Update Date: October 2023

GOLD Project Entry Help Document Contents

Page Nos

- GOLD Overview ---------- **2**
- How to start with entering a new project into GOLD ---------- **3**
- Entering a Genome (Isolate) Sequencing Project (SP) and an associated Analysis Project (AP) ---------- **4**
- Entering a Single Cell Sequencing Project (SP) and an associated Analysis Projects (AP) ---------- **16**
- Entering a Metagenome or a Metatranscriptome Sequencing Project (SP) and associated Analysis Projects (APs) ---------- **24**
- Creating a Combined Assembly Analysis Project (AP):
    - Creating a metagenomic Combined Assembly Analysis Project (AP) ---------- **35**
    - Creating a metagenomic and single cell Combined Assembly Analysis Project (AP) ---------- **37**
    - Creating a single cell Combined Assembly Analysis Project (AP) ---------- **39**
    - Creating combined assembly Analysis Project (AP) for sequencing projects from a single organism ---------- **40**

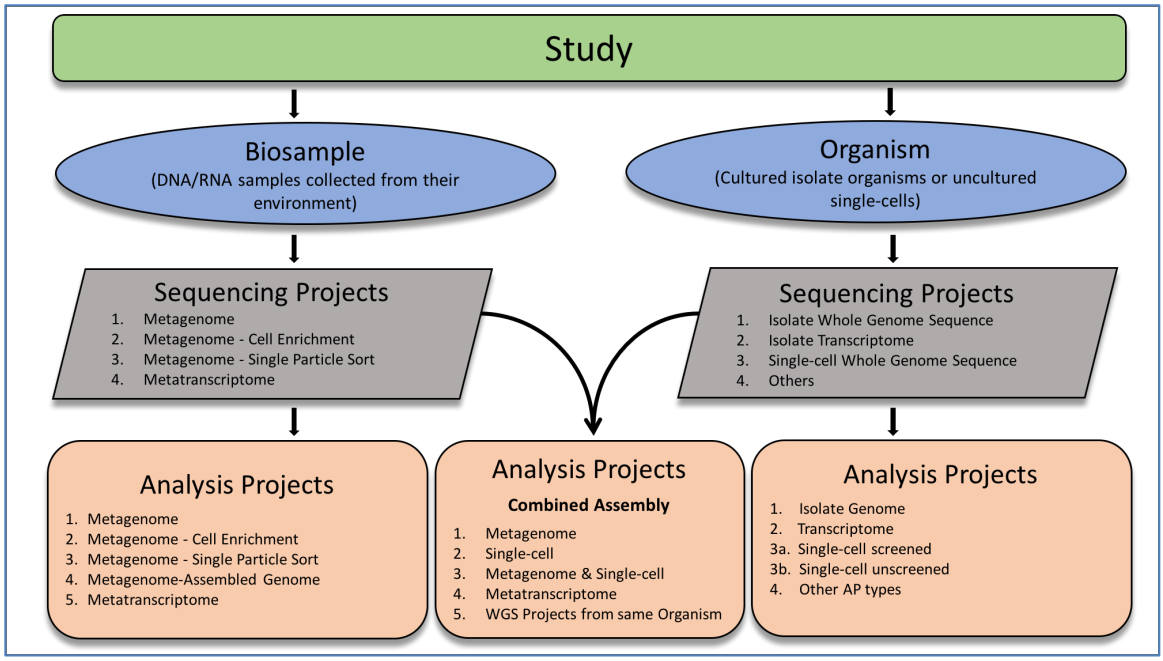
GOLD Home Page: https://gold.jgi.doe.gov/index

GOLD Help Page: https://gold.jgi.doe.gov/help

GOLD Citation: Supratim Mukherjee, Dimitri Stamatis, Cindy Tianqing Li, Galina Ovchinnikova, Jon Bertsch, Jagadish Chandrabose Sundaramurthi, Mahathi Kandimalla, Paul A. Nicolopoulos, Alessandro Favognano, I-Min A. Chen , Nikos C. Kyrpides and T.B.K. Reddy. Twenty-five years of Genomes OnLine Database (GOLD): data updates and new features in v.9. *Nucl. Acids Res. (2022) doi: doi.org/10.1093/nar/gkac974*

GOLD Overview

The revamped GOLD database GOLD v.9 is based on a four-level project classification system. These are Studies, Biosamples/Organisms, Sequencing Projects (SPs) and Analysis Projects (APs).



Study is an umbrella project that broadly defines the overall goal of a research proposal and outlines the key objective of its underlying projects. It contains a list of sequencing projects that are a part of the original proposal. 'Proposal' is a synonym for 'Study' (e.g., HMP study, GEBA study, etc.).

Biosample is a description of the environment from where a DNA/RNA sample was collected. A Biosample contains metadata such as a habitat, an ecosystem, a geographical location, a latitude/longitude, etc. Defining a Biosample entity is essential for creating a GOLD Sequencing Project with Metagenome and Metatranscriptome sequencing strategy. DNA and RNA extracted from the same physical sample can be used for a metagenome and a metatranscriptome project, respectively.

Organism is a description of an individual entity such as a bacterium, a fungus, a plant, an animal, or a virus. It can be a cultured isolate of a pure strain of a bacterium or an uncultured single cell (e.g., isolated by using cell sorting). It can also be an uncultured non-living organism such as a Metagenome-Assembled Genome (MAG). All Organisms in GOLD are required to have basic taxonomic information such as a phylum, a genus, a species, a strain, and an NCBI taxonomy ID. Defining an Organism entity is essential for creating GOLD Sequencing

Projects with Whole Genome Sequencing (WGS) and Transcriptome sequencing strategies.

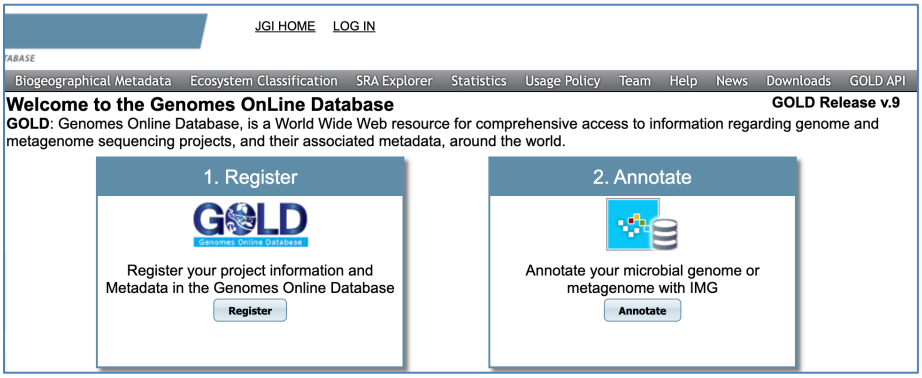
Sequencing Project is a description of the sequencing output from an individual Organism or Biosample. An individual Sequencing Project may be composed of more than one sequencing reaction and/or sequencing technology. A Sequencing Project may be for an isolate genome, a single cell, a metagenome, a metatranscriptome, etc.

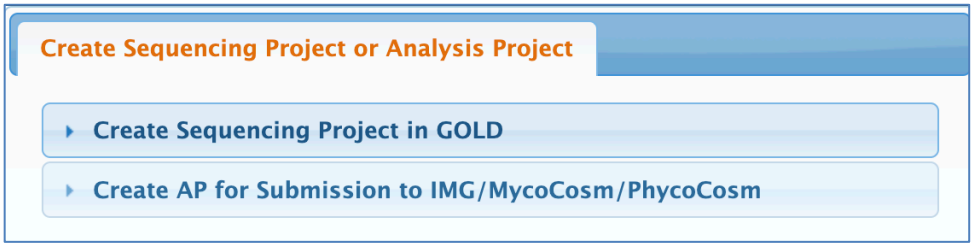
Analysis Project is a description of the informatics processing of a sequence associated with a corresponding Sequencing Project. It specifies how the assembly and annotation of the sequence were performed. Defining an AP entity is essential for submitting dataset to IMG for analysis. An individual submission in IMG represents an individual Analysis Project in GOLD.

Please refer to our publication titled "Twenty-five years of Genomes OnLine Database (GOLD): data updates and new features in v.9" in the Nucleic Acids Research annual Database Issue for more details about this four-level organization.

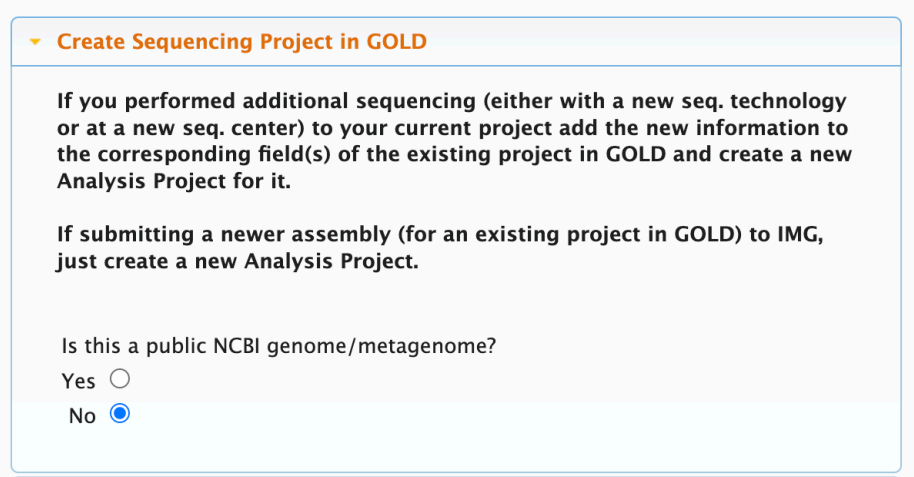
How to start with entering a new project into GOLD

From the GOLD home page, click the "**Register**" button to go to the landing page for entering a Sequencing and/or Analysis Project.





If your Sequencing Project is not yet in GOLD, you will start by clicking the “Create Sequencing Project in GOLD” tab to define a new Sequencing Project.

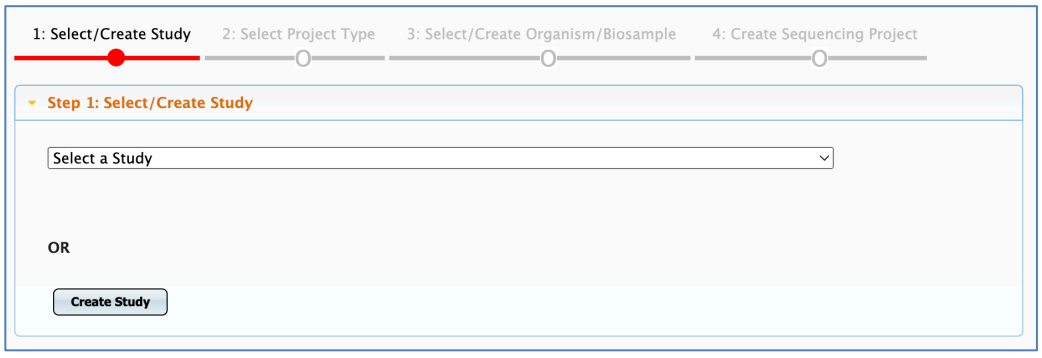


GOLD **does not accept** manually entered public NCBI genomes/metagenomes. If you want to analyze a public genome/metagenome, follow detailed instructions available here: Guidance on submitting public NCBI genomes and metagenomes into GOLD and IMG. Otherwise, select the radio button ‘No’ to move on to the next step.

1) Entering a Genome (Isolate) Sequencing Project (SP) and an associated Analysis Project (AP).

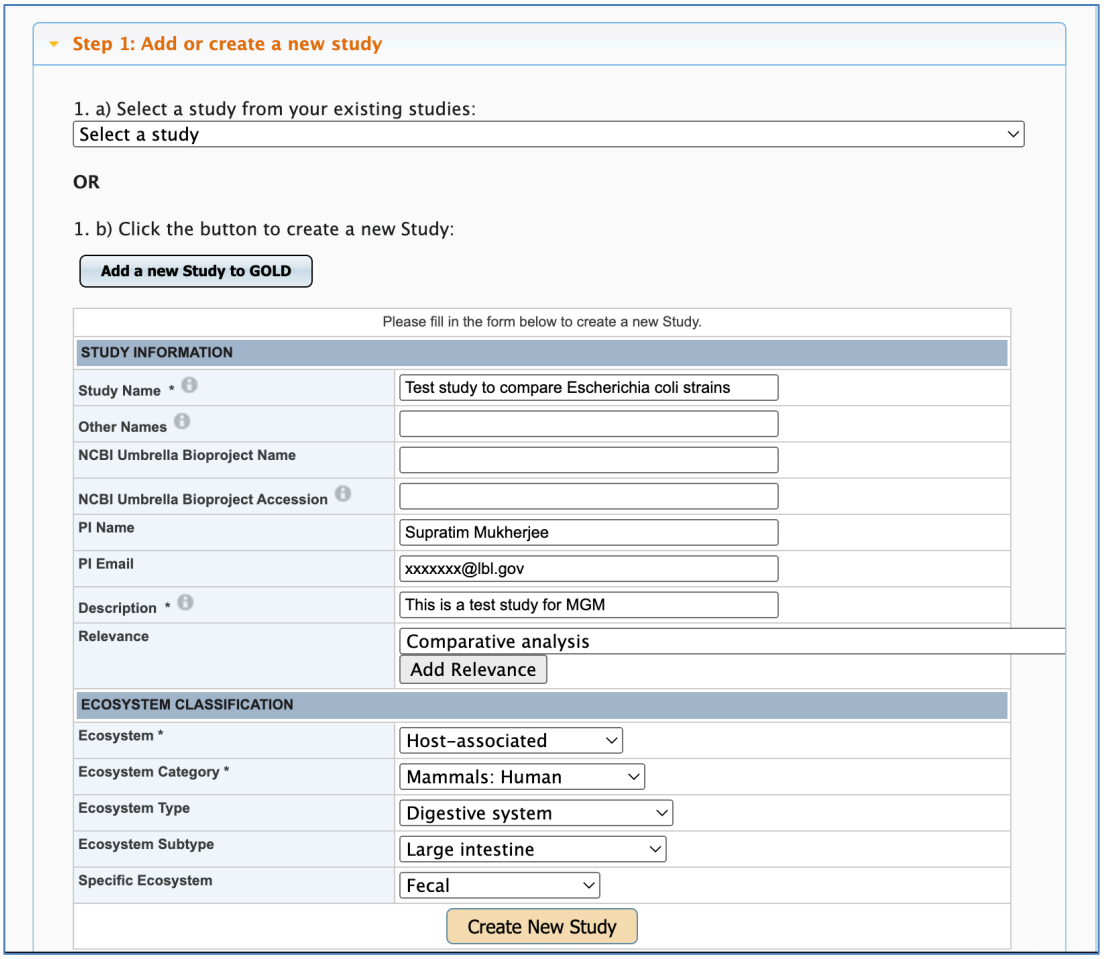
Step 1: Study → Step 2: Select Project Type → Step 3: Organism (Cultured/Isolate) → Step 4: Sequencing Project → Step 5: Analysis Project

The first step to create a Sequencing Project is to define your Study. After confirming that you are NOT entering a public NCBI genome/metagenome, you will come to the page (see the screen below) where you will be able either to select an existing study, under which you want to carry out this new sequencing project, or to define a new study. When applicable, choose a matching study from the drop-down menu.



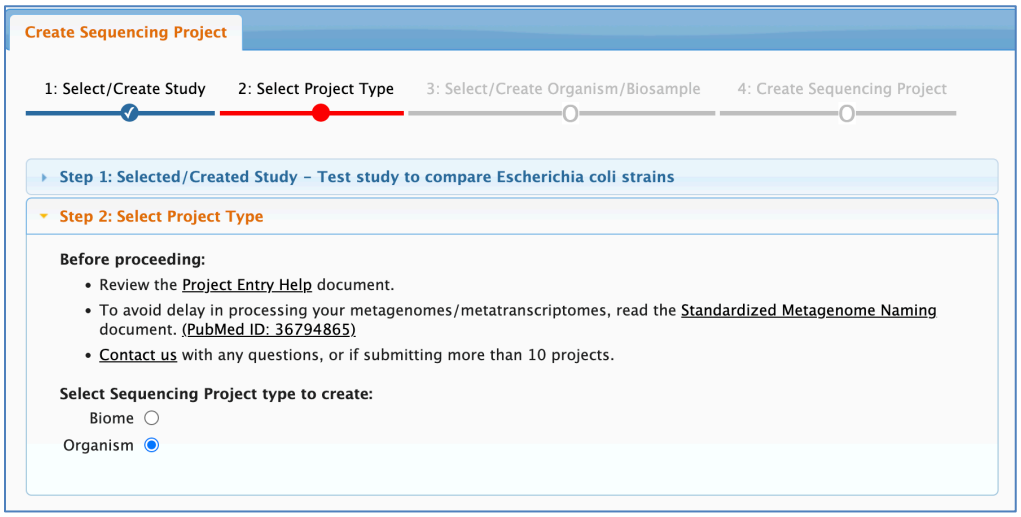
Entering a new Study:

If your new sequencing project is pursued under a new study, click the “Create Study” button to go to the Study entry form. Fill in at least all required fields (*) as shown in the screenshot below. If there is an NCBI BioProject comprising multiple other BioProjects, enter its Name, ID, and Accession into the corresponding NCBI Umbrella BioProject fields. For example, PRJNA28331 is an Umbrella BioProject for Human Microbiome Project (HMP), and it contains about 3000 of the individual BioProjects. If you do not have an umbrella BioProject, leave these fields blank. Do not enter NCBI BioProject IDs or Accessions for individual BioProjects here at the Study level.

