Integrated Omics in Systems Biology: The New Frontier for Environmental Biotechnology

VIMSS

ESPP Environmental Stress Pathway Project

PCAP Protein Complex Analysis Project

IPTAI

jbei Joint BioEnergy Institute

Berkeley Lab 75 Years of World-Class Science 1931-2006

Terry C. Hazen

DOE BER Distinguished Scientist

Head, Ecology Department

Head, Center for Environmental Biotechnology

Co-Director, Virtual Institute for Microbial Stress and Survival

Director, Microbial Communities Dept, Joint BioEnergy Institute

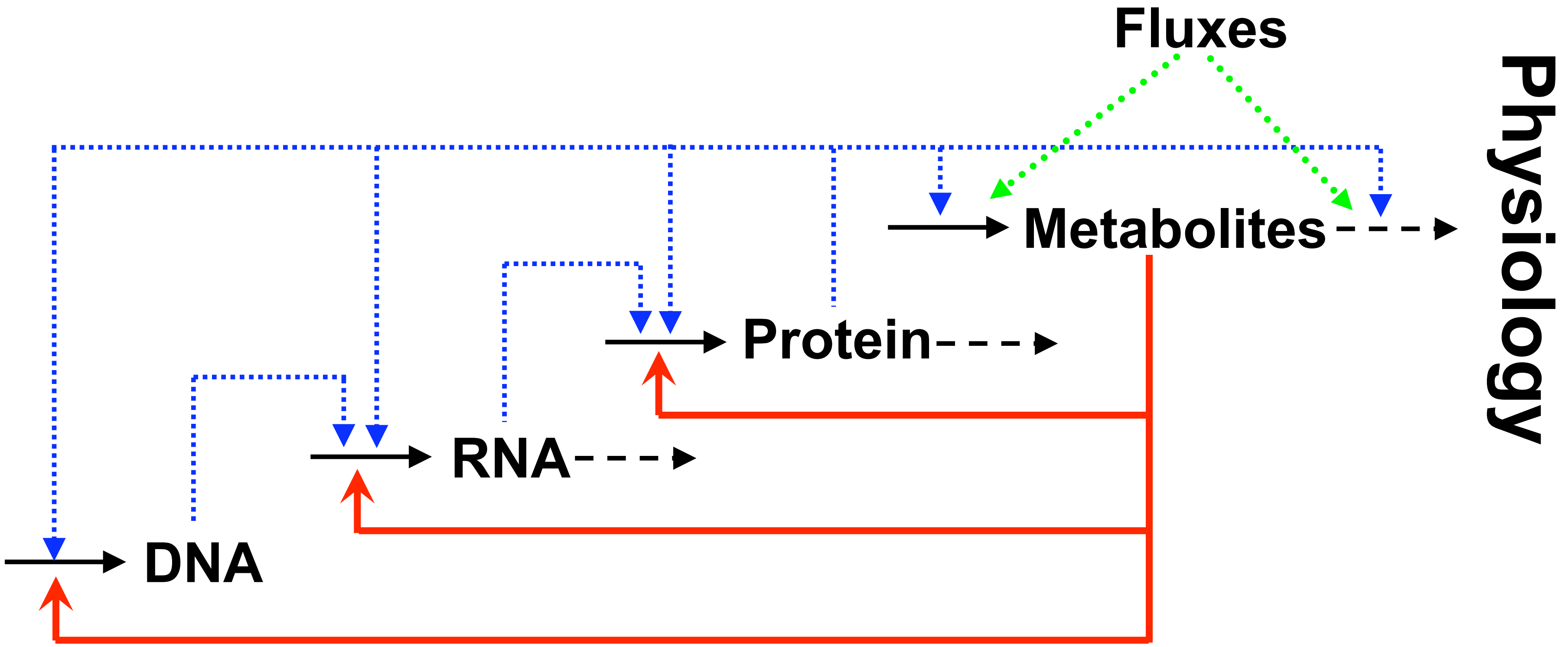
Lawrence Berkeley National Laboratory, University of California

tchazen@lbl.gov *http://vimss.lbl.gov*

"…The Plane, The Plane…"

- There are 3 million parts in a 777 provided by more than 900 suppliers from 17 countries around the world.
- Completed genomes provide "parts lists" for many microbes, the sequence is little more then the blueprint for each part (protein) in the organism.
- Having a blueprint for the parts of a 777 jet gives few clues as to how each part is made, how it assembles into devices and systems, much less how it flies which is its essence.
- genes, proteins, metabolites, and multimolecular assemblies ("molecular machines") interact in an intricate labyrinth of pathways and networks to create, sustain, and reproduce the system we call the living cell—complexity well beyond the engineering and essence of a 777.
- Systems biology will transform biology from an empirical and descriptive science to a more quantitative and predictive science
- Enabling us to manipulate and use living systems and their components.

The Omics!!



Fluxes

Physiology

Metabolites

Protein

RNA

DNA

Genomics

Transcriptomics

Proteomics

Metabolomics
Glycomics
Lipidomics
Fluxomics

Phenomics

Systems Biology Approach

Ecosystem

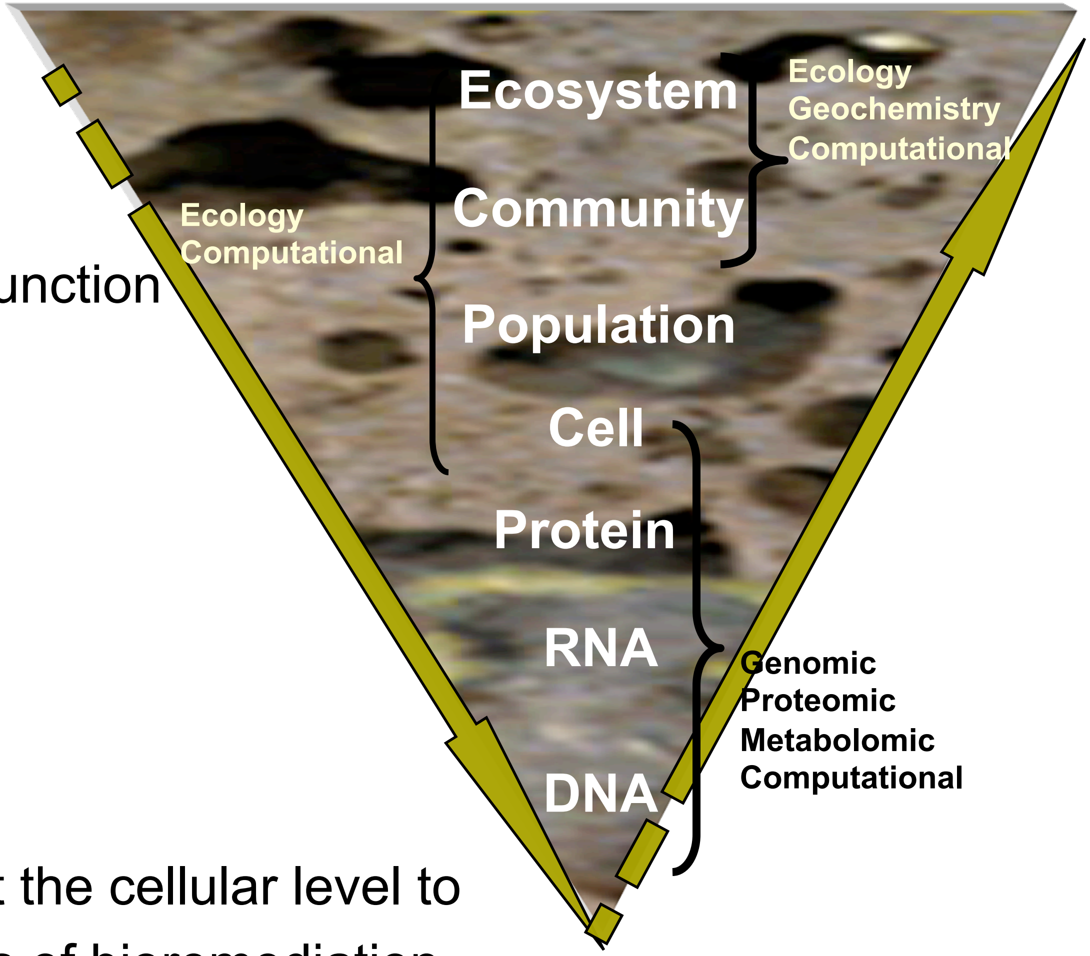
Identify key factors (i.e., stresses) that drive community structure and composition and impact the survival and efficacy of heavy metal-reducers

Community and Population

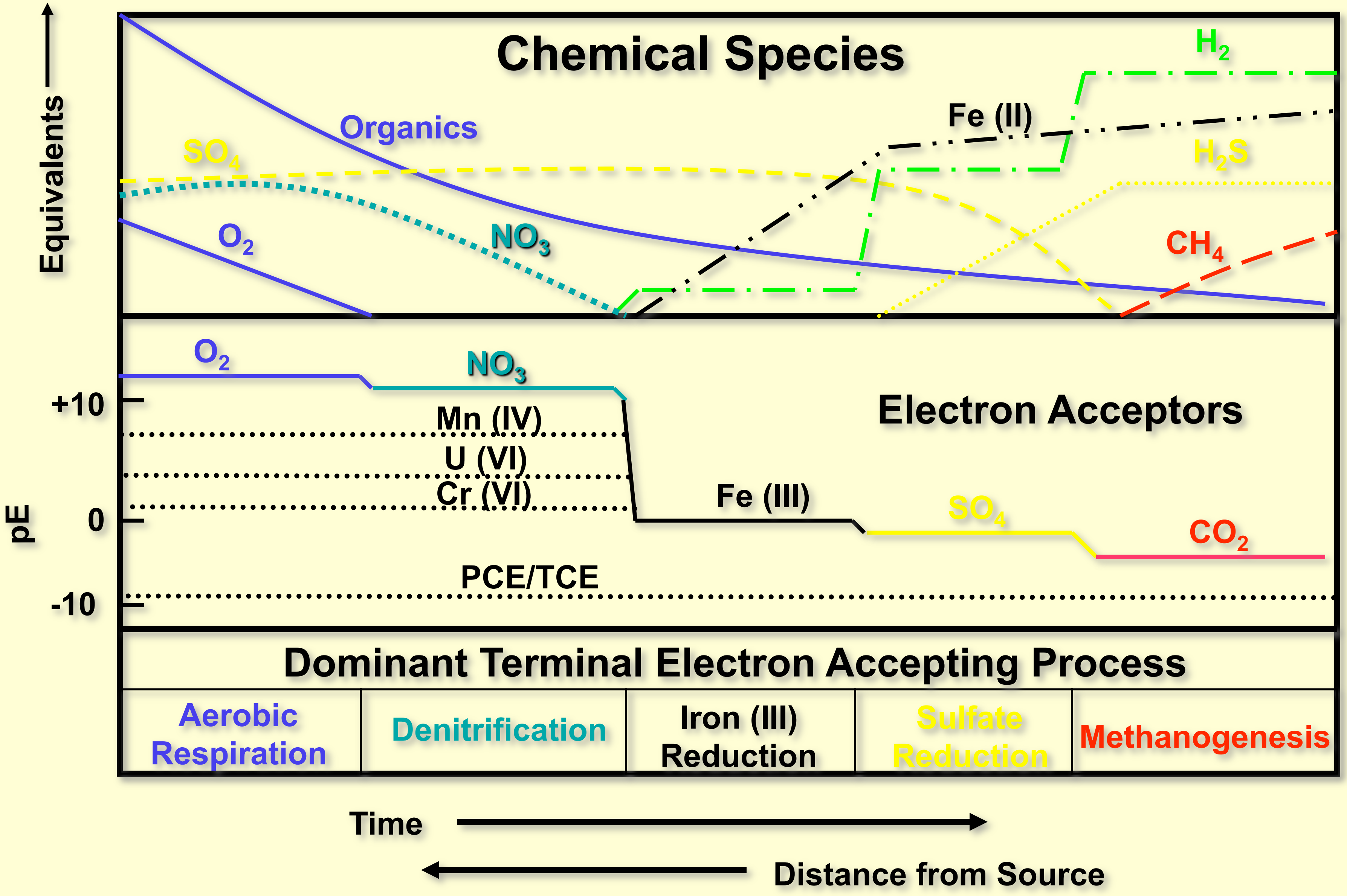
Understand impacts on structure/function relationships

Cell

Analyze DNA, RNA, and protein at the cellular level to understand cellular effects in terms of bioremediation



Critical Biogeochemistry



Ecogenomics & Transcriptomics

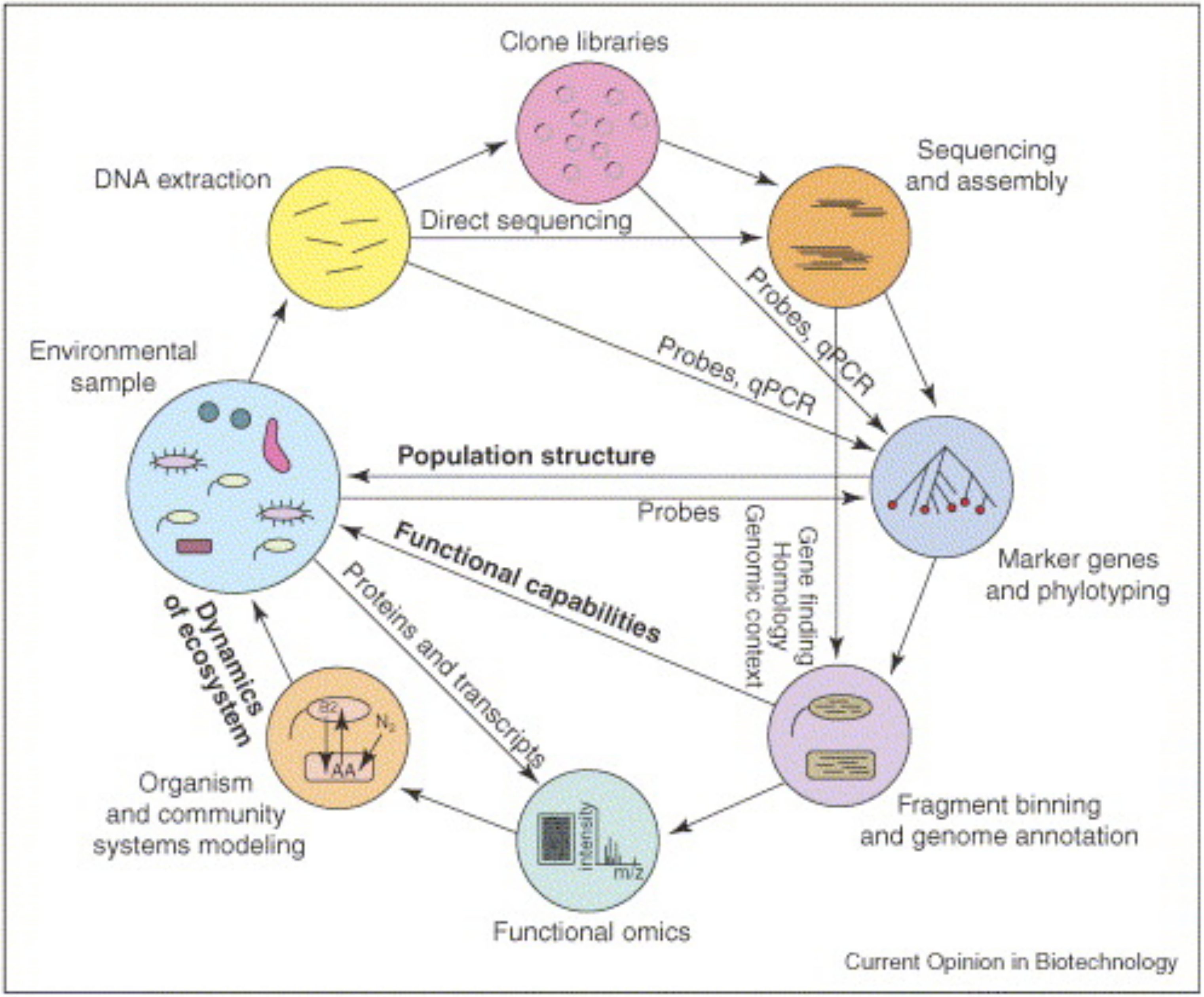
Ecogenomics - studies of genomes in an environmental context

- 16s rDNA microarrays for community analyses
- T-RFLP - terminal restriction fragment length polymorphisms
- Metagenome sequencing
- Annotation of sequences for environmental context
- Microbial Source Tracking for Pathogens

Transcriptomics - gene expression

- mRNA expression arrays of one organism or functional group
- Real-time PCR analyses

