UC San Diego NEWS RELEASE

Launching the Global Community Cyberinfrastructure for Advanced Marine Microbial Research and Analysis (CAMERA)

UC San Diego Makes Center Institute's Global Oean Sampling (GOS) Expedition Microbial Metagenomic Data and Computational Tools Available to Scientists Worldwide

March 13, 2007

Doug Ramsey

Scientists and engineers at the University of California, San Diego and the J. Craig Venter Institute (JCVI) have flipped the virtual switch on the first cyberinfrastructure customized to serve the marine microbial metagenomics community. At the heart of the cyberinfrastructure is a new, high-performance computer and storage complex funded by the Gordon and Betty Moore Foundation and located in UC San Diego's Atkinson Hall, headquarters of the California Institute for Telecommunications and Information Technology (Calit2), a partnership of UC San Diego and UC Irvine.

The computer complex enables analysis of a vast array of biocomplexity data housed in the Community Cyberinfrastructure for Advanced Marine Microbial Research and Analysis (CAMERA).

It includes environmental metagenomic and genomic sequence data, associated environmental parameters ("metadata"), precomputed search results, and cross-analysis of environmental samples. While end users can manipulate the data over the web or over dedicated optical circuits, CAMERA permits scientists to connect their local laboratory computers directly to the CAMERA database and tools using the National LambdaRail, Internet2's NewNet, or international optical circuits, resulting in up to a hundred-fold increase in bandwidth over the conventional shared Internet.

CAMERA has been in beta testing since January and today launched the first production version of its database and computational resources. Simultaneously, at a news conference held in Washington D.C., researchers announced the first scientific findings based on sequences and metadata deposited in CAMERA by JCVI's Global Ocean Sampling (GOS) Expedition. The findings are published as articles in the Oceanic Metagenomics collection in the March 2007 issue of PLoS Biology, including "CAMERA: A Community Resource for Metagenomics," a four-page introduction to the project by JCVI's Rekha Seshadri, Saul Kravitz and Marvin Frazier, with Calit2's Larry Smarr and Paul Gilna [PLoS Biology, March 2007 | Volume 5 | Issue 3 | e75].

[The studies detail the discovery of millions of new genes, thousands of new protein families and specifically the characterization of thousands of new protein kinases from ocean microbes using whole environment shotgun sequencing and new computational tools. For details, read today's companion release from JCVI at http://www.calit2.net/newsroom/release.php?id=1068.]

Calit2 director Larry Smarr is principal investigator on the CAMERA project

"A new cyberinfrastructure architecture is required to support the field of genomics as it transitions to the study of metagenomics," said CAMERA principal investigator Larry Smarr, a professor of computer science and engineering at UC San Diego and director of Calit2. "The infrastructure will create a virtual domain for global data and knowledge sharing by this emerging research community."

The Gordon and Betty Moore Foundation funded the CAMERA project in January 2006 with $24.5 million over seven years, building on its previous support for Venter's ambitious program to sequence marine microbes. "We asked Calit2 to join with us on this project because Larry Smarr and his National Science Foundation-funded OptIPuter team were already pioneering and prototyping infrastructure for large-scale, distributed scientific collaboration," said JCVI founder and chairman, Craig Venter. "CAMERA's database and computational tools are truly global resources, and they will be accelerating and broadening what the community learns from the GOS Expedition and future metagenomic research efforts.

UCSD alumnus Craig Venter (left) on board the Socrerer II duting the two-year odyssey around the world collecting samples every 200 miles.

"We are proud that UC San Diego can bring together the strengths of our world-renowned research units to create CAMERA's state-of-the-art cyberinfrastructure to support this important new scientific discipline," said the university's Chancellor, Marye Anne Fox. CAMERA is being developed by Calit2 at UC San Diego in collaboration with the JCVI, the university's Center for Earth Observations and Applications (anchored by the Scripps Institution of Oceanography), the San Diego Supercomputer Center, and the University of California, Davis.

The CAMERA database is different from most other genomic repositories because it was designed to accommodate environmental metadata as well as the sequence data derived from DNA samples. Even before the production version went online, a beta release of the database had been accessed by over 240 research scientists and students at more than 40 U.S. institutions, as well as users in at least ten foreign countries. Access to the CAMERA resources is free; users who register agree to abide by the terms of the Convention on Biological Diversity, in recognition of the many international sources of the data housed in CAMERA.

