



## CAMERA 1.0 Backgrounder

Genomic DNA sequence data from samples generated from the Sorcerer II Expedition are publicly available to researchers worldwide through a new database for metagenomic data, CAMERA (Community Cyberinfrastructure for Advanced Marine Microbial Ecology Researcher and Analysis). CAMERA, funded by a grant of \$24.5 million over seven years from the Gordon and Betty Moore Foundation, is a state-of-the-art computational resource with software tools to decipher the genetic code of communities of microbial life in the world's oceans.

The new resource will help scientists understand how microbes function in their natural ecosystems, enable studies on the effect humans are having on the environment, as well as permit insight into the evolution of life on Earth. This invaluable new resource has been developed by UC San Diego Division of the California Institute for Telecommunications and Information Technology (Calit2) who will lead the project in partnership with J. Craig Venter Institute (JCVI) and UCSD's Center for Earth Observations and Applications (CEOA) at Scripps Institution of Oceanography.

### Overview of Capabilities

The initial release of CAMERA provides the research community with unprecedented capabilities to access and search metagenomic datasets, while addressing CBD-related issues of concern to the international community. The capabilities include:

- Access to the datasets of GOS and other metagenomic projects
- Browsing of sample and sequence metadata for all samples
- Metadata-driven download of metagenomics datasets
- Ability to quickly perform large searches against metagenomic and genomic datasets on CAMERA's dedicated cyberinfrastructure
- Integration of sample and sequence metadata with search results

### Data Content

The initial release of CAMERA includes data from 3 projects:

- **JCVI Global Ocean Sampling Expedition:** Although all of the GOS data that was amenable to Genbank submission has been provided to Genbank, the metadata-driven access available with CAMERA will make it more accessible to metagenomics researchers.
- **SDSU Marine Virome data set:** This large-scale metagenomic survey of marine viral organisms collected from sites around the North American continent has been collected by CAMERA SAB member Forest Rohwer and his research team at San Diego State University.
- **MIT C-More's Hawaii Ocean Time Series data set:** This vertical profile of marine microbial communities collected at the Hawaii Ocean Time-series (HOTS) station ALOHA has been collected by SAB member Ed DeLong and his research team at MIT.

In addition, CAMERA leverages the following datasets assembled from the National Institutes of Health public database, Genbank and other public sources using JCVI's PANDA system

- Non-identical amino acid database, with associated nucleotides
- Finished and unfinished microbial genomes including the publicly available genomes from the Moore Microbial Sequencing Project
- Collections of Viral, Fungal, and microbial eukaryote genomes and peptides

## Functionality

- **BLAST Search:** A new, high-performance, parallel implementation of NCBI's BLAST search tool is deployed on CAMERA's high-performance cyberinfrastructure allows users to quickly perform large sequence-based searches against genomic and metagenomic data collections.
- **Project/Publication and Sample Frameworks:** This unique framework for distribution of content based on projects and papers should dramatically reduce the effort required by researchers to reproduce or leverage published results. CAMERA staff have worked with investigators to provide rich descriptions of their research, and access to supporting data sets typically not directly available to the research community.

## Future Directions

The initial release of CAMERA is just a start. Future releases of CAMERA will dramatically expand its capabilities, including the following

- **Fragment Recruitment:** an interactive fragment recruitment viewer, integrated with sample metadata, allowing researchers to graphically explore the relationship between metagenomic and genomic data.
- **Protein Clusters:** interactive tools for searching and navigating the content and annotations of the protein clusters.
- **Sample Annotation:** sample-level annotation of the proteins and organisms represented within a sample or collection of samples.
- **Integration of Remote Sensing Data:** Remote sensing metadata will be associated to metagenomic sequence and made accessible to researchers.

The screenshot shows the CAMERA website interface in a Mozilla Firefox browser. The page is titled "Publications" and displays a list of publications. The selected publication is "The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific" by Douglas B. Rusch, Aaron L. Halpern, Oranger Sutton, Karla B. Heidelberg et al. The page includes a "Details" tab, a "Downloads" section with 4 items, and a "More Information" section with links to download the publication, project samples, and view the project website. The main content area features a map of the expedition route, a "Sampling Sites" section, and a "Primary Dataset" section. The "Primary Dataset" section describes a metagenomic dataset generated from the Sorcerer II Global Ocean Sampling Expedition, providing details on the number of sequences, read lengths, and taxonomic content. It also mentions the use of the CAMERA assembly pipeline and the availability of sequence reads for download and search.



CAMERA's BLAST search output displays search results with geographic context. The flags on the map represent sites from which search results are derived.

The screenshot displays the CAMERA web interface in a Mozilla Firefox browser window. The page is titled "Job Results - Mozilla Firefox" and shows the following sections:

- Read Detail:**
  - Read ID: JCVL\_READ\_1093016180716
  - Seq Direction: reverse
  - Template Acc: JCVL\_TMP\_L1061002524866
  - Library Acc: JCVL\_LIB\_OS-19-01-01-1P3-1P6kB
  - Seq Length: 1076
  - Seq Type: DNA
  - Sequence: GGCAGGGTTGGCAGATCGGCTTTTTGAGATTAGCATCTATCGTAGGTAGCAACCCCTTTGTACGGACCATGTAGCAAGTGTAGCCCTGGTAAAGGGCCA TGATGACTGAGCTCGTCCCGGCTCCCGCAGTTTCTCCTGCGCATCTCTAAAGTCCCGGTTAACCGCTGGCAATAAGAAAAGGGTTCGCTCCTT GCGGACCTTAAAGCAATCTCAGCAGCGAGCTGACGACAGCAATGCGACCTGTATTAAGTCCCGAAGCAATTCATCTCTGGAAAGTTCTTACTAT
- Intellectual Property Notice:**

This genetic information downloaded from <http://camera.calit2.net> may be considered to be part of the genetic patrimony of Panama, the country from which the sample was obtained. Users of this information agree to: (1) acknowledge Panama as the country of origin in any publications where the genetic information is presented and (2) contact the CEO local point identified on the CEO website (<http://www.bioinfo.org/cebio-centre/cebio-centre>) if they intend to use the genetic information for commercial purposes.
- Sample Metadata:**
  - Sample Name: G8019
  - Sample Acc: JCVL\_SMP\_L1103283000025
  - Experiment ID: JCVL\_SITE\_G8019
  - Fiber Size: 0.1 - 0.8 μm
- Site Metadata:**
  - Site ID: JCVL\_SITE\_G8019
  - Project: CAM\_PROJ\_G08
  - Leq: 3rd
  - Data Time: 2004-01-12
  - Habitat Type: Coastal
  - Geographic Location: Caribbean Sea
  - Sample Location: Northeast of Colon
  - Country: Panama
  - Latitude: 10d 42'59"n
  - Longitude: 80d 15'18"w
  - Sample Depth: 1.7 m
  - Water Depth: 3326 m
- Sequence Geography:**

A map of Central America showing the location of the sample site, JCVL\_SITE\_G8019, marked with a red pin. The map includes labels for Mexico, Central America, and the Caribbean Sea. A tooltip for the site reads "JCVL\_SITE\_G8019 Northeast of Colon".