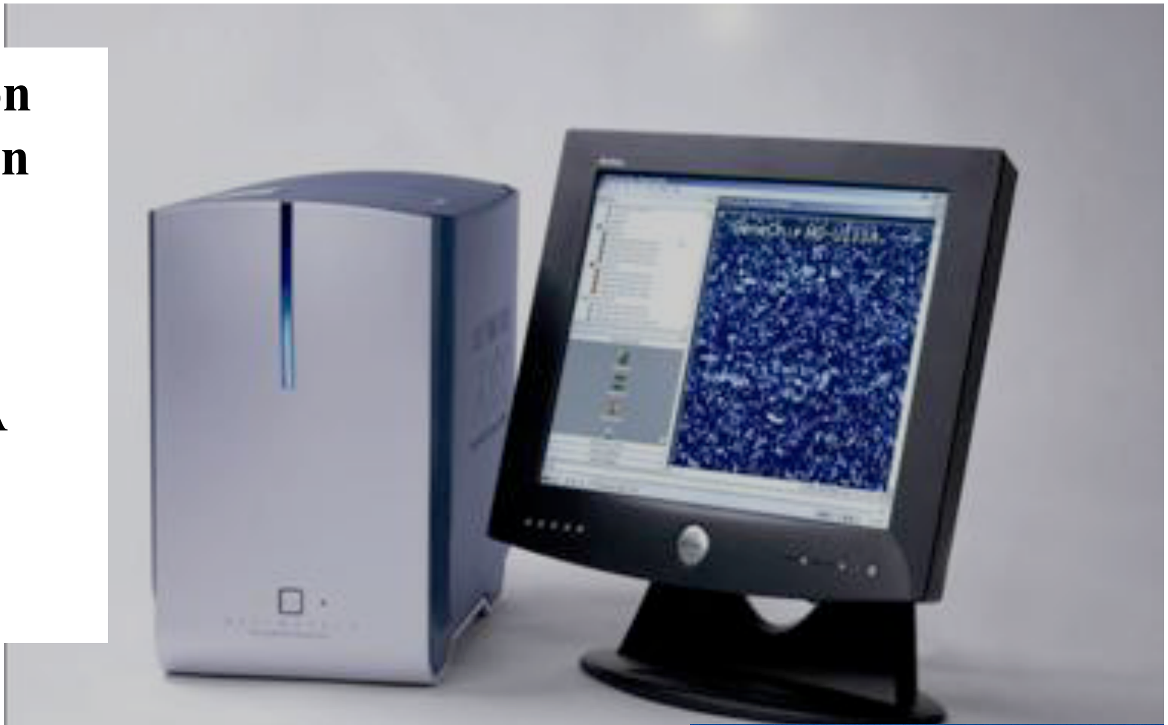
Ecogenomics



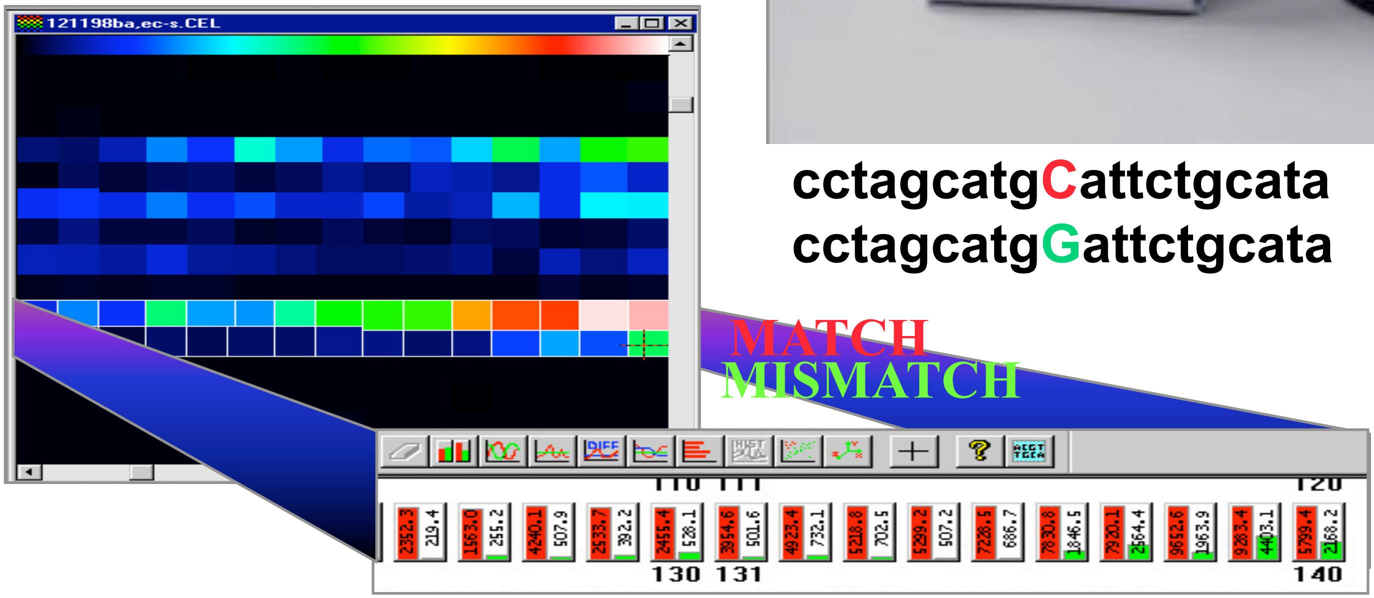
Current Opinion in Biotechnology

DOE 16s rDNA microarray

- Rapidly detect the composition and diversity of microbes in an environmental sample
- Massive parallelism - 550,000 probes in a 1.28 cm^2 array
- all 9,900 species in 16S rDNA database
- Single nucleotide mismatch resolution

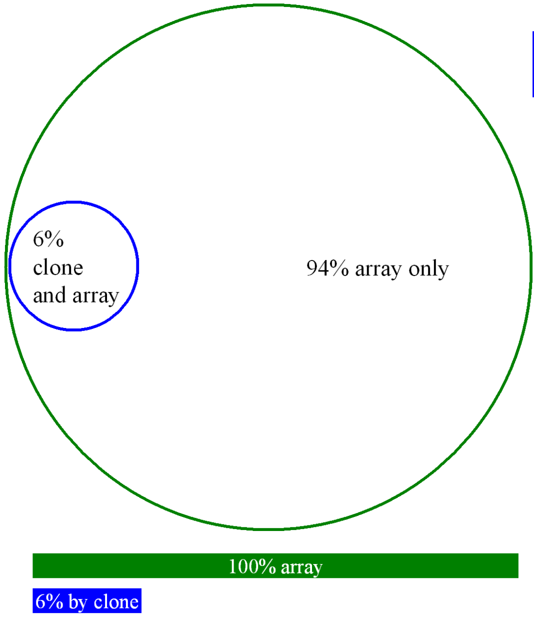


cctagcatgCattctgcata
cctagcatgGattctgcata



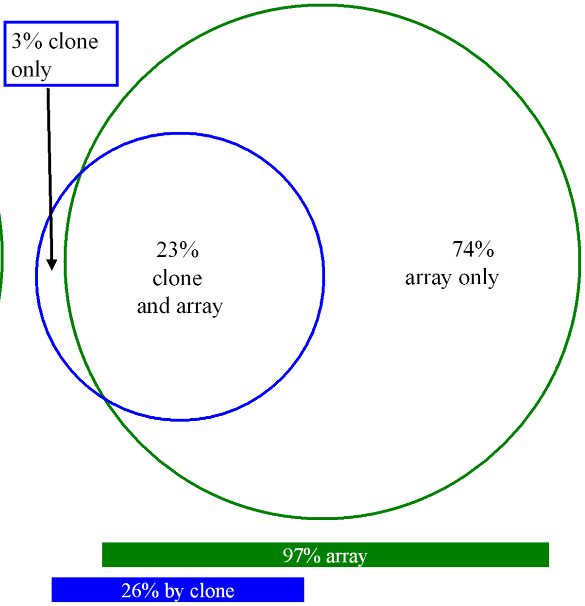
Accuracy V Clone libraries

Cr(VI) groundwater



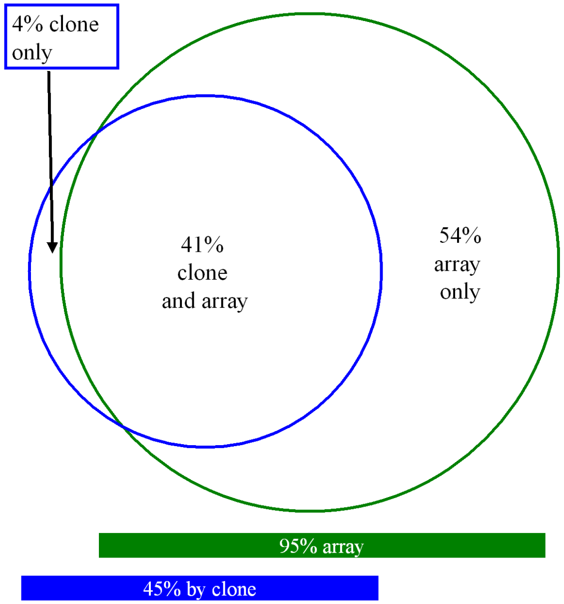
384 clones

U(VI) subsurface soil



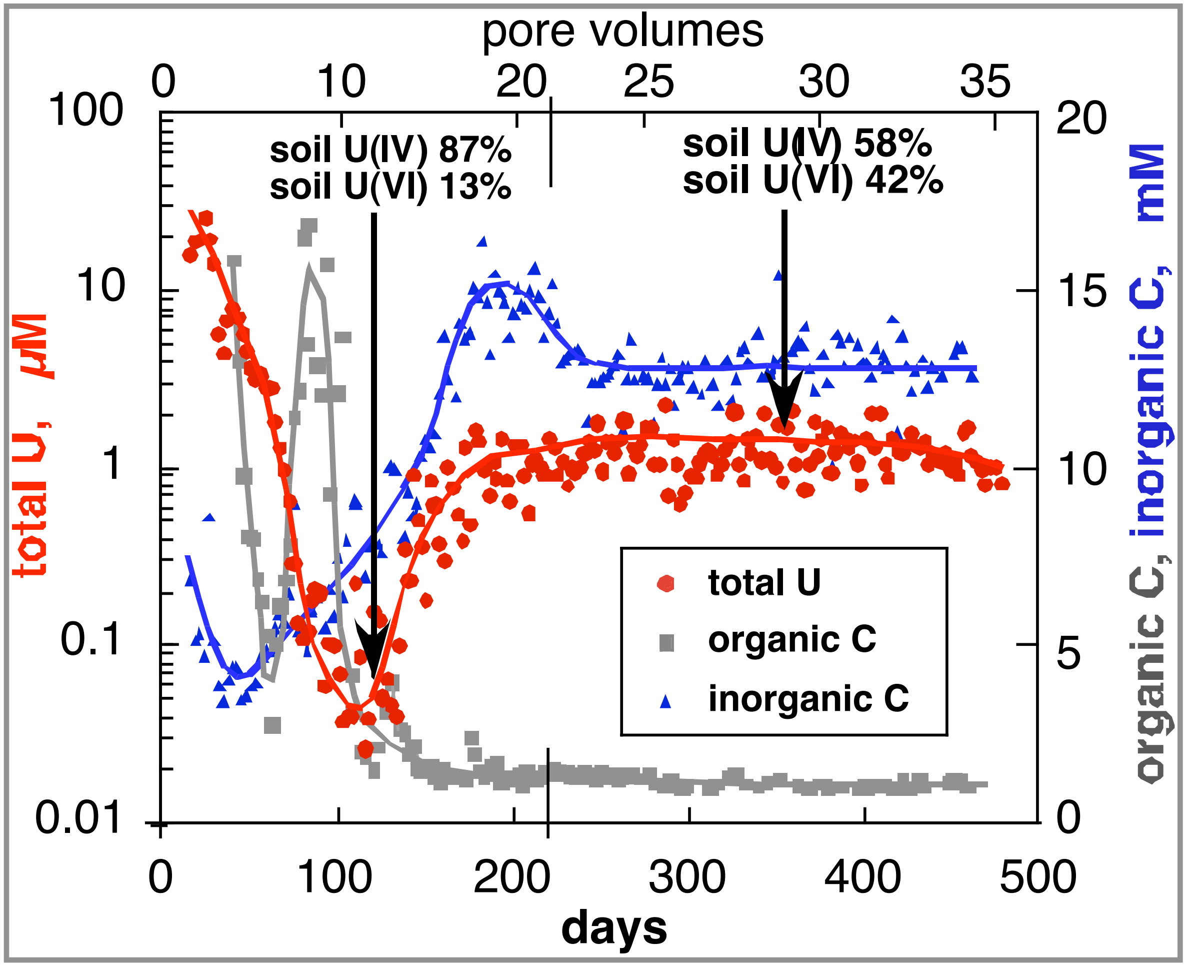
768 clones

Urban aerosol



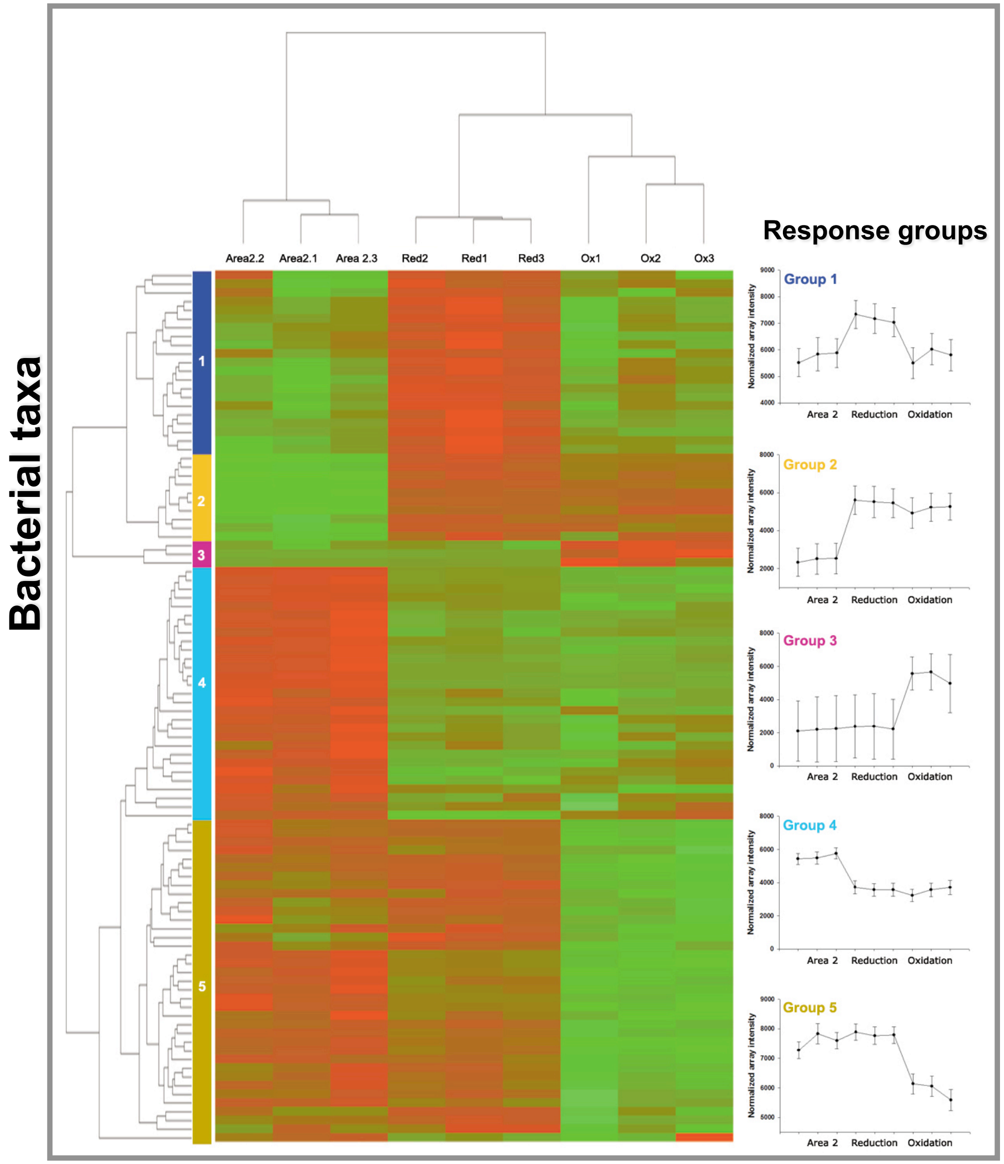
768 clones

Uranium Anaerobic Reoxidation



Microbial metabolism - CO_2 produced increasing dissolved IC

Bi-directional clustering of array data



Column sediment samples

Major components of response groups

- *Arthrobacter sp.*
- Alpha-proteobacteria
- *Azoarcus sp.*

- *Geothrix*
- *Geobacter*
- *Anaeromyxobacter*

Uranium reducers

- *Acidobacteria*
- *Desulfovibrio*

Syntrophy?

- *Actinobacteria*
- *Firmicutes*
- *Alpha-proteobacteria*

- *Actinobacteria*
- *Alpha-proteobacteria*

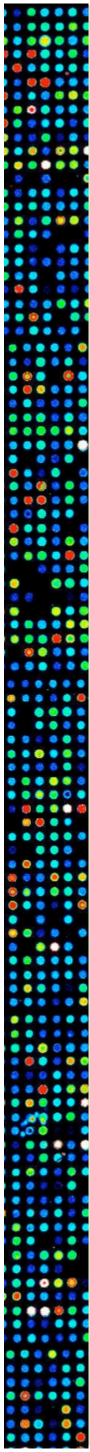
Brodie, E. L., T. Z. DeSantis, D. C. Joyner, S. M. Baek, J. T. Larsen, G. L. Andersen, T. C. Hazen, P. M. Richardson, D. J. Herman, T. K. Tokunaga, J. M. Wan, and M. K. Firestone. 2006. Bacterial population dynamics during uranium reduction and re-oxidation: Application of a novel high density oligonucleotide microarray approach. Appl. Environ. Microbiol. 72:6288-6298

MycoChip Performance

- Tested prototype array at LLNL Livermore Microarray Center.
- Fungal probes corresponding to test sample showed strong response above background.
- Probes showed large dynamic range after background subtraction.
- Low noise detected, and predicted array control intensity.

