The Emerging Global Community of Microbial Metagenomics Researchers

Opening Talk
Metagenomics 2007
Calit2@UCSD
July 11, 2007

Dr. Larry Smarr
Director, California Institute for Telecommunications and Information Technology
Harry E. Gruber Professor,
Dept. of Computer Science and Engineering
Jacobs School of Engineering, UCSD

UCSD
UCIrvine

it²

Abstract

Calit2, the J. Craig Venter Institute, and UCSD's SDSC and Scripps Institution of Oceanography, is creating a metagenomic Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis (CAMERA), funded by the Gordon and Betty Moore Foundation. The CAMERA computational and storage cluster, which contains multiple ocean microbial metagenomic datasets, as well as the full genomes of ~166 marine microbes, is actively in use. End users can access the metagenomic data either via the web or over novel dedicated 10 Gb/s light paths (termed "lambdas") through the National LambdaRail. The end user clusters are reconfigured as "OptIPortals," providing the end user with local scalable visualization, computing, and storage. Currently over 1000 users from over 40 countries are CAMERA registered users, with over a dozen remote OptIPortal sites becoming active. This CAMERA connected community sets the stage for creating a software system to support a social network of metagenomic researchers--a "MySpace" for scientists. We look forward to gathering ideas from Metagenomics 2007 participants for the functional requirements of such a system.

UCSD
UCIrvine

Calit2 Brings Computer Scientists and Engineers Together with Biomedical Researchers

- **Some Areas of Concentration:**
  - Algorithmic and System Biology
  - Bioinformatics
  - Metagenomics
  - Cancer Genomics
  - Human Genomic Variation and Disease
  - Proteomics
  - Mitochondrial Evolution
  - Computational Biology
  - Multi-Scale Cellular Imaging
  - Information Theory and Biology
  - Telemedicine

CASB

INSTITUTE FOR GENOMICS AND BIOINFORMATICS
University of California, Irvine

NanoTumor Center

camera
Marine Microbial Ecology

ICSI

JCSG
Joint Center for Structural Genomics

NBCR
National Biomedical Computation Resource
an NIH supported resource center

NCMIR
NATIONAL CENTER for MICROSCOPY and IMAGING RESEARCH
at San Diego, an NIH supported resource center

BIOMEDICAL INFORMATICS RESEARCH NETWORK

Information Theory and Applications

WISARD

UCSD
UCIrvine

Southern California Telemedicine Learning Center (TLC)

it²

CAMERA: Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis



Paul Gilna Ex. Dir.

Philip Papadopoulos, SDSC/Calit2
2pm Friday

National LambdaRail
Direct Connect
Computation and Storage Complex

PI Larry Smarr

Funded by: Gordon and Betty Moore Foundation

Joint Partnership of:

camera
Marine Microbial Ecology

Announced January 17, 2006
$24.5M Over Seven Years

CAMERA 1.1 is Up and Running!

camera

CONTACT US LOG IN Search

HOME ABOUT CAMERA METAGENOMICS RESEARCH NEWS EVENTS DISCUSSION FORUMS



PLoS Biology Ocean Metagenomics Collection



New CAMERA production server, 512-CPU, 5-Teraflops, 200-Terabyte cluster

WELCOME TO CAMERA!

CAMERA stands for Community Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis. The aim of this project is to serve the needs of the microbial ecology research community by creating a rich, distinctive data repository and a bioinformatics tools resource that will address many of the unique challenges of metagenomic analysis.

WHAT DO YOU WANT TO DO?

- Register with Camera to access services (first time users only)
  - step-by-step registration tutorial
- Run Blast against the available datasets
- Download available datasets
- Ask questions or get help in the user forums
- Report a problem
- Suggest new features (via forums or email)
- Learn more about the Camera project

CURRENT STATUS

- Current application version - 1.1 (release notes) deployed

Funded by the Moore Foundation

MOORE FOUNDATION

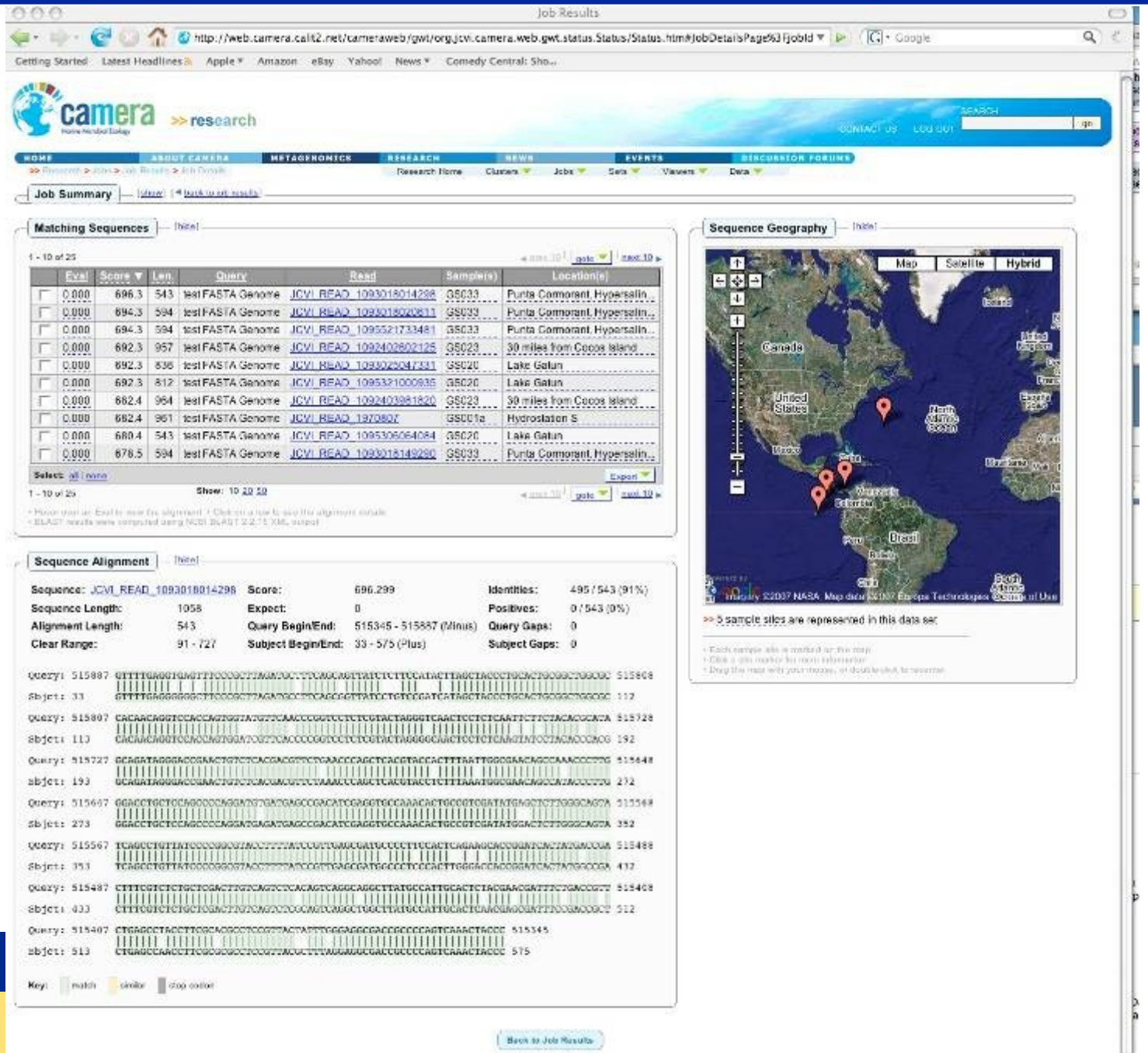
Upcoming Events

- Metagenomics 2007 July 11th, 2007 – Download Flyer
- Gordon Research Conference in Applied and Environmental Microbiology July 15th, 2007

