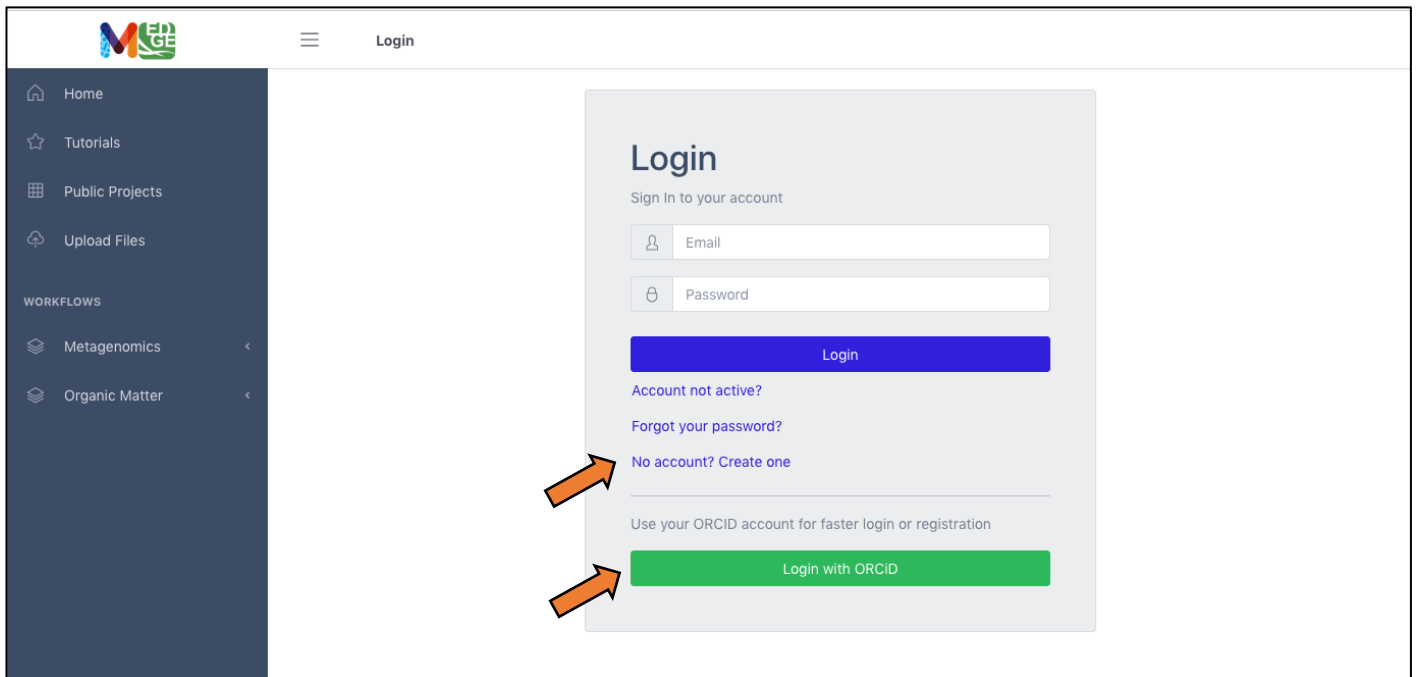
