The Emerging Global Collaboratory for Microbial Metagenomics Researchers

Invited Talk
Delivered From Calit2@UCSD
Monash University MURPA Lecture
Melbourne, Australia
July 30, 2008

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

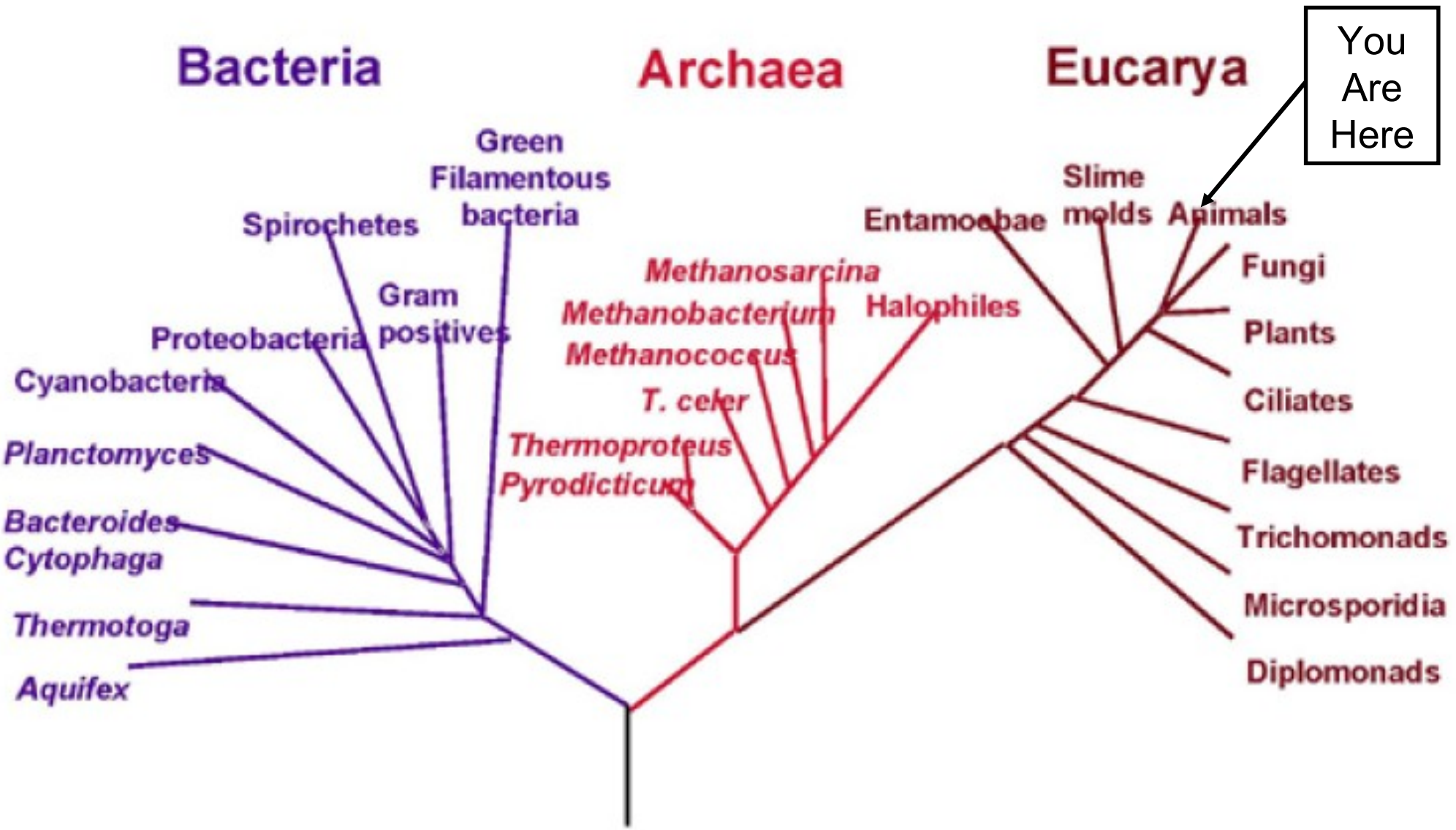
UCSD
UCIrvine

Abstract

Calit2, in collaboration with the J. Craig Venter Institute, 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. Currently over 2000 users from over 50 countries are CAMERA registered users.

