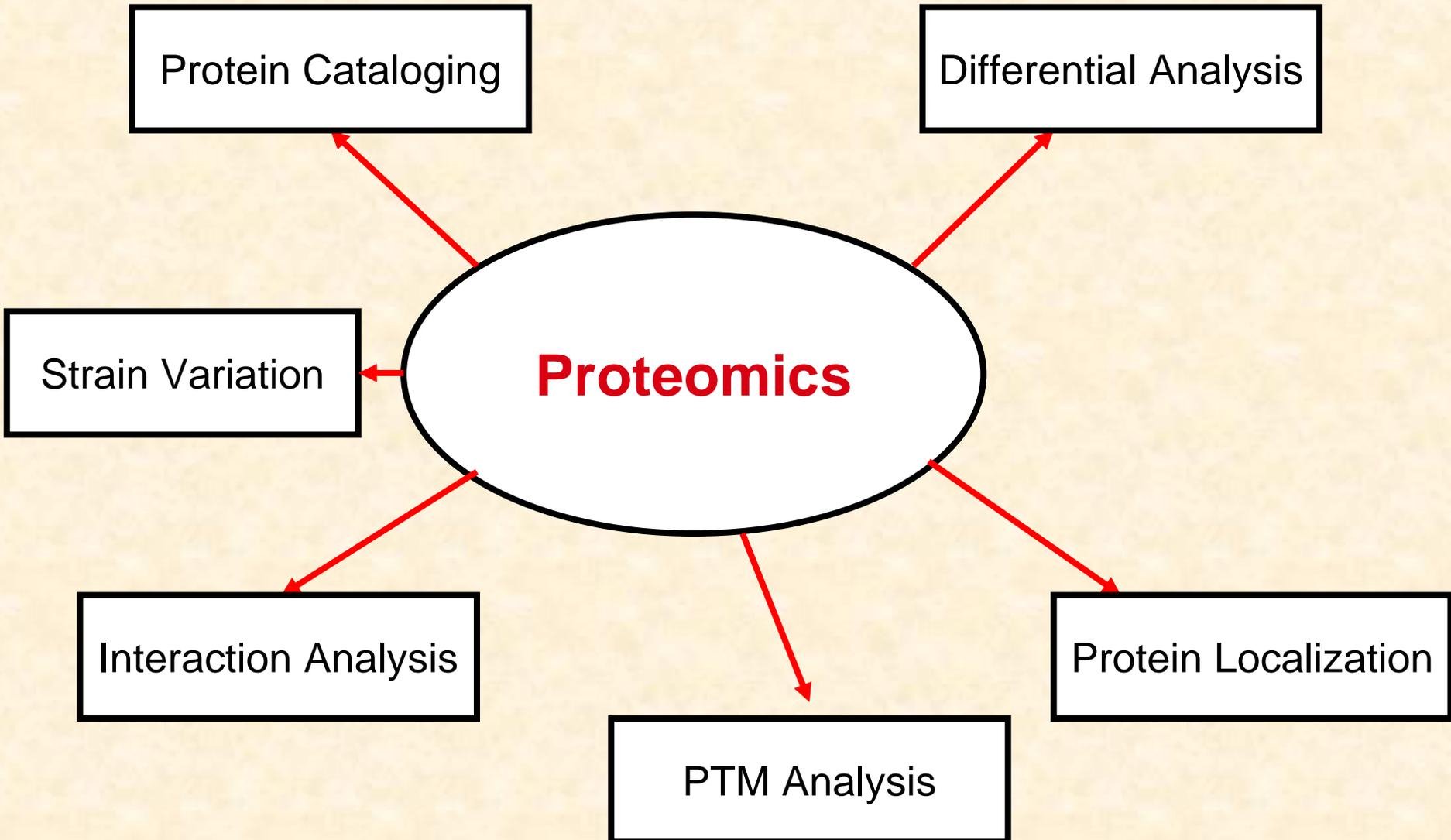


# **Proteomics Approaches for Characterizing Microbial Proteomes from Communities to Environments**

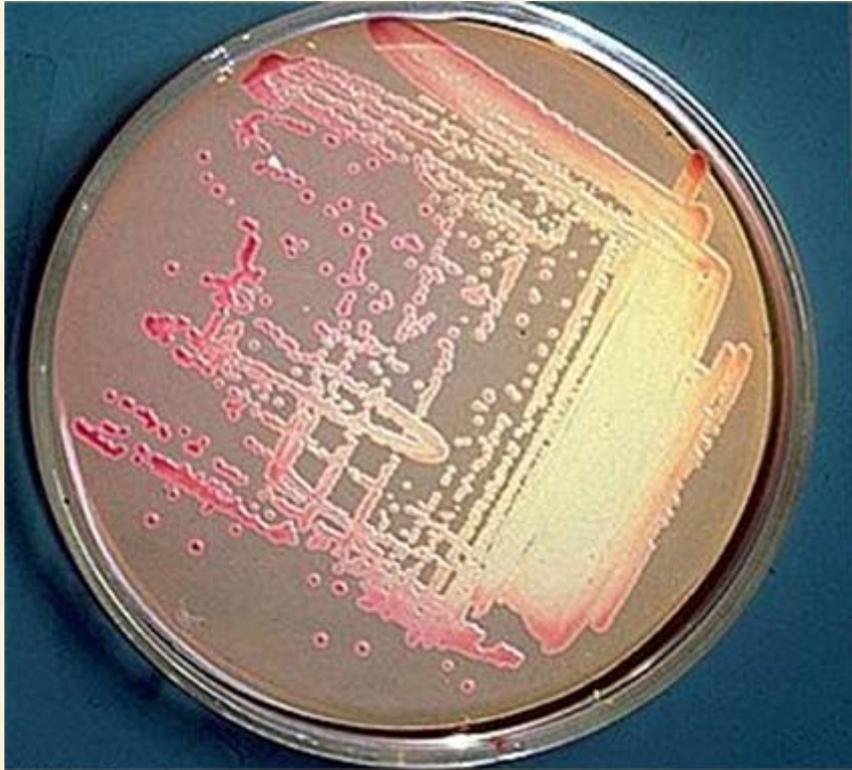
**Nathan VerBerkmoes**

**Organic and Biological Mass Spectrometry Group  
Oak Ridge National Laboratory**

# Major Applications of MS-based Proteomics

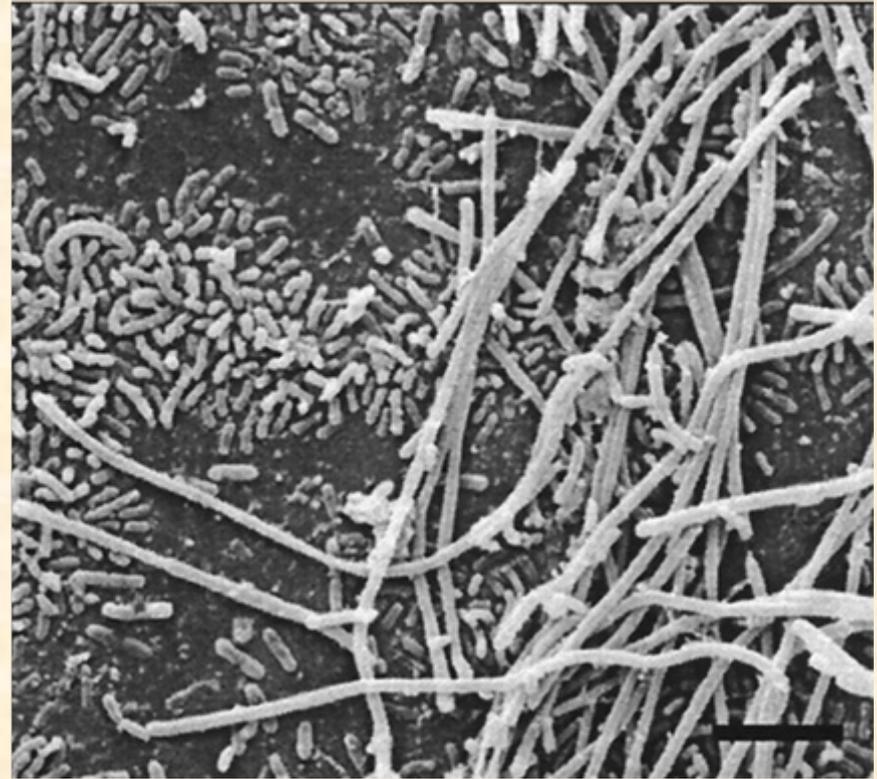


## Isolates



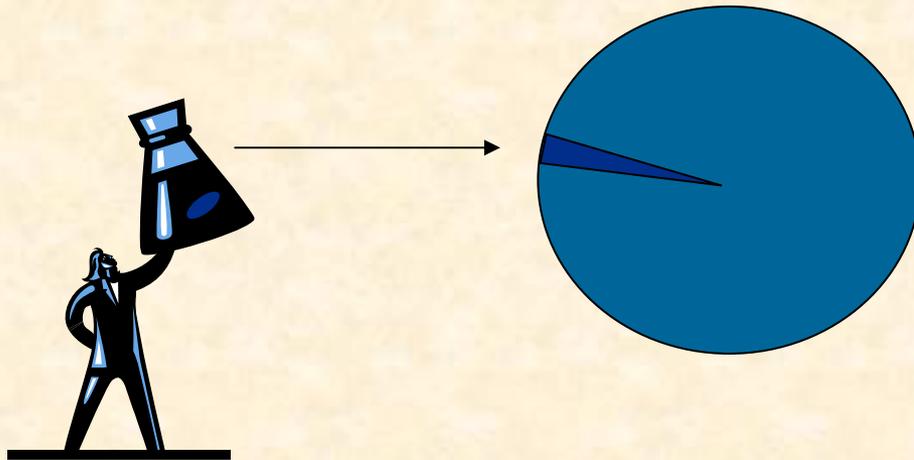
≠

## Communities



Historically, study of microorganisms required their growth in the laboratory, but this can be impossible!

