



JOINT GENOME INSTITUTE
UNITED STATES DEPARTMENT OF ENERGY



<https://gold.jgi.doe.gov/>

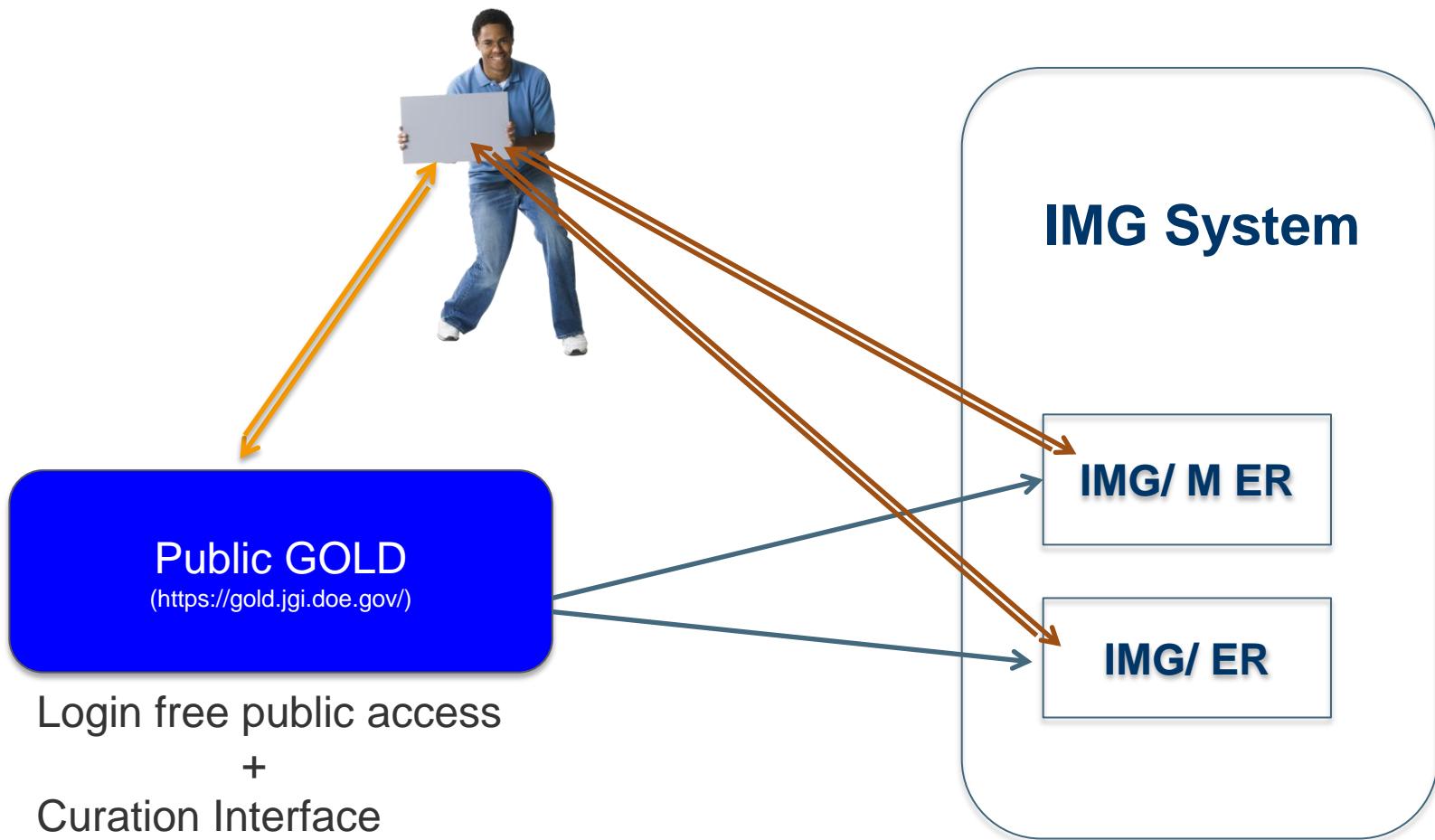
Genomes on Line Database **(GOLD)**

Supratim Mukherjee
Genomic Standards Group
supratimmukherjee@lbl.gov

What is ahead?

- How GOLD and IMG systems are related
- Introduction to GOLD
- GOLD terminology
- Four level GOLD classification/organization system
- A stroll through public GOLD & some of its features
- Entering a Sequencing Project in GOLD
 - Defining Study -> Organism -> Sequencing Project -> Analysis Project
- Standardized Metagenome Nomenclature
 - Case study: Creating a combined assembly Analysis Project

The Big Picture (from a user's perspective)



IMG ER- For submitting genome projects to IMG pipeline

IMG M ER – For submitting metagenome projects IMG pipeline

The Genomes OnLine Database (GOLD) is a centralized resource for the continuous monitoring of sequencing projects from around the world, along with their associated metadata.

Two Aspects of GOLD:

1. Login free public GOLD: Currently holds more than **97000** public projects, whose metadata can be viewed by the general public at <https://gold.jgi.doe.gov>

2. Data can be entered through the GOLD curation interface by users. Both new data submission and edits are supported through a secure login based system.
 - Private projects in GOLD can NOT be viewed by the general public.

 - Even if your projects are public in GOLD, any sequence data you submit to IMG will remain private subject to IMG's data release policies.

Key GOLD Terms



Study: Study is an umbrella Project and broadly represents the overall goal of a research proposal that a researcher sets out to explore. **E.g.** HMP study, GEBA study.

Biosample: The description of the environment from where the sample for a sequencing project was collected. Biosample is nothing more than metadata on the Sequencing Project. **E.g.** *Groundwater sediment microbial communities from Colorado, USA*. **A GOLD Biosample is applicable for metagenome/metatranscriptome projects only.**

Organism: An organism corresponds to any living biological material (bacteria, virus, fungi, plant, animal) that is associated to a Sequencing Project. **Applicable for isolate genome projects only.**

Sequencing Project: Sequencing Project represents individual sequencing deliverables from an Organism or Biosample that is targeted for sequencing. Whole Genome Sequencing, Metagenome, Transcriptome, 16S survey etc.

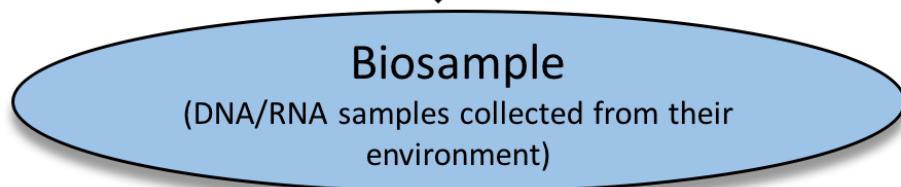
Analysis Project: Analysis Project is the informatics processing of a Sequencing Project. It describes how the assembly and annotation of a Sequencing Project were performed. **Each submission to IMG represents an individual Analysis Project.**

GOLD Overview

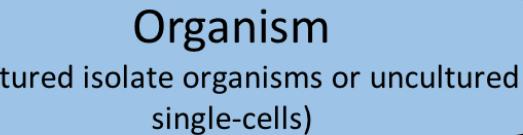
Gs

Study

Gb



Gp



Go

Sequencing Projects

- 1. Metagenome (MetaG)
- 2. Metatranscriptome (MetaT)
- 3. 16S Survey

Analysis Projects

- 1. Metagenome
- 1a. Combined Assembly (MetaG)
- 1b. Genome from Metagenome
- 2. Metatranscriptome
- 2a. Combined Assembly (MetaT)
- 3. Metagenome (iTag)

Sequencing Projects

- 1. Isolate Whole Genome Sequence
- 2. Isolate Transcriptome
- 3. Single-cell Whole Genome Sequence
- 4. Others

Analysis Projects

- 1. Isolate Genome
- 2. Transcriptome
- 3a. Single-cell screened
- 3b. Single cell unscreened
- 4. Other AP types

Analysis Projects

- Combined Assembly
(Metagenome & Single-cell genome)

GOLD Home Page

<https://gold.jgi.doe.gov/>

Home Search Distribution Graphs Biogeographical Metadata Statistics References Team Help News

Studies	26,121
Biosamples	15,883
Sequencing Projects	97,209
Analysis Projects	79,035
Organisms	239,105

[Download Excel Data file](#)

File last generated: 21 Sep, 2016

Welcome to the Genomes OnLine Database

GOLD Release v.5

GOLD: Genomes Online Database, is a World Wide Web resource for comprehensive access to information regarding genome and metagenome sequencing projects, and their associated metadata, around the world.

1. Register



Register your project information and Metadata in the Genomes Online Database

[Register](#)

2. Annotate



Annotate your microbial genome or metagenome with IMG/ER or IMG/MER

[Annotate](#)

3. Publish



Standards in Genomic Sciences

Publish your genome or metagenome in open access standards-supportive journal.