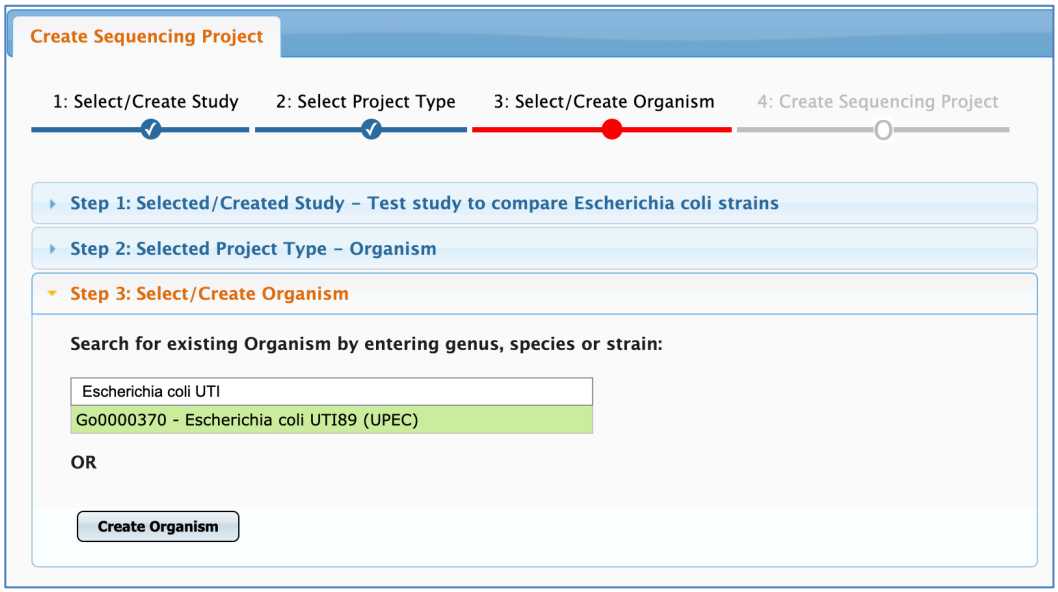


Click the ‘Create New Study’ button and proceed to the next step, i.e., selecting what kind of project you are about to create.

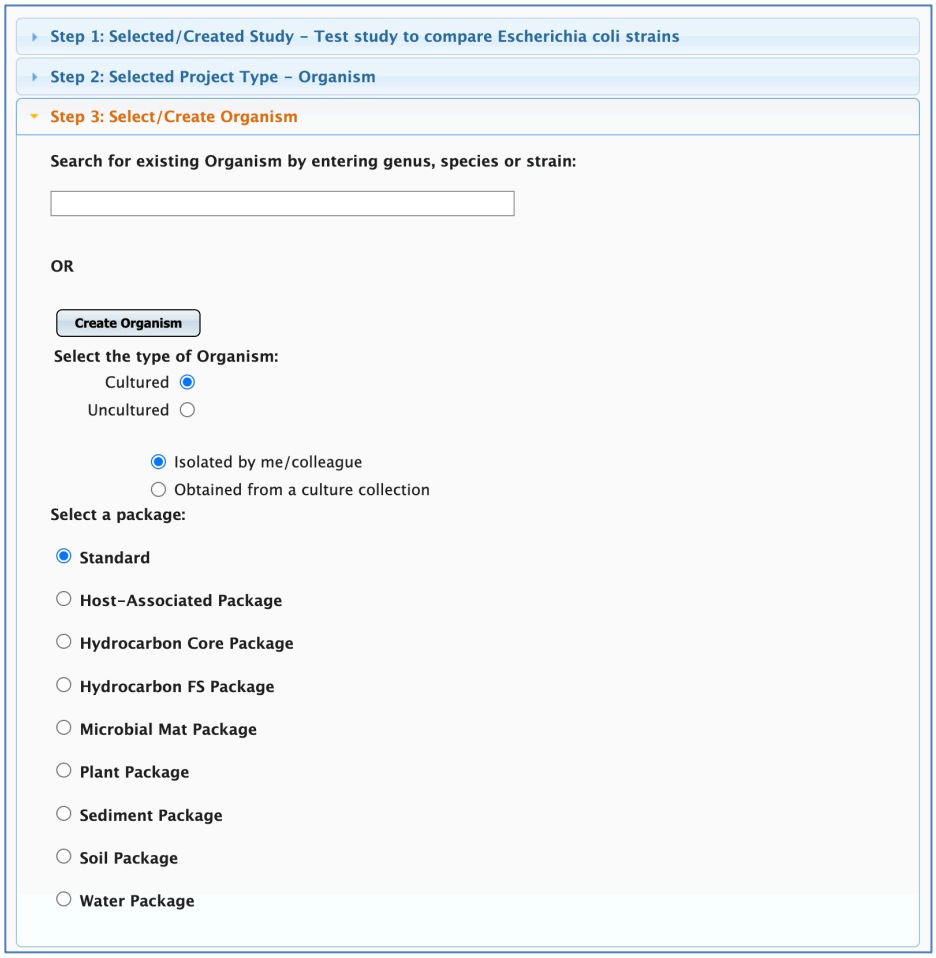
Selecting Project Type:



Select the radio button 'Organism' since you are creating an Isolate genome sequencing project. To create an isolate genome sequencing project, you need either to select an existing Organism or to define a new Organism. Start by searching for a specific Organism strain in GOLD. Once you start typing any part of the Organism name (preferably, a strain or a culture collection id, e.g., DSM 1234 or ATCC 4567), you will get a list of the matching organisms to choose from. Select the exact strain if it is available.



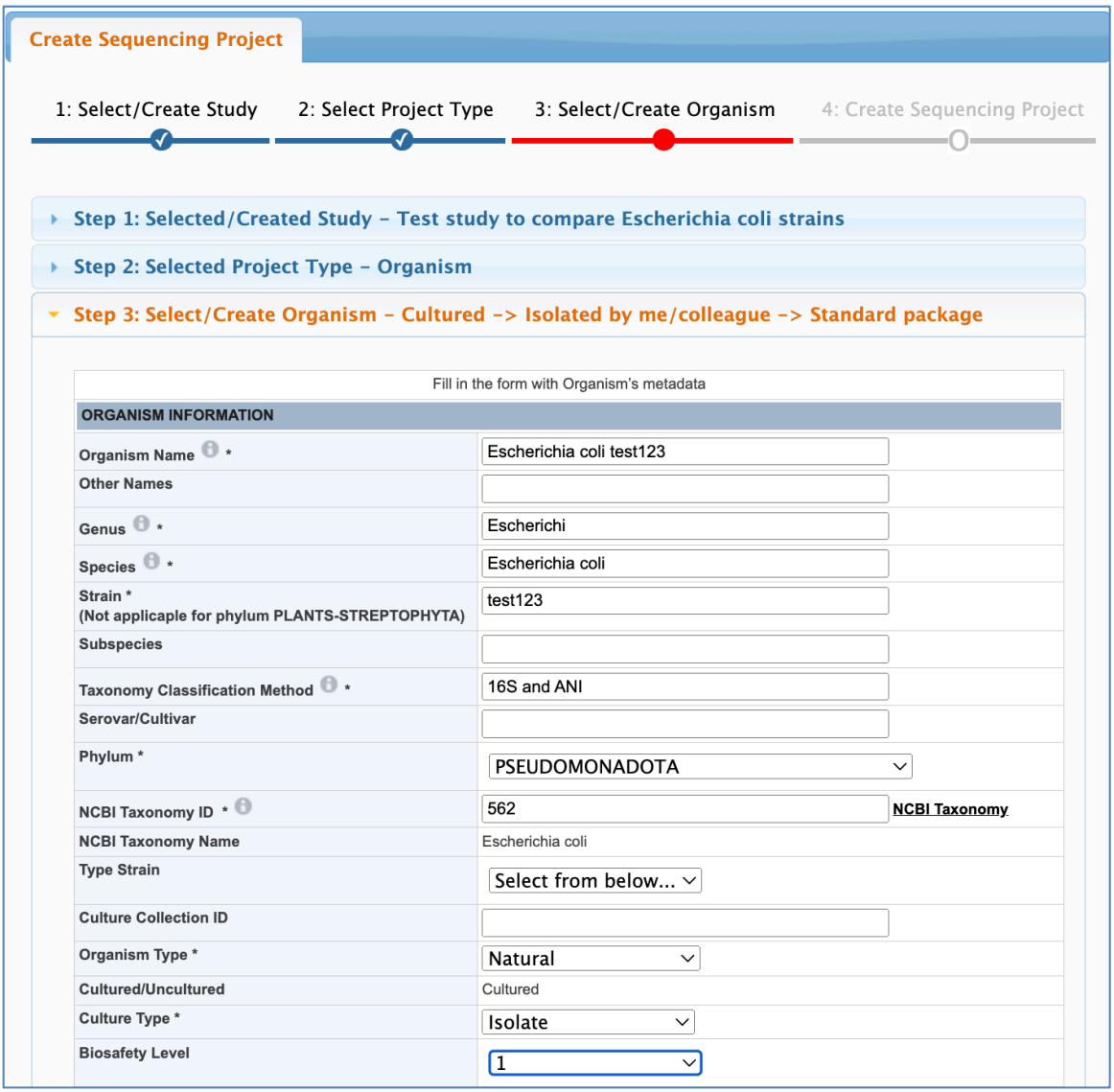
If the exact strain of the organism you are sequencing is not available or you have a new isolate, click the "Create Organism" button to proceed to enter a new Organism. In the next page, you have to indicate whether your organism is cultured or uncultured, and whether it was sequenced by you/your lab/your colleague or it was obtained from a culture collection (ATCC, DSM, etc.)



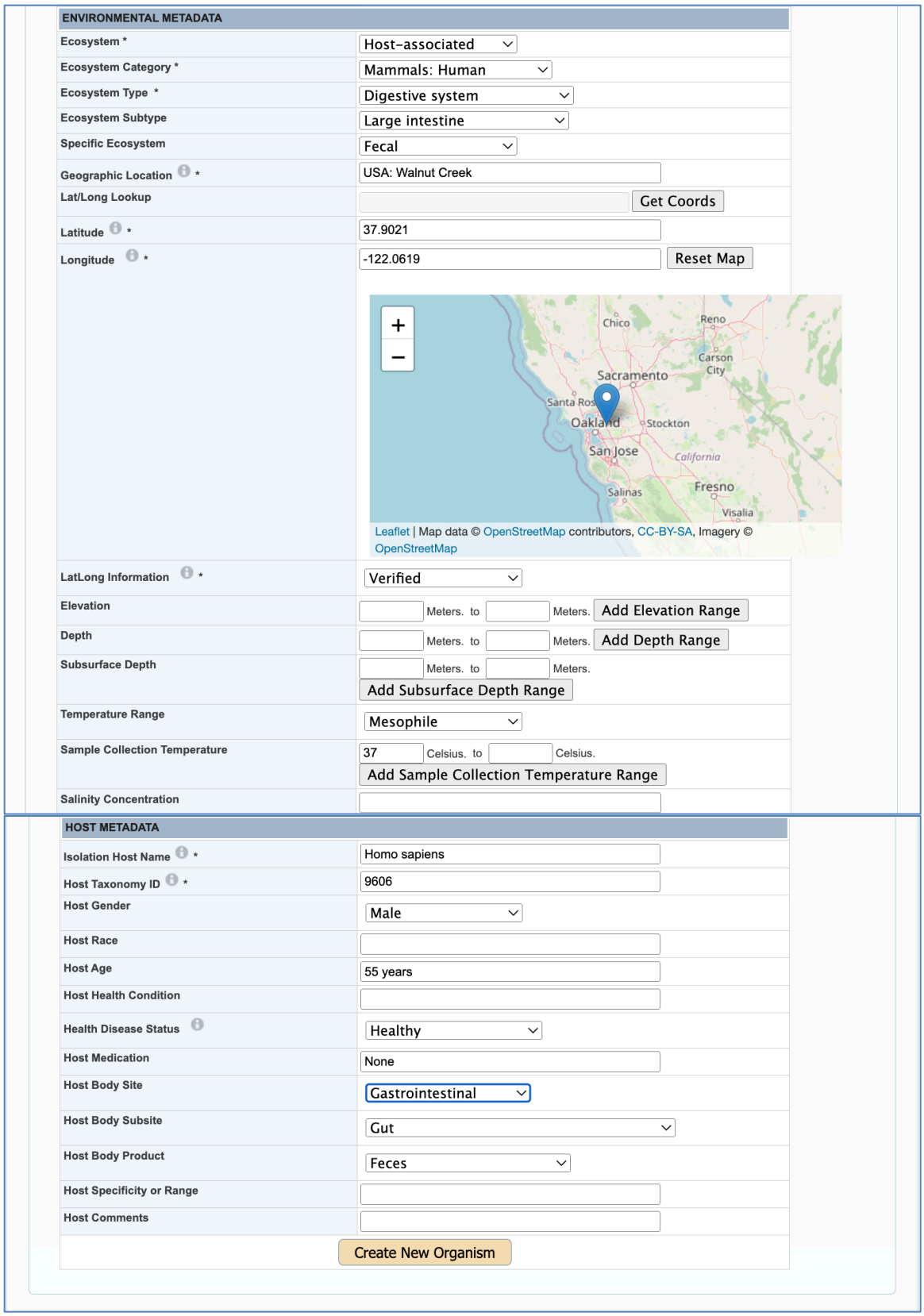
If the Organism is isolated from a host-associated, soil, aquatic or any other environment listed above, you have an option to select the appropriate package and get access to the form with additional metadata fields specific to the respective environment. Otherwise, choose the Standard package to access the default Organism entry form.

Enter organism-specific information such as the organism's name, genus, species, strain, NCBI taxonomy id, culture collection id, etc. into the form. Please note that for the Species field, we require to use the binomial nomenclature (two-term organism naming system) as per the nomenclature conventions followed by NCBI taxonomy. For example, enter "*Escherichia coli*" into the Species field as shown in the example below. If only a genus is defined for your organism, use a format "*Genus sp. Strain*", e.g., "*Pseudomonas sp. 12ab*" to populate the field Species. Fill in the rest of the form.

Though not all fields are mandatory, we encourage you to enter all available metadata.


