Introduction to Bioinformatics
10th practice
Microbiome analysis

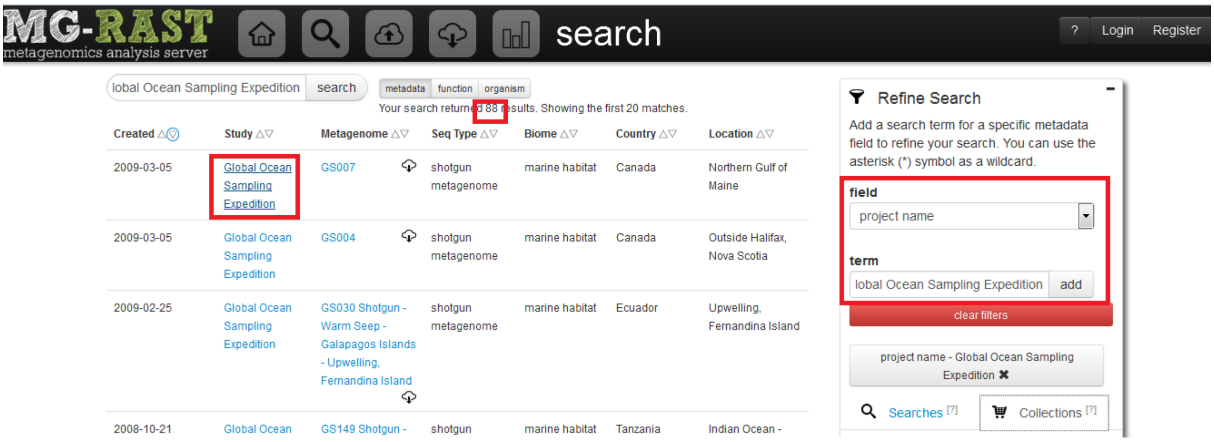
During the previous practices we processed NGS data. At first we assembled reads to contigs and scaffolds, and then we annotated some sequences based on different databases. In this practice session we will analyse microbial NSG data.

I. MG-RAST metagenome analysis server:

1. Visit the MG-RAST metagenome analysis server (it works best with Firefox):
http://metagenomics.anl.gov/
How many metagenomes are currently on the server?

You can upload your NGS sequencing results here and the server process will process it. You can also make your data available for others, and one can analyse, compare and download public datasets. The server has FTP site and command line API too.

2. Browse metagenomes:
Click on the search tab.
Search for the project named “Global Ocean Sampling Expedition”! How many metagenomes are in this project? Who was the principal investigator of the project?
Click on the name of the study for the details!



lobal Ocean Sampling Expedition

Your search returned 88 results. Showing the first 20 matches.

Created	Study	Metagenome	Seq Type	Biome	Country	Location
2009-03-05	Global Ocean Sampling Expedition	GS007	shotgun metagenome	marine habitat	Canada	Northern Gulf of Maine
2009-03-05	Global Ocean Sampling Expedition	GS004	shotgun metagenome	marine habitat	Canada	Outside Halifax, Nova Scotia
2009-02-25	Global Ocean Sampling Expedition	GS030 Shotgun - Warm Seep - Galapagos Islands - Upwelling, Fernandina Island	shotgun metagenome	marine habitat	Ecuador	Upwelling, Fernandina Island
2008-10-21	Global Ocean	GS149 Shotgun -	shotgun	marine habitat	Tanzania	Indian Ocean -

Refine Search

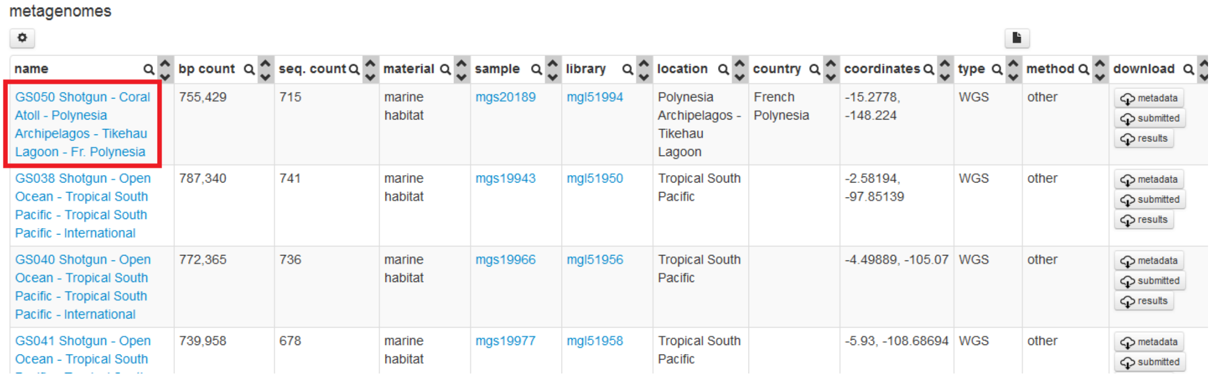
Add a search term for a specific metadata field to refine your search. You can use the asterisk (*) symbol as a wildcard.