More Events

Contact Us

CAMERA Combines Genomic and Metagenomic Tools

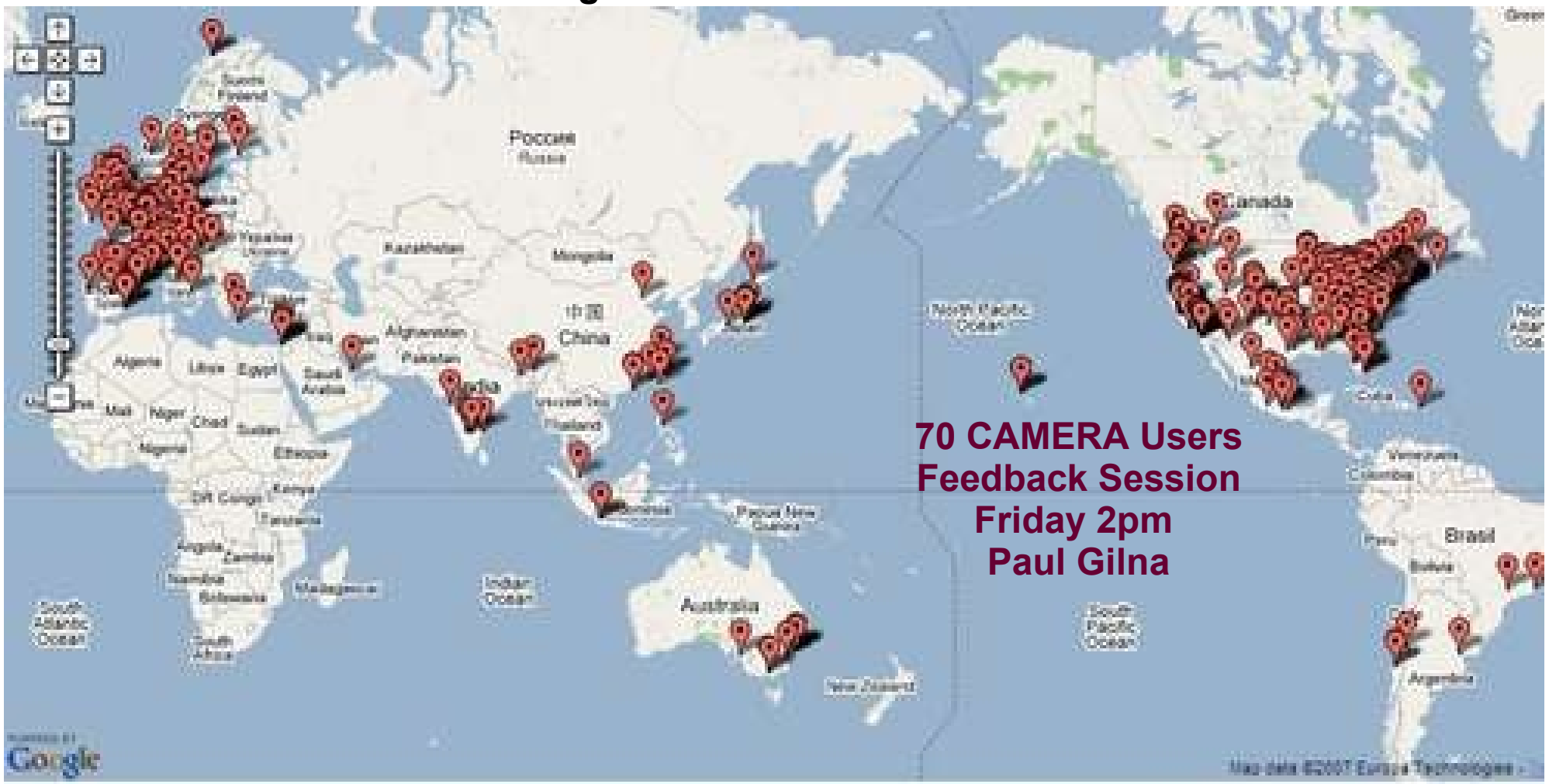


UCSD
UCIrvine

it²

Can We Create a "My Space" for Science Researchers? Microbial Metagenomics as a Cyber-Community

Over 1000 Registered Users From 45 Countries



70 CAMERA Users
Feedback Session
Friday 2pm
Paul Gilna

USA	583
United Kingdom	46
Canada	35
France	35
Germany	32

camera
Marine Microbial Ecology

it²



Calit2 is Prototyping Social Networks for Reseachers

Research Intelligence Project

- ri.calit2.net

Add in:

- MyProteins
- MyMicrobes
- MyEnvironments
- MyPapers
- MyGenomes

UCSD
UCIrvine

The Open Protein Structure Annotation Network

www.topsan.org

Introduction

The Joint Center for Structural Genomics (JCSG) is one of the four large-scale structural genomics centers funded by the U.S. National Institutes of Health (NIH) as part of the Protein Structure Initiative (PSI). PSI centers have determined high-resolution structures of several novel, previously uncharacterized protein families for which associated biochemical and biological research are not available. Consequently, the vast majority of these structures are not described in high impact, peer-reviewed manuscripts. Here, we propose TOPSAN, a radically novel way to collect, share and distribute information about protein three-dimensional structures.

JCSG technology pipeline







Annotating data from HT centers

Annotation:
- Live feeds from public databases
- Expert curated information (collective intelligence)

Content management:
- Wiki-style editing
- Page-level access control
- Online WYSIWYG editor
- Email notification
- Structured fields + free text

Quality control & authorship:
- Encourage community collaboration
- JCSG scientists & invited peers
- Many authors - No contribution too small
- Lead authors (editors) in charge of releases

Publication tools:
- Instant publication of "development" versions
- Releases of polished editions of the articles
- Open for comments at any time

Structural Genomics of *Thermotoga maritima*



Perspectives

TOPSAN is a pilot project towards creating an open annotation platform that will enable us to collect, share and distribute information about proteins in the form of expert-curated annotations combined with information obtained from a network of collaborating resources, enabling us to understand the function and roles of these proteins in their respective organisms.

Acknowledgments

TOPSAN is a members of the JCSG
Grant sponsor: National Institutes of Health, Protein Structure Initiative
Grant numbers: P50 GM64598 (and other NIH)

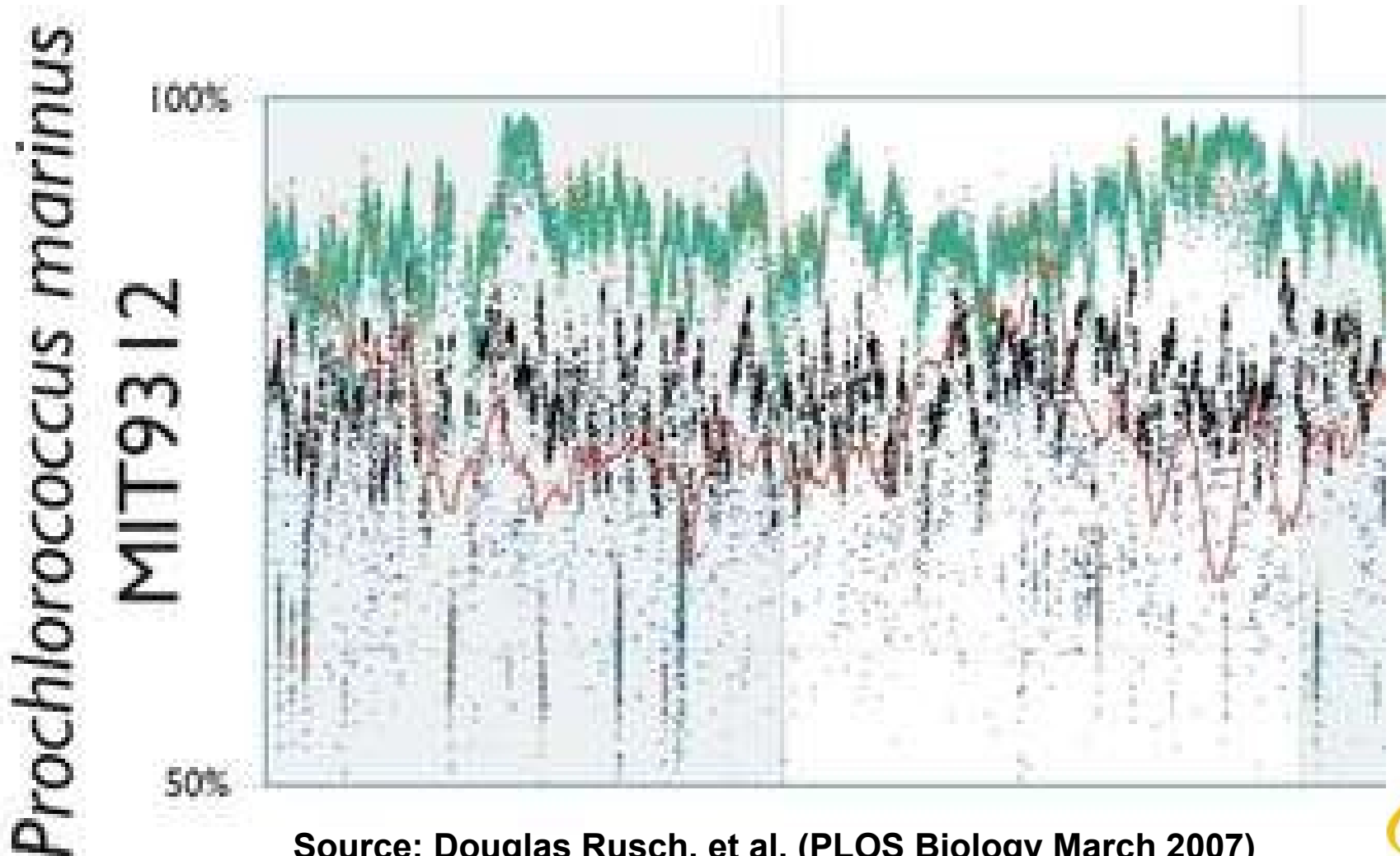
Emerging Capabilities That Tie Together Metagenomics Researchers

