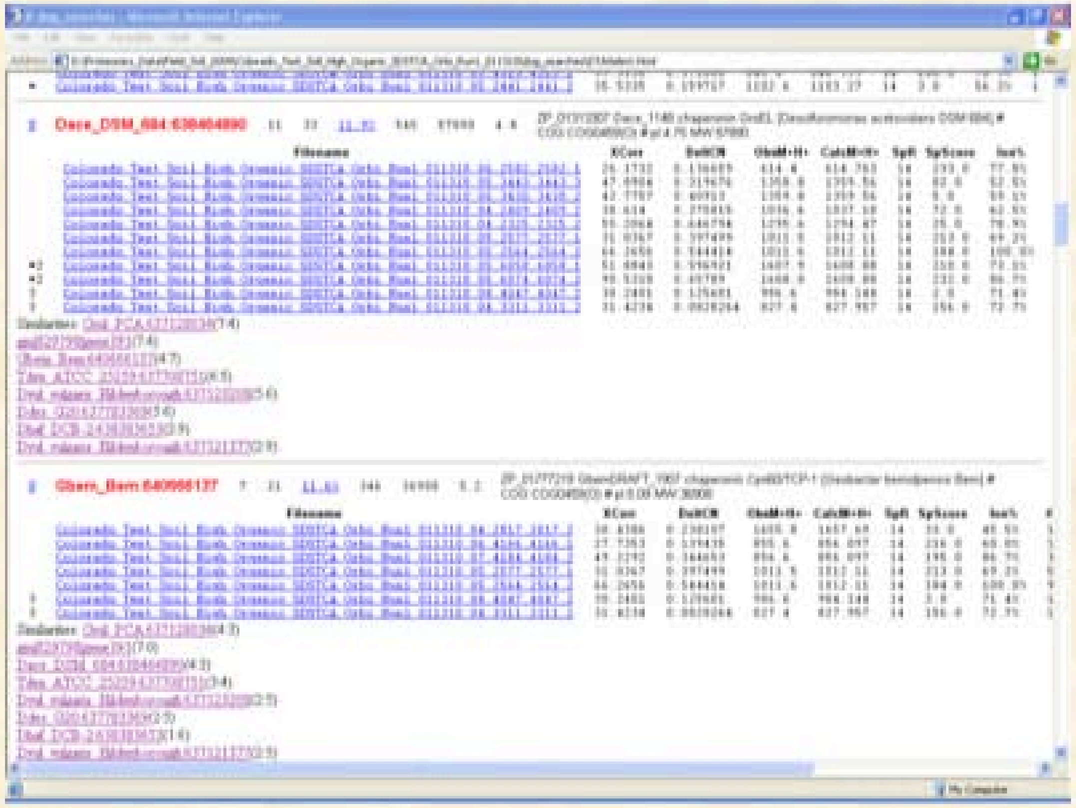


High Organic Colorado Soil









~80 Proteins Identified, ~200 Peptides

Rifle Integrated Field Research Challenge (IFRC) Site

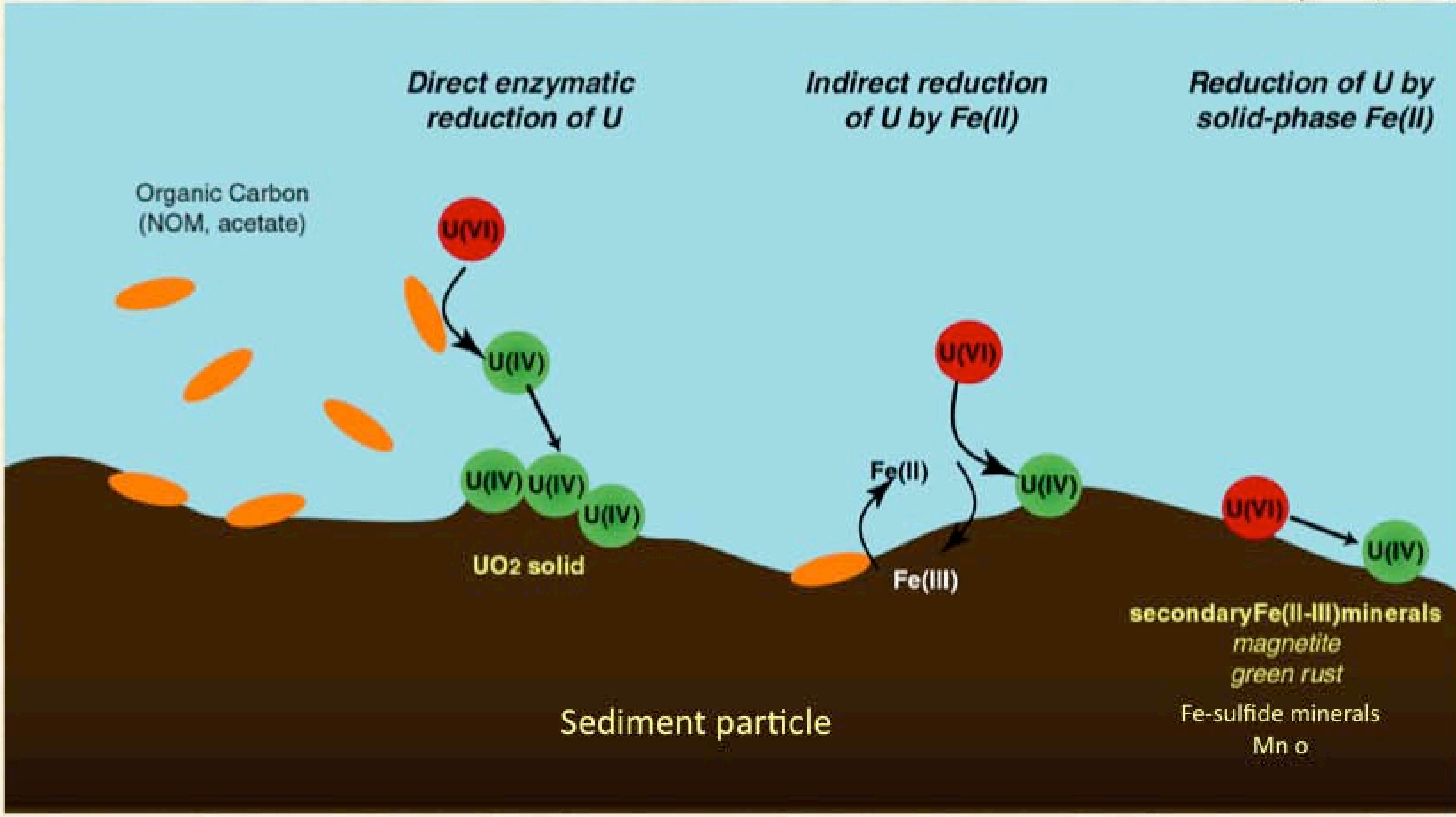
- Over 500,000 m^3 tailings were originally present at the site
- Tailings were removed to the New Rifle site, before being removed to the Estes Gulch storage cell
- Surface layer excavation and clay-rich cap installed: *1996*

1998: Persistent U(VI) plume
(2008 distribution unchanged)

Uranium Biogeochemistry

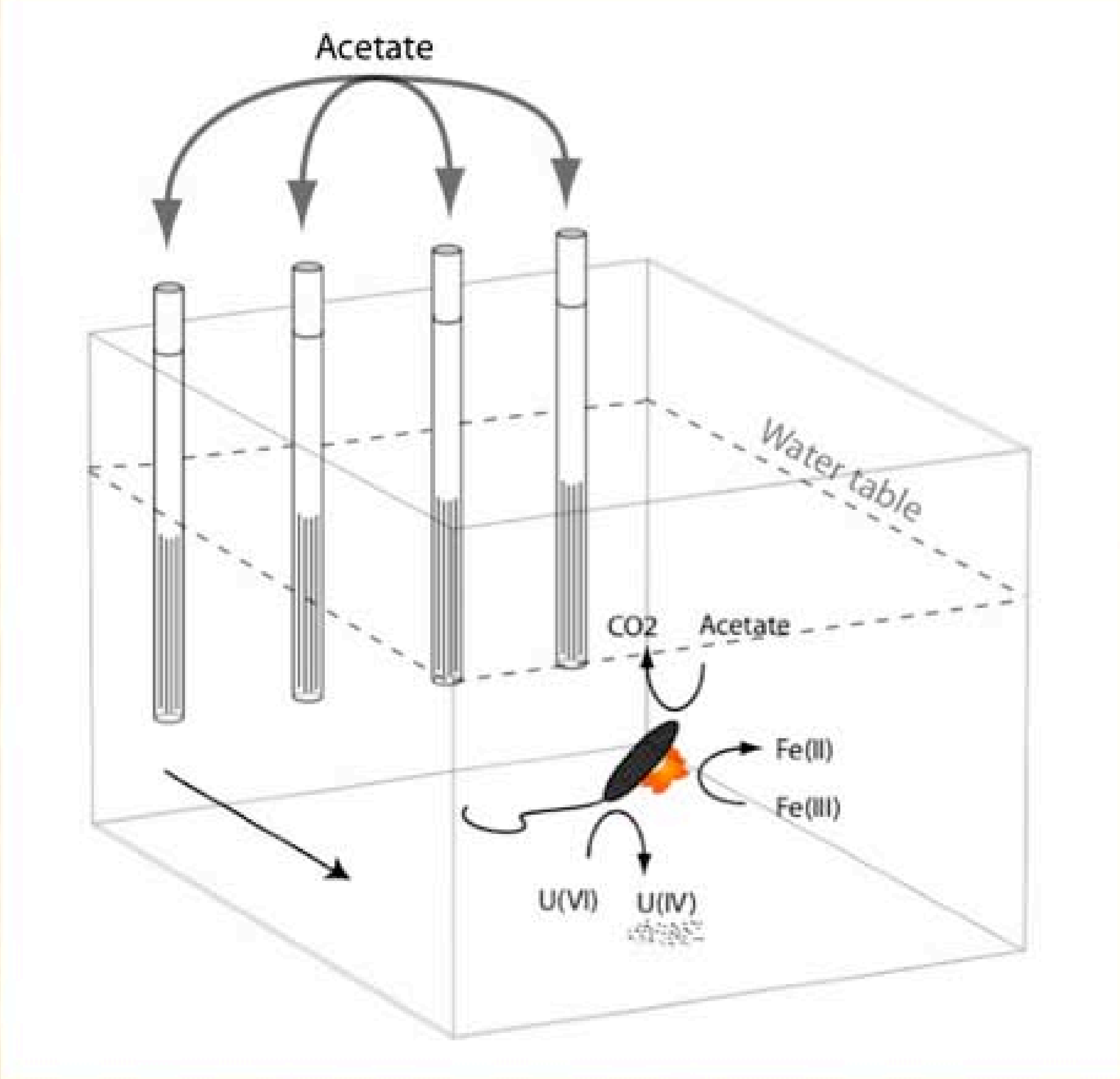
Biostimulation: U Sequestration

from Kate Campbell (USGS)



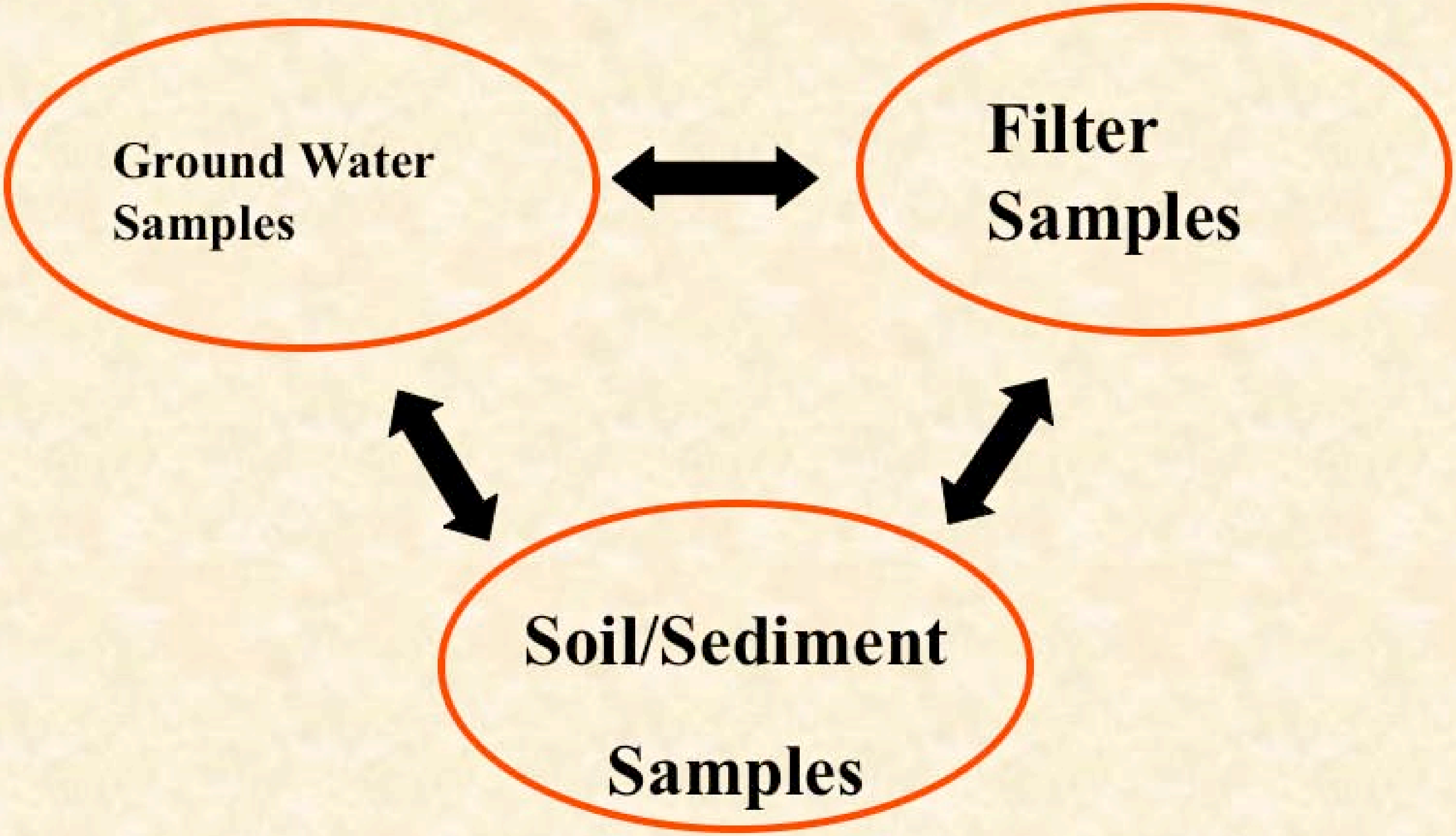
Selected references: Lovley et al., 1991; Suzuki et al., 2002; Suzuki et al., 2003; Singer et al., 2007; Gu and Chen 2003; Anderson et al., 2003; Liger et al., 1999; Jeon et al. 2005; Wersin et al. 1994; O'Loughlin et al., 2003; Missana et al., 2003; Scott et al., 2005; Dodge et al., 2002

Schematic model of bioreduction at the Rifle IFRC site

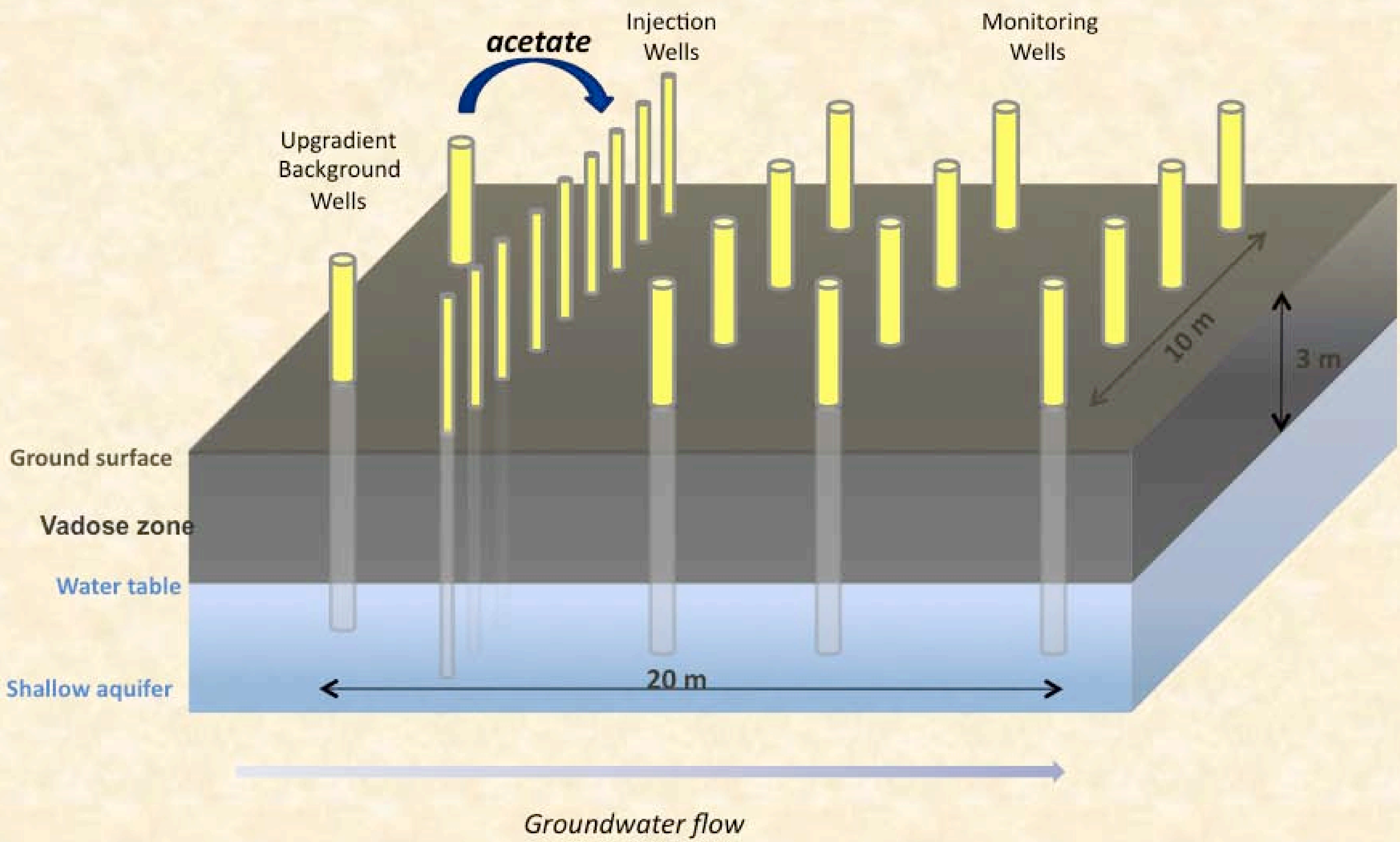


At Rifle, studies utilizing a number of different techniques (PLFA, RNA, DNA) have demonstrated that acetate amendment is followed by a bloom in *Geobacter* populations that reduce Fe(III) and U(VI)

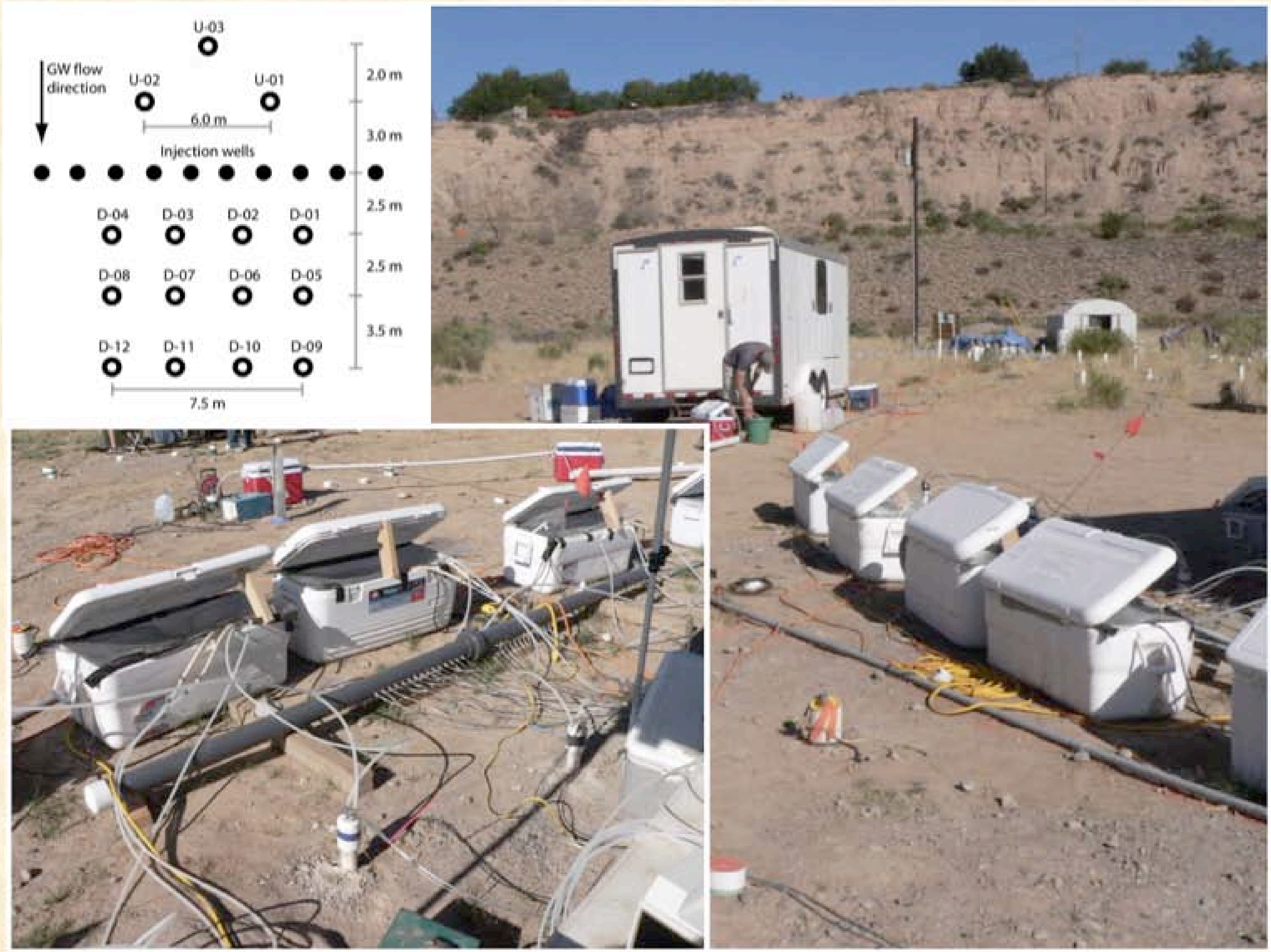
Three sample types from Rifle that we would like to be able to integrate for a comprehensive picture of the bioremediation process



