



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, browse the project, and view the project website. The main content area features a map of the expedition route, a "Sampling Sites" section, and a detailed abstract. The abstract describes the expedition, the metagenomic dataset generated, and the taxonomic content and intra-ribotype diversity. It also mentions the use of the CAMERA assembly pipeline and the availability of sequence reads for download and search.

CAMERA's project- and publication-based organization of content allows researchers to explore metagenomic data with additional context provided by publication authors.

Global Ocean Sampling Expedition

Principal Investigator: J. Craig Venter
Funded By: Moore Foundation, DOE, Venter Science Foundation
Organization: J. Craig Venter Institute

The broad objective of the global ocean sampling expedition is to expand our understanding of the microbial world by studying the gene complement of marine microbial communities. Marine microbes influence the cycling of carbon (and other elements) in the world's oceans, acting as a biological conduit that transports carbon dioxide from the surface to the deep oceanic realms. By sequestering carbon from the atmosphere, marine microorganisms (eukaryotes, prokaryotes and viruses) may significantly affect global climate. How they do so, however, is poorly understood, and our attempts to study their activities are limited by our inability to culture the vast majority of them.

One avenue of exploration is to sequence the genomes of marine microbes using a metagenomics approach. In Spring of 2002, the J. Craig Venter Institute conducted a whole environment shotgun sequencing project to study marine microorganisms in the nutrient-poor Sargasso Sea near Bermuda. This study revealed an unforeseen breadth and depth of microbial diversity - about 1,800 different microbial species encoding over 1.2 million genes were discovered, nearly doubling the number of prokaryotic genes available in public databases. Notably, this study expanded our knowledge of ocean photobiology and nutrient pools. Results from the pilot study were reported in *Science* in 2004.

This pilot study served as the springboard for launching a more comprehensive survey of the bacterial, archaeal and viral diversity of the world's oceans. A global circumnavigation aboard the *Sorcerer's* sailing yacht began in August 2003, starting in Halifax, Canada and samples were collected at sites along the U.S. east coast, Gulf of Mexico, Galapagos Islands, central and south Pacific Oceans, Australia, Indian Ocean, South Africa, across the Atlantic back to the U.S., and was completed in January 2008. An initial analysis of the microbial data from the first leg of the trip - Halifax to the Galapagos Islands - was reported in *PLoS Biology* in January 2008. Shotgun sequencing and deep sequencing of 16S and 18S rRNA continues on the samples from the Pacific and Indian Ocean during 2007.

These studies have produced the largest catalogue of genes to date from thousands of new species, with no apparent slowing of the rate of discovery (i.e., attaining saturation of data). These data have potentially far-reaching implications for biological energy production, bioremediation, and creating solutions for reduction/management of greenhouse gas levels in our biosphere. For more information, see the J. Craig Venter Institute's [Global Ocean Sampling Project](#).

CAMERA provides background on each project, linking research results and datasets for ease of use by investigators.

Matching Sequences

Eval	Len. A	Query	Read	Sample(s)	Location(s)
1.897E-121	248	b3851 Escherichia c.	JCVI_READ_92679	OS000a	Sargasso Station 11, Sargasso Sea
7.067E-118	250	b3851 Escherichia c.	JCVI_READ_564916	OS000a	Sargasso Station 11, Sargasso Sea
7.067E-118	250	b3851 Escherichia c.	JCVI_READ_563590	OS000a	Sargasso Station 11, Sargasso Sea
7.067E-118	250	b3851 Escherichia c.	JCVI_READ_553385	OS000a	Sargasso Station 11, Sargasso Sea
7.067E-118	250	b3851 Escherichia c.	JCVI_READ_549535	OS000a	Sargasso Station 11, Sargasso Sea
7.067E-118	250	b3851 Escherichia c.	JCVI_READ_542109	OS000a	Sargasso Station 11, Sargasso Sea
7.067E-118	250	b3851 Escherichia c.	JCVI_READ_535938	OS000a	Sargasso Station 11, Sargasso Sea
7.067E-118	250	b3851 Escherichia c.	JCVI_READ_528385	OS000a	Sargasso Station 11, Sargasso Sea
7.067E-118	250	b3851 Escherichia c.	JCVI_READ_526728	OS000a	Sargasso Station 11, Sargasso Sea
7.067E-118	250	b3851 Escherichia c.	JCVI_READ_518736	OS000a	Sargasso Station 11, Sargasso Sea