- Advanced Computing Techniques
- Broad Coverage of Complete Microbe Genomes
  - Moore Foundation
  - DOE JGI
- Proteomics of Microbes
- Cellular Network Models

UCSD
UCIrvine

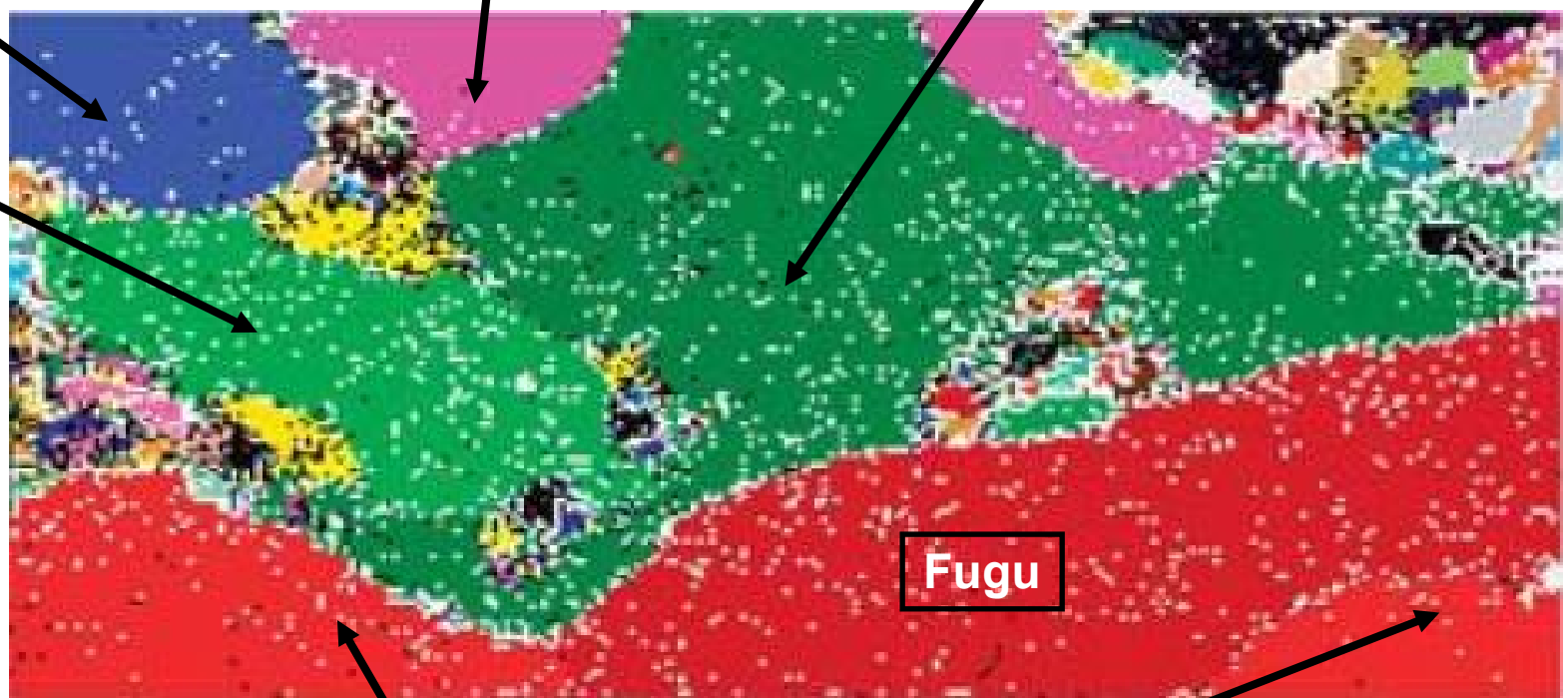
Metagenomic Challenge--Enormous Biodiversity: Very Little of GOS Metagenomic Data Assembles Well

- **Use Reference Genomes to Recruit Fragments**
  - **Compared 334 Finished and 250 Draft Microbial Genomes**
- **Only 5 Microbial Genera Yielded Substantial and Uniform Recruitment**
  - **Prochlorococcus, Synechococcus, Pelagibacter, Shewanella, and Burkholderia**



Source: Douglas Rusch, et al. (PLOS Biology March 2007)

Use of Self Organizing Maps to Identify Species Massive Computation on the Japanese Earth Simulator



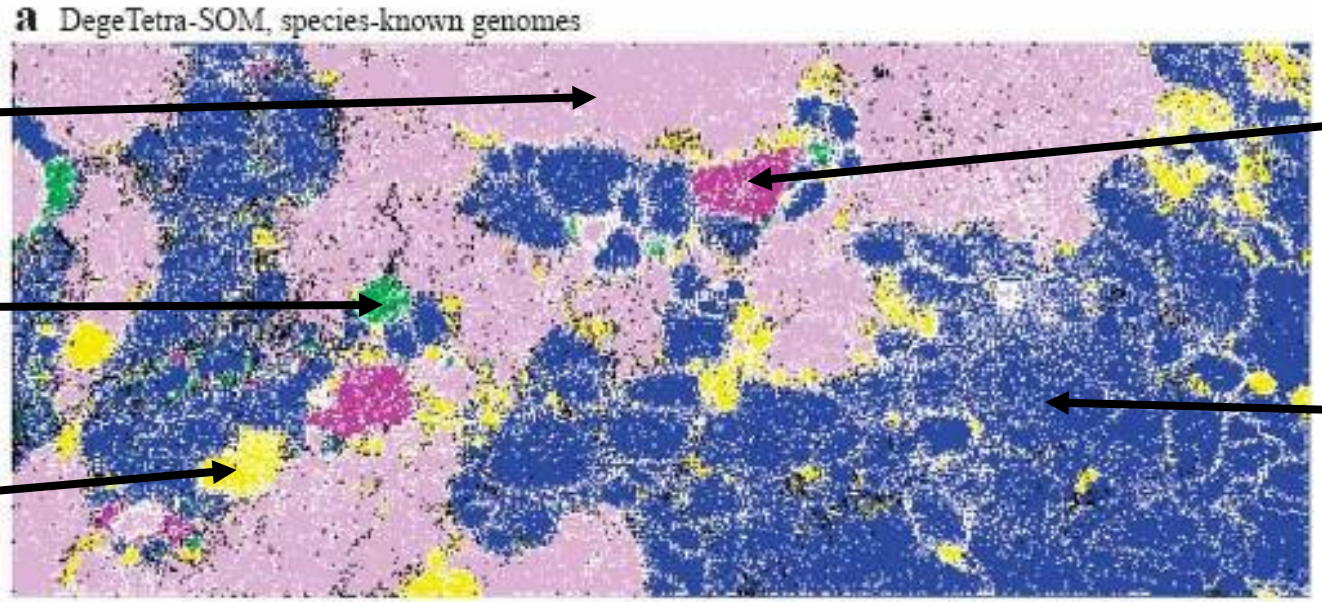
SOM Created from an Unsupervised Neural Network Algorithm to Analyze Tetranucleotide Frequencies in a Wide Range of Genomes

