

Genome Search

Introduction

[Home](#) > [Find Genomes](#) > [Genome Search](#)

Genome Search

Genome Search is another way to for selecting and browsing genomes in IMG.

Genomes can be searched via implicit attributes using the **Quick Genome Search** box at the top-right corner of every IMG page, as shown in Figure 1. A detailed search of genomes can be carried out using **Genome Search** on the second-level menu of **Find Genomes**.

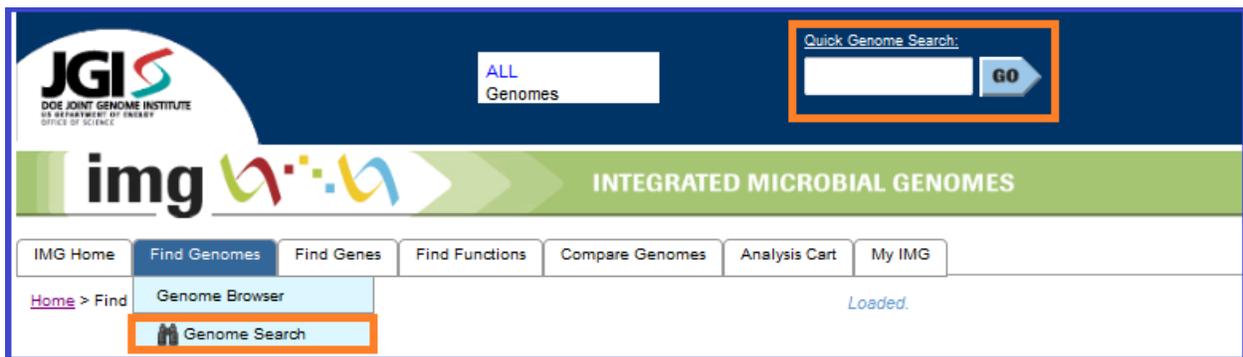


Figure 1: Genome Search: quick search, keyword field search, and metadata attribute search.

The result of **Genome Search** is a **list of genomes**, such as that shown in Figure 3 and Figure 5, which can be examined individually, or selected and saved for reducing the genome context for future analysis.

By Fields

Genome Search can use *genome fields* as search filter to find genomes, as shown in Figure 2.

Genome Search

[by Fields](#)
[by Metadata](#)

by Fields by Metadata

Genome Search by Fields

Find genomes by keyword or substring.

Keyword:

Filters:

Additional Output Columns

hint: Except for "exact" filters, all searches treat the keyword as a substring (a word or part of a word).
The search should contain some alphanumeric characters.
Use an underscore (_) as a single-character wildcard.
Use % to match zero or more characters.
All matches are case insensitive.

Examples

- "pseudomonas" as Genome Name retrieves all genomes with the substring "pseudomonas" such as "Pseudomonas syringae B728a".
- "62977" for NCBI Taxon ID retrieves "Acinetobacter sp. AD1".
- "NC_008009, NC_008010, NZ_AAKW01000001" as Scaffold External Accession will retrieve genomes that have scaffolds with those accession ID's.
- "arch" for Domain retrieves all Archaea.
- "proteo" for Phylum retrieves all proteobacteria.
- "finished" for Sequencing Status retrieves all finished genomes.
- "draft" for Sequencing Status retrieves all unfinished genomes.
- "JGI" for Sequencing Center retrieves all genomes sequenced at JGI.
- "2.0" for IMG Release retrieves all genomes with IMG release 2.0

Figure 2: Genome Search by Fields

Searchable fields are available in the pull-down filter list, including "Genome Name," "NCBI Taxon ID", "RefSeq Project ID", "GenBank Project ID", "Scaffold External Accession", "Taxon Object ID", "Domain," "Phylum", "Sequencing Status", "Sequencing Center", "Funding Agency", "IMG release" and "All field names*".

The "All field names*" is aka 'Quick Genome Search'. In addition to searches on genome name, taxonomy ranks, certain attributes and various IDs, it also covers area such as "RefSeq Project ID", "GenBank Project ID", "Is Public", "IMG Release", "IMG Product Assignment", "Gold ID in IMG Database" and "Add Date" (in format: yyyy-mm-dd).

