CAMERA

A Metagenomics Resource for Microbial Ecology

Saul A. Kravitz
Rekha Seshadri
J. Craig Venter Institute

DOE GTL PI Meeting
February 12, 2008

Presentation Outline

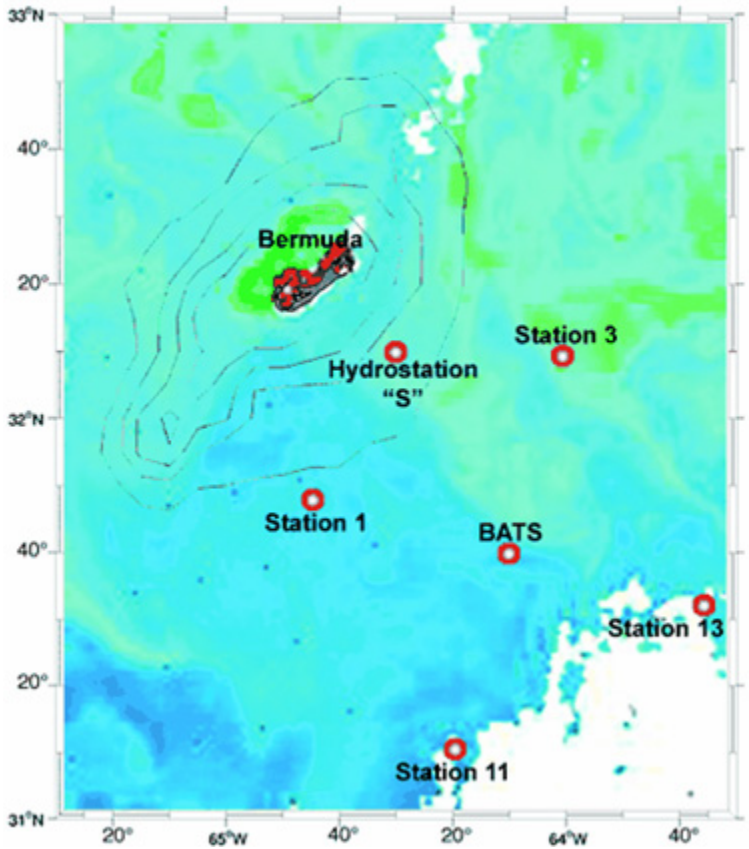
- Introduction to Metagenomics
- Global Ocean Sampling (GOS) Expedition
- Challenges of Metagenomics
- The CAMERA Project
- Live Demonstration of CAMERA

Genomics vs Metagenomics

- Genomics – 'Old School'
  - Study of a single organism's genome
  - Genome sequence determined using shotgun sequencing and assembly
    - ~1300 microbes sequenced, first in 1995
  - DNA usually obtained from pure cultures (<1%)
- Metagenomics
  - Application of genome sequencing methods to environmental samples (no culturing)
  - Environmental shotgun sequencing is the most widely used approach
  - Environmental Metadata provides key context

Metadata

- Metagenomics
  - Genomics + Metadata
- Environmental Metadata
  - Time and location (lat, long, depth) of sample collection
  - Correlate w/remote sensing data
  - Physico-chemical properties (e.g. temperature, nutrients, salinity)



MODIS-Aqua satellite image of ocean chlorophyll in the Sargasso Sea grid about the BATS site from 22 February 2003

Metagenomic Questions

- Within an environment
  - What biological functions are present (absent)?
  - What organisms are present (absent)
- Compare data from (dis)similar environments
  - What are the fundamental rules of microbial ecology
- How do organisms adapt to environmental conditions?
- Search for novel proteins and protein families

Complexity of Microbial Communities

- Simple (e.g., AMD, gutless worm)
  - Few species present (<10)
  - Diverse

  ➔ Variations on standard genomics techniques
- Complex (e.g., Soil or Marine)
  - Many species present (>10, often >1000)
  - Many closely related

  ➔ New techniques

JCVI Global Ocean Sampling Expedition
Largest Metagenomic Study to Date

DOE GENOMICS:GTL ACCELERATING DISCOVERY FOR ENERGY AND ENVIRONMENT
OFFICE OF SCIENCE U.S. DEPARTMENT OF ENERGY

camera Marine Microbial Ecology

J. Craig Venter INSTITUTE

it²

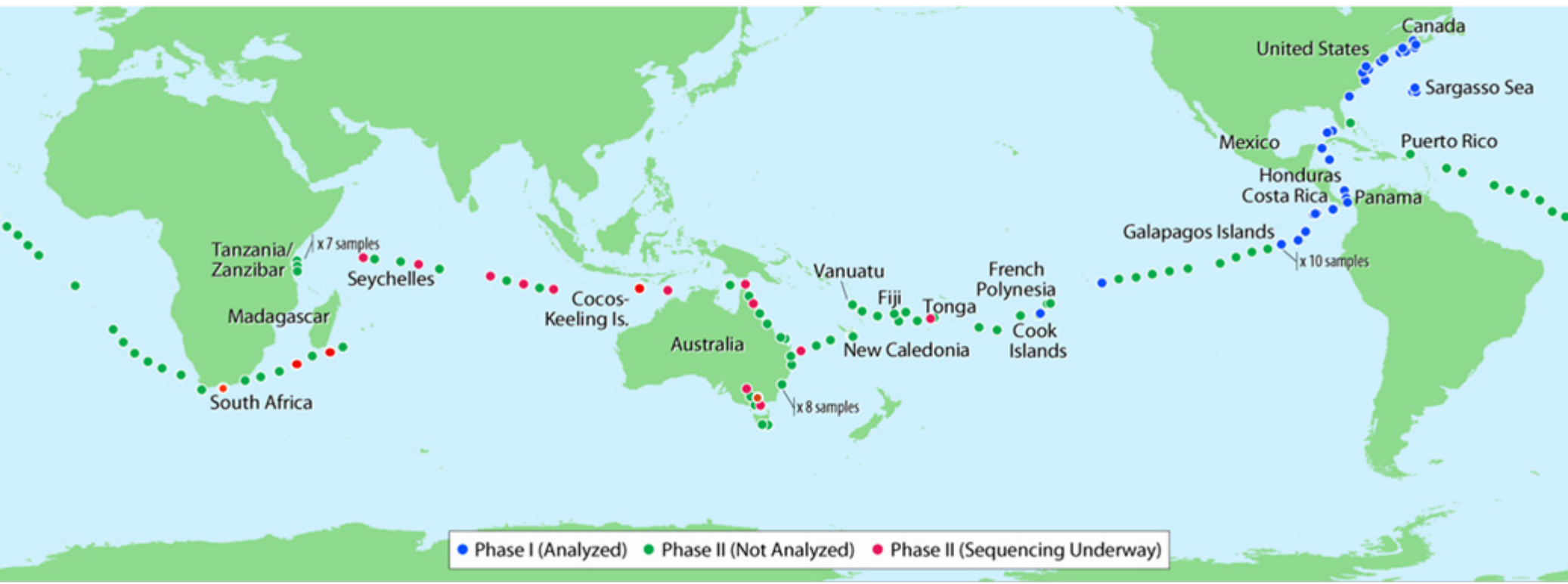
Global Ocean Sampling (GOS)

178 Total Sampling Locations

Phase 1: 41 samples, 7.7M reads, >6M proteins

Diverse Environments

Open ocean, estuary, embayment, upwelling, fringing reef, atoll, warm seep, mangrove, fresh water, biofilms, sediments, soils



DOE GENOMICS:GTL ACCELERATING DISCOVERY FOR ENERGY AND ENVIRONMENT
OFFICE OF SCIENCE U.S. DEPARTMENT OF ENERGY

camera Marine Microbial Ecology

J. Craig Venter INSTITUTE

it²

Challenges of Complex Metagenomic Datasets

- Most sequence reads are unique
  - Very limited assembly
  - Most sequences not taxonomically anchored
  - Annotation of annotated reads required
  - Relating shotgun data to reference genomes
- Annotation challenging
  - Which ORFs code for predicted proteins?
  - Lack of genomic context
  - Large data volume

Motivations for CAMERA

- Significant investment in sequencing
  - Only accessible to bioinformatics elite
  - Diversity of user sophistication and needs
- Bioinformatics and Computation Challenges
  - Assembly, annotation, comparative analysis, visualization
  - Dedicated compute resources
- Importance of Metadata
  - Metadata required for environmental analysis
  - Need to drive standards
- Compliance with Convention on Biodiversity

CAMERA Team

- J. Craig Venter Institute
  - Marv Frazier, co-PI
  - Saul Kravitz, Program Manager
  - Rekha Seshadri, Chief Scientist
- UCSD/Calit2
  - Larry Smarr, co-PI
  - Paul Gilna, Executive Director
- UC Davis
  - Jonathan Eisen, co-investigator
- Gordon and Betty Moore Foundation
  - David Kingsbury and Mary Maxon

Demo of CAMERA 1.3

- Data and Metadata Collections
  - ALL metagenomic nt sequence and metadata
  - Coding sequences functionally annotated
- BLAST with Environmental Metadata
  - Large query and subject sizes
  - 454 FLX run or microbe VS All Metagenomic
- Fragment Recruitment Viewer
  - Based on Rusch's GOS viewer
- Text Search of Annotations

camera Marine Microbial Ecology >> research

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

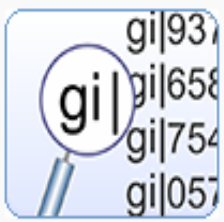
>> Research > Research Home

Research Home | Search | Tools | Projects & Data | Viewers

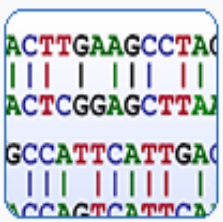
Research Home

help

Search



BLAST

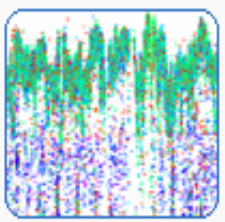


Projects

Publications

Samples

Fragment Recruitment



Recent Searches

[hide]

- lipase no matches
- glycerol kinase (about 4,239 matches)
- Cellulase (about 3,163 matches)
- ATP synthase epsilon (about 2,856 matches)
- dehalogenase (about 4,963 matches)

Recent BLAST Results

[hide]

Job Name	Submit Date	Status	# Hits	Subject Sequences
test11	02/27/07 02:09 PM	completed	2	GOS: Assembled Sequences (N)
tst10	02/27/07 02:09 PM	completed	0	GOS: Combined Assembly ORF Peptides (P)
test09	02/27/07 02:08 PM	completed	1	GOS: All Metagenomic Sequence Reads (N)
test06	02/27/07 02:05 PM	completed	50	All Metagenomic Sequence Reads (N)

View All BLAST Results

This is the "Projects" page. A synopsis of the selected project is displayed

Projects

- Acid Mine Drainage Metagenome
- Alvinella pompejana Epibiont Metagenome
- Chesapeake Bay Virioplankton Metagenome (Data Coming Soon)
- Global Ocean Sampling Expedition
- Mediterranean Gutless Worm Metagenome
- Mediterranenan Bathypelagic Habitat Metagenome
- Metagenome of Marine NaCl-Saturated Brine
- Microbial Community Genomics at the HOT/ALOHA
- Moore Marine Microbial Sequencing
- Ocean Viruses
- Phosphorous Removing (EBPR) Sludge Community
- Viral Metagenomes from Terrestrial Hot Springs
- Waseca County Farm Soil Metagenome
- Whale Fall Metagenome

Details

Global Ocean Sampling Expedition

Principal Investigator: J. Craig Venter

Funded By: Moore Foundation, DOE, Venter Science Foundation

Organization: J. Craig Venter Institute

More Information

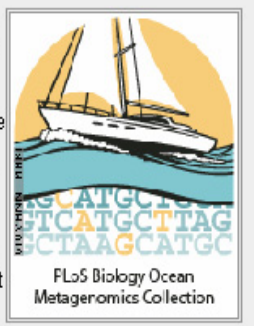
>> Samples
>> Publications & Data
>> Website
>> Contact

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.

The Sorcererll Research Vessel Courtesy - J.Craig Venter Institute
Click on the image for more information.

One avenue of exploration is to sequence the genomes of marine microbes using a metagenomics approach. In Spring of 2003, 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.



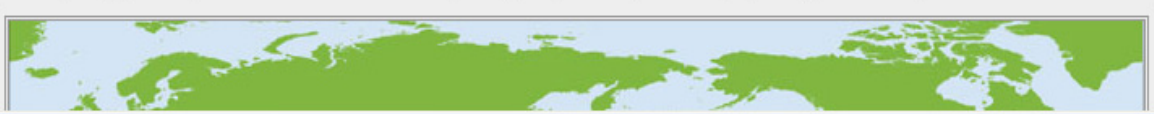
Click on the image for full collection of articles for Global Ocean Sampling.

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 II 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 2006. 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 2007. 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.

JCVI's GOS data is also available through NCBI's Genban k

JCVI has had many requests for GOS materials (clones, templates, DNA). Unfortunately, managing this large collection of materials is an expensive endeavor, and JCVI is not curently equipped to provide this service. JCVI is exploring the possibility of a 3rd party taking on this role, but there are no definitive plans as of yet.



