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**.

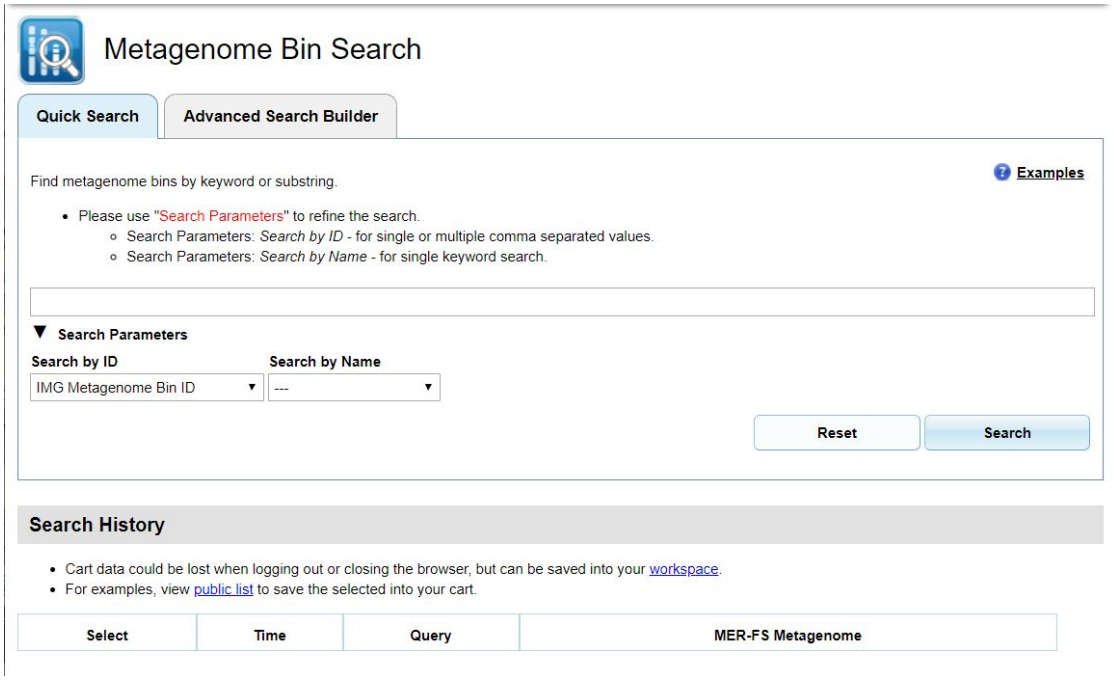

Figure 1. Metagenome Bin Search main page

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.

