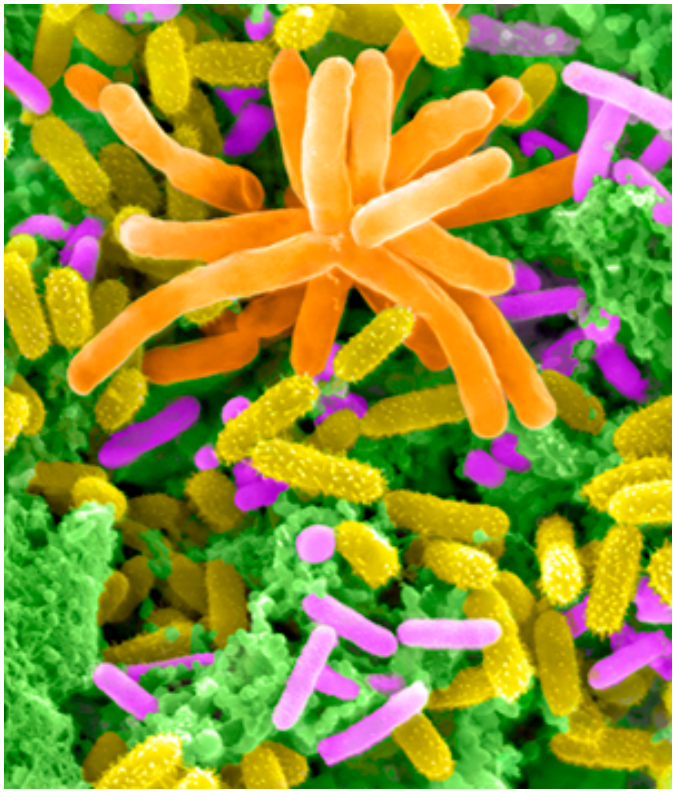
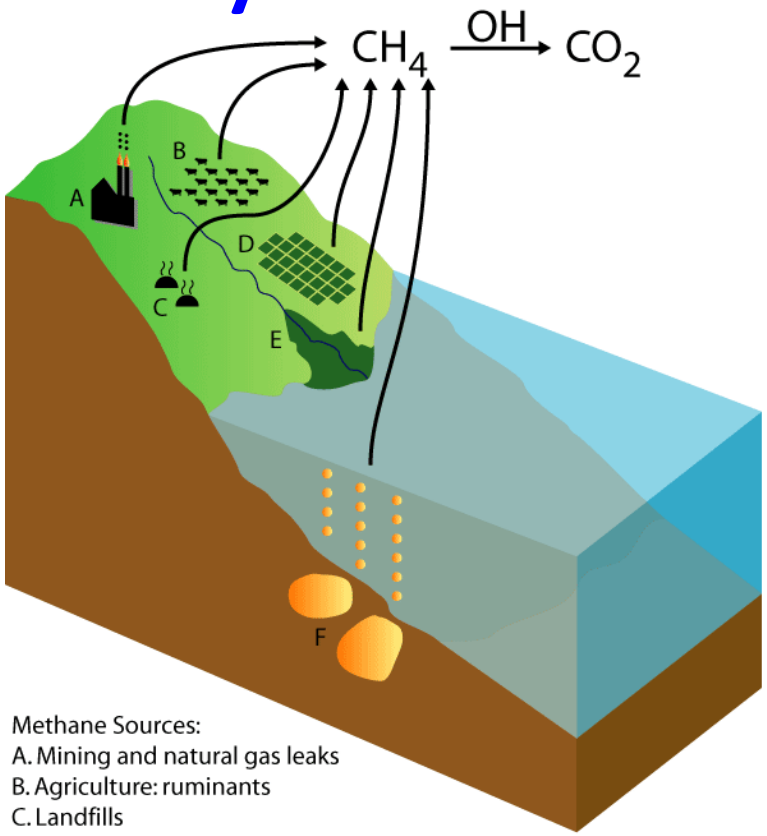
Linking Microbes and Biogeochemistry



Methylamine-enriched community of Lake Washington sediment featuring Methylotenera cells.
Photo © Dennis Kunkel Microscopy, Inc. (Color by Ekaterina Latypova)
http://www.jgi.doe.gov/News/news_8_18_08.htmll



Methane Sources:
A. Mining and natural gas leaks
B. Agriculture: ruminants
C. Landfills
D. Agriculture: rice paddies
E. Natural wetlands
F. Hydrates

http://www.giss.nasa.gov/research/features/200409_methane/schematic.gif

V. Rich
Feb 14th, 2014
SWES 410/510

A way *I* think about *some* of the methods involved in connecting them:

1) Indirect molecular inference:
    a. 16S -> functional guilds
    b. gene ecology (of known functional genes)
    c. transcriptional and translational activity (of known functional genes)

2) direct molecular inference:
    a. SIP

3) direct physiological measurement:
    a. culture based charactrization
    b. Ecolog plate substrate utilization
    c. enzyme assays

4) indirect physiologial measurement:
    a. biomolecule isotopic signatures
    b. emitted product isotopic signatures

5) modeled linkages - Moira Hough will tell us about this

Paul Bodelier's way:

Figure 1 in Bodelier 2011 (optional reading on D2L): "…important elements in elucidating the role of microbial diversity in ecosystem functioning.

"Crucial element is the application of a Functional Biodiversity concept to link microbial diversity to ecosystem functioning. This approach will facilitate predictive ecosystem modeling and will be fostered by omics techniques. However, to make this conceptual step the mechanistic insight into what is going on in the "black box" being the structure and functioning of microbial communities and underlying populations and cells, needs to be elucidated. Application of ecological theory, conceptual experimental design, novel methodology, and mathematical modeling will be the key to gain access to the knowledge in the "Black box."

Eugene Madsen's way:

From Madsen 2011 (optional reading on D2L)

" …five [four] stages of environmental microbiological inquiry leading to advances in biogeochemistry:

- Stage 1. Discovery of new microbiological process. Prove that microorganisms are capable of catalyzing the process of interest. This is achieved via laboratory incubation of environmental samples and/or via chemical or biomarker assays performed on complex, uncharacterized microbial communities accompanied by materials from soils, sediments, or waters.

- Stage 2. Validation of the discovery by finding representative microbiological agents. Refine the test system by isolating a single microorganism capable of catalyzing the process or obtaining a simplified, highly enriched consortium of microbial populations exhibiting the process or via a convincing combination of biomarkers and physiological evidence."

• Stage 3. Characterization of agents and the physiological, biochemical, and/or genomic mechanisms of the biogeochemical process(es) they catalyze. Use of controlled laboratory incubations, chemical assays, isotopic tracers, biomarkers,… bioinformatics to define metabolites, metabolic pathways, enzymatic reactions, and the genetic basis of cellular processes.

• Stage 4. Field verification of ecological relevance of agents and/or their biogeochemical impact. Apply the tools, insights, biomarker analyses from Stage 3 to real-world field sites where microbiological agents (specific taxa and/or their functional genes) are influencing ecological conditions.

• Stage 5. Biotechnological innovation and/or improved site management based on understanding biogeochemical process mechanisms. In some instances microbial-mediated processes can be transplanted from their ecosystem contexts to human-engineered settings for commercial or industrial applications.

Eugene Madsen's take on env. microbiologist toolbox

Five key approaches in environmental microbiology "tool box"

Site geochemistry: Analytical chemistry proves presence of compounds indicative of microbial process (reinforced by flux data and isotopic fractionation patterns).

Cultivation: Provision of appropriate nutrients in liquid or solid media allows isolation of microorganisms catalyzing process of interest (e.g. denitrification or benzene biodegradation).

Incubations: Placing environmental samples or pure cultures in sealed, laboratory vessels allows documentation of physiological changes (e.g. methane generation or consumption) effected by microorganisms.

Biomarkers: Extraction and analysis of key cellular constituents. These provide insights into the taxonomic composition and/or functional potential of microorganisms by focusing upon phospholipid fatty acids, DNA, ribosomal RNA, messenger RNA, or proteins followed by GC/MS, LC/MS and/or various molecular biology procedures ranging from small-scale sequencing to high-throughput meta-genomics, meta-transcriptomics, and meta-proteomics.

Microscopy: Allows images of microorganisms and cell associations to be obtained from site samples or laboratory incubations. Depending upon analytical approach and staining targets, information yielded includes enumeration, identity (e.g. via small subunit rRNA FISH probes), localization of biomarkers within cells, and cell-specific substrate incorporation (e.g. via secondary ion mass spectrometry).

Box 1 From Madsen 2011 (optional reading on D2L)

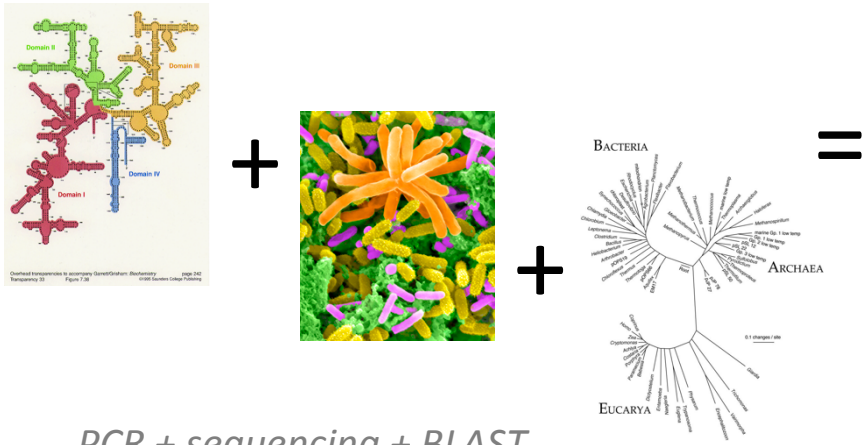
1. Indirect molecular inference

a. 16S -> functional guilds

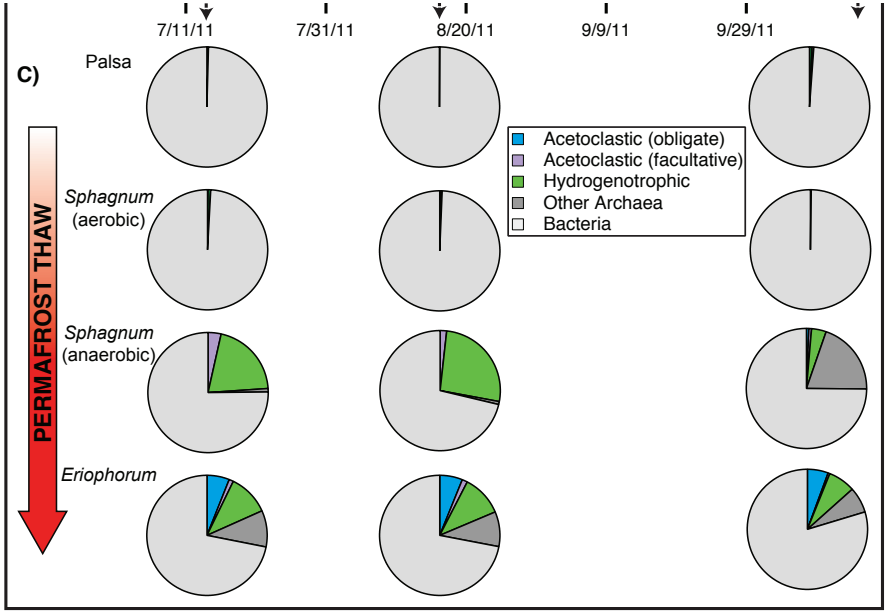
Wikipedia defines ecological guilds as:

- "any group of species that exploit the same resources, often in related ways"
- they may or may not occupy the similar niches
- "defined by locations, attributes, and activities of component species; e.g. mode of acquiring nutrients, mobility, and habitat zones"
- "does not typically have strict, or even clearly defined boundaries. A broadly-defined guild will practically always have constituent guilds; for example, grazing guilds will have some species that concentrate on coarse, plentiful forage, while others concentrate on low-growing, finer plants."

e.g. McCalley et al *submitted…*



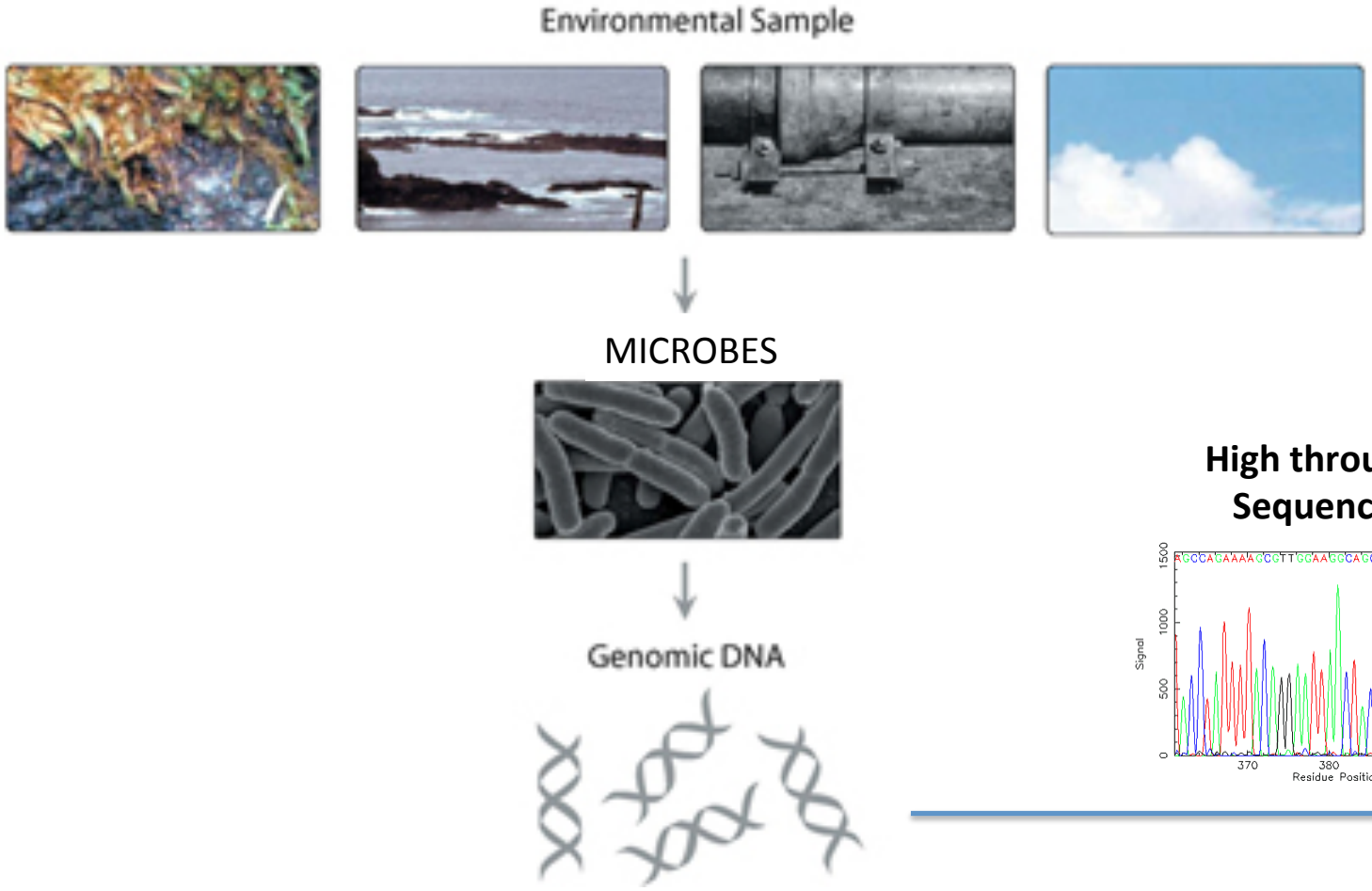
PCR + sequencing + BLAST

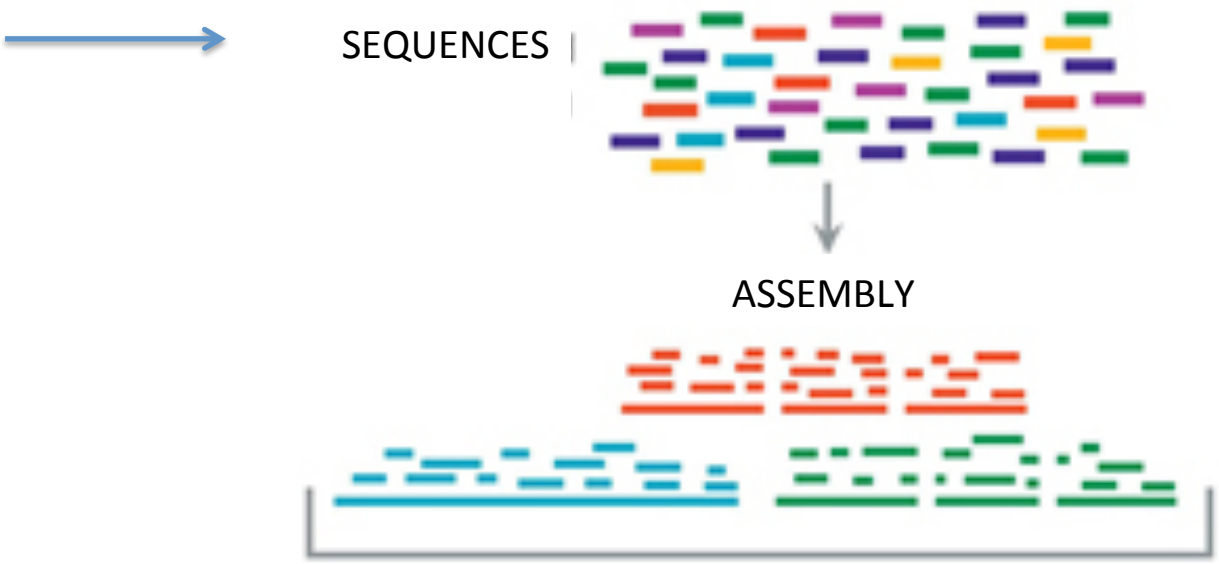


1. Indirect molecular inference

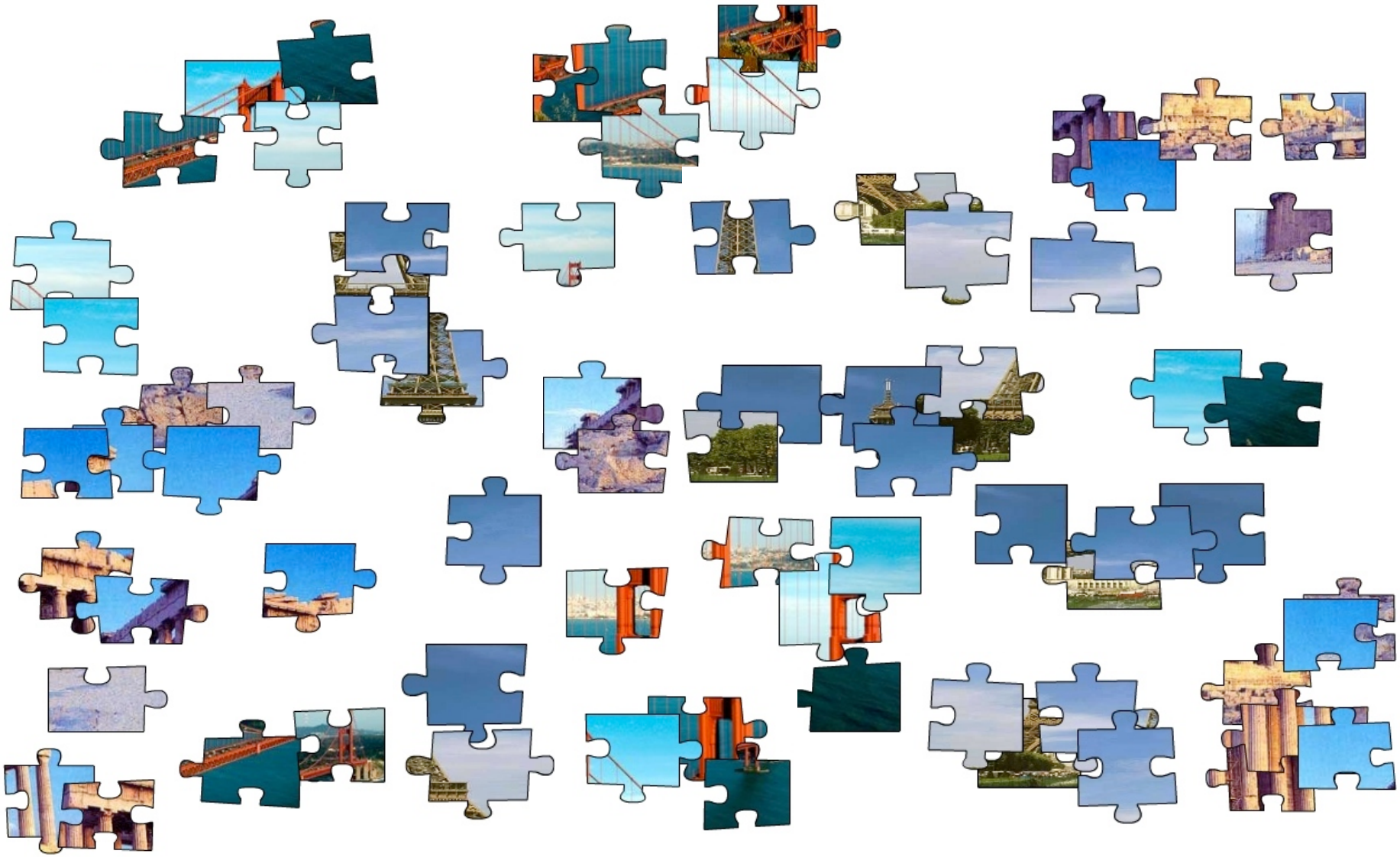
b. gene ecology (of known functional genes)