Project Samples - Windows Internet Explorer

http://web.camera.calit2.net/cameraweb/gwt/org.jcvi.camera.web.gwt.download.ProjectSamplesPage/ProjectSamplesPage.htm

File Edit View Favorites Tools Help

Project Samples

camera Marine Microbial Ecology >> research

SEARCH CONTACT US LOG OUT go

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

>> Research > Data > Browse Samples

Research Home | Clusters | Jobs | Sets | Viewers | Data

Project Samples

Project: Global Ocean Sampling Expedition [view project] [export to Excel]

1 - 20 of 59 Advanced Sort prev 20 | goto | next 20

Sample Dataset ▲	Size	Habitat Type	Geographic Location	Sample Location	Country	Filter Size	Latitude	Longitude	Depth
GS000a Shotgun	1,106 MB	Open Ocean	Sargasso Sea	Sargasso Station 13	Bermuda (UK)	0.1 - 0.8	31d 10'30"n	64d 19'27.6"w	5 m
GS000a Shotgun	1,106 MB	Open Ocean	Sargasso Sea	Sargasso Station 11	Bermuda (UK)	0.1 - 0.8	31d 10'30"n	64d 19'27.6"w	5 m
GS000b Shotgun	541.6 MB	Open Ocean	Sargasso Sea	Sargasso Station 13	Bermuda (UK)	0.22 - 0.8	31d 10'30"n	64d 19'27.6"w	5 m
GS000b Shotgun	541.6 MB	Open Ocean	Sargasso Sea	Sargasso Station 11	Bermuda (UK)	0.22 - 0.8	31d 10'30"n	64d 19'27.6"w	5 m
GS000c Shotgun	628 MB	Open Ocean	Sargasso Sea	Sargasso Stations 3	Bermuda (UK)	0.22 - 0.8	32d 10'29.4"n	64d 00'36.6"w	5 m
GS000d Shotgun	566.8 MB	Open Ocean	Sargasso Sea	Sargasso Station 13	Bermuda (UK)	0.22 - 0.8	31d 10'30"n	64d 19'27.6"w	5 m
GS001a Shotgun	242.2 MB	Open Ocean	Sargasso Sea	Hydrostation S	Bermuda (UK)	3 - 20	32d 10'00"n	64d 3	
GS001b Shotgun	154.5 MB	Open Ocean	Sargasso Sea	Hydrostation S	Bermuda (UK)	0.8 - 3	32d 10'00"n	64d 3	
GS001c Shotgun	156.8 MB	Open Ocean	Sargasso Sea	Hydrostation S	Bermuda (UK)	0.1 - 0.8	32d 10'00"n	64d 3	
GS002 Shotgun	150.2 MB	Coastal	North American Ea...	Gulf of Maine	USA	0.1 - 0.8	42d 30'11"n	67d 1	
GS003 Shotgun	108.4 MB	Coastal	North American Ea...	Browns Bank, Gulf...	Canada	0.1 - 0.8	42d 51'10"n	66d	
GS004 Shotgun	92.8 MB	Coastal	North American Ea...	Outside Halifax, ...	Canada	0.1 - 0.8	44d 8'14"n	63d 3	
GS005 Shotgun	107.1 MB	Embayment	North American Ea...	Bedford Basin, No...	Canada	0.1 - 0.8	44d 41'25"n	63d 3	
GS006 Shotgun	104.8 MB	Estuary	North American Ea...	Bay of Fundy, Nov...	Canada	0.1 - 0.8	45d 6'42"n	64d 5	
				Northern Gulf of ...	Canada	0.1 - 0.8	43d 37'56"n	66d 5	
				Newport Harbor, RI	USA	0.1 - 0.8	41d 29'9"n	71d 21'4"w	1 m
				Block Island, NY	USA	0.1 - 0.8	41d 5'28"n	71d 36'8"w	1 m
				Cape May, NJ	USA	0.1 - 0.8	38d 56'24"n	74d 41'6"w	1 m
				Delaware Bay, NJ	USA	0.1 - 0.8	39d 25'4"n	75d 30'15"w	1 m
				Chesapeake Bay, MD	USA	0.1 - 0.8	38d 56'49"n	76d 25'2"w	13.2 m

Advanced Sort

Sort By: Sample Dataset — Ascending / Descending

Then By: — Ascending / Descending

Then By: — Ascending / Descending

Sort Cancel

Download File

File description: *GS006 Shotgun, Bay of Fundy, Nova Scotia*

>> Download as .zip

>> Download as .gz

Clicking the "Publications and Data" link on the project page invokes a "Publications" page

Publications

Project: Global Ocean Sampling Expedition

- **The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific**
  Douglas B. Rusch, Aaron L. Hal...
- **Structural and functional diversity of the microbial kinome**
  Natarajan Kannan, Susan S. Tay...
- **The Sorcerer II Global Ocean Sampling Expedition: Expanding the Universe of Protein Families**
  Shibu Yooseph, Granger Sutton,...
- **Environmental Genome Shotgun Sequencing of the Sargasso Sea**
  J. Craig Venter, Karin Remingt...
- **Assessing diversity and biogeography of aerobic anoxygenic phototrophic bacteria in surface waters of the Atlantic and Pacific Oceans using the Global Ocean Sampling expedition metagenomes**
  Natalya Yutin, Marcelino T. Su...
- **Prokaryotic Gene Finding and Protein Classification via Incremental Clustering**
  Shibu Yooseph, Weizhong Li, Gr...
- **Viral Photosynthetic Reaction Centre Genes and Transcripts in the Marine Environment**
  Oded Beja, Yael Mandel-Gutfreu...

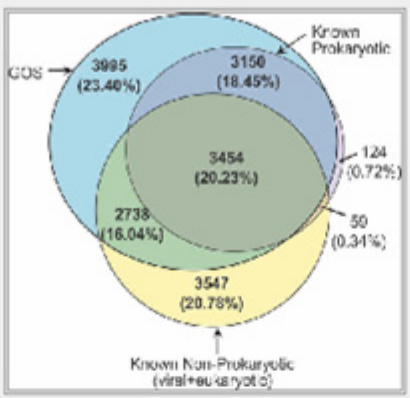
Details | Downloads [◄width]

The Sorcerer II Global Ocean Sampling Expedition: Expanding the Universe of Protein Families

Citation: The Sorcerer II Global Ocean Sampling Expedition: Expanding the Universe of Protein Families
Shibu Yooseph, Granger Sutton, Douglas B. Rusch, Aaron L. Halpern et al.
PLoS Biology Vol. 5, Issue 3, e16 doi:10.1371/journal.pbio.0050016
CAMERA Identifier CAM_PUB_Yooseph07a

More Information
- >> Download this publication
- >> Download publication data
- >> Browse the project
- >> Browse project samples
- >> View the project website

Synopsis: The sequence similarity based clustering approach described in this paper was used both to predict proteins from the Global Ocean Sampling data and to cluster the predicted proteins in the context of an inclusive collection of proteins from public data sources. The clustering results then serve as a resource to address various questions on protein family evolution, diversity, and rate of novel family discovery.



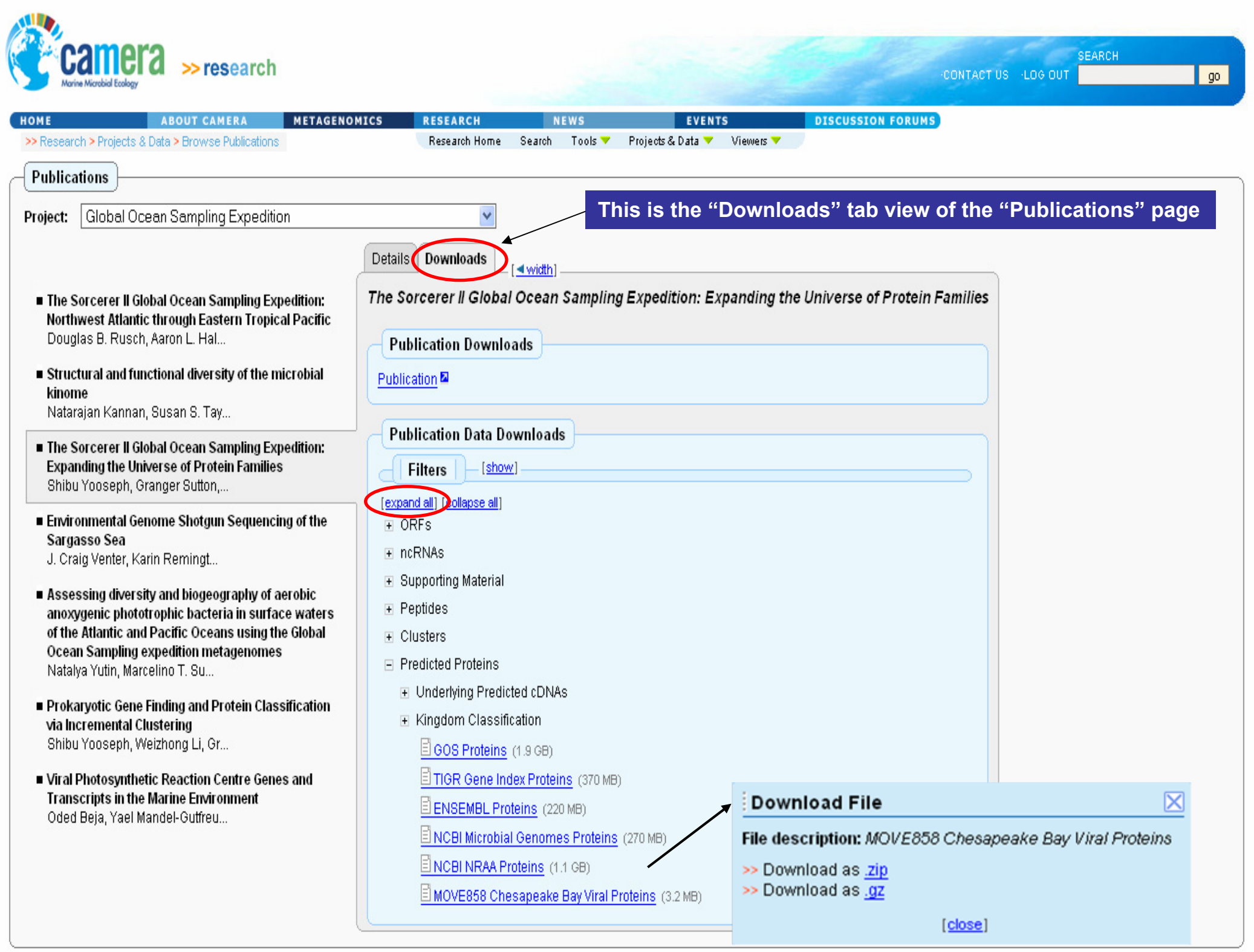
Venn Diagram Showing Breakdown the 17,067 Medium and Large Clusters by three Categories GOS, Known Prokaryotic and Known Non-Prokaryotic
Click to enlarge.

The versions of the following publicly available data sets from February 10, 2005, were utilized in the clustering effort reported in the paper: NCBI's non-redundant protein database (NR) , NCBI Prokaryotic Genomes (PG), TIGR Gene Indices (TGI-EST), and Ensembl (ENS). These data sets are available for download and search. (CAMERA note: Weizhong Li and Shibu Yooseph are currently in the process of building a pipeline for incrementally recomputing the clusters and annotations. Starting in the 2nd quarter of 2007, CAMERA will be updating the clusters periodically incorporating new data, including data from other genomics/metagenomics studies.)

Reads from the Sargasso Sea study were pooled together with the reads from the Sorcerer II Global Ocean Sampling Expedition and the combined set is referred to as the GOS data set. Assembly of the GOS data set was performed using the Celera Assembler. All of the reads and assemblies are available for download and search, and a mapping of GOS ORFs to scaffolds in the GOS assembly is provided in gff format.

Open Reading Frames(ORFs) were identified from the DNA sequences in the PG, TGI-EST, ENS, and GOS data sets. An ORF is commonly defined as a translated DNA sequence that begins with a start codon and ends with a stop codon. To accommodate partial DNA sequences, this definition was extended to allow an ORF to be bracketed by either a start codon or the start of the DNA sequence, and by either a stop codon or the end of the DNA sequence. While generating ORFs from the PG and TGI-EST data sets, the appropriate codon usage table for the DNA sequence was used. GOS ORFs were generated from the assembled sequences only using translation table 11 (the code for bacteria, archaea, and prokaryotic viruses). Only ORFs of length 60 amino acids (aa) were considered for all data sets. Not all ORFs are proteins. In this study, ORFs that have reasonable evidence for being proteins, are denoted as predicted proteins; other ORFs are denoted as spurious ORFS. All of these ORF sets are available for download and search.

