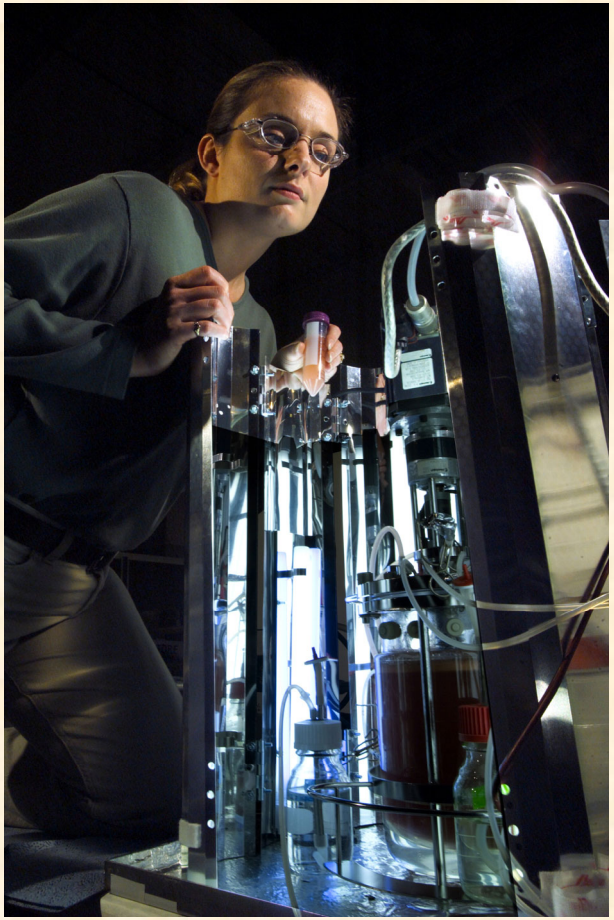
Proteomics of the Microbial Community in an Acid Mine Drainage Biofilm

N. C. VerBerkmoes[1], R. J. Ram[2], M. Shah[1], M. P. Thelen[2], G. W. Tyson[2], B. J. Baker[2], J. F. Banfield[2], R. L. Hettich[1]

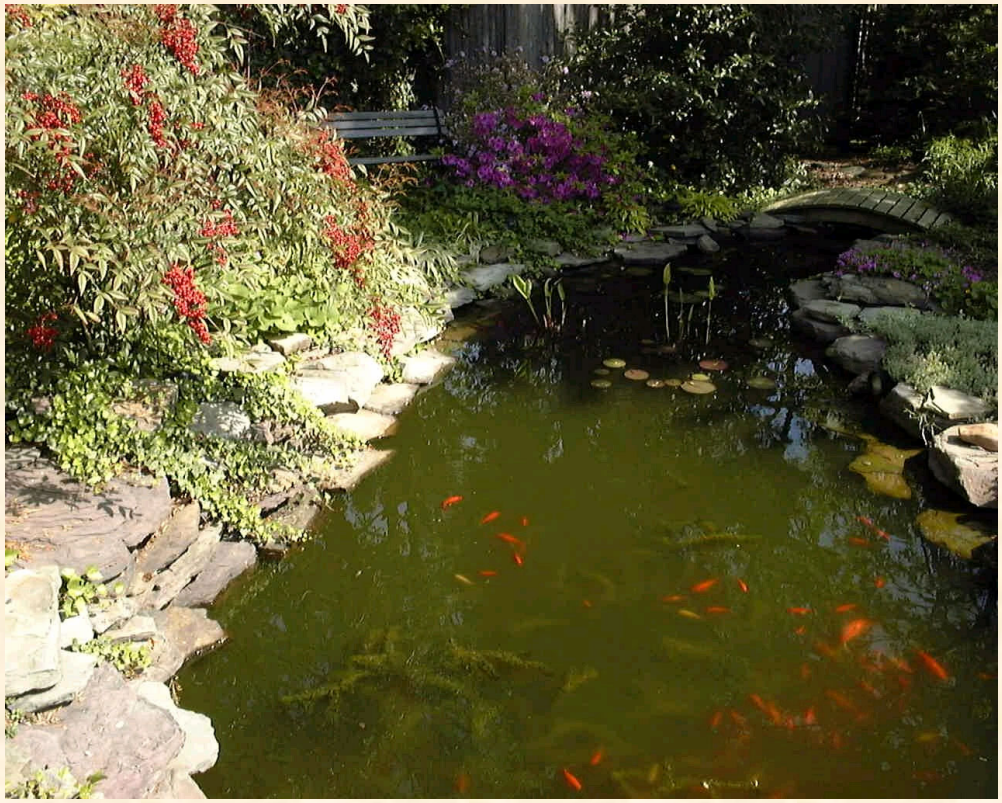
1) Oak Ridge National Laboratory, Oak Ridge, TN
2) University of California, Berkeley, Berkeley, CA

Proteomics from the lab into the world

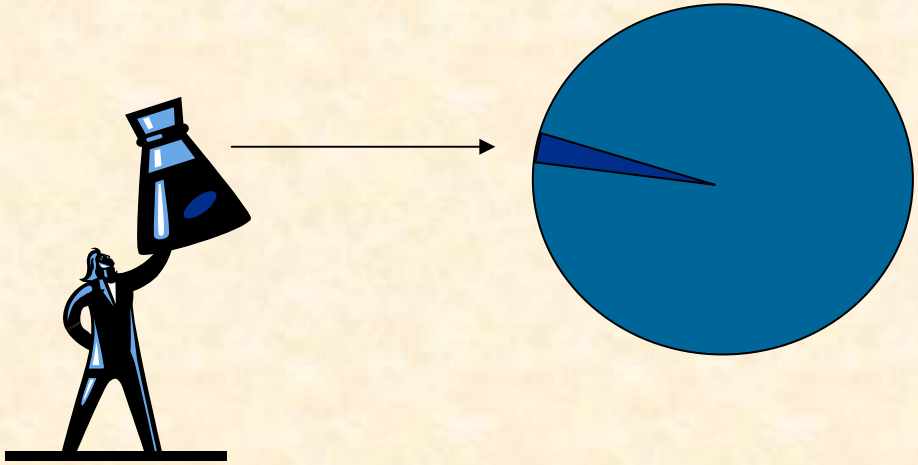
Pure microbial cultures, well-controlled, clean



Mixed undefined microbial communities, messy, changing, dirty



Why study microbes in the environment?



Most microorganisms are not cultured

Microbes in isolation vs. in a natural community

Who is doing what?

How are they doing it?

Challenges for Proteome Analysis of Microbial Communities

Proteome analysis of any microbial community will be difficult with any current technology.

The primary theoretical and practical concerns are:

1) The level of DNA sequence information and quality annotation available on the community.

2) The level of diversity and dynamic range associated with the species of interest in the community.

3) The quantity of biomass available for study.

4) The level of interrelatedness and/or diversity at the base pair level amongst members of the same species in the community and between species in the community.

