# Microbial Metagenomics and Human Health

Invited Talk
Health Sciences Advisory Board
School of Medicine
University of California, San Diego
May 8, 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**


UCSD
UCIrvine

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



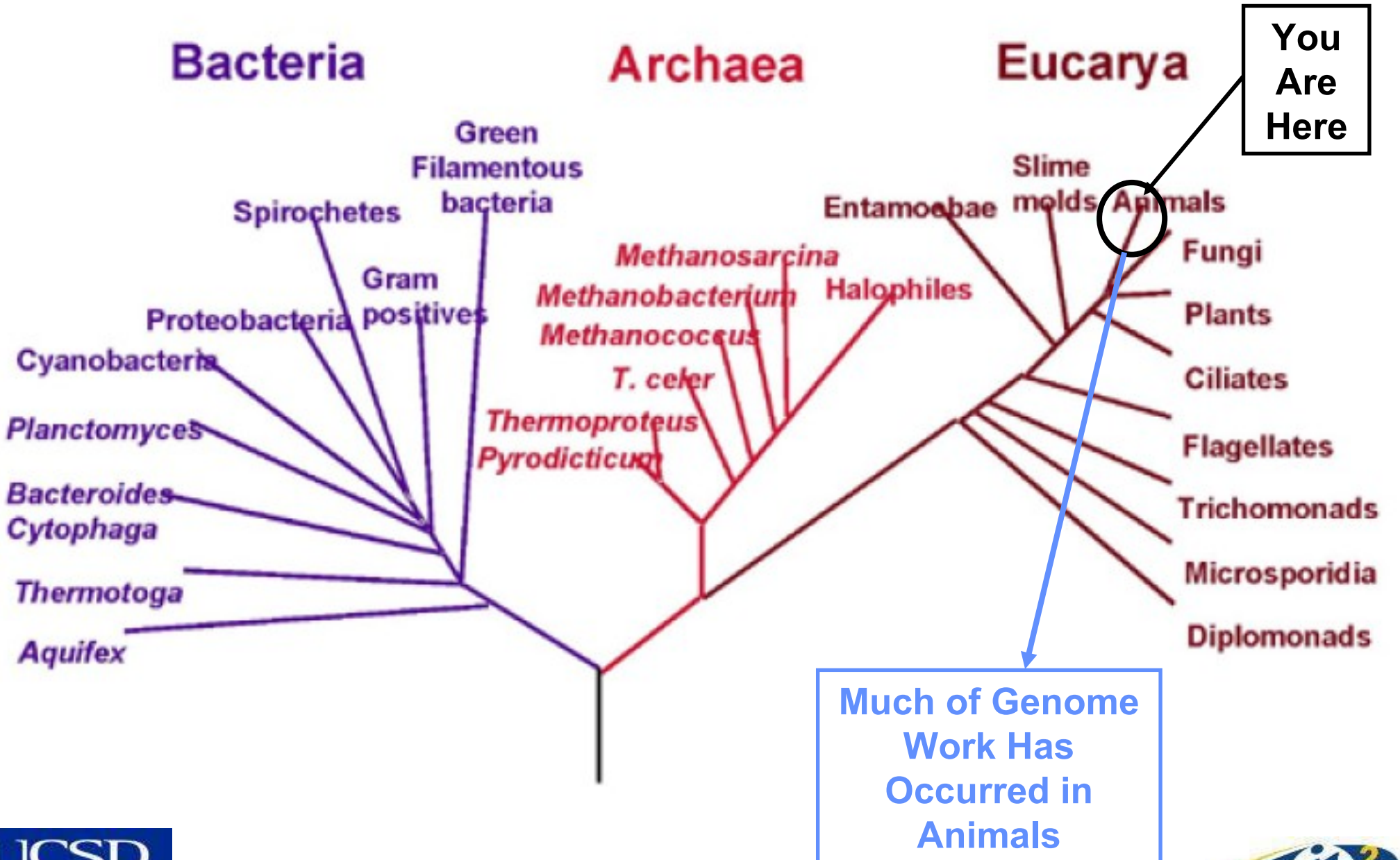




**1200 Researchers in Two Buildings**

UCSD
UCIrvine

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

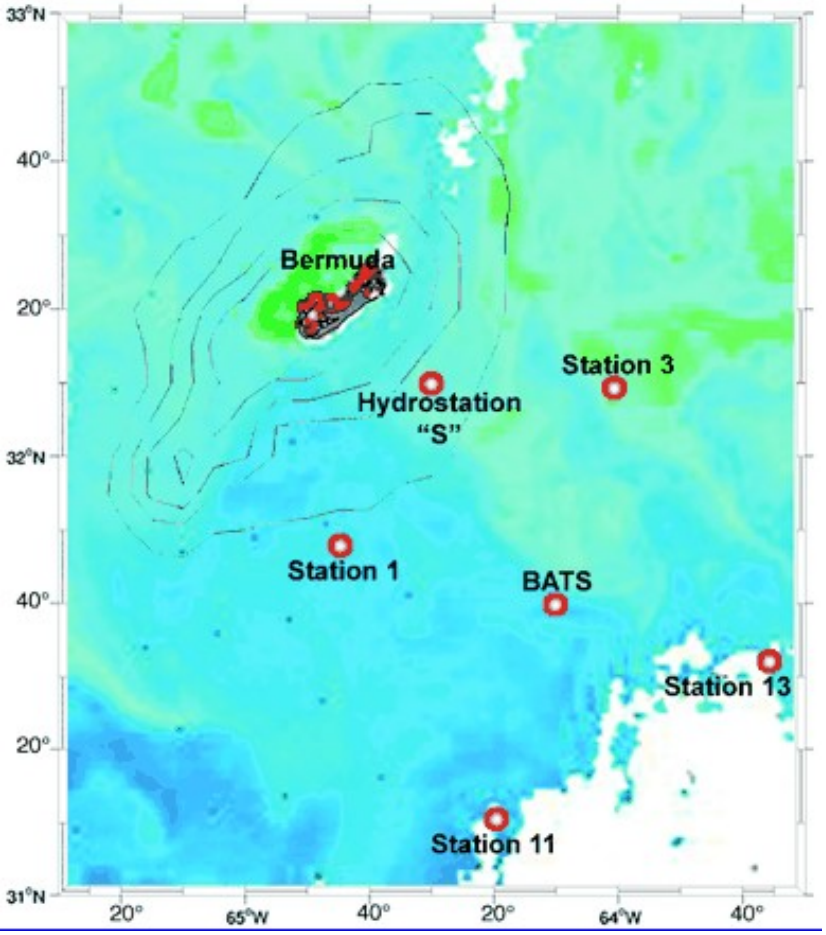



Source: Carl Woese, et al

UCSD
UCIrvine

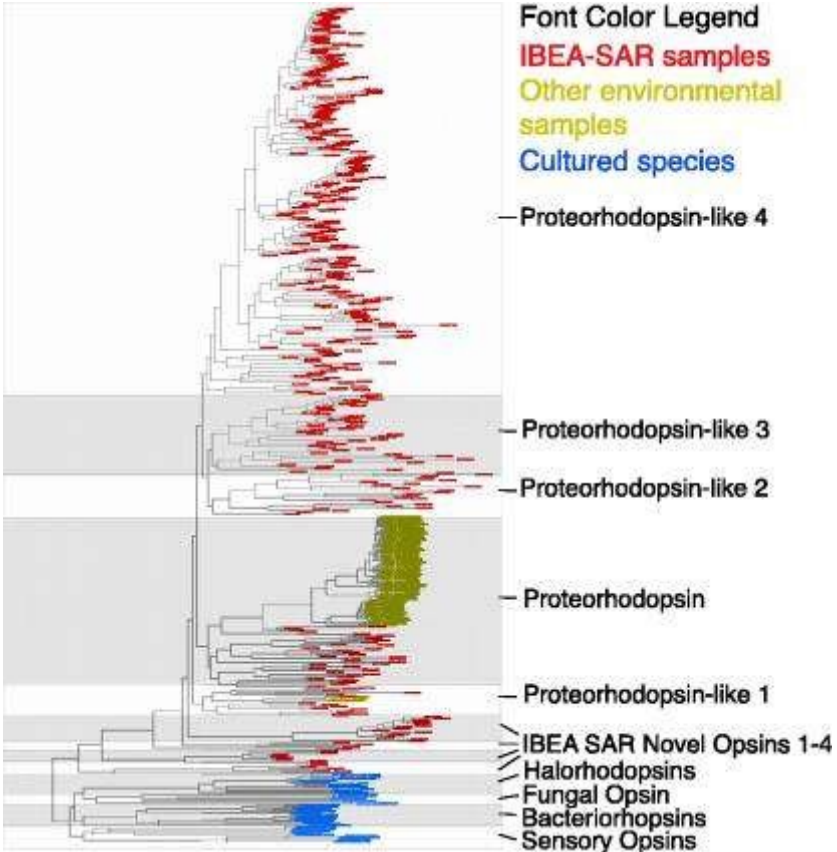
it2

# 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

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

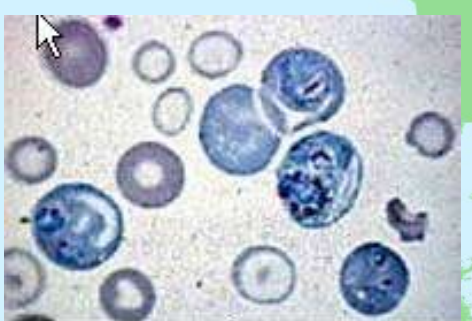
Joint Partnership of:

