

Metagenome Bin Search User Guide

5/6/2020

IMG has developed a new Metagenome Bin Search functionality. This new feature is available from IMG/M or IMG/M ER main menu: **Find Genomes -> Metagenome Bins -> Bin Search**.

Metagenome Bin Search

Quick Search | Advanced Search Builder

Find metagenome bins by keyword or substring. [Examples](#)

- Please use "Search Parameters" to refine the search.
 - Search Parameters: *Search by ID* - for single or multiple comma separated values.
 - Search Parameters: *Search by Name* - for single keyword search.

Search Parameters

Search by ID: IMG Metagenome Bin ID | Search by Name: ---

Reset | Search

Search History

- Cart data could be lost when logging out or closing the browser, but can be saved into your [workspace](#).
- For examples, view [public list](#) to save the selected into your cart.

Select	Time	Query	MER-FS Metagenome
--------	------	-------	-------------------

Figure 1. Metagenome Bin Search main page

The new search allows IMG users to perform quick search by IDs or names, or to search metagenome bins using a more advanced query builder.

Quick Search

Quick search allows users to search metagenome bins associated with the following ID fields:

- IMG Metagenome Bin ID
- IMG Metagenome ID
- IMG Submission ID
- GOLD Analysis Project ID
- GOLD Sequencing Project ID
- GOLD Study ID
- ITS AP ID (JGI ITS Analysis Project ID)

- ITS SP ID (JGI ITS Sequencing Project ID)
- Bin Scaffold ID

And the following name fields:

- Bin Taxonomy
 - NCBI Domain
 - NCBI Phylum
 - NCBI Class
 - NCBI Order
 - NCBI Family
 - NCBI Genus
 - NCBI Species
- GTDBTK Taxonomy
 - GTDBTK Domain
 - GTDBTK Phylum
 - GTDBTK Class
 - GTDBTK Order
 - GTDBTK Family
 - GTDBTK Genus
 - GTDBTK Species
- Sequencing Assembly Annotation
 - Sequencing Center
 - Funding Agency

For example, in order to find all Cyanobacteria bins, select “NCBI Phylum” in the **Search by Name** dropdown list and enter “Cyanobacteria” in the search keyword field.

Metagenome Bin Search

Quick Search Advanced Search Builder

Find metagenome bins by keyword or substring. [Examples](#)

- Please use "Search Parameters" to refine the search.
 - Search Parameters: *Search by ID* - for single or multiple comma separated values.
 - Search Parameters: *Search by Name* - for single keyword search.

Cyanobacteria

▼ Search Parameters

Search by ID Search by Name

--- NCBI Phylum

Reset Search

Figure 2: Quick Search: Cyanobacteria metagenome bin search
 After clicking the **Search** button, the result shows 719 bins from 538 metagenomes:

The screenshot shows the 'Metagenome Bin Search Results' page in a web browser. The query is '(NCBI Phylum [Cyanobacteria])'. It displays 719 entries, with the first 10 shown. The table includes columns for Bin ID, Genome Name, IMG Genome ID, Bin Quality, Bin Lineage, GTDBTK Lineage, Bin Methods, Created By, Date Added, and Bin Completeness.

