Complex environments

Permafrost soil

Human gut

Native prairie soils

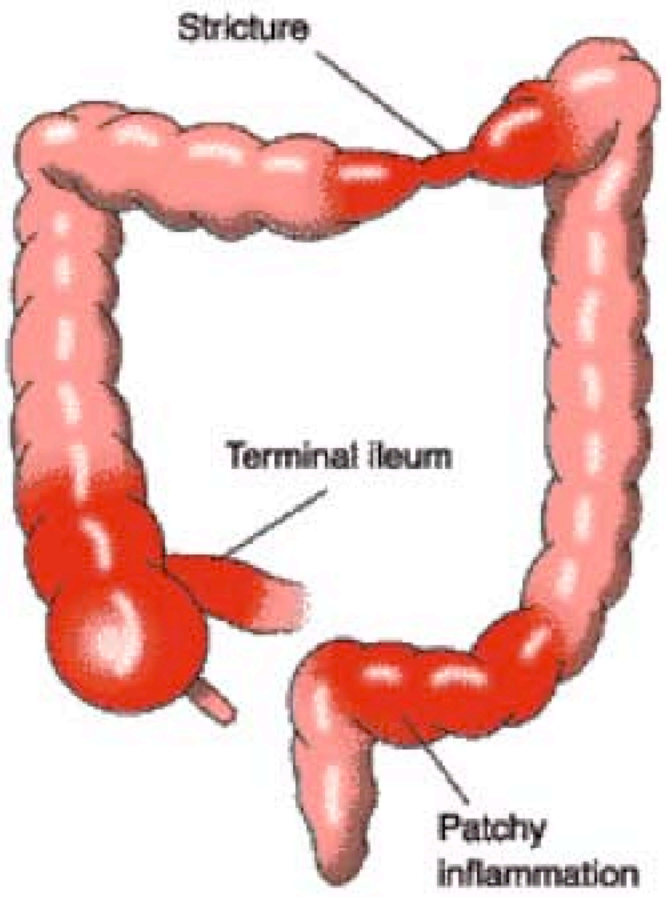
Microbial diversity in nature is enormous

-1 gram of soil contains >10^9 microbial cells and >10^4 species

-1 gram intestinal material contains >10^{11} microorganisms and >10^3 species

Crohn's disease

- 0.1% of Western population
- Inflammation of small bowel (ileum) or larger bowel (colon)
- Genetic and environmental factors

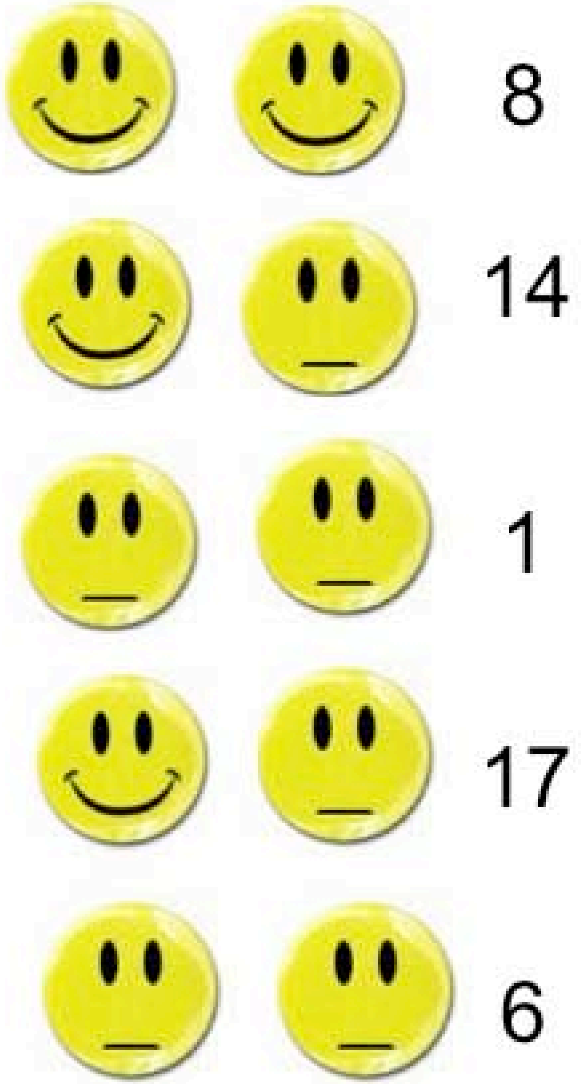


Aim: Study discordant twins (one is healthy and one is sick)

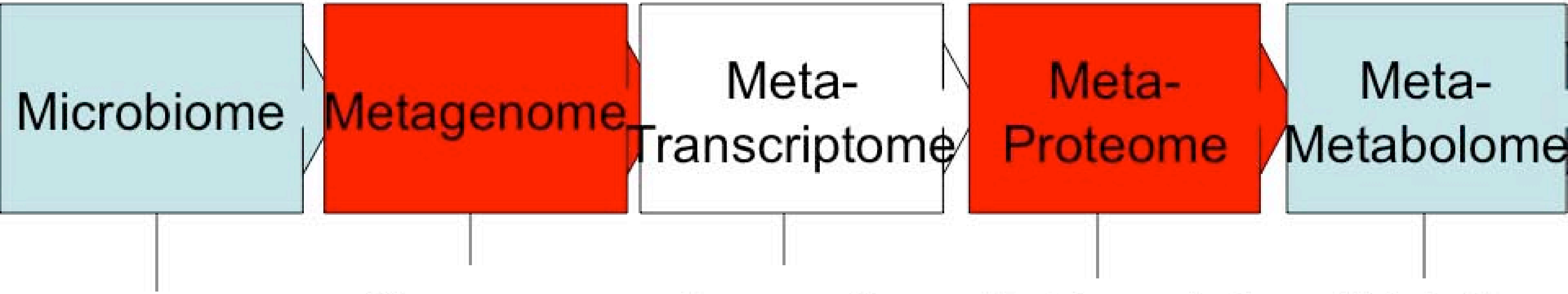
Swedish Twin Cohort

46 twin pairs (92 individuals)

- Healthy twin pairs
- Ulcerative colitis (UC)
  - 14 discordant (9 MZ and 5 DZ)
  - 1 concordant (MZ)
- Crohn's disease
  - 17 discordant (11 MZ and 6 DZ)
  - 6 concordant (all MZ)



Levels of information



Microbiome

Metagenome

Meta-Transcriptome

Meta-Proteome

Meta-Metabolome

Community composition

454 pyrotag sequencing

250,000 sequences

Gene composition (including functional genes)

NIH- HMP

Claire Fraser-Liggett

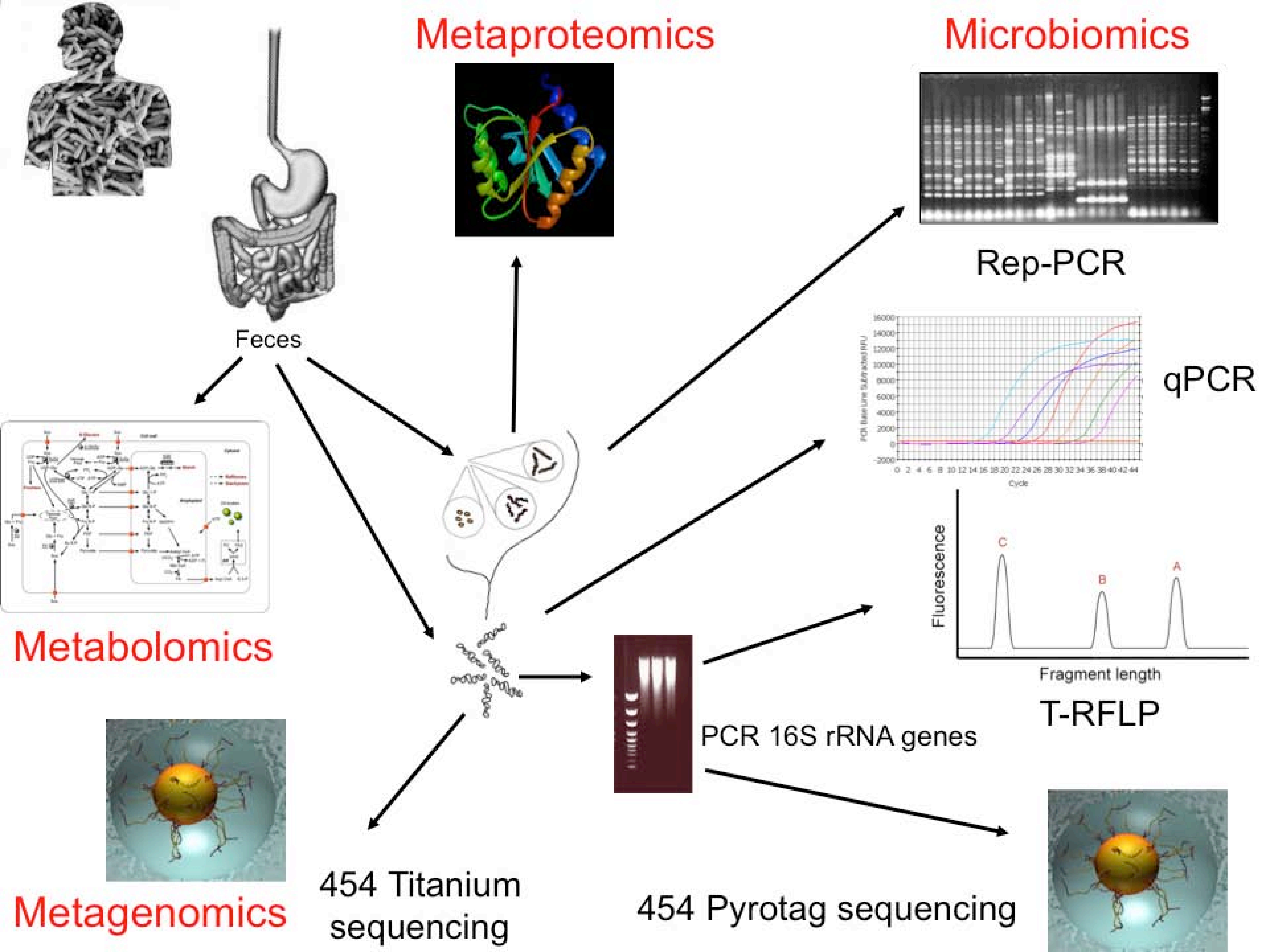
Expressed genes

Protein products

Matched samples

Crohn’s twins

Metabolites produced



Metaproteomics

Microbiomics

Rep-PCR

Feces

qPCR

Metabolomics

T-RFLP

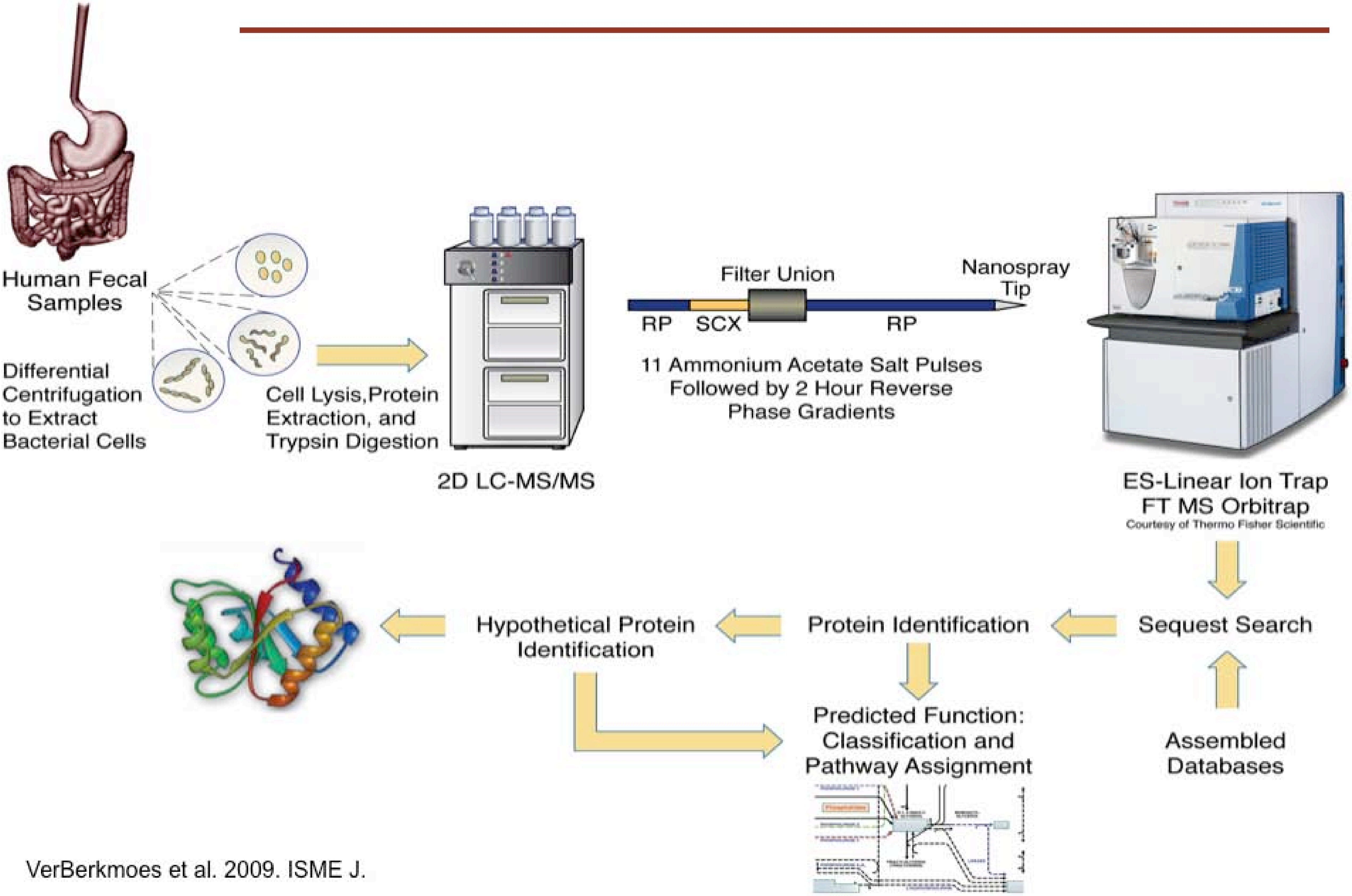
PCR 16S rRNA genes

Metagenomics

454 Titanium sequencing

454 Pyrotag sequencing

Metaproteomics of the Human Gut Microbiome



Human Fecal Samples

Differential Centrifugation to Extract Bacterial Cells

Cell Lysis, Protein Extraction, and Trypsin Digestion

2D LC-MS/MS

Filter Union

Nanospray Tip

RP

SCX

RP

11 Ammonium Acetate Salt Pulses Followed by 2 Hour Reverse Phase Gradients

ES-Linear Ion Trap FT MS Orbitrap

Courtesy of Thermo Fisher Scientific

Sequest Search

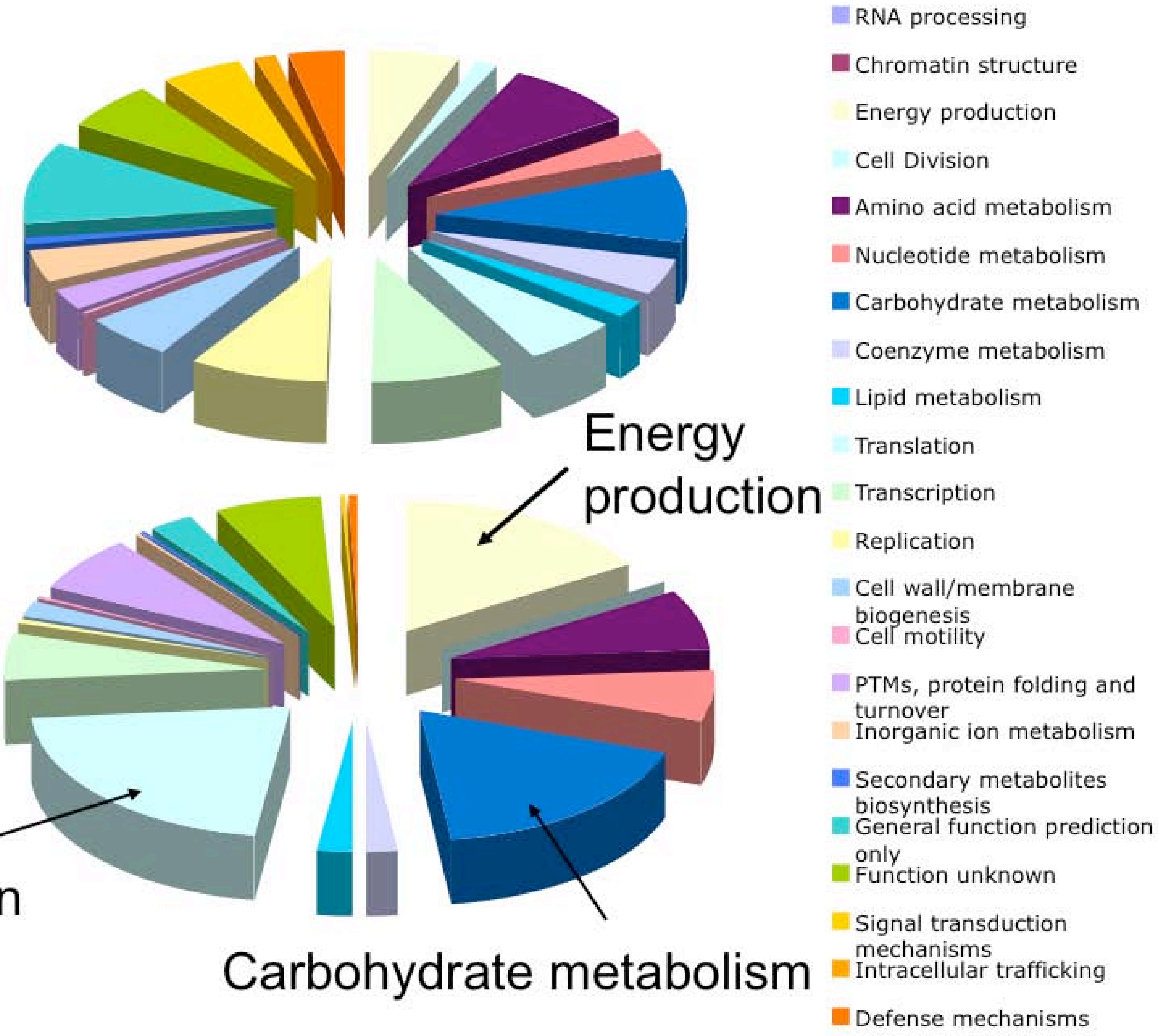
Assembled Databases

Protein Identification

Hypothetical Protein Identification

Predicted Function: Classification and Pathway Assignment

VerBerkmoes et al. 2009. ISME J.