The primary input to the clustering process was pairwise sequence similarity, computed using an all-against-all BLAST search. The sequences were clustered in four steps. In the first step, a non-redundant set of sequences from the entire data set was identified. Only pairwise matches with 98% similarity involving 95% length of the shorter sequence, were considered. Only non-redundant sequences were considered for further steps in the clustering. In the second step, core sets of similar sequences were identified. Only matches between two sequences involving 80% of the length of the longer sequence were considered. In the third step, these core sets were transformed into profiles and a profile-profile method was used to merge related core sets. In the final step, sequences were recruited using sequence-profile matching (PSI-BLAST) and using BLAST matches to core-set members. The recruitment required the match to involve 60% of the length of the sequence being recruited.


camera
Marine Microbial Ecology
>> research
CONTACT US
LOG OUT
SEARCH
go
HOME
ABOUT CAMERA
METAGENOMICS
RESEARCH
NEWS
EVENTS
DISCUSSION FORUMS
>> Research > Projects & Data > Browse Publications
Research Home
Search
Tools
Projects & Data
Viewers
Publications
Project:
Global Ocean Sampling Expedition
This is the “Downloads” tab view of the “Publications” page
Details
Downloads
[width]
The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific
Douglas B. Rusch, Aaron L. Hal...
Structural and functional diversity of the microbial kinome
Natarajan Kannan, Susan S. Tay...
The Sorcerer II Global Ocean Sampling Expedition: Expanding the Universe of Protein Families
Shibu Yooseph, Granger Sutton,...
Environmental Genome Shotgun Sequencing of the Sargasso Sea
J. Craig Venter, Karin Remingt...
Assessing diversity and biogeography of aerobic anoxygenic phototrophic bacteria in surface waters of the Atlantic and Pacific Oceans using the Global Ocean Sampling expedition metagenomes
Natalya Yutin, Marcelino T. Su...
Prokaryotic Gene Finding and Protein Classification via Incremental Clustering
Shibu Yooseph, Weizhong Li, Gr...
Viral Photosynthetic Reaction Centre Genes and Transcripts in the Marine Environment
Oded Beja, Yael Mandel-Gutfreu...
The Sorcerer II Global Ocean Sampling Expedition: Expanding the Universe of Protein Families
Publication Downloads
Publication
Publication Data Downloads
Filters
[show]
[expand all] [collapse all]
ORFs
ncRNAs
Supporting Material
Peptides
Clusters
Predicted Proteins
Underlying Predicted cDNAs
Kingdom Classification
GOS Proteins (1.9 GB)
TIGR Gene Index Proteins (370 MB)
ENSEMBL Proteins (220 MB)
NCBI Microbial Genomes Proteins (270 MB)
NCBI NRAA Proteins (1.1 GB)
MOVE858 Chesapeake Bay Viral Proteins (3.2 MB)
Download File
File description: MOVE858 Chesapeake Bay Viral Proteins
>> Download as .zip
>> Download as .gz
[close]

Details

Moore Marine Microbial Sequencing

Principal Investigator: Various

Funded By: The Gordon and Betty Moore Foundation

Organization: The Gordon and Betty Moore Foundation

Affiliation: The J. Craig Venter Institute

More Information

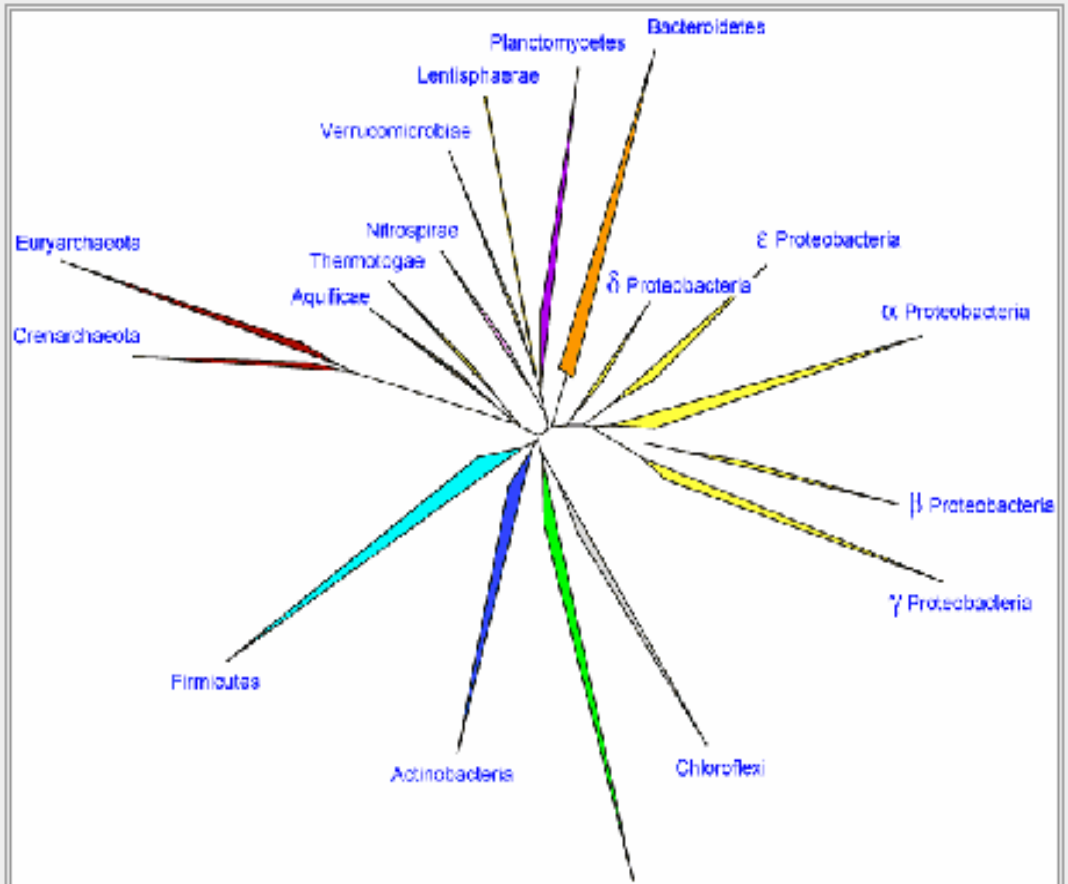
>> Samples

>> Publications & Data

>> Website

>> Contact

The overarching goal of the marine microbial sequencing project is to explore the distribution, function and ecological role of marine microbes, and their role in the overall health of oceans and climates. The field of marine microbiology has made amazing progress in recent years through advances in genomic and metagenomic technologies, and is causing us to reevaluate many of the basic precepts of microbial ecology and evolution. Nonetheless, out of the 300+ prokaryotic genomes that have been completely sequenced to date, only about 18 are from the marine environment. In an attempt to bridge this gap, we are sequencing the draft genomes of over 155 diverse, ecologically relevant marine microorganisms that were selected for sequencing based on the following criteria:



1.Open ocean and coastal bacterioplankton isolates most closely related to uncultured marine clades abundant in ribosomal RNA surveys.

2.Open ocean and coastal bacterioplankton isolates for .cluster sequencing. from ecologically relevant marine clades.

3.Bacterioplankton isolates representing broad physiological diversity (photoautotrophs, photoheterotrophs, nitrifiers, methanotrophs, etc.).

4.Isolates from diverse environments/habitats (greater depths, more extreme temperatures, specialized habitats).

These genome data not only hold the promise of delivering tremendous biological insight through individual and broad comparative genome analyses, but also provide a key reference collection for culture-independent metagenomic surveys like the global ocean sampling expedition and many others.

The project was initiated in Fall 2004 with a grant to the J. Craig Venter Institute. Auto-annotated genome sequences have been deposited in GenBank. For further information, see the J.Craig Venter Institute's Marine Microbial Genome Sequencing Project.

Publications and Data

Project: Moore Marine Microbial Sequencing

This project has no papers.

Marine Organisms

1 - 20 of 168 Advanced Sort ◀ prev 20 | goto | next 20 ▶

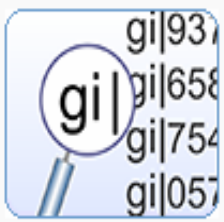
Organism	Investigator	Release Date	Status
Acaryochloris sp. CCMEE 5410	Michelle Wood		Not Yet Received
Aciduliprofundum boonei T469	Anna-Louise Reysenbach	03/07/06	Released to Genbank
AEGEAN_169 clade (SAR11-like) HIMB37	Michael Rappé		Not Yet Received
Alcanivorax sp. DG881	Christopher Bolch	04/07/06	Released to Genbank
Algoriphagus sp. PR1	Nicole King	12/14/05	Released to Genbank
alpha proteobacterium BAL199	Åke Hagström	04/06/06	Released to Genbank
alpha proteobacterium HTCC2255	Stephen J. Giovannoni, Jang-Cheon Cho	08/09/05	Released to Genbank
Alteromonadales TW-7	Farooq Azam	11/22/05	Released to Genbank
Alteromonas macleodii 'Deep ecotype'	Francisco Rodriguez-Valera	11/19/04	Released to Genbank
Aurantimonas sp. SI85-9A1	Bradley Tebo	08/17/04	Released to Genbank
Bacillus sp. B14905	Rob Edwards	11/15/05	Released to Genbank
Bacillus sp. NRRL B-14911	Janet Siefert	08/24/04	Released to Genbank
Bacillus sp. SG-1	Bradley Tebo	11/29/05	Released to Genbank
Beggiatoa sp. 'Orange Guaymas'	Andreas Teske	10/10/07	16S QC (PCR -- Analysis)
beta proteobacterium KB13	Michael Rappe	02/20/07	Delivered to Researcher
Blastopirellula marina DSM 3645	Rudolf Amann	09/01/04	Released to Genbank
Brevundimonas sp. BAL3	Åke Hagström	04/06/06	Released to Genbank
Caldanaerobacter DSM 12653	Juan Gonzalez, Tatyana Sokolova	05/17/06	Delivered to Researcher
Calothrix sp. SC01	Jonathan Zehr		Not Yet Received
Caminibacter mediatlanticus TB-2	Costa Vetriani	12/13/05	Released to Genbank

1 - 20 of 168 **Show:** 10 20 50 ◀ prev 20 | goto | next 20 ▶

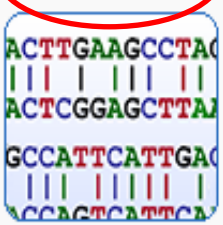
Research Home

help

Search



BLAST

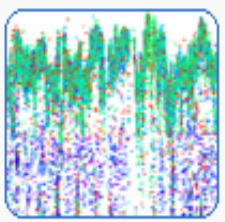


Projects

Publications

Samples

Fragment Recruitment



Recent Searches [hide]

- lipase no matches
- glycerol kinase (about 4,239 matches)
- Cellulase (about 3,163 matches)
- ATP synthase epsilon (about 2,856 matches)
- dehalogenase (about 4,963 matches)

Recent BLAST Results [hide]

Job Name	Submit Date	Status	# Hits	Subject Sequences
test11	02/27/07 02:09 PM	completed	2	GOS: Assembled Sequences (N)
tst10	02/27/07 02:09 PM	completed	0	GOS: Combined Assembly ORF Peptides (P)
test09	02/27/07 02:08 PM	completed	1	GOS: All Metagenomic Sequence Reads (N)
test06	02/27/07 02:05 PM	completed	50	All Metagenomic Sequence Reads (N)

View All BLAST Results

Specify Query Sequence

Specify a nucleotide or peptide query sequence (in FASTA format):

Back Next

New Sequence | Previous Sequence

Sequence Name:

User sequence 02/08/08 05:50 AM

Enter sequence in fasta format:

Upload fasta sequence file

Browse... Upload

Specify Query Sequence

Specify a nucleotide or peptide query sequence (in FASTA format):

Back | Next

New Sequence | **Previous Sequence**

1 - 15 of 40 ◂ prev 15 | goto | next 15 ▸

	Sequence Name	Added ▼	Type	File size	Action
✘	De-Dehalogenase [edit]	02/07/08 12:04 PM	peptide	497	Export
✘	NasA-uncultured-gamma [edit]	02/07/08 11:36 AM	peptide	897	Export
✘	AtzB [edit]	02/07/08 11:32 AM	peptide	363	Export
✘	User sequence 01/11/08 03:15 PM [edit]	01/11/08 03:15 PM	nucleotide	502	Export
✘	User sequence 01/07/08 03:20 PM [edit]	01/07/08 03:21 PM	peptide	363	Export
✘	NasA nitrate reductase [edit]	01/07/08 02:08 PM	peptide	897	Export
✘	Coxiella sterol reducta01/07/08 01:55 PM [edit]	01/07/08 01:56 PM	peptide	448	Export
✘	Arthrobacter AtzD [edit]	01/07/08 01:38 PM	peptide	363	Export
✘	User sequence 12/31/07 10:39 AM [edit]	12/31/07 10:39 AM	nucleotide	1,499	Export
✘	User sequence 12/31/07 10:18 AM [edit]	12/31/07 10:18 AM	nucleotide	1,388	Export
✘	User sequence 12/10/07 10:23 AM [edit]	12/10/07 11:13 AM	peptide	128	Export
✘	User sequence 11/06/07 08:51 AM [edit]	11/06/07 08:52 AM	peptide	554	Export
✘	User sequence 11/06/07 07:41 AM [edit]	11/06/07 07:42 AM	peptide	149	Export
✘	User sequence 11/06/07 07:39 AM [edit]	11/06/07 07:41 AM	peptide	149	Export
✘	User sequence 10/18/07 11:15 AM [edit]	10/18/07 11:17 AM	peptide	520,053	Export

1 - 15 of 40 **Show:** 10 20 50 ◂ prev 15 | goto | next 15 ▸

• Click once to select. • Double click to select and apply.

Select Reference Datasets to Search Against

Read Datasets | Assembled Datasets | External Datasets | All Datasets

Description	Length	Sequence Type
All Metagenomic ncRNAs (N)	33,370,426	nucleotide
All Metagenomic ORF Peptides (P)	4,439,529,911	peptide
All Metagenomic ORFs (N)	13,318,589,733	nucleotide
All Metagenomic Sequence Reads (N)	8,393,828,709	nucleotide
FLAS: All Metagenomic Sequence Reads (N)	2,380,900	nucleotide
GOS: All Metagenomic Sequence Reads (N)	8,072,051,361	nucleotide
GOS: Translated peptide sequences of ORFs predicted from metagenomic sequence reads from the GOS Expedition		e
GOS: All ORF Peptides (P)	4,395,420,221	peptide
GOS: All ORFs (N)	13,186,260,663	nucleotide
GOS: GS000a Reads from 0.1-0.8 Sargasso Sea Stations 11 and 13 (N)	658,755,696	nucleotide
GOS: GS000b Reads from 0.22-0.8 Sargasso Sea Stations 11 and 13 (N)	321,026,307	nucleotide
GOS: GS000c Reads from 0.22-0.8 Sargasso Sea Station 3 (N)	371,688,861	nucleotide
GOS: GS000d Reads from 0.22-0.8 Sargasso Sea Station 13 (N)	335,939,509	nucleotide
GOS: GS001a Reads from 3.0-20.0 Sargasso Sea Hydrostation S (N)	143,316,448	nucleotide
GOS: GS001b Reads from 0.8-3.0 Sargasso Sea Hydrostation S (N)	90,955,161	nucleotide
GOS: GS001c Reads from 0.1-0.8 Sargasso Sea Hydrostation S (N)	92,688,958	nucleotide
GOS: GS002 Reads from 0.1-0.8 Gulf of Maine USA (N)	128,761,768	nucleotide

• Click once to select. • Double click to select and apply.

