“Metagenomics Over Lambdas: Update on the CAMERA Project"

Invited Talk

6th Annual ON*VECTOR International Photonics Workshop

Calit2@ UCSD

February 27, 2007

Dr. Larry Smarr

Director, California Institute for Telecommunications and Information Technology

Harry E. Gruber Professor,

Dept. of Computer Science and Engineering

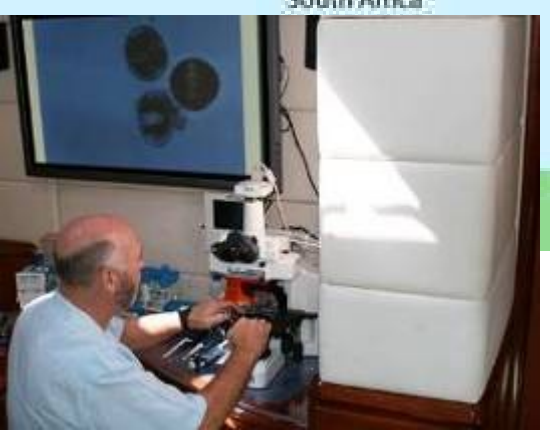
Jacobs School of Engineering, UCSD

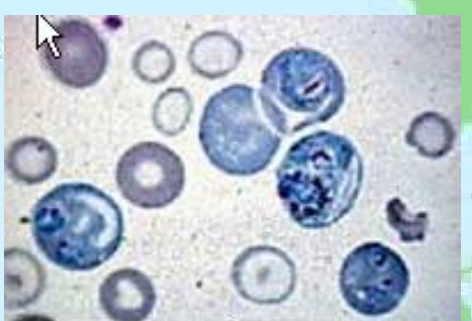
OptIPuter

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
New Caledonia
Australia
x 8 samples
Analyzed
Sampled, but not Analyzed

Sorcerer II Data Will Double Number of Proteins in GenBank!