UCSD
UCIrvine

Most of Evolutionary Time Was in the Microbial World



Tree of Life Derived from 16S rRNA Sequences

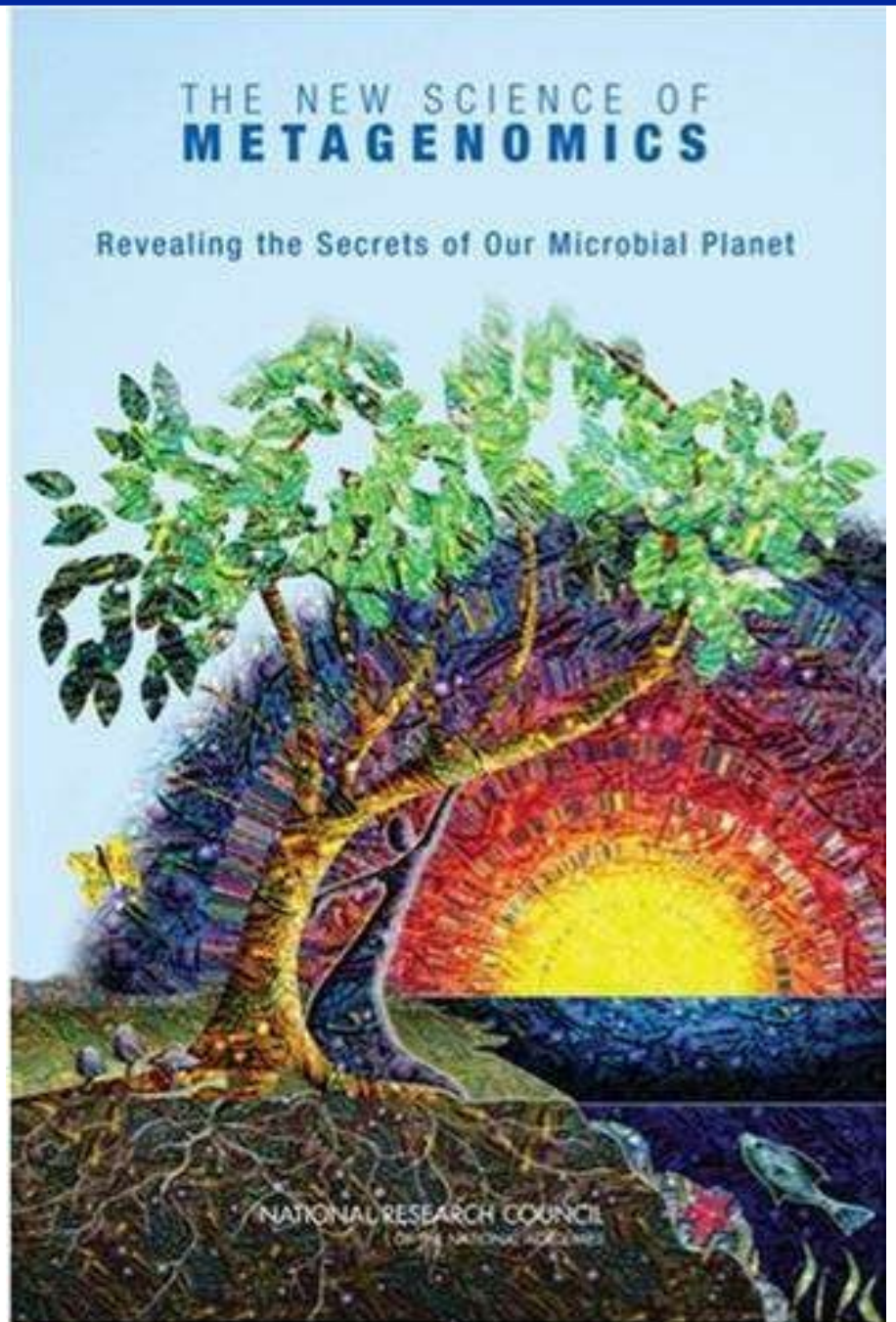
Source: Carl Woese, et al

UCSD
UCIrvine

The New Science of Metagenomics

NRC Report:

Metagenomic data should be made publicly available in international archives as rapidly as possible.

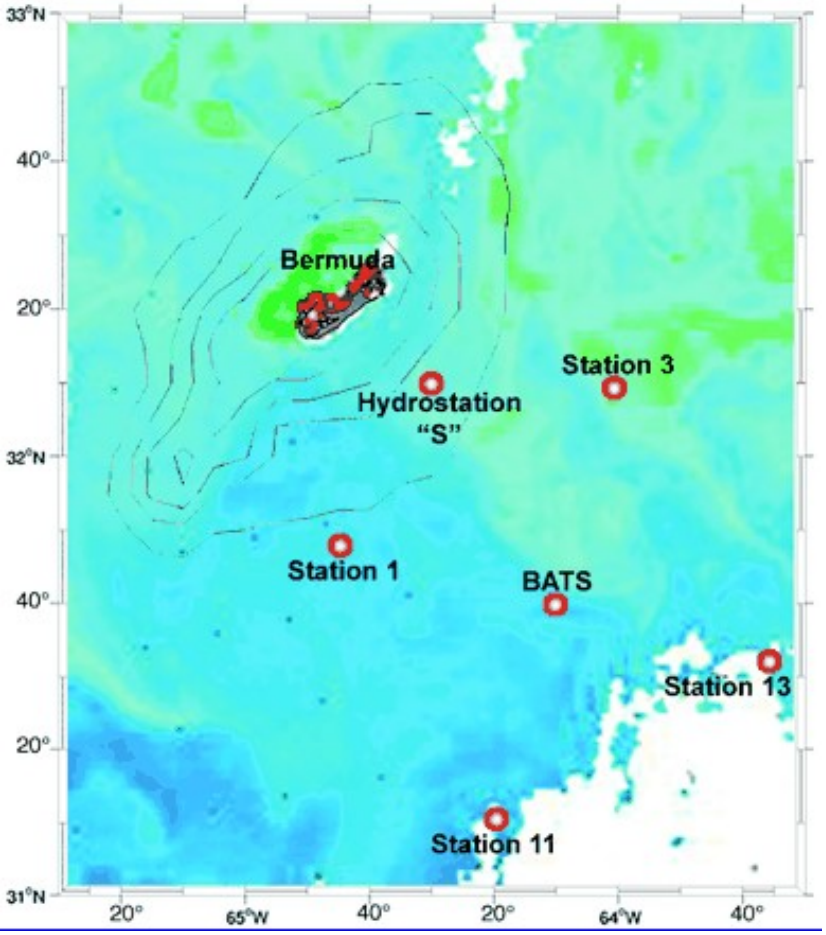


"The emerging field of metagenomics, where the DNA of entire communities of microbes is studied simultaneously, presents the greatest opportunity -- perhaps since the invention of the microscope – to revolutionize understanding of the microbial world." –

National Research Council
March 27, 2007

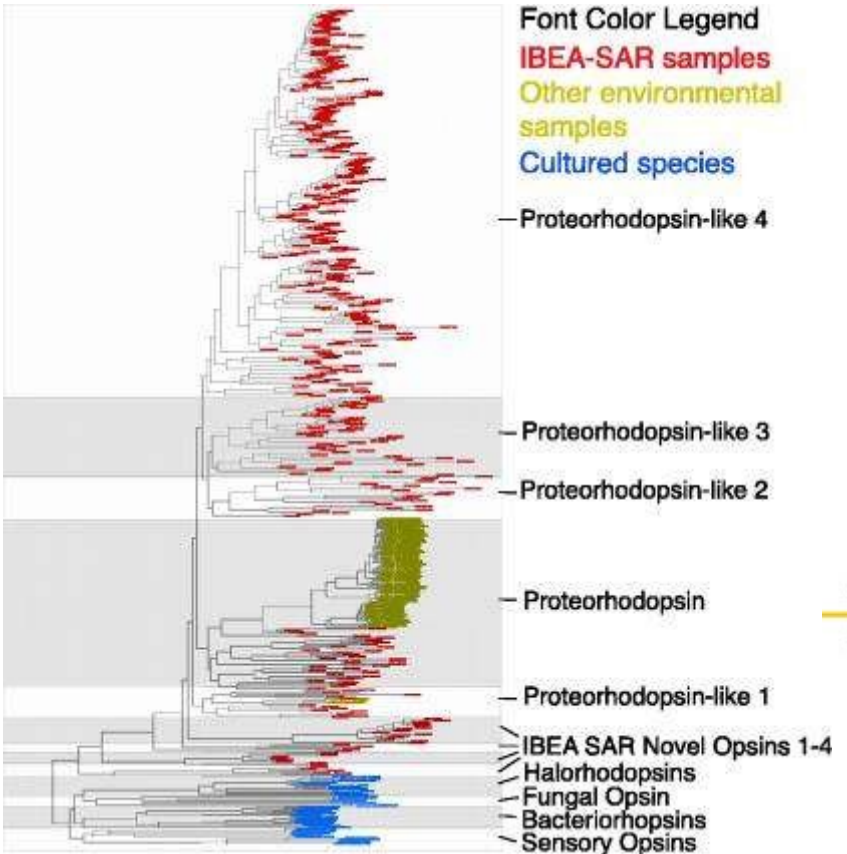
UCSD
UCIrvine

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

UCSD
UCIrvine

J. Craig Venter INSTITUTE

it²

Marine Genome Sequencing Project – Measuring the Genetic Diversity of Ocean Microbes