# Why study microbes in the environment?



**Most microorganisms  
are not culturable**

## Research Questions:

Which factors control community assembly at the species and strain level?

How are metabolic activity levels and partitioning of metabolic tasks affected by community makeup and the physical and chemical surroundings?

How are carbon, nitrogen, and energy resources allocated into metabolic pathways?

# How might we study microbial consortia?

- 1) Choose an appropriate model ecosystem
- 2) Obtain comprehensive genomic information via cultivation-independent methods (**Metagenomics**)
- 3) Identify important functions for each community member simultaneously, and analyze the ecosystem dynamics, *in situ* (**Metaproteomics**)
- 4) Target key proteins of unknown function for functional analysis
- 5) Extend the approach to more complex systems

# Too many terms for the same thing?

- 1) Metaproteomics
- 2) Community Proteomics
- 3) Proteogenomics

# 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:*

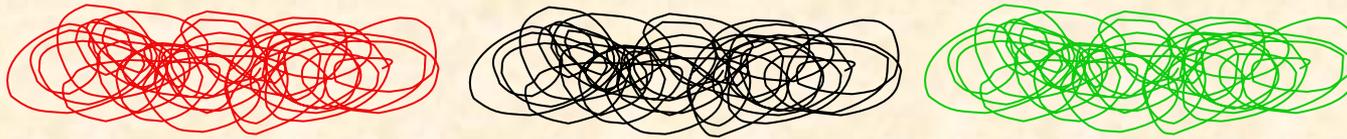
- 1) The level of DNA sequence information and quality annotation available on the community.**
- 2) The level of diversity and dynamic range associated with the species of interest in the community.**
- 3) The quantity of biomass available for study.**
- 4) The level of interrelatedness and/or diversity at the base pair level amongst members of the same species in the community and between species in the community.**

## What can we learn from community proteomics?

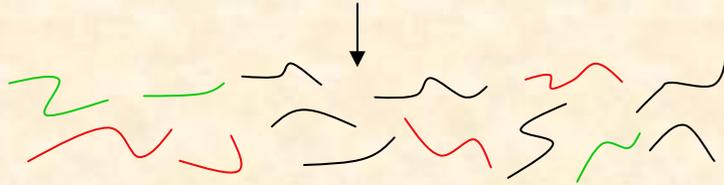
- The microbial community make up, who is there and what are their abundances.
- Who is doing what (nitrogen fixation, carbon fixation, etc)
- What precise metabolic pathways are active, how are energy resources allocated through these pathways
- How are the microbial communities effecting their environments, (Sulfur reduction, Uranium reduction etc) and which key proteins and pathways are involved.
- How are the members of the microbial communities interacting, exchanging molecular information, resources and energy (i.e. expression of nanowires).
- What unknown or hypothetical proteins are abundant and important to the microbial communities and their effects on the environment.



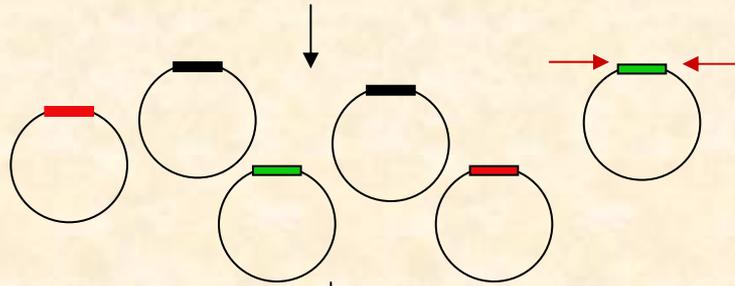
# Environmental Microbial Community Genomics:



Extract DNA from natural sample

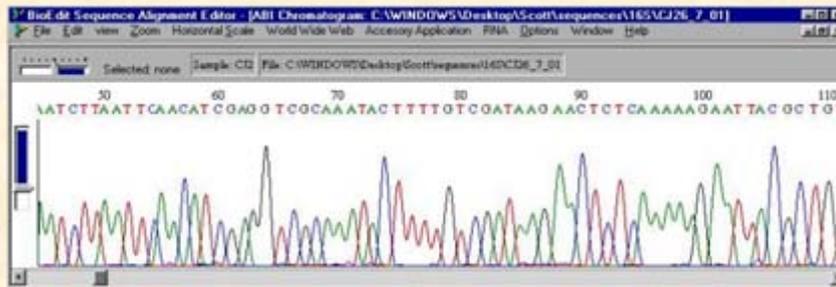


Shear



3 – 4 kb shotgun library

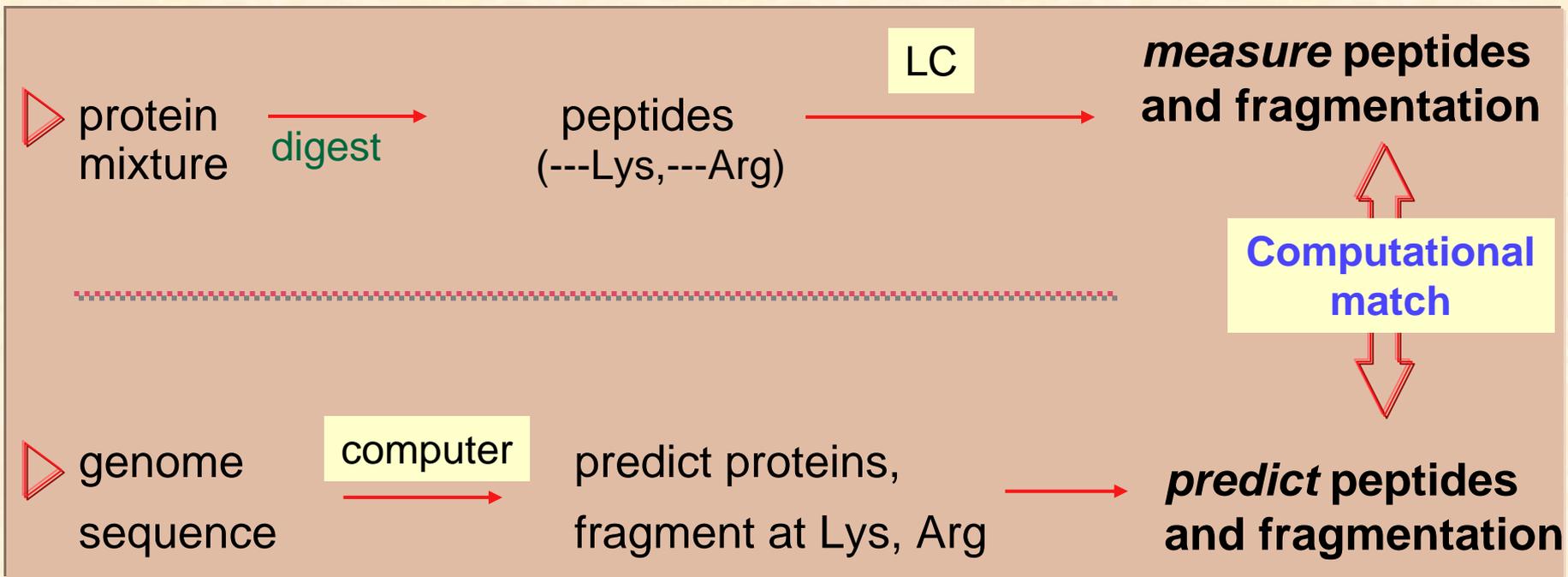
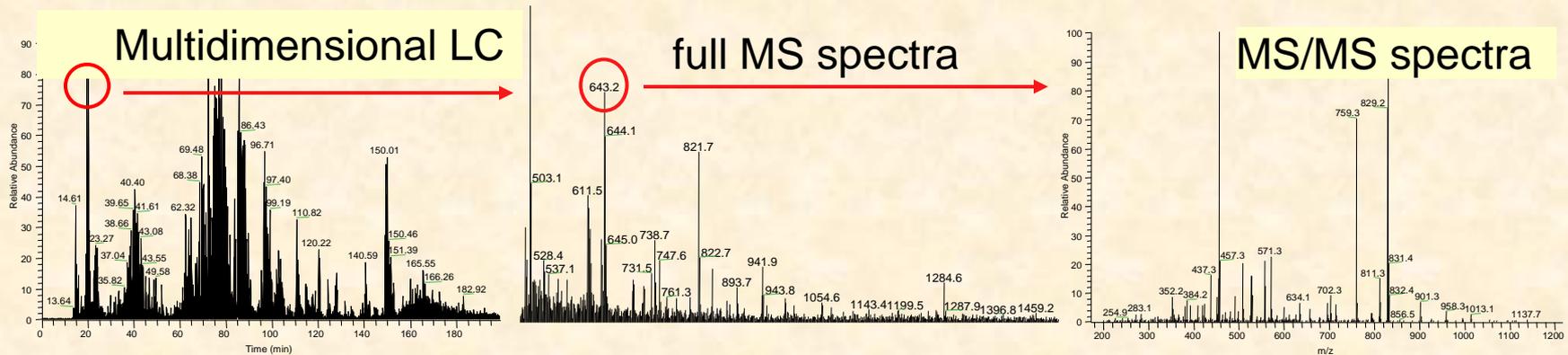
Tyson et al. Nature 2004



End sequence clones

...ACGGCTGCGTTACATCGATCAT  
ACATCGATCATTACGATACCATTG...