Have to back up a sec. to recap Metagenomics

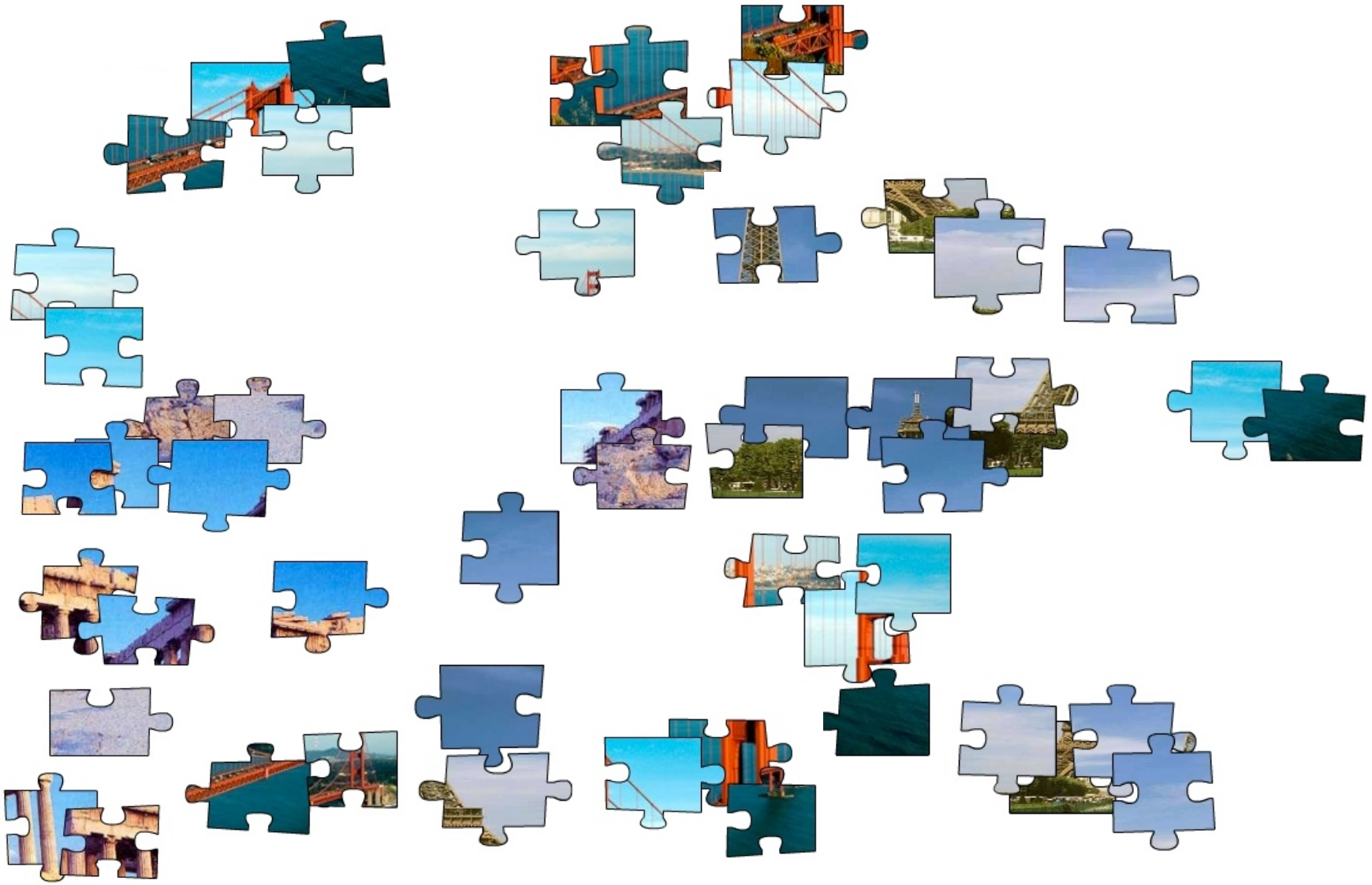



SEQUENCES
ASSEMBLY

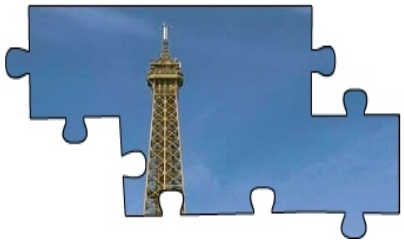
Metagenome assembly

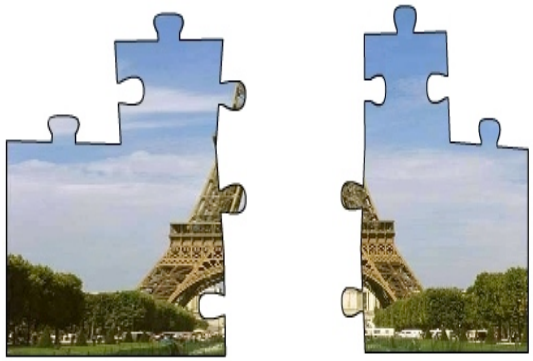


Metagenome assembly



Metagenome assembly





SEQUENCES

ASSEMBLY

Predict & Analyze Genes

Metabolism & Transport

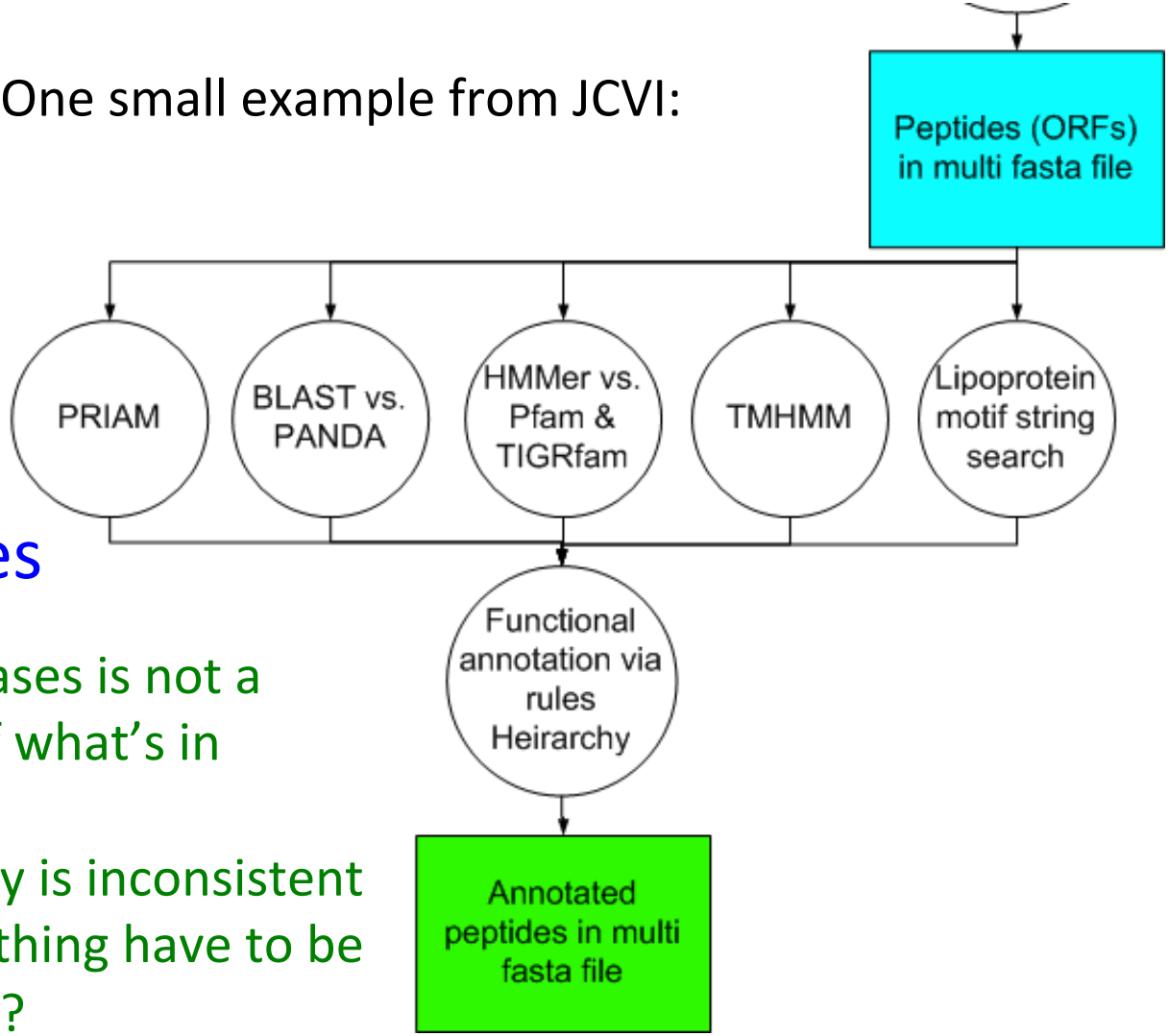
Lateral Gene Transfer?

Phylogeny

Compare Genes from Environments

Annotation is no small feat

One small example from JCVI:



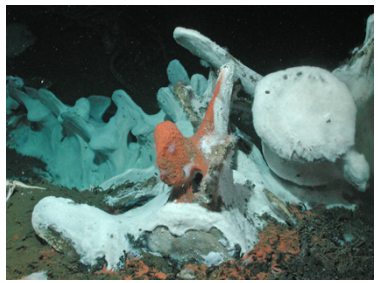
There are many many many many reference databases

- What is IN those databases is not a random broad swath of what's in nature…
- The annotation accuracy is inconsistent
- How similar does something have to be to infer similar function?
- Higher-level functional organization is spotty

1. Indirect molecular inference

b. gene ecology (of known functional genes)