Sequence Alignment

Sequence: JCVI_READ_92679 Score: 444.64 Identifiers: 242/248 (97%)
 Sequence Length: 1002 Expect: 1.85687e-121 Positives: 0/248 (0%)
 Alignment Length: 248 Query Begin/End: 1295-1542 (Plus) Query Gaps: 0
 Clear Range: 42-897 Subject Begin/End: 100-347 (Plus) Subject Gaps: 0

Query: 1295 TCCTAGTCCGGATTGGAGTCTGCAACTCGACTCCAGTGAAGTCGGAAATCGCTAGTAATCGCTGGACAGAAATGCCAGCGTGA 1374
 Sbjct: 100 TCCTAGTCCGGATTGGAGTCTGCAACTCGACTCCAGTGAAGTCGGAAATCGCTAGTAATCGCTGGACAGAAATGCCAGCGTGA 179
 Query: 1375 ATACGTTCCCGGGCTGTACACACCCCGCTCAGCAAGTGGAGTGGTTCGCAAGAGTGTAGTAACTTCGCG 1454
 Sbjct: 180 ATACGTTCCCGGGCTGTACACACCCCGCTCAGCAAGTGGAGTGGTTCGCAAGAGTGTAGTAACTTCGCG 259
 Query: 1455 GAGGGCCCTTACGACTTTGTATTTCATGCTGGGGTGAAGTCTGTAACAAAGTAACTAGGGGAACTTCGGTGGATCA 1534
 Sbjct: 260 GAGGGCCCTTACGACTTTGTATTTCATGCTGGGGTGAAGTCTGTAACAAAGTAACTAGGGGAACTTCGGTGGATCA 339
 Query: 1535 CCTCCCTTA 1542
 Sbjct: 340 CCTCCCTTA 347

Sequence Geography

30 sample sites are represented in this data set. Each sample site is marked on the map. Click a site marker for more information. Drag the map with your mouse, or double-click to recenter.

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 shows a web browser window displaying the CAMERA research portal. The page is titled "Job Results - Mozilla Firefox" and features the CAMERA logo and navigation tabs for HOME, ABOUT CAMERA, METAGENOMICS, RESEARCH, NEWS, EVENTS, and DISCUSSION FORUMS. The main content area is divided into several sections:

- Read Detail:**
 - Read ID: JCVL_READ_1093016180716
 - Seq Direction: reverse
 - Template Acc: JCVL_TEMPL_1061002524866
 - Library Acc: JCVL_LIB_OS-19-01-01-1P3-1P6kB
 - Seq Length: 1076
 - Seq Type: DNA
 - Sequence: `GGCAGGGTTGGCAGATCGGCTTTTTGAGATTAGCATCTATCGTAGGTAGCAACCCCTTTGTACGGACCATGTAGCAAGTGTAGCCCTGGTGTAAAGGGCCA
TGATGACTGAGCTCGTCCCGGCTCCCTCCAGTTTCTCCTGCGCATCTCTTAAAGTCCCGGTTAACCCGCTGGCAATAAGAAAAGGGTTCGCTCCTT
GGCGACTTAAACCCAGCATCTCAGCAGCGAGCTGACGACAGCCATGCGACCTGTATTAAGTCCCGAAGGCAATCCATCTCTGGAAAGTTCTTACTAT`
- 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 focal 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_L103283000025
 - 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 and the Caribbean region showing the location of the sample site, JCVL_SITE_G8019, marked with a red pin. The map includes labels for countries like Mexico, Guatemala, El Salvador, Honduras, Nicaragua, Costa Rica, Panama, and Colombia, as well as cities like Mexico City, Tegucigalpa, Managua, San Jose, and Panama. A legend indicates that 1 sample site is represented in this data set.