UCSD UCIrvine

**Announced January 17, 2006**
**$24.5M Over Seven Years**

# 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
French Polynesia
Vanuatu
Fiji
Tonga
Cook Islands
New Caledonia
x 17 samples
Australia
Cocos-Keeling Is.
Tanzania/ Zanzibar
x 7 samples
Seychelles
Madagascar
South Africa
x 8 samples
Analyzed
Sampled, but not Analyzed

**CAMERA will include**
**All Sorcerer II Metagenomic Data**

MARINE MICROBIOLOGY INITIATIVE

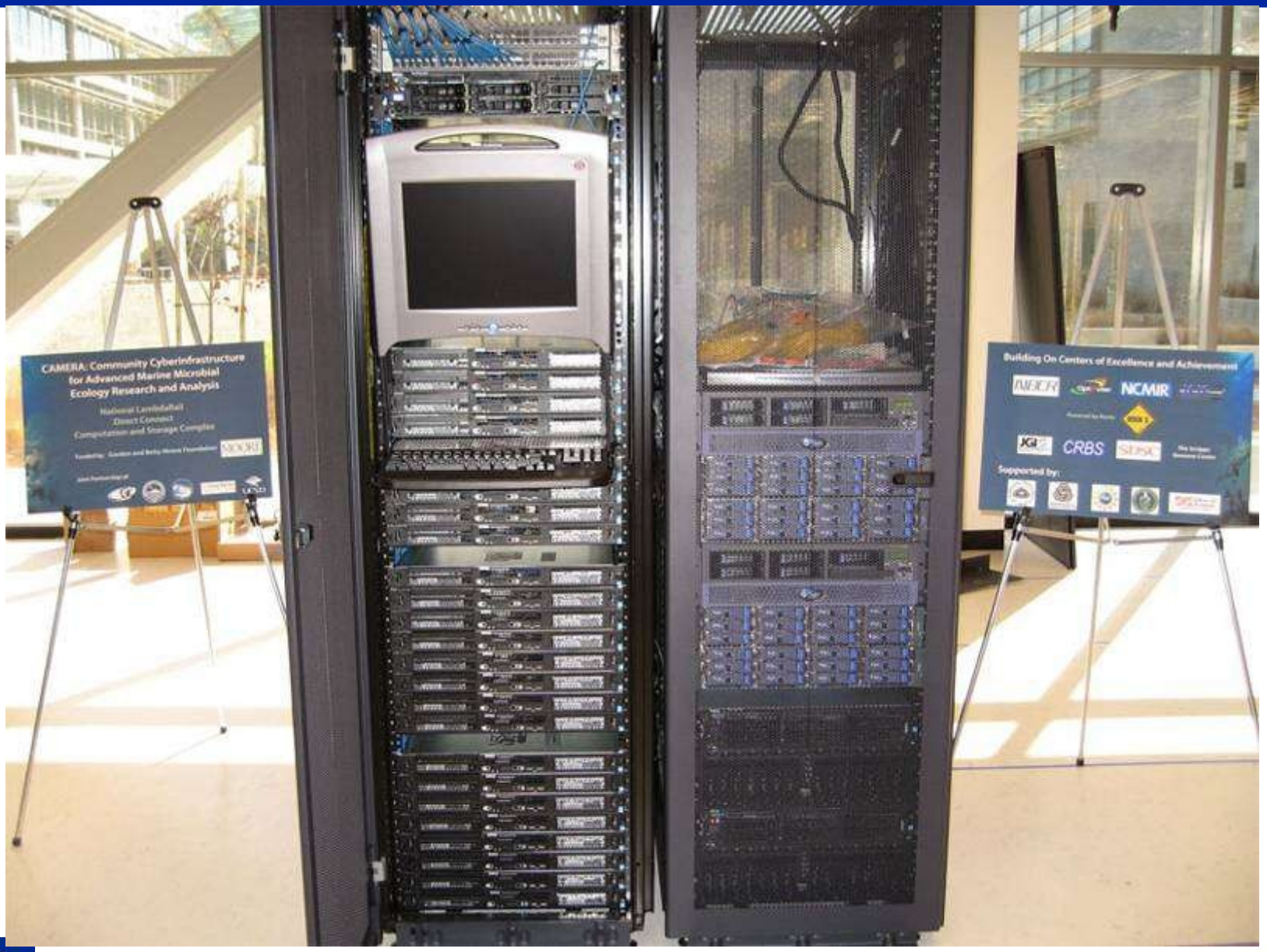
Gordon and Betty MOORE FOUNDATION

UCSD
UCIrvine

J. Craig Venter INSTITUTE

it²

# First Implementation of the CAMERA Complex– 1/10 of Final Scale




Compute

Database & Storage

UCSD
UCIrvine

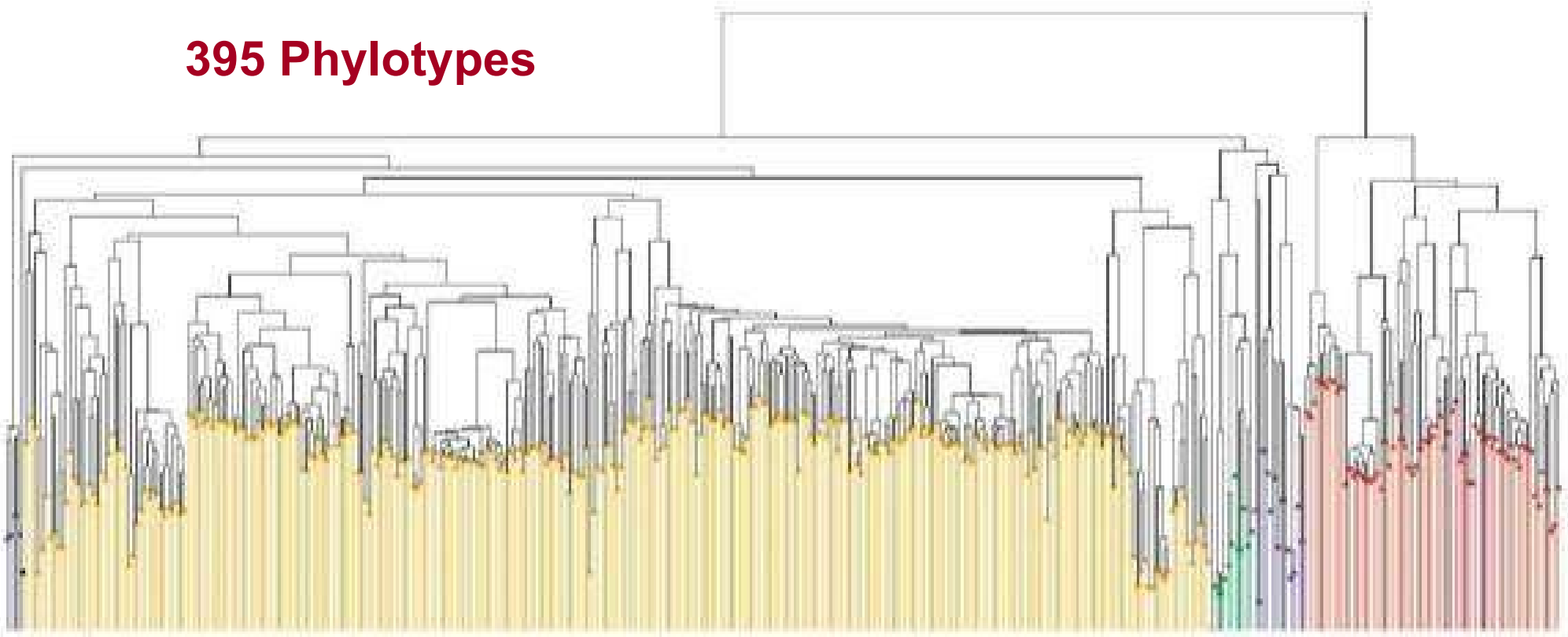
# Paul Gilna Has Just Been Recruited from Los Alamos to Become Executive Director of CAMERA