ORNL Proteome Pipeline for Microbial Species

Microbe Stocks

↓

Cell Growth

↓

Protein extraction/Sample Preperation

↓

Liquid Chromatography

↓

Mass Spectrometry

↓

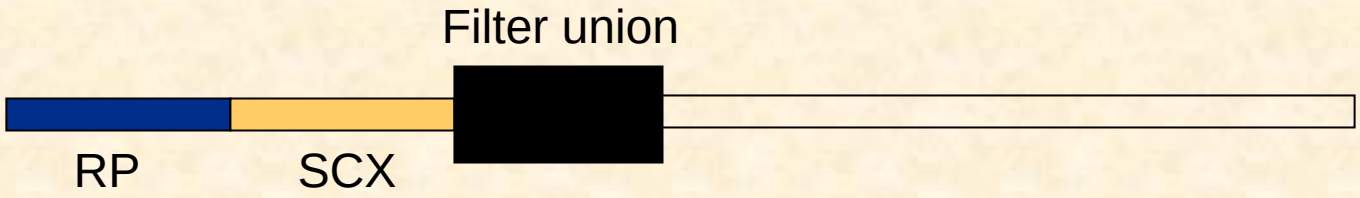
Proteome Bioinformatics

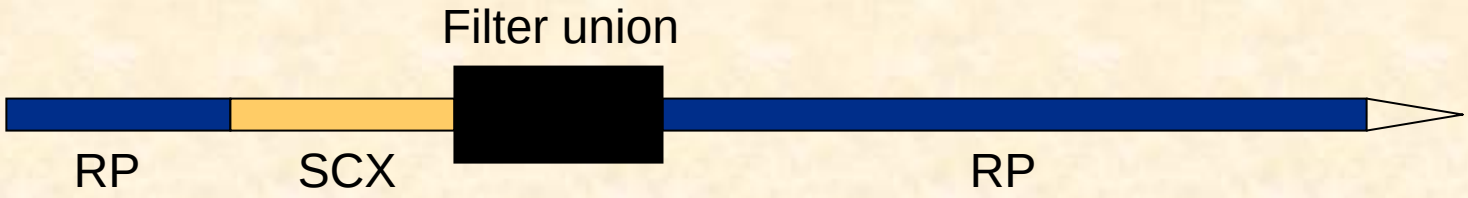
↓

Biology

Split-phase MudPIT
(rapid pressure cell-packing and loading)

described in McDonald et al. IJMS 2002

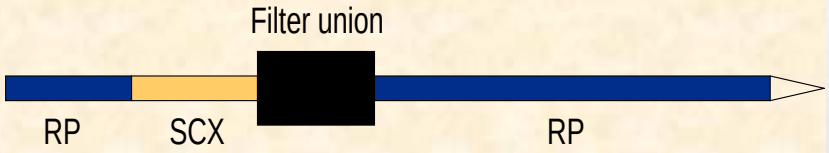




**Link *et al.* 1999 *Washburn *et al.* 2001

LC/LC-MS/MS

Ultimate HPLC



ES-ion trap MS

ORNL Proteomics Website

ORNL - UC, Berkeley AMD (Acid Mine Drainage) Community Proteome Study - Microsoft Internet Explorer

File Edit View Favorites Tools Help

ORNL - UC, Berkeley AMD (Acid Mine Drainage) Community Proteome Study

Home Genomic Databases MS Analysis Supplemental AMD - UC Berkeley

LTQ Analysis

Global Contrast ReadMe

p1_simple		p1_verbose		p2_simple		p3_simple	
HTML	Text	HTML	Text	HTML	Text	HTML	Text

Interrun Contrast ReadMe

Samples	Interrun Contrast							
	p1_simple		p1_verbose		p2_simple		p2_verbose	
Memb_A	HTML	Text	HTML	Text	HTML	Text	HTML	Text
Memb_N	HTML	Text	HTML	Text	HTML	Text	HTML	Text
Sample_A	HTML	Text	HTML	Text	HTML	Text	HTML	Text
Sample_B	HTML	Text	HTML	Text	HTML	Text	HTML	Text
Sample_C	---	---	---	---	---	---	---	---

Pairwise Contrast ReadMe

Sample Pairs	Pairwise Contrast							
	p1_simple		p1_verbose		p2_simple		p2_verbose	
Memb_A-Memb_N	HTML	Text	HTML	Text	HTML	Text	HTML	Text
Sample_A-Sample_B	HTML	Text	HTML	Text	HTML	Text	HTML	Text

Internet

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

Some Highlights of the AMD community

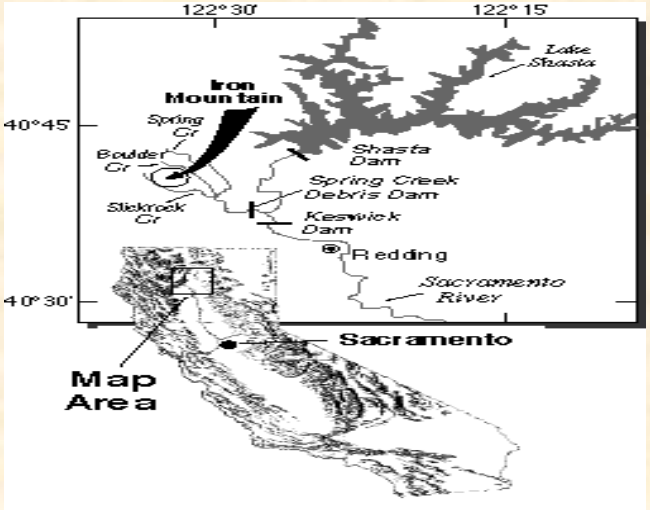
- Acid Mine Drainage has become a serious environmental problem since the advent of modern mining technologies.

- During deep mining projects, large quantities of the sulfide mineral in most rocks (pyrite FeS_2) become exposed to weathering and erosion by air and water. Mining increases the surface area of the sulfide ores exposed to air and water, thus increases the acid generation. In mining areas where the rocks have very low buffering capacity, the build up of acid and heavy metals in the water can become so great that the streams are actually toxic acidic heavy metal solutions called acid mine drainage, or AMD.

- The generation of AMD sites can theoretically occur by just geochemical events, but it is hypothesized that the unique microbial communities found to populate these harsh environments may actually be greatly increasing the generation of AMD (reviewed by Baker and Banfield, 2003).