Create Sequencing Project
1: Select/Create Study
2: Select Project Type
3: Select/Create Organism
4: Create Sequencing Project
Step 1: Selected/Created Study - Test study to compare Escherichia coli strains
Step 2: Selected Project Type - Organism
Step 3: Select/Create Organism - Cultured -> Isolated by me/colleague -> Standard package
Fill in the form with Organism's metadata
ORGANISM INFORMATION
Organism Name *
Escherichia coli test123
Other Names
Genus *
Escherichi
Species *
Escherichia coli
Strain *
(Not applicaple for phylum PLANTS-STREPTOPHYTA)
test123
Subspecies
Taxonomy Classification Method *
16S and ANI
Serovar/Cultivar
Phylum *
PSEUDOMONADOTA
NCBI Taxonomy ID *
562
NCBI Taxonomy
NCBI Taxonomy Name
Escherichia coli
Type Strain
Select from below...
Culture Collection ID
Organism Type *
Natural
Cultured/Uncultured
Cultured
Culture Type *
Isolate
Biosafety Level
1

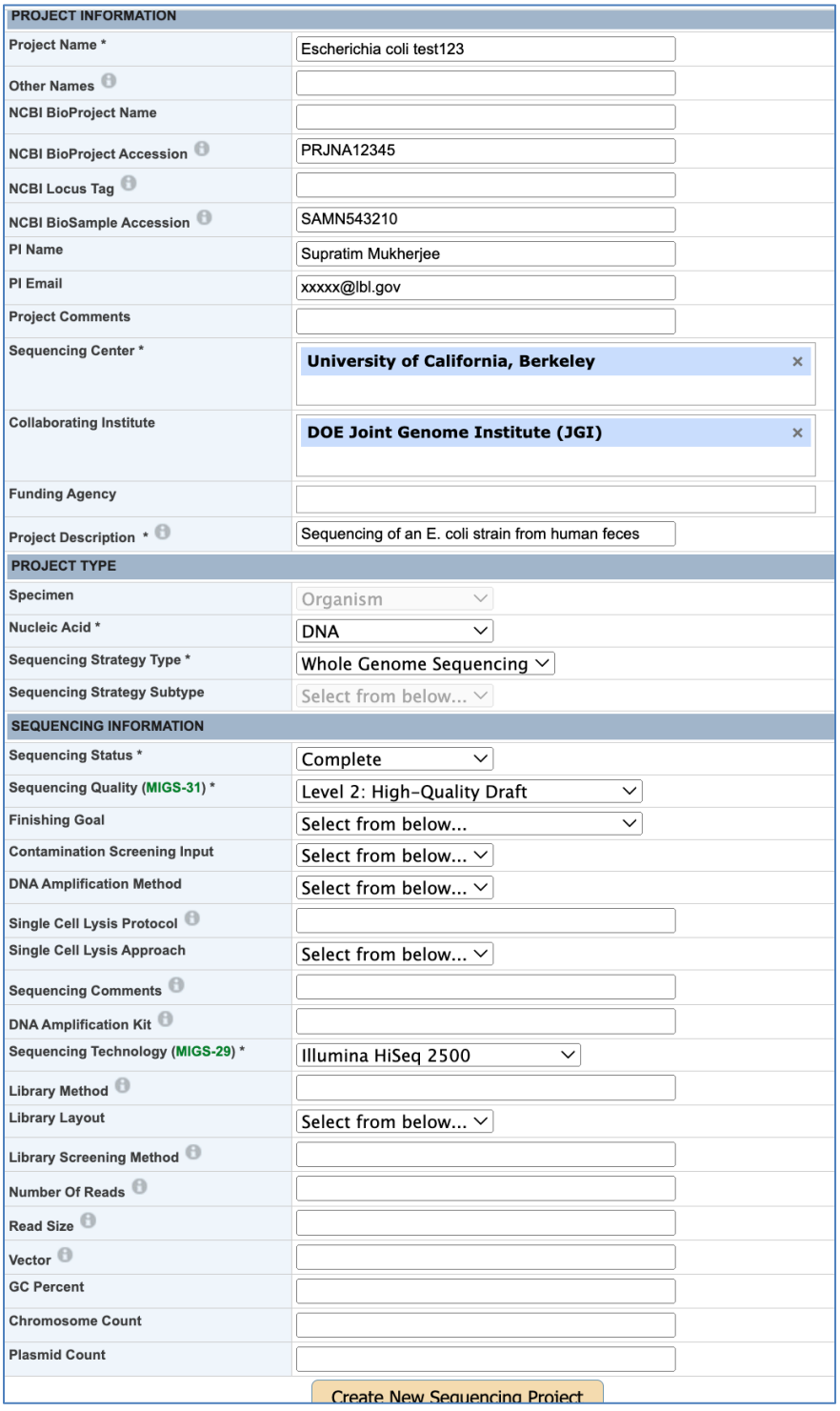
ORGANISM METADATA	
Cell Shape	Rod-shaped
Color	Cream
Gram Stain	Gram-
Motility	Motile
Oxygen Requirement (MIGS-22)	Facultative anaerobe
pH	4.5 to 9 Add pH Range
Salinity (MIGS-6.3)	Undefined
Pressure	
Sporulation	Nonsporulating
Carbon Source	Glucose, Lactose
Growth Temperature	20 Celsius. to 50 Celsius. Add Growth Temperature Range
Cell Arrangements	Singles
Diseases (MIGS-15)	Gastroenteritis
Habitats (MIGS-6)	Human gastrointestinal tract
Metabolism	Hydrocarbon degrading
Phenotypes	Fast growing
Energy Sources	Heterotroph
Body Products	Feces
Biotic Relationships (MIGS-14)	Free living
Propagation	
ISOLATION METADATA	
Sample Collection Date *	Month (mm) * : 01 Day (dd): 15 Year (yyyy) * : 2020
Sample Collection Site *	Human feces
Isolation Country/Ocean *	USA
Sample Isolation Comments	
Collection Method	Serial dilution
Contact Name *	Supratim Mukherjee
Contact Email *	xxxxxxxxxx@lbl.gov



Once you select an existing Organism or have defined a new Organism, you can proceed to the next step, i.e., creating a new Sequencing Project.

Entering a Sequencing Project:

To define a new Sequencing Project, populate at least all required fields (*) such as Sequencing Center, Project Description, Sequencing Strategy Type, Sequencing Technology, etc. as shown in the screenshot below. Enter NCBI BioProject and Biosample Accessions ONLY if they are not public in NCBI.

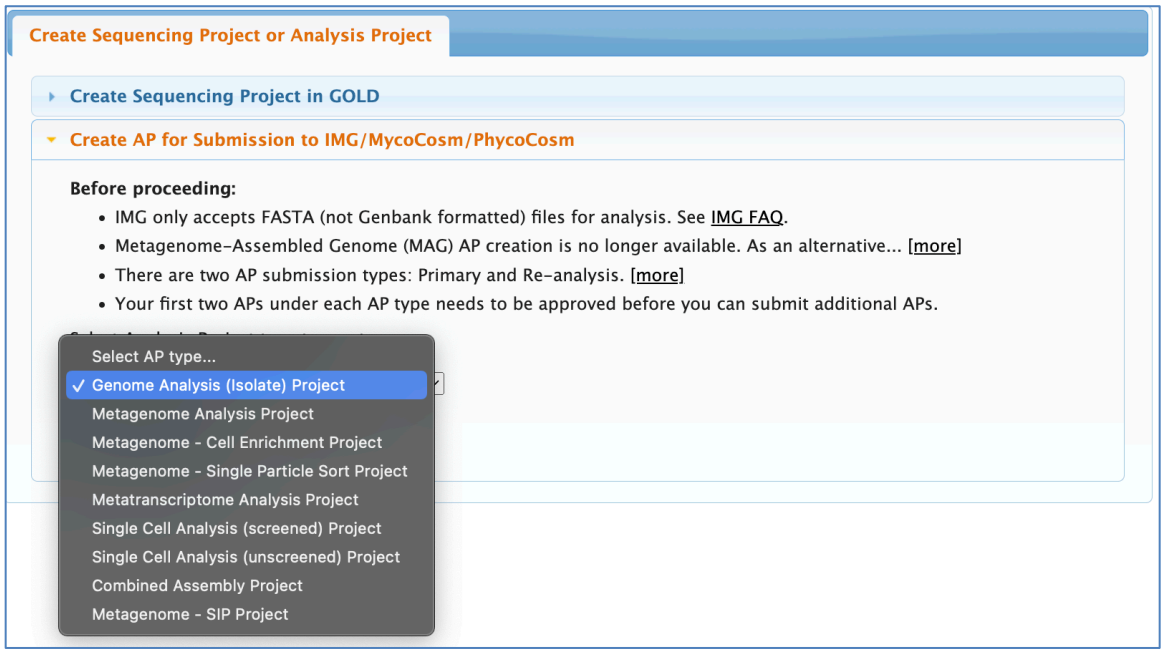


Once you enter all relevant information, click the “Create New Sequencing Project” button.

Now you can proceed to create a GOLD Analysis Project, or you can leave and come back later (when you are ready to submit data to IMG) to create a GOLD AP.

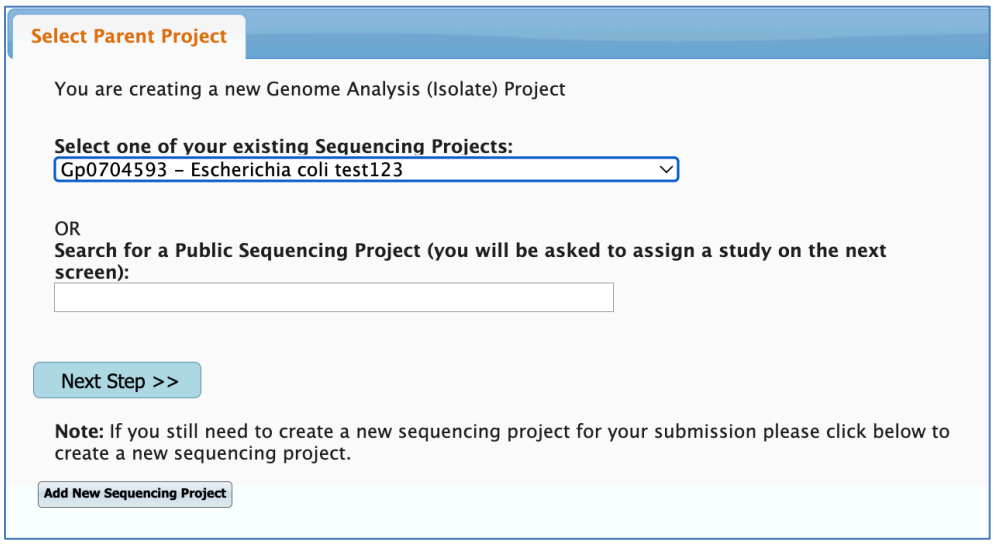
Entering a Genome Analysis Project for an isolate genome:

Go to the main project submission page and click the “Create Analysis Project for Submission to IMG/MycoCosm/PhycoCosm” button. Select “Genome Analysis (Isolate) Project” from the drop-down menu and proceed.

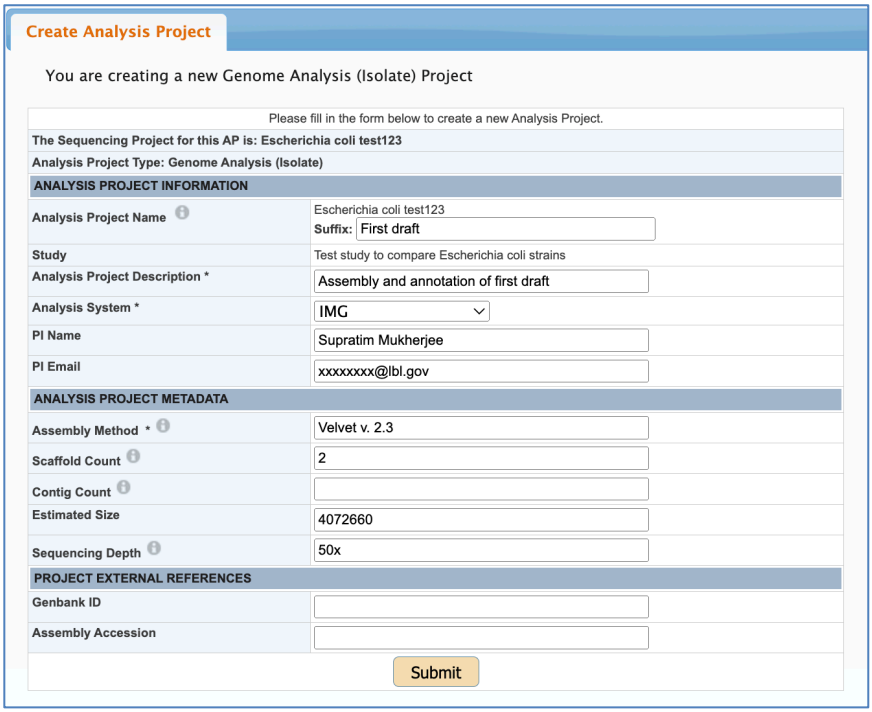


You can create a Genome Analysis Project for your own Sequencing Project that you have previously created, or you can create a Genome AP for any public sequencing project if you have reassembled the associated public sequences and want to re-annotate them.

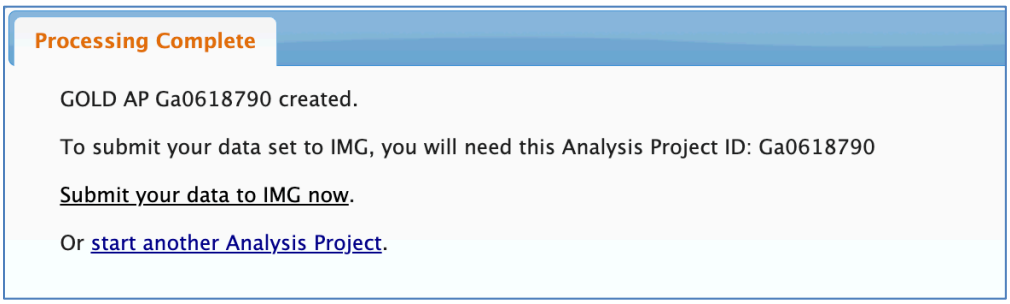
To create an Analysis Project for your own Sequencing Project, select the corresponding SP from the drop-down menu and go to the Analysis Project creation page.



Fill in at least the required fields as shown below. If appropriate, provide a suffix to AP name: a version number, a draft number, etc. anything that will allow you to distinguish between versions in case you have re-sequenced the same organism or reassembled the same sequence. Click the “Submit” button to finish.

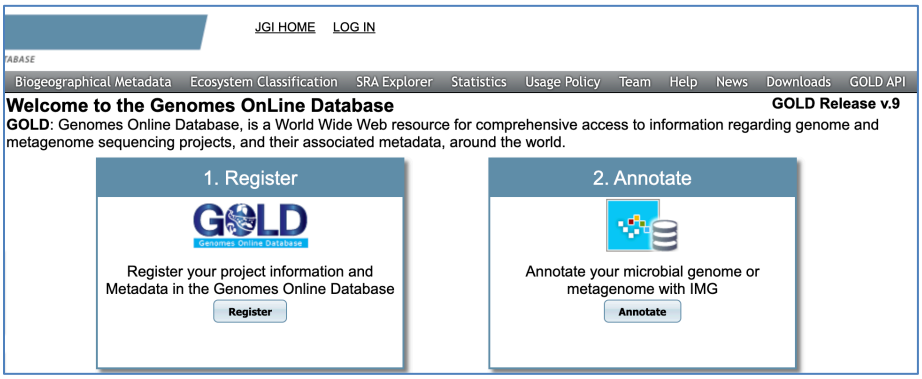


Once you successfully create an Analysis Project, you will see the following window:

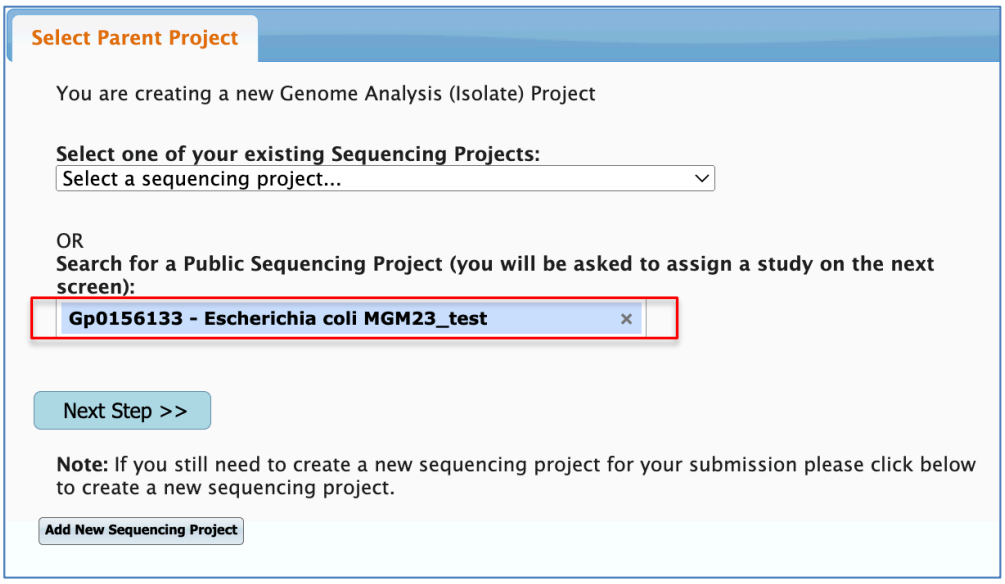


Click the link to proceed to IMG for data submission. In the future, you can create additional APs on the same Sequencing Project if, for example, you have a new assembly or an additional sequence to include.