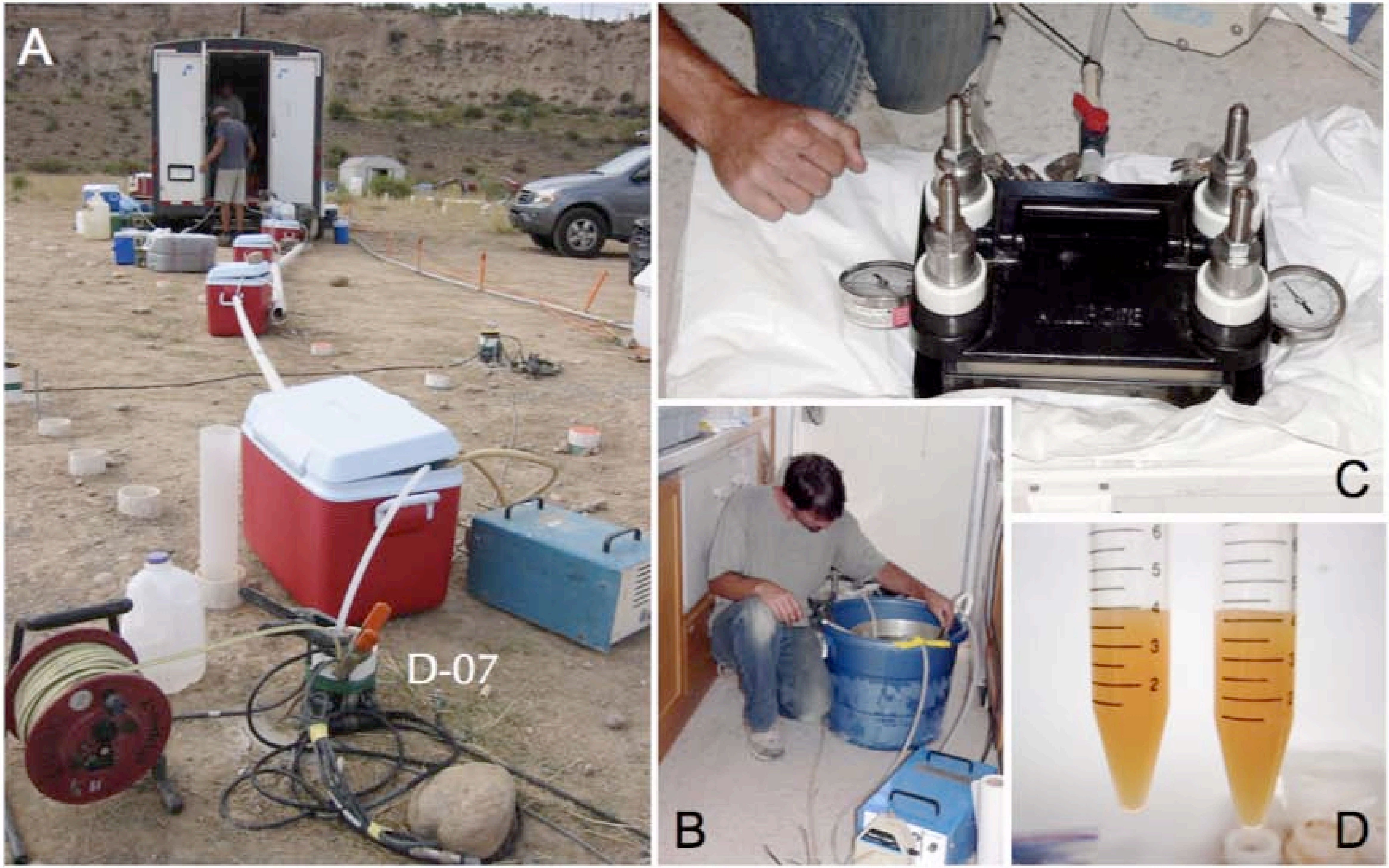
Field biostimulation experiments:
In situ removal of U(VI) from groundwater at Rifle



Stimulated bioreduction at Rifle

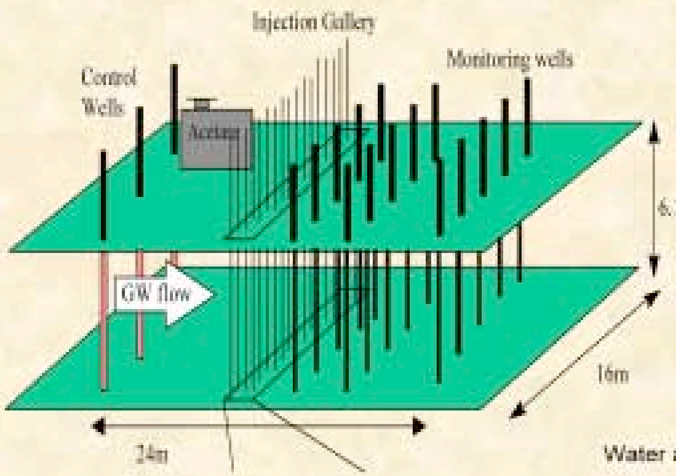


2007 Proteogenomic sampling

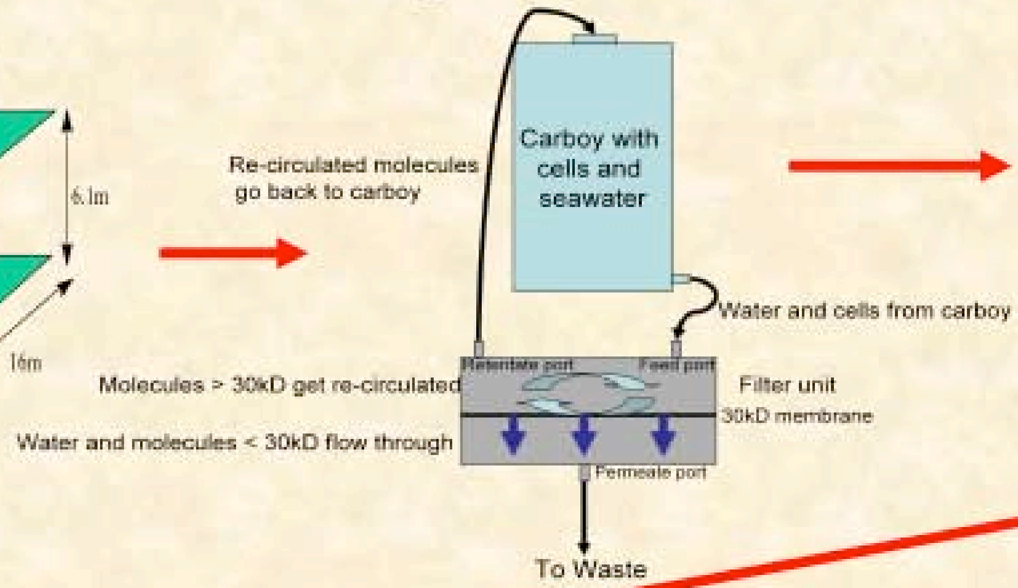


(A) Groundwater, pumped from the surface, passes through chilled ice-buckets on the way to the TFF, (B) Re-circulation bucket in field-trailer, (C) Tangential-flow filtration system, and (D) Concentrated biomass

Initial Proteome Study of the 2007 Gallery TFF enriched ground water samples



Tangential Flow filtration

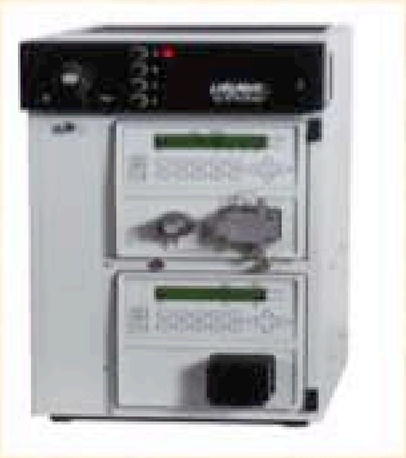


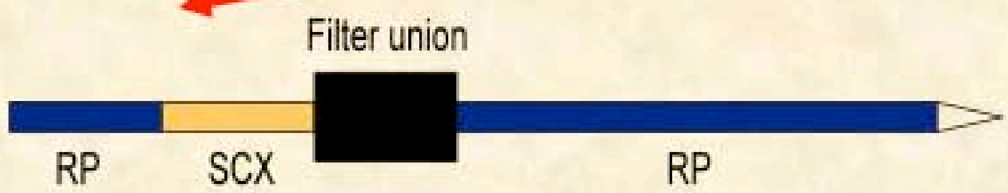
Filtered Cell Pellets

Single Tube Small Sample Processing

1. Cell lyses via 6M Guanidine
2. Protein denaturation, reduction
3. Protein Digestion via Trypsin
4. Peptide de-salting and Concentration

HPLC





Split phase 2D-LC

ES-Linear Ion Trap FT-MS Orbitrap

Three different ground water samples all under iron reducing conditions.

Triplicate 2d-LC-MS/MS analyses of all three samples

Searched against Geobacter isolate database as well as over 800 isolate genomes from JGI

All samples were dominated by Geobacter proteins with 95% of all unique peptides arising from Geobacter proteins.

For each sample ~2000-2500 proteins identified with low false positive rate (~1%)

How to build a "isolate" metagenome from sequenced isolate genomes.

Initial data search completed using all annotated genomes published by JGI.

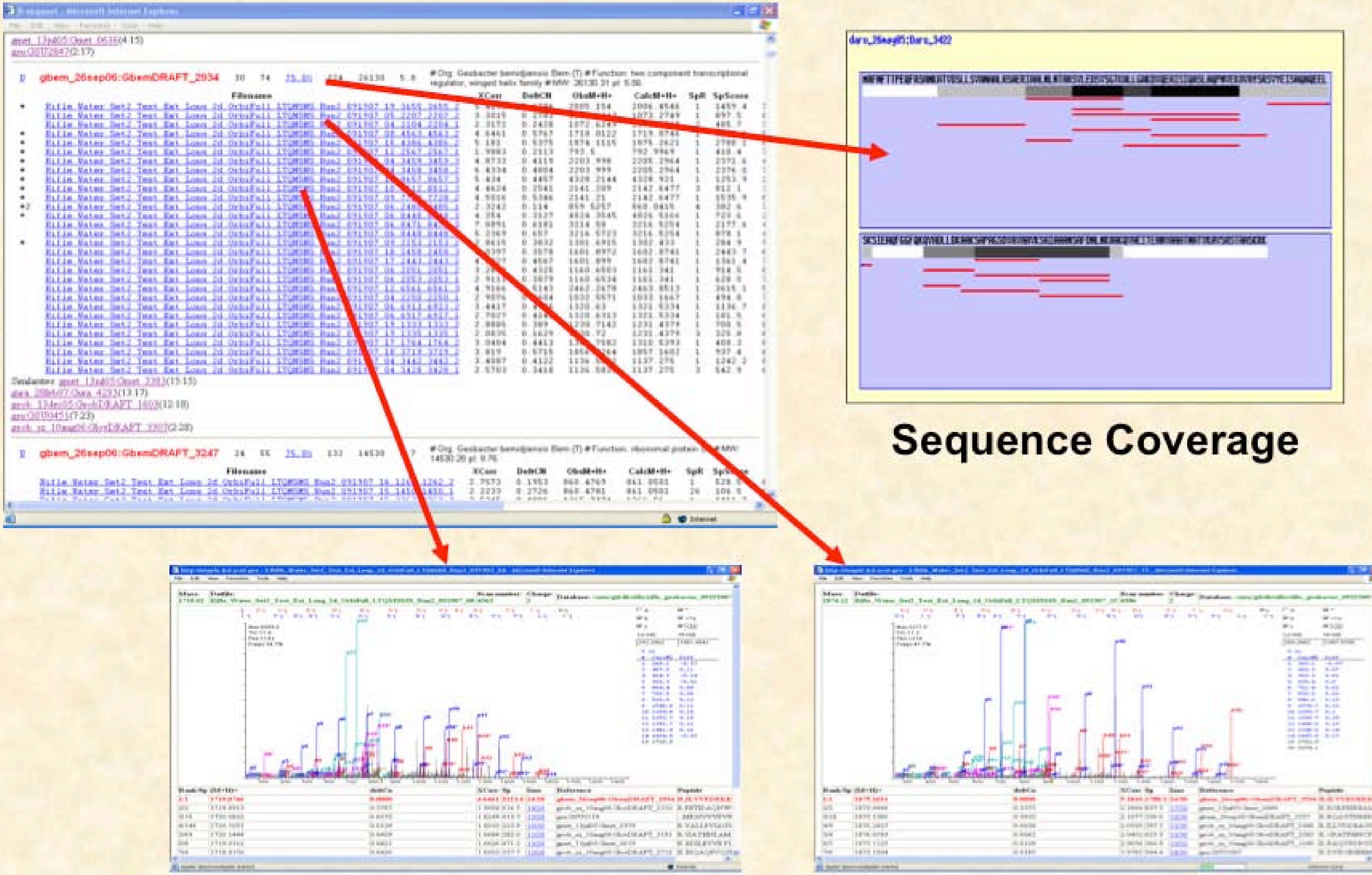
Almost all unique peptides matched *Geobacter* species, other peptides detected were non-unique. e.g. closely related *Pelobacter* species

Subsequent searches used seven *Geobacter* genomes acting as a "mini metagenome":-

G. metallireducens
G. sulfurreducens
G. uraniireducens
G. bemidjiensis
G. M21
G. lovleyi
G. FRC-32

Identified Geobacter protein

Two component transcriptional regulator, winged helix family

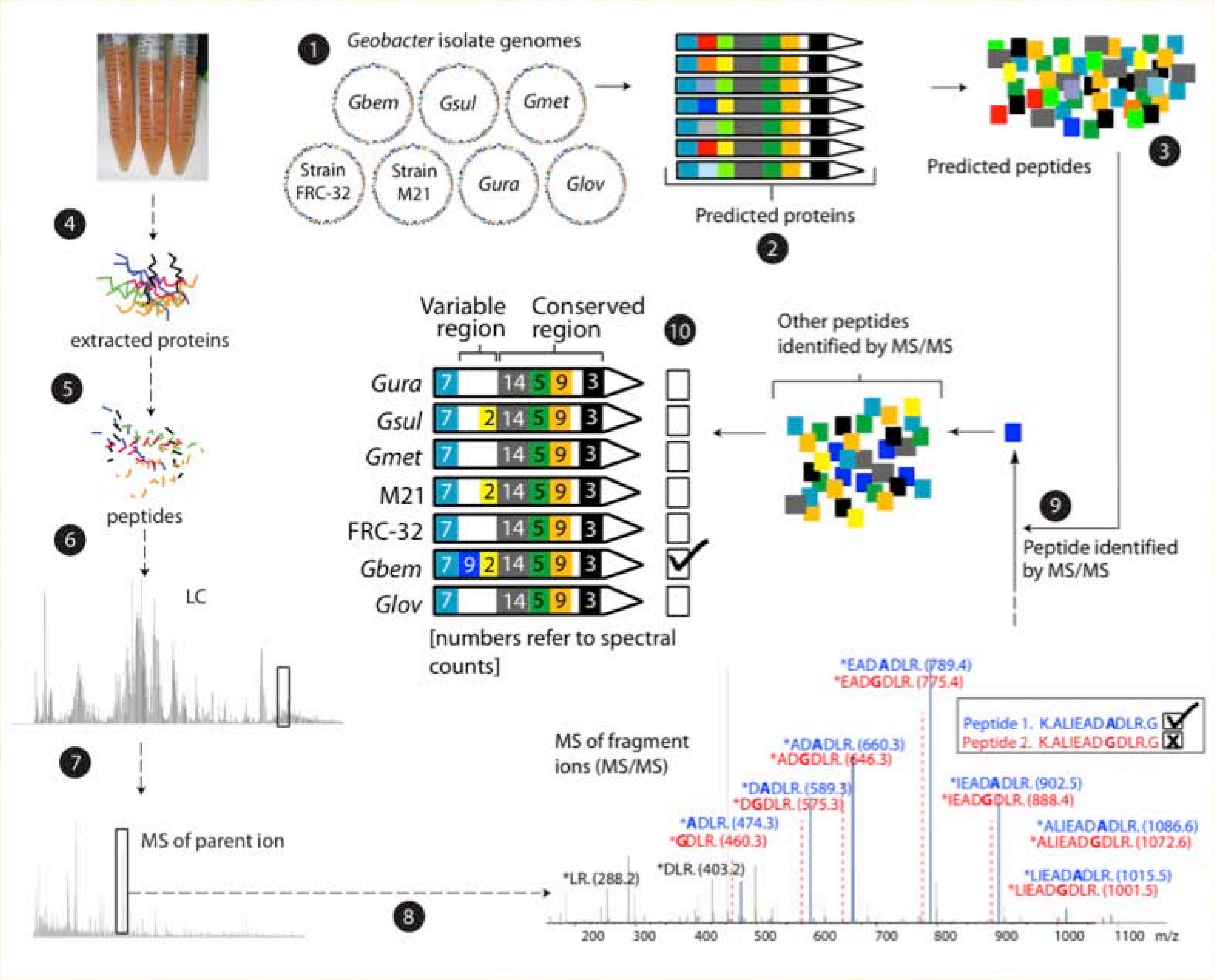


Sequence Coverage

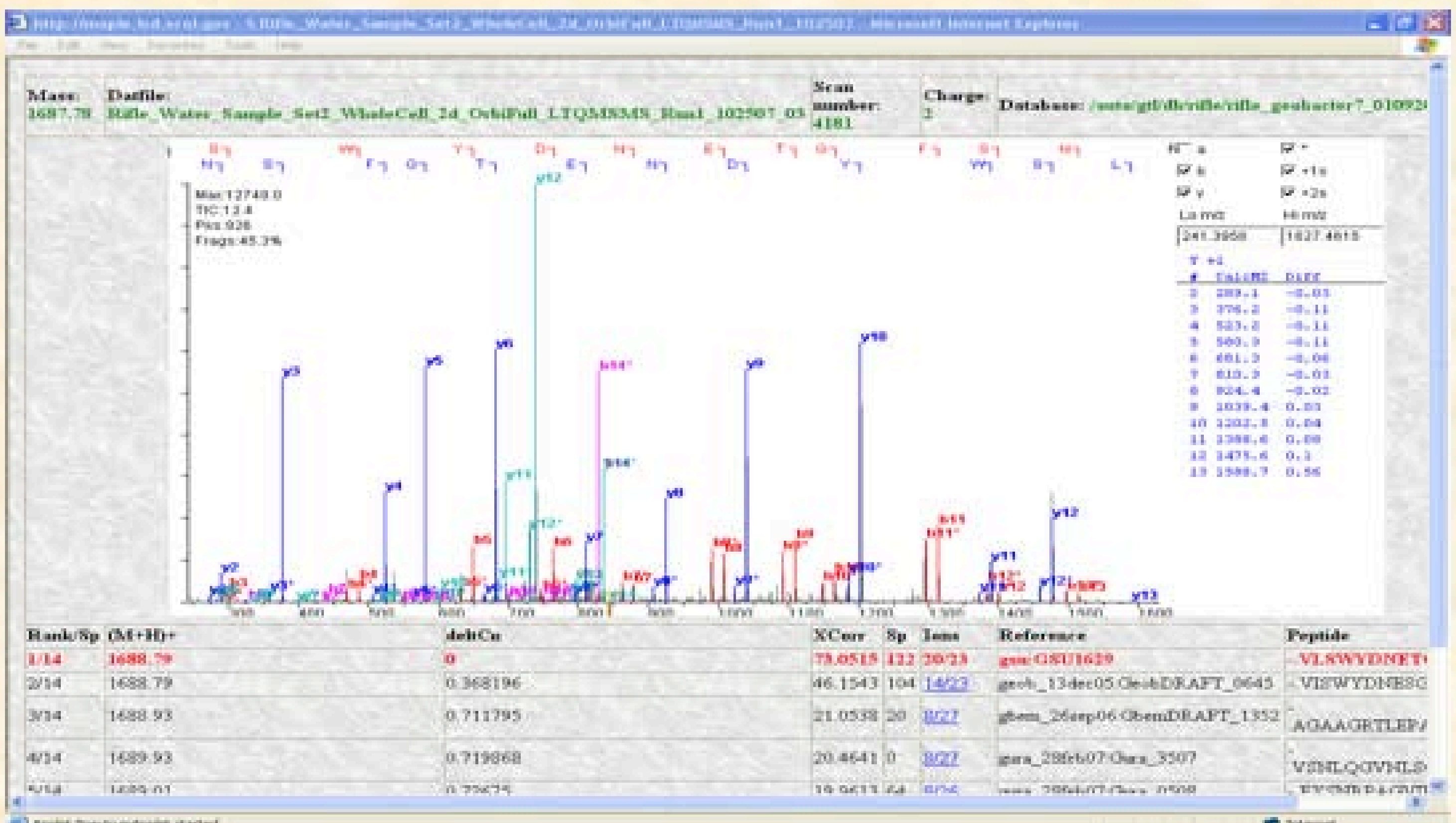
Identified Peptide

Identified Peptide

Measuring strain variation in the Rifle field site



K.VLSWYDNETGFSHR.V
PPM: 1.4739
Delta AMU: 0.0025



K.VLSWYDNETGFSNR.V
PPM: 2.4307
Delta AMU: 0.0041

Protein:
Glyceraldehyde-3-phosphate dehydrogenase

Wilkins_Rifle_Geobacter_AEM_2009_Online_Version.pdf - Adobe Reader

File Edit View Document Tools Window Help

Proteogenomic monitoring of *Geobacter* physiology during stimulated uranium bioremediation

(Proteomic monitoring of *Geobacter* during U reduction)

Michael J. Wilkins[1], Nathan C. VerBerkmoes[2], Kenneth H. Williams[3], Stephen J. Callister[4], Paula J. Mouser[5], Hila Elifantz[5], A. Lucie N'Guessan[5], Brian C. Thomas[1], Carrie D. Nicora[4], Manesh B. Shah[2], Paul Abraham[2,6], Mary S. Lipton[4], Derek R. Lovley[5], Robert L. Hettich[2], Philip E. Long[7], Jillian F. Banfield[*1,3]