Keywords or substrings are used to search names and descriptions. ID values should be exact. As an example, in order to perform a general search of all genomes added in a particular IMG release, such as 1.2, select the "IMG Release" filter, type "1.2" in the keyword box, and click on the "Go" button. To view all genomes sequenced at JGI, select "Sequencing Center" filter, type "JGI" in the keyword box, and click on the "Go" button.

As shown in Figure 3, the result page now displays "Domain" and "Genome Completion" as default. In addition, through the selection in search page, the result can display a variety of "Additional Output Columns":

1. Taxon Object ID; NCBI Taxon ID; RefSeq Project ID; GenBank Project ID
2. Phylum; Class; Order; Family; Genus; Species; Strain

3. Sequencing Center; Funding Agency
4. Add Date; Is Public
5. IMG Submission ID; IMG Release; IMG Product Assignment; GOLD ID in IMG Database

Genome Search (i)

by Fields
by Metadata

by Fields **by Metadata**

Genome Search by Fields

Find genomes by keyword or substring.

Keyword:

Filters:

Additional Output Columns

- Taxon Object ID
- NCBI Taxon ID
- RefSeq Project ID
- GenBank Project ID
- Phylum
- Class
- Order
- Family
- Genus
- Species
- Strain

Genome Field Search Results (ii)

112 genomes retrieved.

Save Selections Select All Clear All

hint: Selections do not take effect until you save them. You must select at least one genome. Go to [Preferences](#) to show or hide plasmids and viruses. Go to home page statistics under [IMG Genomes](#) to select individual phylogenetic domains or all genomes.

Domains(D): B=Bacteria, A=Archaea, E=Eukarya, P=Plasmids, V=Viruses.
Genome Completion(C): F=Finished, P=Permanent Draft, D=Draft.

Click on column name to sort.

Select Page Deselect Page Search column: D Search term:

Export Page 1 of 2 << first < prev 1 2 next > last >> 100

| Select | D | C | Genome Name | Taxon Object ID | NCBI Taxon ID | RefSeq Project ID |
|--------------------------|---|---|---|-----------------|---------------|-------------------|
| <input type="checkbox"/> | B | D | Pseudomonas aeruginosa 2192 | 638341158 | 350703 | 16171 |
| <input type="checkbox"/> | B | D | Pseudomonas aeruginosa C3719 | 638341159 | 350704 | 16170 |
| <input type="checkbox"/> | B | F | Pseudomonas aeruginosa LESB58 | 643348568 | 557722 | 31101 |
| <input type="checkbox"/> | B | F | Pseudomonas aeruginosa PA7 | 640753042 | 381754 | 16720 |
| <input type="checkbox"/> | B | D | Pseudomonas aeruginosa PACS2 | 638341161 | 388272 | 16851 |
| <input type="checkbox"/> | B | F | Pseudomonas aeruginosa PAO1 | 637000218 | 208964 | 331 |
| <input type="checkbox"/> | B | D | Pseudomonas aeruginosa PAb1 | 647000299 | 509633 | 49007 |

Figure 3: Example of Genome Search by Fields and results.

By Metadata

Genome Search also can use *metadata attributes* as search filter to find genomes, as shown in Figure 4.

Genome Search

[by Fields](#)
[by Metadata](#)

Select category search values.