Job Criteria

Job Name:* xyz

Query Sequence: De-Dehalogenase [change]

Subject Datasets: GOS: All ORF Peptides (P) [change]

Job Options

Program: BLASTP (prot/prot)

Basic Options

database alignments: 10000

filter low-complexity seq: (•) True ( ) False

evalue exponent (1Ex): -5

lower case filtering: ( ) True (•) False

Advanced Options [hide]

believe defline: ( ) True (•) False

database size: 0.0

gap extend cost: -1

gapped alignment: (•) True ( ) False

hit extension threshold: 11

multihit window size: 40

show GIs: ( ) True (•) False

wordsize: 3

best hits to keep: 0

final gapped dropoff: 25.0

gap open cost: -1

gapped alignment dropoff: 15

matrix: BLOSUM62

search size: 0.0

ungapped extension dropoff: 7.0

Back Submit Job

Job Results

1 - 20 of 38 Advanced Sort

prev 20 | goto | next 20

	Job Name	Submit Date ▼	Status	Program	# Hits	Subject Sequences	Actions
✘	testMoore [edit]	02/08/08 02:15 PM	completed	tblastn	10	MarineMicrobes: Moore Foundation Marine Microbi...	Results / Job / Export
✘	De-dehaloVsGOS [edit]	02/07/08 12:03 PM	completed	blastp	82	GOS: All ORF Peptides (P)	Results / Job / Export
✘	NasAvsGOS [edit]	02/07/08 11:51 AM	completed	blastp	1,019	GOS: All ORF Peptides (P)	Results / Job / Export
✘	NasAvsGOS [edit]	02/07/08 11:36 AM	completed	blastp	4,060	GOS: All ORF Peptides (P)	Results / Job / Export
✘	sdfs [edit]	01/11/08 03:14 PM	completed	blastn	25	GOS: All ncRNAs (N)	Results / Job / Export
✘	testg [edit]	01/07/08 02:15 PM	completed	blastp	1,000	All Metagenomic ORF Peptides (P)	Results / Job / Export
✘	testf [edit]	01/07/08 02:08 PM	completed	blastp	1,000	GOS: Combined Assembly ORF Peptides (P)	Results / Job / Export
✘	TESTaaa [edit]	12/10/07 11:36 AM	completed	tblastn	2,615	All Metagenomic ORFs (N)	Results / Job / Export
✘	test [edit]	11/06/07 07:41 AM	completed	blastp	318	GOS: All ORF Peptides (P)	Results / Job / Export
✘	ntsars02vsnraa [edit]	10/18/07 11:17 AM	completed	blastp	3,180	CAMERA's Non-Identical Peptide Sequences (P)	Results / Job / Export
✘	ntsars1vsnraa [edit]	10/18/07 10:46 AM	completed	blastp	2,447	CAMERA's Non-Identical Peptide Seque	/ Export
✘	ttt [edit]	08/01/07 01:25 PM	completed	blastp	25	Microbial draft genome ORF peptide sequences (P)	Results / Job / Export
✘	ess [edit]	07/20/07 09:58 AM	completed	tblastn	230	GOS: All Metagenomic Sequence Reads (N)	Results / Job / Export
✘	atz [edit]	07/17/07 05:05 PM	completed	blastp	318	GOS: All ORF Peptides (P)	
✘	atzAA [edit]	07/17/07 04:57 PM	completed	blastp	409	GOS: All ORF Peptides (P)	
✘	atzD [edit]	07/17/07 04:47 PM	completed	blastp	11	GOS: All ORF Peptides (P)	Results / Job / Export
✘	atzC [edit]	07/17/07 04:45 PM	completed	blastp	200	GOS: All ORF Peptides (P)	Results / Job / Export
✘	atzB [edit]	07/17/07 04:44 PM	completed	blastp	200	GOS: All ORF Peptides (P)	Results / Job / Export

Re-run this job w/ new parameters
Show parameters

All Matching Query Sequences as FASTA
All Matching Subject Sequences as FASTA
All BLAST Results with Metadata as CSV
All BLAST Results as NCBI XML

Job Summary

[show] [back to job results]

Matching Sequences

[hide]

1 - 10 of 82

prev 10 | goto | next 10

	Eval	Score	Len.	Query	Subject	Sample(s)	Location(s)
☐	5.142E-37	161.0	276	USER_PEP_1202414661201	JCVI_PEP_1105132741961	GS011	Delaware Bay, NJ
☐	2.097E-22	112.5	144	USER_PEP_1202414661201	JCVI_PEP_1105156975833	GS011	Delaware Bay, NJ
☐	2.482E-15	89.0	206	USER_PEP_1202414661201	JCVI_PEP_1105131342681	GS008	Newport Harbor, RI
☐	2.748E-14	85.5	226	USER_PEP_1202414661201	JCVI_PEP_1105118692855	GS000c	Sargasso Stations 3
☐	6.119E-14	84.3	241	USER_PEP_1202414661201	JCVI_PEP_1105116171163	GS021	Gulf of Panama
☐					78257	GS006	Bay of Fundy, Nova Scotia
☐					05957	GS031	Upwelling, Fernandina Island
☐					94839	GS004	Outside Halifax, Nova Scotia
☐					67667	GS033	Punta Cormorant, Hypersalin...
☐					39069	GS033	Punta Cormorant, Hypersalin...

Select | Advanced Sort | Export

- Matching Query Sequences as FASTA
- Matching Subject Sequences as FASTA
- BLAST Results with Metadata as CSV
- BLAST Results as NCBI XML

1 - 10 d

prev 10 | goto

- Hover
- BLAS

```
Query: 261 ANESQIMYAANLSRYRLFGNIQNSIQEFIRSLGYTCYGYA 300
Sbjct: 77  SGDDWIAVAQSMRAYLRFSLLGGVIAKQIRNLGYSAKAHT 116

Query: 301 SPFSGIMPAIASAQLTGITEGNRNNGFCTSPEYGPVLGVF 340
Sbjct: 117 STDGDVLQP-PLLLLSGLGEVSRIGEVILNPLLGPRLKSG 155

Query: 341 SLVTDMPLEPTNPIDAGIWRFCQTCTKCADACPVNAIPKD 380
Sbjct: 156 VVTTDMPMLFDKPIDFGLQAFCESCNKCARECPAGAITAG 195

Query: 381 HEPTWDLPNIYGKADIVHSPGRKQYWTNAVDCWLF-LTEY 419
Sbjct: 196 PKLMFN--------------GYEIWKSDSQKCTTYRLTNE 221

Query: 420 NG--CGACMATCTFNTNTAPIHELVRATLATTPTLNTFLW 457
Sbjct: 222 GGAMCGRCMKTCPWNLEGLFKEKPFRWLASNVPSVAPVLA 261

Query: 458 QADKFFGY-GLNEDKEEWWDLSLPVYGFDSAATSSHGGYD 496
Sbjct: 262 KLDDRLGNGGLNEIKKWWWDLEL----------KDDGGYK 291

Query: 497 K 497
Sbjct: 292 K 292
```

Seq

Seque: 99 / 276 (35%)

Seque: 46 / 276 (16%)

Alignı ps: 8

Clear aps: 12

```
Query: 218 GKAYVFEDVDVGYEGATTFVLPANKQLYEFCFTVPMSKDMFRTA---NESQIMYAANLSRYRLFGNIQNSIQEFIRSLGY 294
           GK   FEDV+  YE  T  V+P N + +   FTV MS+++ +       +    AA  + Y     I +  Q F+  LGY
Sbjct: 28  GKRLDFEDVEKPYEEETRRVIPNNAR-WVIVFTVQMSEELLKRRVGKTPTPFSAAAAGNAYARSRVIFDDTQIFLHCLGY 106

Query: 295 TCYGY-ASPFSGIMPAIASAQLTGITEGNRNNGFCTSPEYGPVLGVFSLVTDMPLEPTNPIDAGIWRFCQTCTKCADACP 373
             YG   A  F+G+  A A   + G+ E +R N   T PEYGP++ VF ++T++PL P  PIDAGI RFC+TC KCA  CP
Sbjct: 107 --YGLMAYWFNGLGIAPALGIMAGLGELSRLNRMIT-PEYGPMVRVFRMITNLPLAPNKPIDAGIMRFCKTCKKCATECP 183

Query: 374 VNAIPKDHEPTWDLPNIYGKADIVHSPGRKQYWTNAVDC---WLFLTEYNGCGACMATCTFN-TNTAPIHELVRATLATT 449
             +  D EP+W++   +      ++PG K ++ ++  C   WL  T    C  C + C F+  + + +H  V AT+A T
Sbjct: 184 SGTLSLDTEPSWEVKGPW------NNPGHKTWYEDSPKCMAYWLKATA--SCATCFSVCPFSKKDKSFMHHFVEATIAKT 255

Query: 450 PTLNTFLWQADKFFGYGLNEDKEEWWDLSLPVYGFD 485
```

Sequence Geography

[hide]

>> 29 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

Peptide Detail

[hide] [◄back to job details]

PeptideAcc: JCVI_PEP_1105132741961

Def Line: JCVI_PEP_1105132741961 /read_id=JCVI_READ_1092351470431 /begin=0 /end=900 /orientation=1 /5_prime_stop=0 /3_prime_stop=TAG /orf_id=JCVI_ORF_1105132741960 /ttable=11 /length=300 /ergatis_id=7720 /read_defline="/library_id=JCVI_LIB_GS-11-01-01-1P3-1P8KB /template_id=JCVI_TMPL_1061001659921 /sequencing_direction=forward /sample_id=JCVI_SMPL_1103283000017 /clr_range_begin=194 /clr_range_end=1073 /full_length=1124"

Seq Length: 300

Seq Type: AA

Sequence

```
ERSGTSHVAFLELDEYNKKFIYSVDNDGKRLDFEDVEKPYEEETRRVIPNNARWVIVFTVQMSEELLKRRVGKTPTPFSAAAAGNAYARSRVIFDDTQIFLHCLG
YYGLMAYWFNGLGIAPALGIMAGLGELSRLNRMITPEYGPMVRVFRMITNLPLAPNKPIDAGIMRFCKTCKKCATECPSGTLSLDTEPSWEVKGPWNNPGHKTWY
EDSPKCMAYWLKATASCATCFSVCPFSKKDKSFMHHFVEATIAKTPIFNSLFTKMDTLLGYDIPRDPESWWGTNMPPYGIDGLSGTRLE*
```

Intellectual Property Notice

None

Metadata & Geography | Sequence Features

Sample Metadata

[hide]

Sample Name: GS011

Sample Acc: JCVI_SMPL_1103283000017

Site Id: JCVI_SITE_GS011

Filter Size: 0.1 - 0.8 µm

Site Metadata

[hide] [remove scroll]

site id:	JCVI_SITE_GS011
project:	CAM_PROJ_GOS
location:	Delaware Bay, NJ
region:	North American East Coast
comment:	2nd leg
country:	USA
latitude:	39d 25'4"n
longitude:	75d 30'15"w
habitat:	Estuary
sample depth:	1 m
chlorophyll density:	4.8 mg/M3 (sample month)
leg:	2nd leg

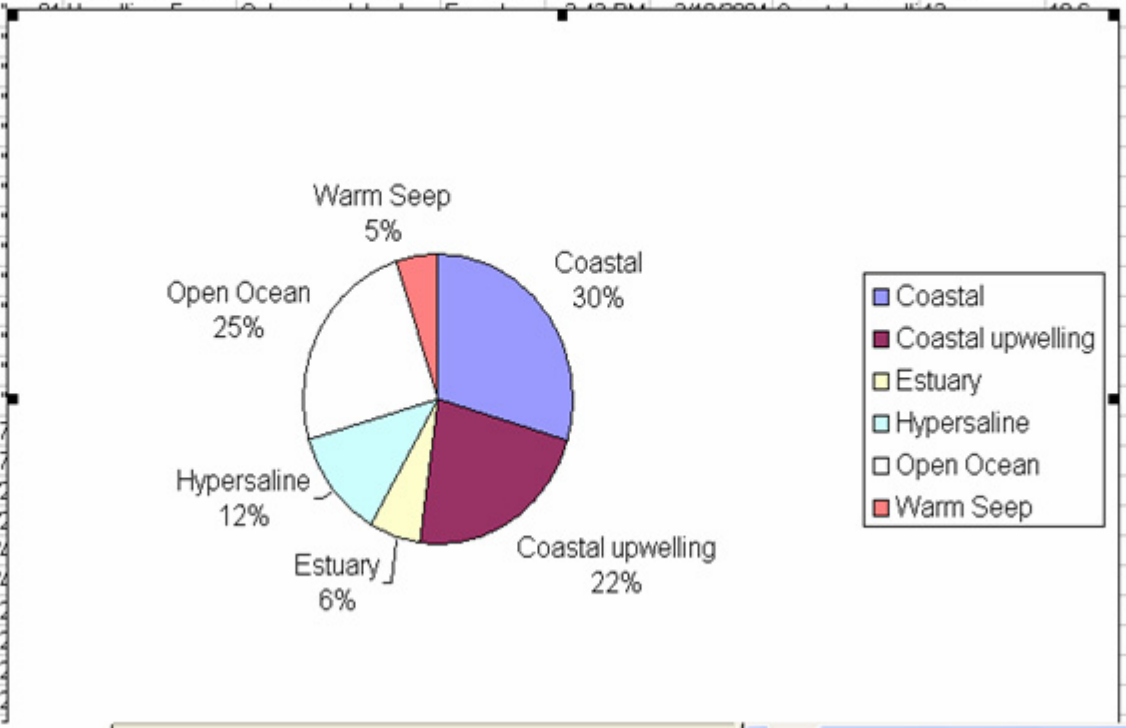
Sequence Geography

[hide]

