Genomics at the Speed of Light: Understanding the Living Ocean

Invited Talk
JASON Summer Program
La Jolla, CA
July 12, 2006

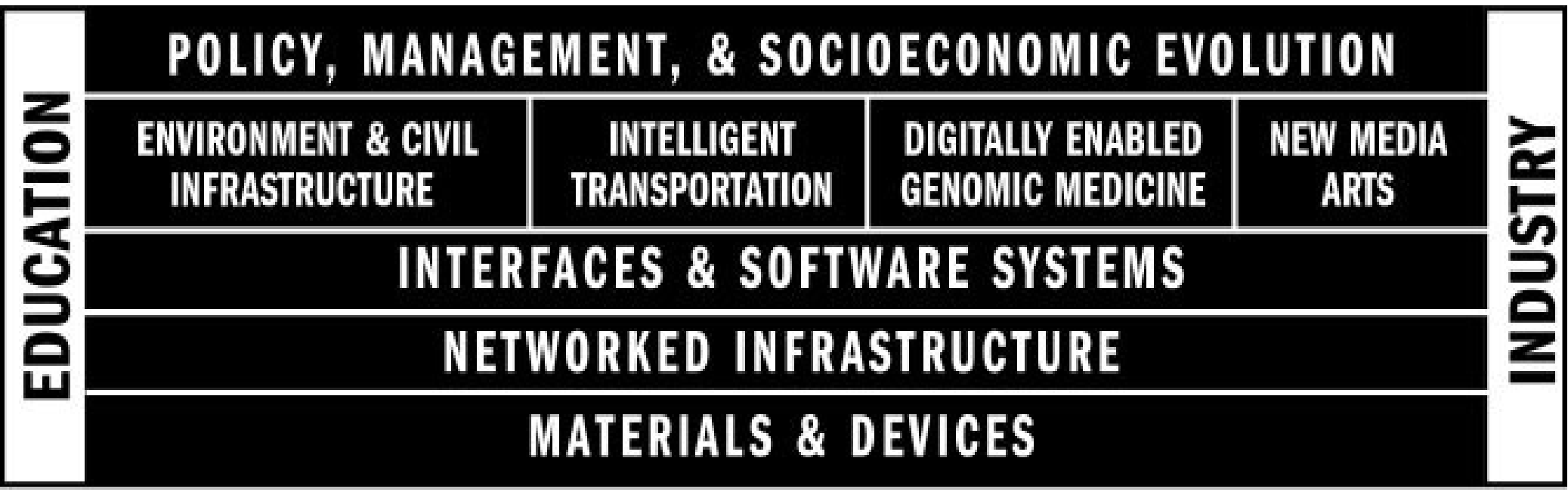
Dr. Larry Smarr
Director, California Institute for Telecommunications and Information Technologies
Harry E. Gruber Professor,
Dept. of Computer Science and Engineering
Jacobs School of Engineering, UCSD

OptIPuter

it²

Calit2 -- Research and Living Laboratories on the Future of the Internet

UC San Diego & UC Irvine Faculty
Working in Multidisciplinary Teams
With Students, Industry, and the Community



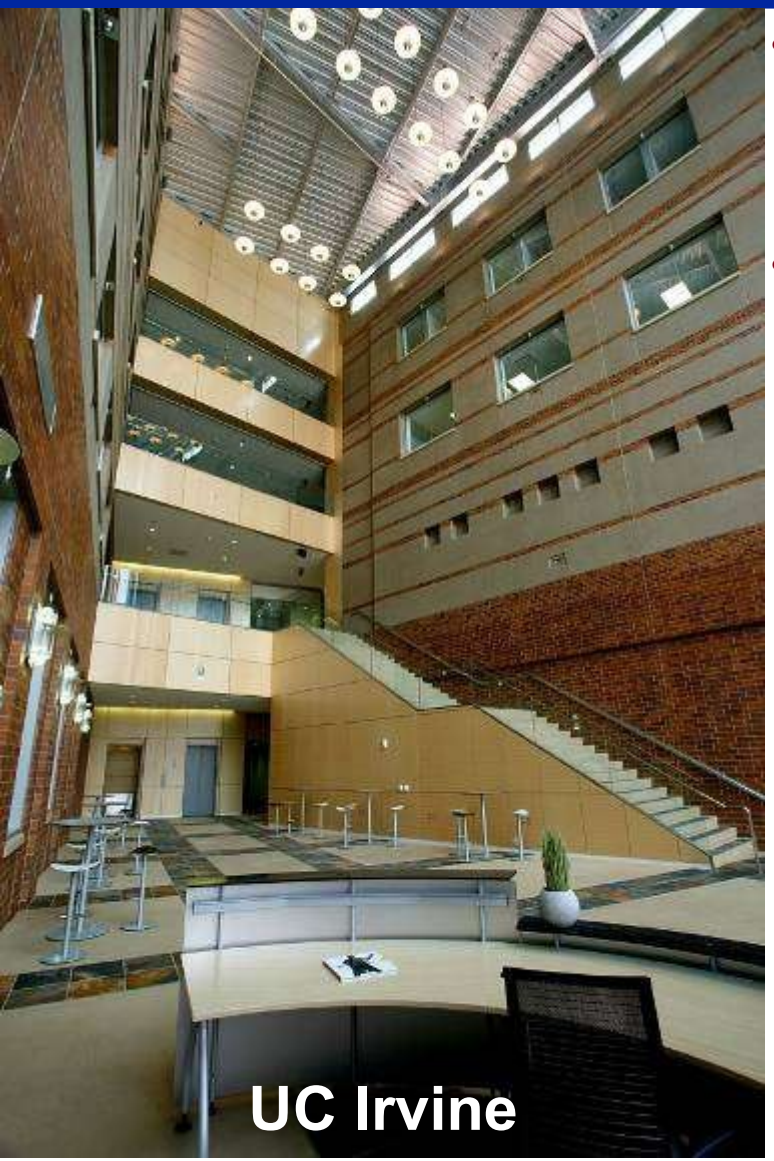
www.calit2.net

California Institute for Telecommunications and Information Technology . Cal-(IT)^2



Extending the Internet Throughout the Physical World

Two New Calit2 Buildings Will Provide Major New Laboratories to Their Campuses



- New Laboratory Facilities
  - Nanotech, BioMEMS, Chips, Radio, Photonics, Grid, Data, Applications
  - Virtual Reality, Digital Cinema, HDTV, Synthesis
- Over 1000 Researchers in Two Buildings
  - Linked via Dedicated Optical Networks
  - International Conferences and Testbeds

Richard C. Atkinson Hall Dedication Oct. 28, 2005



www.calit2.net

OptIPuter

Preparing for an World in Which Distance Has Been Eliminated…

it2

Calit2 Brings Computer Scientists and Engineers Together with Biomedical Researchers

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



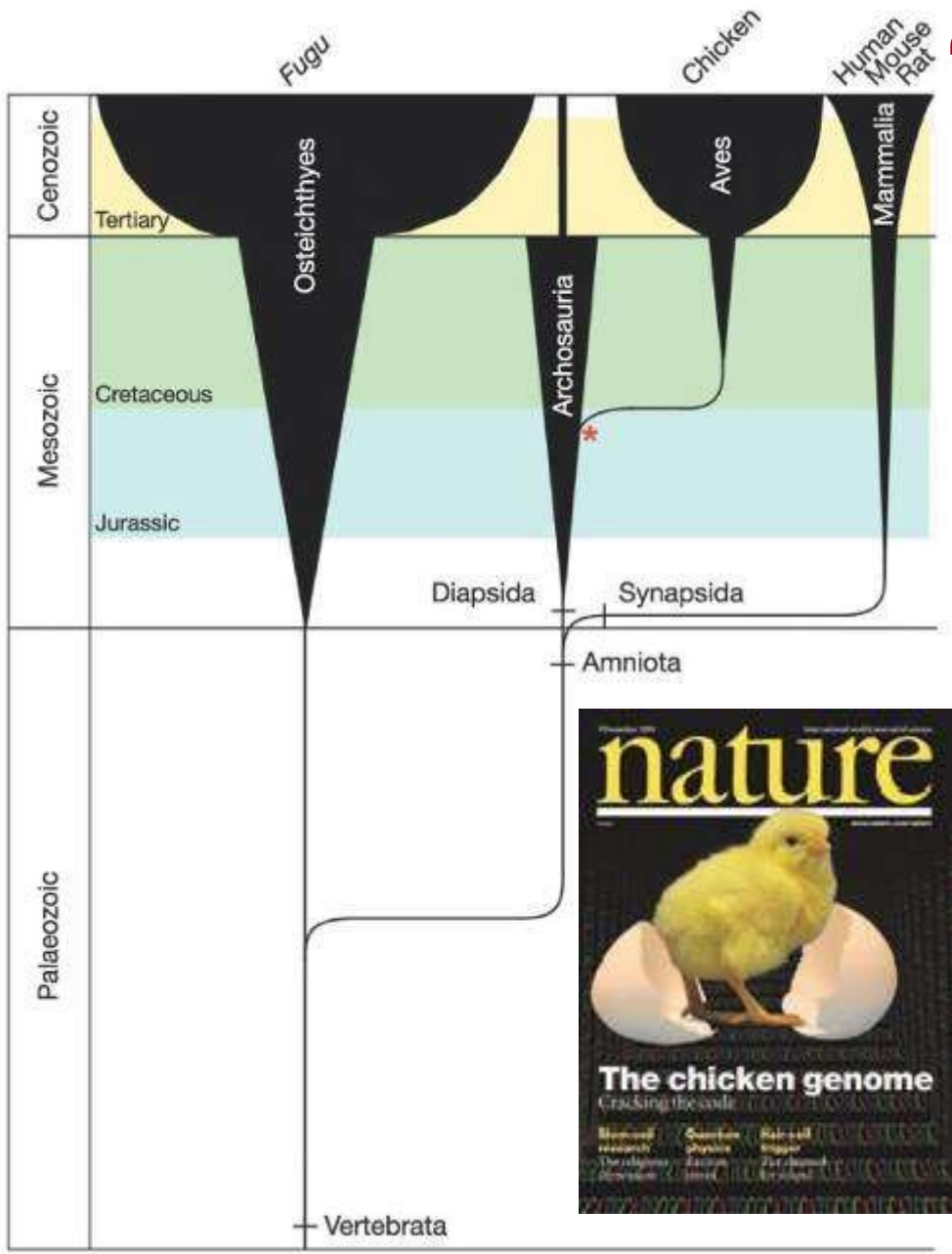


1200 Researchers in Two Buildings

OptIPuter

it2

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



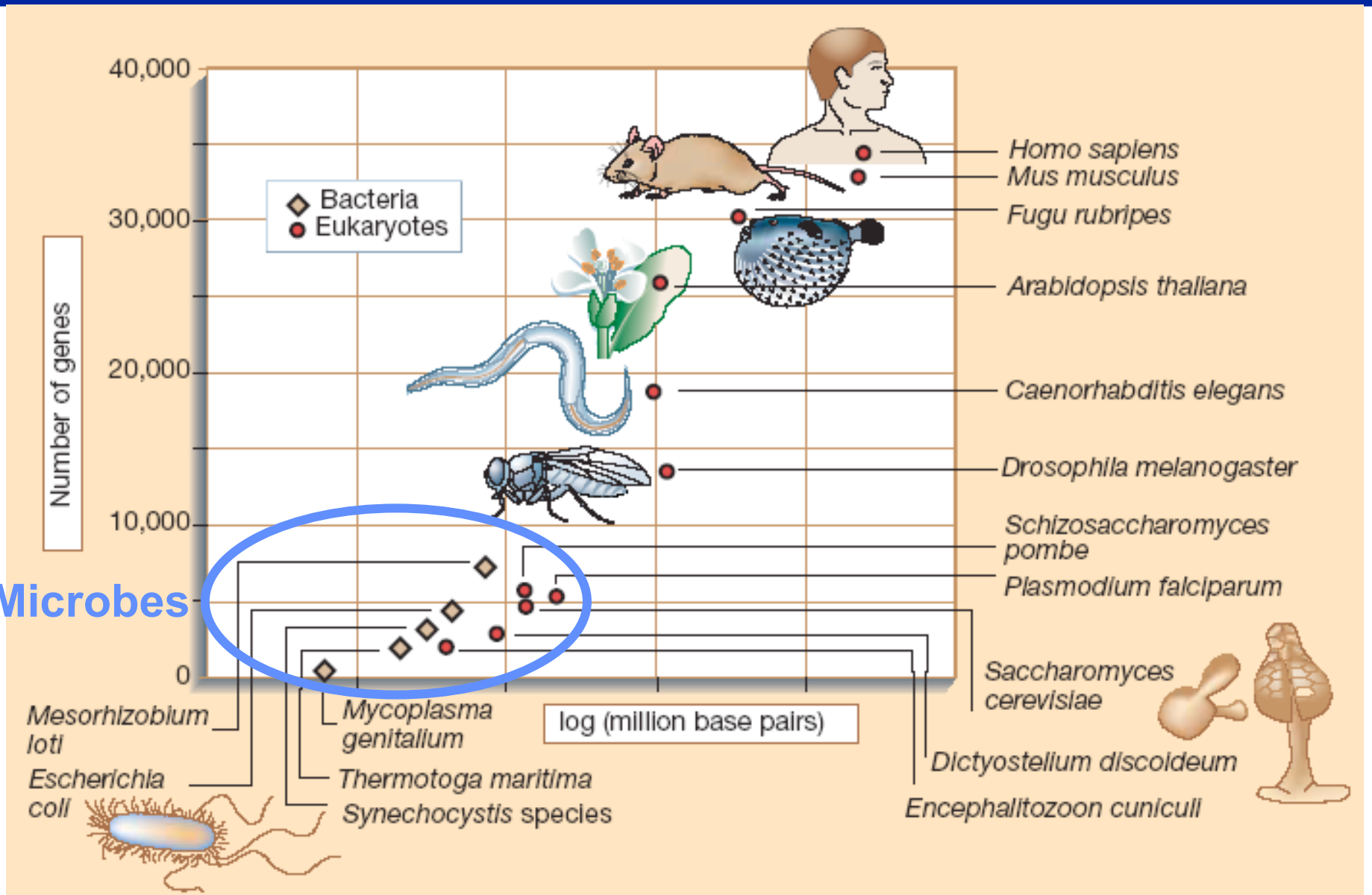
"Many of the chicken–human aligned, non-coding sequences occur far from genes, frequently in clusters that seem to be under selection for functions that are not yet understood."

