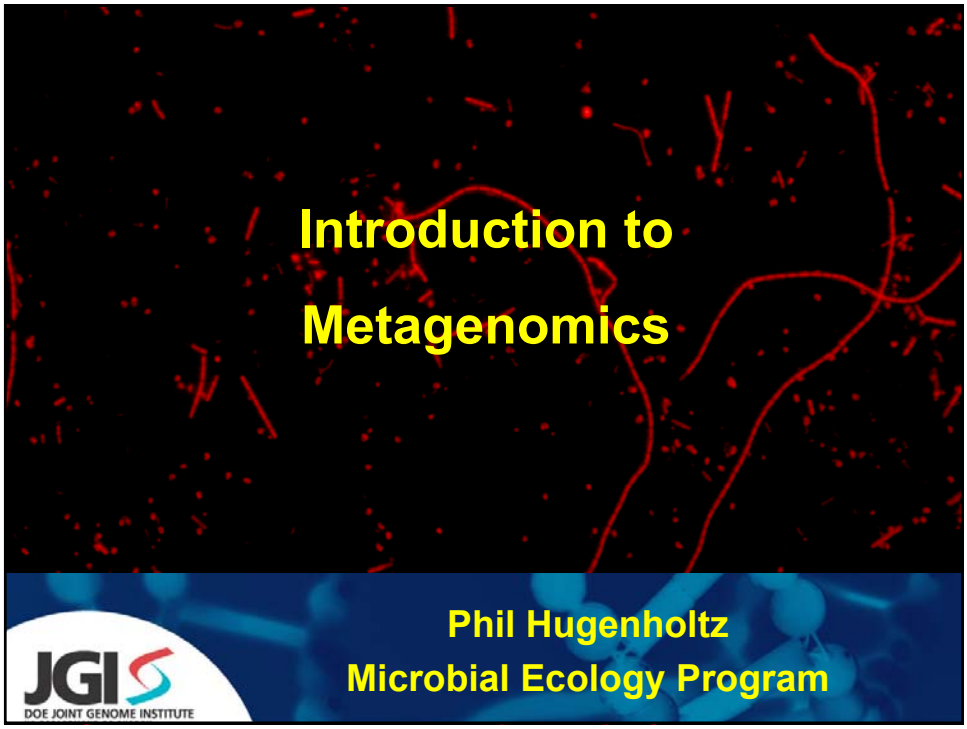
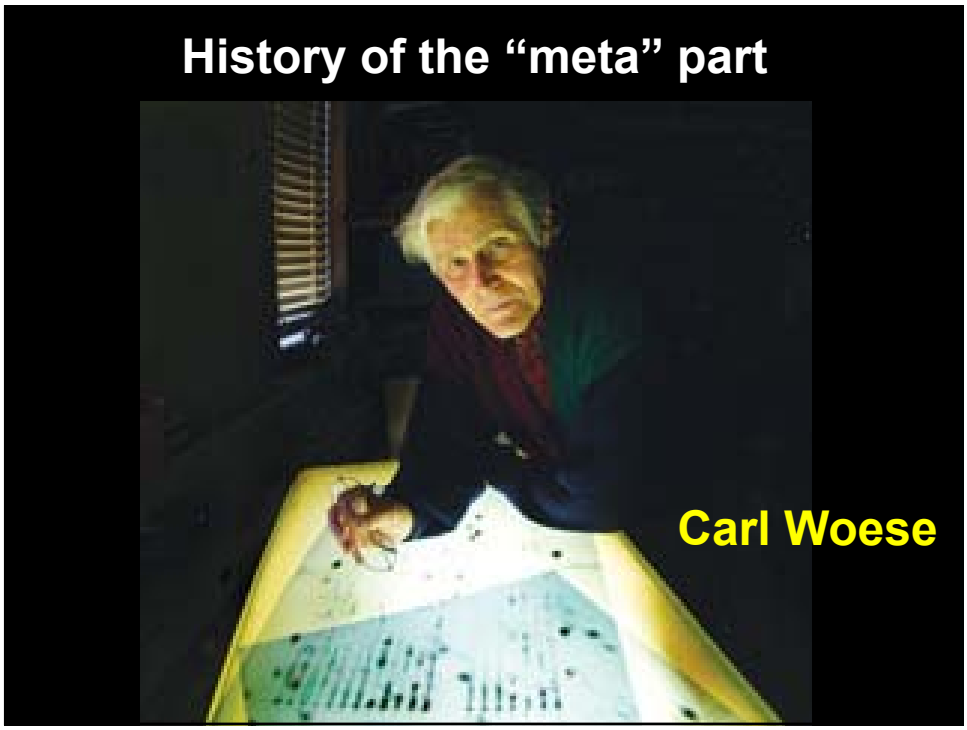
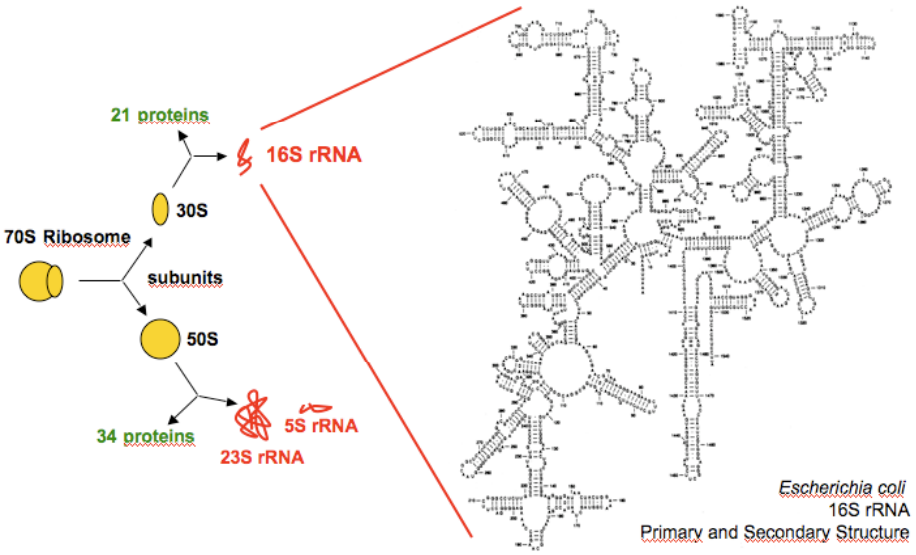
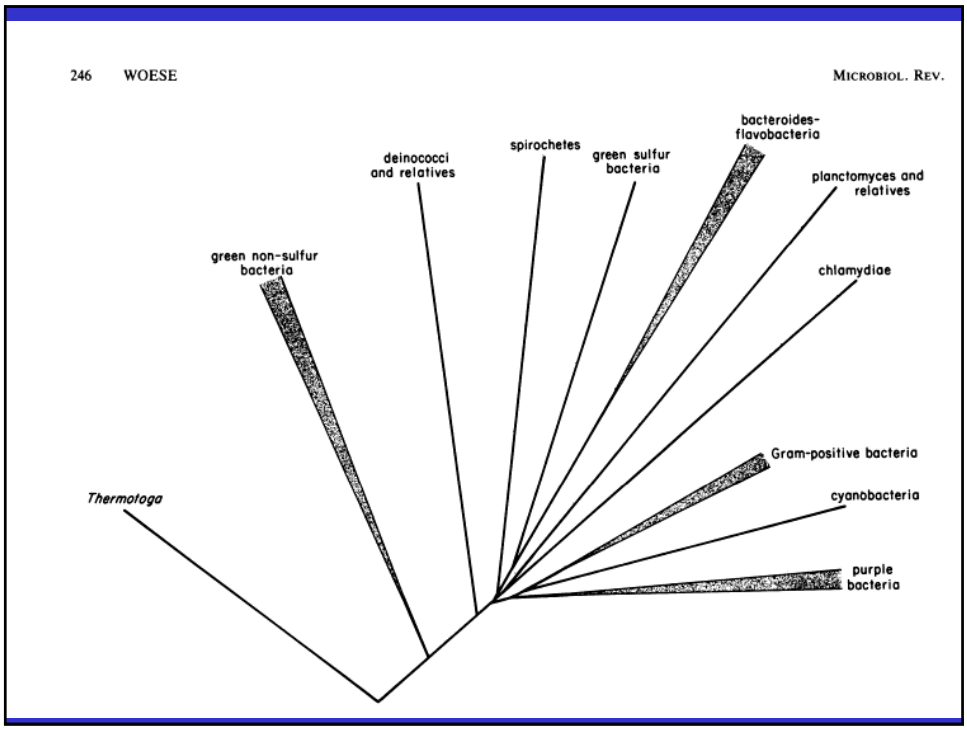