MARINE MICROBIOLOGY INITIATIVE

Gordon and Betty MOORE FOUNDATION

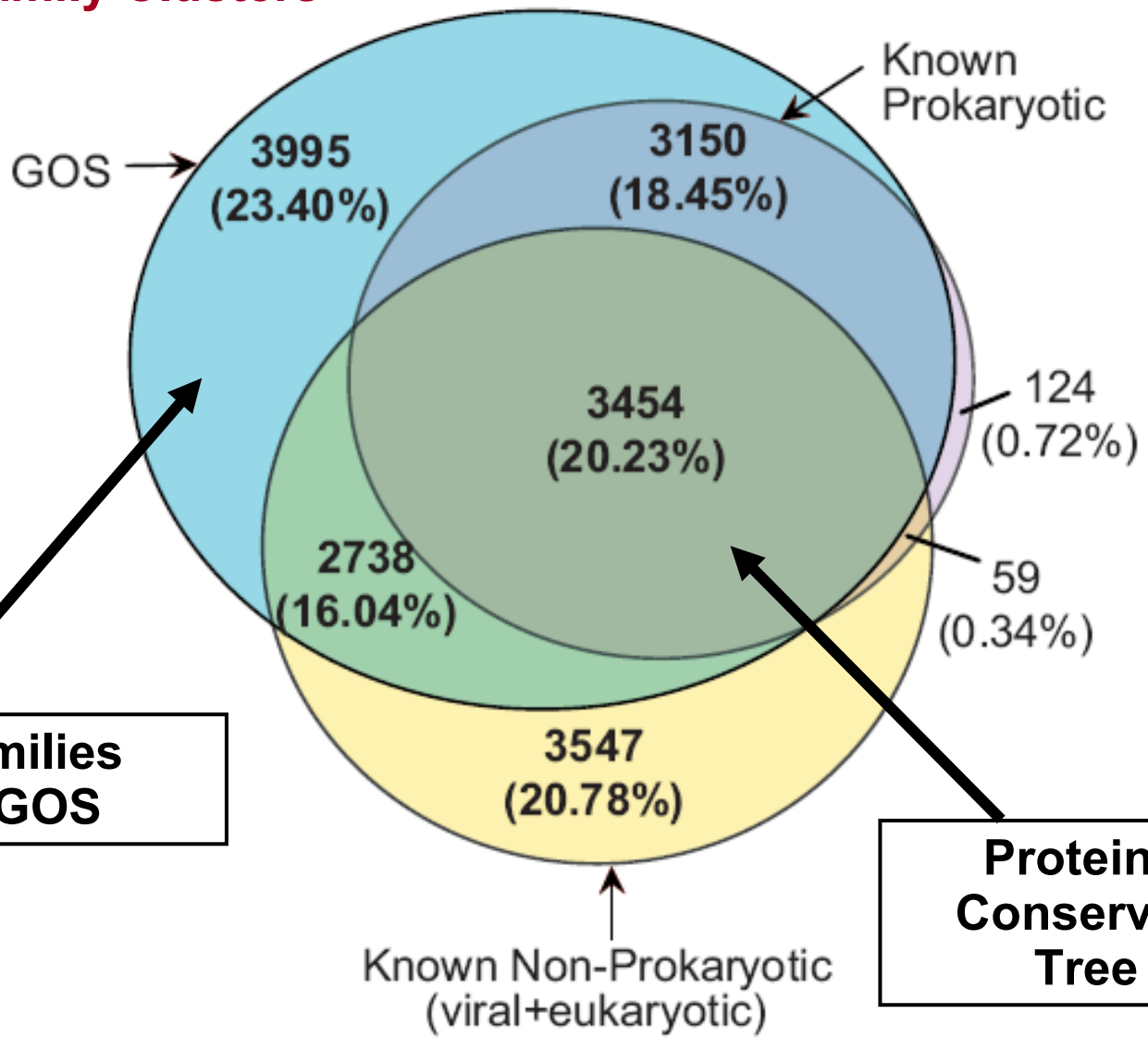
OptIPuter

J. Craig Venter INSTITUTE

it²

Current Universe of Medium/ Large Protein Families

17,067 Protein Family Clusters



Protein Families Unique to GOS

Protein Families Conserved Across Tree of Life

OptIPuter

Source: Shibu Yooseph, et al. (PLOS Biology in press 2006)

it2

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:

OptIPuter

Announced January 17, 2006
$24.5M Over Seven Years

CAMERA Has Development and Production Compute and Storage Clusters On-Line

camera
Marine Microbial Ecology



Development Server
64 Processors
12 Terabytes Storage

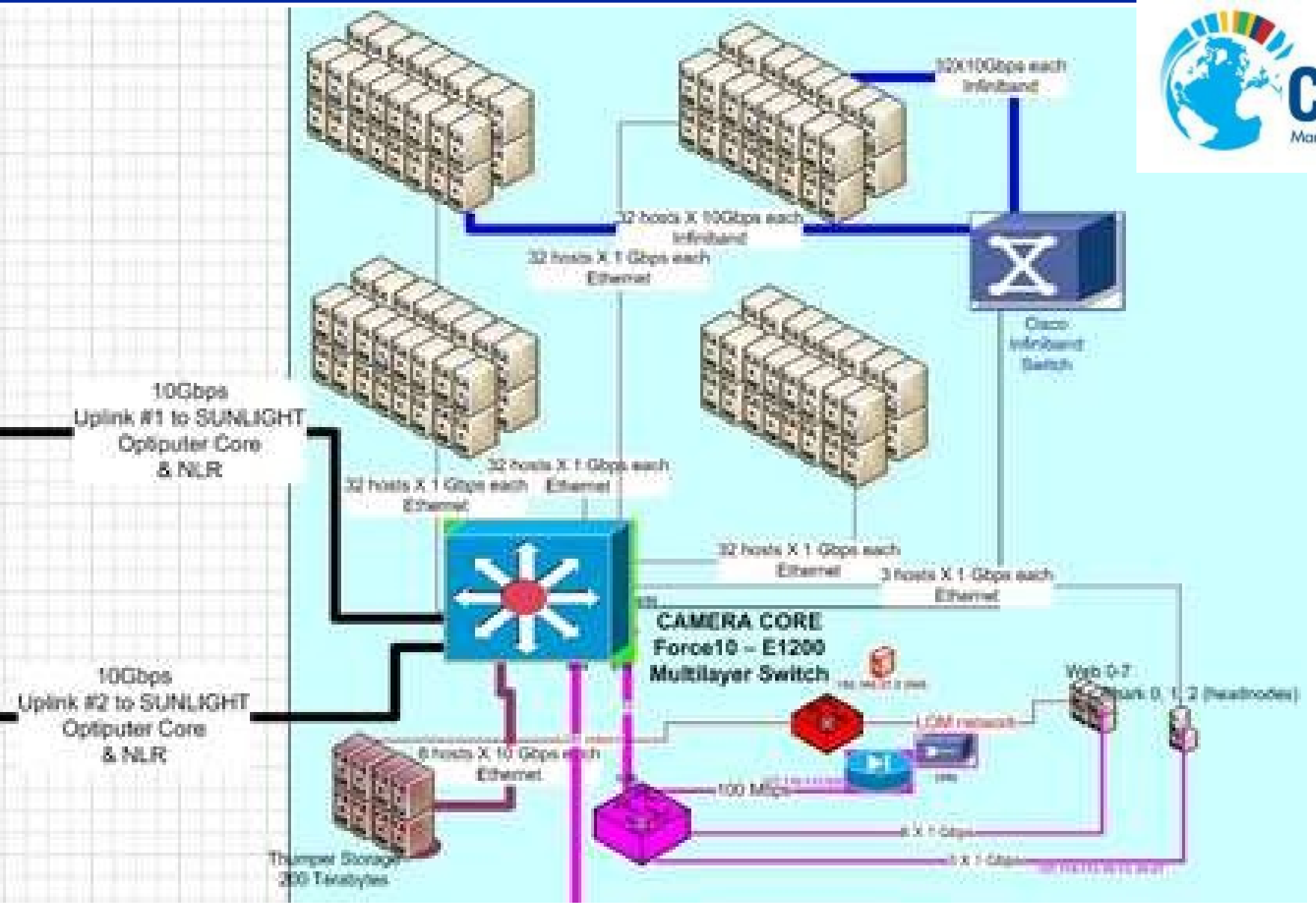


Production Server
512 Processors
~5 Teraflops
~ 200 Terabytes Storage

OptIPuter

it2

The CAMERA Compute and Server Core is Connected Internally with a 1 and 10 Gbps Fabric