Energy production

Translation

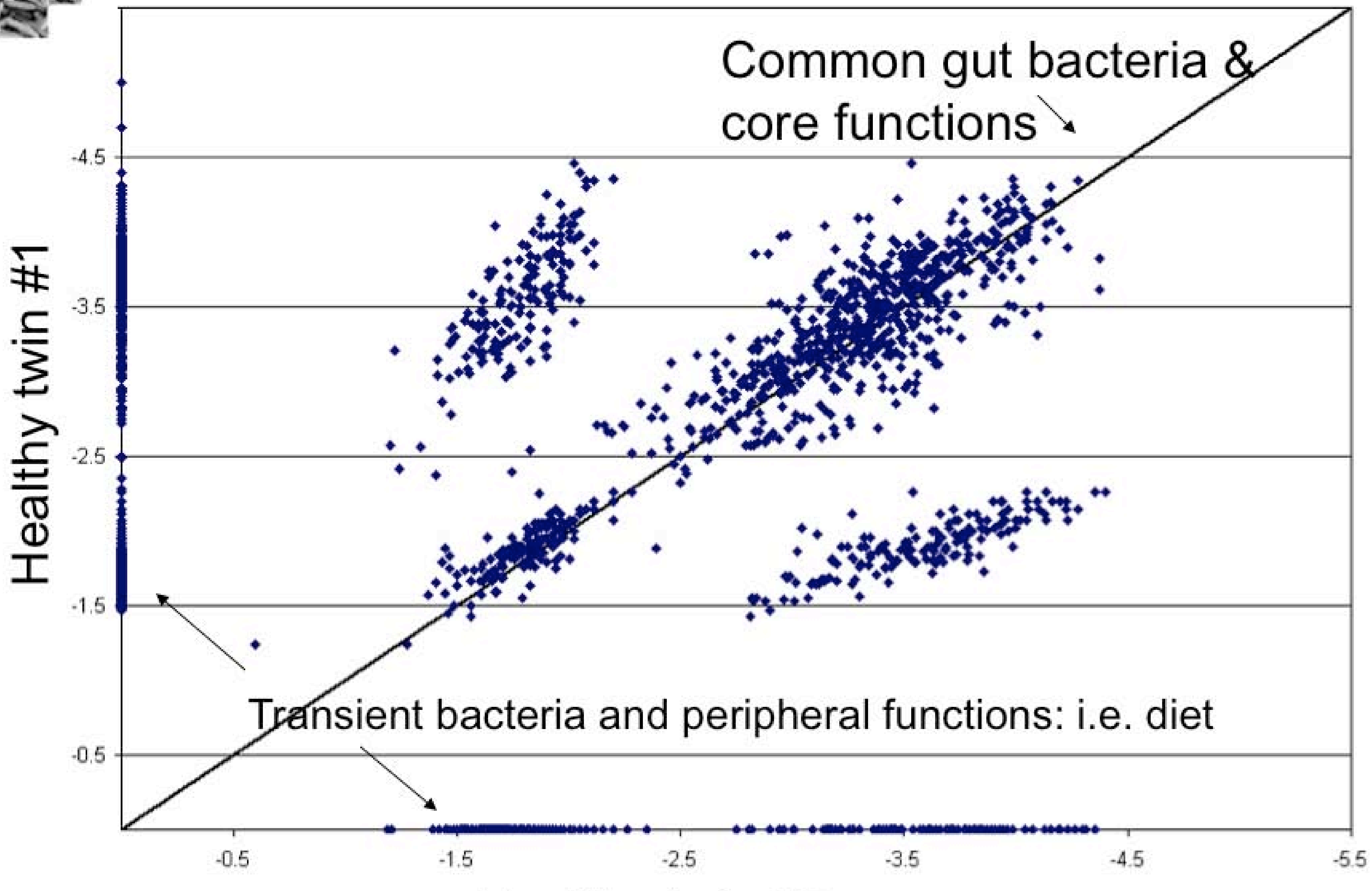
Carbohydrate metabolism

VerBerkmoes et al. 2009. ISME J.

Comparison of average metaproteme to average metagenome (healthy)

- Fewer proteins expressed:
  - Inorganic ion metabolism
  - Cell wall synthesis, membrane biogenesis and cell division
  - Secondary metabolite biosynthesis
- Both have high abundance of genes and proteins for butyrate and host-derived fucose metabolism
- Different H_2 disposal routes
  - Metaproteomes: Formyltetrahydrofolate synthetase (key enzyme in acetyl-coA pathway of acetogens)
  - Metagenomes: methanogenesis (20-50% of western population)

Comparison of normalized proteomes



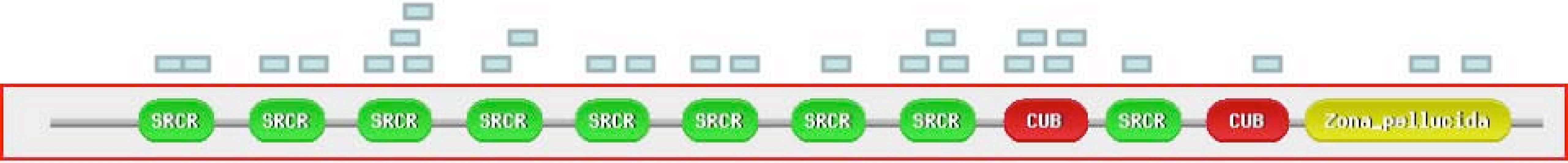
Common gut bacteria & core functions

Transient bacteria and peripheral functions: i.e. diet

VerBerkmoes et al. 2009. ISME J.

Almost 30% of proteins were human!

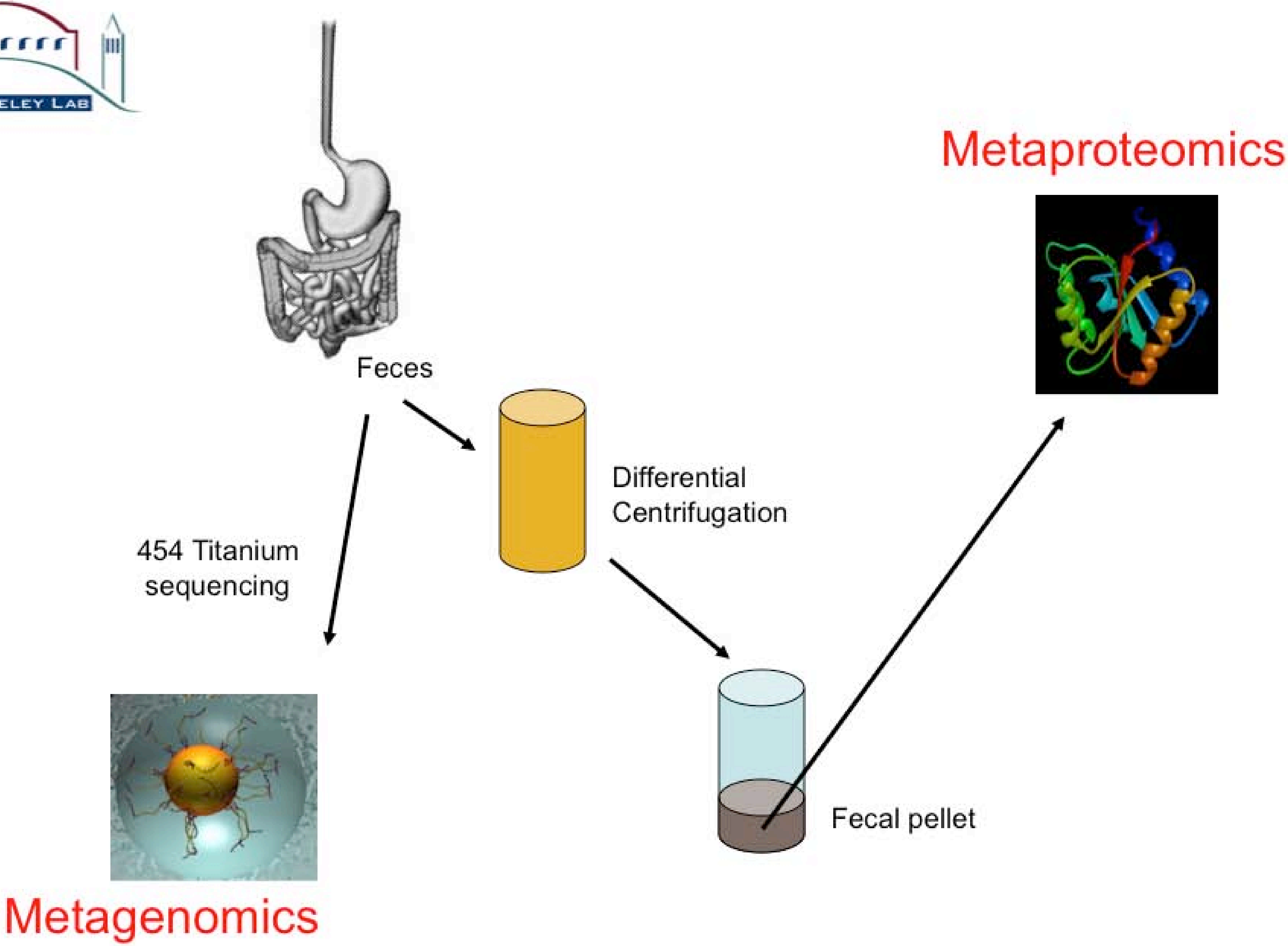
- Two largest groups:
    - Digestive enzymes (elastase, amylase, etc)
    - structural cell adhesion such as actin and myosin
- Third largest category were human immunity proteins, **including antimicrobial peptides**
- DMBT1 (not detected in malignant brain tumors) protein was one of the most abundant proteins
    - Plays a role in cellular immune response in the human digestive tract.



VerBerkmoes et al. 2009. ISME J.

Matched metagenomes and metaproteomes from Crohn's twins

- Twin pairs:
  - 1 healthy twin pair
  - 2 concordant ileal CD
  - 2 discordant ileal CD
  - 1 concordant colonic CD
- Metagenome sequences: 454 Titanium
- Metaproteomes:Shotgun metaproteomics
- Databases Searched
  - 2 American healthy metagenomes (Gill et al., 2006)
  - 13 Japanese metagenomes (Kurokawa et al., 2007)
  - Human
  - Matched metagenomes from Twin pairs (this study)



Feces

454 Titanium
sequencing

Metagenomics

Differential
Centrifugation

Fecal pellet

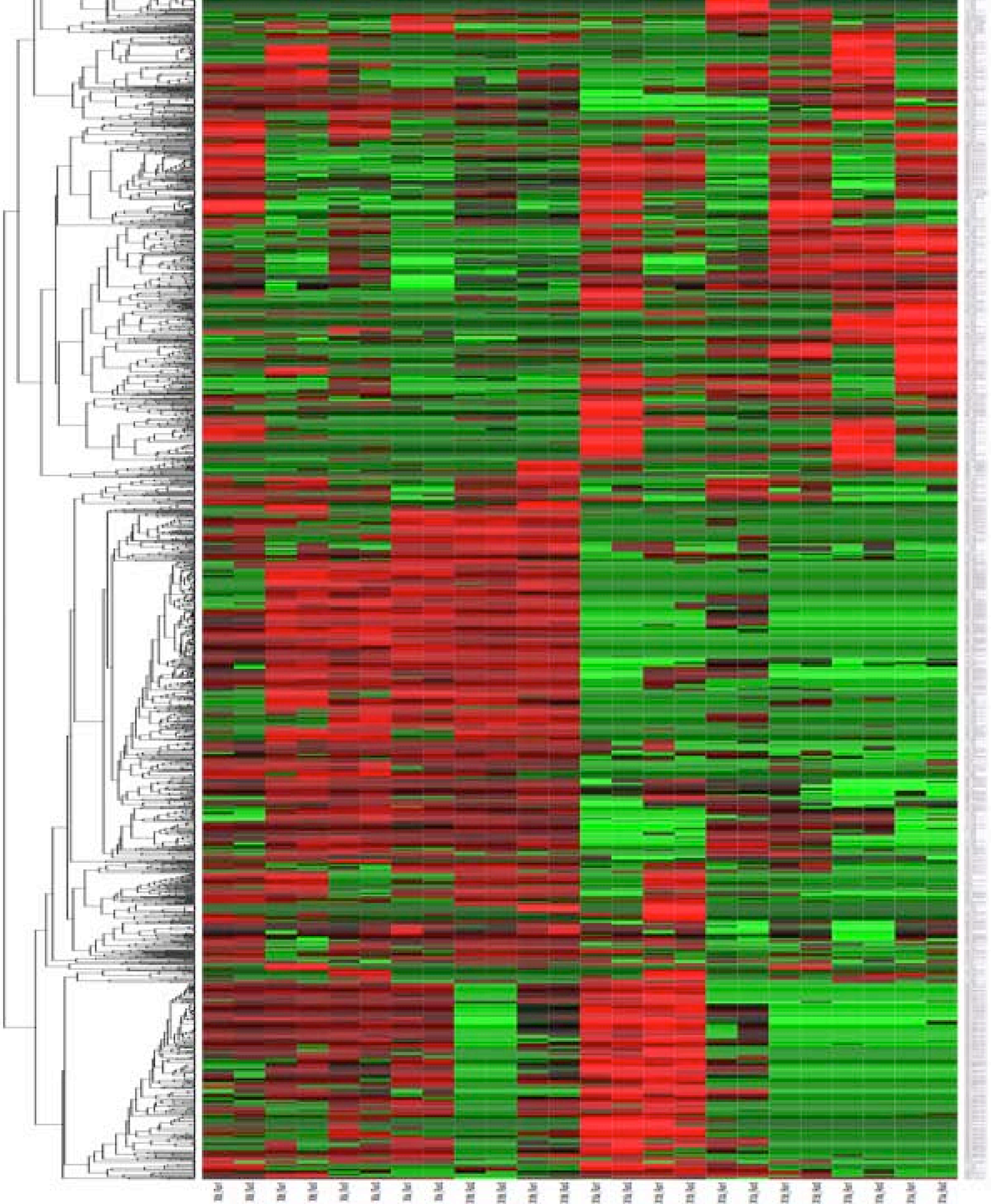
Metaproteomics

Preliminary Metaproteome Results

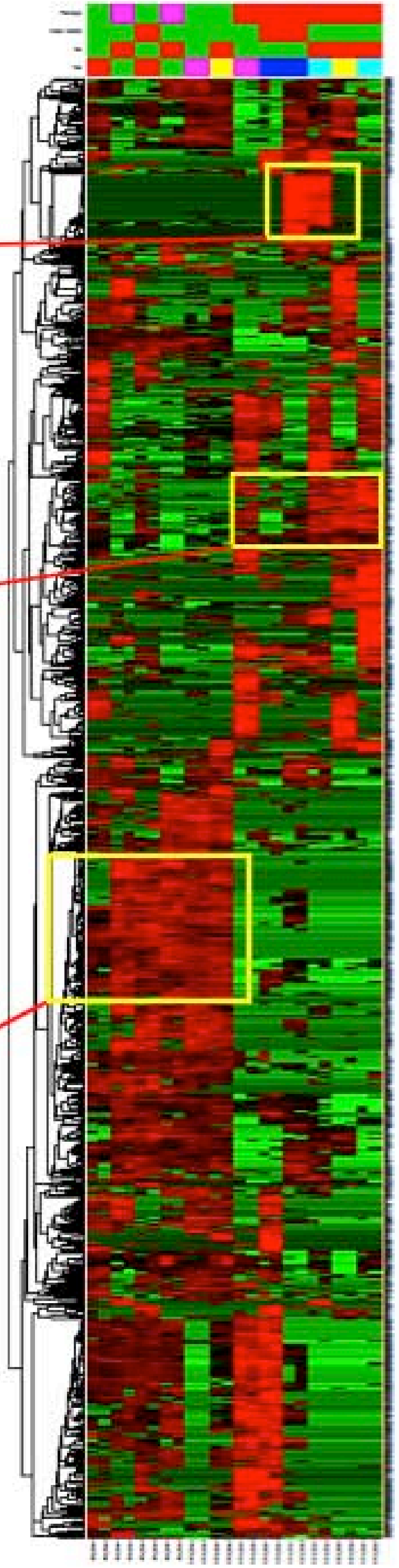
Subject	Avg. Protein Count (duplicate MS runs)
Healthy	3257
Healthy	3058
Crohns	2772
Crohns	2553
Crohns	2814
Crohns	2116
Crohns	1826
Crohns	2352
Crohns	1872
Healthy	4074
Crohns	2228
Healthy	4060

LTQ-Orbitrap Results

Initial observation: Fewer proteins are identified in subjects with CD (~2300) as compared to healthy individuals (~3600)



Protein expression patterns across all samples

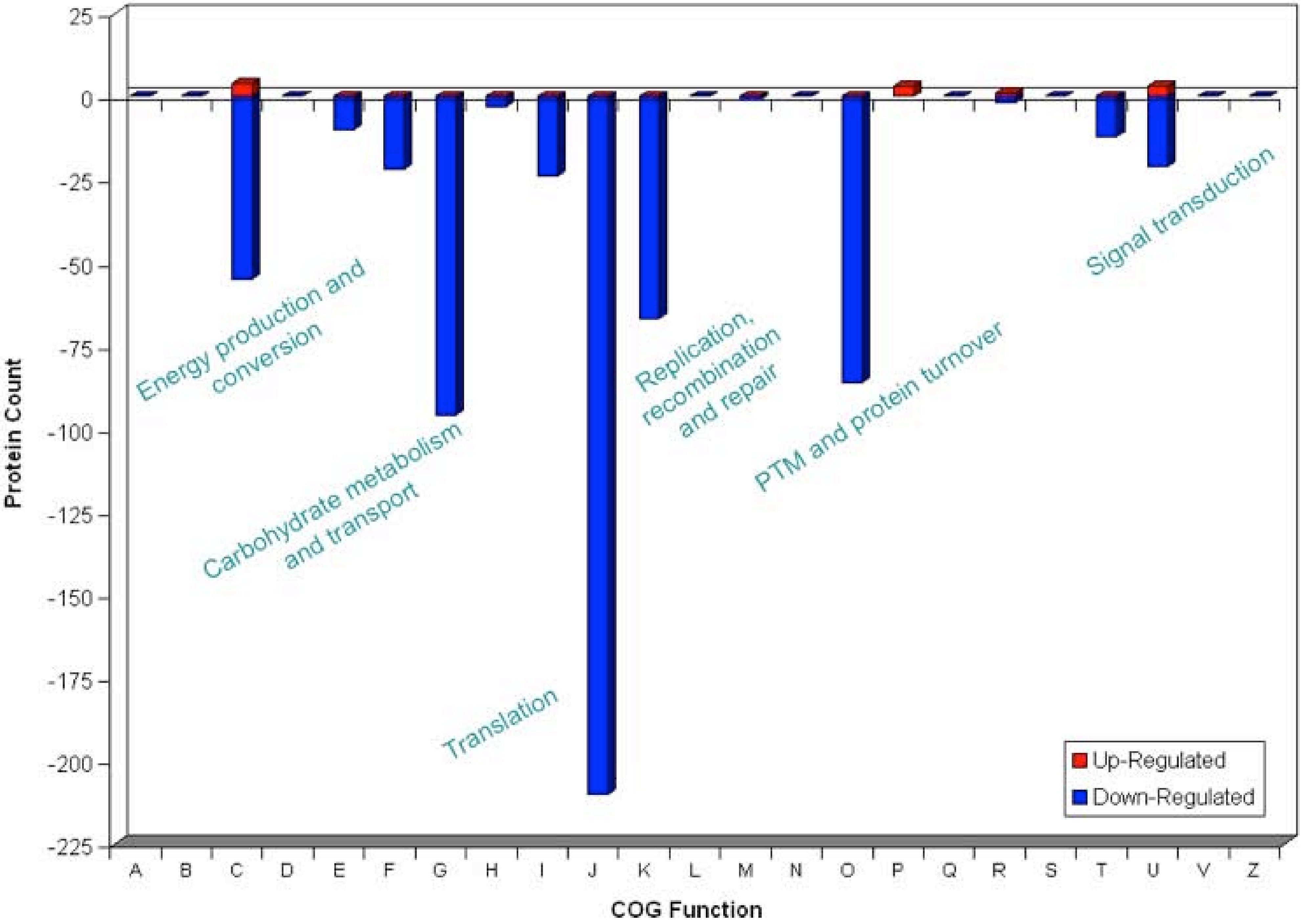


Cluster of Proteins Highly Expressed in CD:

Set of “core” proteins in ileal CD:

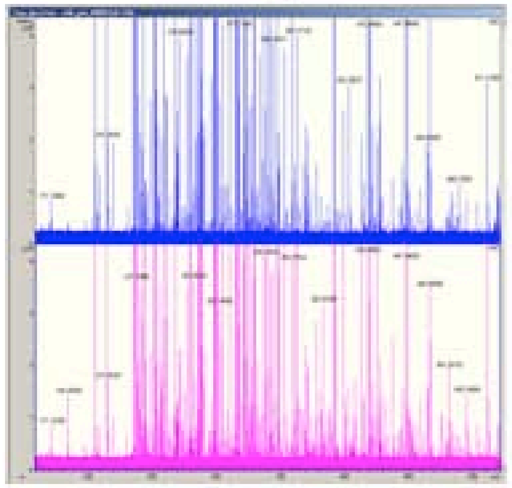
Cluster of Proteins Highly Expressed in Healthy

COG classification of differentially expressed proteins in CD relative to healthy