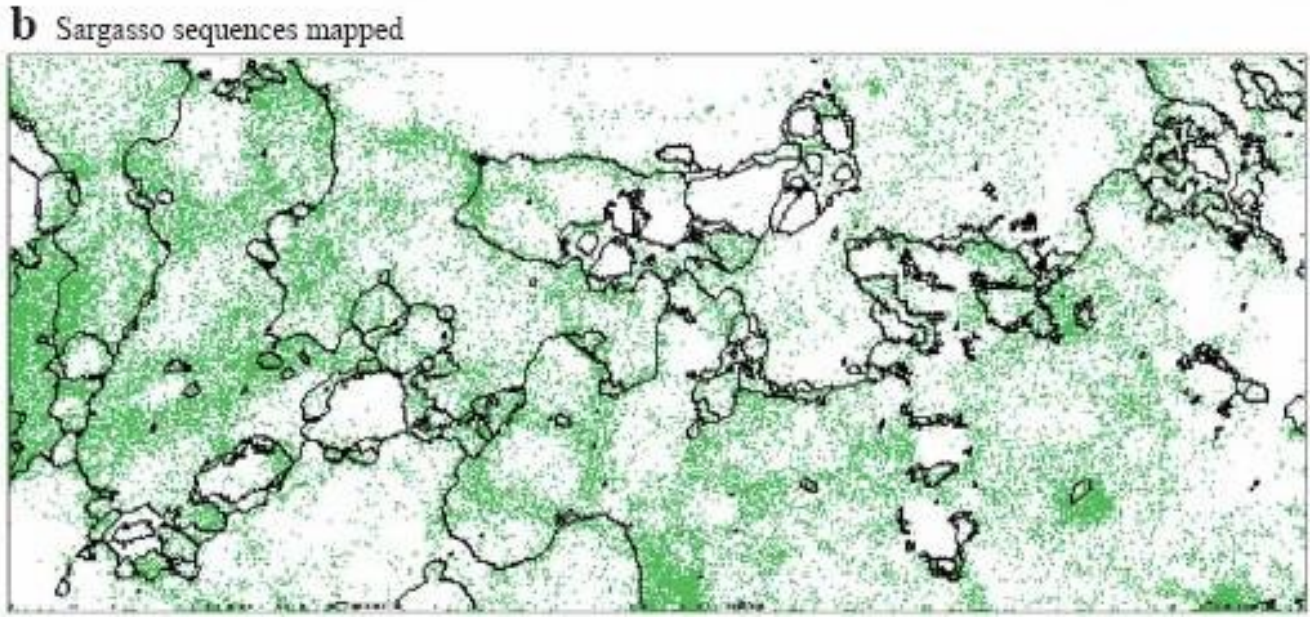
10kb Moving Window

UCSD
UCIrvine

T. Abe, H. Sugawara, S. Kanaya, T. Ikemura
Journal of the Earth Simulator, Volume 6, October 2006, 17–23
www.es.jamstec.go.jp/publication/journal/jes_vol.6/pdf/JES6_22-Abe.pdf

Using SOM, Sargasso Sea Metagenomic Data Yields 92 Microbial Genera !





Input Genomes:

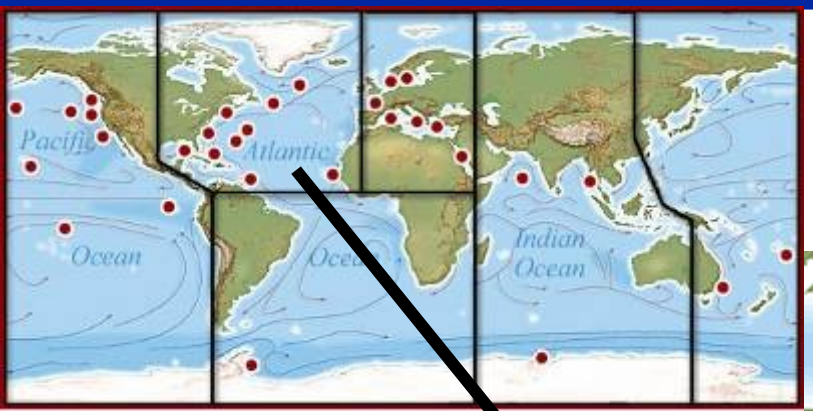
1500 Microbes
40 Eukaryotes
1065 Viruses
642 Mitochondria
42 Chloroplasts

UCSD
UCIrvine

T. Abe, H. Sugawara, S. Kanaya, T. Ikemura
Journal of the Earth Simulator, Volume 6, October 2006, 17–23

it²

Moore Microbial Genome Sequencing Project
Selected Microbes Throughout the World's Oceans



Microbes Nominated by Leading Ocean Microbial Biologists

North Atlantic
Proteobacteria: Oligobacter sp.
Proteobacteria: Sphingomonas sp.
Proteobacteria: Loktanella vestfoldensis
Proteobacteria: Photobacterium sp.

NE U.S. Coastal Waters
Cyanobacteria: Microcoleus chthonoplastes
Cyanobacteria: Synechococcus sp.
Nitrospira: Nitrospira marina
Proteobacteria: Arcobacter sp.
Proteobacteria: Vibrio alginolyticus
Proteobacteria: Vibrio splendidus

NW Atlantic
Proteobacteria: Erythrobacter sp.

SE U.S. Coastal Waters
Chloroflexi: Dehalobium sp.
Proteobacteria: Roseobacter sp.
Proteobacteria: Roseovarius sp.
Proteobacteria: Sagittula stellata
Proteobacteria: Sulfitobacter sp.

Mid-Atlantic
Archea: Pyrolobus sp.
Archea: Thermococcus sp.
Cyanobacteria: Prochlorococcus sp.
Cyanobacteria: Synechococcus sp.
Proteobacteria: Caminibacter sp.
Proteobacteria: Sulfitobacter sp.

Bermuda BATS Station
Proteobacteria: Parvularcula bermudensis
Proteobacteria: Fulvimarina pelagi
Proteobacteria: Oceanicola granulosus
Proteobacteria: Oceanicola batsensis
Proteobacteria: not yet validly described
Proteobacteria: Roseovarius sp.
Proteobacteria: Oceanicaulis aliexandrii
Proteobacteria: Erythrobacter literalis
Bacteroidetes: Croceibacter atlanticus
Bacteroidetes: Robiginitalea biformata
Actinobacteria: Janibacter sp.

Caribbean Sea
Proteobacteria: Roseovarius nubinhibens
Cyanobacteria: Gloeothece sp.

Gulf of Mexico
Cyanobacteria: Synechococcus sp.
Cyanobacteria: Synechococcus sp.
Firmicutes: Bacillus sp.
Firmicutes: Bacillus sp.
Proteobacteria: Vibrionaceae sp.

Western Tropical Atlantic
Proteobacteria: Nitrosococcus oceani
Nitrospira: Nitrospina gracilis
Cyanobacteria: Trichodesmium thiebautii

Eastern Tropical Atlantic
Proteobacteria: Nitrobacter sp.
Cyanobacteria: Synechococcus sp.

