Cyberinfrastructure for Advanced Marine Microbial Ecology Research and Analysis (CAMERA)

Invited Keynote
Annual Meeting CENIC 2006
Oakland, CA
March 13, 2006

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

OptIPuter

it2

Calit2 Brings Computer Scientists and Engineers Together with Biomedical Researchers

- **Some Areas of Concentration:**
  - Metagenomics
  - Genomic Analysis of Organisms
  - Evolution of Genomes
  - Proteomics
  - Mitochondrial Evolution
  - Computational Biology
  - Cancer Genomics
  - Human Genomic Variation and Disease
  - Information Theory and Biological Systems



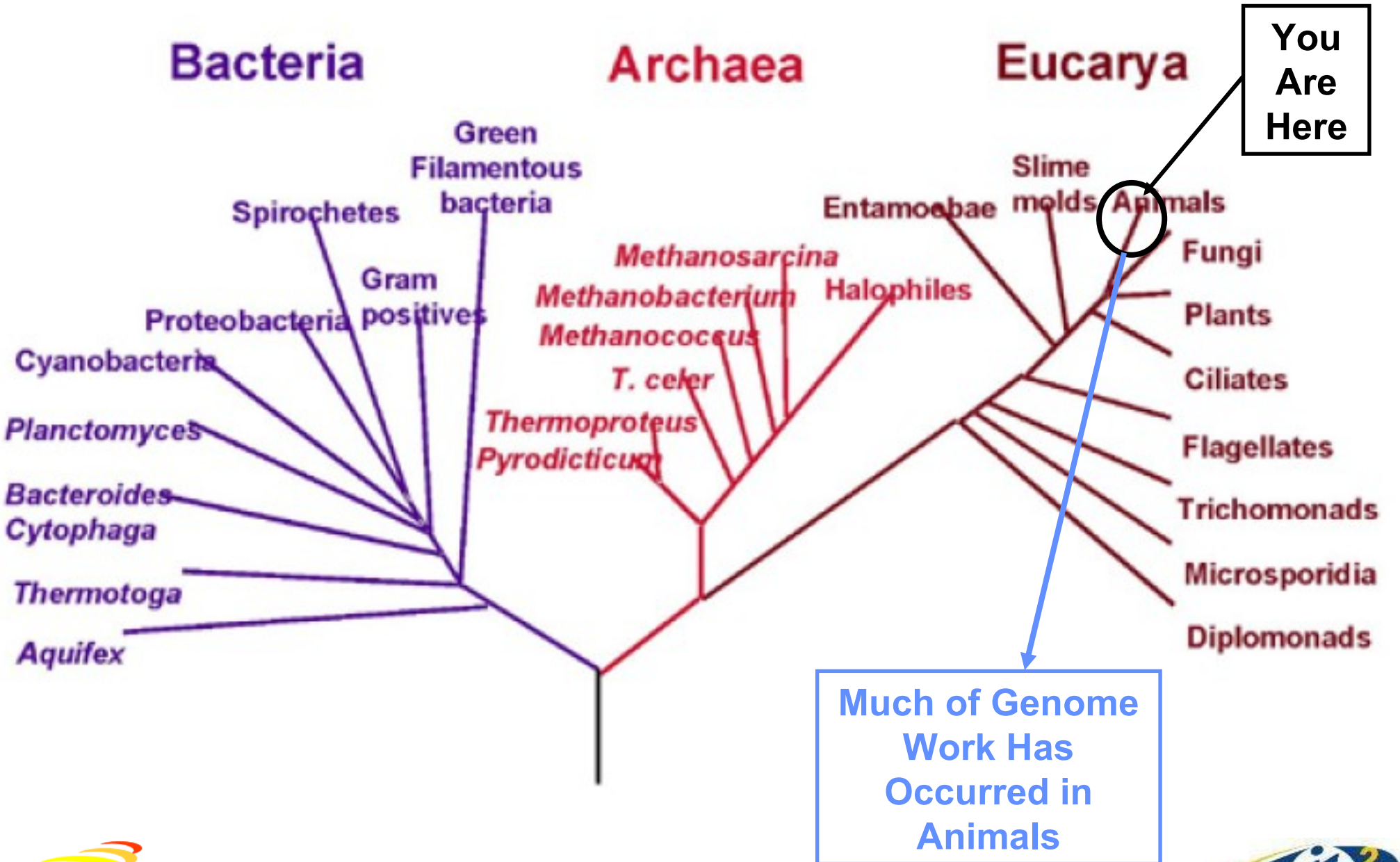


1200 Researchers in Two Buildings

OptIPuter

it2

Evolution is the Principle of Biological Systems: Most of Evolutionary Time Was in the Microbial World



Source: Carl Woese, et al

Comparative Genomics Can Reveal Biological Facts That Are Not Visible Within a Species

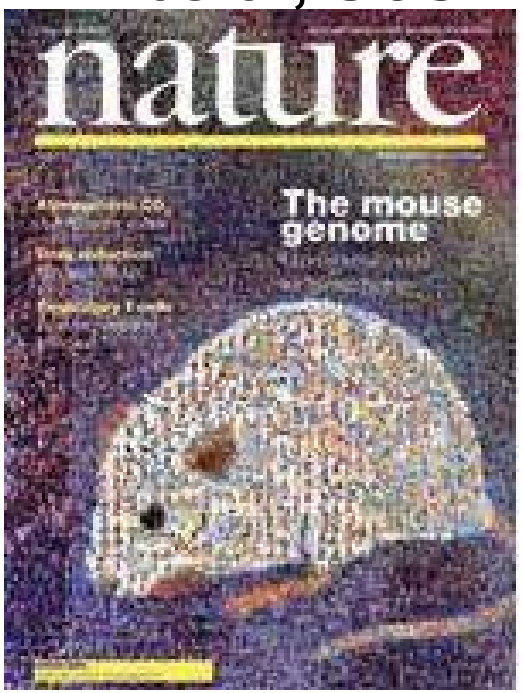
Co-Authors Pavel Pevzner and Glenn Tesler, UCSD



December 9, 2004



April 1, 2004



December 05, 2002

"After sequencing these three genomes, it is clear that substantial rearrangements in the human genome happen only once in a million years, while the rate of rearrangements in the rat and mouse is much faster."

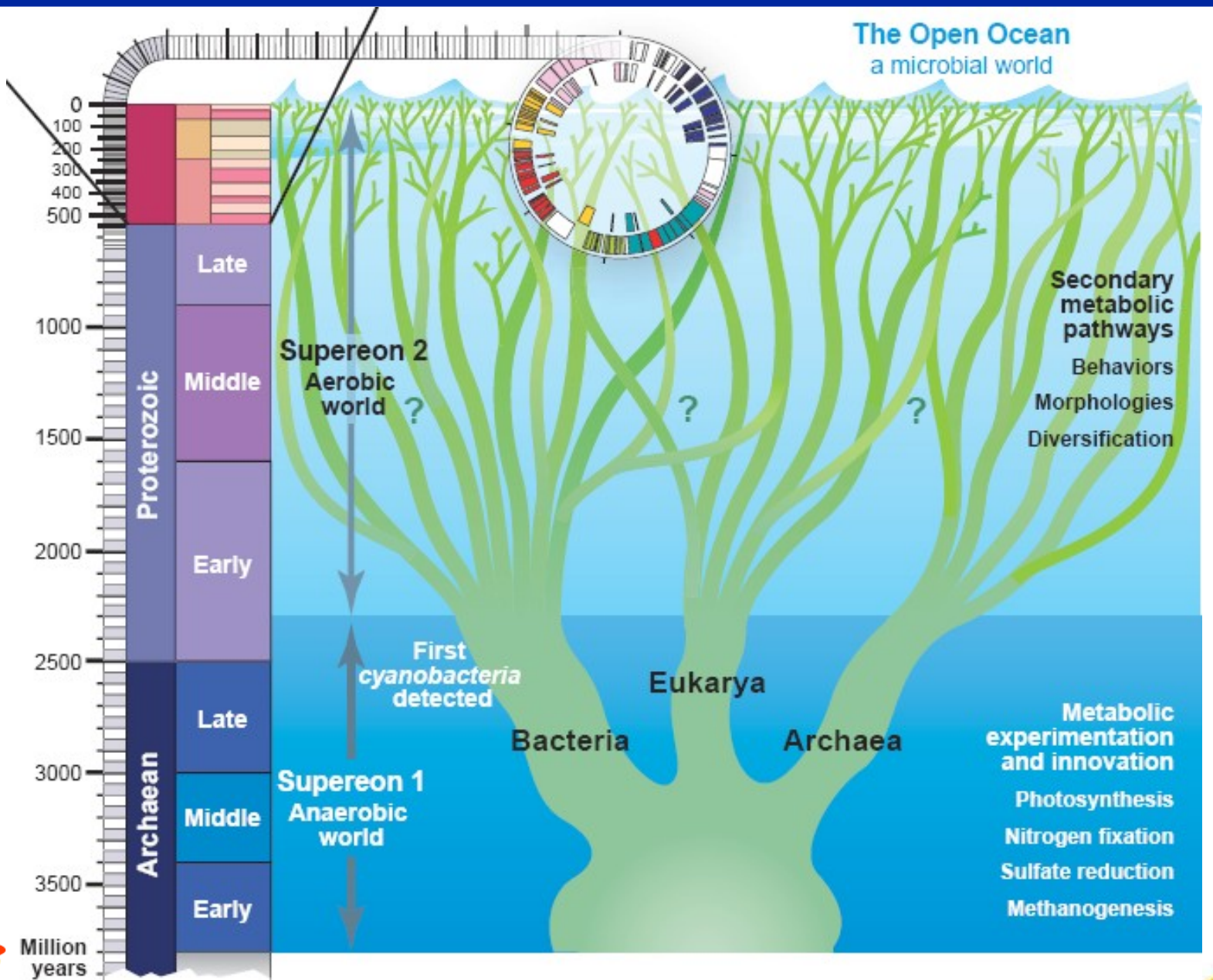
--Glenn Tesler, UCSD Dept. of Mathematics

OptIPuter

www.calit2.net/culture/features/2004/4-1_pevzner.html

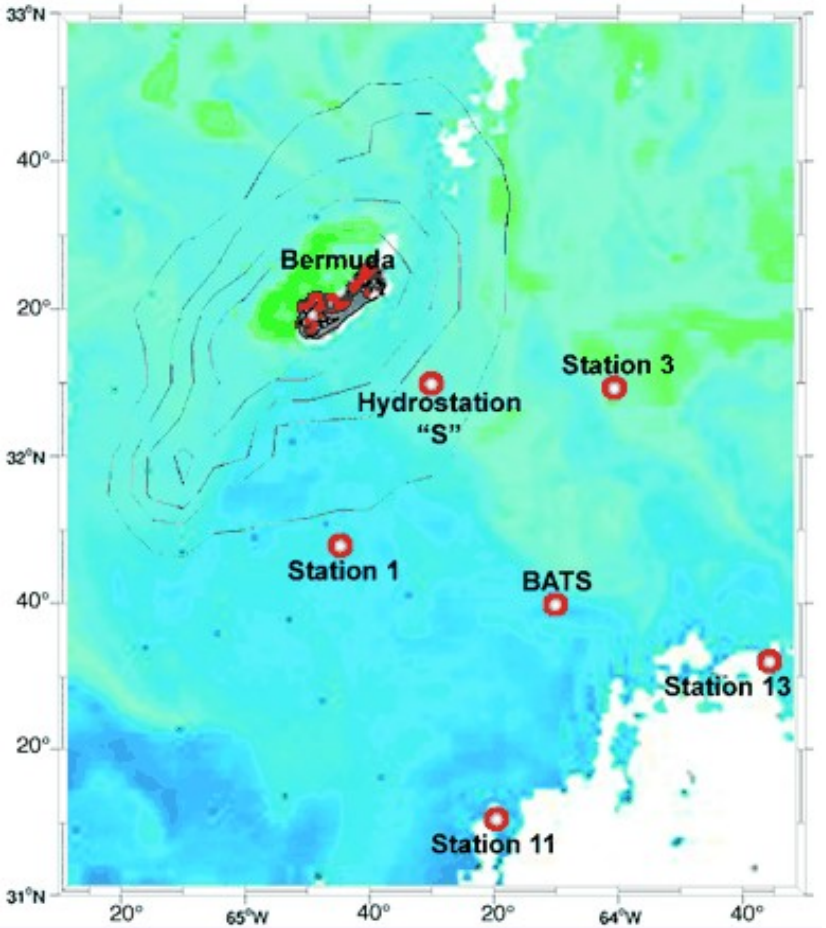
it2

Looking Back Nearly 4 Billion Years In the Evolution of Microbe Genomics



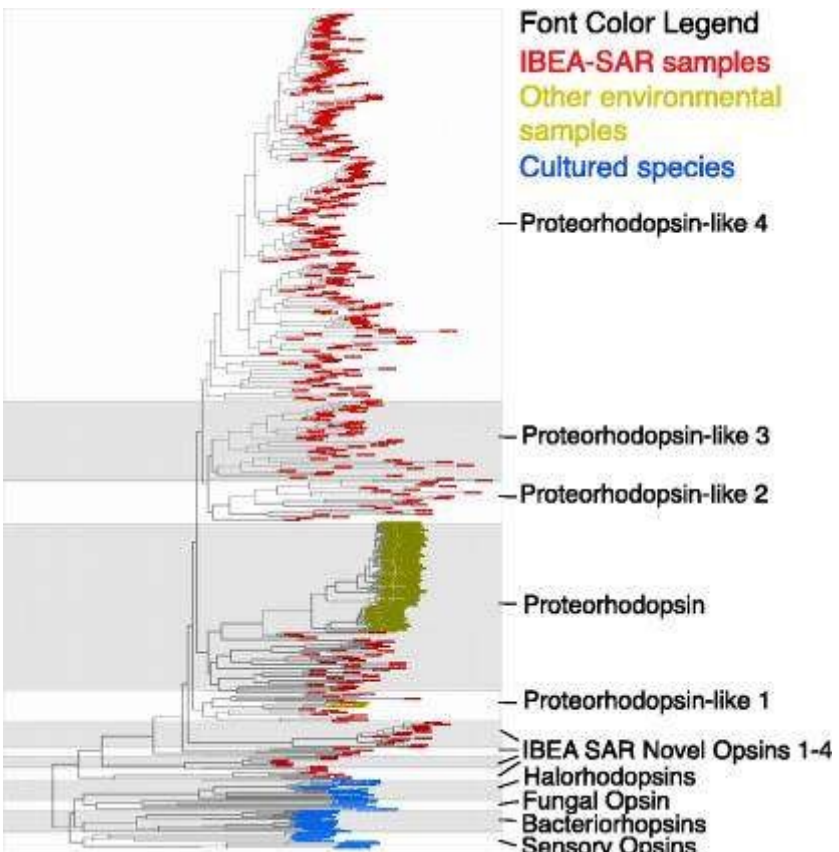
Science Falkowski and Vargas 304 (5667): 58