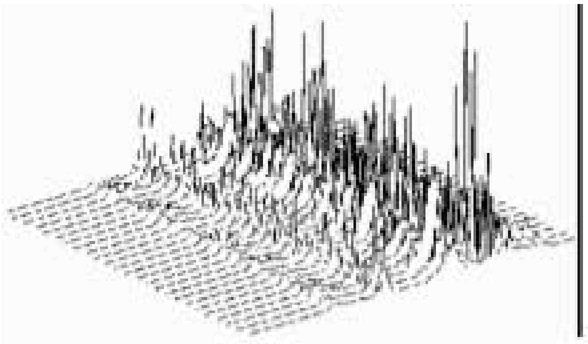


Metametabolomics - fecal water

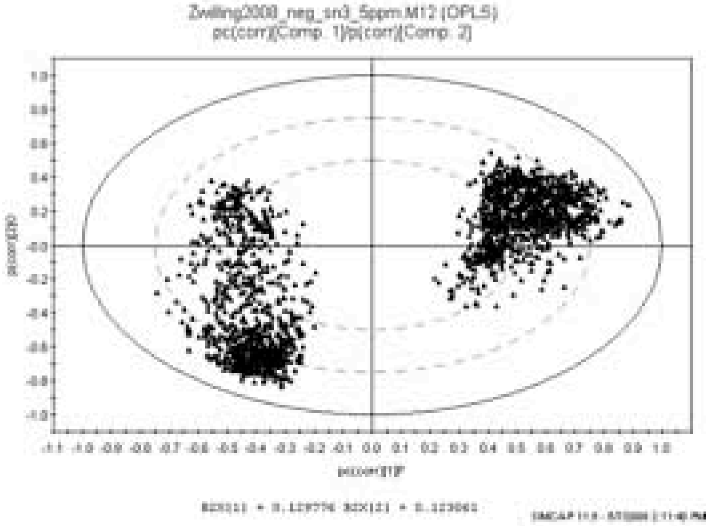
Data quality
& Allignment

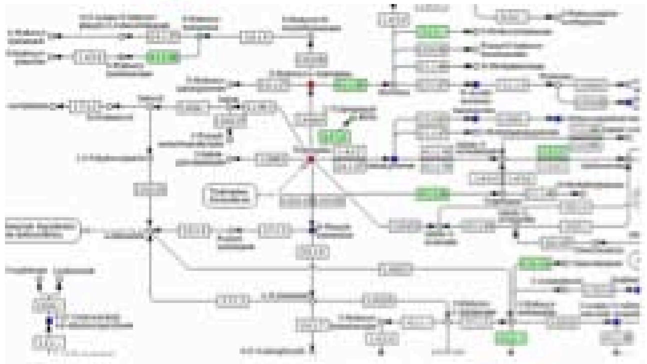


Data pre-processing



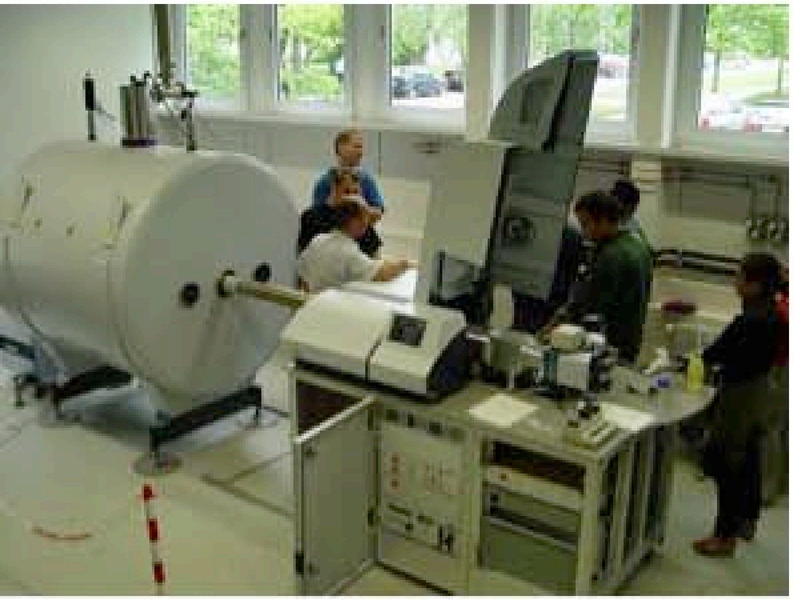
Multivariate analysis





Lists of discriminant m/z
Annotation of the metabolites in MassTRIX

Lists of important
Pathways & biomarkers

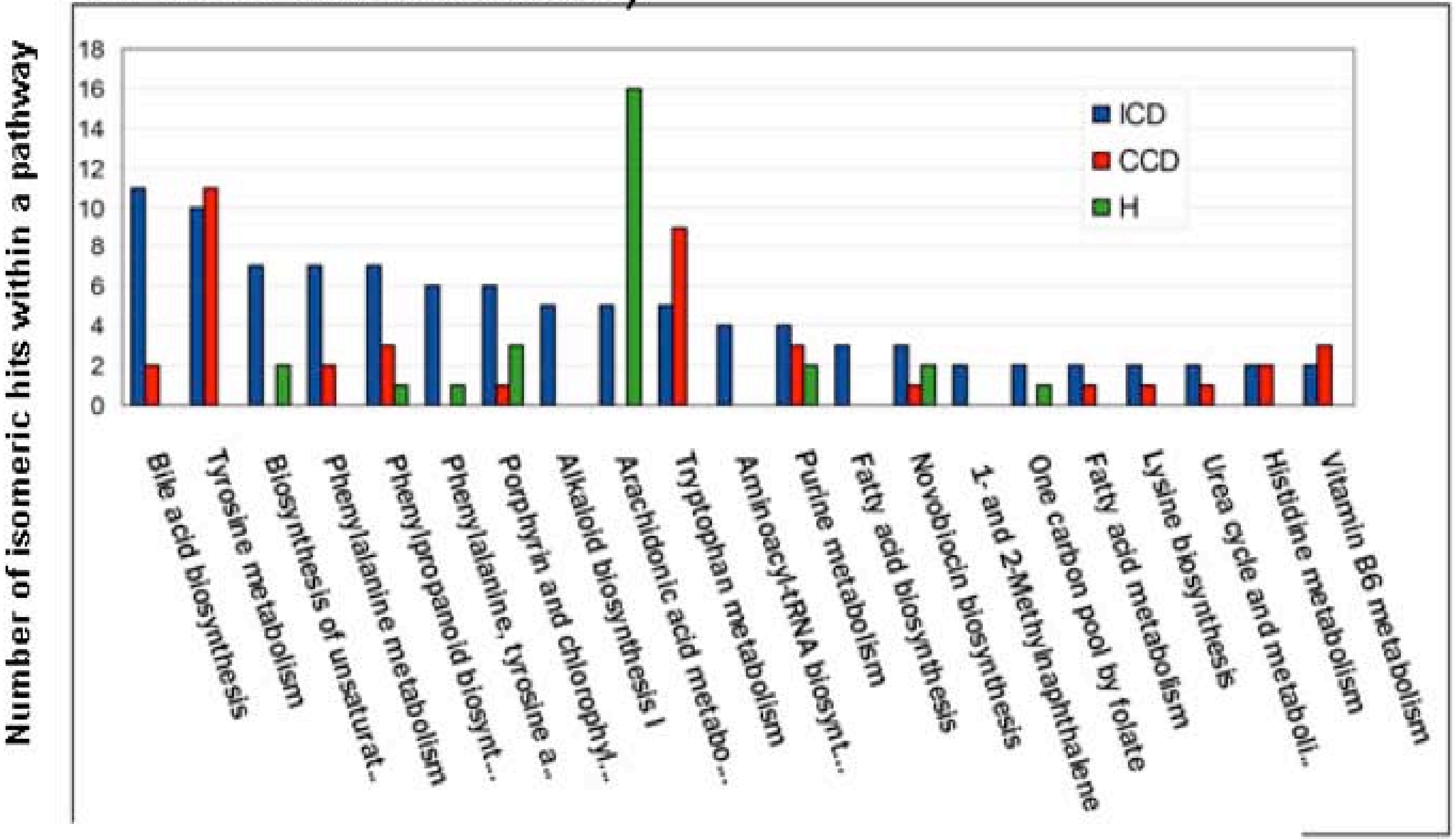


ESI-ICR-FT/MS

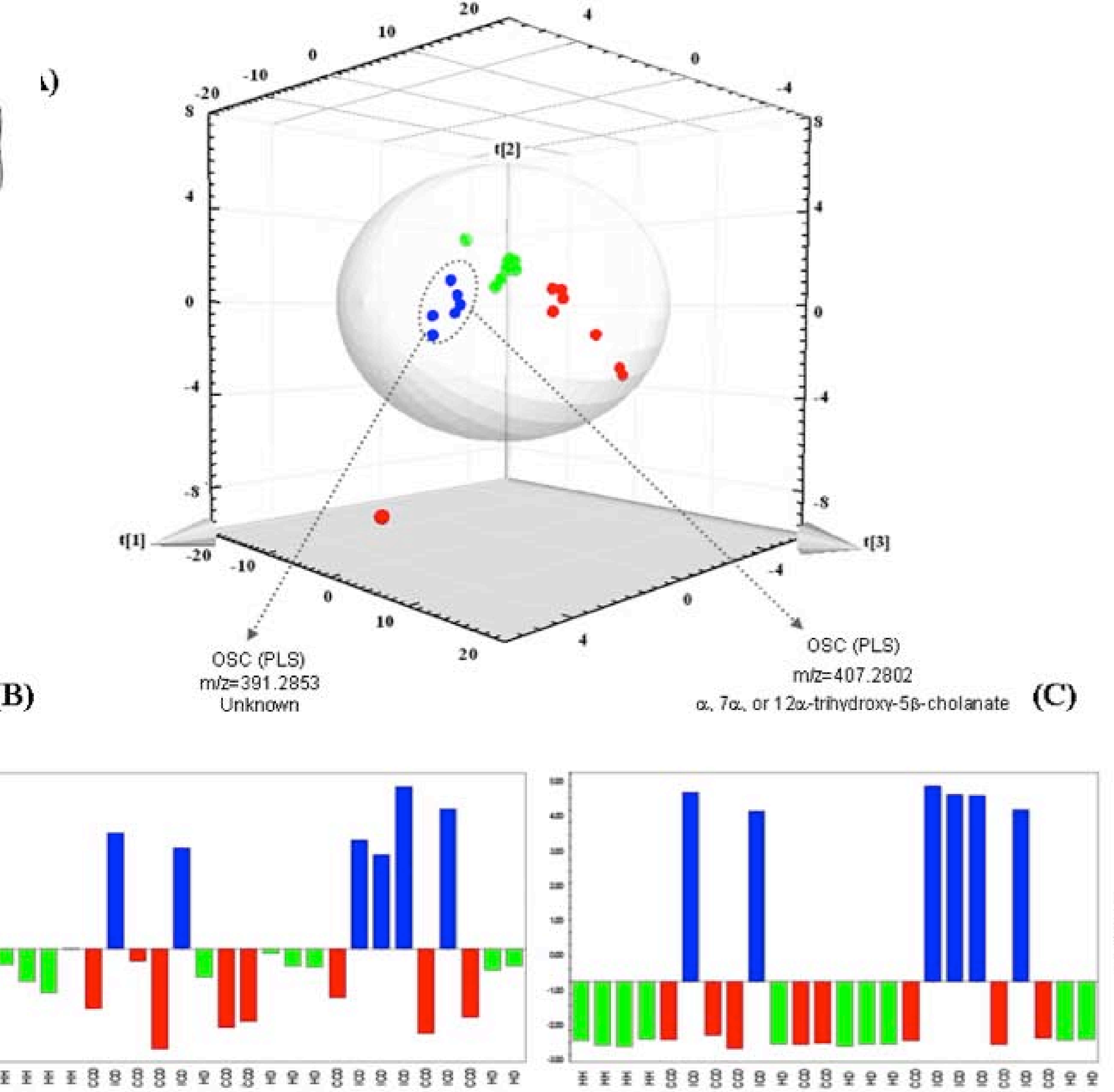
Philippe Schmitt-Kopplin

Total masses: 18706

7919 were discriminating: 2155 for ICD, 3113 for CCD and 2650 for healthy (most were unidentified in current databases!)



Jansson et al. (2009) PLoS One

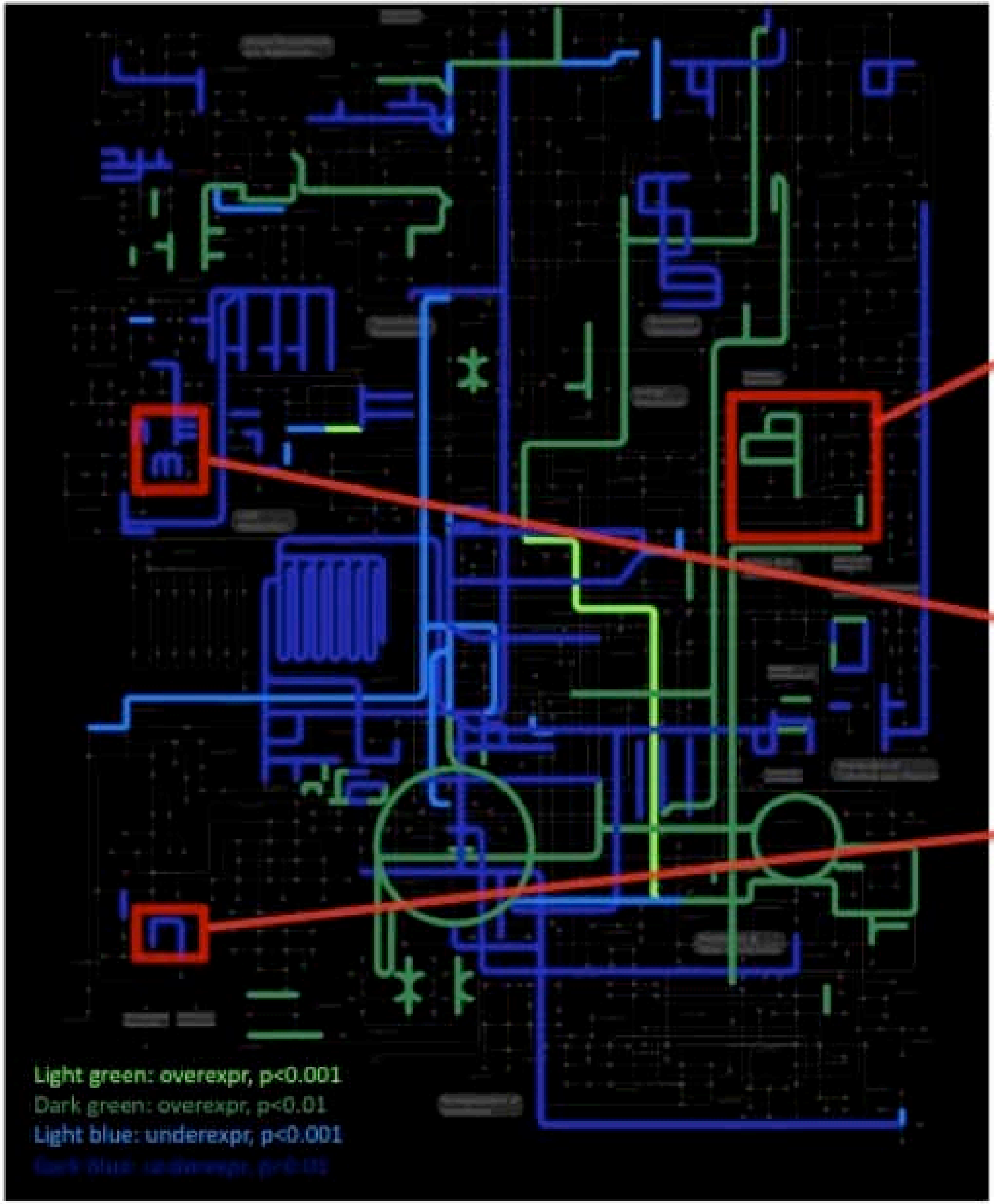


Jansson et al. (2009) PLoS One

Current challenge- Correlation of multi-omics data sets

- 454 pyrotag data: 250,000 16S rRNA gene sequences
- Metaproteome data: >4000 proteins
- Metabolite data: 18,000 metabolites
- Metagenome data: >43 Mb sequence

Pathway analysis metaproteomics data explains accumulated metabolite markers



Tyrosine & tryptophan biosynthesis

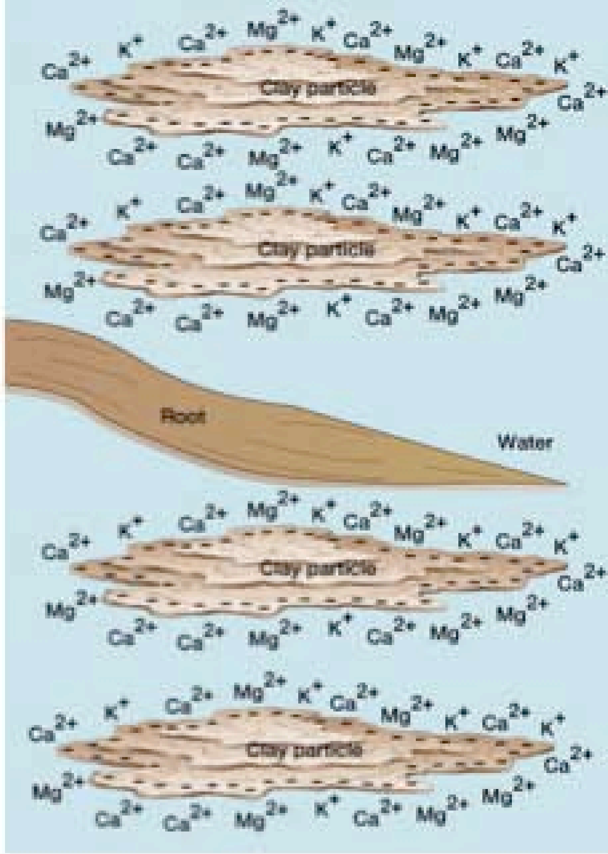
Linoleic & arachidonic acid metabolism

Chenodeoxycholate (bile acid) metabolism

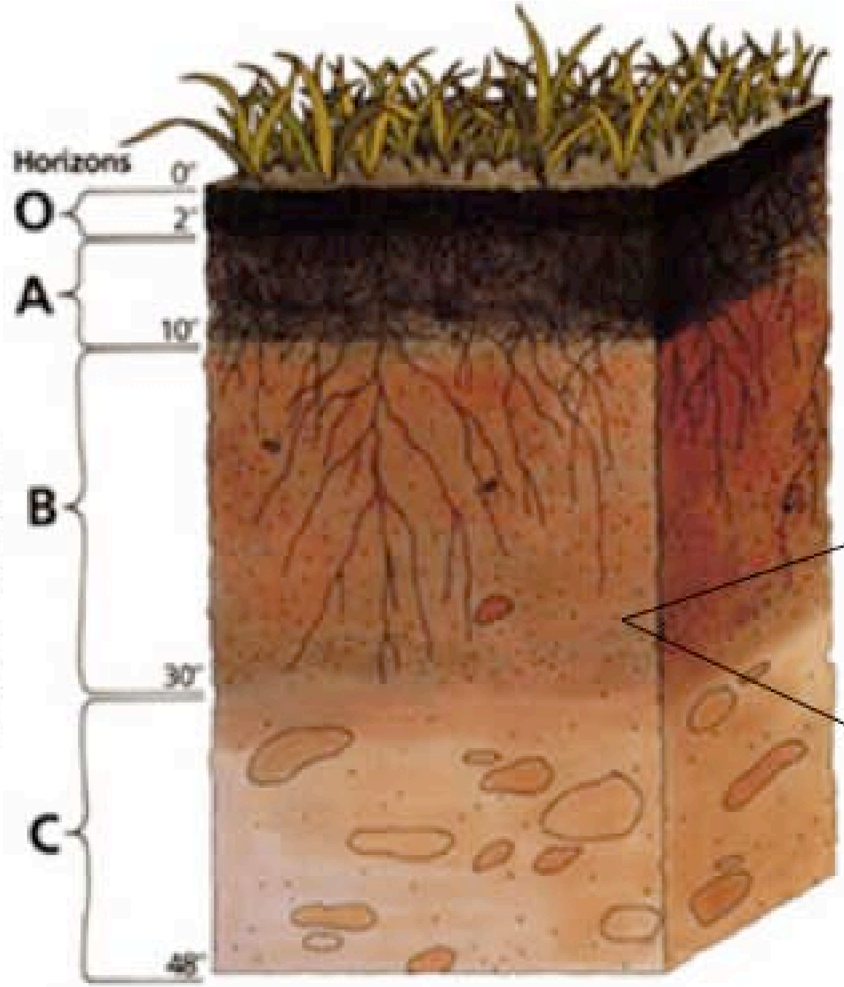
Red: metabolites overrepr in CCD/ICD fecal samples (Jansson et al. PLoS One 2009)

