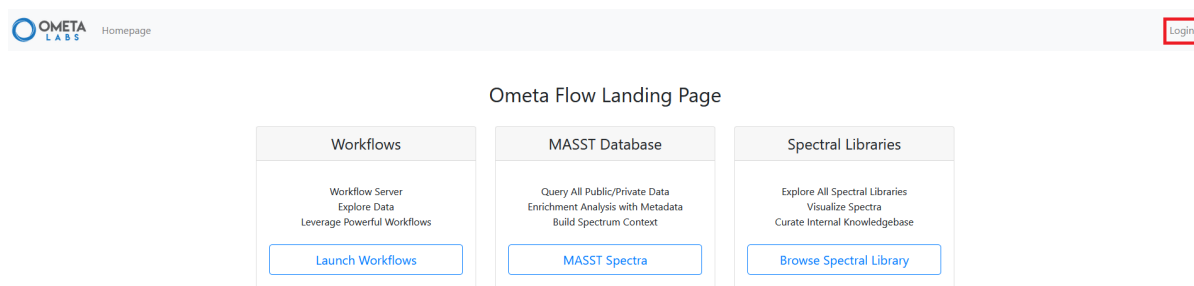


Tutorial: creating your first molecular network

The following tutorial provides step-by-step instructions for creating molecular networks as described in the accompanying white paper. It assumes you have a basic familiarity with MS/MS data and molecular networking. If you are getting lost (or just want to brush up on terminology), check out the metabolomics resources on our website.

1. Log in to Ometa Flow

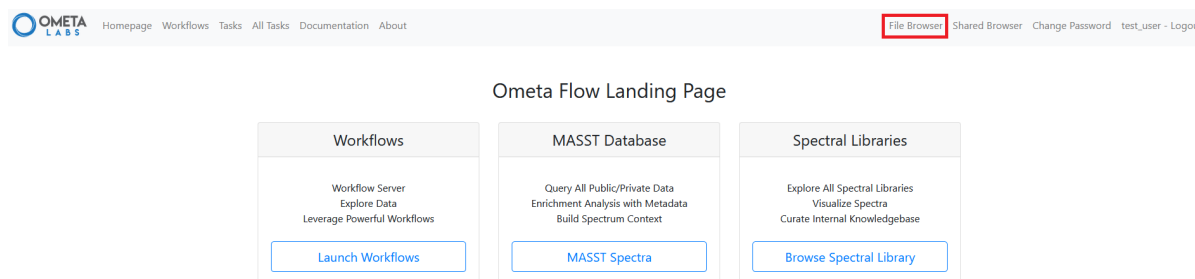
To log in to Ometa Flow, you must first be assigned a username and password by your admin account. Once your username and password have been created, you can login by clicking the button here.



The screenshot shows the Ometa Flow Landing Page. At the top left is the Ometa Labs logo and the text "Homepage". At the top right is a "Login" button, which is highlighted with a red box. Below the header, the page is titled "Ometa Flow Landing Page" and contains three main sections: "Workflows", "MASST Database", and "Spectral Libraries". Each section has a list of features and a corresponding button: "Launch Workflows", "MASST Spectra", and "Browse Spectral Library".

2. Upload your Data

In order to analyze your data in Ometa Flow, you must first upload it to the server. To do so, click on the File Browser tab, which will open up your own personal file browser.



The screenshot shows the Ometa Flow Landing Page with a different navigation menu. At the top left is the Ometa Labs logo and the text "Homepage Workflows Tasks All Tasks Documentation About". At the top right is a "File Browser" button, which is highlighted with a red box, along with "Shared Browser", "Change Password", "test_user", and "Logout". Below the header, the page is titled "Ometa Flow Landing Page" and contains three main sections: "Workflows", "MASST Database", and "Spectral Libraries". Each section has a list of features and a corresponding button: "Launch Workflows", "MASST Spectra", and "Browse Spectral Library".

Here, you can create and organize folders and upload data using either drag/drop or the upload button shown here:





Shared Browser

If you want your data to be accessible to everyone on your server, upload it to the Shared Browser. This is an easy way to share data with colleagues within your organization.

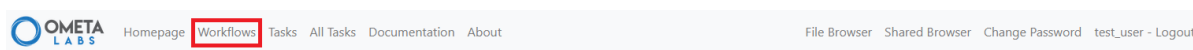
For this tutorial, download the provided zipped file, extract the files, and upload them to your File Browser. This zipped folder contains nine .mzML files from a public dataset containing fecal samples from vegan and omnivore volunteers. The files are organized into folders for procedural blanks, omnivore samples, and vegan samples, along with a Metadata_NIST_omnivore.txt file.

Accepted file types

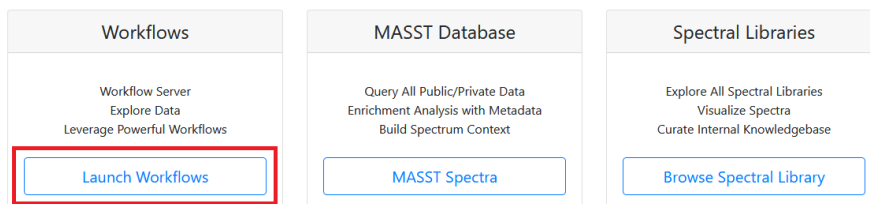
Omata Flow accepts MS/MS files in .mzML, .mzXML, and .mgf formats. For more information, see our [data compatibility documentation](#).

3. Launch a workflow

Omata Flow offers many ways to analyze mass spectrometry data, which we organize into workflows. To access a workflow, click on the Workflows tab at the top right or the Workflows box on the landing page.



Omata Flow Landing Page



This pulls up all the workflows available through Omata Flow. Today, we'll be using one of our



most popular workflows, classical molecular networking. You can find this workflow by looking for classic_networking in the workflow name, or typing it into the search bar on the top right. Once you've found classic_networking, click on the Launch Workflow button to begin.