The Sargasso Sea Experiment
The Power of Environmental Metagenomics



MODIS-Aqua satellite image of ocean chlorophyll in the Sargasso Sea grid about the BATS site from 22 February 2003

- Yielded a Total of Over 1 billion Base Pairs of Non-Redundant Sequence
- Displayed the Gene Content, Diversity, & Relative Abundance of the Organisms
- Sequences from at Least 1800 Genomic Species, including 148 Previously Unknown
- Identified over 1.2 Million Unknown Genes



J. Craig Venter, et al.
Science
2 April 2004:
Vol. 304.
pp. 66 - 74

OptIPuter

it2

Marine Genome Sequencing Project
Measuring the Genetic Diversity of Ocean Microbes



CAMERA will include
All Sorcerer II Metagenomic Data

MARINE MICROBIOLOGY INITIATIVE

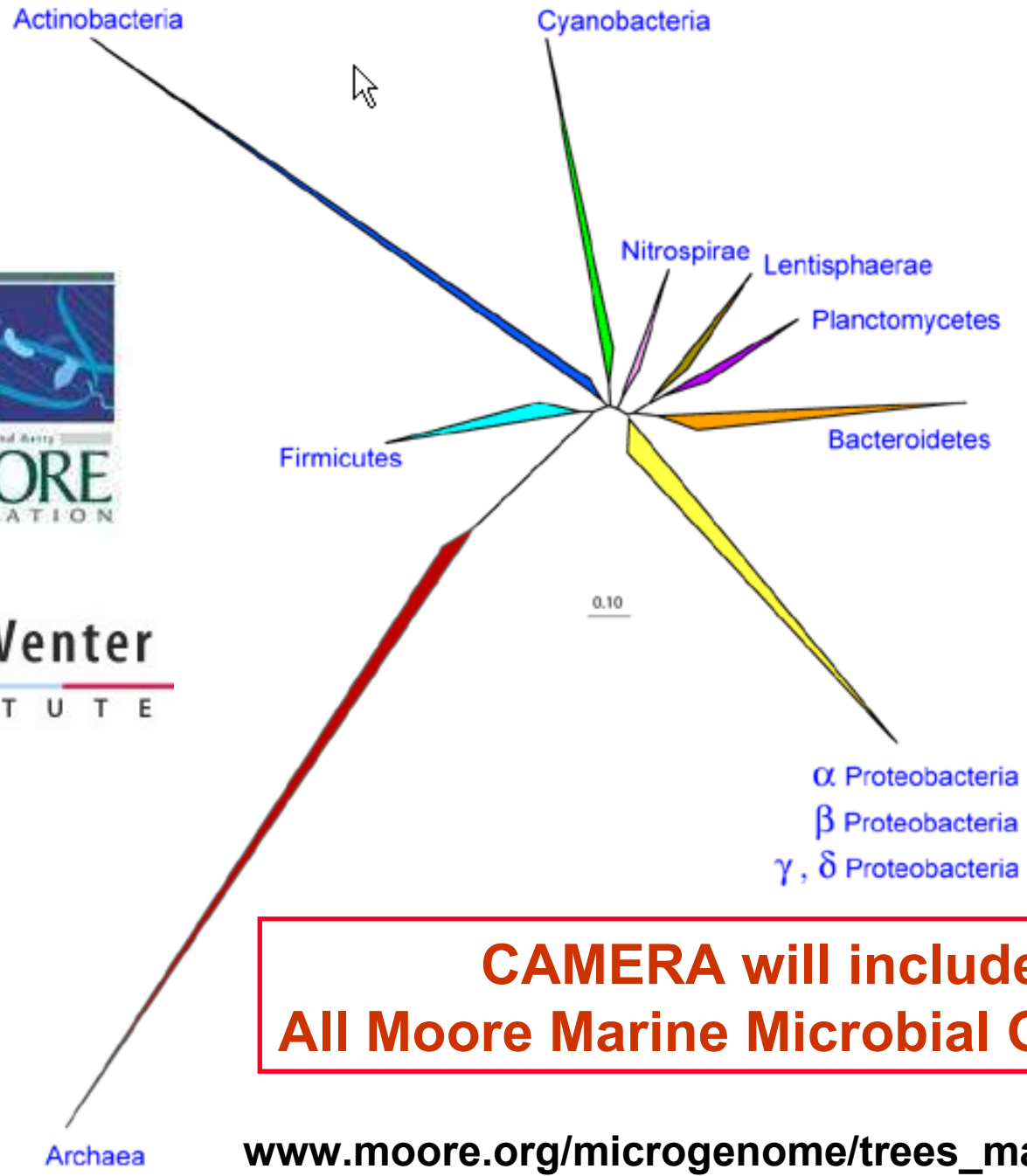
Gordon and Betty MOORE FOUNDATION

OptIPuter

J. Craig Venter INSTITUTE

it²

Moore Foundation Funded the Venter Institute to Provide the Full Genome Sequence of 150 Marine Microbes



MARINE MICROBIOLOGY INITIATIVE

Gordon and Betty MOORE FOUNDATION

J. Craig Venter INSTITUTE

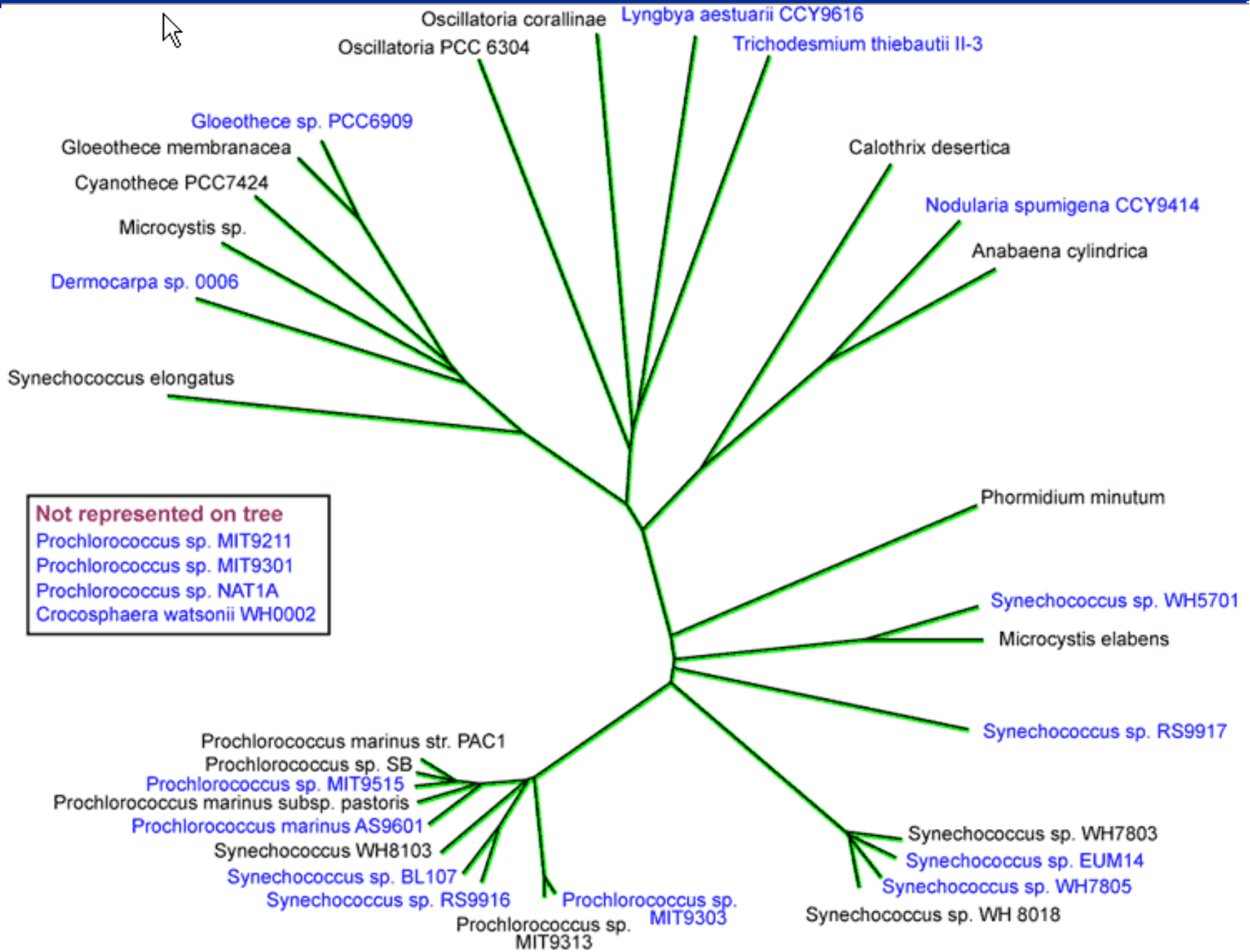
CAMERA will include
All Moore Marine Microbial Genomes

OptIPuter

www.moore.org/microgenome/trees_main.asp

it2

Moore Microbial Genome Sequencing Project: Cyanobacteria Being Sequenced by Venter Institute



OptIPuter

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

National LambdaRail
Direct Connect
Computation and Storage Complex

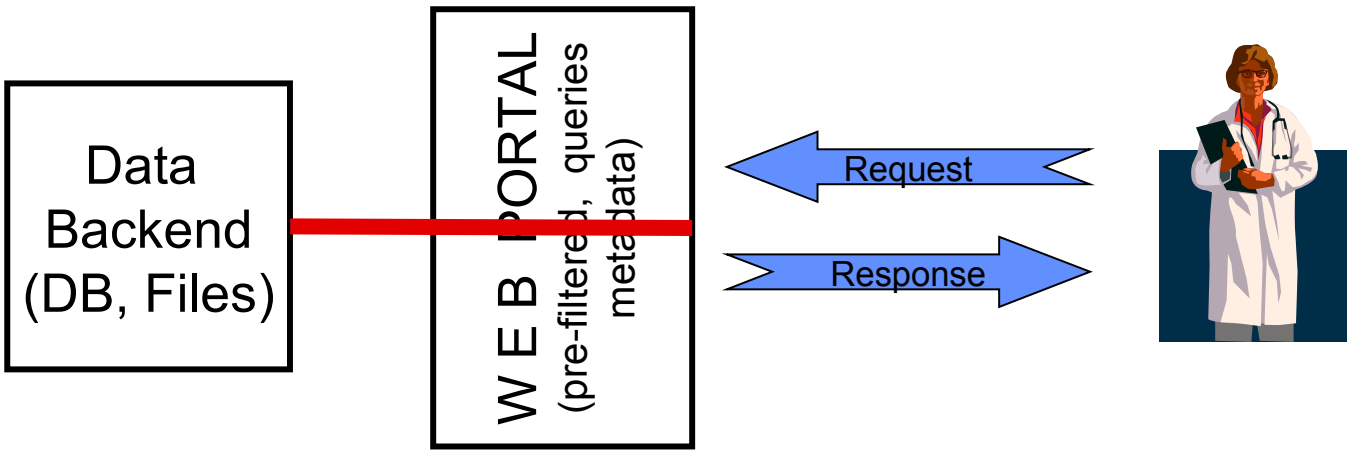
Funded by: Gordon and Betty Moore Foundation