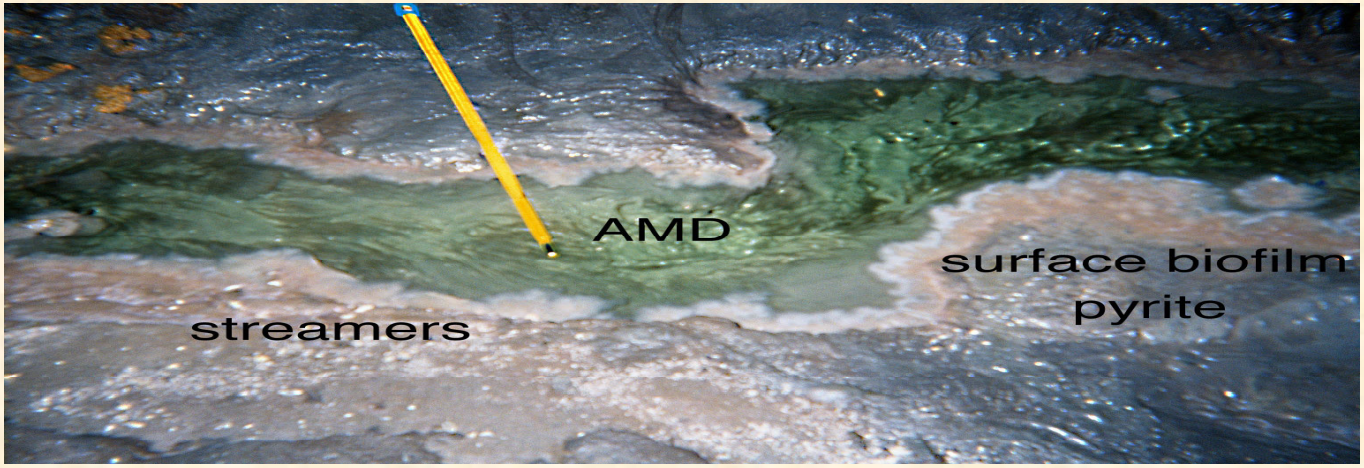
- Recent acquisition of genomic data directly from natural AMD samples has begun to reveal the genetic potential of this community (G. W. Tyson et al., Nature 428, 37, 2004). This has also enabled other systems level studies such as proteomics. **Database contains 12,148 proteins from 5 dominant species.**

The Acid Mine Drainage Site at Iron Mountain Mine, CA



The Location

The AMD streams



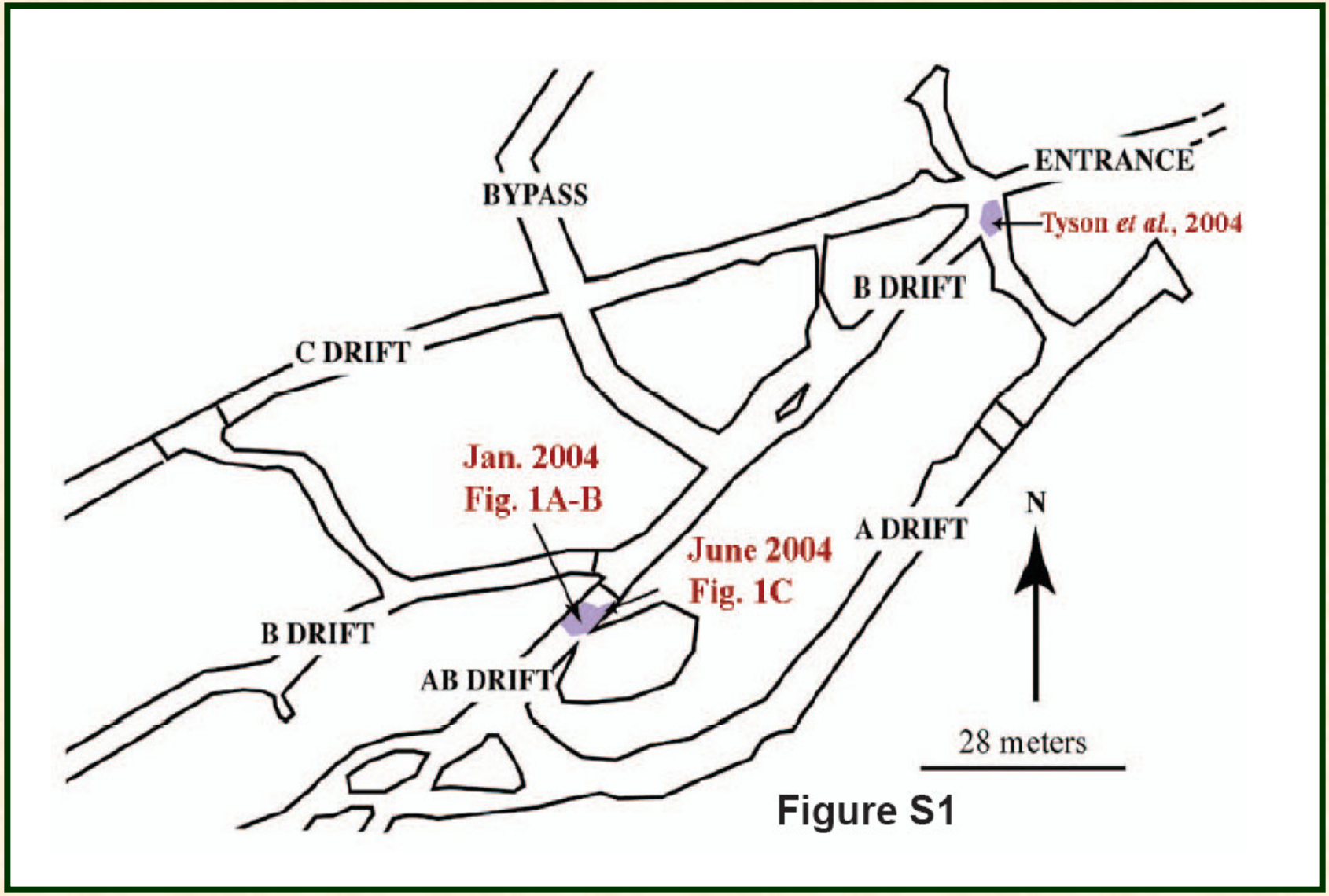
The AMD streams enhanced

Photos by KJ Edwards

And this is what you get…

Photo by KJ Edwards

AMD Collection Site:

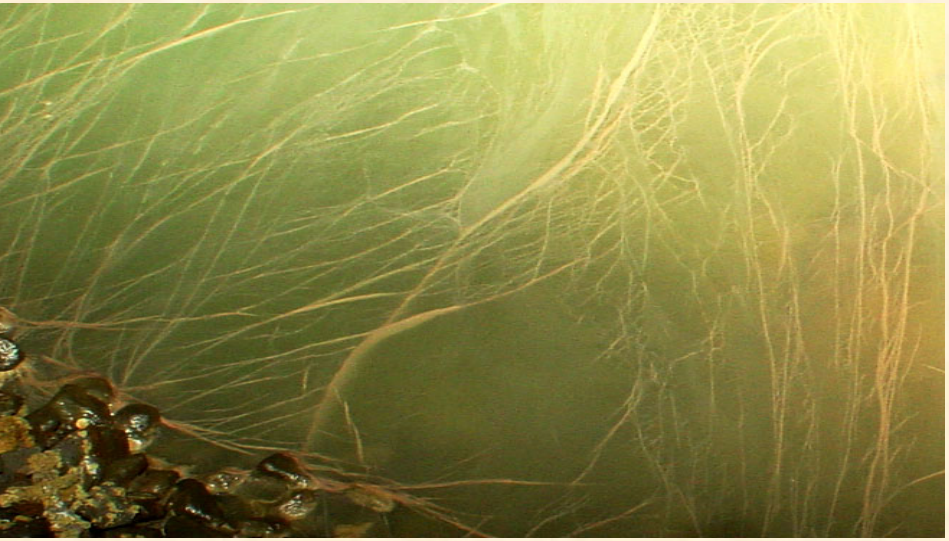


Map courtesy of Jill Banfield

Biofilm Selection

An actively growing biofilm
Similar organisms relative to genome biofilm

Before harvesting

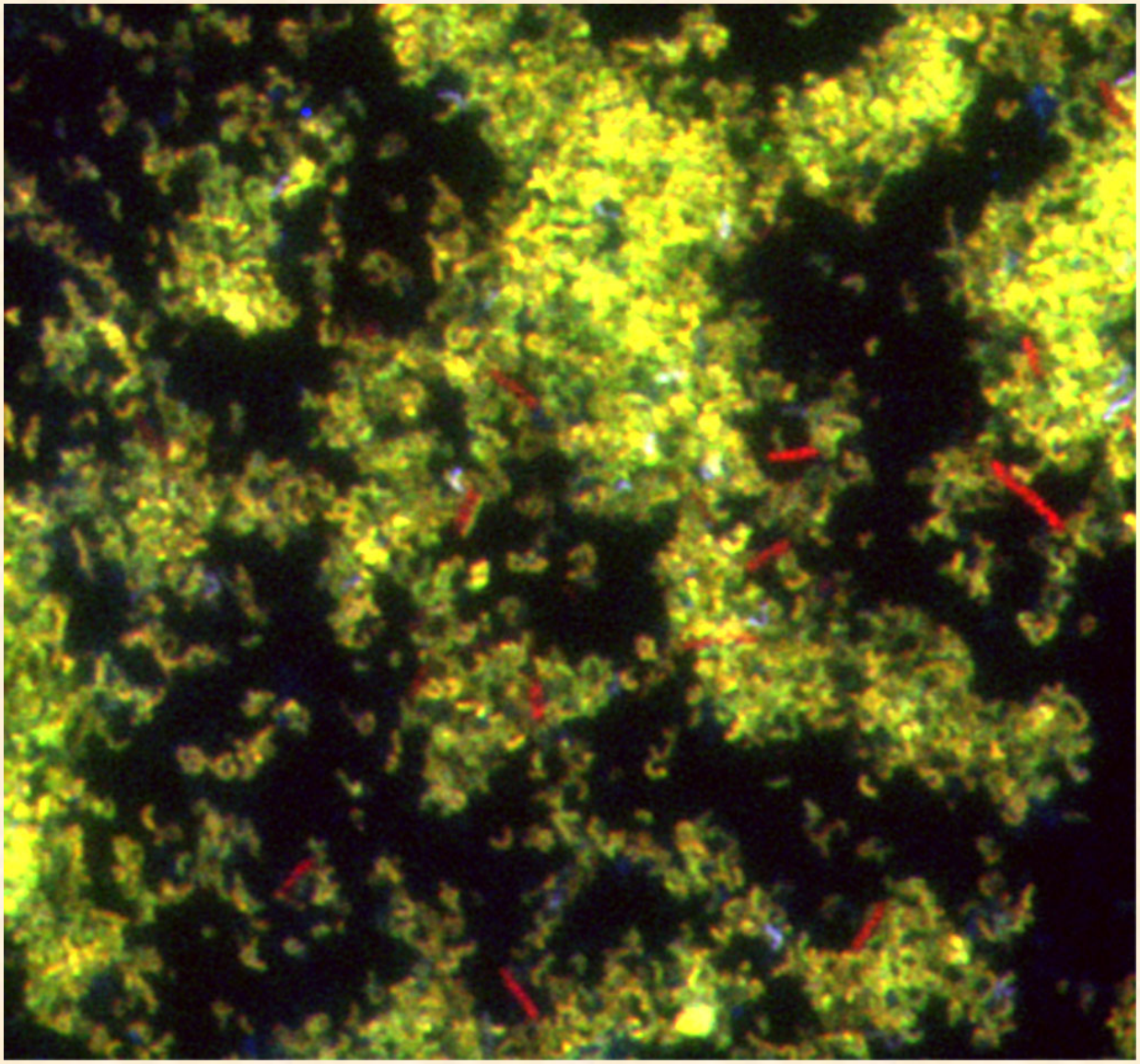


During harvesting

6 months later…

After re-harvesting

Biofilm Composition (FISH)



Archaea

***Leptospirillum* Group III**

***Leptospirillum* Group II**

All other bacteria

Vast majority are *Leptospirillum* Group II

Protein Extraction for Mass Spectrometry

Biofilm Sample

1. Wash in H_2SO_4, pH 1.1 (mine water pH)
2. Low speed centrifugation

P → **Cellular Fraction**

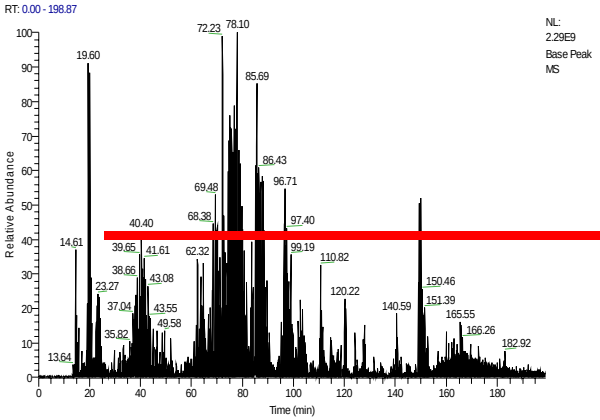
S → **Extracellular fraction**

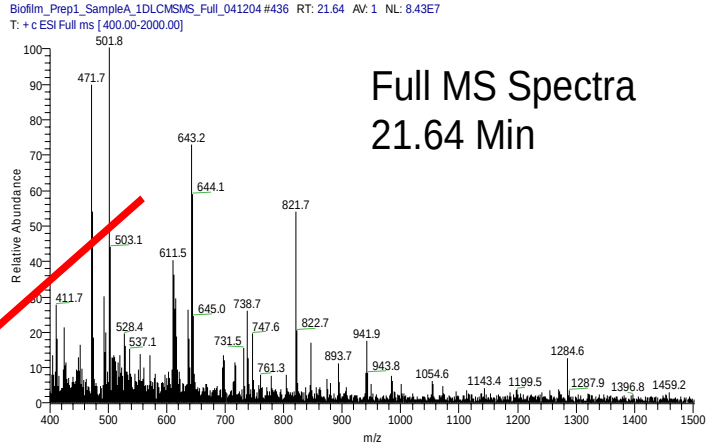
1. Sonicate to lyse
2. Low speed spin (remove sediment/unbroken cells)
3. High speed centrifugation