[1] Department of Earth and Planetary Science and Environmental Science, Policy, and Management, University of California, Berkeley, CA 94720; [2] Chemical Sciences and Biosciences Divisions, Oak Ridge National Laboratory, Oak Ridge, TN 37831; [3] Earth Science Division, Lawrence Berkeley National Laboratory, Berkeley CA 94720; [4] Biological Sciences Division, Pacific Northwest National Laboratory, Richland WA 99353; [5] Department of Microbiology, University of Massachusetts, Amherst MA 01002; [6] Graduate School of Genome Science and Technology, University of Tennessee, Knoxville, TN, 37830; [7] Energy and Environment Directorate, Pacific Northwest National Laboratory, Richland WA 99353

[*]Corresponding Author: jbanfield@berkeley.edu

Isolate genomes used in search has a large effect on data analysis

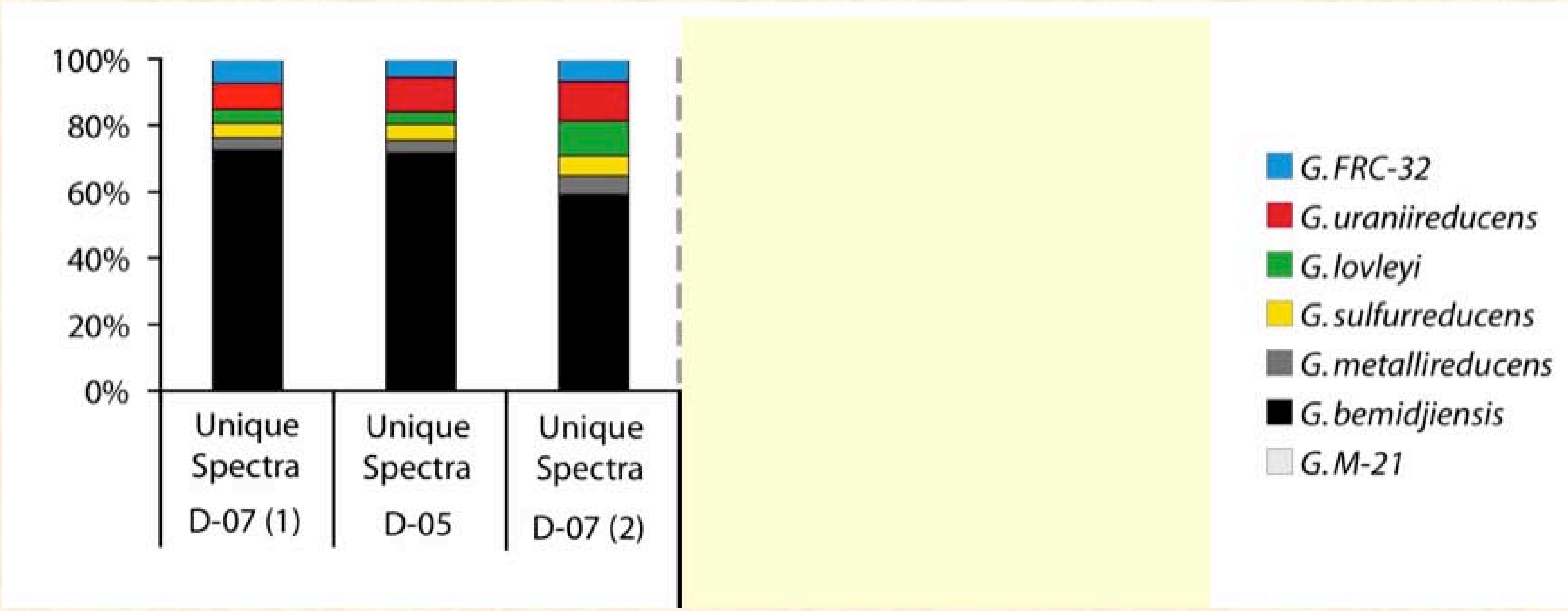
	G. M21	*G. bemidjiensis*	*G. metallireducens*	*G. sulfurreducens*	*G. uraniireducens*	*G. FRC-32*	*G. lovleyi*
G. M21	-	94 %	65 %	64 %	67 %	66 %	60 %
G. bemidjiensis	94 %	-	65 %	64 %	67 %	66 %	60 %
G. metallireducens	65 %	65 %	-	76 %	68 %	67 %	62 %
G. sulfurreducens	64 %	64 %	76 %	-	67 %	66 %	62 %
G. uraniireducens	67 %	67 %	68 %	67 %	-	74 %	61 %
G. FRC-32	66 %	66 %	67 %	66 %	74 %	-	61 %
G. lovleyi	60 %	60 %	62 %	62 %	61 %	61 %	-

Average ortholog % aa similarity

Closely related species (such as *G. bemidjiensis* and strain M21) reduce the number of unique spectra generated by a data search; HOWEVER, they increase the resolution of the analysis – with this analysis, one can now differentiate unique peptides between these two strains.

These low-level strain differences may play an important role in the Rifle subsurface

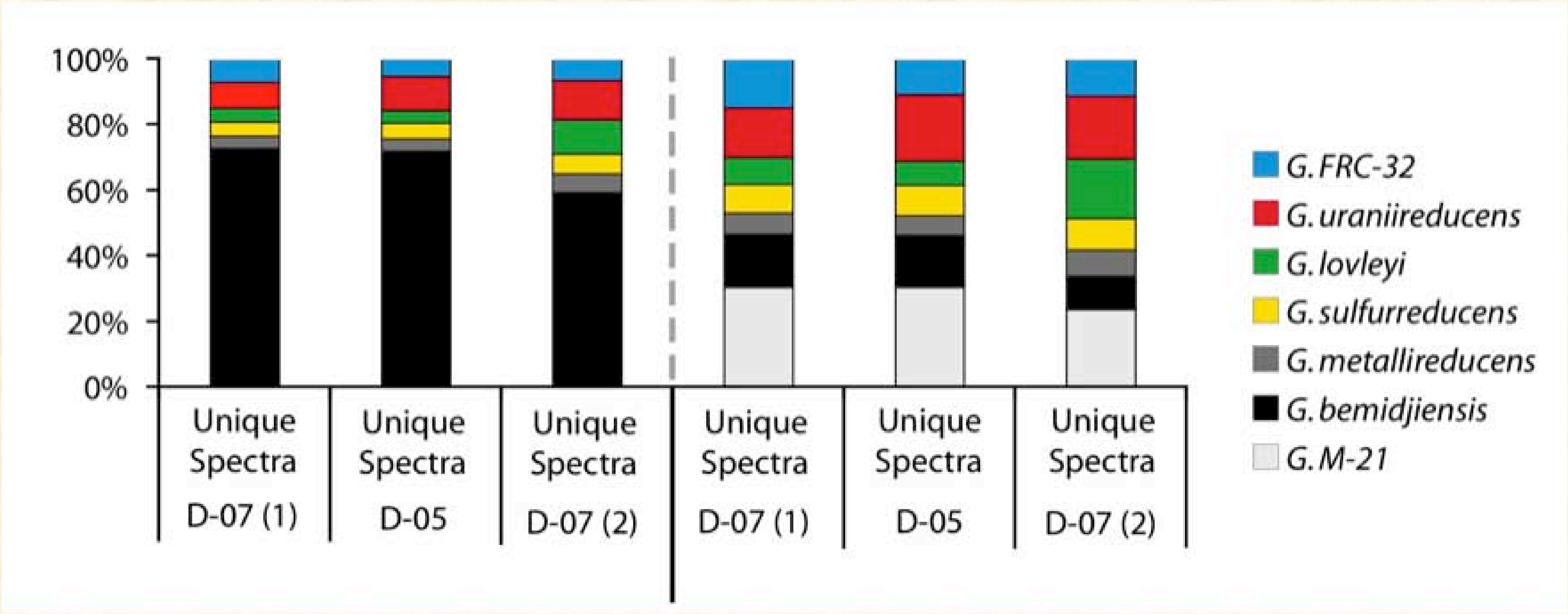
Unique spectral patterns for 2007 samples



Data searched initially using 6 *Geobacter* genomes - *G. M21* not used initially. Evolutionary distance relatively similar between these six strains

In all three samples, unique spectra matching *G. bemidjiensis* dominate the samples. Increase in unique spectra from other strains in D-07(2); the greatest increase is in unique spectra matching *G. lovleyi*

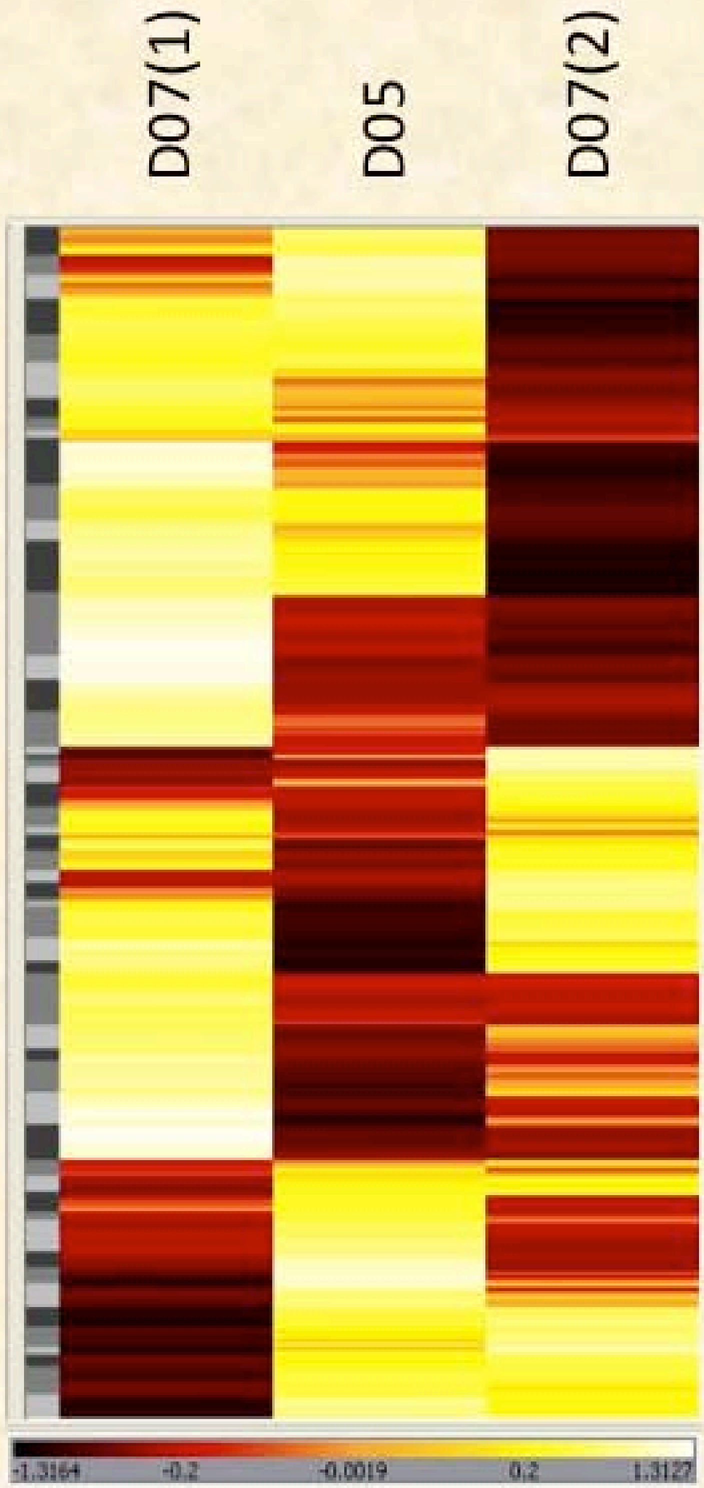
Unique spectral patterns for 2007 samples



Addition of closely related *G. M21* greatly reduces number of *G. bemidjiensis* unique spectra. However, increases in unique spectra matching G. lovleyi are still occurring between samples D07 (1) and D07(2)

Are communities dominated by one strain / species or several?

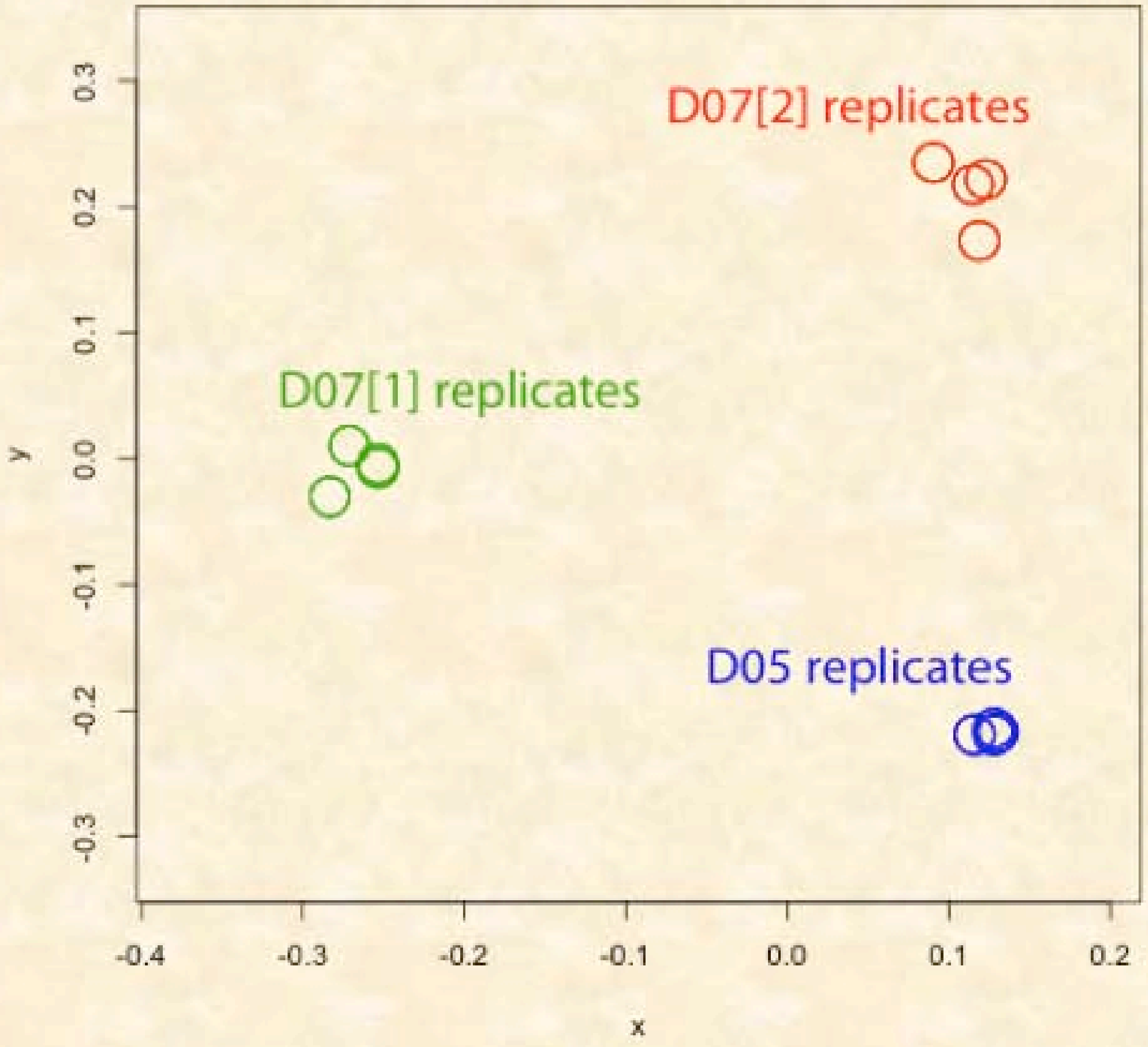
Which isolate is the dominant strain / species most closely related to? Greatest % of unique spectra attributed to *Geobacter* strain M21, a bacteria that was isolated from the Rifle site during biostimulation.



Heat map illustrates similarities and differences between the three samples recovered from the Rifle groundwater

Greatest differences appear between the two samples from well D07. Temporal scale potentially has a greater effect over the spatial scale (D07[1] vs D05) in this instance

Heat map generated using averaged Z scores generated from NSAF protein abundance values (log transformed, and mean centered; presence of peptides in two of three replicates necessary)

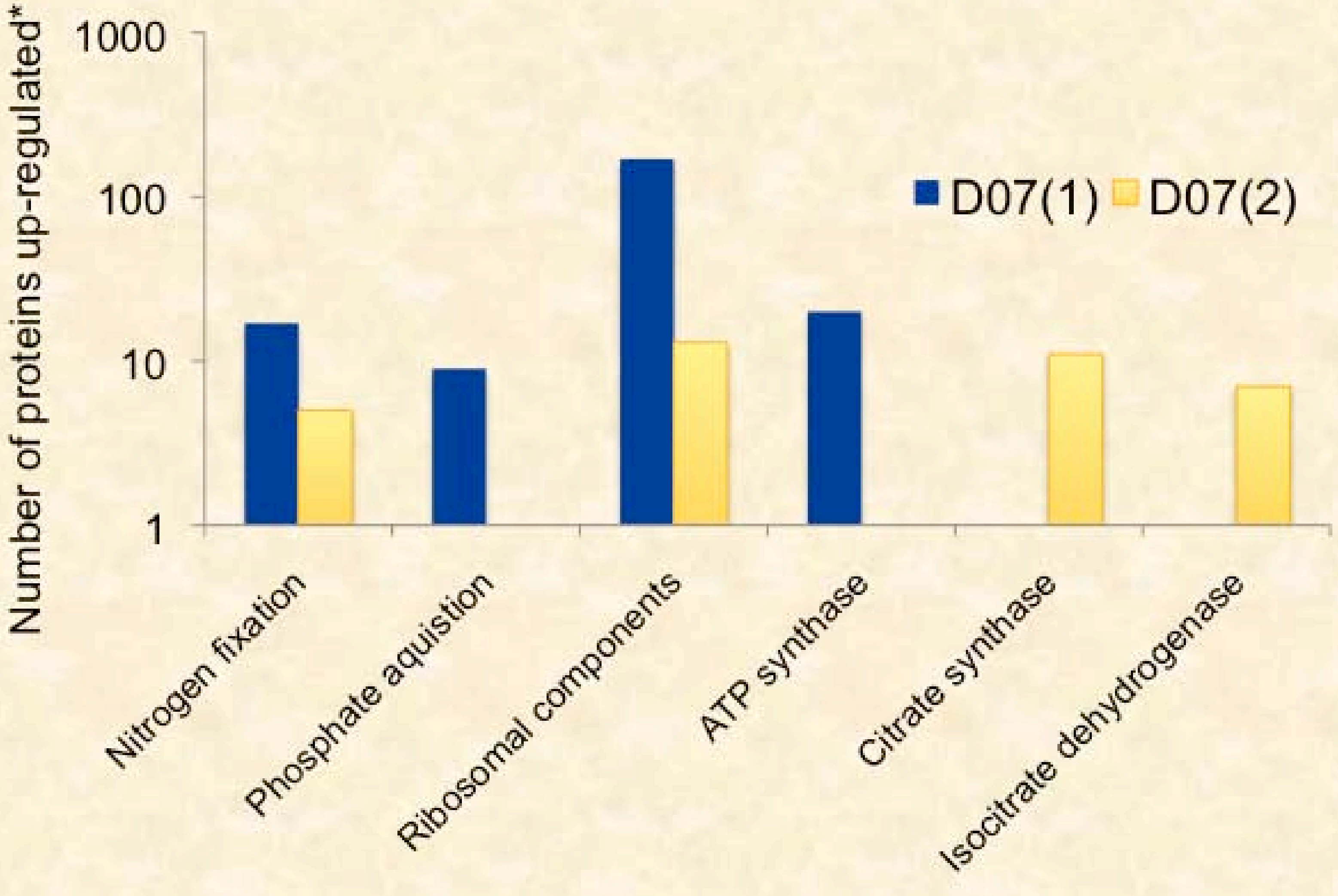


The visual differences seen in the heat map are replicated when the technical replicates for each sample are clustered.

Three technical replicates for each of the three samples were analyzed using non-metric multidimensional scaling.

Clustering patterns indicate that all three samples are different at the protein abundance level.
Tight clustering between technical replicates indicates successful reproducibility between LC-MS/MS runs

Protein abundance trends

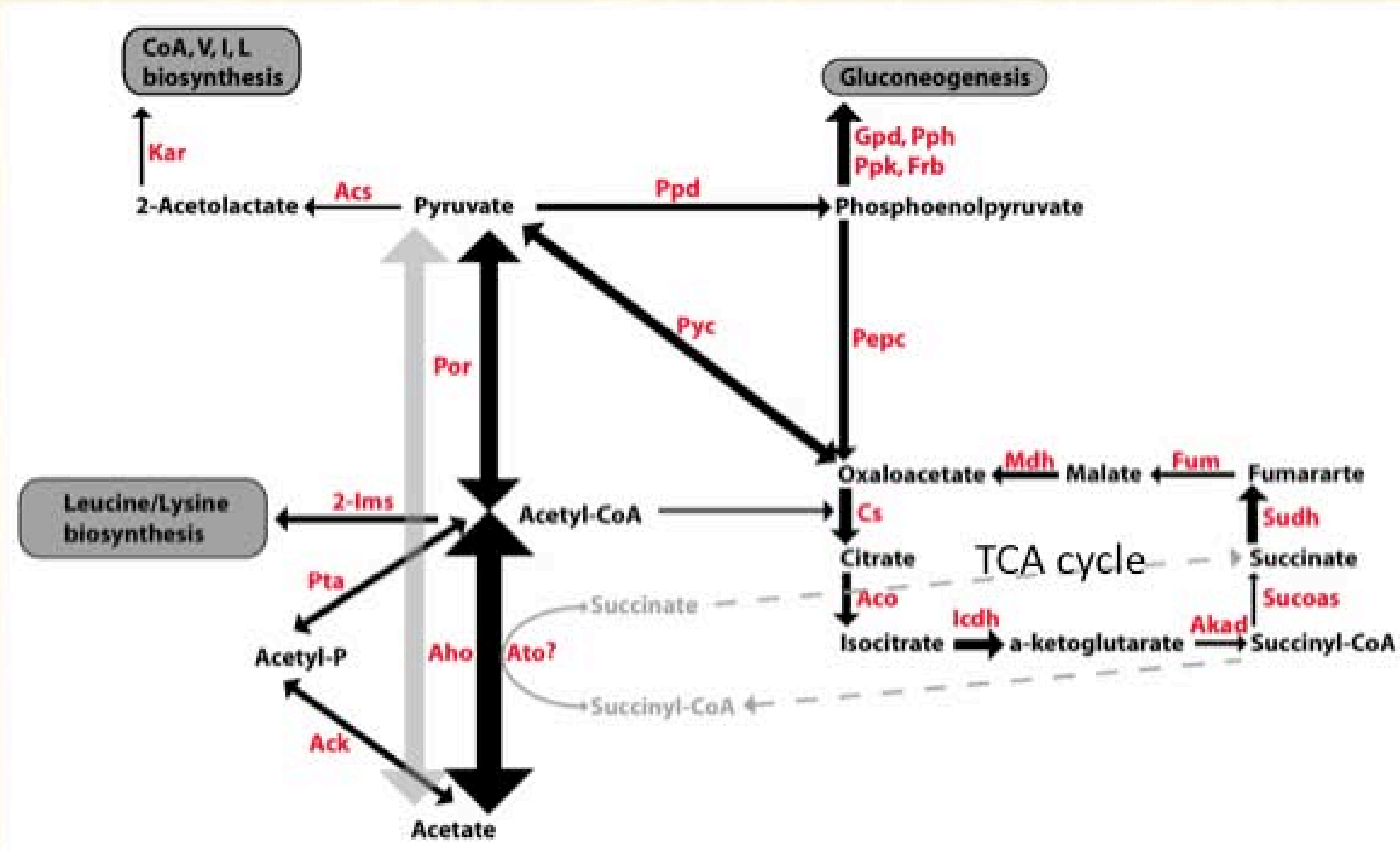


Evidence for slowing cell growth later during biostimulation; decreasing abundances of proteins associated with growth for *Geobacter* strains (ribosomal proteins, ATP synthase subunits). Also evidence for less phosphate nutrient stress later during biostimulation (nitrogenase, phosphate acquisition).