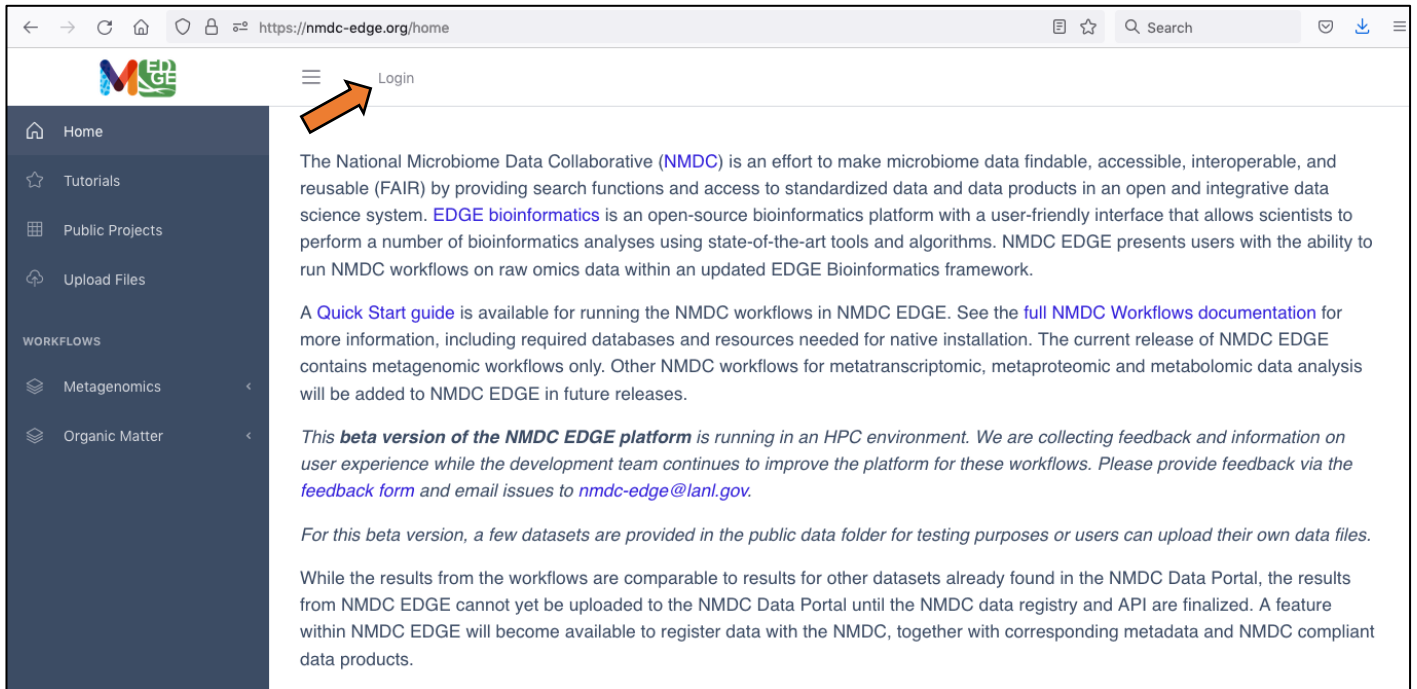


