

## New Genome Search

(7/17/2018)

IMG has developed a new **Genome Search** feature (Figure 1) that is more powerful yet easy to use. For users who wish to use the old Genome Search feature, it is now under **Original Genome Search of Find Genomes**.

If you are looking for IMG's **original** Genome Search [click here](#).

### Genome Search

**Quick Search** | **Advanced Search Builder**

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

- The default choice is "All Fields" with a single keyword or substring.
- Please use "Search Parameters" to refine the search.
  - Search Parameters: *Search by Name* - for single keyword search.
  - Search Parameters: *Search by ID* - for single or multiple comma separated values.

► Search Parameters

#### Search History

- Cart data will be lost when logging out or closing the browser, but can be saved into your [workspace](#).

Select	Time	Query	MER-FS Metagenome

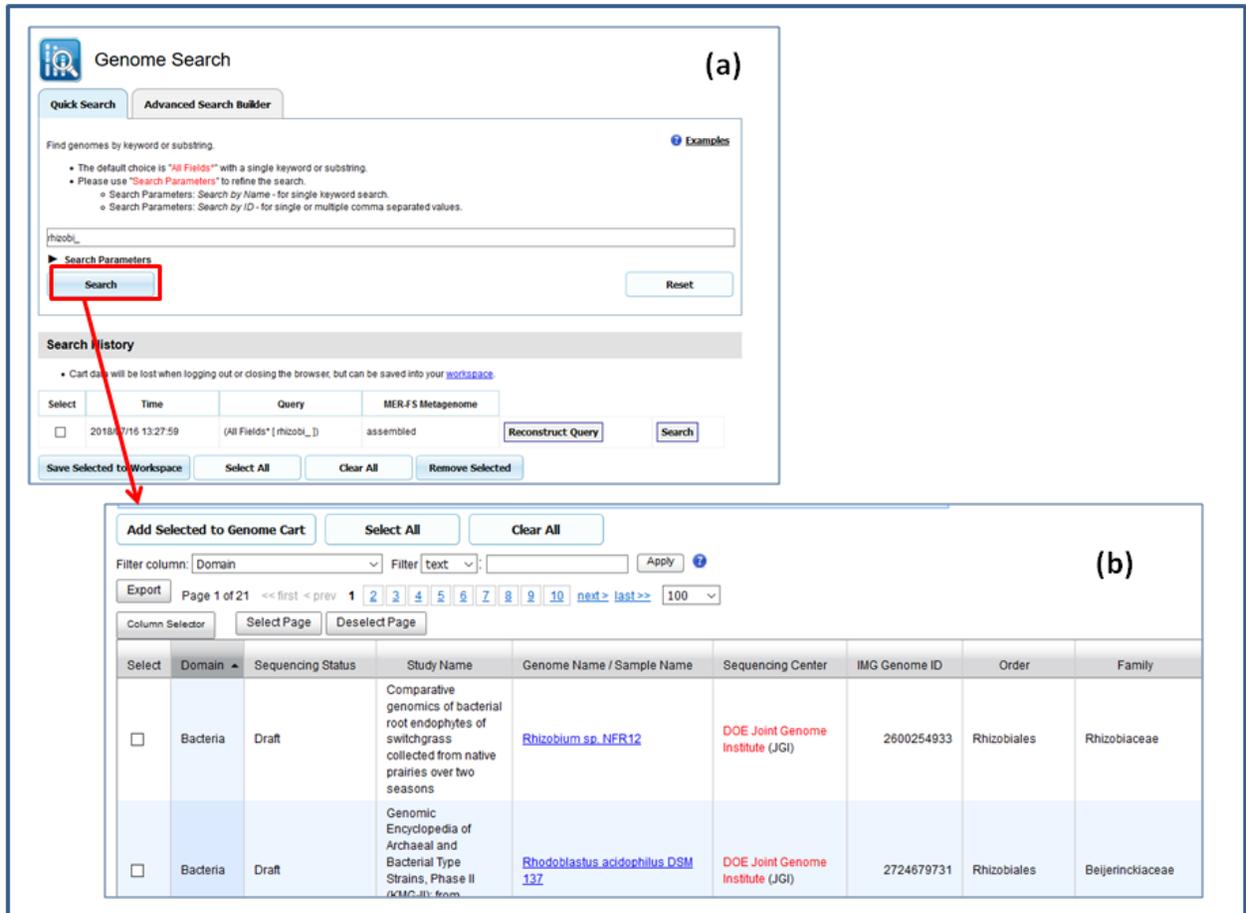
**Figure 1.** New Genome Search main page

The new genome search feature allows IMG users to perform quick search or to search IMG genomes using a more advanced query builder.