Plus 155 Marine Microbial Genomes

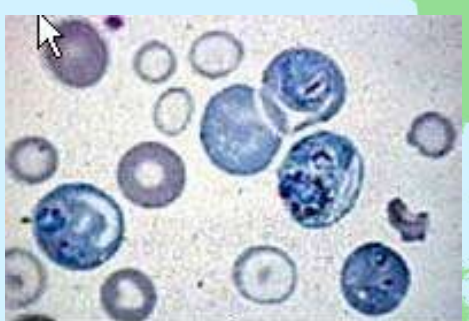


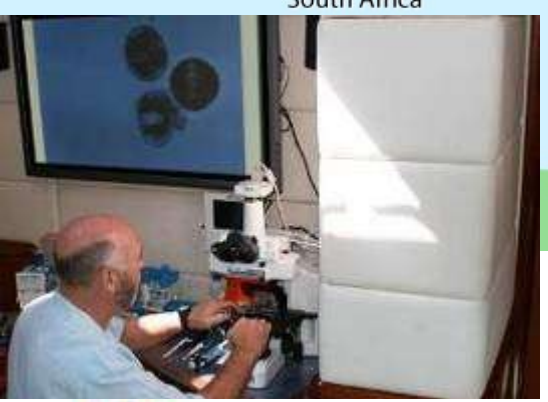
Each Sample ~2000 Microbial Species

Specify Ocean Data

Canada
United States
Mexico
Puerto Rico
Honduras
Costa Rica
Panama
Galapagos Islands
x 10 samples
French Polynesia
Cook Islands
Tonga
Fiji
Vanuatu
x 17 samples
Australia
New Caledonia
x 8 samples
Cocos-Keeling Is.
Tanzania/ Zanzibar
x 7 samples
Seychelles
Madagascar
South Africa

Analyzed
Sampled, but not Analyzed





MARINE MICROBIOLOGY INITIATIVE

camera
Marine Microbial Ecology

Gordon and Betty MOORE FOUNDATION

Sorcerer II Data Will Double Number of Proteins in GenBank!

J. Craig Venter INSTITUTE

it²

Enormous Increase in Scale of Known Genes Over Last Decade

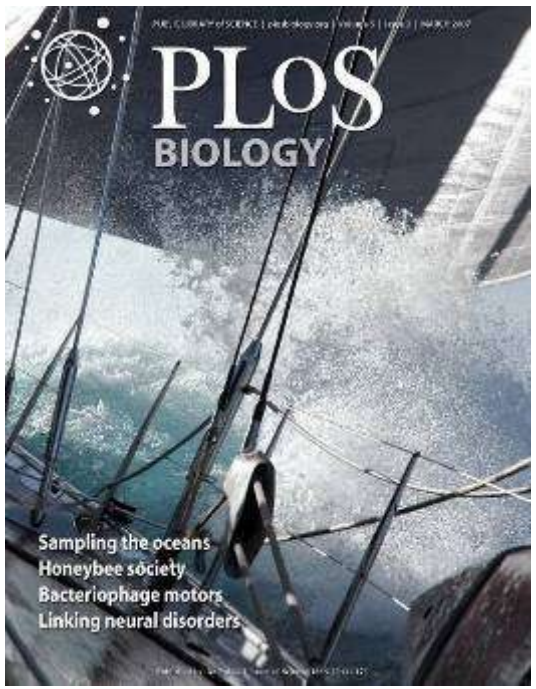
1995
First Microbe Genome



1.8 Million Bases
1749 Genes

J. Craig Venter
INSTITUTE

2007
Ocean Microbial Metagenomics



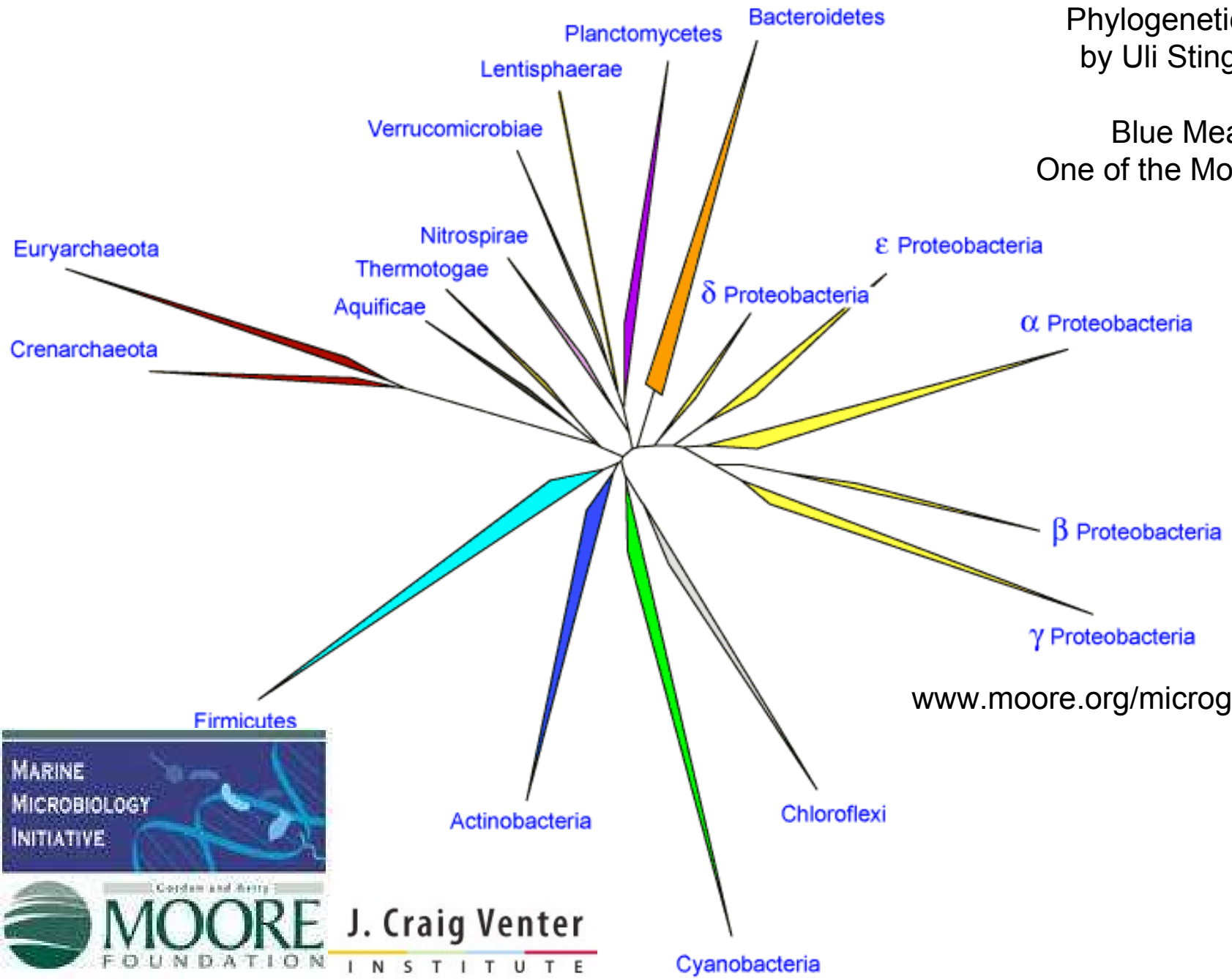
6.3 Billion Bases
5.6 Million Genes

~3300x

UCSD
UCIrvine

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

MOORE FOUNDATION

J. Craig Venter INSTITUTE

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



Paul Gilna Ex. Dir.