Assemble reads by alignment identity



***Leptospirillum* group II\_scaffold\_14\_GENE\_20**

MNKWAGAVLGTVTLGLLSATAYS AELDILKPNRVPADQIAAAKAMKPPFPVTA  
 AVIAKGKEVFNGAGTCYTCHGVGGK **GDGPGAAGMDPSR**FTNHQFDQVRTAGE  
 MVWVVSNGSPLQPAMVGFVSAGITDKQAWAVMYERSLGC GGDMDC.....

# ORNL Proteome Informatics Pipeline

- 1) Extraction of MS, MS/MS spectra from instrument capture files into text files
- 2) Assignment of precursor ion charge state(s) to each tandem mass spectrum
- 3) Identification of peptide for each tandem mass spectrum with SEQUEST search engine (or DBDigger, Inspect, MyriMatch)
- 4) Assembly and filtering of SEQUEST output files with DTASelect
- 5) If high resolution data is available comparison of observed mass and calculated mass (p.p.m difference)
- 6) Comparisons of run-to-run variations and sample-to-sample changes with Contrast and statistical approaches (Zhang, Journal of Proteome Research 2007)
- 7) Creation of web-based data sets and extractions to give easy access to collaborators and general public after publication.  
[http://compbio.ornl.gov/biofilm\\_and\\_recombination/](http://compbio.ornl.gov/biofilm_and_recombination/)
- 8) Import of data into MySQL database

# ORNL Proteome Informatics Pipeline

ORNL - UC Berkeley AMD (Arid Mine Drainage) Community Proteome Study - Results Summary - Results Profile

Sample Set	Environment	Year	Image	pH	Temp	Classification	DOC	Number	Notes
1	AD Event	January 2004		1.07	42.3 °C	TBA	0-1	Final/Control (Final/Control)	EMM2 (Final/Control)
2	AD Event	June 2004		0.99	39 °C	TBA	2(3)	Final	2 (Final)
3	AD Event	June 2004		0.99	39 °C	TBA	2(3)	Final	2 (Final)
4	TBA (Final/Control)	June 2004		1.20	30 °C	TBA	3(4)	Final	3 (Final)
5	Control	June 2004		-	-	TBA	-	Final	Partial proteome
6-022	AD Event (022)	November 2006		0.84	43 °C	Conductivity 95.5 uS/cm	2	Final	2 (Final)
6	AD Event (Final)	November 2006		0.84	43 °C	Conductivity 95.5 uS/cm	2	Final	2 (Final)
7	Nonreactive	November 2006		0.95	17 °C	TBA	0-1	Final	Partial proteome

ORNL - UC Berkeley AMD (Arid Mine Drainage) Community Proteome Study - Results Summary - Results Profile

ORNL [www.ornl.gov](http://www.ornl.gov) ORNL 06/06/06

Global Summary

| Sample |
|--------|--------|--------|--------|--------|--------|--------|--------|
| HTM    | Test   | HTM    | Test   | HTM    | Test   | HTM    | Test   |

DTASelect

| Sample |
|--------|--------|--------|--------|--------|--------|--------|--------|
| HTM    | Test   | HTM    | Test   | HTM    | Test   | HTM    | Test   |

Peptides

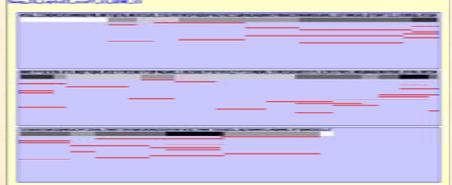
[Final](#) [Exclude Control](#) [Non-Subset Control](#)

ORNL - UC Berkeley AMD (Arid Mine Drainage) Community Proteome Study - Results Summary - Results Profile

Rank	Protein	Score
1	... (Protein Name)	...
2	...	...
3	...	...

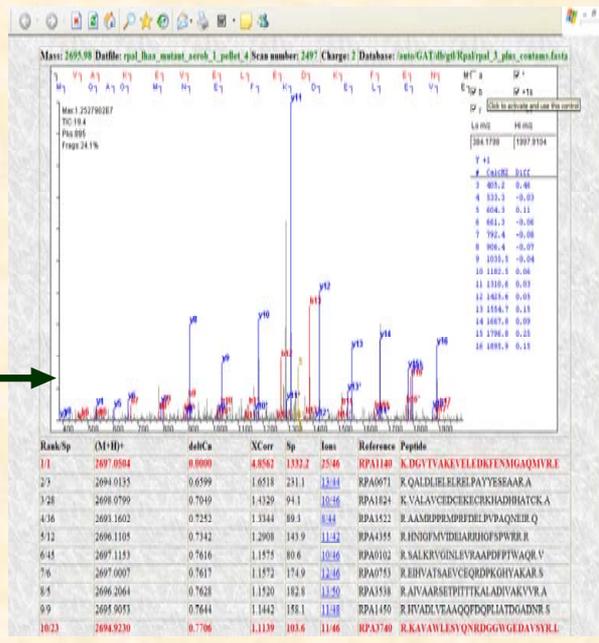
ORNL - UC Berkeley AMD (Arid Mine Drainage) Community Proteome Study - Results Summary - Results Profile

Rank	Protein	Score
4	...	...
5	...	...
6	...	...



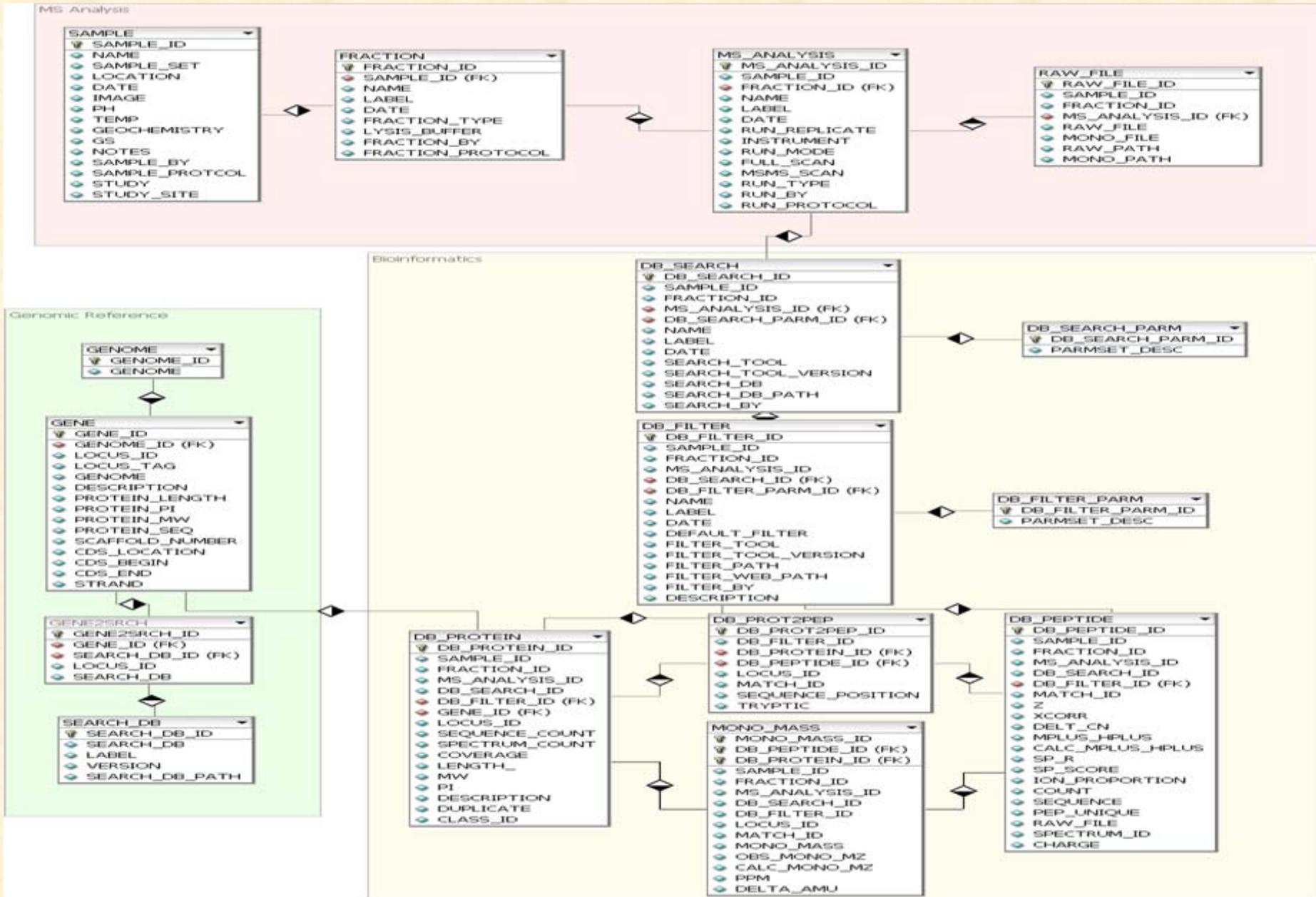
ORNL - UC Berkeley AMD (Arid Mine Drainage) Community Proteome Study - Results Summary - Results Profile