MARINE MICROBIOLOGY INITIATIVE

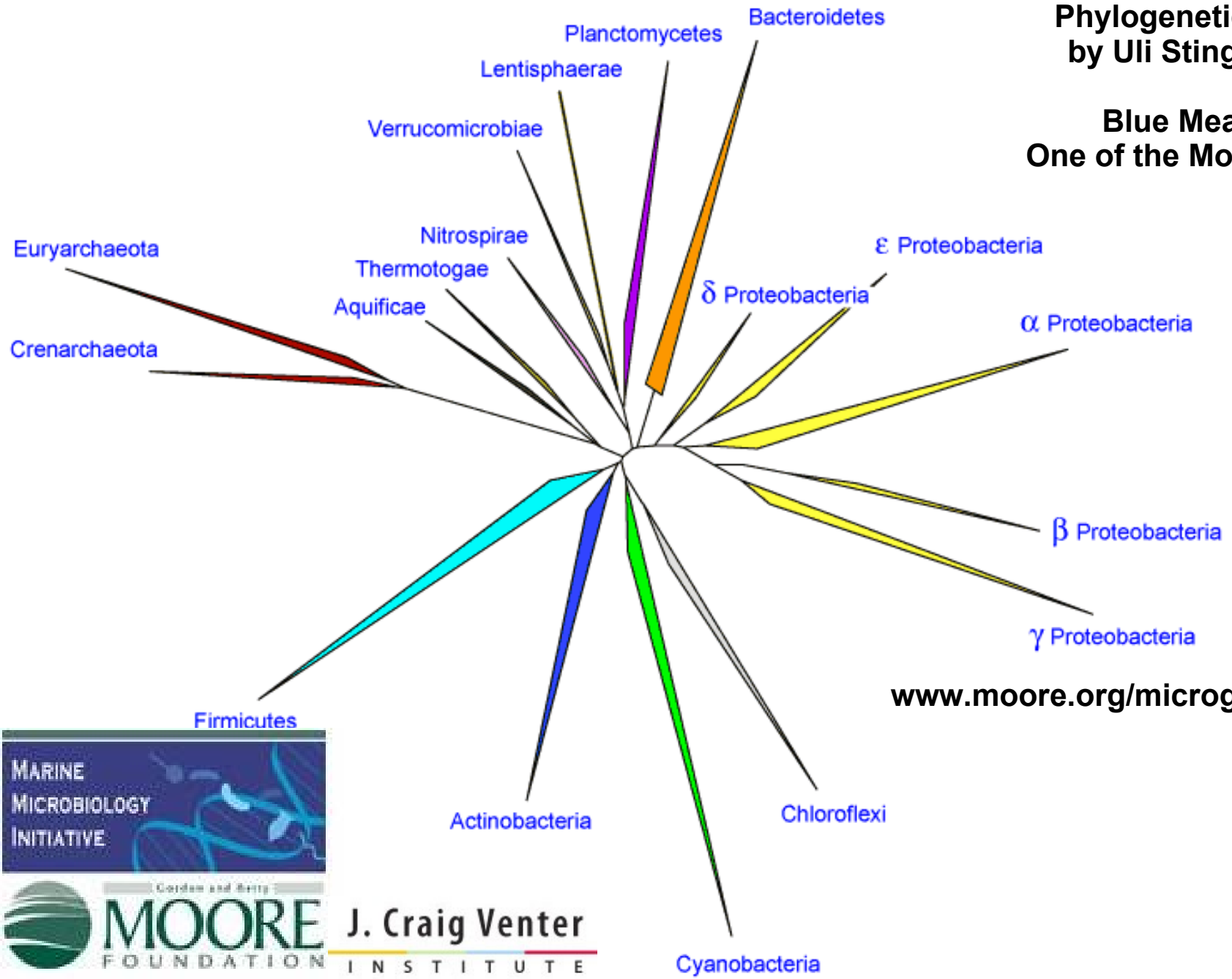
Gordon and Betty MOORE FOUNDATION

UCSD UCIrvine

www.moore.org/microgenome/worldmap.asp

it²

Moore Foundation Funded the Venter Institute to Provide the Full Genome Sequence of 155 Marine Microbes



Phylogenetic Trees Created by Uli Stingl, Oregon State

Blue Means Contains One of the Moore 155 Genomes

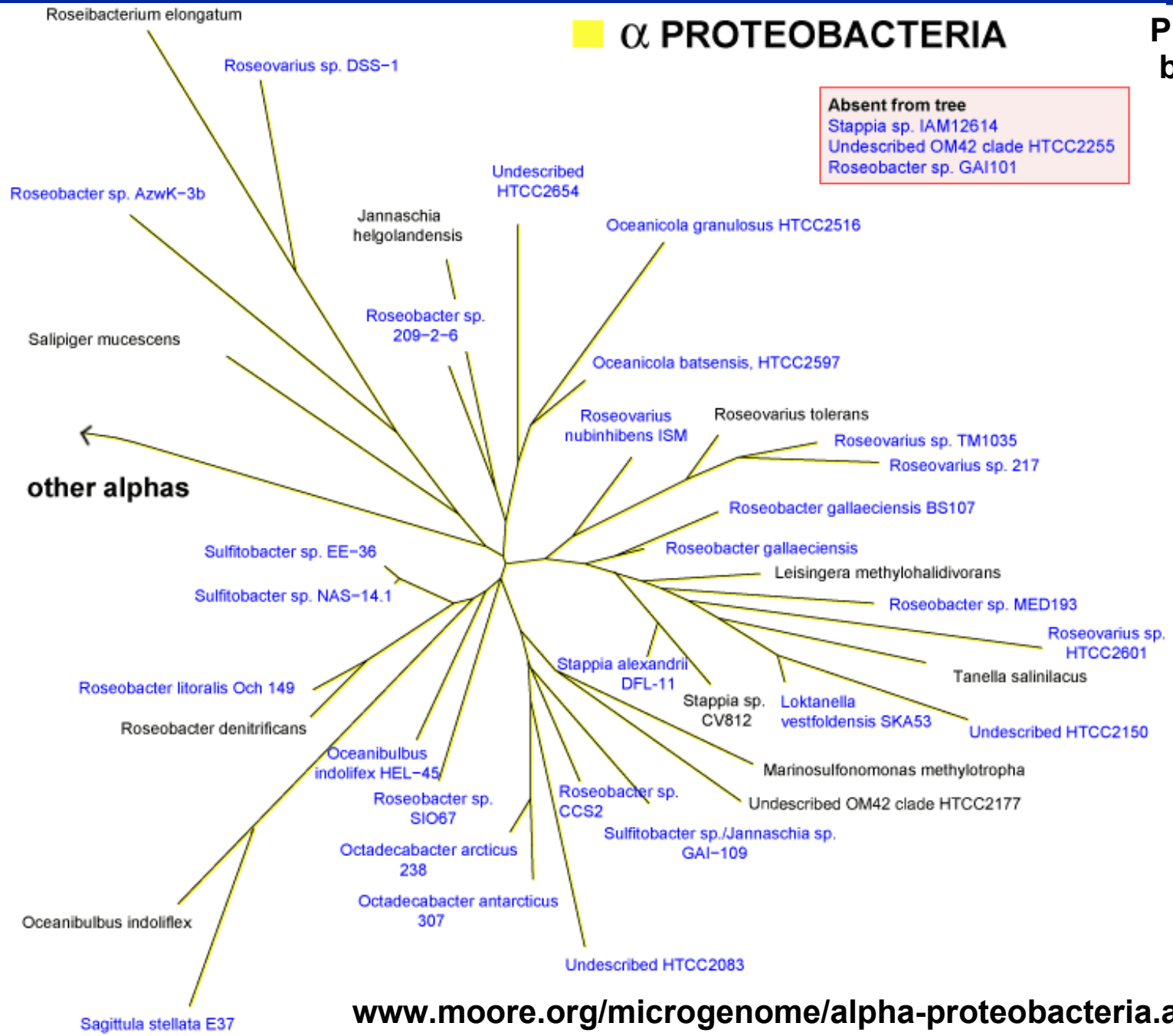
www.moore.org/microgenome/trees.aspx

MARINE MICROBIOLOGY INITIATIVE

Gordon and Betty MOORE FOUNDATION

J. Craig Venter INSTITUTE

Moore 155 Marine Microbial Genomes Gives Broad Coverage of Microbial "Tree of Life"



Phylogenetic Trees Created by Uli Stingl, Oregon State

www.moore.org/microgenome/alpha-proteobacteria.aspx

MARINE MICROBIOLOGY INITIATIVE

Gordon and Betty MOORE FOUNDATION

J. Craig Venter INSTITUTE

it2

Joint Genome Institute is a Leading Microbial Genomic Source

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

All genomes selected

Quick Genome Search : GO

img INTEGRATED MICROBIAL GENOMES

IMG Home | Find Genes | Find Genomes | Find Functions | Compare Genomes | MyIMG | Analysis Carts | About IMG | Using IMG | News