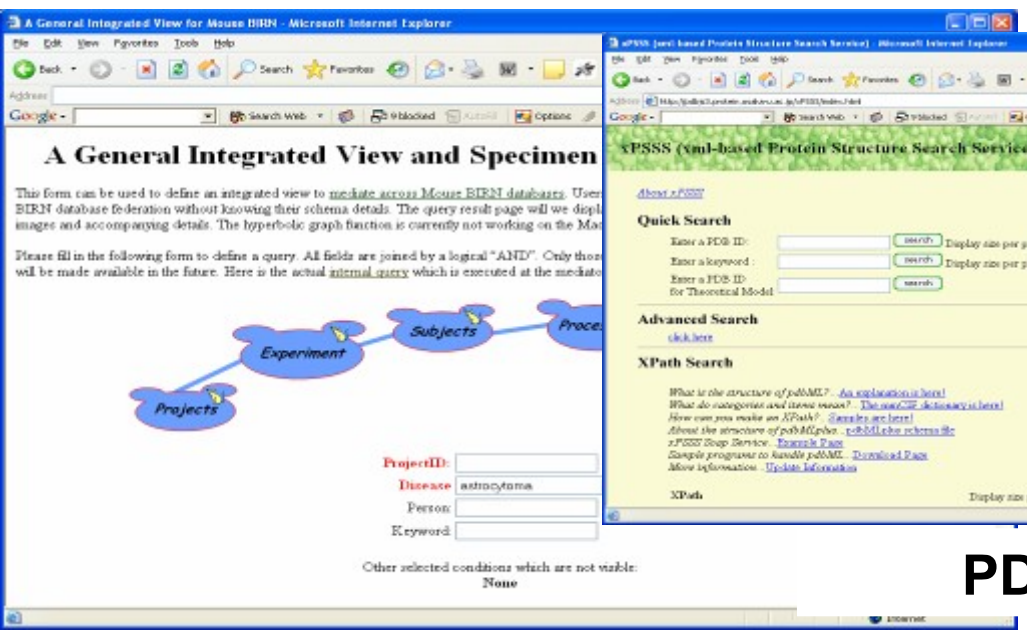
Joint Partnership of:

OptIPuter

PI Larry Smarr

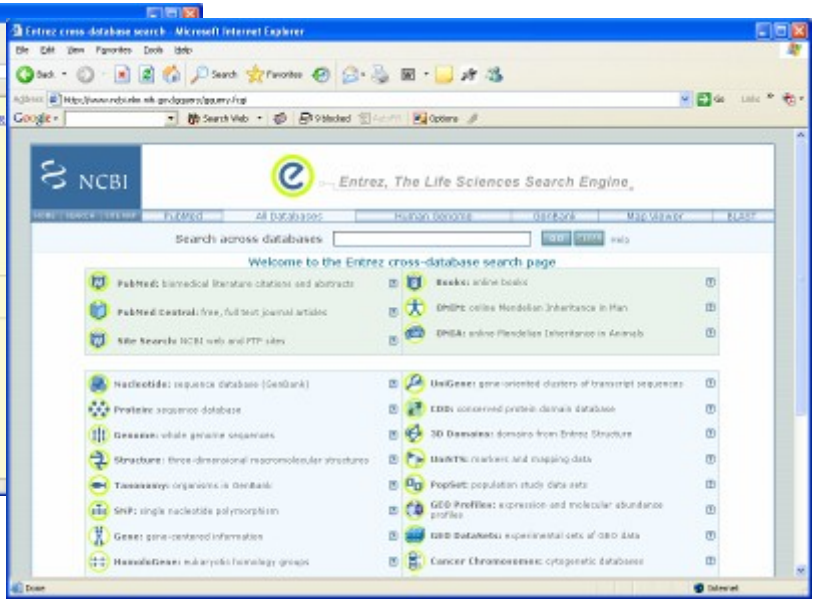
Calit2 Intends to Jump Beyond Traditional Web-Accessible Databases





BIRN

PDB



NCBI Genbank

+ many others

Source: Phil Papadopoulos, SDSC, Calit2

Calit2's Direct Access Core Architecture Will Create Next Generation Metagenomics Server

Sargasso Sea Data

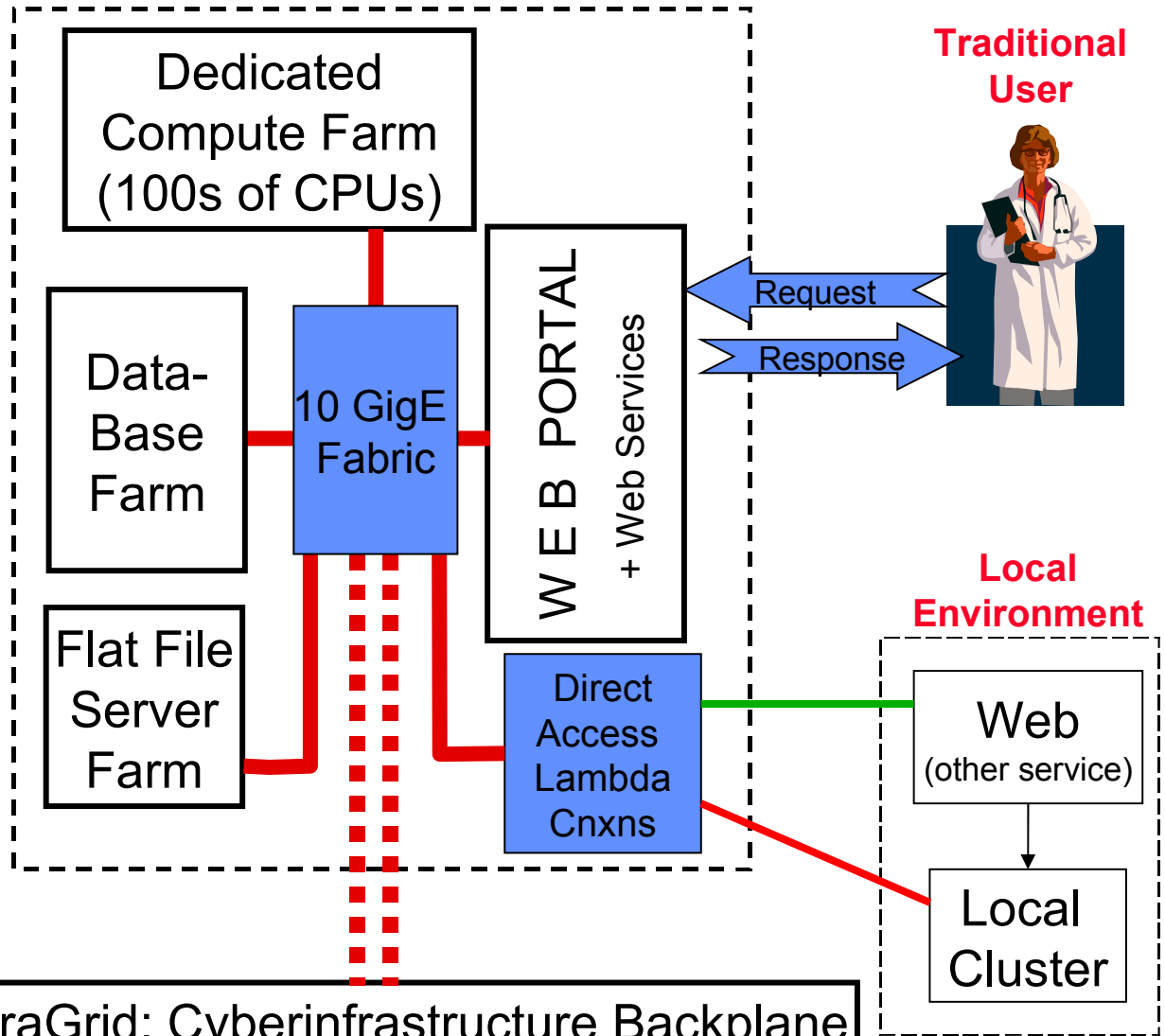
Sorcerer II Expedition (GOS)

JGI Community Sequencing Project

Moore Marine Microbial Project

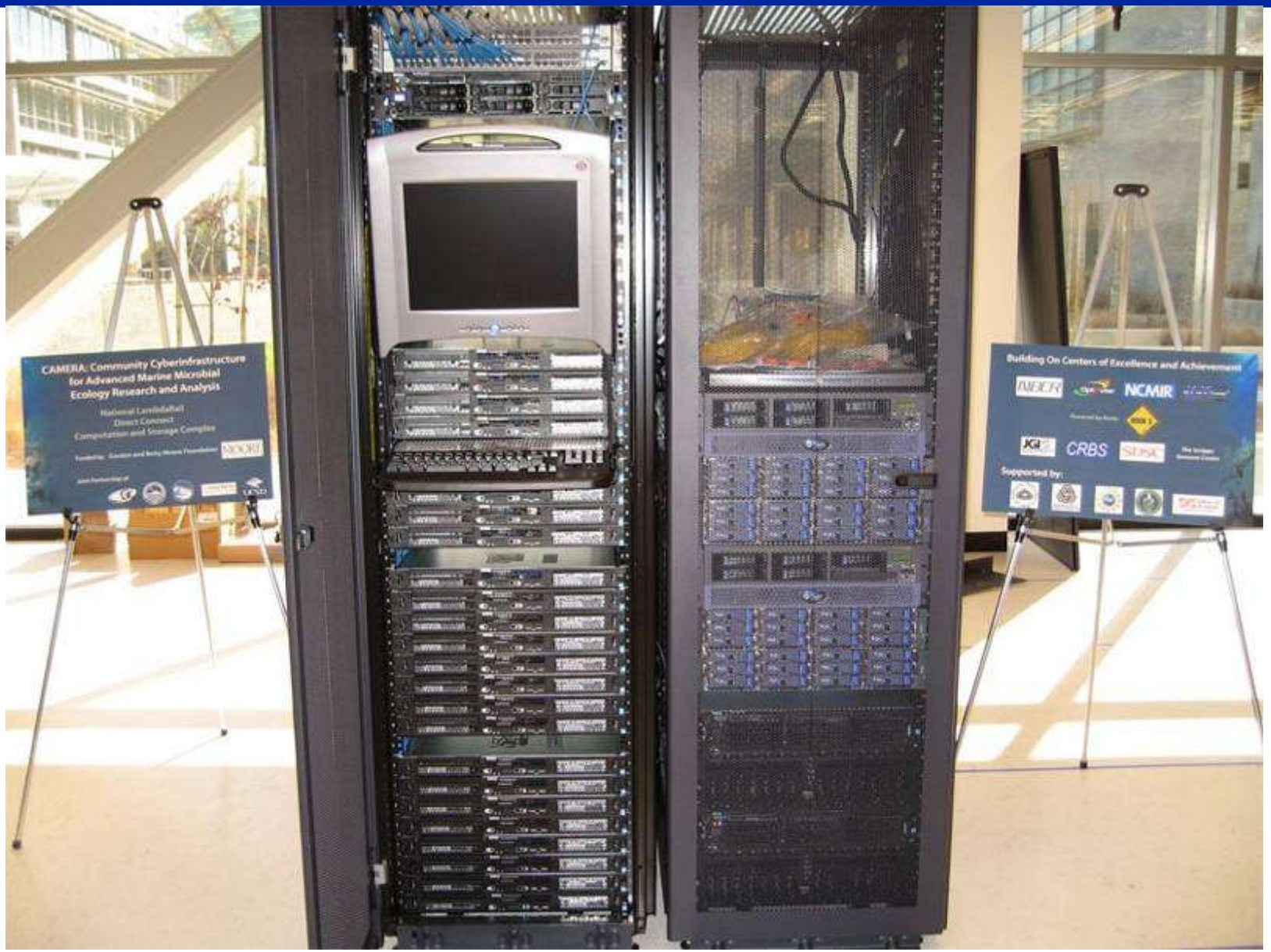
NASA Goddard Satellite Data

Community Microbial Metagenomics Data



Source: Phil Papadopoulos, SDSC, Calit2

First Implementation of the CAMERA Complex



Compute

Database & Storage

OptIPuter

it2

CAMERA Timeline

- **Release 1: Mid-2006**
  - Majority of GOS + Moore Microbe Genome Data
    - 6 Gbp Has Been Assembled
  - Initial Versions of Core Tools
    - BLAST, Reference Alignment Viewer
- **Release 2: Early-2007**
  - Additional Data
  - Additional/Improved Tools
  - Improved Usability
- **Subsequent**
  - Move Towards Semantic DB, Direct Access
  - Additional Tools & Data Based on Community Feedback

OptIPuter

it2

Building On Centers of Excellence and Achievement









Powered by Rocks







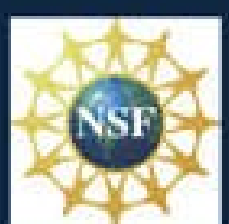


The Scripps Genome Center

Supported by:











OptIPuter

Announced January 17, 2006

it2

CAMERA Builds on Cyberinfrastructure Grid, Workflow, and Portal Projects in a Service Oriented Architecture