Introduction to
Metagenomics
Phil Hugenholtz
Microbial Ecology Program
JGI
DOE JOINT GENOME INSTITUTE


History of the "meta" part
Carl Woese

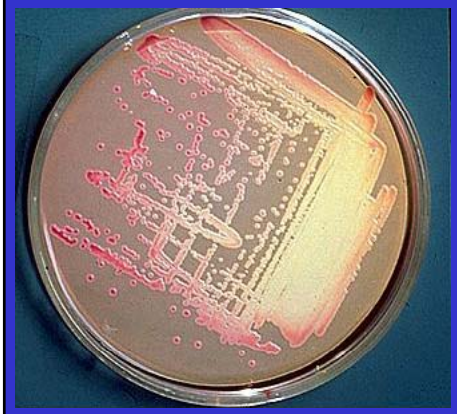
16S ribosomal RNA as a phylogenetic marker



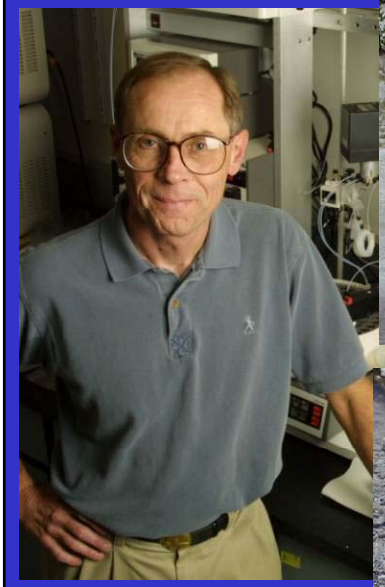
Falk Warnecke



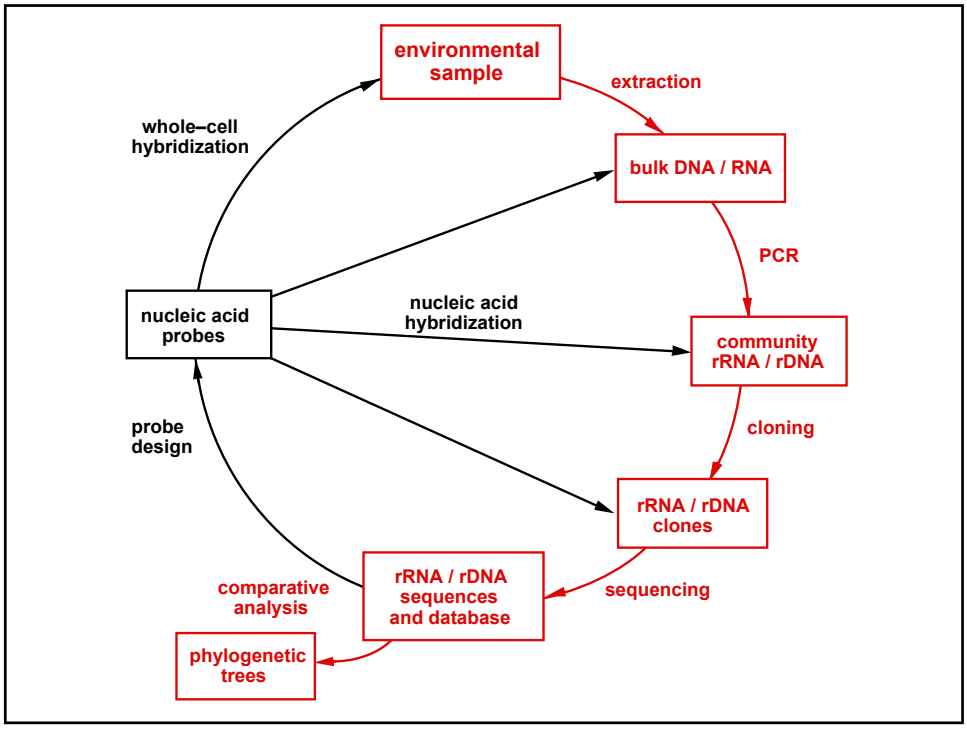
tame ≠ **wild**

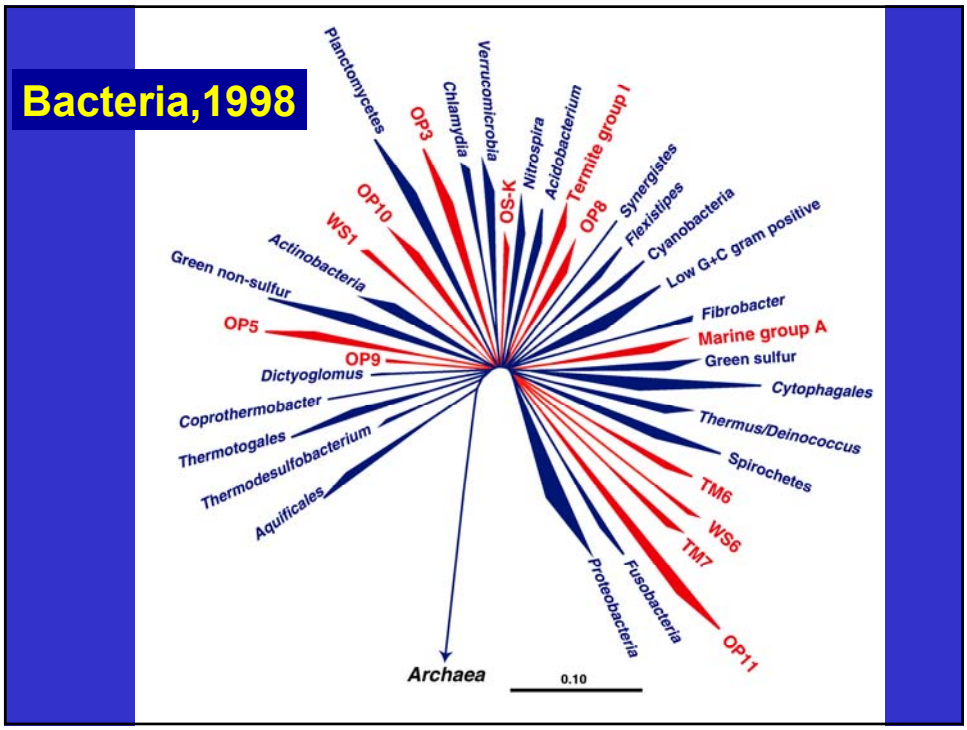


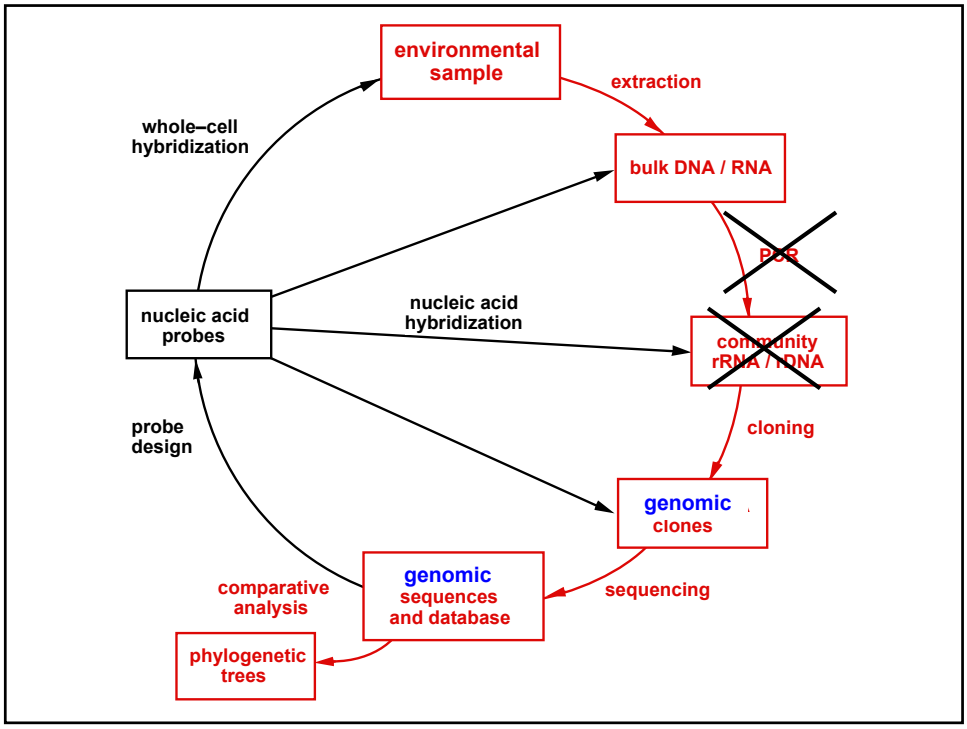
most bacteria don't grow on plates
"the great plate count anomaly"

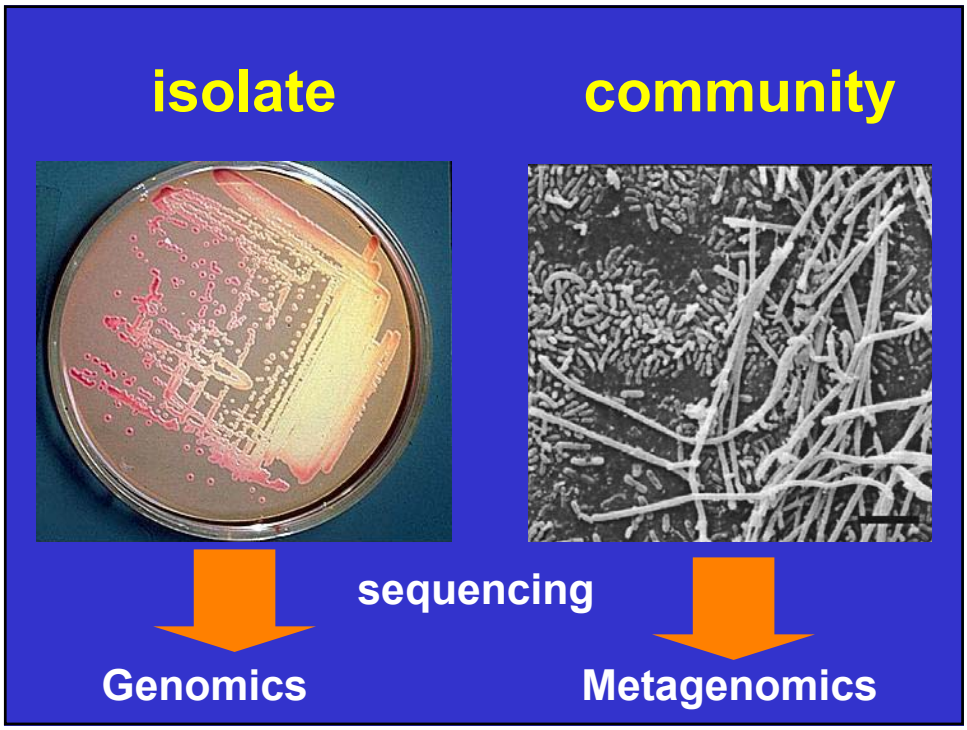


Norm Pace


environmental sample
extraction
bulk DNA / RNA
PCR
community rRNA / rDNA
cloning
rRNA / rDNA clones
sequencing
rRNA / rDNA sequences and database
comparative analysis
phylogenetic trees
probe design
nucleic acid probes
nucleic acid hybridization