P → **Membrane protein-enriched fraction**

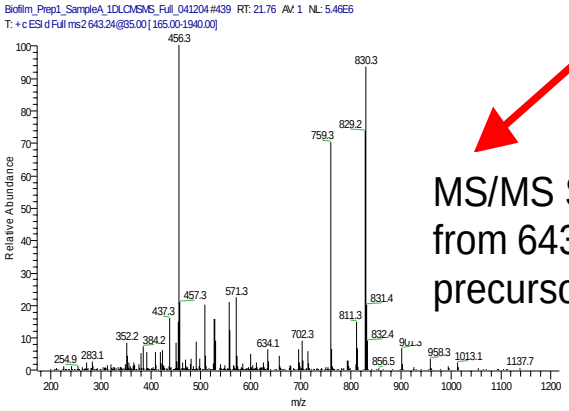
S → **Soluble protein-enriched fraction**

Preliminary Proteome Analysis of AMD sample





Full MS Spectra
21.64 Min



MS/MS Spectra
from 643m/z
precursor

SEQUEST

LeptoII_scaffold_14-GENE-20
MNKWAGAVLGTVTLGLLSATAYSAELDILKP
RVPADQIAAAKAMKPPFPVTAAVIAKGKEVFN
GAGTCYTCHGVGGKGDGPGAAGMDPSPRN
FTNHQFDQVRTAGEMVWVVSNGSPLQPAM
VGFVSAGQITDKQAWEAVMYERSLGCGGDM
DCVTGSADWVGKQPVHEEAASSLKPEYIGV
ASAH

Proteins identified in biofilm samples

Filtering Level	LCQ Dataset	LTQ Dataset	Combined Dataset
Liberal filters*	3127	5,534	5,994
Conservative filters**	1160	2,077	2,146
Ultra-conservative filters***	837	1,419	1,435

*Liberal filters requiring at least 1 peptide per gene;
**Conservative filters requiring at least 2 peptides per gene;
***Ultra Conservative filters requiring at least 3 peptides per gene.
Xcorrs of at least 1.8 (+1), 2.5 (+2) 3.5 (+3) were used in all cases.

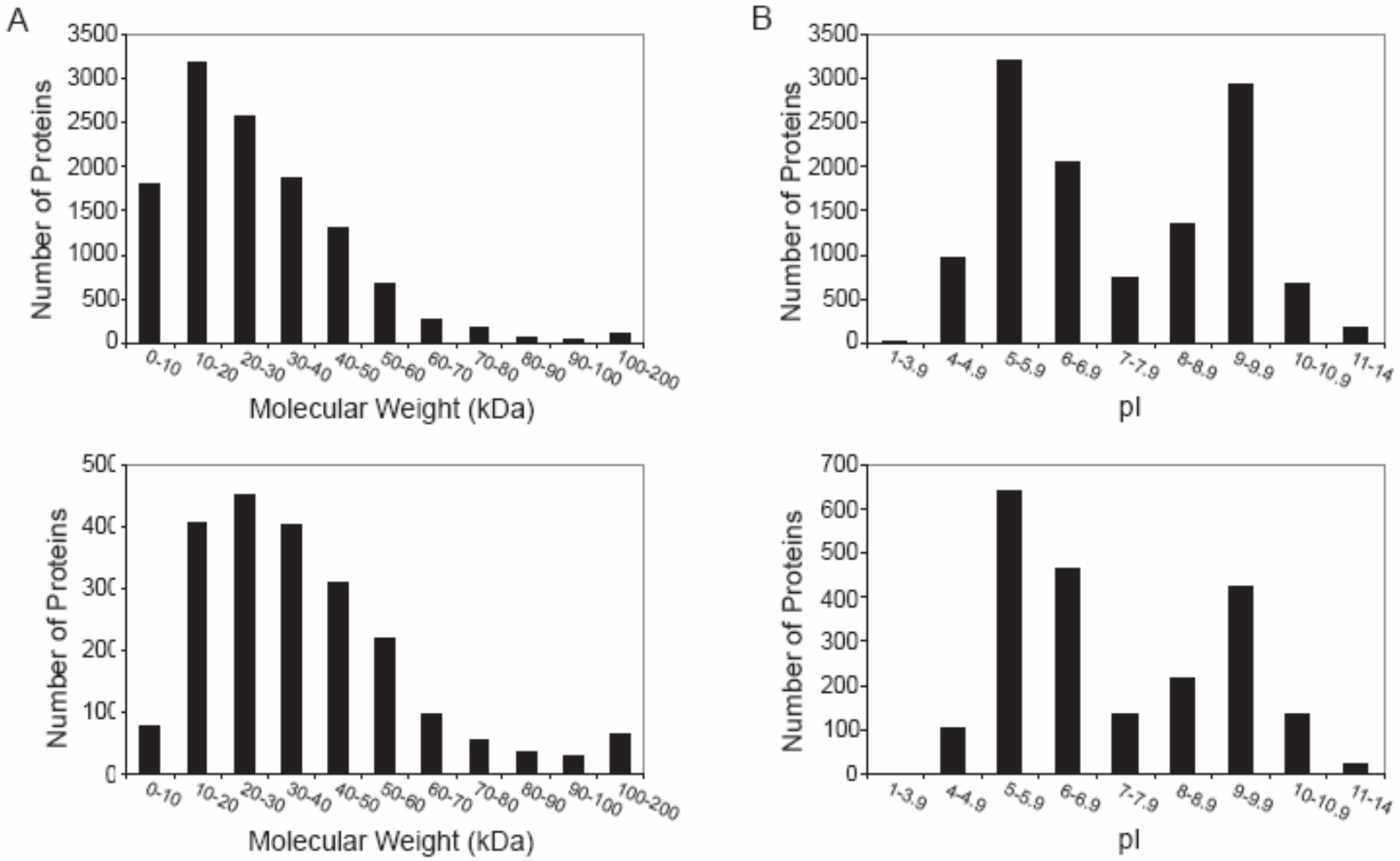
After removal of duplicate genes we confidently identified 2,036 proteins

Flat Xcorr filtering of LTQ dataset at Xcorr 2.0 with 2 peptides gave 9,169 total proteins This contains 1000s of false positives.