RP1A140	74	422	88.83	547	57624	5.5	gprot1 (chaperonin Gln1.1, gprot1.128783.128949) mouse MV 07626
1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9
10	10	10	10	10	10	10	10
11	11	11	11	11	11	11	11
12	12	12	12	12	12	12	12
13	13	13	13	13	13	13	13
14	14	14	14	14	14	14	14
15	15	15	15	15	15	15	15
16	16	16	16	16	16	16	16
17	17	17	17	17	17	17	17
18	18	18	18	18	18	18	18
19	19	19	19	19	19	19	19
20	20	20	20	20	20	20	20
21	21	21	21	21	21	21	21
22	22	22	22	22	22	22	22
23	23	23	23	23	23	23	23
24	24	24	24	24	24	24	24
25	25	25	25	25	25	25	25
26	26	26	26	26	26	26	26
27	27	27	27	27	27	27	27
28	28	28	28	28	28	28	28
29	29	29	29	29	29	29	29
30	30	30	30	30	30	30	30
31	31	31	31	31	31	31	31
32	32	32	32	32	32	32	32
33	33	33	33	33	33	33	33
34	34	34	34	34	34	34	34
35	35	35	35	35	35	35	35
36	36	36	36	36	36	36	36
37	37	37	37	37	37	37	37
38	38	38	38	38	38	38	38
39	39	39	39	39	39	39	39
40	40	40	40	40	40	40	40
41	41	41	41	41	41	41	41
42	42	42	42	42	42	42	42
43	43	43	43	43	43	43	43
44	44	44	44	44	44	44	44
45	45	45	45	45	45	45	45
46	46	46	46	46	46	46	46
47	47	47	47	47	47	47	47
48	48	48	48	48	48	48	48
49	49	49	49	49	49	49	49
50	50	50	50	50	50	50	50
51	51	51	51	51	51	51	51
52	52	52	52	52	52	52	52
53	53	53	53	53	53	53	53
54	54	54	54	54	54	54	54
55	55	55	55	55	55	55	55
56	56	56	56	56	56	56	56
57	57	57	57	57	57	57	57
58	58	58	58	58	58	58	58
59	59	59	59	59	59	59	59
60	60	60	60	60	60	60	60
61	61	61	61	61	61	61	61
62	62	62	62	62	62	62	62
63	63	63	63	63	63	63	63
64	64	64	64	64	64	64	64
65	65	65	65	65	65	65	65
66	66	66	66	66	66	66	66
67	67	67	67	67	67	67	67
68	68	68	68	68	68	68	68
69	69	69	69	69	69	69	69
70	70	70	70	70	70	70	70
71	71	71	71	71	71	71	71
72	72	72	72	72	72	72	72
73	73	73	73	73	73	73	73
74	74	74	74	74	74	74	74
75	75	75	75	75	75	75	75
76	76	76	76	76	76	76	76
77	77	77	77	77	77	77	77
78	78	78	78	78	78	78	78
79	79	79	79	79	79	79	79
80	80	80	80	80	80	80	80
81	81	81	81	81	81	81	81
82	82	82	82	82	82	82	82
83	83	83	83	83	83	83	83
84	84	84	84	84	84	84	84
85	85	85	85	85	85	85	85
86	86	86	86	86	86	86	86
87	87	87	87	87	87	87	87
88	88	88	88	88	88	88	88
89	89	89	89	89	89	89	89
90	90	90	90	90	90	90	90
91	91	91	91	91	91	91	91
92	92	92	92	92	92	92	92
93	93	93	93	93	93	93	93
94	94	94	94	94	94	94	94
95	95	95	95	95	95	95	95
96	96	96	96	96	96	96	96
97	97	97	97	97	97	97	97
98	98	98	98	98	98	98	98
99	99	99	99	99	99	99	99
100	100	100	100	100	100	100	100

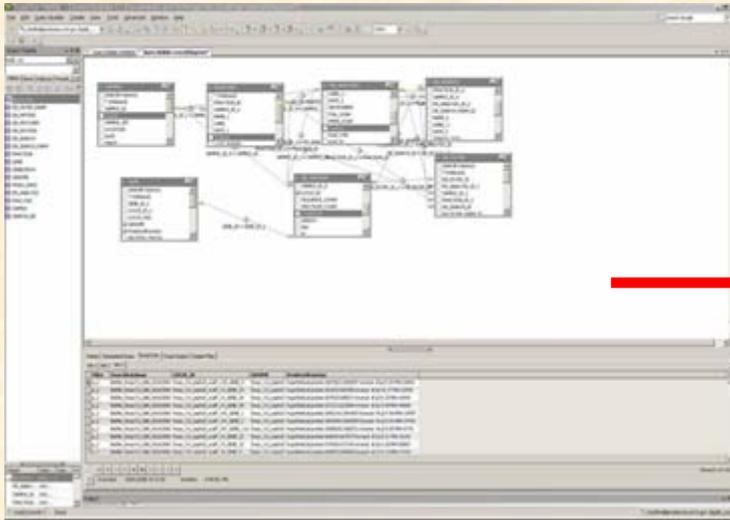


[http://compbio.ornl.gov/amd\\_gtl\\_ms\\_results](http://compbio.ornl.gov/amd_gtl_ms_results)

# MySQL Database Construction



# MySQL Database Queries



Sample	LOCUS_ID	SEQUENCE_COUNT	SPECTRUM_COUNT	COVERAGE	PredictedFunction
ABmuckGS2	5way_CG_Leptoll_scaff_630_GENE_6	26	871	35.1	hypothetical protein 2710915.2712627 reverse #pI5.11 MW61312
ABmuckGS2	5way_CG_Leptoll_scaff_630_GENE_6	27	799	33.7	hypothetical protein 2710915.2712627 reverse #pI5.11 MW61312
ABfrott2	5way_CG_Leptoll_scaff_8_GENE_1	25	709	62	hypothetical protein 502204.502657 forward #pI10.92 MW9687
ABfrott2	UBA_Leptoll_Scaffold_8634_GENE_180	25	709	62	Hypothetical protein 168070.168423 forward #pI10.92 MW9687
ABfrott2	5way_CG_Leptoll_scaff_8_GENE_1	29	682	75	hypothetical protein 502204.502657 forward #pI10.92 MW9687
ABfrott2	UBA_Leptoll_Scaffold_8634_GENE_180	29	682	75	Hypothetical protein 168070.168423 forward #pI10.92 MW9687
ABfrott2	5way_CG_Leptoll_scaff_630_GENE_6	31	681	35.1	hypothetical protein 2710915.2712627 reverse #pI5.11 MW61312
ABmuckGS2	5way_CG_Leptoll_scaff_630_GENE_6	29	680	35.1	hypothetical protein 2710915.2712627 reverse #pI5.11 MW61312
ABfrott2	5way_CG_Leptoll_scaff_630_GENE_6	32	653	29.8	hypothetical protein 2710915.2712627 reverse #pI5.11 MW61312
ABfrott3	5way_CG_Leptoll_scaff_8_GENE_1	32	642	78.3	hypothetical protein 502204.502657 forward #pI10.92 MW9687
ABfrott3	UBA_Leptoll_Scaffold_8634_GENE_180	32	642	78.3	Hypothetical protein 168070.168423 forward #pI10.92 MW9687
ABmuckFriable	5way_CG_Leptoll_scaff_630_GENE_6	22	617	33.7	hypothetical protein 2710915.2712627 reverse #pI5.11 MW61312
ABfrott3	5way_CG_Leptoll_scaff_8_GENE_1	26	576	78.3	hypothetical protein 502204.502657 forward #pI10.92 MW9687
ABfrott3	UBA_Leptoll_Scaffold_8634_GENE_180	26	576	78.3	Hypothetical protein 168070.168423 forward #pI10.92 MW9687
ABfrott3	5way_CG_Leptoll_scaff_630_GENE_6	24	550	24.7	hypothetical protein 2710915.2712627 reverse #pI5.11 MW61312
ABfrott3	5way_CG_Leptoll_scaff_630_GENE_6	27	525	29.8	hypothetical protein 2710915.2712627 reverse #pI5.11 MW61312
ABfrott3	5way_CG_Leptoll_scaff_630_GENE_6	26	519	29.8	hypothetical protein 2710915.2712627 reverse #pI5.11 MW61312
ABmuckFriable	5way_CG_Leptoll_scaff_630_GENE_6	26	493	35.1	hypothetical protein 2710915.2712627 reverse #pI5.11 MW61312
ABmuckFriable	5way_CG_Leptoll_scaff_630_GENE_6	27	470	33.7	hypothetical protein 2710915.2712627 reverse #pI5.11 MW61312
ABfrott3	5way_CG_Leptoll_scaff_630_GENE_6	27	440	24.7	hypothetical protein 2710915.2712627 reverse #pI5.11 MW61312
ABfrott3	5way_CG_Leptoll_scaff_630_GENE_6	25	435	29.8	hypothetical protein 2710915.2712627 reverse #pI5.11 MW61312
ABfrott3	5way_CG_Leptoll_scaff_45_GENE_15	12	423	44.5	hypothetical protein 1338454.1340110 forward #pI9.91 MW23967
ABfrott3	5way_CG_Leptoll_scaff_630_GENE_6	23	417	23.3	hypothetical protein 2710915.2712627 reverse #pI5.11 MW61312
ABfrott3	5way_CG_Leptoll_scaff_6_GENE_19	46	405	58	hypothetical protein 353626.354226 reverse #pI9.70 MW22551