Youssef Darz & Jeroen Raes, VIB

Grand challenge: Soil Metaproteomics



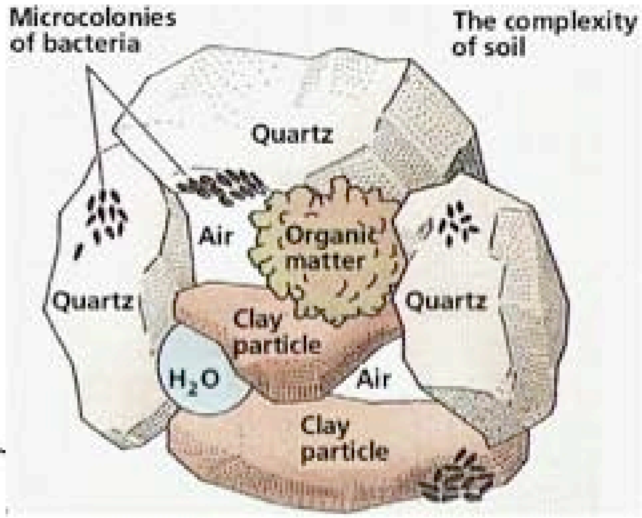
Complexity of soil



cm to m

← cm to km →

Particle size distribution
And aggregate formation
Results in complex 3D structure

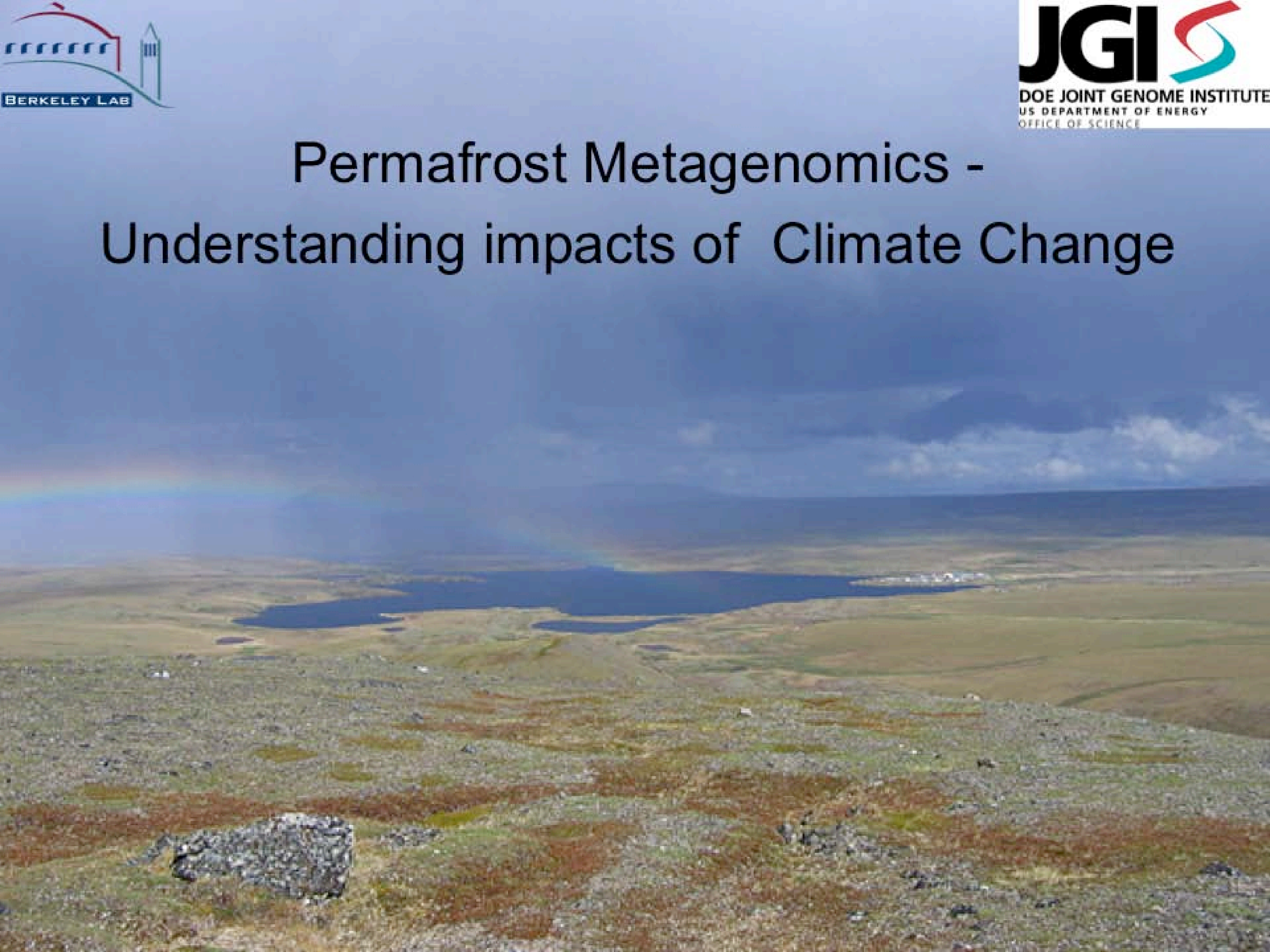


← µm to mm →

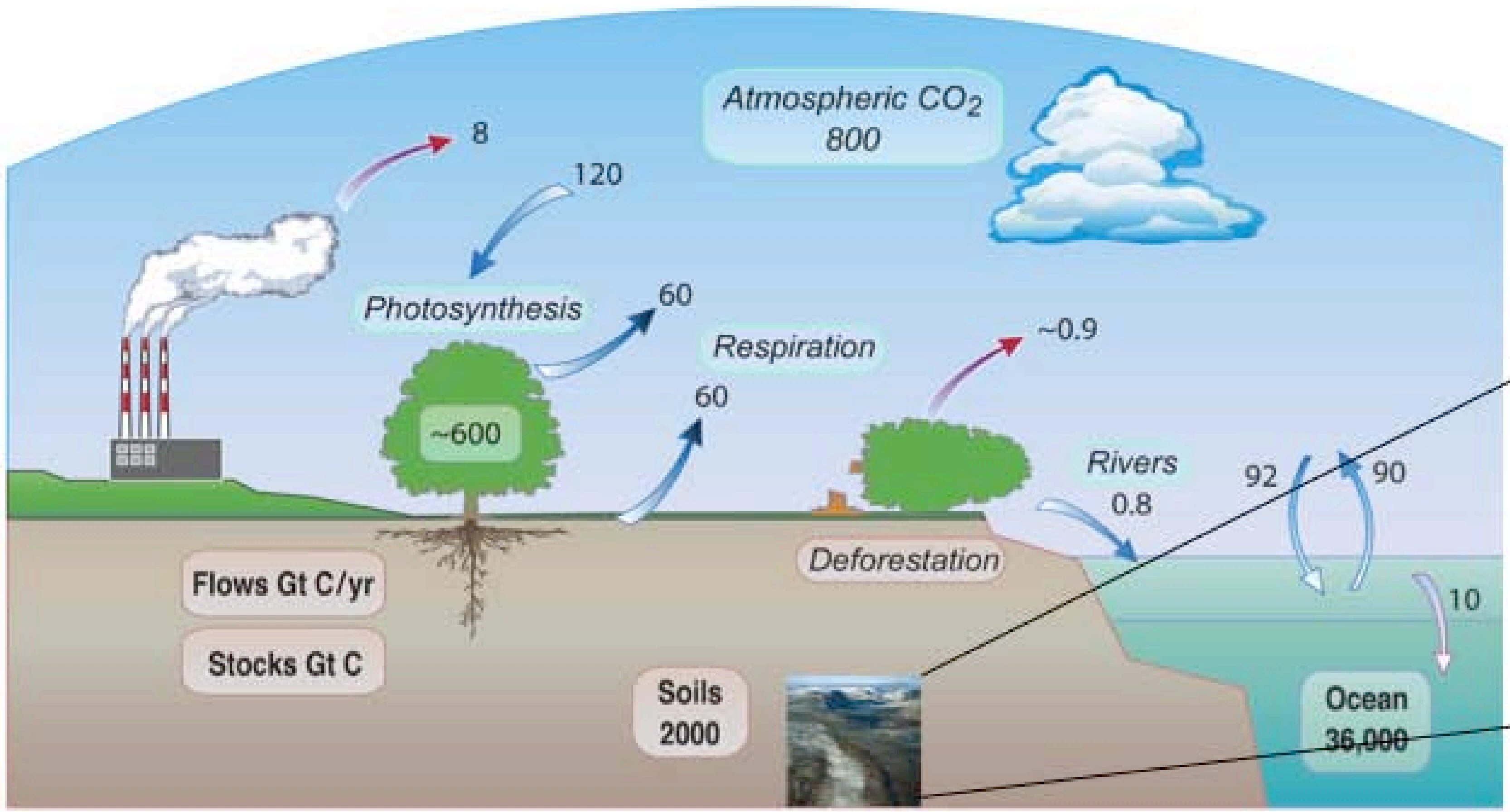
Current bottlenecks in soil metaproteomics

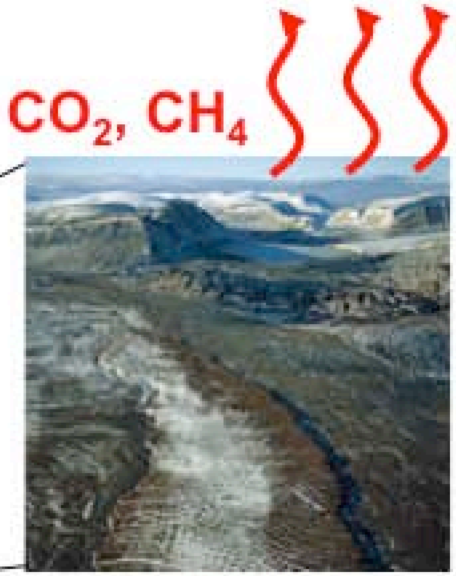
- Lack of good extraction methods
  - Direct & Indirect
- Low yield
  - Only dominant community members represented
- Lack of matching metagenome databases to search for protein IDs

Permafrost Metagenomics - Understanding impacts of Climate Change



Permafrost soils store vast amounts of carbon



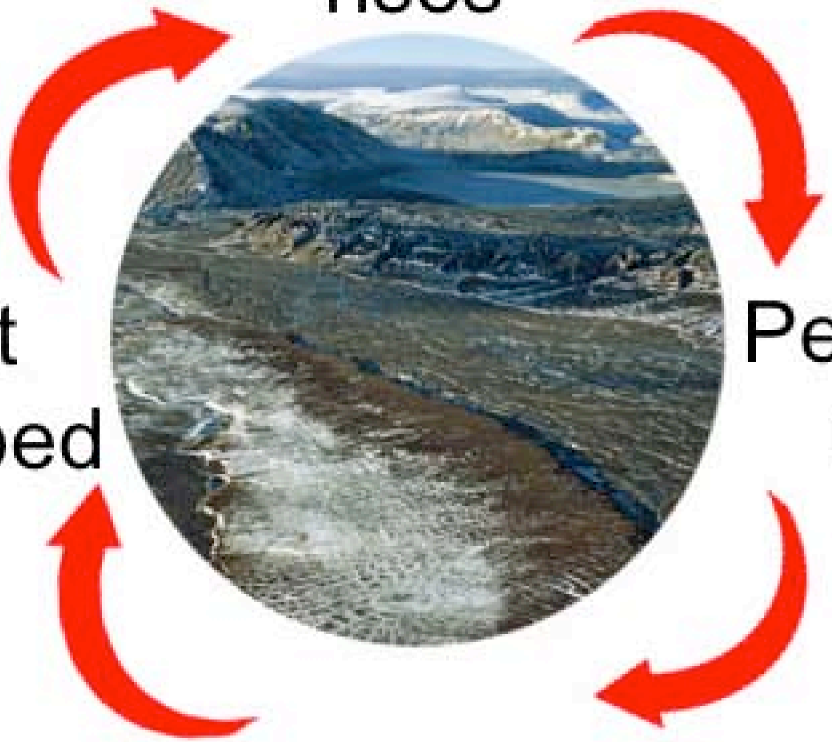


Permafrost: 960 Gt C.

Thaw-induced microbial decomposition releases CO_2 and CH_4

Permafrost Thaw and the Global Carbon Cycle

Positive Feedback Loop



Temperature rises

Permafrost thaws

Methane and CO_2 released

Heat absorbed

Why Study the Permafrost Metagenome?