Ometa Labs Flow - All Workflows

Show entries Search:

Workflow Name	Description	Version	Type	Launch
adduct_detection	Search for adducts from RT and Delta Mass relationships.	15.41	SYSTEM	Launch Workflow
bioactivity_network	bioactivity_network	15.41	SYSTEM	Launch Workflow
classic_networking	LC - Classical Molecular Networking	15.35	SYSTEM	Launch Workflow

This will pull up the workflow input page for classic_networking. If you start to scroll down, you'll notice there are a lot of parameters you can set for this workflow (don't worry, most workflows aren't nearly as complex). If you have questions about what each parameter means, check out the workflow documentation, which you can find by clicking the documentation link at the top of the page. If you have a question about a specific parameter, you can hover over the question mark next to it to get a little more information about it.

Workflow Input - classic_networking

Workflow Version - 15.35

[Workflow Documentation](#)

LC - Classical Molecular Networking - Molecular networks are visual displays of the chemical space present in tandem mass spectrometry (MS/MS) experiments. This visualization approach can detect sets of spectra from related molecules (molecular networks), even when the spectra themselves are not matched to any known compounds. Classical molecular networking is well suited for discovery and can be analyzed directly from raw mass spectrometry files.

Job Description

Enter Job Description

Workflow Presets

Low Resolution High Resolution Demo Data

File Selection

File Selection - Input Data Folder - Group

Documentation

If you have a question that isn't answered in our documentation, [let us know](#). We're always trying to improve our documentation and appreciate your feedback.



Enter a job description in the Job Description text box. This can be anything (it can even be left blank), but from personal experience we recommend you make it something informative that will help you remember what you were doing when you come back to this in six months.

4. Select your data

After you've entered a job description, you can start selecting what data you want to analyze. Clicking one of the green buttons in the File Selection section will open up a new window where you can select your data.

File Selection

File Selection - Input Data Folder - Group 1 ?



Here, you can click through your user's file browser or the shared browser to select samples. For this tutorial, navigate to the Ometa_Flow_Demo folder you uploaded and find the folders corresponding to Blanks, Vegan, and Omnivore samples. Select the folder labeled "Blanks" and click "Add Selected File to Input Data Folder - Group 1".





Next, select the Omnivore folder and click "Add Selected File to Input Data Folder - Group 2".
Select the Vegan folder and click "Add Selected File to Input Data Folder - Group 3".





Selecting Files

You can select data as folders or individual files. If you select a folder, all files in that folder will be included. If you have data you want to analyze as separate groups, we suggest uploading or reorganizing those data into separate folders so you can select the folder instead of clicking on each individual file.

Close out of the window by clicking "Close" on the bottom left. You'll now see your selections for each group. If you added something in error, you can always click the "Remove File Selection" button to remove a file or folder and then click on one of the green buttons to select the right file. Your File Selection should look like this:

File Selection

File Selection - Input Data Folder - Group 1 ?

Remove File Selection USERUPLOAD/sydney/CMN_Training_Demo/Blanks

Select Input Data Folder - Group 1

Show/Hide Manual File Selection

File Selection - Input Data Folder - Group 2 ?

Remove File Selection USERUPLOAD/sydney/CMN_Training_Demo/Omnivore

Select Input Data Folder - Group 2

Show/Hide Manual File Selection

File Selection - Input Data Folder - Group 3 ?

Remove File Selection USERUPLOAD/sydney/CMN_Training_Demo/Vegan

Select Input Data Folder - Group 3

Show/Hide Manual File Selection

Finally, let's select our metadata. Scroll down to "File Selection - Metadata Path" and click the green button. This will bring up the same file selection window. Navigate to your Ometa_Flow_Demo folder and select the "Metadata_NIST_omnivore.txt" file. Click on "Add Selected File to Metadata Path" to add your metadata.



File Selector For Analysis ✕

- └─ sydney
 - └─ CMN_Training_Demo
 - └─ Blanks
 - └─ Metadata_NIST_omnivore.txt
 - └─ Omnivore
 - └─ Vegan
 - └─ GNPS-NIH-CLINICALCOLLECTION1.mgf
 - └─ NIST_Omnivore
 - └─ mgf_test.mgf
 - └─ query_table.tsv
 - └─ shared
 - └─ Datasets
 - └─ Task
 - └─ Libraries

Add Selected File to Input Data Folder - Group 1

Add Selected File to Input Data Folder - Group 2

Add Selected File to Input Data Folder - Group 3

Add Selected File to Input Data Folder - Group 4

Add Selected File to Input Data Folder - Group 5

Add Selected File to Input Data Folder - Group 6

Add Selected File to Spectral Library Folder

Add Selected File to Metadata Path

Add Selected File to Select Exact Mass Database File (tsv or xls)

Manage Task Imports Visualize First Selected MS Data File

Close



Metadata

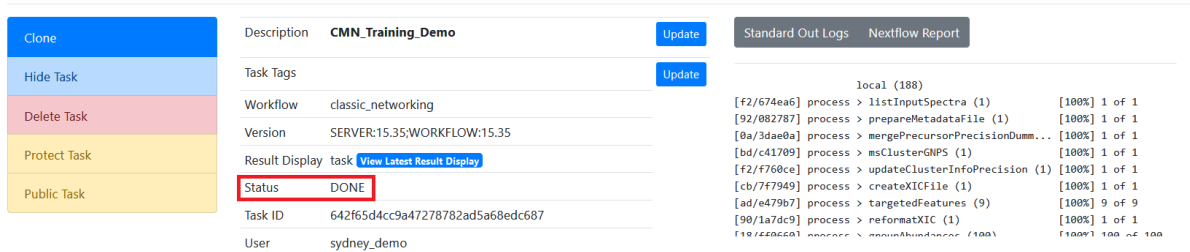
For more information on how to format your metadata and to see an example metadata template, see our [metadata documentation](#).

5. Start a workflow

For today, we're going to keep all the other parameters as their default values. Scroll down to the bottom of the page and click the "Submit Workflow" button to begin the analysis.

This will bring up a task status page, which provides information on the analysis and results once it's finished. To check its progress, check the Status box (highlighted below). Right now it should either say "QUEUED" or "RUNNING". Once your analysis is finished, it will say "DONE".