### Quick Search

A user can simply types "rhizobi\_" in the search field (where "\_" is the wildcard character) and clicks the **Search** button as shown in Figure 2(a). The default option is to search all fields. The result shows thousands of genomes satisfying the search condition. For example, the first genome in the list ***Rhizobium sp. NFR12*** has *Genome Name, Order, Family*, etc. containing keyword "rhizobi\_" and all these fields are included in the search result as shown in Figure 2(b). The user can select a subset of the result to be added to the Genome Cart.

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



**Figure 2.** Quick search using keyword "rhizobi\_"

If a user only wishes to search names or IDs in a certain field, then he/she can click the triangle next to **Search Parameters**.

The Search by Name option allows the user to select:

- All fields
- Genome Name / Sample Name
- Study Name
- Domain
- Phylum
- Sequencing Status
- Sequencing Center
- Funding Agency
- IMG Release

The Search by ID option allows the user to select:

- IMG Internal IDs

- IMG Genome ID (IMG Taxon ID)
- IMG Submission ID
- JGI Project ID / ITS PID
- GOLD Study ID
- GOLD Sequencing Project ID
- GOLD Analysis Project ID
- Scaffold ID
- External IDs
  - NCBI Taxon ID
  - RefSeq Project ID
  - GenBank Project ID
  - Scaffold External Accession

For example, search by GOLD sequencing ID "Gp0104829" will show that there is only one genome in IMG associated with this GOLD SP ID (Figure 3).

(a)

Genome Search

Quick Search | Advanced Search Builder

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

- The default choice is "All Fields" with a single keyword or substring.
- Please use "Search Parameters" to refine the search.
  - Search Parameters: Search by Name - for single keyword search.
  - Search Parameters: Search by ID - for single or multiple comma separated values.

Gp0104829

Search Parameters

Search by Name: --- Search by ID (list): GOLD Sequencing Project ID

Search Reset

(b)

Add Selected to Genome Cart | Select All | Clear All

Filter column: Domain Filter: text Apply

Export Page 1 of 1 << first < prev 1 next > last >> All

Column Selector | Select Page | Deselect Page

Select	Domain	Sequencing Status	Study Name	Genome Name / Sample Name	Sequencing Center	IMG Genome ID	GOLD Sequencing Project ID
<input type="checkbox"/>	B	D	Comparative genomics of bacterial root endophytes of switchgrass collected from native prairies over two seasons	<a href="#">Rhizobium sp. NFR12</a>	DOE Joint Genome Institute (JGI)	2600254933	Gp0104829

Figure 3. Search by GOLD Sequencing Project ID

## Advanced Search Builder

**Advanced Search Builder** allows IMG users to form a more advanced query such as: find all soil metagenome samples at depth up to 10 cm in Wisconsin or Michigan not from agricultural soils.

To construct this query (see Figure 4(a)):

1. first select "Add new builder line" in **Advanced Search Builder**, and select *Environmental Classification:Ecosystem Type* to be "soil"
2. then click on "Add new builder line" and select *Geographic Metadata:Depth* and enter "<= 0.1" (which means up to 0.1 m)
3. add another builder line and select *Geographic Metadata:Geographic Location* and enter "Wisconsin, Michigan" (which means Wisconsin or Michigan)
4. add the last builder line and select *Study Dataset Names:Genome Name / Sample Name*, change "AND" to "AND NOT" and enter " agricultural"

The screenshot shows the Advanced Search Builder interface. On the left, a query builder panel (a) contains several rows of criteria:

- Row 1: Environmental Classification: Ecosystem Type (Soil)
- Row 2: Geographic Metadata: Depth (Range: -10 to 10970 meters) (<= 0.1)
- Row 3: Geographic Metadata: Geographic Location (Wisconsin, Michigan)
- Row 4: Study Dataset Names: Genome Name / Sample Name (agricultural)

The logical operators between rows are AND, AND, and AND NOT. A red arrow points from the 'Add new builder line' button to the first row. Below the builder panel are 'Search' and 'Evaluate Query' buttons. A red arrow points from the 'Evaluate Query' button to the results section (b).