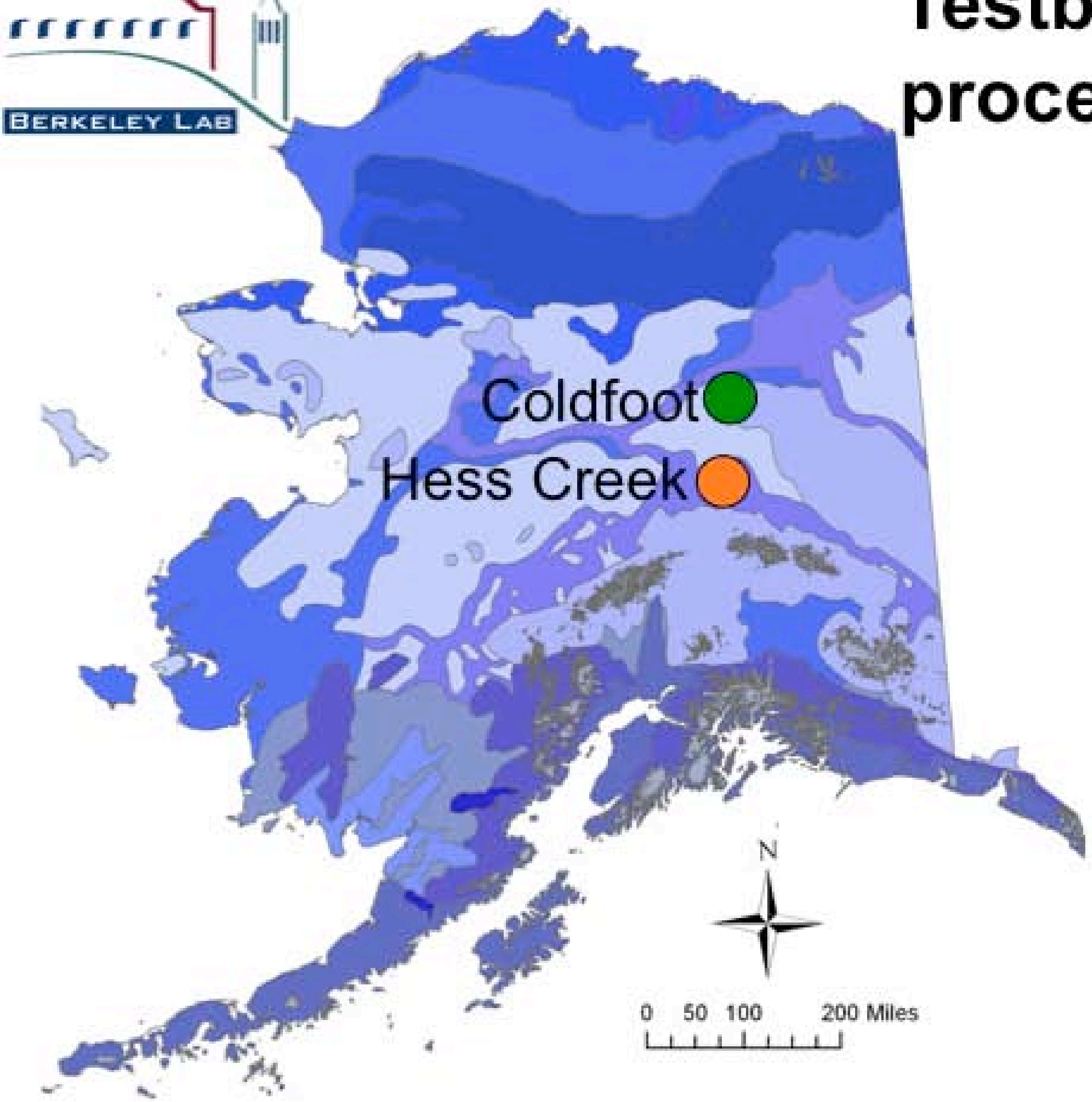
Understand

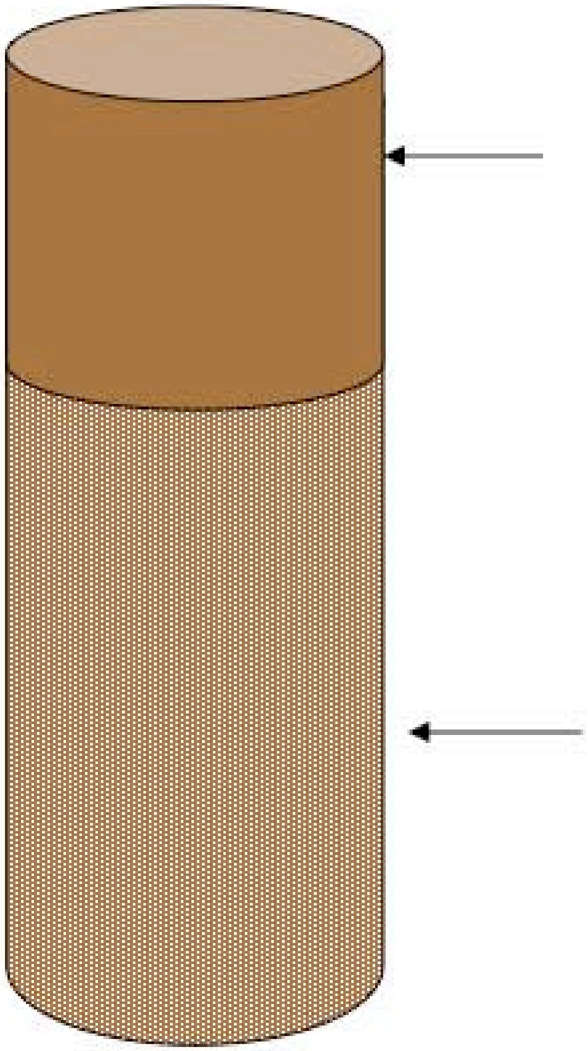
Predict

Mitigate

BERKELEY LAB

Testbed for understanding microbial processes during permafrost thaw



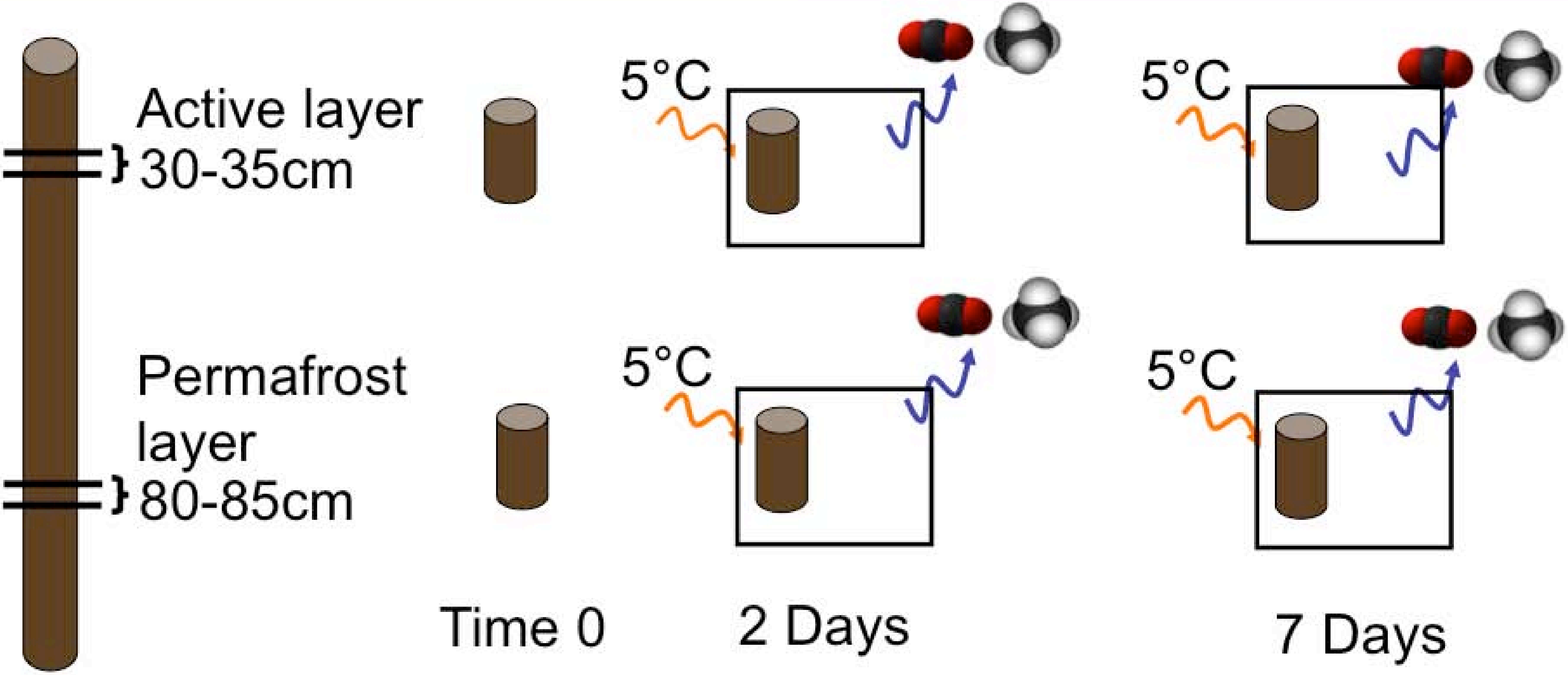


Active layer: Seasonally freezen and thawed

Permafrost: Permanently frozen rock or soil material

3 Cores per site:
Divide into active layer and permafrost layer

Experimental design



Active layer
30-35cm

Permafrost layer
80-85cm

5°C

5°C

5°C

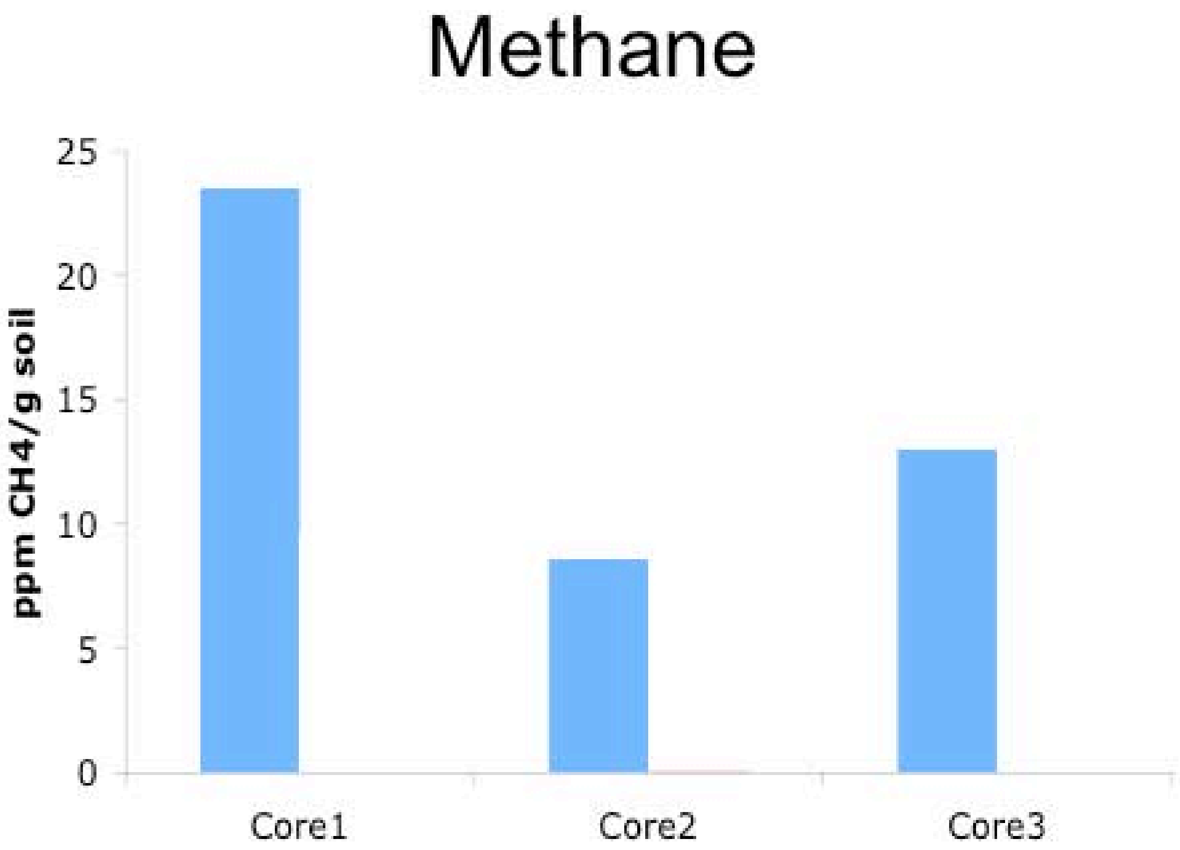
5°C

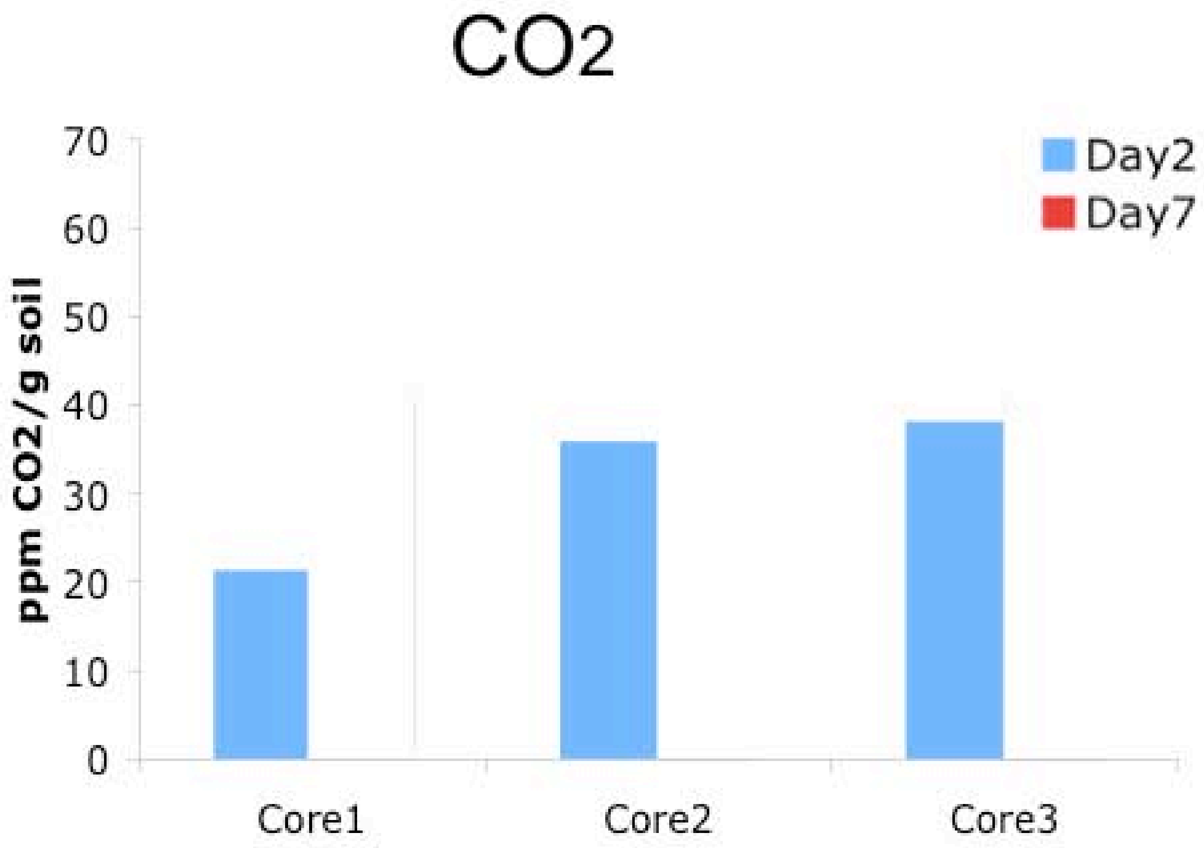
Time 0

2 Days

7 Days

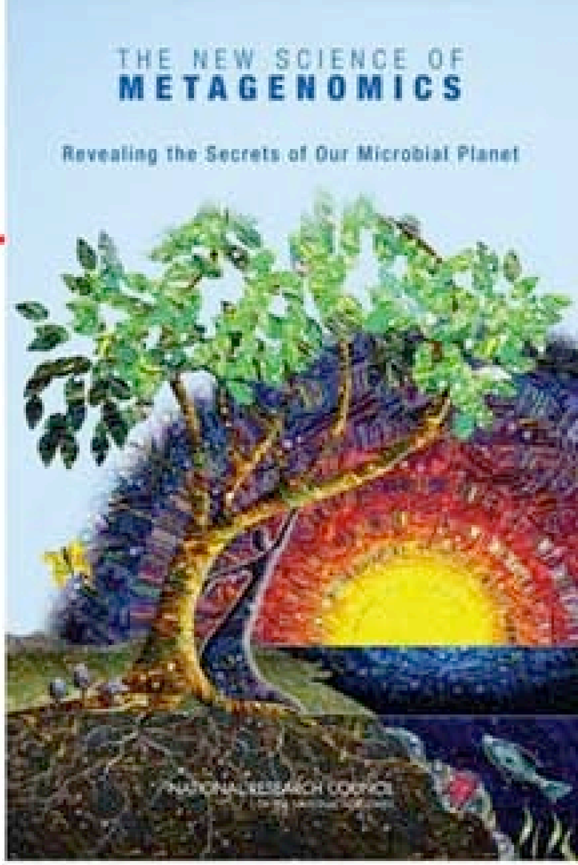
Methane and CO2 measurements


Methane


CO2

BERKELEY LAB

Metagenomics


THE NEW SCIENCE OF
METAGENOMICS
Revealing the Secrets of Our Microbial Planet
NATIONAL RESEARCH COUNCIL

JGI
DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

In collaboration with Eddy Rubin & Rachel Mackelprang at JGI

& Mark Waldrop at the USGS

→ Extract all DNA from permafrost samples

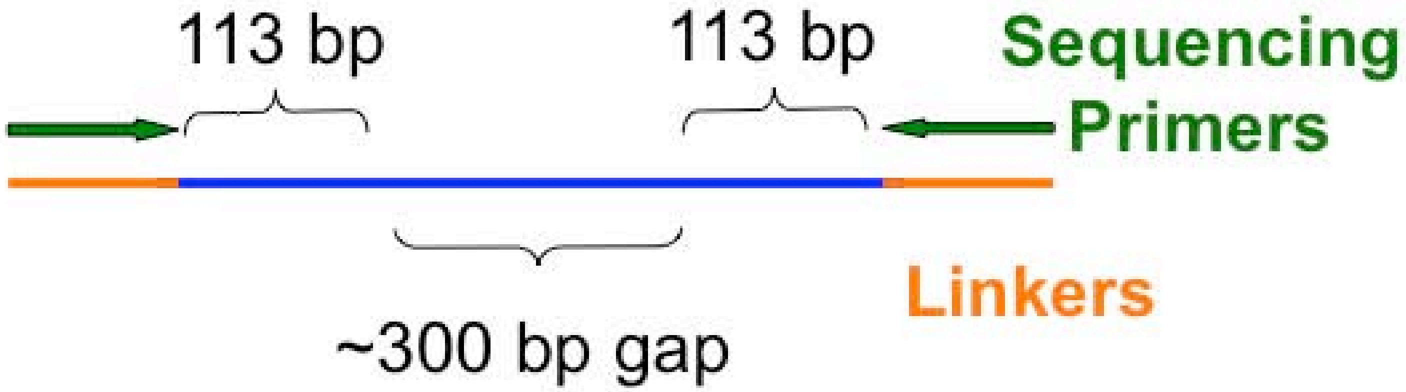
(before and after thawing)

→ Sequence community DNA

(both phylogenetic and functional genes)

Provides information on microbial community composition and function, including representatives that have never been cultivated (>90% have never been isolated!)

Library Characteristics



113 bp

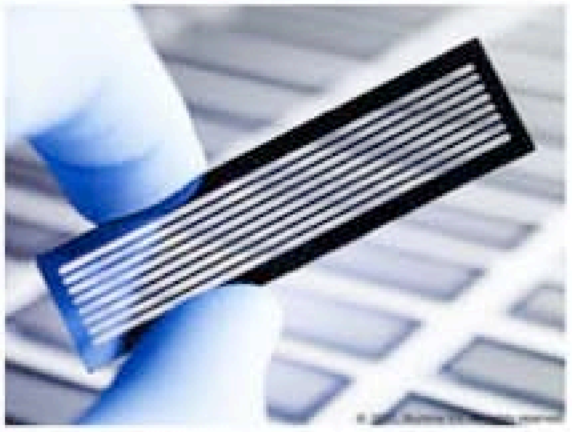
113 bp

Sequencing Primers

Linkers

~300 bp gap

Flow Cell (8 Lanes)



AL0	AL2	AL7	PL0	PL2	PL7
12,936,330	12,209,016	11,860,629	16,629,647	17,164,291	16,334,711

<u>12 Day Paired End Run</u>

Each read has 226 bp

226bp x 87M reads = 19.6 Gbp

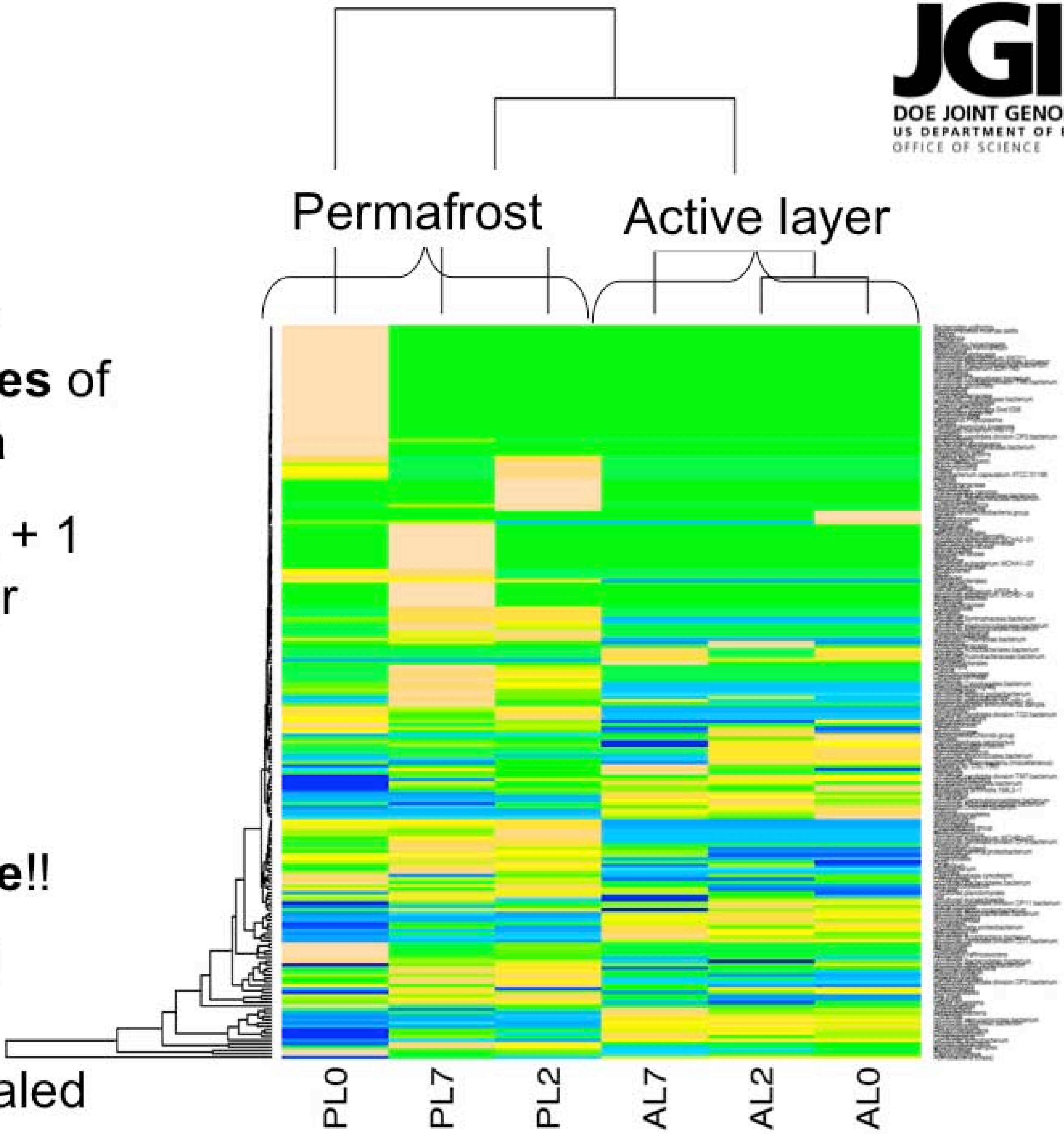
6 samples:

Approximately **20 billion bases** of sequence data

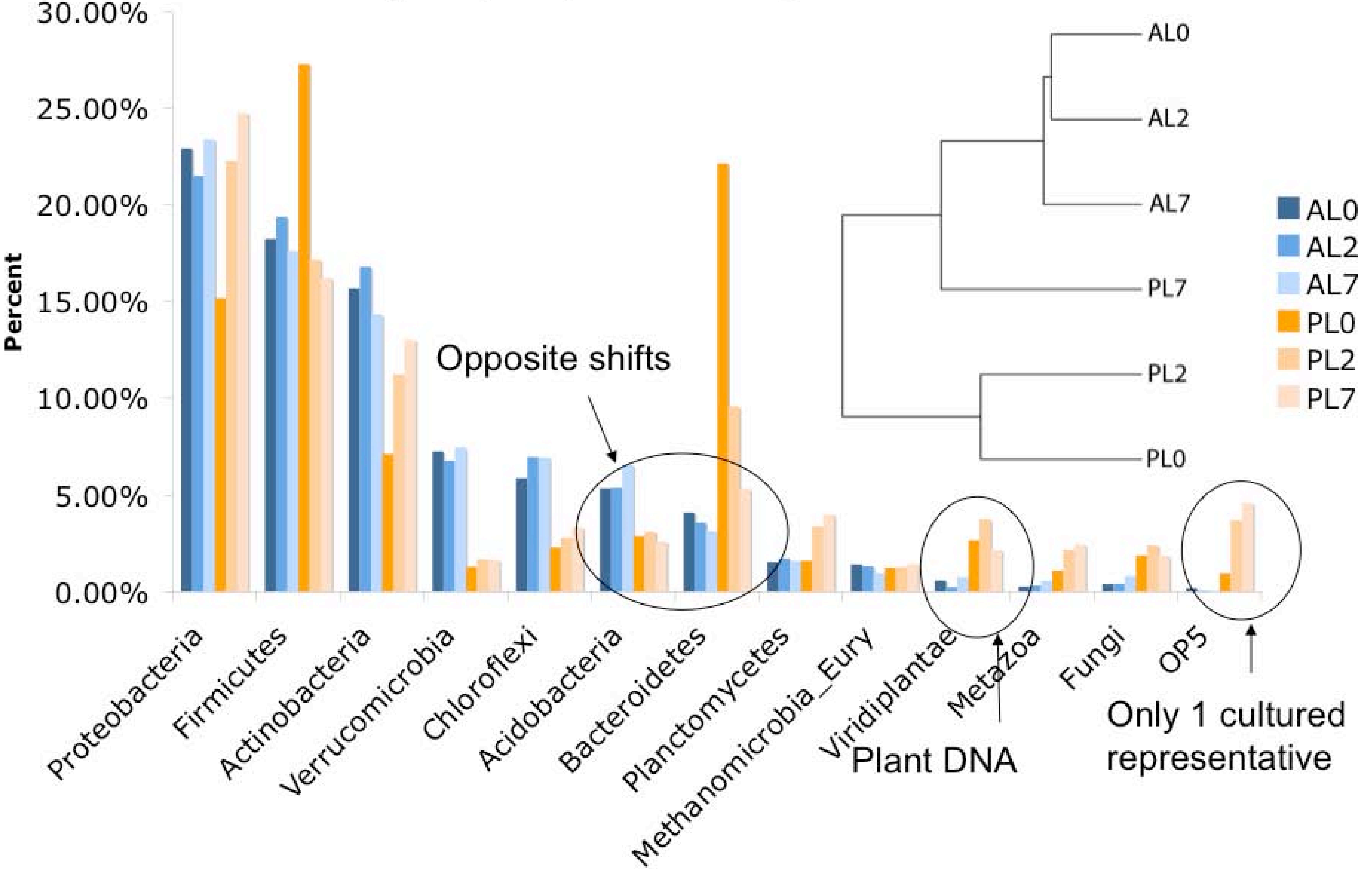
12 days to run + 1 week computer time

Largest sequencing project to date!!