Bin ID	Genome Name	IMG Genome ID	Bin Quality	Bin Lineage	GTDGTK Lineage	Bin Methods	Created By	Date Added	Bin Completeness
2014031008_1	Hot spring microbial communities from Yellowstone National Park, Wyoming, USA - YNP7 Chocolate Pots	2014031008	MQ	Bacteria; Cyanobacteria; unclassified; Synechococcales; Synechococcaceae; Synechococcus; Synechococcus sp. JA-2-3B/a(2-13)	Bacteria; Cyanobacteriota; Cyanobacteria; Euryococcales; Leptococccaceae; Leptococcus; GCF_000013225.1	MetaBAT version 0.32.4, CheckM v1.0.11, GTDB database release 86, GTDB-ik version v0.1.6	IMG_PIPELINE	2018-08-15	86.6
2049941002_1	Sinkhole freshwater microbial communities from Lake Huron, US, Sample 419	2049941002	MQ	Bacteria; Cyanobacteria; unclassified; Oscillatoriales	Bacteria; Cyanobacteriota; Cyanobacteria; Cyanobacteriales; Phormidiaceae; Tychonema	MetaBAT version 0.32.4, CheckM v1.0.11, GTDB database release 86, GTDB-ik version v0.1.6	IMG_PIPELINE	2018-08-15	90.66
2049941002_2	Sinkhole freshwater microbial communities from Lake Huron, US, Sample 419	2049941002	MQ	Bacteria; Cyanobacteria; unclassified; Oscillatoriales; Microcoleaceae;	Bacteria; Cyanobacteriota; Cyanobacteria; Cyanobacteriales; Phormidiaceae; Planktotothrix;	MetaBAT version 0.32.4, CheckM v1.0.11, GTDB database release 86, GTDB-ik version v0.1.6	IMG_PIPELINE	2018-08-15	78.88

Figure 3: Cyanobacteria bin result page

To search all metagenome bins obtained from Tara Oceans metagenomes, simply enter “Tara Oceans” in the search keyword field, and select “Metagenome Name” in the **Search by Name** dropdown list.

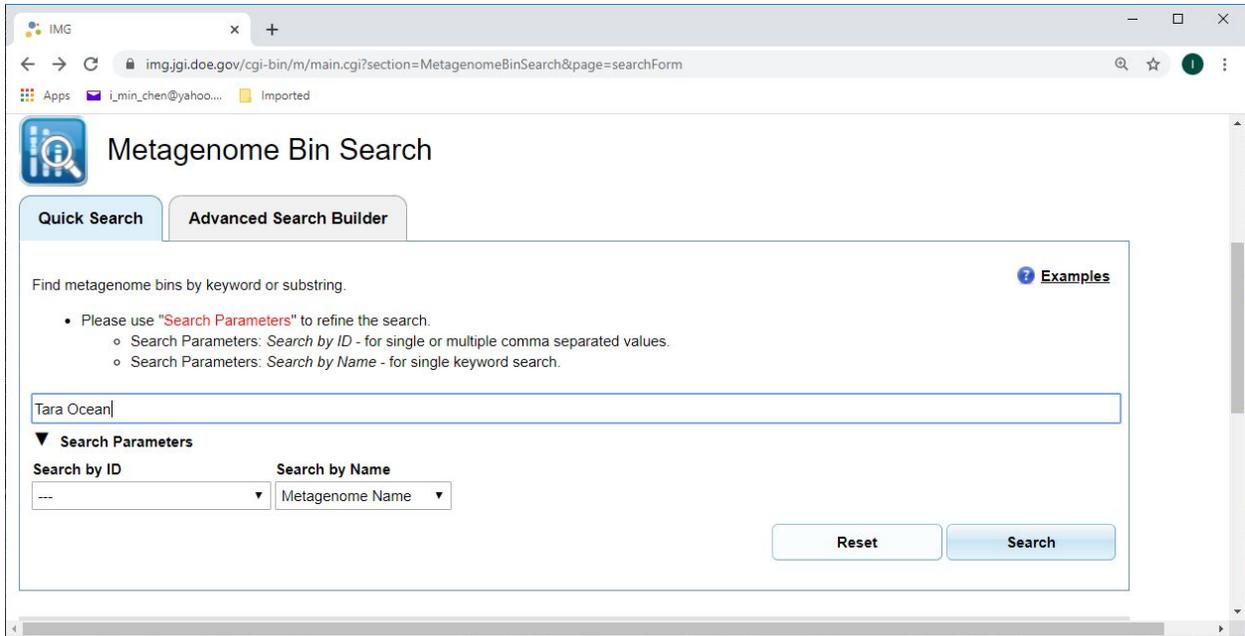


Figure 4: Quick Search: bins obtained from Tara Oceans metagenomes
After clicking the **Search** button, result will show 1,956 bins obtained from 240 Tara Oceans metagenomes.