[Publish](#)

Studies

Metagenomic [965](#)
Non-Metagenomic [25,156](#)

Biosamples

Classification
Ecosystems
Host-associated [5,665](#)
Engineered [2,750](#)
Environmental [7,492](#)

Sequencing Projects

Complete Projects [9,050](#)
 Permanent Drafts [42,673](#)
 Incomplete Projects [43,064](#)
 Targeted Projects [1,462](#)

Analysis Projects

Genome Analysis [56,396](#)
Metagenome Analysis [11,272](#)
Combined Assembly [108](#)
Genome from Metagenome [1,713](#)
Metatranscriptome Analysis [1,683](#)
Single Cell (Screened) [1,185](#)
Single Cell (Unscreened) [838](#)
Transcriptome Analysis [12](#)

Organisms

Organisms [239,105](#)
Archaea [1,997](#)
Bacteria [218,766](#)
Eukarya [13,699](#)
Viruses [4,614](#)

Special Projects

Type Strain Projects [5,596](#)
GEBA Projects [2,776](#)
HMP Projects [2,916](#)

JGI Projects

JGI Studies [1,135](#)
JGI Biosamples [6,851](#)
JGI Sequencing Projects [33,831](#)
JGI Analysis Projects [20,301](#)

Projects with Genbank Data

Seq. Projects [52,740](#)
Archaeal Projects [708](#)
Bacterial Projects [44,549](#)

Navigating GOLD

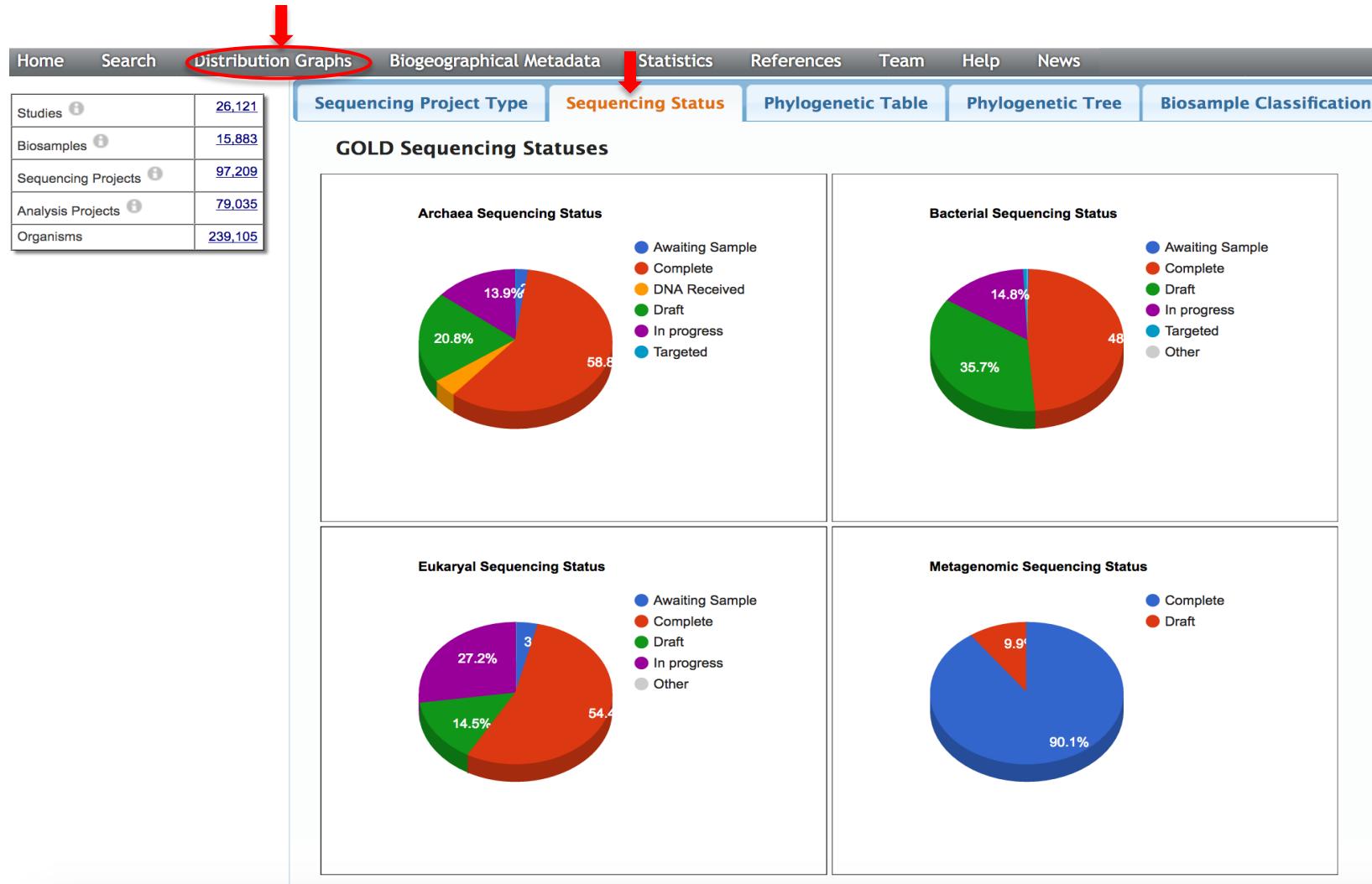
<https://gold.jgi.doe.gov/>

The screenshot shows the GOLD (Genomes Online Database) homepage. Key features highlighted include:

- Header Navigation:** JGI logo, GOLD logo, "GENOMES ON-LINE DATABASE", "JGI HOME", "LOG IN", "Home", "Search", "Distribution Graphs", "Biogeographical Metadata", "Statistics", "References", "Team", "Help", "News".
- Data Summary Table:** Shows counts for Studies (26,121), Biosamples (15,883), Sequencing Projects (97,209), Analysis Projects (79,035), and Organisms (239,105). A red box highlights this table.
- Welcome Message:** "Welcome to the Genomes OnLine Database".
- GOLD Release v.5:** Announced on the right.
- Three-Step Workflow:** 1. Register (with GOLD logo), 2. Annotate (with database icon), 3. Publish (with SIGS logo).
- Project Categories:** Studies, Biosamples, Sequencing Projects, Analysis Projects, Organisms, Special Projects, JGI Projects, and Projects with Genbank Data. Each category has a red circle around its title.
- Download Options:** "Download Excel Data file" (File last generated: 21 Sep, 2016).