Both microbial identities and functions revealed



Community Composition: (Phyla) Greengenes

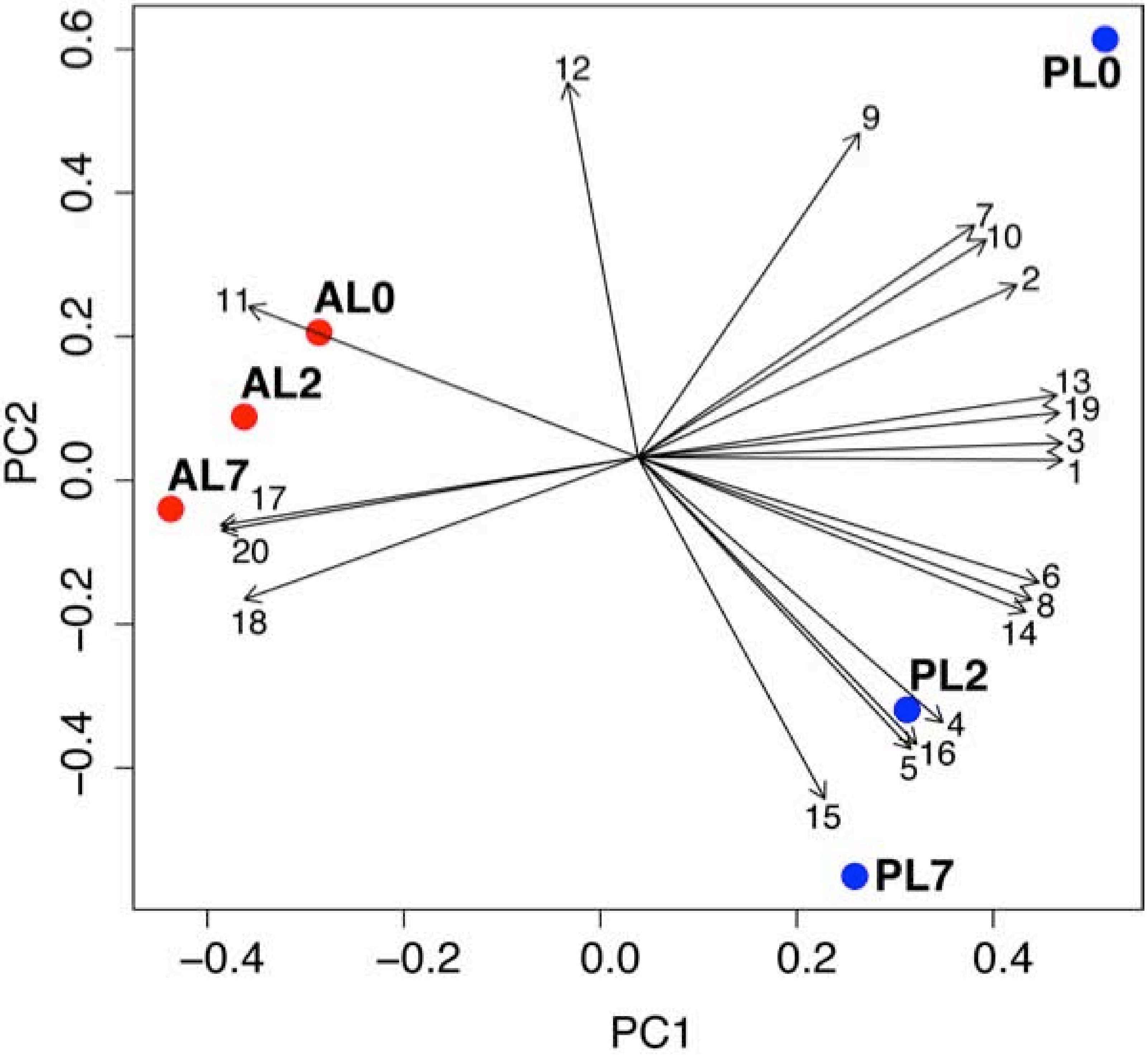


Opposite shifts

Plant DNA

Only 1 cultured representative

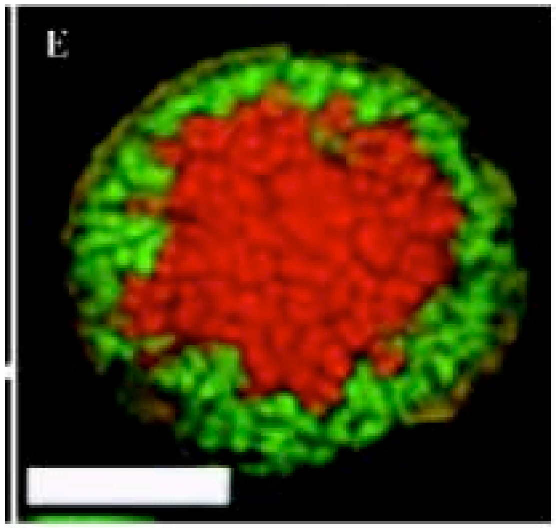
COGs

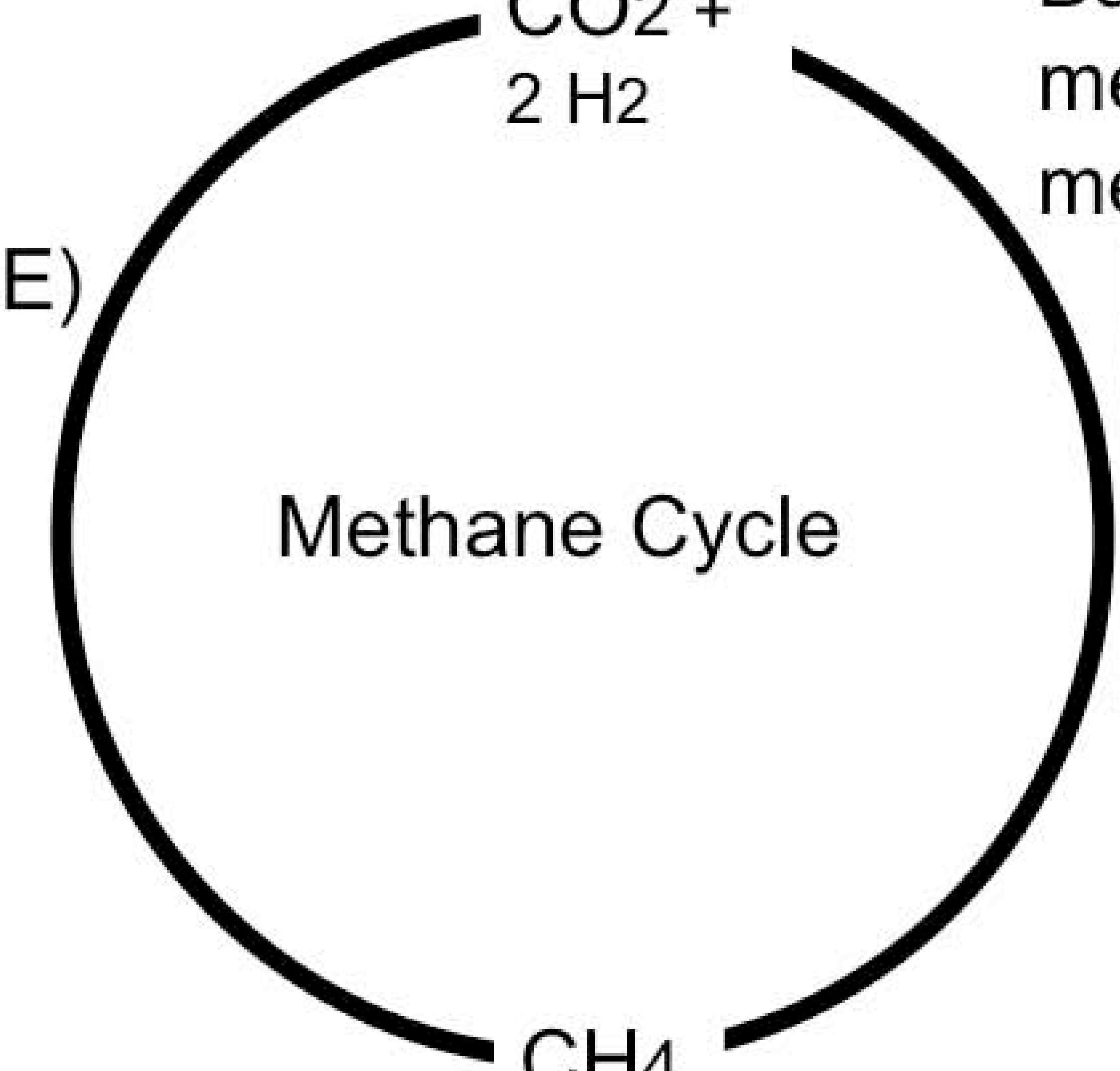


Rank	Group
1	Chemotaxis protein; stimulates methylation of MCP proteins
2	Fe2+ transport system protein B
3	Transcriptional accessory protein
4	Predicted sulfurtransferase
5	Polyketide synthase modules and related proteins
6	Uncharacterized membrane protein
7	Aspartate ammonia-lyase
8	Sigma54-dependent transcription regulator containing regulator containing on AAA-type ATPase domain and a DNA-binding domain
9	Tryptophan synthase beta chain
10	Helicase subunit of the DNA excision repair complex
11	NADH:ubiquinone oxireductase 49kD subunit 7
12	ABC-sugar transport system, permease component
13	Tetrahydromethanopterin S-methyltransferase, subunit E
14	Uncharacterized conserved protein
15	Uncharacterized membrane protein (homolog of Drosophila rhomboid)
16	Cytochrome c2
17	Acyl-CoA dehydrogenases
18	Gamma-glutamyltransferase
19	Phenylalanyl-tRNA synthetase alpha subunit
20	NADPH:quinone reductase and related Zn-dependent oxidoreductases

Methanotrophs: $CH_4 + O_2 \dashrightarrow CO_2 + 2H_2$

Archaea (Anaerobic methane oxiders (ANME)



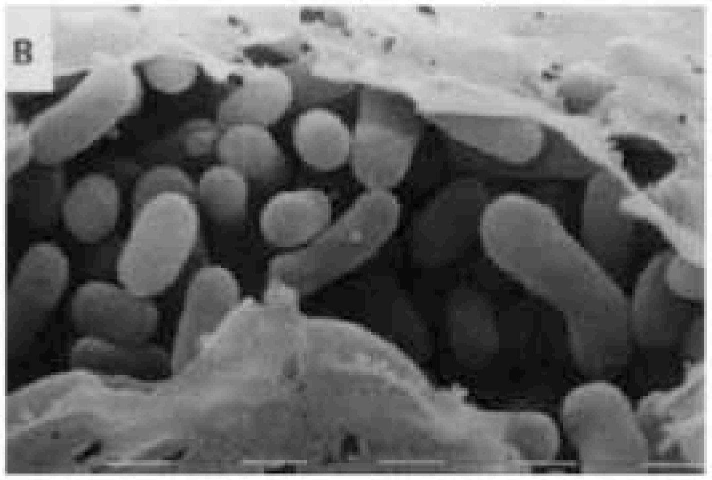


CO_2 + 2 H_2

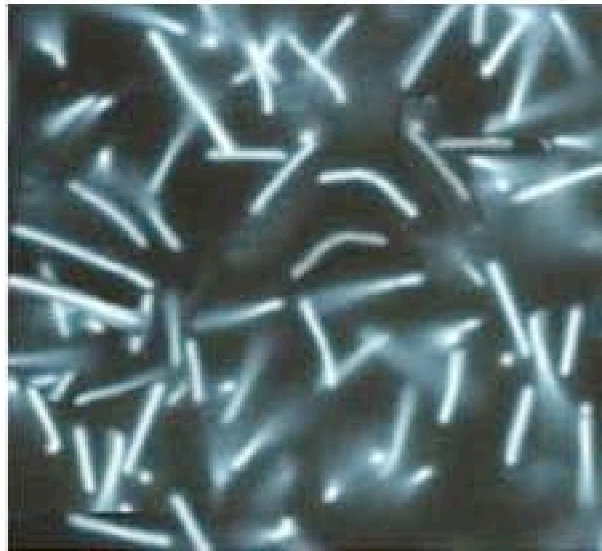
Methane Cycle

CH_4

Bacteria (aerobic methane oxiders, methanotrophs)

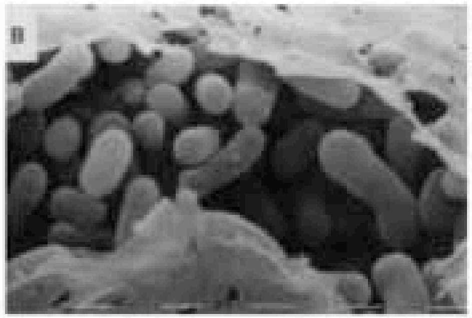


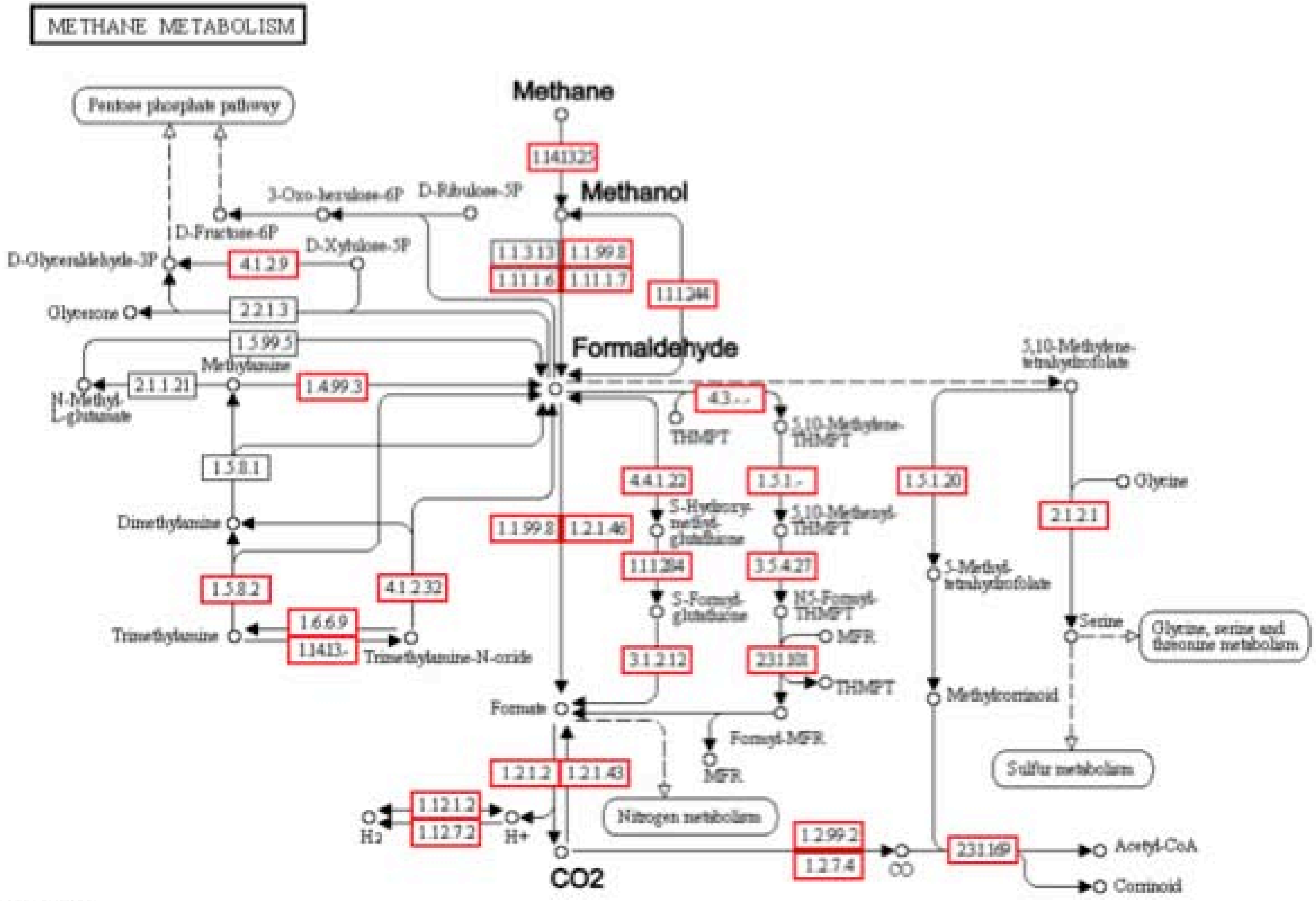
Methanogens: $CO_2 + 2H_2 \dashrightarrow CH_4 + O_2$