The screenshot shows the 'Metagenome Bin Search Results' page. The query is '(undefined [Tara Ocean])'. It shows a table with 10 columns: Bin ID, Genome Name, IMG Genome ID, Bin Quality, Bin Lineage, GTDBTK Lineage, Bin Methods, Created By, Date Added, and Bin Completeness. The first three rows are visible.

Bin ID	Genome Name	IMG Genome ID	Bin Quality	Bin Lineage	GTDGTK Lineage	Bin Methods	Created By	Date Added	Bin Completeness
3300020237_2	Marine microbial communities from Tara Oceans - TARA_A100001011 (ERX291767-ERR318621)	3300020237	MQ	Bacteria	Bacteria; Actinobacteriota; Acidimicrobia; TMED189, TMED189	MetaBAT version 0.32.4, CheckM v1.0.11, GTDB database release 96, GTDB-tk version v0.1.6	IMG_PIPELINE	2018-08-15	59.4
3300020239_6	Marine microbial communities from Tara Oceans - TARA_B100000003 (ERX555909-ERR598959)	3300020239	MQ	Bacteria; Cyanobacteria; unclassified; Chroococcales; unclassified; Synechococcus, sp	Bacteria; Cyanobacteriota; Synechococcales_A; Cyanobiaceae; Synechococcus_C; GCA_001628295.1	MetaBAT version 0.32.4, CheckM v1.0.11, GTDB database release 96, GTDB-tk version v0.1.6	IMG_PIPELINE	2018-08-15	62.5
3300020241_3	Marine microbial communities from Tara Oceans - TARA_A100001011 (ERX291768-ERR318619)	3300020241	MQ	Bacteria	Bacteria; Actinobacteriota; Acidimicrobia; TMED189, TMED189	MetaBAT version 0.32.4, CheckM v1.0.11, GTDB database release 96, GTDB-tk version v0.1.6	IMG_PIPELINE	2018-08-15	52.99

Figure 5: Tara Ocean metagenome bin search result

Note that after a search is done, it will be added to the **Search History** shown at the bottom of Figure 1. We will discuss this feature later at the **Search History** section.

Advanced Search Builder

Advanced Search Builder allows IMG users to search metagenome bins based on the following categories and fields:

- Bin Taxonomy
 - NCBI Domain
 - NCBI Phylum
 - NCBI Class
 - NCBI Order
 - NCBI Family
 - NCBI Genus
 - NCBI Species
 - GTDBTK Domain
 - GTDBTK Phylum
 - GTDBTK Class
 - GTDBTK Order
 - GTDBTK Family
 - GTDBTK Genus
 - GTDBTK Species
 - Any Field
- Metagenome Bin IDs
 - IMG Metagenome Bin ID (list)
 - IMG Metagenome ID (list)
 - IMG Submission ID (list)
 - GOLD Analysis Project ID (list)
 - GOLD Sequencing Project ID (list)
 - GOLD Study ID (list)
 - GPTS Proposal ID (list)
 - ITS AP ID (list)
 - ITS SP ID (list)
 - ITS Proposal ID (list)
 - PMO Project ID (list)
 - Bin Scaffold ID (list)
 - Any Field
- Sequencing Assembly Annotation
 - Add Date
 - Assembly Method
 - Funding Agency
 - Funding Program
 - GOLD Analysis Project Type