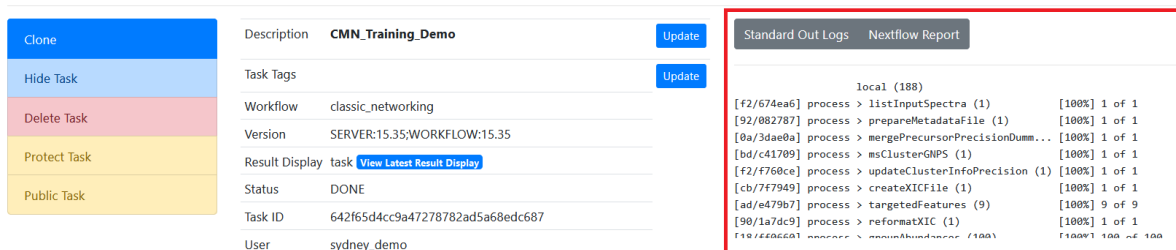
Ometa Labs Task Status Page



The screenshot shows the Omata Labs Task Status Page for a task named "CMN_Training_Demo". On the left, there is a vertical menu with buttons: Clone (blue), Hide Task (light blue), Delete Task (pink), Protect Task (yellow), and Public Task (yellow). The main content area displays task details: Description (CMN_Training_Demo), Task Tags, Workflow (classic_networking), Version (SERVER:15.35;WORKFLOW:15.35), Result Display task (View Latest Result Display), Status (DONE, highlighted with a red box), Task ID (642f65d4cc9a47278782ad5a68edc687), and User (sydney_demo). On the right, there are buttons for Standard Out Logs and Nextflow Report, and a terminal window showing process logs with progress indicators like [100%] 1 of 1.

If something went wrong, the Status box will read "FAILED". If this happens, check the right hand text box to get a description of the error (you can click the Nextflow Report button to open the message in a window that's easier to read). If you ever need to contact us about an error, please copy/paste this error message text in your email to us so that we can help you troubleshoot.

Omata Labs Task Status Page



This screenshot is similar to the previous one, but the "Status" box is not highlighted. Instead, the "Standard Out Logs" and "Nextflow Report" buttons and the terminal window below them are highlighted with a red box. The terminal window shows the same process logs as in the previous screenshot.

There are a couple of other useful options at the top of the page here. The first is the Clone button. This button will create a new workflow with exactly the same parameters and file selections as your current workflow. This is especially useful if you want to tweak a single



parameter without changing anything else.

Ometa Labs Task Status Page

Description	CMN_Training_Demo	Update
Task Tags		Update
Workflow	classic_networking	
Version	SERVER:15.35;WORKFLOW:15.35	
Result Display	task View Latest Result Display	
Status	DONE	
Task ID	642f65d4cc9a47278782ad5a68edc687	
User	sydney_demo	

```
local (188)
[f2/674ea6] process > listInputSpectra (1) [100%] 1 of 1
[92/082787] process > prepareMetadataFile (1) [100%] 1 of 1
[0a/3dae0a] process > mergePrecursorPrecisionDumm... [100%] 1 of 1
[bd/c41709] process > msClusterGNPS (1) [100%] 1 of 1
[f2/f760ce] process > updateClusterInfoPrecision (1) [100%] 1 of 1
[cb/7f7949] process > createXICFile (1) [100%] 1 of 1
[ad/e479b7] process > targetedFeatures (9) [100%] 9 of 9
[90/1a7dc9] process > reformatXIC (1) [100%] 1 of 1
```

You also have options to Hide, Delete, or Protect Tasks, as well as an option to Public Task. By default, all your tasks are only visible to you and your admin account. If you want to share a task your colleagues, you can click Public Task to make the task visible to anyone on your server. To find all public tasks on your server, go to the Omata Labs homepage, find the Public Tasks box, and click on "Browse Tasks".

Public Tasks

Never fear, public tasks aren't *actually* public. They're only visible to people with access to your server (eg, people within your organization).

You can also share a task by copy/pasting the url of your task status page. This url will allow anyone on your server to access the task. If you don't want this to be the case, you can use the Protect Task button to make the url visible only to you.

This task should take 5-10 minutes to run. While you're waiting, feel free to exit out of this page. You can always come back by clicking the Task tab on the top left corner of your screen. This will pull up a list of all the tasks you've run on the server, as well as their status. To get back to the task status page for any task, click the "View Task" link.

6. Explore your results

Further down on the Task Status page is the Task Result section. This contains links to tables and visualizations from your workflow. There are a lot of results for classical molecular networking, but one of the first you should check out is the "Visualize Full Network in Browser" view.



Task Results Links

[Copy Results to User Space](#) [Import Task for Reanalysis](#) [Download All Results](#) [Browse All Results](#)

Summary Views

[Files List](#) [Clusters Table](#) [Library Identifications](#) [Library Identifications \(Top K\)](#) [View Analysis Metadata](#) [View Upset Plots](#)

Networking Views

[Visualize Full Network in Browser](#) [Visualize Full Network in Browser with Singletons](#) [Visualize Full Network in Browser with XIC Quant](#) [Network Components List](#) [Network Pairs Unfiltered](#) [Network Pairs Filtered](#)