Genome Browser | Genome Search

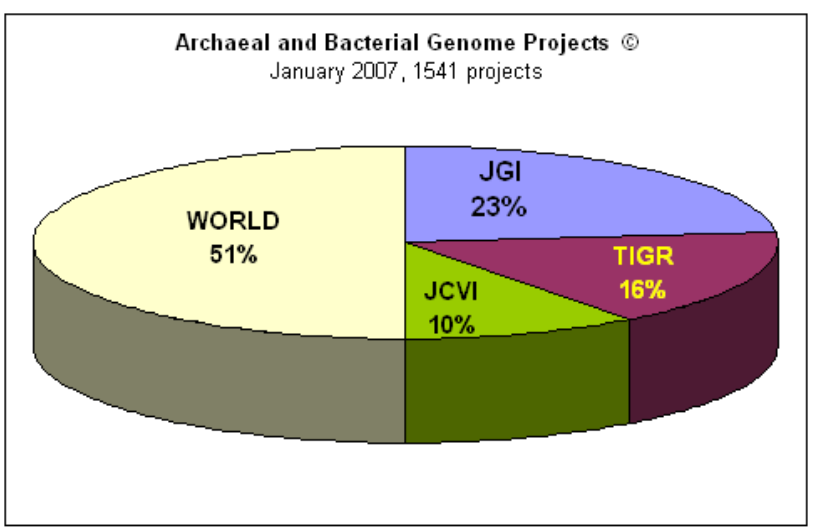
IMG Genomes

finished/draft	JGI	Total
Bacteria	120/92	442/245
Archaea	11/4	37/4
Eukarya	0/0	16/8
Plasmids	0/0	402/0
Viruses	0/0	1661/0
All Genomes	131/96	2558/257
Grand Total	227	2815

MICROBIAL GENOMICS

Home Microbes Assemblies Collaborators IMG Download Help

Select a microbe from: all JGI Microbes and Acidiphilium cryptum JF-5



GOLD

UCSD
UCIrvine

it²

JGI Metagenomics Projects (42 Projects)

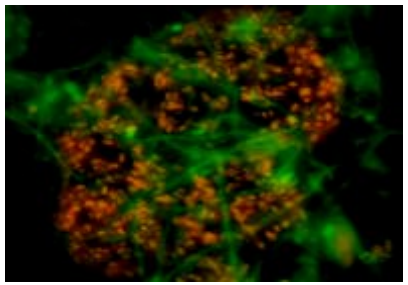
2005



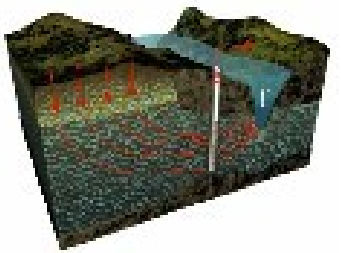
termite hindgut (CalTech)



planktonic archaea (MIT)



EBPR sludge (UW/UQ)



groundwater (ORNL)

2006



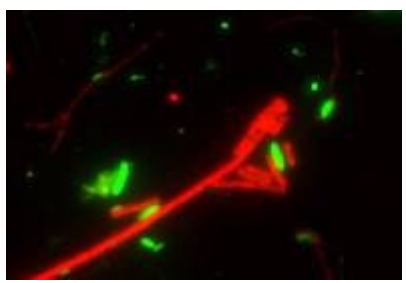
AMD



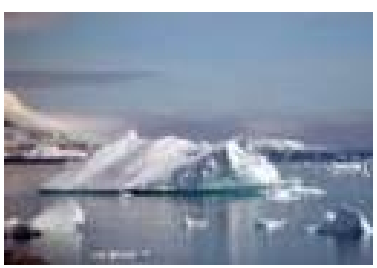
Alaskan soil (UW)



Gutless worm (MPI)



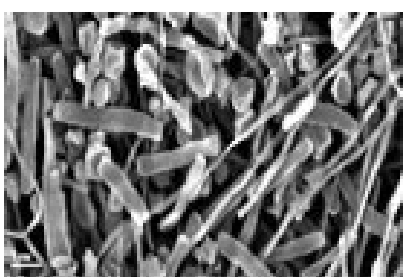
TA-degrading bioreactor (NUS)



Antarctic bacterioplankton (DRI)



hypersaline mats (UCol)



Korarchaeota enrichment



Farm soil (Diversa)

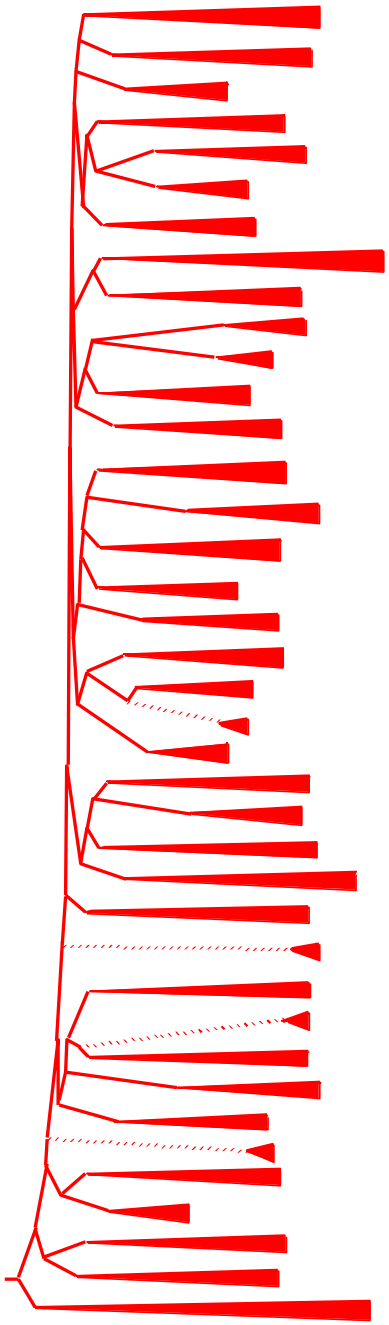
2007 **8 new metagenomic projects**

Source: Eddie Rubin, DOE JGI

UCSD UCIrvine

it²

Key Problem with Analysis of Microbial Metagenomic Data



At Least 40 Phyla of Bacteria,
But Only a Few are Well Sampled

UCSD
UCIrvine

Source: Eddie Rubin, DOE JGI

it²

DOE Genomic Encyclopedia of Bacteria and Archaea (GEBA) / Bergey Solution: Deep Sampling Across Phyla



Well sampled phyla

No cultured taxa

Source: Eddie Rubin, DOE JGI

UCSD
UCIrvine

GEBA / Bergey
Pilot Project at JGI

- **Goal**
  - To Finish ~100 Bacterial and Archaeal Genomes
  - Selected Based on:
    - Phylogeny,
    - Availability of Phenotype Information
    - Community Interest
- **Approach**
  - Select 200 Organisms
  - Order DNA from Culture Collections (DSMZ and ATCC)
  - Sequence 100 for which DNA QC is Received
- **Project Lead (Jonathan Eisen JGI/UC Davis)**
  - Project Management (David Bruce JGI/LANL)
  - Methods for Sequencing in Changing Technology Landscape (Paul Richardson JGI)
  - Linking to educational project (Cheryl Kerfeld JGI)