Nature 432, 695 - 716 (09 December 2004)

OptIPuter

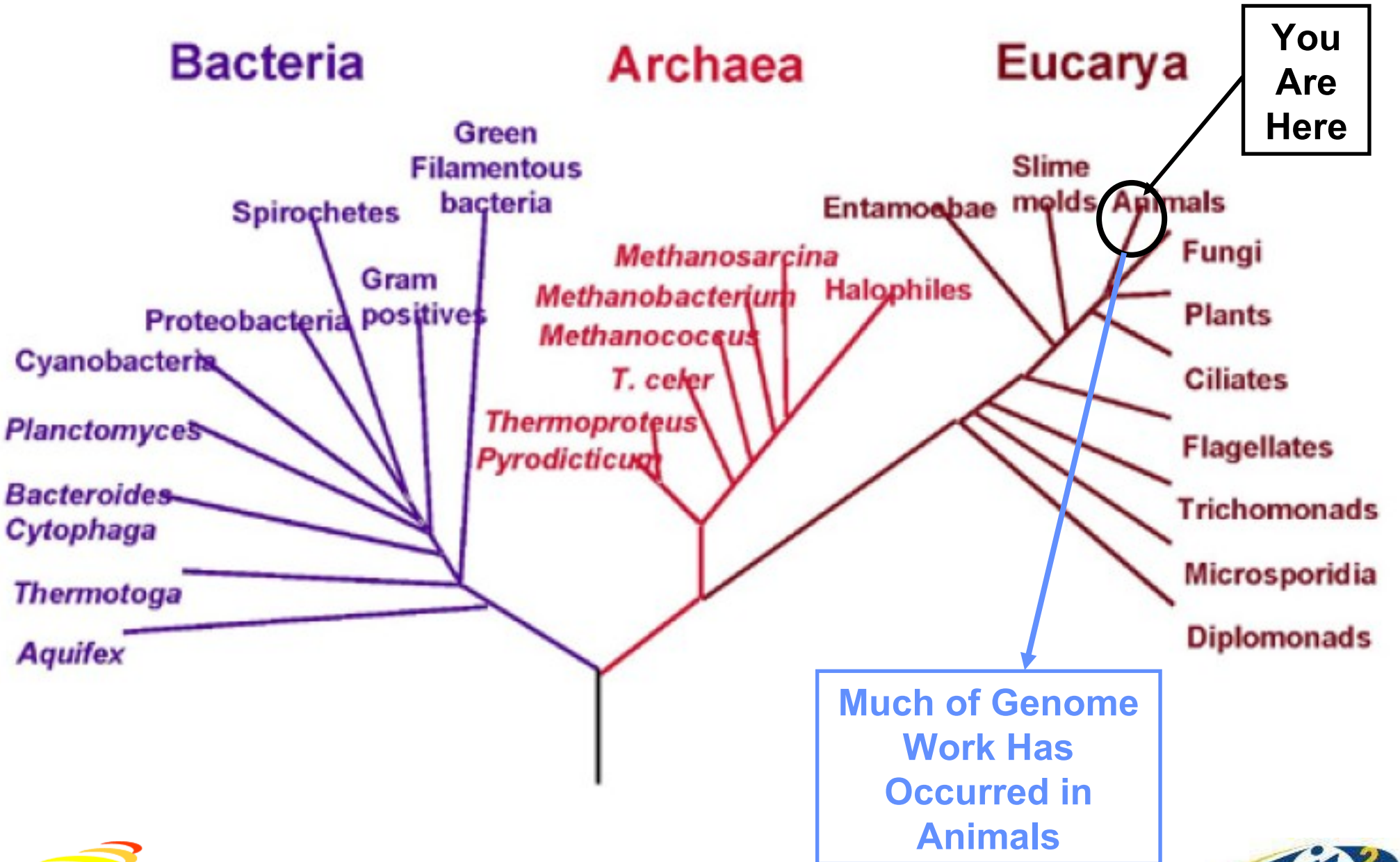
it2

Genomes Range Over Orders of Magnitude in Length



Russell Dolittle, Nature v.419, p. 494 (2002)

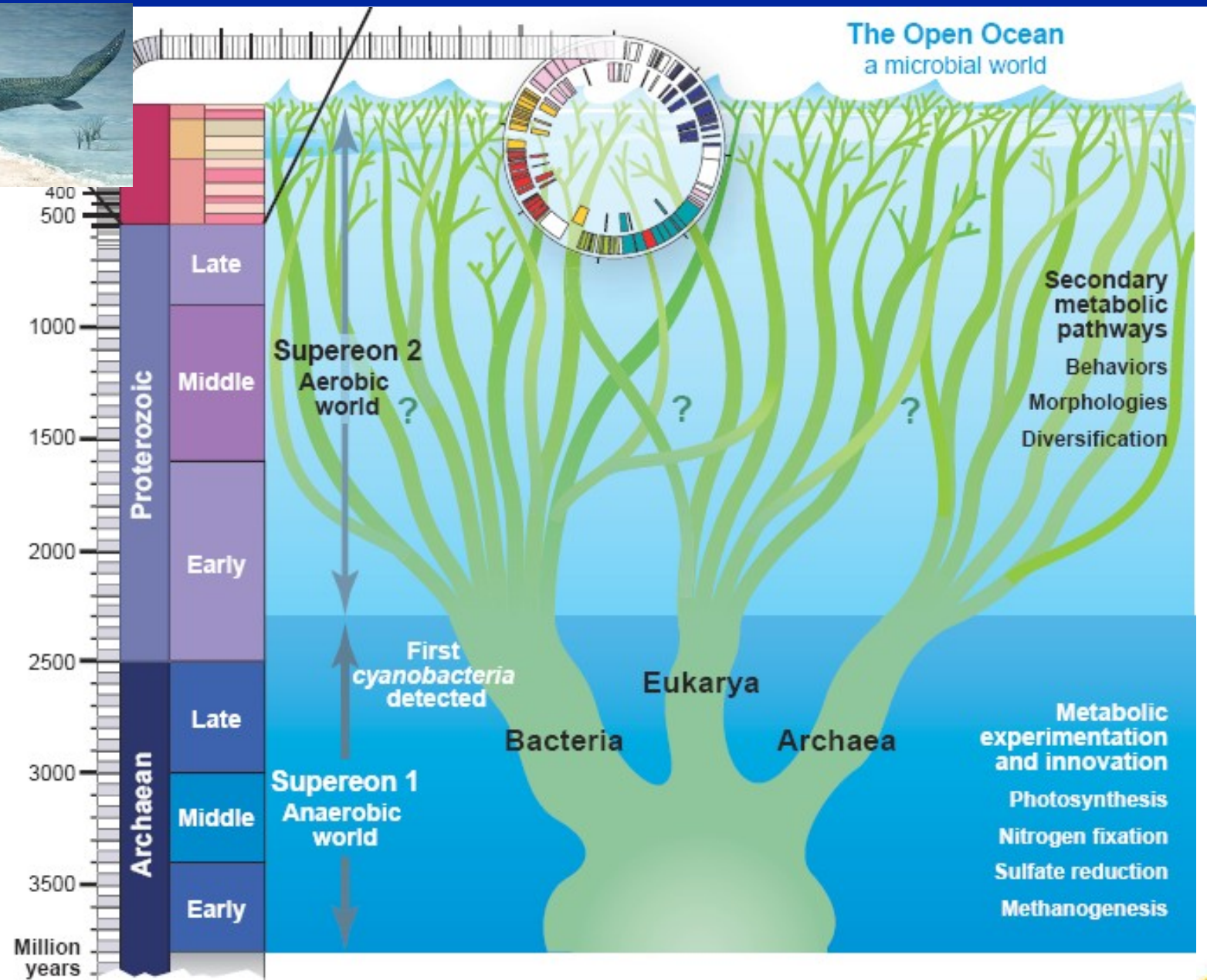
Evolution is the Principle of Biological Systems: Most of Evolutionary Time Was in the Microbial World



Source: Carl Woese, et al

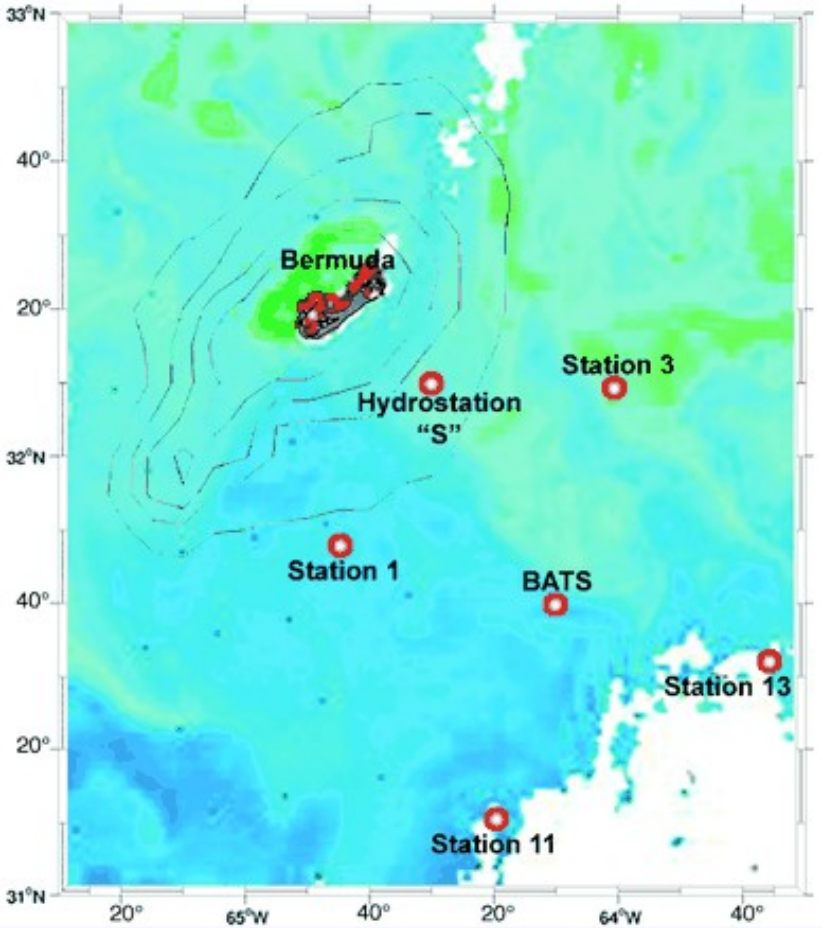
Microbial Genomics Let's Us Look Back Nearly 4 Billion Years In the Evolution of Life





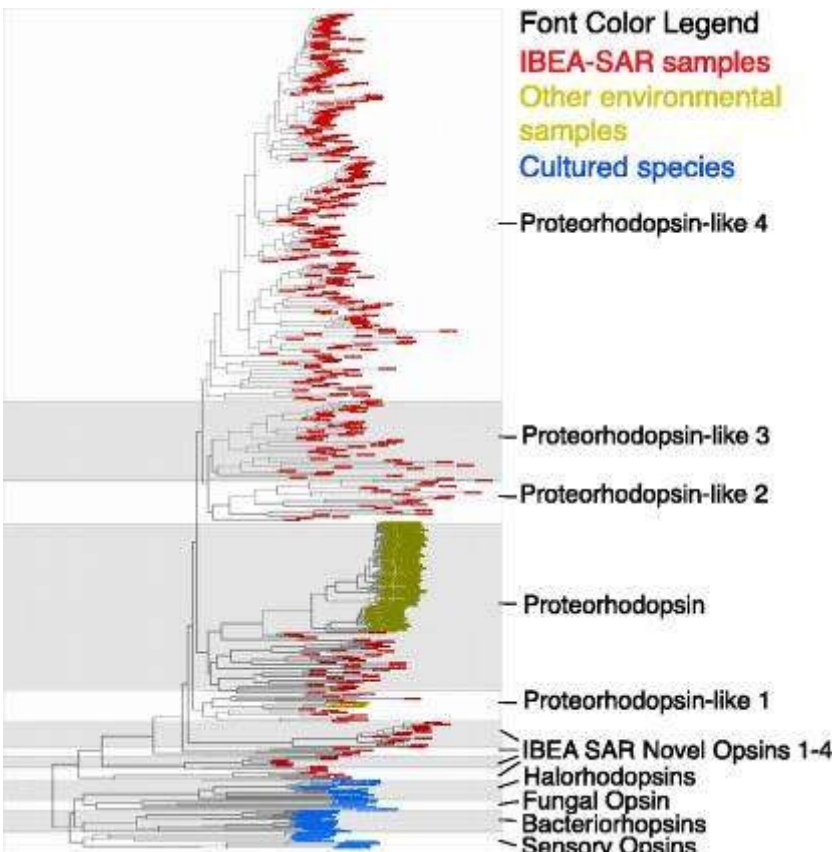
Science Falkowski and Vargas 304 (5667): 58