However, increases in abundance of some TCA cycle proteins later during biostimulation. Not associated with cellular growth. Associated with an as-yet undetermined energy-requiring process.

* Note the log scale used on graph.

Geobacter physiology ; proteomic data was used to piece together the mechanism by which these species are able to efficiently utilize acetate

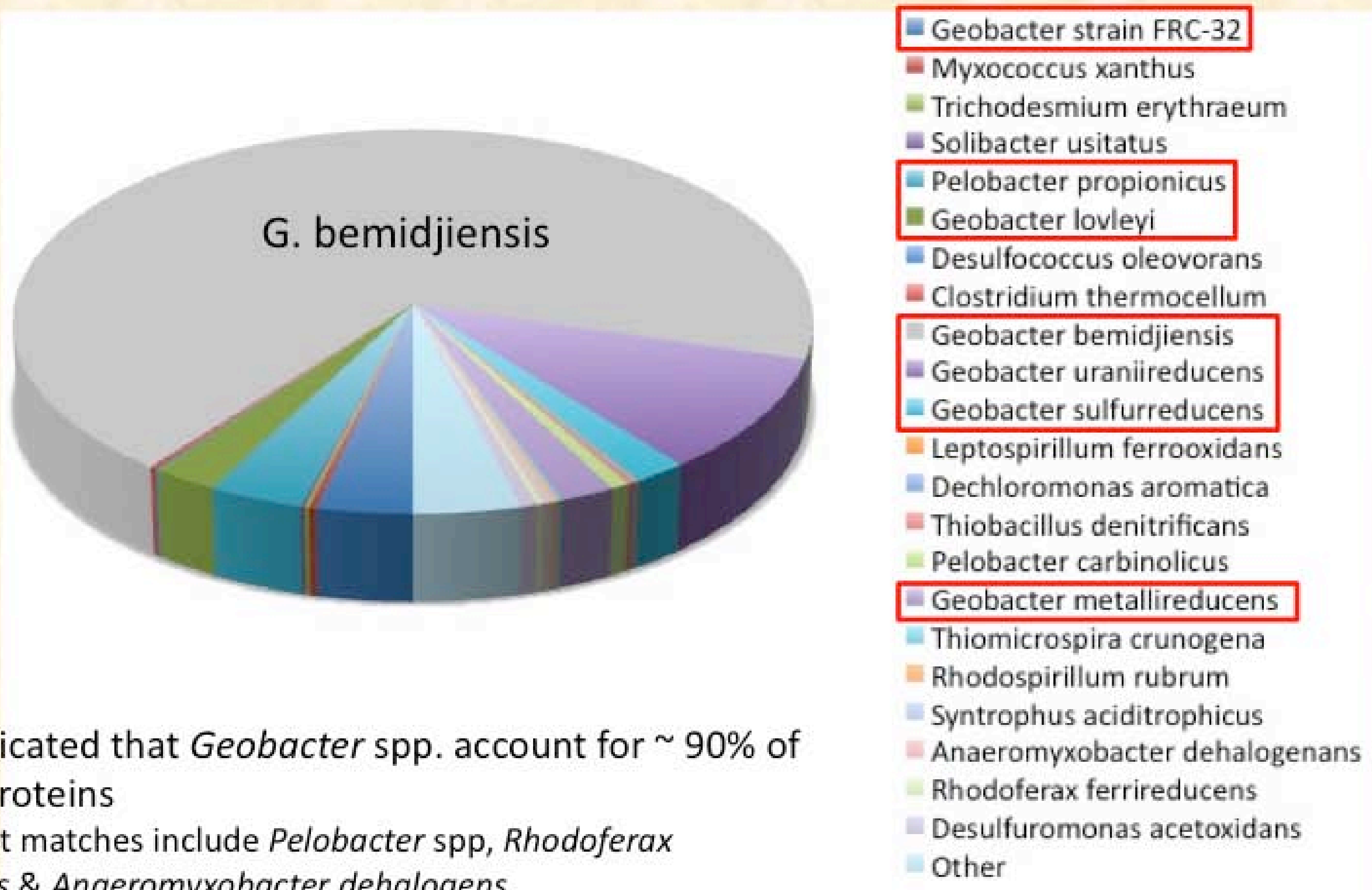


2-Ims - 2-isopropylmalate synthase
Acs - acetolactate synthase
Frb - fructose bisphosphatase
Gpd - glyceraldehyde-3-phosphate dehydrogenase
Kar - ketol acid reductoisomerase
Pepc - phosphoenolpyruvate carboxykinase
Por - pyruvate ferredoxin oxioreductase
Ppd - pyruvate phosphate dikinase
Pph - phosphopyruvate hydratase
Ppk - phosphoglycerate kinase
Pta - phosphotranscetylase
Pyc - pyruvate carboxylase

There are multiple pathways for these species to channel acetate into the TCA cycle for energy generation.

Constraint-based modeling indicates the pathway catalyzed by the POR enzyme (conversion between acetyl-CoA and pyruvate) allows *G. sulfurreducens* to synthesize amino acids more efficiently than *Escherichia coli*

Following the analysis of the 2007 data using isolate *Geobacter* genomes, a metagenome for one sample, D05, became available. The D05 proteomic sample was searched against this metagenome. The metagenomic reads that matched proteomic data were then BlastP searched against the NCBI database to determine what species these reads were most closely related to.



Results indicated that *Geobacter* spp. account for ~ 90% of detected proteins

Other closest matches include *Pelobacter* spp, *Rhodoferax ferrireducens* & *Anaeromyxobacter dehalogens*

Use of metagenomic sequence doubled the number of proteins detected

- From approximately 2000 per sample, to 4000 per sample

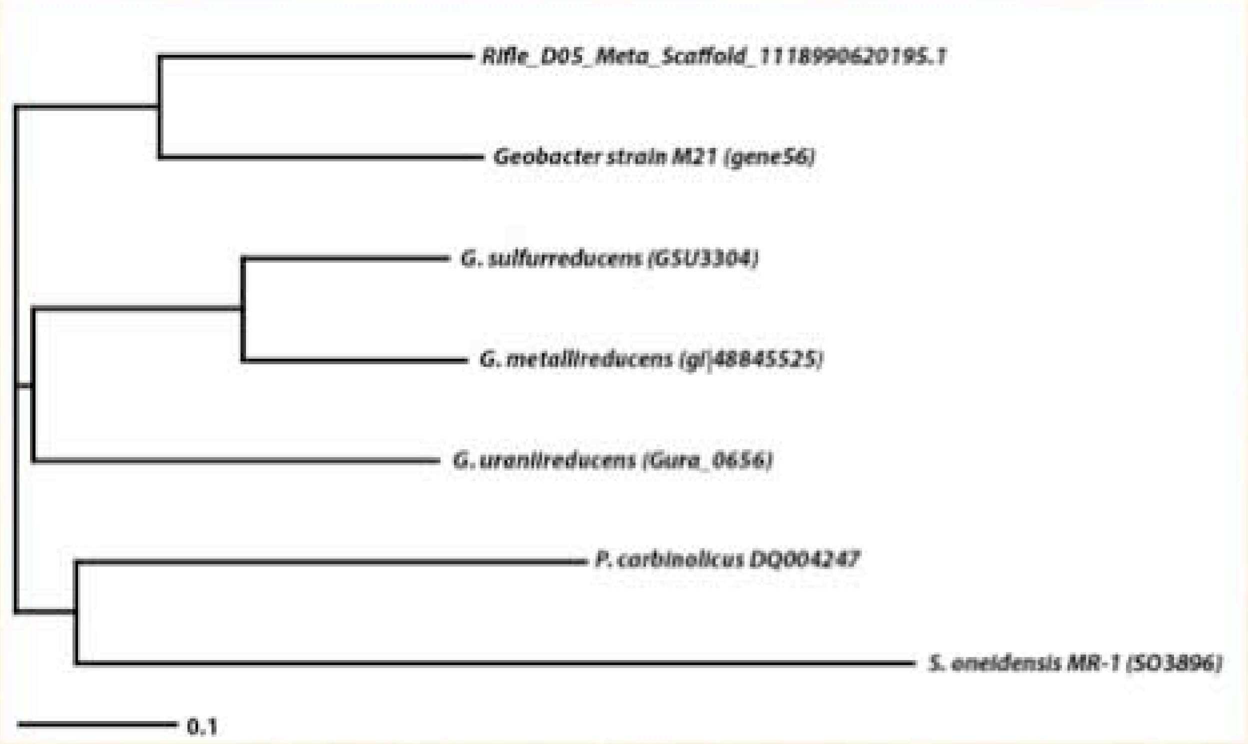
Many proteins had sequence too dissimilar to *Geobacter* isolate sequences – therefore these proteins were only detected when metagenomic sequence added to search database.

Example: OmpJ-like protein, shown to be essential for Fe(III)-reduction in pure culture lab studies.

OmpJ-like protein detected in all three samples. Most abundance OM-associated protein

Low sequence similarity to homologs in sequenced *Geobacter* genomes, hence only detected when metagenomic sequence added into search database.

Best BlastP match = *Geobacter* strain M21, *G. bemidjiensis* [~ 58%]



2008 Rifle proteomic samples ground water Gallery

Well D04 Geobacter Protein Totals and Trypsin Spectral Counts

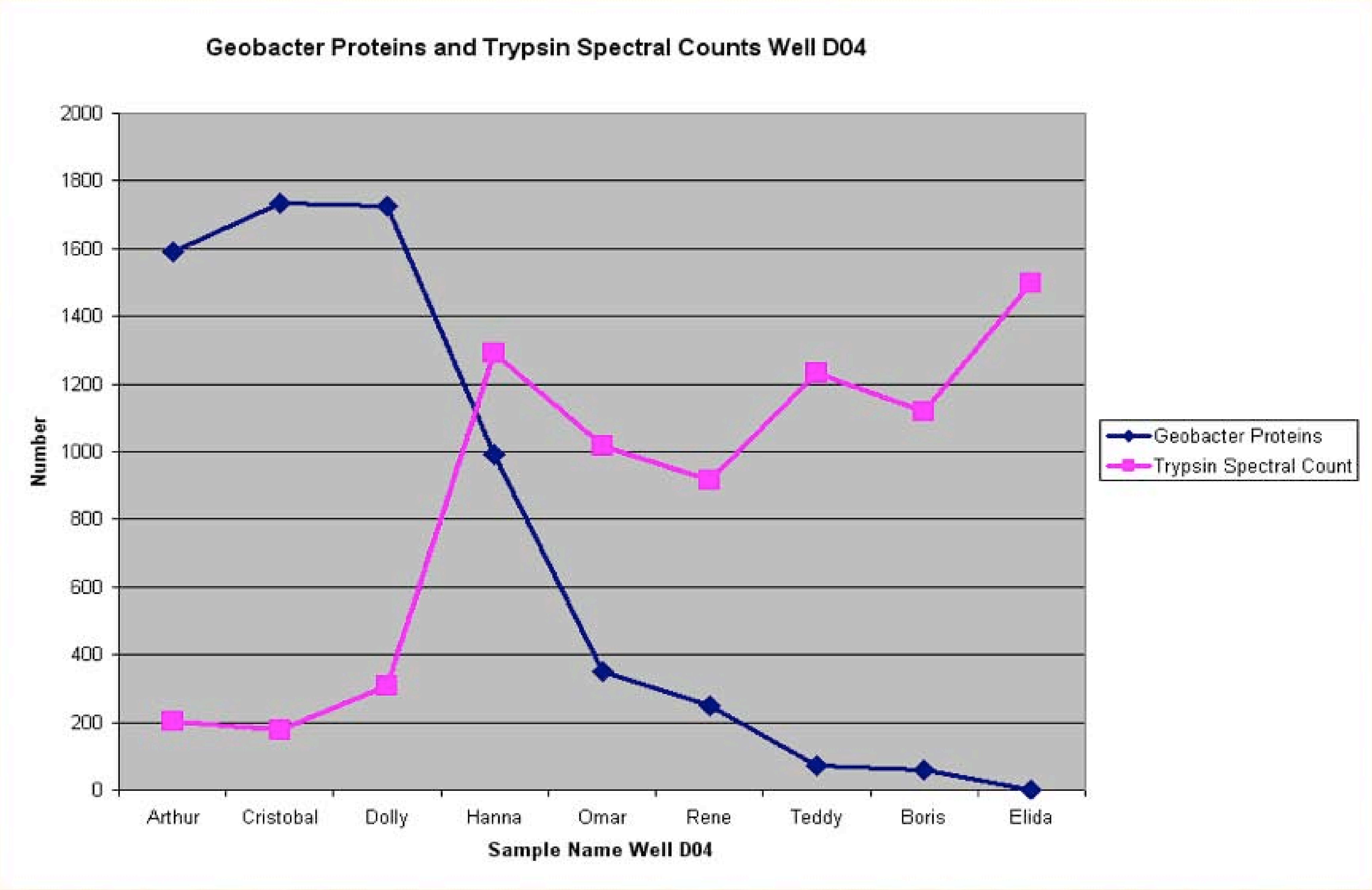
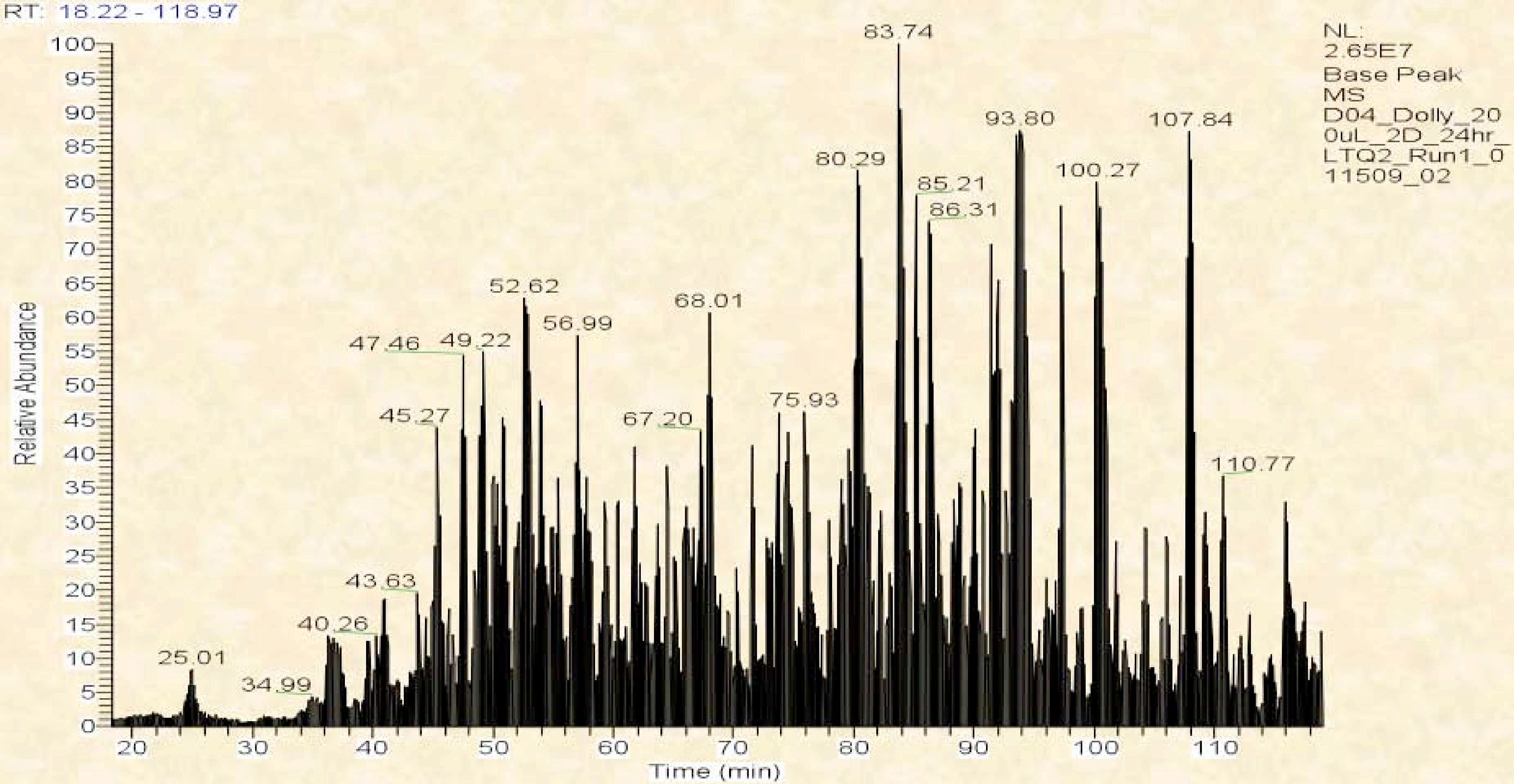


Figure 15: Top ~70 proteins from the D04 Well Geobacter bloom. High reproducibility of expression and overall abundance