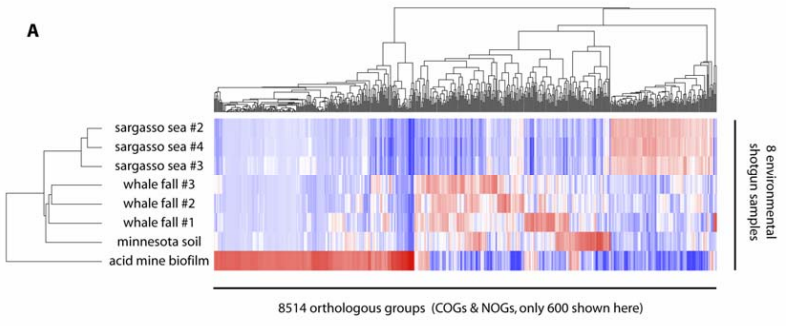
e.g. Tringe et al 2005, Science

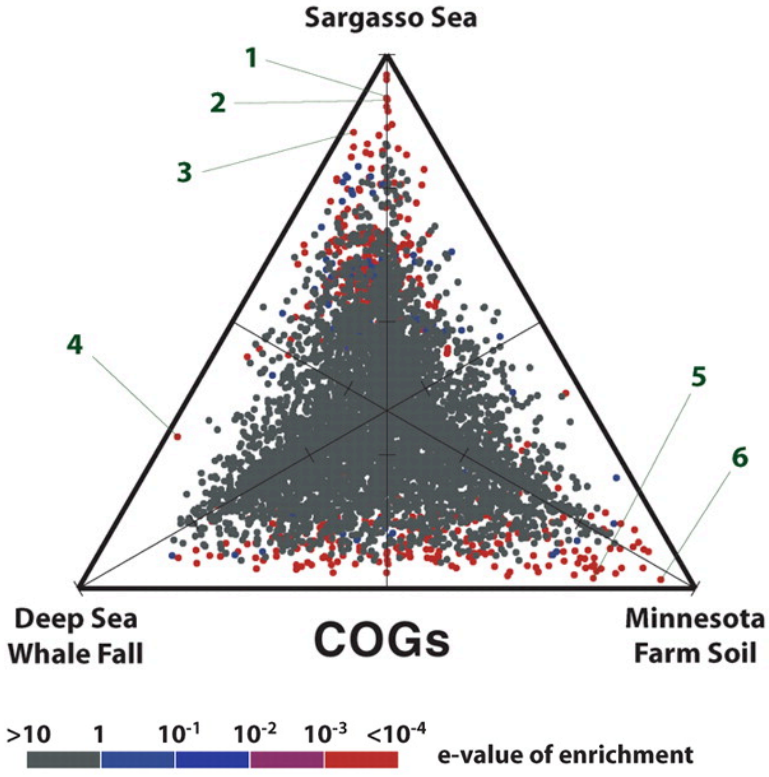


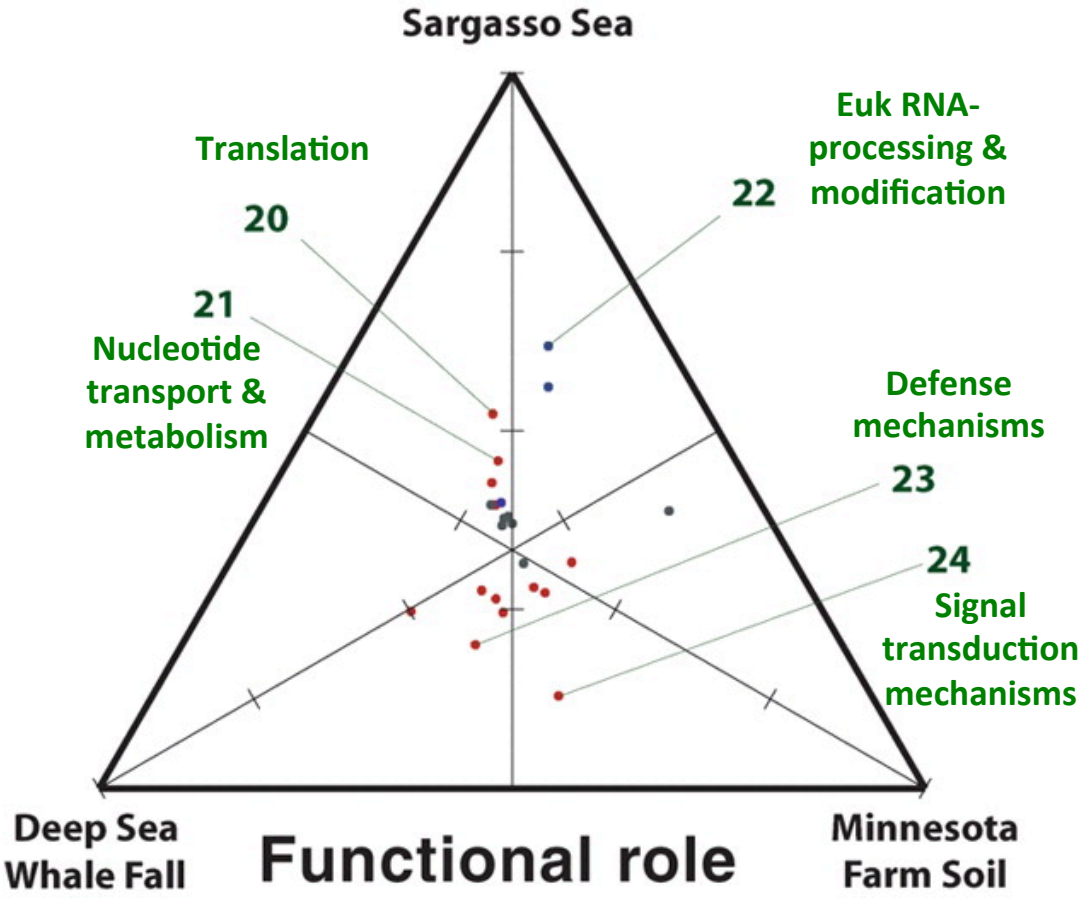
Comparative Metagenomics of Microbial Communities

Susannah Green Tringe,[1,2*] Christian von Mering,[3*] Arthur Kobayashi,[1] Asaf A. Salamov,[1] Kevin Chen,[4] Hwai W. Chang,[5] Mircea Podar,[5] Jay M. Short,[5] Eric J. Mathur,[5] John C. Detter,[1] Peer Bork,[3] Philip Hugenholtz,[1] Edward M. Rubin[1,2†]

The species complexity of microbial communities and challenges in culturing representative isolates make it difficult to obtain assembled genomes. Here we characterize and compare the metabolic capabilities of terrestrial and marine microbial communities using largely unassembled sequence data obtained by shotgun sequencing DNA isolated from the various environments. Quantitative gene content analysis reveals habitat-specific fingerprints that reflect known characteristics of the sampled environments. The identification of environment-specific genes through a gene-centric comparative analysis presents new opportunities for interpreting and diagnosing environments.





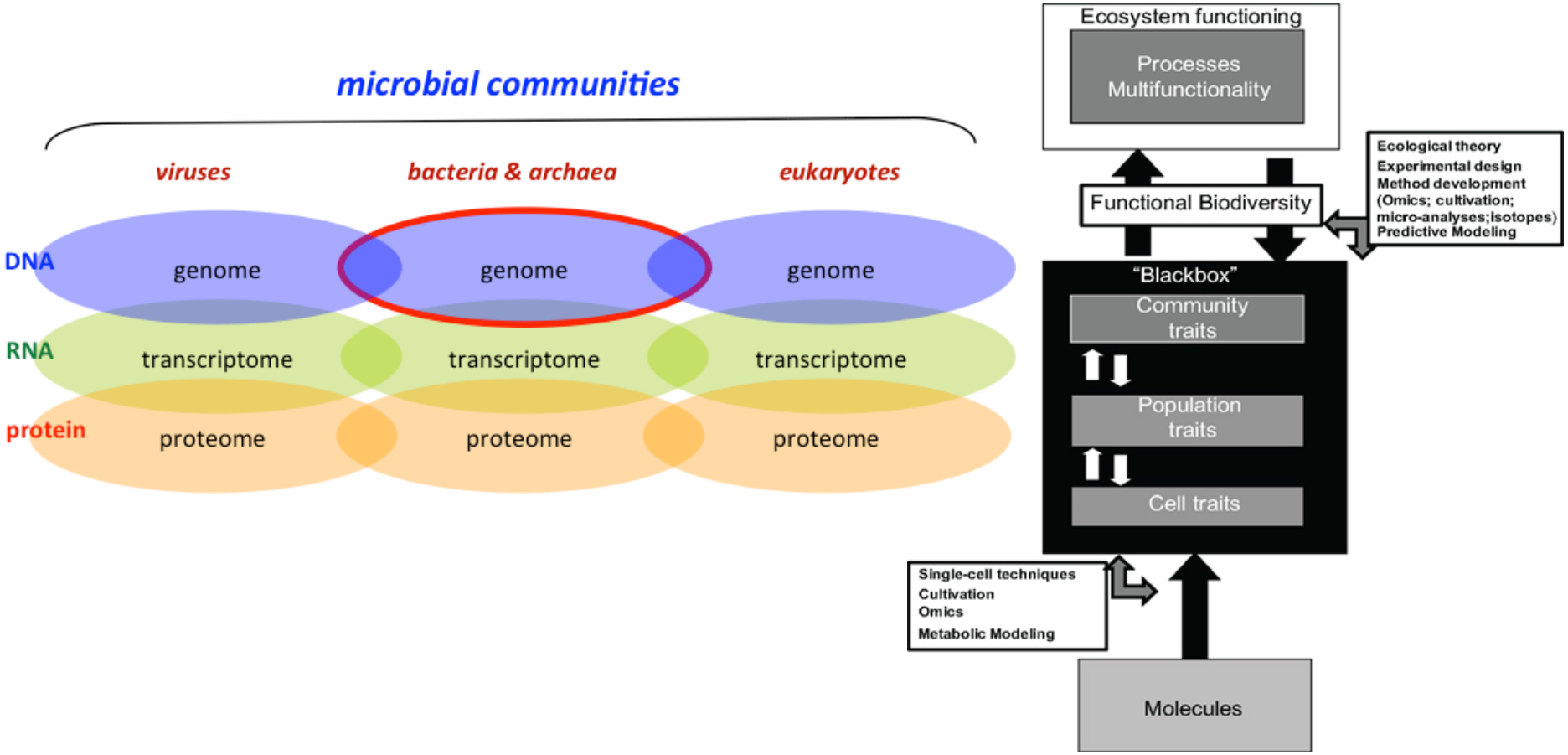


= COG Functional Categories

1. Indirect molecular inference

c. transcriptional and translational activity (of known functional genes)