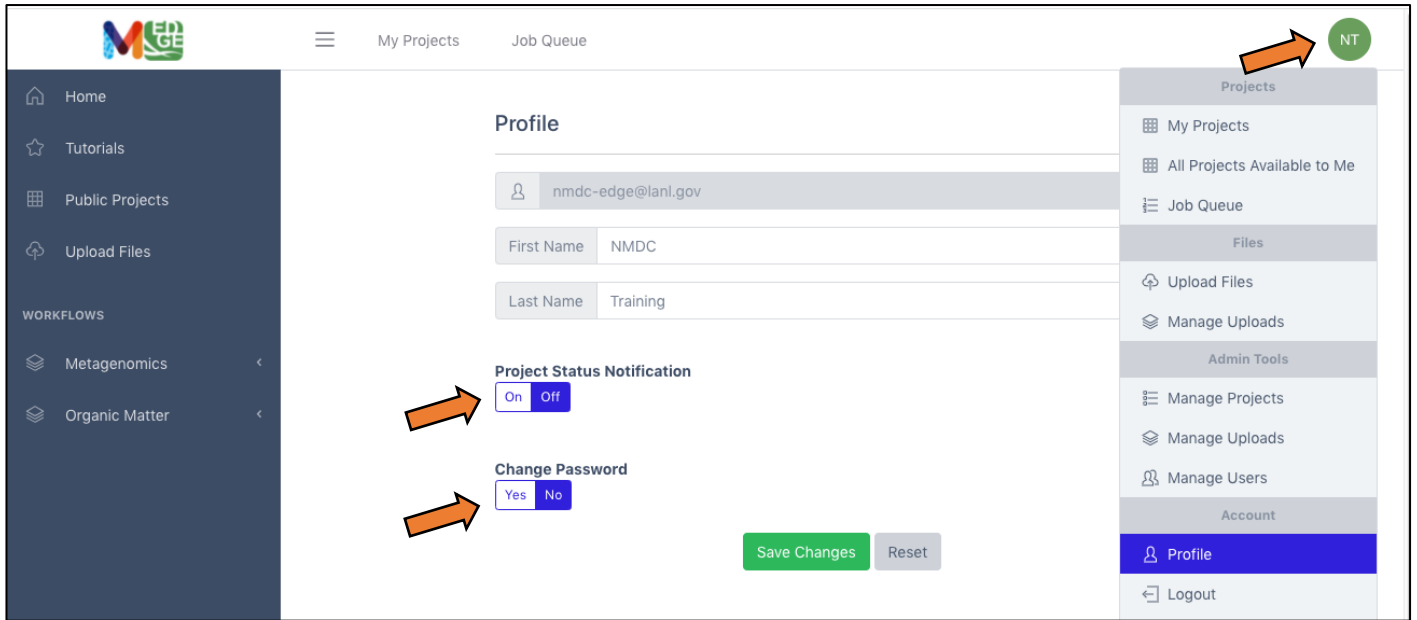
## Register for an account

Users must register for an account within the NMDC EDGE platform or login using the user's ORCID account.



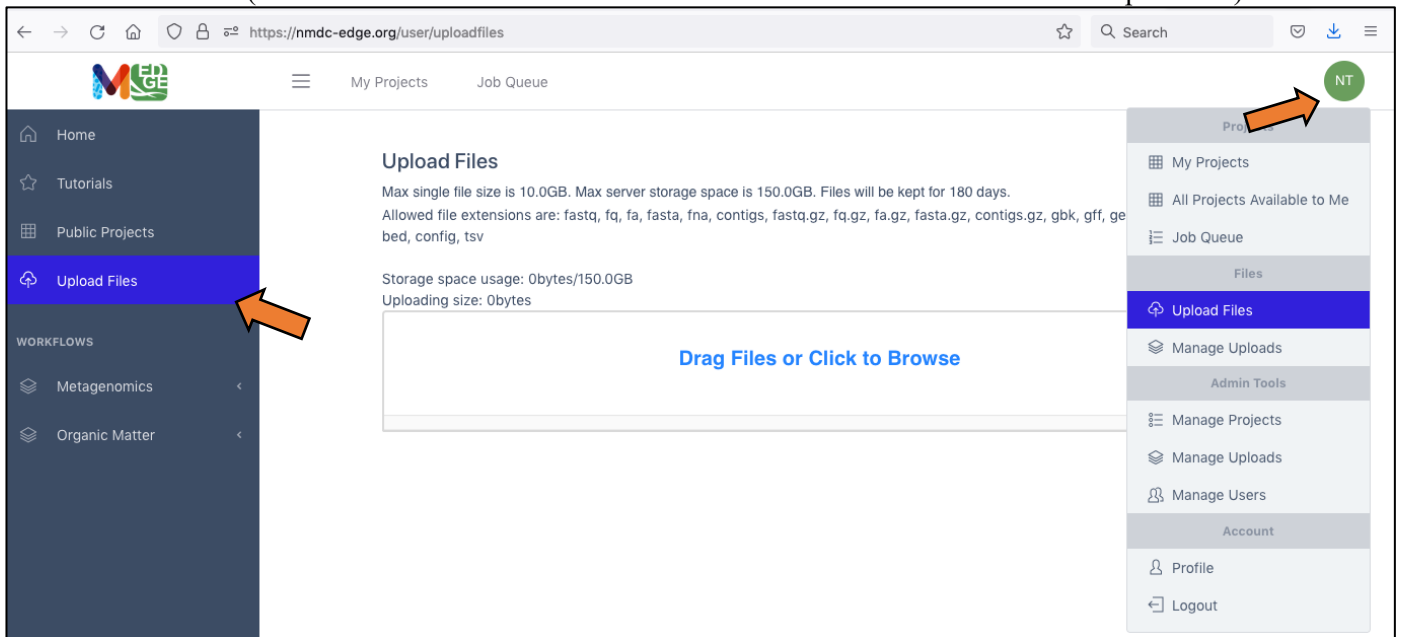
## User Profile

Once logged in, the green button with the user's initials on the right provides a drop-down menu which allows the user to manage their projects and uploads; there is also a button which allows users to edit their profile. On this profile page, there are two options: 1) the option to receive email notification of a project's status (OFF by default) and 2) the option to change the user's password (also OFF by default).