- **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

# Calit2 is Discussing Including Other Metagenomic Data Sets in CAMERA

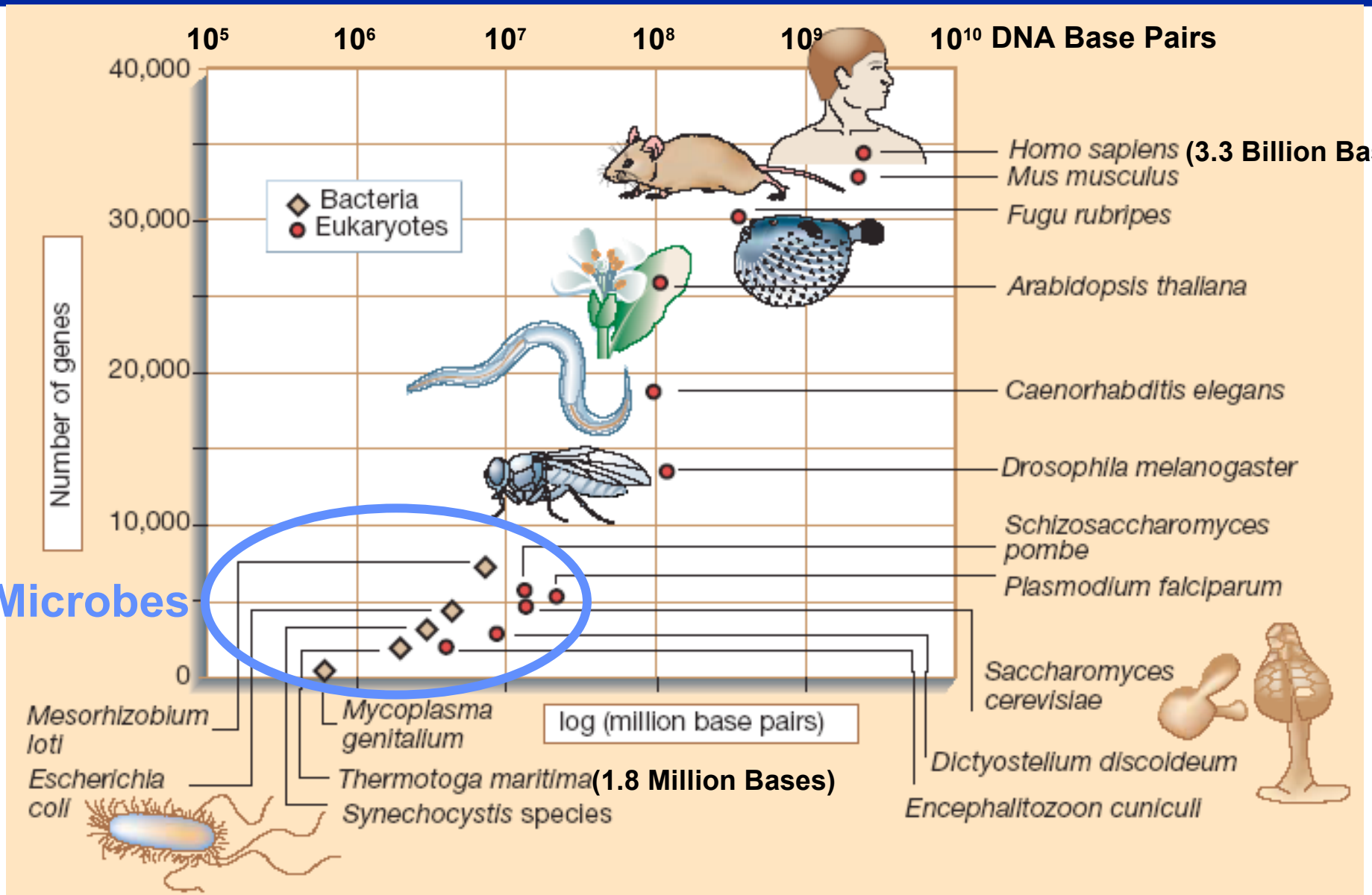
- "A majority of the bacterial sequences corresponded to uncultivated species and novel microorganisms."
- "We discovered significant inter-subject variability."
- "Characterization of this immensely diverse ecosystem is the first step in elucidating its role in health and disease."

**395 Phylotypes**



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

# The Human Genome Is Vastly More Complicated than Microbial Genomes




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

# From Microbial Genomes To Human Disease

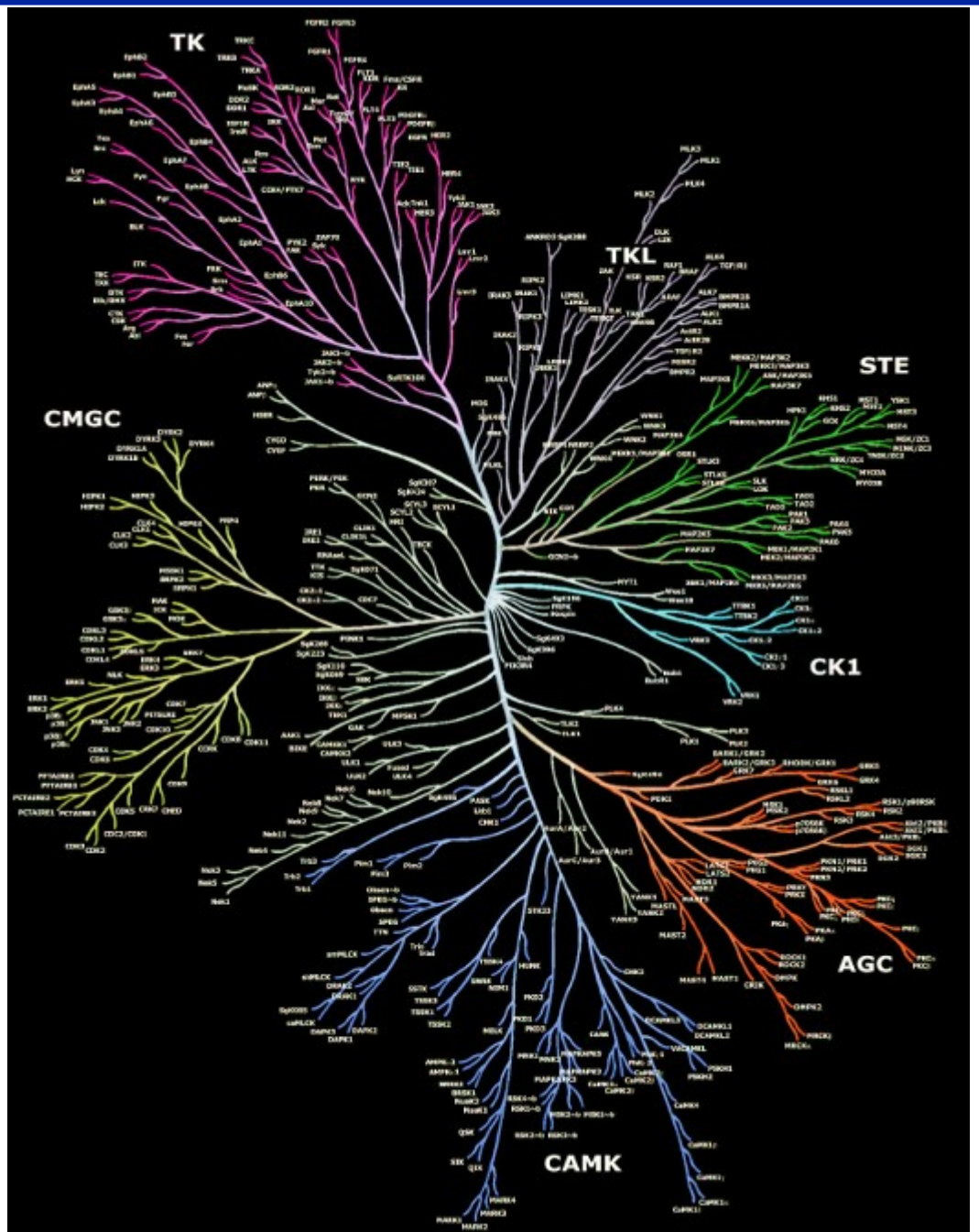
- **Microbes Have a Much Simpler Genome Than Humans**
  - **Human Genome ~ 1000x Longer than Microbial Genome**
- **However, Microbes Share Many of the Core Components of the Molecular Signaling Machinery Used by Humans**
- **Understand Both the Evolution and Regulation of Signaling Systems, First in Microbes and Then in Humans**
- **We Illustrate This Using the Protein Kinase Superfamily**
  - **A Very Large Family That is Implicated in Numerous Human Diseases**