Protein	Arthur_200ul_Run1	Cristobal_LTQ_Run1	Cristobal_Orbi_Run1	Dolly_LTQ_Run1	Dolly_Orbi_Run1	Hanna_Orbi_Run1	Average	Description
gbem_26sep06:GbemDRAFT_3222	626	403	476	342	361	269	412.83	# Org: Geobacter bemidjiensis Bem (T) # Function: translation elongation factor Tu # MW: 42996.43 pI: 5.20
cont_gi|136429|sp|P00761|TRYP_PIG	204	176	177	306	309	1290	410.33	Trypsin precursor
gbem_26sep06:GbemDRAFT_3233	592	395	439	317	349	249	390.17	# Org: Geobacter bemidjiensis Bem (T) # Function: translation elongation factor Tu # MW: 43026.40 pI: 5.19
gm|829141|gene2605	530	304	342	247	274	217	319.00	# Org: Geobacter M21 # Function: GTPases - translation elongation factors # MW:28675.00 pI:4.74
gura_28feb07:Gura_1053	298	330	259	241	242	232	267.00	# Org: Geobacter uraniumreducens RF4 # Function: translation elongation factor Tu # MW: 42998.44 pI: 5.25
gura_28feb07:Gura_1066	298	330	259	241	242	232	267.00	# Org: Geobacter uraniumreducens RF4 # Function: translation elongation factor Tu # MW: 42998.44 pI: 5.25
gbem_26sep06:GbemDRAFT_0731	338	282	281	258	292	127	263.00	# Org: Geobacter bemidjiensis Bem (T) # Function: ATP synthase F1, beta subunit # MW: 51057.59 pI: 4.92
gm|829113|gene3309	319	261	265	241	275	125	247.67	# Org: Geobacter M21 # Function: F0F1-type ATP synthase, beta subunit # MW:51015.50 pI:4.92
gm|829798|gene391	272	272	215	246	196	172	228.83	# Org: Geobacter M21 # Function: Chaperonin GroEL (HSP60 family) # MW:58374.72 pI:5.14
geob_13dec05:GeobDRAFT_0688	264	286	220	208	207	167	225.33	# Org: Geobacter sp. FRC-32 # Function: translation elongation factor Tu Small GTP-binding protein domain # MW: 431
gura_28feb07:Gura_4307	229	250	184	238	164	167	205.33	# Org: Geobacter uraniumreducens RF4 # Function: chaperonin GroEL # MW: 58631.66 pI: 5.10
gbem_26sep06:GbemDRAFT_2454	311	195	213	147	191	164	203.50	# Org: Geobacter bemidjiensis Bem (T) # Function: histone family protein DNA-binding protein # MW: 9779.45 pI:10.42
gm|829365|gene3580	311	195	213	147	191	164	203.50	# Org: Geobacter M21 # Function: Bacterial nucleoid DNA-binding protein # MW:9779.45 pI:10.42
geob_13dec05:GeobDRAFT_0219	248	214	201	201	240	86	198.33	# Org: Geobacter sp. FRC-32 # Function: ATP synthase F1, beta subunit # MW: 51033.40 pI: 4.89
geob_13dec05:GeobDRAFT_1140	227	240	179	218	154	161	196.50	# Org: Geobacter sp. FRC-32 # Function: chaperonin GroEL # MW: 58381.51 pI: 5.14
gmet_13jul05:Gmet_3406	248	206	188	193	218	86	189.83	# Org: Geobacter metallireducens GS-15 # Function: ATP synthase F1, beta subunit # MW: 51033.38 pI: 4.97
gsu:GSU0113	247	206	188	190	223	86	189.83	# Org: Geobacter sulfurreducens # Function: atpD; ATP synthase F1, beta subunit [EC:3.6.3.14] [KO:K02112] # MW: 5
gmet_13jul05:Gmet_0611	199	218	164	156	163	179	179.83	# Org: Geobacter metallireducens GS-15 # Function: Translation elongation factor Tu # MW: 43207.47 pI: 5.20
gmet_13jul05:Gmet_0624	199	218	164	156	163	179	179.83	# Org: Geobacter metallireducens GS-15 # Function: Translation elongation factor Tu # MW: 43207.47 pI: 5.20
gbem_26sep06:GbemDRAFT_0457	262	190	155	212	163	60	173.67	# Org: Geobacter bemidjiensis Bem (T) # Function: Citrate (Si)-synthase # MW: 49331.72 pI: 6.86
geob_sz_10aug06:GlovDRAFT_1965	229	191	178	156	193	77	170.67	# Org: Geobacter lovleyi SZ # Function: ATP synthase F1, beta subunit # MW: 51139.51 pI: 4.98
gsu:GSU3340	190	213	153	197	133	128	169.00	# Org: Geobacter sulfurreducens # Function: groeL; 60 kDa chaperonin [KO:K04077] # MW: 58149.16 pI: 5.19
gmet_13jul05:Gmet_0029	195	220	166	192	134	116	168.83	# Org: Geobacter metallireducens GS-15 # Function: chaperonin GroEL # MW: 58014.98 pI: 5.11
geob_sz_10aug06:GlovDRAFT_1700	193	208	164	156	157	115	165.50	# Org: Geobacter lovleyi SZ # Function: translation elongation factor Tu # MW: 42983.18 pI: 5.18
gbem_26sep06:GbemDRAFT_1902	283	174	195	112	137	42	157.17	# Org: Geobacter bemidjiensis Bem (T) # Function: thiamine pyrophosphate enzyme domain protein TPP-binding # MW
gm|826531|gene1422	223	171	135	202	149	60	156.67	# Org: Geobacter M21 # Function: Citrate synthase # MW:49341.77 pI:6.86
gm|828923|gene2041	280	169	194	110	137	42	155.33	# Org: Geobacter M21 # Function: Thiamine pyrophosphate-requiring enzymes [acetolactate synthase, pyruvate dehydro
gura_28feb07:Gura_4407	253	136	171	94	150	127	155.00	# Org: Geobacter uraniumreducens RF4 # Function: histone family protein DNA-binding protein # MW: 9714.41 pI:10.60
gm|828918|gene2037	252	151	171	148	122	75	153.17	# Org: Geobacter M21 # Function: Monomeric isocitrate dehydrogenase # MW:81463.43 pI:6.23
gm|829595|gene135	242	144	259	115	139	19	153.00	# Org: Geobacter M21 # Function: Acetyl-CoA hydrolase # MW:58401.29 pI:6.13
gbem_26sep06:GbemDRAFT_2114	243	142	257	114	141	20	152.83	# Org: Geobacter bemidjiensis Bem (T) # Function: Acetyl-CoA hydrolase # MW: 58404.35 pI: 6.23
gsu:GSU2859	169	197	139	138	141	107	148.50	# Org: Geobacter sulfurreducens # Function: tuf-1; translation elongation factor Tu [EC:3.6.5.3] [KO:K02358] # MW: 432
gsu:GSU2871	169	197	139	138	141	107	148.50	# Org: Geobacter sulfurreducens # Function: tuf-2; translation elongation factor Tu [EC:3.6.5.3] [KO:K02358] # MW: 432
gura_28feb07:Gura_2190	266	158	176	102	126	44	145.33	# Org: Geobacter uraniumreducens RF4 # Function: thiamine pyrophosphate enzyme domain protein # MW: 29329.08 p
geob_13dec05:GeobDRAFT_0226	255	151	180	94	123	40	140.50	# Org: Geobacter sp. FRC-32 # Function: Thiamine pyrophosphate enzyme, C-terminal TPP-binding # MW: 29329.13 pI:
gmet_13jul05:Gmet_1363	249	151	168	87	125	51	138.50	# Org: Geobacter metallireducens GS-15 # Function: Thiamine pyrophosphate enzyme, C-terminal # MW: 29447.36 pI: 8
gbem_26sep06:GbemDRAFT_1901	203	115	217	113	120	57	137.50	# Org: Geobacter bemidjiensis Bem (T) # Function: pyruvate ferredoxin/flavodoxin oxidoreductase # MW: 19242.23 pI: 5.
gm|828924|gene2042	203	115	217	113	120	57	137.50	# Org: Geobacter M21 # Function: Pyruvate ferredoxin oxidoreductase and related 2-oxoacid ferredoxin oxidoreductases
geob_sz_10aug06:GlovDRAFT_2458	244	142	163	85	125	51	135.00	# Org: Geobacter lovleyi SZ # Function: thiamine pyrophosphate enzyme-like TPP-binding # MW: 29386.25 pI: 8.89
gura_28feb07:Gura_4263	163	144	134	140	168	58	134.50	# Org: Geobacter uraniumreducens RF4 # Function: ATP synthase F1, beta subunit # MW: 50848.22 pI: 4.88
gura_28feb07:Gura_4029	199	107	229	86	117	17	125.83	# Org: Geobacter uraniumreducens RF4 # Function: Acetyl-CoA hydrolase # MW: 58323.09 pI: 6.13
gbem_26sep06:GbemDRAFT_3287	157	129	176	98	115	47	120.33	# Org: Geobacter bemidjiensis Bem (T) # Function: Phosphopyruvate hydratase # MW: 46491.98 pI: 4.68
gm|830021|gene2528	157	129	176	98	115	47	120.33	# Org: Geobacter M21 # Function: Enolase # MW:46491.98 pI:4.68
gsu:GSU0490	186	107	224	79	108	14	119.67	# Org: Geobacter sulfurreducens # Function: acetyl-CoA hydrolase/transferase family protein # MW: 58603.38 pI: 6.13
geob_13dec05:GeobDRAFT_3394	130	124	135	114	108	88	116.50	# Org: Geobacter sp. FRC-32 # Function: Small GTP-binding protein domain # MW: 24985.58 pI: 4.88
gm|827001|gene3816	182	102	213	73	105	18	115.50	# Org: Geobacter M21 # Function: Acetyl-CoA hydrolase # MW:60030.09 pI:6.22
gbem_26sep06:GbemDRAFT_3229	167	125	115	139	89	54	114.83	# Org: Geobacter bemidjiensis Bem (T) # Function: DNA-directed RNA polymerase, beta' subunit # MW:153860.91 pI: 6
geob_sz_10aug06:GlovDRAFT_2454	173	107	146	105	94	58	113.83	# Org: Geobacter lovleyi SZ # Function: isocitrate dehydrogenase, NADP-dependent # MW: 81837.87 pI: 6.01
gmet_13jul05:Gmet_3044	165	102	212	76	106	15	112.67	# Org: Geobacter metallireducens GS-15 # Function: Acetyl-CoA hydrolase/transferase # MW: 58493.42 pI: 6.08
gm|829130|gene2596	162	123	115	137	85	54	112.67	# Org: Geobacter M21 # Function: DNA-directed RNA polymerase, beta' subunit/160 kD subunit # MW:153897.94 pI:6.2
gura_28feb07:Gura_2194	171	102	147	103	101	52	112.67	# Org: Geobacter uraniumreducens RF4 # Function: isocitrate dehydrogenase, NADP-dependent # MW: 82067.28 pI: 6.
gbem_26sep06:GbemDRAFT_1590	170	101	209	70	104	18	112.00	# Org: Geobacter bemidjiensis Bem (T) # Function: Acetyl-CoA hydrolase # MW: 59917.01 pI: 6.22
gbem_26sep06:GbemDRAFT_3228	166	112	107	112	108	60	110.83	# Org: Geobacter bemidjiensis Bem (T) # Function: DNA-directed RNA polymerase, beta subunit # MW:152913.47 pI: 5.
gsu:GSU1469	202	109	134	57	96	36	105.67	# Org: Geobacter sulfurreducens # Function: keto/oxoacid ferredoxin oxidoreductase, beta subunit [EC:1.2.7.3] [KO:K00
gura_28feb07:Gura_2189	149	92	184	86	87	30	104.67	# Org: Geobacter uraniumreducens RF4 # Function: pyruvate ferredoxin/flavodoxin oxidoreductase # MW: 19079.97 pI: 5
geob_sz_10aug06:GlovDRAFT_0147	128	114	82	111	93	95	103.83	# Org: Geobacter lovleyi SZ # Function: chaperonin GroEL # MW: 58621.56 pI: 5.11
geob_13dec05:GeobDRAFT_2091	144	111	77	135	101	52	103.33	# Org: Geobacter sp. FRC-32 # Function: Citrate (Si)-synthase # MW: 49221.45 pI: 6.84
geob_13dec05:GeobDRAFT_0227	146	91	180	82	86	31	102.67	# Org: Geobacter sp. FRC-32 # Function: 2-oxoglutarate synthase # MW: 19122.01 pI: 5.88
gm|828919|gene2038	152	114	96	96	104	42	102.33	# Org: Geobacter M21 # Function: Malate/lactate dehydrogenases # MW:28126.97 pI:9.49
gbem_26sep06:GbemDRAFT_0733	133	118	74	124	82	71	100.33	# Org: Geobacter bemidjiensis Bem (T) # Function: ATP synthase F1, alpha subunit # MW: 54106.58 pI: 5.55
gm|829116|gene3311	133	118	74	124	82	71	100.33	# Org: Geobacter M21 # Function: F0F1-type ATP synthase, beta subunit # MW:54120.61 pI:5.56
gura_28feb07:Gura_3711	72	81	156	79	142	71	100.17	# Org: Geobacter uraniumreducens RF4 # Function: putative cold-shock DNA-binding domain protein # MW: 7108.98 pI:
gsu:GSU1470	141	86	178	78	83	25	98.50	# Org: Geobacter sulfurreducens # Function: keto/oxoacid ferredoxin oxidoreductase, gamma subunit [EC:1.2.7.3] [KO:
geob_13dec05:GeobDRAFT_1862	129	110	169	75	89	28	98.33	# Org: Geobacter sp. FRC-32 # Function: enolase # MW: 46655.31 pI: 4.80
gbem_26sep06:GbemDRAFT_1903	96	155	77	122	79	45	95.50	# Org: Geobacter bemidjiensis Bem (T) # Function: pyruvate flavodoxin/ferredoxin oxidoreductase domain protein # MW:
gm|828922|gene2040	94	154	77	121	79	45	95.00	# Org: Geobacter M21 # Function: Pyruvate ferredoxin oxidoreductase and related 2-oxoacid ferredoxin oxidoreductases
gm|827071|gene3884	153	93	123	74	79	40	93.67	# Org: Geobacter M21 # Function: Aspartate oxidase # MW:70821.16 pI:8.12
gbem_26sep06:GbemDRAFT_2603	152	93	123	74	79	40	93.50	# Org: Geobacter bemidjiensis Bem (T) # Function: succinate dehydrogenase or fumarate reductase, flavoprotein subuni
gura_28feb07:Gura_1780	116	103	148	74	85	24	91.67	# Org: Geobacter uraniumreducens RF4 # Function: Phosphopyruvate hydratase # MW: 46474.15 pI: 4.78
gura_28feb07:Gura_1062	131	104	88	101	74	46	90.67	# Org: Geobacter uraniumreducens RF4 # Function: DNA-directed RNA polymerase, beta' subunit # MW:153038.44 pI:
geob_13dec05:GeobDRAFT_0348	208	59	131	32	70	35	89.17	# Org: Geobacter sp. FRC-32 # Function: ribosomal protein L7/L12 # MW: 13088.21 pI: 4.76
gbem_26sep06:GbemDRAFT_1907	99	104	70	107	85	65	88.33	# Org: Geobacter bemidjiensis Bem (T) # Function: chaperonin Cpn60/TCP-1 # MW: 36907.92 pI: 5.09

Dolly: Excellent High Quality Chromatograms



Paloma Medium Quality Chromatogram

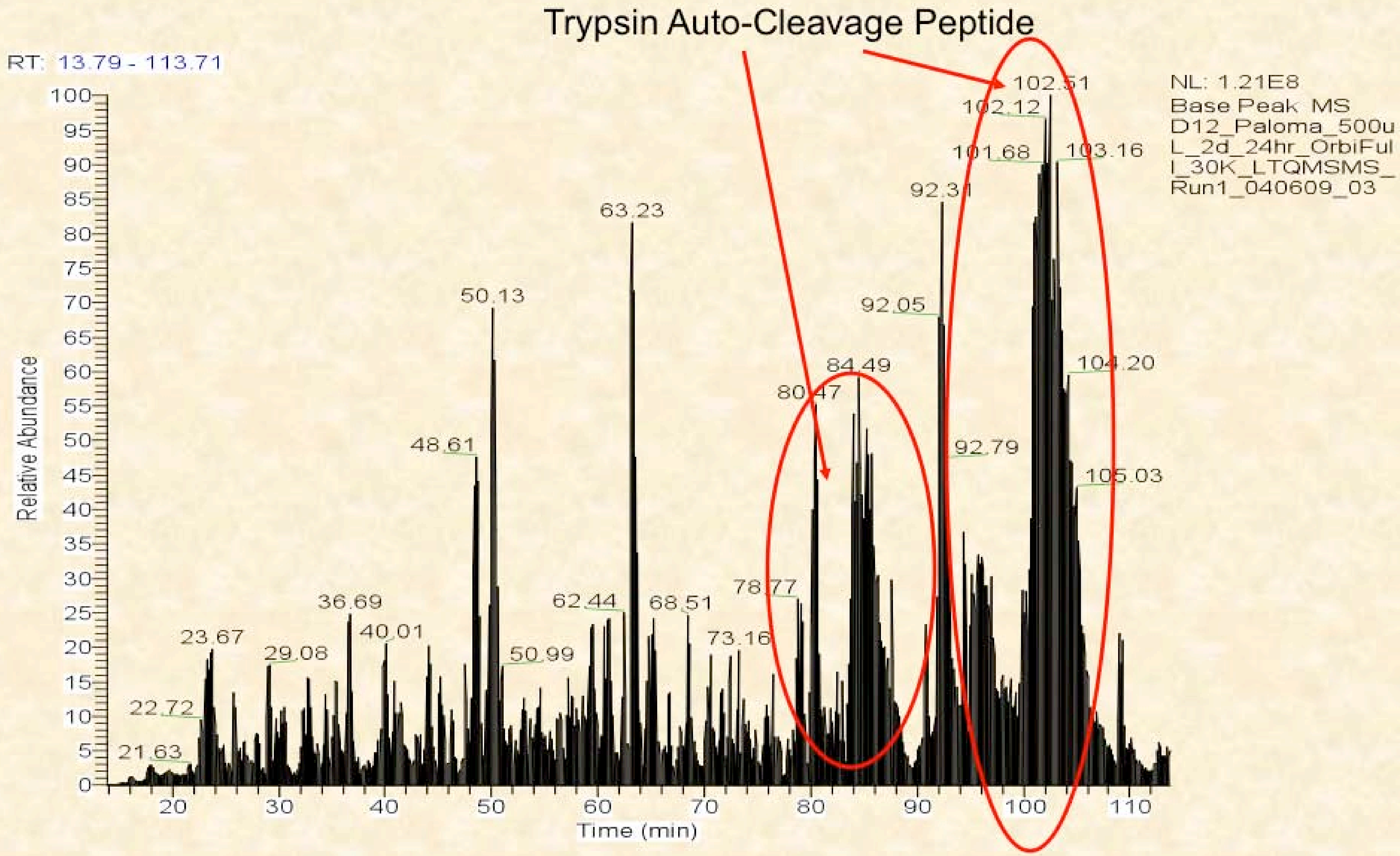


Figure 9 Vicky: Low Quality Chromatogram

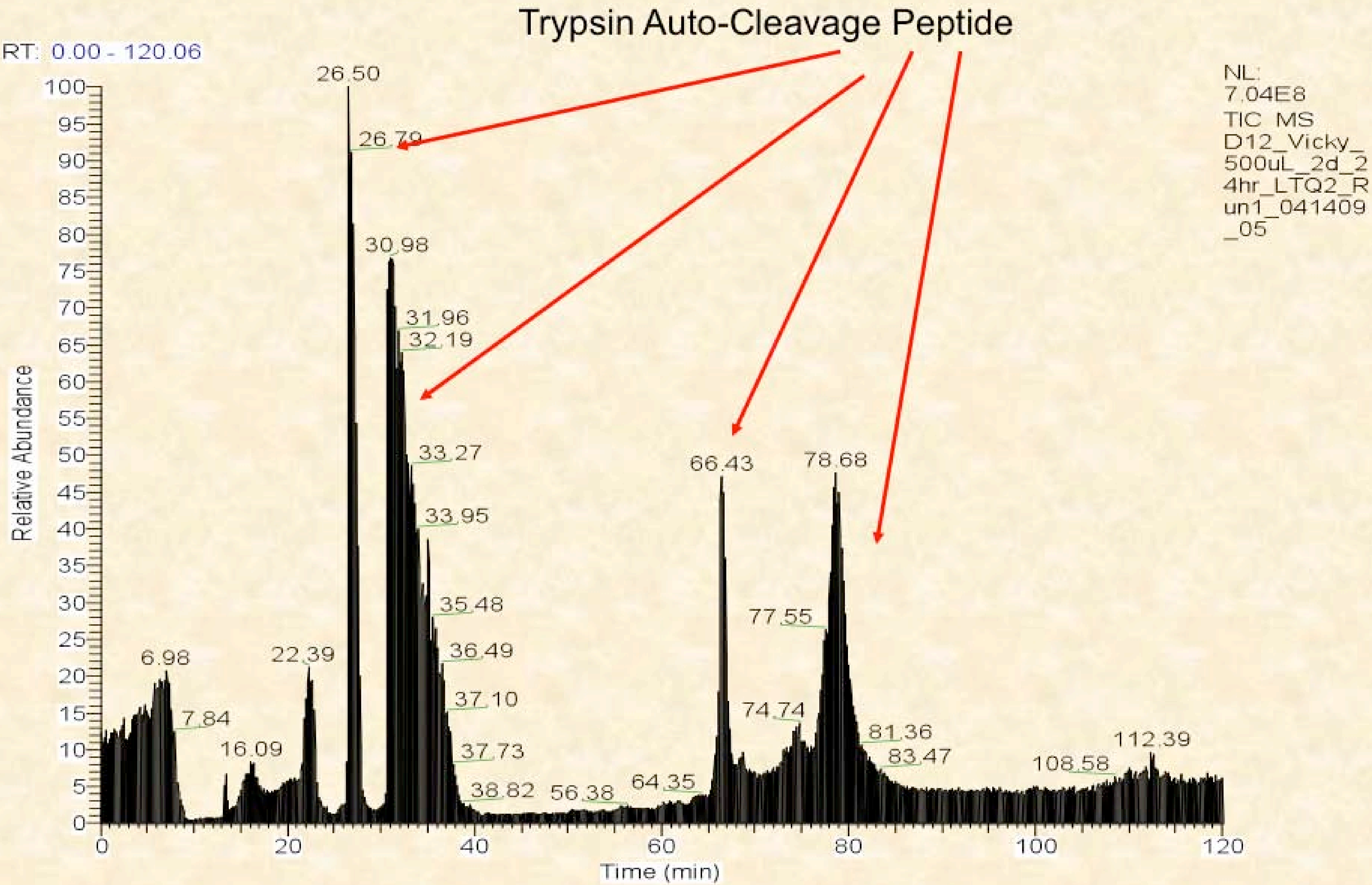


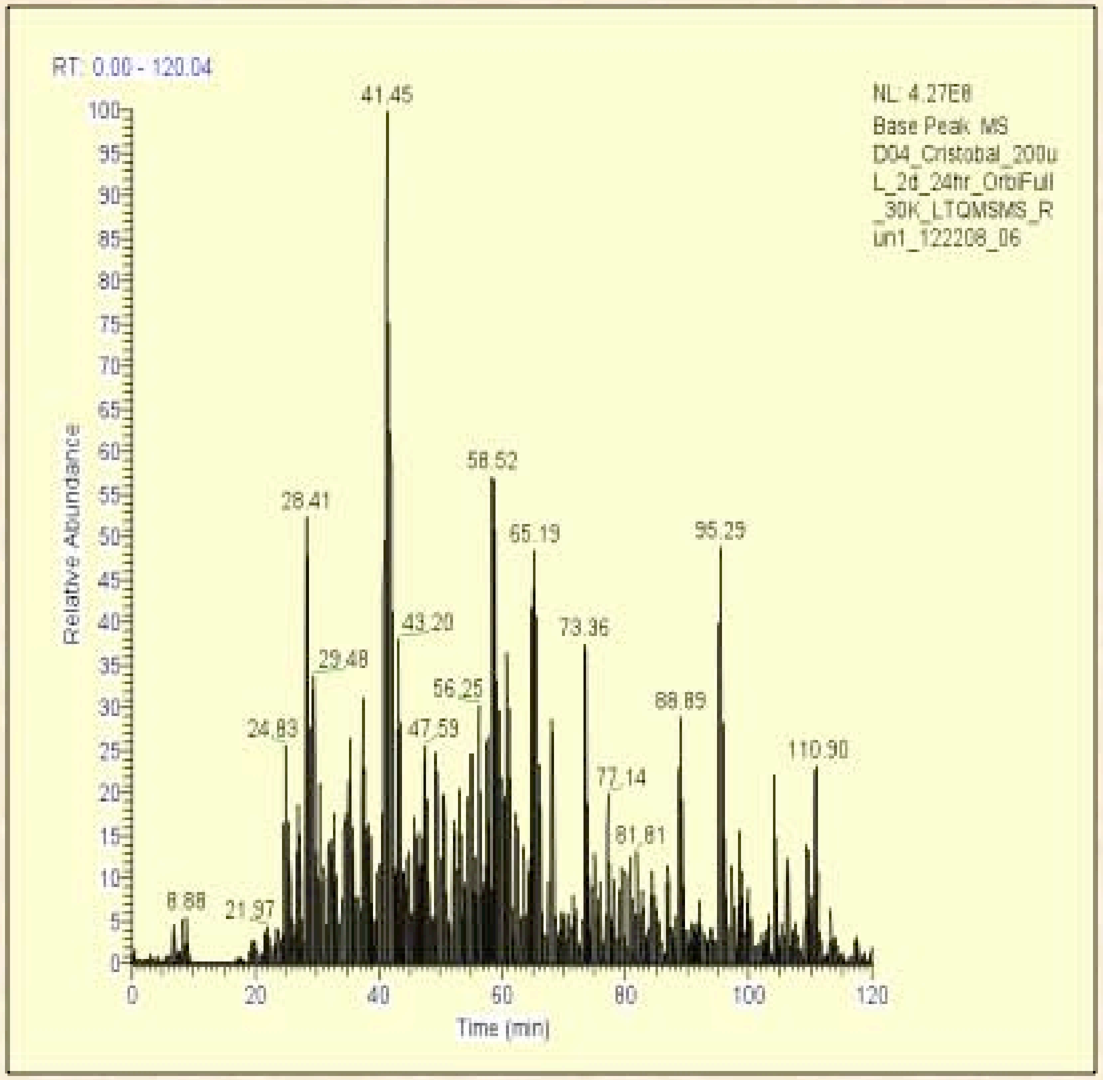
Figure 10: Typical sample showing Geobacter dominance.