NBCR

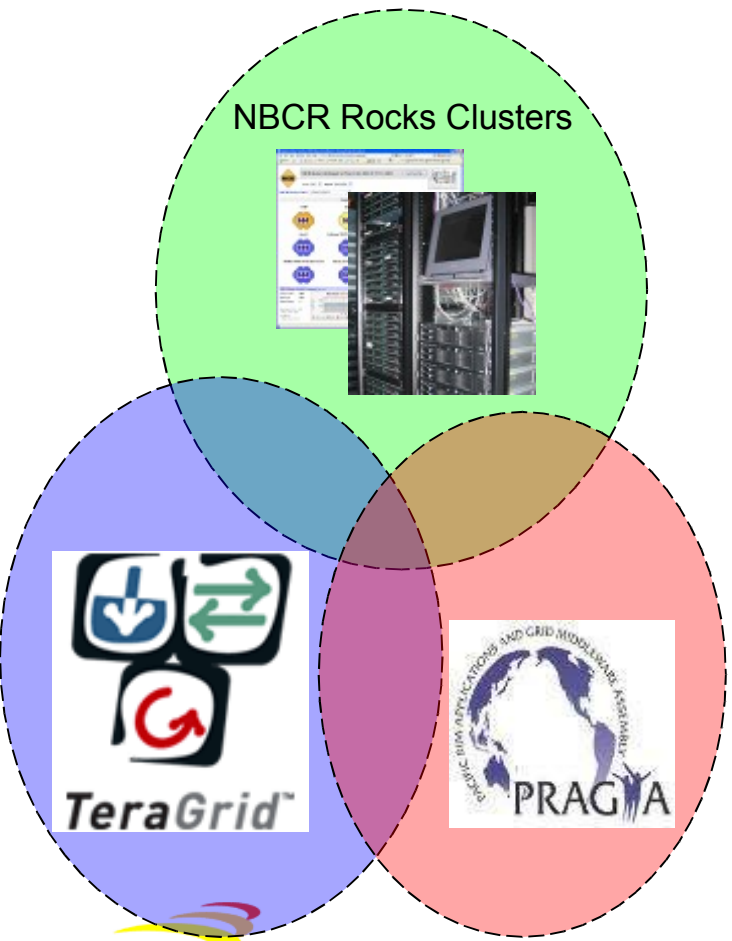
National Biomedical Computation Resource

an NIH supported resource center

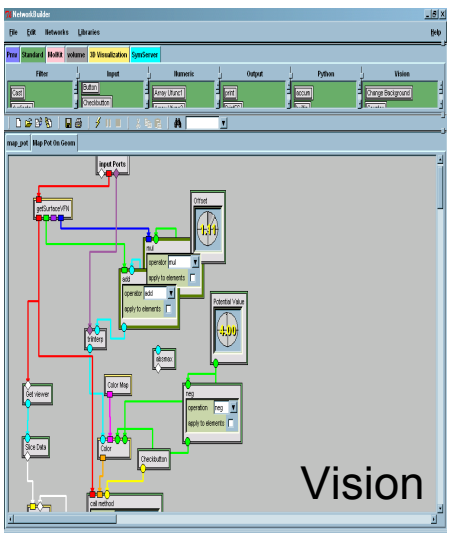
Located in Calit2@UCSD Building

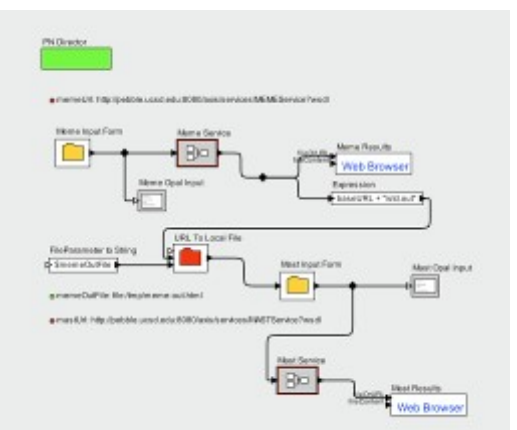
yberinfrastructure: Raw Resources, Middleware & Execution Environment

Virtual Organizations



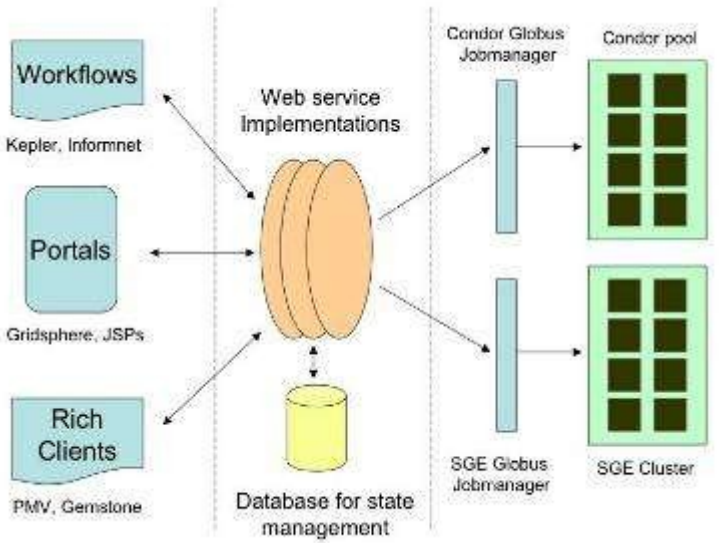
Workflow Management





KEPLER

Web Services



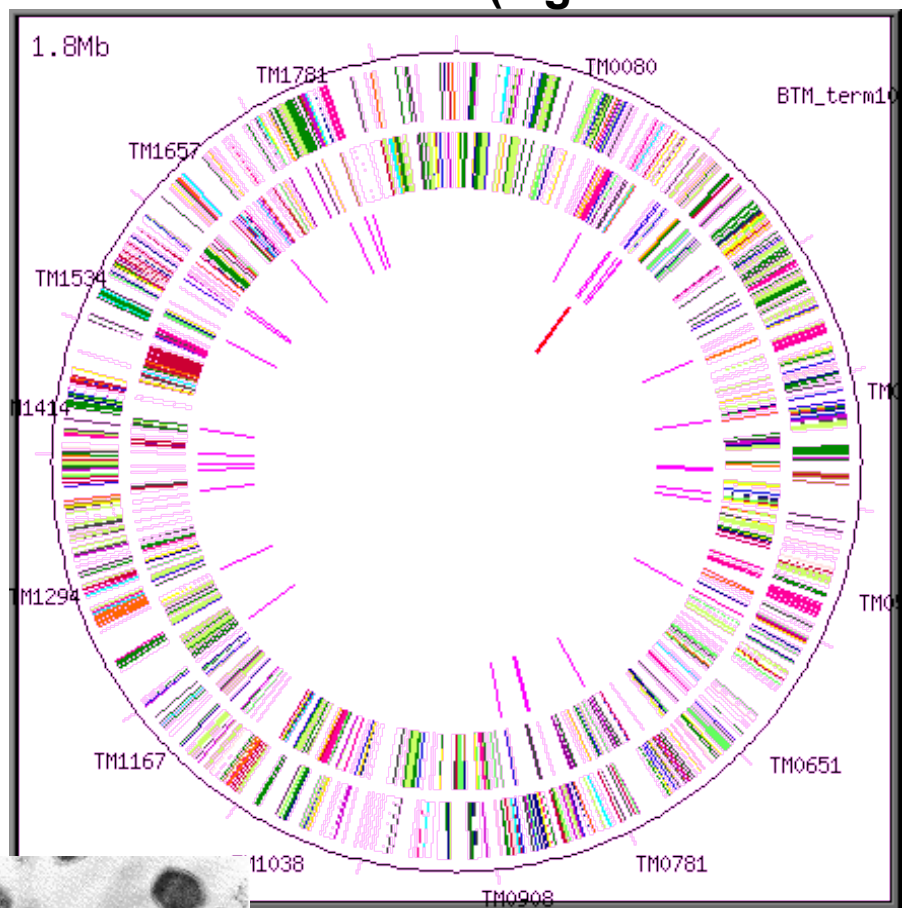
Telescience Portal

TS telescience project

it2

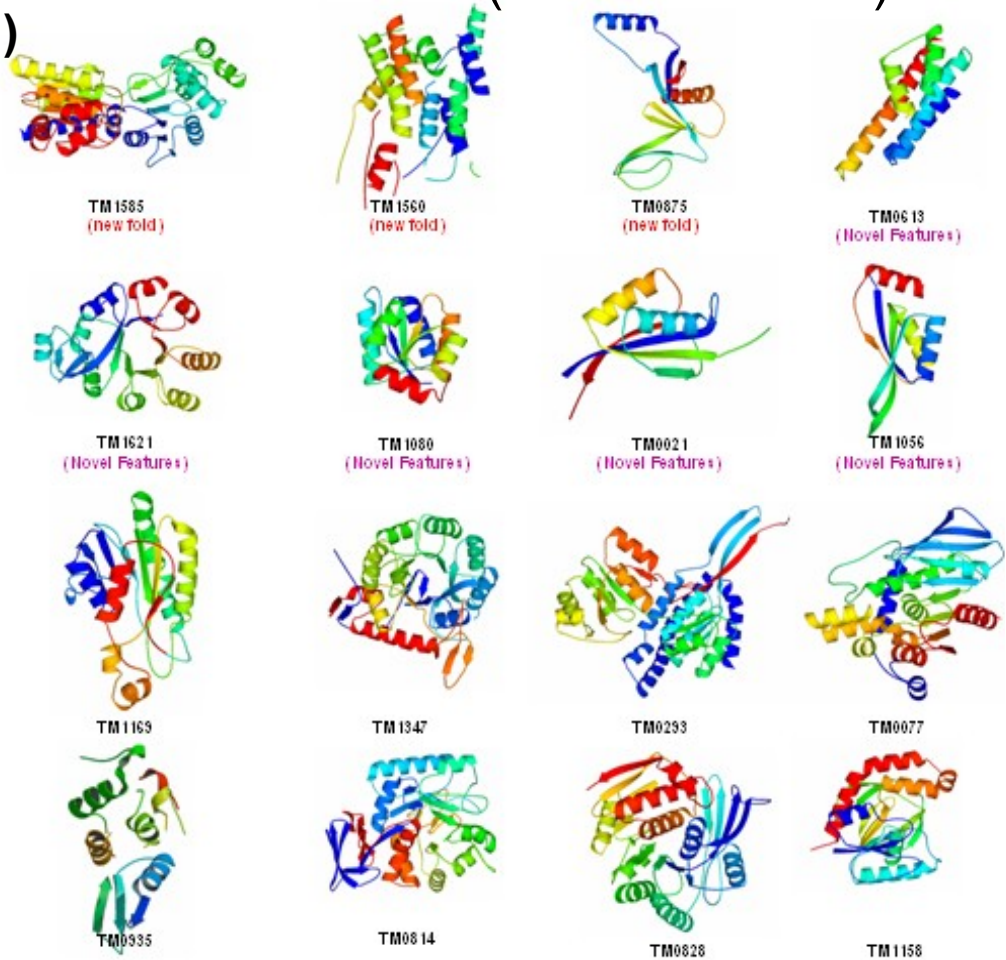
The Bioinformatics Core of the Joint Center for Structural Genomics will be Housed in the Calit2@UCSD Building

Extremely Thermostable -- Useful for Many Industrial Processes (e.g. Chemical and Food)





173 Structures (122 from JCSG)

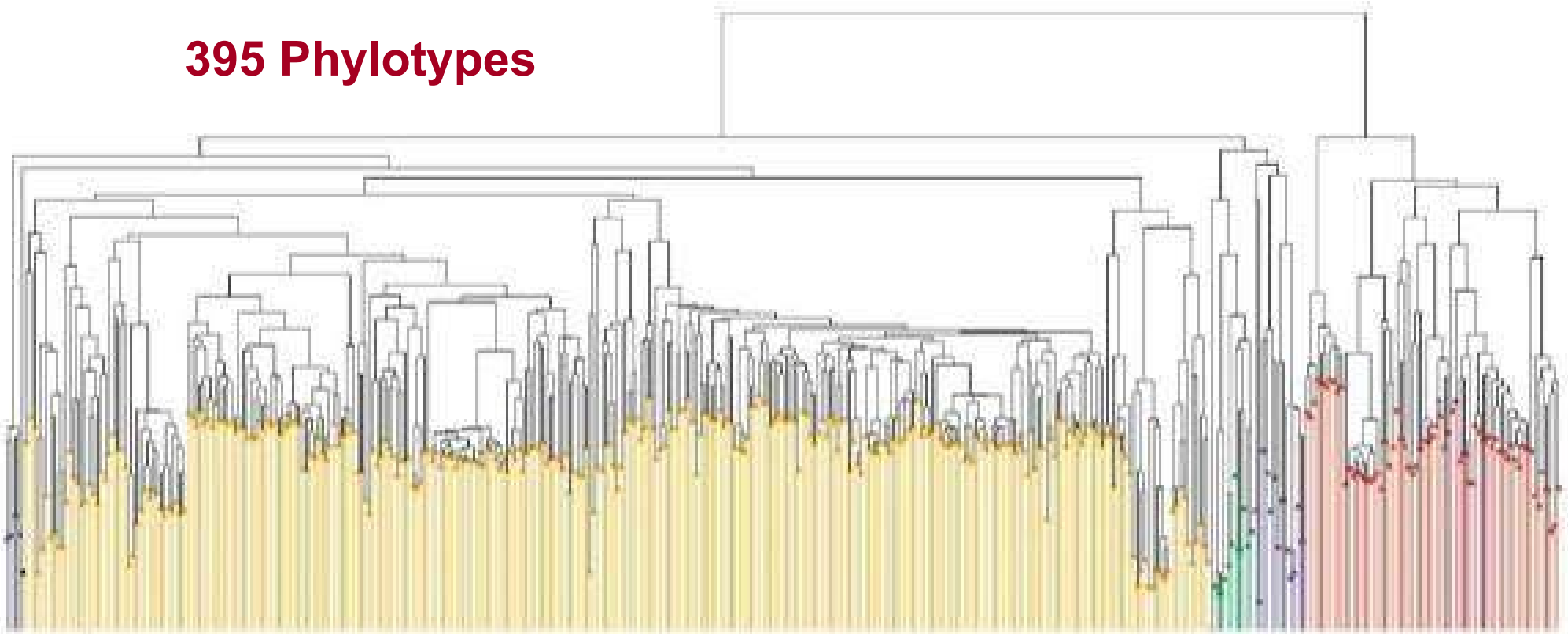


- Determining the Protein Structures of the Thermotoga Maritima Genome
- 122 T.M. Structures Solved by JCSG (75 Unique In The PDB)
- Direct Structural Coverage of 25% of the Expressed Soluble Proteins
- Probably Represents the Highest Structural Coverage of Any Organism

Source: John Wooley, UCSD

Calit2 is Discussing Including Other Metagenomic Data Sets

- A majority of the bacterial sequences corresponded to uncultivated species and novel microorganisms.
- We discovered significant intersubject variability.
- Characterization of this immensely diverse ecosystem is the first step in elucidating its role in health and disease.