UCSD
UCIrvine

**Source: Susan Taylor, SOM, UCSD**

it²

# The Human Kinome




**Over 500 Protein Kinases**
**2% of the Human Genome**

**Many splice variants**

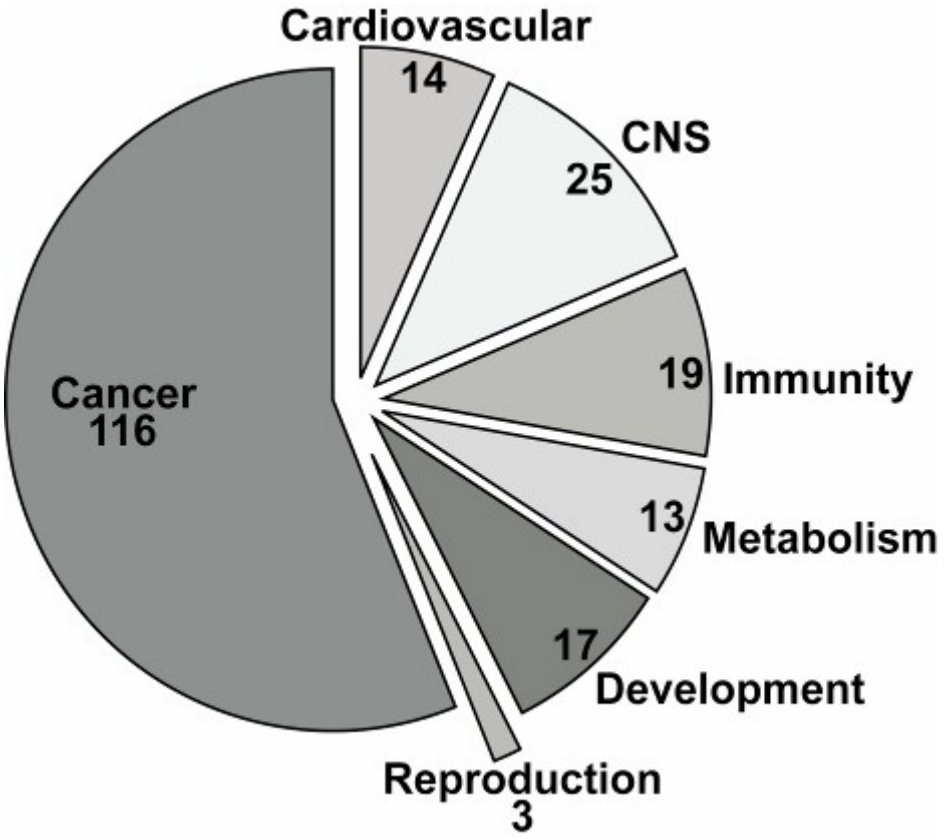
**Manning, et al (2002) *Science* 298:1912**

**Source: Susan Taylor, SOM, UCSD**

UCSD
UCIrvine

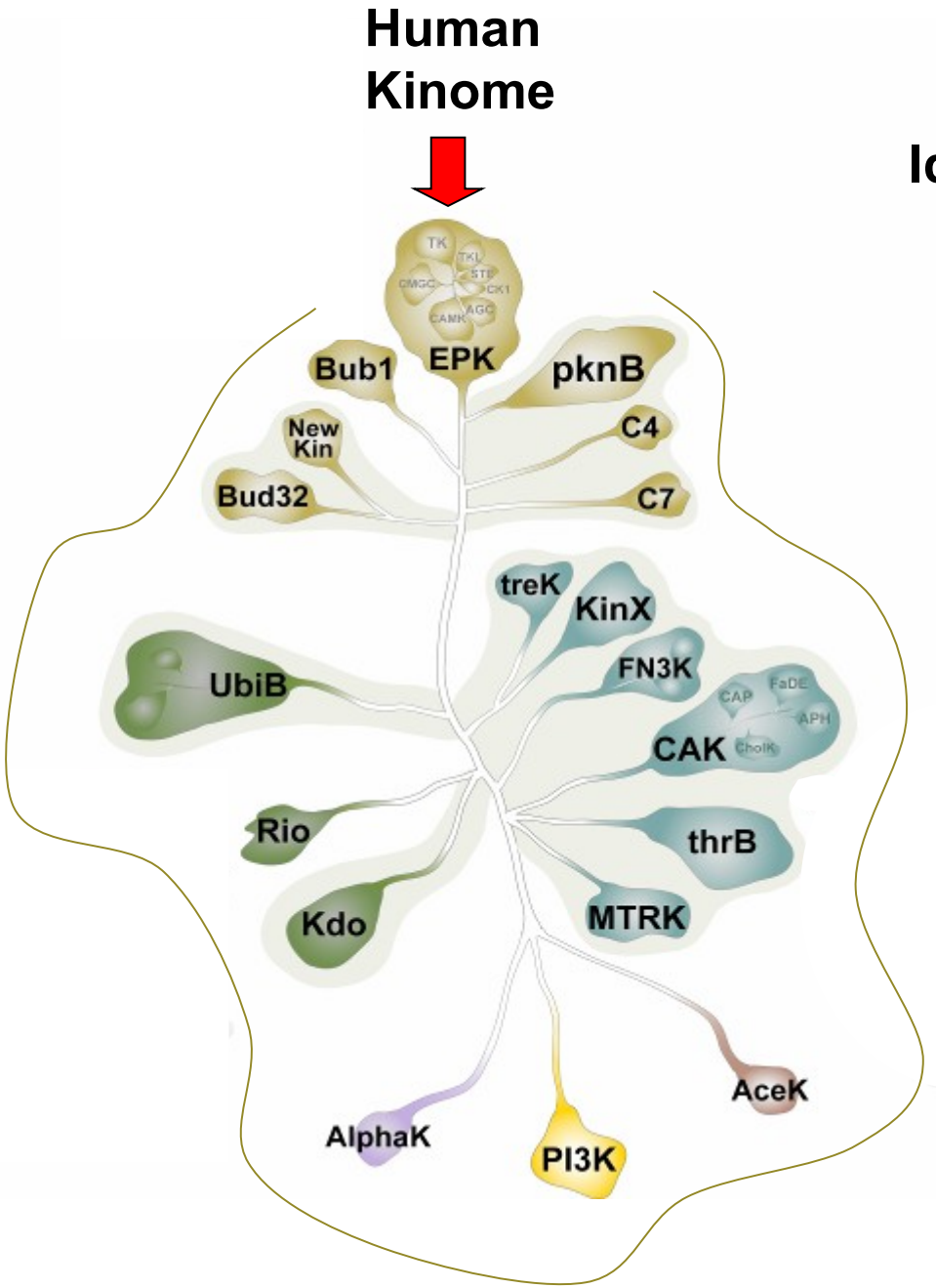
# Kinases and Diseases: Molecular Switches that Regulate Cell Function

- 30% Of Protein Kinases Published are Implicated in Various Diseases
- Many More are Likely to Follow, From Expression, SNP Analyses, Genetics and Functional Genomics
- Kinases are Tractable Drug Targets with Several Approved Drugs and Large Development Efforts




Source: Susan Taylor, SOM, UCSD

# The Human Kinome is a Small Part of the Kinome Tree Across All Living Creatures




**Identified 15,000 New Kinases In Venter Global Ocean Sampling Data**

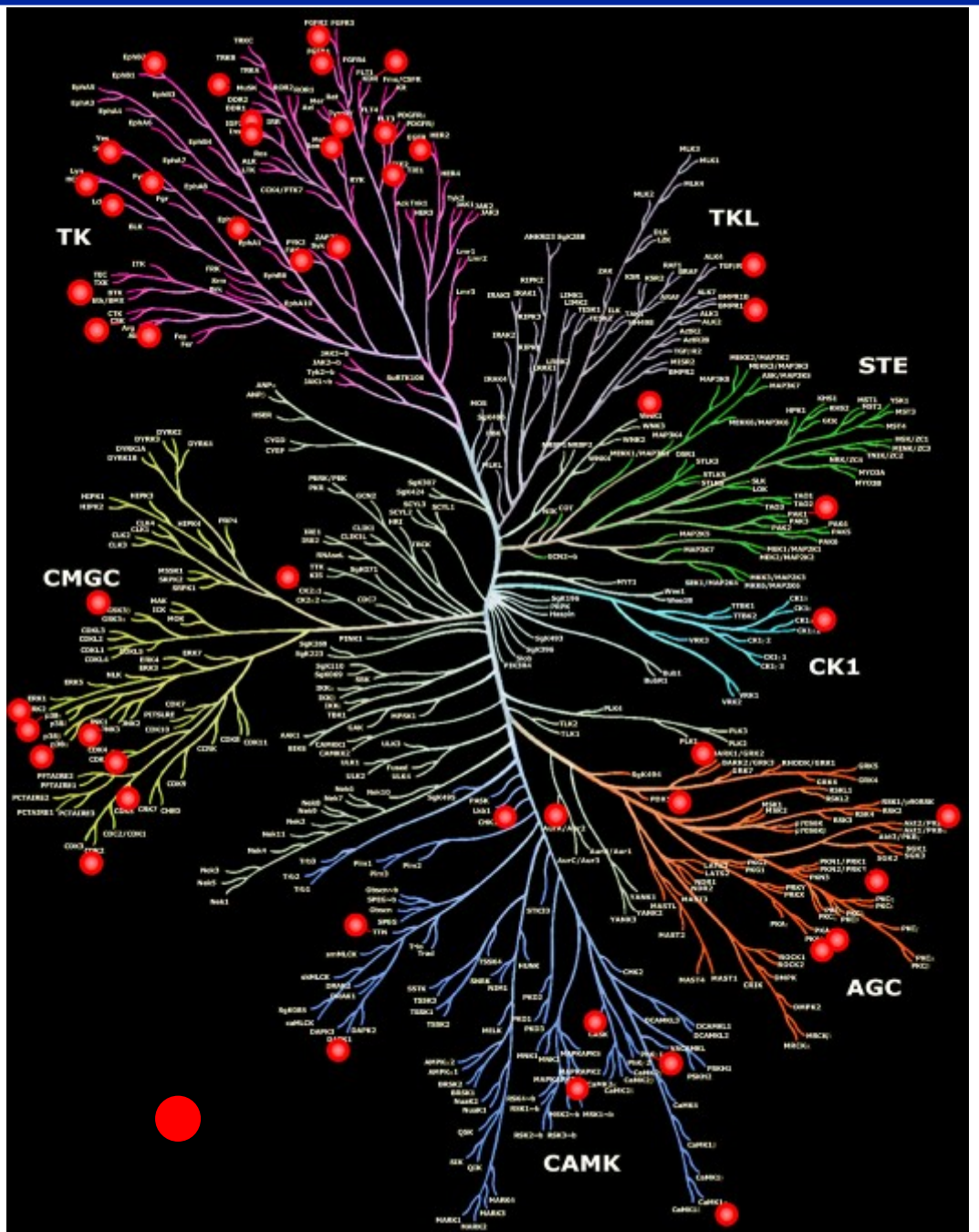
**Defines the Evolution of the Eukaryotic Protein Kinases**