National LambdaRail Direct Connect Computation and Storage Complex

PI Larry Smarr

Funded by: Gordon and Betty Moore Foundation

Joint Partnership of:

Announced January 17, 2006
$24.5M Over Seven Years

Building On Centers of Excellence and Achievement









Powered by Rocks







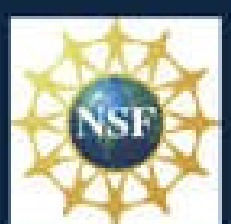


The Scripps Genome Center

Supported by:







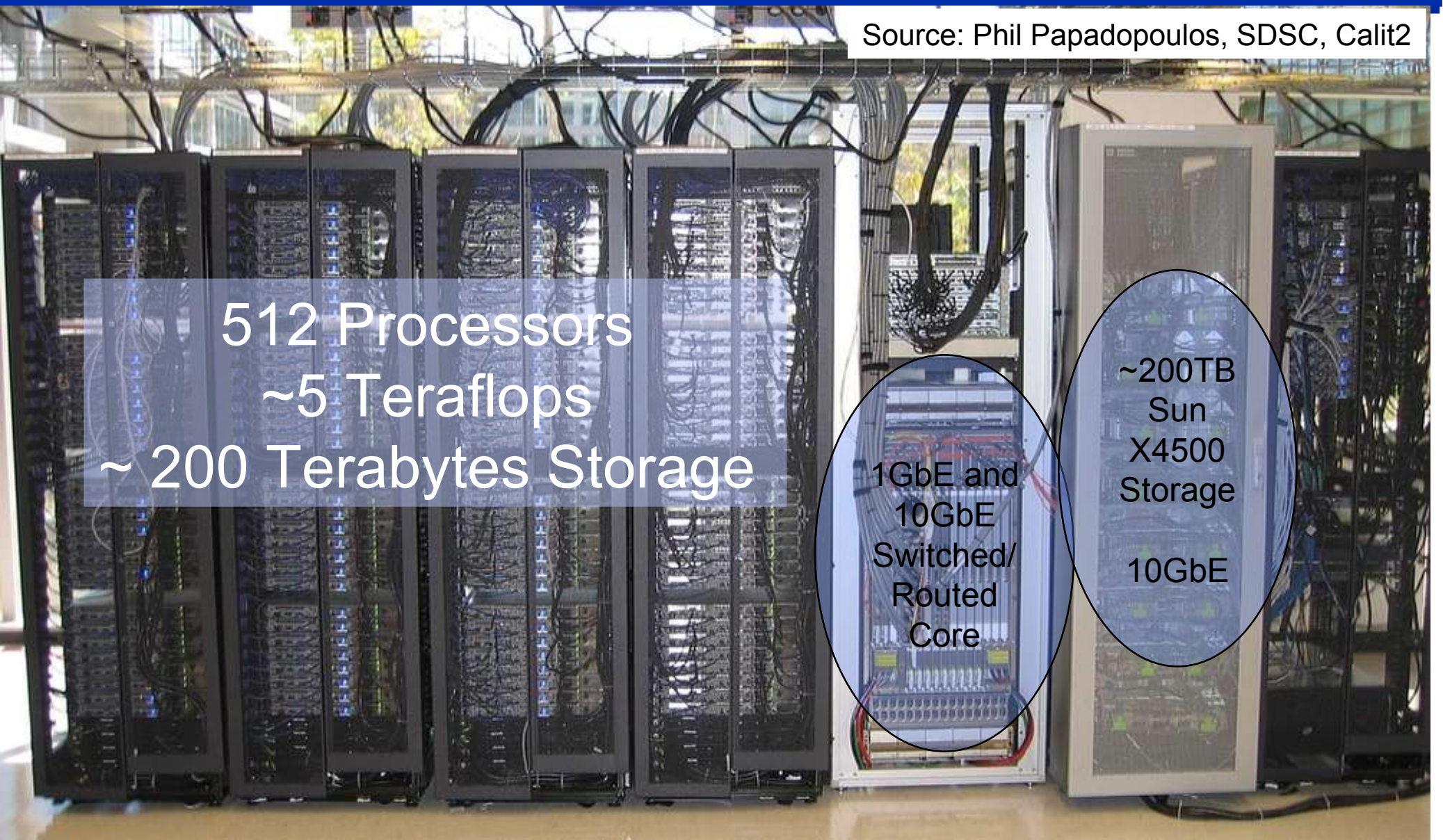








Calit2 Microbial Metagenomics Cluster- Next Generation Optically Linked Science Data Server



UCSD UCIrvine SDSC ROCKS camera Marine Microbial Ecology J. Craig Venter INSTITUTE it2