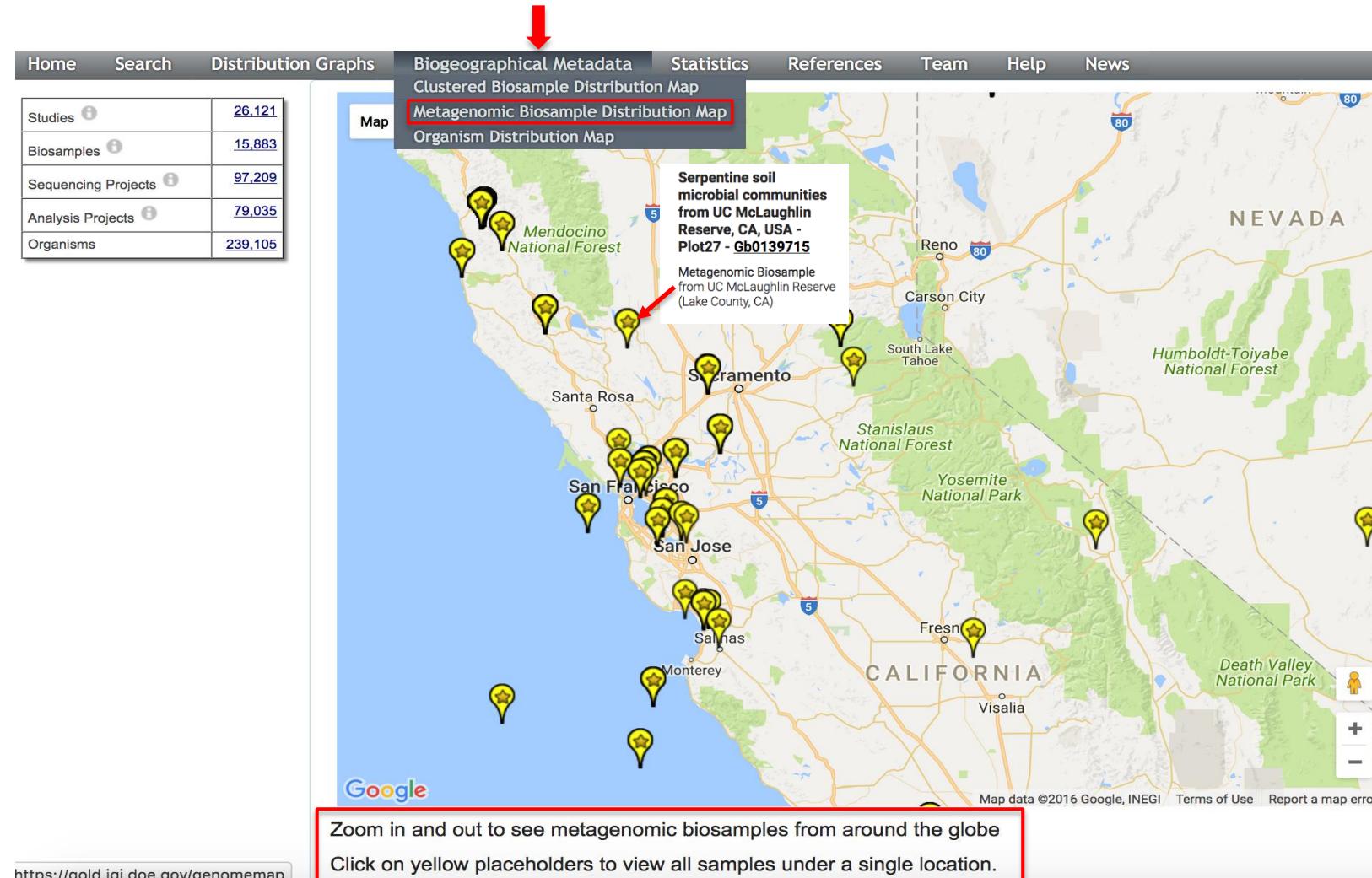
Navigating GOLD: Distribution Graphs

- Data summary of different types of Sequencing Projects, Sequencing Status, Organism Phylogenetic Classification, Biosample Ecosystem Classifications etc., are provided as pre-computed pie charts and tables



Navigating GOLD: Biogeographical Metadata

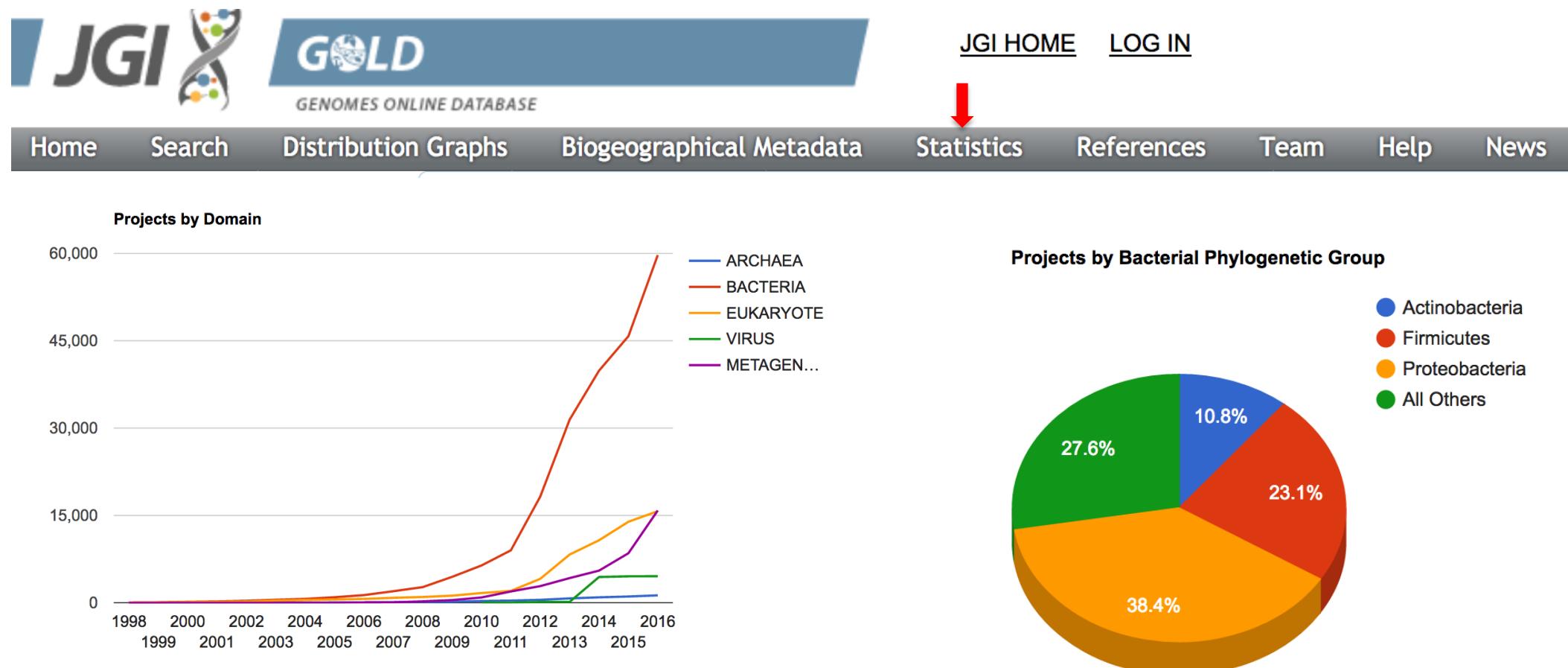
- The Biogeographical Metadata section displays the geographic location of GOLD Biosamples and Organisms using the map and terrain components of Google map.



The interactive maps in this segment can be zoomed in or out to focus on a particular area in the map and search for Biosamples from that region.

Navigating GOLD: A Few Stats..

- Precomputed graphs and charts such as No. of Projects by year and Domain, No. of genomes by year, Phylogenetic Distribution of Bacterial Genome Projects, Distribution of Projects by Sequencing Centers etc.



Studies	26,121
Biosamples	15,883
Sequencing Projects	97,209
Analysis Projects	79,035
Organisms	239,105

[Download Excel Data file](#)

File last generated: 21 Sep, 2016

Welcome to the Genomes OnLine

GOLD: Genomes Online Database, is a Web-based resource for sharing genomic and metagenomic data and associated metadata, around the world.

1. Register



Register your project information and
Genomes Online Database

[Register](#)
Studies

 Metagenomic [965](#)

 Non-Metagenomic [25,156](#)
Organisms

 Organisms [239,105](#)

 Archaea [1,997](#)

 Bacteria [218,766](#)

 Eukarya [13,699](#)

 Viruses [4,614](#)

Select Columns for Table

GOLD Study ID	Study Name	Add Date
Gs0121530	Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics	2016-09-09
Gs0121491	Freshwater microbial communities from high-altitude lakes in Yosemite National Park, California, USA	2016-08-24
Gs0121435	Terrestrial oil reservoir microbial communities from Alaska	2016-08-02
Gs0121351	Human host and surrounding environment bacterial communities from Fiji	2016-07-09
Gs0121350	Ectoparasite and host associated microbial communities from various locations - Smithsonian Institution, USA	2016-07-09
Gs0121349	Sediment and sludge enrichment culture microbial communities from Biogas Institute, Beijing, China	2016-07-09
Gs0121348	Pacific white shrimp tissue microbial communities imported from various locations to Iowa, USA	2016-07-09
Gs0121347	Human stomach microbial communities from New York City, USA - Weill Cornell Medical College	2016-07-09
Gs0121346	Lettuce associated viral communities from produce distribution center in Michigan, USA	2016-07-09
Gs0121345	Mouse gut microbial communities before and after antibiotic treatment from Singapore	2016-07-09
Gs0121344	Human fecal microbial communities from Human Longevity Inc., USA	2016-07-09
Gs0121343	Human fecal and colon tissue microbial communities from Howard University Cancer Center, USA	2016-07-09
Gs0121342	Nasal microbial communities from cattle with respiratory disease in Kansas and Nebraska, USA	2016-07-09
Gs0121341	Human sputum microbial communities from cystic fibrosis patients in Italy	2016-07-09
Gs0121340	Nursing home wastewater effluent microbial communities near Freiburg, Germany	2016-07-09
Gs0121339	Ancient human oral microbial communities from various locations	2016-07-09
Gs0121338	Marine microbial communities from Great Barrier Reef, Australia to study anthropogenic effects	2016-07-09
Gs0121337	Uranium tailings and control soil microbial communities from Southeast China	2016-07-09
Gs0121336	Activated sewage sludge microbial communities from Geneva, Switzerland	2016-07-09
Gs0121335	Soil fungal communities from apple orchard in Beijing, China	2016-07-09
Gs0121334	Maize phyllosphere microbial communities from Texas and California to study drought stress	2016-07-09

GOLD Release v.5
provides access to over 239,000 organisms, and their associated

genomic and metagenomic data in the sciences. It is a peer-reviewed journal.