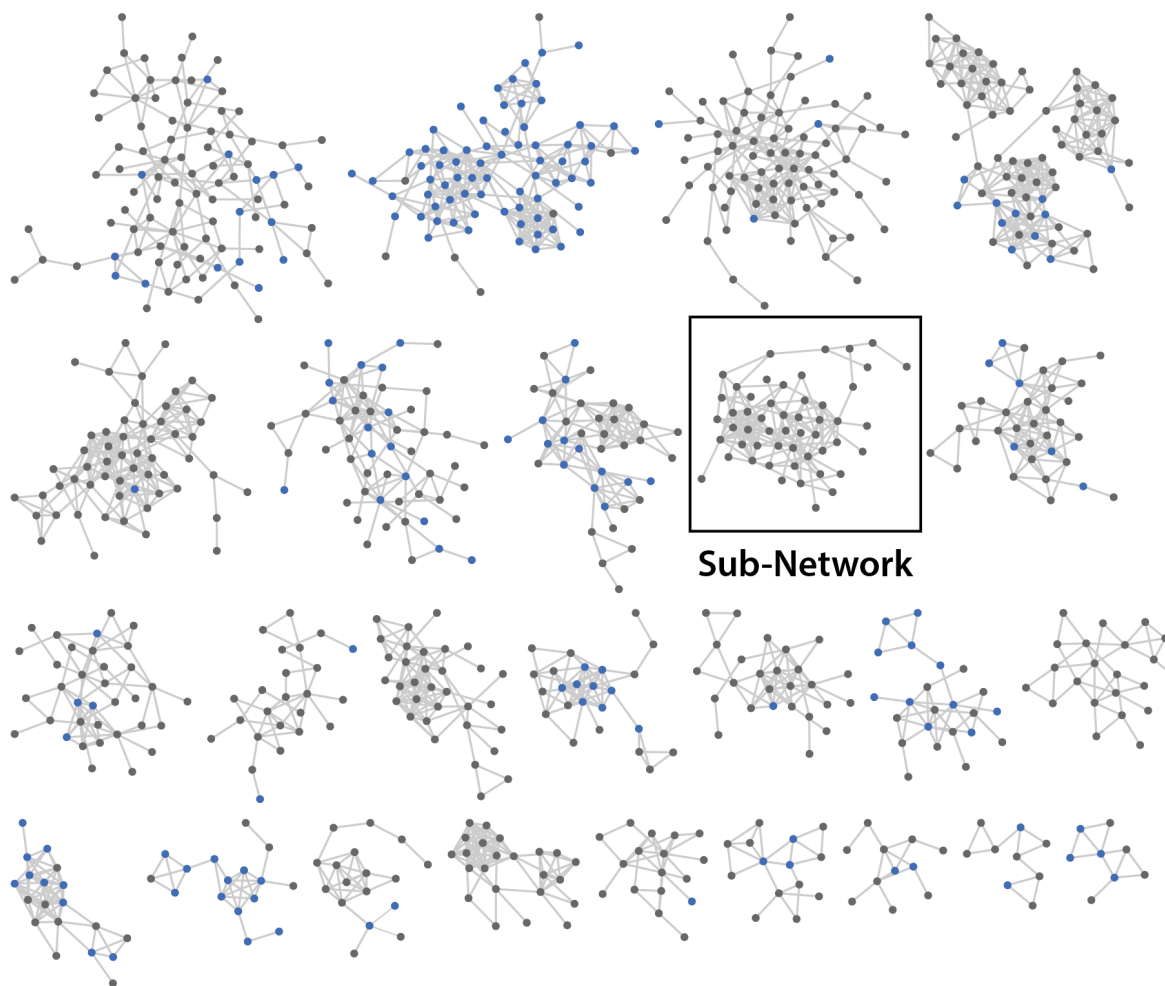
All Results

If you would like to see all the results in file/folder format instead of clicking through different views, check out the blue "Browse All Results" button on the right.

This will bring up the entire molecular network from your experiment. In basic terms, the molecular network has grouped the metabolites from our samples into sub-networks based on their structural similarity. Each sub-network usually contains a single class of compounds (eg, lipids or sugars).



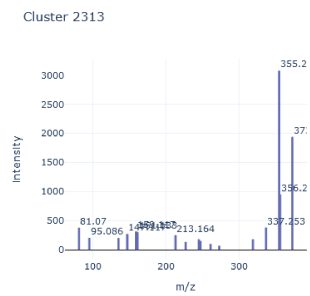
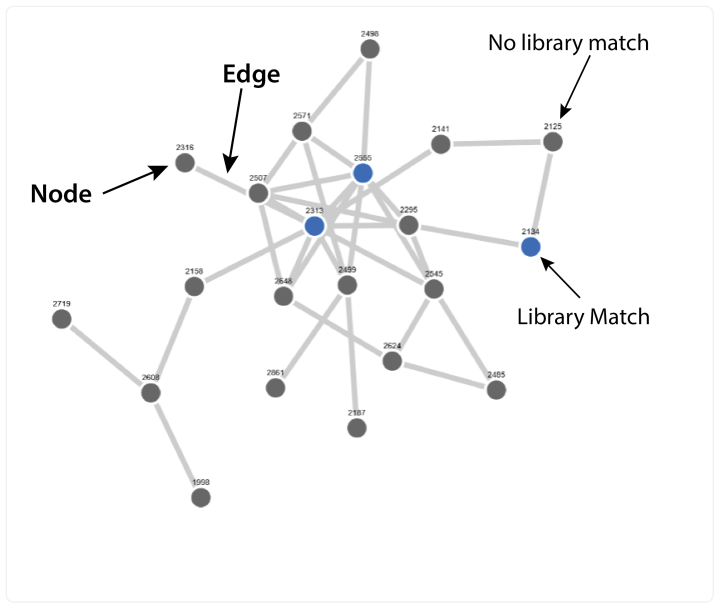
Full Molecular Network



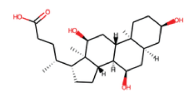
If you zoom in on one of these sub-networks, you'll notice they're made up of **nodes** and **edges**. Nodes represent each metabolite, and the edges between them indicate that there is structural similarity between those metabolites. Any nodes that are colored blue matched a library spectra, while gray nodes had no library matches. If you click on a node, you'll notice the spectrum for that metabolite will come up on the right hand side of the screen, along with information on m/z, retention time, and library match (if any).



