

MultiAlign – Update

3-7-2014

Brian LaMarche

MultiAlign

- Updated feature finding algorithm to build XIC's for LC-MS features using the RAW data
 - Better abundance calculation
- Improved traceback algorithm to link to MS/MS spectra to new LC-MS features
 - Also integrated MSGF+ reading
- Indexed database to increase database access speed
- Fixed a few bugs in some alignment wrapper code

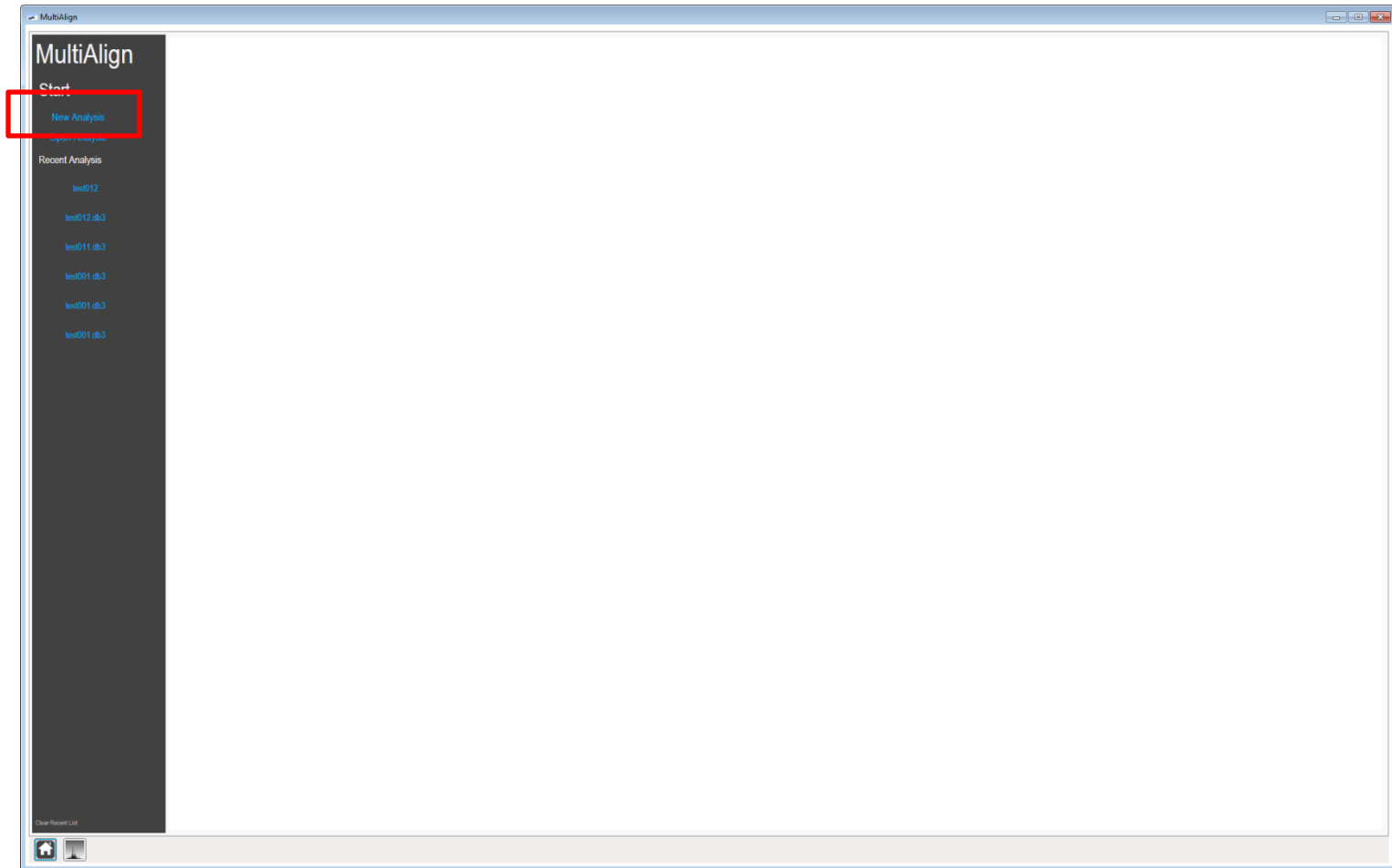
MultiAlign GUI Improvements

- Wizard Updates
 - Analysis wizard updated
 - AMT Tag Database Selection Window

MultiAlign GUI Improvements

- Filtering for searching
 - “Show me clusters with features across 4/5 datasets”
 - “Show me clusters that have at least 2 MS/MS spectra”
 - “Show me clusters within a given mono mass and NET range”
 - “Show me cluster Id’s 4, 515, 629,”
- Display
 - Display of parent spectra for a given feature
 - Display of all MS/MS spectra related to a cluster

Start An Analysis



Enter your folder path

MultiAlign

Local Files **DMS**

Add Folder Path

☐ Include sub-directories

Add Input File

Add Single File

Select:

0 / 0 Datasets

Dataset ID	Dataset Name	Instrument File	Peaks File	Sequence File	Parent Folder
------------	--------------	-----------------	------------	---------------	---------------

Press Add

MultiAlign

Local Files | DMS

Add Folder Path: M:\data\proteomics\TestData\QC-Shew-Annotated Browse Add

☐ Include sub-directories

Add Input File Browse Add

Add Single File Browse Add

Remove Selected Select: All None 0 / 2 Datasets

	Dataset ID	Dataset Name	Instrument File	Peaks File	Sequence File	Parent Folder
Modify	0	QC_Shew_13_04_1a_18Sep13_Cougar_13-06-14	Yes		Yes	M:\data\proteomics\TestData\QC-Shew-Annotated