Searching GOLD

GOLD
GENOMES ONLINE DATABASE

Home Search Distribution Graphs Biog...
Studies 2,221
Biosamples 1,883
Sequencing Projects 97,209
Analysis Projects 79,035
Organisms 239,105

[Download Excel Data file](#)
File last generated: 21 Sep, 2016

Welcome to GOLD: Genomes Online Database, around metadata, around their associated

Register

Studies
Metagenomic
Non-Metagenomic
Organisms
Archaea 1,951
Bacteria 218,700
Eukarya 13,699
Viruses 4,614

HMP Projects 2,916 JGI Sequencing Projects 33,631 JGI Analysis Projects 20,301 Bacterial Projects 44,549

JGI HOME LOG IN

<https://gold.jgi.doe.gov/>

Select Columns for Table			
GOLD Project ID	Project Name	Project Status	Add Date
Gp0156133	Escherichia coli MGM23_te	incomplete	2016-09-22
Gp0156106	Streptomyces agglomeratu	incomplete	2016-09-20
Gp0155985	Yangia sp. CCB-MM3	Complete and Published	2016-09-13 Open access
Gp0155984	Mouse cecal microbial communities from Vienna, Austria - N-Acetylglucosamine amended D2O_June	incomplete	2016-09-13
Gp0155983	Mouse cecal microbial communities from Vienna, Austria - N-Acetylneuramnic amended D2O_June	incomplete	2016-09-13
Gp0155982	Mouse cecal microbial communities from Vienna, Austria - N-Acetylneuramnic and N-Acetylglucosamine D2O microcosms_June	incomplete	2016-09-13
Gp0155981	Mouse cecal microbial communities from Vienna, Austria - N-Acetylglucosamine amended D2O_March	incomplete	2016-09-13
Gp0155980	Mouse cecal microbial communities from Vienna, Austria - N-Acetylneuramnic and N-Acetylglucosamine D2O microcosms_March	incomplete	2016-09-13
Gp0155979	Mouse cecal microbial communities from Vienna, Austria - N-Acetylglucosamine amended with D2O_GlcNAcF	incomplete	2016-09-13
Gp0155978	Mouse cecal microbial communities from Vienna, Austria - N-Acetylglucosamine amended with D2O_GlcNAcE	incomplete	2016-09-13
Gp0155977	Mouse cecal microbial communities from Vienna, Austria - N-Acetylglucosamine amended with D2O_GlcNAcC	incomplete	2016-09-13

Sort column: ▼ ▲
Search:

Searching GOLD

Q. What if you are interested in **Metagenome** projects from **Freshwater** environment, whose analysis was completed after **January 1, 2014**?

JGI **GOLD**
GENOMES ONLINE DATABASE

[JGI HOME](#) [LOG IN](#)

[Home](#) [Search](#) [Distribution Graphs](#) [Biogeographical Metadata](#) [Statistics](#) [References](#) [Team](#) [Help](#) [News](#)

[Advanced Search](#) [Metadata Search](#) (21)

Studies	15,883
Biosamples	15,883
Sequencing Projects	97,209
Analysis Projects	79,035
Organisms	239,105

[Download Excel Data file](#)
File last generated: 21 Sep, 2016

Welcome to the Genomes OnLine Database GOLD Release v.5

GOLD: Genomes Online Database, is a World Wide Web resource for comprehensive access to information regarding genome and metagenome sequencing projects, and their associated metadata, around the world.

1. Register  Register your project information and Metadata in the Genomes Online Database [Register](#)

2. Annotate  Annotate your microbial genome or metagenome with IMG/ER or IMG/MER [Annotate](#)

3. Publish  Standards in Genomic Sciences Publish your genome or metagenome in open access standards-supportive journal. [Publish](#)

Advanced Search



Studies	26,121
Biosamples	15,883
Sequencing Projects	97,209
Analysis Projects	79,035
Organisms	239,105

Advanced Search

Advanced Search allows you to search across different levels (Study, Biosample/Organisms, Projects and Analysis Projects) in GOLD.

For example using this advanced search wizard, you may select Complete and Published, Whole Genome Sequencing projects of Finished quality for Gram negative organisms with GenBank sequence data. To perform the above search you would select filters as shown below:

Organism.Gram Stain → Gram-
Project.Sequencing Strategy → Whole Genome Sequencing
Project.Project Status → Complete and Published
Project.Sequencing Quality → Level 6: Finished
Analysis Project.Genbank ID → true

Current Filters:: None Set

Choose Filters (Click on + to expand and select fields of interest for filtering)

+ Study Fields

+ Biosample Fields

+ Organism Fields

+ Project Fields

+ Analysis Project Fields

Biosample.Ecosystem → Environmental

Biosample.Ecosystem Category → Aquatic

Biosample.Ecosystem Type → Freshwater

Project.Sequencing Strategy → Metagenome

Analysis Project.Completion Date → >01-01-2014

Submit Search

Advanced Search: Results..

Your search results are below:

Studies	Biosamples	Organisms	Sequencing Projects	Analysis Projects
44	651	0	651	614

Current Filters::

- Biosample.Ecosystem Category → Aquatic [X](#)
Analysis Project.Completion Date → >01-01-2014 [X](#)
Biosample.Ecosystem → Environmental [X](#)
Biosample.Ecosystem Type → Freshwater [X](#)
Project.Sequencing Strategy → Metagenome [X](#)

[Clear All Filters](#)

Choose Filters (Click on + to expand and select fields of interest for filtering)

[+ Study Fields](#)

[+ Biosample Fields](#)

[+ Organism Fields](#)

[+ Project Fields](#)

[+ Analysis Project Fields](#)

[Submit Search](#)

Modifying Search..

Your search results are below:

Studies	Biosamples	Organisms	Sequencing Projects	Analysis Projects
44	651	0	651	614

Current Filters::

Biosample.Ecosystem Category → Aquatic  
Analysis Project.Completion Date → >01-01-2014  
Biosample.Ecosystem → Environmental 
Biosample.Ecosystem Type → Freshwater  
Project.Sequencing Strategy → Metagenome 

[Clear All Filters](#)

Choose Filters (Click on + to expand and select fields of interest for filtering)

- + Study Fields
- + Biosample Fields
- + Organism Fields
- + Project Fields
- + Analysis Project Fields

[Submit Search](#)

Your search results are below:

Studies	Biosamples	Organisms	Sequencing Projects	Analysis Projects
356	3,944	0	3,953	4,325

Current Filters::

Biosample.Ecosystem Category → Aquatic 
Biosample.Ecosystem → Environmental 
Project.Sequencing Strategy → Metagenome 

[Clear All Filters](#)

Choose Filters (Click on + to expand and select fields of interest for filtering)

- + Study Fields
- + Biosample Fields
- + Organism Fields
- + Project Fields
- + Analysis Project Fields

[Submit Search](#)

Exploring Further..

Your search results are below:

Studies	Biosamples	Organisms	Sequencing Projects	Analysis Projects
356	3,944	0	3,953	4,325

Current Filters::

Biosample.Ecosystem Category → Aquatic 
Biosample.Ecosystem → Environmental 
Project.Sequencing Strategy → Metagenome 

[Clear All Filters](#)

Choose Filters (Click on + to expand and select fields of interest for filtering)

[+ Study Fields](#)

[+ Biosample Fields](#)

[+ Organism Fields](#)

[+ Project Fields](#)

[+ Analysis Project Fields](#)

[Submit Search](#)

Your current search results are:

[Studies](#) [Biosamples](#) [Organisms](#) [Sequencing Projects](#) [Analysis Projects](#)

356 3,944 1 3,953 4,325

Current Filters::

Project.Sequencing Strategy → Metagenome 
Biosample.Ecosystem → Environmental
Biosample.Ecosystem Category → Aquatic

[Refine Search Filters](#)

[Clear All Filters](#)

[New Search](#)

[Select Columns for Table](#)

GOLD Biosample ID	Biosample Name	Add Date	Ecosystem	Ecosystem Category
Gb0143532	Groundwater sediment microbial communities from an aquifer in Colorado, USA - East MetaG ERMGT842_2	2016-09-13	Environmental	Aquatic
Gb0143531	Groundwater sediment microbial communities from an aquifer in Colorado, USA - East MetaG ERMGT828_2	2016-09-13	Environmental	Aquatic
Gb0143530	Groundwater sediment microbial communities from an aquifer in Colorado, USA - East MetaG ERMGT814_2	2016-09-13	Environmental	Aquatic
Gb0143529	Groundwater sediment microbial communities from an aquifer in Colorado, USA - East MetaG ERMGT800_2	2016-09-13	Environmental	Aquatic
Gb0143528	Groundwater sediment microbial communities from an aquifer in Colorado, USA - East MetaG ERMGT769_2	2016-09-13	Environmental	Aquatic
Gb0143527	Groundwater sediment microbial communities from an aquifer in Colorado, USA - East MetaG ERMGT746_2	2016-09-13	Environmental	Aquatic
Gb0143526	Groundwater sediment microbial communities from an aquifer in Colorado, USA - East MetaG ERMGT723_2	2016-09-13	Environmental	Aquatic
Gb0143525	Groundwater sediment microbial communities from an aquifer in Colorado, USA - East MetaG ERMGT700_2	2016-09-13	Environmental	Aquatic

Selecting More Columns..