whole–cell hybridization

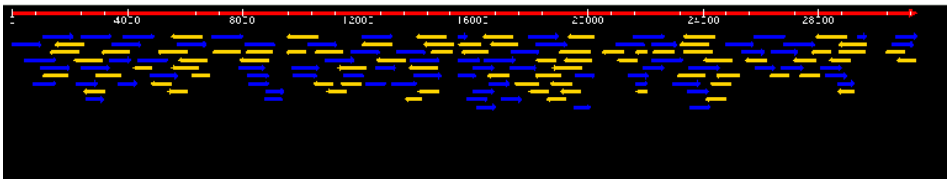

Bacteria,1998
Planctomycetes
OP3
Chlamydia
Verrucomicrobia
OS-K
Nitrospira
Acidobacterium
Termite group I
OP8
Synergistes
Flexistipes
Cyanobacteria
Low G+C gram positive
Fibrobacter
Marine group A
Green sulfur
Cytophagales
Thermus/Deinococcus
Spirochetes
TM6
WS6
TM7
OP11
Fusobacteria
Proteobacteria
Archaea
0.10
OP10
WS1
Actinobacteria
Green non-sulfur
OP5
OP9
Dictyoglomus
Coprothermobacter
Thermotogales
Thermodesulfobacterium
Aquificales


environmental sample
extraction
whole–cell hybridization
bulk DNA / RNA
PCR
nucleic acid probes
nucleic acid hybridization
community rRNA / rDNA
cloning
probe design
genomic clones
genomic sequences and database
sequencing
comparative analysis
phylogenetic trees