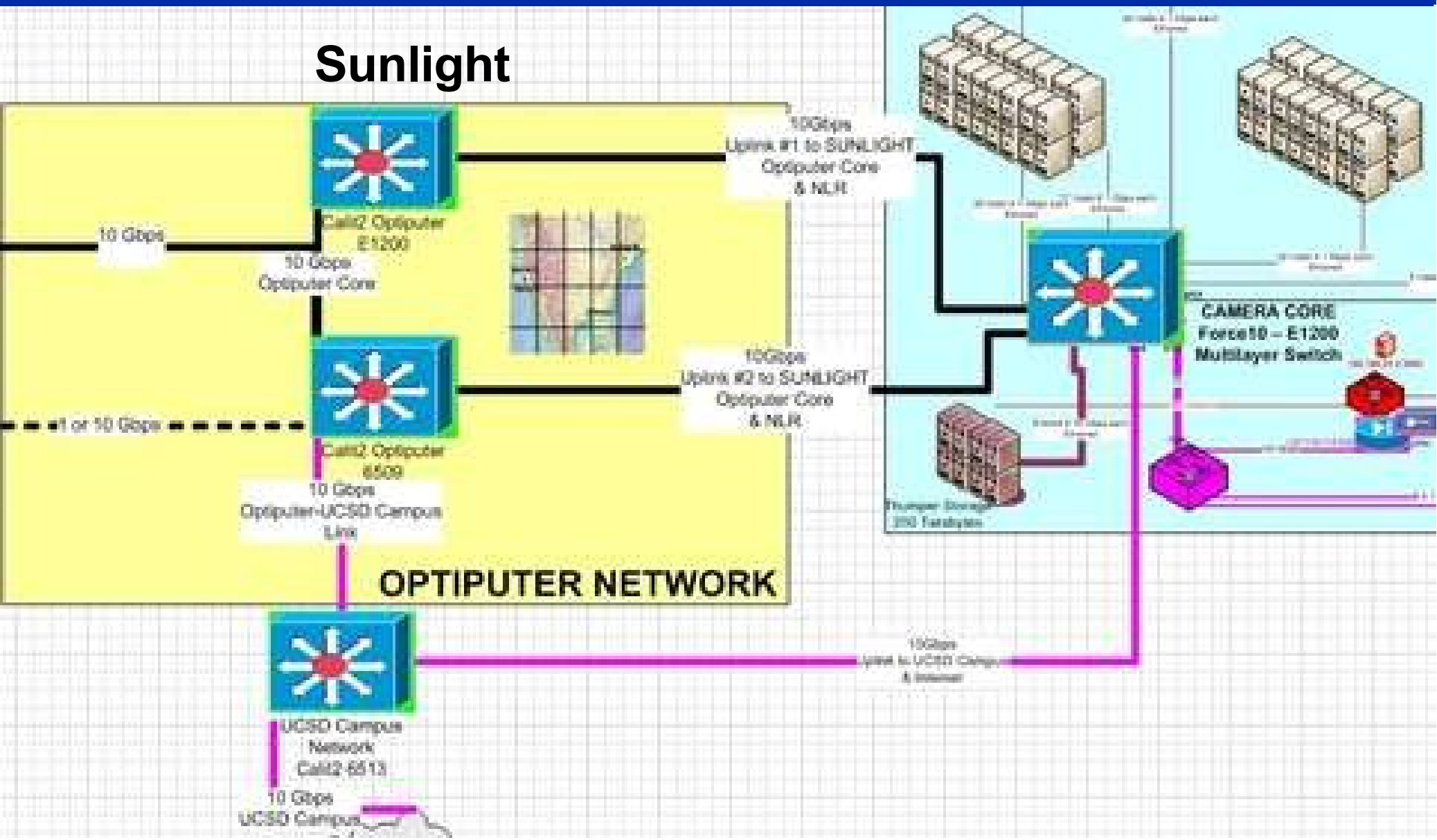
camera
Marine Microbial Ecology

OptIPuter

it2

Source: Brian Dunne, Calit2

The CAMERA Server Complex is Connected via Multiple 10 Gbps Links to Calit2 Sunlight



Source: Brian Dunne, Calit2

Calit2 is Now OptIPuter Connecting Remote OptIPortal Moore-Funded Microbial Researchers via NLR