- GOLD Sequencing Depth
- GOLD Sequencing Quality
- JGI Analysis Product Name
- JGI Analysis Project Type
- Is Published
- Is Public
- Release Date
- Sequencing Center
- Sequencing Method
- Submission Type
- Any Field
- Study Dataset Names
 - Genome Name / Sample Name
 - Study Name
 - Any Field
- Environmental Classification
 - Ecosystem
 - Ecosystem Category
 - Ecosystem Subtype
 - Ecosystem Type
 - Habitat
 - Host Name
 - Specific Ecosystem
 - Any Field
- Geographic Metadata
 - Altitude in Meters
 - Depth in Meters
 - Geographic Location
 - Isolate Country
 - Latitude
 - Longhurst Code
 - Longhurst Description
 - Longitude
 - Any Field
- Physicochemical Metadata
 - Chlorophyll Concentration
 - Nitrate Concentration
 - Oxygen Concentration
 - pH
 - Pressure
 - Salinity
 - Salinity Concentration
 - Sample Collection Date

- Sample Collection Temperature
 - Any Field
- Human Microbiome Project Metadata
 - HMP ID (list)
 - Host Gender
 - Medical Record Number (list)
 - Sample Body Site
 - Sample Body Subsite
 - Visits
 - Any Field
- Bin Statistics Metadata
 - Bin Size
 - Gene Count
 - Scaffold Count
 - Completeness %
 - Contamination %
 - 5S rRNA Count
 - 16S rRNA Count
 - 23S rRNA Count
 - tRNA Count
 - Bin Method
 - Bin Quality
 - Any Field

Advanced search allows users to search metagenome bins using more complicated queries. For example, in order to search all cyanobacteria from Tara Oceans with > 80% completeness that were found in “deep” samples, we need to add 4 new builder line with the following:

1. Bin Taxonomy - All Field - Cyano
2. Study Dataset Names - Genome Name / Sample Name - Tara Oceans
3. Bin Statistics Metadata - Completeness - >80%
4. Geographic Metadata - Depth in meters - >100

Metagenome Bin Search

Find Genomes by constructing a query using keywords, substrings, and AND/(AND NOT)/OR operators.

- The 'AND' / 'AND NOT' query operator(s) that combine builder lines will be processed sequentially.
- Range queries enabled with keyword 'to' indicate range searches. For example, input '1 to 10' in the field of Geographic Metadata->Depth will retrieve datasets collected at depth between 1 and 10 meters.
- Mathematical operators '>', '>=', '<', '<=' can be applied to a range search. For example, input '>=10' in the field of Geographic Metadata->Depth will retrieve datasets collected at depth of at least 10 meters.
- Comma separated keywords are treated as multiple values joined by 'OR' operators. For example, an input of '2007309000_2, 2012990006_1' in the field of Metagenome Bin ID->IMG Metagenome Bin ID retrieves 2 bins.
- If 'Any Field' selected, search will be performed on each field in the same category, and results will be pooled.
- Each field can only be searched only once as a specific field and multiple times as part of 'Any Field' selection.
- The maximum number of builder lines is limited to 5.
- Click on "Evaluate Query" button to see the results of each specific subquery and the overall count. Click on "Search" to see the overall results.

Bin Taxonomy: Any Field | Cyano - Remove

AND | Study Dataset Names | Genome Name / Sample Name | Tara Oceans - Remove

AND | Bin Statistics Metadata | Completeness % (Range: 50 to 100) | > 80 - Remove

AND | Geographic Metadata | Depth In Meters (Range: -10 to 10970) | > 100 - Remove

+ Add new builder line

Constructed Query: Use the builder above to create a search query.

Reset Evaluate Query Search

Hide All Categories

Figure 6: Advanced Search: Deep sea cyanobacteria bins from Tara Oceans metagenome with at least 80% completeness

The above search will return 3 bins as shown below:

Final Combination: 3 count(s)

Showing 1 to 3 of 3 entries

First Previous 1 Next Last Export Select All Clear All Select - page Deselect - page Show 10

