Scaffold Search User Guide

(9/28/2020)

Introduction

IMG has developed a new Scaffold Search feature under Find Genomes (Figure 1). Similar to other new search functions in IMG, the new Scaffold Search feature allows IMG users to perform quick search using IDs, or to search IMG scaffolds using a more advanced query builder.

Quick Search

Quick search allows users to search scaffolds using IMG scaffold IDs, external accessions or scaffold names. Note that metagenome scaffolds can only be searched using IMG IDs.

Example 1:
Suppose a user is interested in finding a scaffold with ID “SRS018774_WUGC_scaffold_235” from HMP study, but does not know the corresponding metagenome OID. The user can simply use the Scaffold ID search option in Quick Search:

Enter the scaffold ID in the blank, and click the **Search** button, the following result will appear:

The user can select the scaffold and add it to Scaffold Cart for further analysis.

**Advanced Search**

Advanced Search Builder allows IMG users to form a more advanced search using a combination of any of the following fields:

- Scaffold Taxonomy
  - Domain
  - Phylum
  - Class
- Order
- Family
- Genus
- Species

- Function IDs
  - COD ID (list)
  - KOG ID (list)
  - Pfam ID (list)
  - TIGRfam ID (list)
  - SMART ID (list)
  - SUPERFam ID (list)
  - CATH FunFam ID (list)
  - KO ID (list)
  - Enzyme ID (list)
  - IMG Term ID (list)

- Function Names
  - COD Name
  - KOG Name
  - Pfam Name
  - TIGRfam Name
  - SMART Name
  - SUPERFam Name
  - CATH FunFam Name
  - KO Name
  - Enzyme Name
  - IMG Name

- Scaffold Statistics
  - Scaffold Topology
  - Scaffold Nucleotide Length
  - Scaffold GC Percentage
  - Scaffold Read Depth
  - Scaffold Gene Count
  - Scaffold CDS Gene Count
  - Scaffold tRNA Gene Count
  - Scaffold 5S rRNA Gene Count
  - Scaffold 16S rRNA Gene Count
  - Scaffold 18S rRNA Gene Count
  - Scaffold 23S rRNA Gene Count
  - Scaffold 28S rRNA Gene Count
  - Scaffold other rRNA Gene Count

Example 2:
Suppose a user has two marine microbial communities from Delaware Coast in the Genome Cart (IMG taxon IDs: 3300000101, 3300000116):

To find all scaffolds with both 16s and 23s rRNAs in these two metagenomes, the user can use the Advanced Search Builder with the following two conditions:

- **Scaffold Statistics -- Scaffold 16S rRNA Gene Count * (Range) >= 1**
- **Scaffold Statistics -- Scaffold 23S rRNA Gene Count * (Range) >= 1**

Then add both metagenomes in the cart to search.
Click the Search button, and the following result with 91 scaffolds will be displayed:
Again, scaffolds in the result table can be selected to add to the Scaffold Cart for further analysis. The user can also add any additional attributes in the Table Configuration and click the Redisplay button to add the additional fields to the table display.

**Example 3:**

Suppose the user is interested in finding all scaffolds in the same two metagenomes satisfying the following 3 conditions:

- Having genes annotated with either pfam00543 (P-II - Nitrogen regulatory protein P-II) or pfam00909 (Ammonium_transp - Ammonium Transporter Family)
- Scaffold lineage assigned to the *Proteobacteria* phylum
- With length greater than or equal to 10,000 bp

The user can use the Advanced Search Builder with the following 3 builder lines:

- Function IDs -- Pfam ID (list) *: pfam00543, pfam00909
- Scaffold Taxonomy -- Phylum: Proteobacteria
- Scaffold Statistics -- Scaffold Nucleotide Length * (Range): >= 10000

And add the 2 metagenomes to search.
The result shows that 68 scaffolds satisfying the condition:

- **Table:**

<table>
<thead>
<tr>
<th>Scaffold ID</th>
<th>Scaffold Name</th>
<th>External Accession</th>
<th>Genes Count</th>
<th>Sequence Length (bp)</th>
<th>GC Content</th>
<th>Read Depth</th>
<th>Lineage Domain</th>
<th>Lineage Phylum</th>
</tr>
</thead>
<tbody>
<tr>
<td>DemM100201_c100000002</td>
<td>Marine microbial communities from Delaware Coast</td>
<td></td>
<td>127</td>
<td>146253</td>
<td>0.30</td>
<td>207</td>
<td>Bacteria</td>
<td>Proteobacteria</td>
</tr>
<tr>
<td>DemM100201_c100000009</td>
<td>Marine microbial communities from Delaware</td>
<td></td>
<td>79</td>
<td>80061</td>
<td>0.30</td>
<td>61</td>
<td>Bacteria</td>
<td>Proteobacteria</td>
</tr>
<tr>
<td>DemM100201_c100000074</td>
<td>Marine microbial communities from Delaware Coast</td>
<td></td>
<td>99</td>
<td>84561</td>
<td>0.30</td>
<td>231</td>
<td>Bacteria</td>
<td>Proteobacteria</td>
</tr>
</tbody>
</table>
Example 4:

Now suppose the user wishes to find *Proteobacteria* scaffolds with length greater than or equal to 10,000 bp, and having annotated with both pfam00543 and pfam00909. Since IMG does not allow users to add the same condition filter more than once, the user will have to perform the searches twice: once with pfam00543, and the other time with pfam00909. After each search, the user will have to save the result to a workspace scaffold set, and then get the intersection of both scaffold sets at the end to get the desired result.

However, the user is smart enough to figure out that he can use Pfam ID (for pfam00543) and Pfam Name (for pfam00909) to achieve the result he wants in a single query. That is, the user uses the query builder with the following 4 builder lines:

- **Function IDs -- Pfam ID (list)**: pfam00543
- **Scaffold Taxonomy -- Phylum**: Proteobacteria
- **Scaffold Statistics -- Scaffold Nucleotide Length (Range)**: >= 10000
- **Function Names -- Pfam Name**: Ammonium Transporter

After adding the same two metagenome to search and click the **Search** button, the result shows that 19 scaffolds satisfying the search condition:
Search History

All the scaffold search history will be recorded in the Search History section in reverse chronicle order:
Save to Workspace

Search history is like analysis carts: Any queries shown in the history could 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 the Scaffold 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.