Reset Zoom/Layout Export PNG Export SVG + -



View Spectrum US! View Node LCMS Annotate Spectrum



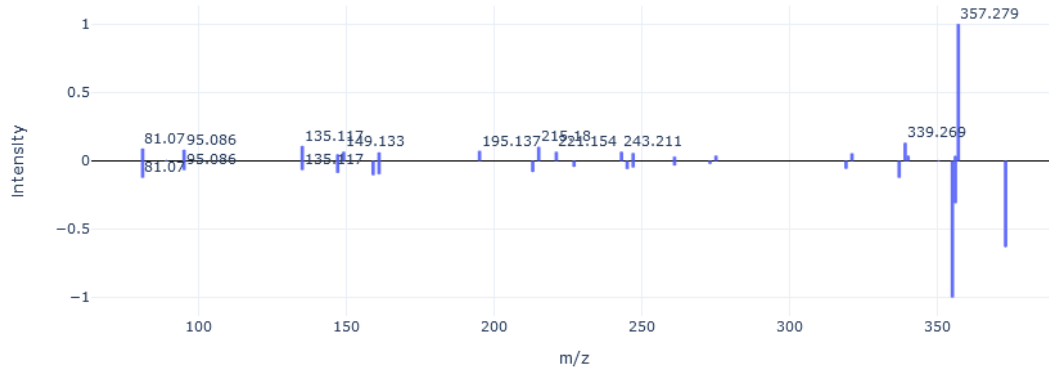
Name	Value
Cluster	2313
m/z	426.322
rt seconds	242.84
rt min	4.0473
Charge	1
Component	25
Number Spectra	10
Library	Cholic acid

Network Tables

Clicking on the edge between two nodes will bring up both spectra connected to that edge. If you scroll down, you'll notice the Spectrum Comparison window on the right with a mirror plot comparing the two spectra. To see this mirror plot in a larger format, click the "View Mirror USI" button. This will bring up a new window with a larger mirror plot and a peak list for each spectrum.



Spectrum Comparison 2158 vs 2313

[View Mirror US!](#)

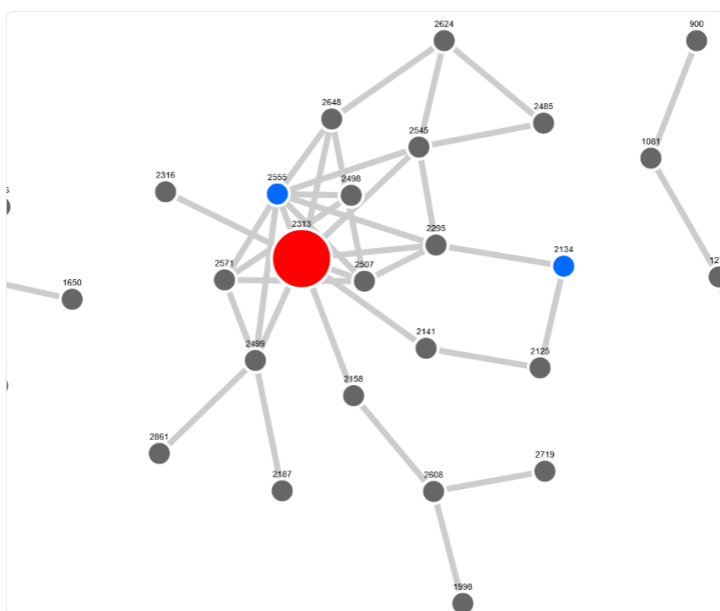
Name	Value
DeltaMZ	-15.995
score	0.729
matched_peaks	-1

7. Visualize your metadata

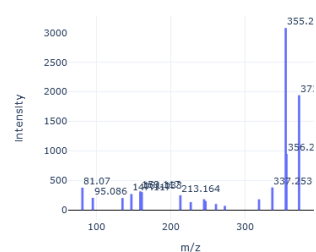
Now let's look at some specific nodes. Scroll down to the Network Controls and type "2313" into the Node Search box. This will highlight node 2313 in your network. Click on this node (you can delete the text in the node search box to get rid of the red highlight). It should be a library match to Cholic Acid.



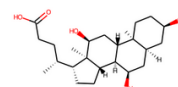
Reset Zoom/Layout Export PNG Export SVG + -



Cluster 2313



View Spectrum USI View Node LCMS Annotate Spectrum



Name	Value
Cluster	2313
m/z	426.322
rt seconds	242.84
rt min	4,0473
Charge	1
Component	25
Number Spectra	10
Library	Cholic acid

Network Tables

i Node IDs

If 2313 isn't showing up, or it isn't a match to Cholic acid, you may have a slightly different node ID. You can check which ID matches to Cholic acid by scrolling down to the Network Tables and typing "Cholic" into the library_compound_name column (note, search is case-sensitive). This should bring up a row with the correct ID.

Cholic acid is a bile acid, a class of compounds that share a core structure but are extensively modified by the microbiome. The other annotated nodes in this sub-network correspond to other bile acids, but you'll also see a few unannotated nodes as well. These are likely modified bile acids that don't exist in our spectral libraries.

To get a little more information, let's layer some metadata onto this network. To do so, scroll down to the Pie Configuration box. Click on the Attribute Selection dropdown and select "ATTRIBUTE_Diet". You'll notice that the groups change to assign colors to blank, omnivore, and vegan samples. Click on the Show Pie Charts dropdown and select "Yes".



Pie Configuration ?

Attribute Selection ?

■ Group 1 ?

■ Group 2 ?

■ Group 3 ?

■ Group 4 ?

■ Group 5 ?

■ Group 6 ?

Show Pie Charts ?