OptIPortal–
Termination Device for the OptIPuter Global Backplane

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

evl electronic visualization laboratory





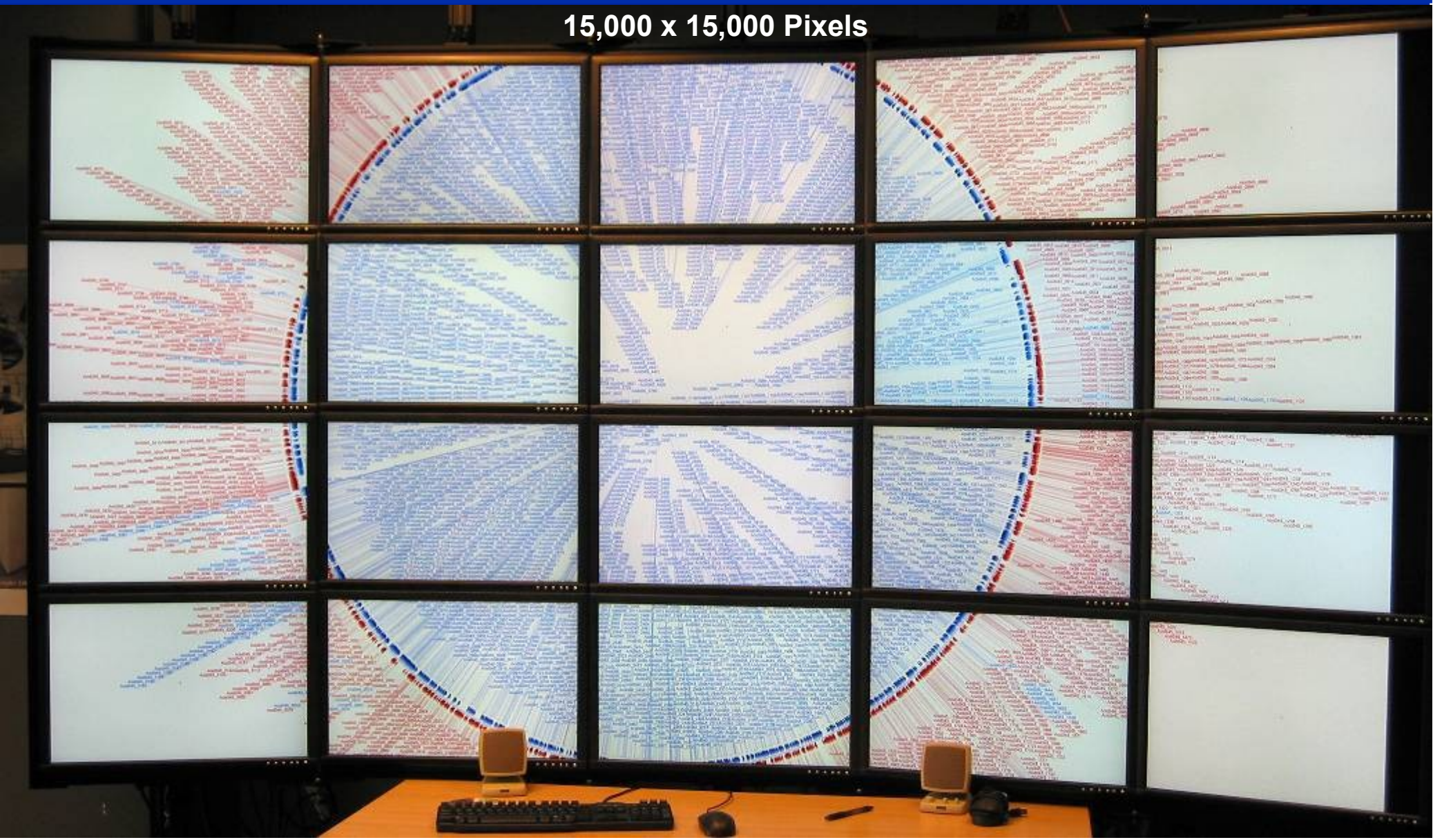
OptIPuter

Powered by Rocks ROCKS

Source: Phil Papadopoulos SDSC, Calit2

it2

Use of OptIPortal to Interactively View Microbial Genome



OptIPuter

Source: Raj Singh, UCSD

Acidobacteria bacterium Ellin345 (NCBI)
Soil Bacterium 5.6 Mb

it²

Use of OptIPortal to Interactively View Microbial Genome



OptIPuter

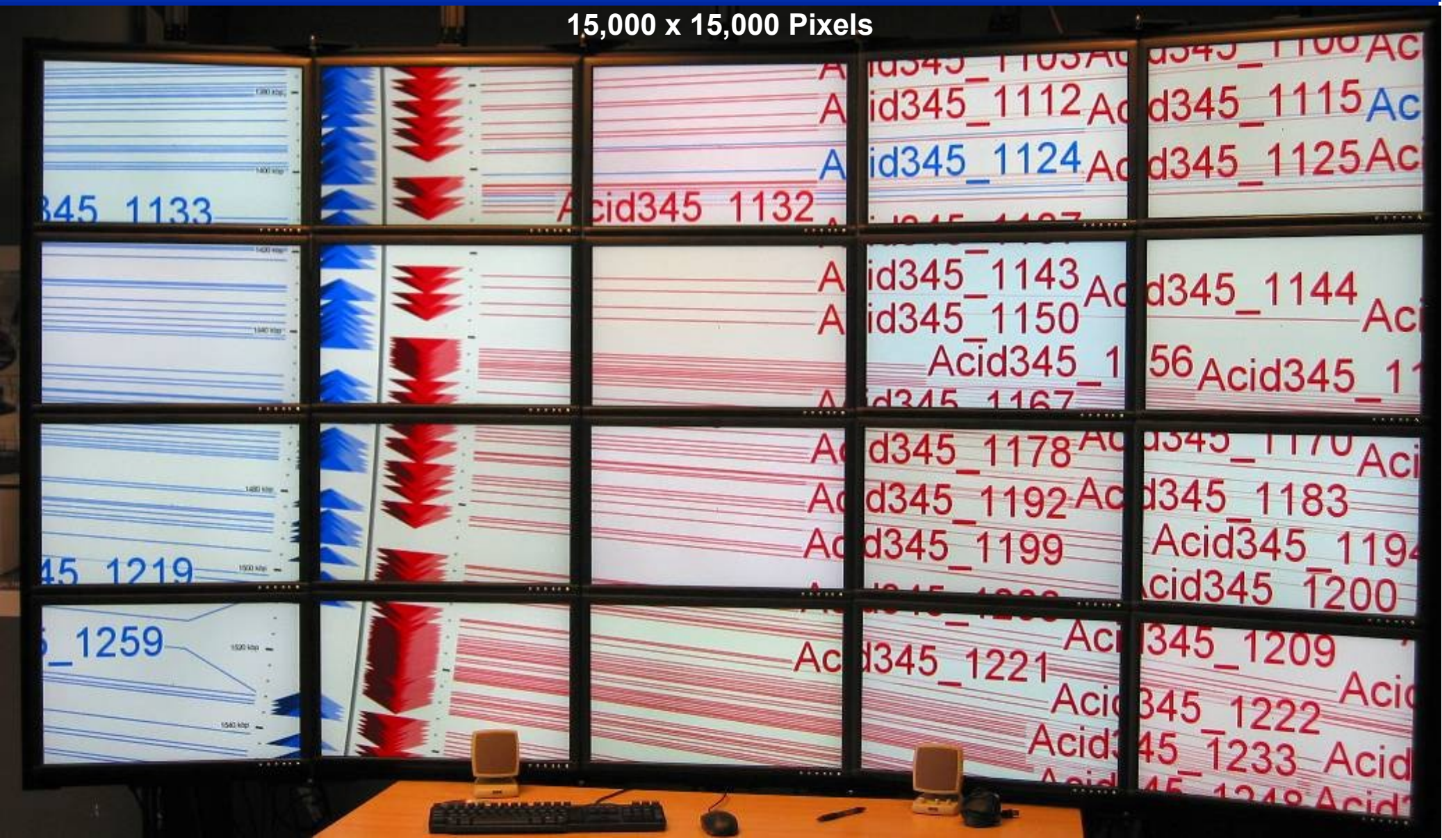
Source: Raj Singh, UCSD

Acidobacteria bacterium Ellin345 (NCBI)
Soil Bacterium 5.6 Mb

it²

Use of OptIPortal to Interactively View Microbial Genome

15,000 x 15,000 Pixels



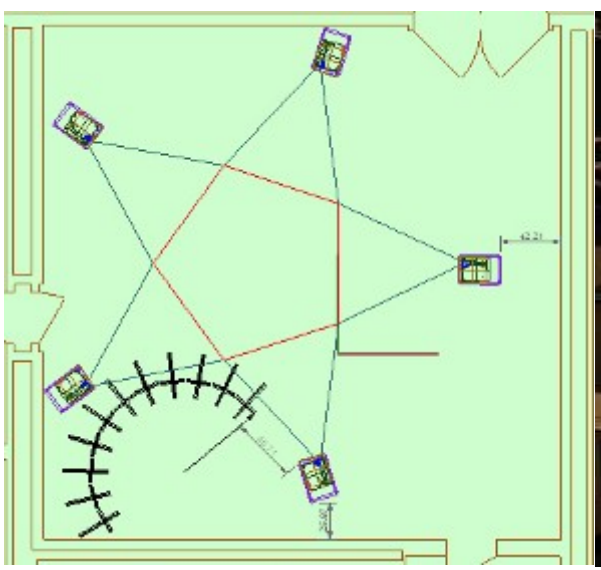
Source: Raj Singh, UCSD

Acidobacteria bacterium Ellin345 (NCBI)
Soil Bacterium 5.6 Mb

Partnering with UIC Electronic Visualization Lab to Create Next Generation OptIPortals

- Varrier Autostereo Virtual Reality