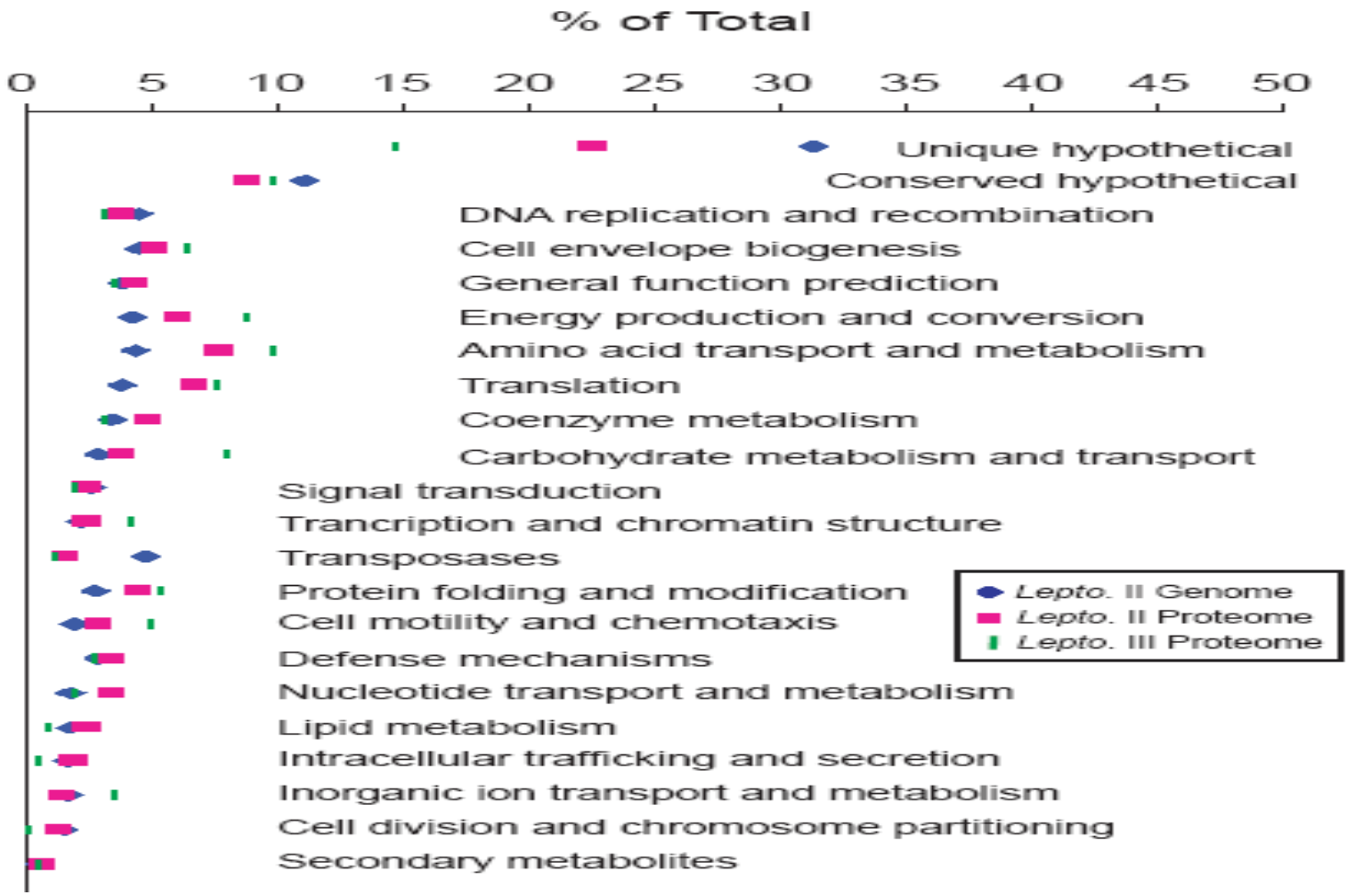
Proteins identified in biofilm samples by species type

- 1,387 proteins from *Leptospirillum* group II, representing 48% of the estimated 2,877 genes in this organism
- 268 *Leptospirillum* group III
- 84 *Ferroplasma* type I
- 99 *Ferroplasma* type II
- 120 *Gplasma* proteins were detected.
- In addition 34 proteins were detected on unassigned archaeal scaffolds and 39 on unassigned bacterial scaffolds