- To make use of the MySQL database, queries need to be built that can interrogate the data in a meaningful format.
- This can be done through command lines programs or through GUI interfaces, such as the Freeware Toad for MySQL.
- The following examples use queries built in the Toad GUI; the results can be easily extracted into Excel.
- The first step is to load the entire database into Toad. This step produces a large list of **135,123 protein entries** when filtered at 2 peptides per proteins. When filtered for only unique entries, **6,723 proteins** are identified from the entire dataset.
- A first simple query was done for all conserved hypothetical and hypothetical proteins. This retrieved 1,327 entries.
- Retrieval of the most abundant hypothetical proteins, as estimated by spectral counts (>100), gave 247 total protein entries.
- One abundant hypothetical protein (5way\_CG\_Leptoll\_scaff\_98\_GENE\_16) was sub-queried to determine exactly what samples it was detected in and to what level.
- The protein was found in all field samples, but at the highest abundance level in the AB Muck field sample



# Mass Spectrometry for Quantitation

*Mass Spectrometry is considered to be one of the most accurate analytical methods for quantitation. So why is protein/peptide quantitation so difficult?*

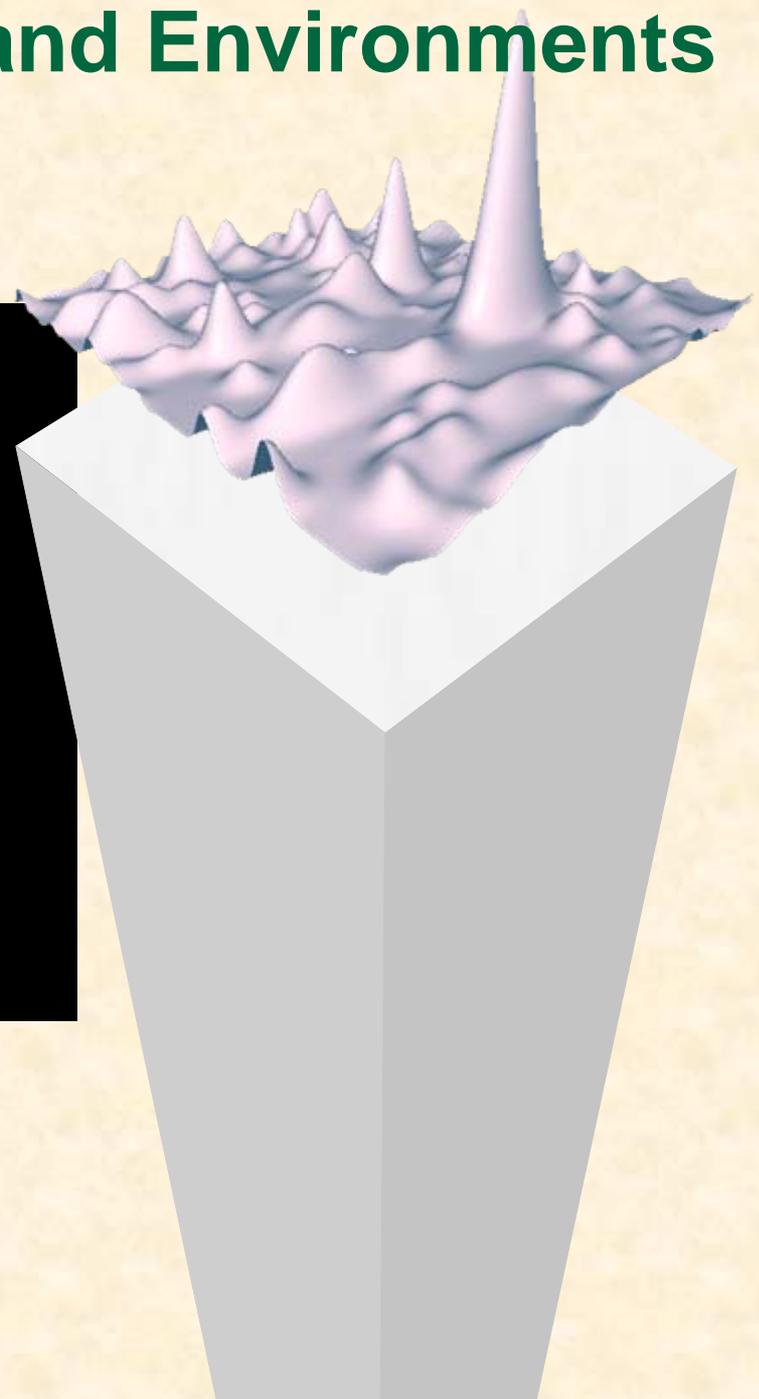
- Complexity of proteome samples
- Dynamic range necessary for quantitation of proteins
- Difficulties in preparing complex proteome samples exactly the same
- Differential digestion problems
- Losses in peptide/protein separations
- Matrix effects upon ionization for both MALDI and ESI
- Protein synthesis, protein pools, protein degradation
- Absolute vs. Relative quantification

**Problem even worse in microbial communities**

*The signal of proteins up and down regulation is mixed with changing concentrations of microbial species.*

# Complex Communities and Environments

<b>Ecosystem</b>	<b>Estimated number of expressed proteins</b>
<b>AMD biofilm</b>	$1.8 \times 10^4$
<b>Activated sludge</b>	$5.1 \times 10^4$ Š $8 \times 10^5$
<b>Ocean water (1 ml)</b>	$4.8 \times 10^5$
<b>Sargasso Sea (1,730 l)</b>	$5.5 \times 10^6$ Š $1.4 \times 10^8$
<b>Soil (1 g)</b>	$3 \times 10^9$