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



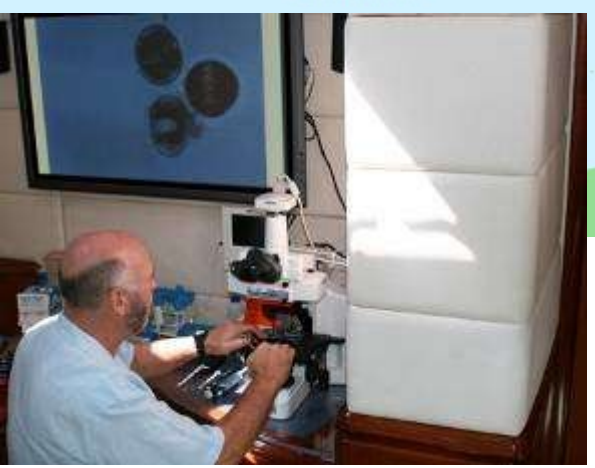
J. Craig Venter, et al.
Science
2 April 2004:
Vol. 304.
pp. 66 - 74

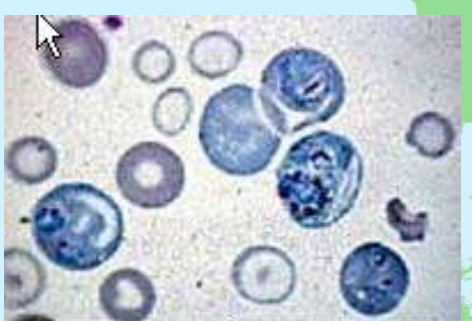
OptIPuter

it2

Marine Genome Sequencing Project
Measuring the Genetic Diversity of Ocean Microbes







Canada
United States
Mexico
Puerto Rico
Honduras
Costa Rica
Panama
Galapagos Islands
x 10 samples
Tanzania/
Zanzibar
x 7 samples
Seychelles
Madagascar
South Africa
Cocos-
Keeling Is.
Vanuatu
x 17 samples
Fiji
Tonga
French
Polynesia
Cook
Islands
Australia
New Caledonia
x 8 samples
Analyzed
Sampled, but not Analyzed

Sorcerer II Data Will Double Number of Proteins in GenBank!

OptIPuter

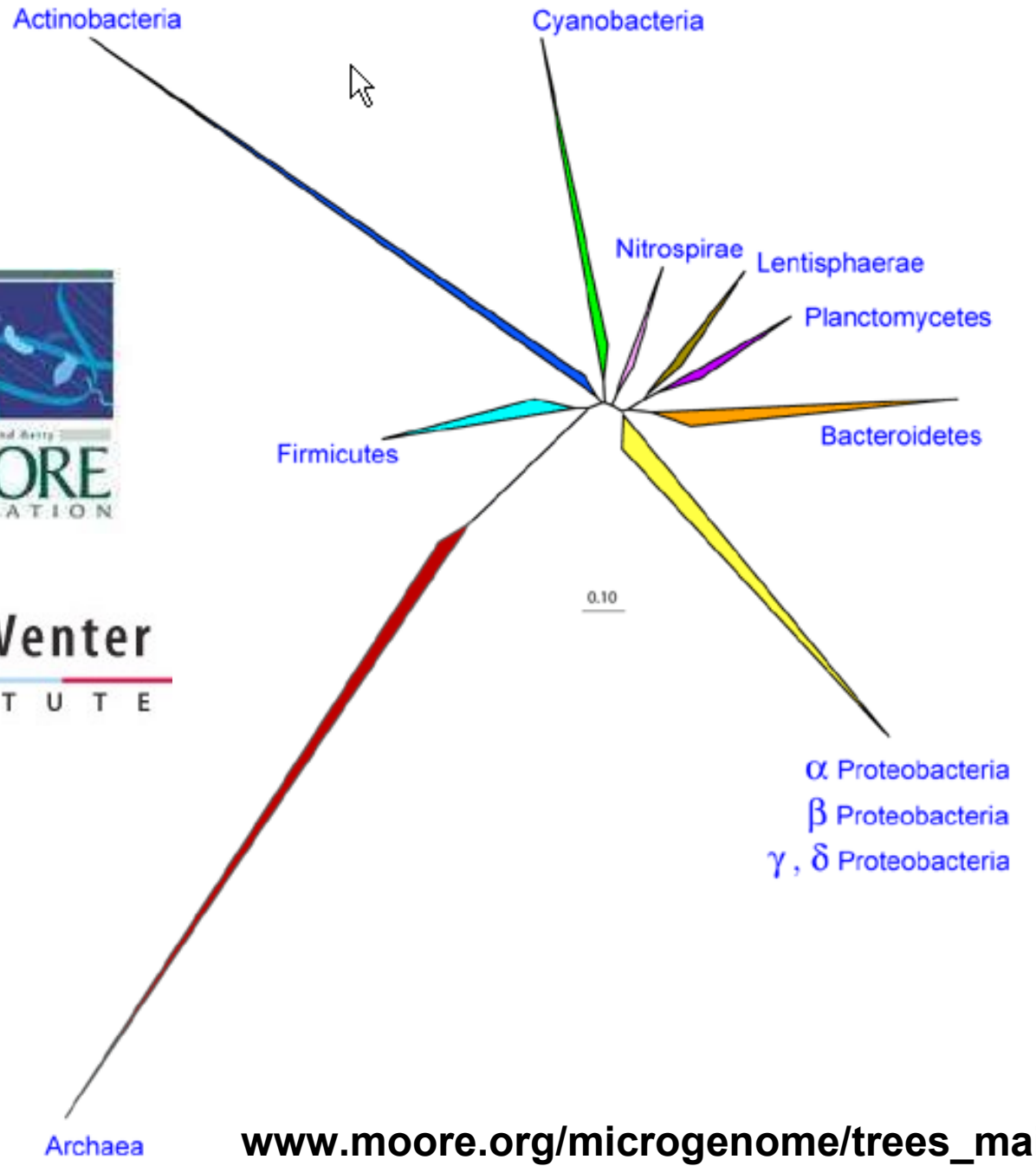
MARINE MICROBIOLOGY INITIATIVE

Gordon and Betty MOORE FOUNDATION

J. Craig Venter INSTITUTE

it2

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



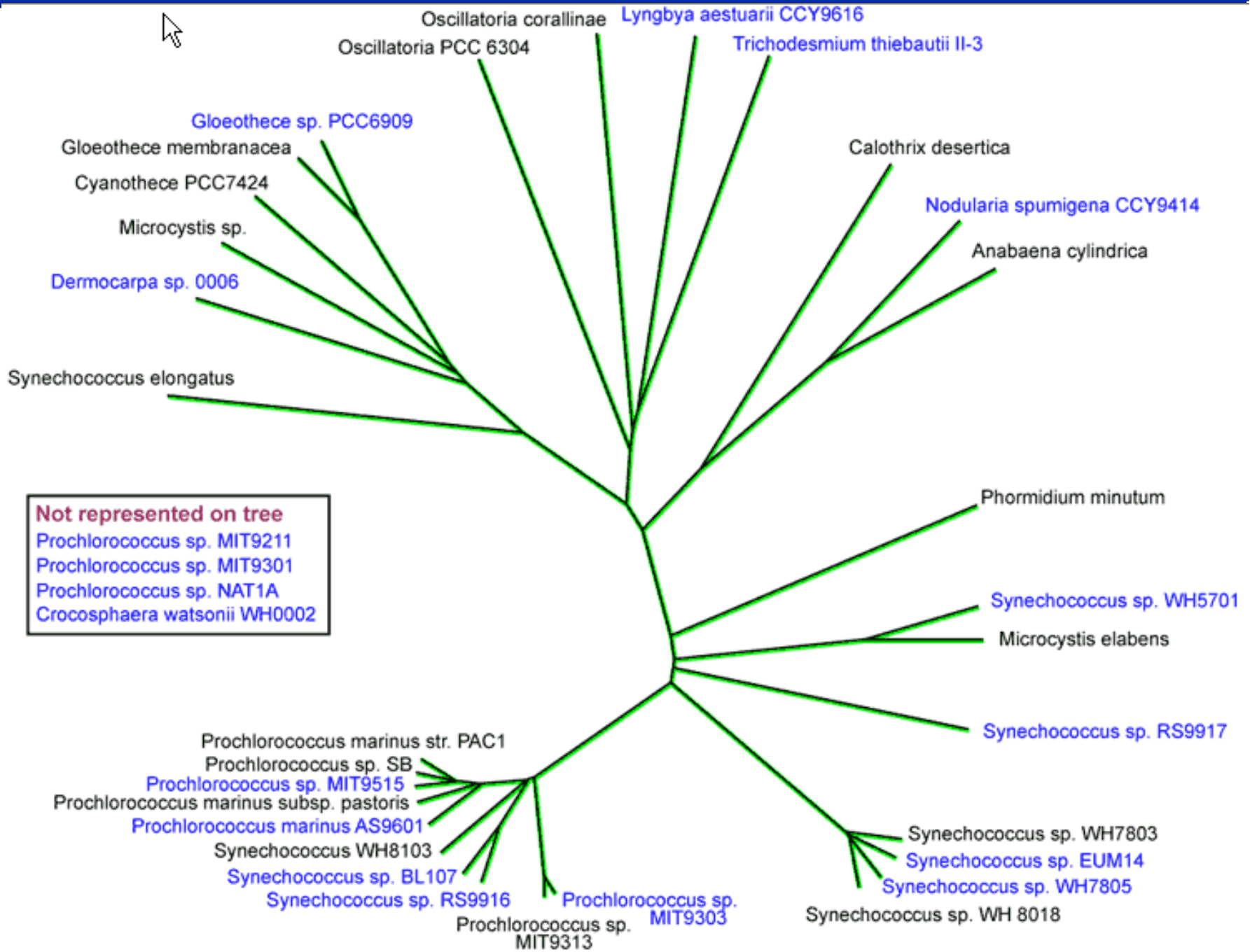
MARINE MICROBIOLOGY INITIATIVE

Gordon and Betty MOORE FOUNDATION

J. Craig Venter INSTITUTE

www.moore.org/microgenome/trees_main.asp

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



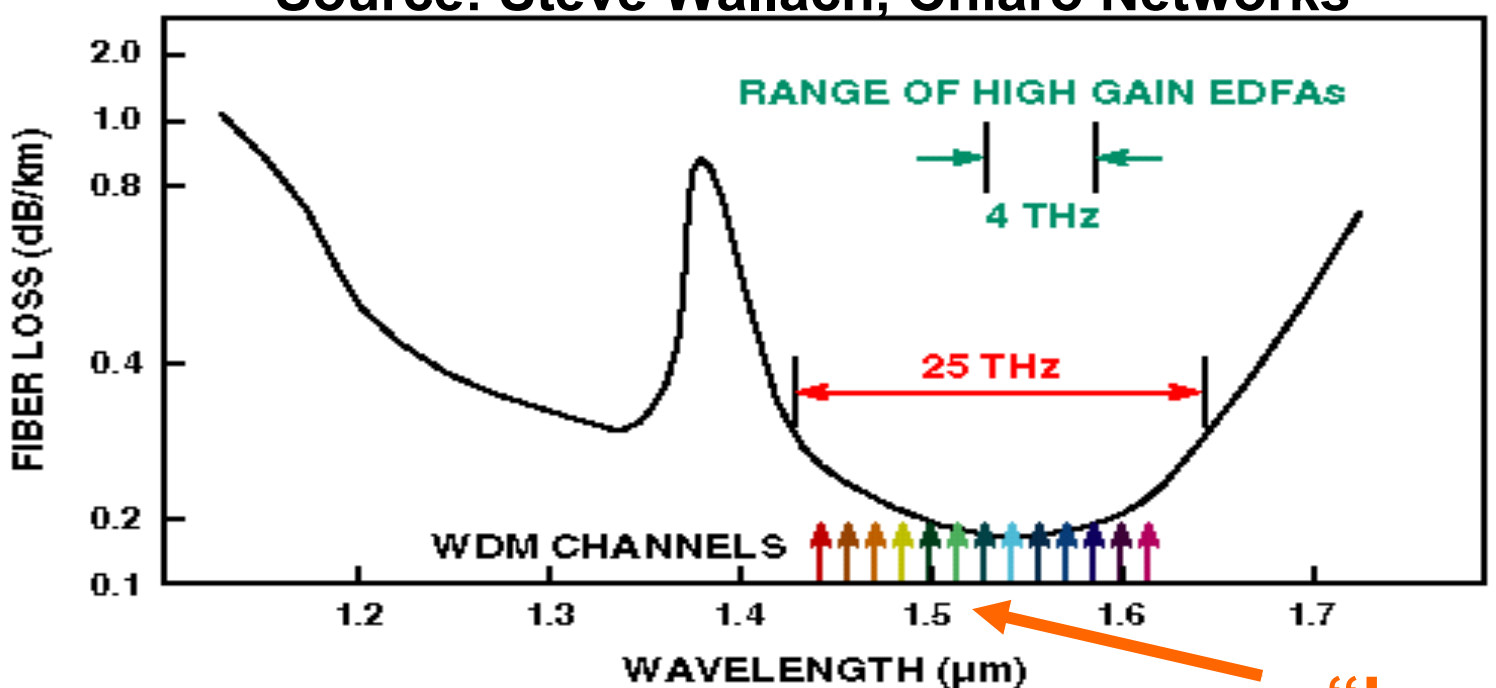
OptIPuter

Dedicated Optical Channels Makes High Performance Cyberinfrastructure Possible

WAVELENGTH DIVISION MULTIPLEXING (WDM)