>> 1 sample site is represented in this data set

- Each sample site is marked on the map.
- Click a site marker for more information on Site Metadata panel

	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Filter Size (	Length	Site	Lat/Long	Smpl Pos	Geo Pos	Nation	Time	Date	Habitat	Smpl Depth	Water De	Temperat	Salinity	Chlorophyll Densi A
2	0.1 - 0.8	237	JCVI_SITE_GS008	41d 29'9"n ; 7	Newport Harbor, F	North American East	USA	4:45 PM	11/16/2003	Coastal	1 m	12 m	9.4 C	26.5 ppt	2.2 mg/M3 (sample 1
3	0.1 - 0.8	360	JCVI_SITE_GS021	8d 7'45"n ; 79	Gulf of Panama	Eastern Tropical Pac	Panama	4:48 PM	1/19/2004	Coastal	1.6 m	76 m	27.6 C	30.7 ppt	0.50 mg/M3 (sampl 0
4	0.1 - 0.8	326	JCVI_SITE_GS004	44d 8'14"n ; 6	Outside Halifax, N	North American East	Canada	5:25 AM	8/22/2003	Coastal	2 m	142 m	17.3 C	28.3 ppt	0.4 mg/M3 (sample 0
5	0.1 - 0.8	325	JCVI_SITE_GS003	42d 51'10"n ;	Browns Bank, Gu	North American East	Canada	11:50 AM	8/21/2003	Coastal	1 m	119 m	11.7 C	29.9 ppt	1.4 mg/M3 (sample 1
6	0.1 - 0.8	228	JCVI_SITE_GS021	8d 7'45"n ; 79	Gulf of Panama	Eastern Tropical Pac	Panama	4:48 PM	1/19/2004	Coastal	1.6 m	76 m	27.6 C	30.7 ppt	0.50 mg/M3 (sampl 0
7	0.1 - 0.8	336	JCVI_SITE_GS008	41d 29'9"n ; 7	Newport Harbor, F	North American East	USA	4:45 PM	11/16/2003	Coastal	1 m	12 m	9.4 C	26.5 ppt	2.2 mg/M3 (sample 1
8	0.1 - 0.8	371	JCVI_SITE_GS008	41d 29'9"n ; 7	Newport Harbor, F	North American East	USA	4:45 PM	11/16/2003	Coastal	1 m	12 m	9.4 C	26.5 ppt	2.2 mg/M3 (sample 1
9	0.1 - 0.8	303	JCVI_SITE_GS010	38d 56'24"n ;	Cape May, NJ	North American East	USA	4:30 AM	11/18/2003	Coastal	1 m	10 m	12 C	31 ppt	2.0 mg/M3 (sample 2
10	0.1 - 0.8	359	JCVI_SITE_GS013	36d 0'14"n ; 7	Off Nags Head, N	North American East	USA	6:28 AM	12/19/2003	Coastal	2.1 m	20 m	9.3 C		3.0 mg/M3 (sample 2
11	0.1 - 0.8	220	JCVI_SITE_GS014	32d 30'25"n ;	South of Charlest	North American East	USA	5:12 PM	12/20/2003	Coastal	1 m	31 m	18.6 C		1.70 mg/M3 (sampl 1
12	0.1 - 0.8	308	JCVI_SITE_GS034	0d 22'59"s ; 9	North Seamore Isl	Galapagos Islands	Ecuador	5:06 PM	2/19/2004	Coastal	2.1 m	35 m	27.5 C		0.36 mg/M3 (sampl 0
13	0.1 - 0.8	338	JCVI_SITE_GS021	8d 7'45"n ; 79	Gulf of Panama	Eastern Tropical Pac	Panama	4:48 PM	1/19/2004	Coastal	1.6 m	76 m	27.6 C	30.7 ppt	0.50 mg/M3 (sampl 0
14	0.1 - 0.8	337	JCVI_SITE_GS021	8d 7'45"n ; 79	Gulf of Panama	Eastern Tropical Pac	Panama	4:48 PM	1/19/2004	Coastal	1.6 m	76 m	27.6 C	30.7 ppt	0.50 mg/M3 (sampl 0
15	0.1 - 0.8	338	JCVI_SITE_GS008	41d 29'9"n ; 7	Newport Harbor, F	North American East	USA	4:45 PM	11/16/2003	Coastal	1 m	12 m	9.4 C	26.5 ppt	2.2 mg/M3 (sample 1
16	0.1 - 0.8	179	JCVI_SITE_GS008	41d 29'9"n ; 7	Newport Harbor, F	North American East	USA	4:45 PM	11/16/2003	Coastal	1 m	12 m	9.4 C	26.5 ppt	2.2 mg/M3 (sample 1
17	0.1 - 0.8	141	JCVI_SITE_GS019	10d 42'59"n ;	Northeast of Color	Caribbean Sea	Panama	9:03 AM	1/12/2004	Coastal	1.7 m	3336 m	27.7 C	35.4 ppt	0.23 mg/M3 (sampl 0
18	0.1 - 0.8	318	JCVI_SITE_GS027	1d 12'58"s ; 9	Devil's Crown, Flo	Galapagos Islands	Ecuador	11:41 AM	2/4/2004	Coastal	2.2 m	2.3 m	25.5 C	34.9 ppt	0.40 mg/M3 (sampl 0
19	0.1 - 0.8	341	JCVI_SITE_GS013	36d 0'14"n ; 7	Off Nags Head, N	North American East	USA	6:28 AM	12/19/2003	Coastal	2.1 m	20 m	9.3 C		3.0 mg/M3 (sample 2
20	0.1 - 0.8	316	JCVI_SITE_GS028	1d 13'1"s ; 90	Coastal Floreana	Galapagos Islands	Ecuador	3:47 PM	2/4/2004	Coastal	2 m	156 m			0.35 mg/M3 (sampl 0
21	0.1 - 0.8	117	JCVI_SITE_GS035	1d 23'21"n ; 9	Wolf Island	Galapagos Islands	Ecuador	4:44 PM	3/1/2004	Coastal	1.7 m	71 m	21.8 C	34.5 ppt	0.28 mg/M3 (sampl 0
22	0.1 - 0.8	219	JCVI_SITE_GS016	24d 10'29"n ;	Gulf of Mexico	Caribbean Sea	USA	2:15 PM	1/8/2004	Coastal Sea	2 m	3333 m	26.4 C	35.8 ppt	0.16 mg/M3 (sampl 0
23	0.1 - 0.8	288	JCVI_SITE_GS016	24d 10'29"n ;	Gulf of Mexico	Caribbean Sea	USA	2:15 PM	1/8/2004	Coastal Sea	2 m	3333 m	26.4 C	35.8 ppt	0.16 mg/M3 (sampl 0
24	0.1 - 0.8	204	JCVI_SITE_GS031	0d 18'4"s ; 91	Upwelling, Fernan	Galapagos Islands	Ecuador	2:43 PM	2/10/2004	Coastal upwelli	12 m	19.6 m	18.6 C		0.35 mg/M3 (sampl 0
25	0.1 - 0.8	354	JCVI_SITE_GS031	0d 18'4"s ; 91	Upwelling, Fernan	Galapagos Islands	Ecuador	2:43 PM	2/10/2004	Coastal upwelli	12 m	19.6 m	18.6 C		0.35 mg/M3 (sampl 0
26	0.1 - 0.8	340	JCVI_SITE_GS031	0d 18'4"s ; 91	Upwelling, Fernan	Galapagos Islands	Ecuador	2:43 PM	2/10/2004	Coastal upwelli	12 m	19.6 m	18.6 C		0.35 mg/M3 (sampl 0
27	0.1 - 0.8	314	JCVI_SITE_GS031	0d 18'4"s ; 91	Upwelling, Fernan	Galapagos Islands	Ecuador	2:43 PM	2/10/2004	Coastal upwelli	12 m	19.6 m	18.6 C		0.35 mg/M3 (sampl 0
28	0.1 - 0.8	293	JCVI_SITE_GS031	0d 18'4"									18.6 C		0.35 mg/M3 (sampl 0
29	0.1 - 0.8	198	JCVI_SITE_GS031	0d 18'4"									18.6 C		0.35 mg/M3 (sampl 0
30	0.1 - 0.8	241	JCVI_SITE_GS031	0d 18'4"									18.6 C		0.35 mg/M3 (sampl 0
31	0.1 - 0.8	226	JCVI_SITE_GS031	0d 18'4"									18.6 C		0.35 mg/M3 (sampl 0
32	0.1 - 0.8	159	JCVI_SITE_GS031	0d 18'4"									18.6 C		0.35 mg/M3 (sampl 0
33	0.1 - 0.8	262	JCVI_SITE_GS031	0d 18'4"									18.6 C		0.35 mg/M3 (sampl 0
34	0.1 - 0.8	156	JCVI_SITE_GS031	0d 18'4"									18.6 C		0.35 mg/M3 (sampl 0
35	0.1 - 0.8	294	JCVI_SITE_GS031	0d 18'4"									18.6 C		0.35 mg/M3 (sampl 0
36	0.1 - 0.8	279	JCVI_SITE_GS031	0d 18'4"									18.6 C		0.35 mg/M3 (sampl 0
37	0.1 - 0.8	327	JCVI_SITE_GS031	0d 18'4"									18.6 C		0.35 mg/M3 (sampl 0
38	0.1 - 0.8	356	JCVI_SITE_GS031	0d 18'4"									18.6 C		0.35 mg/M3 (sampl 0
39	0.1 - 0.8	307	JCVI_SITE_GS031	0d 18'4"									18.6 C		0.35 mg/M3 (sampl 0
40	0.1 - 0.8	258	JCVI_SITE_GS031	0d 18'4"									18.6 C		0.35 mg/M3 (sampl 0
41	0.1 - 0.8	145	JCVI_SITE_GS031	0d 18'4"									18.6 C		0.35 mg/M3 (sampl 0
42	0.1 - 0.8	338	JCVI_SITE_GS051	15d 8'37									27.3 C	34.2 ppt	
43	0.1 - 0.8	349	JCVI_SITE_GS051	15d 8'37									27.3 C	34.2 ppt	
44	0.1 - 0.8	296	JCVI_SITE_GS005	44d 41'2									15 C	30.2 ppt	6 mg/M3 (sample n 6
45	0.1 - 0.8	337	JCVI_SITE_GS005	44d 41'2									15 C	30.2 ppt	6 mg/M3 (sample n 6
46	0.1 - 0.8	300	JCVI_SITE_GS011	39d 25'4									11 C		4.8 mg/M3 (sample 9
47	0.1 - 0.8	145	JCVI_SITE_GS011	39d 25'4									11 C		4.8 mg/M3 (sample 9
48	0.1 - 0.8	367	JCVI_SITE_GS006	45d 6'42									11.2 C		2.8 mg/M3 (sample 1
49	0.1 - 0.8	306	JCVI_SITE_GS006	45d 6'42									11.2 C		2.8 mg/M3 (sample 1
50	0.1 - 0.8	315	JCVI_SITE_GS006	45d 6'42									11.2 C		2.8 mg/M3 (sample 1
51	0.1 - 0.8	309	JCVI_SITE_GS033	1d 13'42									37.6 C	63.4 ppt	

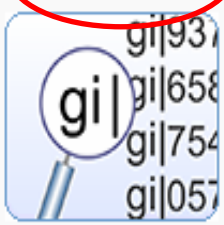


camera >> research
Marine Microbial Ecology

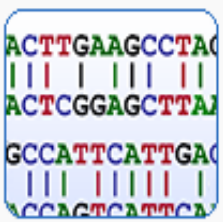
Research Home

? help

Search



BLAST

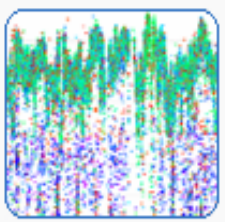


Projects

Publications

Samples

Fragment Recruitment



Recent Searches [hide]

- lipase no matches
- glycerol kinase (about 4,239 matches)
- Cellulase (about 3,163 matches)
- ATP synthase epsilon (about 2,856 matches)
- dehalogenase (about 4,963 matches)

Recent BLAST Results [hide]

Job Name	Submit Date	Status	# Hits	Subject Sequences
test11	02/27/07 02:09 PM	completed	2	GOS: Assembled Sequences (N)
tst10	02/27/07 02:09 PM	completed	0	GOS: Combined Assembly ORF Peptides (P)
test09	02/27/07 02:08 PM	completed	1	GOS: All Metagenomic Sequence Reads (N)
test06	02/27/07 02:05 PM	completed	50	All Metagenomic Sequence Reads (N)

View All BLAST Results

Search

search help

Keywords: dehalogenase Search

Examples | Find all words | Find any word | Find phrase

Search All	Accessions	Proteins	Clusters	Publications	Projects	Samples	Website
		6301	69	0	0	0	

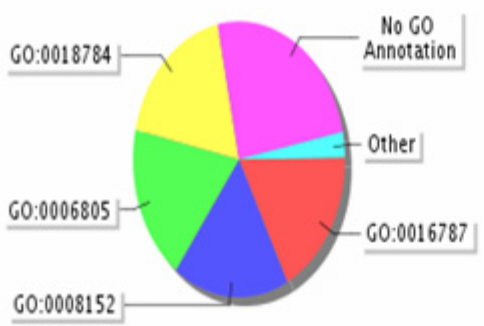
Search Results

Proteins

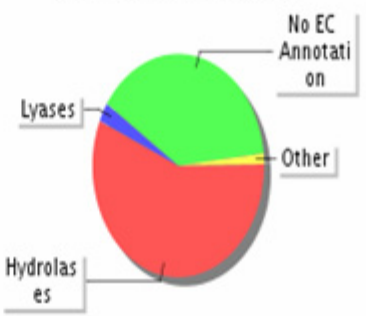
6301

[detail view]