**Source:**
**Susan Taylor,**
**SOM, UCSD**

UCSD
UCIrvine

# The Human Kinome: 3D Protein Structures

**Crystal Structures**

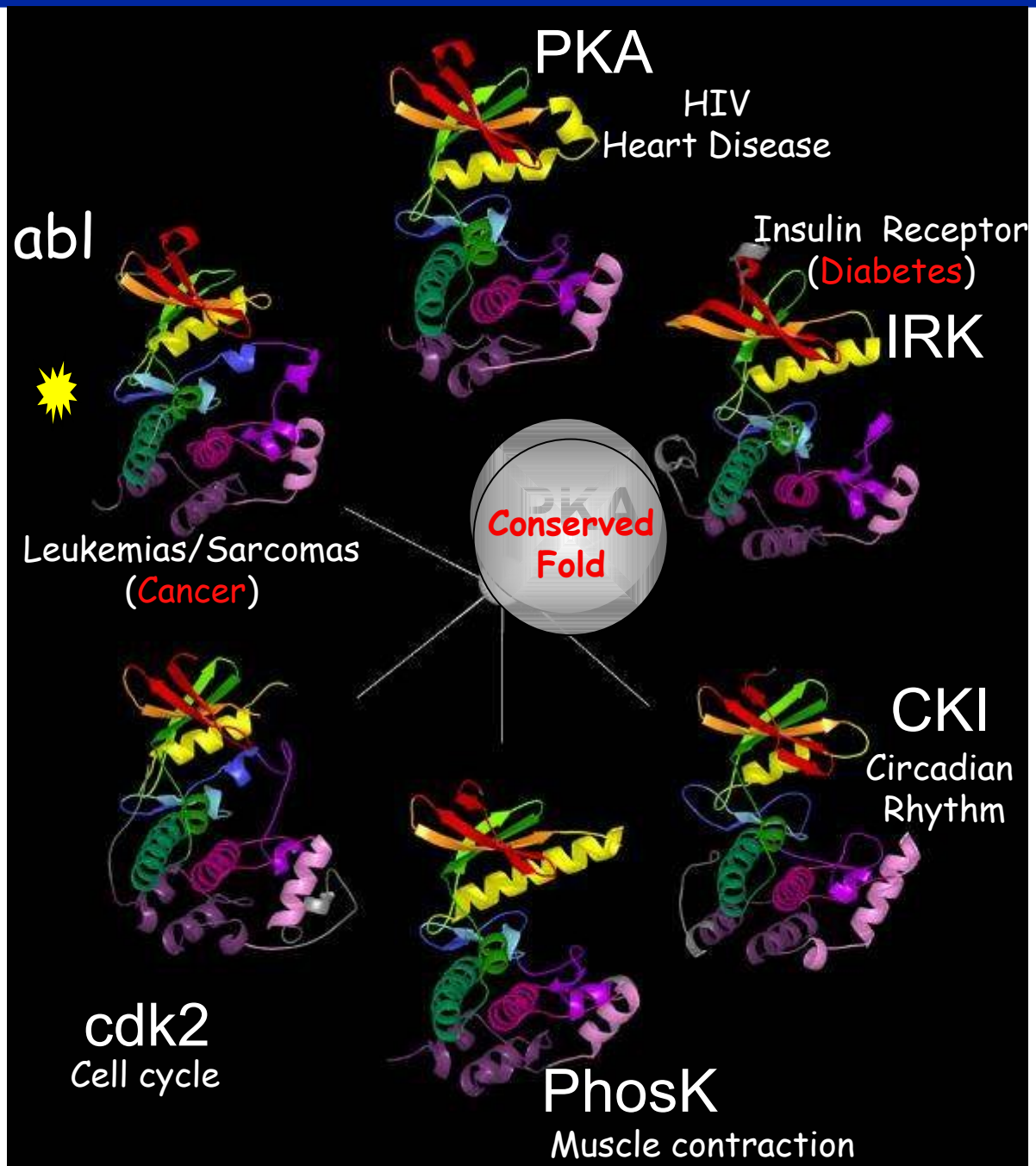



**Source: Susan Taylor, SOM, UCSD**

UCSD
UCIrvine

# 3D Kinase Protein Structures That are Implicated in Disease

**The Anti-Cancer Drug Gleevac Targets abl**

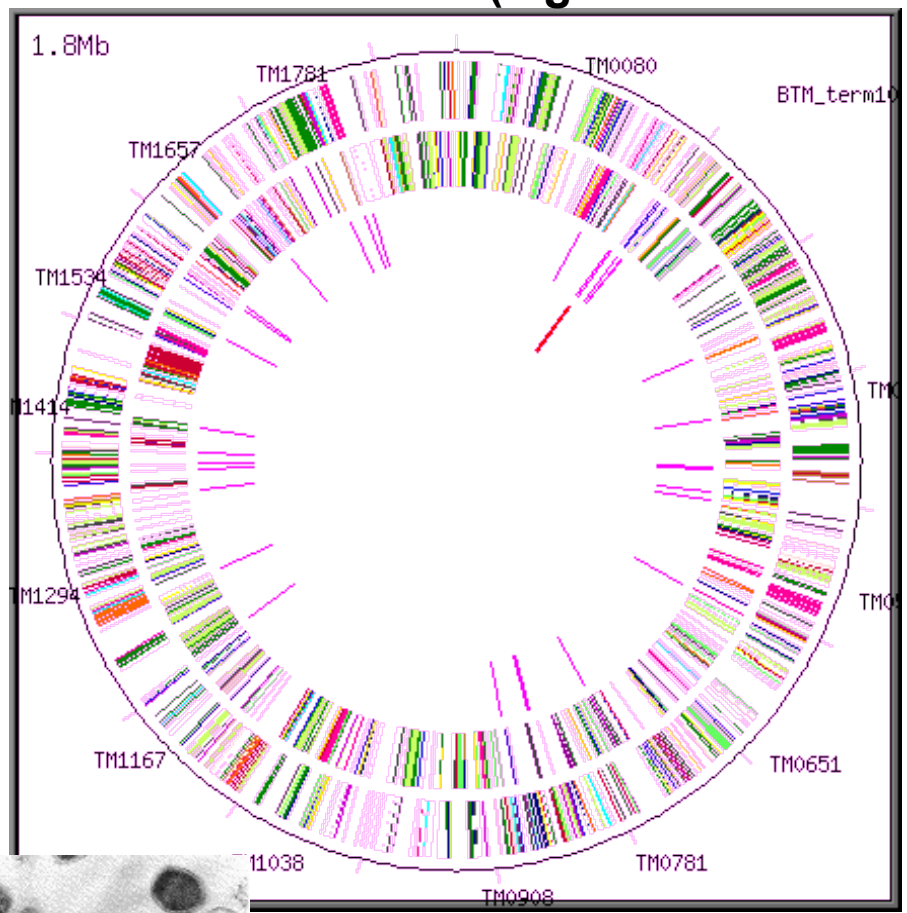



**Source: Susan Taylor, SOM, UCSD**

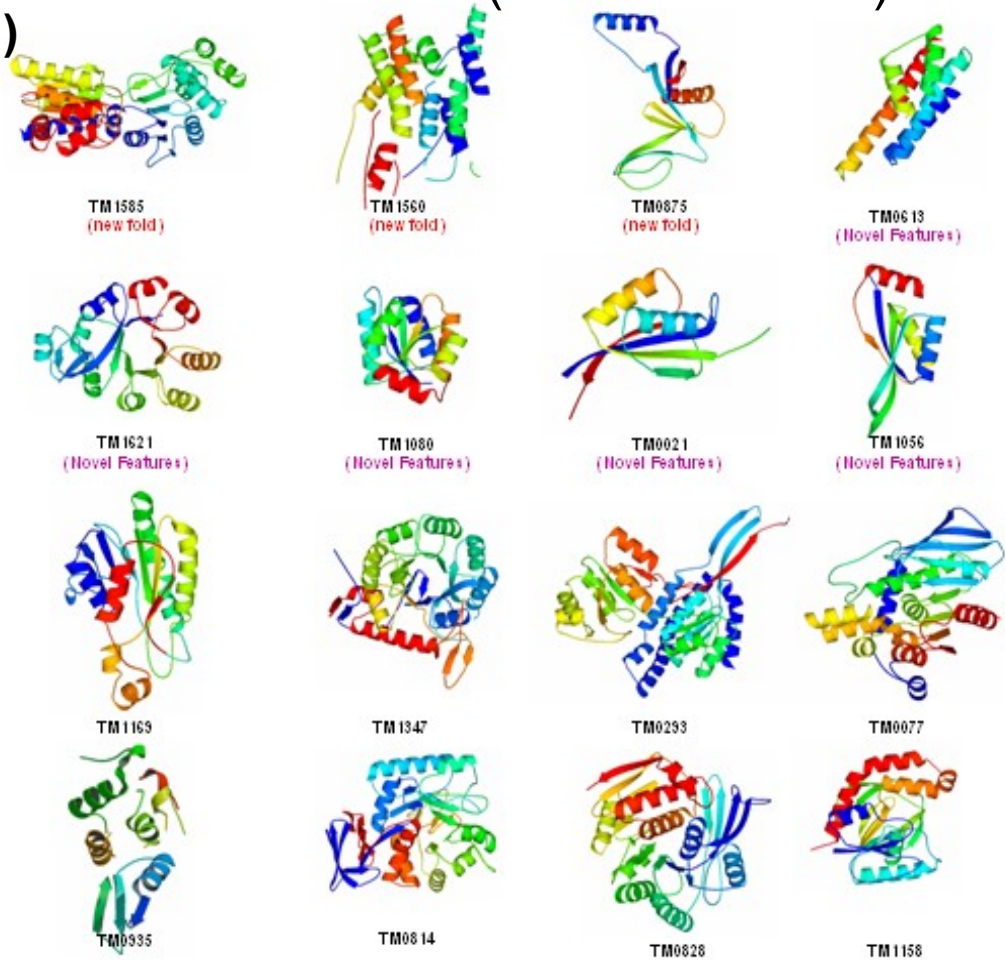
UCSD
UCIrvine

# 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)





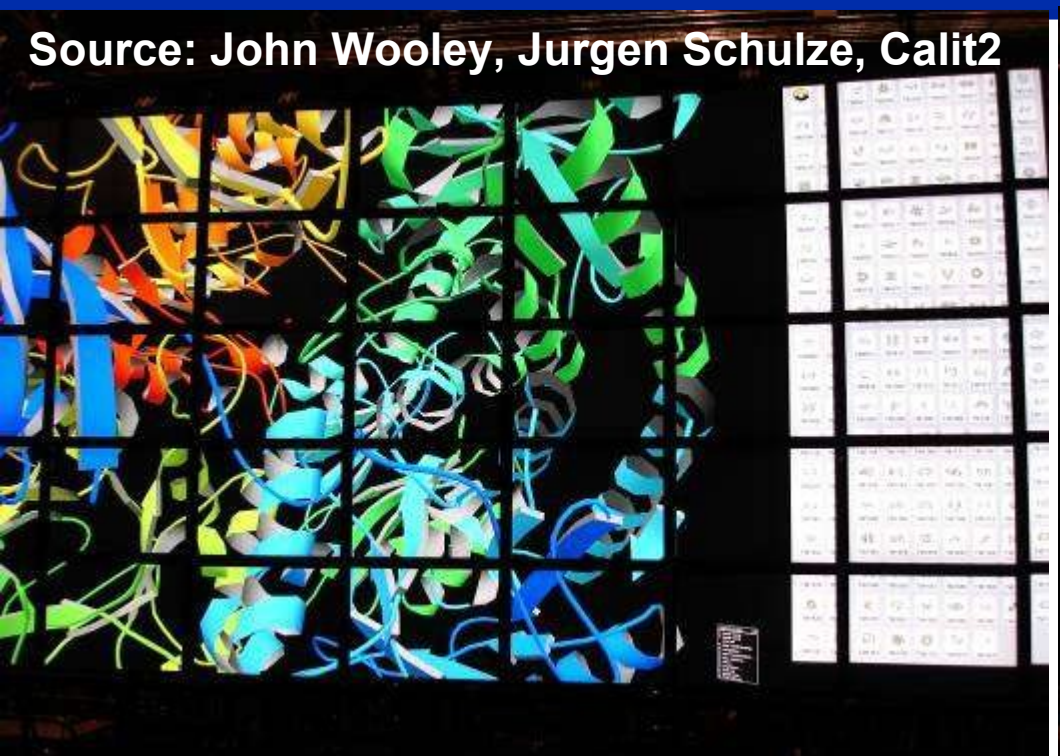

- Determining the Protein Structures of the Thermotoga Maritima Genome
- 122 T.M. Structures Solved by JCSG (75 Unique In The PDB)
- Direct Structural Coverage of 25% of the Expressed Soluble Proteins
- Probably Represents the Highest Structural Coverage of Any Organism