isolate
community
sequencing
Genomics
Metagenomics

The catch…. it comes in small pieces

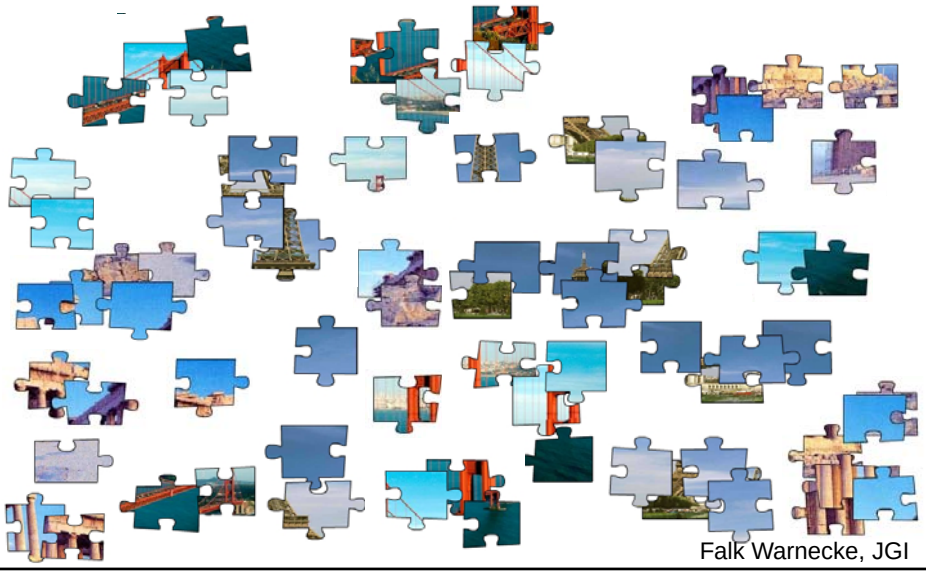
Average Sanger read length – 750 bases



Must assemble the reads together

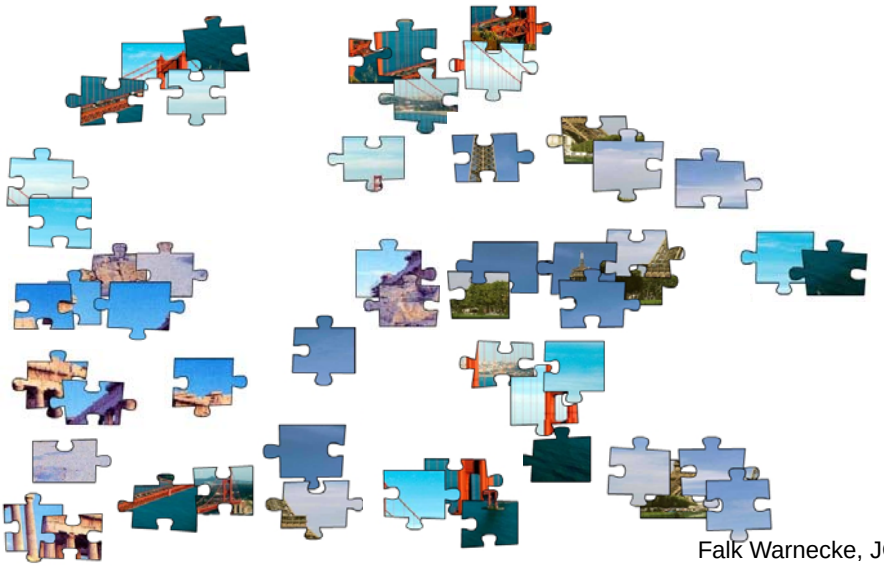
Giant jigsaw puzzles

Metagenome assembly – like putting together several jigsaw puzzles



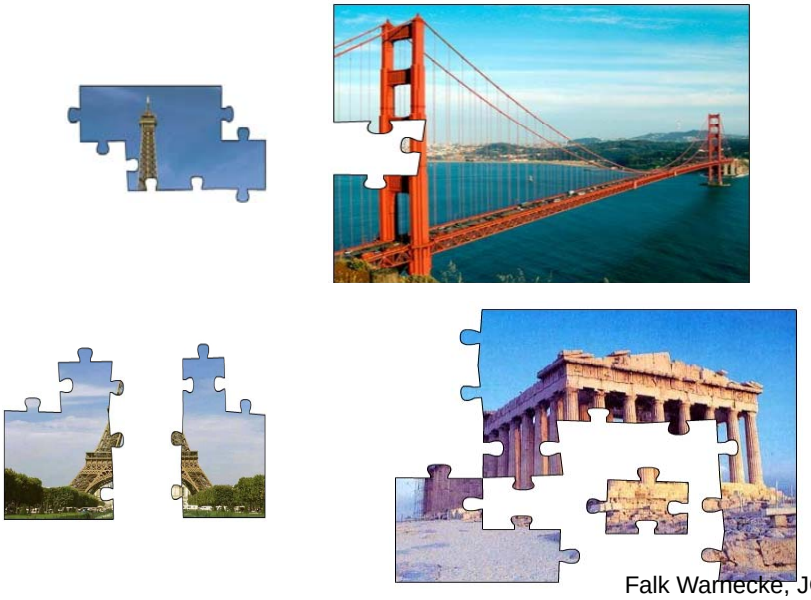
Falk Warnecke, JGI

. . . with some pieces missing



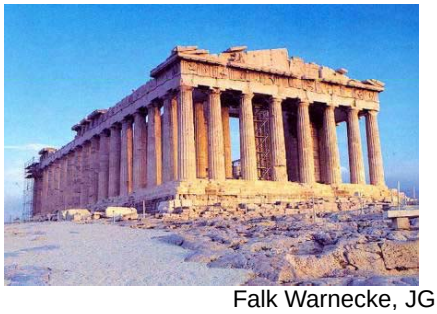
Falk Warnecke, JGI

Can we still reconstruct?



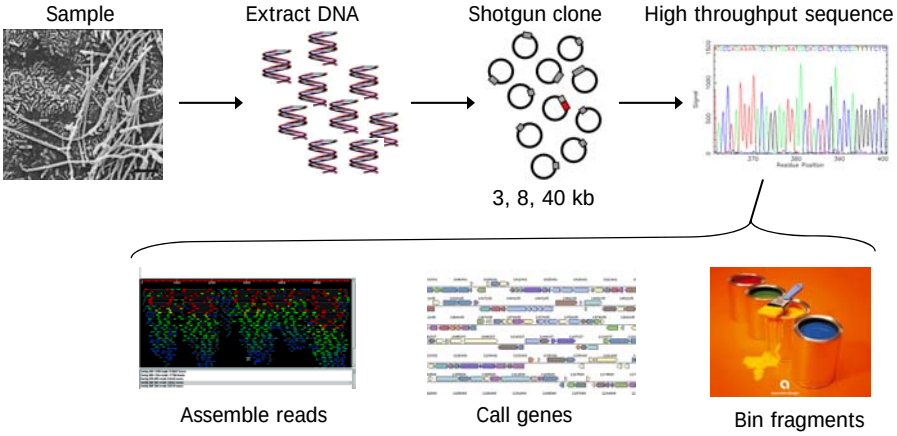
Falk Warnecke, JGI

Can we still reconstruct?



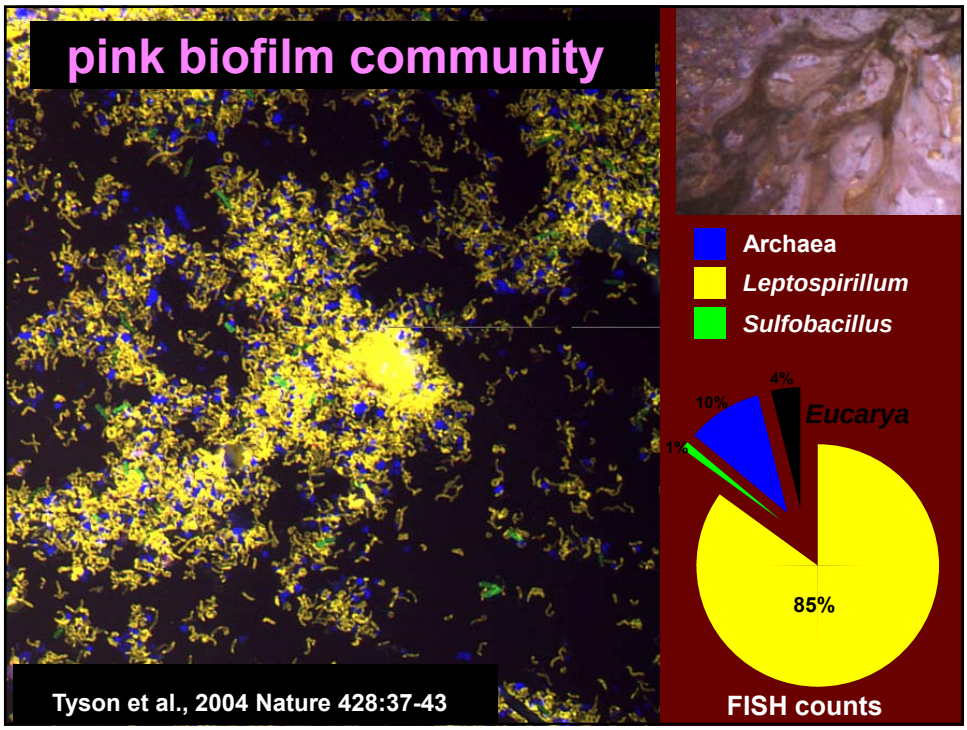
Falk Warnecke, JGI

Making metagenomes



Will this really work??

- need to test whole genome shot-gun sequencing on a very simple community to see if it will work!