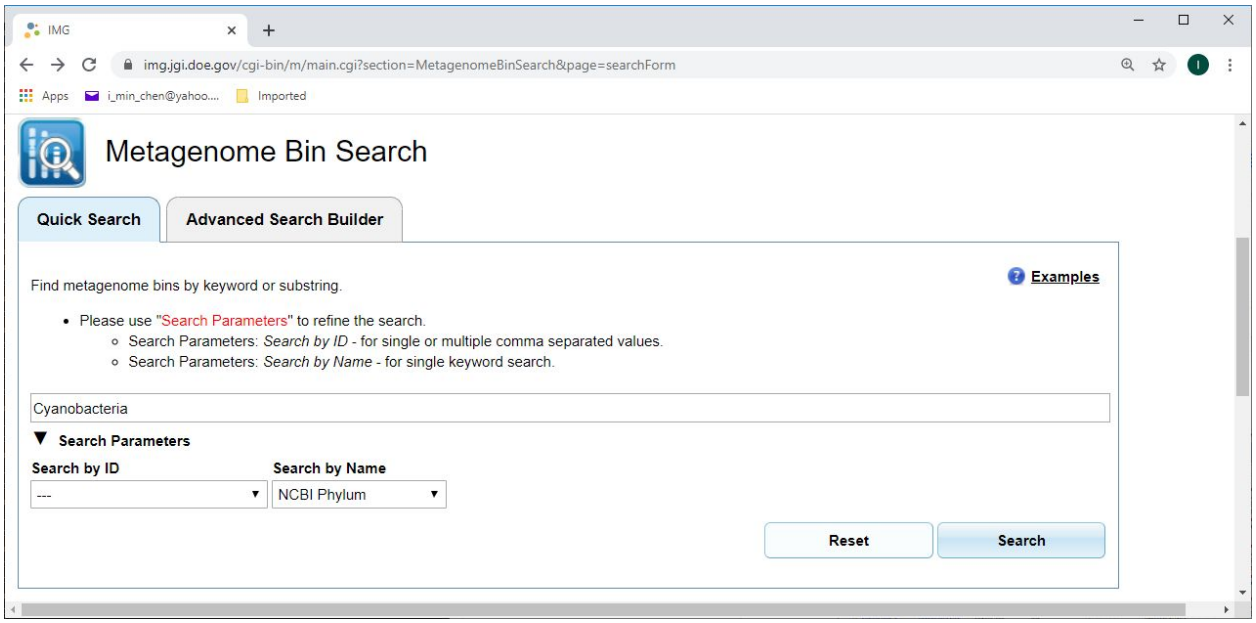

Figure 2: Quick Search: Cyanobacteria metagenome bin search
After clicking the **Search** button, the result shows 719 bins from 538 metagenomes:

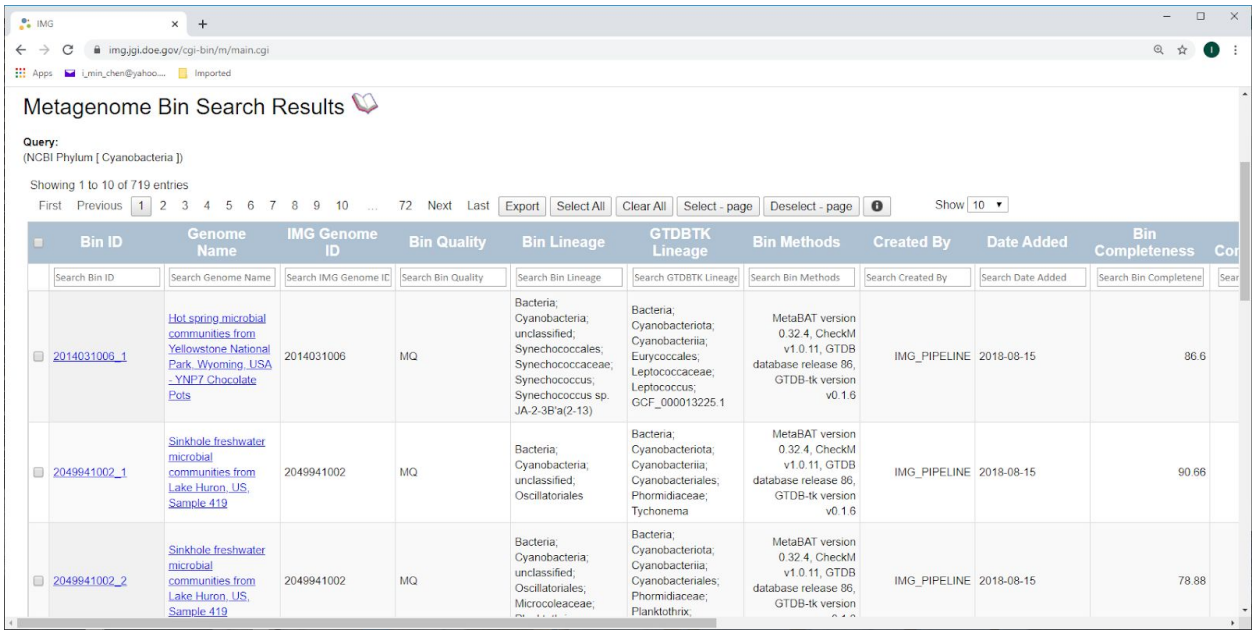

Figure 3: Cyanobacteria bin result page

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.