- EXPLOITS
  - ENORMOUS BANDWIDTH OF SILICA FIBER
  - HIGH-GAIN WIDEBAND OPTICAL AMPLIFIERS

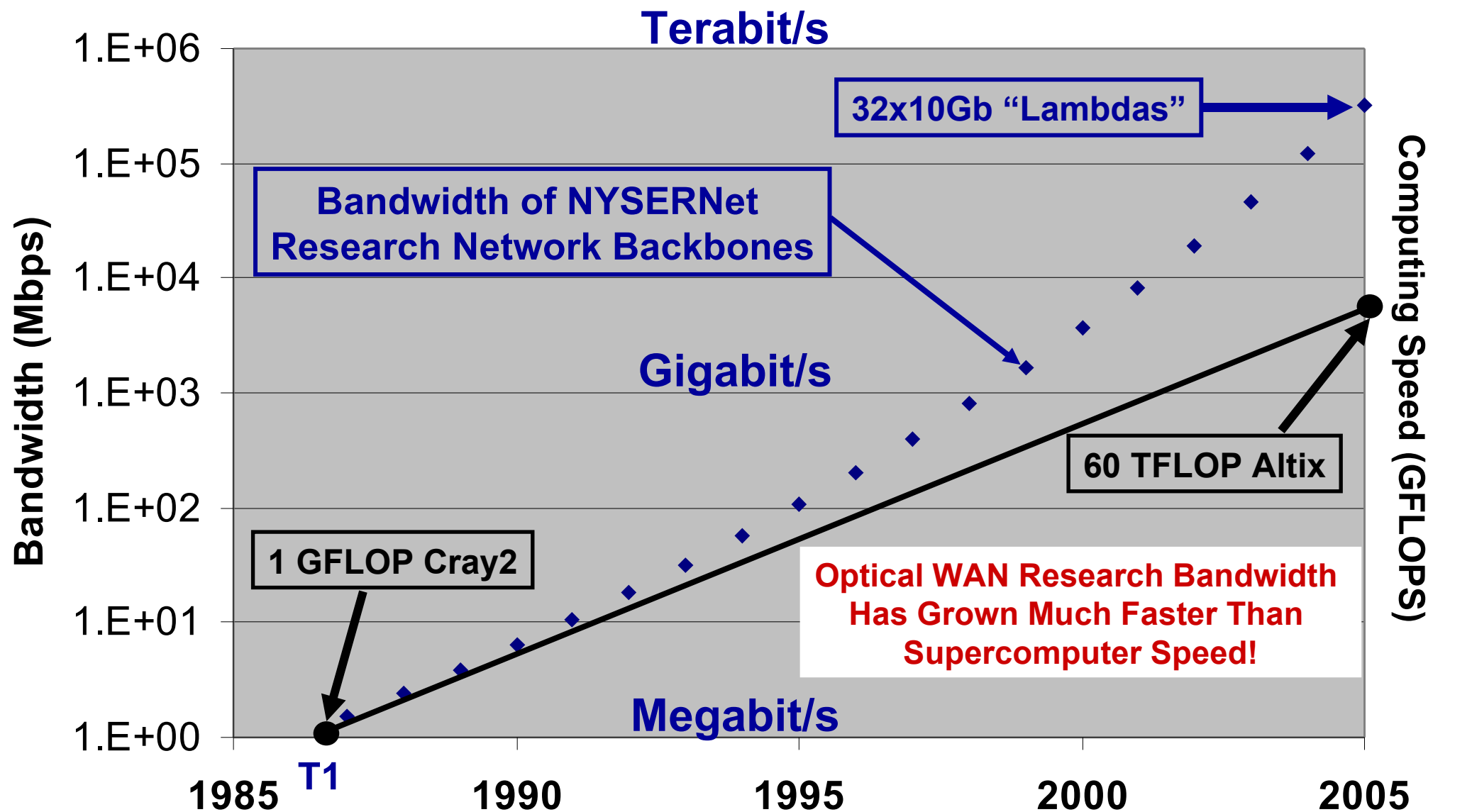
Source: Steve Wallach, Chiaro Networks



"Lambdas"

Parallel Lambdas are Driving Optical Networking The Way Parallel Processors Drove 1990s Computing

From "Supercomputer–Centric" to "Supernetwork-Centric" Cyberinfrastructure



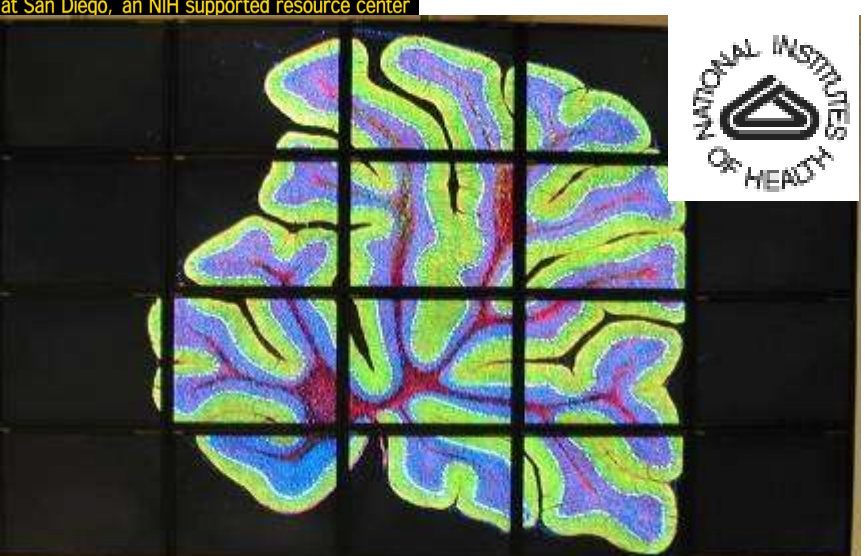
Network Data Source: Timothy Lance, President, NYSERNet

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

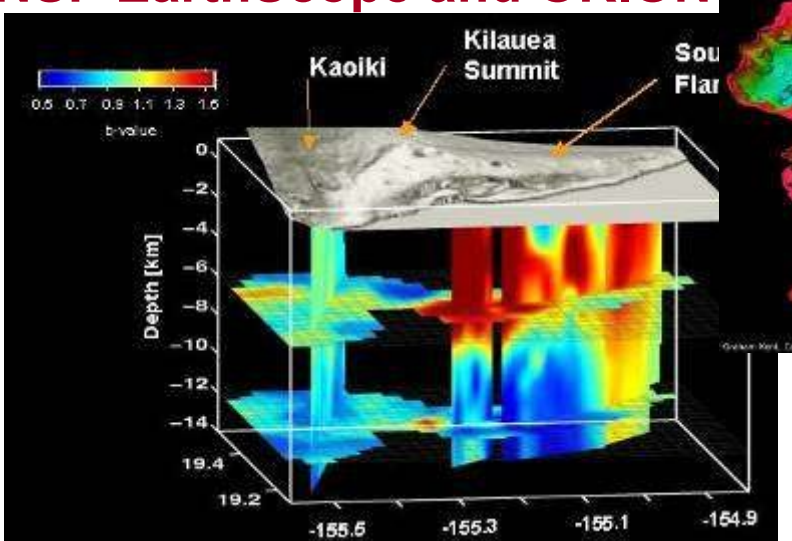
- **NSF Large Information Technology Research Proposal**
  - Calit2 (UCSD, UCI) and UIC Lead Campuses—Larry Smarr PI
  - Partnering Campuses: SDSC, USC, SDSU, NCSA, NW, TA&M, UvA, SARA, NASA Goddard, KISTI, AIST, CRC(Canada), CICESE (Mexico)
- **Industrial Partners**
  - IBM, Sun, Telcordia, Chiaro, Calient, Glimmerglass, Lucent
- **$13.5 Million Over Five Years—Now In the Fourth Year**

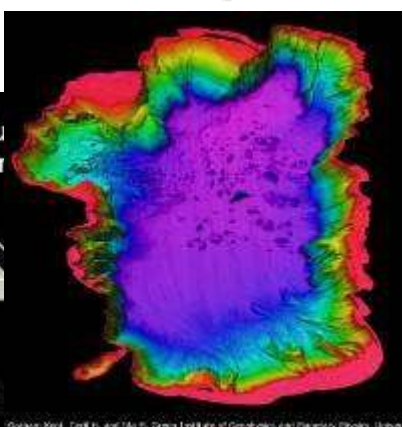
NIH Biomedical Informatics Research Network