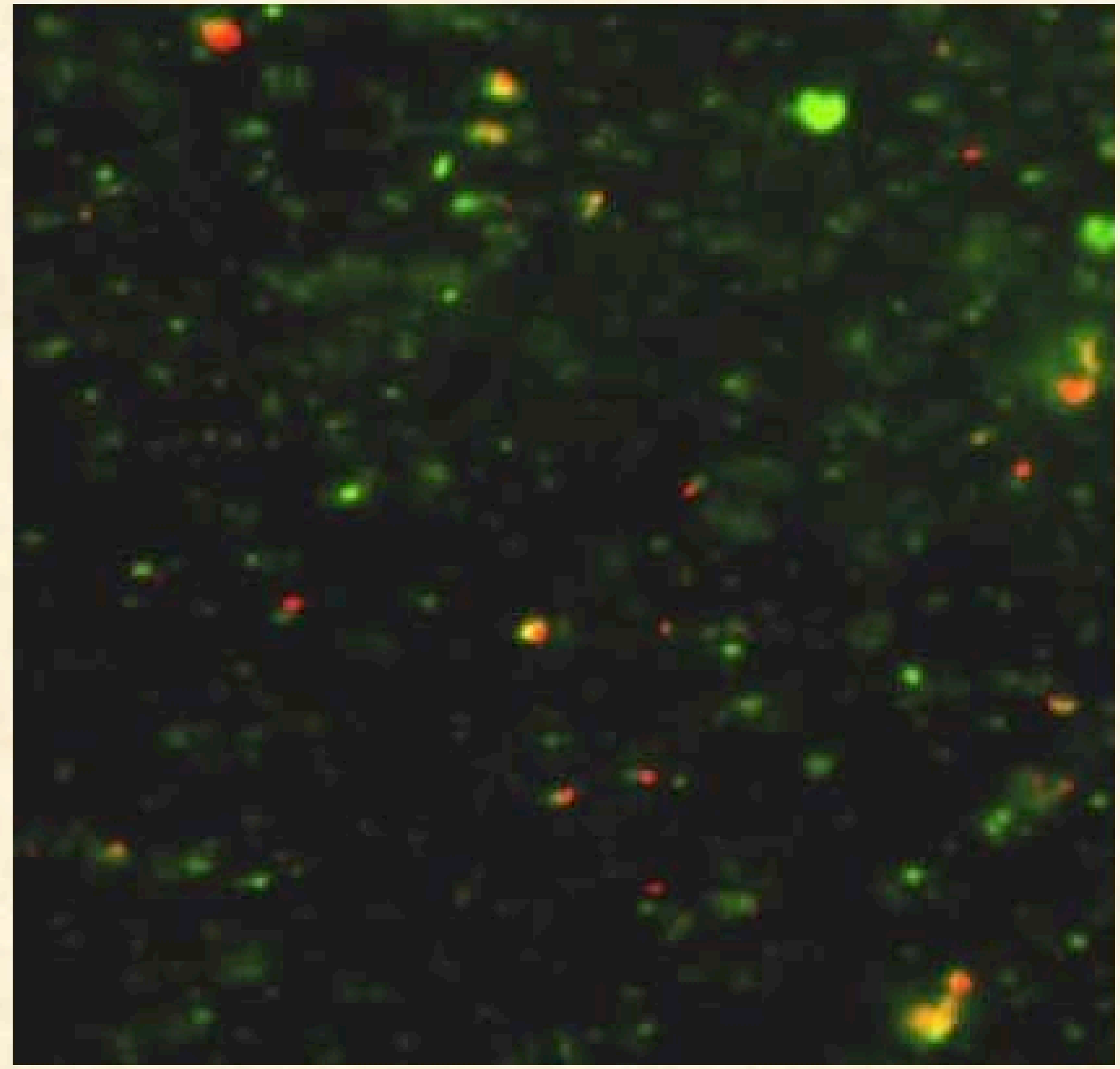
ORNL Mass Spec Sequence Coverage - ERSP_Rifle Samples - Windows Internet Explorer

http://compbio.ornl.gov/cgi-bin/mspipeline/seqcvg/dtaprtns_SeqCvg.cgi?dtadir=mspipeline/ersp_rifle/analysis/Groundwater_

#	Protein	Stats											Description
	Name	Length	Mol Wt	pI	% SeqCvg	SeqCnt	SeqUnique	SeqNon-Unique	SpecCnt	Unique	Non-Unique	NSAF	
1	gm|828920|gene2087	72	7855	4.5	100.0%	8	8	0	41	41	0	0.00324	# Org: Geobacter M21 # Function: Malate/lactate dehydrogenases
2	geob_13dec05:GeobDRAFT_0097	66	7227	5.3	100.0%	7	2	5	20	5	15	0.00173	# Org: Geobacter sp. FRC-32 # Function: Cold-shock protein, DNA-binding
3	gbem_26sep06:GbemDRAFT_2454	94	9779	10.4	91.5%	26	0	26	311	0	311	0.01885	# Org: Geobacter bemidjiensis Bem (T) # Function: histone family protein DNA-binding protein
4	gm|829365|gene3580	94	9779	10.4	91.5%	26	0	26	311	0	311	0.01885	# Org: Geobacter M21 # Function: Bacterial nucleoid DNA-binding protein
5	gm|829486|gene4063	66	7040	4.9	89.4%	4	3	1	14	11	3	0.00121	# Org: Geobacter M21 # Function: Cold shock proteins
6	gbem_26sep06:GbemDRAFT_3246	179	19874	9.7	86.6%	22	0	22	40	0	40	0.00127	# Org: Geobacter bemidjiensis Bem (T) # Function: ribosomal protein L5
7	gm|830124|gene2578	179	19860	9.7	86.6%	22	0	22	40	0	40	0.00127	# Org: Geobacter M21 # Function: Ribosomal protein L5
8	gm|829133|gene2599	126	12983	4.7	84.9%	12	1	11	51	15	36	0.00231	# Org: Geobacter M21 # Function: Ribosomal protein L7/L12
9	gbem_26sep06:GbemDRAFT_3239	93	10403	10.3	81.7%	11	0	11	42	0	42	0.00257	# Org: Geobacter bemidjiensis Bem (T) # Function: ribosomal protein S19
10	gm|830132|gene2585	93	10403	10.3	81.7%	11	0	11	42	0	42	0.00257	# Org: Geobacter M21 # Function: Ribosomal protein S19
11	gbem_26sep06:GbemDRAFT_2049	82	8745	8.4	79.3%	11	1	10	23	2	21	0.00160	# Org: Geobacter bemidjiensis Bem (T) # Function: thioredoxin
12	gbem_26sep06:GbemDRAFT_0731	470	51057	5.0	77.2%	51	6	45	338	19	319	0.00410	# Org: Geobacter bemidjiensis Bem (T) # Function: ATP synthase F1, beta subunit
13	gbem_26sep06:GbemDRAFT_1901	182	19242	6.2	76.9%	20	0	20	203	0	203	0.00636	# Org: Geobacter bemidjiensis Bem (T) # Function: pyruvate ferredoxin/flavodoxin oxidoreductase
14	gm|828924|gene2042	182	19242	6.2	76.9%	20	0	20	203	0	203	0.00636	# Org: Geobacter M21 # Function: Pyruvate:ferredoxin oxidoreductase and related 2-oxoacid:ferredoxin oxidoreductases, gamma subunit
15	gbem_26sep06:GbemDRAFT_3244	122	13111	10.2	76.2%	17	0	17	91	0	91	0.00425	# Org: Geobacter bemidjiensis Bem (T) # Function: ribosomal protein L14
16	gm|830126|gene2580	122	13111	10.2	76.2%	17	0	17	91	0	91	0.00425	# Org: Geobacter M21 # Function: Ribosomal protein L14
17	gbem_26sep06:GbemDRAFT_3247	132	14530	9.7	75.0%	27	0	27	74	0	74	0.00319	# Org: Geobacter bemidjiensis Bem (T) # Function: ribosomal protein S8
18	gm|830123|gene2577	132	14530	9.7	75.0%	27	0	27	74	0	74	0.00319	# Org: Geobacter M21 # Function: Ribosomal protein S8

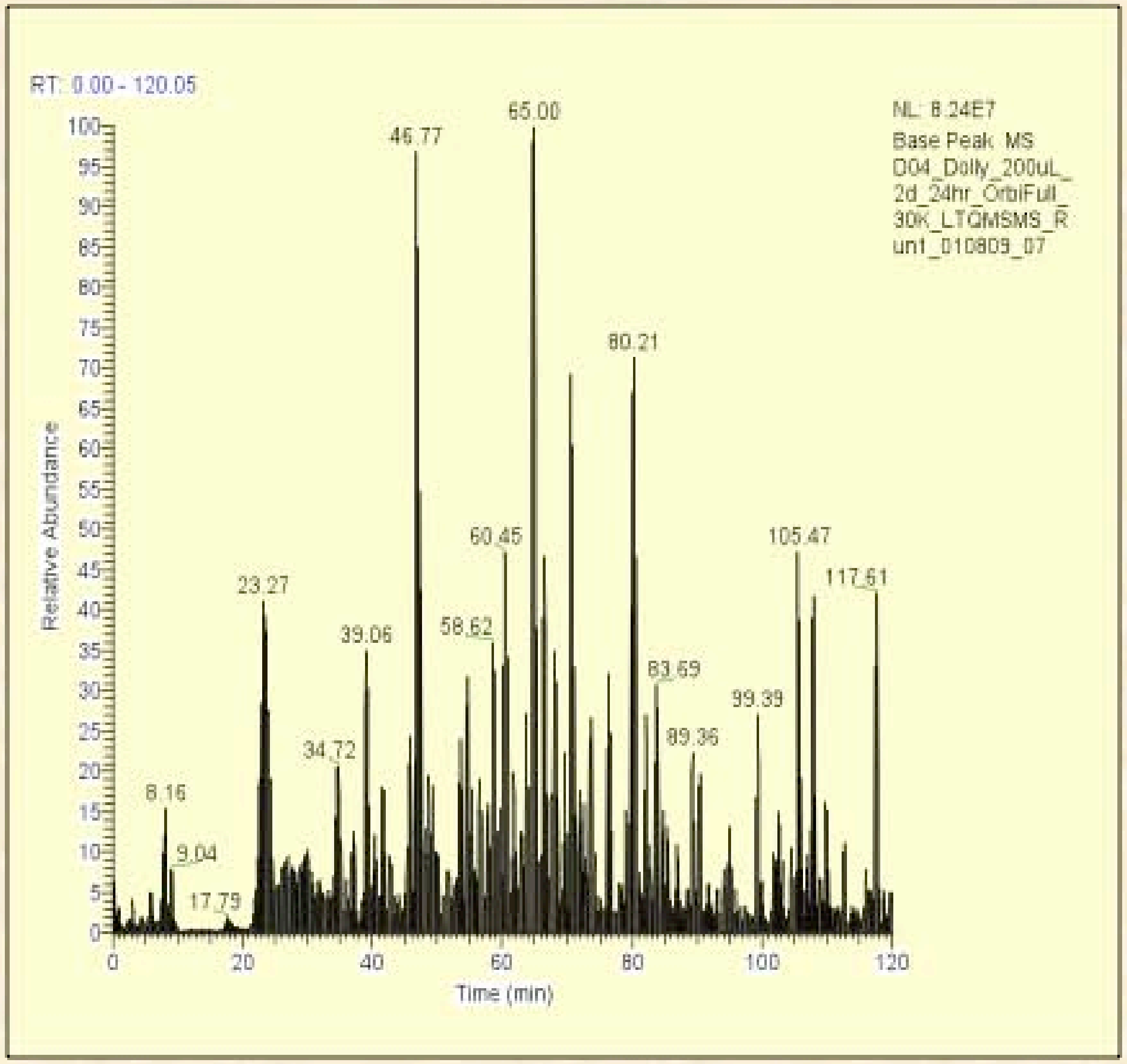
Local intranet 75%

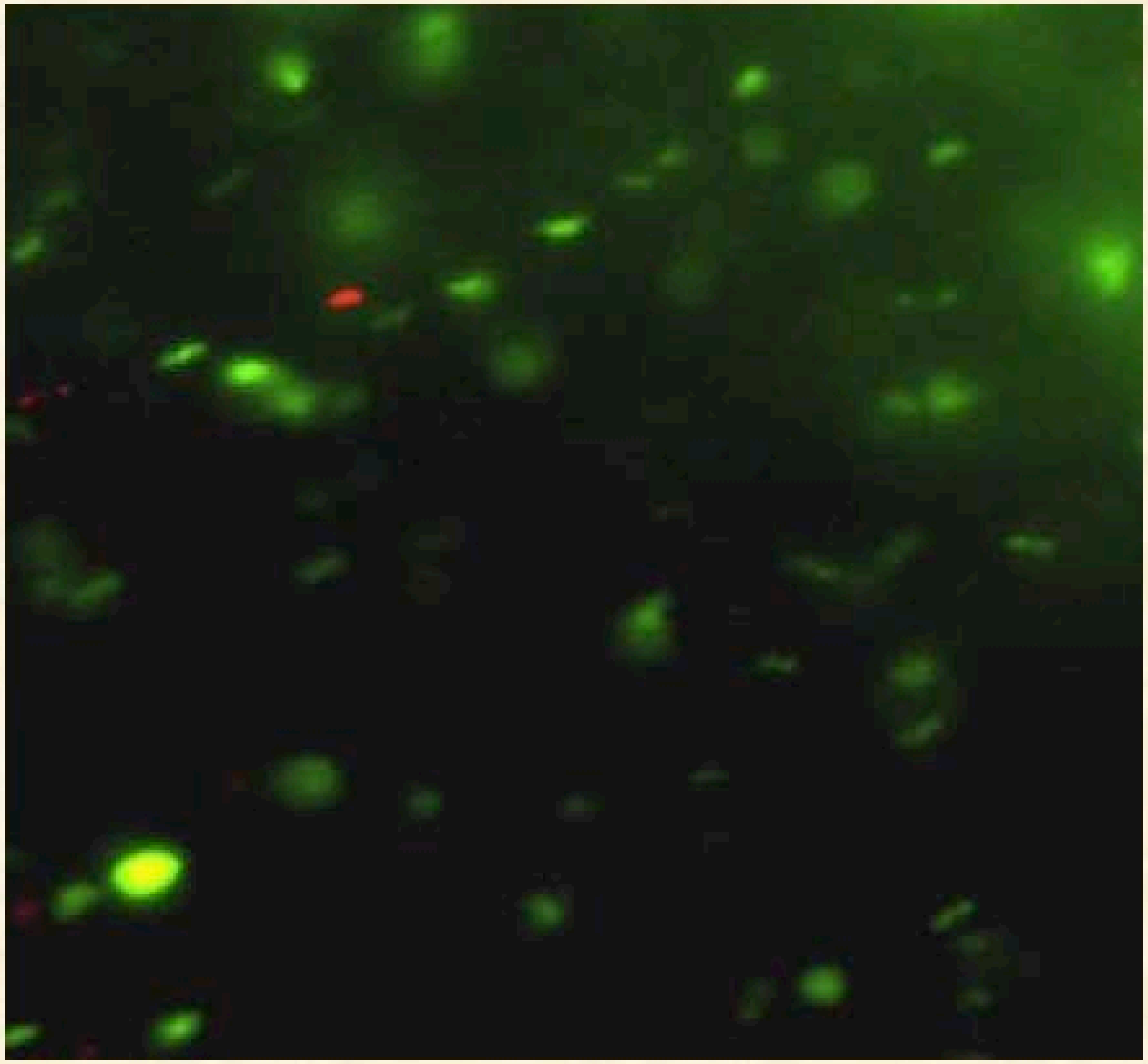




Cristobal: Lots of cells, mostly motile. More red (dead cells) than Arthur. Occasional precipitates seen.

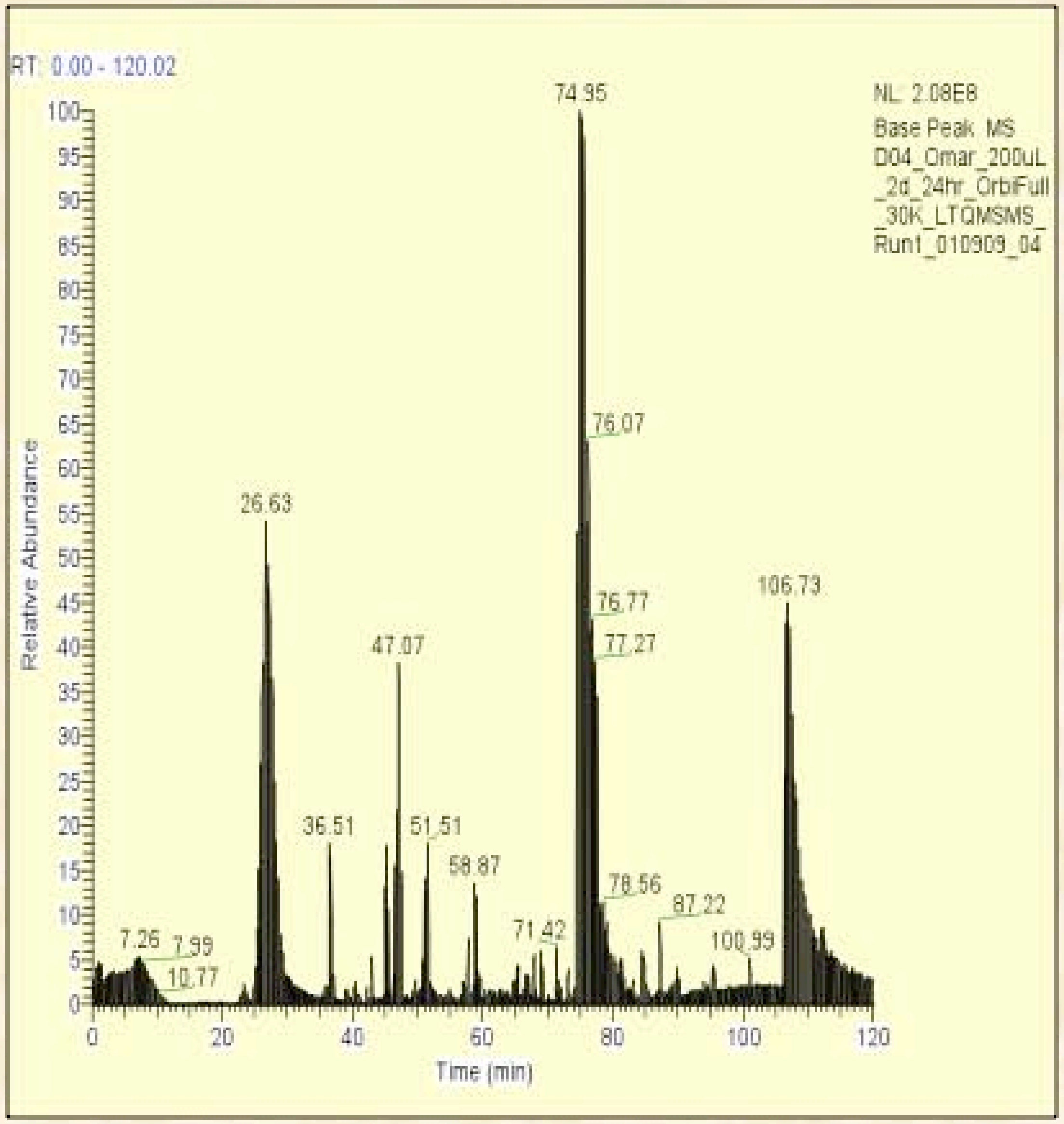
Proteome dominated by M21 and bemidjiensis, but more diverse hits to other species
1735 Proteins, 5501Peptides

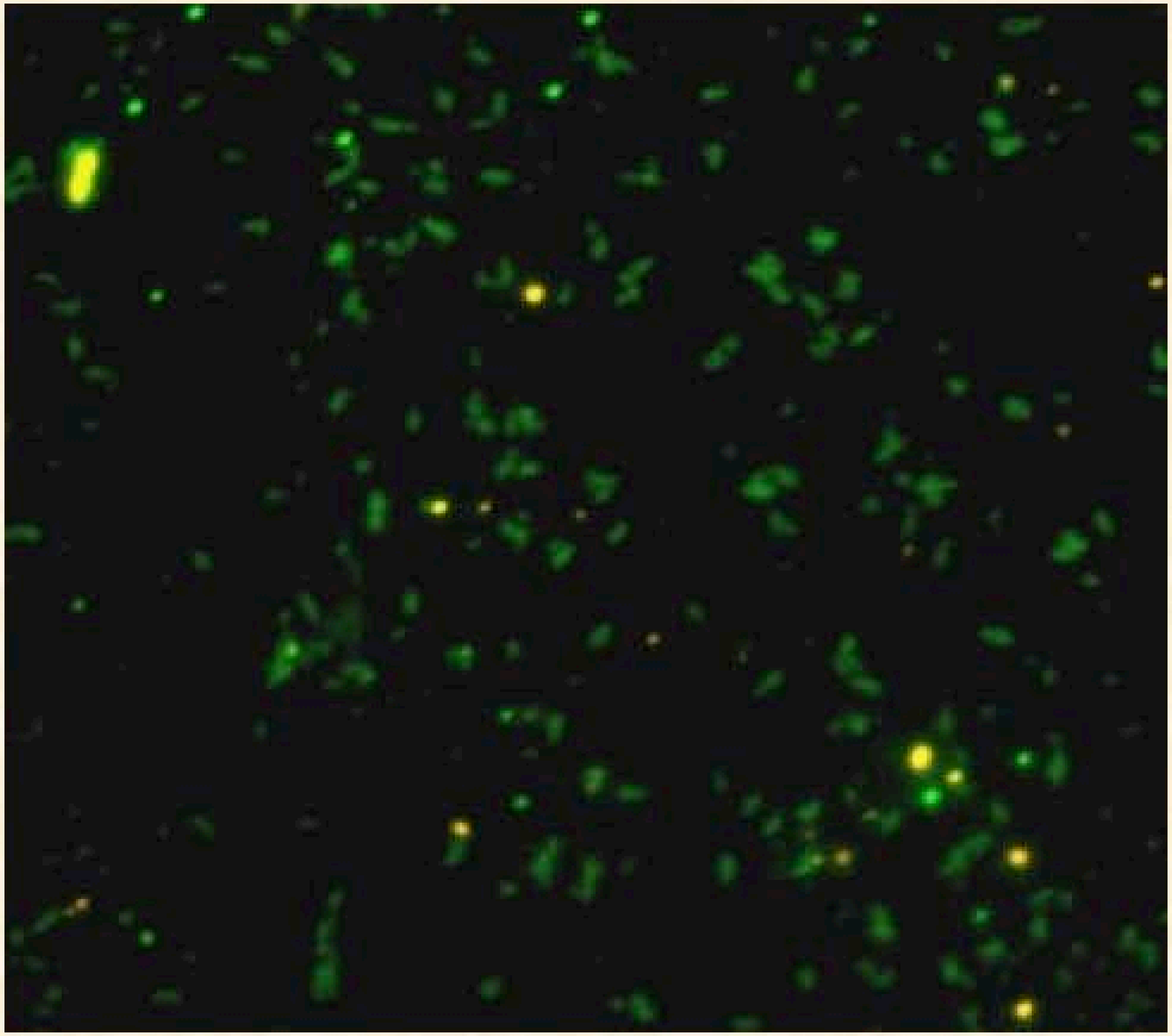




Dolly: Lots of motile cells, few dead cells, some precipitates. Cells counts appear to have decreased in comparison to Arthur and Cristobal.