**Dynamic range of protein copies: 1-  $10^6$**

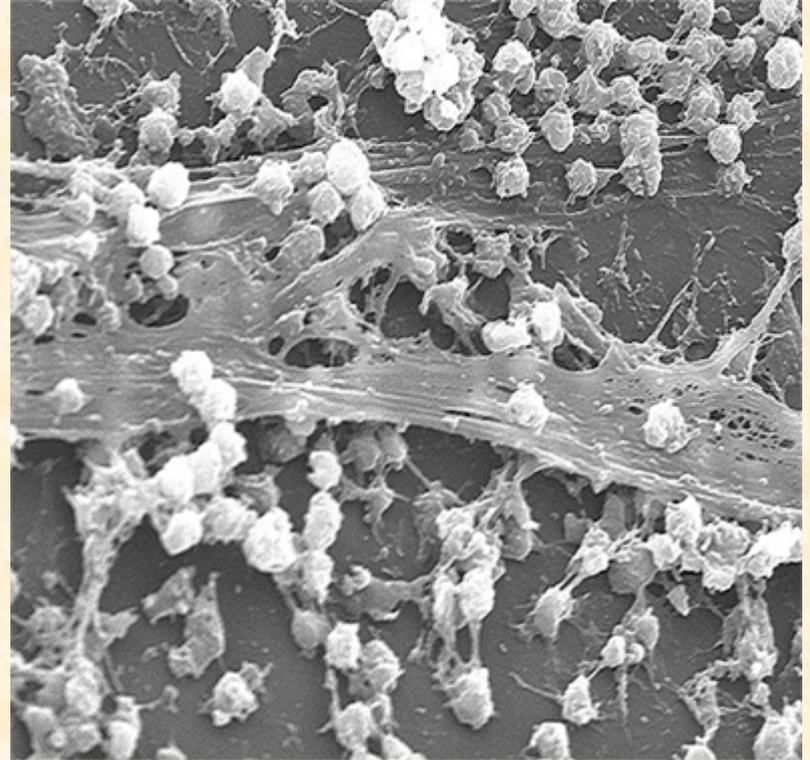
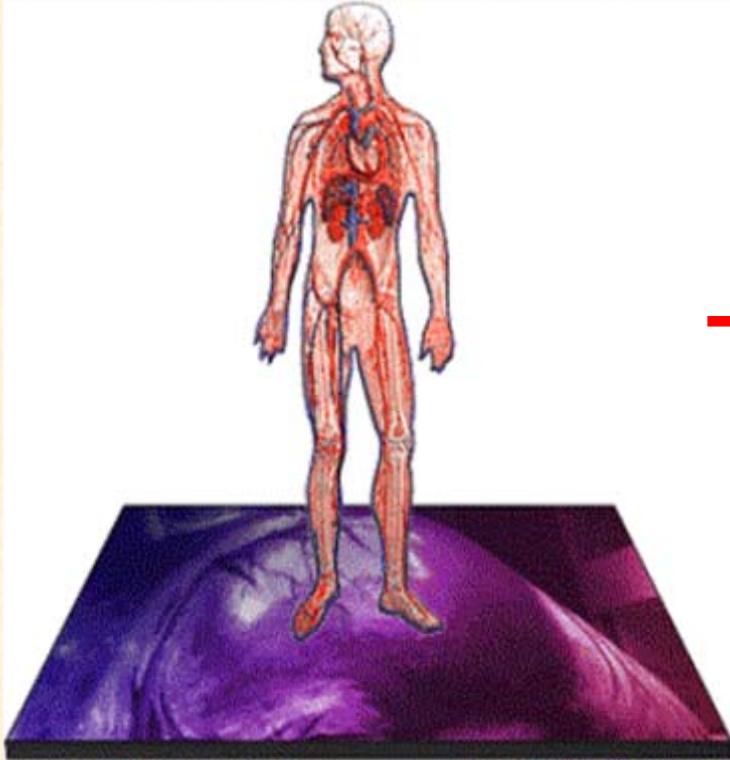
# Community Proteomics Needs

- Better sensitivity, wider dynamic range
  - “deeper and wider”
- Solid genomic foundation (accurately curated databases)
- Less ambiguity of peptide identifications
- Characterization of strain variations (peptide and protein)
- Better quantitative methods
- Advanced sequencing tagging and *de-novo* sequencing
- Miniaturization for fine scale resolution

## MS Analytical needs

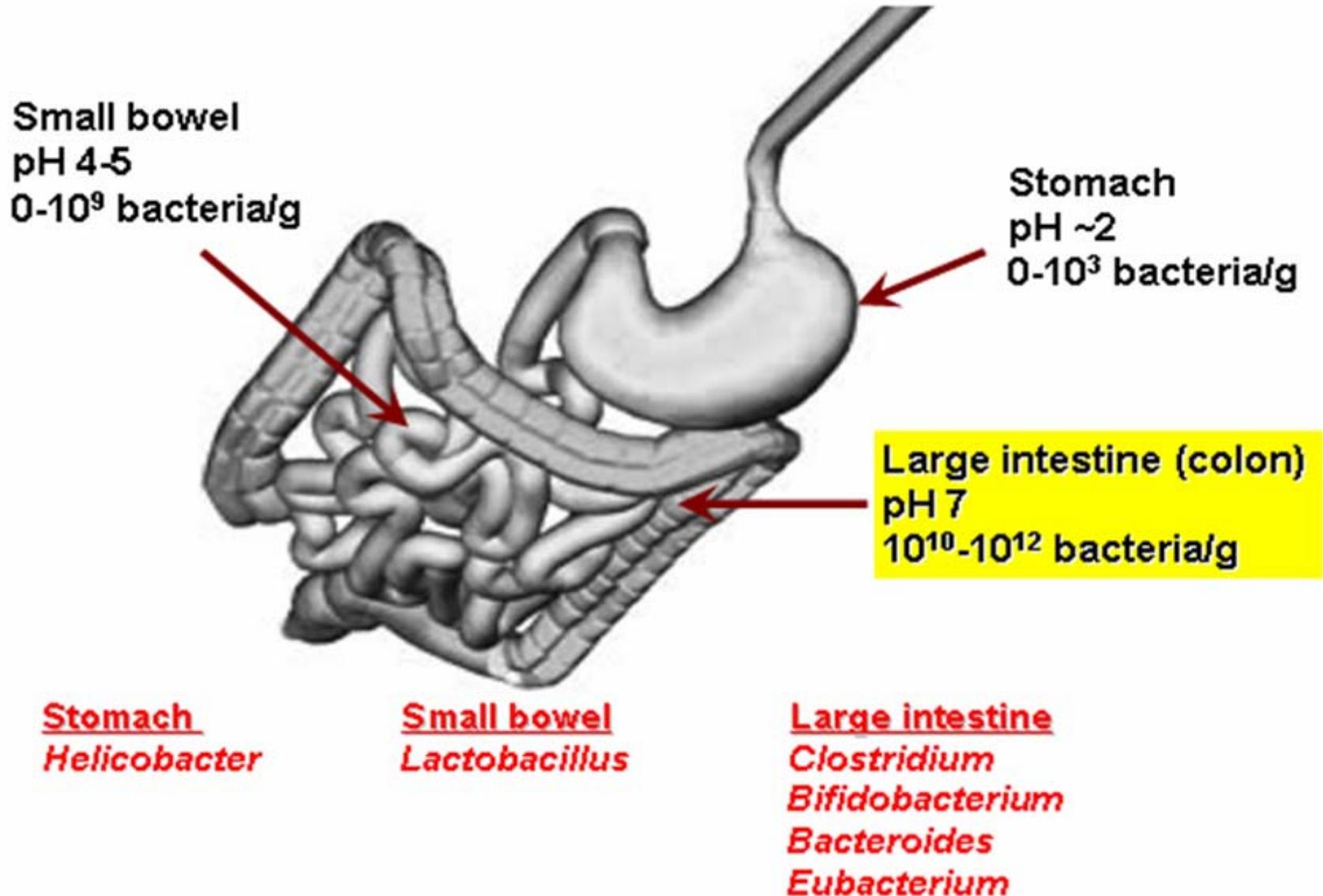
- High-throughput
- Sophisticated chromatographic separations
- High resolution and accurate mass measurements of both Full scan and MS/MS spectra on liquid chromatography time scales.
- Alternate MS/MS dissociation methods
- Examination of both peptides and intact proteins