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



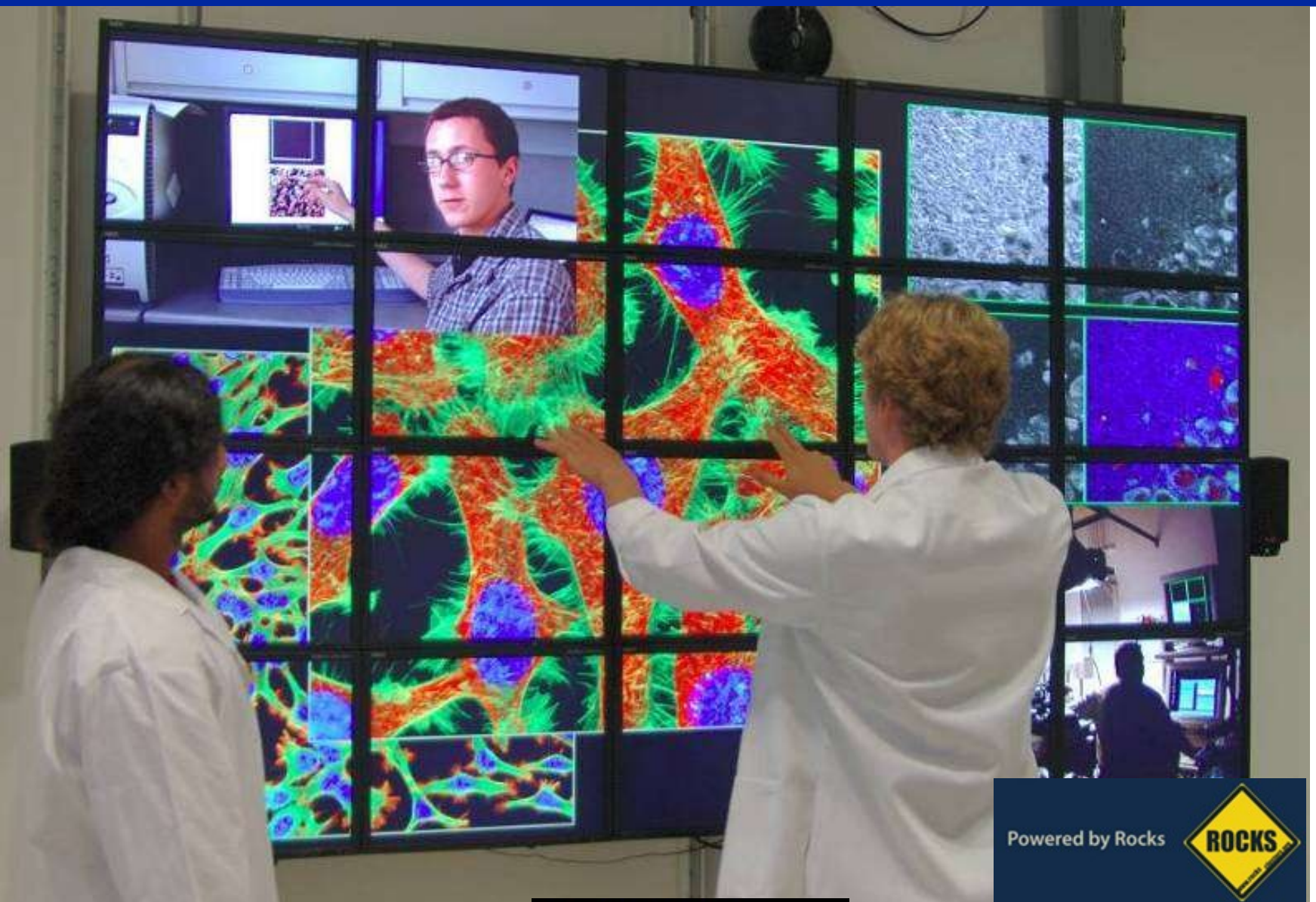
NSF EarthScope and ORION





SCRIPPS INSTITUTION OF OCEANOGRAPHY

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



OptIPortal– Termination Device for the OptIPuter Global Backplane

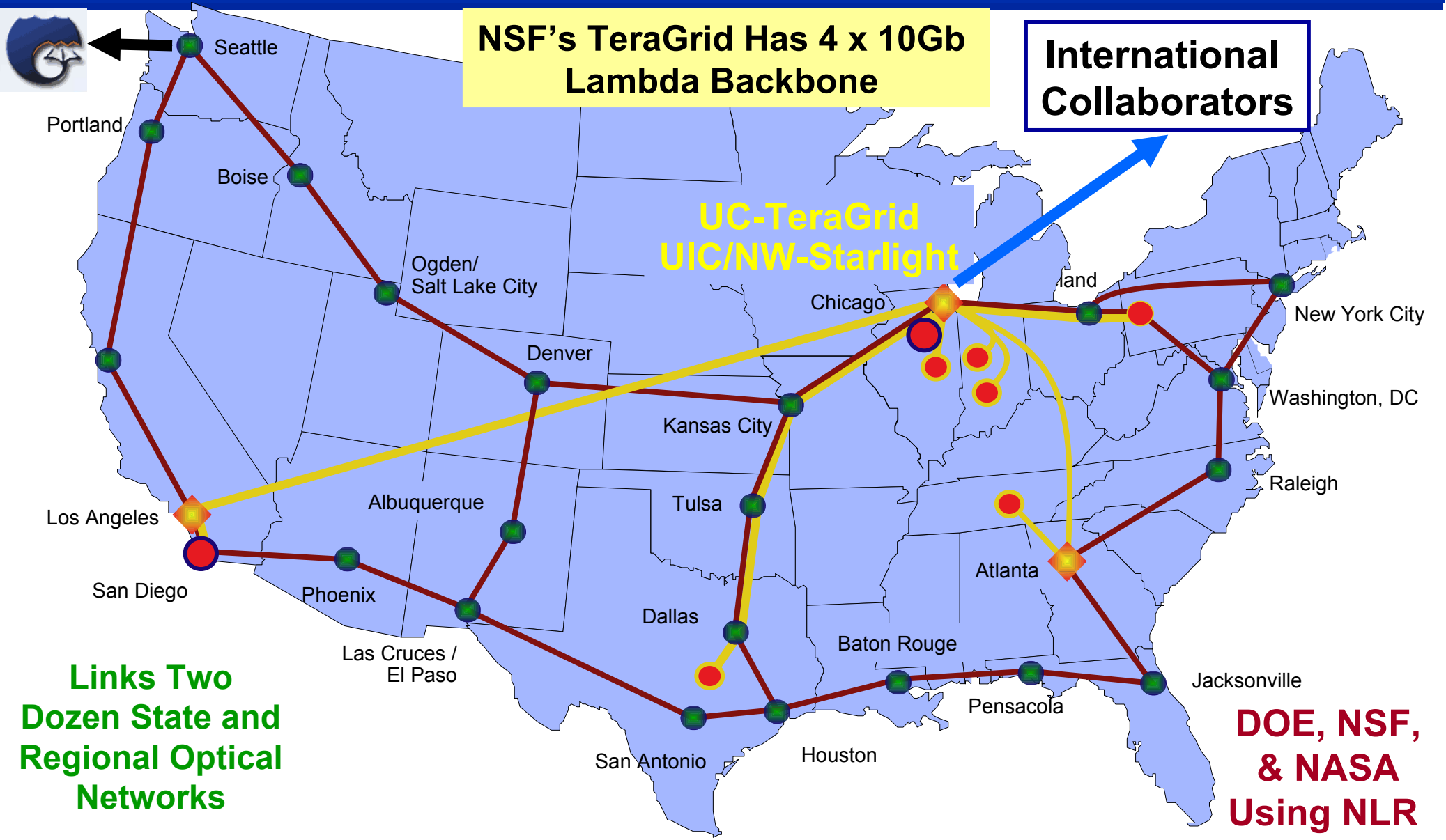
evl electronic visualization laboratory

OptIPuter

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

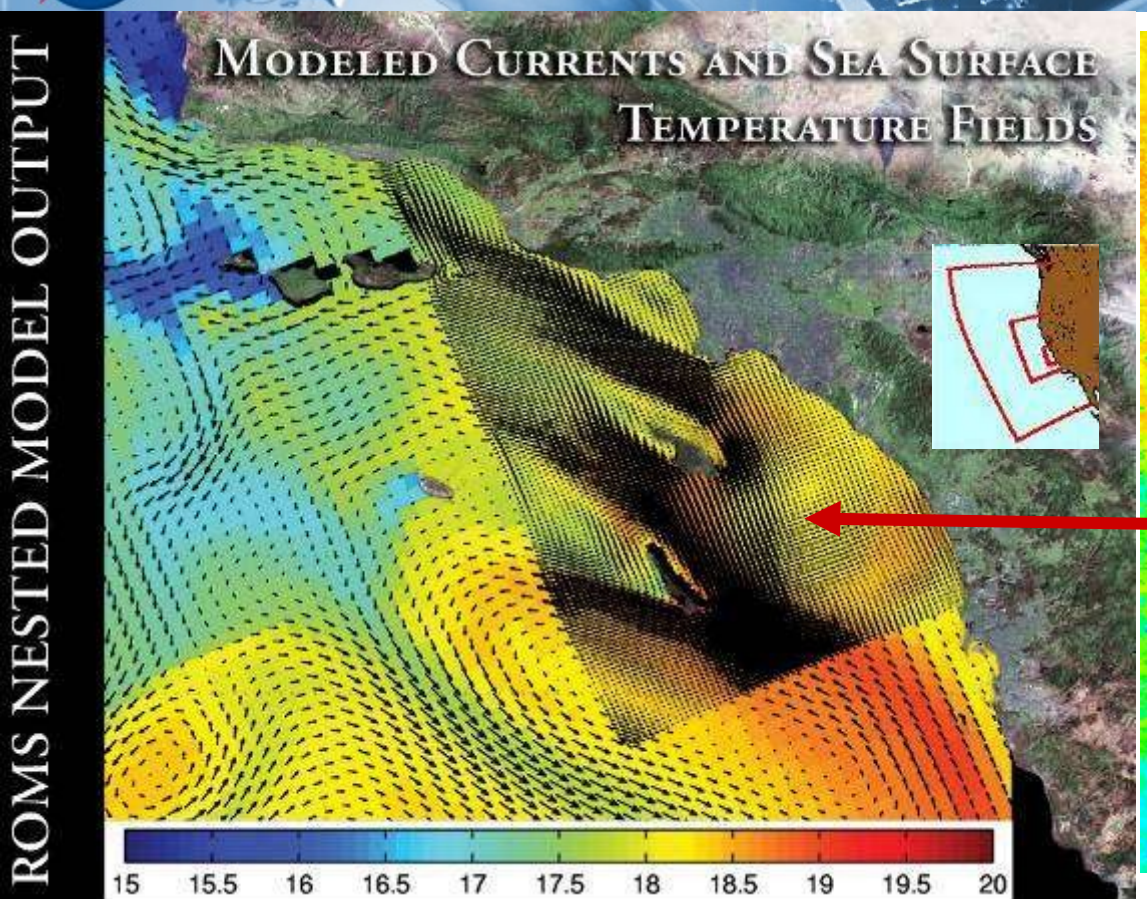
it2

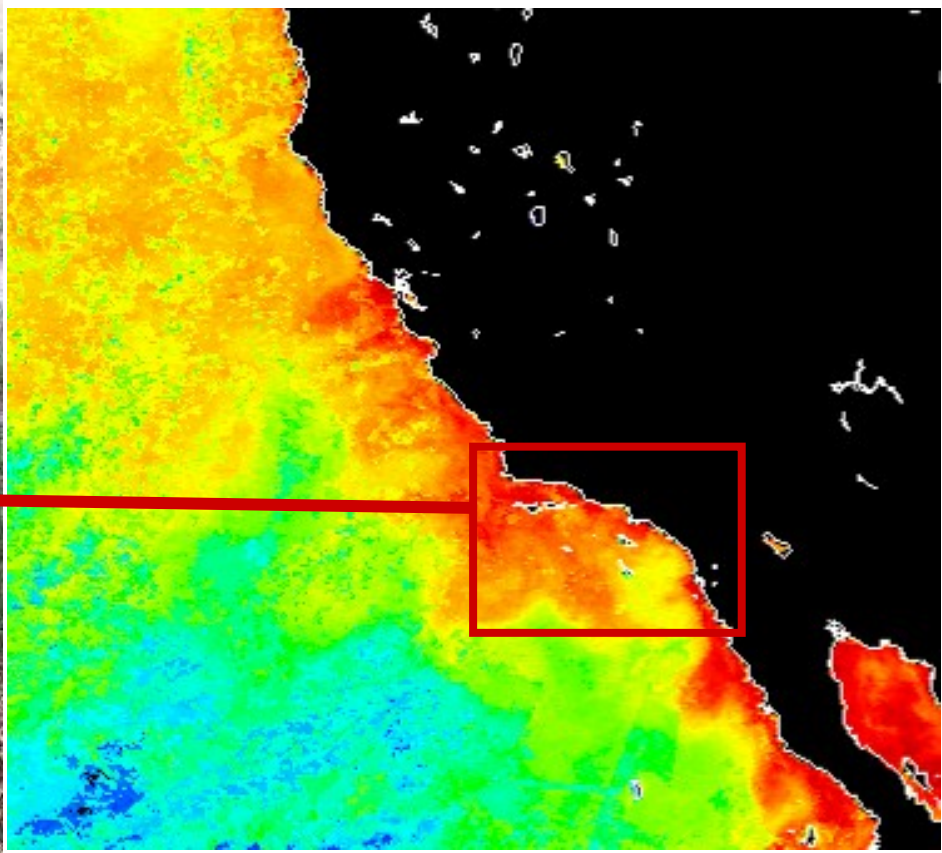
National Lambda Rail (NLR) and TeraGrid Provides Cyberinfrastructure Backbone for U.S. Researchers



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

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

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

J. Craig Venter INSTITUTE

UCSD

OptIPuter

Announced January 17, 2006
$24.5M Over Seven Years

Paul Gilna Has Just Been Recruited from Los Alamos to Become Executive Director of CAMERA

- **Formerly**
  - Former Director of the Department of Energy's Joint Genome Institute (JGI) Operations at Los Alamos National Laboratory (LANL)
  - Group Leader of Genomic Science and Computational Biology in LANL's Bioscience Division
- **JGI**
  - A $70-million-per-Year collaboration that teams the expertise:
    - Lawrence Berkeley,
    - Lawrence Livermore,
    - Los Alamos,
    - Oak Ridge, and
    - Pacific Northwest
    - and the Stanford Human Genome Center
  - Working at The Frontiers of Genome Sequencing and Biosciences

camera
Marine Microbial Ecology

OptIPuter

Embargoed till Press Announcement This Week!

it2

Building On Centers of Excellence and Achievement

NBCR

OptIPuter

NCMIR

JCSG

Powered by Rocks

ROCKS

JGI DOE JOINT GENOME INSTITUTE

CRBS

SDSC

The Scripps Genome Center

Supported by:

National Institutes of Health

National Center for Research Resources

NSF

Office of Science U.S. Department of Energy

OptIPuter

camera Marine Microbial Ecology

Announced January 17, 2006

it2

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

Sargasso Sea Data

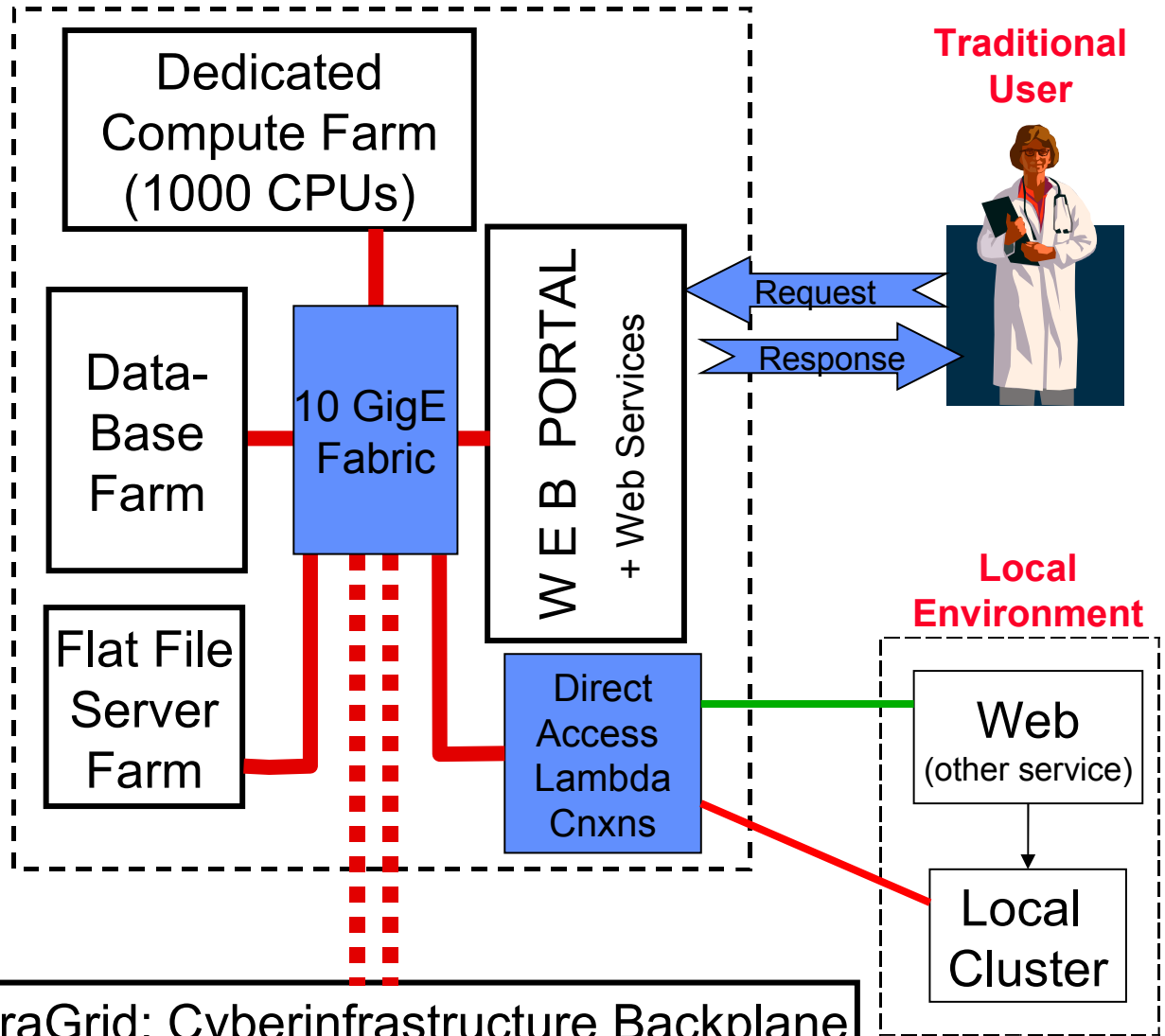
Sorcerer II Expedition (GOS)