  - Head-Tracked No Need for Glasses
  - 65 LCD Tiles
  - 45 Mpixels/eye of Visual Stereo
- PentaCAVE— High Definition Surround
  - Working Prototype 4 Mpixel Wall
  - Full Scale PentaCAVE Being Built
    - 6 JVC HD2K Projectors Per Wall
    - 30 Mpixel/eye of Stereo w/5-Walls







evl electronic visualization laboratory

OptIPuter

Dan Sandin, Greg Dawe, Tom Peterka, Tom DeFanti, Jason Leigh, Jinghua Ge, Javier Girado, Bob Kooima, Todd Margolis, Lance Long, Alan Verlo, Maxine Brown, Jurgen Schulze, Qian Liu, Ian Kaufman, Bryan Glogowski

it2

The CAMERA Beta Site is Now Active

camera

HOME | ABOUT CAMERA | METAGENOMICS | RESEARCH | NEWS | EVENTS



J. Craig Venter, UCSD chancellor Marye Anne Fox, Larry Smarr, and John Orcutt at the news conference announcing award of the CAMERA project to CalIT2 UCSD January 17, 2006, at the Birch Aquarium at UCSD.



Members of the CAMERA Science Advisory Board at a meeting on December 13, 2006, to review

WELCOME TO CAMERA BETA RELEASE CANDIDATE 0.7!