## Upload data

Two options are available for users to upload their own data to process through the workflows. The first is using the button in the left menu bar. The second is through the drop-down menu shown when clicking the green button with the user's initials on the right. Either button will open a window which allows the user to drag and drop files or browse for the user's data files. (There are also some datasets in the Public Data folder for users to test the platform.)



## Running a single workflow

To run a workflow, the user must provide:

1. A unique Project/Run Name with no spaces (underscores are fine).
2. A description is optional, but recommended.
3. The user then selects the workflow desired from the drop-down menu.
4. For metagenomic/metatranscriptomic data, the user must also select if the input data is interleaved or separate files for the paired reads.
5. Then the input file(s) from the available list of files.
6. The user should click “Submit”.

The screenshot shows the 'Run a Single Workflow' form in the Metagenomics section. The form includes the following fields and elements:

- Project/Run Name:** A required text input field (callout 1).
- Description:** An optional text input field (callout 2).
- Workflow:** A dropdown menu currently set to 'ReadsQC' (callout 3).
- Input Section:**
  - Input Raw Reads:** A section with a question 'Is interleaved?' and 'Yes'/'No' radio buttons (callout 4).
  - Input paired fastq:** A button labeled 'Add paired-end fastq'.
  - Pair-1 FASTQ #1:** A file selection input field with a menu icon (callout 5).
  - Pair-2 FASTQ #1:** A file selection input field with a menu icon (callout 5).
  - Remove:** A button to remove the selected file.
- Submit:** A blue button at the bottom right (callout 6).

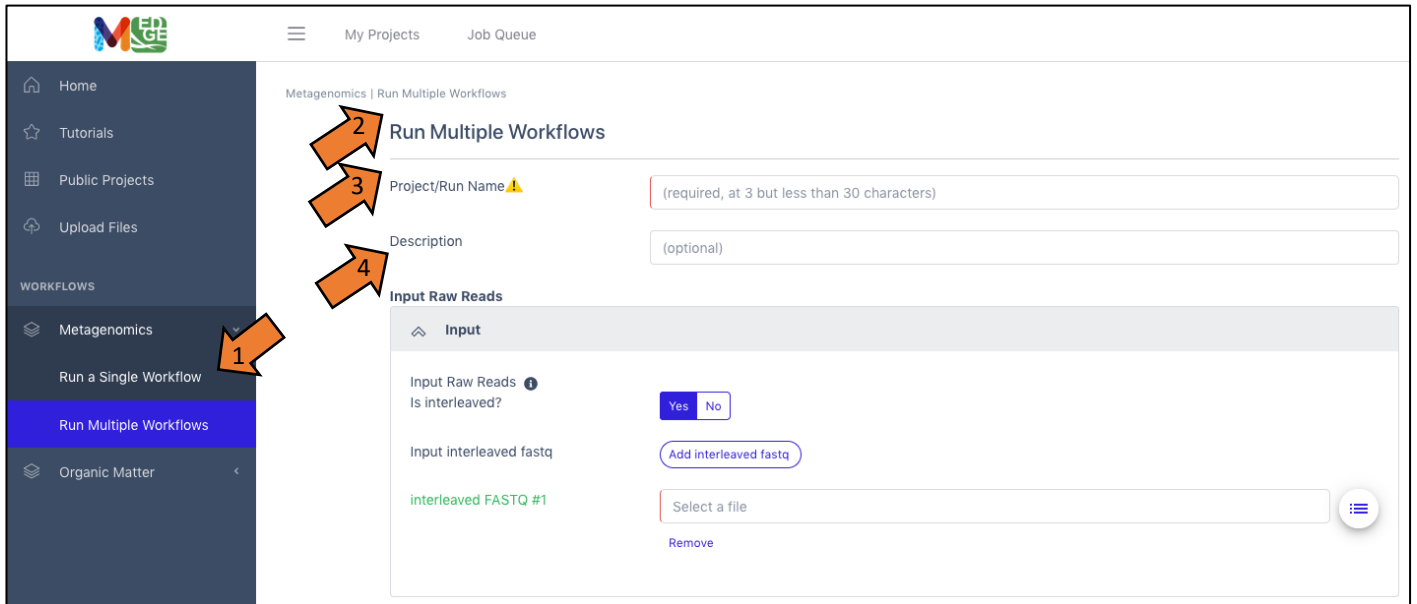
Note: Clicking on the buttons to the right of the data input blanks opens a box called “Select a file” to allow the user to find the desired files (shown in purple) from previously run projects, the public data folder, or user uploaded files.