	Controls	**Test Sample**	**All**
Positive Calls	36	2729	2765
Total Probes	1424	129540	130964
Mean Intensity	1340.4	258.8	272.9
Median Intensity	130.2	241.8	241.0
Min	84.0	95.3	84.0
Max	22503.5	799.8	22503.5

DOE Bioenergy Research Centers



GeoChip3.0 development and construction

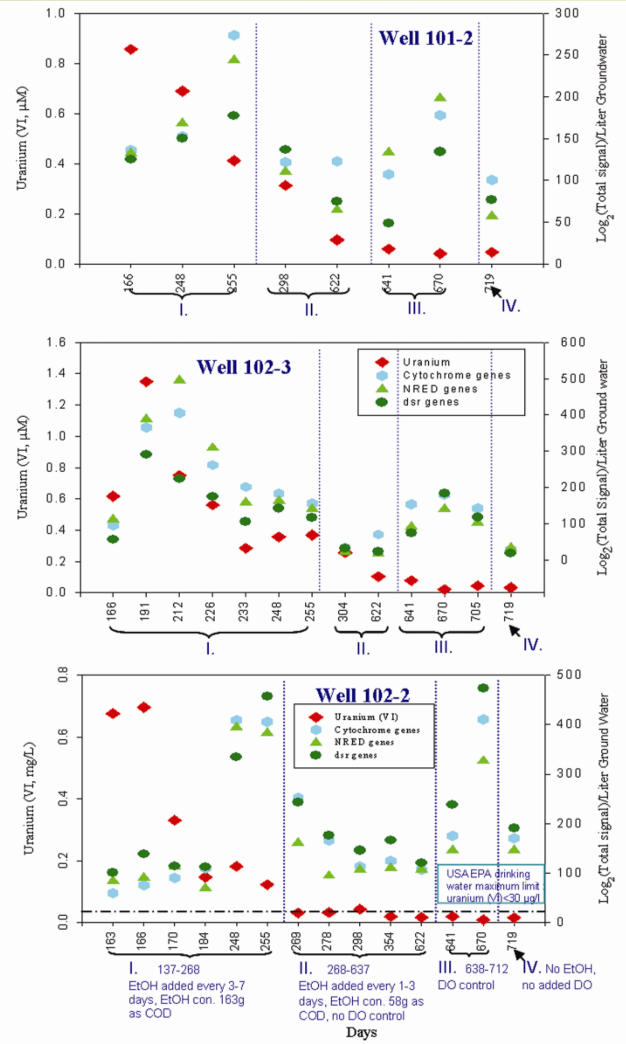
Gene category	# Genes	# probes	# unique probes	# group probes	# covered CDS
Carbon deg	31	2727	1167	1560	4737
Carbon fix	5	898	182	716	1806
Energy procession	2	413	377	36	449
Mathane	3	254	171	83	434
Metal Resistance	43	4917	609	4308	10458
Nitrogen	13	3561	2105	1456	6892
Organic Remediation	190	8815	2244	6571	16948
gyrB	1	1164	629	535	2251
Phosphorus	3	599	183	416	1212
Sulphur	3	1328	952	376	1773
Total Summary	294	24676	8619	16057	46960

GeoChip 3.0 have been used for analyses of soil samples from BioCON (biodiversity, CO_2 and nitrogen) and warming sites.

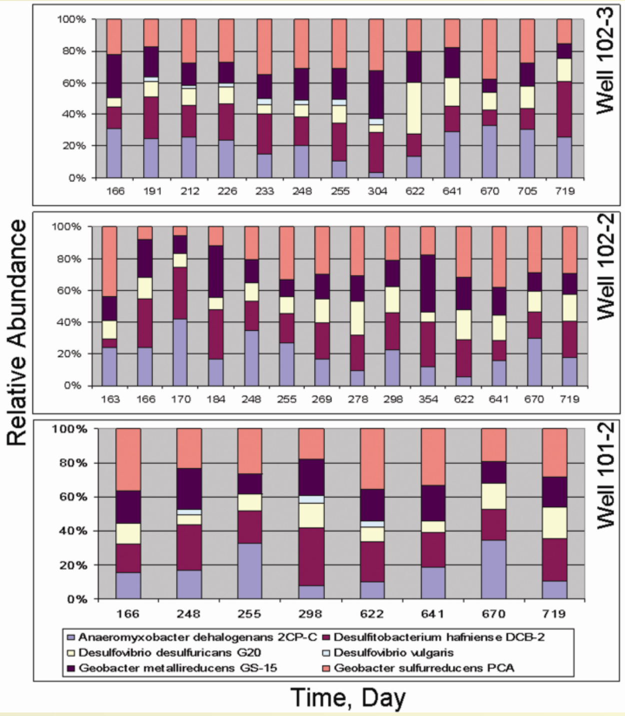
INSTITUTE FOR ENVIRONMENTAL GENOMICS
UNIVERSITY OF OKLAHOMA

Population Changes During the Active U(VI) Bioremediation Phase (operational days 163-719)

◆ Uranium ◆ NRED genes
◆ Cytochrome genes ◆ *dsr* genes



Bacterial populations increased initially, reaching a peak around day 255, then dropping as U(IV) concentrations decreased.

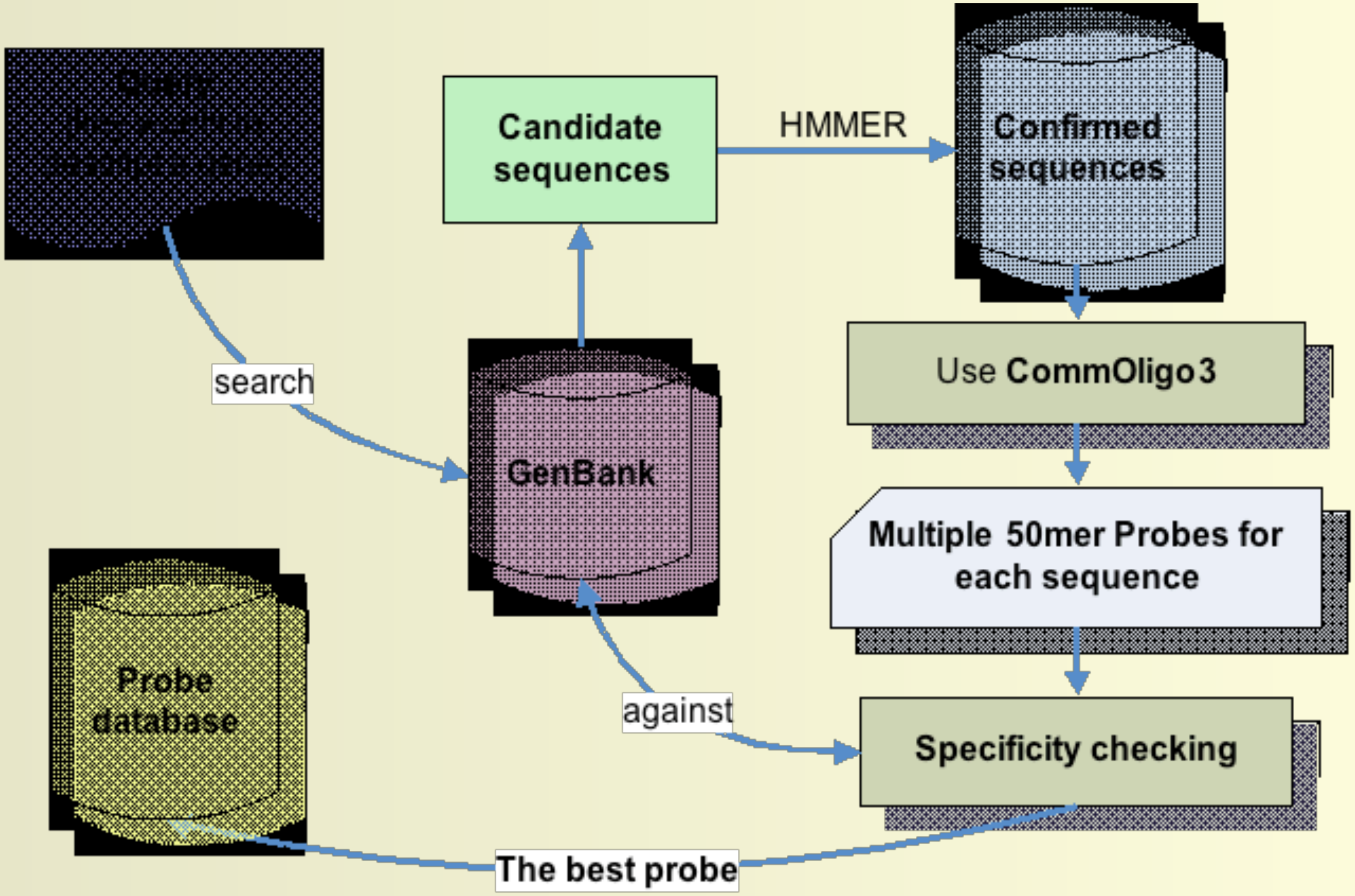


Relative abundance of cytochrome C genes from different metal reducing bacteria.

The relative proportions of the different species changed during bioremediation of U(VI)

- The microbial communities responded to the bioremediation treatment
- The communities remained somewhat spatially distinct although the populations changed temporally
- Manuscript in preparation

Future plan: ESPP stress chip design and construction



Candidate sequences

HMMER

Confirmed sequences

search

GenBank

Use CommOligo3

Multiple 50mer Probes for each sequence

against

Specificity checking

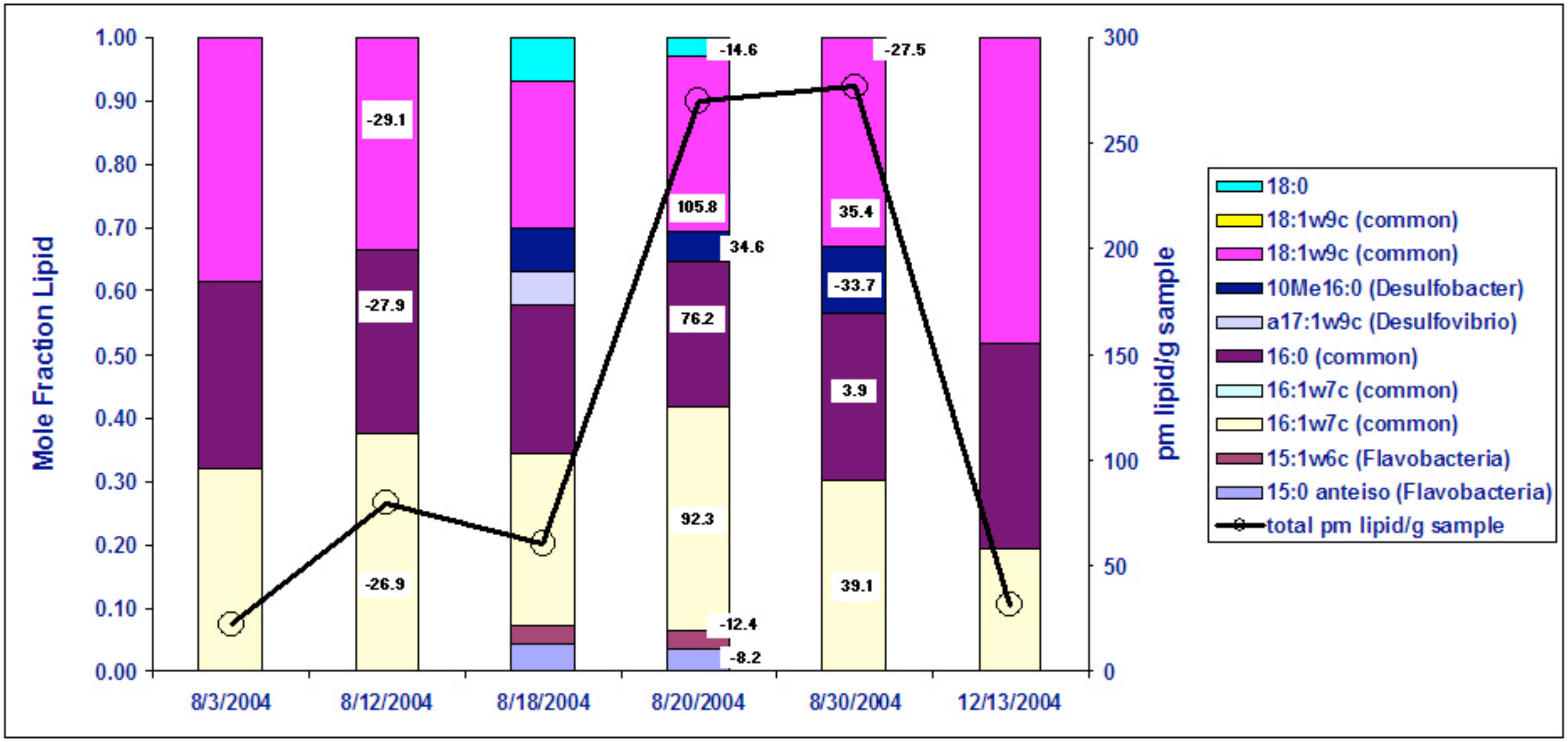
Probe database

The best probe

What genes/sequences should be put on the array?

INSTITUTE FOR ENVIRONMENTAL GENOMICS
UNIVERSITY OF OKLAHOMA

^{13}C Phospholipid Analysis



- General bacterial biomarkers indicate rapid enrichment in ^{13}C
- ^{13}C ratio is greater than expected (overall spiked HRC ratio was 15 per mil)
  - ^{13}C polylactate used as spike it is not esterified to glycerol backbone
  - it is released and consumed more rapidly
- Biomarkers for *Flavobacteriaceae* increased following injection but showed minimal enrichment with ^{13}C.
  - *Flavobacteria* do NOT typically utilize lactate, but may use glycerol (backbone, unlabeled)

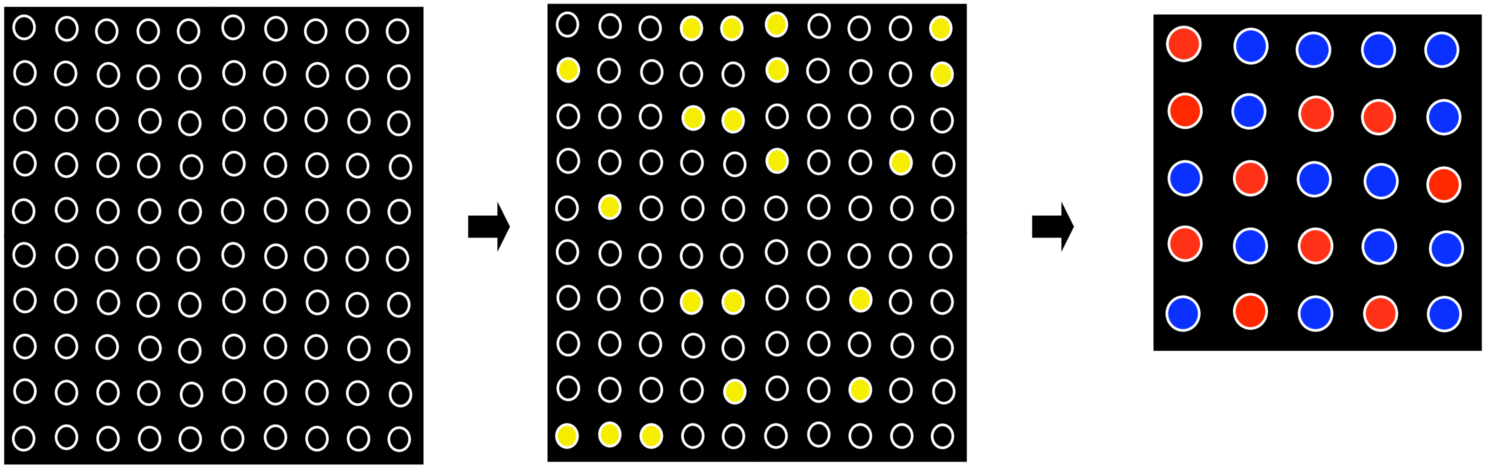
NanoSIMS + microarray indicates active organisms

(Hoeprich, Pett-Ridge, Brodie, et al. New Genomics:GTL project)

Array spotted with
universal 16S
probe set

Affymetrix PhyloChip
Fluorescence
Scan

System-specific
NimbleGen chip
+ NanoSIMS
^{13}C:^{12}C analysis



----100 µm----

rRNA profile shows
species that are present
in the community

Indicates subset of active
community that consumed
^{13}C-label substrate

● ^{12}C rRNA

● ^{13}C rRNA

➡ **“Chip-SIP” yields identity and function from the same sample**

Environmental Whole-Genome Amplification To Access Microbial Populations in Contaminated Sediments