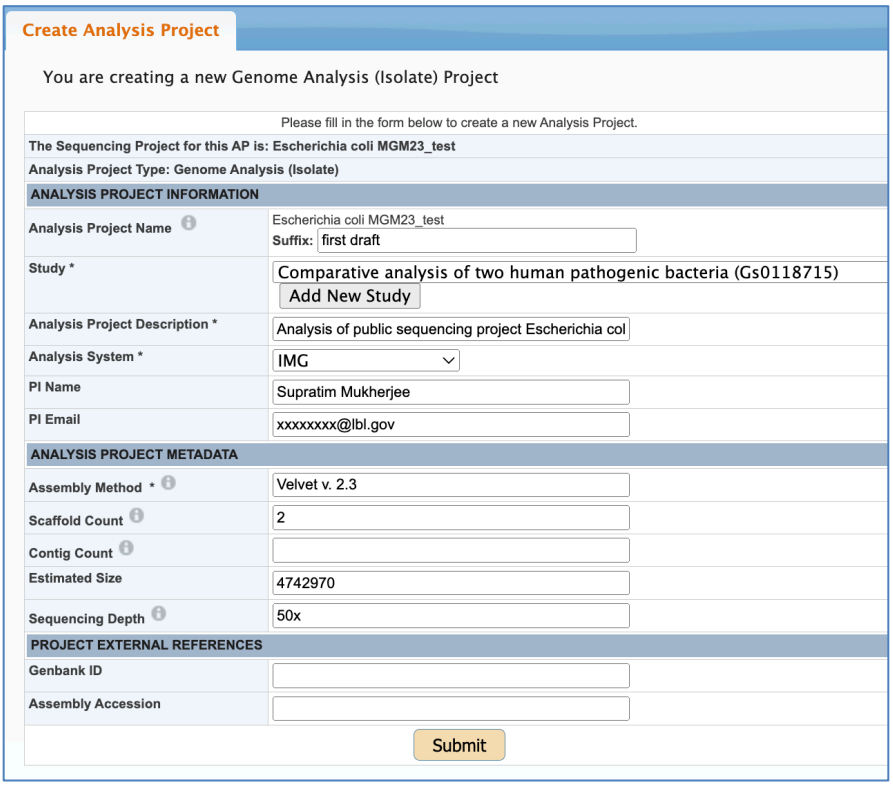
If you are not ready to submit your data to IMG yet, you can return later and use your AP for IMG submission. When you click the “Annotate” button on the GOLD home page, you will be redirected to the IMG submission page.



To create an AP for a public project, use the corresponding search box to select a project of your interest by typing a part of the organism’s name or a strain and proceed to the Analysis Project creation page.



Fill in at lease the required Analysis Project form's fields as shown below. When you are creating an AP on a public SP, you must choose the appropriate study for your AP in case you have more than one study in GOLD. You also have the option of creating a new Study at this point. Click the "Submit" button to finish.



2) Entering a Single Cell Sequencing Project (SP) and an associated Analysis Project (AP).

Step 1: Study → Step 2: Select Project Type → Step 3: Organism (Uncultured/Single Cell) → Step 4: Sequencing Project → Step 5: Analysis Project

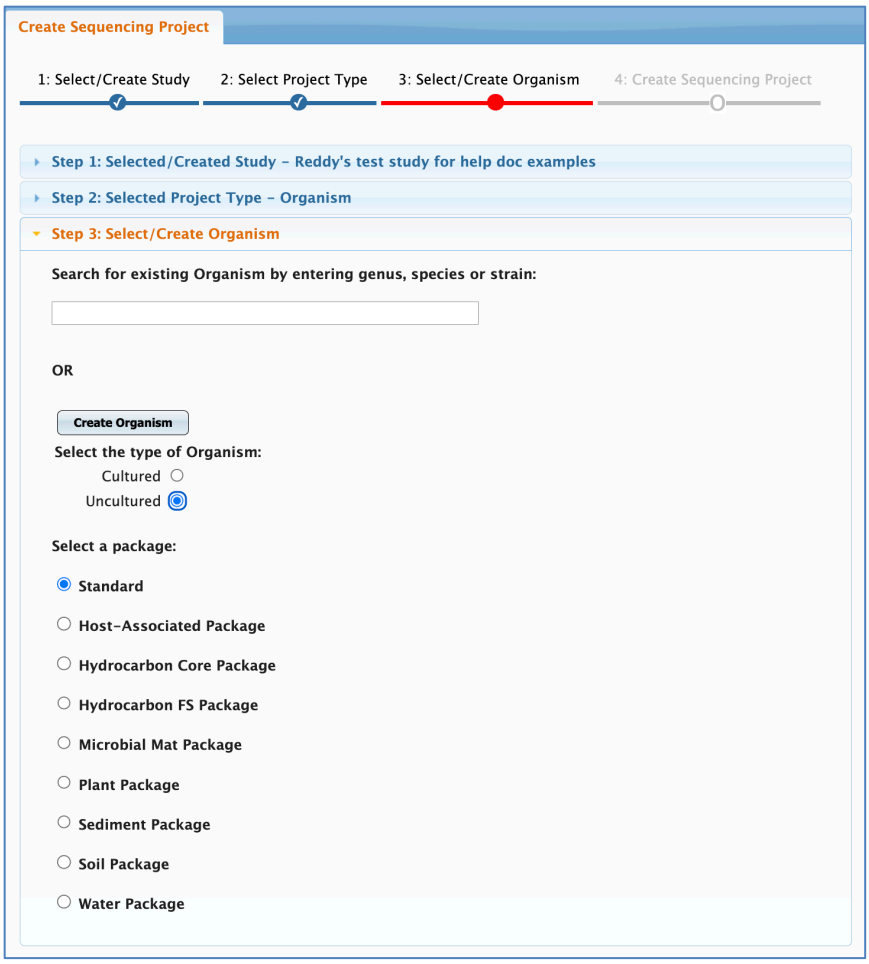
Creating a Study and Selecting Project Type:

The first two steps for entering a Single Cell project, namely, selecting or entering a Study and selecting an SP type are identical to the ones that were described earlier on pages 3-6 of this document.

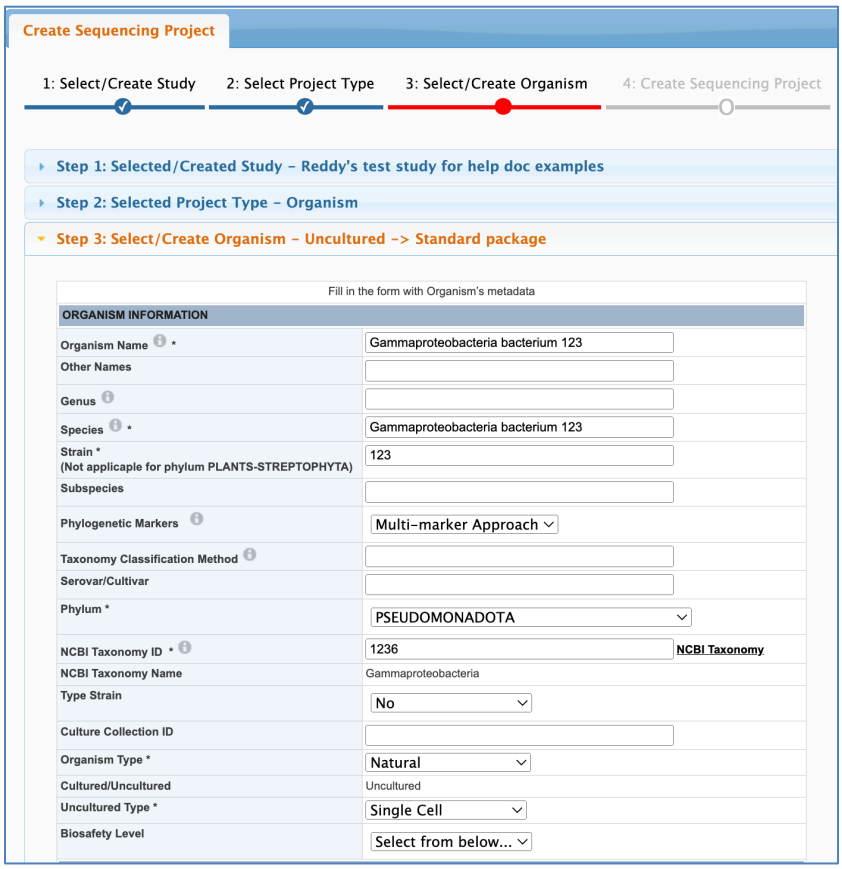
Entering a new Organism:

It is very unlikely that you will find a single cell organism entry already in GOLD to choose, as no one else can possibly possess the same single cell you isolated from an environment. The only time you may be able to use the same/existing organism entry is if you happen to enter this organism before. In all other cases, you will be creating a new organism for a single cell.

Entering a single cell organism is similar to entering an isolate organism, which is described above. The main difference is in defining an organism as a single cell. When defining an organism for single cell projects you will select the “Uncultured” radio button as shown below. If the Organism is isolated from a host-associated, a soil, an aquatic or any other environment listed below, you have an option to select the appropriate package and to get access to the form with additional metadata fields specific to the respective environment. Otherwise, choose the Standard package to access the default Organism entry form.



After that, you will proceed to enter a new single cell organism. The first section of the Organism entry form is shown below. The form is similar to the isolate Organism form. The only difference is that the Cultured/Uncultured field is set to the Uncultured one by default, and you need to select 'Single Cell' from the Uncultured Type field's menu.



Once you define an organism, you will proceed to the next step, i.e., creating a Sequencing Project.

Entering a Sequencing Project:

The process of entering a Sequencing Project for a single cell organism is identical to the one described on pages 10-11 for an isolate organism.

Now you can proceed to create a GOLD Analysis Project, or you can leave and come back later (when you are ready to submit data to IMG) to create a GOLD AP.

Entering a Single Cell Analysis Project:

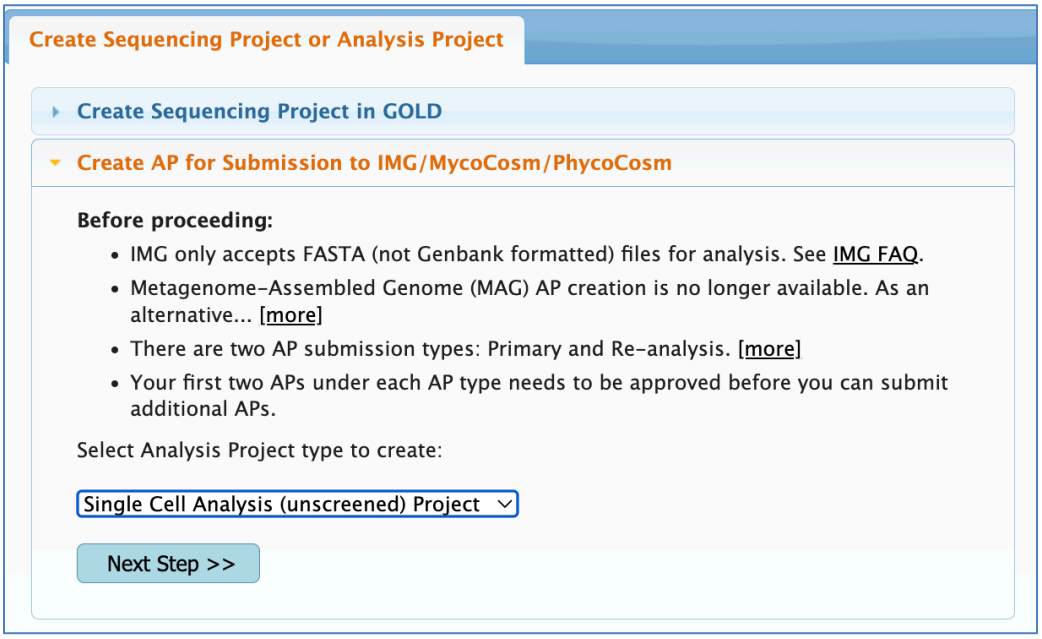
There are two types of APs for a single-cell-based SP. They are Single Cell Analysis Project (unscreened) and Single Cell Analysis Project (screened).

Typically, a Single Cell Analysis (unscreened) AP is used when no additional screening is done besides the standard contamination screen.

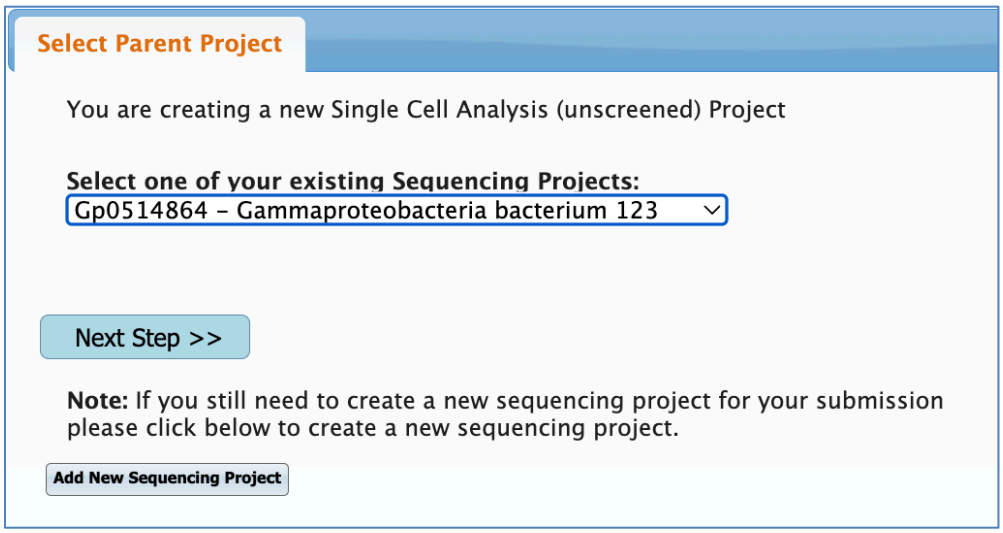
Single Cell Analysis (screened) is used when an extensive manual or automatic screening is performed to remove the sequence that does not belong to the single cell of your interest.

Creating a Single Cell Analysis (unscreened) Project:

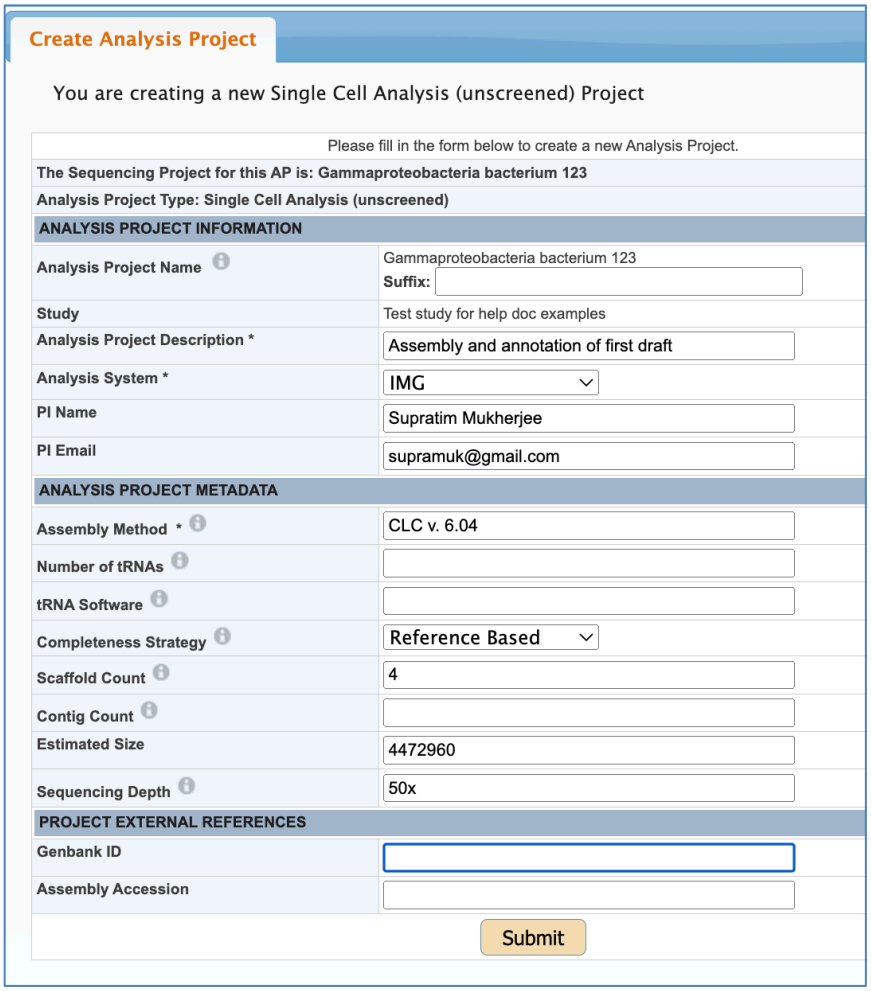
Go to the main project submission page and click the “Create Analysis Project for Submission to IMG/MycoCosm/PhycoCosm” button. Select “Single Cell Analysis (unscreened) Project” from the drop-down menu and proceed to the next step.