**Source: John Wooley, UCSD**

UCSD
UCIrvine

# Interactive Visualization of Thermatoga Proteins at Calit2

Source: John Wooley, Jurgen Schulze, Calit2

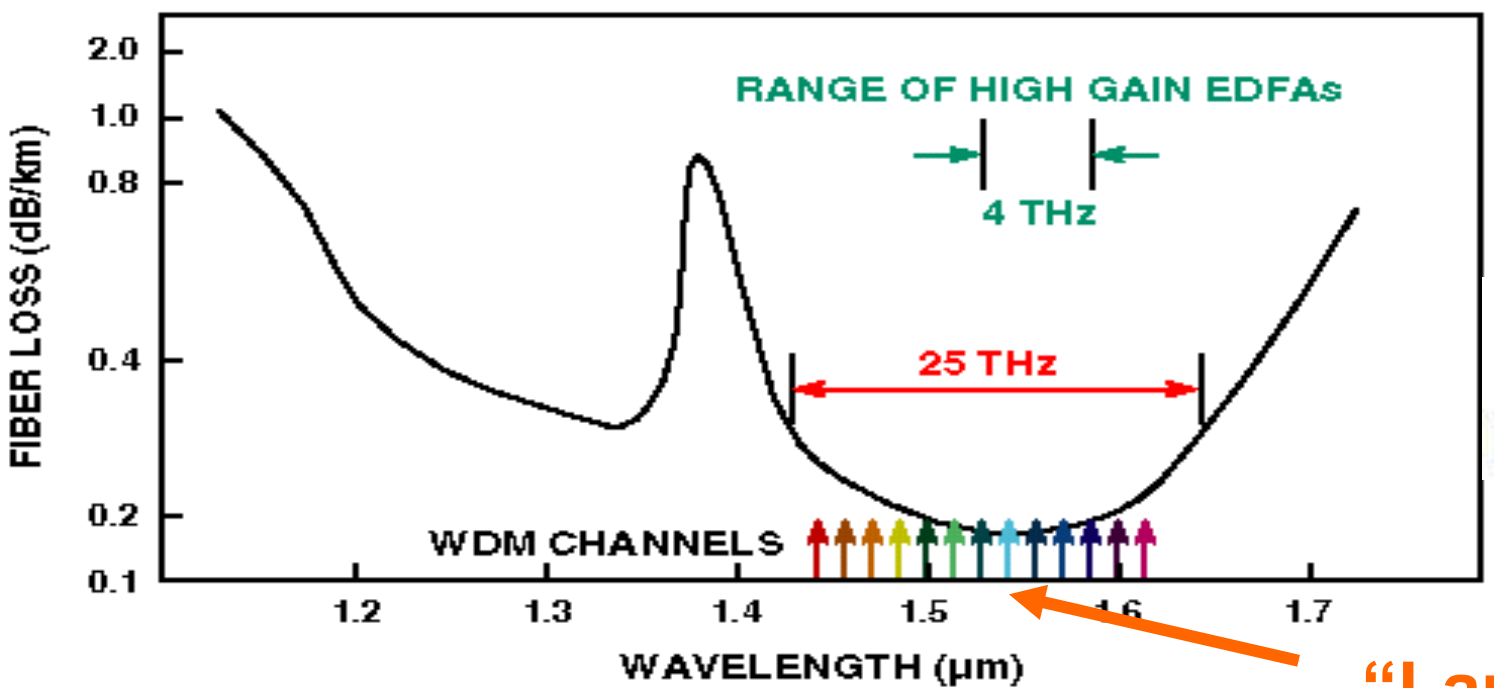




# End Users Can Direct Connect to CAMERA Using Lambdas-- Individual 1 or 10Gbps Dedicated Lightpaths

## WAVELENGTH DIVISION MULTIPLEXING (WDM)

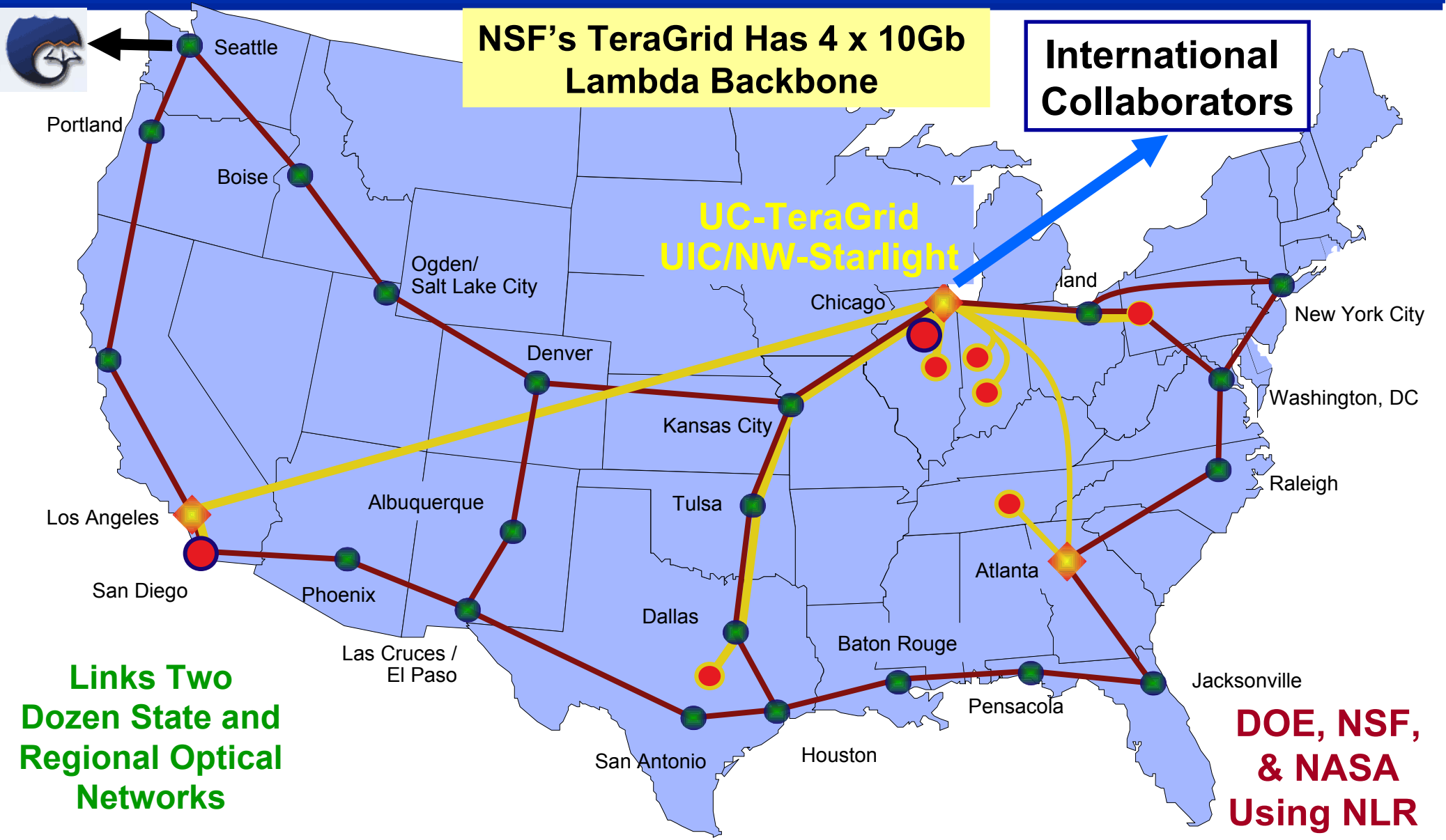
- EXPLOITS
  - ENORMOUS BANDWIDTH OF SILICA FIBER
  - HIGH-GAIN WIDEBAND OPTICAL AMPLIFIERS




"Lambdas"