Your current search results are:

Studies Biosamples Organisms Sequencing Projects Analysis Projects

356 3,944 1 3,953 4,325

Current Filters::
Project.Sequencing Strategy → Metagenome 
Biosample.Ecosystem → Environmental
Biosample.Ecosystem Category → Aquatic

[Refine Search Filters](#)

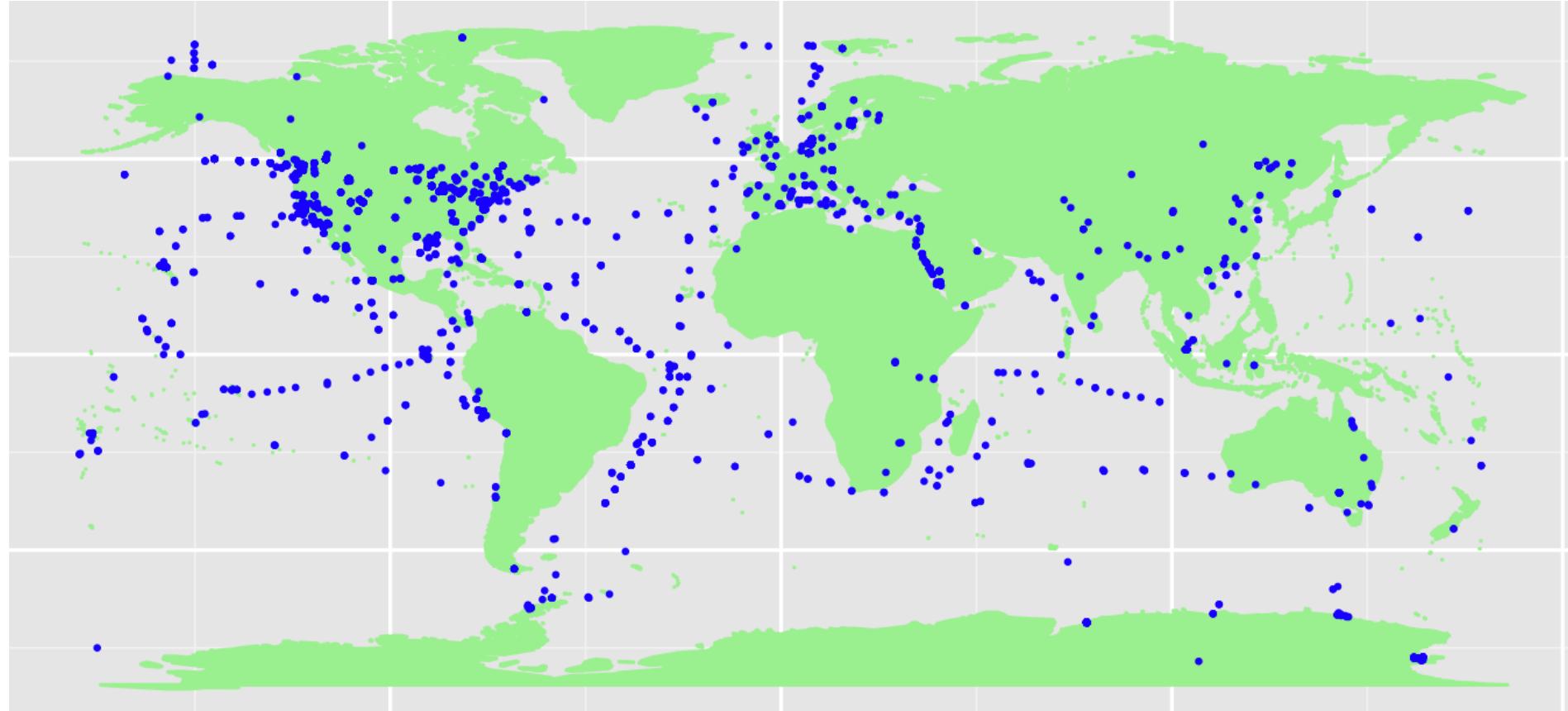
[Clear All Filters](#) [New Search](#)

[Select Columns for Table](#) 

GOLD Biosample ID	Biosample Name	Add Date	Ecosystem	Ecosystem Category	Geographic Location	Latitude	Longitude
Gb0143532	Groundwater sediment microbial communities from an aquifer in Colorado, USA - East MetaG ERMGT842_2	2016-09-13	Environmental	Aquatic	USA: Colorado	38.9779	-107.0020
Gb0143531	Groundwater sediment microbial communities from an aquifer in Colorado, USA - East MetaG ERMGT828_2	2016-09-13	Environmental	Aquatic	USA: Colorado	38.9778	-107.0022
Gb0143530	Groundwater sediment microbial communities from an aquifer in Colorado, USA - East MetaG ERMGT814_2	2016-09-13	Environmental	Aquatic	USA: Colorado	38.9777	-107.0023
Gb0143529	Groundwater sediment microbial communities from an aquifer in Colorado, USA - East MetaG ERMGT800_2	2016-09-13	Environmental	Aquatic	USA: Colorado	38.9776	-107.0024
Gb0143528	Groundwater sediment microbial communities from an aquifer in Colorado, USA - East MetaG ERMGT769_2	2016-09-13	Environmental	Aquatic	USA: Colorado	38.9779	-107.0023
Gb0143527	Groundwater sediment microbial communities from an aquifer in Colorado, USA - East MetaG ERMGT746_2	2016-09-13	Environmental	Aquatic	USA: Colorado	38.9778	-107.0025
Gb0143526	Groundwater sediment microbial communities from an aquifer in Colorado, USA - East MetaG ERMGT723_2	2016-09-13	Environmental	Aquatic	USA: Colorado	38.9777	-107.0027
Gb0143525	Groundwater sediment microbial communities from an aquifer in Colorado, USA - East MetaG ERMGT700_2	2016-09-13	Environmental	Aquatic	USA: Colorado	38.9775	-107.0029

Geographic Distribution of Aquatic Biosamples

(Figure generated in R)



Before Entering a Sequencing Project and/or Analysis Project in GOLD..



Home Search Distribution Graphs Biogeographical Metadata Statistics References Team **Help** News

Studies	26,121
Biosamples	15,883
Sequencing Projects	97,209
Analysis Projects	79,035
Organisms	239,105

Help Page

▼ GOLD Documentation

[**GOLD Project Entry Help Document**](#)

► [**Gold Analysis Projects SOP**](#)

► [**Email contact**](#)

► [**GOLD Terminology**](#)

► [**GOLD Sequencing Projects**](#)



➤ https://gold.jgi.doe.gov/resources/project_help_doc.pdf

Entering an Isolate Sequencing Project

Studies	28,849
Biosamples	25,696
Sequencing Projects	115,701
Analysis Projects	112,366
Organisms	253,233

[Download Excel Data file](#)
File last generated: 21 Sep, 2016

Welcome to the Genomes OnLine Database

GOLD Release v.6

GOLD: Genomes Online Database, is a World Wide Web resource for comprehensive access to information regarding genome and metagenome sequencing projects, and their associated metadata, around the world.

1. Register



Register your project information and Metadata in the Genomes Online Database

[Register](#)

2. Annotate



Annotate your microbial genome or metagenome with IMG/ER or IMG/MER

[Annotate](#)

3. Publish



Standards in Genomic Sciences

Publish your genome or metagenome in open access standards-supportive journal.

[Publish](#)

My Data:

Home Search Distribution Graphs Biogeographical Metadata Statistics References Team Help News My Profile My Past Searches

Studies	28,849
Biosamples	25,696
Sequencing Projects	115,701
Analysis Projects	112,366
Organisms	253,233

Create Sequencing Projects, Analysis Projects, Studies and Biosamples

For information on the new structure of GOLD please [review the GOLD Project Entry Help Document](#)

▼ Create a new Sequencing Project in GOLD

PLEASE NOTE: We import public genomes from NCBI. Please make sure the genome you want to enter is not already in GOLD. If you have newer assembly to annotate, you only need to create a new Analysis Project. Please [contact us](#) with any questions.

Please select the type of sequencing project you wish to create

1. Select the type of submission you wish to make:

Biome Organism

[Next step >>](#)

► Create a new Analysis Project for submission to IMG

- Specify if the sequencing project type is Organism (for isolates) or Biome (for metagenomes)

Step 1. Enter a New Study

Select study

1: Add a study 2: Add an Organism 3: Complete Sequencing Project

1. a) Search for study:

Enter study name...
OK

1. b) Click the button to create a new Study:
Add a new Study to GOLD

2. Click "Next" to add an organism

Next >>

- Select an existing Study if you already have one
- Otherwise define a new Study and fill in all the relevant fields

▼ Step 1 – Add or create a new study

Study

OR

1. b) Click the button to create a new Study:

Add a new Study to GOLD

Please fill in the form below to create a new Study.