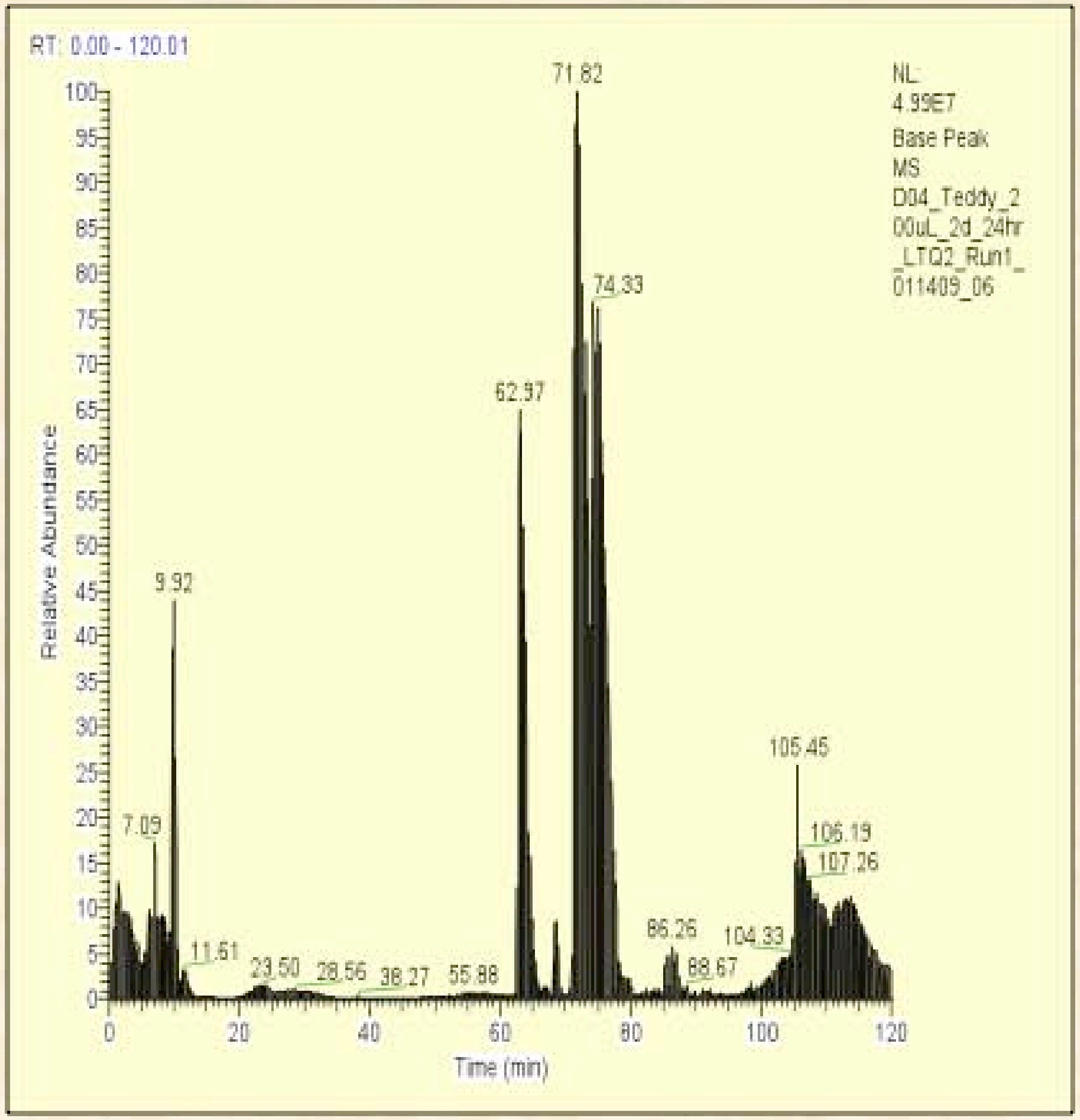
Proteome dominated by M21 and bemidjiensis
1726 Proteins, 5568 Peptides.

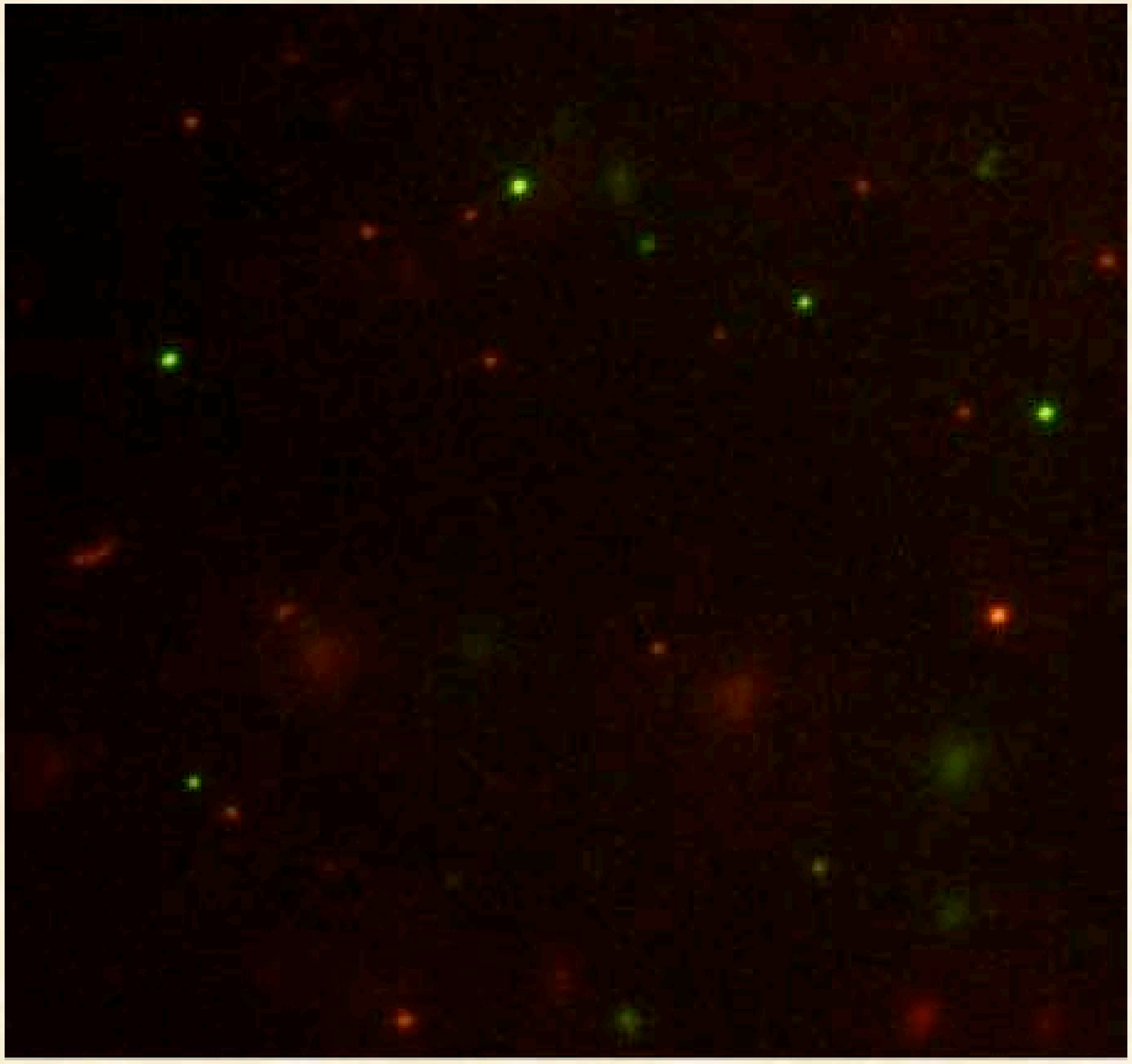




Omar: Lots of live cells. 10X more live cells than Hanna but still much less than Arthur, Cristobal, Dolly. Occasional precipitate seen.

Omar Proteome dominated by M21 and bemidjiensis
352 Proteins, 818Peptides. Excess trypsin/keratins

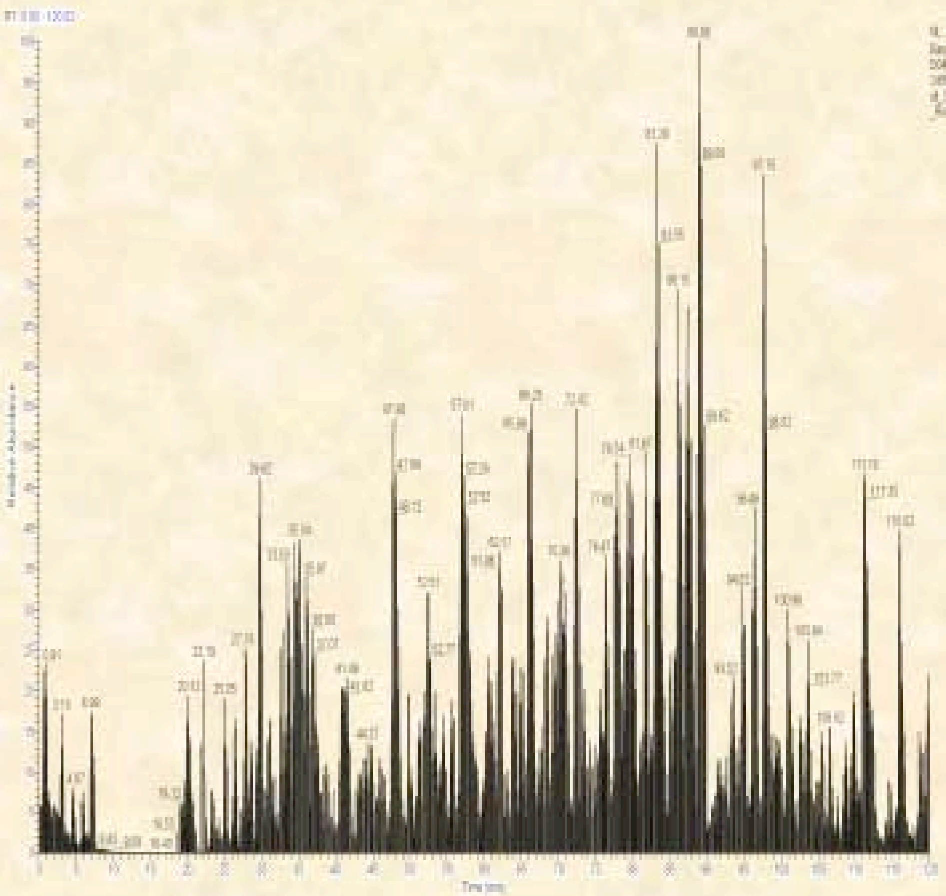


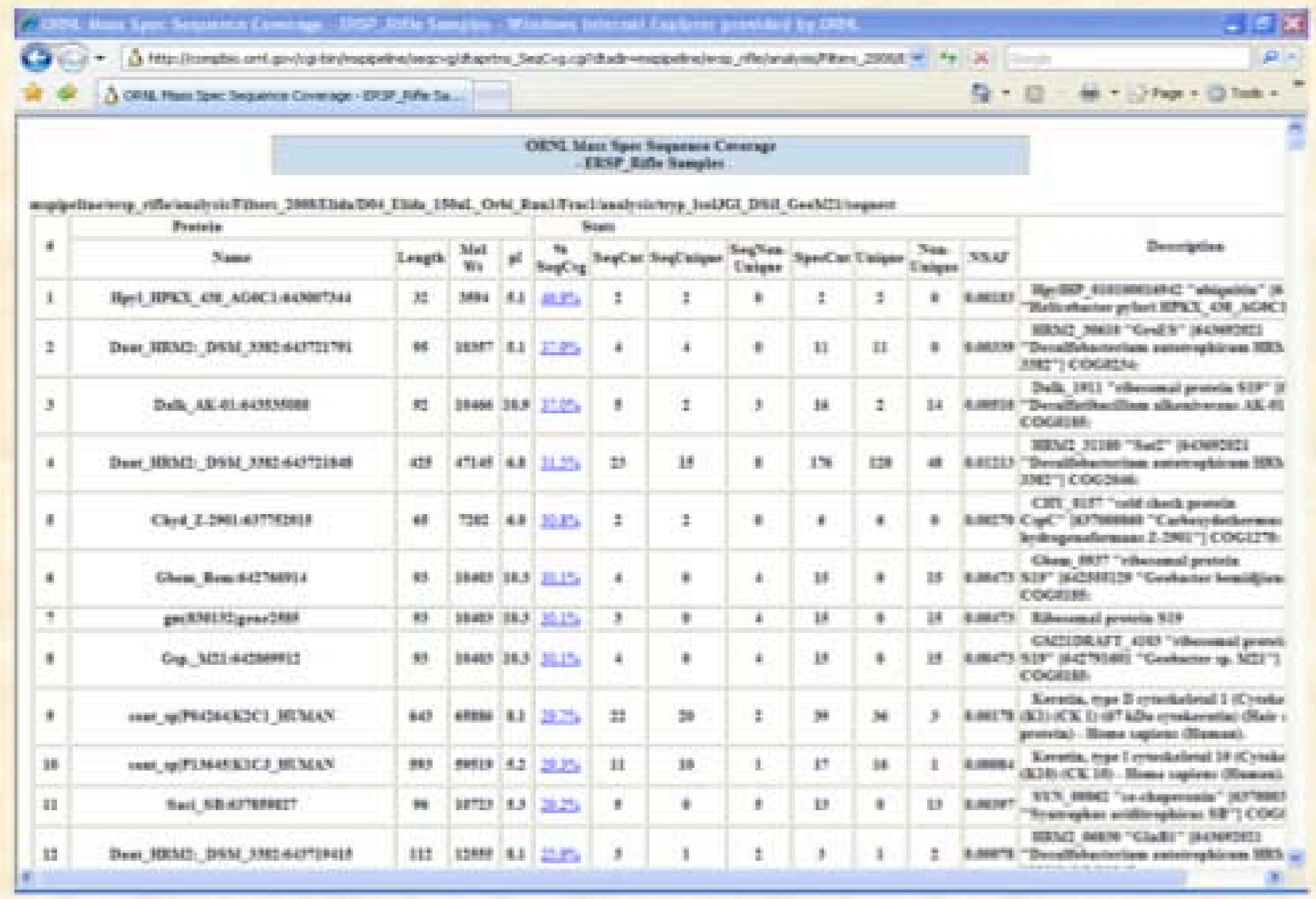


Teddy: Very few cells overall. Many dead cells seen.

Teddy Proteome dominated by M21 and bemidjiensis only proteins seen with high confidence, GroEL/GroES, Ef-TU etc.
60 Proteins, 119Peptides. Excess trypsin/keratins

Elida Filters-Need metagenomes

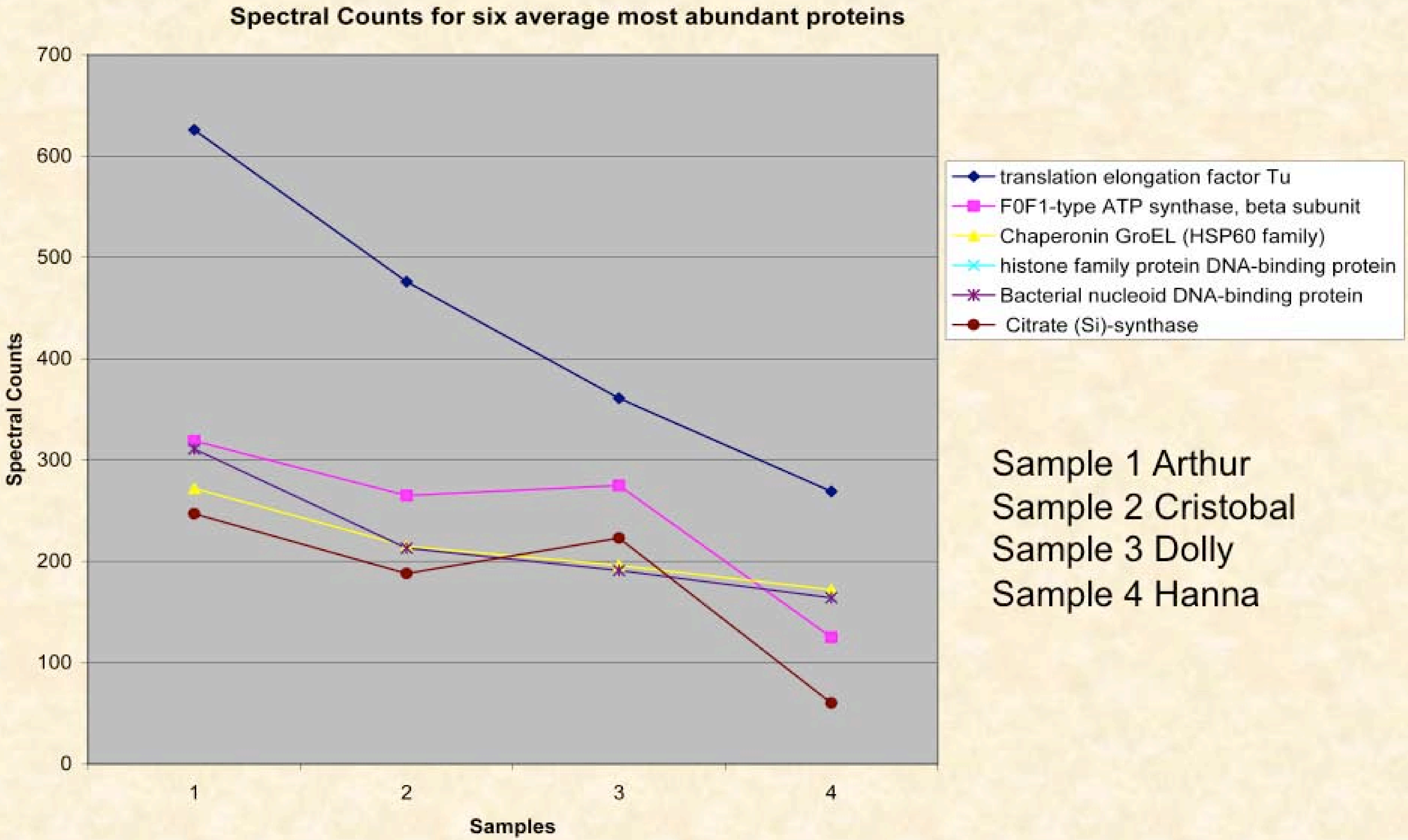




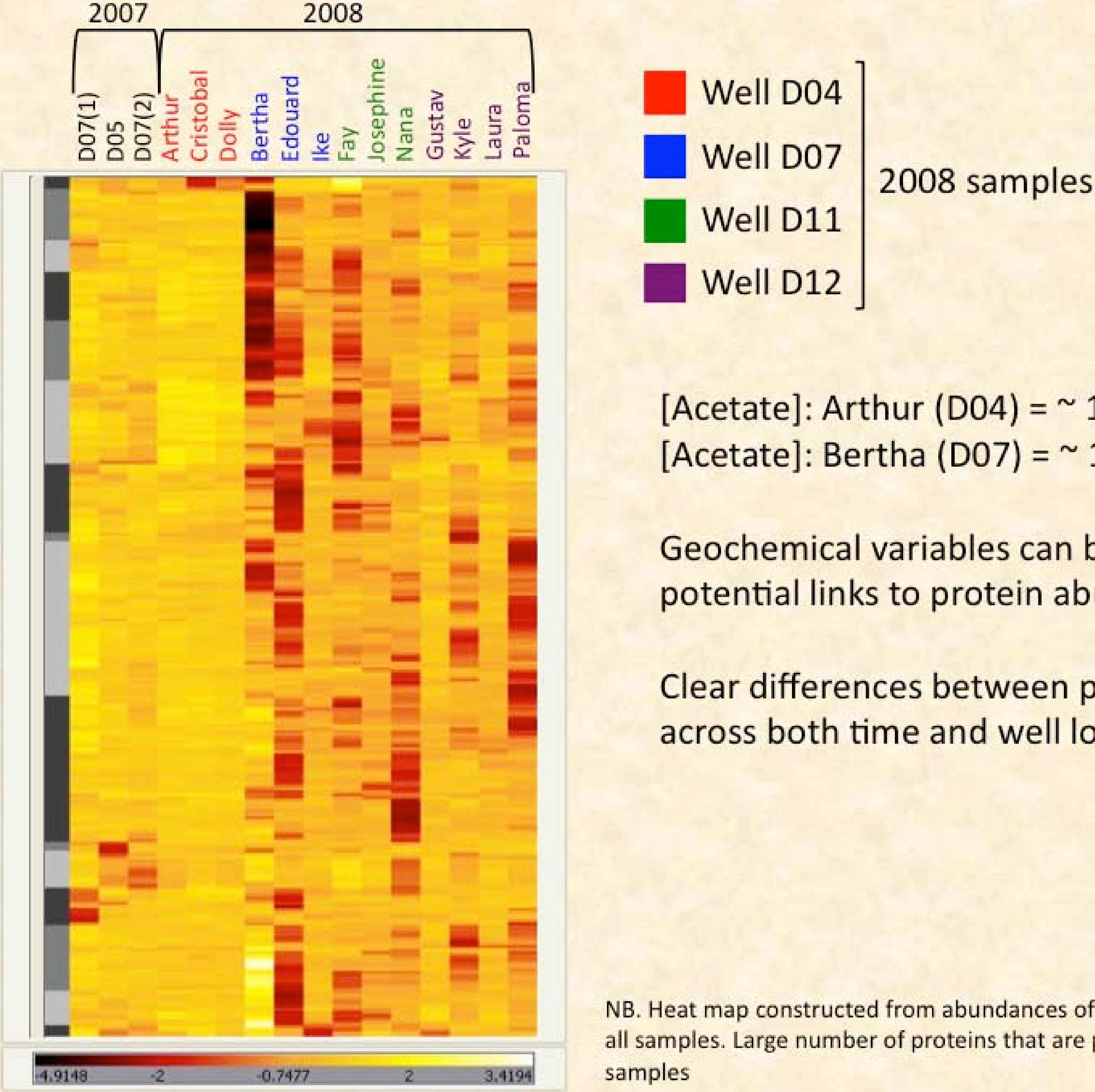
1200 Proteins but very low coverage

Dpsy_LSv54:637528384	15	30	23.80%	DP1105 "adenylylsulfate reductase" [637000094 "Desulfotalea psych	
Unique Tag	Xcorr	DelCN	PPM	Spectral Count	Peptide
*	3.6899	0.1693	-1000000	1	R.SGAVAQGLSAINTYIGENKIEDYVK.M
*	4.3637	0.3142	0.0103	1	R.NDLMGVVREDLIYDCGR.H
*	3.6313	0.2757	1851.447	1	R.HVDESVKLFEEWGLPVWK.K
*	2.6657	0.1116	0.3994	1	K.LFEEWGLPVWKK.D
	3.764	0.3082	0.4363	1	R.TGKWQIMINGESYK.P
*	4.6484	0.3173	-0.2105	1	K.LLLDKNKPNQIAGAAGFSTR.D
*	5.7301	0.4863	1.5404	1	K.LLLDKNKPNQIAGAAGFSTR.D
*	3.618	0.2917	0.138	6	K.NKPNQIAGAAGFSTR.D
*	3.0731	0.2284	0.9068	1	K.AAMVACGGAVNIFRPR.S
	5.1192	0.3687	1.8276	5	MTTVDGLFTAGDGVGASGHKFSSGSHAEGR.I
	3.7985	0.4893	2.6112	2	MTTVDGLFTAGDGVGASGHKFSSGSHAEGR.I
	2.9524	0.2951	-2.8661	6	K.FSSGSHAEGR.I
	2.3966	0.2821	2.7462	1	K.FSSGSHAEGR.I
	2.5772	0.1104	-2.2965	1	R.IWCVETHIR.H
	2.601	0.1248	-0.8729	1	K.ESRYPGFYYR.S

Figure 16: Spectral Counts from Abundant Geobacter proteins



Sample 1 Arthur
Sample 2 Cristobal
Sample 3 Dolly
Sample 4 Hanna



[Acetate]: Arthur (D04) = ~ 1 mM
[Acetate]: Bertha (D07) = ~ 150 µM

Geochemical variables can be assessed for potential links to protein abundances

Clear differences between protein abundances across both time and well locations.