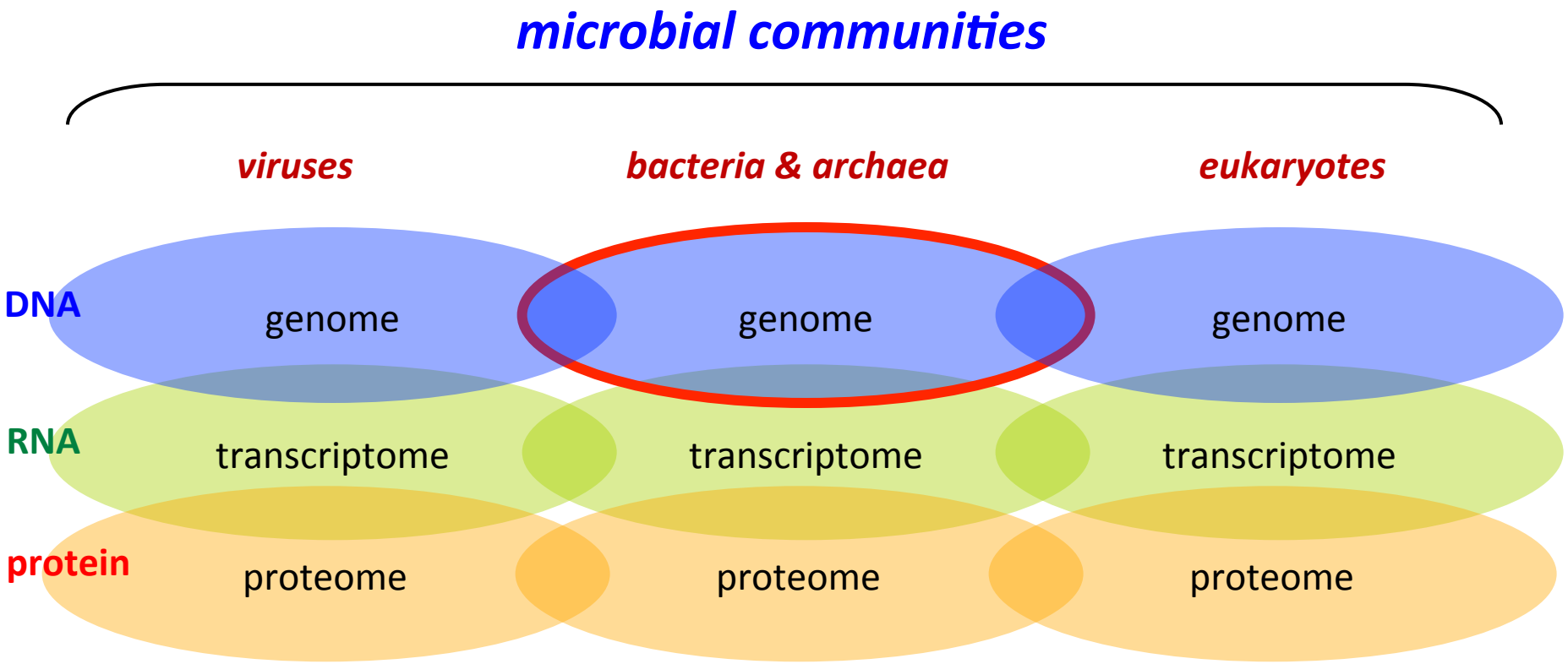
Metagenomics is just the first level



1. Indirect molecular inference

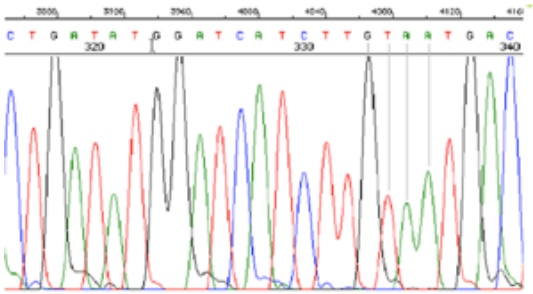
c. transcriptional and translational activity (of known functional genes)

Metagenomics is just the first level



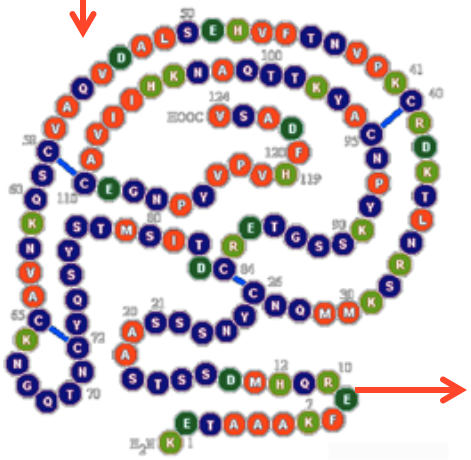
Shotgun community proteomics

DNA

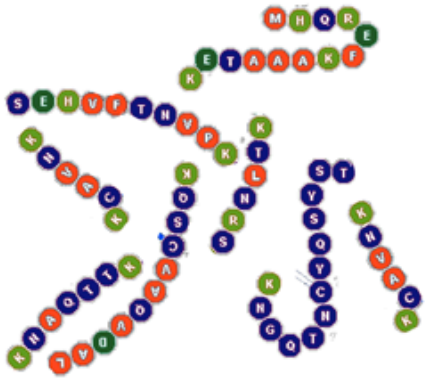


DNA sequence

environmental sample

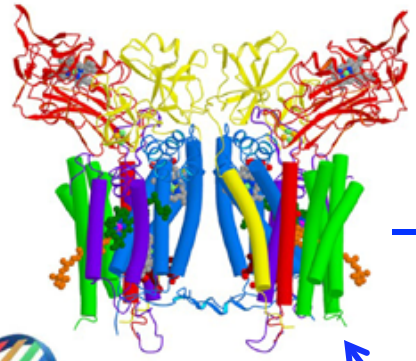


predicted protein sequence

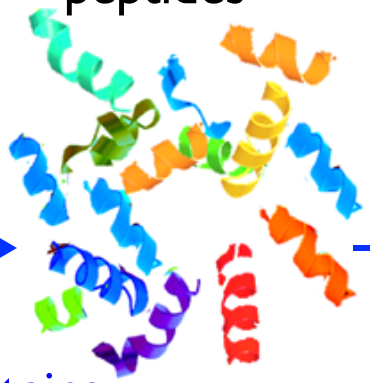


predicted peptides

Shotgun community proteomics



peptides



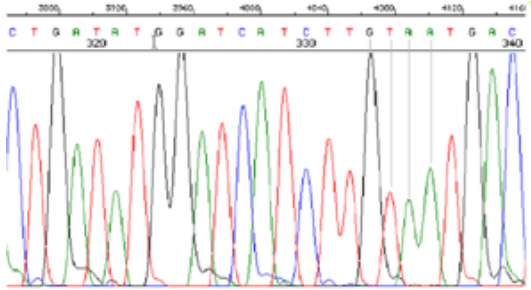
LC-MS/MS

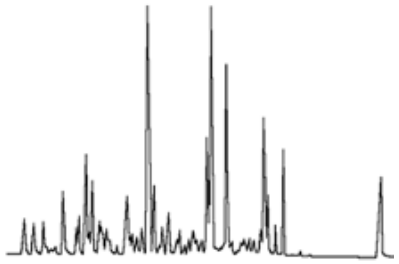
proteins

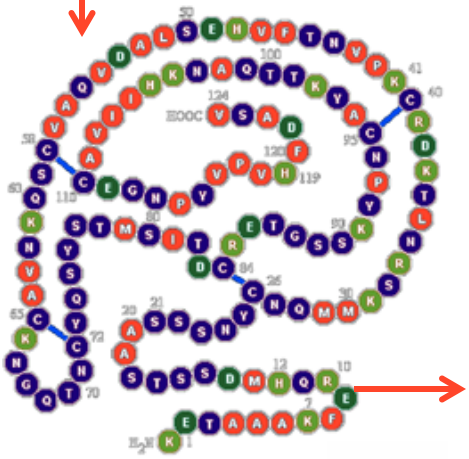
DNA

environmental sample

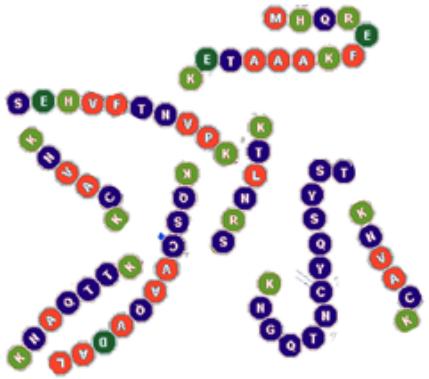
DNA sequence



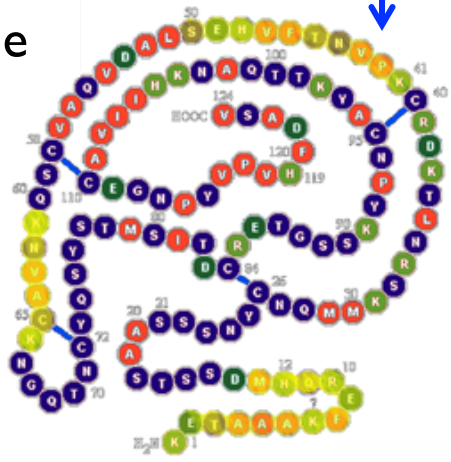




predicted protein sequence

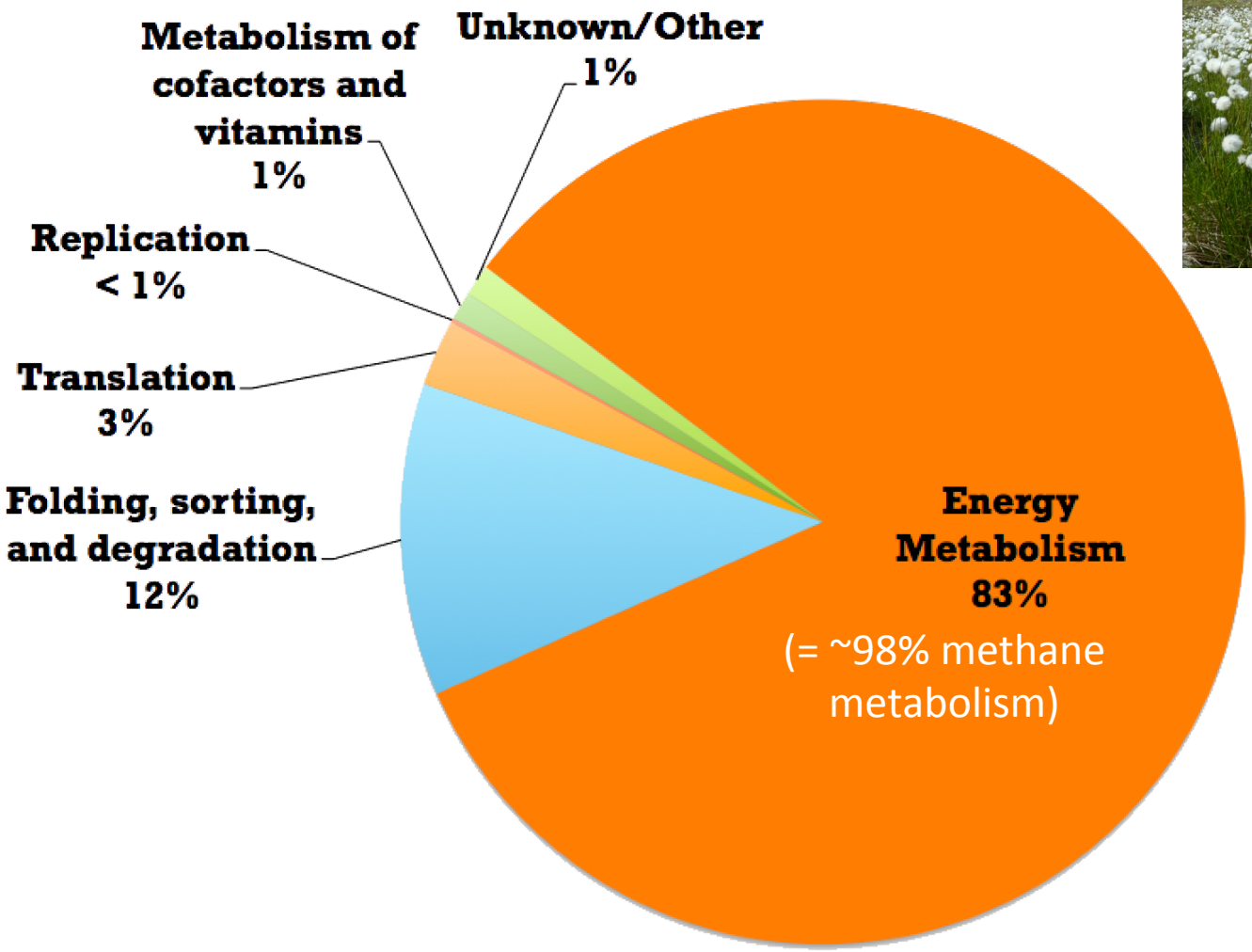


predicted peptides



protein identification

An example from the site of McCalley et al



The proteome of the methanogens present

2. Direct molecular inference

a. Stable isotope probing ("SIP")