Section (b) displays the 'Constructed Query' and its results:

**Constructed Query:** (Environmental Classification -- Ecosystem Type [ Soil ]) AND (Geographic Metadata -- Geographic Location [ Wisconsin, Michigan ]) AND NOT (Study Dataset Name [ agricultural ])

- (Environmental Classification -- Ecosystem Type [ Soil ]): 8269 count(s).
- (Geographic Metadata -- Depth (Range: -10 to 10970 meters) [ <= 0.1 ]): 5917 count(s).
- (Geographic Metadata -- Geographic Location [ Wisconsin, Michigan ]): 2879 count(s).
  - Wisconsin: 1746 count(s).
  - Michigan: 1133 count(s).
- (Study Dataset Names -- Genome Name / Sample Name [ agricultural ]): 547 count(s).

**Final Combination:** 53 count(s).

Section (c) shows a table of results:

Select	Domain	Sequencing Status	Study Name	Genome Name / Sample Name
<input type="checkbox"/>	Metagenome	Permanent Draft	Soil microbial communities from Arlington Agricultural Research Station in Wisconsin and Kellogg Biological Station in Michigan, replicating the bioenergy cropping systems trials (BCSTs)	<a href="#">Soil microbial communities from Kellogg Biological Station, Michigan, USA - Nitrogen cycling UWRJ-G10K1-12</a>
<input type="checkbox"/>	Metagenome	Permanent Draft	Soil microbial communities from Arlington Agricultural Research Station in Wisconsin and Kellogg Biological Station in Michigan, replicating the bioenergy	<a href="#">Soil microbial communities from Kellogg Biological Station, Michigan, USA - Nitrogen cycling UWRJ-G06K3-12 (SPAdes)</a>

**Figure 4.** Advance Search Builder example

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

Clicking the **Evaluate Query** button near the end of the page will show the constructed query, count of genomes satisfying each query condition, and count of genomes satisfying the constructed query (Figure 4(b)). To view the actual result, simply select the **Search** button. The result table (Figure 4(c)) will also include additional fields (e.g., Depth) used in the query condition.

## Search History

After the above three searches, the **Search History** section will have 3 entries recorded in reverse order, with the most recent search on top (see Figure 5(a)). Users will be able to save and/or re-use any queries.

**(a)**

Select	Time	Query	MER-FS Metagenome	Buttons
<input checked="" type="checkbox"/>	2018/07/17 13:20:24	(Environmental Classification -- Ecosystem Type [ Soil ]) AND (Geographic Metadata -- Depth (Range: -10 to 10970 meters) [ <= 0.1 ]) AND (Geographic Metadata -- Geographic Location [ Wisconsin, Michigan ]) AND NOT (Study Dataset Names -- Genome Name / Sample Name [ agricultural ])	assembled	<b>Reconstruct Query</b> Search
<input type="checkbox"/>	2018/07/17 13:17:35	(GOLD Sequencing Project ID (list) [ Gp0104829 ])	assembled	Reconstruct Query Search
<input type="checkbox"/>	2018/07/17 13:16:45	(All Fields* [ rhizobi_ ])	assembled	Reconstruct Query Search

**(b)**

Genome Search History List

Group sharing is not displayed. Please go to [Preferences](#) to change the sharing display options.

Search History Import & Export

Filter column: File Name Filter: text Apply

Export Page 1 of 1 << first < prev 1 next > last >> All

Column Selector Select Page Deselect Page

Select	File Name	Time	Query	MER-FS Metagenome	File Size
<input type="checkbox"/>	20180717_132024	2018/07/17 13:20:24	(Environmental Classification -- Ecosystem Type [ Soil ]) AND (Geographic Metadata -- Depth (Range: -10 to 10970 meters) [ <= 0.1 ]) AND (Geographic Metadata -- Geographic Location [ Wisconsin, Michigan ]) AND NOT (Study Dataset Names -- Genome Name / Sample Name [ agricultural ])	assembled	612 B

Export Page 1 of 1 << first < prev 1 next > last >> All

**Annotations:**

- To view and to revise the soil query (points to 'Reconstruct Query' button in row 1)
- To run the rhizobi\_ query again (points to 'Search' button in row 3)
- To save the selected query to workspace (points to 'Save Selected to Workspace' button)

**Figure 5.** Search History

## Save to Workspace

Search history is like analysis carts: Any queries shown in the 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 **Genome Search History** submenu under the **Workspace** menu item (see Figure 5(b)).

## Reconstruct Query

The **Reconstruct Query** button next to each query allows users to view and to revise a previously constructed query (see Figure 5).

## Rerun Query

The **Search** button next to the Reconstruct Query button allows users to rerun a previously constructed query (see Figure 5).