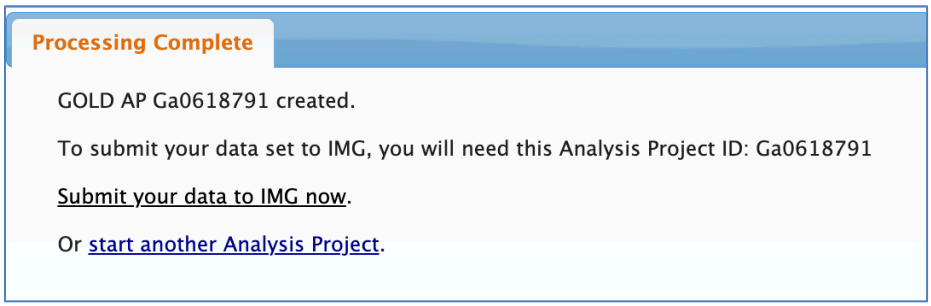
Choose the corresponding single cell SP from the drop-down menu.



Fill in and submit the following form to create a Single Cell (unscreened) AP.

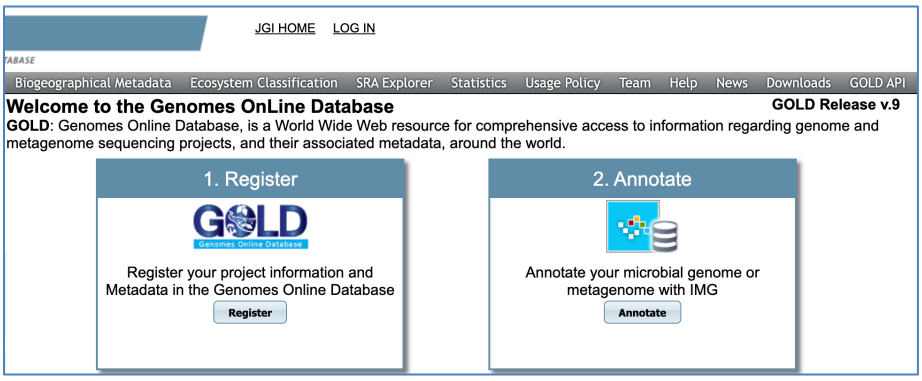


Once an AP is created, you will see the following window:



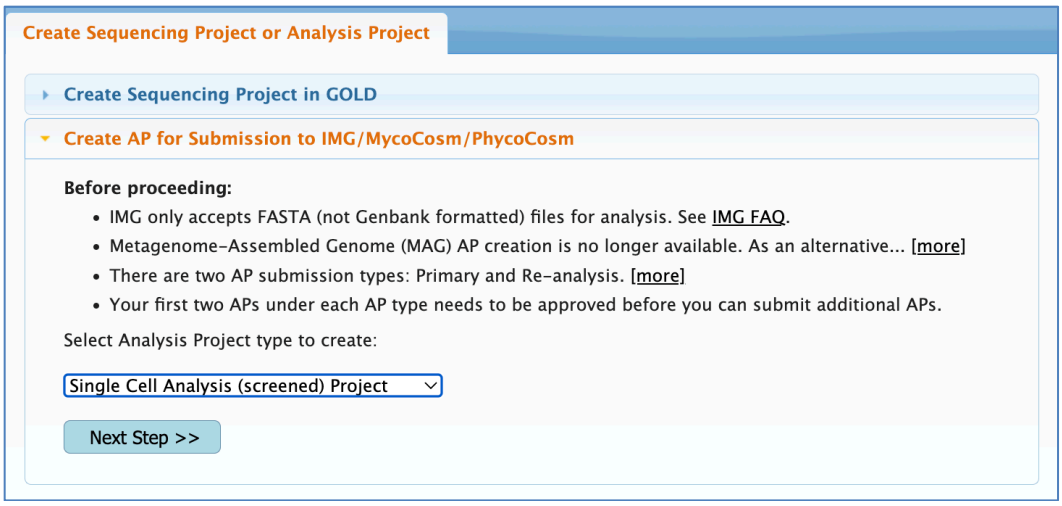
Click the link to proceed to IMG for data submission.

If you are not ready to submit your data to IMG yet, you can return later and use your AP for IMG submission. You will be redirected to the IMG submission page when you click the “Annotate” button on the GOLD home page.

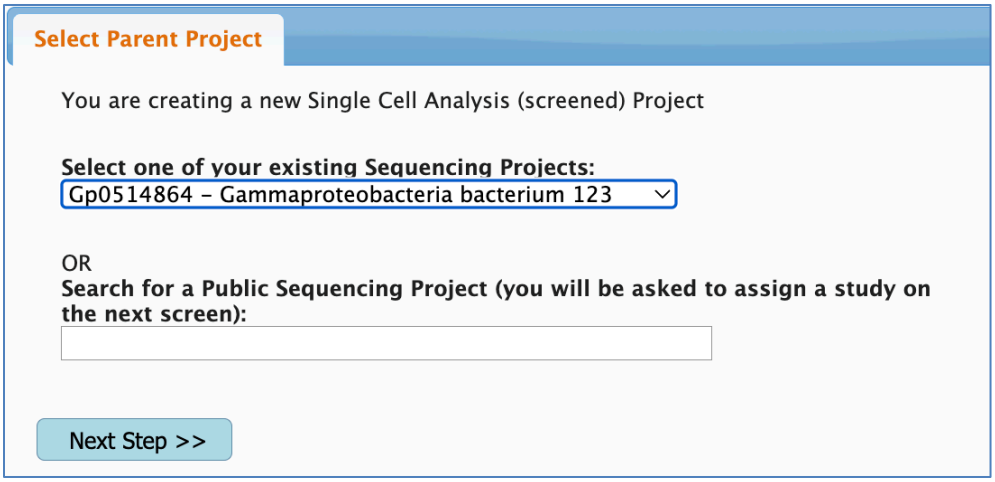


Creating a Single Cell Analysis (screened) Project:

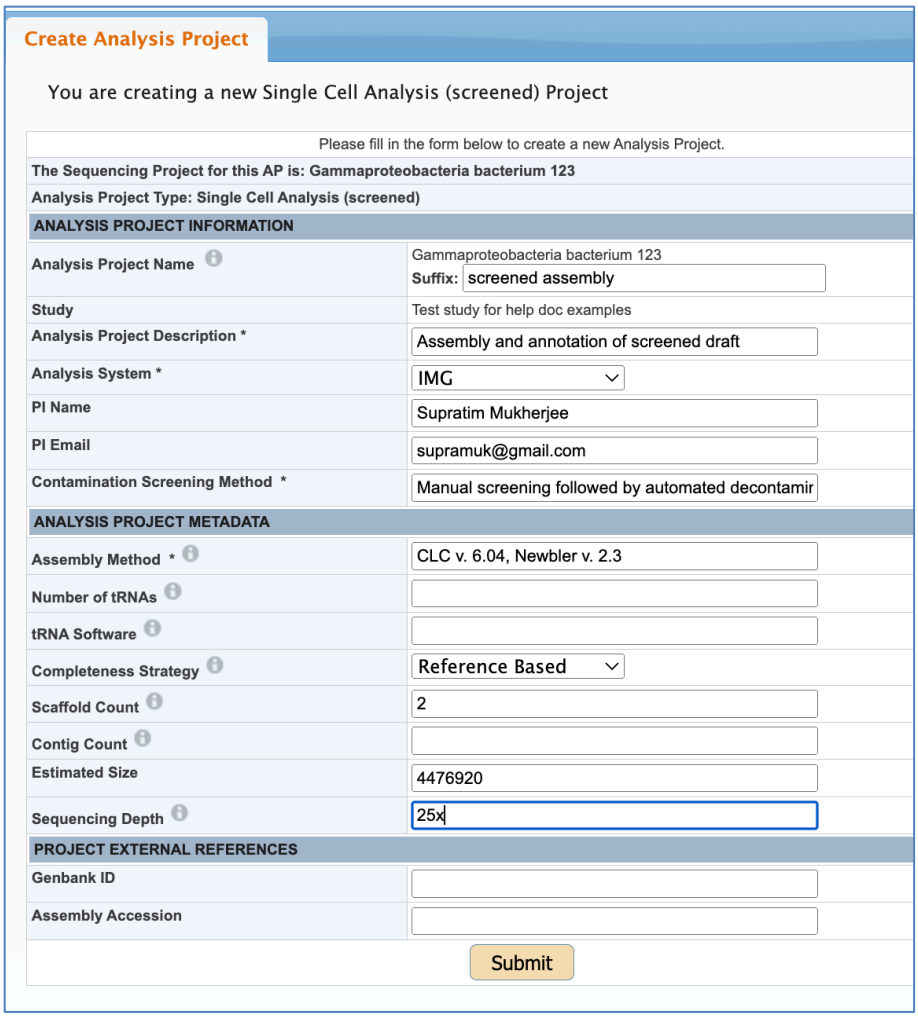
Go to the main project submission page and click the “Create Analysis Project for Submission to IMG/MycoCosm/PhycoCosm” button, select “Single Cell Analysis (screened) Project” from the drop-down menu, and proceed to the next step.



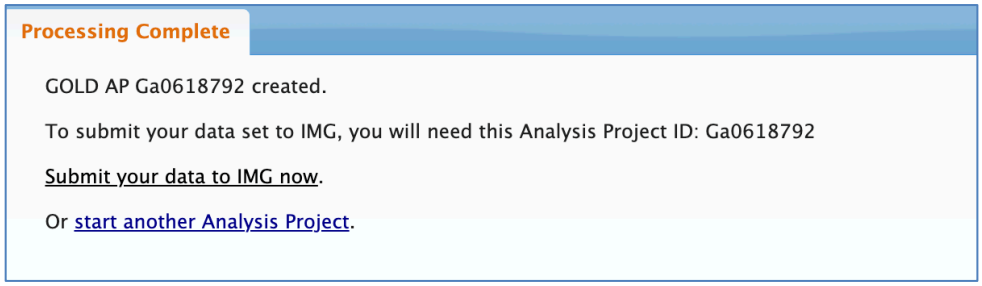
Choose the corresponding single cell SP from the drop-down menu.



Fill in and submit the following form to create a Single Cell (screened) AP.

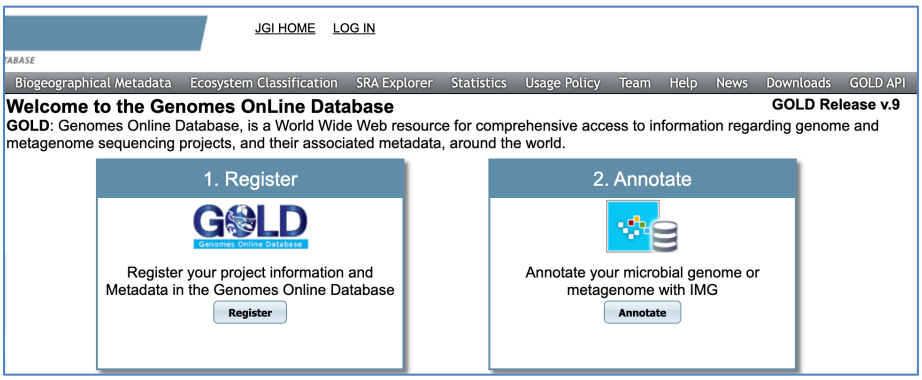


Once an AP is created, you will see the following window:



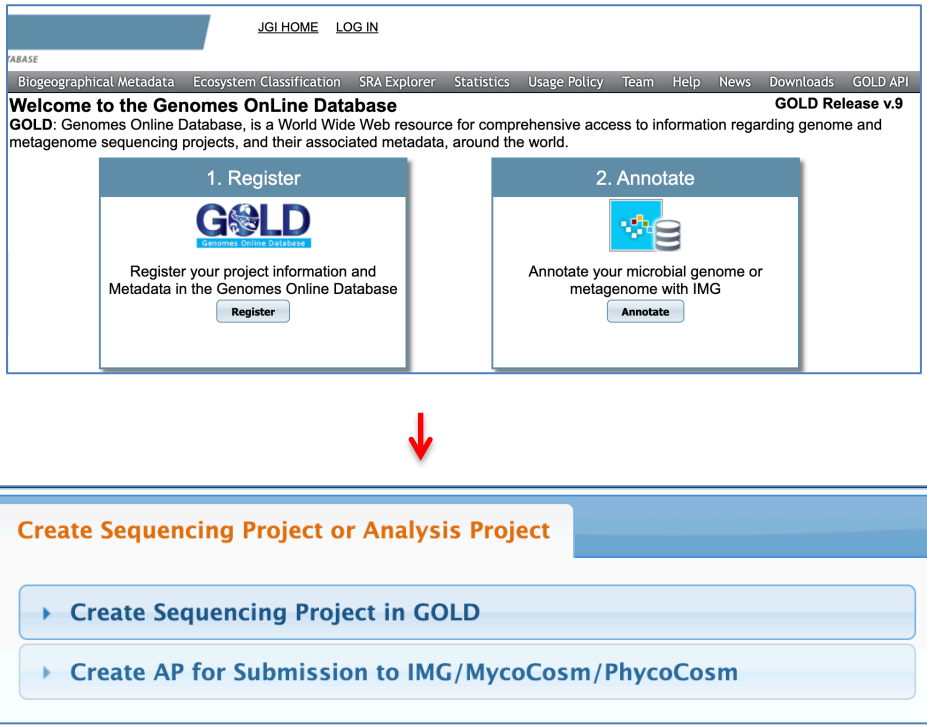
Click the link to proceed to IMG for data submission.

If you are not ready to submit your data to IMG yet, you can return later and use your AP for IMG submission. You will be redirected to the IMG submission page when you click the “Annotate” button on the GOLD home page.

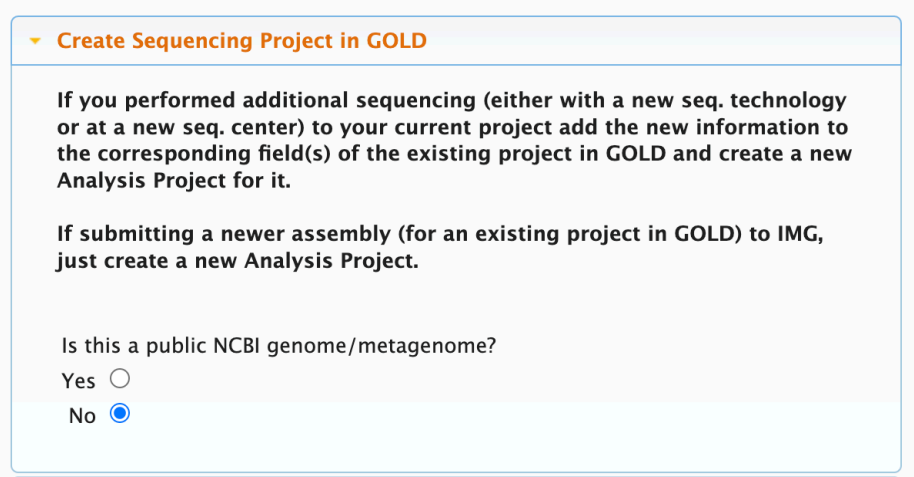


3) Entering a Metagenome or a Metatranscriptome Sequencing Project (SP) and associated Analysis Projects (APs).

From the GOLD home page, click the "**Register**" button to go to the landing page for entering a Sequencing and/or an Analysis Project.



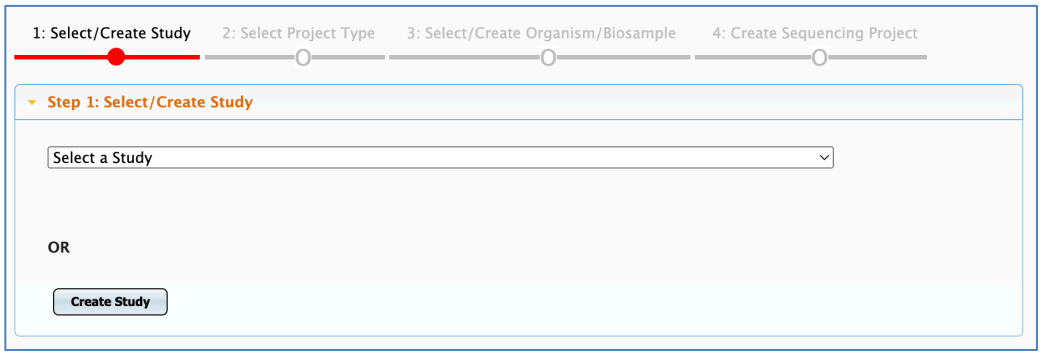
If your Sequencing Project is not yet in GOLD, you will start by clicking the "Create Sequencing Project in GOLD" button to define a new Sequencing Project.



GOLD **does not accept** manually entered public NCBI genomes/metagenomes. If you want to analyze a public genome/metagenome, follow detailed instructions available here: [Guidance on submitting public NCBI genomes and metagenomes into GOLD and IMG]. Otherwise, select the radio button 'No' to proceed.