The screenshot shows the 'Select a file' dialog box overlaid on the workflow form. The dialog box displays a list of files and folders with columns for 'File', 'Size', and 'Last Modified'.

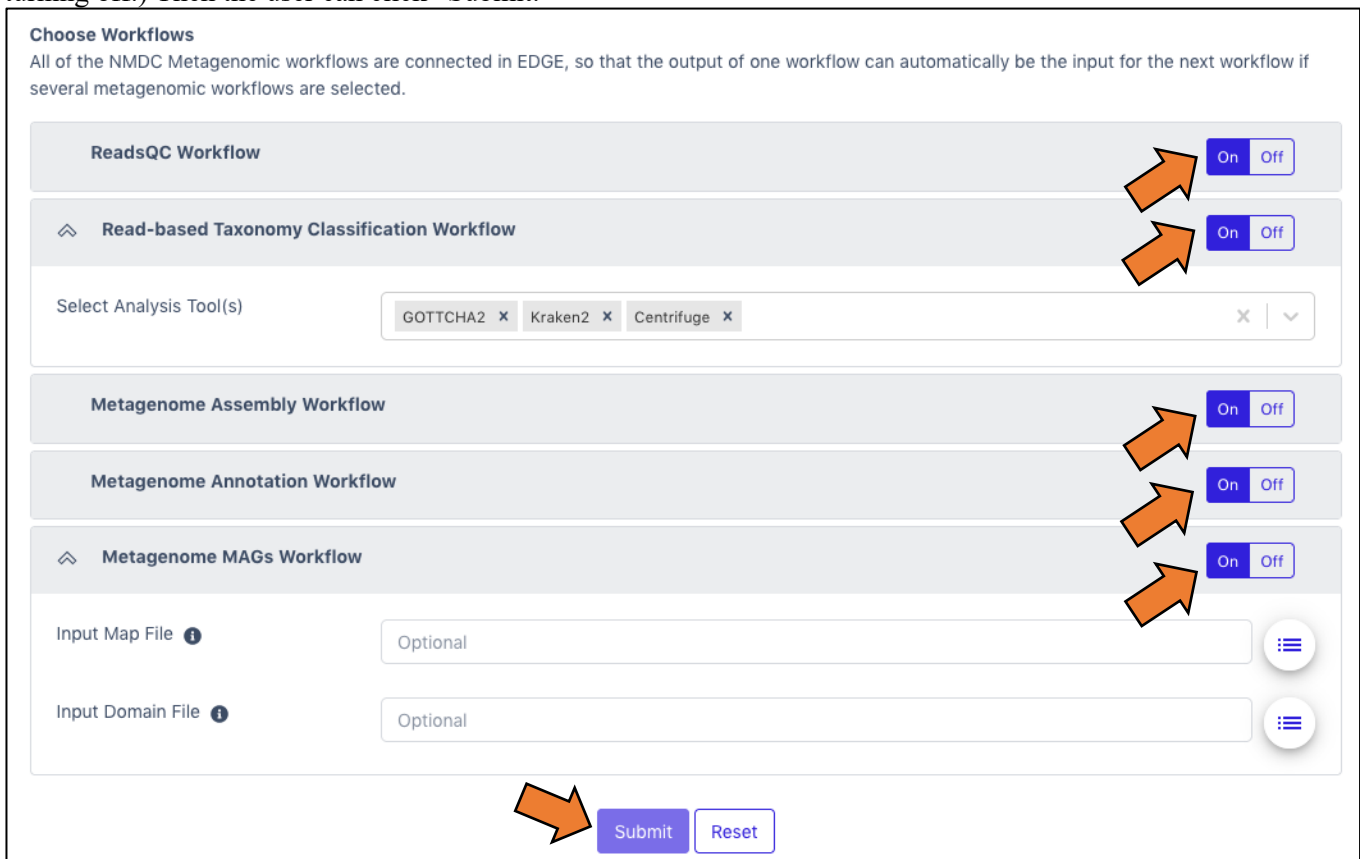
File	Size	Last Modified
projects		
nmdc-edge@lanl.gov		
Pipeline_Test (Metagenome Pipeline, 2021-10-13, 10:49:56 PM)		
output		
ReadsQC		
SRR7877884-int		
SRR7877884-int.anqpht.fq.gz	1713.59 MB	5 days ago
ReadsQC_Test (ReadsQC, 2021-10-18, 5:20:43 PM)		
ReadsQC_Test (ReadsQC, 2021-10-13, 8:32:26 PM)		
publicdata		
uploads		

