


MG-RAST Tutorial

Getting started

- Web Page: <http://metagenomics.anl.gov/>
 - Register to run you own analysis
 - Publicly available datasets can be explored without registering.




Web Interface

- Home Page
 - The home page has search box and options to upload, download and analyze data
 - Three icons on top right corner are available on all pages
 - The search box allows users to search publicly available datasets from the repository with extensive options to filter.
- Search
 - Simply click on the search button to browse the repository.
 - Will list all datasets; 20 records at time
 - The repository has 50,154 projects.
 - Provide specific search criteria to limit the results to you interest area.
 - The "field" dropdown lists the different features you can search against
 - Search all projects that has samples taken from "Nebraska"
 - Filter more to include samples collected from "nine-mile parairie"
 - The table shows the filtered results based on the supplied criteria.
 - The settings icon () on the top left corner of the table allows user to select the fields to be displayed on the table.
 - Choose "material" in the settings icon options to check the material sequenced for the samples.
 - Click the link (phosphorous removing (ebpr) sludge community) on the study to go to the details of the study.
 - Data descriptions including PI, funding source etc. are uploaded as metadata during analysis process.
 - Map on the right shows the geographical location where the samples were collected from.
 - The table lists all the samples in the study.
 - The links under "MG-RAST ID" and "name" takes to the detailed analysis results for each sample.
 - The "download column" lets user download the fasta sequences and metadata that were submitted during analysis.
 - Click a link under "MG-RAST ID"
 - Try to answer following questions:
 - How many sequences failed QC?
 - What are the percentages of known proteins/ rRNA ?
 - What was the sequencing method used?
 - Check other analysis statistics.

- Check the Nucleotide Histogram.
- Check number of hits in different database.
- Check for functional category hits distribution (COG, NOG, KEGG etc.)
- Check for Taxonomic distributions (Domain, Phylum, Order etc.)
- See options to download the data sets. What format are they downloaded in?
- Check bar plots for abundance.
- Understand Rarefaction Curve
- Check out the metadata table at the bottom of the page.

- **Analyze metagenomic data**




- From the top bar select () to go to analysis page.
- The page gives basic description on how it works.
- Start by selecting the database that you want to run analysis against. The "available databases" drop down lists the database that are available. Select the ones that are relevant and add it to the "selected databases"
- Choose the metagenome you are interested in by provided the filter criteria. For example, project id = mgp43
- Add the metagenomes to the right side and click on the green tick mark ().
 - Selecting multiple metagenome allows you to compare them.
 - Try analyzing single metagenome, two metagenomes and multiple metagenomes
- It will start the download of the profiles for each selected metagenome for all the selected databases.
- You can also upload previously generated profiles by selecting "load from disk" under () icon on top right corner of the page.
- Try various options under "Analysis" section to compare the two metagenomes.

- **Uploading your own data**

- Register and Login



- Click on the upload button ()
- Upload page has a table with two columns and allows you to upload data to your write only inbox.
- User need to prepare a metadata file before uploading fastaq sequences.
 - Metadata file
 - User can use MetaZen tool or template excel file to create metadata.

- Upload metadata file
- Upload fastq files
 - Upload both pairs of files.
 - Merge paired end files
 - Select the read 1 file and click "join paired ends" button on the right panel to initiate the join. This will allow you to give custom name for you joined file.
 - The merged file name should match with the metagenome_name field or the file_name field in the metadata file
 - Click "next" to proceed
- Select metadata file
- Give a name for the project if new (same as in metadata file). If you already have existing project, you will get an option to add the new sample to the project
- Select the merged sequence file
- Choose pipeline options
- Submit
 - Based on if/when you want to make the data public, you will get priority for processing your data.