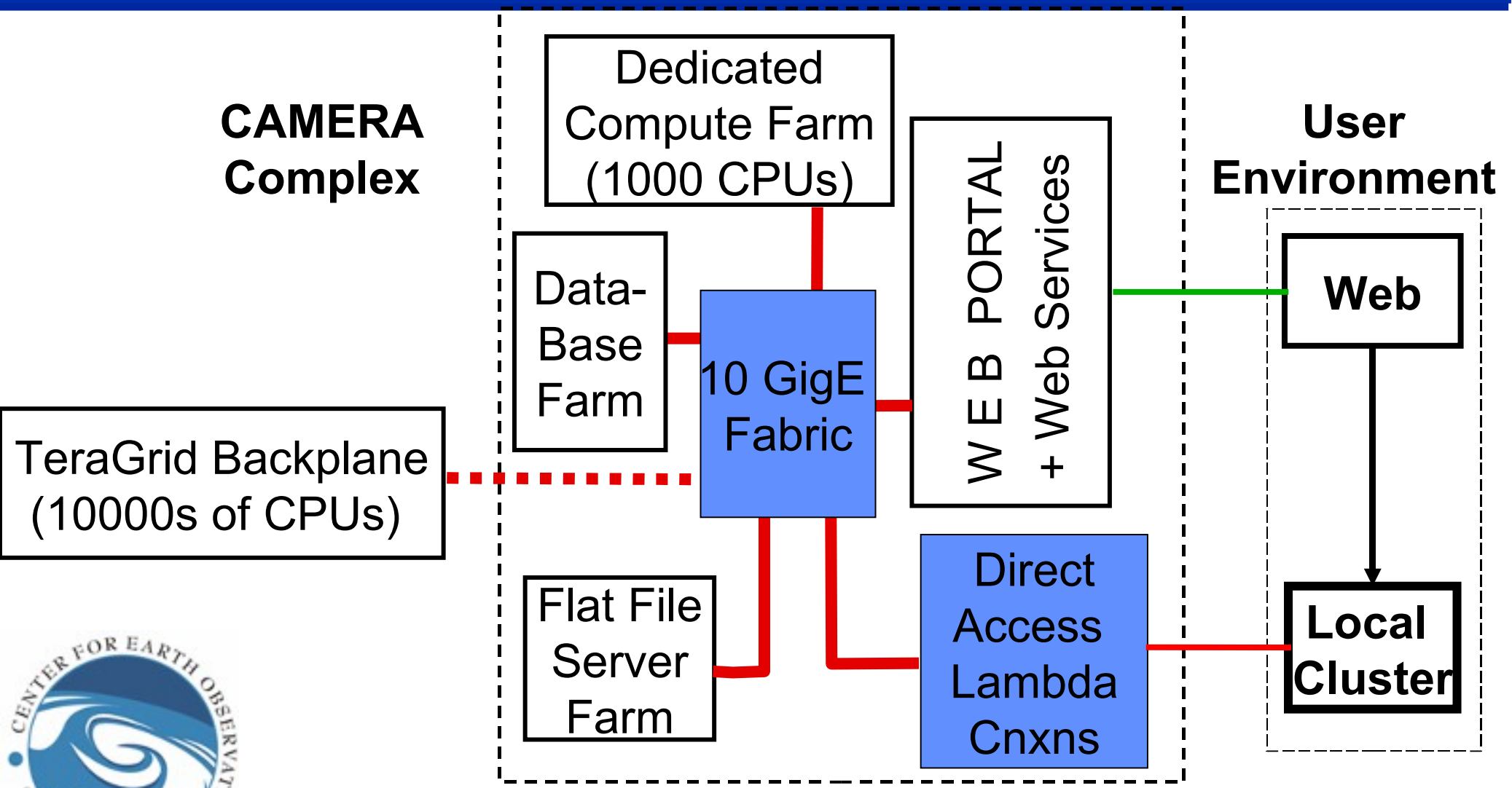
Source: Steve Wallach, Chiaro Networks

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



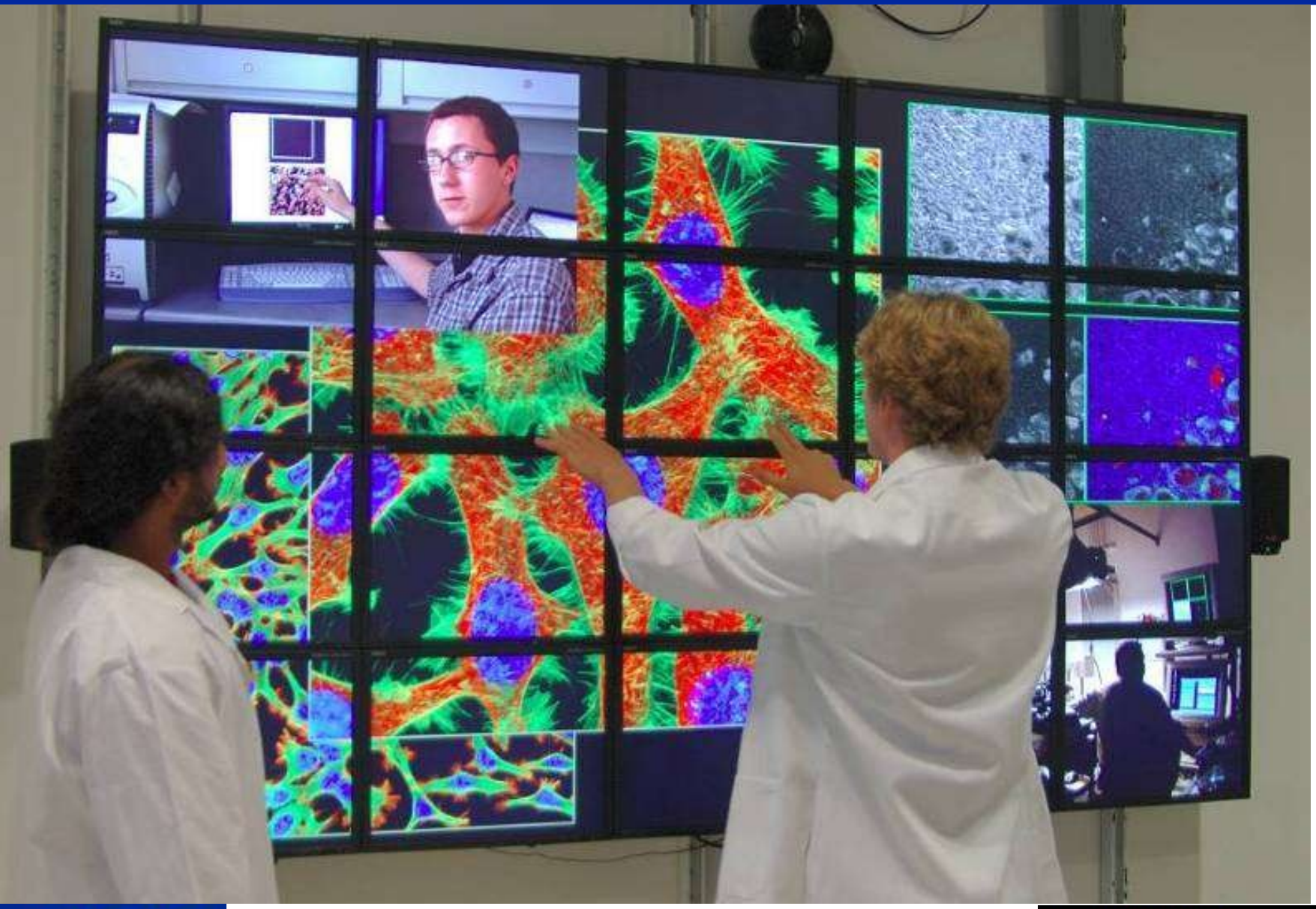

UCSD
UCIrvine

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




Source: Phil Papadopoulos, SDSC, Calit2

# Combining High Definition Video Streams with Large Scale Image Display Walls



**Large Scale Images of He-La Cancer Cells**

UCSD UCIrvine

evl electronic visualization laboratory