field: project name

term: lobal Ocean Sampling Expedition

project name - Global Ocean Sampling Expedition ✖

Choose a metagenome from this project, click on it and study the properties of the sample.


metagenomes

name	bp count	seq. count	material	sample	library	location	country	coordinates	type	method	download
GS050 Shotgun - Coral Atoll - Polynesia Archipelagos - Tikehau Lagoon - Fr. Polynesia	755,429	715	marine habitat	mgs20189	mgl51994	Polynesia Archipelagos - Tikehau Lagoon	French Polynesia	-15.2778, -148.224	WGS	other	metadata submitted results
GS038 Shotgun - Open Ocean - Tropical South Pacific - Tropical South Pacific - International	787,340	741	marine habitat	mgs19943	mgl51950	Tropical South Pacific		-2.58194, -97.85139	WGS	other	metadata submitted results
GS040 Shotgun - Open Ocean - Tropical South Pacific - Tropical South Pacific - International	772,365	736	marine habitat	mgs19966	mgl51956	Tropical South Pacific		-4.49889, -105.07	WGS	other	metadata submitted results
GS041 Shotgun - Open Ocean - Tropical South	739,958	678	marine habitat	mgs19977	mgl51958	Tropical South Pacific		-5.93, -108.68694	WGS	other	metadata submitted

Find the metagenomes with the following metagenome IDd: mgm4562213.3, mgm4457774.3, mgm4441488.3!

3. Answer the following questions about your samples!

What is the source of each dataset? Where are they taken from? (source organism, habitat)

What was the sequencing technique?

Which dataset is the biggest, which is the smallest (based on the number of sequences)?

How long are the reads on average?

Is there a significant difference in the shapes of the nucleotide diagrams?

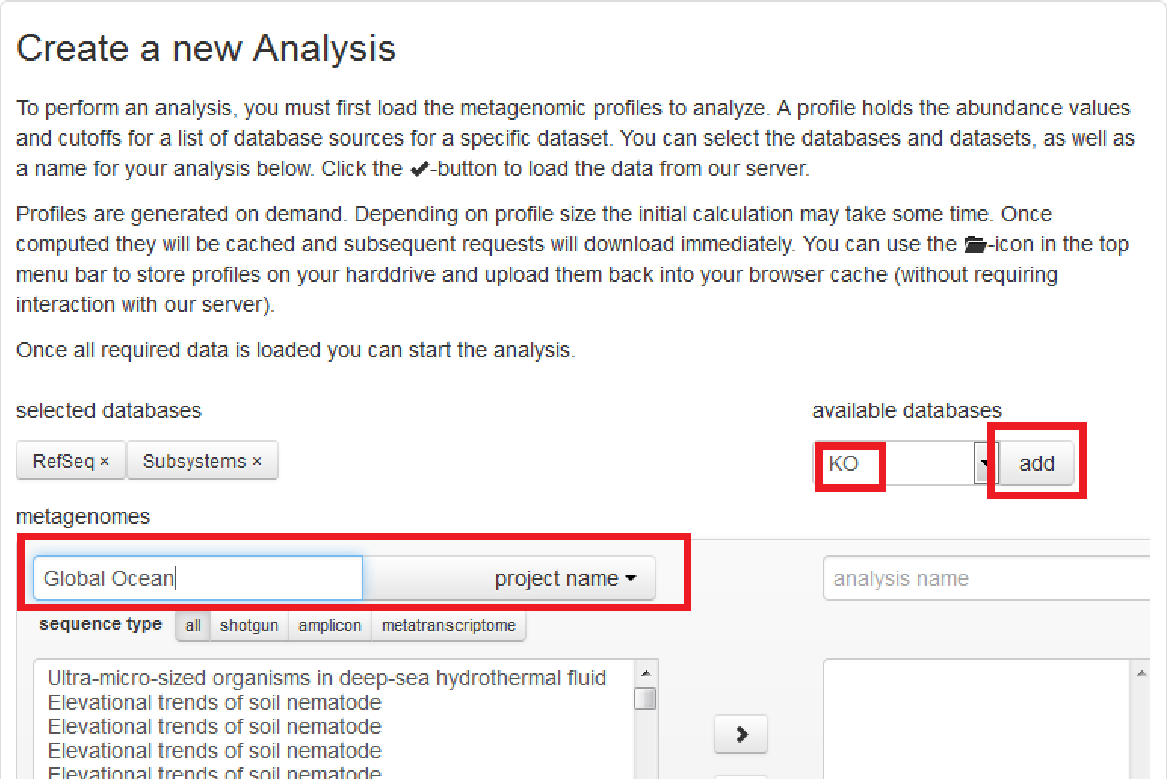
Which sample is the better characterised based on the “sequence breakdown” piecharts?

Explain the meaning of alpha-diversity and rarefaction curve of the samples! Which environment could be more diverse? Which was sequenced more fully?

Which are the most abundant domains, phyla in each datasets in the datasets?