| Category | Logical Operator (or: intra-category union) (and: inter-category intersection) |
|---|--|
| <input type="button" value="Expand All"/> <input type="button" value="Collapse All"/> | |
| <input type="checkbox"/> Biotic Relationships | or |
| | and |
| <input type="checkbox"/> Body Site | or |
| | and |
| <input type="checkbox"/> Body Subsite | or |
| | and |
| <input type="checkbox"/> Cell Arrangement | or |
| | and |
| <input type="checkbox"/> Cell Shape | or |
| | and |
| <input type="checkbox"/> Diseases | or |
| | and |
| <input type="checkbox"/> Energy Source | or |
| | and |
| <input type="checkbox"/> Ecosystem | or |
| | and |
| <input type="checkbox"/> Ecosystem Category | or |
| | and |
| <input type="checkbox"/> Ecosystem Type | or |
| | and |
| <input type="checkbox"/> Ecosystem Subtype | or |
| | and |
| <input type="checkbox"/> Specific Ecosystem | or |
| | and |
| <input type="checkbox"/> Gram Staining | or |
| | and |
| <input type="checkbox"/> Host Name | or |
| | and |
| <input type="checkbox"/> Motility | or |
| | and |
| <input type="checkbox"/> Metabolism | or |
| | and |
| <input type="checkbox"/> Oxygen Requirement | or |
| | and |
| <input type="checkbox"/> Phenotype | or |
| | and |
| <input type="checkbox"/> Relevance | or |
| | and |
| <input type="checkbox"/> Salinity | or |
| | and |
| <input type="checkbox"/> Sporulation | or |
| | and |
| <input type="checkbox"/> Temperature Range | or |

Include category data in the result display (slow)

Max number of returned search result:

Figure 4: Genome Search by Metadata.

Genomes also can be searched using specific metadata values. Each metadata category is displayed as one tree. User can click-open the tree, or use provided buttons to expand or collapse all trees.

Select one or more values for each metadata category. Metadata categories without a selected value will be ignored in the search. A logical "OR" is used when searching multiple values

selected for a single metadata category (intra-category). A logical "AND" is used when searching for values across multiple metadata categories (inter-category). Once the selection is completed, click on the "Go" button to see the search results, as illustrated in Figure 3(ii), which shows the result of searching for genomes with "Blood" or "Bone" in "Body Site".

The screenshot is divided into two main panels: (i) Genome Search and (ii) Genome Metadata Search Results.

Panel (i) Genome Search: This panel shows the search configuration. It has tabs for "by Fields" and "by Metadata". Under "by Metadata", there are "Expand All" and "Collapse All" buttons. A section titled "Select category search values." contains a tree view of categories. The "Body Site" category is expanded, showing a list of sub-categories: Airways, Bladder, Blood, Bone, Brain, Ear, Eye, Gastrointestinal tract, Heart, Liver, Lymph nodes, Oral, Plant-root, Plant-stem, and Root. The "Blood" and "Bone" sub-categories are checked. To the right, a "Logical Operator" section shows "and" selected, with a note: "(or: intra-category union) (and: inter-category intersection)".

Panel (ii) Genome Metadata Search Results: This panel shows the search condition "Body Site: Blood, Bone" and indicates that 56 genomes were retrieved. It includes buttons for "Save Selections", "Select All", and "Clear All". A "hint" box provides instructions: "Selections do not take effect until you save them. You must select at least one genome. Go to Preferences to show or hide plasmids and viruses. Go to home page statistics under IMG Genomes to select individual phylogenetic domains or all genomes." Below this, domain and completion codes are listed: "Domains(D): B=Bacteria, A=Archaea, E=Eukarya, P=Plasmids, V=Viruses. Genome Completion(C): F=Finished, P=Permanent Draft, D=Draft." There are also buttons for "Select Page", "Deselect Page", and a search field. A table of results is shown below, with columns for "Select", "D", "C", and "Genome Name".

| Select | D | C | Genome Name |
|--------------------------|---|---|--|
| <input type="checkbox"/> | B | F | Acinetobacter baumannii AB0057 |
| <input type="checkbox"/> | B | F | Acinetobacter baumannii AB307-0294 |
| <input type="checkbox"/> | B | F | Anaerococcus prevotii DSM 20548 |
| <input type="checkbox"/> | B | F | Anaplasma marcinale Florida |
| <input type="checkbox"/> | B | F | Arcanobacterium haemolyticum DSM 20595 |
| <input type="checkbox"/> | B | F | Bacillus cereus_03BR102 |
| <input type="checkbox"/> | B | D | Bacillus cereus AH1273 |

Figure 5: Example of Genome Search by Metadata and results.