No significant bias in the size or pI of detected proteins



Functional Category analysis shows dominance of hypothetical proteins

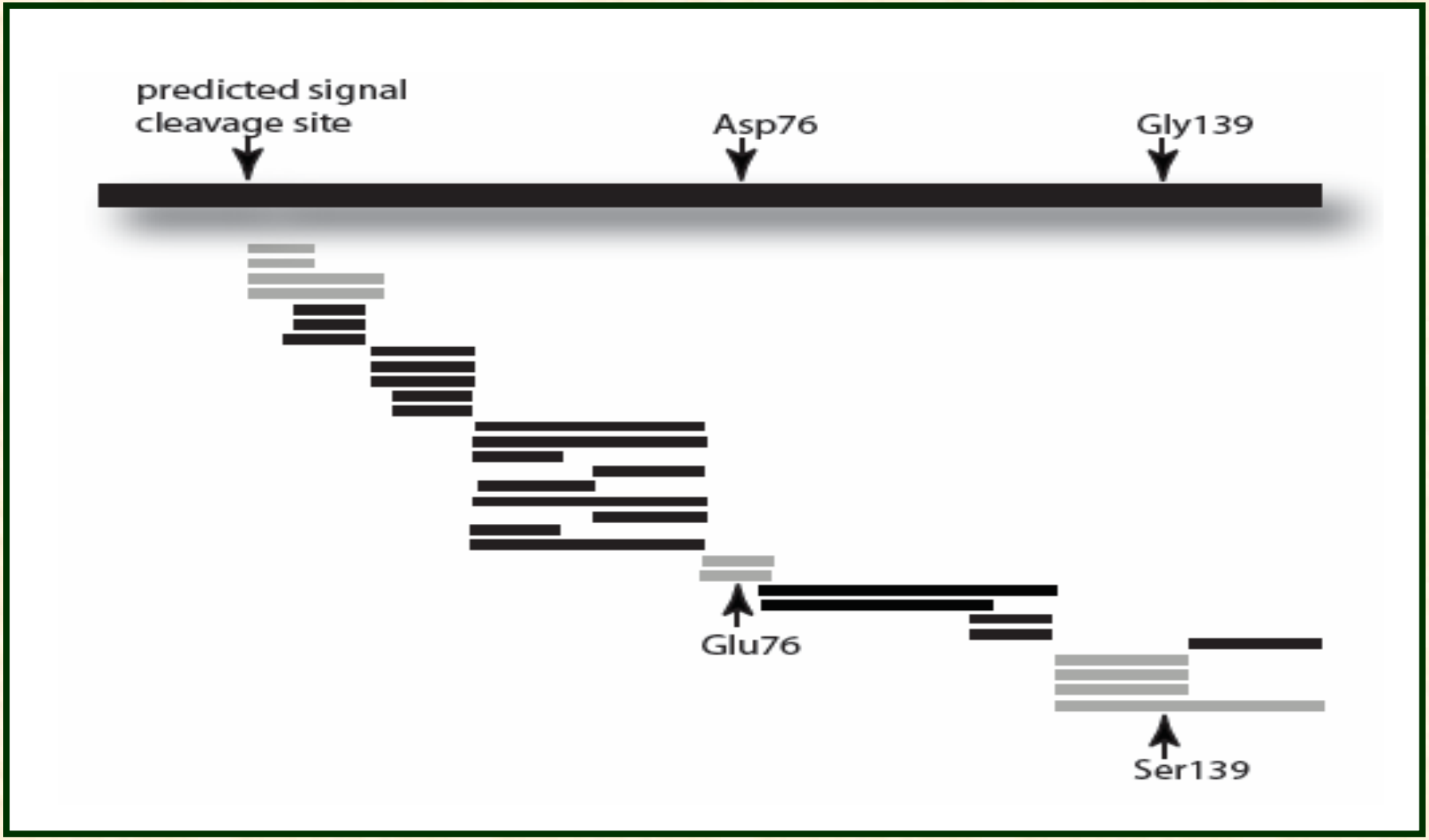


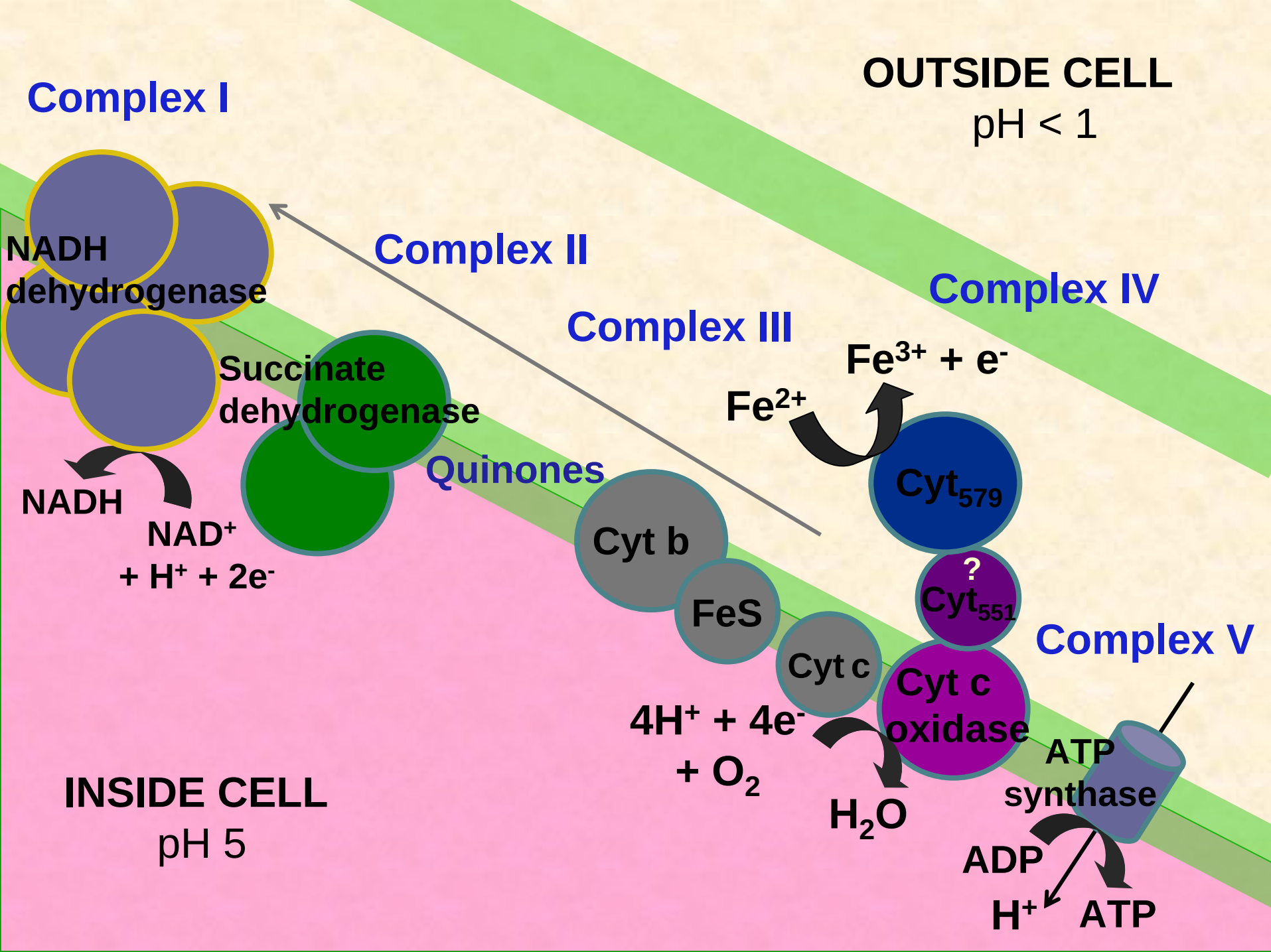
Identification of abundant *Leptospirillum* group II proteins enriched in specific fractions

very weak homology to cyt c & Fe/Pb permeases

		(% coverage)		
		Extracell.	Membr.	Soluble
Acid-stable	"unique" hypothetical confirmed	65.2	10.9	5.4
	conserved hypothetical confirmed	57.0	11.5	11.5
	"unique" hypothetical confirmed	57.1	-	-
	"unique" hypothetical confirmed	52.1	-	-
	phosphate-binding periplasmic protein	50.1	16.5	12.0
	ATP synthase, alpha subunit	-	45.3	22.8
	ATP synthase, beta subunit	-	42.1	10.5
	membrane protease	3.1	41.6	13.6
	succinyl-coA synthase subunit	-	41.3	17.3
	"unique" hypothetical confirmed	-	41.0	13.7
	conserved hypothetical confirmed	-	14.9	79.3
	conserved hypothetical confirmed	-	6.3	62.4
	bacterioferretin	-	12.8	61.4
	conserved hypothetical confirmed	-	9.4	56.8
	carbon storage regulator CsrA	-	-	47.1

Characterization of Scaff_14_GENE_20 An unknown protein from *Leptospirillum* Group II




OUTSIDE CELL
pH < 1
Complex I
NADH dehydrogenase
NADH
NAD+ + H+ + 2e-
Complex II
Succinate dehydrogenase
Quinones
Complex III
Cyt b
FeS
Cyt c
Complex IV
Fe3+ + e-
Fe2+
Cyt579
?
Cyt551
Cyt c oxidase
4H+ + 4e- + O2
H2O
Complex V
ATP synthase
ADP
H+
ATP
INSIDE CELL
pH 5

Conclusions

- We confidently identified 2,036 unique proteins from the Acid Mine Drainage biofilm community. This is the first example of a large scale proteome analysis of a natural microbial community.

- The large number of detected unique and conserved novel proteins underscores the importance of proteins of unknown function in the community. Novel proteins that were detected and abundant may be targeted for purification from biofilm samples and subsequent functional analysis.

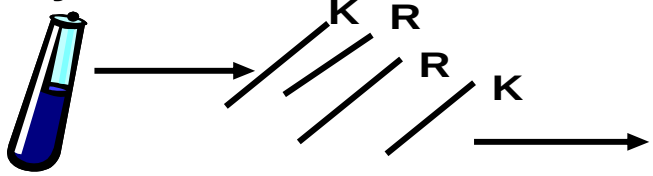
- This study illustrates the power of an integrated community genomic and proteomic approach and paves the way for future analysis of microbial communities.