We are proud to announce that the CAMERA (Community Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis) website is now available to the public for a Beta testing period. CAMERA is a user-driven site dedicated to providing the scientific community with metagenomics data and software analysis tools. This public release of the CAMERA system marks the beginning of a two-month Beta test period for CAMERA software capabilities, data, and website. We invite you to browse and help us test the functionality of the site. If you discover any problems or have suggestions to help us prioritize our work, please e-mail us. For further information about the project, see About CAMERA.

We invite you to register for access to the CAMERA beta site; your registration will remain in place once the Beta test period expires.

Register Now!

If you are already registered, click on the "Log In" button at the top right of this page.

BLAST AVAILABLE

We have implemented a BLAST tool for querying your favorite sequences against a suite of microbial whole genome and metagenomic sequence databases. In particular, the output has been customized to give users direct access to the metadata associated with

Funded by the Moore Foundation

Upcoming Events

- Microbial Genomics 2007 April 11th, 2007
- Bio-IT World April 30th, 2007
- American Society for Microbiology General 107th Meeting May 21st, 2007
- Environmental Genomics: From Individual Genomes to Genomes of Complex Communities June 9th, 2007
- Microbial Community Networks (Biogeco9) June

OptIPuter

http://camera.calit2.net/

J. Craig Venter INSTITUTE

it2

CAMERA
Global Ocean Sampling Page



Sample ▲	Size	Habitat Type	Geographical Location	Sample Location	Country	Latitude	Longitude	Sample Depth
GS000a	770.4 MB	Open Ocean	Sargasso Sea	Sargasso Station 13	Bermuda (UK)	31d 32'06"n	63d 35'42"w	5 m
GS000b	376.4 MB	Open Ocean	Sargasso Sea	Sargasso Station 13	Bermuda (UK)	31d 32'06"n	63d 35'42"w	5 m
GS000c	436 MB	Open Ocean	Sargasso Sea	Sargasso Stations 3	Bermuda (UK)	32d 10'29.4"n	64d 00'36.6"w	5 m
GS000d	393.8 MB	Open Ocean	Sargasso Sea	Sargasso Station 13	Bermuda (UK)	31d 32'06"n	63d 35'42"w	5 m
GS001a	168.1 MB	Open Ocean	Sargasso Sea	Hydrostation S	Bermuda (UK)	32d 10'00"n	64d 30'00"w	5 m
GS001b	107.1 MB	Open Ocean	Sargasso Sea	Hydrostation S	Bermuda (UK)	32d 10'00"n	64d 30'00"w	5 m
GS001c	108.8 MB	Open Ocean	Sargasso Sea	Hydrostation S	Bermuda (UK)	32d 10'00"n	64d 30'00"w	5 m
GS002	150.2 MB	Coastal	North American East Coast	Gulf of Maine	USA	42d 30'11"n	67d 14'24"w	1 m
GS003	77.7 MB	Coastal	North American East Coast	Browns Bank, Gulf of Maine	Canada	42d 51'19"n	66d 13'2"w	1 m
GS004	66.4 MB	Coastal	North American East Coast	Outside Halifax, Nova Scotia	Canada	44d 8'14"n	63d 38'40"w	2 m
GS005	78.7 MB	Embayment	North American East Coast	Bedford Basin, Nova Scotia	Canada	44d 41'25"n	63d 38'14"w	1 m
GS006	75.1 MB	Estuary	North American East Coast	Bay of Fundy, Nova Scotia	Canada	45d 8'42"n	64d 58'48"w	1 m
GS007	84.3 MB	Coastal	North American East Coast	Northern Gulf of Maine	Canada	43d 37'56"n	66d 50'50"w	1 m
GS008	160.8 MB	Coastal	North American East Coast	Newport Harbor, RI	USA	41d 29'9"n	71d 21'4"w	1 m
GS009	99.6 MB	Coastal	North American East Coast	Block Island, NY	USA	41d 5'26"n	71d 35'6"w	1 m
GS010	96.2 MB	Coastal	North American East Coast	Cape May, NJ	USA	38d 55'24"n	74d 41'8"w	1 m