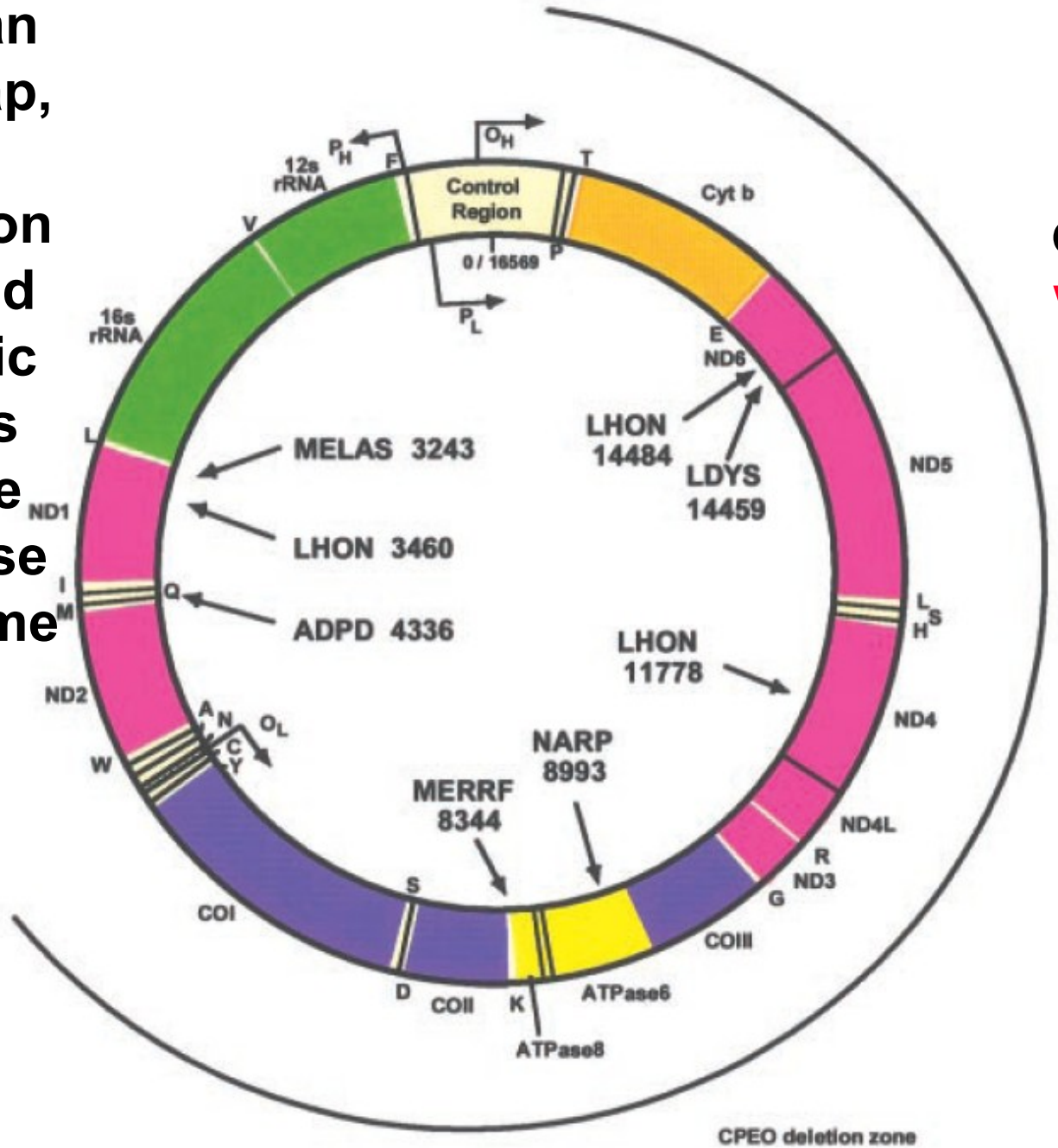
OptIPuter

"Diversity of the Human Intestinal Microbial Flora"
Paul B. Eckburg, et al *Science* (10 June 2005)

it2

Calit2 is Collaborating with Douglas Wallace-- Planning to Bring MITOMAP into Calit2 Domain

The Human mtDNA Map, Showing the Location of Selected Pathogenic Mutations Within the 16,569-Base Pair Genome



MITOMAP: A Human Mitochondrial Genome Database. www.mitomap.org, 2005

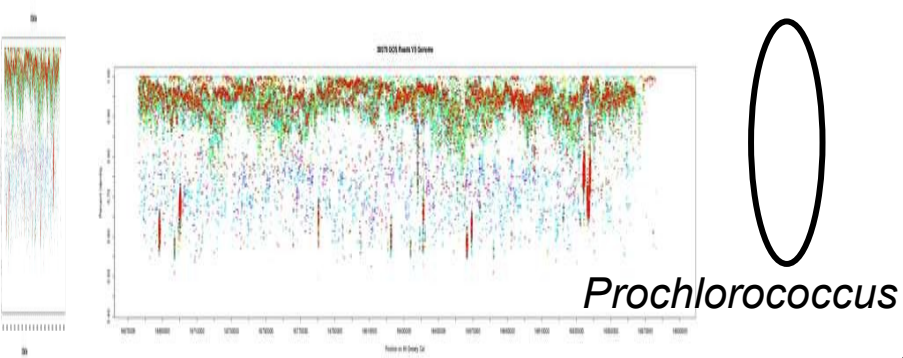


5 March 1999

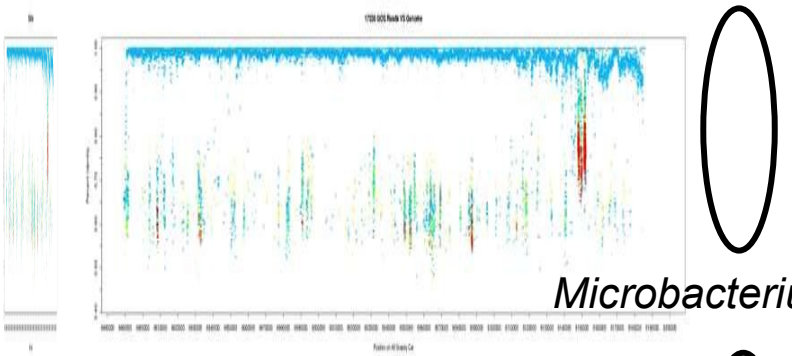
OptIPuter

it2

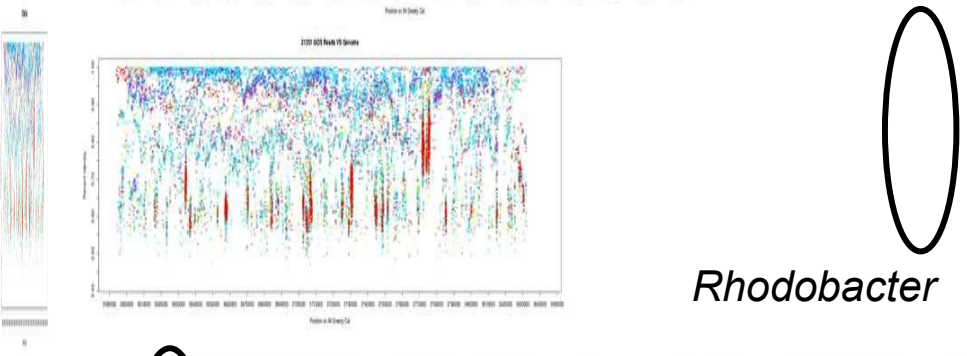
Metagenomics "Extreme Assembly" Requires Large Amount of Pixel Real Estate