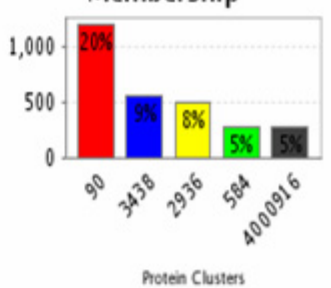
Protein search hits per Gene Ontology



Protein EC Summary

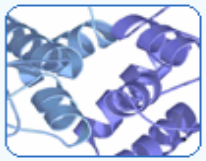


Protein Cluster Membership



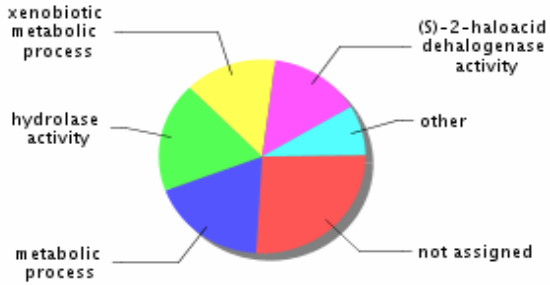
Protein Search Details — [hide] [back to full search results]

Proteins

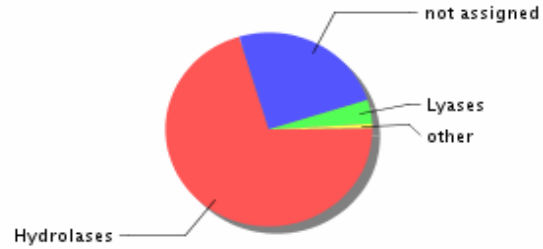


4915 matches

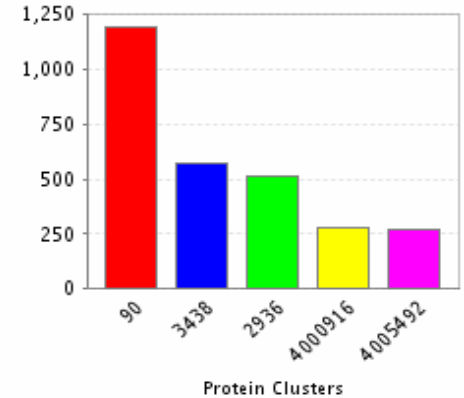
Gene Ontology Summary



Enzyme Summary



Protein Cluster Membership



All Matching Proteins

1 - 20 of 4915 Advanced Sort prev 20 | goto | next 20

Accession	External Acc	NCBI GI	Headline	DefLine	Length	Final
JCVI_PEP_1096665714389			... Accession: CAM_CNAM_102139 c...	JCVI_PEP_1096665714389 /source_dna_id=JCVI_ORF_...	147	CAM_C
JCVI_PEP_1096666313063			... Accession: CAM_CNAM_102139 c...	JCVI_PEP_1096666313063 /source_dna_id=JCVI_ORF_...	147	CAM_C
JCVI_PEP_1096668602421			... Accession: CAM_CNAM_102139 c...	JCVI_PEP_1096668602421 /source_dna_id=JCVI_ORF_...	155	CAM_C
JCVI_PEP_1096673421409			... Accession: CAM_CNAM_102139 c...	JCVI_PEP_1096673421409 /source_dna_id=JCVI_ORF_...	148	CAM_C
JCVI_PEP_1096676798831			... Accession: CAM_CNAM_102139 c...	JCVI_PEP_1096676798831 /source_dna_id=JCVI_ORF_...	90	CAM_C
JCVI_PEP_1096680464767			... Accession: CAM_CNAM_102139 c...	JCVI_PEP_1096680464767 /source_dna_id=JCVI_ORF_...	147	CAM_C
JCVI_PEP_1096680522357			... Accession: CAM_CNAM_102139 c...	JCVI_PEP_1096680522357 /source_dna_id=JCVI_ORF_...	160	CAM_C
JCVI_PEP_1096691663293			... Accession: CAM_CNAM_102139 c...	JCVI_PEP_1096691663293 /source_dna_id=JCVI_ORF_...	146	CAM_C
JCVI_PEP_1096692584989			... Accession: CAM_CNAM_102139 c...	JCVI_PEP_1096692584989 /source_dna_id=JCVI_ORF_...	100	CAM_C
JCVI_PEP_1096697966445			... Accession: CAM_CNAM_102139 c...	JCVI_PEP_1096697966445 /source_dna_id=JCVI_ORF_...	151	CAM_C
JCVI_PEP_1105173466959			... Accession: CAM_CNAM_102139 c...	JCVI_PEP_1105173466959 /read_id=HOT_READ_857589...	127	CAM_C
NCBI_PEP_37702690	AAR00932.1	37702690	... Accession: CAM_CNAM_102139 c...	>NCBI_PEP_37702690/ACCESSION=AAR00932.1/GI=3770...	150	CAM_C
JCVI_PEP_1096675606411			... Accession: CAM_CNAM_104264 c...	JCVI_PEP_1096675606411 /source_dna_id=JCVI_ORF_...	132	CAM_C
NCBI_PEP_111022131	YP_705103.1	111022131	... Accession: CAM_CNAM_104264 c...	NCBI_PEP_111022131 /offset=0 /translation_table...	276	CAM_C
JCVI_PEP_1096665787299			... Accession: CAM_CNAM_104669 c...	JCVI_PEP_1096665787299 /source_dna_id=JCVI_ORF_...	74	CAM_C
JCVI_PEP_1096670067379			... Accession: CAM_CNAM_104669 c...	JCVI_PEP_1096670067379 /source_dna_id=JCVI_ORF_...	71	CAM_C
JCVI_PEP_1096673037697			... Accession: CAM_CNAM_104669 c...	JCVI_PEP_1096673037697 /source_dna_id=JCVI_ORF_...	185	CAM_C
JCVI_PEP_1096674592091			... Accession: CAM_CNAM_104669 c...	JCVI_PEP_1096674592091 /source_dna_id=JCVI_ORF_...	175	CAM_C
JCVI_PEP_1096676803799			... Accession: CAM_CNAM_104669 c...	JCVI_PEP_1096676803799 /source_dna_id=JCVI_ORF_...	178	CAM_C
JCVI_PEP_1096688757069			... Accession: CAM_CNAM_104669 c...	JCVI_PEP_1096688757069 /source_dna_id=JCVI_ORF_...	121	CAM_C

Export

1 - 20 of 4915 **Show:** 10 20 50

Protein search results as CSV
Protein sequences as FASTA

Peptide Details

[hide] [back to search details] help

CAMERA Accession:	JCVI_PEP_1105173466959
Entity Type:	Protein
Sample:	HF10_10-07-02 - HOT station ALOHA, 10 m
ORF:	JCVI_ORF_1105173466958
Source DNA:	HOT_READ_85758924
Translation begin:	64
Translation end:	445
Orientation:	forward
Translation table:	11
5' Stop:	TAA
3' Stop:	TAA
Defline:	JCVI_PEP_1105173466959 /read_id=HOT_READ_85758924 /begin=64 /end=445 /orientation=1 /5_prime_stop=TAA /3_prime_stop=TAA /orf_id=JCVI_ORF_1105173466958 /ttable=11 /length=127 /ergatis_id=7702 /sample_id=JGI_SMPL_HF10_10-07-02 /sample_name=HF10_10-07-02 /number_of_sites=1 /site_id_1=HAWAII_SITE_HOT_10 /location_1="Hawaii Ocean Time-series station ALOHA, 10 m" /region_1="North Pacific Subtropical Gyre" /country_1=USA /site_depth_1="10 m" /biomass_1="7.21 +/-2.68 ug/kg" /chlorophyll_density_1="0.08 ug/kg" /dissolved_inorganic_carbon_1="1,967.60 umol/kg" /dissolved_inorganic_phosphate_1="41 nmol/kg" /dissolved_organic_carbon_1="78 umol/kg" /dissolved_oxygen_1="204.6 nmol/kg" /nitrate+nitrite_1="1 nmol/kg" /salinity_1="35.08 psu" /temperature_1="26.4 C" /water_depth_1="4750 m" /core_cluster_id=CAM_CRCL_4671 /final_cluster_id=CAM_CL_10705
Seq Length:	127
Seq Type:	AA
Sequence	IANDITNTHSKFTGANTFFAQVIFQKNPKNSHFMGGKLVKTKEIFLNGQIRAGRTSKVKKQLILGLRKILIKNTKLQSDFIWVYLEDLLPHQMIEYGEVLPKSGQ EKKWFNSLSSSLRKRLKKMEK*

Downloads

Download Protein sequence

Intellectual Property Notice

None

Metadata & Geography | Sequence Features | Annotations

Sample Metadata

[hide]

Sample Name:	HF10_10-07-02
Sample Accession:	JGI_SMPL_HF10_10-07-02
Site Id:	HAWAII_SITE_HOT_10
Filter Size:	0.22 - 1.6 µm

Site Metadata

[hide] [remove scroll]

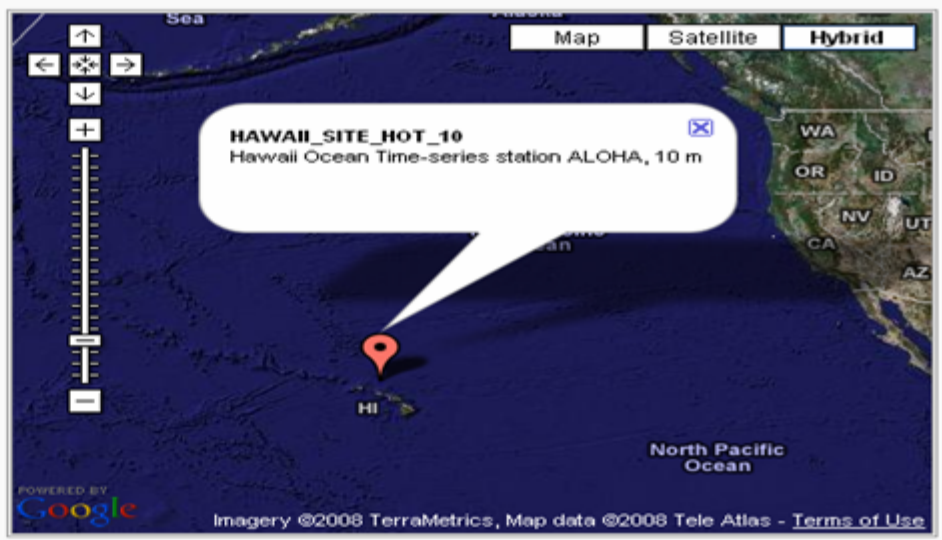
region:	North Pacific Subtropical Gyre
country:	USA
latitude:	22°45'n
longitude:	158°w
habitat:	upper euphotic
sample depth:	10 m
dissolved inorganic phosphate:	41 nmol/kg
nitrate+nitrite:	1 nmol/kg
water depth:	4750 m
biomass:	7.21 +/-2.68 ug/kg (+/-SE)
dissolved oxygen:	204.6 nmol/kg

Sample Data Download

HF10_10-07-02 16s, ALOHA/HOT, 10 m
HF10_10-07-02 Shotgun, ALOHA/HOT, 10 m

Sequence Geography

[hide]



>> 1 sample site is represented in this data set

camera Marine Microbial Ecology >> research

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

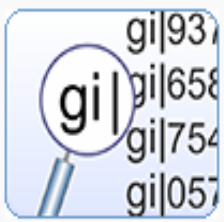
>> Research > Research Home

Research Home | Search | Tools | Projects & Data | Viewers

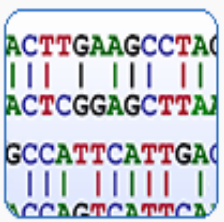
Research Home

? help

Search



BLAST

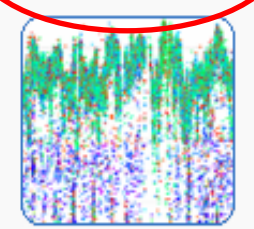


Projects

Publications

Samples

Fragment Recruitment



Recent Searches [hide]

- lipase no matches
- glycerol kinase (about 4,239 matches)
- Cellulase (about 3,163 matches)
- ATP synthase epsilon (about 2,856 matches)
- dehalogenase (about 4,963 matches)

Recent BLAST Results [hide]

Job Name	Submit Date	Status	# Hits	Subject Sequences
test11	02/27/07 02:09 PM	completed	2	GOS: Assembled Sequences (N)
tst10	02/27/07 02:09 PM	completed	0	GOS: Combined Assembly ORF Peptides (P)
test09	02/27/07 02:08 PM	completed	1	GOS: All Metagenomic Sequence Reads (N)
test06	02/27/07 02:05 PM	completed	50	All Metagenomic Sequence Reads (N)

View All BLAST Results

Controls [hide] [FRV help]

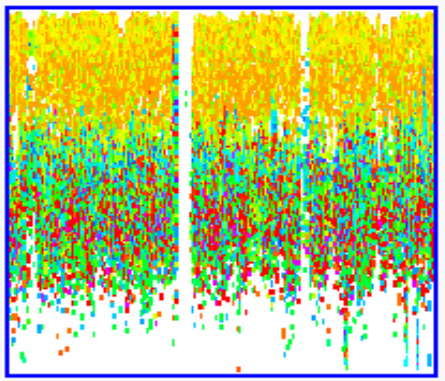
Data | Sample Filters | Mate Pair Filters | Annotation Filters

