

User Guide

The NMDC Read-based Analysis

Introduction

The read-based analysis workflow taxonomically classifies shotgun metagenome sequence data using three profiling tools: Centrifuge, Kraken, and GOTCHA2. Additional information can be found here: <https://github.com/microbiomedata/ReadbasedAnalysis>. In IMG, these data can be accessed two ways:

1. Access via the Metagenome List

One way to access the read-based analysis output is directly from the “Metagenome List” at the bottom of the NMDC landing page (highlighted in the red box).

The read-based analysis workflow can accommodate raw sequence data, however to ensure high quality profiling results, raw data is first processed through the quality control module to remove artifacts, adapters, contaminants, phiX genome, and host genomes sequences (<https://github.com/microbiomedata/ReadsQC>). The reference databases for each profiler tool used in this workflow are available for download from each website of the tools respectively. The databases are mainly built using bacteria, archaeal and viral nucleotide sequences from [NCBI RefSeq Release 90](#). The output of these analyses can be viewed interactively in IMG/M through Krona plots [\[5\]](#) as well as accessing the raw output files through the [Genome Portal](#). The Krona plot provides an interactive visualization of the hierarchical taxonomic content of the metagenome data at each level. The display resembles a pie chart that presents the read-count as the relative magnitude of identified taxonomies and enables users to explore these multi-layered pie charts.

Additional details about this workflow can be found here:

<https://github.com/microbiomedata/ReadbasedAnalysis>.

Metagenome List

User Guide

References:

- [1] Wood, Derrick E., Jennifer Lu, and Ben Langmead. "Improved metagenomic analysis with Kraken 2." *Genome biology* 20.1 (2019): 257. <https://doi.org/10.1186/s13059-019-1891-0> ↗
- [2] Kim, Daehwan, et al. "Centrifuge: rapid and sensitive classification of metagenomic sequences." *Genome research* 26.12 (2016): 1721-1729. <https://doi.org/10.1101/gr.210641.116> ↗
- [3] Freitas, Tracey Allen K., et al. "Accurate read-based metagenome characterization using a hierarchical suite of unique signatures." *Nucleic acids research* 43.10 (2015): e69-e69. <https://doi.org/10.1093/nar/gkv180> ↗
- [4] Heng Li, "Minimap2: pairwise alignment for nucleotide sequences", *Bioinformatics*, 34. 18 (2018): 3094-3100 <https://doi.org/10.1093/bioinformatics/bty191> ↗
- [5] Ondov, B.D., Bergman, N.H. & Phillippy, A.M. "Interactive metagenomic visualization in a Web browser". *BMC Bioinformatics* 12, 385 (2011). <https://doi.org/10.1186/1471-2105-12-385> ↗

A full list of available metagenomes with read-based analysis results is displayed, and can be filtered and sorted based on Metagenome ID, Metagenome Name, Krona plots, and JGI Data Download. Clicking on the Centrifuge, Kraken, or GOTCHA2 hyperlinks will display the interactive Krona plots for viewing. Further details on Krona plot navigation are available below. Additionally, the raw outputs are available for download through the JGI Data Portal.



 Quick Genome Search: Go Hi Emiley Eloe-Fadrosch | Logout (JGI SSO) 105610

My Analysis Carts: 0 Genomes | 0 Scaffolds | 0 Functions | 0 Genes | 2 Genome Search History | 0 Gene Search History | 0 Scaffold Search History | 0 Bin Search History | 0 BGC |

Home | IMG/MER | Find Genomes | Find Genes | Find Functions | Compare Genomes | OMICS | Workspace | My IMG | Collaborations | Help

Home > Find Genomes loaded.

NMDC Metagenome List

All the NMDC's read analysis data sets will be available for download. Any metagenome without a "JGI Data Portal" link. The data set bundle is still being prepared and it will be made available for download. Please try again later.