Marine Microbial Metagenomics is a Global Scientific Research Cyber-Community

Over 2100 Registered Users From 50 Countries

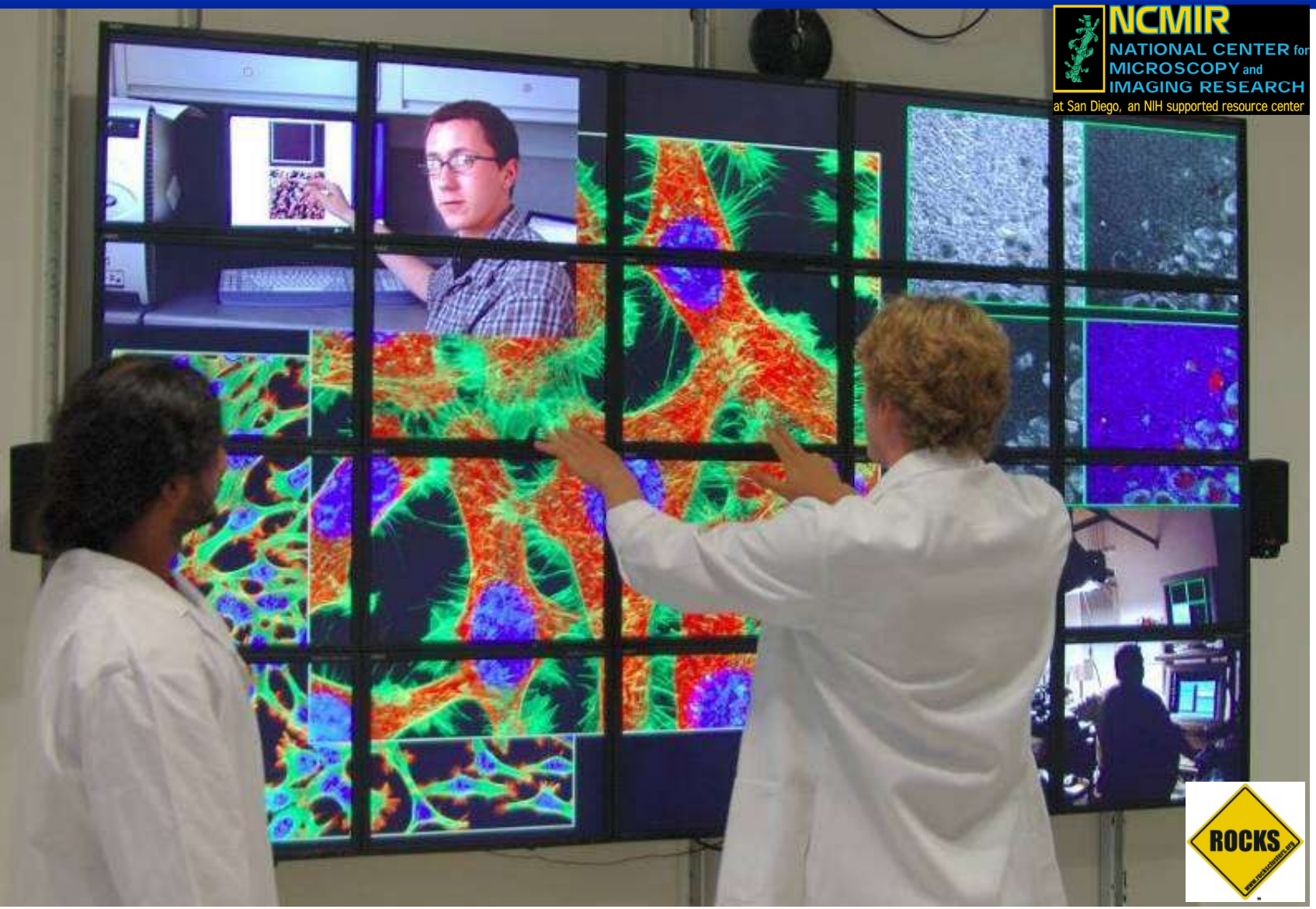


camera
Marine Microbial Ecology

J. Craig Venter
INSTITUTE

it2

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



Scalable Adaptive Graphics Environment (SAGE)

evl electronic visualization laboratory

NSF

Now in Sixth and Final Year

Picture Source: Mark Ellisman, David Lee, Jason Leigh

UCSD
UCIrvine

Calit2 (UCSD, UCI), SDSC, and UIC Leads—Larry Smarr PI
Univ. Partners: NCSA, USC, SDSU, NW, TA&M, UvA, SARA, KISTI, AIST
Industry: IBM, Sun, Telcordia, Chiaro, Calient, Glimmerglass, Lucent

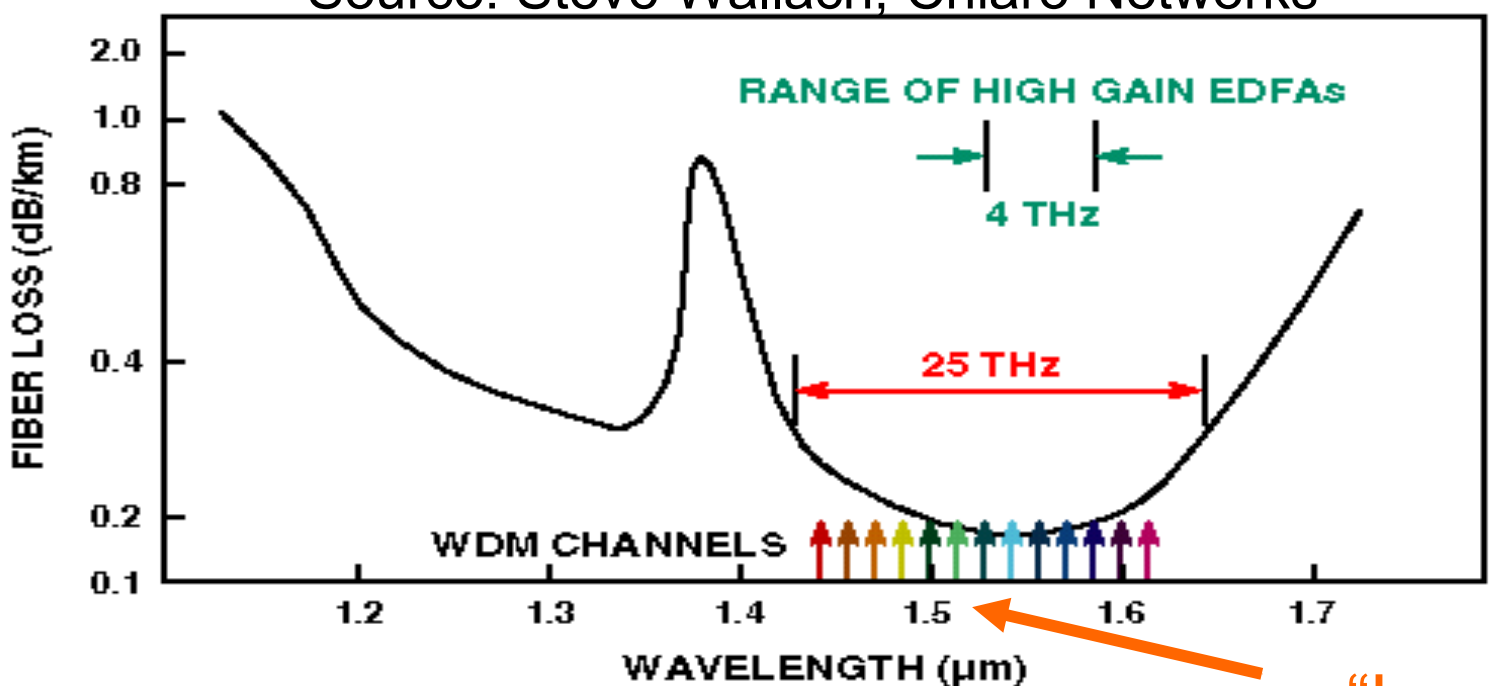
it2

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

UCSD
UCIrvine

it²

My OptIPortal™ – Affordable Termination Device for the OptIPuter Global Backplane