OptIPuter

J. Craig Venter INSTITUTE

it2

CAMERA is Already in Use Worldwide

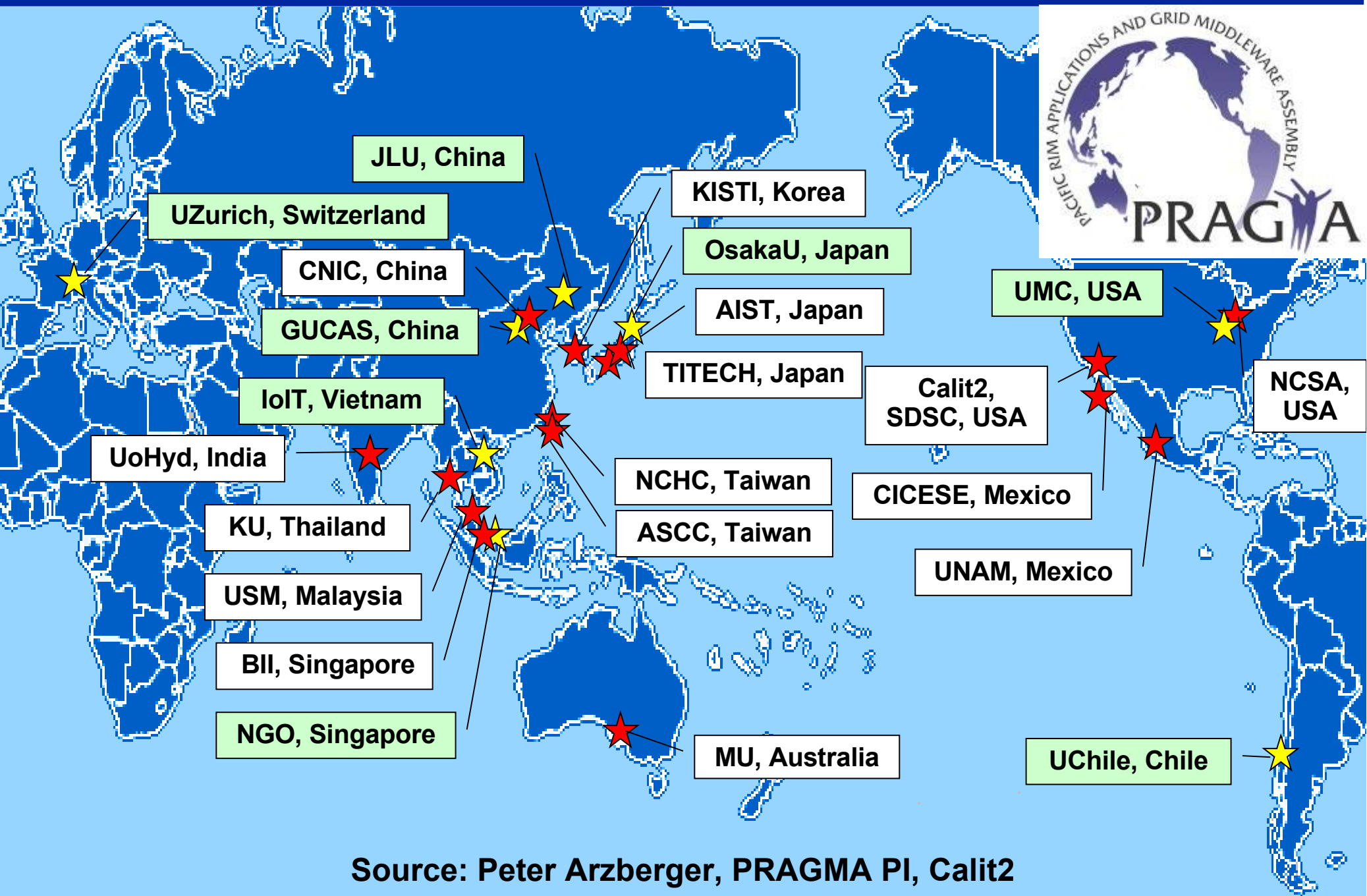
- **Users from over 40 Institutions in North America, Europe, and the South Pacific**
  - **Including Australia, Canada, France, Germany, Mexico, the Netherlands, Spain, Sweden, Switzerland, and the U.K.**
  - **> 120 Research Scientists, Postdocs, and Students**

- **In the First Two Weeks:**
  - **133 Blast Jobs Were Submitted**
  - **Consumed Over 600 CPU Hours**
  - **Generated 183 Gigabyte of Output Data**

OptIPuter

it2

Pacific Rim Applications and Grid Middleware Assembly Are Exploring LambdaGrids to Support Metagenomics



Source: Peter Arzberger, PRAGMA PI, Calit2