- Recovery of adequate amounts of DNA for molecular analyses can often be challenging in stressed microbial environments.
- Developed multiple displacement amplification (MDA) methods for unbiased, isothermal, amplification of DNA
- Subsequently applied these technologies to understand stressed, low biomass, populations in multiple sediments contaminated with Uranium on the Oak Ridge Reservation
- Over 4000 clones were end sequenced. 5% of all clones were identified as belonging to Deltaproteobacteria (primarily, Geobacter and Desulfovibrio-like)
- Significant overabundance of proteins (COGs) associated with: 1) Carbohydrate transport & metabol. 2) Energy production & conversion, 3) Postranslational modification, protein turnover, & chaperones. --- All of which may be important in adaptation to environmental stressors such as low pH, high contaminate loads, and oligotrophic nature of the subsurface environment

Statistics on amplified metagenome library end-sequences

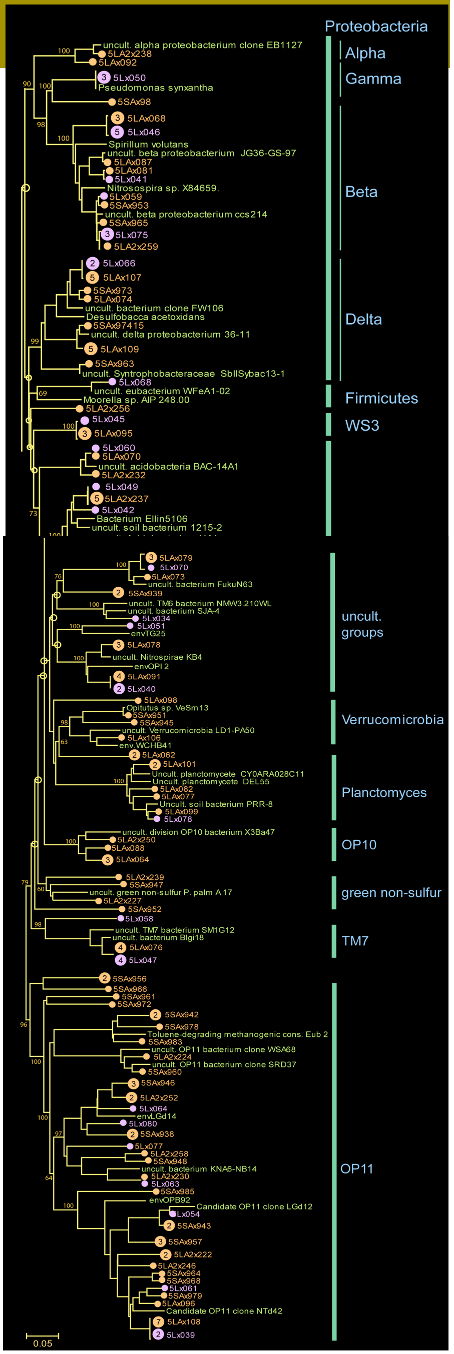
Library	Area 3, Deep	%	Area 3, Shallow	%	Area 2	%	Total	%
Number of clones sequenced	960		864		864			
Sequences generated	1,920		1,728		1,728			
Quality sequences [a]	1,394	100	1,118	100	1,509	100	4,021	100
Sequences that form contigs	370	26.5	152	13.6	141	9.3	663	16.5
Number of contigs assembled	101		53		54		208	
Sequences with similarities to known proteins [b]	928	66.6	692	61.9	990	65.6	2,610	64.9
Highest similarity to bacterial proteins	901	64.6	629	56.3	890	59.0	2,420	60.2
Highest similarity to Deltaproteobacteria proteins	35	2.5	23	2.1	155	10.3	213	5.3
Highest similarity to archaeal proteins	12	0.9	43	3.8	79	5.2	134	3.3
Highest similarity to eukaryotic proteins	12	0.9	18	1.6	21	1.4	51	1.3

a. Sequences >400nt in length

b. e-values <1e-10 from BLASTX searches against the NCBI protein database

Abulencia, C.B., Wyborski, D.L., Garcia, J., Podar, M., Chen, W., Chang, S. H., Chang, H.W., Watson, D., Brodie, E.I., Hazen, T.C. and Keller, M. (2006) Environmental Whole-Genome Amplification to Access Microbial Populations in Contaminated Sediments. *Appl. Environ. Microbiol.* **72**(5):3291-3301

[download pdf]

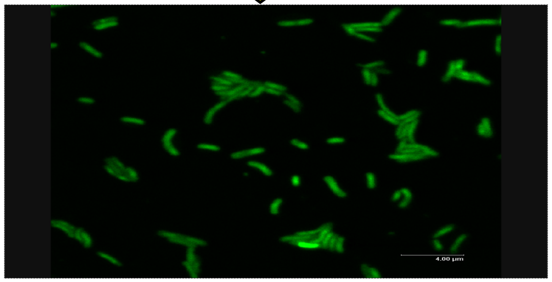


Metagenomic Analysis of NABIR FRC Groundwater Community

Data: Jizhong Zhou et al.

Metagenomic sequencing:

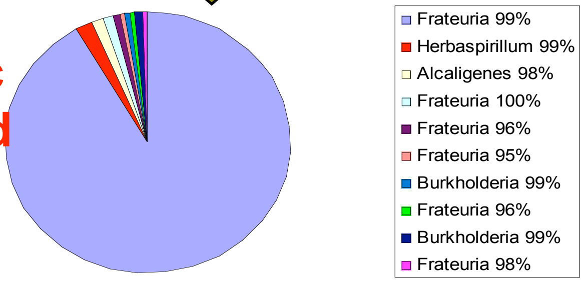
- Almost like a mono-culture
- 52.44 Mb raw data assembled into contigs totaling ~5.5 Mb
- 224 scaffolds (largest 2.4 Mb)
- Genes important to the survival and life style in such environment were found



Extremely low diversity:

- Dominated by *Frateuria*-like organism
- At least 2 *Frateuria* phylotypes
- Azoarcus species: less abundant

These results suggest that contaminants have dramatic effects on the groundwater microbial communities, and these populations are well adapted to such environments.



Life in the slow lane

Lin et al. "Long-Term Sustainability of a High-Energy, Low-Diversity Crustal Biome" Science *10/06*

Chivian et al. Science (in final review)

Radioactivity splits water and regenerates system

Life on/under MARS?

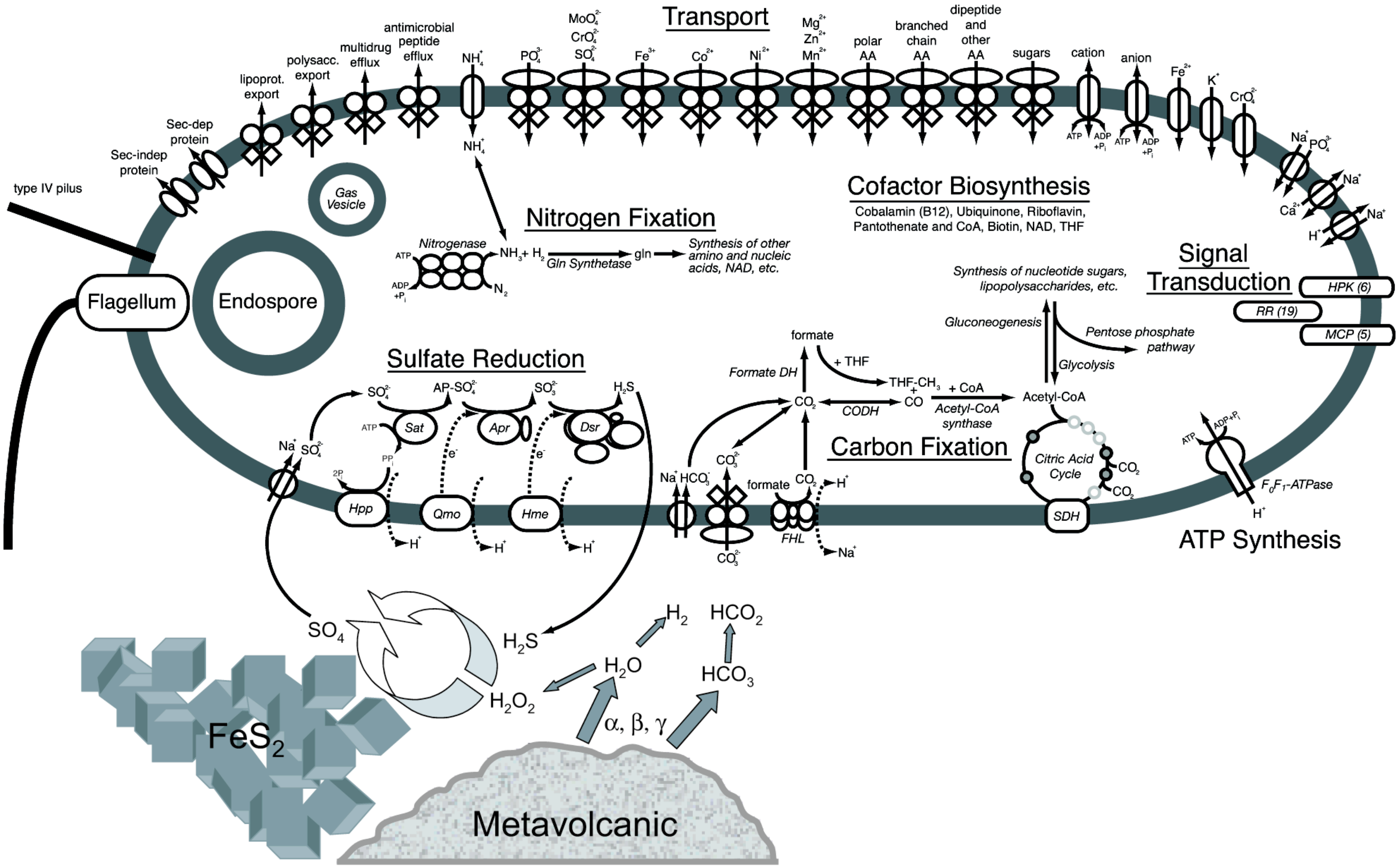
Despite knowing most of the chemical conditions,
We could not grow
Desulforudis audaxviator

But it is the major organism there:
94+% (16S PCR) to 99.9+% (metagenomics) of the bacteria

So how do we study it?

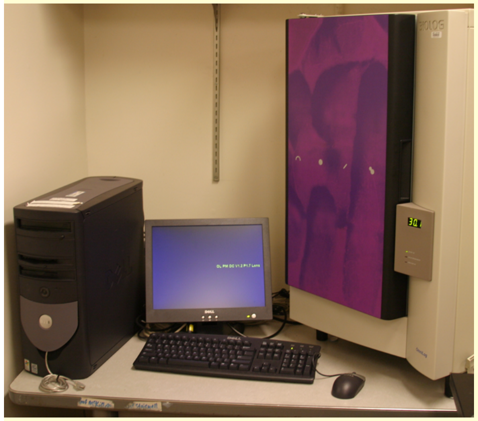
Homestake Mine
8,000 ft Lead, SD

Model of Single Organism Ecosystem

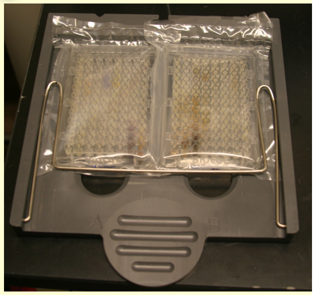


Phenomics, Proteomics & Lipidomics

- Phenomics - phenotype expression & physiology
  - Phenotypic microarrays
  - Real-time analyses using FTIR, etc
- Proteomics - protein expression
  - ICAT - Isotope-Coded Affinity Tags
  - DIGE - Differential In-Gel Electrophoresis
- Lipidomics - lipid/fatty acid expression especially as it relates to membranes and cell walls
  - FAME - Fatty Acid Methyl Ester
  - PLFA - Phospholipid Fatty Acid