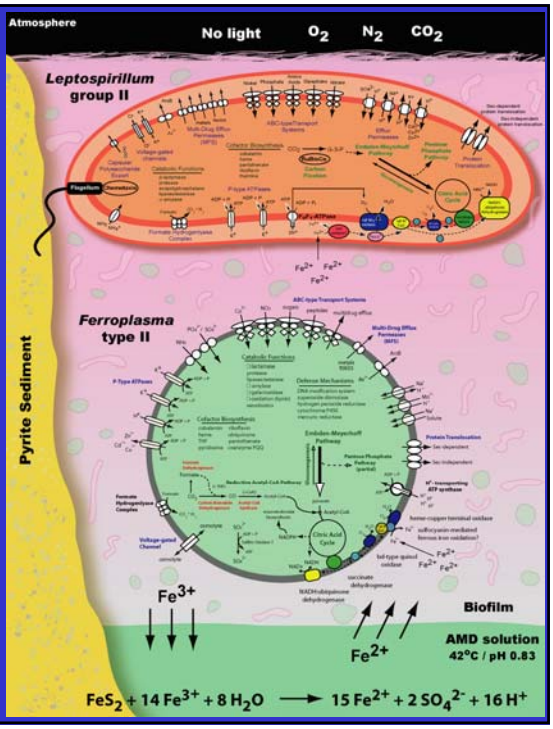
Metabolism

Expected

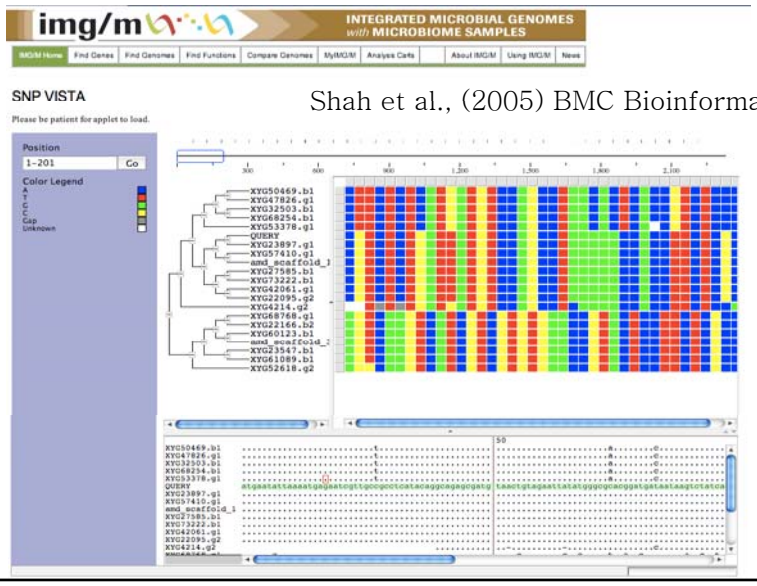
- proton pumps for pH
- metal resistance genes
- novel cytochromes for Fe^{2+} oxidation
- chemotaxis genes for Fe^{2+} and O_2
- genes for CO_2 and N_2 fixation*
- lots of amino acid and sugar transporters in the archaea

Unexpected

- *only one set of *nif* genes
- DNA photolyase genes
- cellulose synthase in *Leptospirillum* - biofilm bouyancy



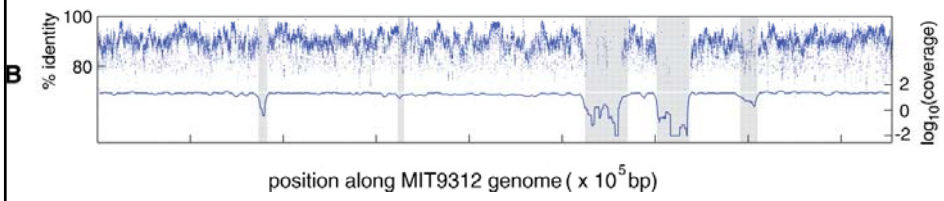
Fine-scale population structure



Shah et al., (2005) BMC Bioinformatics 6:29

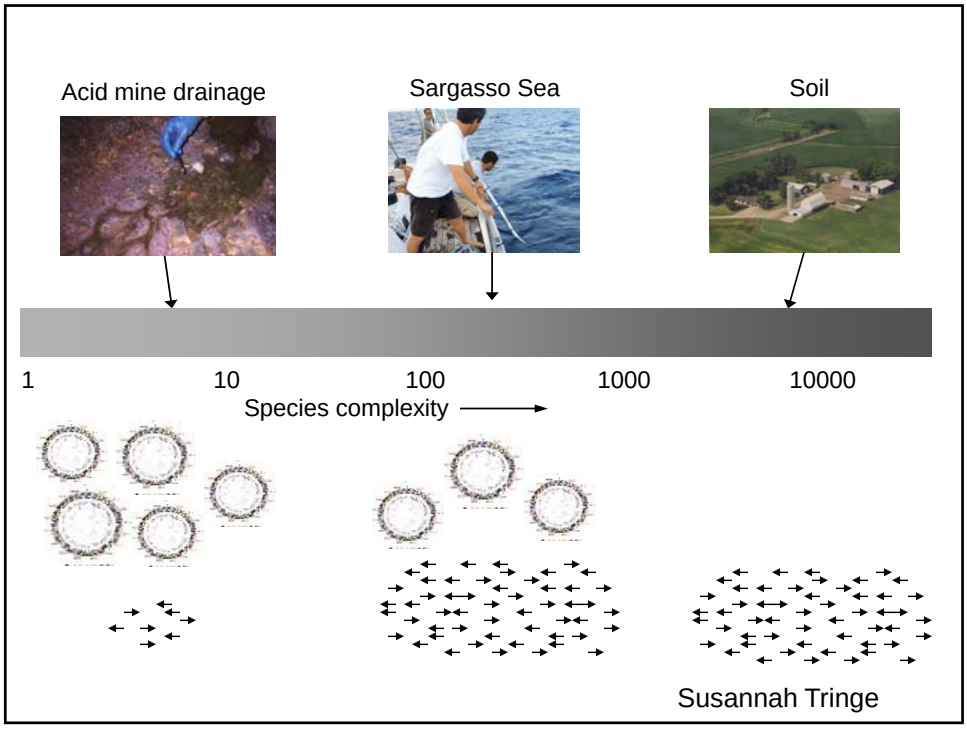
Gross-scale population structure

- align metagenomic reads against reference genomes or genome fragments
- color-code to enhance visualization
- recently implemented in IMG/M