Step 1: Study → Step 2: Select Project Type → Step 3: Biosample → Step 4: Sequencing Project → Step 5: Analysis Project

The first step to create a Sequencing Project is to define your Study. After confirming that you are NOT entering a public NCBI genome/metagenome, you will come to the page (see the screen below) where you will be able either to select an existing study, under which you want to carry out this new sequencing project, or to define a new study. When applicable, choose a matching study from the drop-down menu.

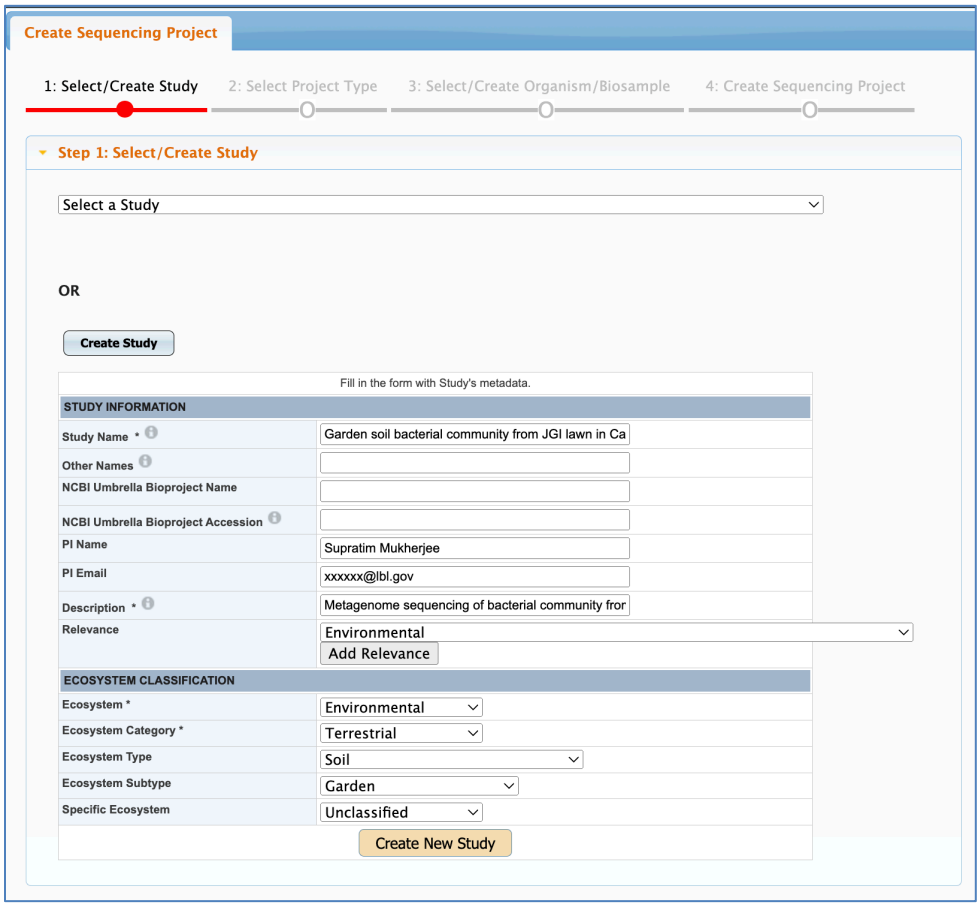


Entering a new Study:

If your new sequencing project is pursued under a new study, click the "Create Study" button to go to the Study entry form. Fill in at least all required fields (*) as shown in the screenshot below. If there is an NCBI BioProject comprising multiple

other BioProjects, enter its Name, ID, and Accession into the corresponding NCBI Umbrella BioProject fields. For example, PRJNA28331 is an Umbrella BioProject for Human Microbiome Project (HMP), and it contains about 3000 of the individual BioProjects. If you do not have an umbrella BioProject, leave these fields blank. Do not enter NCBI BioProject IDs or Accessions for individual BioProjects here at the Study level. .

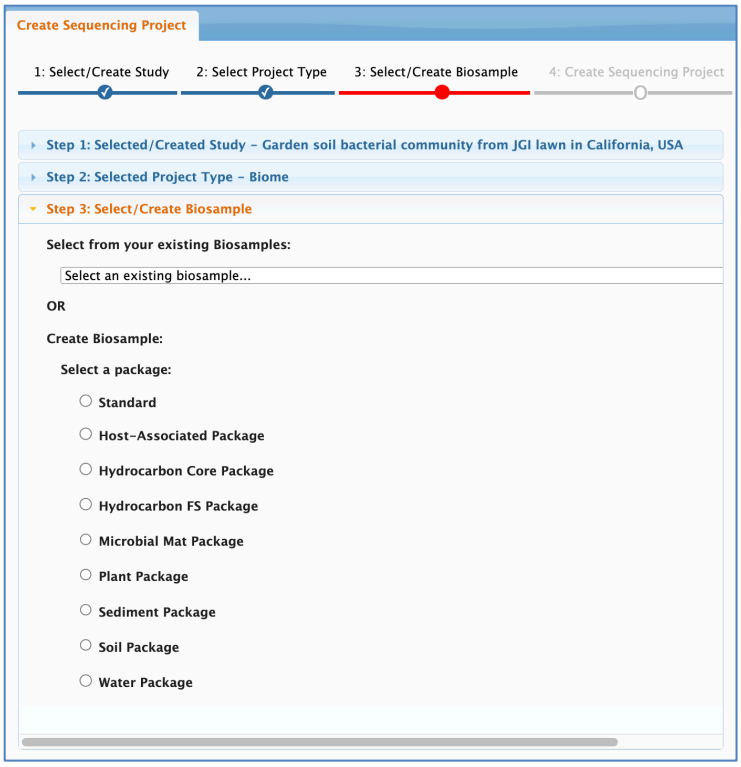
Note that GOLD follows a standardized, canonical naming system for microbiomes which applies to a metagenome Study Name. More information on canonical naming can be found here:
https://gold.jgi.doe.gov/resources/Standardized_Metagenome_Naming.pdf



Click the ‘Create New Study’ button and proceed to the next step, i.e., selecting what kind of project you are about to create.

Selecting an existing Biosample or entering a new Biosample:

If you have not created a Biosample for a specific sample yet, proceed to enter a Biosample. If the sample is from a host-associated, a soil, an aquatic or any other environment listed below, you have an option to select the appropriate package and to get access to the form with additional metadata fields specific to the respective environment. Otherwise, choose the Standard package to access the default Biosample entry form. Note that like an Organism in GOLD represents one specific strain, a Biosample in GOLD represents one specific environmental sample. You cannot create a general Biosample for multiple environmental samples even if they are from the same location.



GOLD follows a standardized, canonical naming system for Biosamples. Detailed information on canonical naming can be found here:
https://gold.jgi.doe.gov/resources/Standardized_Metagenome_Naming.pdf

Fill in at least all required fields in the Biosample entry form. For a metagenome, the fields Habitat, Location, Ecosystem Classification, Sample Collection Site, and few others are mandatory and must be populated to proceed to the next step. But we encourage you to populate the other fields too if you have metadata for them. See the example of the Biosample form for the Soil Package below. The fields specific to the Soil package are highlighted in blue.

Fill in the form with Biosample's metadata

Before proceeding, please read the document GOLD Standardized Metagenome Naming Document and follow the instructions

BIOSAMPLE INFORMATION

Biosample Name *	Garden soil bacterial communities from JGI lawn in Berkeley, California, USA – A1 Habitat / Community / Location / Identifier Garden soil bacterial communities from JGI lawn in Berkeley, California, USA - A1
Other Names	
Habitat *	Garden soil
Community *	bacterial communities
Location *	from JGI lawn in Berkeley, California
Identifier *	A1
NCBI Taxonomy ID *	410658 NCBI Metagenome Taxonomy
NCBI Taxonomy Name	soil metagenome
Biosample Description *	Soil sample collected after two days of rainfall from la

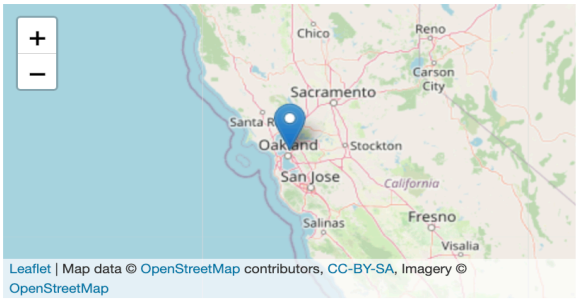
ECOSYSTEM CLASSIFICATION

Ecosystem *	Environmental
Ecosystem Category *	Terrestrial
Ecosystem Type *	Soil
Ecosystem Subtype	Garden
Specific Ecosystem	Unclassified

BIOSAMPLE ISOLATION

Sample Collection Site *	garden soil
Sample Collection Date *	Month (mm): 01 * Day (dd): 20 Year (yyyy): 2020 * Add collection time: Hour (hh): Minute (mm):
Sample Isolation Comments	
Size Fraction	
Sample Isolation Country/Ocean *	USA
Sample Collection Method	
Sample Contact Name *	Supratim Mukherjee
Sample Contact Email *	xxxxxxxx@lbl.gov

ENVIRONMENTAL INFORMATION

Geographic Location *	USA: Berkeley
Lat/Long Lookup	Get Coords
Latitude *	37.877
Longitude *	-122.2456 Reset Map
	 Leaflet | Map data © OpenStreetMap contributors, CC-BY-SA, Imagery © OpenStreetMap
LatLong Information *	Verified
Elevation	129 Meters. to Meters. Add Elevation Range
Depth	Meters. to Meters. Add Depth Range
Subsurface Depth	Meters. to Meters. Add Subsurface Depth Range
Sample Collection Temperature	55 Celsius. to Celsius. Add Sample Collection Temperature Range
Salinity (MIGS-6.3)	Select from below...
Pressure	
pH	8 to Add pH Range
Salinity Concentration	

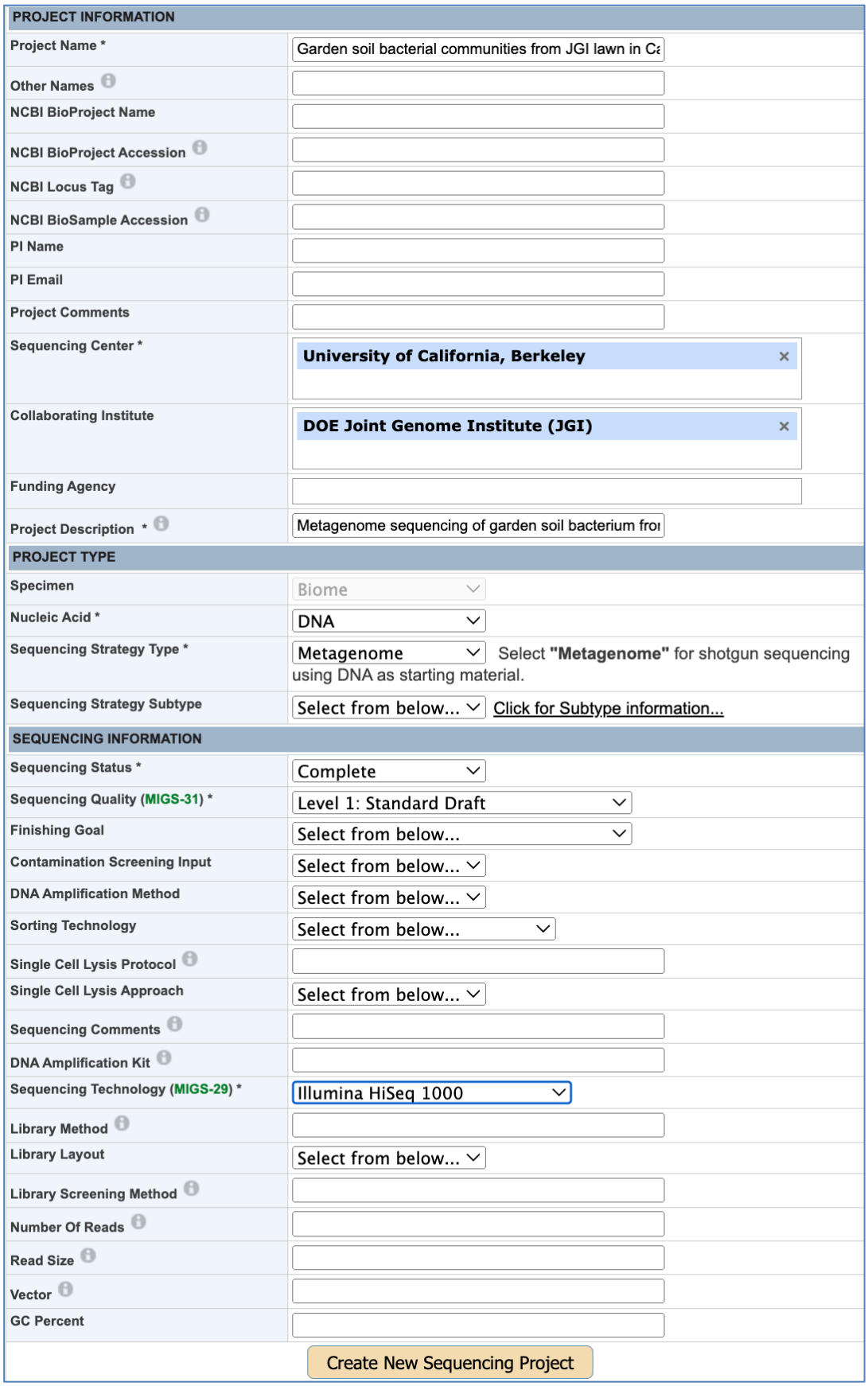
Create New Biosample

SOIL ENVIRONMENTAL INFORMATION	
Mean Annual Precipitation	25
Mean Seasonal Precipitation	
Mean Annual Temperature	15.4
Mean Seasonal Temperature	
Link Climate Info	
LAND USE & HISTORICAL DATA	
Crop Rotation	
Current Land Use	Shrub Land
Current Vegetation	Perrenial Grass and Coastal Shrubs
Current Vegetation Method	
Fire	
Flooding	No
Extreme Event	
Previous Land Use	
Previous Land Use Method	
PHYSICOCHEMICAL PROPERTIES	
Drainage Class	Moderately Well
FAO Class	Select from below...
Horizon	A Horizon
Horizon Method	
Link Class Info	
Local Class	
Local Class Method	
Profile Position	Backslope
Sieving	
Slope Aspect	W-SW
Slope Gradient	32

After defining a Biosample, proceed to the next step, i.e., creating a metagenome SP.

Entering a Metagenome Sequencing Project:

To define a new metagenome Sequencing Project, choose DNA as a nucleic acid and Metagenome as a sequencing strategy type and populate at least all required fields (*) such as Sequencing Center, Project Description, Sequencing Strategy Type, Sequencing Technology, etc. as shown in the screenshot below. After that, click the “Create New Sequencing Project” button.



PROJECT INFORMATION	
Project Name *	Garden soil bacterial communities from JGI lawn in Ca
Other Names	
NCBI BioProject Name	
NCBI BioProject Accession	
NCBI Locus Tag	
NCBI BioSample Accession	
PI Name	
PI Email	
Project Comments	
Sequencing Center *	University of California, Berkeley
Collaborating Institute	DOE Joint Genome Institute (JGI)
Funding Agency	
Project Description *	Metagenome sequencing of garden soil bacterium fro
PROJECT TYPE	
Specimen	Biome
Nucleic Acid *	DNA
Sequencing Strategy Type *	Metagenome — Select **"Metagenome"** for shotgun sequencing using DNA as starting material.
Sequencing Strategy Subtype	Select from below... — Click for Subtype information...
SEQUENCING INFORMATION	
Sequencing Status *	Complete
Sequencing Quality (MIGS-31) *	Level 1: Standard Draft
Finishing Goal	Select from below...
Contamination Screening Input	Select from below...
DNA Amplification Method	Select from below...
Sorting Technology	Select from below...
Single Cell Lysis Protocol	
Single Cell Lysis Approach	Select from below...
Sequencing Comments	
DNA Amplification Kit	
Sequencing Technology (MIGS-29) *	Illumina HiSeq 1000
Library Method	
Library Layout	Select from below...
Library Screening Method	
Number Of Reads	
Read Size	
Vector	
GC Percent	

Create New Sequencing Project

In addition to the traditional metagenome sequencing project, there are two subtypes of the metagenome SPs that are available to a user to choose from, based on how the sample was isolated. These two subtypes are:

i) **Metagenome - Cell Enrichment:** A draft metagenome assembly derived from a cell enrichment (> 1 cell) sample. A cell enrichment is generally obtained by physical separation of a biologically relevant unit, such as microcolonies. Due to the low biomass for cell enrichments, the extracted DNA is typically amplified using whole-genome amplification prior to sequencing.