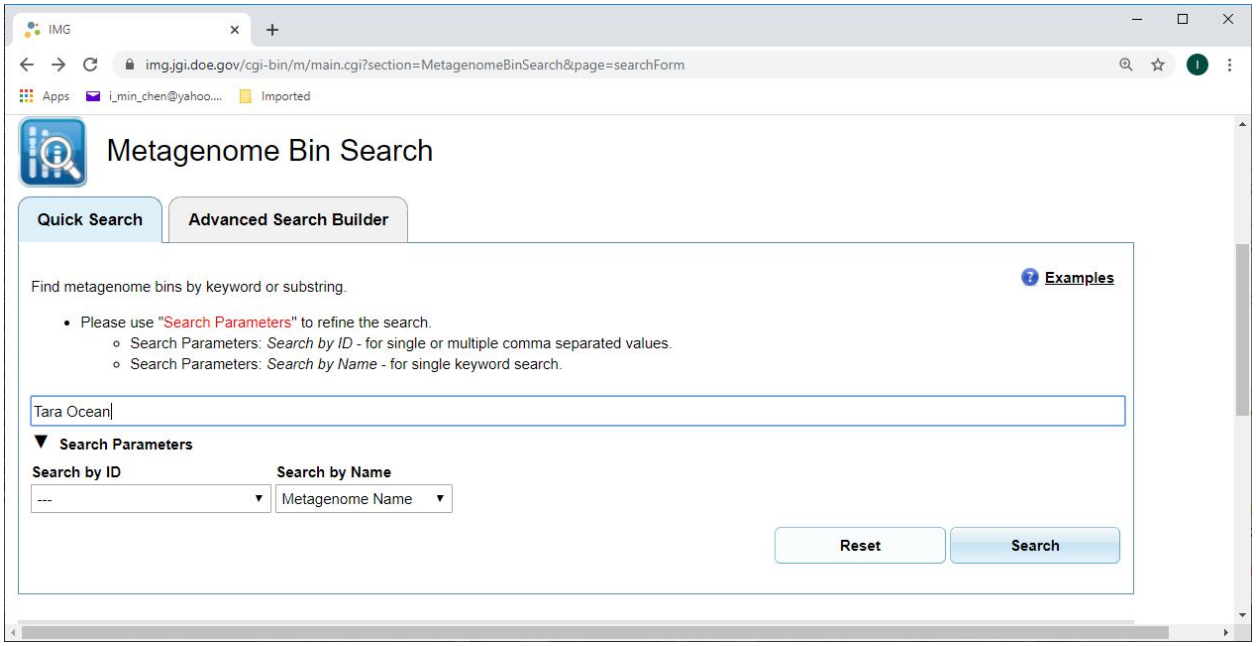

Figure 4: Quick Search: bins obtained from Tara Oceans metagenomes
After clicking the **Search** button, result will show 1,956 bins obtained from 240 Tara Oceans metagenomes.