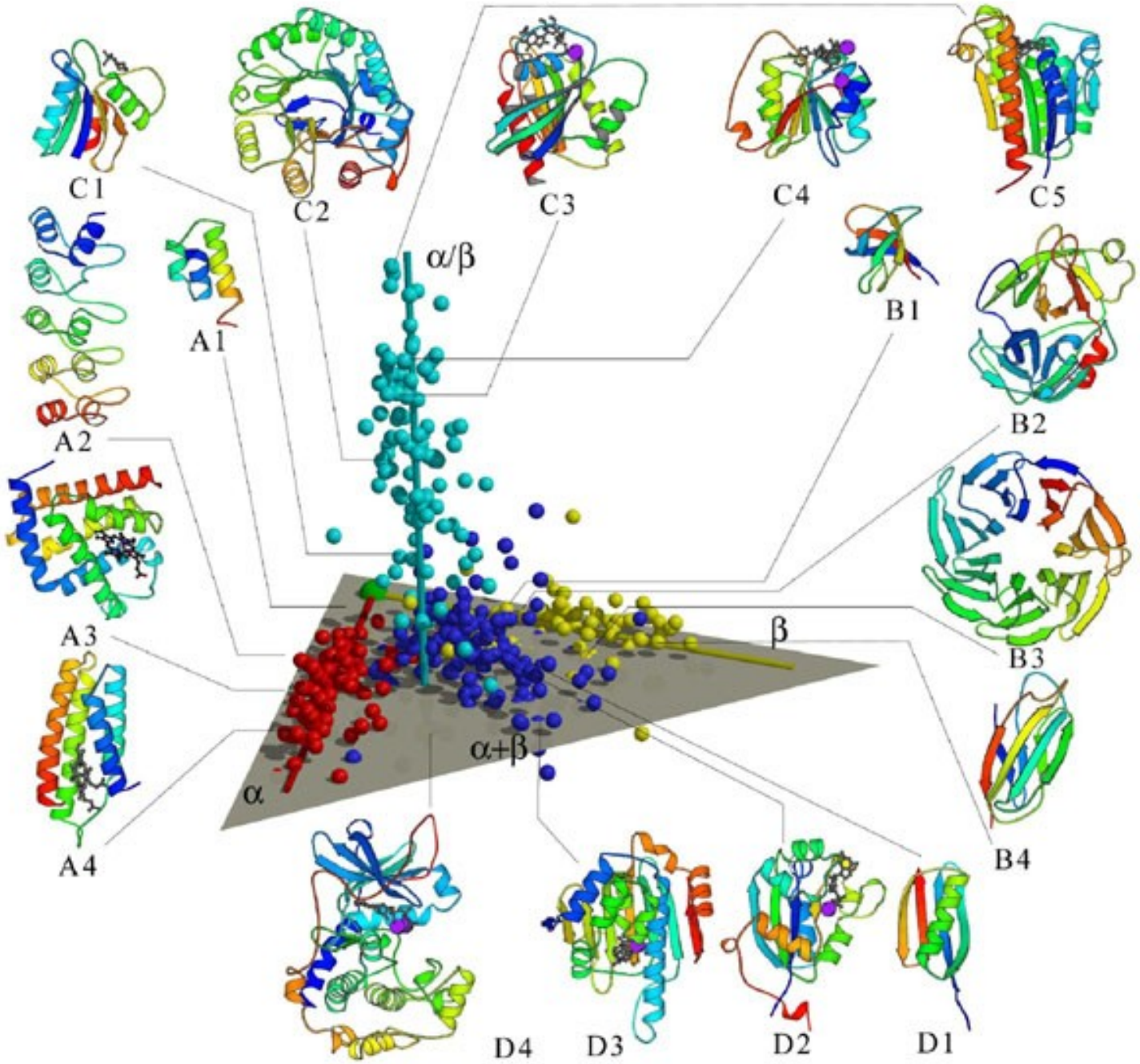
Input / Interactions with:
Community Advisory Group ,
ASM,
Academy of Microbiology,
Etc…