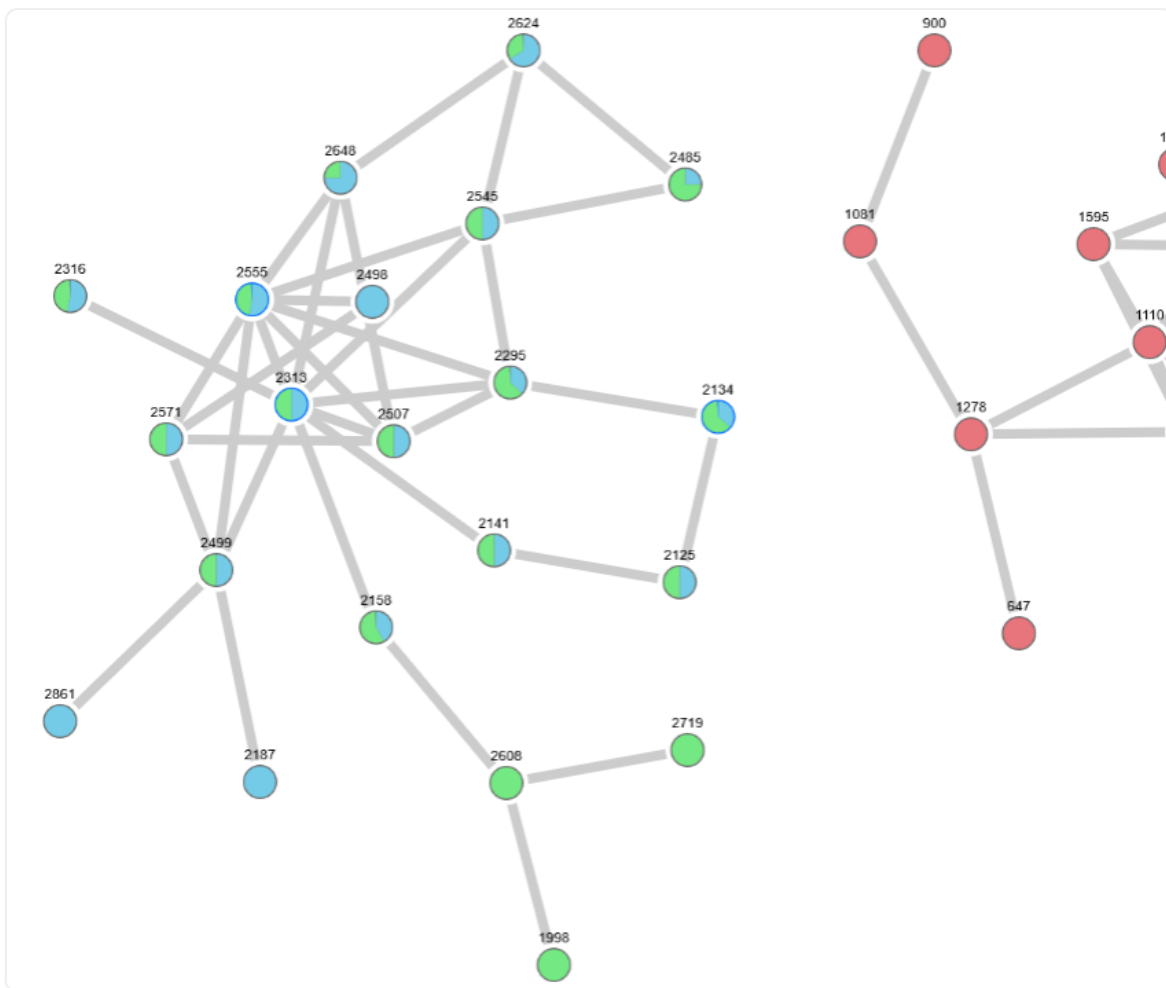
807 Nodes, 1333 Edges

Advanced Network Coloring

Scroll back up to the network, and you'll notice that the nodes are different colors. The colors correspond to how many samples in a group contained that specific metabolite. The nodes in our bile acid sub-network are blue and green, indicating that they were found in vegan and omnivore samples. Some nodes in other sub-networks are completely pink, indicating that they were only found in blanks. This visualization is a quick and easy way to identify metabolites that



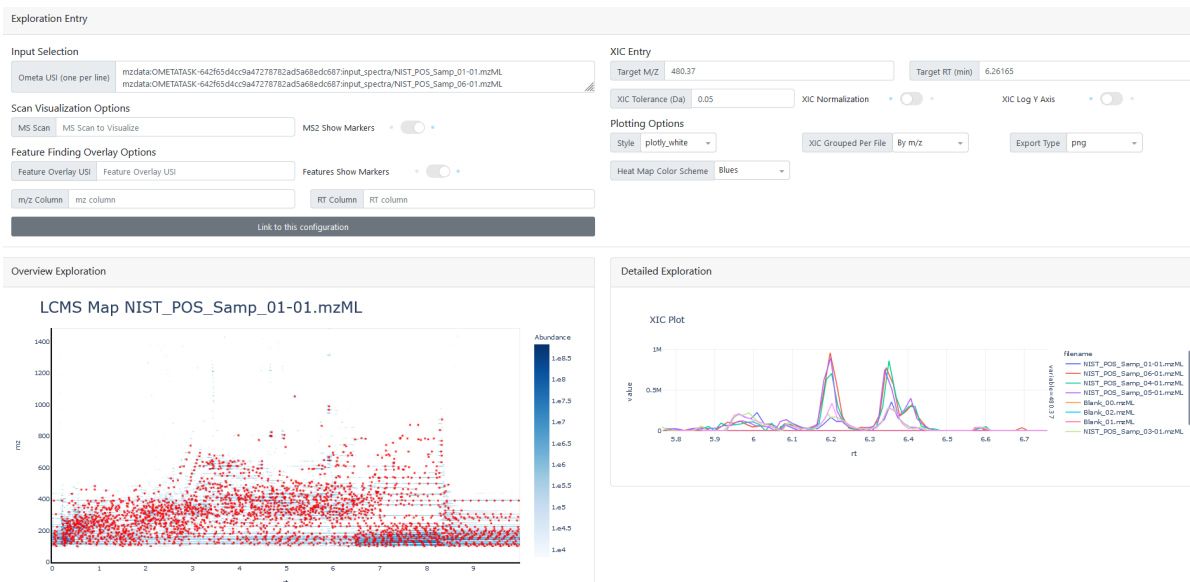
are unique to a specific group, or metabolites that show up everywhere.



Let's click on Node 2608, which only shows up in vegan samples. To get a more in-depth look at this peak, go to the Spectrum Details box and click on "View Node LCMS". This will bring up a new window with an interactive XIC plot.

The Overview Exploration plots on the left provide an overview of your entire sample. You'll find an LCMS map that plots all m/z values measured across the entire MS/MS run, and plots red x's whenever an MS2 event occurred (this is a quick and easy way to check that you got good MS2 coverage across your run). You'll also find the TIC trace for the full sample. The Detailed Exploration on the right draws the XIC plot for your metabolite of interest. You'll notice that the XIC Entry was auto-populated with the m/z and target RT of your metabolite.