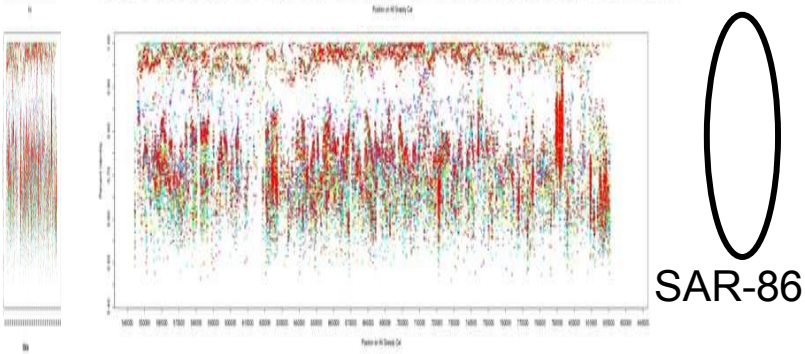
Prochlorococcus



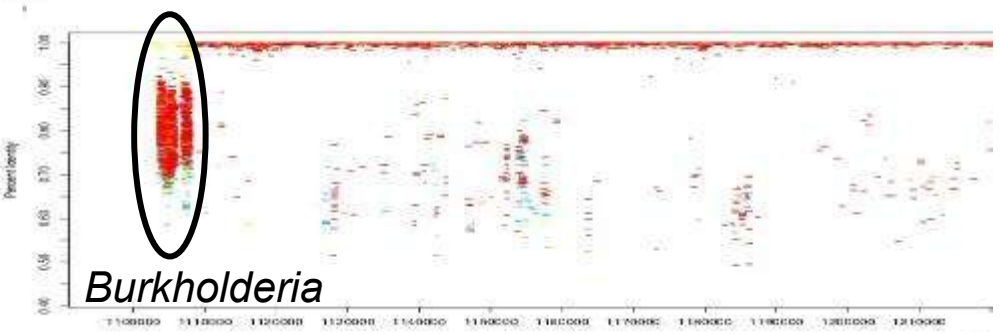
Microbacterium



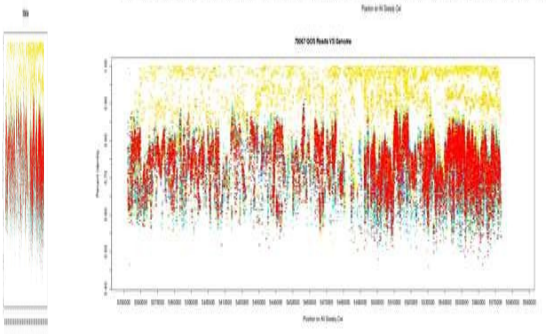
Rhodobacter



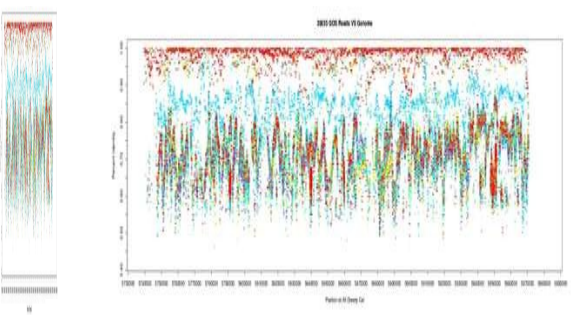
SAR-86



Burkholderia



unknown

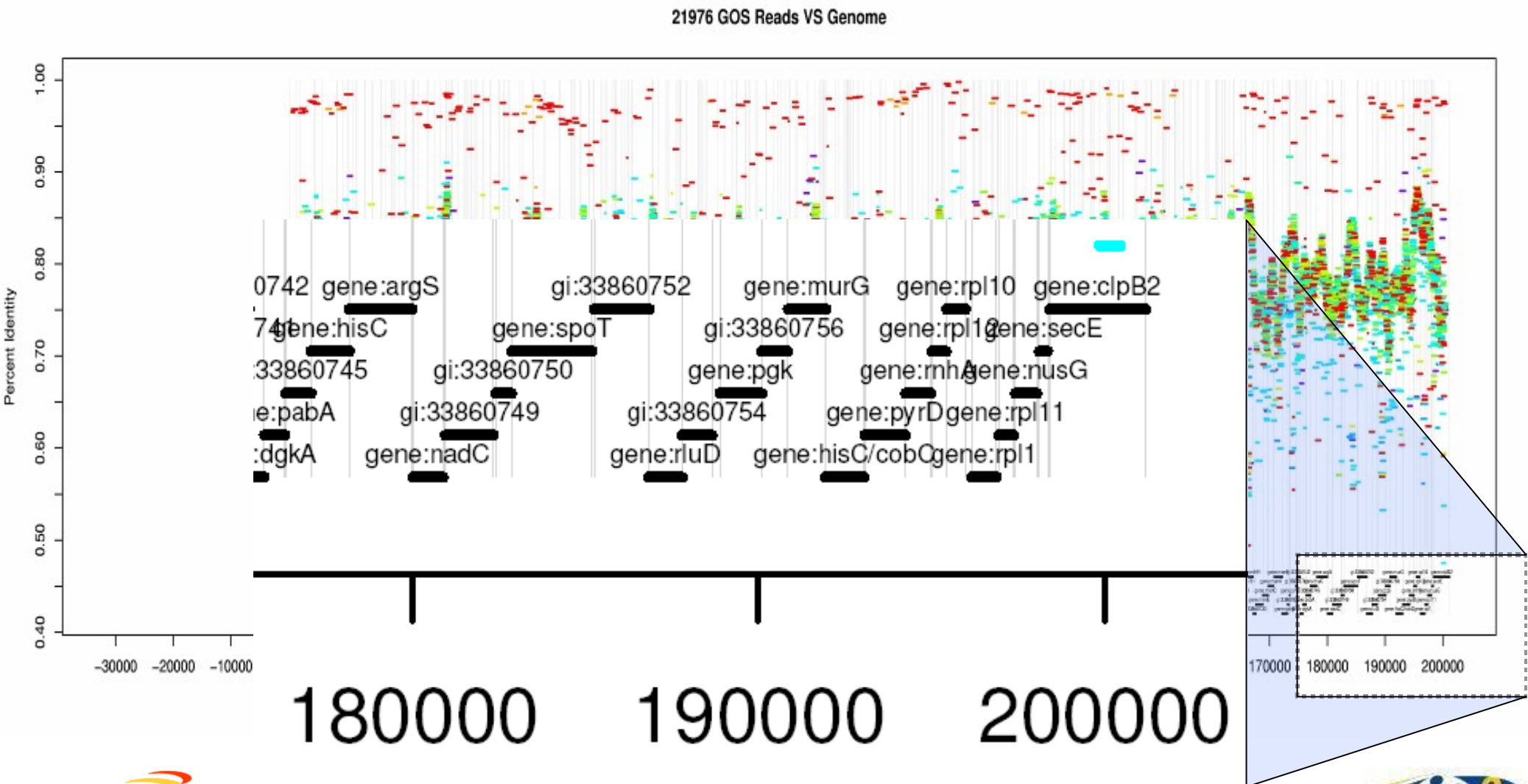


unknown

Source: Karin Remington
J. Craig Venter Institute

Metagenomics Requires a Global View of Data and the Ability to Zoom Into Detail Interactively

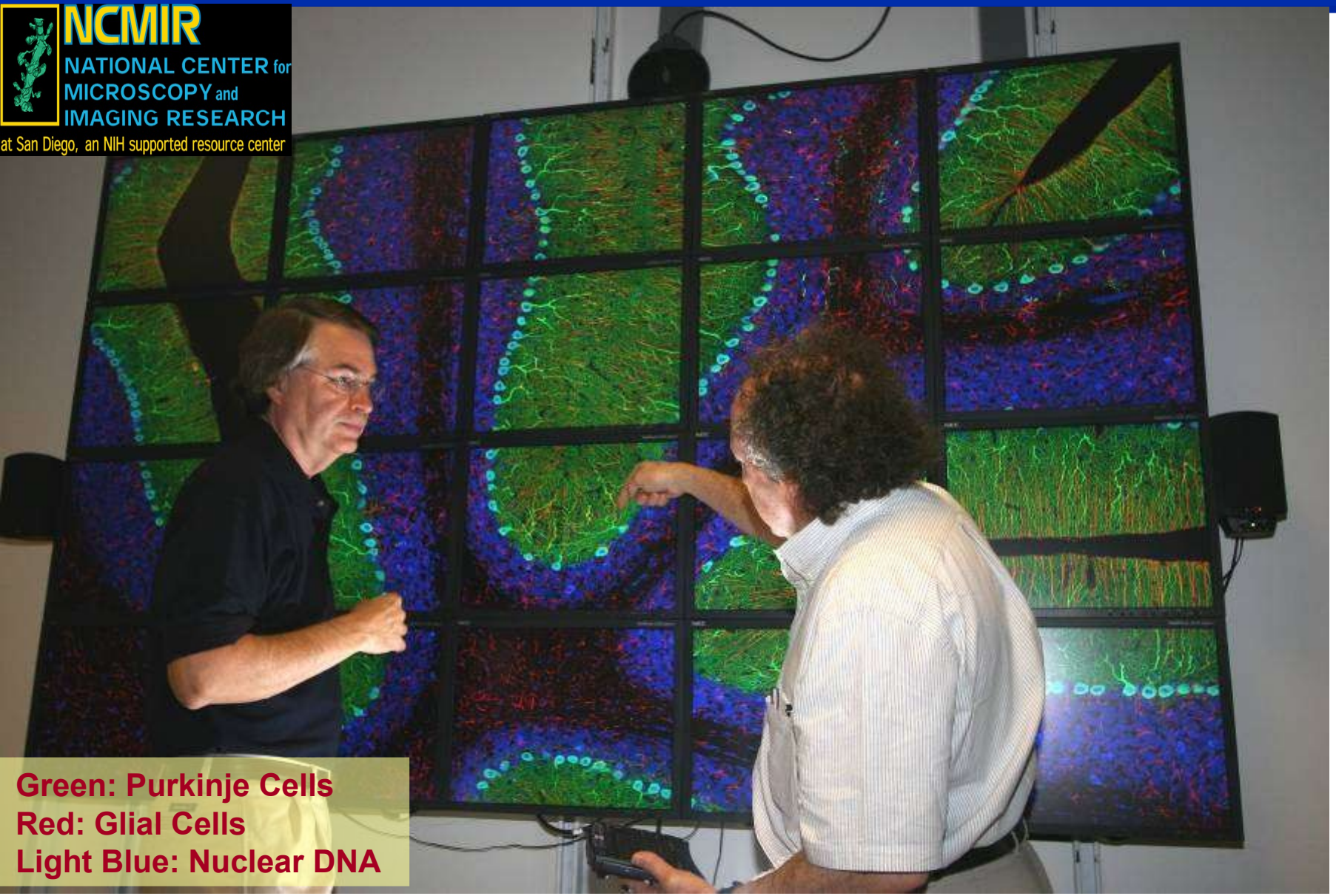
Overlay of Metagenomics Data onto Sequenced Reference Genomes (This Image: Prochloroccocus marinus MED4)



Source: Karin Remington
J. Craig Venter Institute

The OptIPuter – Creating High Resolution Portals Over Dedicated Optical Channels to Global Science Data

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



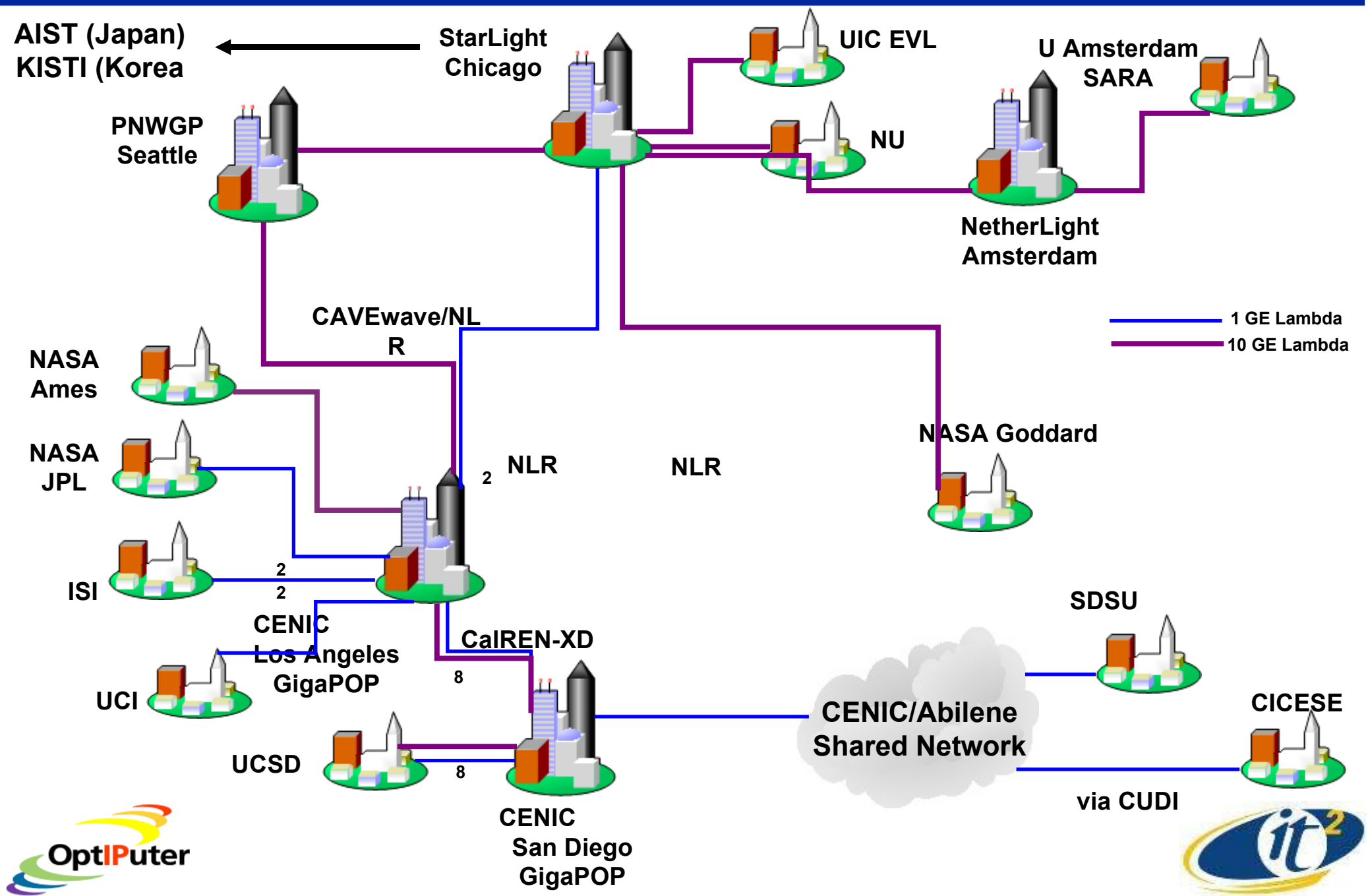
Green: Purkinje Cells
Red: Glial Cells
Light Blue: Nuclear DNA

Source: Mark Ellisman, David Lee, Jason Leigh

OptIPuter

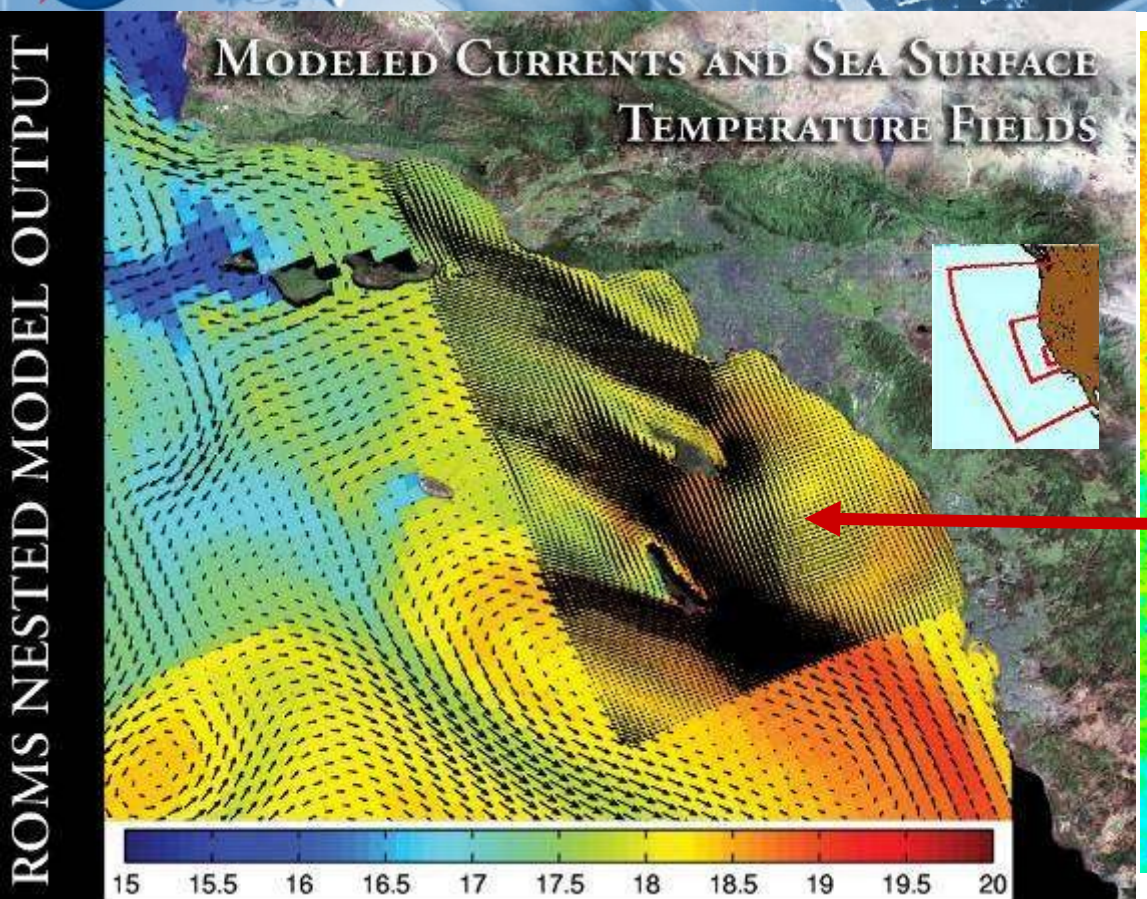
Calit2 (UCSD, UCI) and UIC Lead Campuses—Larry Smarr PI
Partners: SDSC, USC, SDSU, NW, TA&M, UvA, SARA, KISTI, AIST