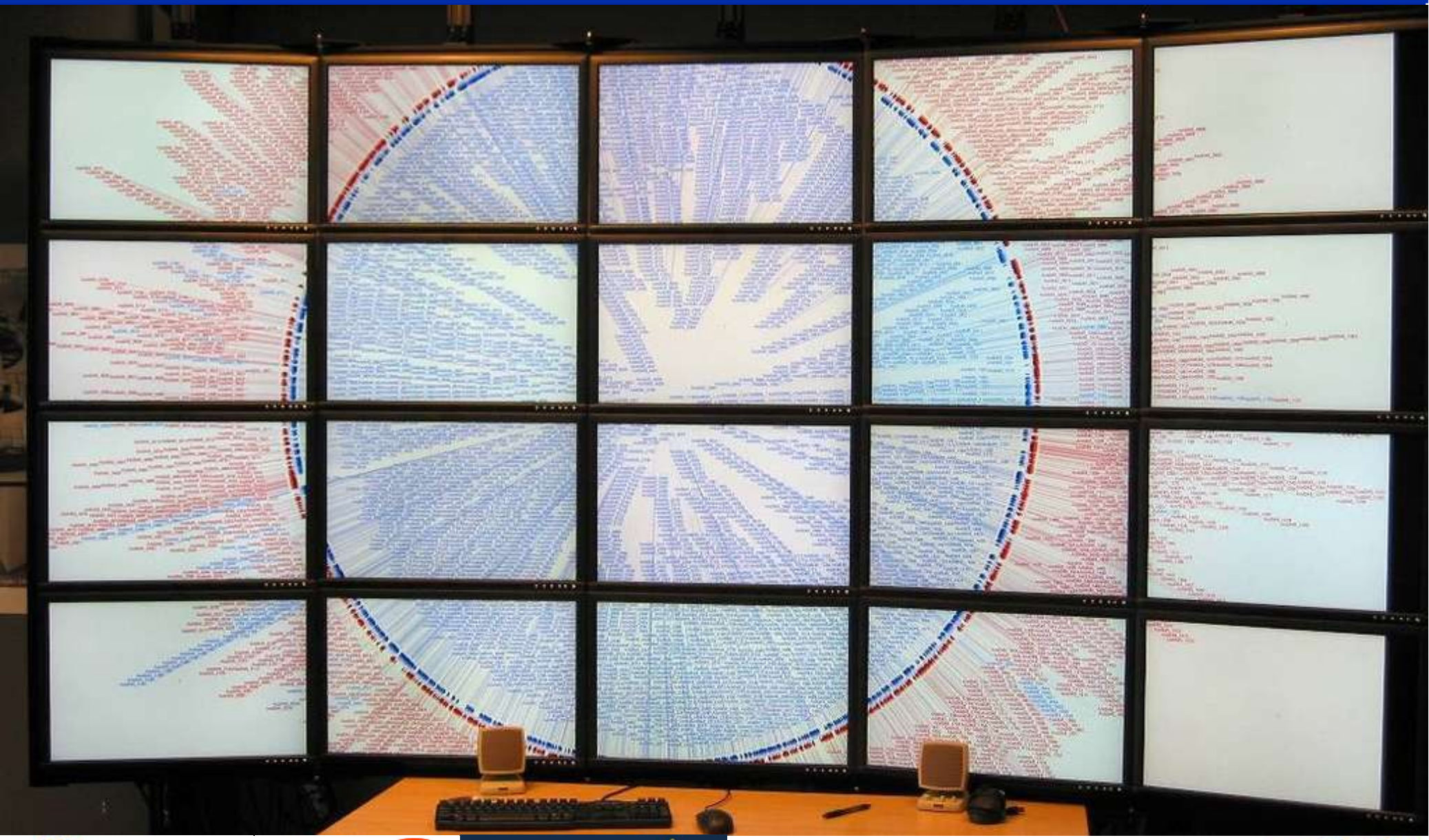
- 20 Dual CPU Nodes, Twenty 24” Monitors, ~$50,000
- 1/4 Teraflop, 5 Terabyte Storage, 45 Mega Pixels--Nice PC!
- Scalable Adaptive Graphics Environment (SAGE) Jason Leigh, EVL-UIC