4. Comparative analysis:
Click on the analysis tab!

Choose 3 samples from the “Global Ocean Sampling Expedition” project (You can choose metagenomes from different places by sorting based on location). Choose the following databeses: RefSeq, Subsystems and KO. Load them.



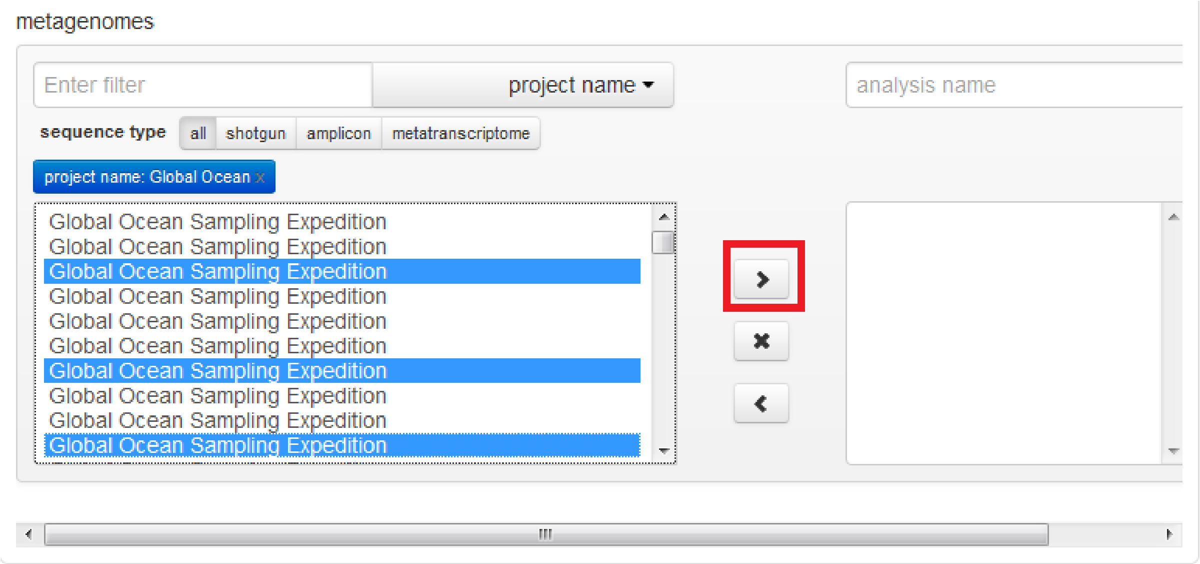
Create a new Analysis

To perform an analysis, you must first load the metagenomic profiles to analyze. A profile holds the abundance values and cutoffs for a list of database sources for a specific dataset. You can select the databases and datasets, as well as a name for your analysis below. Click the ✔-button to load the data from our server.

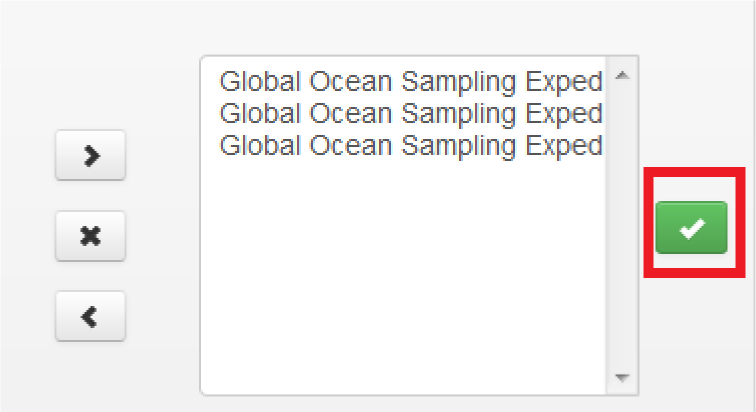
Profiles are generated on demand. Depending on profile size the initial calculation may take some time. Once computed they will be cached and subsequent requests will download immediately. You can use the 📂-icon in the top menu bar to store profiles on your harddrive and upload them back into your browser cache (without requiring interaction with our server).

Once all required data is loaded you can start the analysis.

Ultra-micro-sized organisms in deep-sea hydrothermal fluid
Elevational trends of soil nematode
Elevational trends of soil nematode
Elevational trends of soil nematode
Elevational trends of soil nematode



Global Ocean Sampling Expedition
Global Ocean Sampling Expedition
Global Ocean Sampling Expedition
Global Ocean Sampling Expedition
Global Ocean Sampling Expedition
Global Ocean Sampling Expedition
Global Ocean Sampling Expedition
Global Ocean Sampling Expedition
Global Ocean Sampling Expedition
Global Ocean Sampling Expedition



Global Ocean Sampling Exped
Global Ocean Sampling Exped
Global Ocean Sampling Exped

Be careful, for amplicon samples (e.g.: 16S rRNA) only RNA databases (e.g.: Greengenes) can be used. It means that you cannot get functional information from these samples just taxonomy!

After loading the required data, you can explore the differences between the metagenomes.
Try the simplest table view, donut charts, and stacked bars. Try Krona viewer and KEGG mapper in the “Plugins” section.