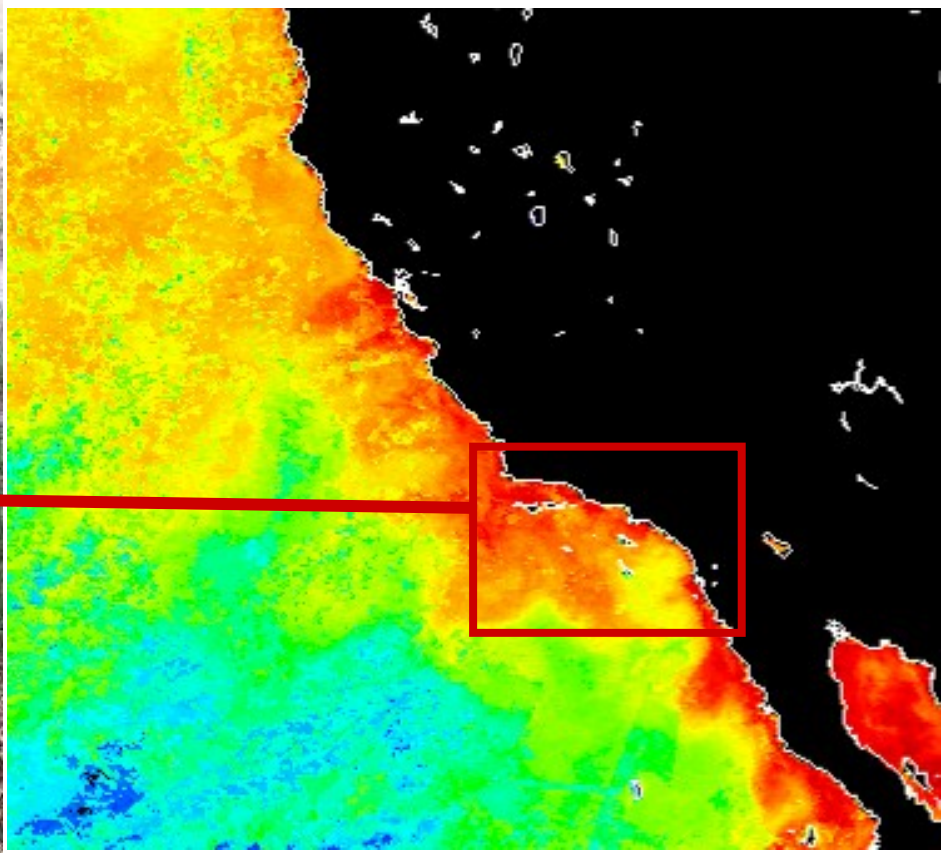
Expanding the OptIPuter LambdaGrid



Using the OptIPuter to Couple Data Assimilation Models to Remote Data Sources Including Biology







NASA MODIS Mean Primary Productivity for April 2001 in California Current System

JPL ONR MBARI

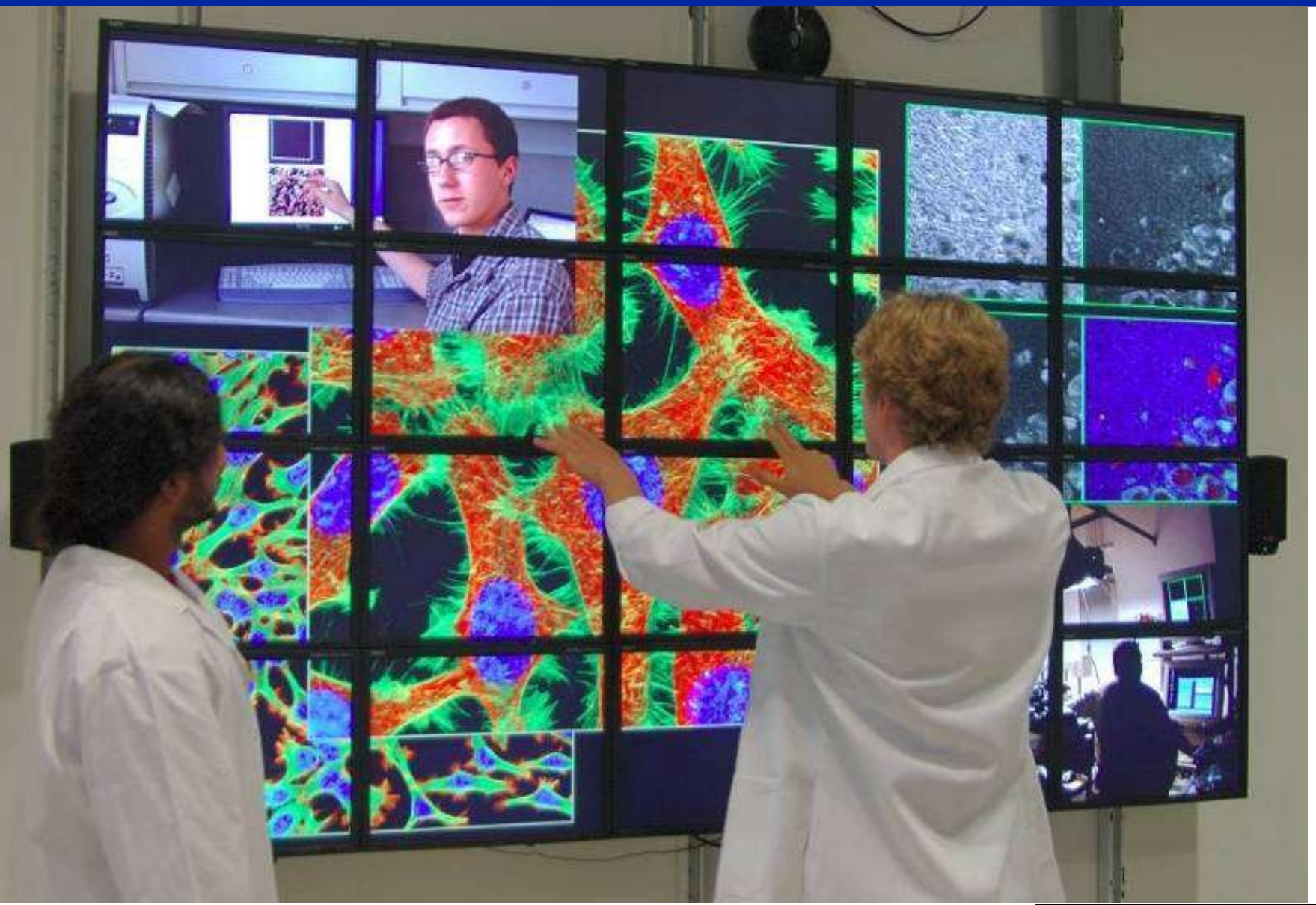
OptIPuter

Regional Ocean Modeling System (ROMS)

http://ourocean.jpl.nasa.gov/

it2

OptIPuter Scalable Adaptive Graphics Environment (SAGE) Allows Integration of HD Streams



OptIPuter

evl electronic visualization laboratory

Source: David Lee, NCMIR, UCSD

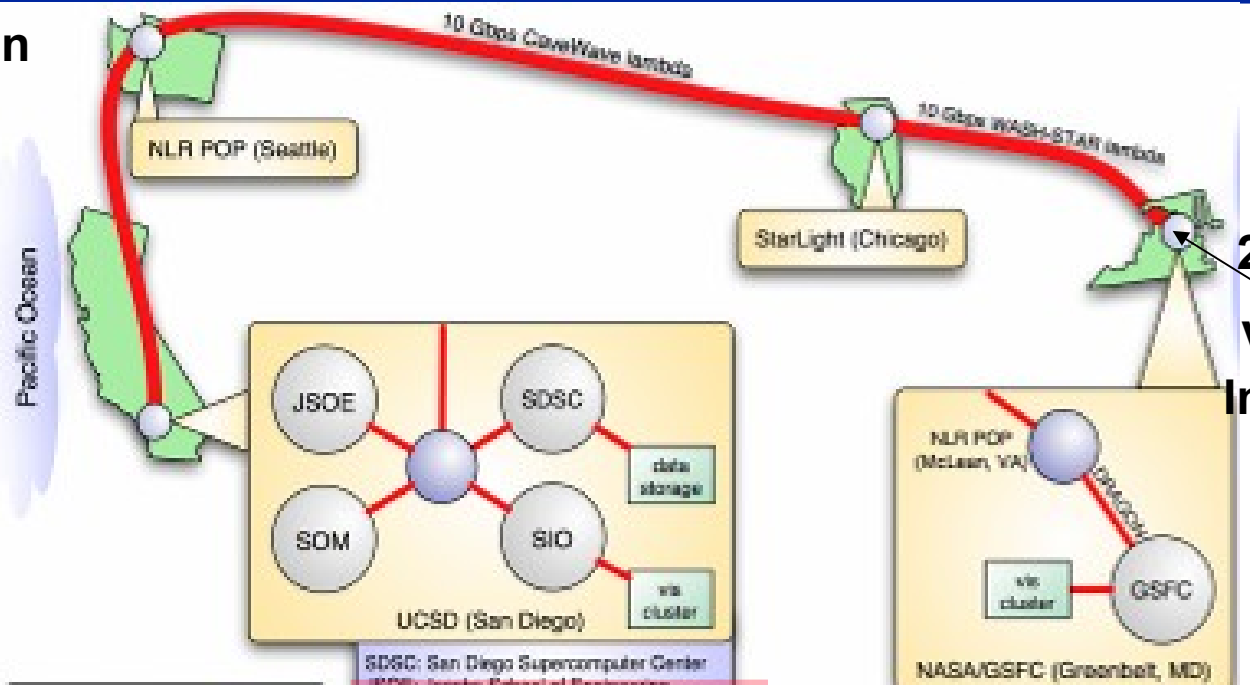
NCMIR NATIONAL CENTER for MICROSCOPY and IMAGING RESEARCH at San Diego, an NIH supported resource center

it2

Calit2 and the Venter Institute Will Combine Telepresence with Remote Interactive Analysis