JGI Community Sequencing Project

Moore Marine Microbial Project

NASA Goddard Satellite Data

Community Microbial Metagenomics Data



Source: Phil Papadopoulos, SDSC, Calit2

The Future Home of the Moore Foundation Funded Marine Microbial Ecology Metagenomics Complex

camera
Marine Microbial Ecology



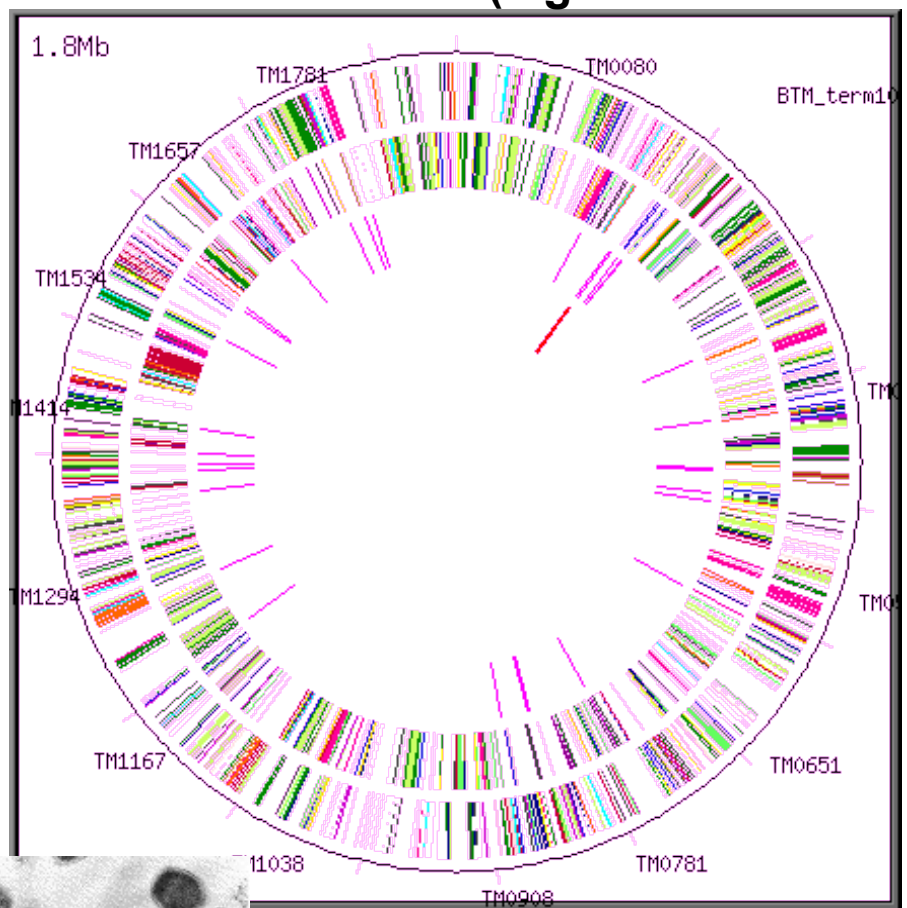
OptIPuter

Photo Courtesy Joe Keefe, Calit2

it2

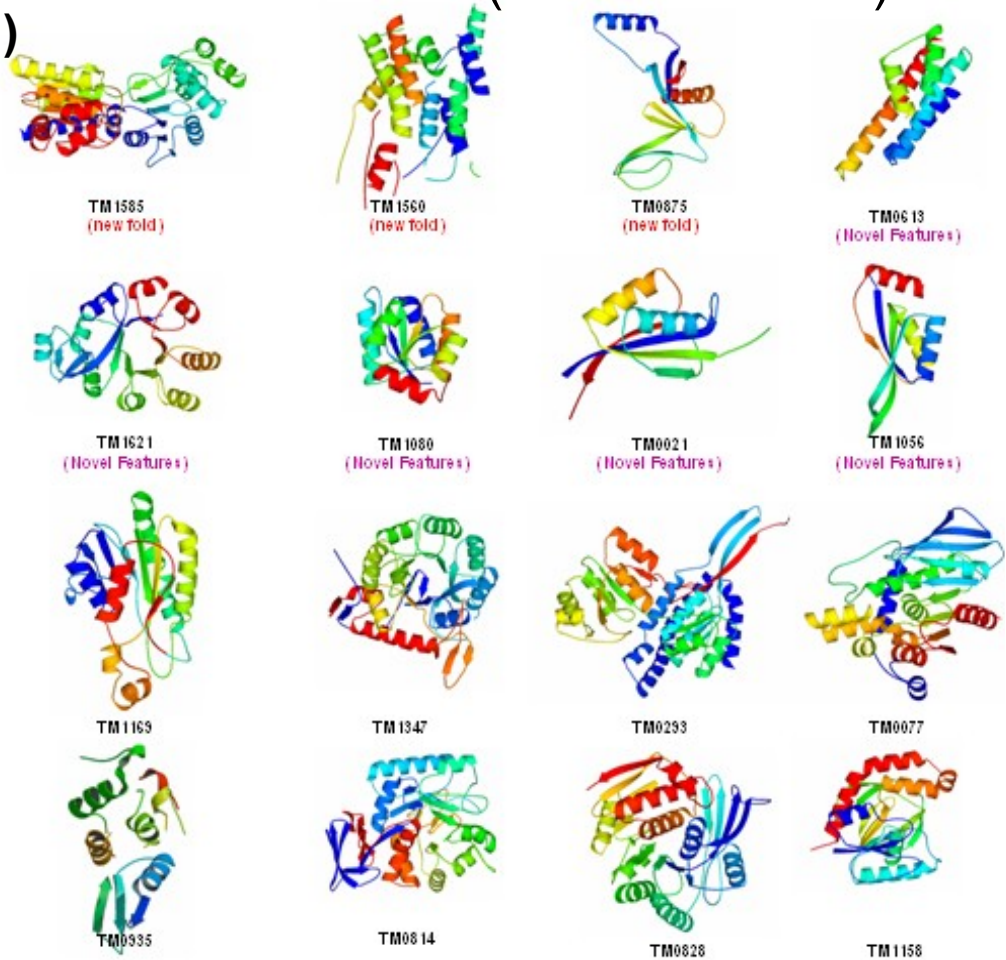
The Bioinformatics Core of the Joint Center for Structural Genomics will be Housed in the Calit2@UCSD Building

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





173 Structures (122 from JCSG)

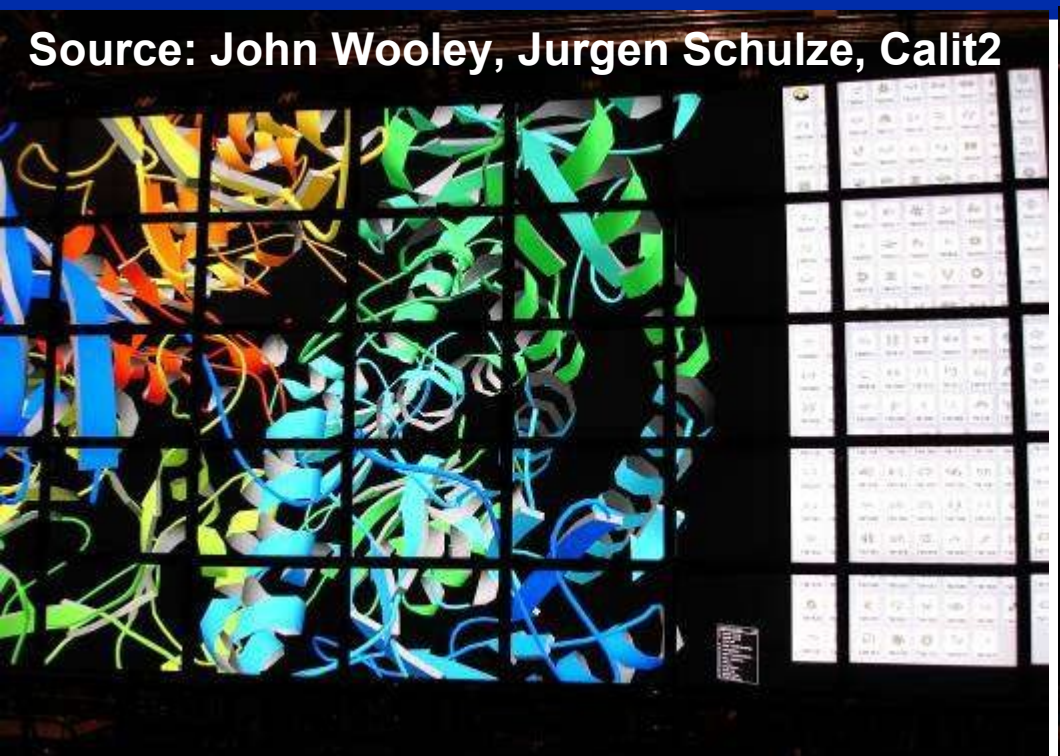


- Determining the Protein Structures of the Thermotoga Maritima Genome
- 122 T.M. Structures Solved by JCSG (75 Unique In The PDB)
- Direct Structural Coverage of 25% of the Expressed Soluble Proteins
- Probably Represents the Highest Structural Coverage of Any Organism

Source: John Wooley, UCSD

Interactive Visualization of Thermatoga Proteins at Calit2

Source: John Wooley, Jurgen Schulze, Calit2





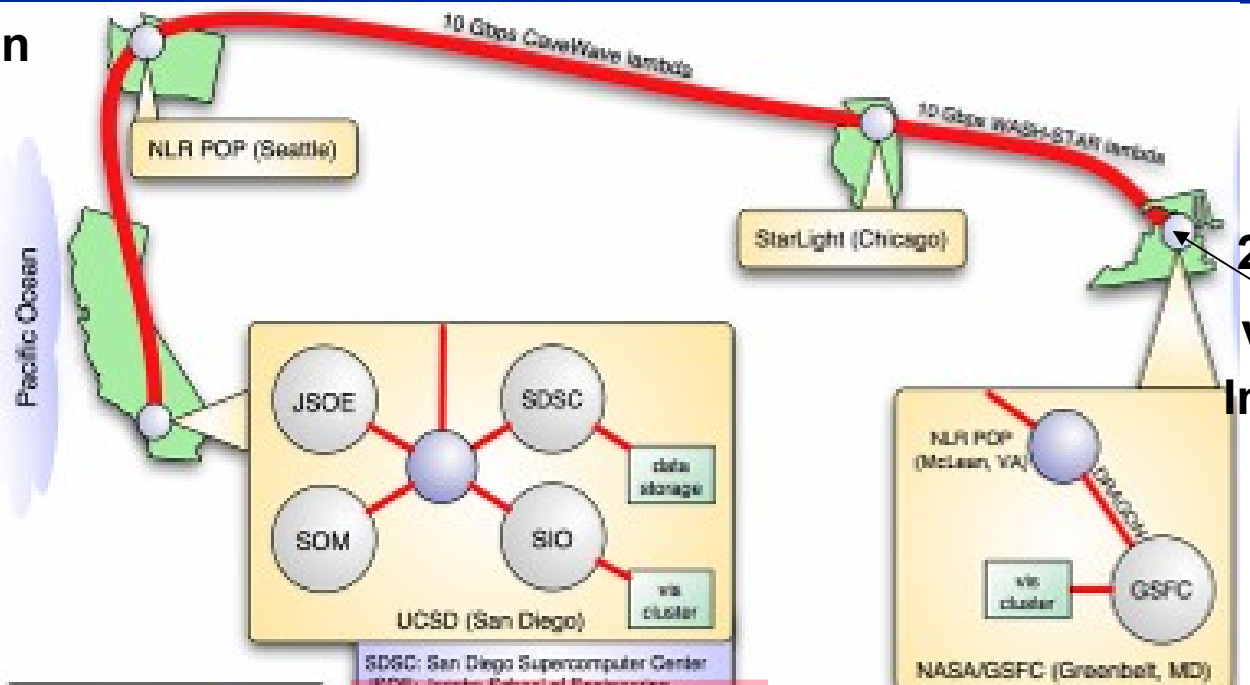




Calit2 and the Venter Institute Will Combine Telepresence with Remote Interactive Analysis