Modify	1	QC_Shew_13_04_1a_6Oct13_Cougar_13-06-14	Yes		Yes	M:\data\proteomics\TestData\QC-Shew-Annotated

Cancel Back Next

Select Next

MultiAlign

Local Files | DMS

Add Folder Path: M:\data\proteomics\TestData\QC-Shew-Annotated Browse Add

☐ Include sub-directories

Add Input File Browse Add

Add Single File Browse Add

Remove Selected Select: All None 0 / 2 Datasets

	Dataset ID	Dataset Name	Instrument File	Peaks File	Sequence File	Parent Folder	
Modify	0	QC_Shew_13_04_1a_18Sep13_Cougar_13-06-14	Yes		Yes	M:\data\proteomics\TestData\QC-Shew-Annotated	
Modify	1	QC_Shew_13_04_1a_6Oct13_Cougar_13-06-14	Yes		Yes	M:\data\proteomics\TestData\QC-Shew-Annotated	

Cancel Back Next

Set your options (presets just modify mass resolution etc)

MultiAlign

Options

Experiment

Experiment Presets: Small Molecules (Lipids, Metabolites)

☐ Ion Mobility Experiment (IMS)?

Instrument

Instrument Presets: Small Molecules (Lipids, Metabolites)
Bottom Up Proteomics (Peptides)

Precursor Mass Resolution: ppm Low Mass (mono):

Fragmentation Window Size: Th. High Mass(mono):

LC-Retention Time: NET Ion Mobility: ms

Algorithms

Filtering

Advanced Load Existing Save

Cancel Back Next

Select Instrument Presets

MultiAlign

Options

Experiment

Experiment Presets: Bottom Up Proteomics (Peptides)

☐ Ion Mobility Experiment (IMS)?

Instrument

Instrument Presets: LTQ Orbitrap

Precursor Mass Resolution: LTQ Orbitrap

Fragmentation Window Size: TOF

LC-Retention Time: Velos

Algorithms

☐ Filtering

ppm Low Mass (mono)

Th. High Mass(mono)

NET Ion Mobility ms


Advanced Load Existing Save

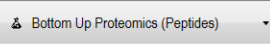
Cancel Back Next

Select Algorithms...if you need to, presets are best.