Query Sequence: Candidatus Pelagibacter ubique HTCC1062, complete genome. Export

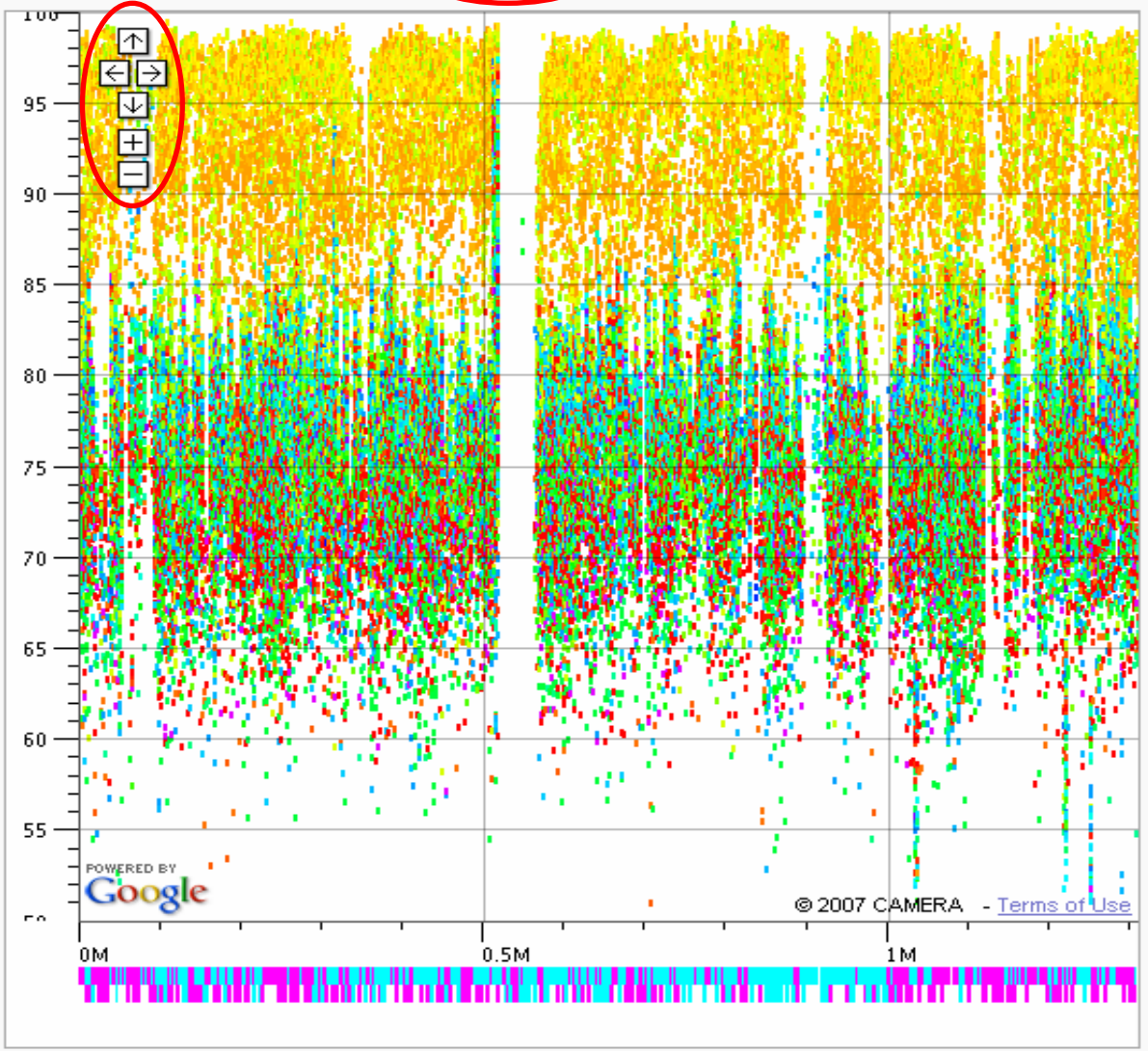
Subject Sequence: All Metagenomic Sequence Reads (N)

Submit | Reset | Save My Work

Overview [hide]



Fragment Recruitment Viewer [show legend] [export region] [select annotations]



Controls [hide] [FRV help]

Data | Sample Filters | Mate Pair Filters | Annotation Filters

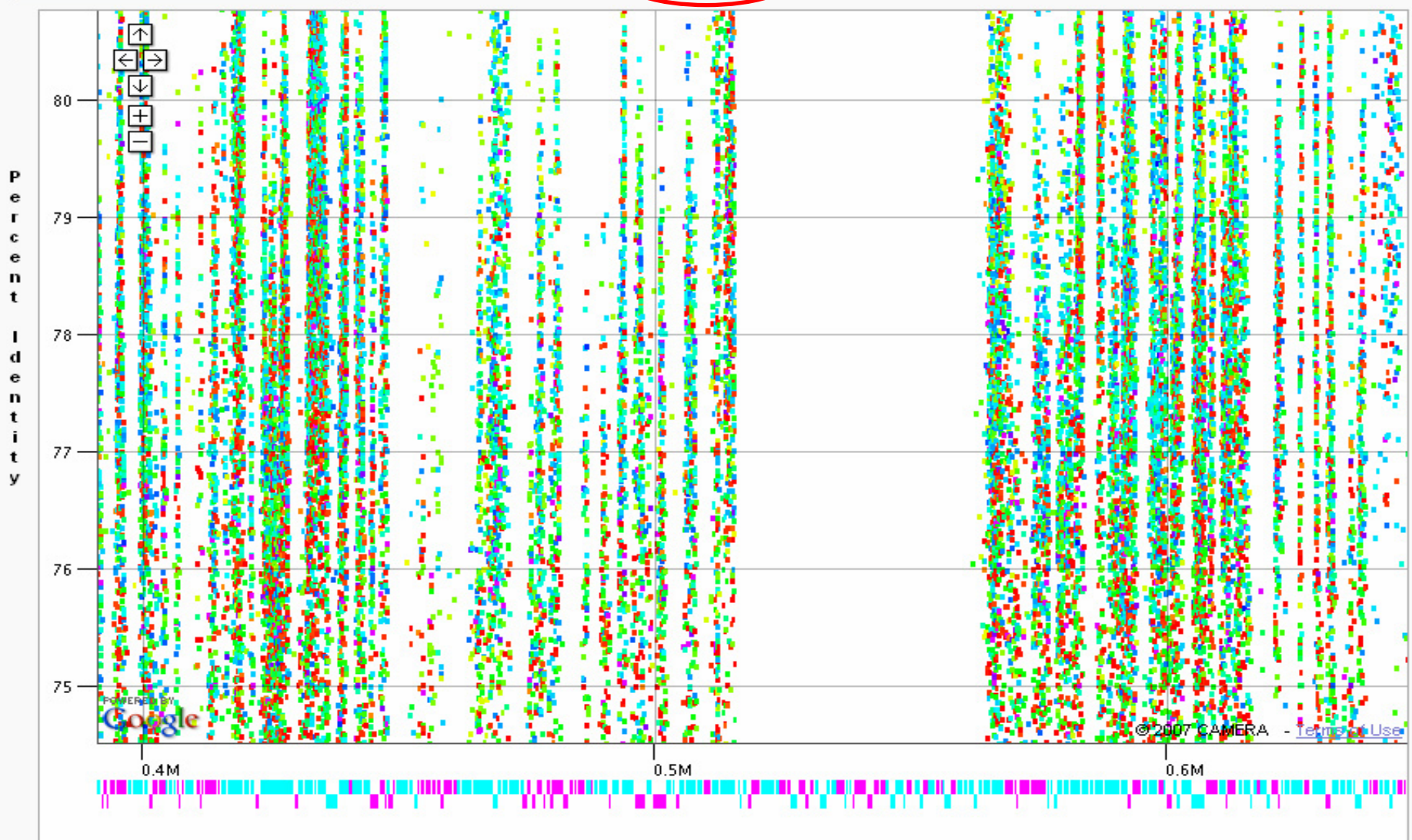
Query Sequence: Candidatus Pelagibacter ubique HTCC1062, complete genome. Export

Subject Sequence: All Metagenomic Sequence Reads (N)

Submit | Reset | Save My Work

Overview [show]

Fragment Recruitment Viewer [show legend] [export region] [select annotations]



Select Annotation Range

Region

Base Pair range: [min] 512,192 - 569,067 [max]

Filter: (None)
Annotations:
1 - 50 of 61

◄ prev 50 | goto | next 50 ►

Begin ▲	End	Strand	Length	GI Number	Protein ID	Product
512931	513007	+	76	GeneID:3517178		tRNA-Ile
513017	513092	+	75	GeneID:3517179		tRNA-Ala
513101	515997	+	2896	GeneID:3517180		23S ribosomal RNA
516442	516975	+	533	GeneID:3517181	YP_265947.1	glucose-6-phosphate isomerase-like protein
516956	517402	+	446	GeneID:3517182	YP_265948.1	hypothetical protein
517419	518273	+	854	GeneID:3517183	YP_265949.1	NAD dependent epimerase/dehydratase family
518270	519133	-	863	GeneID:3517184	YP_265950.1	NAD dependent epimerase/dehydratase family
519133	520869	-	1736	GeneID:3517185	YP_265951.1	acetolactate synthase II large chain
520922	521581	+	659	GeneID:3517186	YP_265952.1	Protein-L-isoaspartate(D-aspartate) O-methyltransferase (PCMT) family
521587	522627	+	1040	GeneID:3517152	YP_265953.1	zinc-containing alcohol dehydrogenase
522650	523831	+	1181	GeneID:3517153	YP_265954.1	cyclopropane-fatty-acyl-phospholipid synthase (CMAS family)
523831	524610	+	779	GeneID:3517154	YP_265955.1	FabG-like protein
524613	525977	-	1364	GeneID:3517155	YP_265956.1	FAD oxidase family
526036	526809	-	773	GeneID:3517156	YP_265957.1	methyltransferase
526790	528139	+	1349	GeneID:3517157	YP_265958.1	aminotransferase
528210	529664	-	1454	GeneID:3517158	YP_265959.1	deoxynucleotidase
529658	530587	-	929	GeneID:3517159	YP_265960.1	transketolase family
530587	531378	-	791	GeneID:3517160	YP_265961.1	transketolase
531418	532356	+	938	GeneID:3517161	YP_265962.1	UDP-glucose 4 epimerase
532349	533857	+	1508	GeneID:3517162	YP_265963.1	cytidylyltransferase/carbohydrate (or pyrimidine) kinase
533858	534580	+	722	GeneID:3517163	YP_265964.1	spsA-like glycosyl transferase
534588	535577	+	989	GeneID:3517164	YP_265965.1	dTDPglucose 4,6-dehydratase
535581	537269	-	1688	GeneID:3517165	YP_265966.1	unknown membrane protein
537246	537701	+	455	GeneID:3517166	YP_265967.1	unknown membrane protein
537702	538163	+	461	GeneID:3517167	YP_265968.1	unknown membrane protein
538166	538732	+	566	GeneID:3517168	YP_265969.1	phosphoheptose isomerase
538725	539216	-	491	GeneID:3517169	YP_265970.1	D-glycero-D-manno-heptose 1,7 bisphosphate phosphatase

Data | **Sample Filters** | Mate Pair Filters | Annotation Filters

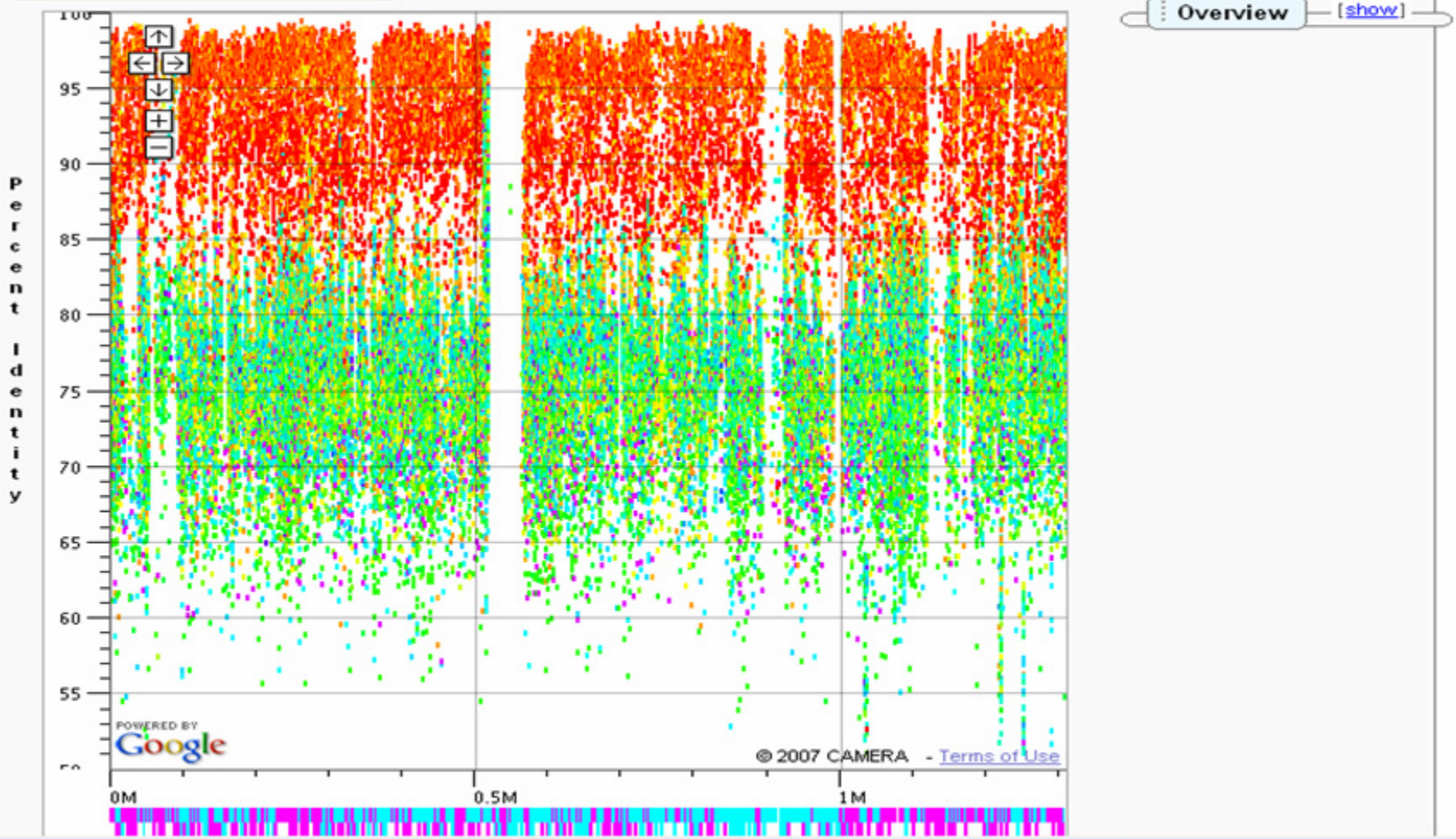
>> **Global Ocean Sampling Expedition** [select all] [select none]