Phenotypic Microarray

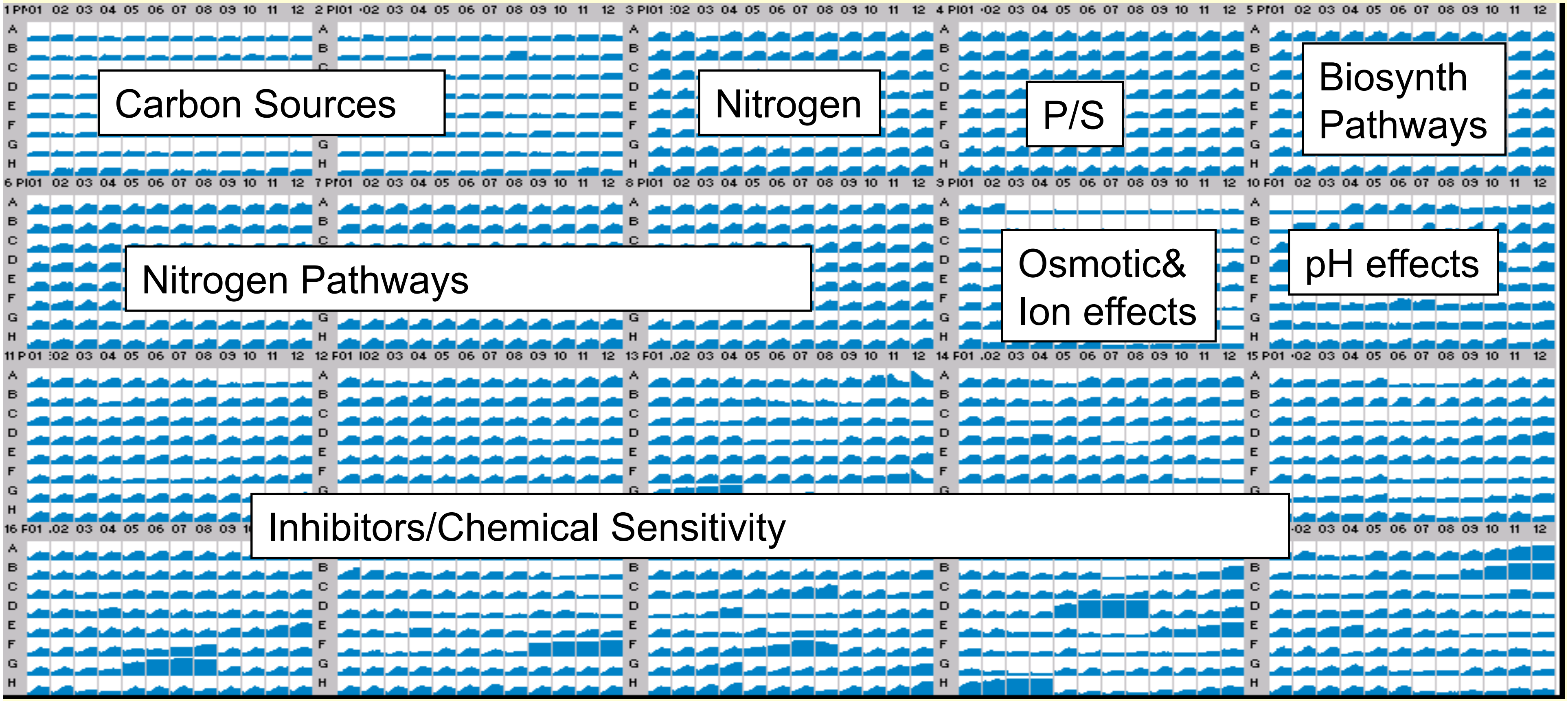


Omnilog System - 2000 assays,

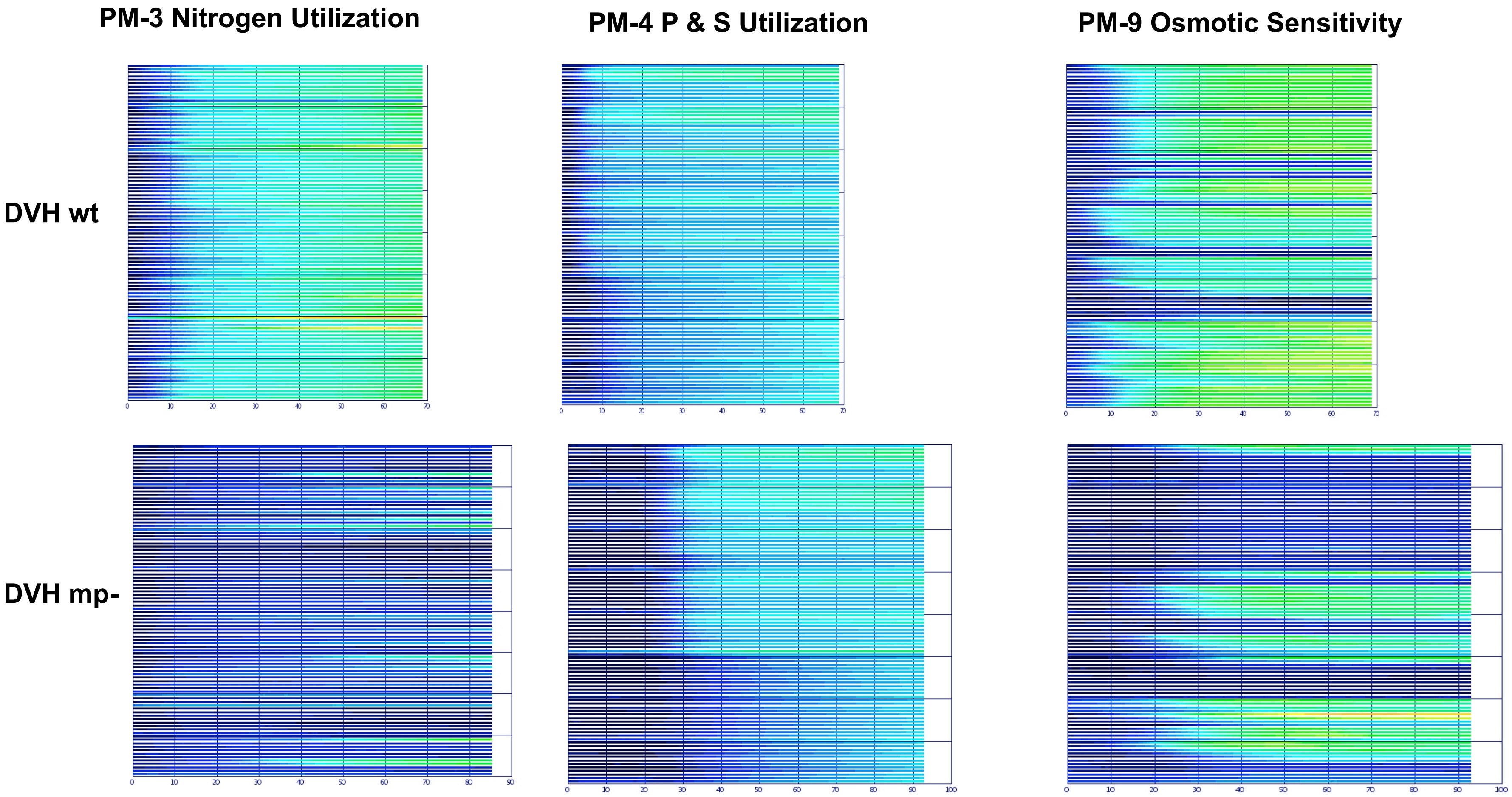
50 - 96-well plates at one time

>750 metabolic assays

239 inhibition/sensitivity assays



Desulfovibrio vulgaris phenomics wildtype vs. megaplasmid minus



PM-3 Nitrogen Utilization

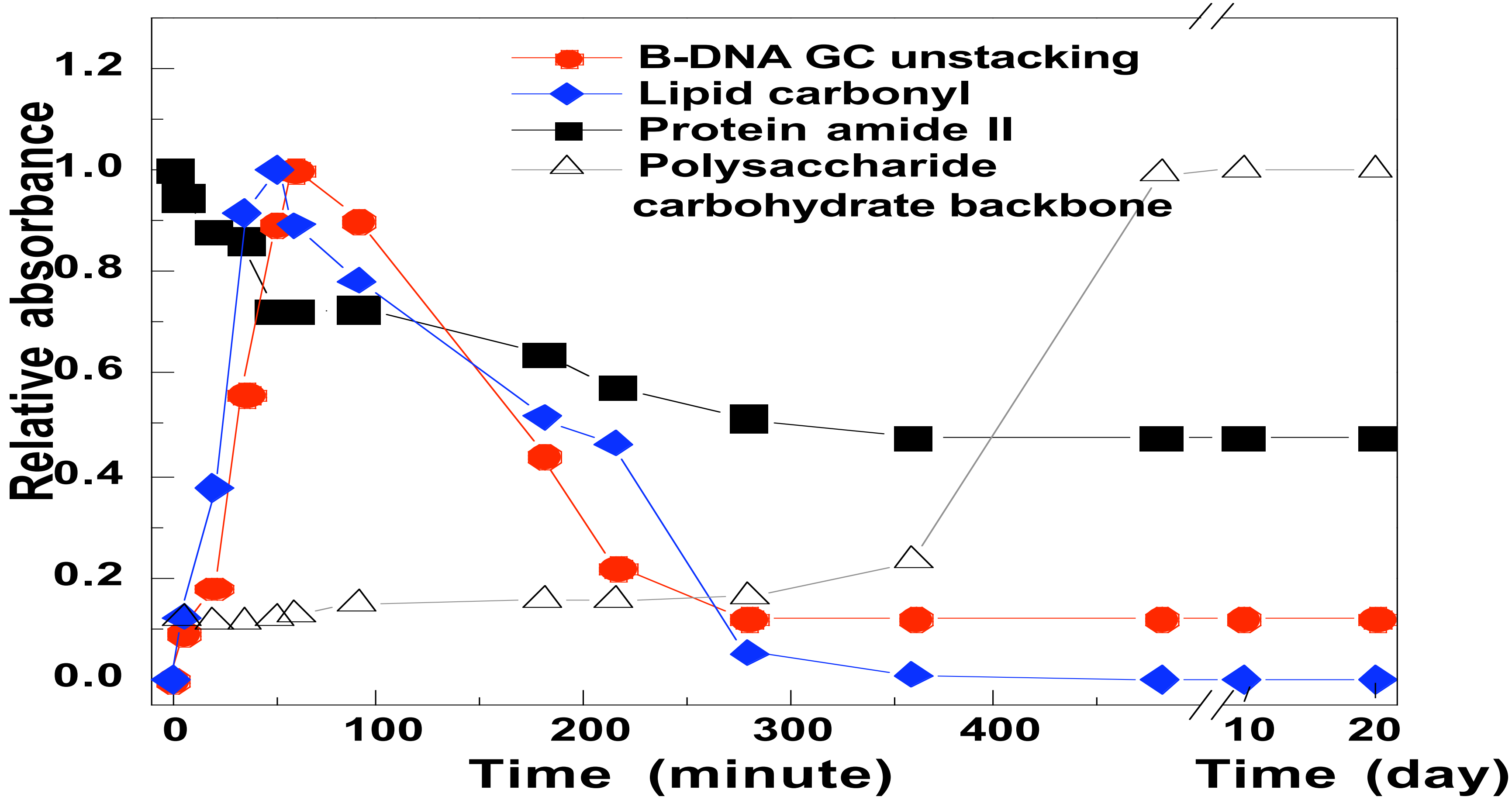
PM-4 P & S Utilization

PM-9 Osmotic Sensitivity

DVH wt

DVH mp-

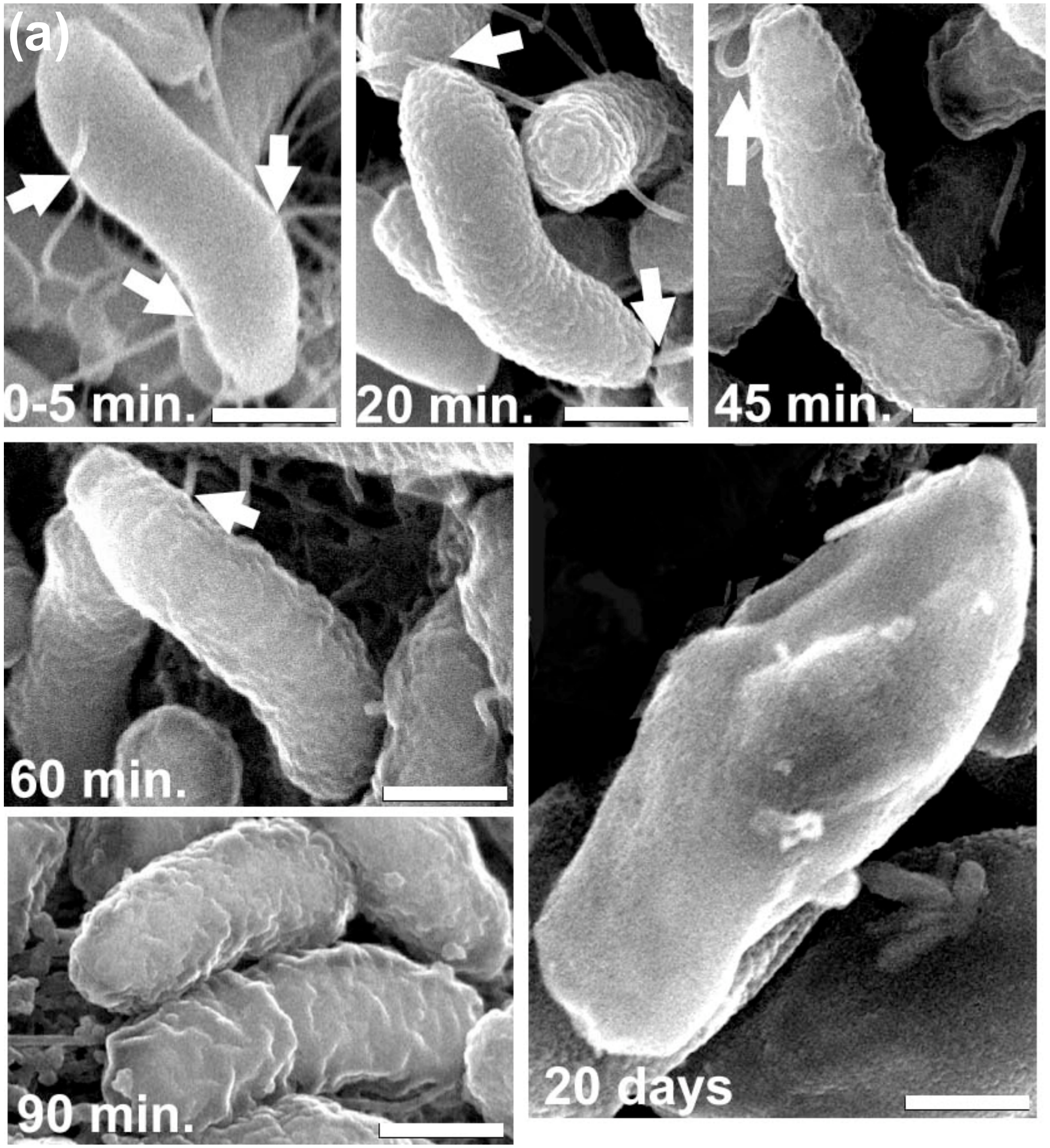
FTIR Profiling

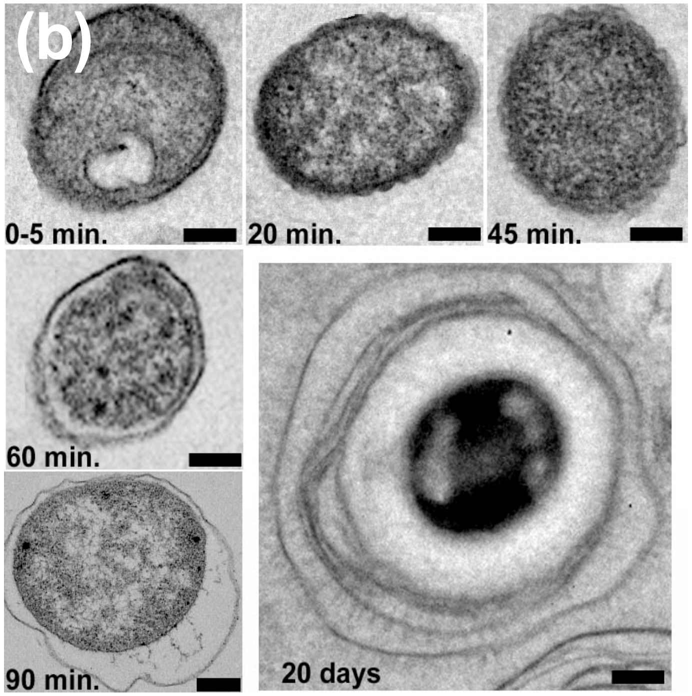


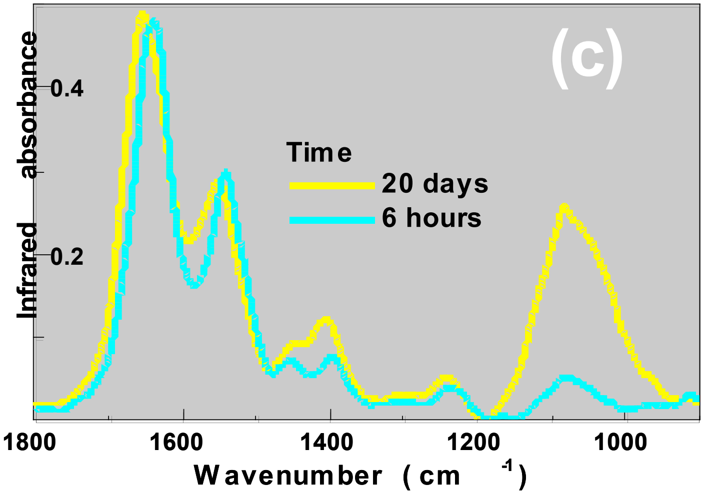
- Synchrotron FTIR time course of infrared absorption intensity, indicative of oxidative stress levels in different biologically important molecules in *Desulfovibrio vulgaris* after exposure to atmospheric oxygen.

- Also found signatures for Cytochrome B hemes

FTIR Profiling





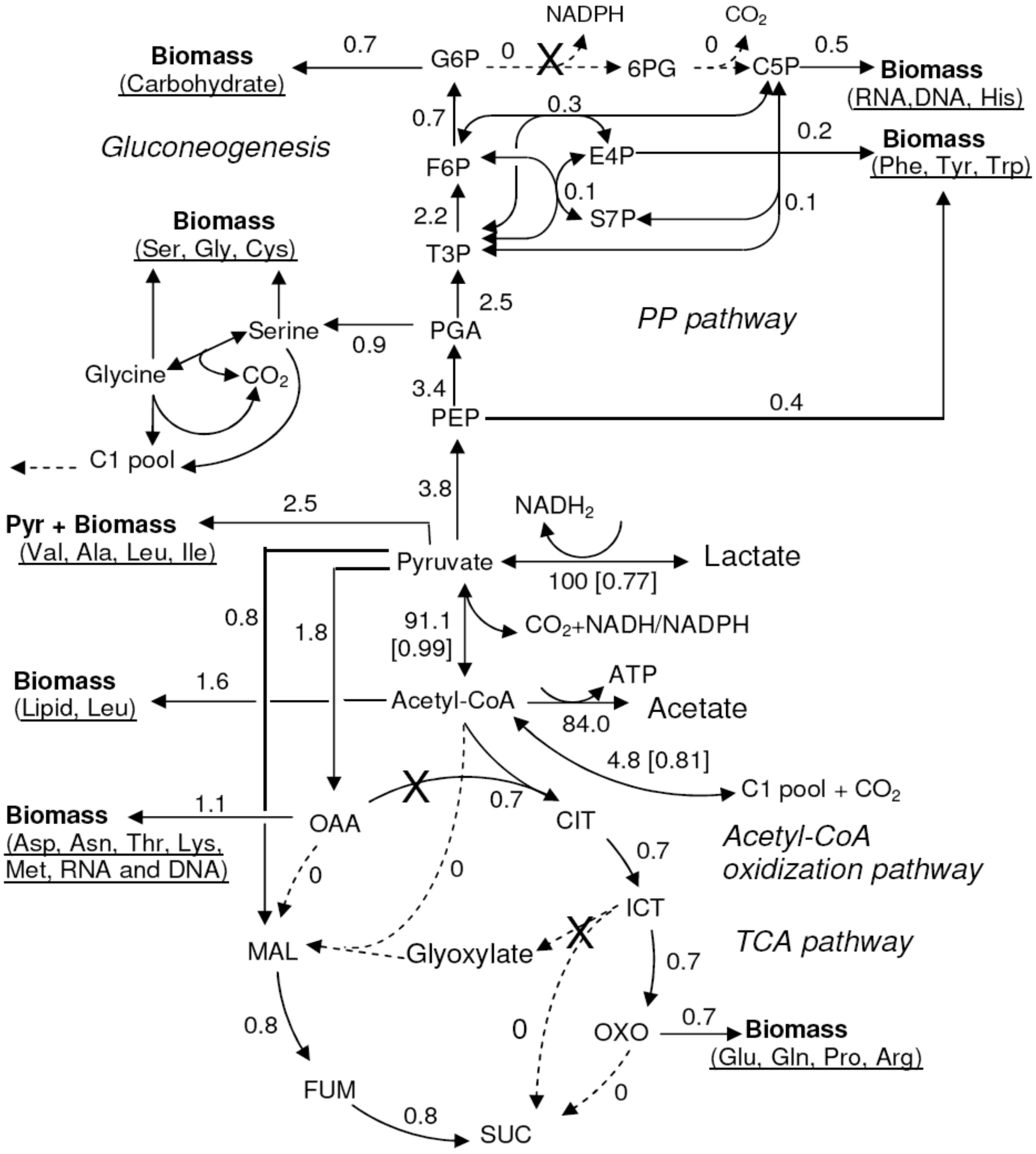


Metabolomics & Fluxomics

- Metabolomics- metabolite expression
    - hydrophilic interaction chromatography technique coupled to MS/MS detection and CE-MS methods for amino acids, nucleosides, nucleotides, organic acid CoAs, redox cofactors and the metabolic intermediates of glycolysis, TCA, and pentose phosphate pathway, etc.
- Fluxomics - studies of rate changes in metabolites
    - Same techniques as above
- These two areas are the newest and least developed, lots of development needs, but lots of breakthrough potential.