Source: Phil Papadopoulos SDSC, Calit2

Use of Tiled Display Wall OptIPortal to Interactively View Microbial Genome



camera Marine Microbial Ecology

OptIPuter

Powered by Rocks ROCKS

Acidobacteria bacterium Ellin345 Soil Bacterium 5.6 Mb

Source: Raj Singh, UCSD

it²

Use of Tiled Display Wall OptIPortal to Interactively View Microbial Genome



camera Marine Microbial Ecology

OptIPuter

Powered by Rocks ROCKS

Source: Raj Singh, UCSD

it2

Use of Tiled Display Wall OptIPortal to Interactively View Microbial Genome



camera Marine Microbial Ecology

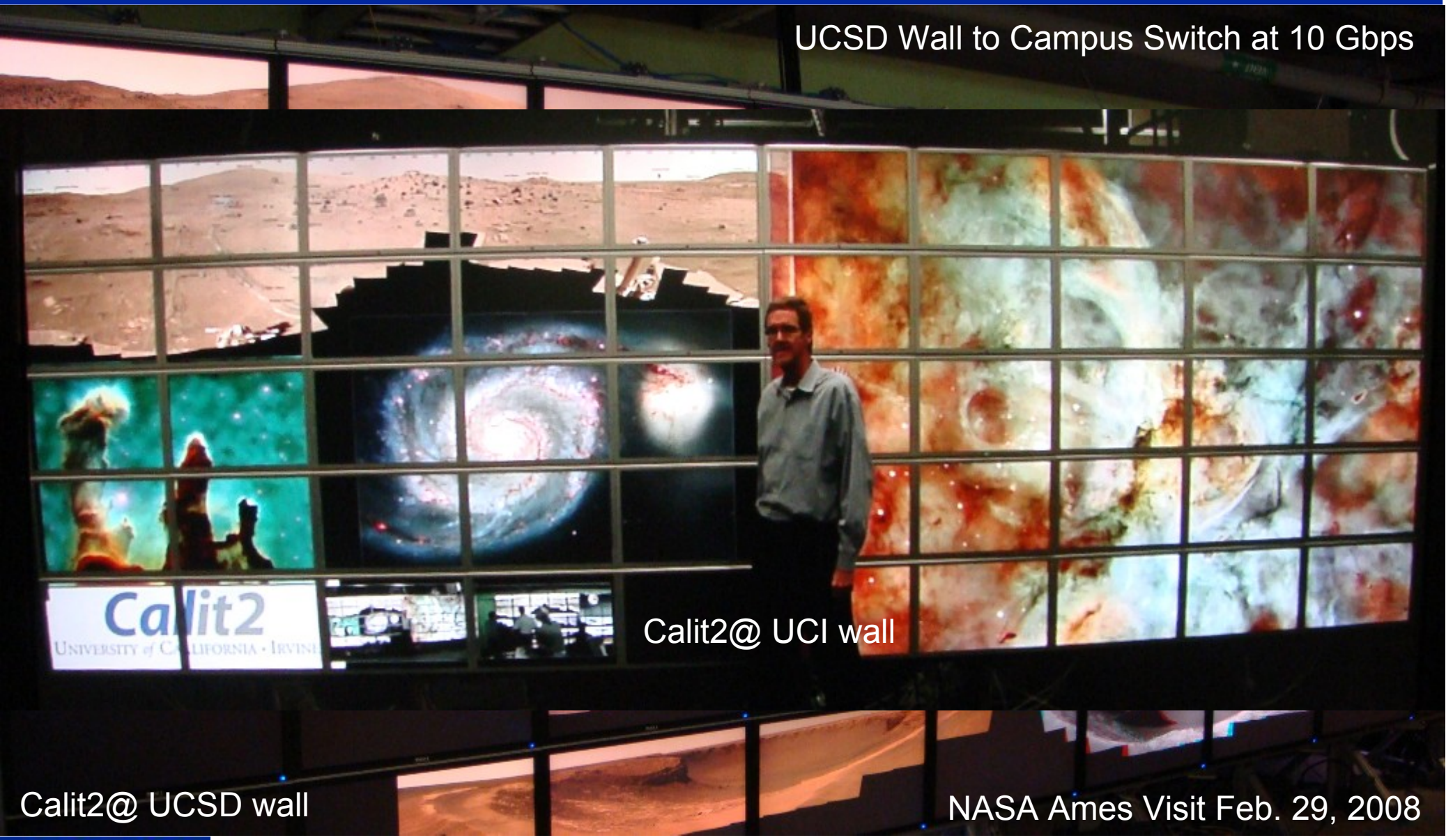
OptIPuter

Powered by Rocks ROCKS

Source: Raj Singh, UCSD

it²

The Calit2 1/4 Gigapixel OptIPortals at UCSD and UCI Are Joined to Form a Gbit/s HD Collaboratory



UCSD
UCIrvine

UCSD cluster: 15 x Quad core Dell XPS with Dual nVIDIA 5600s
UCI cluster: 25 x Dual Core Apple G5

OptIPlanet Collaboratory Persistent Infrastructure Between Calit2 and U Washington



OptIPuter

Ginger Armbrust's Diatoms: Micrographs, Chromosomes, Genetic Assembly

ROCKS



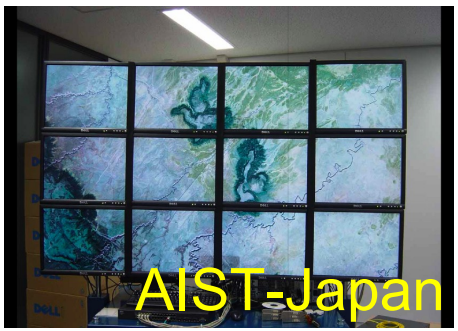




UW's Research Channel Michael Wellings

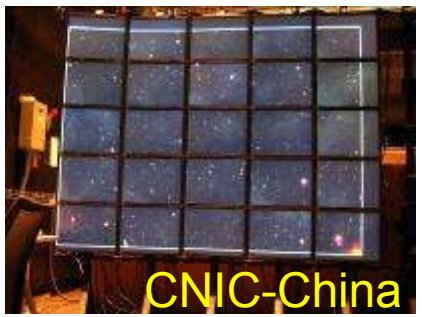
it2

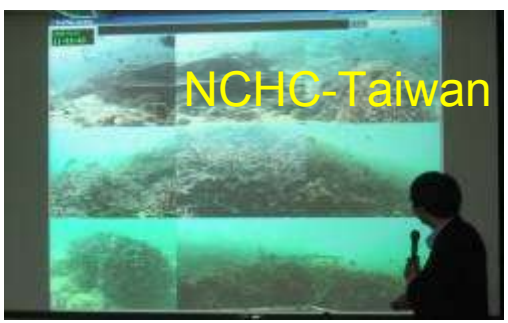
OptIPortals
Are Being Adopted Globally























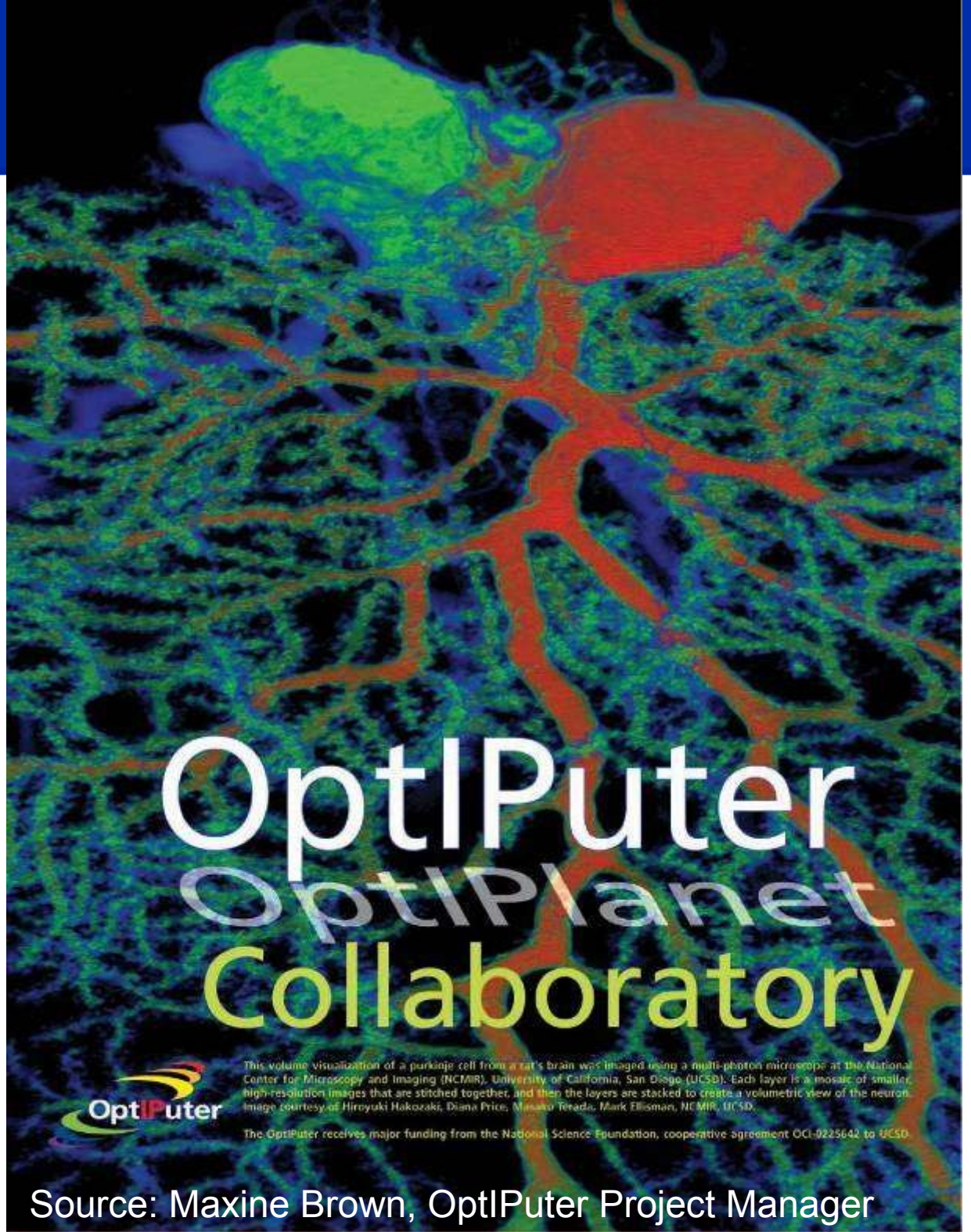


UCSD
UCIrvine

Powered by Rocks ROCKS

evl electronic visualization laboratory

it2



Source: Maxine Brown, OptIPuter Project Manager

Green Initiative:

Can Optical Fiber Replace Airline Travel for Continuing Collaborations?

