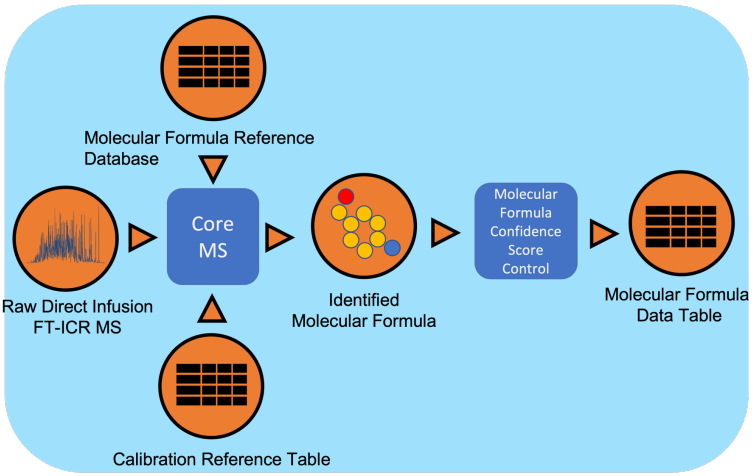
Natural Organic Matter Workflow (v4.1.5)



Overview

This workflow takes FTICR mass spectrometry data collected from organic extracts to determine the molecular formulas of natural organic biomolecules in the input sample.

Running the Workflow

Currently, this workflow can be run in NMDC EDGE or from the command line. (CLI instructions and requirements are found here.)

Input

The input for this workflow is the output from a massSpec experiment (a massSpec list) which includes a minimum of two columns of data: a mass-to-charge ratio (m/z) and a signal intensity (Intensity) column for every feature in the analysis. A calibration file of molecular formula references is also required when running the workflow via command line. (This calibration file is built into NMDC EDGE.)

Acceptable file formats: .raw, .tsv, .csv, .xlsx

Details

Direct Infusion Fourier Transform Ion Cyclotron Resonance mass spectrometry (DI FTICR-MS) data undergoes signal processing and molecular formula assignment leveraging EMSL's CoreMS framework. Raw time domain data is transformed into the m/z domain using Fourier Transform and Ledford equation. Data is denoised followed by peak picking, recalibration using an external reference list of known compounds, and searched against a dynamically generated molecular formula library with a defined molecular search space. The confidence scores for all the molecular formula candidates are calculated based on the mass accuracy and fine isotopic structure, and the best candidate assigned as the highest score. This workflow will not work as reliably with Orbitrap mass spectrometry data.

Software Versions

- CoreMS (2-clause BSD)
- Click (BSD 3-Clause "New" or "Revised" License)

Output

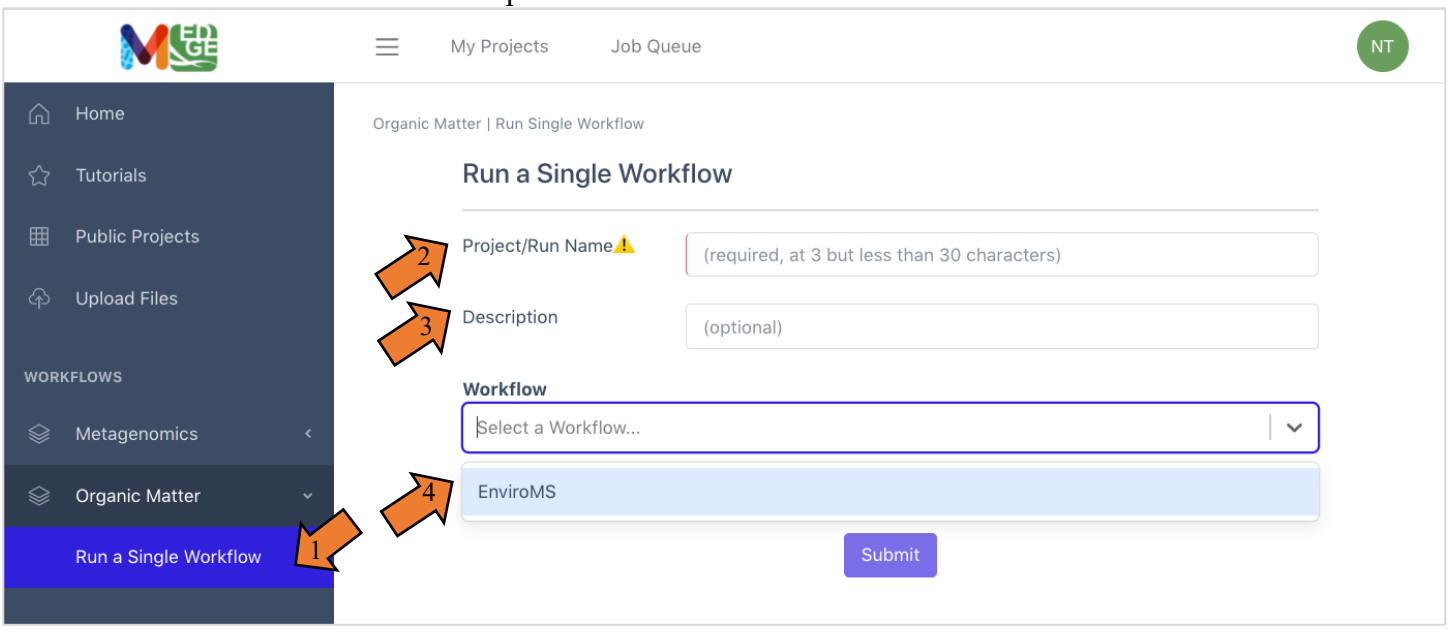
The primary output file is the Molecular Formula Data Table (in a .csv file).