MultiAlign

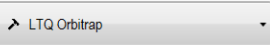
Options

 Experiment

Experiment Presets: 

☐ Ion Mobility Experiment (IMS)?


 Instrument

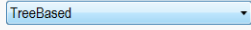
Instrument Presets: 

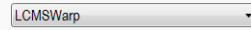
Precursor Mass Resolution: ppm Low Mass (mono):


Fragmentation Window Size: Th. High Mass (mono):

LC-Retention Time: NET Ion Mobility: ms

 Algorithms

Feature Definition: 

LC-MS Alignment: 

Across Dataset Clustering: 

☐ Filtering

Advanced Load Existing Save

Cancel Back Next

Feeling advanced? All options are still available here...

MultiAlign

Options

Experiment

Experiment Presets: **Bottom Up Proteomics (Peptides)**

☐ Ion Mobility Experiment (IMS)?

Instrument

Instrument Presets: **LITQ Orbitrap**

Precursor Mass Resolution: ppm Low Mass (mono):

Fragmentation Window Size: Th. High Mass (mono):

LC-Retention Time: NET Ion Mobility: ms

Algorithms

Feature Definition: **TreeBased**

LC-MS Alignment: **LCMSWarp**

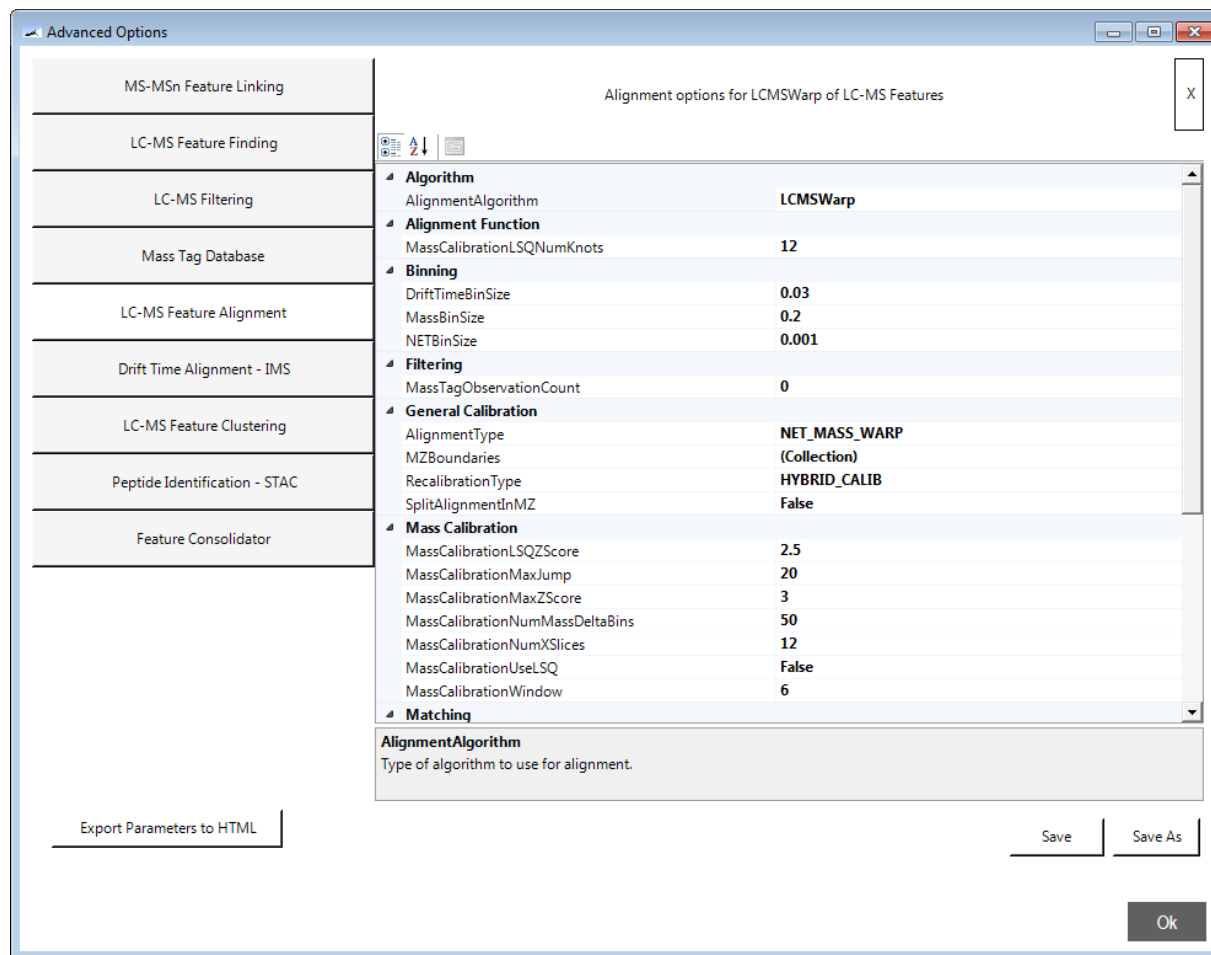
Across Dataset Clustering: **AverageLinkage**