Methane Cycle Pathways

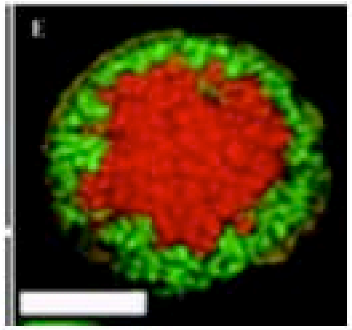
Aerobic/Bacteria

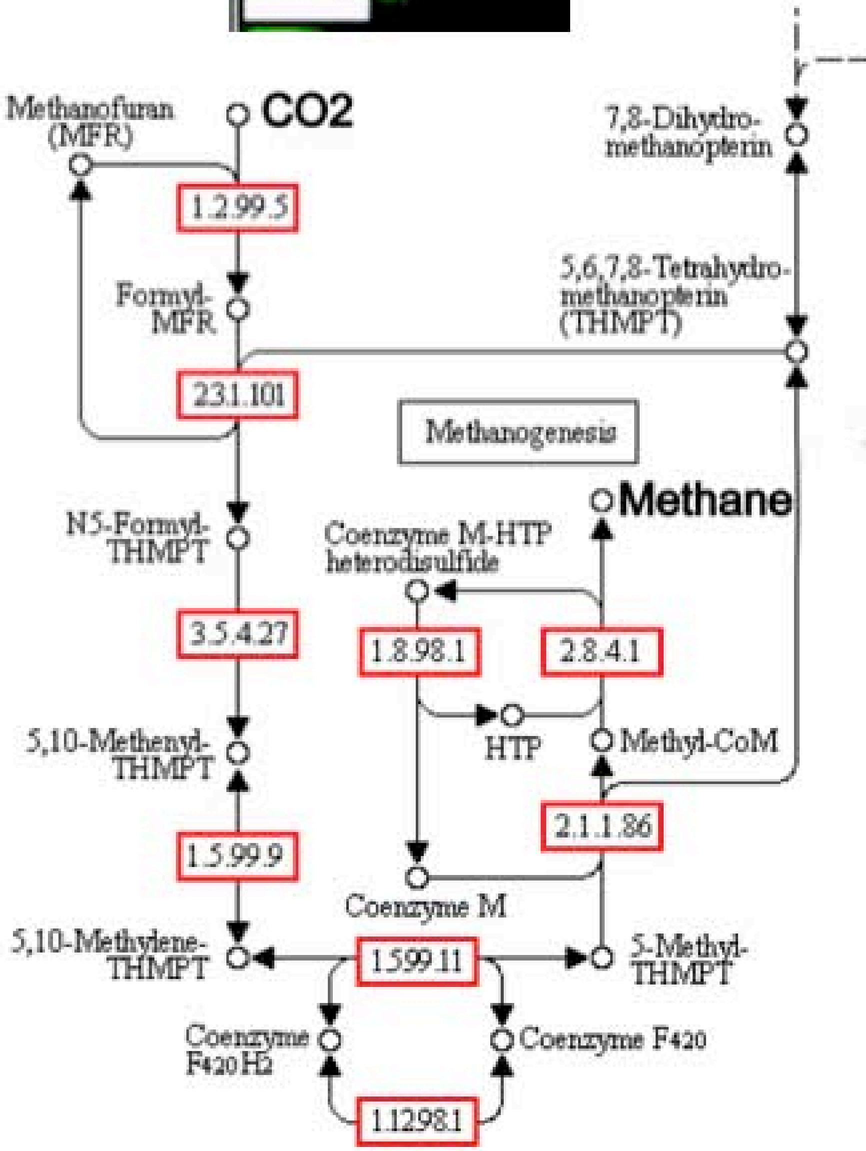




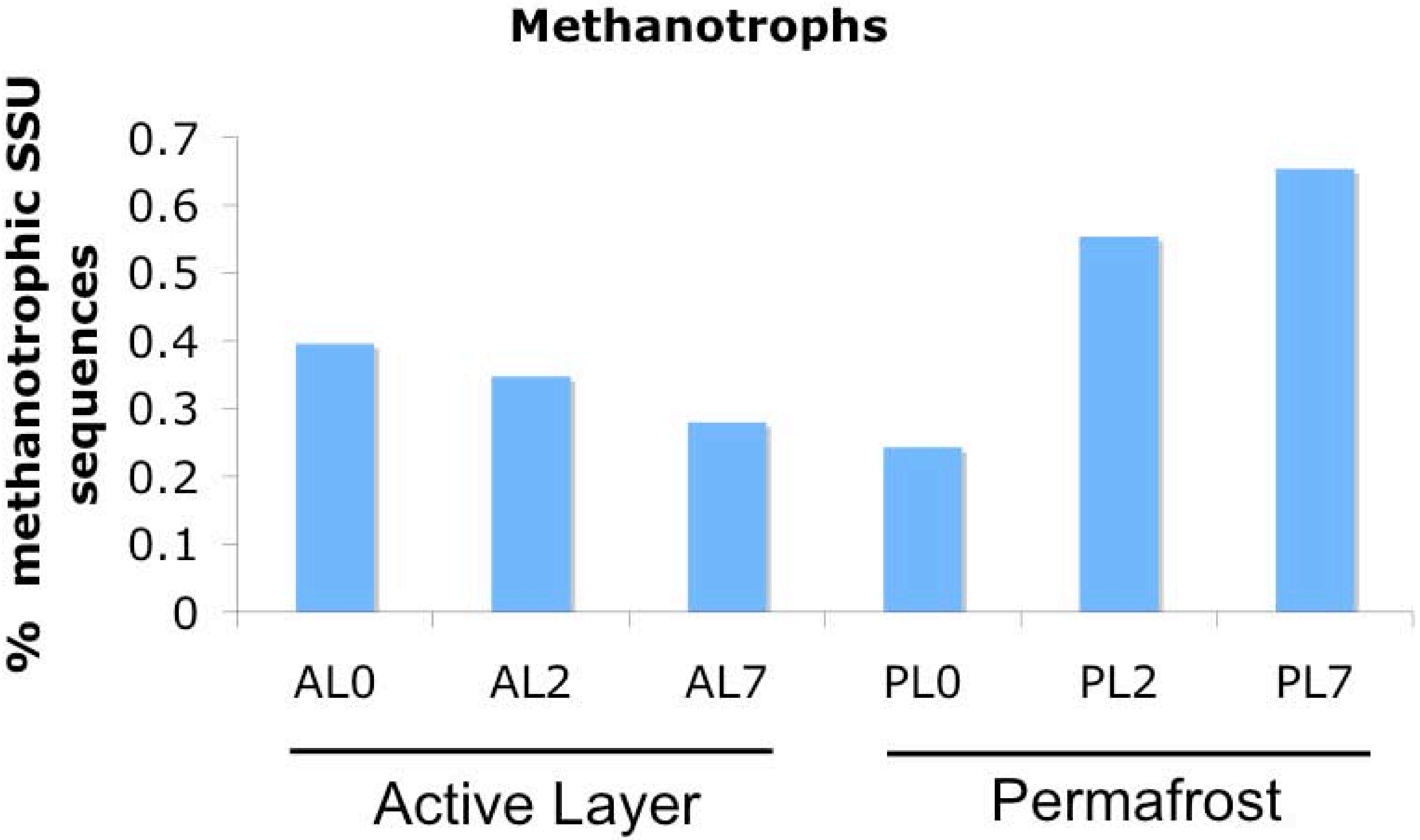
00680 9/1/09
(c) Kanehisa Laboratories

Anaerobic/Archaea

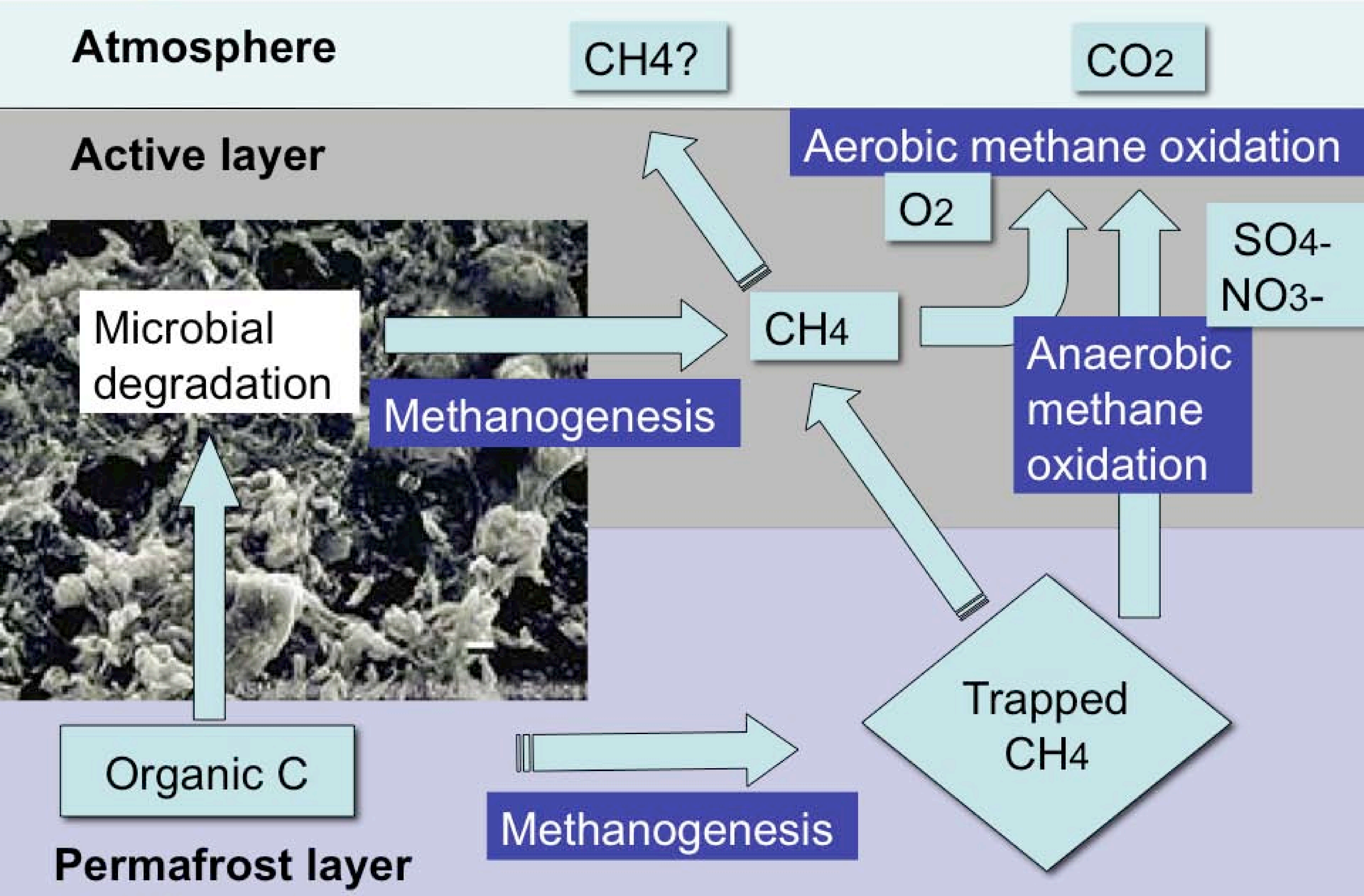




Methanotrophs (bacteria) Increase in Abundance in Permafrost



Methane cycle process in permafrost soil??



Atmosphere

CH_4?

CO_2

Active layer

Aerobic methane oxidation

O_2

SO_4^-
NO_3^-

Microbial degradation

CH_4

Methanogenesis

Anaerobic methane oxidation

Trapped CH_4

Organic C

Methanogenesis

Permafrost layer

BERKELEY LAB

JGI
DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

Pilot project for JGI's Soil Metagenome Initiative

The GREAT PRAIRIE

Coordinated by James Tiedje, Janet Jansson

Plus site collaborators and scientists at JGI

Rationale

- Midwest prairie represents largest expanse of world’s most fertile soils
- Reference site for understanding biological basis and ecosystem services of its microbial community
- Parallel’s the ocean gyre as important ecosystem for primary productivity and biogeochemical cycling
- Sequesters the most carbon of any soil system in the U.S.
- Produces large amounts of biomass annually: Key for biofuels, food security, carbon sequestration
- Central to DOE Mission

Pilot study: The Great Prairie

Sampling sites

- Wisconsin
  - Native prairie
  - Continuous corn
  - Established Switchgrass area
  - Restored prairie (from 1998)
- Iowa
  - Native prairie (Morris prairie)
  - Long term cultivated corn
- Kansas (Konza prairie)
  - Native prairie (Konza prairie)
  - Long term cultivated wheat

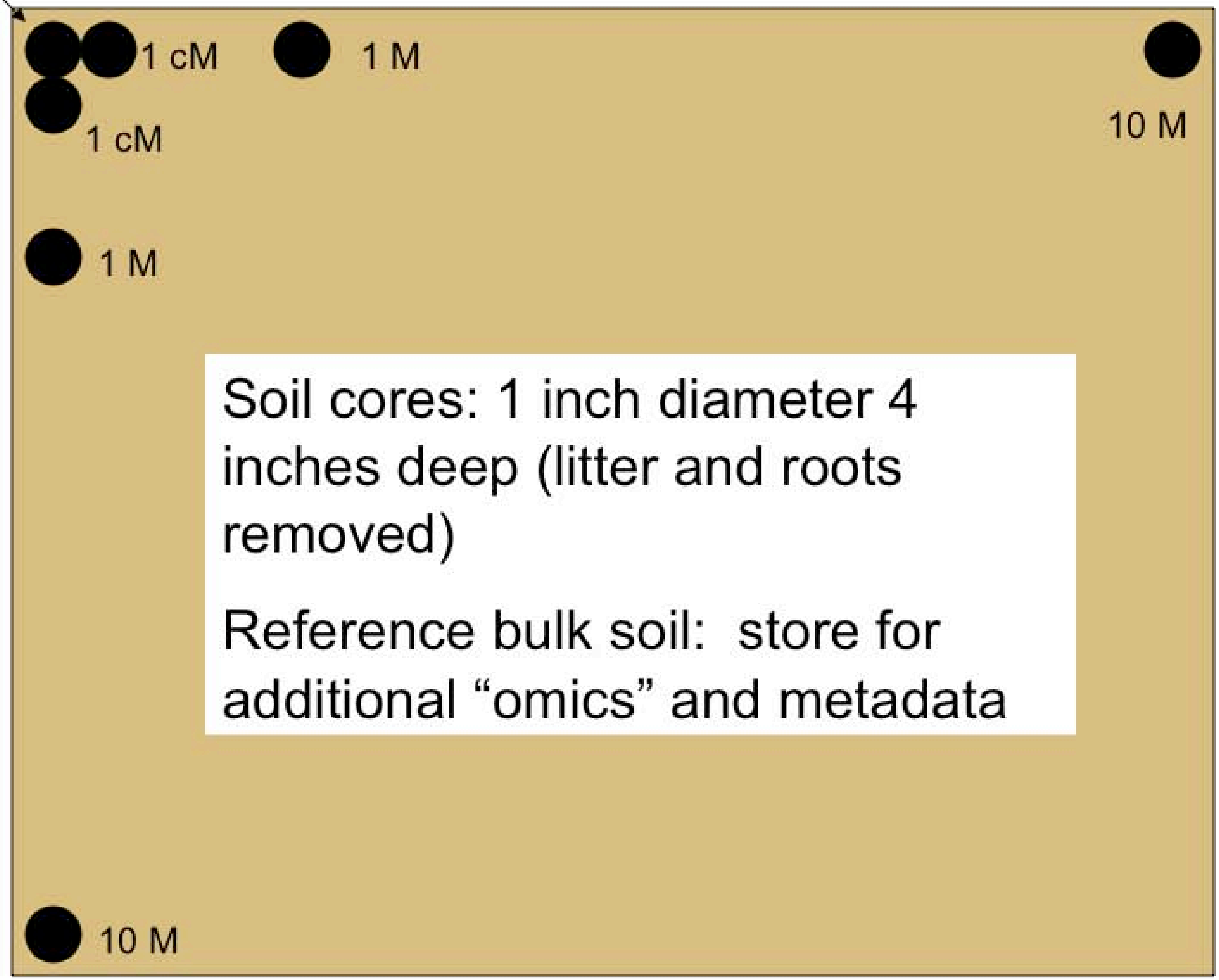

Wisconsin Switchgrass


Wisconsin Restored Prairie


Iowa >100 yrs Corn

Sampling strategy

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

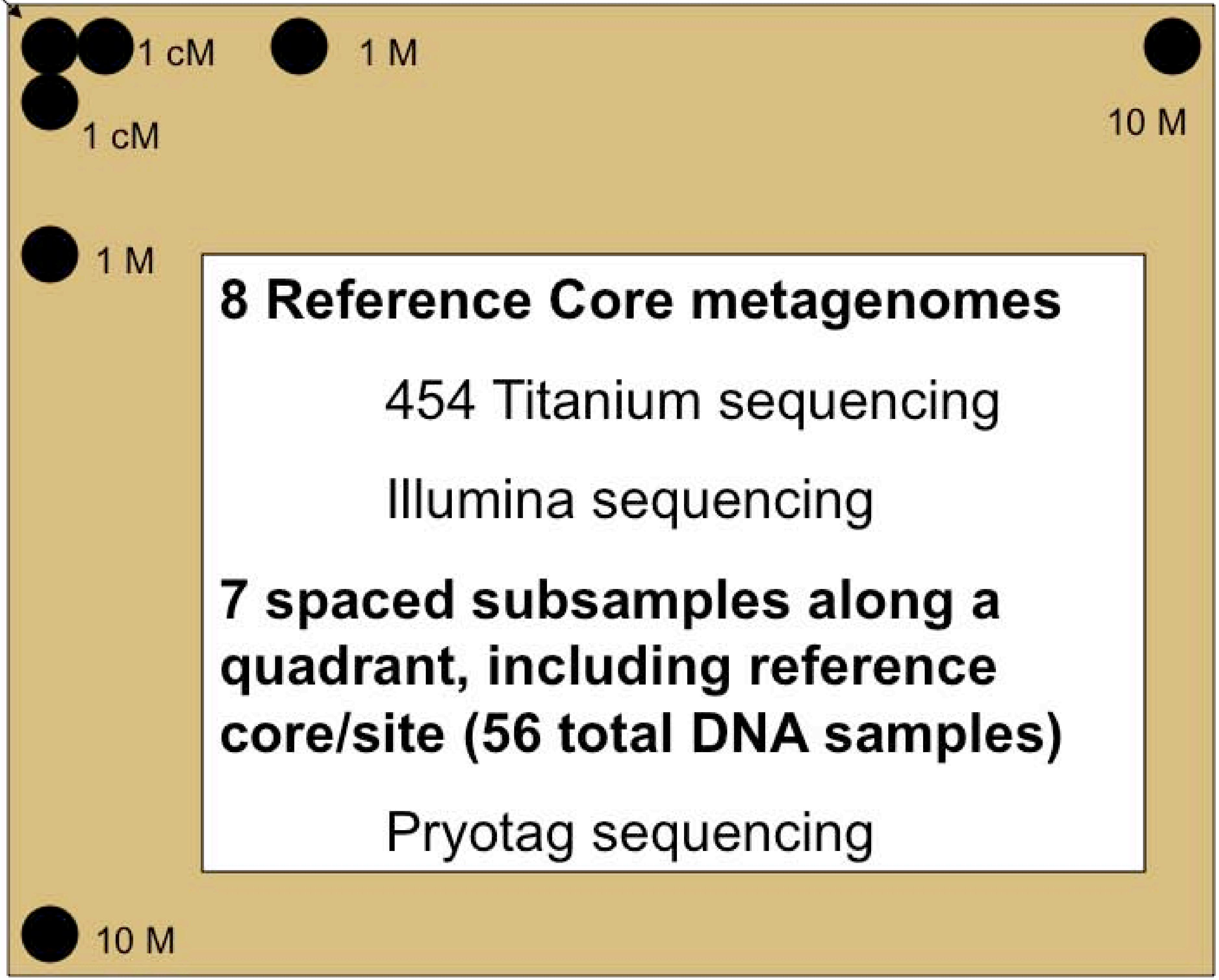


Soil cores: 1 inch diameter 4 inches deep (litter and roots removed)

Reference bulk soil: store for additional “omics” and metadata

Sequencing projects

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE



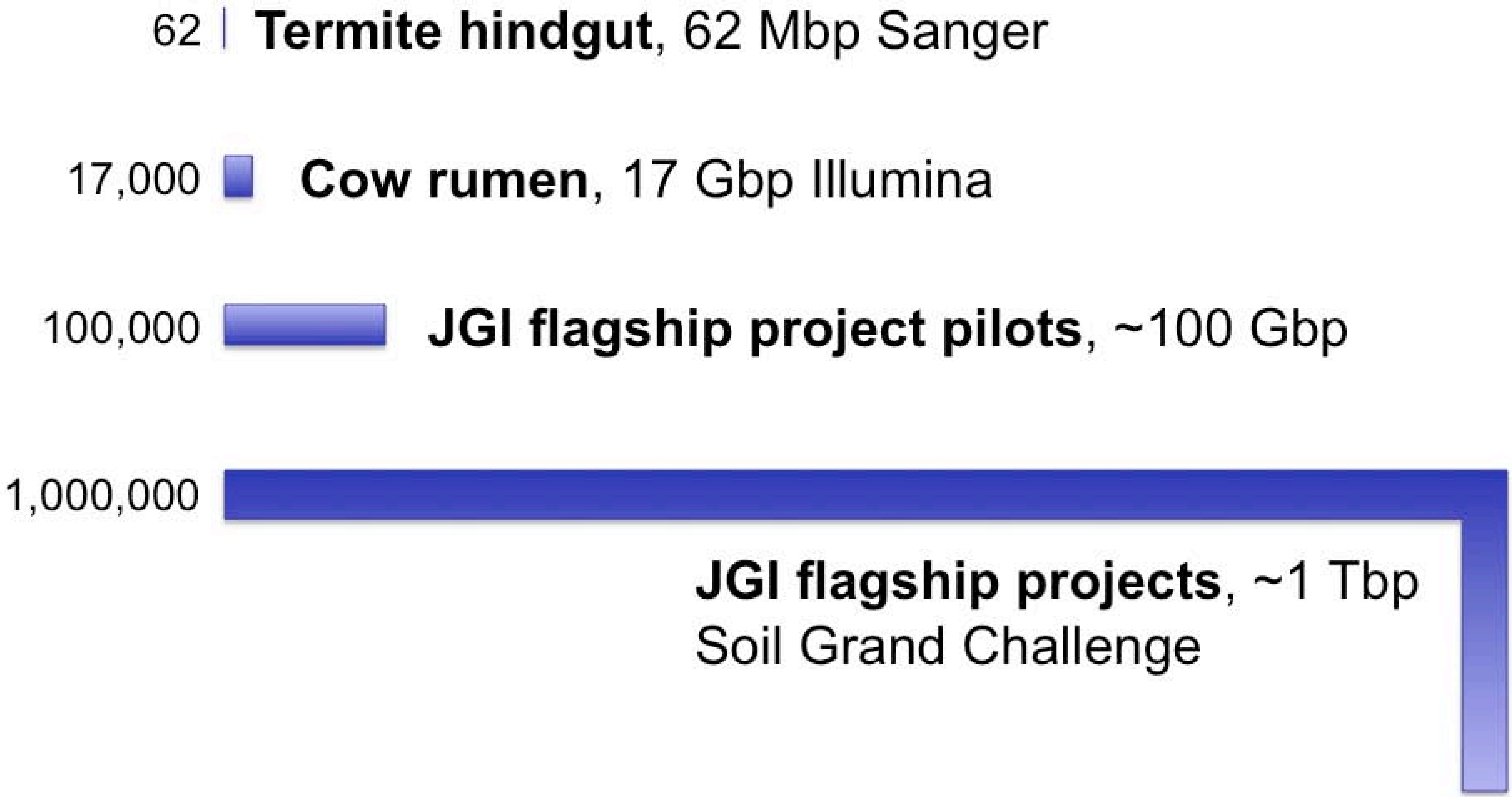
8 Reference Core metagenomes

454 Titanium sequencing

Illumina sequencing

7 spaced subsamples along a quadrant, including reference core/site (56 total DNA samples)