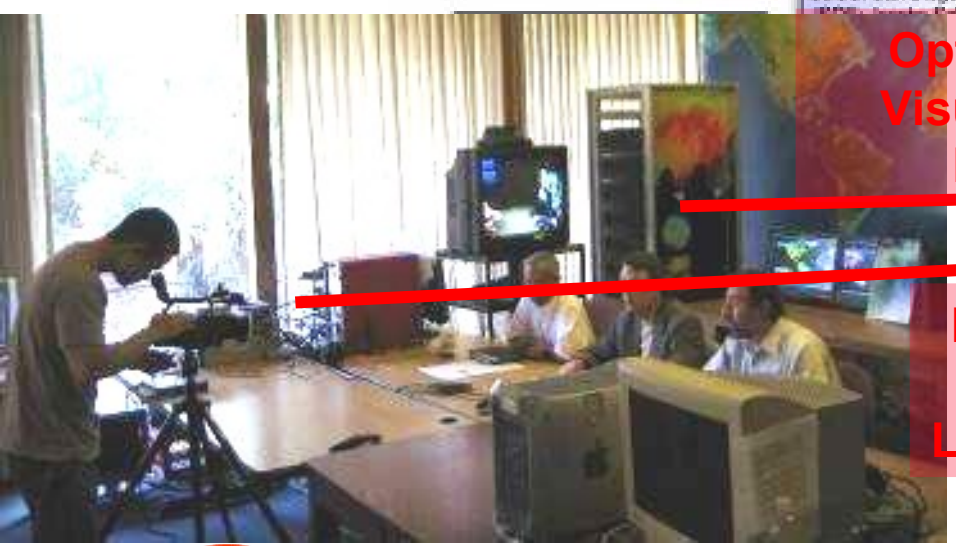
Live Demonstration

of 21st Century National-Scale Team Science



25 Miles

Venter Institute



OptIPuter Visualized Data

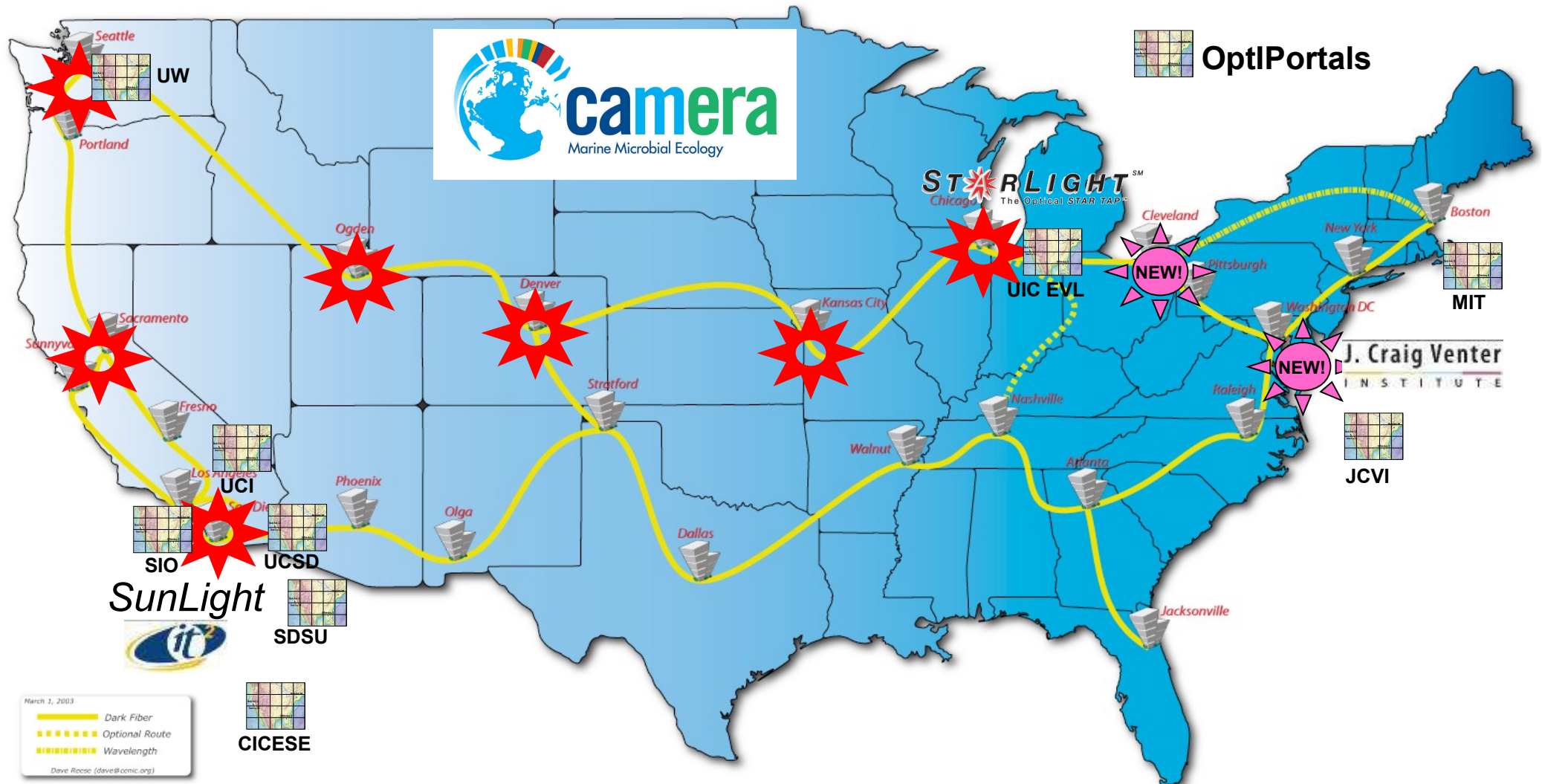
HDTV Over Lambda



OptIPuter

it2

UIC/UCSD 10GE *CAVEWave* on the National LambdaRail Emerging OptIPortal Sites



CAVEWave Connects Chicago to Seattle to San Diego…and Washington D.C. as of 4/1/06

and JCVI as of 5/15/06

OptIPuter

STARLIGHT The Optical STAR TAP

it2

CAMERA Outreach

- SAB Meetings
- Targeted Workshops,
  - User Forums,
  - User Software Testing
  - Viz Tool Brainstorming
- Presence at Scientific Meetings
  - Demonstration Booths, Tutorials, User Forums, Presentations
- Partnerships with Metagenomics Projects
  - JGI, …
- Training
- Policy Study on Convention on Biological Diversity
- User Services Team



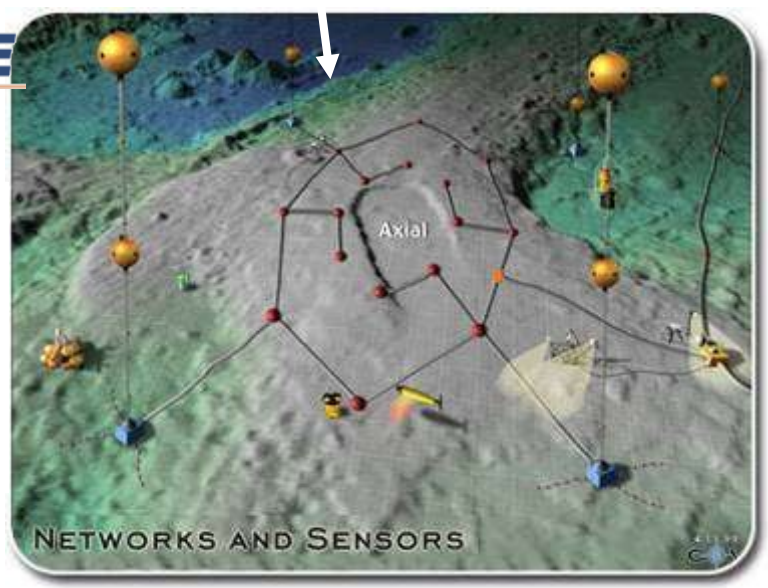
OptIPuter

it2

NSF's Ocean Observatories Initiative (OOI) Envisions Global, Regional, and Coastal Scales



NEPTUNE



LEO15 Inset Courtesy of Rutgers University, Institute of Marine and Coastal Sciences

NSF

ocean research interactive ORION observatory networks

OptIPuter

it2

New OptIPuter Driver: Gigabit Fibers on the Ocean Floor -- Controlling Sensors and HDTV Cameras Remotely



LOOKING is Driven By NEPTUNE CI Requirements

- **National Science Foundation Is Planning a New Generation of Ocean Observatories**
  - Ocean Research Interactive Observatory Networks (ORION)
- **Fibered Observatories Linked to Land Fiber Infrastructure**
- **Laboratory for the Ocean Observatory Knowledge Integration Grid (LOOKING)**
  - Building a Prototype Based on OptIPuter Technologies Plus Web/ Grid Services
  - HDTV Streams Over IP Will be a Major Driver

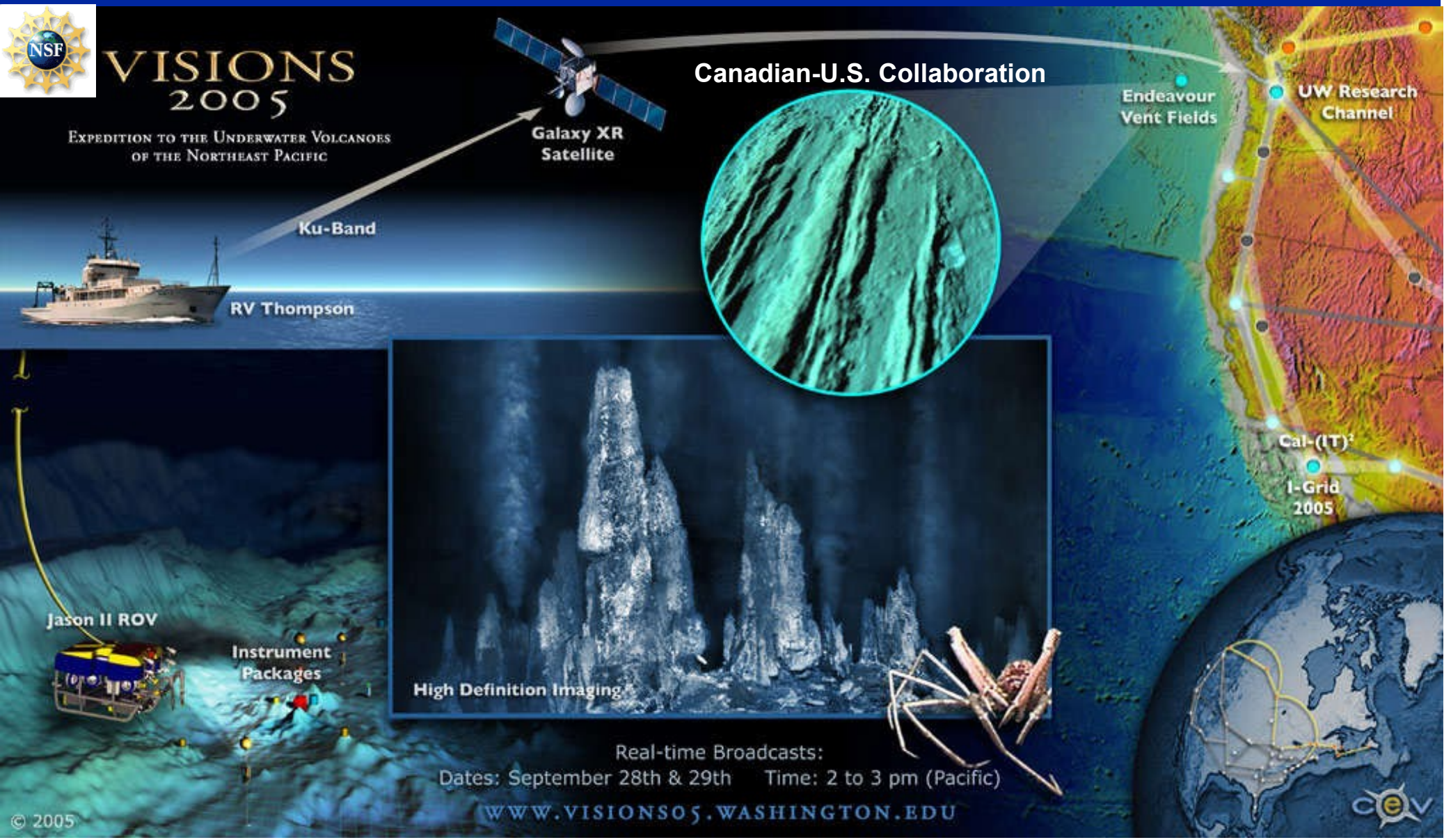
OptIPuter

(Funded by NSF ITR- John Delaney, UWash, PI)