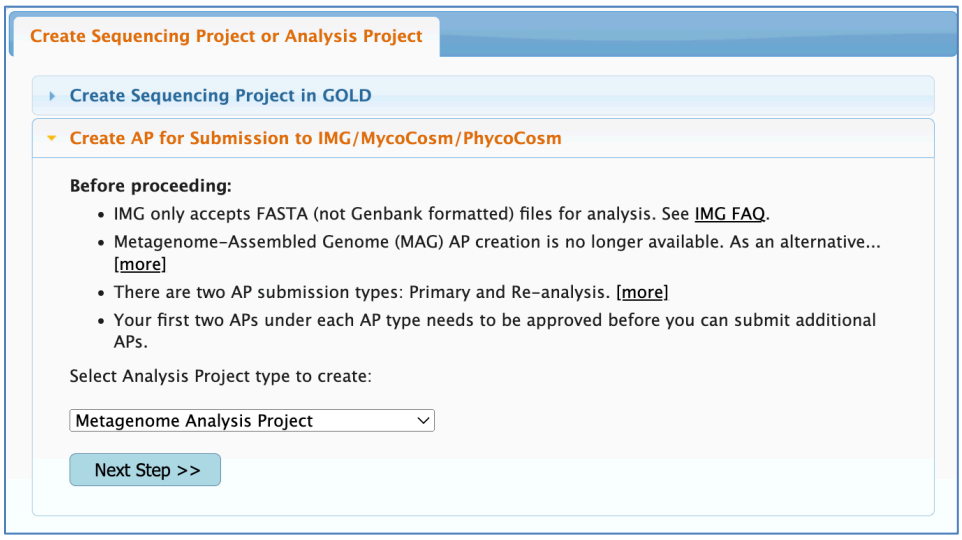
ii) **Metagenome - Single Particle Sort:** A draft genome or metagenome assembly derived from a single particle isolated via flow cytometry. A single particle sort can consist of a single cell or an aggregate of multiple cells, not necessarily of the same phylogenetic background. The extracted DNA is amplified using whole-genome amplification prior to sequencing. No amplicon-based 16S rRNA gene information is available for single particle sorts.

Entering a Metatranscriptome Sequencing Project:

To create a metatranscriptome SP, follow the same steps described above for a metagenome SP, but choose RNA as a nucleic acid and Metatranscriptome as a sequencing strategy type.

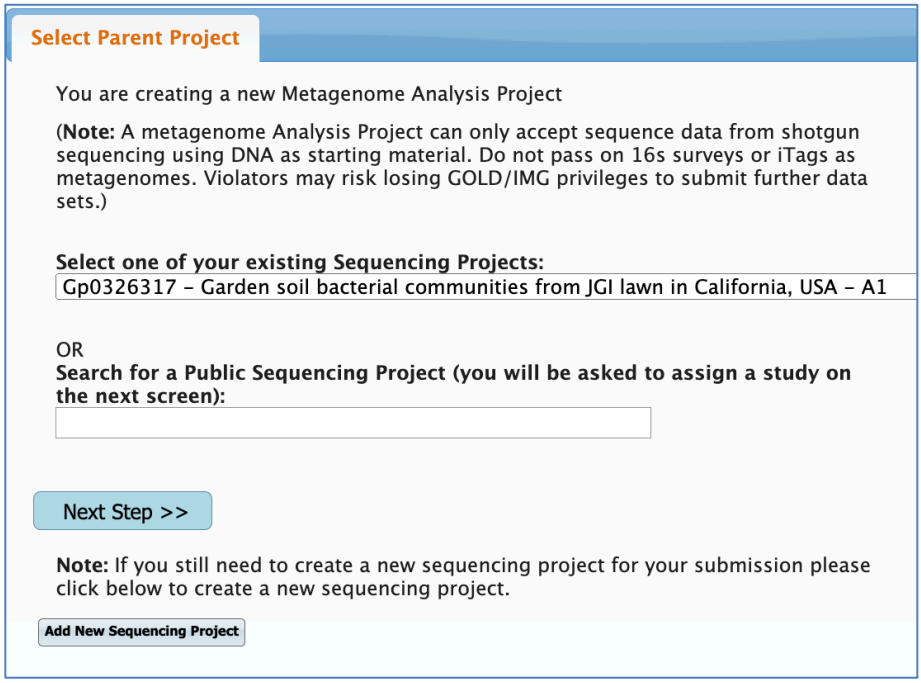
Entering a Metagenome Analysis Project:

Go to the main project submission page and click the “Create AP for Submission to IMG/MycoCosm/PhycoCosm” button. Select “Metagenome Analysis Project” (or one of the subtypes if applicable: Metagenome – Cell Enrichment or Metagenome - Single Particle Sort Analysis Project) from the drop-down menu and click the “Next Step” button.



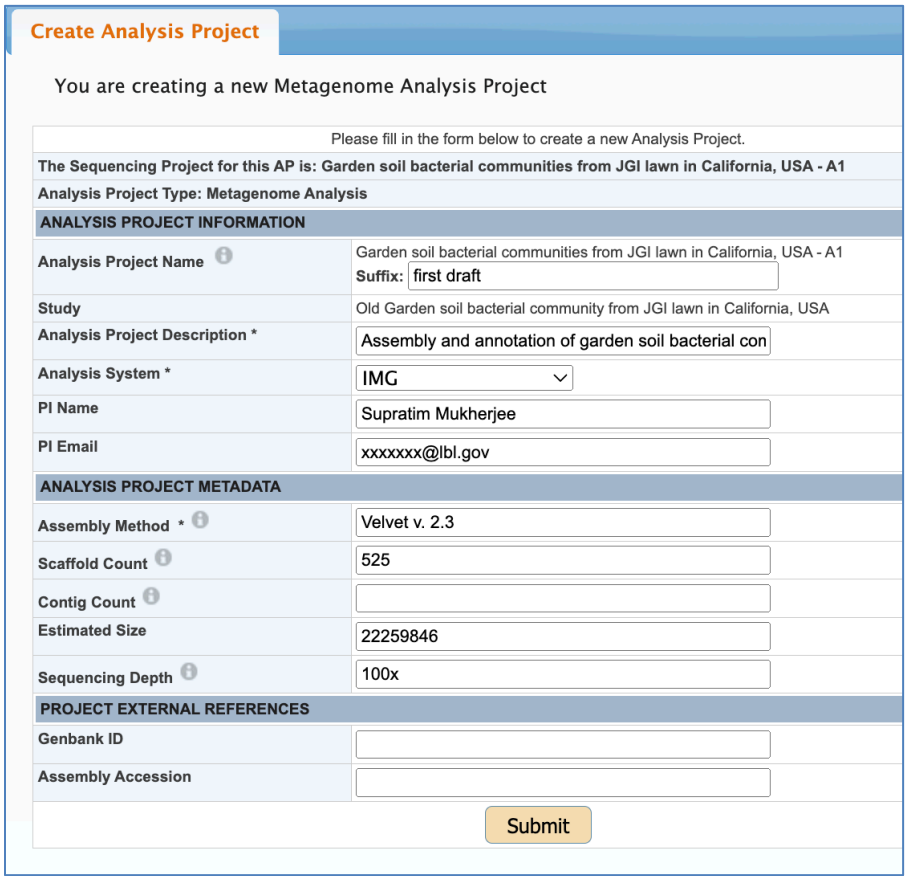
Like in the case of the Genome Analysis Projects described earlier, you can create a Metagenome Analysis Project for your own Sequencing Project that you have created earlier or for a public Sequencing Project in GOLD if you have reassembled the associated public sequences and want to re-annotate them.

To create an AP for your own metagenome sequencing project, select a corresponding SP from the drop-down menu, as shown in the screenshot below and proceed to the next step.

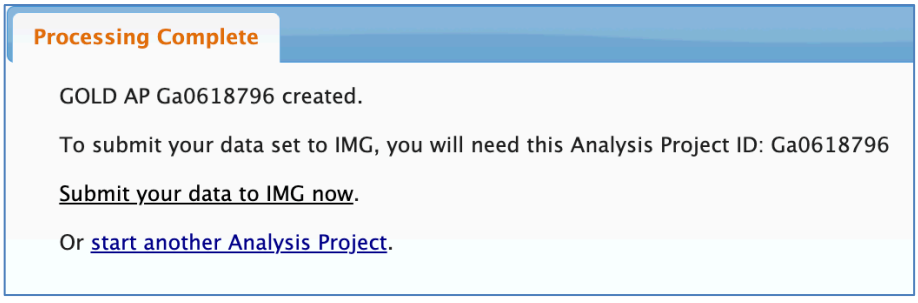


Choose the corresponding metagenome SP from the drop-down menu.

Fill in at least the required fields of the Analysis Project form to define an AP for your metagenome.

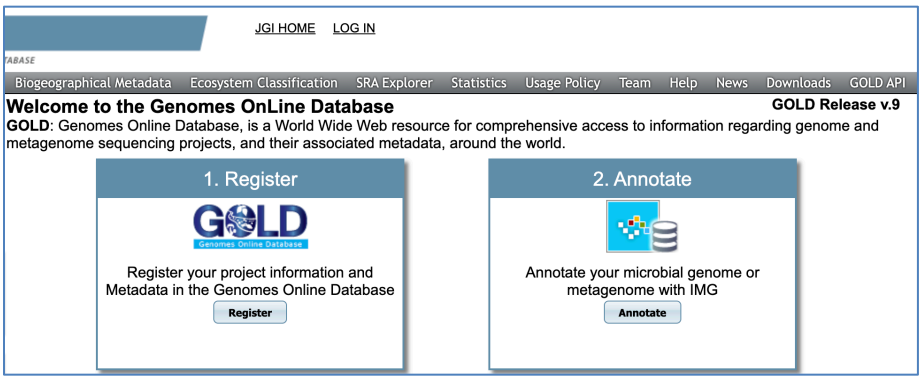


Once you submit the above form, a Metagenome Analysis Project will be created and you will see the following message:



Click the link to proceed to IMG for data submission.

If you are not ready to submit your data to IMG yet, you can return later and use your AP for IMG submission. You will be redirected to the IMG submission page when you click the “Annotate” button on the GOLD home page.



Steps for creating a Metagenome AP for a public metagenome SP are the same as for a public isolate genome SP. See the corresponding section above.

Entering a Metatranscriptome Analysis Project:

The steps for creating a Metatranscriptome AP after you select “Metatranscriptome Analysis Project” from the drop-down menu on the submission page are the same as for creating a Metagenome AP.
Note that Metatranscriptome APs can be created only for your own Metatranscriptome SPs.

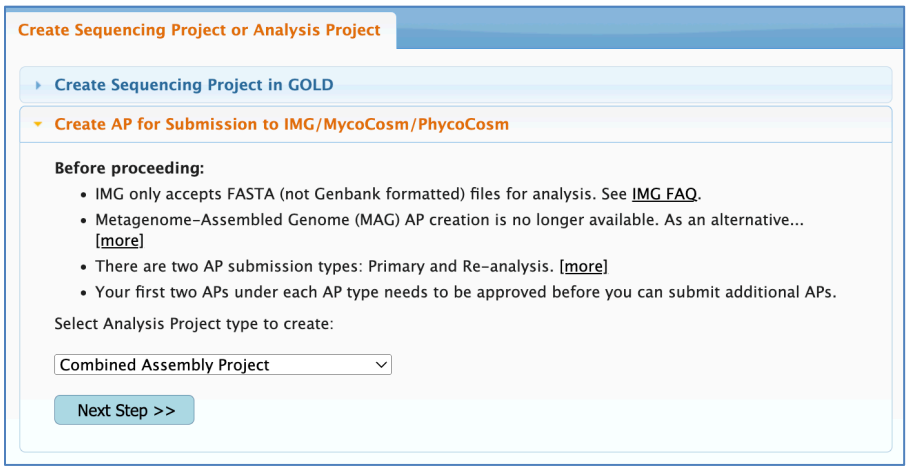
5) Creating a Combined Assembly Analysis Project (AP) in GOLD.

When you combine sequence data from more than one sequencing project and generate a single assembly, it is called combined assembly. Currently we support five types of combined assemblies. They can be between sequencing projects of the same type or between sequencing projects of different types:

1) Metagenomic projects
2) Metagenomic project with Single-Cells
3) Metatranscriptome projects
4) Single-Cell projects
5) Sequencing Projects from a single Organism

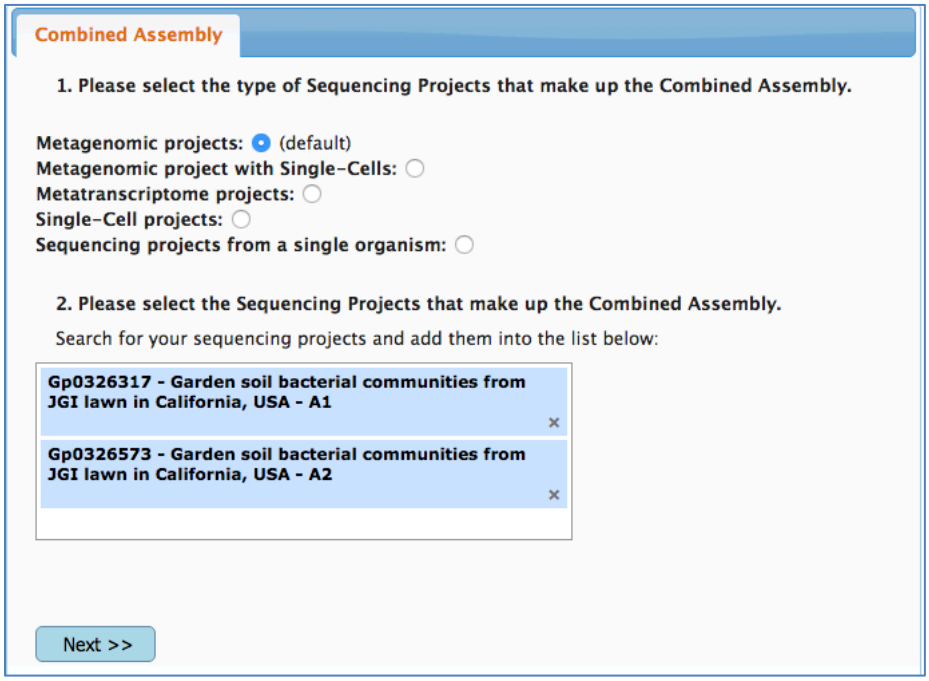
5a) Creating a metagenomic Combined Assembly Analysis Project (AP).

Use case: Say, you have two or more metagenome SPs in GOLD, and you are hoping to get a better assembly by doing a combined assembly of the sequences you have obtained under more than one SP. Then you need to define a Combined Assembly AP in GOLD.
First, select the “Combined Assembly” AP type from the drop-down menu and proceed to the next step.

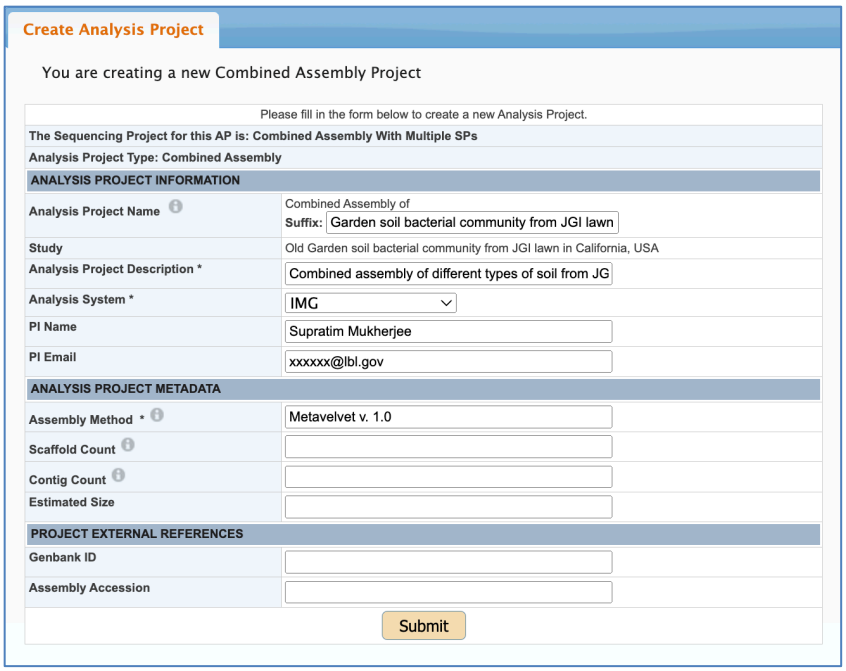


The “Metagenomic projects” radio button is preselected by default.

Using the search box, choose two or more metagenome SPs representing the sequences that you have used for your combined assembly as shown below.



Fill in at least the required fields of the Combined Assembly Analysis Project form. If the SPs you have used in your combined assembly came from two different studies, at this point, you will be able to select the study under which you want your new AP to appear.



Once you create a Combined Assembly AP, you can proceed to IMG for submitting your data, or you can leave and return later to submit your dataset (see page 34-35 for more details).