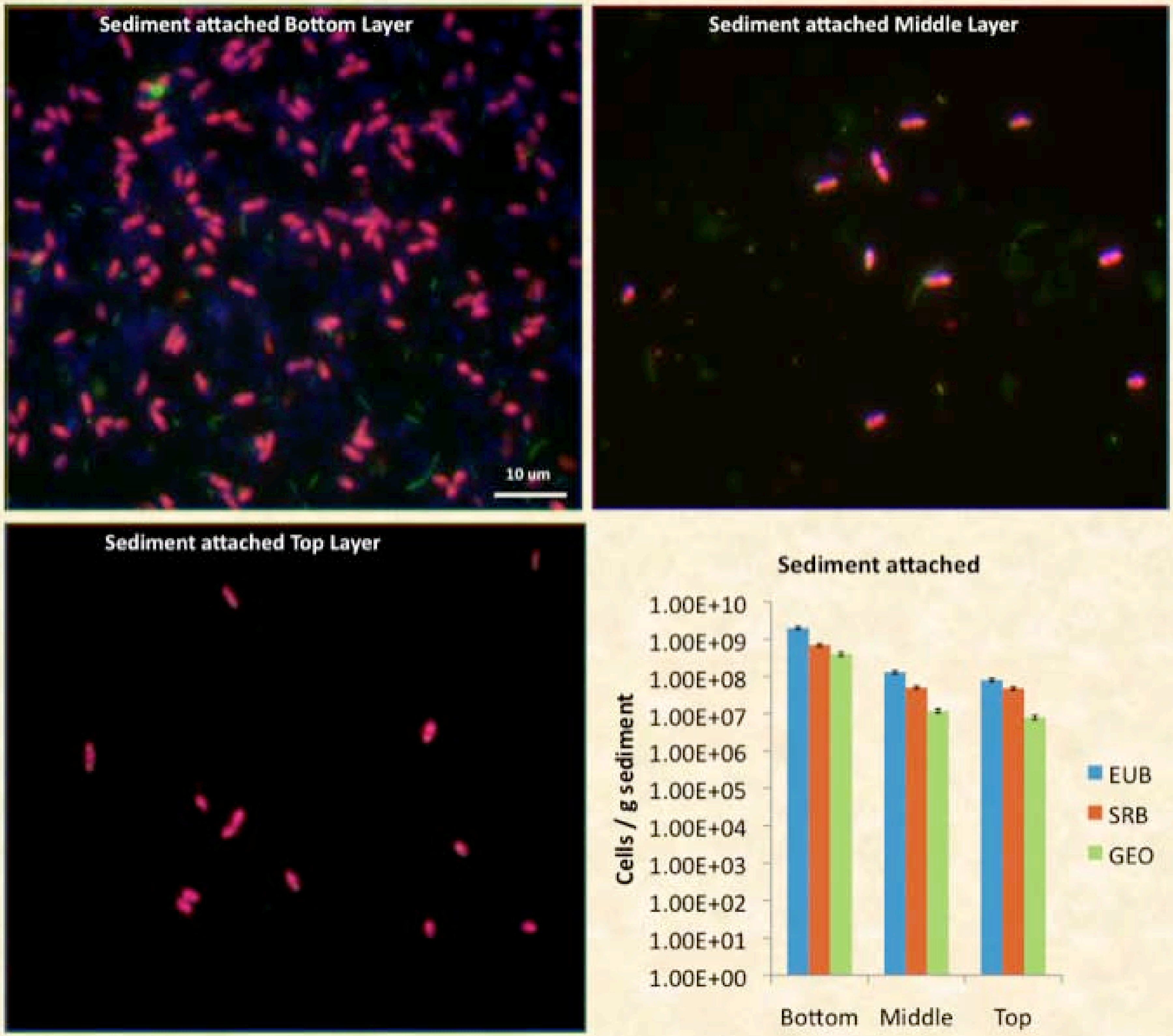
NB. Heat map constructed from abundances of proteins that were present in all samples. Large number of proteins that are present/absent between samples

In-well Column Experiments:

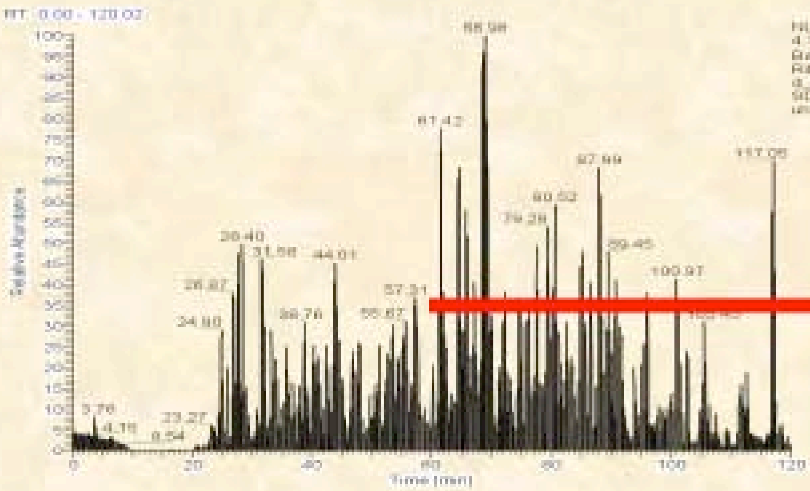
- Located inside borehole P104 during acetate amendment
- Packed with background sediments
- Recovered after 27-days of acetate injection
- Flow was upward

In-well Column Experiments:

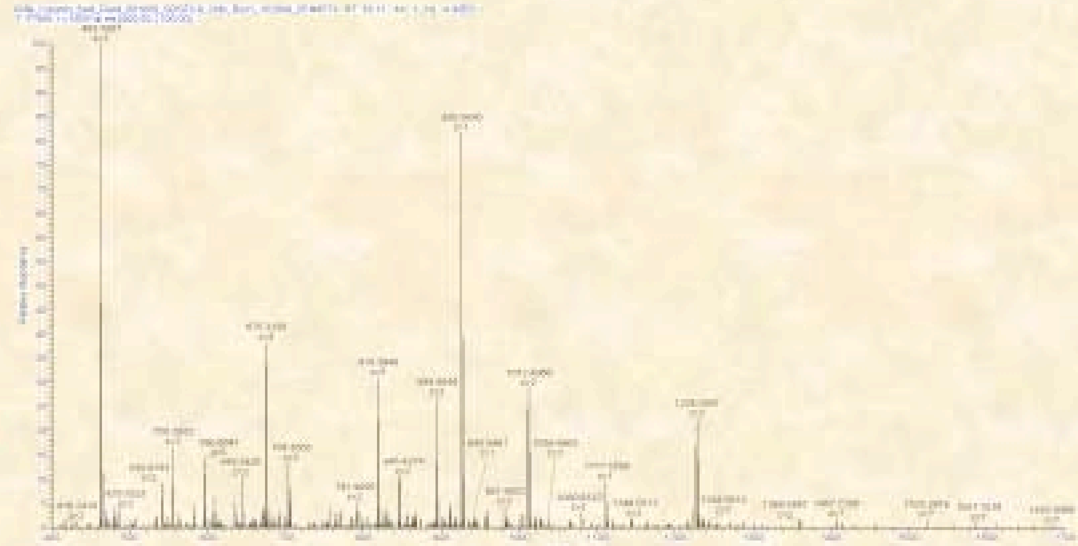
- FISH results [courtesy Dr. Shabir Dar and Prof. Derek Lovley (UMASS)



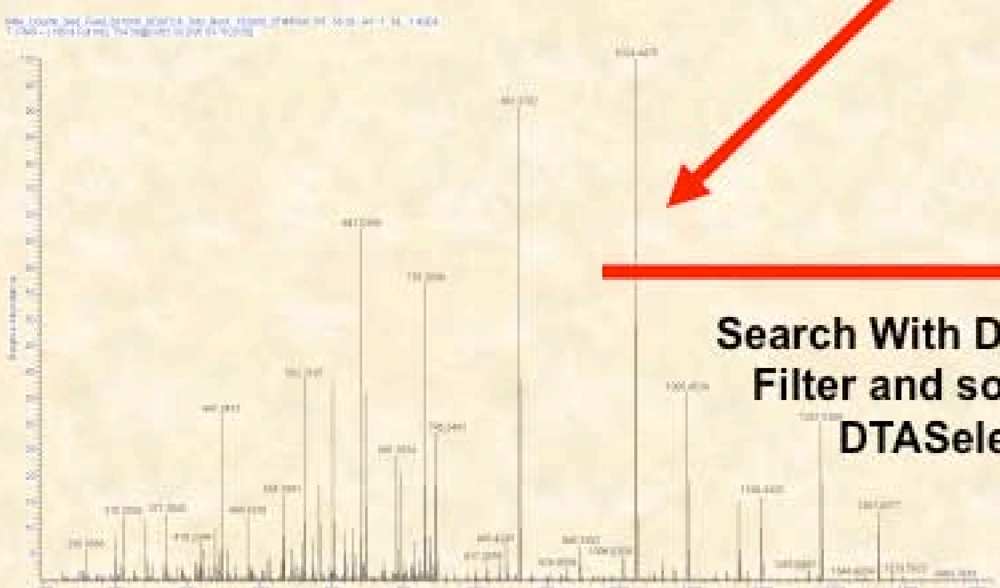
Raw LC-MS/MS data
Rifle In-ground Column Experiment
After Acetate amendment



Base Peak Chromatogram (BPC)
35% Salt Pulse from 2d-LC-MS/MS
This is a rich BPC with 10,000s of peptides
All other chromatograms were similar

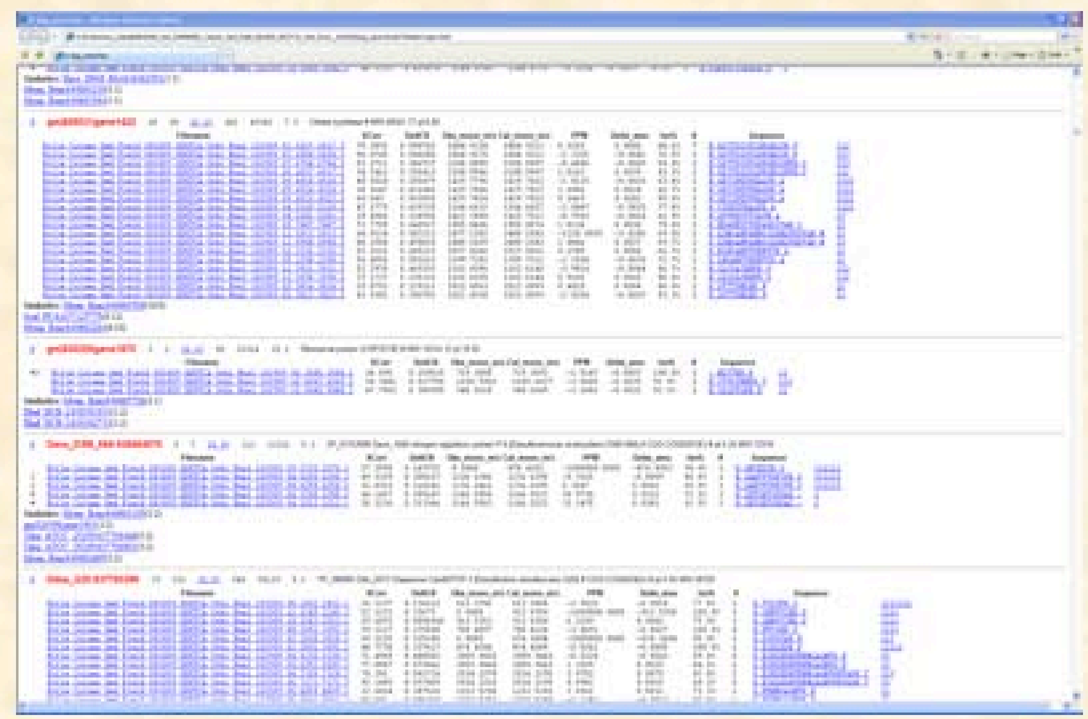


Full Scan in Orbitrap
Time point 56 minutes
16 Strong Peptide Signals at this time point



High quality tandem MS/MS spectra
of 1508 Dalton peptide at 56 minutes

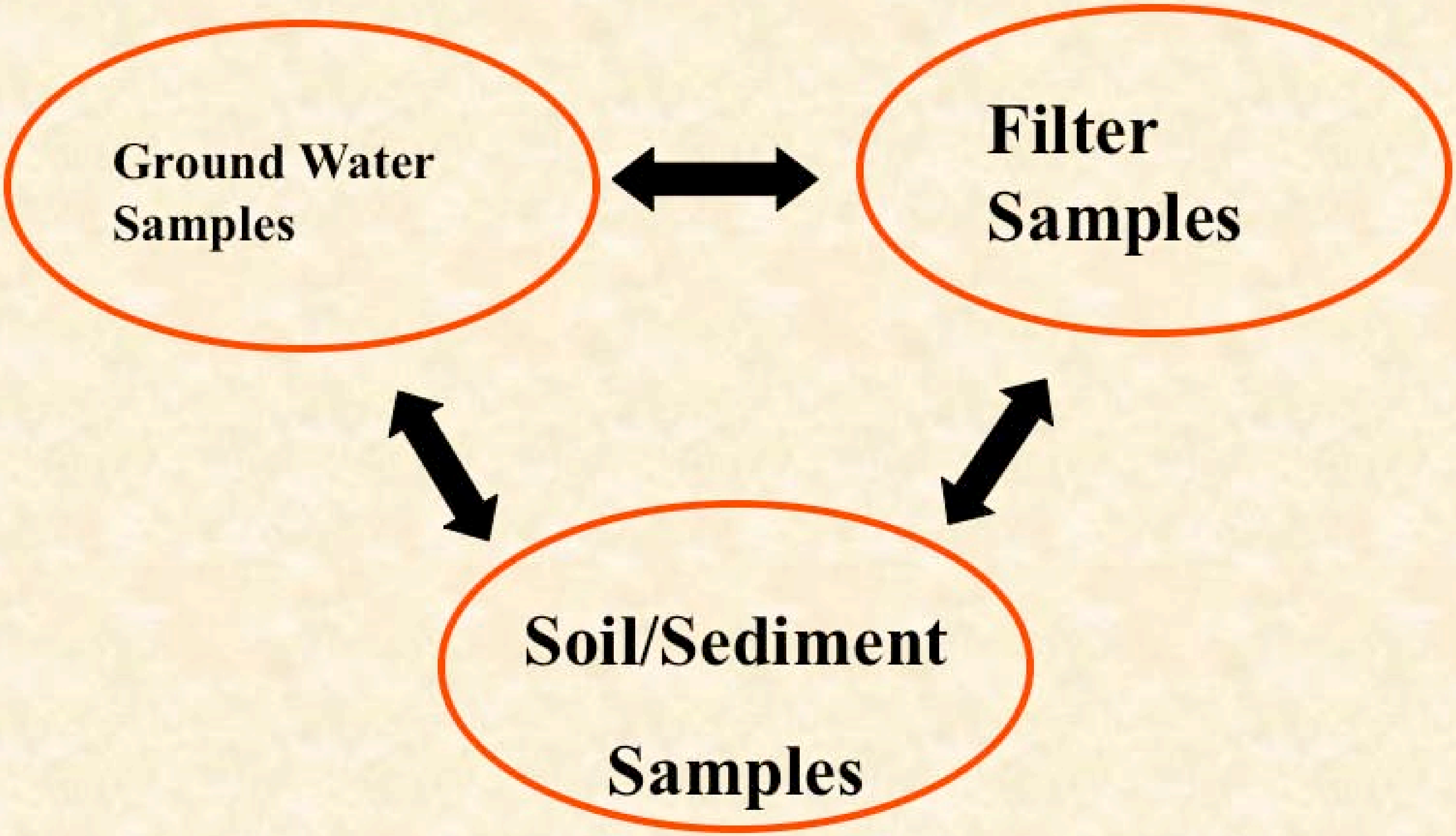
Search With DBDigger
Filter and sort with
DTASelect



DTASelect output over 1000 proteins identified in 24hrs from 1 run from only 5 grams of soil. Pictured here Citrate Synthase from Geobacter M21 with over 20% sequence coverage and all peptides found at high mass accuracy

So what do we need to integrate all three of these sample types??

Metagenomic datasets



Acknowledgements

Office of Science
U.S. DEPARTMENT OF ENERGY

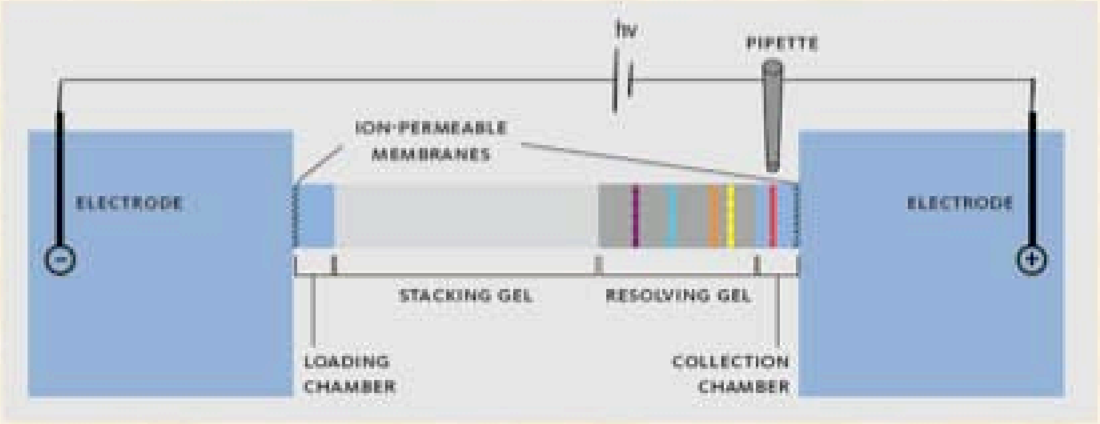
- **Funding:** *Environmental and Remediation Sciences Program (ERSP), Biological and Environmental Research (BER), Office of Science, U.S. Department of Energy*
- *Partial List of Project Participants:*

  Todd Anderson DOE-SC; Dick Dayvault Stoller; Jill Banfield UC Berkeley; Kim Handley, UC Berkeley; Mike Wilkins PNNL; Dawn Holmes U. Mass; Derek Lovely U. Mass; Lucie N'Guessan Umass; Darrell Chandler Akonni; Jim Davis USGS; Patricia Fox USGS; Nathan VerBerkmoes ORNL; Manesh Shah ORNL; Bob Hettich ORNL; Paul Abraham ORNL; Mark Lefsrud ORNL; Peter Jaffe Princeton; Lee Kerkhof Rutgers; Aaron Peacock Microbial Insights; Ken Williams LBNL; Susan Hubbard LBNL; Mary Lipton PNNL; Phil Long PNNL

- **Cooperation:**
  - *City of Rifle, Colorado*
  - *Colorado Department of Public Health and Environment*
  - *U.S. Environmental Protection Agency, Region 8*

New Technologies for MS-based Proteomics

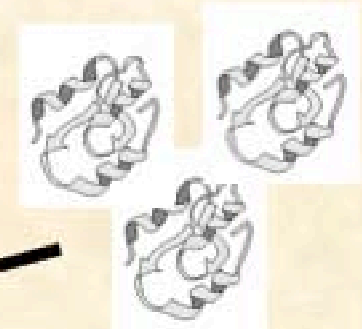
Electrospray
Linear Ion Trap
Orbitrap



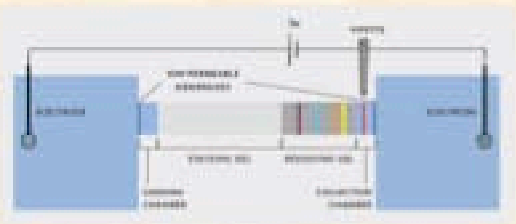
Three dimensional separations for metaproteomics

Complex Microbial Community

Cell lyses,
protein denaturation
and extraction



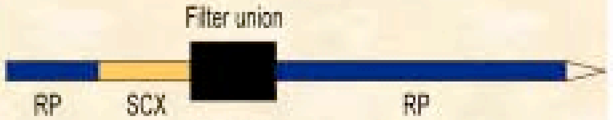
Complex Mixture of 10,000s of proteins



1st Dimension Intact Protein Sep
By MW via the GELFREE8100

Fraction Collection

HPLC



Split phase 2D-LC

LTQ-Velos Orbitrap