UCSD
UCIrvine

New Year's Challenge: Streaming Underwater Video From Taiwan's Kenting Reef to Calit2's OptIPortal



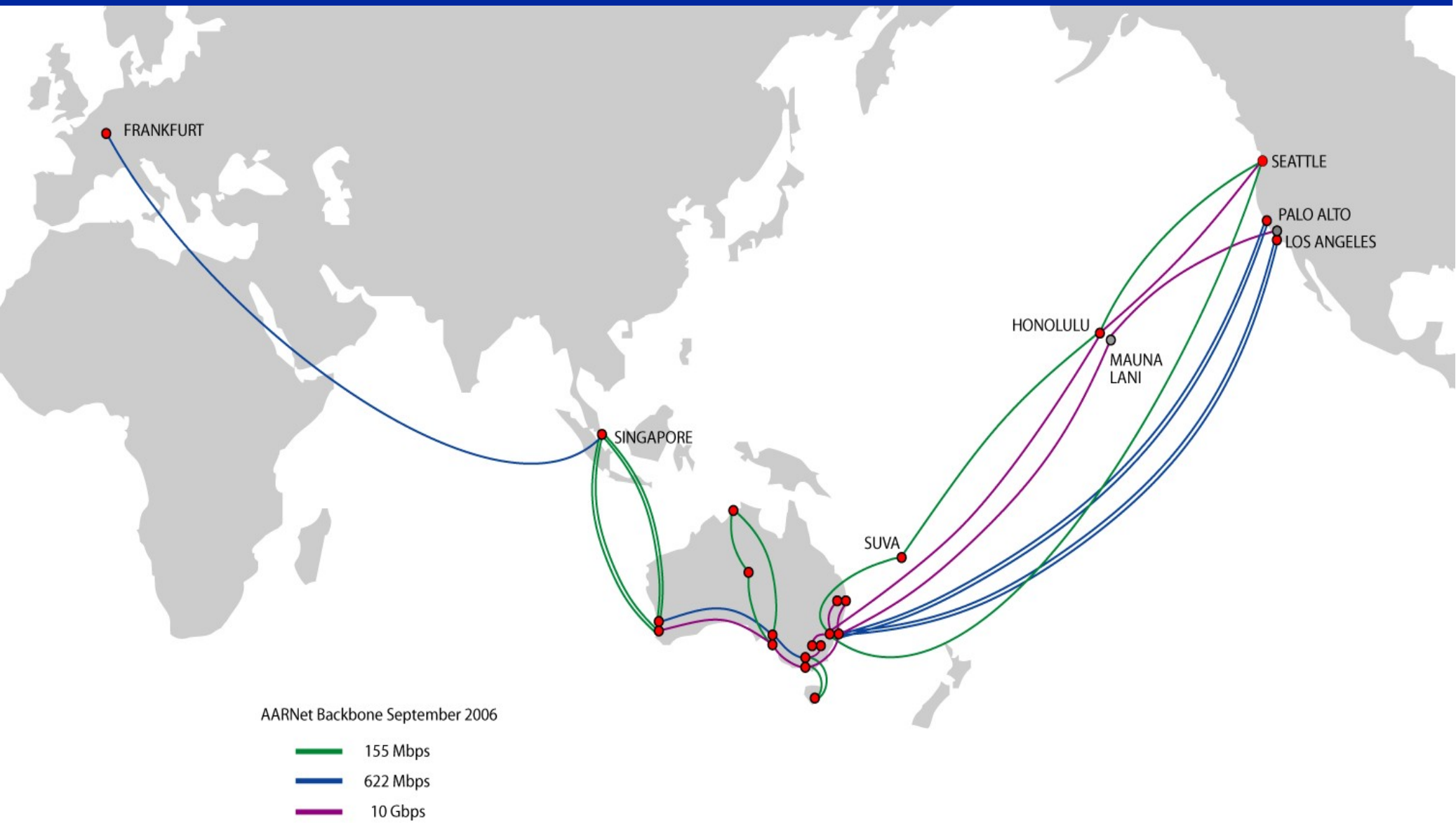
OPEN SOURCE DATA TURBINE INITIATIVE
Empowering the Scientific Community with Streaming Data Middleware

UCSD
UCIrvine

UCSD: Rajvikram Singh, Sameer Tilak, Jurgen Schulze, Tony Fountain, Peter Arzberger
NCHC : Ebbe Strandell, Sun-In Lin, Yao-Tsung Wang, Fang-Pang Lin

it2

AARNet International Network



UCSD
UCIrvine

Launch of the 100 Megapixel OzIPortal Over Qvidium Compressed HD on 1 Gbps CENIC/PW/AARNet Fiber



UCSD
UCIrvine

www.calit2.net/newsroom/release.php?id=1219

Victoria Premier and Australian Deputy Prime Minister Asking Questions



UCSD
UCIrvine

www.calit2.net/newsroom/release.php?id=1219

it²

University of Melbourne Vice Chancellor Glyn Davis in Calit2 Replies to Question from Australia



UCSD
UCIrvine

OptIPuterizing Australian Universities in 2008: CENIC Coupling to AARNet



UMelbourne/Calit2 Telepresence Session May 21, 2008

Augmented by Many Physical Visits This Year
Culminating in Two Week Lecture Tour
of Australian Research Universities
by Larry Smarr October 2008

Phil Scanlan
Founder-
Australian American
Leadership Dialogue
www.aald.org



UCSD
UCIrvine

Draft Schedule
Smarr AALD Lecture Tour October 2008

- Oct 2—University of Adelaide
- Oct 6—Univ of Western Australia
- Oct 8—Monash University
- Oct 9—University of Melbourne
- Oct 10—University of Queensland
- Oct 14—University of New South Wales
- Oct 15—Leadership Dialogue Scholar Oration, Canberra
- Oct 16—CSIRO OptIPortal Dedication
- Oct 16—Sydney University

UCSD
UCIrvine