

MG-RAST Data

If using Chordomics for metagenomics or metatranscriptomics samples one must first upload the data to MG-RAST.

Chordomics allows the user to enter just their MG-RAST dataset ID (beginning with “mgm”) and it will download and process the data automatically. Alternatively users can manually download the appropriate files and upload them manually to Chordomics. This walkthrough will explain how to get them.

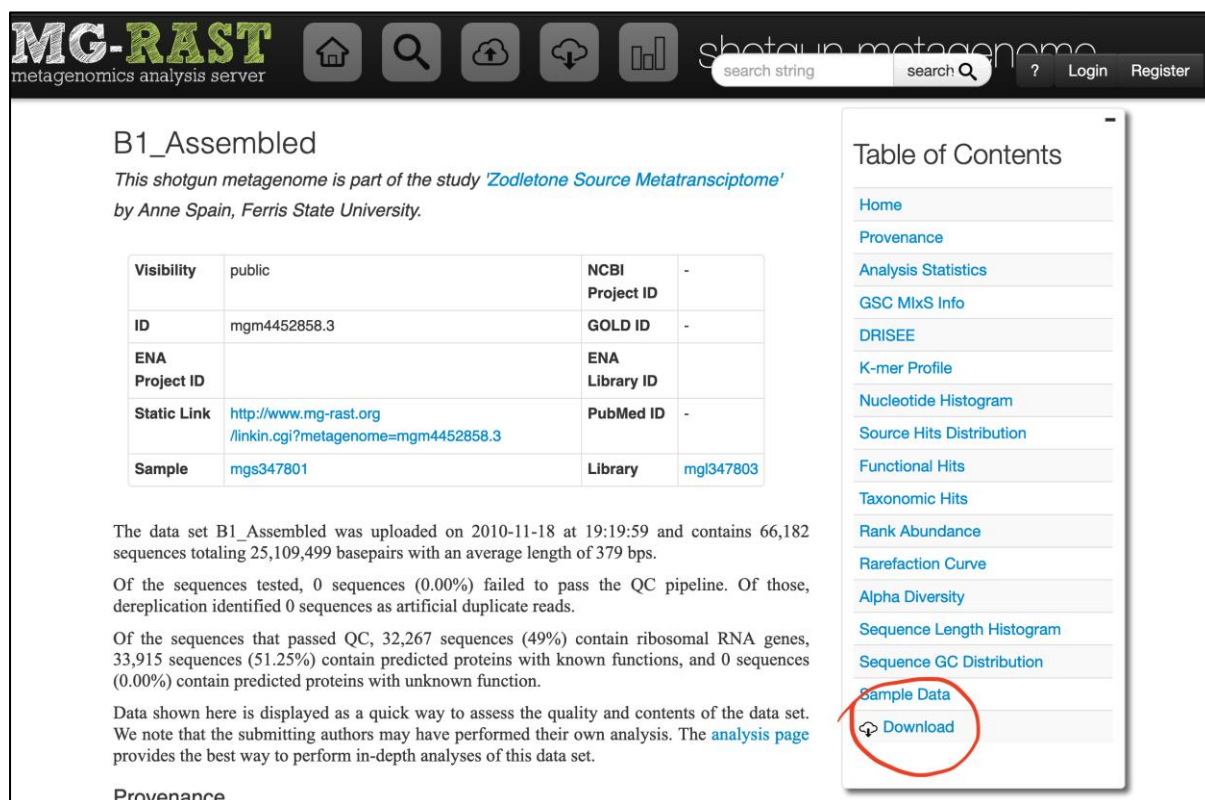
Chordomics requires 3 files:

- COG ontology file
- Organism file e.g. RefSeq
- Original fasta input

Walkthrough

Go to dataset page of the sample you want to download.

Click on the “Download” link on the panel on the right as shown.



MG-RAST
metagenomics analysis server

search string search ? Login Register

B1_Assembled

This shotgun metagenome is part of the study 'Zodletone Source Metatranscriptome' by Anne Spain, Ferris State University.

Visibility	public	NCBI Project ID	-
ID	mgm4452858.3	GOLD ID	-
ENA Project ID		ENA Library ID	
Static Link	http://www.mg-rast.org/linkin.cgi?metagenome=mgm4452858.3	PubMed ID	-
Sample	mgs347801	Library	mg1347803

The data set B1_Assembled was uploaded on 2010-11-18 at 19:19:59 and contains 66,182 sequences totaling 25,109,499 basepairs with an average length of 379 bps.

Of the sequences tested, 0 sequences (0.00%) failed to pass the QC pipeline. Of those, dereplication identified 0 sequences as artificial duplicate reads.

Of the sequences that passed QC, 32,267 sequences (49%) contain ribosomal RNA genes, 33,915 sequences (51.25%) contain predicted proteins with known functions, and 0 sequences (0.00%) contain predicted proteins with unknown function.

Data shown here is displayed as a quick way to assess the quality and contents of the data set. We note that the submitting authors may have performed their own analysis. The [analysis page](#) provides the best way to perform in-depth analyses of this data set.

Provenance

Table of Contents

- Home
- Provenance
- Analysis Statistics
- GSC MixS Info
- DRISEE
- K-mer Profile
- Nucleotide Histogram
- Source Hits Distribution
- Functional Hits
- Taxonomic Hits
- Rank Abundance
- Rarefaction Curve
- Alpha Diversity
- Sequence Length Histogram
- Sequence GC Distribution
- Sample Data
- Download**

Under the heading “Annotation Downloads” select and download the taxonomy file from Annotation Type – organism and your Data Source of choice. We recommend RefSeq.

Similarly select and download the functional file from Annotation Type – ontology and Data Source COG.

MG-RAST
metagenomics analysis server

publication never
length filtering yes
length filter deviation multiplicator 2.0
ambiguous base filtering yes
maximum ambiguous basepairs 5

The computational environment and workflow can be downloaded below:

- [environment](#)
- [workflow document](#)

Annotation Downloads

Annotated reads are available through the [MG-RAST API](#). They are built dynamically based on the chosen annotation type and source.^[7]

Annotation Type: Data Source:

Processing Steps

The input fasta file can be downloaded just below this from Step 0.

0. Upload

This is the original submitted sequence file. This is a sequence file in either fasta or fastq format. It may have been edited to change all end-of-line characters into UNIX format.

mgm4452858.3.050.upload.fna	<input type="button" value="download"/>
filesize	28.8 MB
MD5	f2a6a11d31d23e5acaaebd044d0aa68