Primary Output Files	Description
INPUT_NAME.csv	m/z, Peak height, Peak Area, Molecular Formula IDs, Confidence Score, etc.

Running the Natural Organic Matter Workflow in NMDC EDGE

Select a workflow

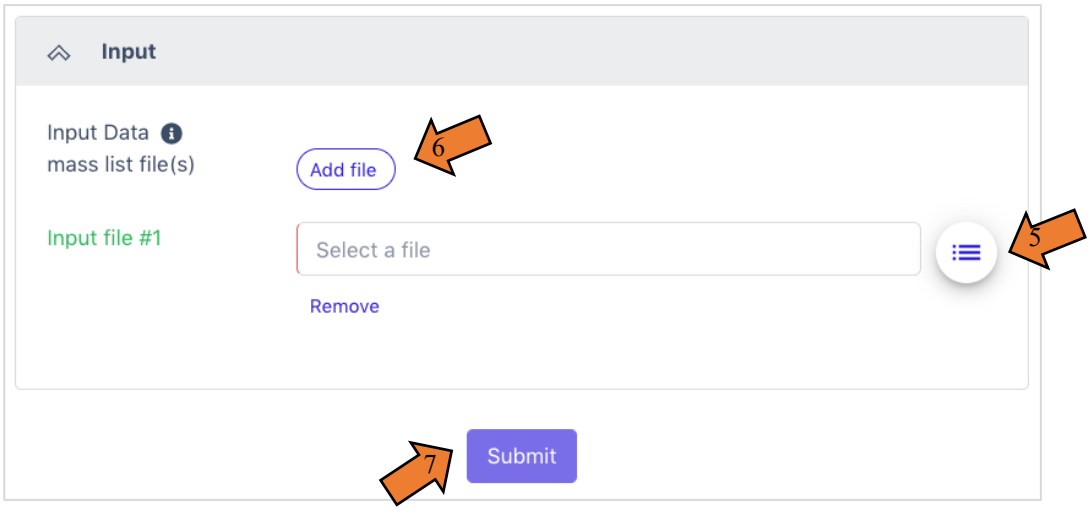
1. From the Organic Matter category in the left menu bar, select 'Run a Single Workflow'.
2. Enter a ***unique*** project name with no spaces (underscores are fine).
3. A description is optional, but helpful.
4. Select 'EnviroMS' from the dropdown menu under Workflow.



Input

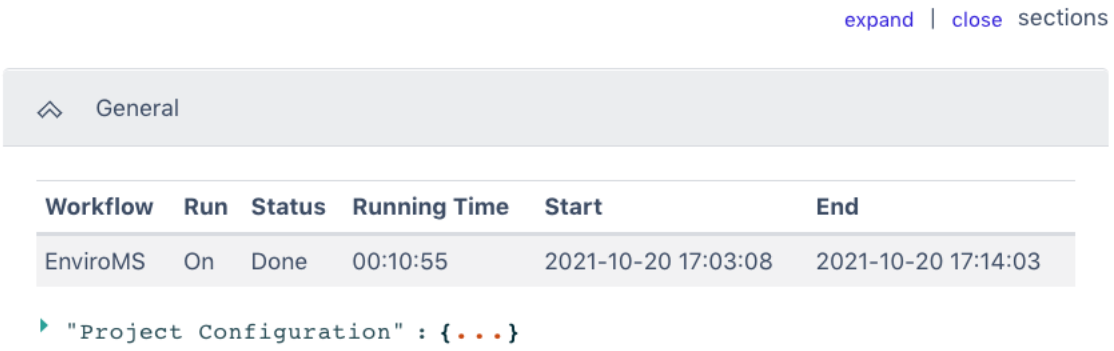
The Natural Organic Matter workflow input is the output from a massSpec experiment (a massSpec list) with a minimum of two columns of data: a mass-to-charge ratio (m/z) and a signal intensity (Intensity) column for every feature in the analysis. **Acceptable file formats:** .tsv, .csv, .raw, .xlsx

5. Click the button to the right of the input blank for data to select the data file for the analysis. (If there are separate files, there will be two input blanks.) A box called 'Select a File' will open to allow the user to find the desired file(s) from the public data folder or files uploaded by the user.
6. Additional input files can be added by clicking the 'Add file' button to create additional input blanks.
7. Once all the input files have been selected, click 'Submit'.



Output

The General section of the output shows which workflow was run and the run time information. The Project Configuration can be seen by clicking the three dots in the bracket.



The Browser/Download Output section provides output files available to download. The primary output files are: the Molecular Formula Data-Table (.csv file) containing m/z measurements, Peak height, Peak Area, Molecular Formula Identification, Ion Type, and Confidence Score.