UCSD
UCIrvine

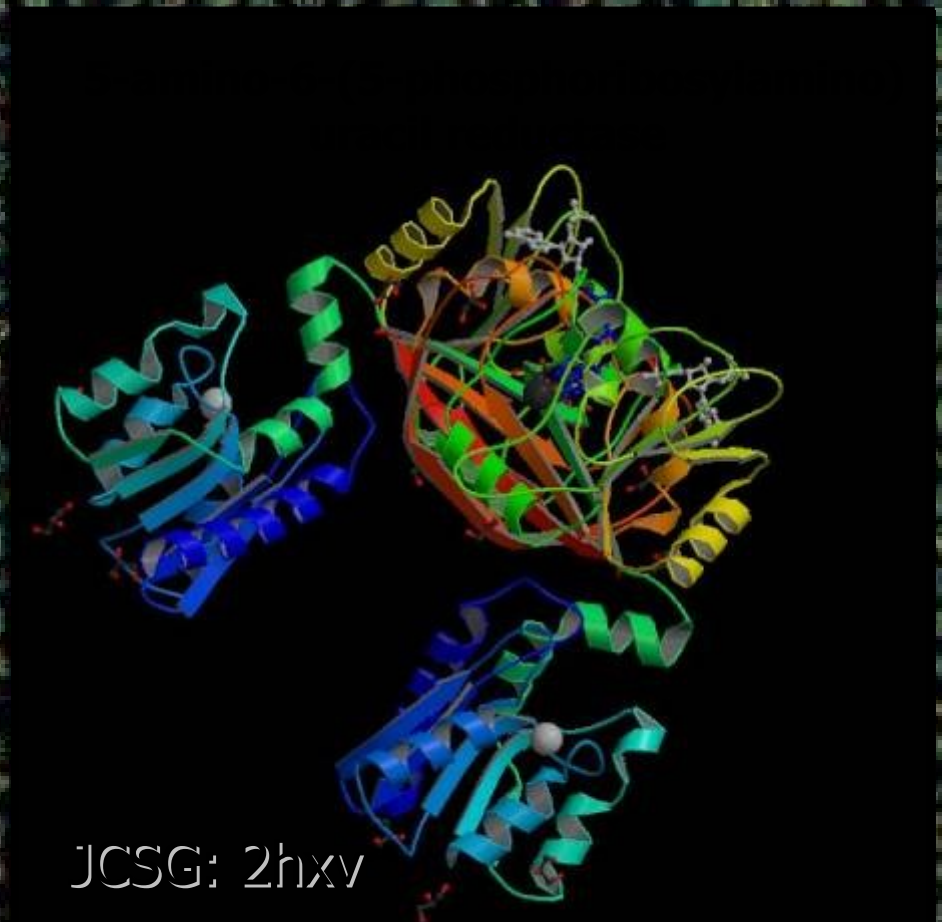
Source: Eddie Rubin,
DOE JGI

it²

Converting Genome Sequences to Protein Fold Space



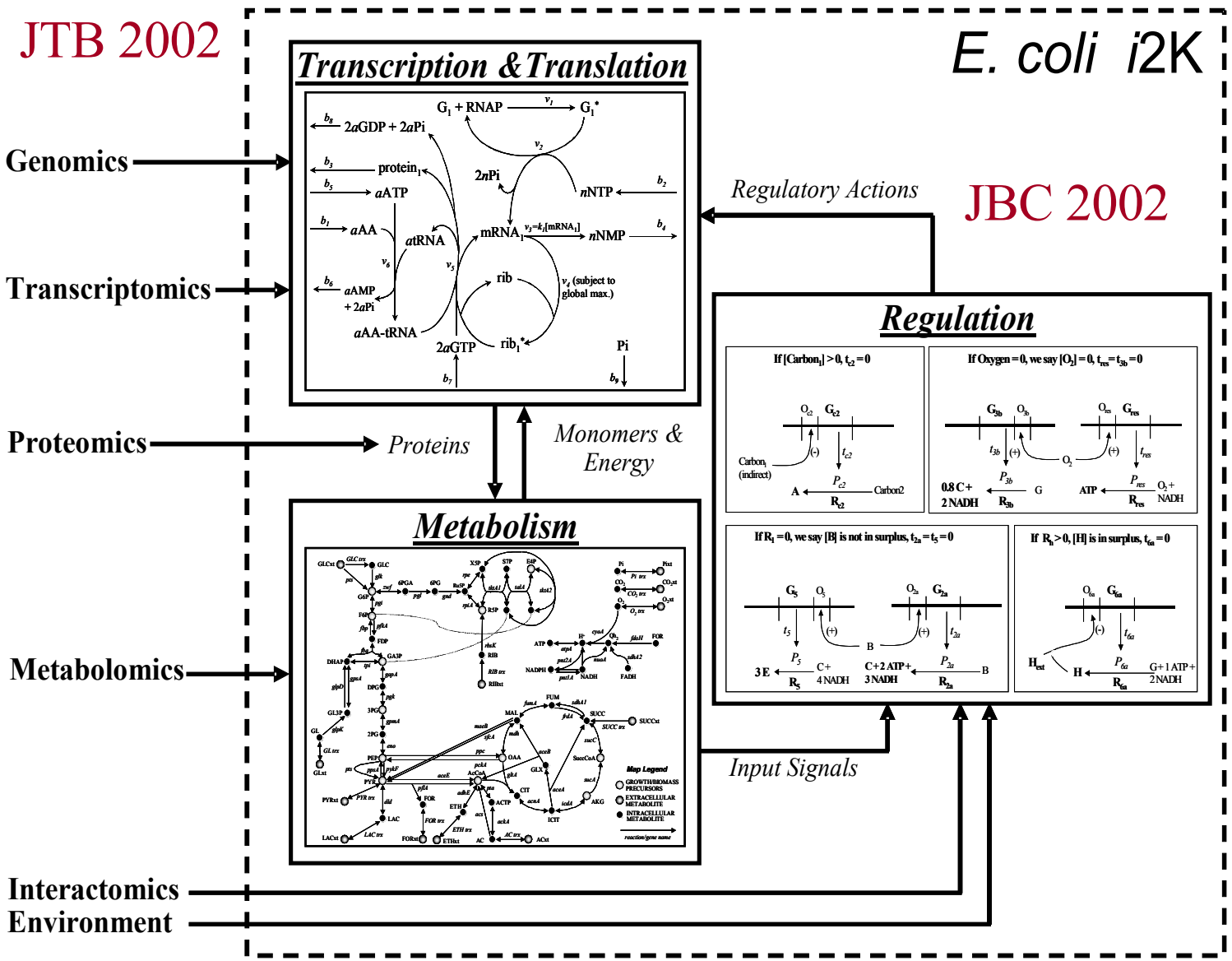
- How many folds?
- How many sequences adopt the same fold?
- How does function vary as sequences diverge within a family?
- Are there still Kingdom-specific families?
- Can we determine function from structure?
- How diverse are metabolic pathways and networks?



Joint Center for Structural Genomics

PSI
Protein Structure Initiative

Building Genome-Scale Models of Living Organisms



- **E. Coli**
  - **Has 4300 Genes**
  - **Model Has 2000!**

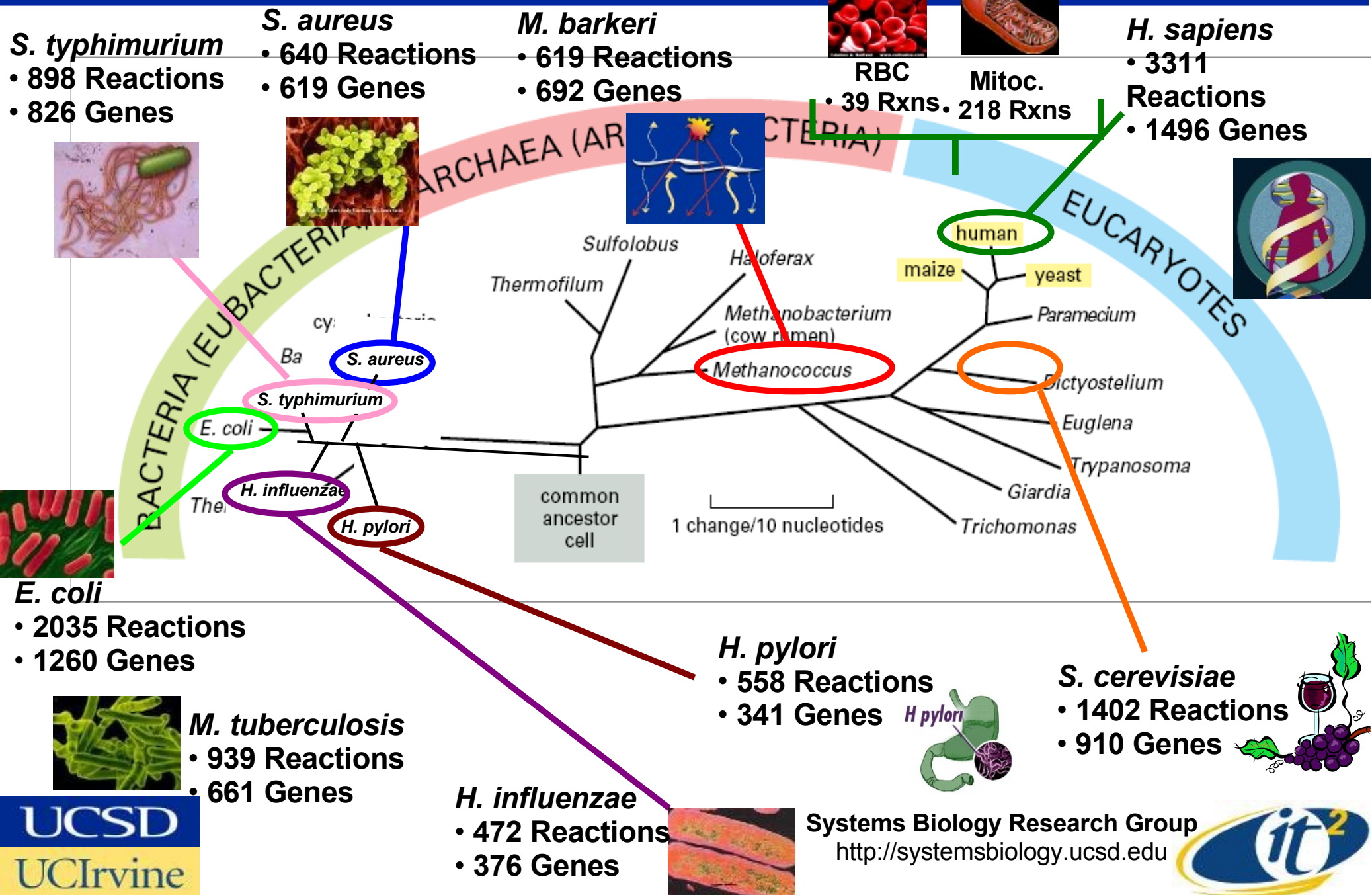
***in Silico* Organisms Now Available 2007:**

- Escherichia coli
- Haemophilus influenzae
- Helicobacter pylori
- Homo sapiens Build 1
- Human red blood cell
- Human cardiac mitochondria
- Methanosarcina barkeri
- Mouse Cardiomyocyte
- Mycobacterium tuberculosis
- Saccharomyces cerevisiae
- Staphylococcus aureus

Source: Bernhard Palsson
UCSD Genetic Circuits Research Group
http://gcrg.ucsd.edu

UCSD
UCIrvine

Biochemically, Genetically and Genomically (BiGG) Genome-Scale Metabolic Reconstructions



UCSD
UCIrvine

Systems Biology Research Group
http://systemsbiology.ucsd.edu

Use of Tiled Display Wall OptIPortal to Interactively View Microbial Genome



OptIPuter

camera
Marine Microbial Ecology

Powered by Rocks
ROCKS

Acidobacteria bacterium Ellin345
Soil Bacterium 5.6 Mb

it2

Use of Tiled Display Wall OptIPortal to Interactively View Microbial Genome



camera
Marine Microbial Ecology

Source: Raj Singh, UCSD

Powered by Rocks ROCKS

OptIPuter

it2

Use of Tiled Display Wall OptIPortal to Interactively View Microbial Genome



camera Marine Microbial Ecology

Source: Raj Singh, UCSD

Powered by Rocks ROCKS

OptIPuter

it2

OptIPortal–Termination Device for the Dedicated Gigabit/sec Lightpaths



evl electronic visualization laboratory

Collaborative Analysis of Large Scale Images of Cancer Cells

Integration of High Definition Video Streams with Large Scale Image Display Walls

UCSD UCIrvine

OptIPuter

Photo Source: David Lee, Mark Ellisman NCMIR, UCSD

NCMIR NATIONAL CENTER for MICROSCOPY and IMAGING RESEARCH at San Diego, an NIH supported resource center

it2

An Emerging High Performance Collaboratory for Microbial Metagenomics