## Running multiple workflows

1. Another option is to select “Run Multiple Workflows” if the user desires to run more than one of the metagenomic workflows or the entire metagenomic pipeline.
2. Enter a **unique** Project/Run Name with no spaces (underscores are fine).
3. A description is optional, but recommended.
4. The user must also select if the input data is interleaved or separate files for the paired reads.



All five of the metagenomic workflows are “ON” by default, but the user can select to turn off any workflows not desired. The pipeline uses the output of each workflow as the input for subsequent workflows. (Note: Some workflows require input data from prior workflows, so turning one workflow off may result in other workflows also automatically turning off.) Then the user can click “Submit.”



## Output

1. The link for 'My Projects' opens the list of projects for that user
2. Links (in the purple circles) are provided to share projects, make projects public, or delete projects
3. The "Status" column shows whether the job is in the queue (gray), submitted (purple), running (yellow), has failed (red) or completed (green). If a project fails, a log will give the error messages for troubleshooting.
4. Clicking on the icon to the left of a project name opens up the results page for that project.

Project	Type	Status	Shared	Public	Created	Updated	Actions
ReadsQC_Test2	ReadsQC	Submitted	No	No	10/19/2021, 8:37:20 AM	10/19/2021, 8:39:00 AM	[Edit]
Taxonomy_Test	Read-based Taxonomy Classification	Running	No	No	10/19/2021, 8:34:32 AM	10/19/2021, 8:39:00 AM	[Edit]
ReadsQC_Test	ReadsQC	Complete	No	No	10/18/2021, 11:20:43 AM	10/18/2021, 1:22:00 PM	[Edit]
MAGs_Test	Metagenome MAGs	Complete	No	No	10/18/2021, 10:50:59 AM	10/18/2021, 11:55:00 AM	[Edit]

## Project Summary (Results)

The project summary page will show three categories. Clicking on the bar or tab opens up the information.

1. General contains the project run information.
2. "Workflow" Result contains the tabular/visual output.
3. Browser/Download Outputs contains all the output files available for downloading. There may be several folders.

**ReadsQC\_Test**

**Project Summary:**

**Description:** This is a test of the ReadsQC workflow for NMDC workflows training  
**Owner:** nmcd-edge@lanl.gov  
**Submission Time:** Mon Oct 18 2021 11:20:43 GMT-0600  
**Status:** Complete  
**Type:** ReadsQC

expand | close sections

- General
- ReadsQC Result
- Browser/Download Outputs

This example shows the results of a ReadsQC workflow run which shows run time under the General tab, the workflow results of quality trimming and filtering under the ReadsQC Results tab, and the files available for download (shown in purple) under the Browser/Download Outputs tab.

The screenshot displays the MAGE workflow interface. On the left is a dark sidebar with navigation options: Home, Tutorials, Public Projects, Upload Files, and WORKFLOWS (Metagenomics, Organic Matter). The main content area is titled 'My Projects' and 'Job Queue'. Three orange arrows point to specific sections:

- General:** Shows a table with workflow details.
 

Workflow	Run	Status	Running Time	Start	End
ReadsQC	On	Done	03:03:03	2021-10-20 20:33:14	2021-10-20 23:36:17

 Below the table is a 'Project Configuration' section with a JSON object: `{...}`.
- ReadsQC Result:** Shows the input 'SRR7877884-int' and a table of reads and bases.
 

Reads	Status
inputReads	44,943,418
inputBases	6,741,512,700
qtrimmedReads	8,583
qtrimmedBases	8,690
qfilteredReads	200,626
qfilteredBases	29,786,796
ktrimmedReads	6,186,690
ktrimmedBases	354,478,706
kfilteredReads	100,360
kfilteredBases	14,684,762
outputReads	33,510,668
outputBases	4,868,925,674
gcPolymerRatio	0.42
- Browser/Download Outputs:** Shows a file list for the 'SRR7877884-int' project.
 