The sequence data from the GOS study have also been deposited in the National Institutes of Health-funded public database GenBank, however CAMERA is designed with metagenomic researchers in mind. "If a scientist queries our database for a particular set of sequence data, he or she would also get back all the metadata associated with each metagenomic sequence read," said Paul Gilna, executive director of CAMERA. "This is a very useful feature because the metadata could provide clues to understanding differences between microbial specimens, especially if you are comparing microbes that live in very different ocean environments. Our metadata, more generally, will serve to advance marine environmental biology and ecology research."

Cyberinfrastructure

CAMERA builds on the NSF funded OptIPuter research project, which is prototyping a global-scale end-to-end cyberinfrastructure backplane, stretching from a high-resolution visualization cluster in the researcher's lab, over dedicated one- or ten-gigabit per second lightpaths on optical fiber, to remote data and compute servers that may be located next door or thousands of miles away. Over the next few years, the dedicated but reconfigurable optical connectivity will provide metagenomics researchers with the freedom to work with data objects that are orders-of-magnitude larger than those transmitted over the conventional shared Internet.

The tiled LCD visualization displays, known as OptIPortals, scale from tens of millions to hundreds of million pixels. They provide the end user with the "pixel real estate" needed to explore the complexities of metagenomic data. Developed by the Electronic Visualization Laboratory at the University of Illinois at Chicago and Calit2, they are already installed at the JCVI and UC San Diego, and more are now being installed in partner metagenomics labs at the University of Washington, San Diego State University, MIT and elsewhere, including several international labs.

New 512-CPU computer cluster in Atkinson Hall contains the newly launched data

The CAMERA data resides on servers located at Calit2's headquarters on the UC San Diego campus, including a large production server consisting of a 512-CPU cluster (approximately 5 trillion floating-point operations per second, or "teraflops") with roughly 200 trillion bytes ("terabytes") of dedicated storage, all built on the SDSC Rocks (www.rocksclusters.org) cluster configuration software. A separate server at Calit2 hosts tools for analyses, tools for transferring large data sets and applications, and also a web-based interface for the user community. **Data and Tools**

In addition to data from the GOS Expedition, the CAMERA database includes metagenomic data from the Marine virome data collection from Forest Rohwer's group at San Diego State University, and the metagenomic data from the Hawaii Ocean Time Series Station ALOHA contributed by Ed DeLong's group at MIT. Today's release will also allow users to access or search 68 completed genomes from the 155 genomes included in the Moore Microbial Sequencing project being conducted at the JCVI. Large reference collections of relevant sequence data are available for search, including non-identical amino acids, microbial, viral, and fungal sequence and peptides, as well as sequences and peptides from microbial eukaryotes. "Since CAMERA is designed to be a community-driven resource, we will routinely introduce additional data sets based on input and priorities set by the metagenomics community," said Marv Fazier, VP for Research at JCVI and co-principal investigator on the CAMERA project. "Our plan is that in year two, the additional data will include further phases of GOS, additional members of the Moore 155 genome sets as they become available, the Department of Energy's Joint Genome Institute [JGI] metagenomic data sets, as well as the data from projects deposited at the National Institutes of Health's National Center for Biotechnology Information [NCBI]."

Researchers working on pre-release versions of the GOS data used initial analysis tools developed at the JCVI, including a number of variants of BLAST nucleotide and amino acid sequence search tools that were allied to a metadata export capability. Tools in the pipeline will allow users to upload metagenomic data and their metadata, do metagenomic and whole microbial genome annotation, and conduct phylogenetic analyses.

Building the User Community

Through its new user software portal, the CAMERA website (http://camera.calit2.net) allows the user community to learn about the project and access applications and data sets. Meanwhile, leaders of the project have attracted a global set of metagenomics researchers by organizing an annual international metagenomics conference series at Calit2 (www.calit2.net/metagenomics2007/). CAMERA has also attracted leaders from the community to form an external Scientific Advisory Board.

With today's initial release of CAMERA capabilities, says principal investigator Smarr, "we have arrived at the beginning. The CAMERA team has built the foundation for an innovative and powerful cyberinfrastructure supporting biological research globally, with plans to enhance its capabilities and build the scientific community it serves. This infrastructure will now serve as a template or model for other large scientific communities requiring such an extensive level of computing, networking, data sharing, and collaboration."

Users interested in exploring and using the CAMERA resource should go to http://camera.calit2.net and follow the instructions for registration. Registration is free and open to all interested in using CAMERA in their research.

Media Contacts: Doug Ramsey, UCSD, 858-822-5825 and Heather Kowlaski Venter Institute

camera
Marine Microbial Ecology