STUDY NAME	
Study Name *	<input type="text" value="Test study to compare two E. coli strains"/>
Other Names	<input type="text"/>
NCBI Umbrella Bioproject Name	<input type="text"/>
NCBI Umbrella Bioproject ID	<input type="text"/>
STUDY DESCRIPTION	
PI Name	<input type="text" value="Supratim Mukherjee"/>
PI Email	<input type="text" value="supratimmukherjee@lbl.gov"/>
Description *	<input type="text" value="This is a test study for MGM-23"/>
Relevance	<input type="text" value="Comparative analysis"/> Add Relevance
GOLD CLASSIFICATION	
Ecosystem	<input type="text" value="Host-Associated"/>
Ecosystem Category	<input type="text" value="Human"/>
Ecosystem Type	<input type="text" value="Digestive system"/>
Ecosystem Subtype	<input type="text" value="Large intestine"/>
Specific Ecosystem	<input type="text" value="Fecal"/>

Create New Study

Step 2. Add (or select an existing) Organism

Add Organism to Sequencing Project

1: Add a study 2: Add an Organism 3: Complete Sequencing Project

► Step 1: Selected Study – Test study to compare two E. coli strains

▼ Step 2: Select or add a new Organism

1. a) Search for an organism by entering genus, species or strain:

Escherichia coli UTI

370 - Escherichia coli UTI89 (UPEC)

OR

1. b) Click the button to create a new Organism:

Add a new Organism to GOLD

2. Click "Next" to add your sequencing project details

Next >>



▼ Step 2: Select or add a new Organism

1. a) Search for an organism by entering genus, species or strain:

OR

1. b) Click the button to create a new Organism:

Add a new Organism to GOLD

Please select the type of organism:

Cultured

Uncultured

Synthetic

Please fill in the form below to create a new Organism

Study

Organism
(Cultured/Isolate/organisms for uncultured single-cells)

Genus *	Escherichia
Species *	
Strain *	
Culture Collection	
Phylum *	PROTEOBACTERIA-GAMMA
NCBI Taxonomy ID *	562
Organism Type	Natural
Cultured/Uncultured	Cultured
Culture Type *	Isolate
GOLD CLASSIFICATION	
Ecosystem *	Host-associated
Ecosystem Category *	Human
Ecosystem Type	Digestive system
Ecosystem Subtype	Large intestine
Specific Ecosystem	Fecal
Ecosystem Suggestion	
ORGANISM ISOLATION	
Sample Collection Site	Human feces
Sample Isolation Comments	
Isolation Country	Select from below...
Collection Method	
Contact Name *	Supratim Mukherjee
Contact Email *	supratimmukherjee@lbl.gov



Entering a New Organism in GOLD

Background Research and Gathering Metadata a.k.a. Homework

The screenshot shows the NCBI Taxonomy Browser interface. At the top, there are links for Entrez, PubMed, and Nucleotide. Below that is a search bar with the placeholder "Search for". Underneath the search bar are buttons for "Display" (set to 3), "levels using filter" (set to none), and checkboxes for "complete name" and "lock". To the right of the search bar are "Go" and "Clear" buttons. The main content area displays information for **Escherichia coli**. It includes the **Taxonomy ID: 562**, which is circled in red. Other details shown include the **Inherited blast name: enterobacteria**, **Rank: species**, **Genetic code: Translation table 11 (Bacterial, Archaeal and Plant Plastid)**, and **Other names:** Enterococcus coli, Bacterium coli commune, Bacterium coli, Bacillus coli. It also lists common names, includes, authorities, and type materials.

<http://www.ncbi.nlm.nih.gov/taxonomy/>

- Search with any possible combination of culture collection id, official strain name, serovar on/off, etc.
- If you don't find the exact taxon id for your strain, you can go one level up and search for the taxon id of the genus/species combination
- E.g. Taxon id for *Escherichia coli* is 562.

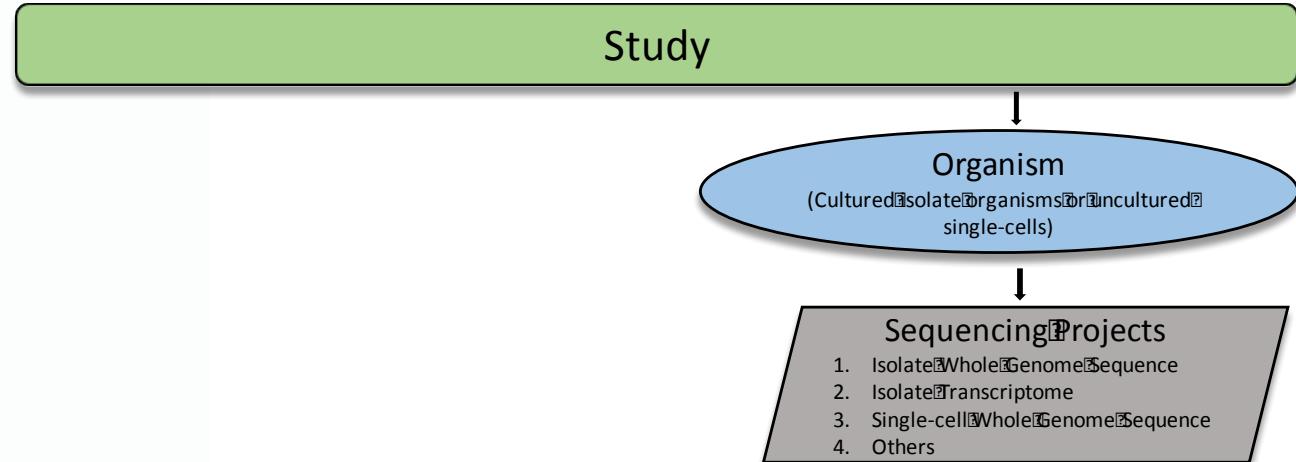
Step 3. Enter Details of Sequencing Project

▼ Step 3: Add sequencing project details

Please do no submit standalone plasmids. Instead, include plasmid sequences as part of your chromosomes/contigs sequence data.

Please fill in the form below to create a new Sequencing Project.
The new sequencing project will remain in draft form until you release it.

PROJECT NAME	
Project Name *	Escherichia coli MGM23_test
Other Names ⓘ	
NCBI BioProject Name	Escherichia coli MGM23_test
NCBI BioProject ID ⓘ	12345
NCBI BioProject Accession ⓘ	PRJNA12345
NCBI BioSample Accession ⓘ	SAMN54321
NCBI Locus Tag ⓘ	
Project Comments	
PROJECT TYPE	
Specimen	Organism
Nucleic Acid *	DNA
Sequencing Strategy *	Whole Genome Sequencing
PROJECT INFORMATION	
Sequencing Center *	University of California, Berkeley
Collaborating Institute	
Funding Agency	U.S. Department of Energy (DOE)
Project Description * ⓘ	Sequencing of an E. coli strain from human gut
SEQUENCING INFORMATION	
Sequencing Status (MIGS-31) *	Complete
Sequencing Quality (MIGS-31)	Level 2: High-Quality Draft
Finishing Goal	Level 2: High-Quality Draft
Sequencing Comments ⓘ	
Sequencing Technology (MIGS-29) *	Illumina HiSeq



Entering a Genome Project in GOLD

Background Research and Gathering Metadata a.k.a. Homework

<http://www.ncbi.nlm.nih.gov/bioproject/>

NCBI Resources How To Sign in to NCBI

BioProject BioProject 50621[uid] Save search Limits Advanced Help

Display Settings: Send to:

Escherichia coli DSM 30083

Escherichia coli DSM 30083 genome sequencing project

Project Data Type: Genome sequencing; **Locus Tag Prefix:** ESCCO

Attributes: Scope: Monoisolate; Material: Genome; Capture: Whole; Method type: Sequencing

Project Data:

Resource Name	Number of Links
SEQUENCE DATA	
SRA Experiments	2
OTHER DATASETS	
BioSample	1

SRA Data Details

Parameter	Value
Data volume, Mbases	81
Data volume, Mbytes	229

Lineage: Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia; Escherichia coli; Escherichia coli DSM 30083 = JCM 1649 [Taxonomy ID: 866789]

Submission:
Registration date: 20-Jul-2010
- DOE Joint Genome Institute

Accession: PRJNA50621 ID: 50621

Related information

BioSample

SRA

Taxonomy

LinkOut to external resources

GOLDCARD: Gi07590 [Genomes On Line Database]

2528311135: Escherichia coli O1:K1:H7 U5/41, DSM 30083 [Integrated Microbial Genomes]

Recent activity

Turn Off Clear

Escherichia coli DSM 30083 BioProject

50621[uid] (1) BioProject

anaerobic digester metagenome taxonomy

Anaerobic digester community_T1 biosample

Aanaerobic digester community_T14 biosample

Step 4. Entering Analysis Project

For information on the new structure of GOLD please [review the GOLD Project Entry Help Docu](#)

▶ [Create a new Sequencing Project in GOLD](#)

▼ [Create a new Analysis Project for submission to IMG](#)

For more information about Analysis Projects please review the [Gold Analysis Projects SOP](#)

To create an analysis project for submission to IMG you need to select from your existing submission please click the button to start the process.

