# Guidance on importing NCBI genomes/metagenomes into GOLD/IMG

For all public isolate genome and metagenome projects at GenBank, we have a process in place to import those projects into GOLD and annotate them in IMG. Project information and metadata are curated in GOLD. GenBank and SRA data are annotated in IMG.

We **do not accept** such projects manually entered by users.

We regularly import isolate and metagenomes from NCBI. However, we may not have all of them in GOLD/IMG yet.

For any public GenBank genome, these are the possible scenarios:

1) The genome is already in IMG.
2) Genome analysis project is defined in GOLD but not yet annotated in IMG.
3) We don't have the specific genome (sequencing project/analysis project) in GOLD yet.

What should you do if you want to analyze a public genome from GenBank?

First of all, you need to check whether we already have the genome of your interest in IMG or an Analysis Project in GOLD. The best way to do that is to perform a search by NCBI BioSample Accession as shown in Figure 1.

Step 1 - Click on the number of Sequencing Projects in the top left menu table;

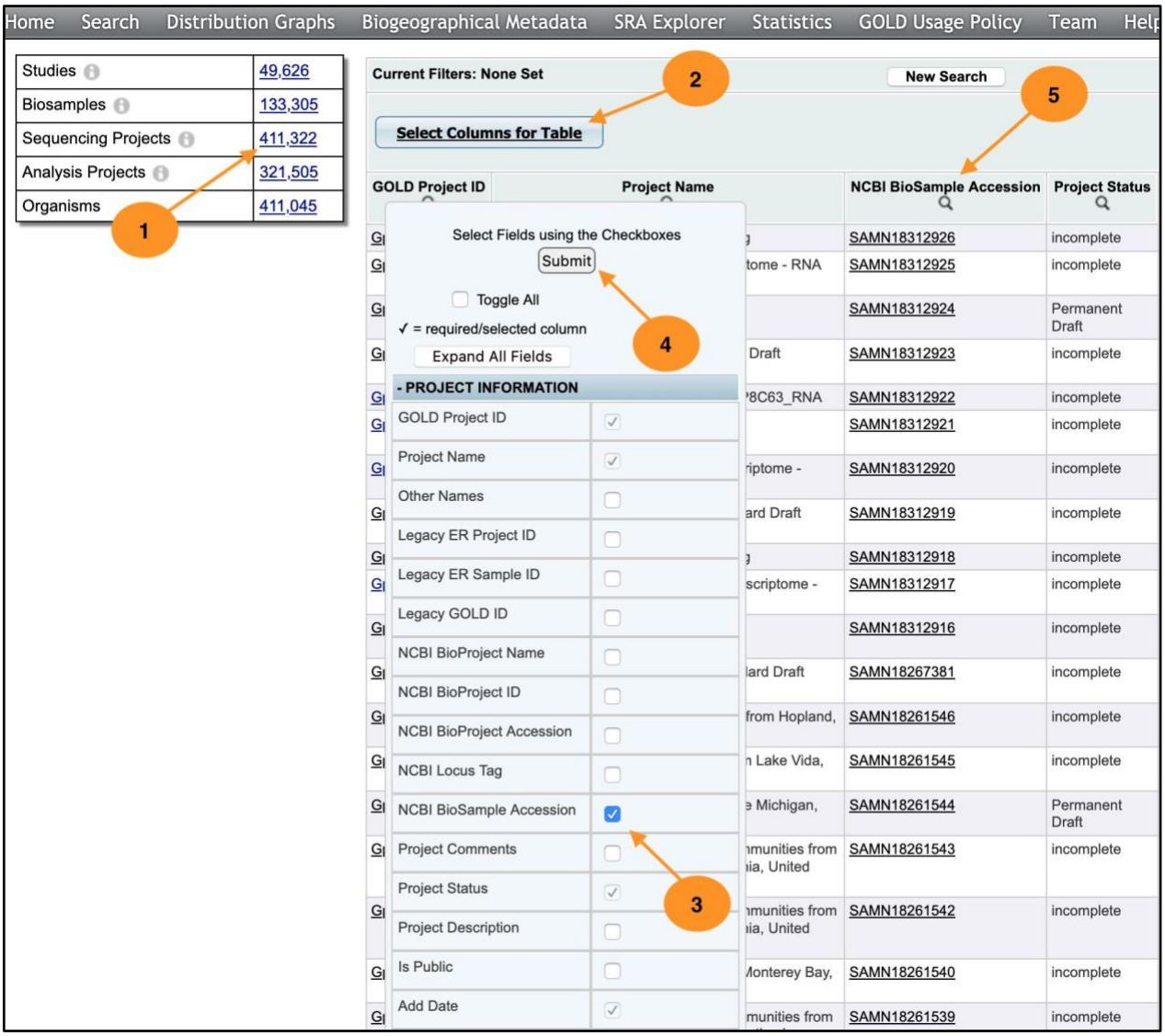
Step 2 - Click on the "Select Columns for Table" button;