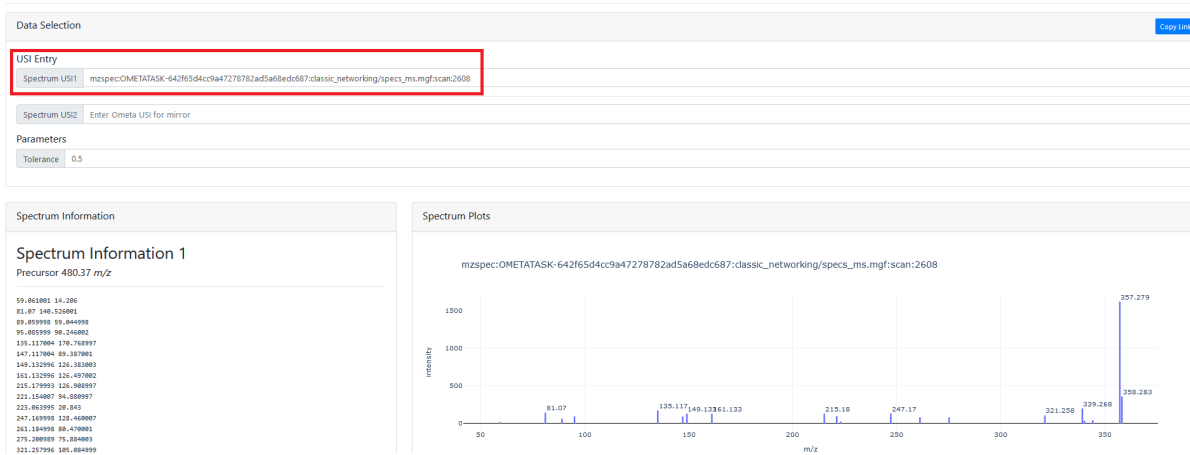


There are other peaks with an m/z of 480.37 besides our peak at 6.2 minutes. This is common for bile acids, which have a lot of isomers. To see how many other peaks match this m/z value across your run, you can change the Target RT to a range of values (say 1-7) and the plot will update to show retention time from 1-7 minutes. Be patient, sometimes the new plot can take some time to load.

8. Search for unknown molecules using MASST

Now that we've identified an interesting unknown compound with good peak shape, let's see where else this metabolite shows up. Go back to your Molecular Networking Browser and click on node 2608. This will bring up the spectrum in the Spectrum Details box. Click on "View Spectrum USI" to pull up a new window with the spectrum, peak list, and Spectrum USI.

Omata Spectrum Resolver

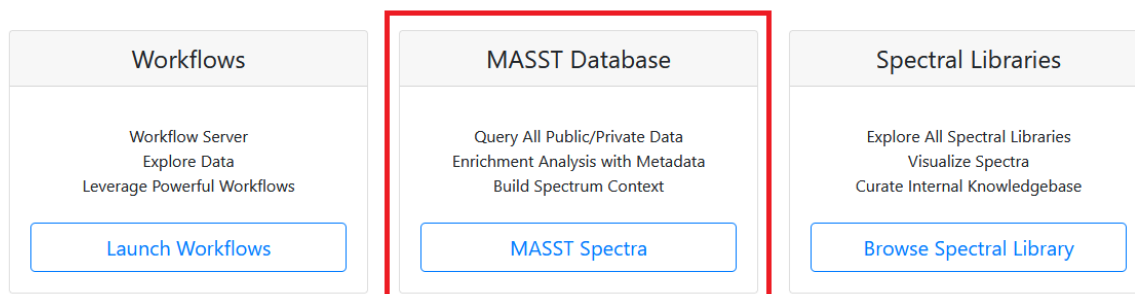


Spectrum USI

USI stands for Universal Spectrum Identifier and is a unique identifier for your spectrum. This unique identifier can be used across the Ometa Flow platform and will always point to this specific spectrum. For more information, check out our [USI documentation](#).

Copy the spectrum USI and navigate to the MASST Database box on the Ometa Flow Homepage.

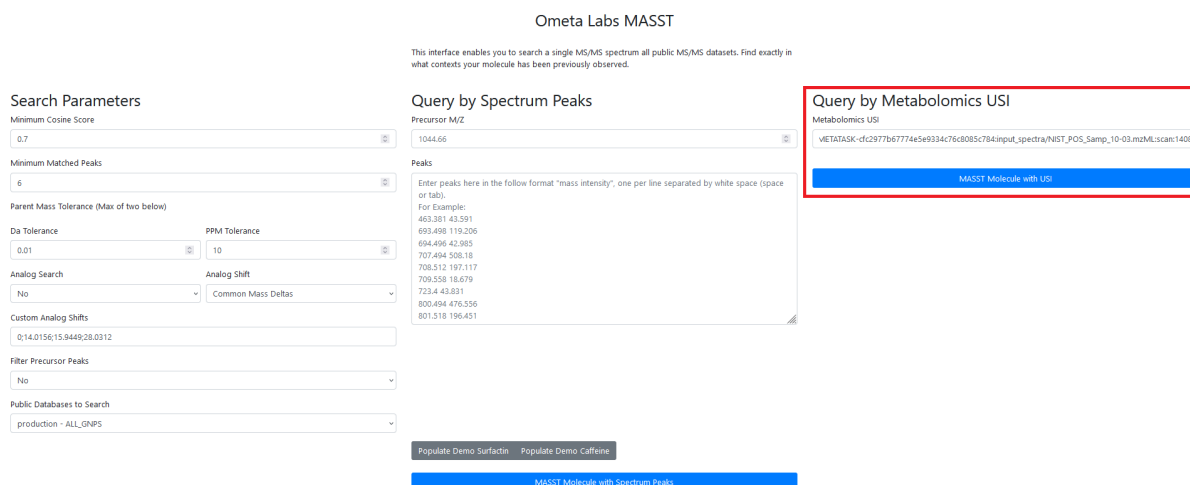
Ometa Flow Landing Page



The landing page features three main sections:

- Workflows:** Includes "Workflow Server", "Explore Data", and "Leverage Powerful Workflows". A button labeled "Launch Workflows" is at the bottom.
- MASST Database:** Includes "Query All Public/Private Data", "Enrichment Analysis with Metadata", and "Build Spectrum Context". A button labeled "MASST Spectra" is at the bottom. This section is highlighted with a red border.
- Spectral Libraries:** Includes "Explore All Spectral Libraries", "Visualize Spectra", and "Curate Internal Knowledgebase". A button labeled "Browse Spectral Library" is at the bottom.