**Source: David Lee, NCMIR, UCSD**

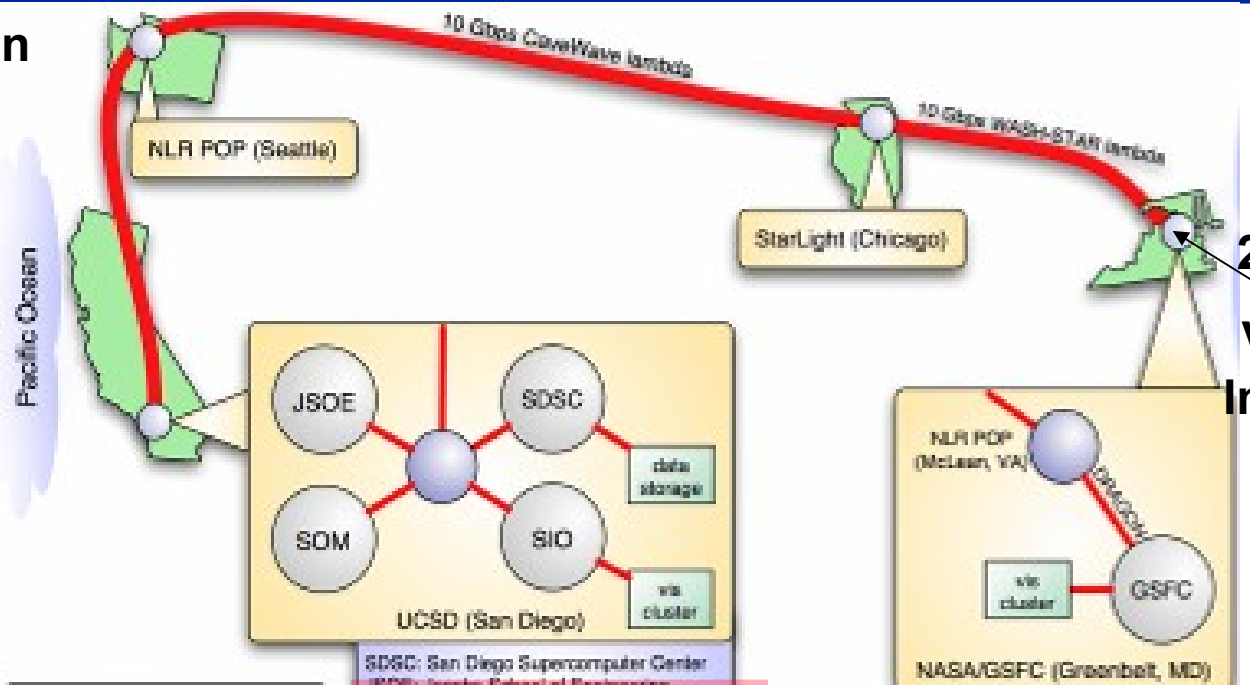
NCMIR NATIONAL CENTER for MICROSCOPY and IMAGING RESEARCH at San Diego, an NIH supported resource center

it2

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

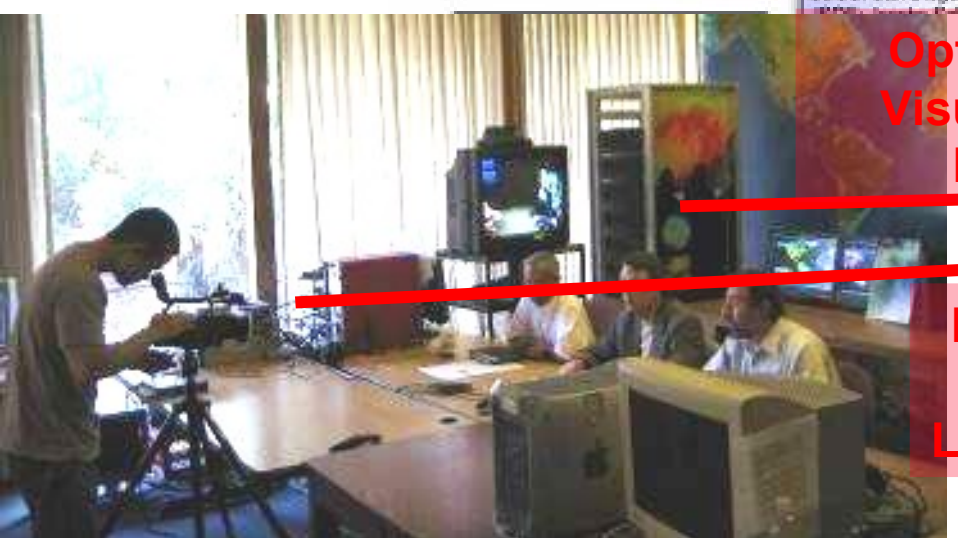
**Live Demonstration**

**of 21st Century National-Scale Team Science**




**25 Miles**

**Venter Institute**



**OptIPuter Visualized Data**

**HDTV Over Lambda**



UCSD

UCIrvine