Making Management of Gigabit Flows Routine

it2

First Remote Interactive High Definition Video Exploration of Deep Sea Vents

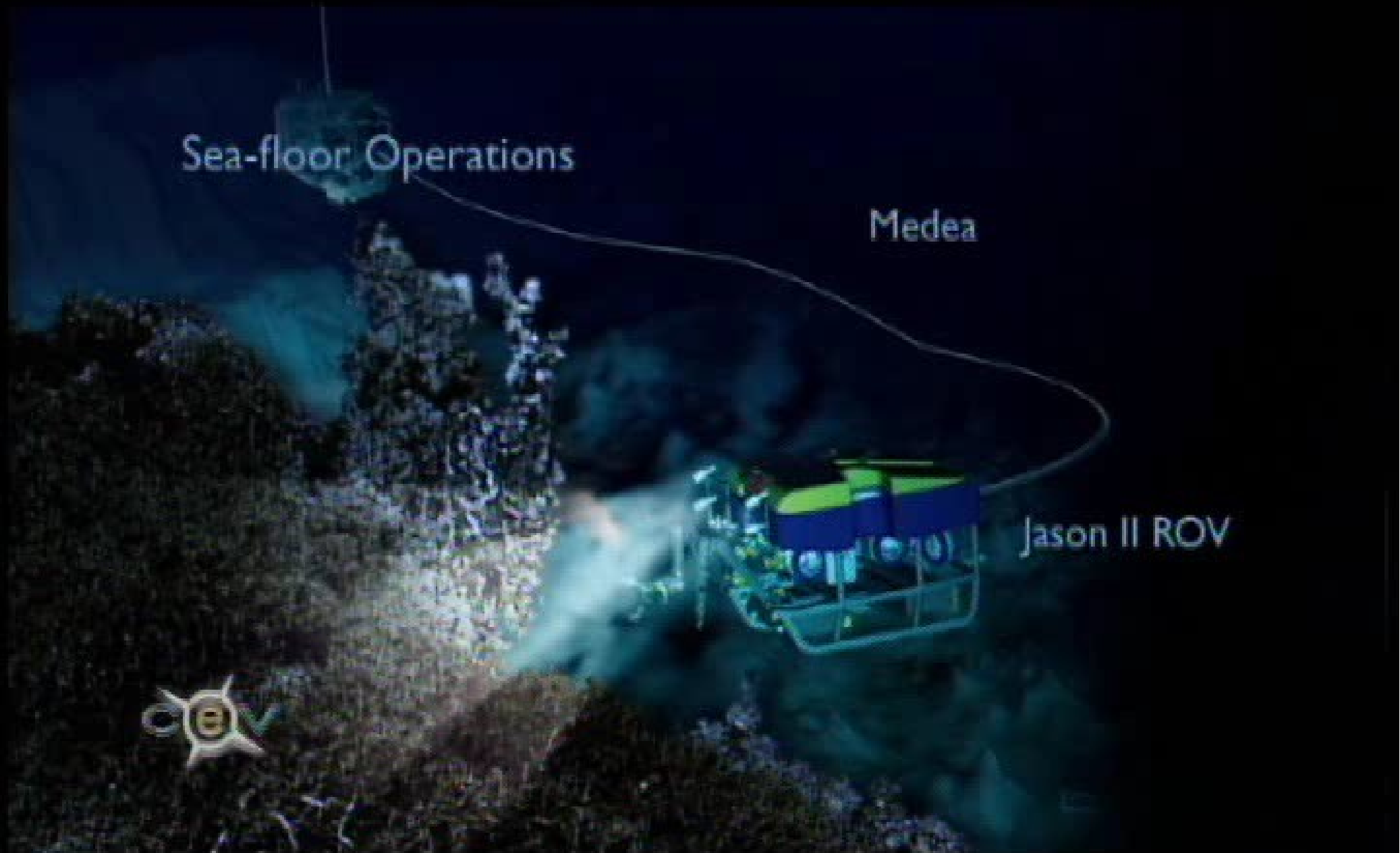


OptIPuter

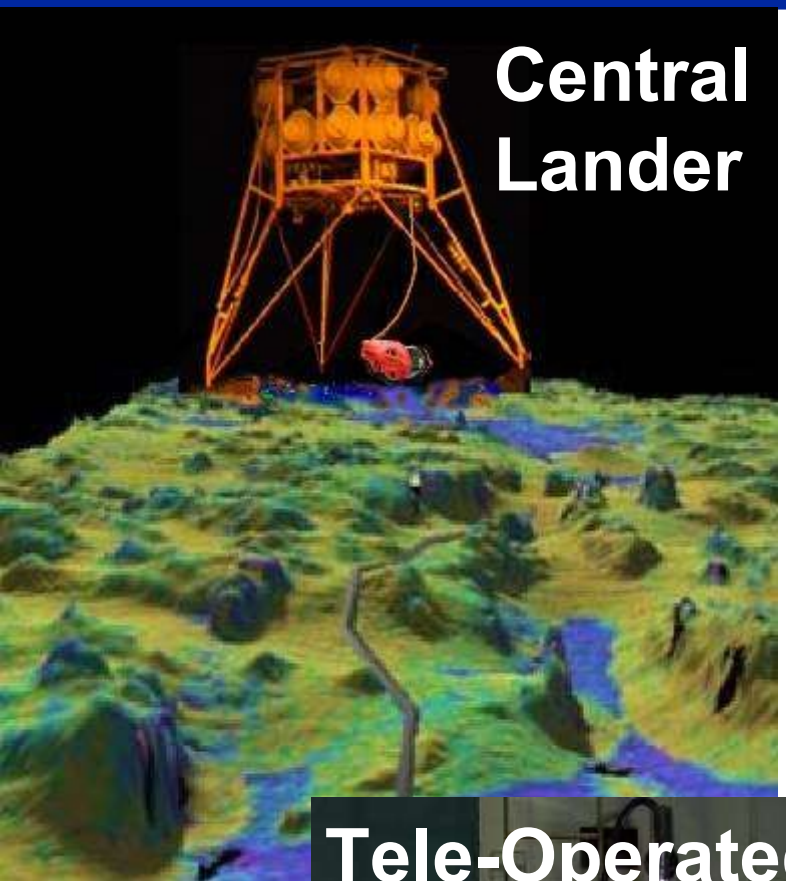
Source John Delaney & Deborah Kelley, UWash

it²

High Definition Video - 2.5 km Below the Ocean Surface



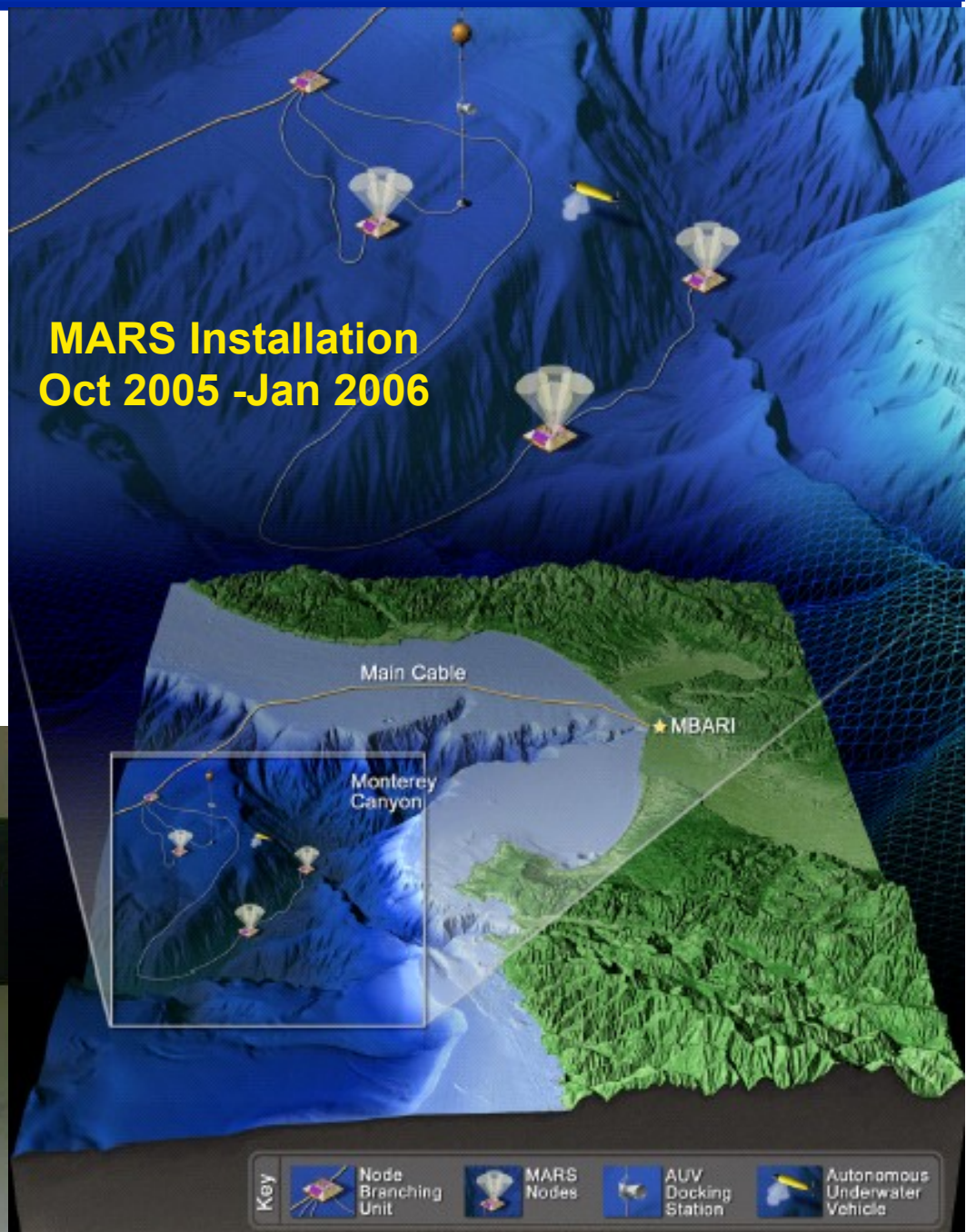
MARS Cable Observatory Testbed – LOOKING "Living Laboratory"





Source:
Jim
Bellingham,
MBARI

OptIPuter

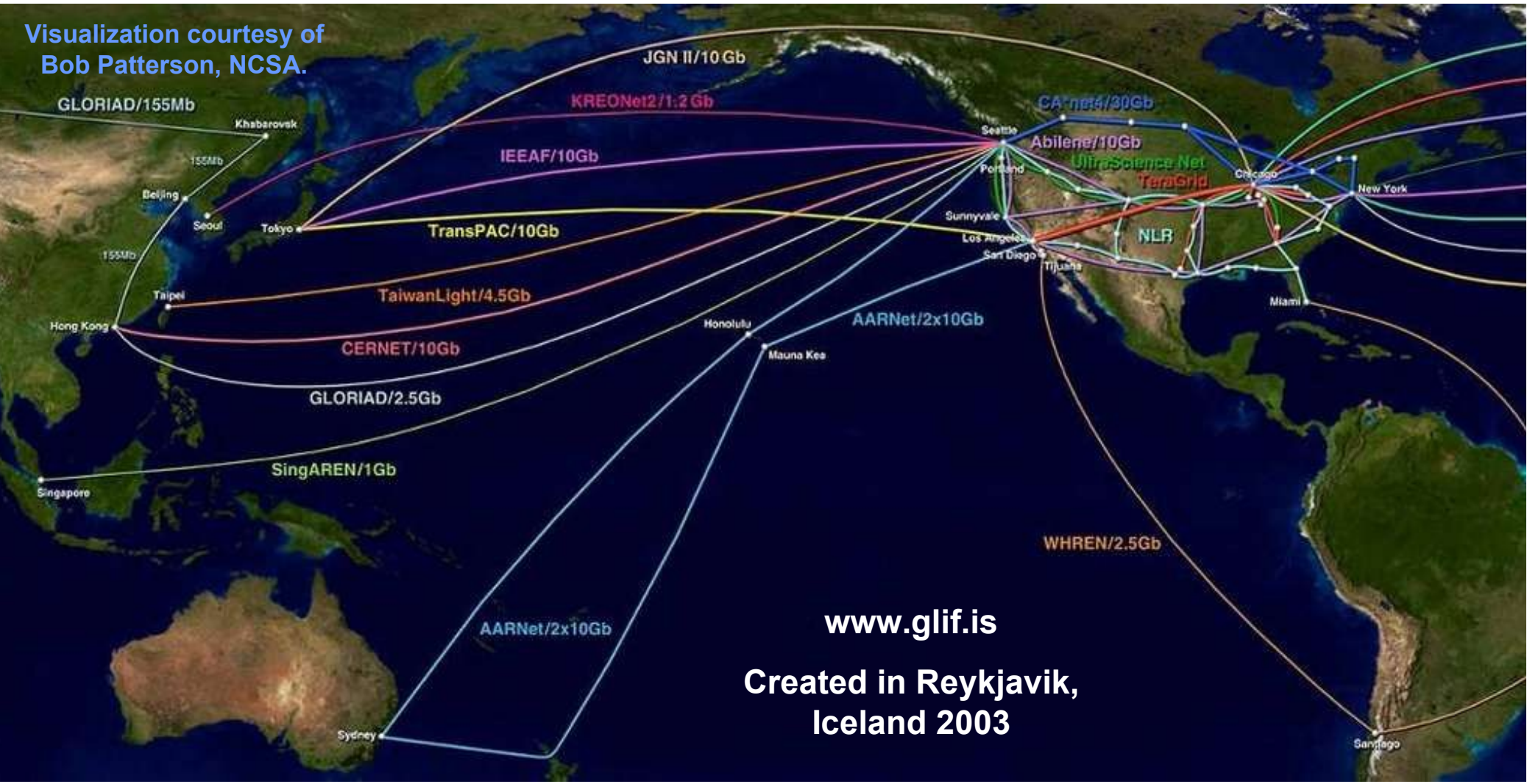


A Near Future Metagenomics Fiber Optic-Enabled Data Generator



Source John Delaney, UWash

Countries are Aggressively Creating Gigabit Services: Interactive Access to CAMERA and LOOKING Systems



OptIPuter

it2