Pryotag sequencing



62 **Termite hindgut**, 62 Mbp Sanger

17,000 **Cow rumen**, 17 Gbp Illumina

100,000 **JGI flagship project pilots**, ~100 Gbp

1,000,000 **JGI flagship projects**, ~1 Tbp
Soil Grand Challenge

Courtesy of Phil Hugenholtz, JGI

Ongoing

- Continued investigation of microbes, genes, pathways in permafrost & active layer
  - (13C-labeled metabolites)
- Sequencing of additional samples for statistical analysis
- New samples collected from bogs will be sequenced at JGI + additional "omics"
  - Transcriptomics, Metabolomics, Proteomics
- Other sites being sequenced with a range of C sequestration potential
  - Puerto Rican Rain forest
  - Great Prairie pilot study at JGI
  - Rothamsted grassland (TerraGenome)

Terragenome

International Soil Metagenome Sequencing Consortium

(http://www.terragenome.org)

Points to address:

- Improvements in protein extraction methods, yield and quality
- Quantification of specific proteins
- Mapping of functional pathways from proteome data
- Improving availability of matching metagenome data
  - Working with 454 and Illumina data

Current challenge- Correlation of multi-omics data sets

- 454 pyrotag data: 250,000 16S rRNA gene sequences
- Metaproteome data: >4000 proteins
- Metabolite data: 18,000 metabolites
- Metagenome data: >43 Mb sequence

Collaborators

JGI
Eddy Rubin
Rachel Mackelprang
Susannah Tringe

Michigan State University
James Tiedje

Illumina
Dirk Evers

Rothamsted Field Station
Penny Hirsch
Ian Clark

Karolinska Institute
Lars Engstrand

Orebro Univ. Hospital
Jonas Halfvarsson

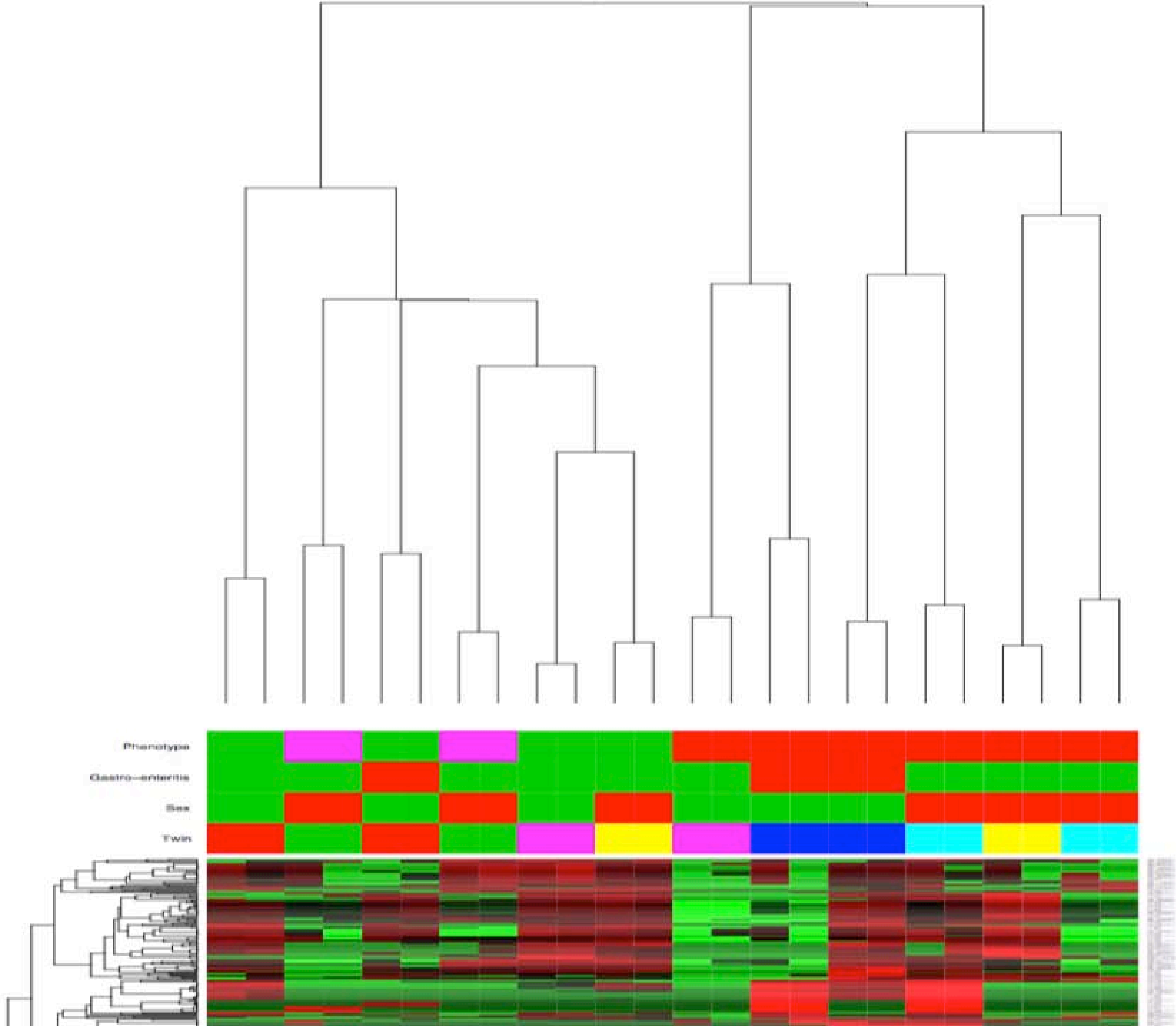
Univ. Maryland School Medicine
Claire Fraser-Liggett
Brandi Cantarel

ORNL
Robert Hettich
Nathan Verberkmoes
Alison Russell

Funding:

NATIONAL INSTITUTES OF HEALTH

DEPARTMENT OF ENERGY
UNITED STATES OF AMERICA

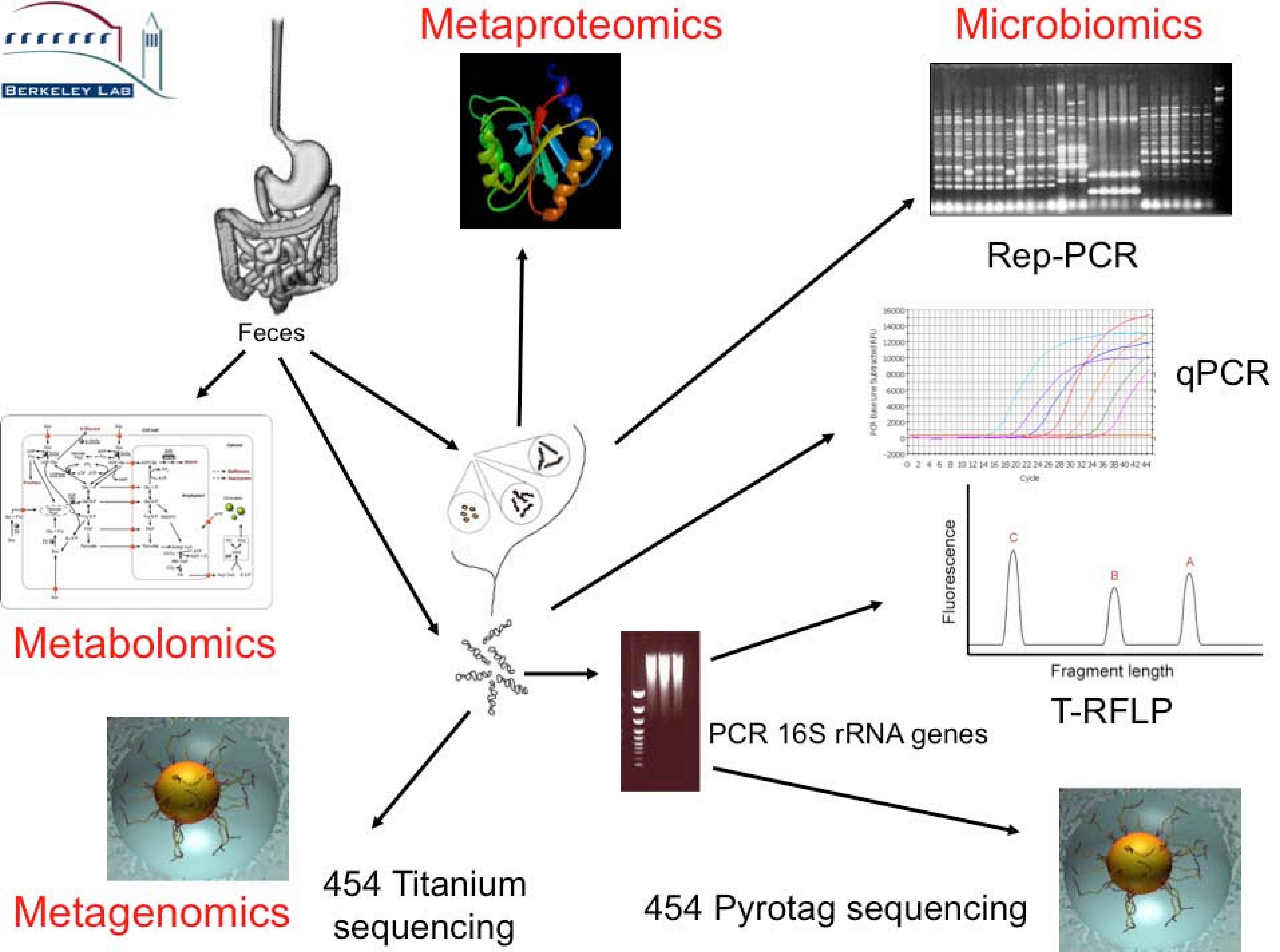

Phenotype
Gastro−enteritis
Sex
Twin

Subject	Ave. Proteins (duplicate runs)	
Healthy	3207	←
Healthy	3057	←
Crohns	2771	
Crohns	2553	
Crohns	2814	
Crohns	2116	
Crohns	1826	
Crohns	2352	
Crohns	1872	
Healthy	4073	←
Crohns	2228	
Healthy	4060	←

DeepSoil

Consortium:

- Dirk Evers (Illumina, Cambridge, UK)
- Penny Hirsch and Ian Clark (Rothamsted Research, UK)
- Folker Meyer (Argonne National Laboratory, USA)
- Janet Jansson (Lawerence Berkeley National Laboratory, USA)
- Jeroen Raes (VIB, Brussels, Belgium)
- Jim Tiedje (Michigan State University, USA)
- Tim Vogel & Pascal Simonet (Ecole Centrale Lyon, France)
- Eric Triplett (University of Florida, USA)



Metaproteomics

Microbiomics

Feces

Rep-PCR

qPCR

T-RFLP

Metabolomics

PCR 16S rRNA genes

Metagenomics

454 Titanium sequencing

454 Pyrotag sequencing

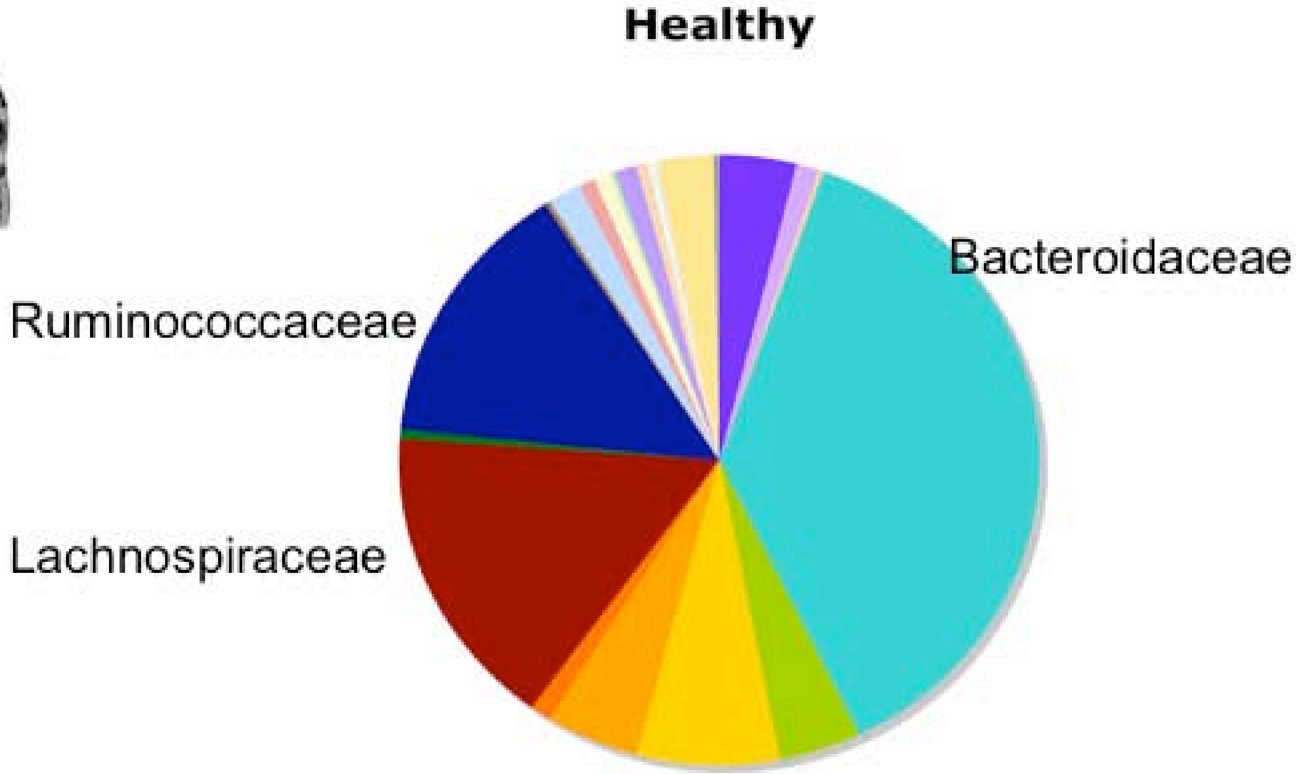
”Meta”genomics vs ”Meta”proteomics

- Genomics/Metagenomics
  - Information on gene content having the potential of being expressed

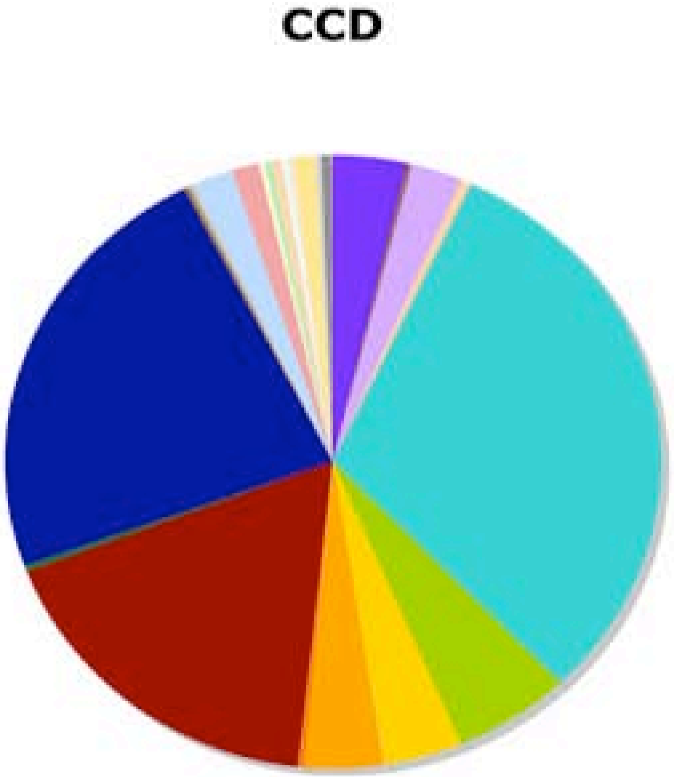
- Proteomics/Metaproteomics
  - Provides evidence of gene expression under a given condition
    i.e. Genes must have been transcribed and translated to produce a protein product
  - Better for assessment of microbial function

454 Pyrotag sequencing of 16S rRNA genes (v5-v6)

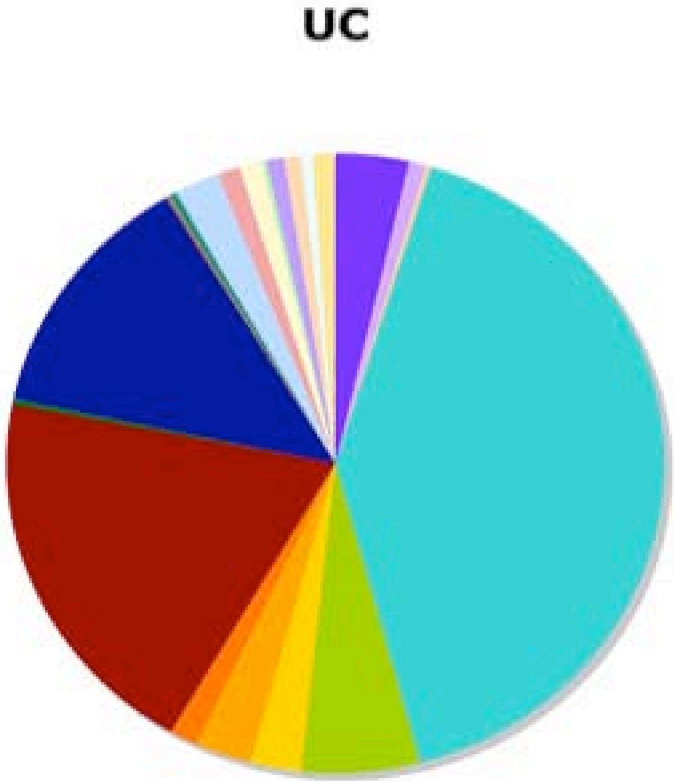
- 40 twin pairs (80 individuals)
- Total of 250,000 sequences, average of 2000 reads per sample
- 6600 OTUs, average of 333 per individual (5% dissimilarity)
- Healthy core microbiota
    - Firmicutes (57)
        - Lachnospiraceaea & Ruminococcaceaea
    - Bacteroidetes (26)
    - Actinobacteria (2)
        - Coriobacteriaceaea (Collinsella aerofaciens)
    - Proteobacteria (2)



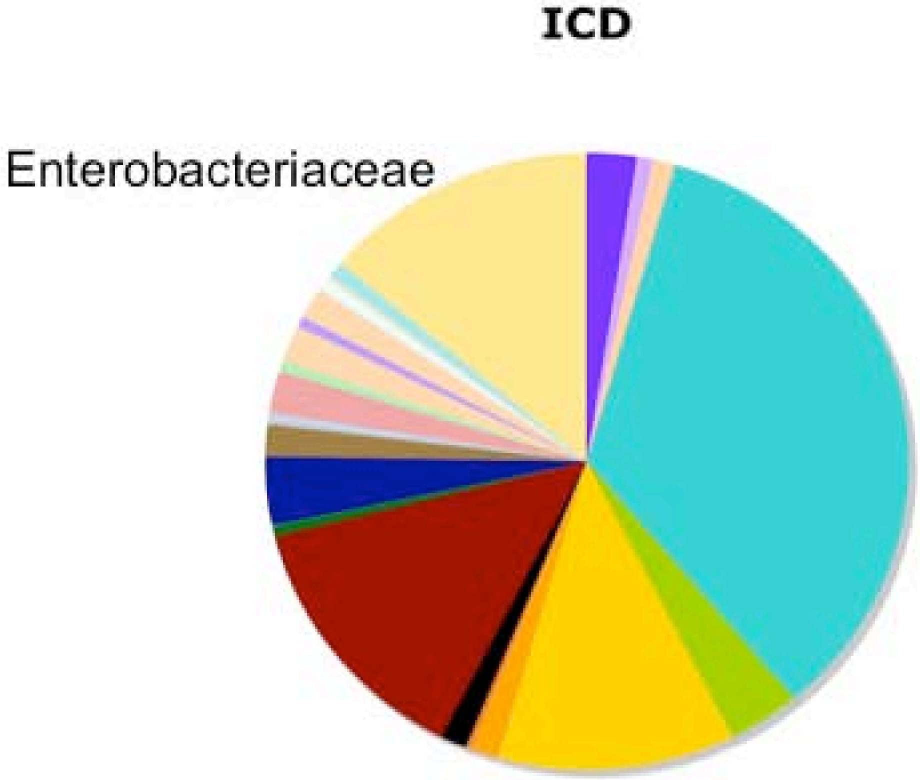
Healthy



CCD



UC



ICD