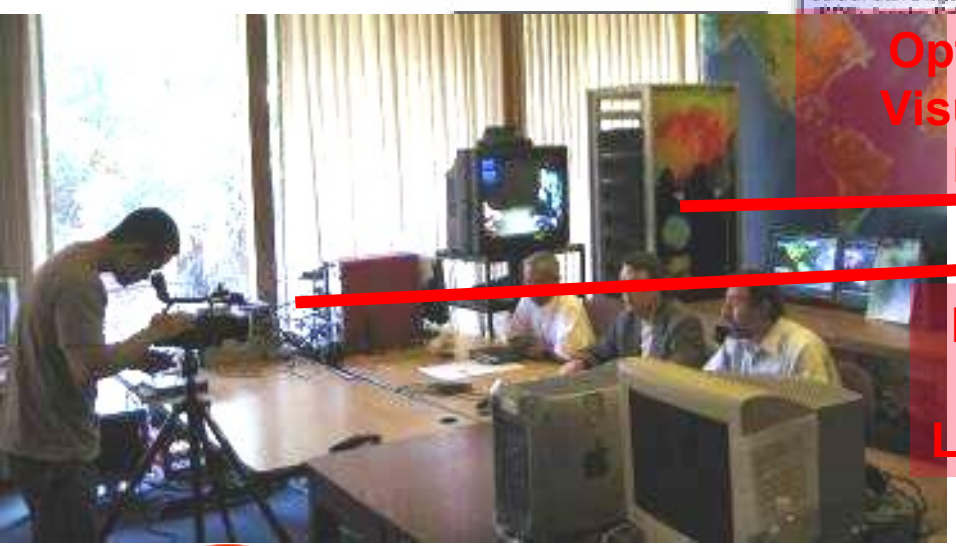
Live Demonstration

of 21st Century National-Scale Team Science



25 Miles

Venter Institute



OptIPuter Visualized Data

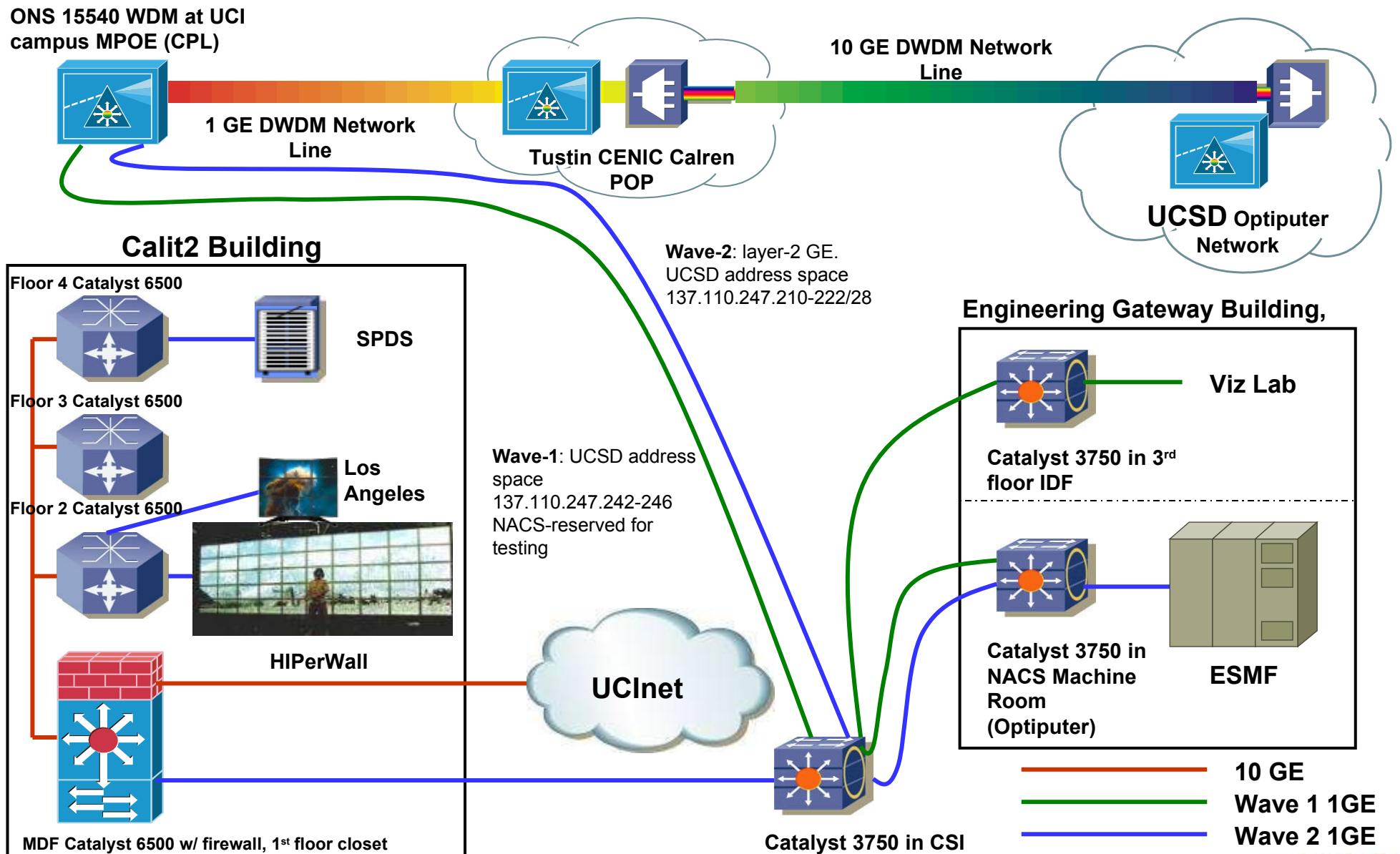
HDTV Over Lambda



OptIPuter

it2

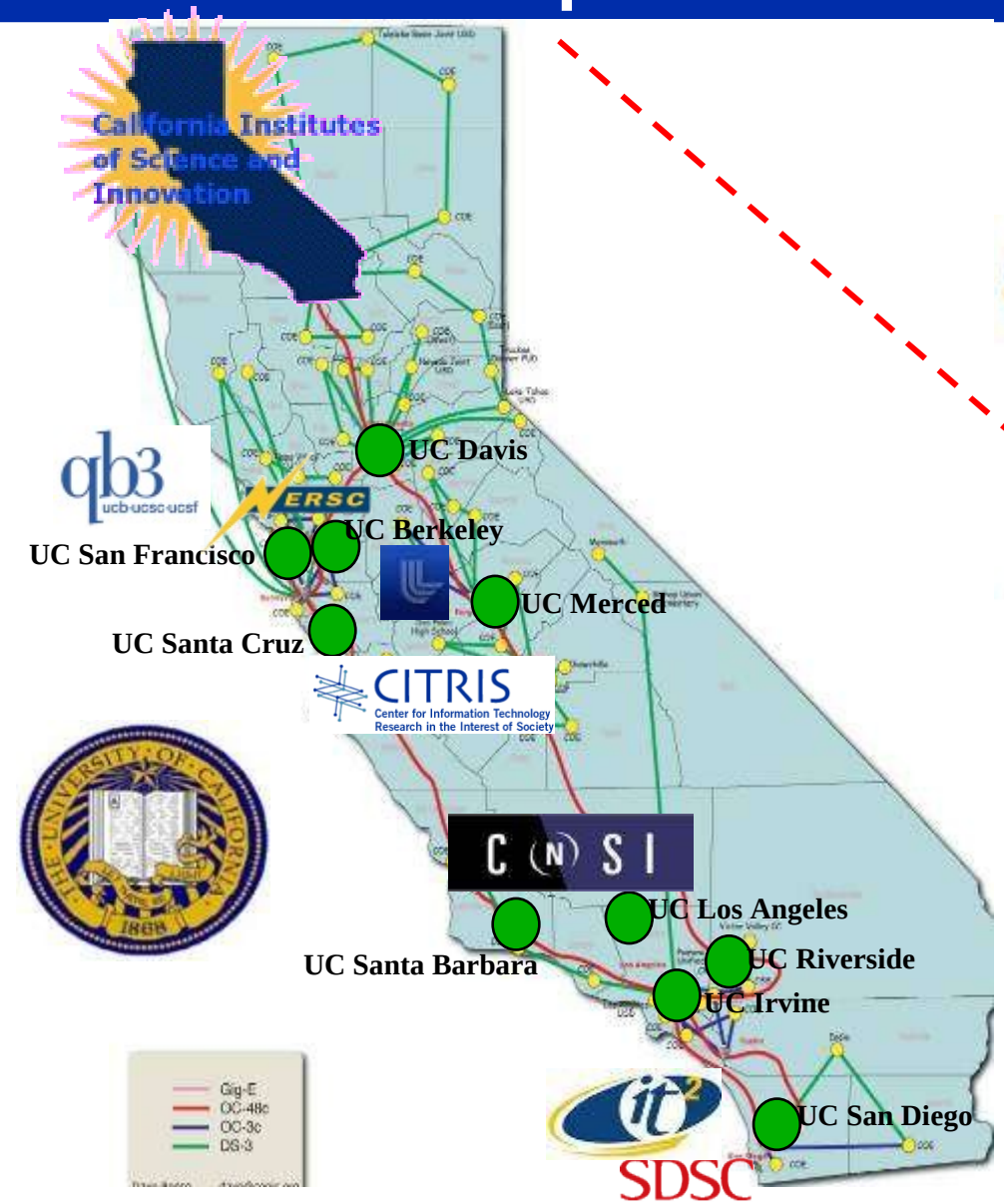
Calit2@UCI Will Be the "Beta-Test" Campus for Accessing CAMERA



Created 09-27-2005 by Garrett Hildebrand

Modified 11-03-2005 by Jessica Yu

Calit2/SDSC Proposal to Create a UC Cyberinfrastructure of "On-Ramps" to National LambdaRail Resources



OptIPuter + CalREN-XD + TeraGrid = "OptiGrid"



Creating a Critical Mass of End Users on a Secure LambdaGrid

OptIPuter

Source: Fran Berman, SDSC , Larry Smarr, Calit2

Lambda Connectivity to CAMERA Will Enable International Scientific Collaboration on Marine Microbial Metagenomics

CICESE

Departament of
BIOLOGICAL OCEANOGRAPHY

Dedicated to Biologican and Fisheries Oceanography

EN Spanish | Graduate Studies | Library | CICESE

Biological Oceanography
- Home
- Mission & vision

Directory
- Scientific staff
- Research assistants
- PhD students
- MSc students

Productivity

WELCOME

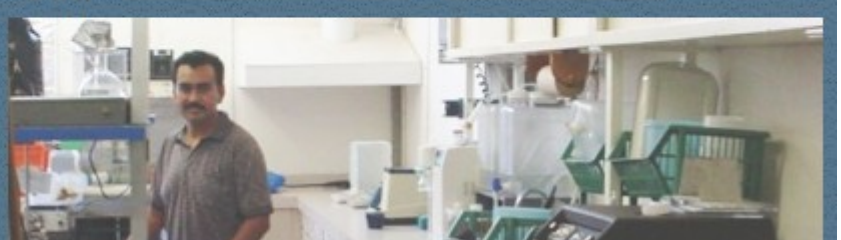


Related sites
- Oceanology Division
- Marine Sciences
- Societies
- IMECOCAL
- FLUCAR
- Agenda 21 Local
- CICESE magazine
- Books & journals

SIO and CICESE Have 30-Year History of Collaboration

Ecology of Marine Plankton

Phytoplankton and bacteria culture; photosynthesis, respiration, in vivo fluorescence and absorption of phytoplankton; inorganic carbon and oxygen in seawater. [ext. 242-86]



OptIPuter

it2

CUDI-CENIC Fiber Dedication at Border Governor's Conference, July 14, 2005



US Mexico

Torreon Conference---Fiber Dedication Linking Mexico and US, crossing at San Diego-Tijuana



- **Shared Security**
- **Energy**
- **Trans-National Crime**
- **Education and Research**
- **Business Development**

Culmination of Three Years of Work Between Calit2, CICESE, CENIC, and CUDI

OptIPuter

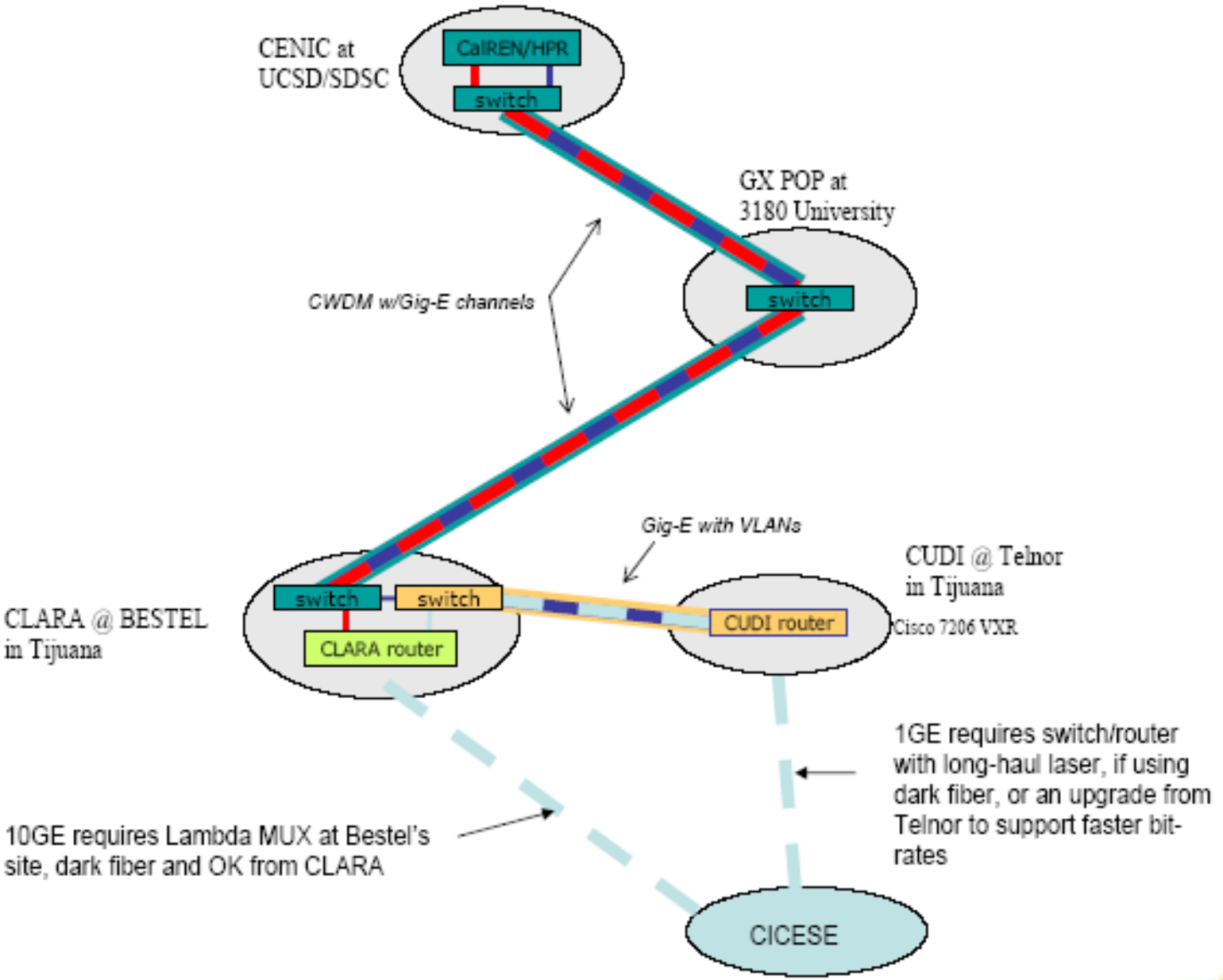


cenic
Corporation for Education Network Initiatives in California

LEADING THE WAY TO TOMORROW'S INTERNET

it2

We are Very Close to Setting Up a Gigabit Lambda Between Calit2 and CICESE



Source: Raúl Hazas, CICESE