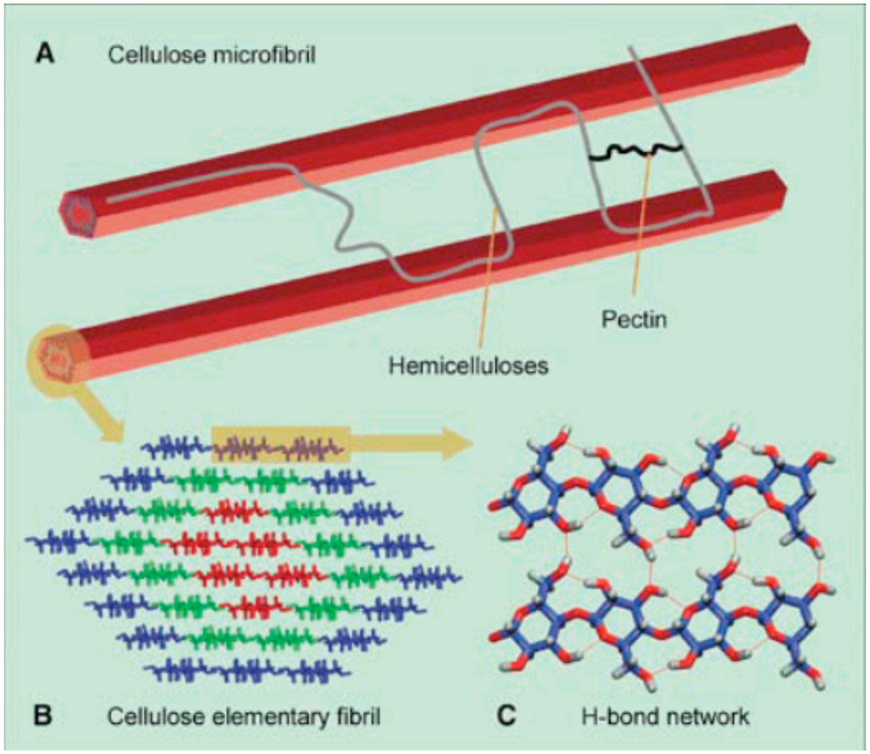
Desulfovibrio vulgaris Flux

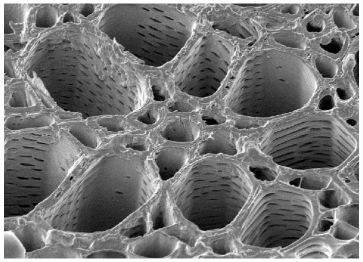
Tang, Y., F. Pingitore, A. Mukhopadhyay, R. Phan, T. C. Hazen, J. D. Keasling. 2007. Pathway confirmation and flux analysis using ^{13}C isotopic labeling of metabolites in Desulfovibrio vulgaris Hildenborough via FT-ICR Mass Spectrometry. J. Bacteriol. 189:940-949.

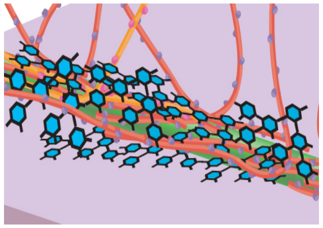


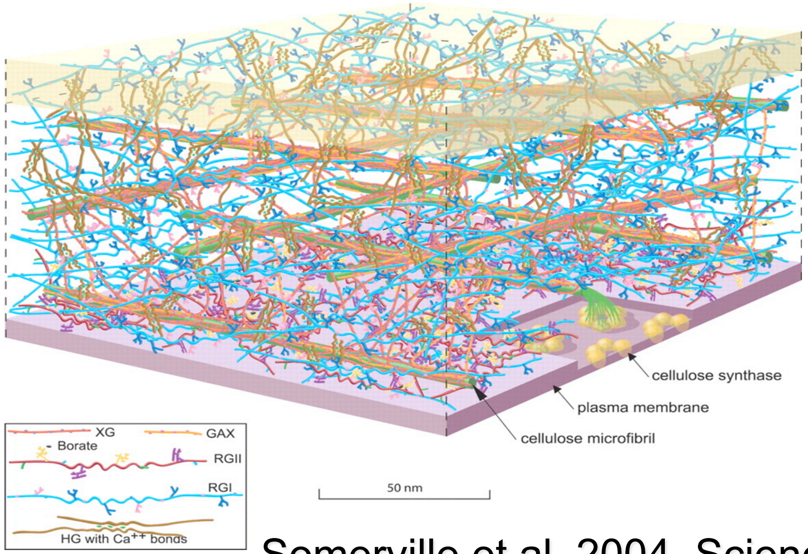
Glycomics

- The study of the entirity of carbohydrates in an organism or biological system
- Extreme interest for Bioenergy
  - Feedstocks, e.g. Lignocellulose degradation
  - Fuel synthesis, e.g. Fermentation for production of ethanol, butanol, etc
  - Medical applications for Diabetes, etc.









Somerville et al. 2004, Science

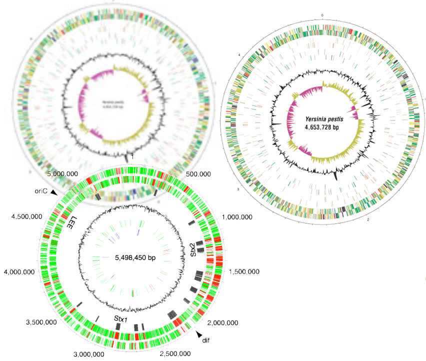
Bioinformatics

- Annotation of sequences
- Comparative genomics
- Integration from Biomolecules to Ecosystems
- Models for environmental biotechnology verification and prediction

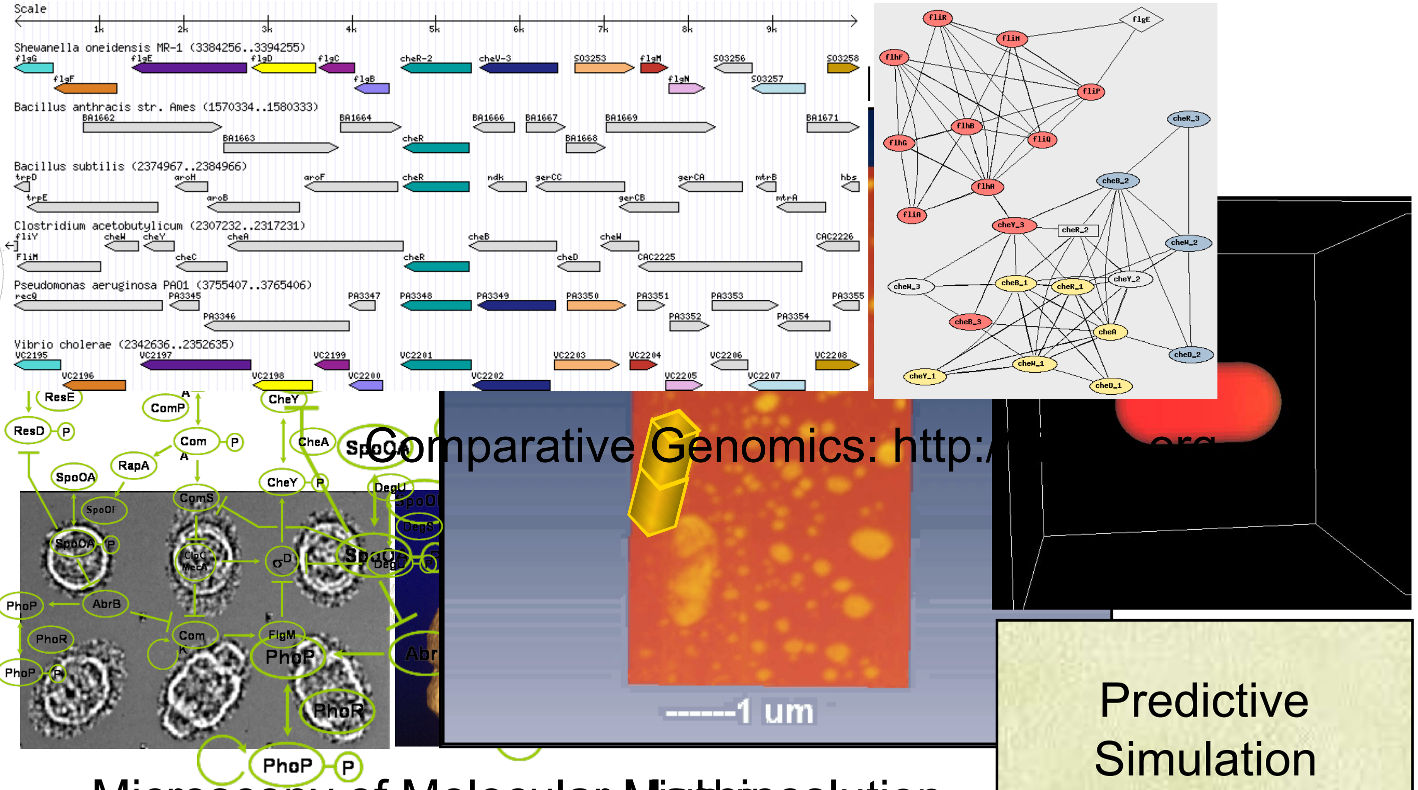
Critical Path

Centralized, Cross-Referenced Databases

- Environmental Characterization
- Environmental Sequence
- Functional Genomics
- Biophysically Characterized Molecules and Machines
- Cellular Imaging



Sequence Data



Comparative Genomics: http:/

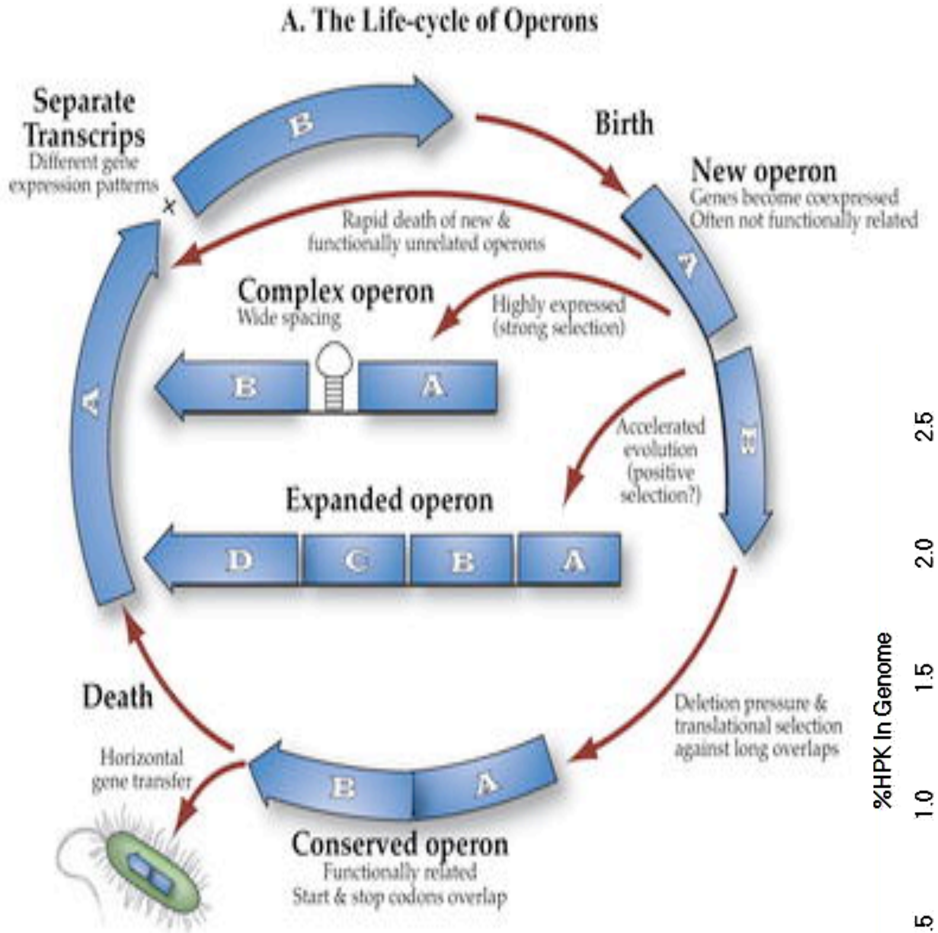
Microscopy of Molecular Machines

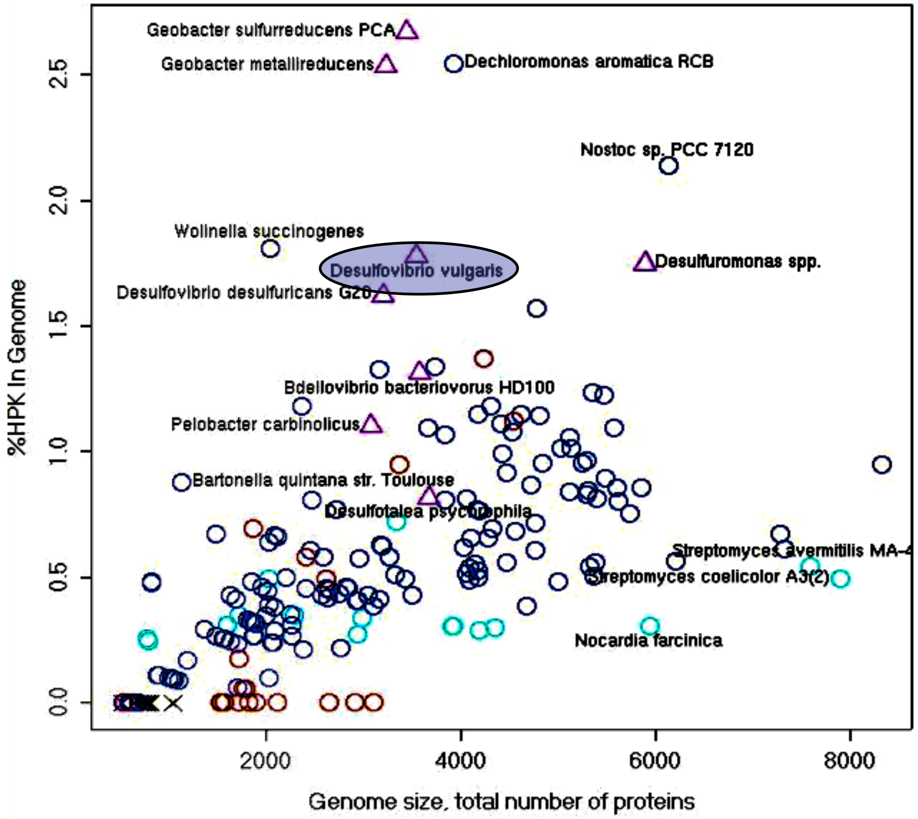
High resolution Cell Imaging

Deduction of pathways, modules and dynamics

Predictive Simulation

Comparative Genomics

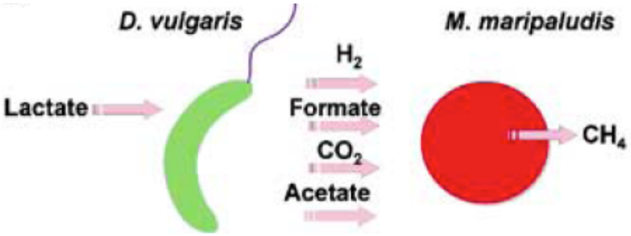




Amazing signal transduction apparatus

Modeling a Microbial Community

Metabolic network of *Desulfovibrio vulgaris* & *Methanococcus maripaludis* syntrophy