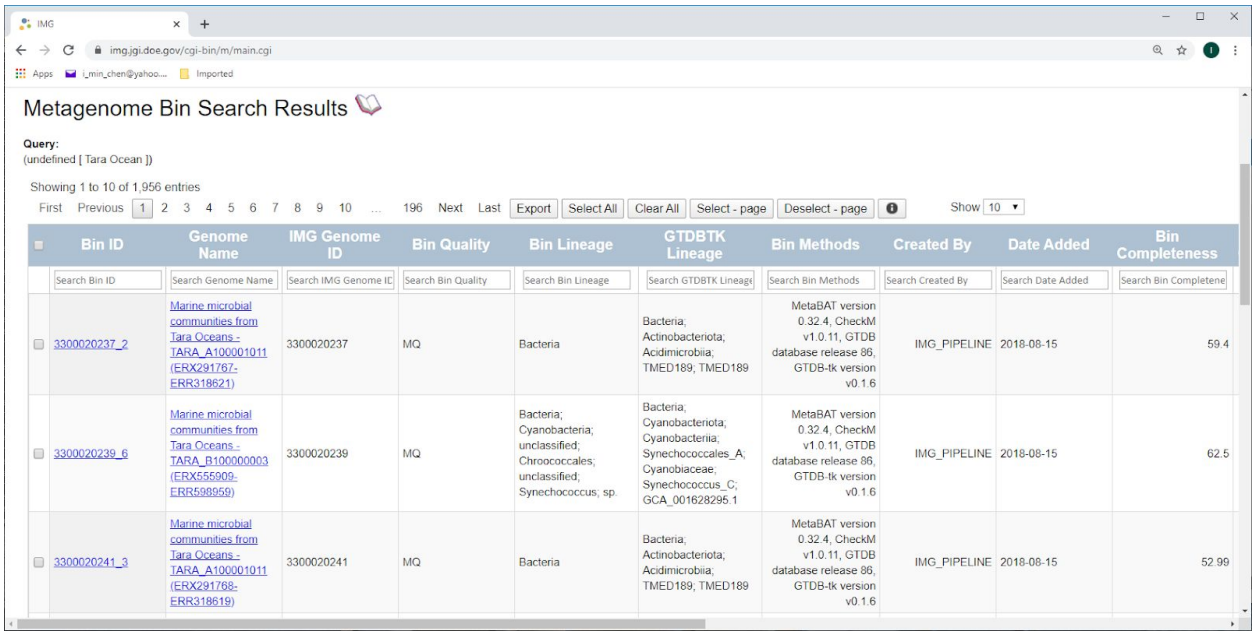

Figure 5: Tara Ocean metagenome bin search result

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

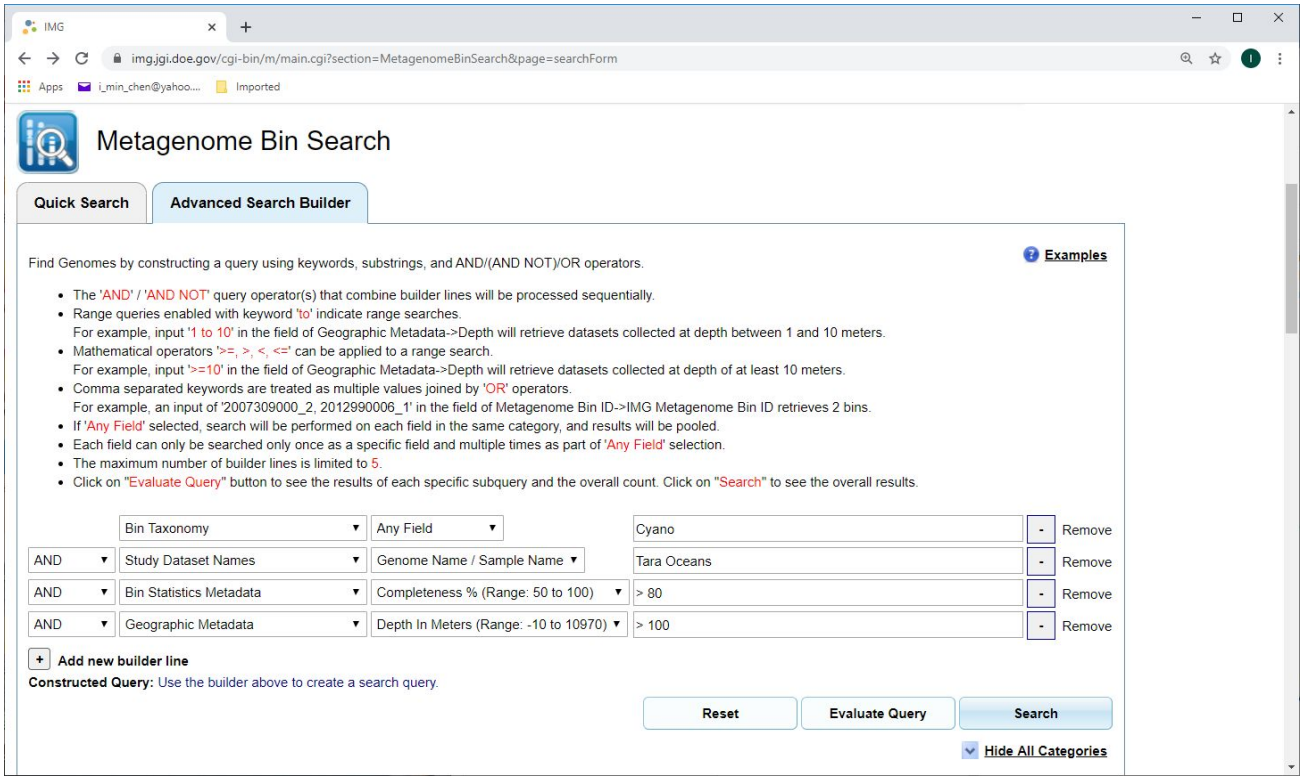

Figure 6: Advanced Search: Deep sea cyanobacteria bins from Tara Oceans metagenome with at least 80% completeness

The above search will return 3 bins as shown below:

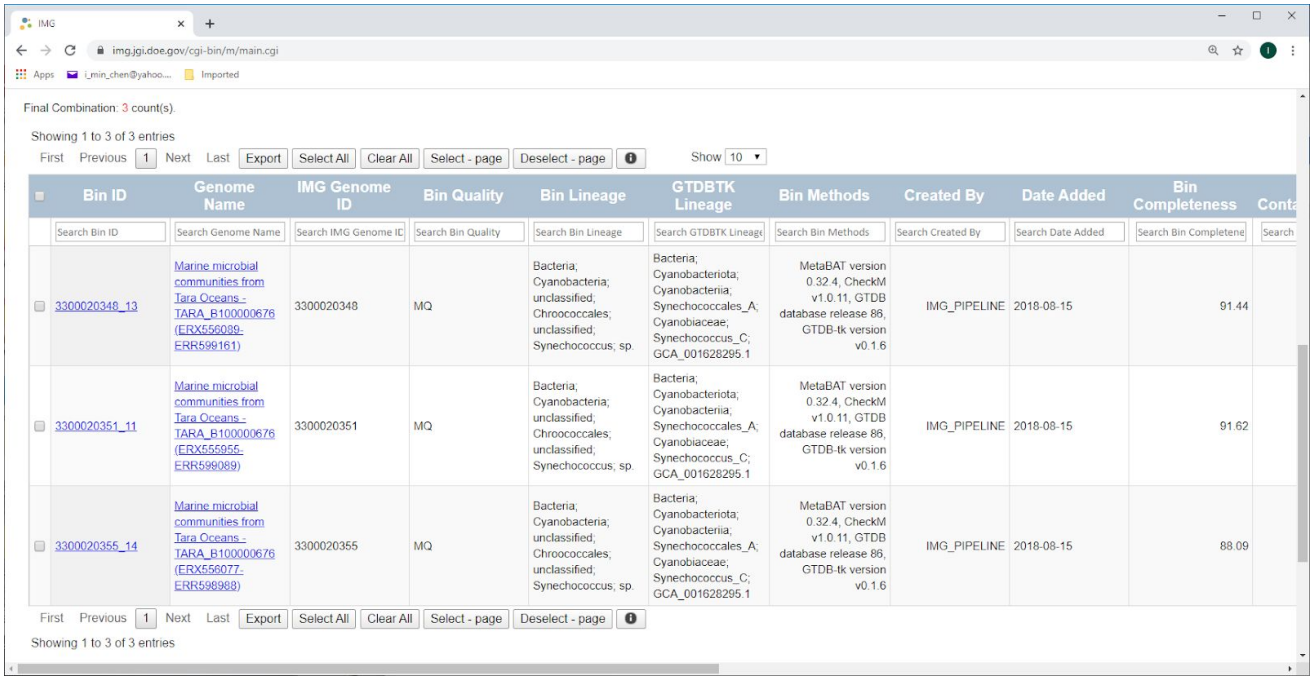

Figure 7: Deep sea Tara Oceans metagenome bin search result

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.

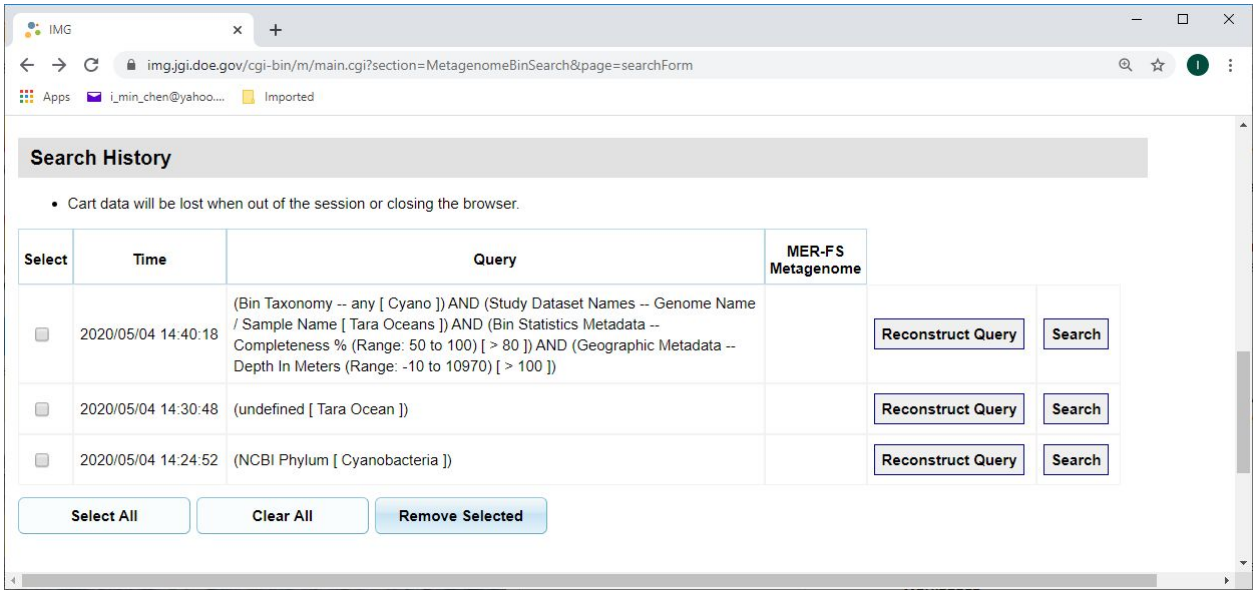

Figure 8: Metagenome Bin Search History

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.