Bin ID	Genome Name	IMG Genome ID	Bin Quality	Bin Lineage	GTDBTK Lineage	Bin Methods	Created By	Date Added	Bin Completeness	Contigs
3300020348_13	Marine microbial communities from Tara Oceans - JARA_B100000676 (ERR556099-ERR598161)	3300020348	MQ	Bacteria; Cyanobacteria; unclassified; Chroococcales; unclassified; Synechococcus, sp.	Bacteria; Cyanobacteriota; Cyanobacteria; Synechococcales_A; Cyanobiaceae; Synechococcus_C; GCA_001628295.1	MetaBAT version 0.32.4, CheckM v1.0.11, GTDB database release 88, GTDB-tk version v0.1.6	IMG_PIPELINE	2018-08-15	91.44	
3300020351_11	Marine microbial communities from Tara Oceans - JARA_B100000676 (ERR555955-ERR598089)	3300020351	MQ	Bacteria; Cyanobacteria; unclassified; Chroococcales; unclassified; Synechococcus, sp.	Bacteria; Cyanobacteriota; Cyanobacteria; Synechococcales_A; Cyanobiaceae; Synechococcus_C; GCA_001628295.1	MetaBAT version 0.32.4, CheckM v1.0.11, GTDB database release 88, GTDB-tk version v0.1.6	IMG_PIPELINE	2018-08-15	91.62	
3300020355_14	Marine microbial communities from Tara Oceans - JARA_B100000676 (ERR556077-ERR598988)	3300020355	MQ	Bacteria; Cyanobacteria; unclassified; Chroococcales; unclassified; Synechococcus, sp.	Bacteria; Cyanobacteriota; Cyanobacteria; Synechococcales_A; Cyanobiaceae; Synechococcus_C; GCA_001628295.1	MetaBAT version 0.32.4, CheckM v1.0.11, GTDB database release 88, GTDB-tk version v0.1.6	IMG_PIPELINE	2018-08-15	88.09	

First Previous 1 Next Last Export Select All Clear All Select - page Deselect - page Show 10

Showing 1 to 3 of 3 entries

Figure 7: Deep sea Tara Oceans metagenome bin search result

To remove any query condition, simply click the “-” **Remove** button at the right.

Clicking the **Evaluate Query** button to show the constructed query, count of bins satisfying each query condition, and count of bins satisfying the constructed query. To view the actual result, simply click the **Search** button. The result table will also include additional fields used in the query condition.

Search History

After the above searches, the Search History section will have all queries recorded in reverse order, with the most recent search on top. Users will be able to save and/or re-use any queries.

Search History

- Cart data will be lost when out of the session or closing the browser.

Select	Time	Query	MER-FS Metagenome
<input type="checkbox"/>	2020/05/04 14:40:18	(Bin Taxonomy -- any [Cyano]) AND (Study Dataset Names -- Genome Name / Sample Name [Tara Oceans]) AND (Bin Statistics Metadata -- Completeness % (Range: 50 to 100) [> 80]) AND (Geographic Metadata -- Depth In Meters (Range: -10 to 10970) [> 100])	<input type="button" value="Reconstruct Query"/> <input type="button" value="Search"/>
<input type="checkbox"/>	2020/05/04 14:30:48	(undefined [Tara Ocean])	<input type="button" value="Reconstruct Query"/> <input type="button" value="Search"/>
<input type="checkbox"/>	2020/05/04 14:24:52	(NCBI Phylum [Cyanobacteria])	<input type="button" value="Reconstruct Query"/> <input type="button" value="Search"/>

Figure 8: Metagenome Bin Search History

Save to Workspace

Search history will be lost after users close the browser, and users can save the data to **Workspace**.

To save any queries to workspace, simply select the queries and click the **Save Selected to Workspace** button. To view the saved query, go to **Bin Search History** submenu under the **Workspace** menu item.

Reconstruct Query

The **Reconstruct Query** button next to each query allows users to view and to revise a previously constructed query.

Rerun Query

The **Search** button next to the Reconstruct Query button allows users to rerun a previously constructed query.

Additional information on IMG's Metagenome Bins can be found in the [Metagenome Bins - Tips & Tricks](#) from the IMG UI Help Menu.