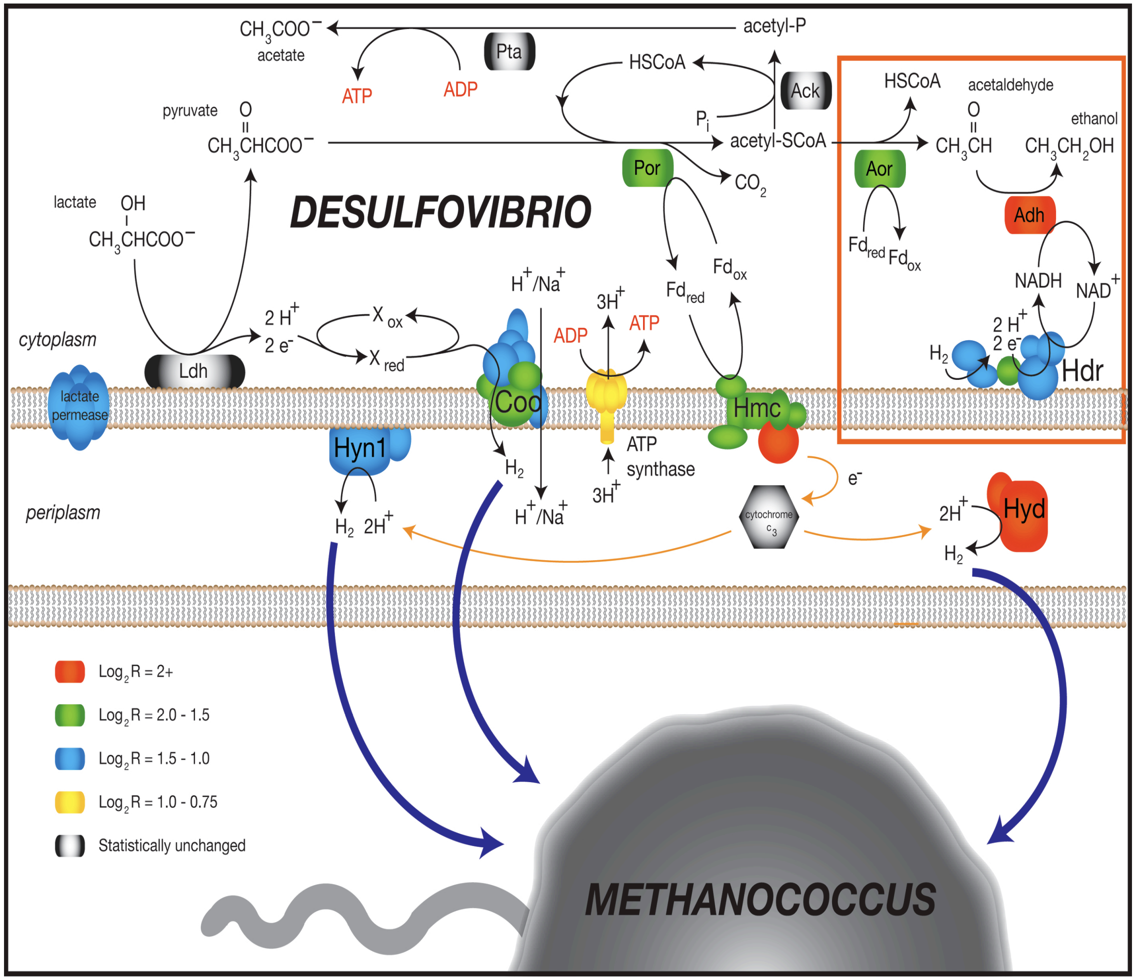


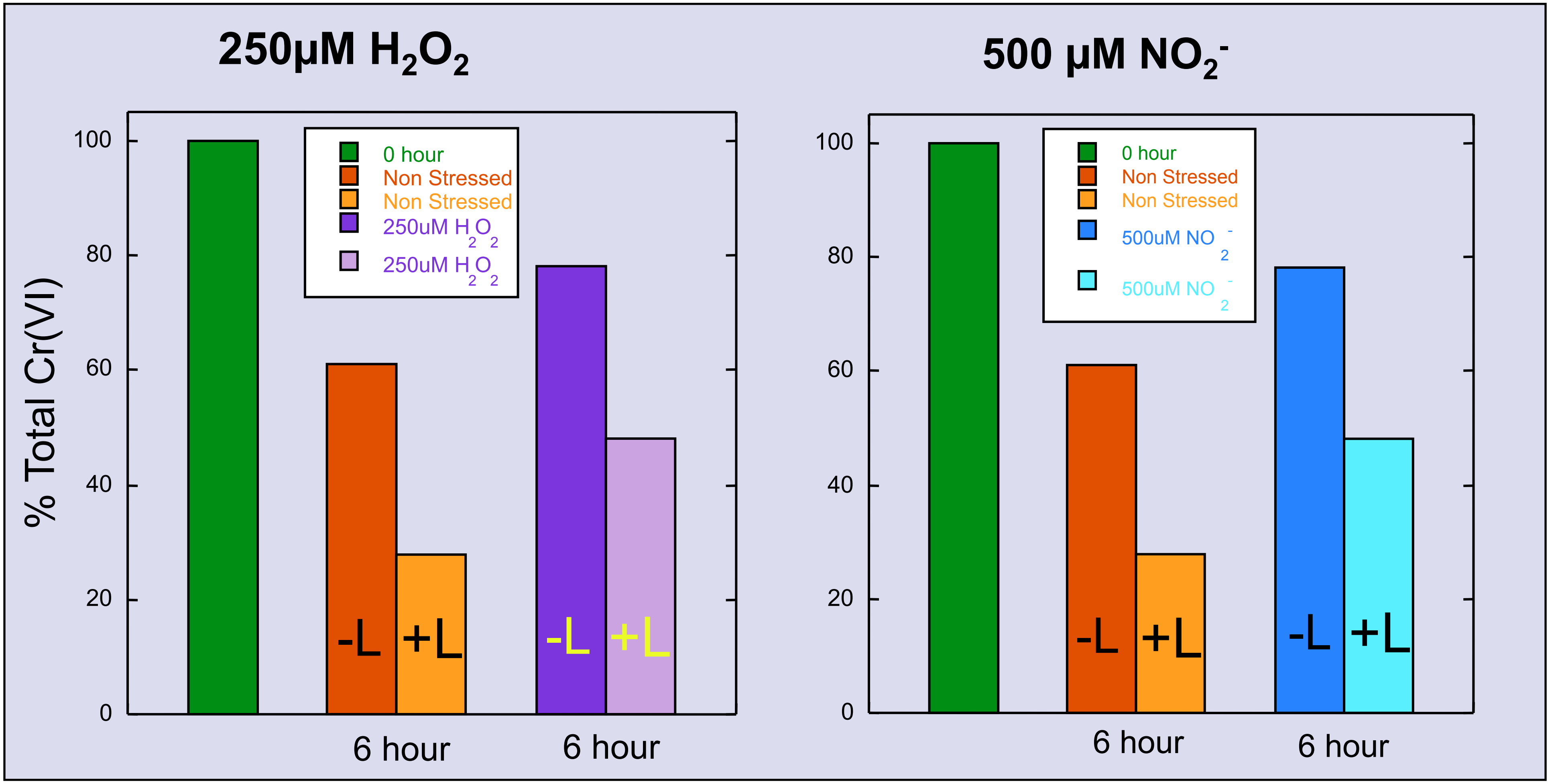
$$S \times v = b$$

S =Stoichiometric matrix
v = Vector of metabolic fluxes
b = Vector of system inputs and outputs

- The model consists of 170 reactions and 147 internal metabolites, mostly involved in central metabolism
- Simulations were performed by constraining the inputs to the system (i.e., lactate, sulfate) and optimizing for *Desulfovibrio* cell growth
- The model and our experimental data suggested that hydrogen transfer was essential for syntrophic growth but that formate was not essential

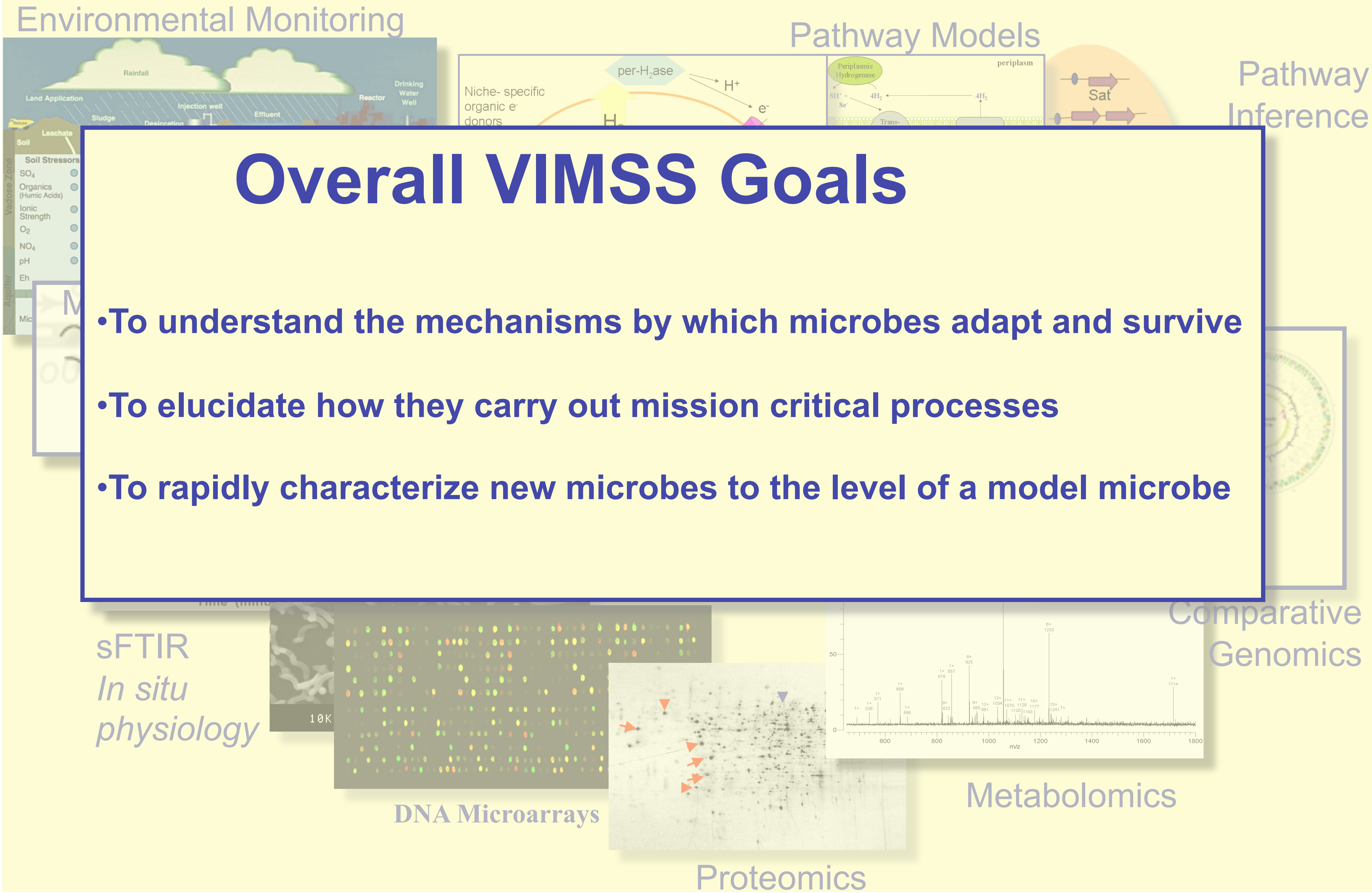
Stolyar, S., Van Dien, S., Hillesland, K.L., Pinel, N., Lie, T.J., Leigh, J.A., and Stahl, D. (2007) Metabolic modeling of a mutualistic microbial community. *Molecular Systems Biology* **3**:92





Coculture treated with 20mM lactate (+L) or not (-L) as an abiotic control. spiked with 200µM Cr(VI). Analyzed by a colorimetric DPC assay.

Virtual Institute of Microbial Stress and Survival



Overall VIMSS Goals

- To understand the mechanisms by which microbes adapt and survive
- To elucidate how they carry out mission critical processes
- To rapidly characterize new microbes to the level of a model microbe

Microbial Communities Department

Mission: *Discovery of Lignocellulolytic Enzymes and Metabolic Pathways*

Approach: *Deep and High-Throughput 'Omics' Analyses of Microbial Communities in potentially High Return Environments*

Initial Source Environments:

- *Green Waste Compost (high nutrient)*
- *Puerto Rican Rain Forest Soil (low nutrient)*

Delieverables: *Identified and isolated lignocellulolytic enzymes, metabolic pathways, whole cell biocatalysts, and mixed cultures for deconstruction*

JBEI Clients: *Deconstruction Division Enzyme Engineering Group and Fuels Synthesis Division*

Research Team Leads:
Terry C. Hazen (Director), LBNL
Jean VanderGheynst, UC Davis
Phil Hugenholtz, JGI
Michael Thelen, LLNL
Patrik D'haeseleer, LLNL
Steve Singer, LLNL

Cross-cutting Technologies

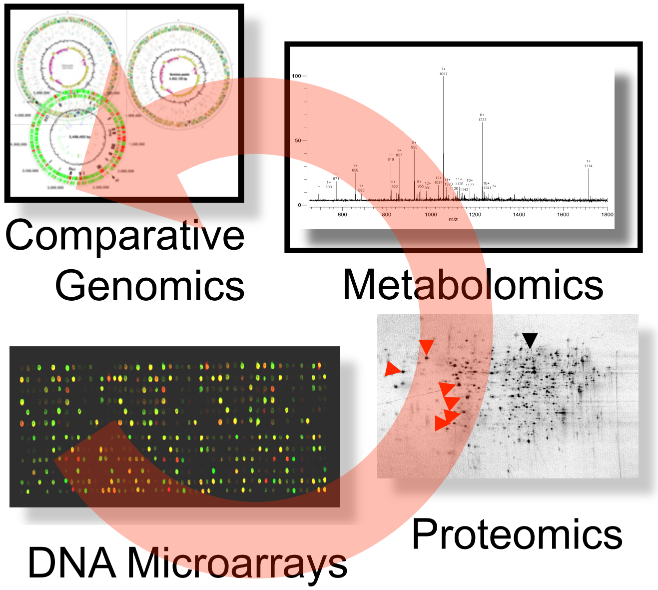
Omics

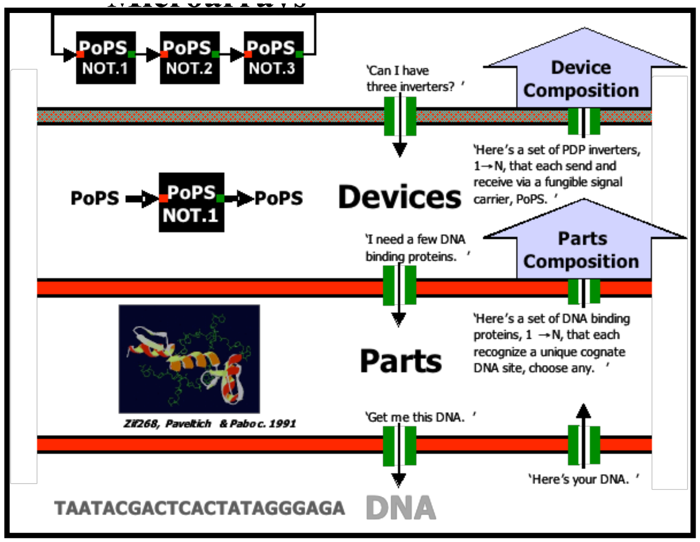
Multi-scale Imaging

Technology: new tools for biofuels research

Deliverables

- High-throughput microfluidics platforms for large scale analysis of plant and microbial enzyme activities
- Ligno- and glyco-arrays for rapid screening of enzymatic function
- ‘Omics’ pipelines for systems biology
- Integrated data capture, analysis and dissemination
- Parts, devices, chassis for synthetic biology