Clicking "MASST Spectra" will bring you to the Ometa Labs MASST page. On the left, you can edit your search parameters and choose which database to search (for more information, check out the [MASST documentation](#)). Paste your USI into the Metabolomics USI box on the far right and click "MASST Molecule with USI".



The interface is titled "Ometa Labs MASST" and includes the following sections:

- Search Parameters:** Fields for "Minimum Cosine Score" (0.7), "Minimum Matched Peaks" (6), "Parent Mass Tolerance" (Max of two below), "Da Tolerance" (0.01), "PPM Tolerance" (10), "Analog Search" (No), "Analog Shift" (Common Mass Deltas), "Custom Analog Shifts" (0:14.0156;15.9449;28.0312), "Filter Precursor Peaks" (No), and "Public Databases to Search" (production - ALL_GNPS).
- Query by Spectrum Peaks:** "Precursor M/Z" field with value 1044.66. A list of peaks is shown: 463.381 43.591, 693.498 119.206, 694.496 42.985, 707.494 508.18, 708.512 197.117, 709.558 18.679, 723.4 43.831, 800.494 476.556, 801.518 196.451.
- Query by Metabolomics USI:** A text box containing the USI: `VIETASK-df2977b67774e5e9334c76c8085c784:input_spectra/MST_POS_Samp_10-03.mzMLscan1408`. A button labeled "MASST Molecule with USI" is below it. This section is highlighted with a red border.

Buttons at the bottom include "Populate Demo Surfactin", "Populate Demo Caffeine", and "MASST Molecule with Spectrum Peaks".

This will bring up a results page with the list of datasets, files, and specific scans that match the spectrum from node 2608. This can give you an idea of how widely distributed this unknown bile acid is across species, tissue types, and even diseases.



MASST Results

Your MASST results will greatly depend on which database you search. The most basic version will provide a list of filenames that contained matched scans. If your database has curated metadata, you may also be able to sort matches by organism, phenotype, sample collection method, etc. For questions on how to get the most out of your MASST searches, [contact us](#).

9. Illuminate substructure information with MassQL

MassQL is a programmatic language for searching MS/MS data. It allows you to flexibly search your data for things such as specific MS2 fragment ions, MS1 isotopic patterns, or neutral losses. Our results so far indicate that our unknown metabolite of interest (node 2608) is likely a modified bile acid. Bile acids share handful of core structures that can be conjugated to a variety of modifications. The core structures differ in their number of hydroxyl groups - for instance, a dihydroxylated bile acid will have a two -OH group attached to its core structure. Since each -OH group will slightly change the mass of the core structure, we can identify specific MS2 peaks that are unique to mono, di, or trihydroxylated bile acids. By searching for those MS2 peaks, we can determine the core structure of our unknown metabolite of interest.

We're not going to do a deep dive into MassQL here, but if you want to learn more, check out our [massql documentation](#) or our [library generation tutorial](#). For now, go back to your Molecular Networking Browser (as a reminder, you can access this by going to the Task Status page for your classic_networking job and clicking "View Full Network in Browser"). Scroll down to the Network Controls and paste this text into the MassQL highlight text box:

```
QUERY scaninfo(MS2DATA) WHERE  
MS2PROD=339.27:TOLERANCEMZ=0.01:INTENSITYPERCENT=5 AND  
MS2PROD=321.26:TOLERANCEMZ=0.01:INTENSITYPERCENT=5 AND  
MS2PROD=161.13:TOLERANCEMZ=0.01:INTENSITYPERCENT=5
```

This text may look daunting, but it's really just searching for three MS2 peaks at m/z 339.27, 321.26, and 161.13. These peaks are unique to dihydroxylated bile acids. The TOLERANCEMZ qualifier indicates the mass tolerance for a match, and the INTENSITYPERCENT qualifier indicates that the peak has to be at least 5% of the highest peak for that spectrum (a good way to filter out noise). After a moment, you should see a green box appear that says "Valid MassQL Query".



Automated MassQL Queries

Don't want to create a query yourself? The networking dashboard can create basic MassQL queries for you using the MS2 m/z highlight text box.

Scroll back up to your network. You'll notice that all nodes that matched your MassQL query are now highlighted in purple - including node 2608! From just your classic_networking results, you've gone from an unknown metabolite to a modified dihydroxylated bile acid that's upregulated in people with a vegan diet. Congrats!

The screenshot displays the networking dashboard interface. On the left, a network graph shows nodes and edges, with several nodes highlighted in purple. In the center, a mass spectrum plot shows relative intensity versus m/z, with a prominent peak at m/z 260.8. On the right, a table lists mass spectrometry data:

Name	Value
Cluster	2608
m/z	448.37
# Records	375,699
# m/z	6,2017
Charge	1
Component	25
Number Spectra	4

Below the spectrum is a 'MassQL Results Table' with columns for m/z, scale, charge, and pvalue. The table lists several rows, with the top row being m/z 2608, scale 1, charge 1, and pvalue 408.527. Below this is a 'Full Network Table' with columns for ID, m/z, pvalue, and component name. The table lists various metabolites, with the top row being ID 2608, m/z 260.8, pvalue 4.0073, and component name 1230mz2_component_name. On the right side of the dashboard, there is a 'Network Controls' panel with various filters and options. A red box highlights the 'MassQL Highlight' section, which contains a text input field with the query: `QUERY pvalue<=0.001 WHERE MS2PROQ=2608 OR MS2PROQ=2608 OR MS2PROQ=2608`. Below the input field is a 'Valid MassQL Query' button.

The data used in this tutorial is publicly available in MassIVE under the accession MSV000086989.