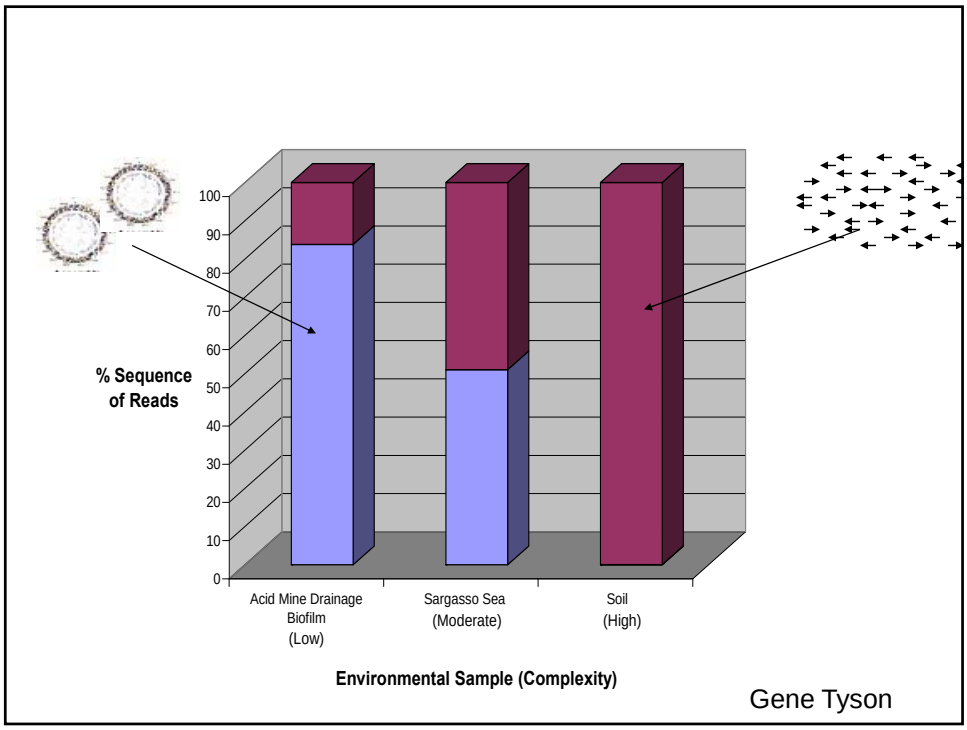


from Fig. 2 Coleman et al. (2006) Science 311(5768):17

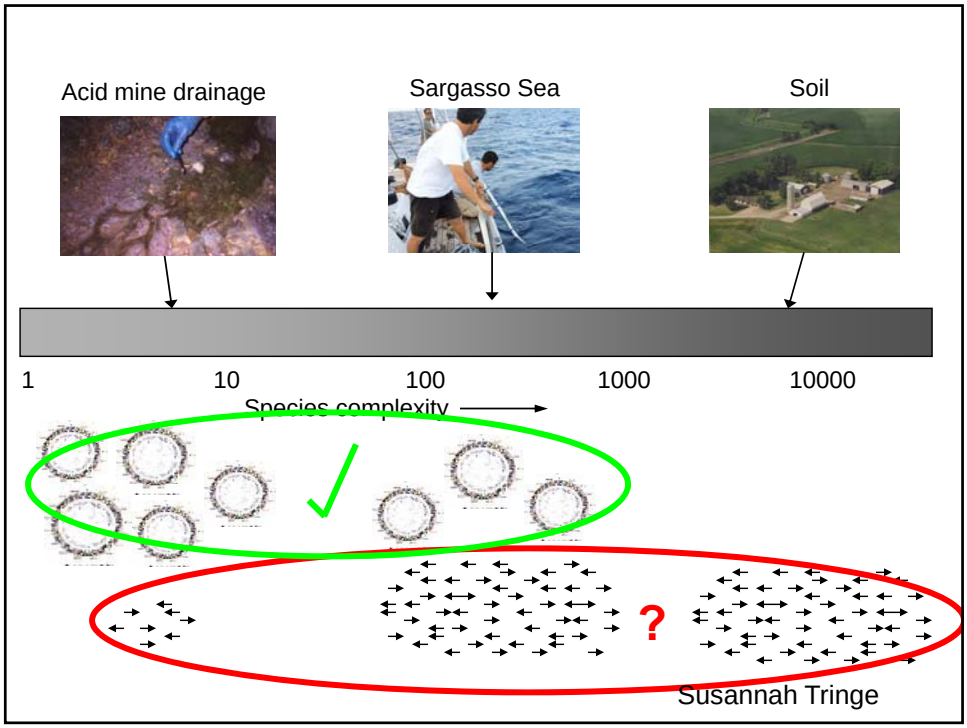
- **does it work with more complex communities?**



Susannah Tringe



Gene Tyson

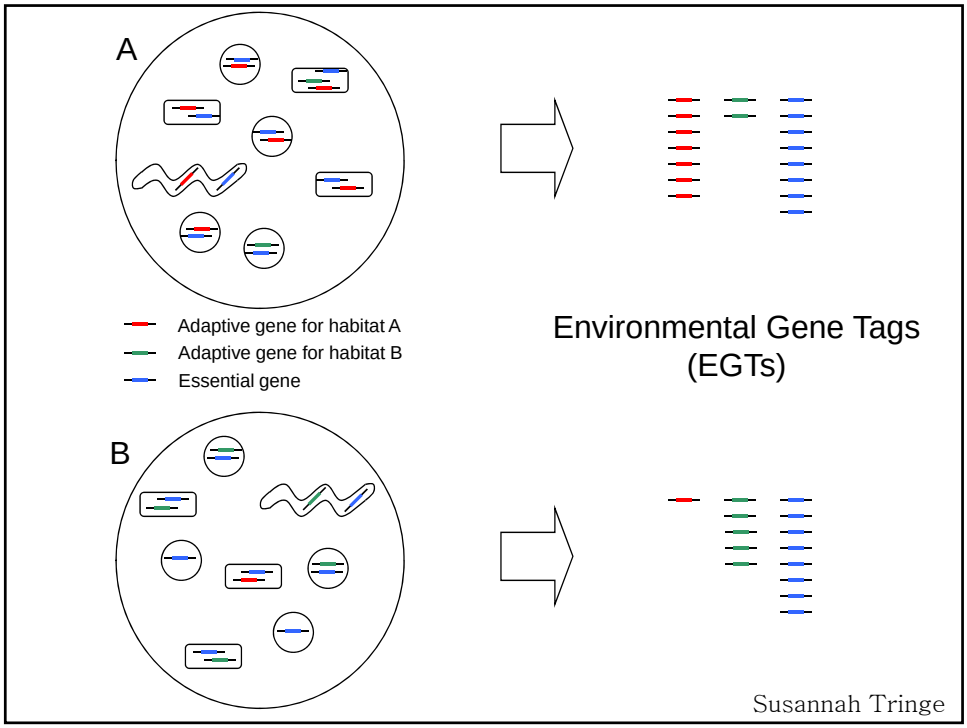


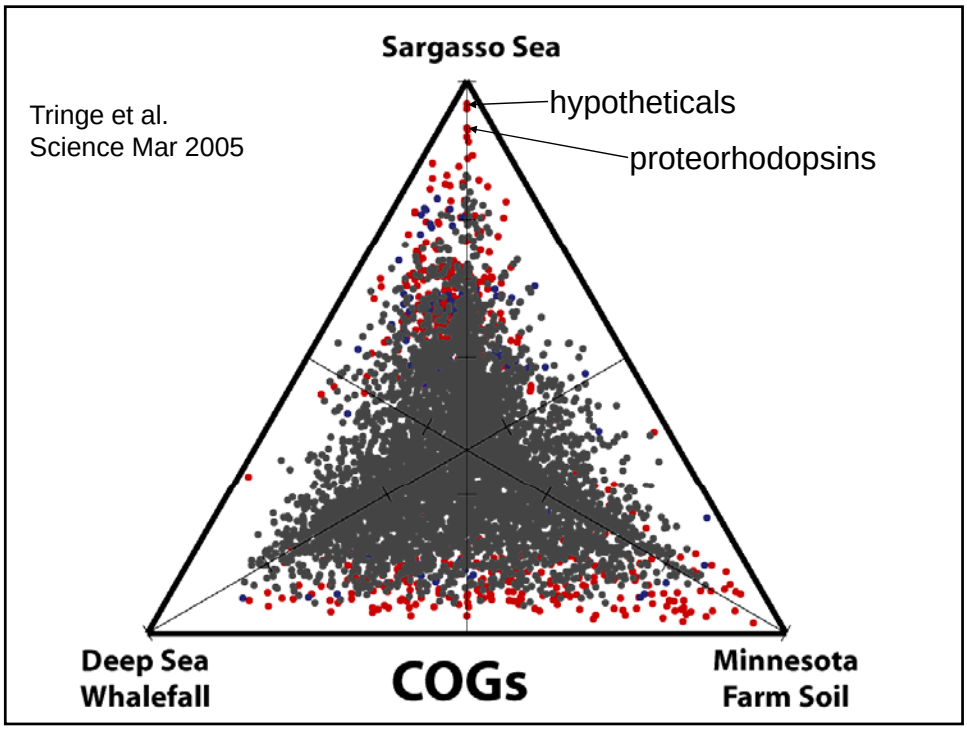
Environmental Gene Tags (EGT)

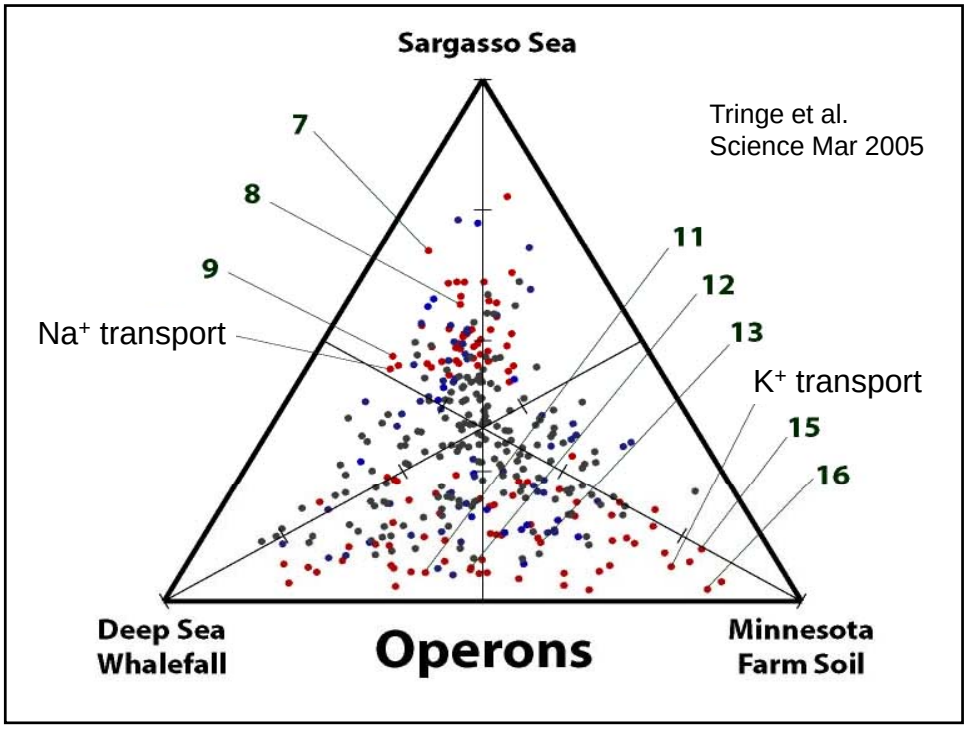
Identify genes in sequence data (contigs to unassembled reads) from *multiple* environmental samples