File	Size	Last Modified
filterStats2.txt	706 B	14 days ago
filterStats.json	337 B	14 days ago
filterStats.txt	287 B	14 days ago
SRR7877884-int.anqpht.fq.gz	1713.08 MB	14 days ago

The full Metagenome pipeline or “Multiple Workflow” run results show the results of each workflow under a separate tab and the associated files available for download are in separate workflow folders under the Browser/Download Outputs tab.

The screenshot displays the NMDC Metagenome Pipeline web interface. The top navigation bar includes the NMDC logo, a hamburger menu, and links for 'My Projects' and 'Job Queue'. A dark sidebar on the left contains navigation options: Home, Tutorials, Public Projects, Upload Files, WORKFLOWS, Metagenomics (with sub-options 'Run a Single Workflow' and 'Run Multiple Workflows'), and Organic Matter. The main content area is titled 'Pipeline\_Test' and features a 'Project Summary' section with the following details: Description: This is a test of the NMDC metagenome pipeline for training purposes; Owner: nmdc-edge@lanl.gov; Submission Time: Wed Oct 13 2021 16:49:56 GMT-0600; Status: Complete; Type: Metagenome Pipeline. Below the summary is a list of workflow results, each with a dropdown arrow: General, ReadsQC Result, Read-based Taxonomy Classification Result, Metagenome Assembly Result, Metagenome Annotation Result, Metagenome MAGs Result, and Browser/Download Outputs. The 'Browser/Download Outputs' section is expanded to show a table of files:

File	Size	Last Modified
MetagenomeAnnotation		
MetagenomeAssembly		
MetagenomeMAGs		
ReadbasedAnalysis		
ReadsQC		

As a second example, the next two figures show the results from the Read-based Taxonomy Classification workflow. The summary includes classified reads and the number of species identified for all of the selected taxonomy classifiers. The top ten organisms identified by each tool at three taxonomic levels is also provided. Tabs for each of the classification tools providing more in-depth results are in the Detail section. Krona plots are generated for the results at each of the three taxonomic levels for each of the tools and can also be found in the Detail section. Full results files (beyond the Top 10) and the graphics are available for download.

Read-based Taxonomy Classification Result

**Summary**

Tool	Classified Reads	Species Reads	Species
gottcha2	89,222,937	89,222,937	9
centrifuge	14,874,315	14,485,925	5,127
kraken2	30,854,417	29,421,033	2,791

**Taxonomy Top 10**

Species Genus Family

Tool	Level	Top1	Top2	Top3	Top4	Top5	Top6	Top7	Top8	Top9	T
gottcha2	species	Pseudomonas aeruginosa	Salmonella enterica	Listeria monocytogenes	Enterococcus faecalis	Lactobacillus fermentum	Bacillus subtilis	Escherichia coli	Staphylococcus aureus	Listeria phage A500	
centrifuge	species	Pseudomonas aeruginosa	Enterococcus faecalis	Bacillus subtilis	Bacillus intestinalis	Listeria monocytogenes	Lactobacillus fermentum	Pseudomonas fluorescens	Pseudomonas sp. AK6U	Salmonella enterica	E
kraken2	species	Pseudomonas aeruginosa	Salmonella enterica	Bacillus subtilis	Listeria monocytogenes	Enterococcus faecalis	Lactobacillus fermentum	Escherichia coli	Staphylococcus aureus	Homo sapiens	B

**Detail**

centrifuge gottcha2 kraken2

Species Genus Family

Level	Taxonomy	Reads	Abundance
species	Pseudomonas aeruginosa	7,026,567	0.001
species	Enterococcus faecalis	5,840,658	0.002
species	Bacillus subtilis	5,224,145	0
species	Bacillus intestinalis	5,080,010	0.002
species	Listeria monocytogenes	4,963,265	0.002
species	Lactobacillus fermentum	4,223,501	0.002
species	Pseudomonas fluorescens	2,735,891	0
species	Pseudomonas sp. AK6U	2,685,240	0
species	Salmonella enterica	2,353,819	0
species	Escherichia coli	776,322	0