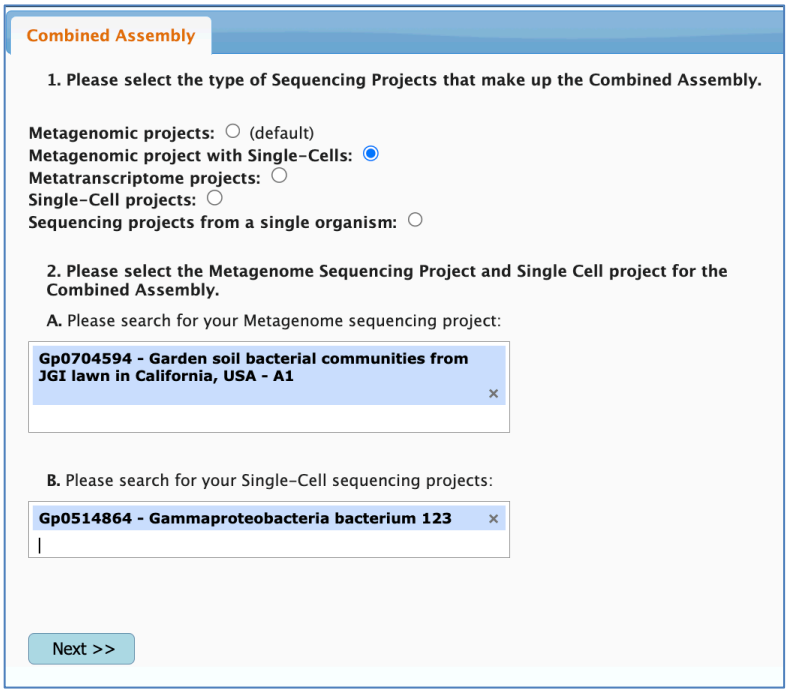
Combined Assembly APs for metatranscriptome projects can be created in a similar manner.

5b) Creating a metagenomic and single cell Combined Assembly Analysis Project (AP).

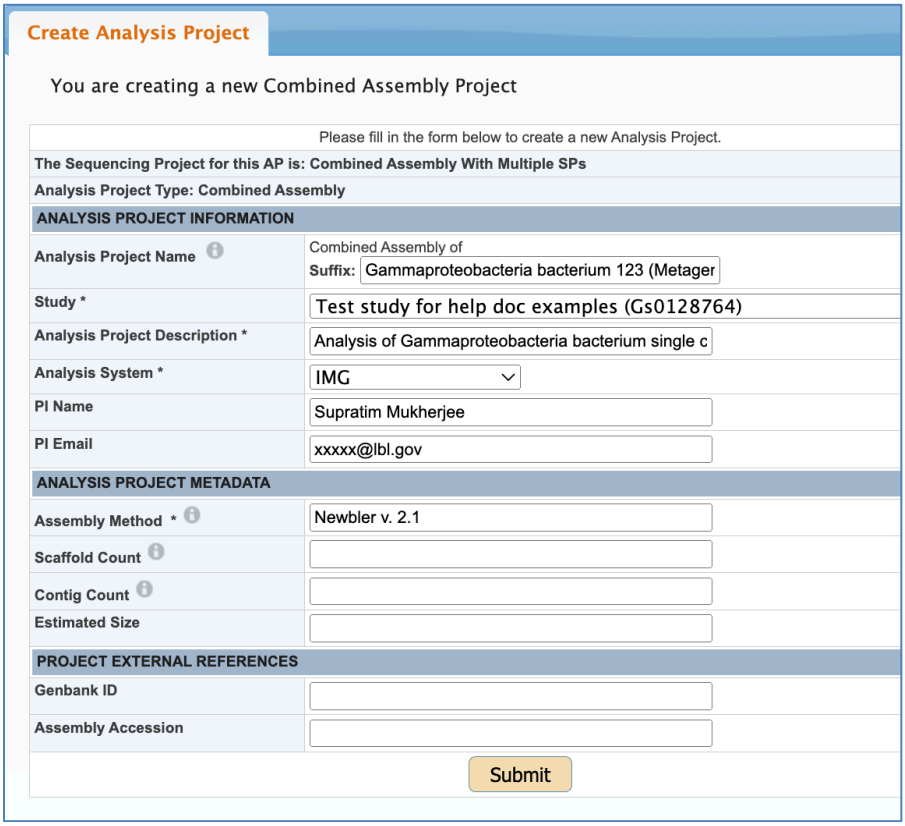
Use case: Say you had sequenced a single cell genome, but it was a partial assembly. Then you found hits in one of your metagenomes to this single cell genome and decided to take advantage of the available metagenome sequence to get a better assembly for your single cell genome. In this case, you need to define a Combined Assembly AP that includes a metagenome and a single cell SP.

First, select the “Combined Assembly” AP type from the drop-down menu and proceed to the next step.

Select the “Metagenomic projects with Single-Cells” radio button as shown in the screen below.
Using the search boxes (separate boxes for metagenome SPs and single cells SPs are provided), choose one or more metagenome and single cell genome sequencing projects that you have used for your combined assembly. Once you select all component SPs of your combined assembly, proceed to the next step.



Fill in at least the required fields. If the SPs you have used in your combined assembly are from two different studies, at this point, you will be able to select the study under which you want your new AP to appear.



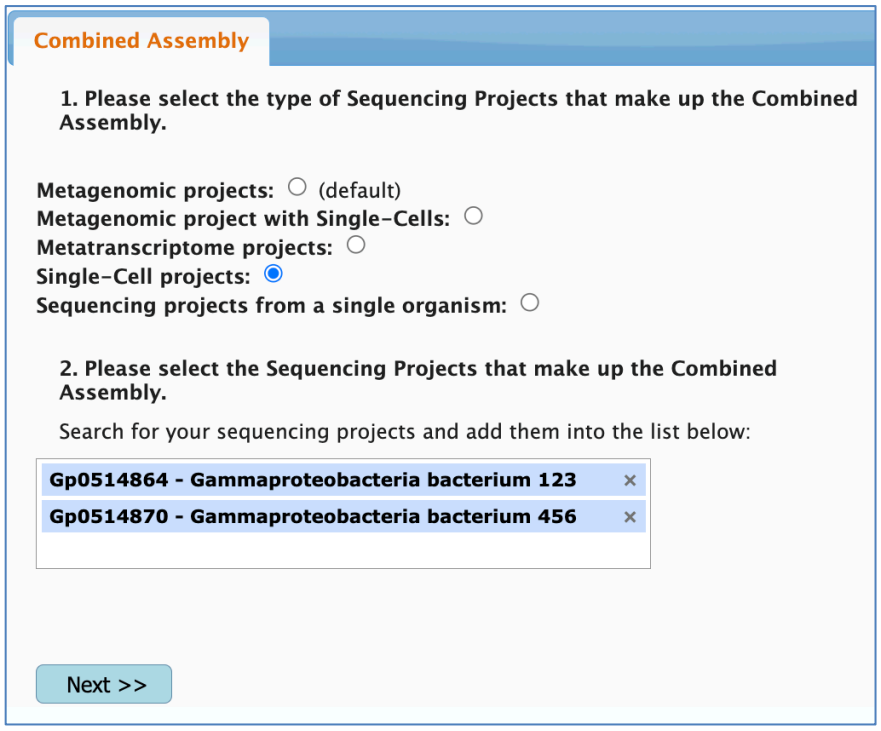
Once you create a Combined Assembly AP, proceed to IMG for submitting your data or return later and submit your dataset, using the “Annotate” button on the GOLD home page (see page 34-35 for more details).

5c) Creating a single cell Combined Assembly Analysis Project (AP).

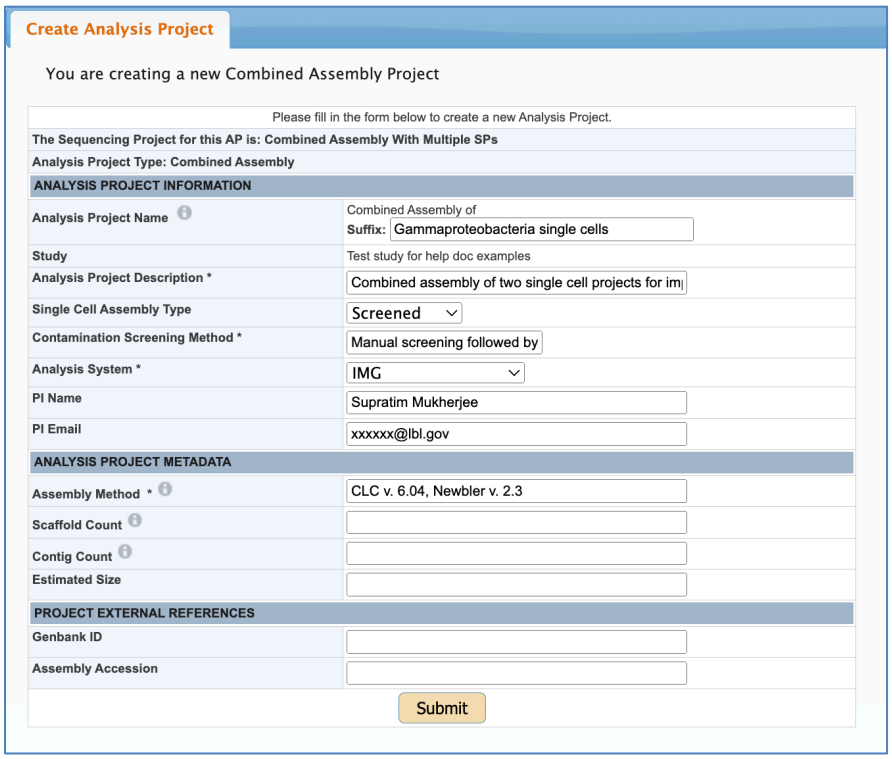
Use case: Say you had sequenced several single cell genomes and assembled them separately. But each one of those ended up being a partial assembly. Then you found that two of the single cell genomes you had sequenced appeared to be very similar. So, you decided to do a combined assembly, using sequence data from two individual single cell genome projects. In this case, you need to define a Combined Assembly AP that includes both single cell SPs.

First, select the “Combined Assembly” AP type from the drop-down menu and proceed to the next step.

Select the “Single-Cell projects” radio button as shown in the screen below.



Using the search box, choose two or more single cell genome sequencing projects that you have used for your combined assembly as shown above. Once you selecte all component SPs of your combined assembly, proceed to the next step.

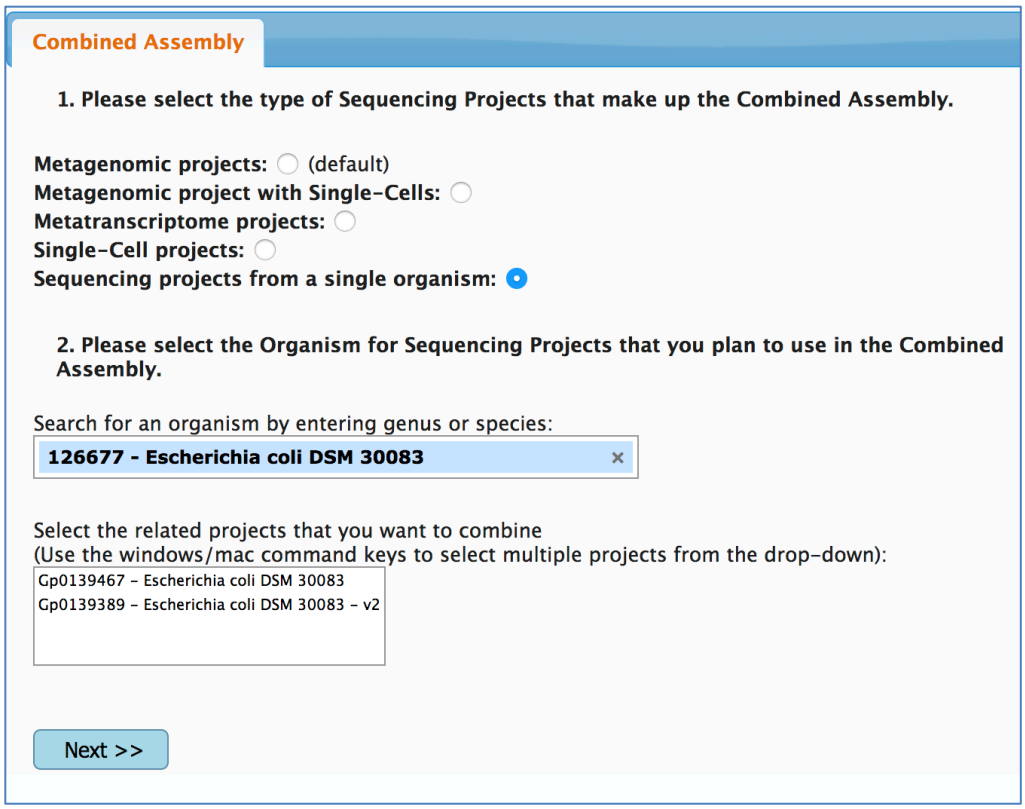


Fill in at least the required fields. If the SPs you have used in your combined assembly are from two different studies, at this point, you will be able to select the study under which you want your new AP to appear.

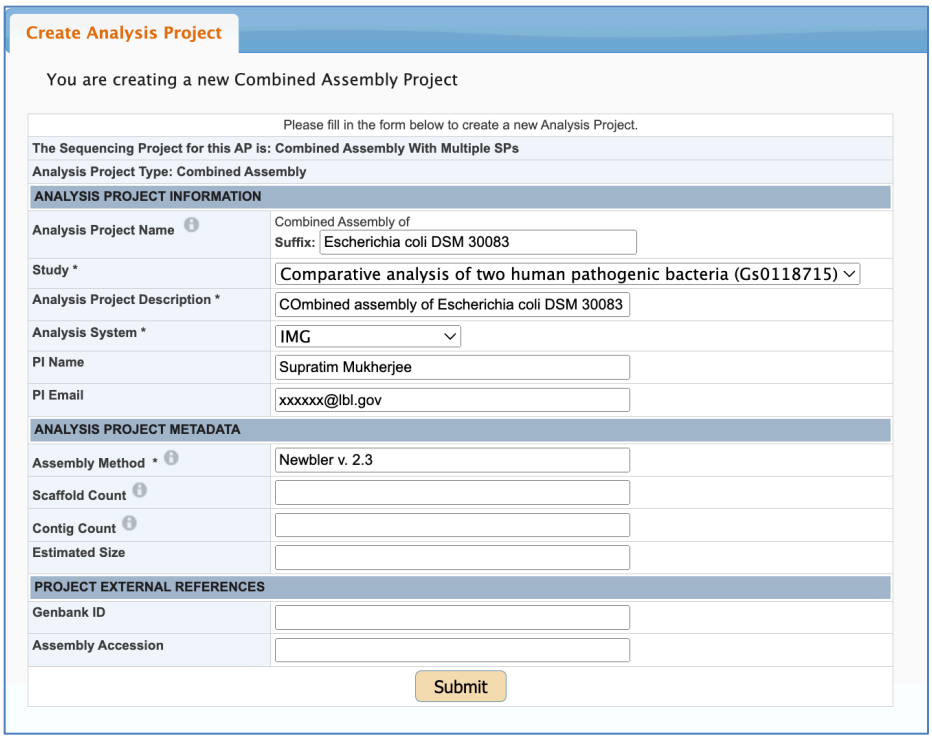
Once you create a Combined Assembly AP, proceed to IMG for submitting your data or return later and submit your dataset, using the “Annotate” button on the GOLD home page (see page 34-35 for more details).

5d) Creating a Combined Assembly Analysis Project (AP) with sequencing projects from a single organism.

Use case: Say you had sequenced a single organism in your lab but did not get a good assembly. You then came across a public SP for the same organism in GOLD. You now have the option to combine sequence data from your SP and the data from the public sequencing project of the same Organism to create a combined assembly AP. This option is only available if there are two or more Sequencing Projects for exactly the same strain of a single GOLD Organism. To define such a combined assembly AP, select the “Combined Assembly” AP type from the drop-down menu and proceed to the next step.



Select the “Sequencing projects from a single organism” radio button as shown in the image above. In the search bar, enter the name of the organism for which you want to create the combined assembly AP. Note that only those Organisms that have two or more Sequencing Projects in your GOLD user account will show up. Once you select the Organism of your choice, the Sequencing Projects associated with it will be displayed in the box below. Select the Sequencing Projects that you want to include in your Combined Assembly AP and proceed to the next step.



Fill in at least the required fields as shown above. If the SPs you have used in the combined assembly are from two different Studies, at this point, you will be able to select the study under which you want your new AP to appear. Click the “Submit” button to create a new AP. Once you create a Combined Assembly AP, proceed to IMG for submitting data or return later and submit your dataset, using the “Annotate” button on the GOLD home page (see page 34-35 for more details).