- Purpose: to identify the organisms responsible for a given chemical transformation

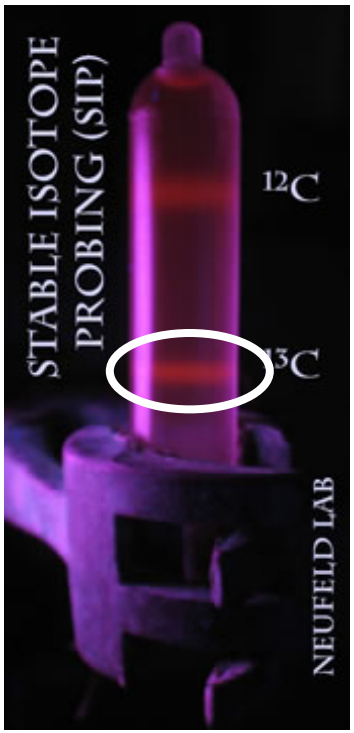
- Method principle: if you incubate a "wild" community with a stable isotope-enriched substrate, the organisms that transform that substrate will become labeled by the stable isotope. You can then separate the labeled biomolecules of those organisms from the rest of the community. These biomolecules can inform you about the identity and genetic content of the chemically active organisms.

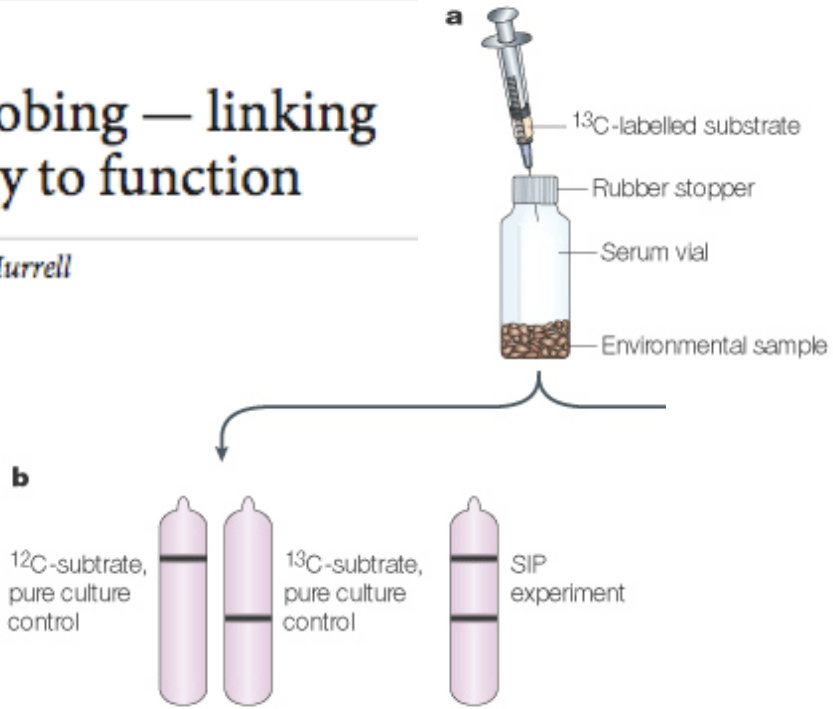
INNOVATION

Stable isotope probing — linking microbial identity to function

Marc G. Dumont and J. Colin Murrell

2005 in NRMicro





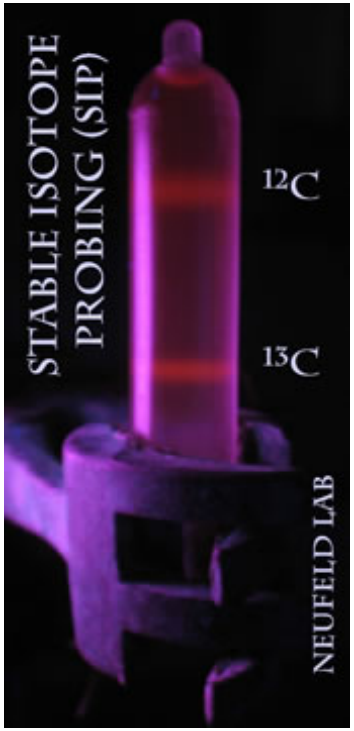
Substrate eg: $^{13}CH_4$
Incubation time eg: 16 days

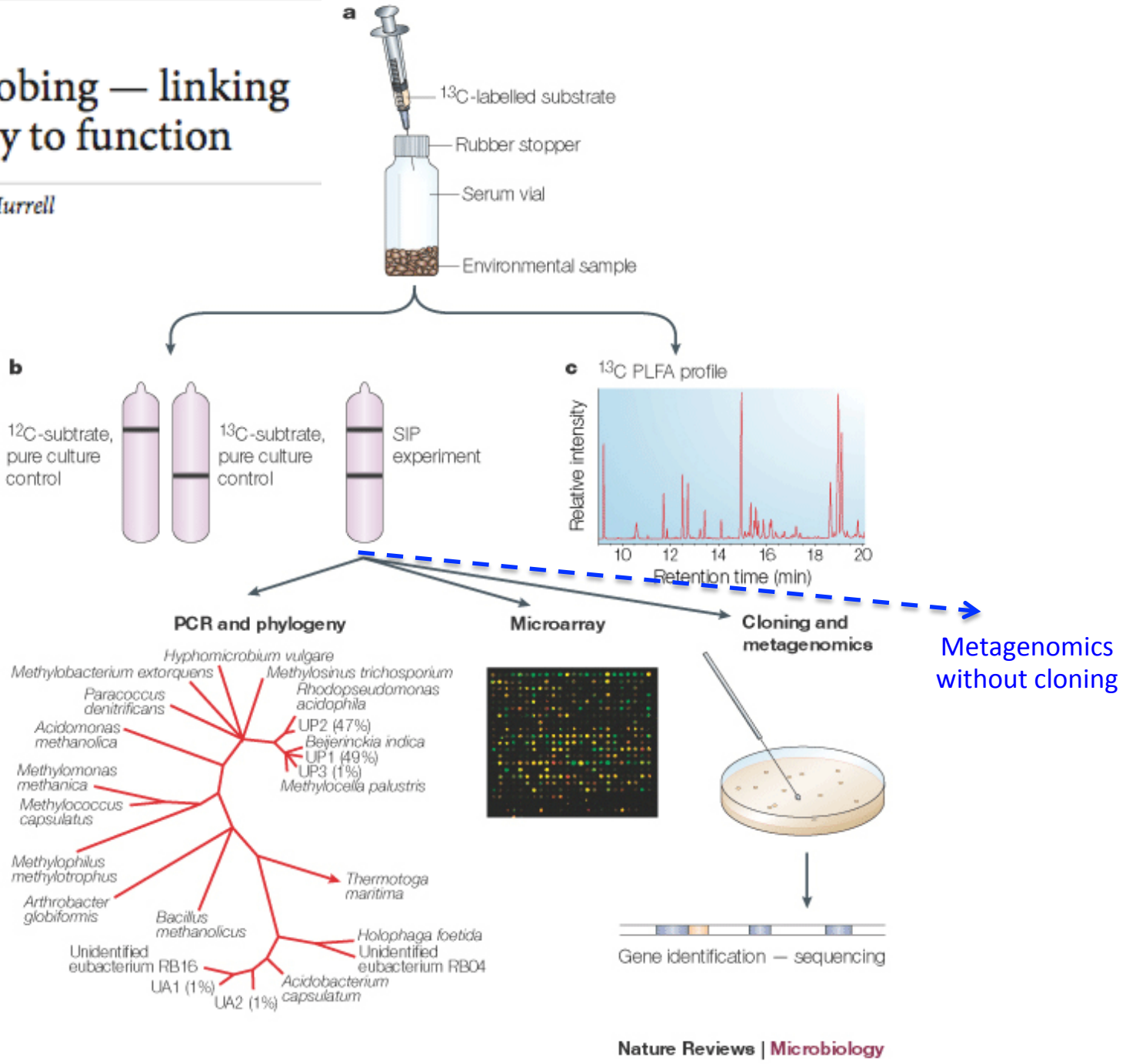
INNOVATION

Stable isotope probing — linking microbial identity to function

Marc G. Dumont and J. Colin Murrell

2005 in NRMicro





Metagenomics without cloning

2. Direct molecular inference

a. Stable isotope probing ("SIP")

•Purpose: to identify the organisms responsible for a given chemical transformation

•Method principle: if you incubate a "wild" community with a stable isotope-enriched substrate, the organisms that transform that substrate will become labeled by the stable isotope. You can then separate the labeled biomolecules of those organisms from the rest of the community. These biomolecules can inform you about the identity and genetic content of the chemically active organisms.