[Add New Sequencing Project](#)

There are two Analysis Project submission types, primary and reanalysis. Your first submission later, [read more...](#)

1. Select the type of Analysis Project you are creating

(depending upon the processing status of your other APs you may not be able to submit a

Create Genome Analysis Project ▾

Please select from below...

- [Create Genome Analysis Project](#)
- [Create Metagenome Analysis Project](#)
- [Create Metatranscriptome Analysis Project](#)
- [Create Single Cell Analysis \(unscreened\)](#)
- [Create Single Cell Analysis \(screened\)](#)
- [Create Genome from Metagenome](#)
- [Create Combined Assembly](#)
- [Create Transcriptome Analysis Project](#)

Study

You are creating a new: Genome Analysis Project

Please fill in the form below to create your Analysis Project

The Sequencing Project for this AP is: Escherichia coli MG

Analysis Project Type: Primary Genome Analysis Project

ANALYSIS PROJECT DESCRIPTION

Analysis Project Name	Escherichia coli MGM23
Suffix:	First draft
Study	Test study to compare t
Analysis Project Description *	Assembly and anno
Public Visibility	Yes

ANALYSIS PROJECT METADATA

Assembly Method *	SOAP 2.3
PI Name	Supratim Mukherjee
PI Email	supratimmukherjee@lbl.gov
Scaffold Count	
Contig Count	
Gene Calling Method	
Gene Count	
Estimated Size	

EXTERNAL REFERENCES

Genbank ID	
Assembly Accession	

Organism
(Cultured, Isolate, Organisms or uncultured, single-cells)

Sequencing Projects

- 1. Isolate Whole Genome Sequence
- 2. Isolate Transcriptome
- 3. Single-cell Whole Genome Sequence
- 4. Others

Analysis Projects

- 1. Isolate Genome
- 2. Transcriptome
- 3a. Single-cell Screened
- 3b. Single-cell Unscreened
- 4. Other AP types

Submit

Entering a Metagenome Sequencing Project

Home Search Distribution Graphs Biogeographical Metadata Statistics References Team Help

Studies	28,849
Biosamples	25,696
Sequencing Projects	115,701
Analysis Projects	112,366
Organisms	253,233

[Download Excel Data file](#)
File last generated: 21 Sep, 2016

My Data:

Home Search Distribution Graphs Biogeographical Metadata Statistics References Team Help News My Profile My

Studies	28,849
Biosamples	25,696
Sequencing Projects	115,701
Analysis Projects	112,366
Organisms	253,233

My Data:

Studies	34
Biosamples	34
Sequencing Projects	64
Analysis Projects	54

Welcome to the Genomes OnLine Database

GOLD:Genomes Online Database, is a World Wide Web resource for comprehensive projects, and their associated metadata, around the world.

1. Register



Register your project information and Metadata in the Genomes Online Database

[Register](#)

2. Annotate



Annotate your microbial with IMG/ER

[Annotate](#)

My Data:

Create Sequencing Projects, Analysis Projects, Studies and Biosamples

For information on the new structure of GOLD please [review the GOLD Project Entry Help Document](#)

▼ Create a new Sequencing Project in GOLD

PLEASE NOTE: We import public genomes from NCBI. Please make sure the genome you want to enter is not you only need to create a new Analysis Project. Please [contact us](#) with any questions.

Please select the type of sequencing project you wish to create

1. Select the type of submission you wish to make:

Biome

Organism

[Next step >>](#)

► Create a new Analysis Project for submission to IMG

Study
Please fill in the form below to create a new Biosample.

BIOSAMPLE NAME	Biosample (DNA/RNA samples collected from their environment)
Other Names	
Habitat *	
Community *	
Location *	
Identifier	
BIOSAMPLE DESCRIPTION	
Biosample Description * <input type="text"/>	
GOLD CLASSIFICATION	
Ecosystem *	Select Ecosystem Type <input type="button" value="Select Ecosystem Type"/>
Ecosystem Category *	<input type="button" value="--"/>
Ecosystem Type *	<input type="button" value="--"/>
Ecosystem Subtype	<input type="button" value="--"/>
Specific Ecosystem	<input type="button" value="--"/>
Ecosystem Suggestion <input type="text"/>	
BIOSAMPLE ISOLATION	
Sample Collection Site *	<input type="text"/>
Sample Collection Date	Month (mm): <input type="text"/> Day (dd): <input type="text"/> Year (yyyy): <input type="text"/>
Add collection time: Hour (hh): <input type="text"/> Minute (mm): <input type="text"/>	
Sample Isolation Comments <input type="text"/>	
Sample Isolation Country	Select from below... <input type="button" value="Select from below..."/>

Standardized Biosample Naming

Four major parts to a Biosample name:

1. Biosample habitat :

- a. Environment (E.g. marine sediment, soil, etc.)
- b. Host-associated (E.g. bovine rumen, human fecal, etc.)
- c. Engineered (E.g. activated sludge, wastewater bioreactor etc.)

2. Biosample community:

Microbial, bacterial, viral, archaeal, eukaryotic etc.

3. Biosample location: geographic location, longitude and latitude for environmental samples are required as MIMS (minimum information about metagenomic sequence/sample)

4. Identifier: Anything that uniquely identifies the project/biosample

Please fill in the form below to create a new Biosample.

BIOSAMPLE NAME	
Biosample Name *	<input type="text"/>
Other Names	<input type="text"/>
Habitat *	<input type="text"/>
Community *	<input type="text"/>
Location *	<input type="text"/>
Identifier	<input type="text"/>

Standardized Metagenome Naming



Emphasis on standardized metagenome biosample and sequencing project naming classification.

Why? With cryptic and esoteric names it is impossible to build a database of metagenomic genes from comparable environments and hosts.

Here's a list of actual metagenome study names we had in GOLD in the past:

- hkbic
- Saliva_contig300
- HK Metagenome T1
- US Sludge
- How many organisms are out there? (ed. Have you tried the Drake equation?)
- (and my favorite)
- --

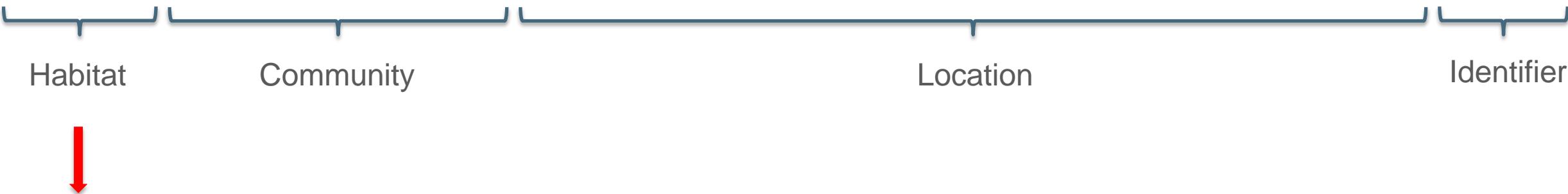
Case Study for Standardized Metagenome Nomenclature



Q. How would you name a project examining microbial communities from sewage samples from a medical facility in Germany?

Maybe.. “German sewage??!! ”

Wastewater microbial communities from medical facility sewage samples near Freiburg, Germany - C1753



GOLD CLASSIFICATION	
Ecosystem	Engineered
Ecosystem Category	Wastewater
Ecosystem Type	Industrial wastewater

Complete Sequencing Project and Analysis Project

Fill in the details of your Sequencing Project:

1: Add a study 2: Add a Biosample 3: Complete Sequencing Project

Step 1: Selected Study – Wastewater microbial communities from medical facility sewage samples near Freiburg, Germany

Step 2: Selected Biosample – Wastewater microbial communities from medical facility sewage samples near Freiburg, Germany – C1753

Step 3: Add sequencing project details

Please fill in the form below to create a new Sequencing Project. The new sequencing project will remain in draft form until you release it.

PROJECT NAME	
Project Name *	Wastewater microbial communities from medical
Other Names	
NCBI BioProject Name	Abundance of antibiotic resistance genes and str
NCBI BioProject ID	264280
NCBI BioProject Accession	PRJNA264280
NCBI BioSample Accession	SAMN03120677
NCBI Locus Tag	
Project Comments	
PROJECT TYPE	
Specimen	Biome
Nucleic Acid	DNA
Sequencing Strategy *	Metagenome

Select "Metagenome" for shotgun sequencing using DNA as starting material. Select "Targeted Gene Survey" for amplicon sequencing including 16S and iTag surveys.