☐ GS000a - Sargasso Stations 11...	☑ GS010 - Cape May, NJ	☑ GS026 - 134 miles NE of Galap...	☑ GS041 - Tropical South Pacific
☐ GS000b - Sargasso Stations 11...	☑ GS011 - Delaware Bay, NJ	☑ GS027 - Devil's Crown, Florea...	☑ GS042 - Tropical South Pacific
☐ GS000c - Sargasso Station 3	☑ GS012 - Chesapeake Bay, MD	☑ GS028 - Coastal Floreana	☑ GS043 - Tropical South Pacific
☐ GS000d - Sargasso Station 13	☑ GS013 - Off Nags Head, NC	☑ GS029 - North James Bay, Sant...	☑ GS044 - 600 miles from F. Pol...
☐ GS001a - Hydrostation S	☑ GS014 - South of Charleston, SC	☑ GS030 - Warm seep, Roca Redonda	☑ GS045 - 400 miles from F. Pol...
☐ GS001b - Hydrostation S	☑ GS015 - Off Key West, FL	☑ GS031 - Upwelling, Fernandina...	☑ GS046 - 300 miles from F. Pol...
☐ GS001c - Hydrostation S	☑ GS016 - Gulf of Mexico	☑ GS032 - Mangrove on Isabella ...	☑ GS047 - 201 miles from F. Pol...
☑ GS002 - Gulf of Maine	☑ GS017 - Yucatan Channel	☑ GS033 - Punta Cormorant, Hype...	☑ GS048 - Moorea, Cooks Bay
☑ GS003 - Browns Bank, Gulf of ...	☑ GS018 - Rosario Bank	☑ GS034 - North Seamore Island	☑ GS049 - Moorea, Outside Cooks...
☑ GS004 - Outside Halifax, Nova...	☑ GS019 - Northeast of Colon	☑ GS035 - Wolf Island	☑ GS050 - Tikehau Lagoon

Submit | Reset | Save My Work

Fragment Recruitment Viewer [show legend] [export region] [select annotations]

Overview [show]



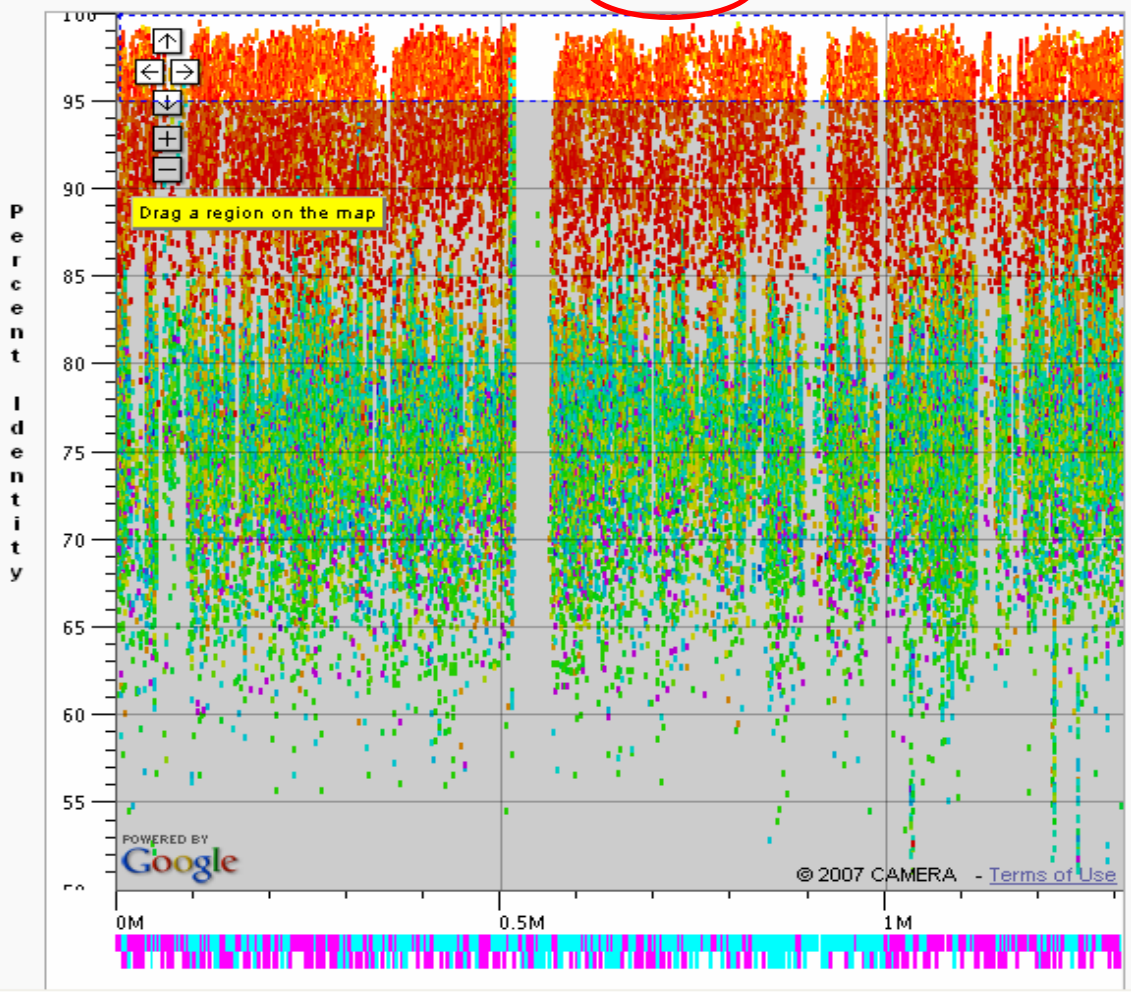
Data | **Sample Filters** | Mate Pair Filters | Annotation Filters

>> **Global Ocean Sampling Expedition** [select all] [select none]

☐ GS000a - Sargasso Stations 11...	☑ GS010 - Cape May, NJ	☑ GS026 - 134 miles NE of Galap...	☑ GS041 - Tropical South Pacific
☐ GS000b - Sargasso Stations 11...	☑ GS011 - Delaware Bay, NJ	☑ GS027 - Devil's Crown, Florea...	☑ GS042 - Tropical South Pacific
☐ GS000c - Sargasso Station 3	☑ GS012 - Chesapeake Bay, MD	☑ GS028 - Coastal Floreana	☑ GS043 - Tropical South Pacific
☐ GS000d - Sargasso Station 13	☑ GS013 - Off Nags Head, NC	☑ GS029 - North James Bay, Sant...	☑ GS044 - 600 miles from F. Pol...
☐ GS001a - Hydrostation S	☑ GS014 - South of Charleston, SC	☑ GS030 - Warm seep, Roca Redonda	☑ GS045 - 400 miles from F. Pol...
☐ GS001b - Hydrostation S	☑ GS015 - Off Key West, FL	☑ GS031 - Upwelling, Fernandina...	☑ GS046 - 300 miles from F. Pol...
☐ GS001c - Hydrostation S	☑ GS016 - Gulf of Mexico	☑ GS032 - Mangrove on Isabella ...	☑ GS047 - 201 miles from F. Pol...
☑ GS002 - Gulf of Maine	☑ GS017 - Yucatan Channel	☑ GS033 - Punta Cormorant, Hype...	☑ GS048 - Moorea, Cooks Bay
☑ GS003 - Browns Bank, Gulf of ...	☑ GS018 - Rosario Bank	☑ GS034 - North Seamore Island	☑ GS049 - Moorea, Outside Cooks...
☑ GS004 - Outside Halifax, Nova...	☑ GS019 - Northeast of Colon	☑ GS035 - Wolf Island	☑ GS050 - Tikehau Lagoon

Submit | Reset | Save My Work

Fragment Recruitment Viewer [show legend] [export region] [select annotations]



Overview [show]

Export Selected Sequences

Region

Base Pair range:	[min]	2,556	- 7,668	[max]
Percent Identity range:	[min]	95.12	- 99.9	[max]

Export | Cancel

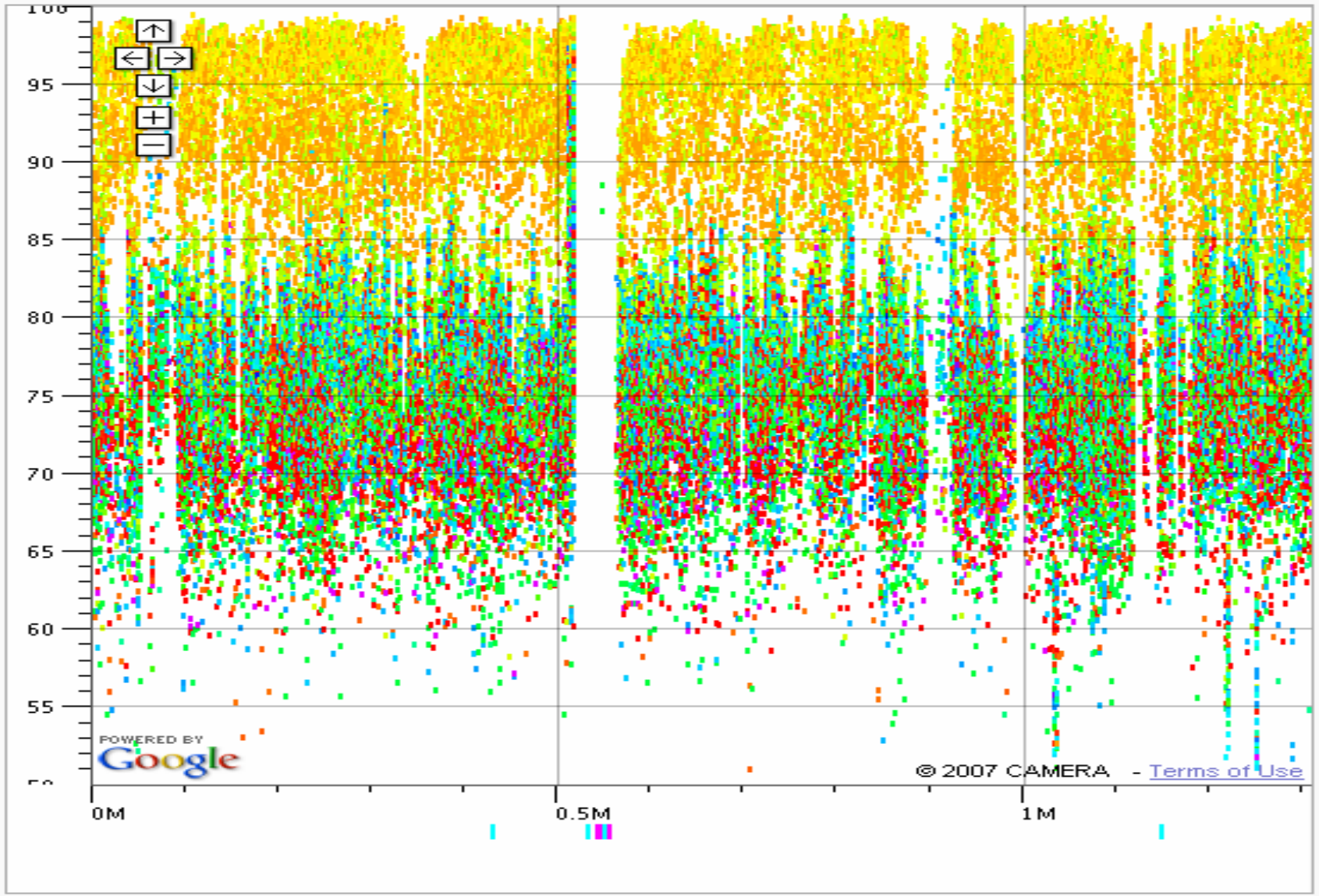
Controls [hide] [FRV help]

Data | Sample Filters | Mate Pair Filters | **Annotation Filters**

Show annotations whose product have text: glycosyl transferase [clear]

Submit | Reset | Save My Work

Fragment Recruitment Viewer [show legend] [export region] [select annotations]



CAMERA
http://camera.calit2.net

- Website: http://camera.calit2.net
- Questions: camera-info@calit2.net

- Saul A. Kravitz skravitz@jcvi.org
- Rekha Seshadri rseshadri@jcvi.org