Talk and poster can be downloaded from:
http://www.ornl.gov/sci/csd/Research_areas/obms_rd.html

Future Plans

- Deeper coverage through 3D-LC and increase dynamic range with FT-ICR.
- Sequence tagging or *de-novo* sequencing approaches.
- Absolute and relative quantitation for exploring spatial and temporal changes in the community.

Deep analysis requires greater dynamic range
Solution: 3D-LC/MS/MS with a LTQ-FT-MS



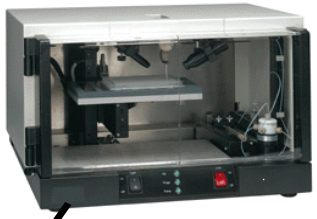
Tryptic Digestion of Mixed Bacterial Lysates

HPLC

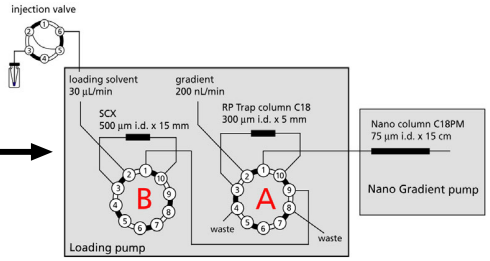
1 mm I.D columns
Flow Rate ~200ul/min

1D-Separation
C18 Reverse Phase
or IEF
or AEX
or Size Exclusion

Probot fraction collection in microtiter plates



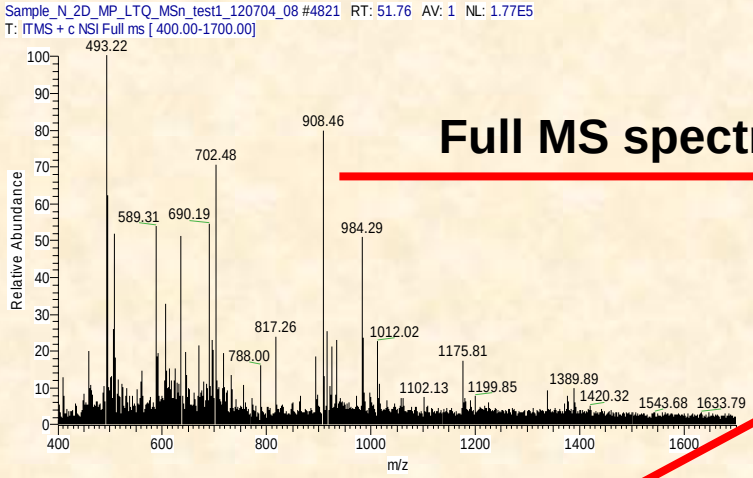
Autosampler makes injections on to the 2D system from fractions in microtiter plates



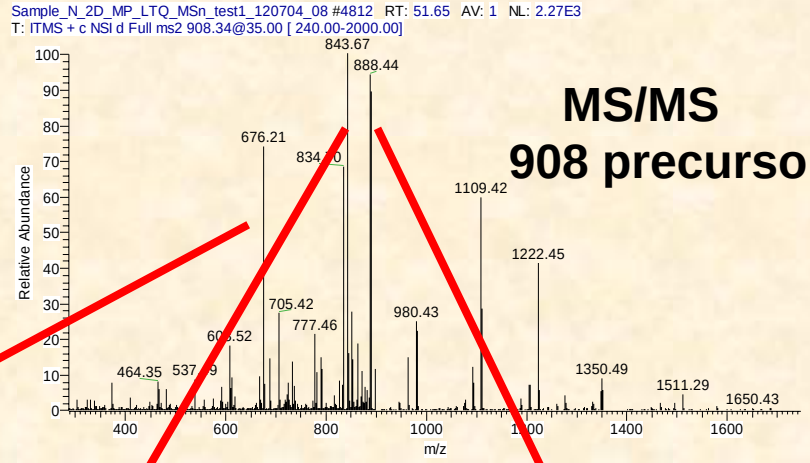
Automated 2D HPLC
1st Dimension: SCX
2nd Dimension: Nano RP C18

ES-LTQ-FT-ICR

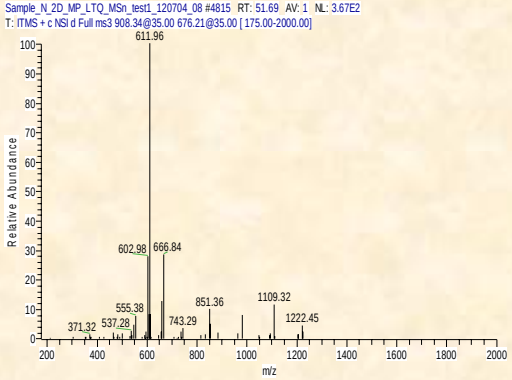
MS^3 for verification of *de-novo* sequencing attempts



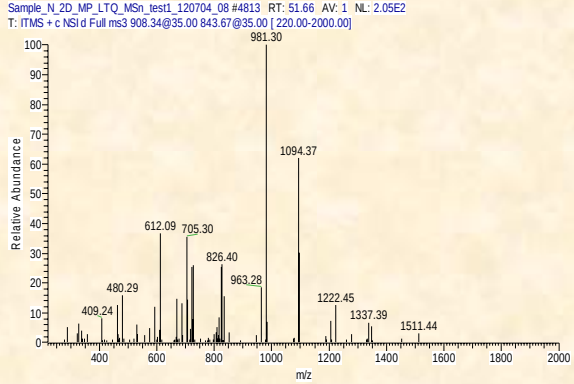
Full MS spectra



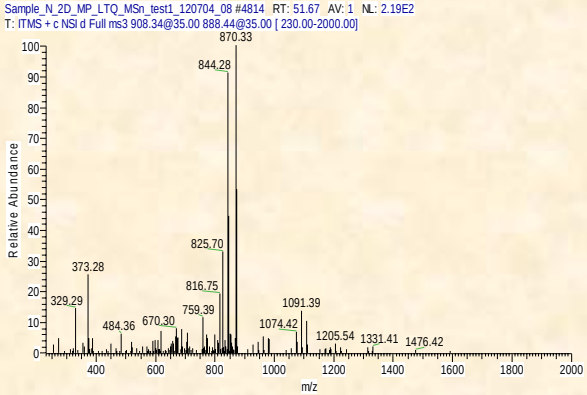
MS/MS
908 precursor



MS^3 of 676 fragment ion



MS^3 of 843 fragment ion



MS^3 of 888 fragment ion

Acknowledgements

- Yates Lab at Scripps and David Tabb for DTASelect/Contrast
- UT-ORNL Graduate program in Genomic Science and Technology
- U.S. Department of Energy Office of Biological and Environmental Research
  Genomes to Life Project
  Microbial Cell Project