# A human microbiome program.



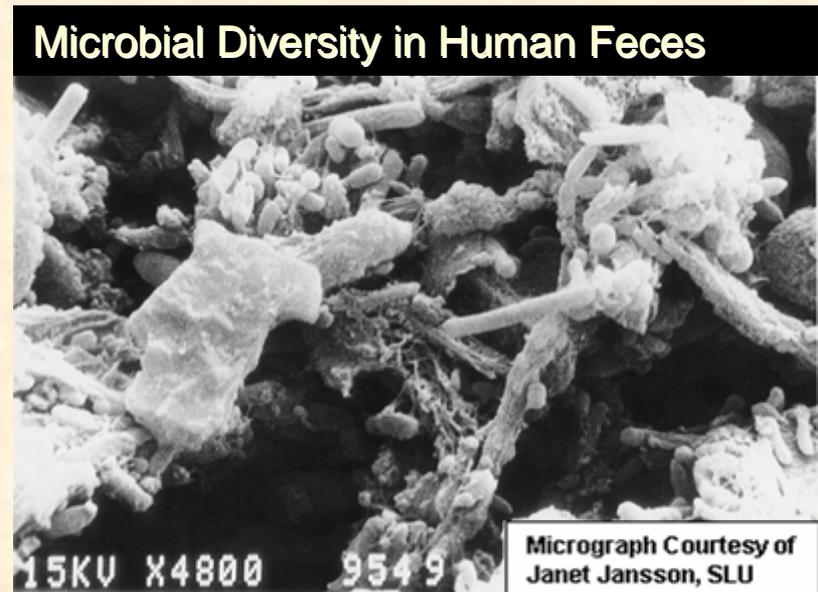
**Are we human or are we microbes?**

# 1<sup>st</sup> large-scale investigation of the human gut microbial metaproteome



# Introduction

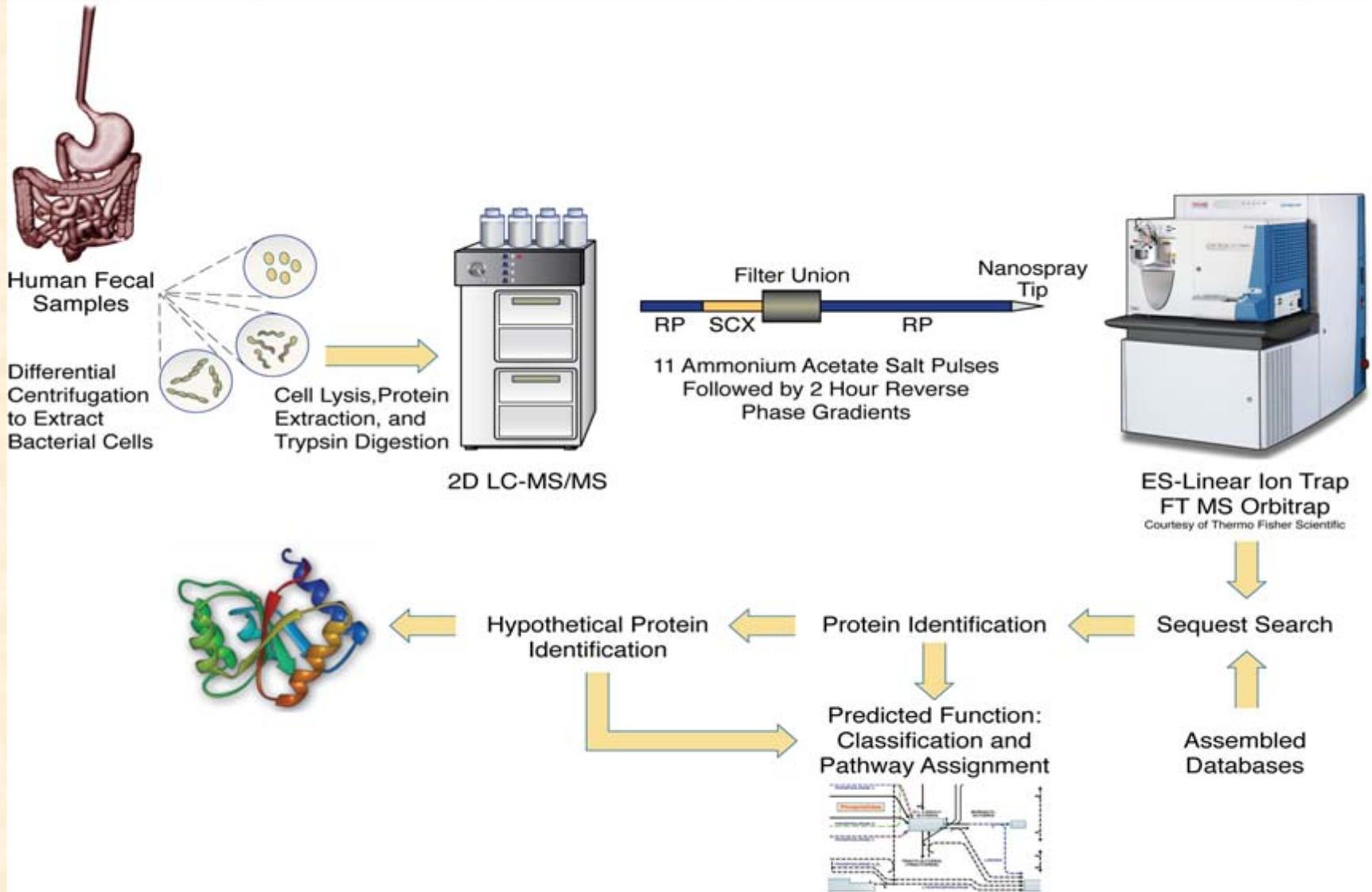
- Human gastrointestinal tract (GI) is a largely unexplored community dependent upon microorganisms for normal gut functioning
- Gut microbiome: the collection of all microbes that inhabit the GI tract
- How diverse is a normal human gut microbiome?
- Human GI is a host for an indefinite # of microorganisms (~  $10^{11}$ /gram feces) (Franks, 1998)
- Recent estimates: 800-1000 different species & >7000 different strains inhabit the GI (Backhead, 2005)
- How can we even begin to determine the composition of microbes without cultivating them?



# Goals

- 1. Biomass quantity**
- 2. Development of MS based approaches**
- 3. Reproducibility in measurements**
- 4. Coverage (deep and wide)**
- 5. Can we use representative metagenomes or isolate genomes**
- 6. Healthy versus Disease (i.e., Crohns disease, ulcerative colitis, colon cancer)**

# Experimental Design for Human Gut Microbiomes



# Databases

- **Ideal would be exact metagenome from same sample.**
- **Relevant reference metagenome- concatenated genomes from two human gut microbial metagenomes (Gill *et. al. Science 2006, human protein database, rice protein database, 33 Human commensals and pathogens, and 26 isolate distracters.***
- **Forward + reverse database- reverse each protein entry and append these reversed sequences onto the original database. Used for determining false positive levels.**

# Samples

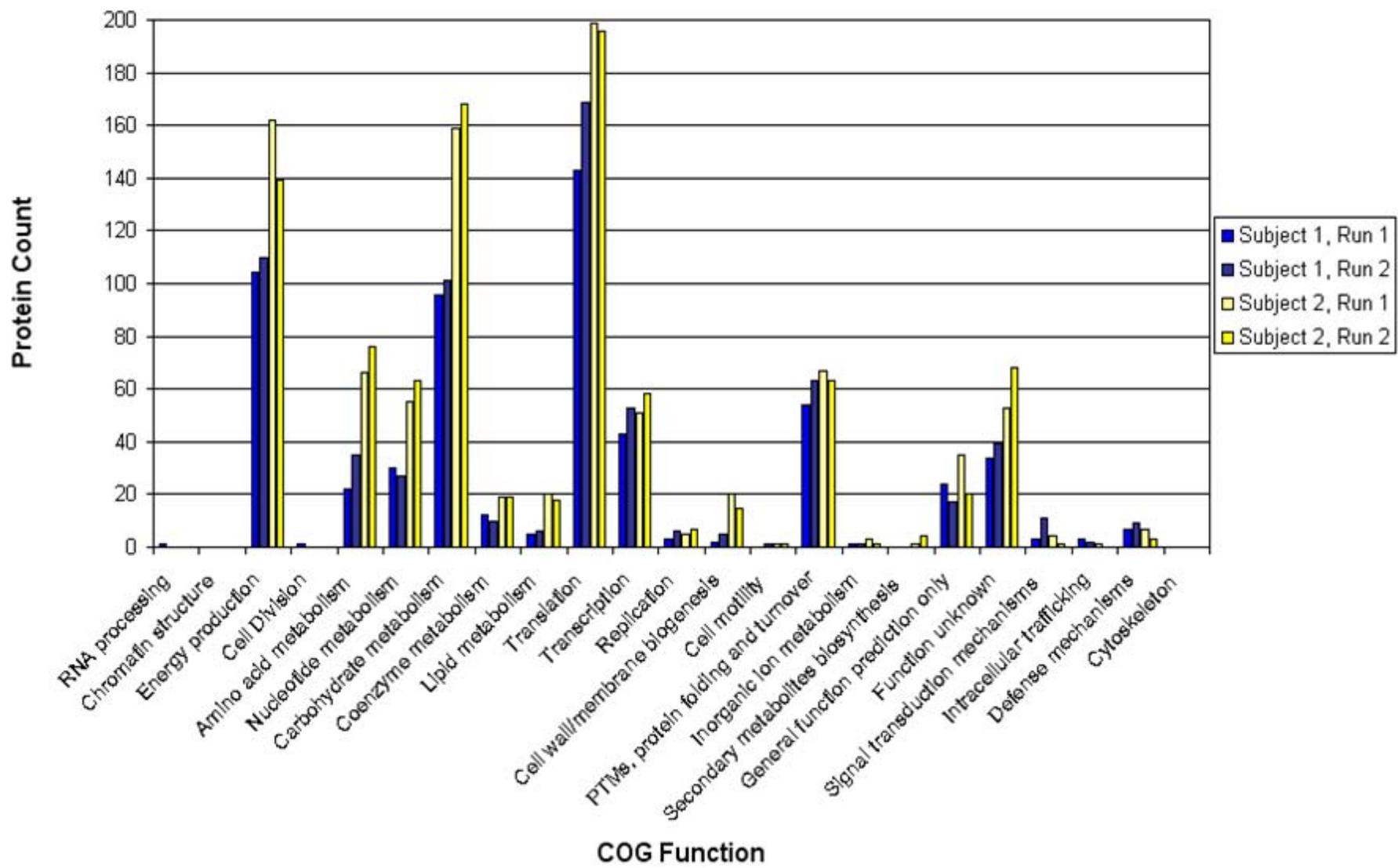
- **A female healthy monozygotic twin pair born in 1951 was invited to take part in the study.**
- **The twin pair was identified through the local twin club. Fecal samples were collected in 20 ml colonic tubes by the twins and sent to Örebro University Hospital on the day of collection, where they were placed at  $-70^{\circ}\text{C}$  and stored.**
- **At the time of sampling each twin filled out a questionnaire with information about diet, including ingestion of live yogurt or alcohol, antibiotic or drug therapy.**
- **The Uppsala County Ethics Committee approved the study.**
- **ORNL human subject samples review board approved the study.**
- **Microbial cell pellets were extracted from the fecal samples (Alimetrics Ltd, Helsinki, Finland)**