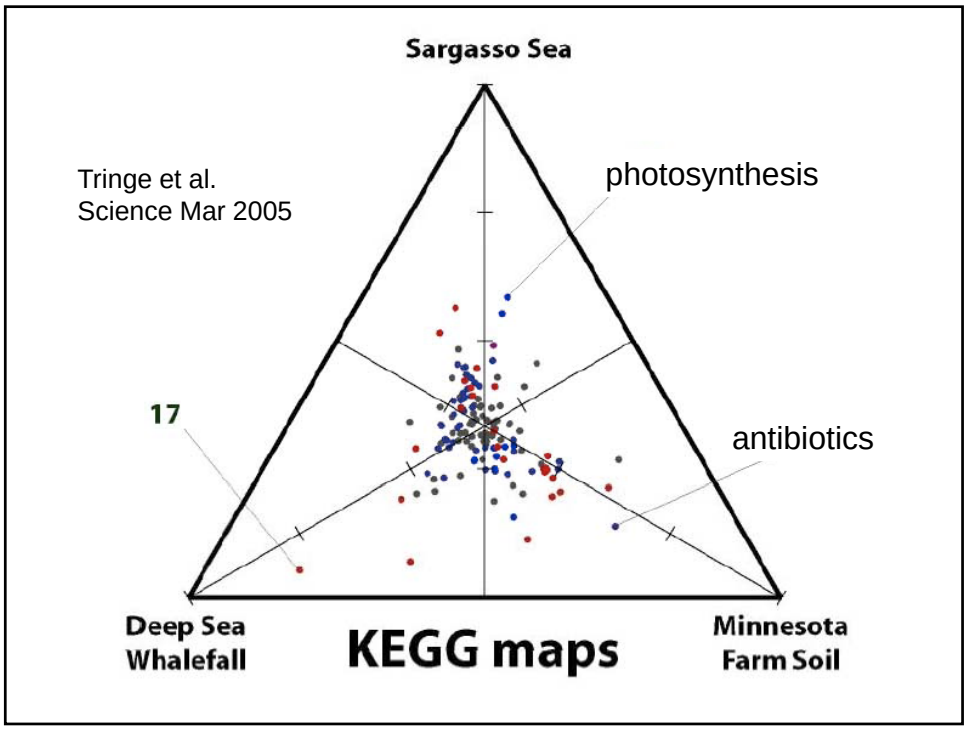
Assign genes to their gene family, or higher level groupings

Compare relative abundance of different gene families according to habitat

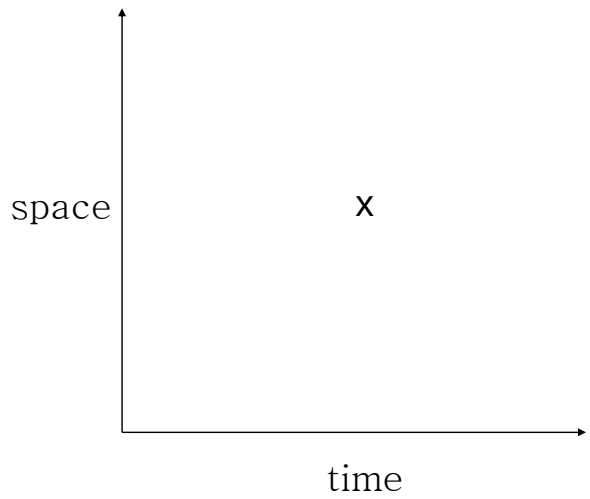

A
Adaptive gene for habitat A
Adaptive gene for habitat B
Essential gene
Environmental Gene Tags
(EGTs)
B
Susannah Tringe


Tringe et al.
Science Mar 2005
Sargasso Sea
hypotheticals
proteorhodopsins
Deep Sea
Whalefall
COGs
Minnesota
Farm Soil


Sargasso Sea
Tringe et al.
Science Mar 2005
7
8
9
Na+ transport
11
12
13
K+ transport
15
16
Deep Sea
Whalefall
Operons
Minnesota
Farm Soil


Sargasso Sea
Tringe et al.
Science Mar 2005
photosynthesis
17
antibiotics
Deep Sea
Whalefall
KEGG maps
Minnesota
Farm Soil

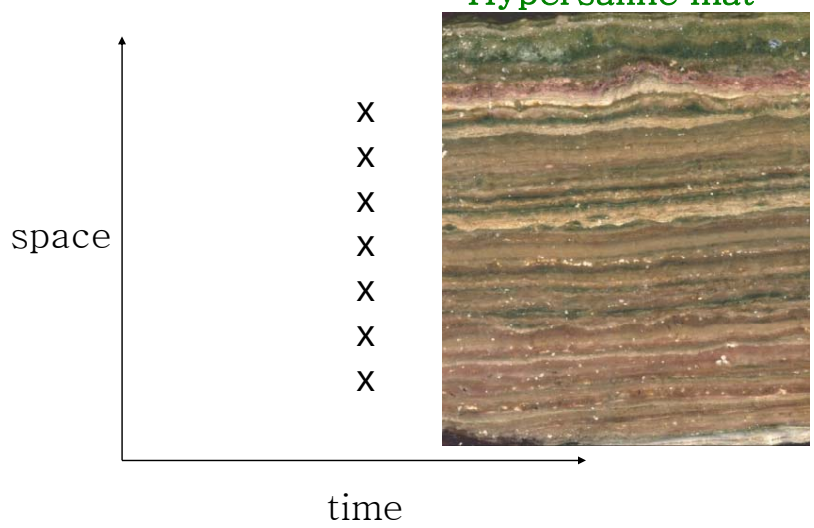
Snapshot studies



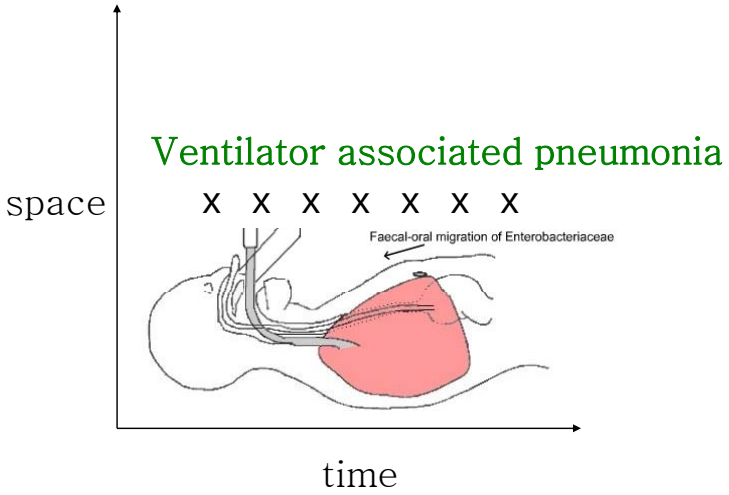
~90% of all current and pending metagenomic projects