Step 3 - Expand Project Information section and select NCBI Biosample Accession field;

Step 4 - Click on the "Submit" button;

Step 5 - Perform the search, using the BioSample accession of your interest.

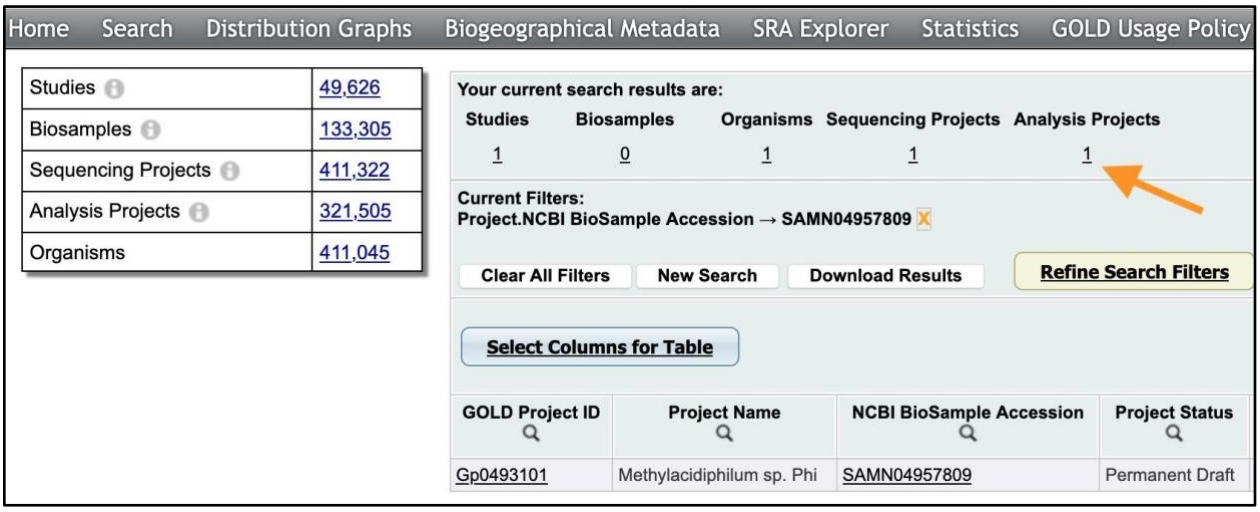
**Figure 1**




For example, you are interested in the public NCBI genome *Methylacidiphilum sp. Phi* with BioSample Accession SAMN04957809. The search returns the seq. project with ID Gp0493101.