## Number of protein, peptide and spectra identifications for Subjects 1 and 2 (replicates)

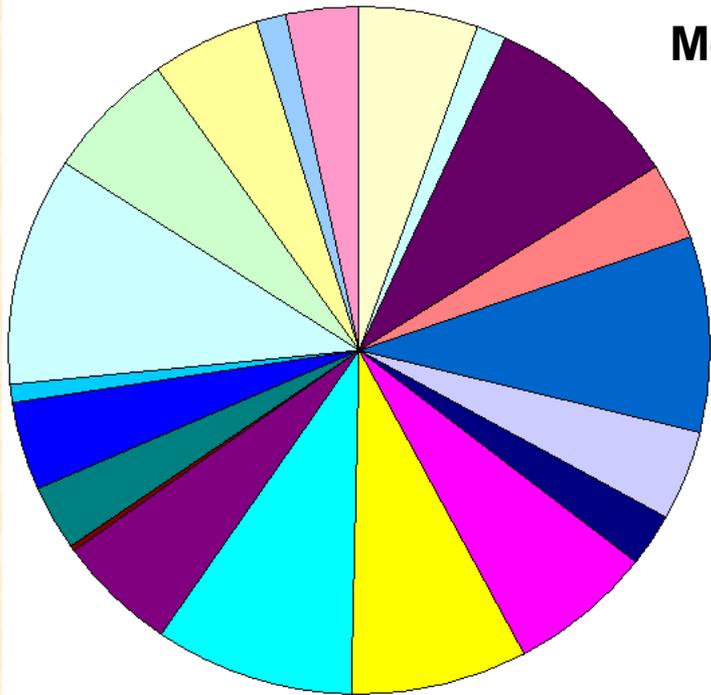
<b>db1 database</b>				
<b>Subject ID</b>	<b>Protein identifications *</b>	<b>Peptide identifications</b>	<b>MS/MS Spectra</b>	<b>Peptides between 10 and -10 ppm</b>
Subject 1, Run 2	722	2253	4440	80.42
Subject 1, Run 3	634	1886	4069	81.7
Subject 2, Run 1	974	3021	5829	83.41
Subject 2, Run 2	983	2948	6131	81.47
<b>metadb database</b>				
Subject 1, Run 2	1098	2977	5364	81.67
Subject 1, Run 3	970	2441	4829	84.47
Subject 2, Run 1	1341	3586	6509	84.71
Subject 1, Run 2	1275	3374	6635	82.92

\*Numbers given are non-redundant

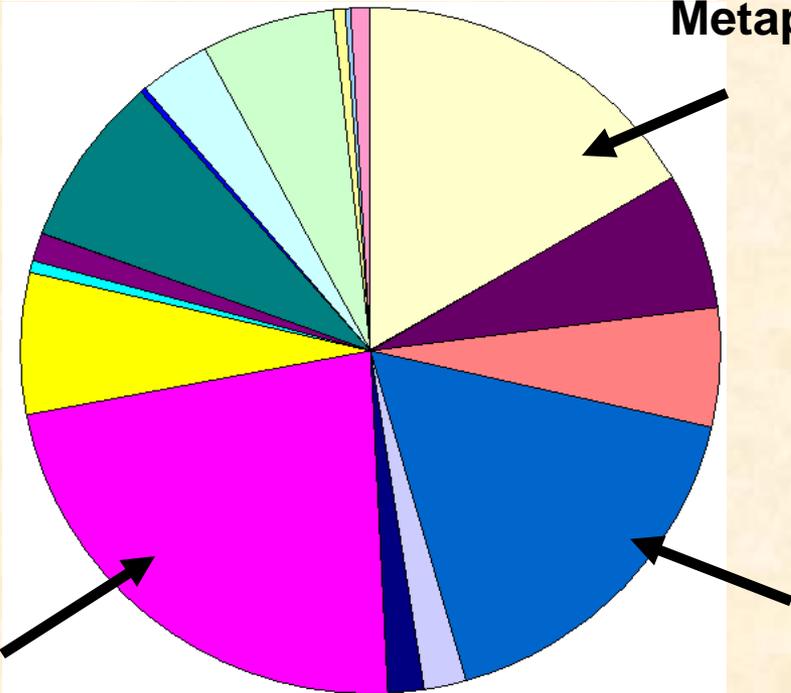
## Proteome Subject 1 and 2 Comparison



## Metagenome



## Metaproteome



- RNA processing
- Chromatin structure
- Energy production
- Cell Division
- Amino acid metabolism
- Nucleotide metabolism
- Carbohydrate metabolism
- Coenzyme metabolism
- Lipid metabolism
- Translation
- Transcription
- Replication
- Cell wall/membrane biogenesis
- Cell motility
- PTMs, protein folding and turnover
- Inorganic ion metabolism
- Secondary metabolites biosynthesis
- General function prediction only
- Function unknown
- Signal transduction mechanisms
- Intracellular trafficking
- Defense mechanisms
- Cytoskeleton

## Abundant Microbial Species

- Microbial proteins were the predominate protein identifications
- Expected gut isolates such as *Bacteroides thetaiotaomicron*, *Bifidobacterium longum*, *Bacteroides fragilis*, and *Bifidobacterium adolescentis* were the most abundant
- ~ 35% of the total spectra matched to *Bacteroides* or *Bifidobacterium* proteins with *Bacteroides* proteins being more prevalent

# Measured abundant human proteins

- Digestive enzymes such as elastase, chymotrypsin C and salivary amylases
- Structural cell adhesion and cell-cell interactions proteins such as actin & myosin
- Innate immunity proteins including antimicrobial peptides, scavenger receptor cysteine-rich (SRCR) proteins and others related to immunity and inflammation response

Subject1_Run1					Subject2_Run1				
Database	Proteins	%	Spectra	Total %	Database	Proteins	%	Spectra	Total %
Gut Isolate Genomes	547	38.17	2926	28.11	Gut Isolate Genomes	604	32.26	3047	22.12
Contams	7	0.49	177	1.70	Contams	5	0.27	85	0.62
Human Proteins	166	11.58	3276	31.48	Human Proteins	232	12.39	4440	32.24
Gill Metagenome Set7	205	14.31	1304	12.53	Gill Metagenome Set7	304	16.24	1835	13.32
Gill Metagenome Set8	328	22.89	1977	19.00	Gill Metagenome Set8	568	30.34	3720	27.01
Rice	45	3.14	226	2.17	Rice	53	2.83	273	1.98
Isolate Distracters	135	9.42	522	5.02	Isolate Distracters	106	5.66	373	2.71
Totals:	1433		10408		Totals:	1872		13773	

Subject1_Run2					Subject2_Run2				
Database	Proteins	%	Spectra	Total %	Database	Proteins	%	Spectra	Total %
Gut Isolate Genomes	600	38.73	3111	27.99	Gut Isolate Genomes	515	30.13	2658	20.08
Contams	6	0.39	154	1.39	Contams	3	0.18	63	0.48
Human Proteins	187	12.07	3752	33.76	Human Proteins	214	12.52	4671	35.28
Gill Metagenome Set7	243	15.69	1477	13.29	Gill Metagenome Set7	303	17.73	1810	13.67
Gill Metagenome Set8	365	23.56	2013	18.11	Gill Metagenome Set8	556	32.53	3535	26.70
Rice	57	3.68	246	2.21	Rice	34	1.99	197	1.49
Isolate Distracters	91	5.87	360	3.24	Isolate Distracters	84	4.92	305	2.30
Totals:	1549		11113		Totals:	1709		13239	

# Human Gut Proteome Summary

- Measured ~ 600 -1400 proteins/run with high reproducibility
- False positive rates range from 1-5% with low mass accuracy and 0.1-1% with high mass accuracy.
- Used protein identifications to verify gene expression using indirect metagenome information
- The majority of proteins detected were involved with translation, carbohydrate metabolism and energy production
  - The community is taking advantage of the abundant store of nutrients for the generation of energy via carbohydrate catabolism
- Future:
  1. Expand on the # of normal human samples
  2. Increase the # of sequenced human metagenomes for database searching (Japanese study)
  3. Characterize the natural microbiome in Crohn's patients...identify an altered microbial signature.

# Conclusions

- **Mass Spectrometry Based Proteomics can be used to understand normal gut microbial community function and dynamics at a systems level.**
- **Suitable biomass extraction from fecal and cecal samples enables comprehensive proteome identification**
- **Challenges remain...**
  - **Depth of proteome measurements**
  - **Protein annotation**
  - **Inherent dynamic range**
  - **Incomplete or partial genome information (lack of metagenomes)**
  - **Obtaining additional human gut samples**

# Acknowledgements human gut metaproteome

- **ORNL Group:**
- **Dr. R. L. Hettich**
- **Dr. M. G. Lefsrud**
- **Alison Russell**
- **M. Shah**
  
- **Collaborators:**
- **Dr. A. Godzik**
- **Dr. J. K. Jansson and Swedish University of Agriculture Sciences group**
- **Orebro University Hospital group (Orebro, Sweden)**
  
- **The ORNL part of this research sponsored by U.S. Department of Energy under contract DE-AC05-00OR22725 with Oak Ridge National Laboratory, managed and operated by UT-Battelle, LLC.**