Study

You are creating a new: Metagenome Analysis Project

Biosample
(DNA/RNA samples collected from their environment)

near Freiburg, Germany - C1753

How to create a new Analysis Project.

Wastewater microbial communities from medical facility sewage samples

Analysis Project Type: Metagenome Reanalysis Project

Sequencing Projects

- 1. Metagenome (MetaG)
- 2. Metatranscriptome (MetaT)
- 3. 16S Survey

Study * Select a Study...

Analysis Project Description *

Public Visibility Yes

ANALYSIS PROJECT METADATA

Assembly Method *

PI Name

PI Email

Scaffold Count

Contig Count

Gene Calling Method

Gene Count

Estimated Size

EXTERNAL REFERENCES

Genbank ID

Assembly Accession

Submit



Different Analysis Projects in GOLD

If you extracted one or more **Genomes from a Metagenome** and want to submit the data to IMG isolate annotation pipeline

Genome Analysis

Metagenome Analysis

Metatranscriptome Analysis

Transcriptome Analysis

Genome from Metagenome

Single Cell Analysis (screened)

Single Cell Analysis (unscreened)

Combined Assembly

Create Projects, Studies and Biosamples

For information on the new structure of GOLD please [review the GOLD Project Entry Help Document](#)

▶ Create a new Sequencing Project in GOLD
▼ Create a new Analysis Project for submission to IMG

For more information about Analysis Projects please review the [Gold Analysis Projects SOP](#)

To create an analysis project for submission to IMG you need to select from your existing sequencing projects. If you need to create a new sequencing project for this submission please click the button to start the process.

Add New Sequencing Project

There are two Analysis Project submission types, primary and reanalysis. Your first submission for any Sequencing Project will default to Primary. You can always update this later, [read more...](#)

1. Select the type of Analysis Project you are creating
(depending upon the processing status of your other APs you may not be able to submit any new APs at this time) :

Create Genome from Metagenome ▾

NEXT >>>

▶ Review your Studies, Biosamples and Sequencing and Analysis Projects

Note: You will NOT be able to directly select a metagenome Sequencing Project here. The metagenome from which you extracted a genome must have at least one Analysis Project, for you to be able to choose from..

Different Analysis Projects in GOLD

Genome Analysis

Metagenome Analysis

Metatranscriptome Analysis

Transcriptome Analysis

Genome from Metagenome

Single Cell Analysis (screened)

Single Cell Analysis (unscreened)

Combined Assembly

▼ Create a new Analysis Project for submission to IMG

For more information about Analysis Projects please review the [Gold Analysis Projects SOP](#)

To create an analysis project for submission to IMG you need to select from your existing sequencing projects. If you need to create a new sequencing project for this submission please click the button to start the process.

[Add New Sequencing Project](#)

There are two Analysis Project submission types, primary and reanalysis. Your first submission for any Sequencing Project will default to Primary. You can always update this later, [read more...](#)

1. Select the type of Analysis Project you are creating
(depending upon the processing status of your other APs you may not be able to submit any new APs at this time):

[Create Single Cell Analysis \(screened\)](#)

[NEXT >>](#)

► Review your Studies, Biosamples and Sequencing and Analysis Projects

Combine sequence data from more than one sequencing project to generate a single assembly

Metagenomic projects: (default)
Metagenomic project with Single-Cells:
Metatranscriptome projects:
Single-Cell projects:

Select Parent Project

You are creating a new Single Cell Analysis (screened) Project

Search for a sequencing project:
Gp0139354 - Gammaproteobacterium single cell #123

OR

Select one of your existing sequencing projects:
[Select a sequencing project...](#)

[Next >>](#)

Enter and Submit Analysis Project Details

You are creating a new: Single Cell Analysis (screened) Project

Please fill in the form below to create a new Analysis Project.
The Sequencing Project for this AP is: Gammaproteobacterium single cell #123
Analysis Project Type: Primary Singlecellscreened Analysis Project

ANALYSIS PROJECT DESCRIPTION

Analysis Project Name	Gammaproteobacterium single cell #123
Suffix:	screened assembly
Study	Reddy's test study for help doc examples
Analysis Project Description	Assembly and annotation of screened draft
Contamination Screening Method	Manual screening following automatic decontami

ANALYSIS PROJECT METADATA

Assembly Method	CLC 6.04, Newbler v. 2.3
PI Name	TBK Reddy
PI Email	xxxx@xxxxx.gov
Scaffold Count	
Contig Count	
Binning Method	
Gene Calling Method	
Gene Count	
Estimated Size	

EXTERNAL REFERENCES

Genbank ID	
------------	--

[Submit](#)

Case Study: Combined Assembly b/w MetaG and Single Cell Genome

Case Study: Say you have sequenced a single cell genome. But it is a partial assembly. Then you found hits in one of your metagenome to this single cell genome and wanted take advantage of the available metagenome sequence to get a better assembly for your single cell genome. Then you need to define a combined assembly AP that includes metagenome and single cell SPs.

Combined Assembly

1. Please select the type of Sequencing Projects that make up the Combined Assembly.

Metagenomic projects: (default)
Metagenomic project with Single-Cells:
Metatranscriptome projects:
Single-Cell projects:

2. Please select the Metagenome Sequencing Project and Single Cell project for the Combined Assembly.

A. Please search for your Metagenome sequencing project:

Gp0110165 - Panda gut microbial communities from San Diego Zoo, California, USA - bamboo shoots x

B. Please search for your Single-Cell sequencing projects:

Gp0139354 - Gammaproteobacterium single cell #123 x

You are creating a new: Combined Assembly Project

Study

Biosample (DNA/RNA samples collected from their environment) →
Organism (Cultured or isolated organisms or uncultured single-cells) →

ANALYSIS PROJECT DESCRIPTION

Sequencing Projects
1. Metagenome (MetaG)
2. Metatranscriptome (MetaT)
3. 16S Survey

ANALYSIS PROJECT METADATA

Phylum * PROTEOBACTERIA-GAMMA
NCBI Taxonomy ID * 1236
Assembly Method * CLC 6.04, N
PI Name TBK Reddy
PI Email xxxxx@xxxxxx
Scaffold Count
Contig Count
Binning Method
Gene Calling Method
Gene Count
Estimated Size

Analysis Projects
Combined Assembly (Metagenome & Single-cell genome)

EXTERNAL REFERENCES

Genbank ID Submit

The Not-To-Do List



- Don't create a new study for each sequencing project.
- Don't create a new sequencing project to submit multiple assemblies of the same genome to IMG. Instead create a new Analysis Project.
- Don't create a new Sequencing Project for a public project that you don't own. You can still choose it for your own Analysis Project
- NCBI Taxonomy ID, BioProject ID, BioSample ID etc. are distinct ids.
- For geo coordinates of isolation, use Google maps format (decimal degrees DD and signs +/-; E.g. Co-ordinates for Walnut Creek would be: Latitude: 37.89, Longitude: -122.04). Do not use degree, minutes, seconds format.

When in doubt, send us an email

mail@genomesonline.org

supratimmukherjee@lbl.gov

tbreddy@lbl.gov

Quote your entries by respective ids for Study, Biosample, Sequence Project, Analysis Project etc.

Additional Resources and Links

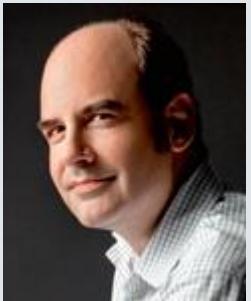


- NCBI Taxonomy (<http://www.ncbi.nlm.nih.gov/taxonomy>)
- StrainInfo (<http://www.straininfo.net/>)
- NamesforLife (<https://services.namesforlife.com/about>)
- NCBI BioProject (<http://www.ncbi.nlm.nih.gov/bioproject>)
- NCBI GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>)
- ATCC (<http://www.atcc.org/>)
- DSMZ (<http://www.dsmz.de/>)
- IMG (<http://img.jgi.doe.gov/cgi-bin/gold/gold.cgi>)

References:

- Reddy TBK, Thomas A, Stamatis D, Bertsch J, Isbandi M, Jansson J, Mallajosyula J, Pagani I, Lobos E and Kyrpides N. **The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification.** *Nucl. Acids Res.* (2014) doi: 10.1093/nar/gku950. PMID: 25348402
- Ivanova N, Tringe SG, Liolios K, Liu WT, Morrison N, Hugenholtz P, Kyrpides NC. **A call for standardized classification of metagenome projects.** *Environ Microbiol.* 2010 Jul;12(7):1803-5. doi: 10.1111/j.1462-2920.2010.02270.x. Review. PubMed PMID: 20653767

GOLD Team



Nikos
Kyrpides



T.B.K.
Reddy



Supratim
Mukherjee



Dimitris
Stamatis



Jon
Bertch



Galina
Ovchinnikova



Olena
Verezemska



Amy
Chen



Krishnaveni
Palaniappan