Showing 21 to 30 of 521 entries

First Previous 1 2 3 4 5 6 7 8 9 10 ... 53 Next Last

Metagenome Id	Metagenome Name	Plots	Download Read Analysis
<input type="checkbox"/> 3300005944	Permafrost soil microbial communities from the Arctic, to analyse light accelerated degradation of dissolved organic matter (DOM) - Organic soil replicate 2 DNA2013-048	Centrifuge Gottcha2 Kraken2	JGI Data Portal
<input type="checkbox"/> 3300005947	Permafrost soil microbial communities from the Arctic, to analyse light accelerated degradation of dissolved organic matter (DOM) - Permafrost soil replicate 2 DNA2013-190	Centrifuge Gottcha2 Kraken2	JGI Data Portal
<input type="checkbox"/> 3300005949	Permafrost soil microbial communities from the Arctic, to analyse light accelerated degradation of dissolved organic matter (DOM) - Organic soil leachate replicate DNA2013-051	Centrifuge Gottcha2 Kraken2	JGI Data Portal
<input type="checkbox"/> 3300005950	Permafrost soil microbial communities from the Arctic, to analyse light accelerated degradation of dissolved organic matter (DOM) - Organic soil replicate 2 DNA2013-047	Centrifuge Gottcha2 Kraken2	JGI Data Portal
<input type="checkbox"/> 3300005952	Permafrost soil microbial communities from the Arctic, to analyse light accelerated degradation of dissolved organic matter (DOM) - Organic soil replicate 1 DNA2013-045	Centrifuge Gottcha2 Kraken2	JGI Data Portal

2. Access via Metagenome Detail Page

If the user is interested in a particular metagenome which has the read-based analysis, the links to the Centrifuge, Kraken, and GOTTCHA2 outputs can be accessed directly from the Metagenome Details Page (highlighted in the red box), along with the link to the JGI Data Portal for download.

Microbiome Details (Assembled and Unassembled Data)

About Genome

- [Overview](#)
- [Statistics](#)
- [Genes](#)

Overview

Study Name (Proposal Name)	Permafrost soil microbial communities from the Arctic, to analyse light accelerated degradation of dissolved organic matter (DOM)
Sample Name	Permafrost soil microbial communities from the Arctic, to analyse light accelerated degradation of dissolved organic matter (DOM) - Permafrost soil replicate 2 DNA2013-191
Taxon Object ID	3300005938
IMG Submission ID	70940
GOLD ID in IMG Database	Project ID: Gp0111286
GOLD Analysis Project Id	Ga0066795
GOLD Analysis Project Type	Metagenome Analysis
Submission Type	Primary
JGI Analysis Product Name	Metagenome Minimal Draft
JGI Analysis Project Type	Metagenome Analysis
NMDC Analysis	 <input type="button" value="Centrifuge Plot"/> <input type="button" value="Gottcha2 Plot"/> <input type="button" value="Kraken2 Plot"/> <input type="button" value="Download Read Analysis"/>
Sequencing Status	Permanent Draft
Sequencing Center	DOE Joint Genome Institute (JGI)
IMG Release/Pipeline Version	
Comment	
Release Date	2016-05-27
Add Date	2015-09-23
Modified Date	2015-09-23

Download tarball

The raw outputs for all three taxonomic profiling tools are available for download through the JGI Data Portal. Clicking the “Download Read Analysis” will transfer to the JGI Genome Portal page for the selected metagenome. A folder titled “NMDC read-based analysis” will be listed and contains three tarballs, one for each of the three analyses. Users may download the tarballs directly for further analyses and review the README documentation.

The screenshot displays the JGI Data Portal interface for downloading data. At the top, there is a navigation bar with tabs: INFO, DOWNLOAD, STATUS, REPORT, HELP!, and ADMIN. Below the navigation bar, there are several informational messages and buttons:

- A message: "Open Downloads as XML" - the document tree is shown within XML file.
- A warning message: "Please be aware that restore requests sourced from tape and of more than 1 file (including Globus downloads) will be delayed until 8PM."
- A button: "Download via Globus (v.2)".
- A section titled "For PMO and Superusers Only" containing buttons: "Inspect" and "Edit display locations (ITS SP)".
- Buttons: "Download Selected Files", "ExpandAll", "CollapseAll", "Rescan", and a checkbox for "Organize By File Type".
- A note: "Please keep in mind that downloading tape files () can take a few minutes."
- A "Legend" button.

The main content area shows a tree view of the data. The "PersoIDNA2013191" folder is expanded, revealing the "NMDC read-based analysis" folder, which is highlighted with a red box. This folder contains three tarball files:

File Name	Size	Downloaded	Cache Expires
Permafrost soil replicate 2 DNA2013-191: Project: 1058849_1782_77878.centrifuge.tar.gz	309 MB	Tue Oct 27 19:07:14 PDT 2020	Fri Oct 29 14:13:37 PDT 2021
Permafrost soil replicate 2 DNA2013-191: Project: 1058849_1782_77878.gottcha2.tar.gz	34 MB	Tue Oct 27 19:07:18 PDT 2020	Fri Oct 29 14:13:38 PDT 2021
Permafrost soil replicate 2 DNA2013-191: Project: 1058849_1782_77878.kraken2.tar.gz	319 MB	Tue Oct 27 19:07:20 PDT 2020	Fri Oct 29 14:13:38 PDT 2021

Below the "NMDC read-based analysis" folder, there are other folders: "QC Filtered Raw Data", "Raw Data", and "Sequencing QC Reports".