•Caveats: (a) incubation time is a bugaboo – cross-feeding of biomolecules can muddy the signal over time. How long depends on the organisms, their growth rates, and their trophic interactions. (b) some reactions are catalyzed without incorporation of the labeled atom (e.g. by exoenzymes). (c) orgs that feed on a diversity of substrates may not be sufficiently labeled (d) important to use ambient concentrations of substrate or you're doing an enrichment…

Environmental Microbiology (2008) 10(10), 2609–2622 doi:10.1111/j.1462-2920.2008.01683.x

Revealing the uncultivated majority: combining DNA stable-isotope probing, multiple displacement amplification and metagenomic analyses of uncultivated *Methylocystis* in acidic peatlands

Yin Chen,[1] Marc G. Dumont,[1†] Josh D. Neufeld,[1‡] Levente Bodrossy,[2] Nancy Stralis-Pavese,[2] Niall P. McNamara,[3] Nick Ostle,[3] Maria J. I. Briones[4] and J. Colin Murrell[1*]

[1]*Department of Biological Sciences, the University of Warwick, Coventry, CV4 7AL, UK.*
[2]*Department of Bioresources, Austrian Research Centre, GmbH, A-2444 Seibersdorf, Austria.*
[3]*Centre for Ecology and Hydrology, Lancaster Environment Centre, Lancaster, LA1 4AP, UK.*
[4]*Departamento de Ecología y Biología Animal, Facultad de Biología, Universidad de Vigo, 36310 Vigo, Spain.*

template (1–5 ng) was used in MDA to circumvent this bias and chimeric artefacts were minimized by using an enzymatic treatment of MDA-generated DNA with S1 nuclease and DNA polymerase I. Screening of the metagenomic library revealed one fosmid containing methanol dehydrogenase and two fosmids containing 16S rRNA genes from these *Methylocystis*-related species as well as one fosmid containing a 16S rRNA gene related to that of *Methylocella*/*Methylocapsa*. Sequencing of the 14 kb methanol dehydrogenase-containing fosmid allowed the assembly of a gene cluster encoding polypeptides involved in bacterial methanol utilization (*mxaFJGIRSAC*). This combina-

A way *I* think about *some* of the methods involved in connecting them:

1) Indirect molecular inference:
    a. 16S -> functional guilds
    b. gene ecology (of known functional genes)
    c. transcriptional and translational activity (of known functional genes)

2) direct molecular inference:
    a. SIP

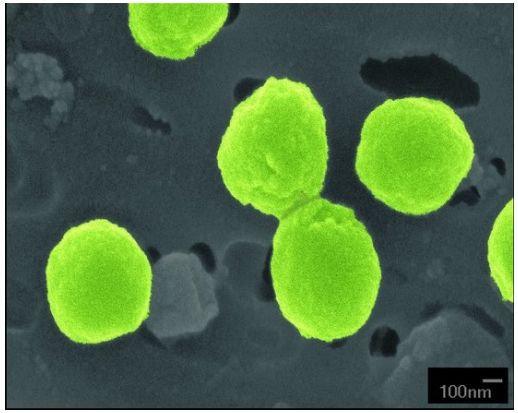
3) direct physiological measurement:
    a. culture based charactrization
    b. Ecolog plate substrate utilization
    c. enzyme assays

4) indirect physiologial measurement:
    a. biomolecule isotopic signatures
    b. emitted product isotopic signatures

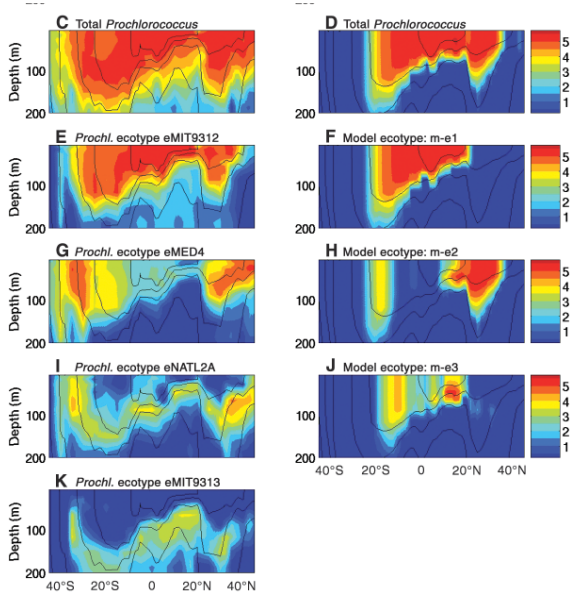
5) modeled linkages - Moira Hough will tell us about this

3. Direct physiological measurement

a. culture-based characterization

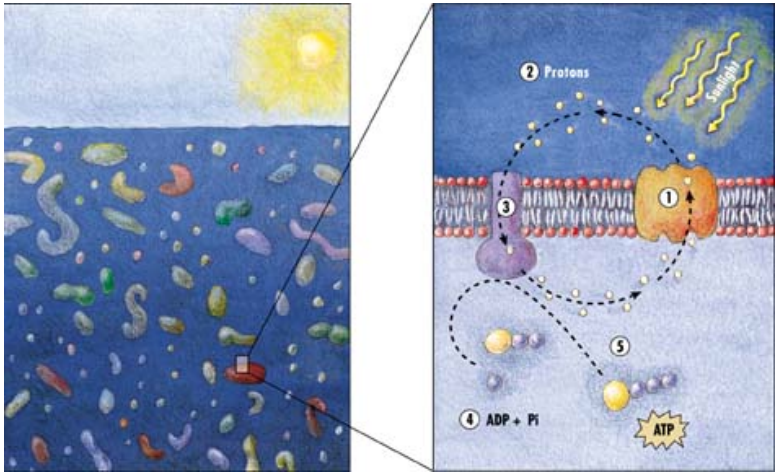


Prochlorococcus!



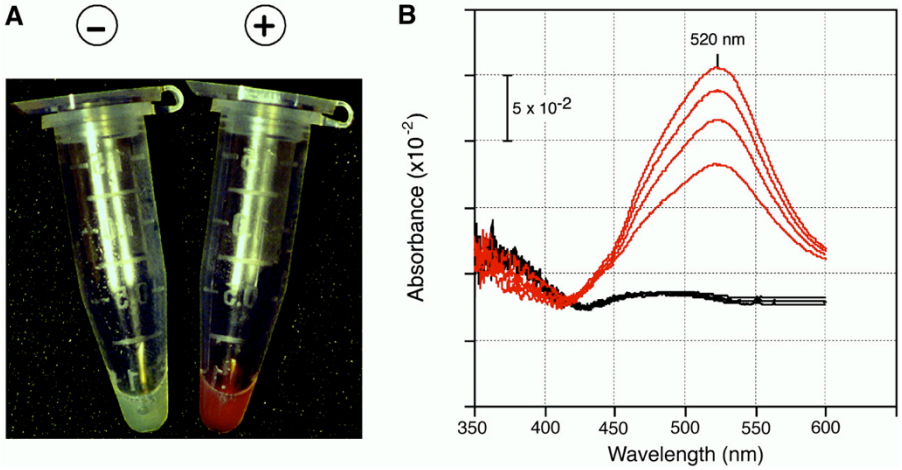
Follows et al 2007

VS.



proteorhodopsin

E. coli +/- the cloned gene



Beja et al 2000

3. Direct physiological measurement

b. Ecolog plate substrate utilization

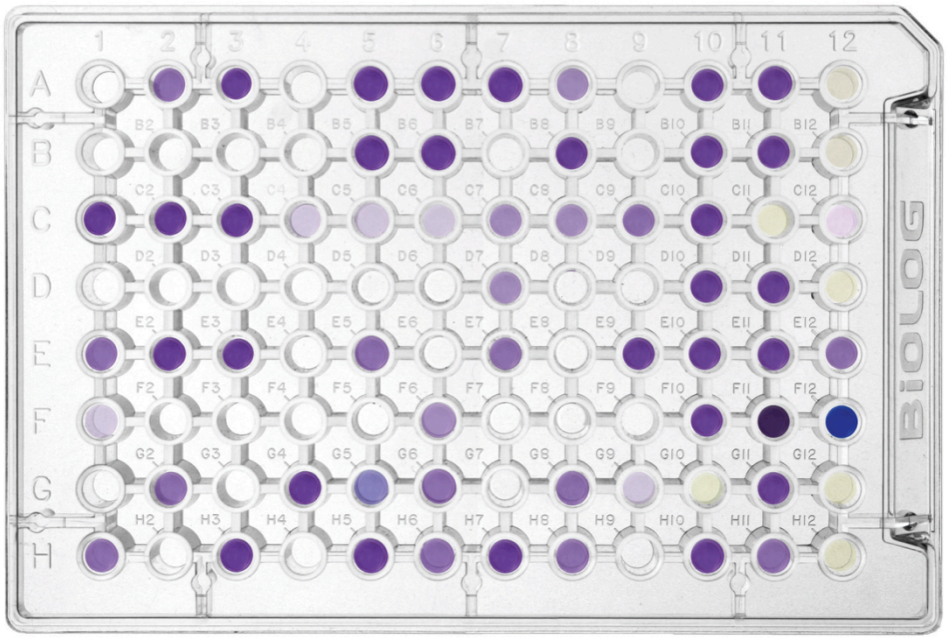
BIOLOG

Microbial Community Analysis

EcoPlate™

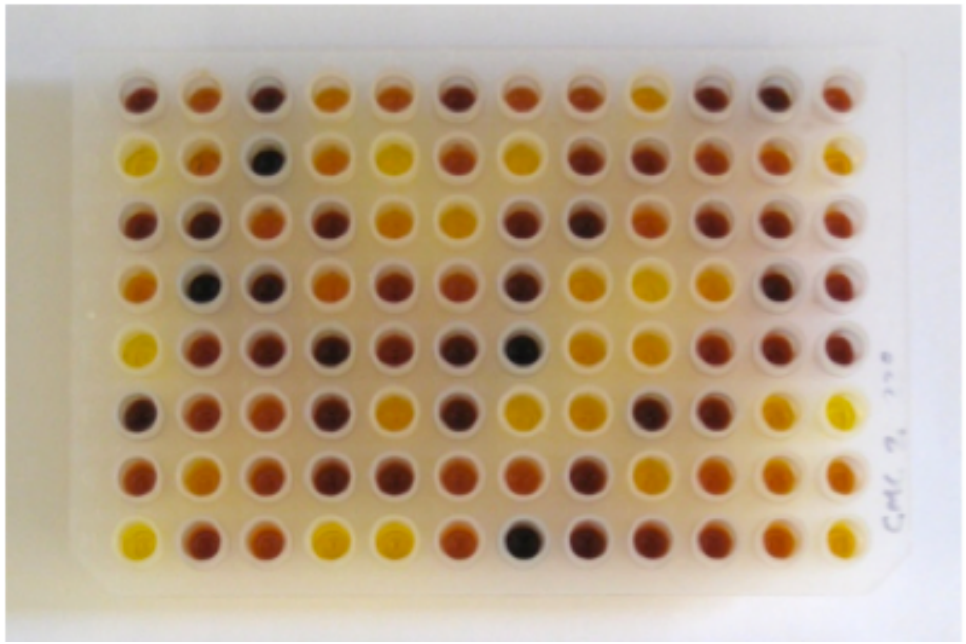
A1 Water	A2 β-Methyl-D-Glucoside	A3 D-Galactonic Acid γ-Lactone	A4 L-Arginine	A1 Water	A2 β-Methyl-D-Glucoside	A3 D-Galactonic Acid γ-Lactone	A4 L-Arginine	A1 Water	A2 β-Methyl-D-Glucoside	A3 D-Galactonic Acid γ-Lactone	A4 L-Arginine
B1 Pyruvic Acid Methyl Ester	B2 D-Xylose	B3 D-Galacturonic Acid	B4 L-Asparagine	B1 Pyruvic Acid Methyl Ester	B2 D-Xylose	B3 D-Galacturonic Acid	B4 L-Asparagine	B1 Pyruvic Acid Methyl Ester	B2 D-Xylose	B3 D-Galacturonic Acid	B4 L-Asparagine
C1 Tween 40	C2 i-Erythritol	C3 2-Hydroxy Benzoic Acid	C4 L-Phenylalanine	C1 Tween 40	C2 i-Erythritol	C3 2-Hydroxy Benzoic Acid	C4 L-Phenylalanine	C1 Tween 40	C2 i-Erythritol	C3 2-Hydroxy Benzoic Acid	C4 L-Phenylalanine
D1 Tween 80	D2 D-Mannitol	D3 4-Hydroxy Benzoic Acid	D4 L-Serine	D1 Tween 80	D2 D-Mannitol	D3 4-Hydroxy Benzoic Acid	D4 L-Serine	D1 Tween 80	D2 D-Mannitol	D3 4-Hydroxy Benzoic Acid	D4 L-Serine
E1 α-Cyclodextrin	E2 N-Acetyl-D-Glucosamine	E3 γ-Hydroxybutyric Acid	E4 L-Threonine	E1 α-Cyclodextrin	E2 N-Acetyl-D-Glucosamine	E3 γ-Hydroxybutyric Acid	E4 L-Threonine	E1 α-Cyclodex			
F1 Glycogen	F2 D-Glucosaminic Acid	F3 Itaconic Acid	F4 Glycyl-L-Glutamic Acid	F1 Glycogen	F2 D-Glucosaminic Acid	F3 Itaconic Acid	F4 Glycyl-L-Glutamic Acid	F1 Glycogen			
G1 D-Cellobiose	G2 Glucose-1-Phosphate	G3 α-Ketobutyric Acid	G4 Phenylethyl-amine	G1 D-Cellobiose	G2 Glucose-1-Phosphate	G3 α-Ketobutyric Acid	G4 Phenylethyl-amine	G1 D-Cellobi			
H1 α-D-Lactose	H2 D,L-α-Glycerol Phosphate	H3 D-Malic Acid	H4 Putrescine	H1 α-D-Lactose	H2 D,L-α-Glycerol Phosphate	H3 D-Malic Acid	H4 Putrescine	H1 α-D-Lacto			