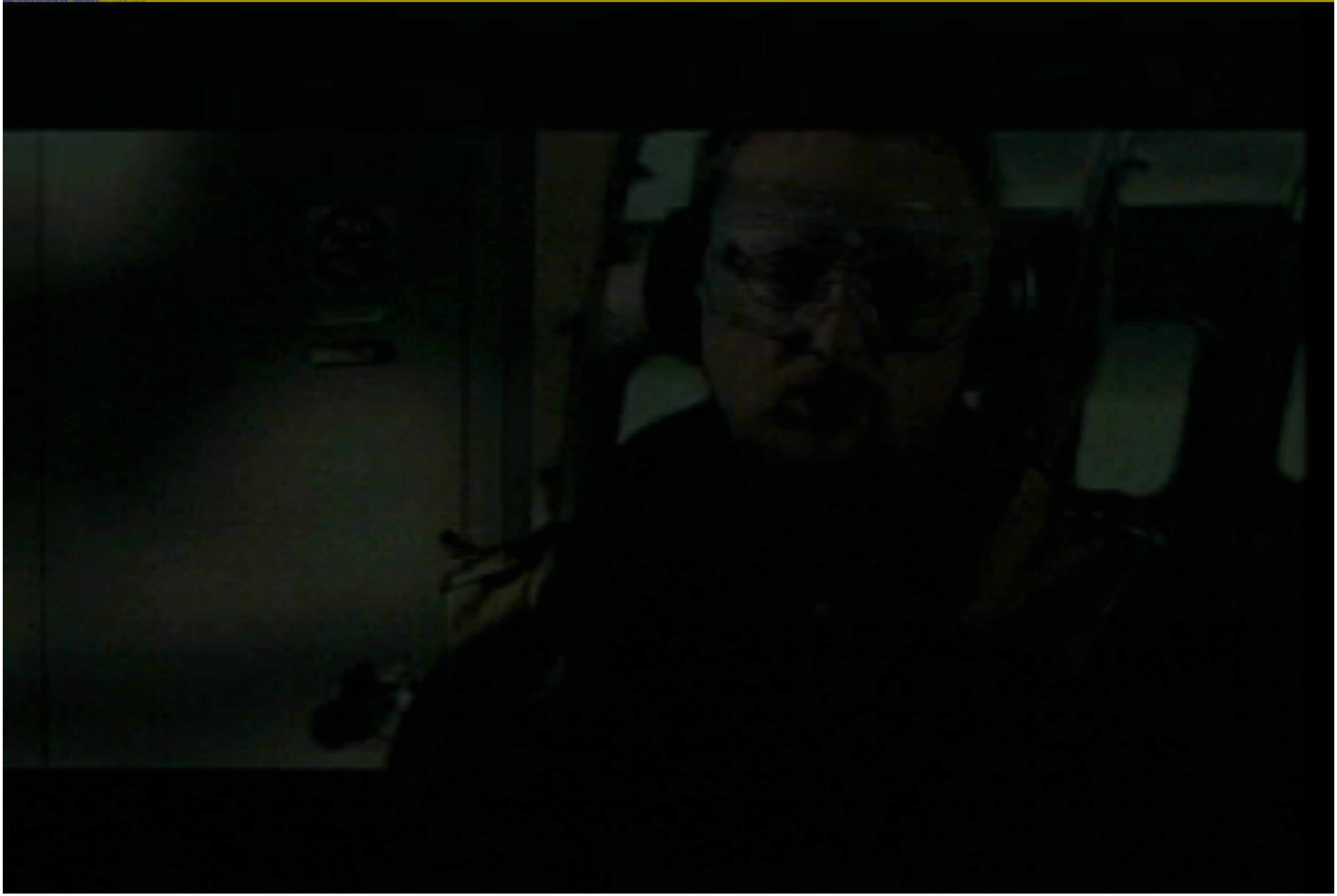


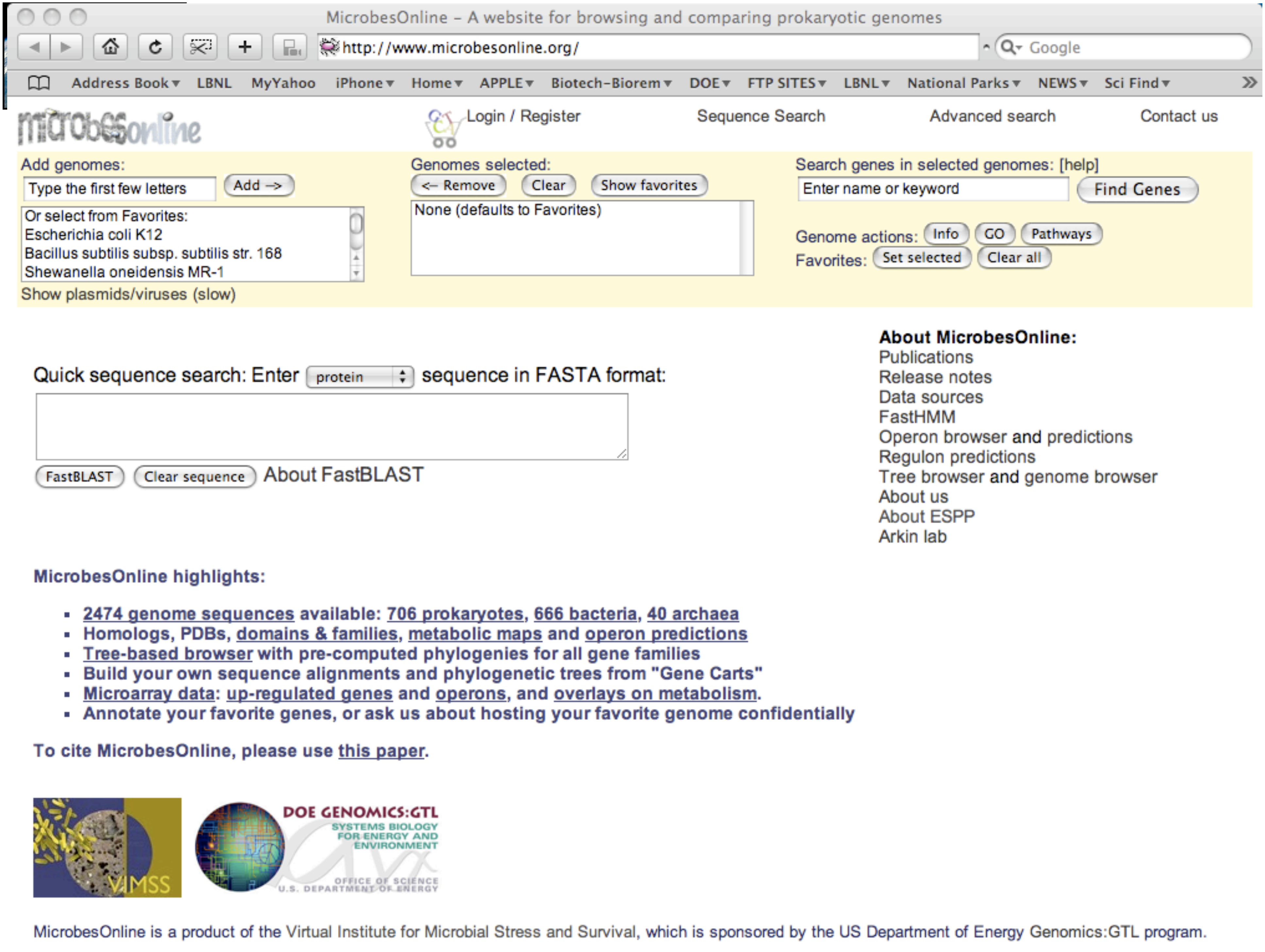


JBEI
JOINT BIOENERGY INSTITUTE

DOE Bioenergy Research Centers

How we feel at times!





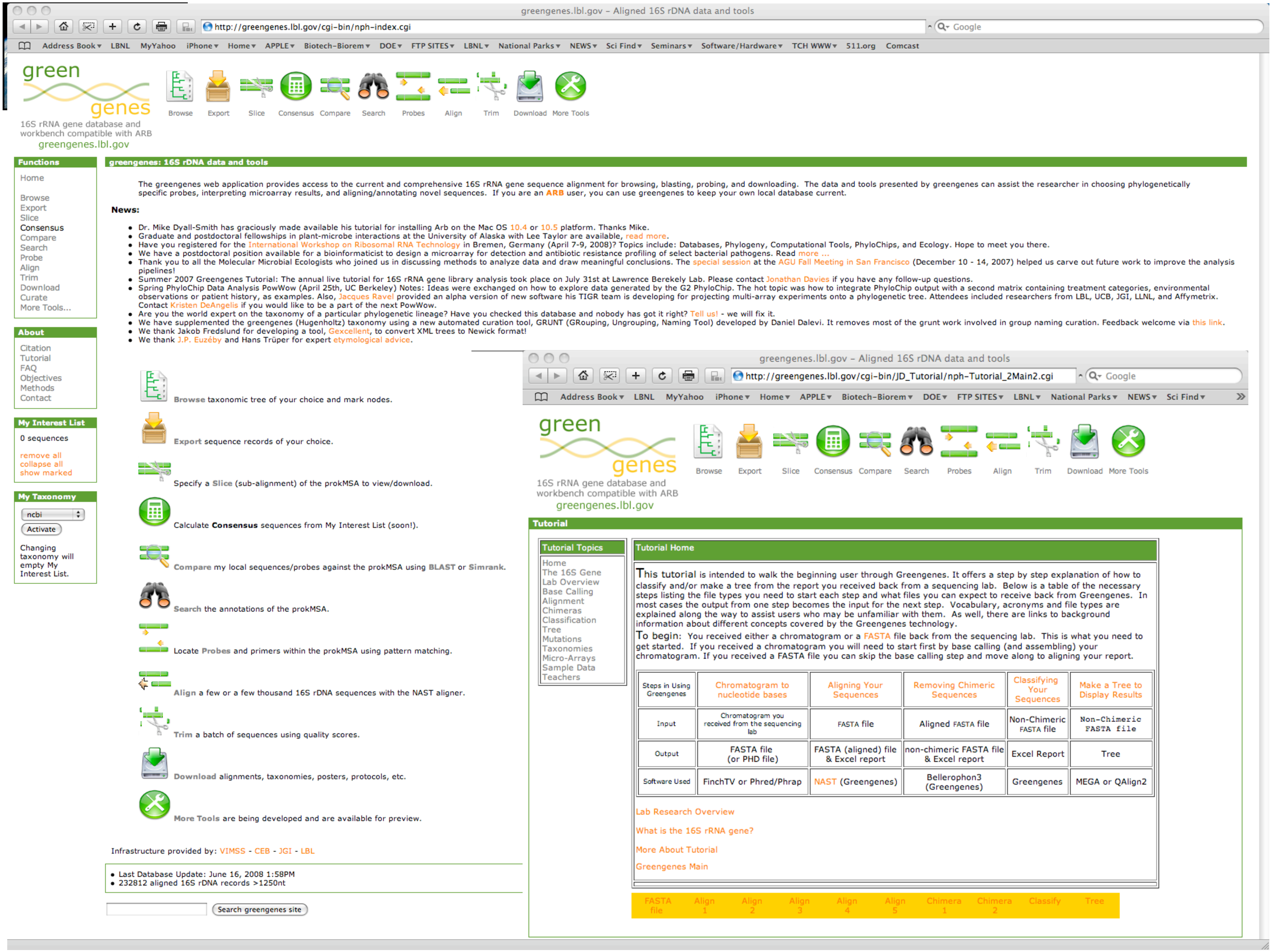
Quick sequence search: Enter protein sequence in FASTA format:

MicrobesOnline highlights:

- 2474 genome sequences available: 706 prokaryotes, 666 bacteria, 40 archaea
- Homologs, PDBs, domains & families, metabolic maps and operon predictions
- Tree-based browser with pre-computed phylogenies for all gene families
- Build your own sequence alignments and phylogenetic trees from "Gene Carts"
- Microarray data: up-regulated genes and operons, and overlays on metabolism.
- Annotate your favorite genes, or ask us about hosting your favorite genome confidentially

To cite MicrobesOnline, please use this paper.

MicrobesOnline is a product of the Virtual Institute for Microbial Stress and Survival, which is sponsored by the US Department of Energy Genomics:GTL program.



greengenes: 16S rDNA data and tools

The greengenes web application provides access to the current and comprehensive 16S rRNA gene sequence alignment for browsing, blasting, probing, and downloading. The data and tools presented by greengenes can assist the researcher in choosing phylogenetically specific probes, interpreting microarray results, and aligning/annotating novel sequences. If you are an **ARB** user, you can use greengenes to keep your own local database current.

News:

- Dr. Mike Dyall-Smith has graciously made available his tutorial for installing Arb on the Mac OS 10.4 or 10.5 platform. Thanks Mike.
- Graduate and postdoctoral fellowships in plant-microbe interactions at the University of Alaska with Lee Taylor are available, read more.
- Have you registered for the International Workshop on Ribosomal RNA Technology in Bremen, Germany (April 7-9, 2008)? Topics include: Databases, Phylogeny, Computational Tools, PhyloChips, and Ecology. Hope to meet you there.
- We have a postdoctoral position available for a bioinformaticist to design a microarray for detection and antibiotic resistance profiling of select bacterial pathogens. Read more ...
- Thank you to all the Molecular Microbial Ecologists who joined us in discussing methods to analyze data and draw meaningful conclusions. The special session at the AGU Fall Meeting in San Francisco (December 10 - 14, 2007) helped us carve out future work to improve the analysis pipelines!
- Summer 2007 Greengenes Tutorial: The annual live tutorial for 16S rRNA gene library analysis took place on July 31st at Lawrence Berekely Lab. Please contact Jonathan Davies if you have any follow-up questions.
- Spring PhyloChip Data Analysis PowWow (April 25th, UC Berkeley) Notes: Ideas were exchanged on how to explore data generated by the G2 PhyloChip. The hot topic was how to integrate PhyloChip output with a second matrix containing treatment categories, environmental observations or patient history, as examples. Also, Jacques Ravel provided an alpha version of new software his TIGR team is developing for projecting multi-array experiments onto a phylogenetic tree. Attendees included researchers from LBL, UCB, JGI, LLNL, and Affymetrix. Contact Kristen DeAngelis if you would like to be a part of the next PowWow.
- Are you the world expert on the taxonomy of a particular phylogenetic lineage? Have you checked this database and nobody has got it right? Tell us! - we will fix it.
- We have supplemented the greengenes (Hugenholtz) taxonomy using a new automated curation tool, GRUNT (GRouping, Ungrouping, Naming Tool) developed by Daniel Dalevi. It removes most of the grunt work involved in group naming curation. Feedback welcome via this link.
- We thank Jakob Fredslund for developing a tool, Gexcellent, to convert XML trees to Newick format!
- We thank J.P. Euzéby and Hans Trüper for expert etymological advice.

Browse taxonomic tree of your choice and mark nodes.

Export sequence records of your choice.

Specify a **Slice** (sub-alignment) of the prokMSA to view/download.

Calculate **Consensus** sequences from My Interest List (soon!).

Compare my local sequences/probes against the prokMSA using **BLAST** or **Simrank**.

Search the annotations of the prokMSA.

Locate **Probes** and primers within the prokMSA using pattern matching.

Align a few or a few thousand 16S rDNA sequences with the NAST aligner.

Trim a batch of sequences using quality scores.

Download alignments, taxonomies, posters, protocols, etc.

More Tools are being developed and are available for preview.

Infrastructure provided by: VIMSS - CEB - JGI - LBL

- Last Database Update: June 16, 2008 1:58PM
- 232812 aligned 16S rDNA records >1250nt

Tutorial Home

This tutorial is intended to walk the beginning user through Greengenes. It offers a step by step explanation of how to classify and/or make a tree from the report you received back from a sequencing lab. Below is a table of the necessary steps listing the file types you need to start each step and what files you can expect to receive back from Greengenes. In most cases the output from one step becomes the input for the next step. Vocabulary, acronyms and file types are explained along the way to assist users who may be unfamiliar with them. As well, there are links to background information about different concepts covered by the Greengenes technology.

To begin: You received either a chromatogram or a FASTA file back from the sequencing lab. This is what you need to get started. If you received a chromatogram you will need to start first by base calling (and assembling) your chromatogram. If you received a FASTA file you can skip the base calling step and move along to aligning your report.

Steps in Using Greengenes	Chromatogram to nucleotide bases	Aligning Your Sequences	Removing Chimeric Sequences	Classifying Your Sequences	Make a Tree to Display Results
Input	Chromatogram you received from the sequencing lab	FASTA file	Aligned FASTA file	Non-Chimeric FASTA file	Non-Chimeric FASTA file
Output	FASTA file (or PHD file)	FASTA (aligned) file & Excel report	non-chimeric FASTA file & Excel report	Excel Report	Tree
Software Used	FinchTV or Phred/Phrap	NAST (Greengenes)	Bellerophon3 (Greengenes)	Greengenes	MEGA or QAlign2

Lab Research Overview

What is the 16S rRNA gene?

More About Tutorial

Greengenes Main

VIMSS

Acknowledgments

Berkeley Lab

The University of California Berkeley

Adam Arkin, Eric Alm, Kat Huang, Dylan Chivian, Janet Jacobson, Jay Keasling, Aindrila Mukhopadhyay, Eoin Brodie, Sharon Borglin, Hoi-Ying Holman, Jil Geller, Lauren Camp, Cindy Wu, Dominique Joyner, Julian Fortney, Romy Chakraborty, Boris Faybishenko, Mark Conrad, Zouping Zheng, Gary Andersen, Todd DeSantis, Tetsu Tokunaga, Jiamin Wan, Susan Hubbard, Ken Williams, John Peterson, Jill Banfield, Tamas Torok, Seung Baek, Don Herman, Mary Firestone, Rebecca Daly, Kate Goldfarb

Pacific Northwest National Laboratory ...delivering breakthrough science and technology — Phil Long, Fred Brockman, et al.

JGI DOE Joint Genome Institute — Paul Richardson, Phil Hugenholtz, Inna Dubchez

Montana State University — Mathew Fields, et. al.

The University of Chicago — Stephen Sutton, Matthew Newville

Jizhong Zhou, et. al.

University of Washington — David Stahl, et. al.

Indiana University — Lisa Pratt, et. al.

Regenesis — Steve Koenigsberg, Ana Willet

Missouri — Judy Wall, et. al.

Massachusetts Institute of Technology — Eric Alm, et. al.

Sandia National Laboratories — Anup Singh, et. al.

Oak Ridge National Laboratory — Martin Keller, et. al.

Southern Illinois University Carbondale — Kelly Bender, et. al.

Princeton — T. C. Onstott, et. al.

ESPP2 is part of the Virtual Institute for Microbial Stress and Survival supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics Program:GTL through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.

Environmental Remediation Sciences

DOE Bioenergy Research Centers

Contacts

Dr. Terry C. Hazen tchazen@lbl.gov

Center for Environmental Biotechnology

http://www-esd.lbl.gov/CEB

Virtual Institute for Microbial Stress and Survival

http://vimss.lbl.gov

http://www.microbesonline.org - Comparative Genomics Database, Genome Browser, Operon Browser, Regulon Browser

Environmental Remediation Technology Program

http://www-esd.lbl.gov/ERT

Ecology Department

http://www-esd.lbl.gov/ECO

DOE Environmental Remediation Sciences Program

http://www.lbl.gov/ERSP

Joint BioEnergy Institute

http://www.jbei.org

DOE Genomics:GTL Program

http://doegenomestolife.org

NASA Indiana Princeton Tennessee Astrobiology Initiative (IPTAI)

http://www.indiana.edu/%7Edeeplife/homepg.html

YouTube: Bioremediation: The Hope and the Hype of Environmental Cleanup

http://www.youtube.com/watch?v=MT0qY3_n1kI&fmt=18