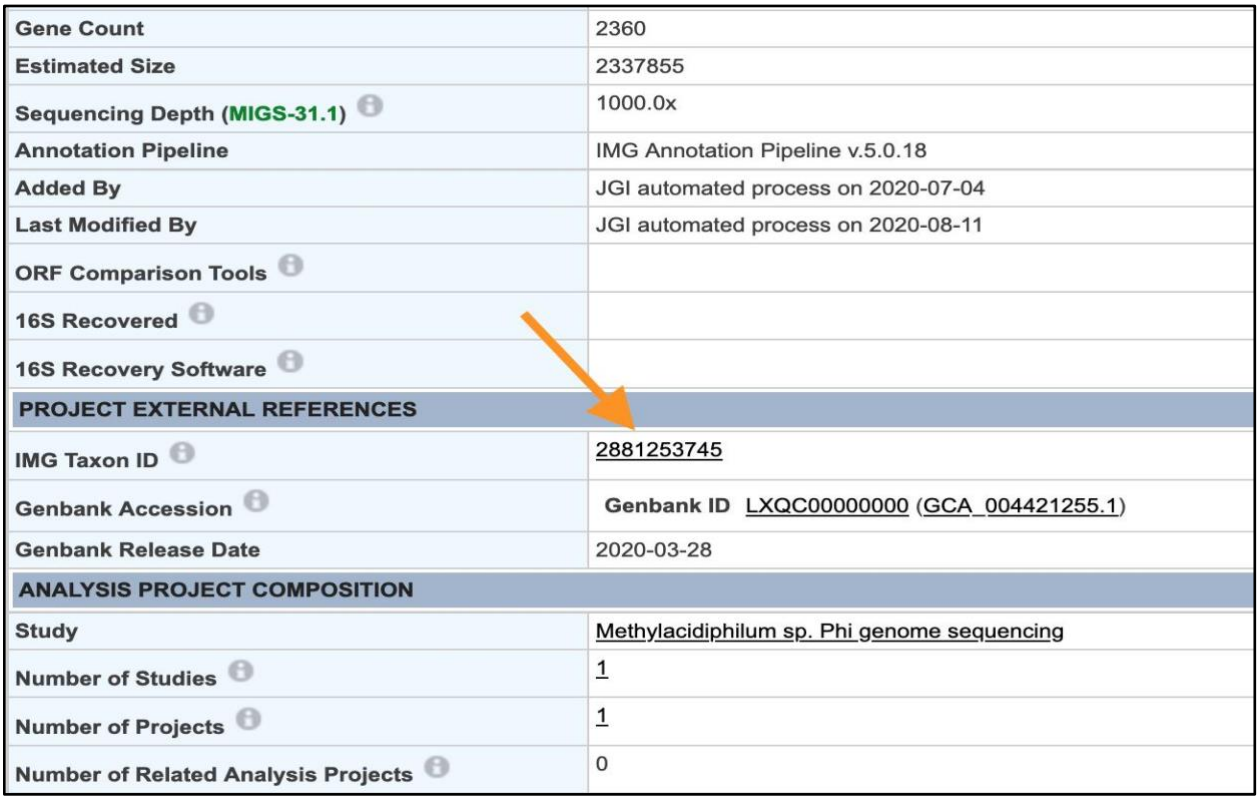
Click on the number of the Analysis Projects under the “Your current search results are:" menu (see Figure 2).

**Figure 2**




Click on Analysis Project ID that will open the Analysis Project page. Locate the field IMG Taxon ID at the bottom of the page (see Figure 3). The field is populated with a number 2881253745. (Case 1. The genome is already in IMG.)

**Figure 3**



| | |
|---|---|
| Gene Count | 2360 |
| Estimated Size | 2337855 |
| Sequencing Depth (MIGS-31.1) | 1000.0x |
| Annotation Pipeline | IMG Annotation Pipeline v.5.0.18 |
| Added By | JGI automated process on 2020-07-04 |
| Last Modified By | JGI automated process on 2020-08-11 |
| ORF Comparison Tools | |
| 16S Recovered | |
| 16S Recovery Software | |
| **PROJECT EXTERNAL REFERENCES** | |
| IMG Taxon ID | 2881253745 |
| Genbank Accession | Genbank ID LXQC00000000 (GCA_004421255.1) |
| Genbank Release Date | 2020-03-28 |
| **ANALYSIS PROJECT COMPOSITION** | |
| Study | Methylacidiphilum sp. Phi genome sequencing |
| Number of Studies | 1 |
| Number of Projects | 1 |
| Number of Related Analysis Projects | 0 |

**Action**: Click on the number, and link will take you to the IMG page for this genome.

Let's say another search returns a sequencing project, but the field IMG Taxon ID of the associated analysis project is blank. It means that though we already have the analysis project in place for this particular genome, we haven't annotated this genome in IMG yet (Case 2).

**Action**: Send us this AP (or APs) via Contact Us form under the Help section of the top menu, and we will prioritize it (them) for you.

There can be a situation where the search does not return a sequencing project. That means that we do not have a sequencing/analysis project in GOLD yet (Case 3).

**Action**: Send us NCBI BioProject and BioSample Accessions for the specific genome (or a list of Accessions for multiple genomes).

If you want to submit **your own assembly** of a public genome/metagenome that is not in IMG, please contact us.