FIGURE 1. Carbon Sources in EcoPlate



3. Direct physiological measurement

c. enzyme assays

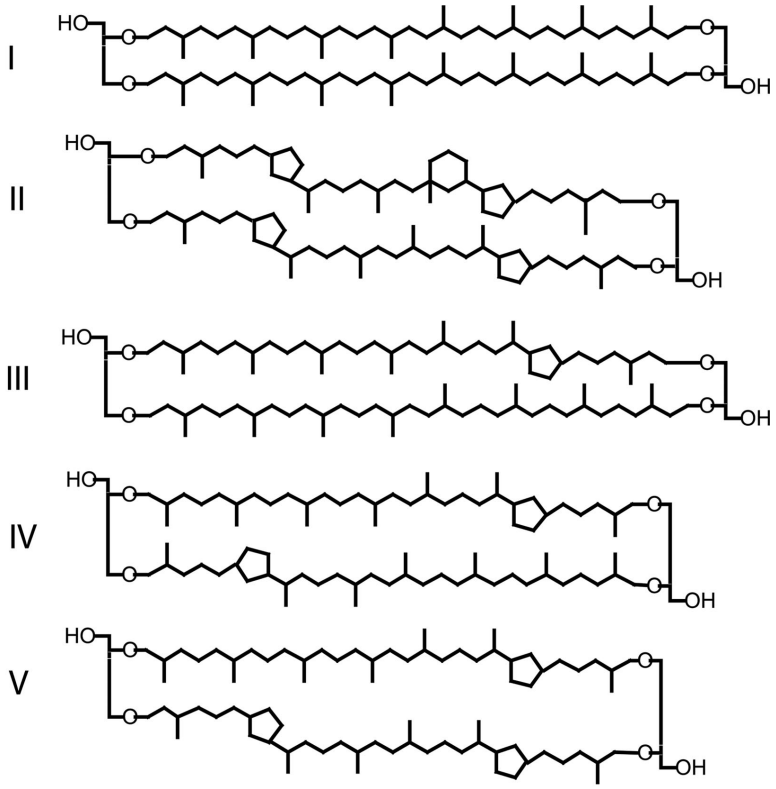


Colorimetric assay for cellulase activity on plant fungi

http://walkerlab.bee.cornell.edu/Enzymes.html

4. Indirect physiological measurement

a. biomolecule isotopic signatures



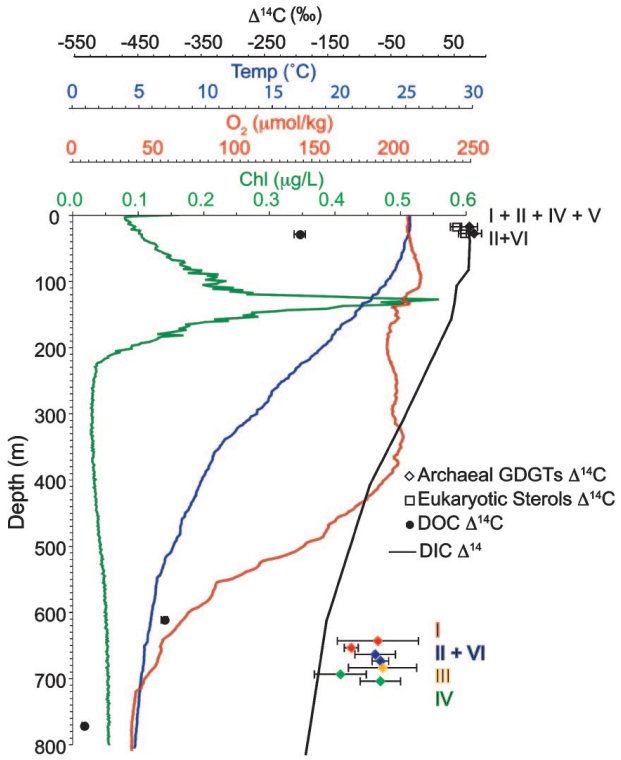
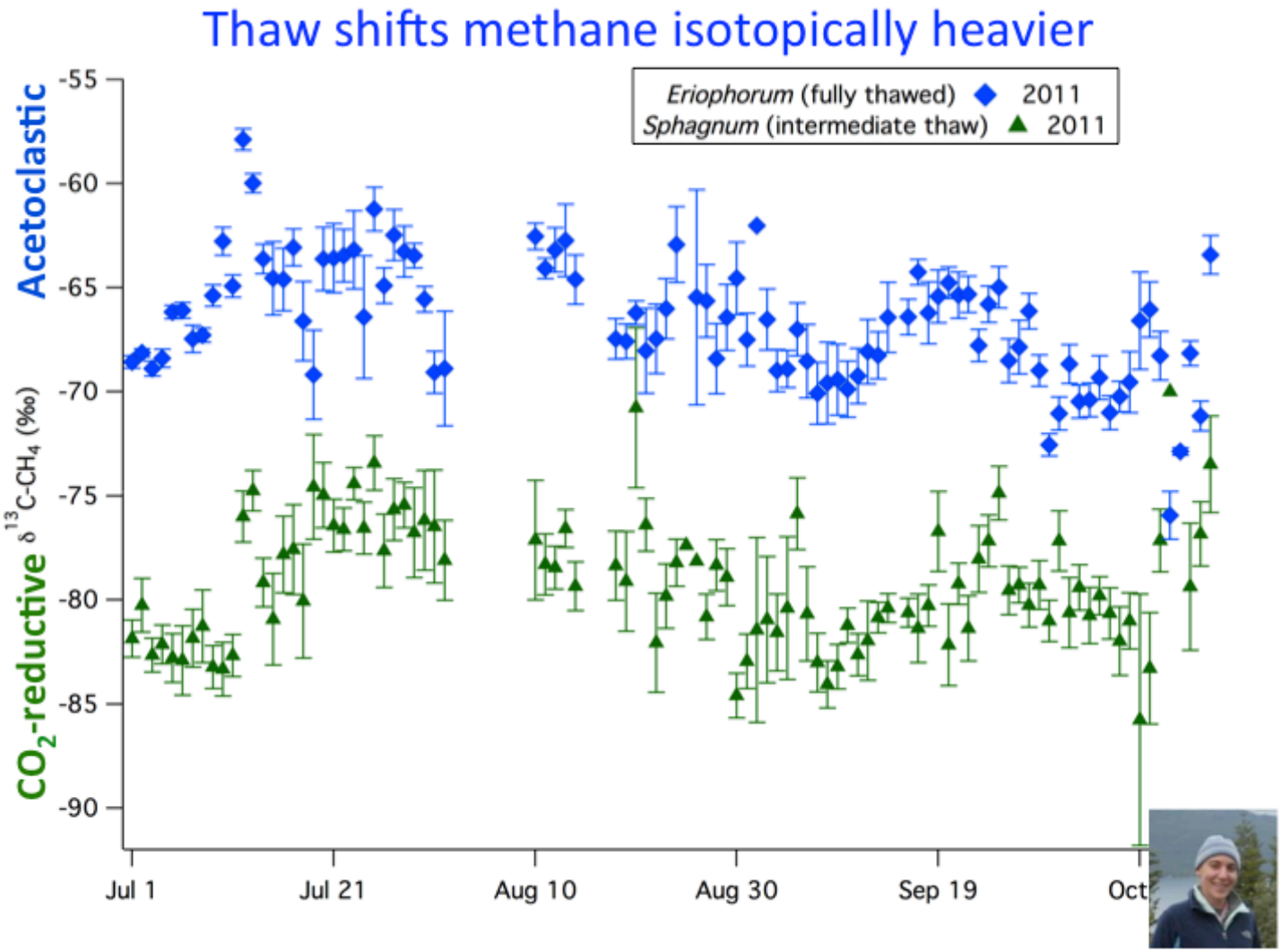


Fig. 2. Water column properties and $\Delta^{14}C$ values for DIC, DOC, sterols, and GDGTs. Chlorophyll, temperature, and dissolved oxygen data are from the Hawaii Ocean Time Series (HOTS) public data collected on May 19, 2004. DOC and DIC $\Delta^{14}C$ data are from refs. 24 and 46. Data points for individual compounds at 670 m have been separated for clarity.

Ingalls et al 2006

4. Indirect physiological measurement

b. emitted product isotopic signatures



McCalley et al submitted

5. Modeled linkages

Moira!

A way *I* think about *some* of the methods involved in connecting them:

1) Indirect molecular inference:
    a. 16S -> functional guilds
    b. gene ecology (of known functional genes)
    c. transcriptional and translational activity (of known functional genes)

2) direct molecular inference:
    a. SIP

3) direct physiological measurement:
    a. culture based charactrization
    b. Ecolog plate substrate utilization
    c. enzyme assays

4) indirect physiologial measurement:
    a. biomolecule isotopic signatures
    b. emitted product isotopic signatures

5) modeled linkages - Moira Hough will tell us about this