Spatial series studies

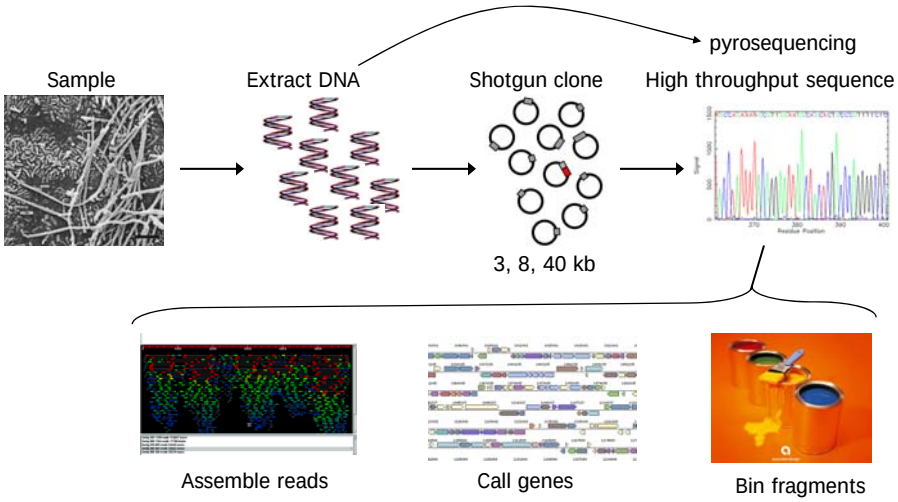
Hypersaline mat

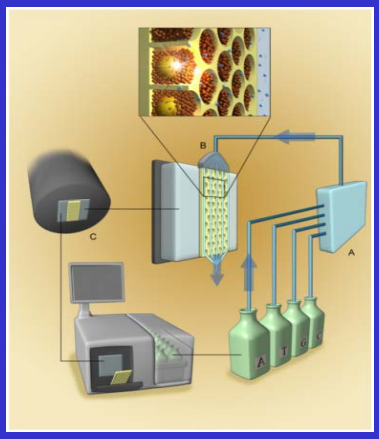


Temporal series studies

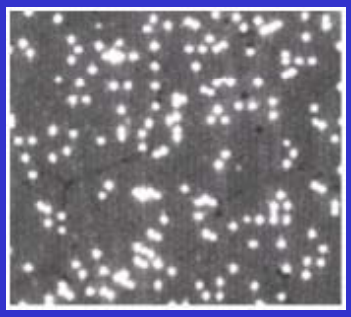


New sequencing technologies

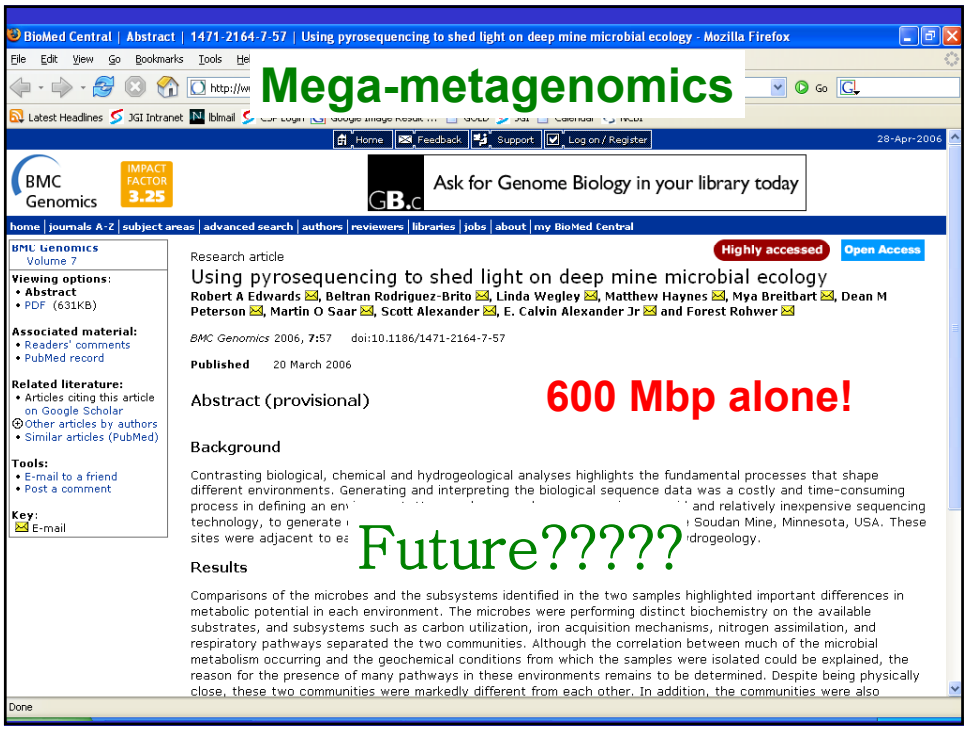




454 pyrosequencer



Sanger	GS20	FLX	XLR
0.07 Mbp	35 Mbp	100 Mbp	400 Mbp per run
700 bp	100 bp	200 bp	400 bp reads
0.1 ¢	0.03 ¢	0.01¢	0.003 ¢ per base



BioMed Central | Abstract | 1471-2164-7-57 | Using pyrosequencing to shed light on deep mine microbial ecology - Mozilla Firefox

Mega-metagenomics

BMC Genomics IMPACT FACTOR 3.25

Ask for Genome Biology in your library today

Research article

Highly accessed | Open Access

Using pyrosequencing to shed light on deep mine microbial ecology

Robert A Edwards, Beltran Rodriguez-Brito, Linda Wegley, Matthew Haynes, Mya Breitbart, Dean M Peterson, Martin O Saar, Scott Alexander, E. Calvin Alexander Jr and Forest Rohwer

BMC Genomics 2006, **7**:57 doi:10.1186/1471-2164-7-57

Published 20 March 2006

Abstract (provisional)

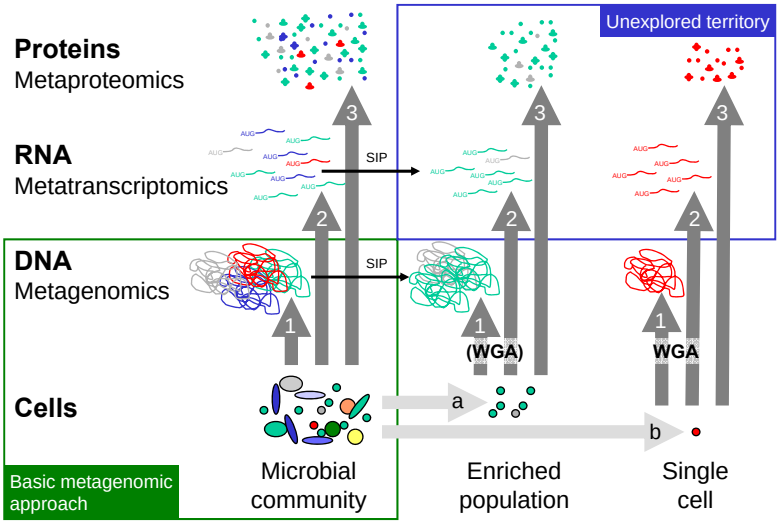
600 Mbp alone!

Background

Contrasting biological, chemical and hydrogeological analyses highlights the fundamental processes that shape different environments. Generating and interpreting the biological sequence data was a costly and time-consuming process in defining an en ... and relatively inexpensive sequencing technology, to generate ... Soudan Mine, Minnesota, USA. These sites were adjacent to ea ... drogeology.

Future?????

Results

Comparisons of the microbes and the subsystems identified in the two samples highlighted important differences in metabolic potential in each environment. The microbes were performing distinct biochemistry on the available substrates, and subsystems such as carbon utilization, iron acquisition mechanisms, nitrogen assimilation, and respiratory pathways separated the two communities. Although the correlation between much of the microbial metabolism occurring and the geochemical conditions from which the samples were isolated could be explained, the reason for the presence of many pathways in these environments remains to be determined. Despite being physically close, these two communities were markedly different from each other. In addition, the communities were also

Enhancing "basic" metagenomics