☐ Filtering

Advanced Load Existing Save

Cancel Back Next

If you need to....



Next...

MultiAlign

Options

Experiment

Experiment Presets: **Bottom Up Proteomics (Peptides)**

☐ Ion Mobility Experiment (IMS)?

Instrument

Instrument Presets: **LTQ Orbitrap**

Precursor Mass Resolution: **8** ppm Low Mass (mono): **200**

Fragmentation Window Size: **0.5** Th. High Mass(mono): **10000**

LC-Retention Time: **0.03** NET Ion Mobility: **50** ms

Algorithms

Feature Definition: **TreeBased**

LC-MS Alignment: **LCMSWarp**

Across Dataset Clustering: **AverageLinkage**

☐ Filtering

Buttons: Advanced Load Existing Save Cancel Back Next

Select to align to a baseline dataset, or to a MTDB

The image shows a screenshot of the MultiAlign software interface. The window title is "MultiAlign". It features two main sections: "Alignment Type" and "Database for Alignment and/or Identification".

Alignment Type: This section contains two radio buttons. The first, "Align Dataset(s) to Dataset", is selected. To its right is a list box containing two entries: "QC_Shew_13_04_1a_18Sep13_Cougar_13-06-14" and "QC_Shew_13_04_1a_6Oct13_Cougar_13-06-14". The second radio button, "Align Dataset(s) to Database", is unselected.

Database for Alignment and/or Identification: This section contains three radio buttons: "Use Local Database" (selected), "Use Server Database", and "Don't use database". To the right of these buttons are two buttons: "Browse" and "Find".

Navigation: At the bottom of the window, there are three buttons: "Cancel", "Back", and "Next".

Align to an AMT Tag DB

The image shows a screenshot of the MultiAlign software window. The window has a title bar with the text 'MultiAlign' and standard window controls. The main content area is divided into two sections. The first section, titled 'Alignment Type', contains two radio buttons. The first radio button is labeled 'Align Dataset(s) to Dataset' and is followed by a large, empty text input field. The second radio button is labeled 'Align Dataset(s) to Database' and is selected. To the right of this radio button is the text 'You must select a database below.' The second section, titled 'Database for Alignment and/or Identification', contains three radio buttons: 'Use Local Database', 'Use Server Database' (which is selected), and 'Don't use database'. To the right of these radio buttons are two buttons: 'Browse' and 'Find'. The 'Find' button is highlighted with a red rectangular box. At the bottom of the window, there are three buttons: 'Cancel', 'Back', and 'Next'. The 'Back' and 'Next' buttons are blue, while the 'Cancel' button is grey. There are also two small icons in the bottom left corner.

MultiAlign

Alignment Type

☐ Align Dataset(s) to Dataset

☒ Align Dataset(s) to Database You must select a database below.

Database for Alignment and/or Identification

☐ Use Local Database

☒ Use Server Database

☐ Don't use database

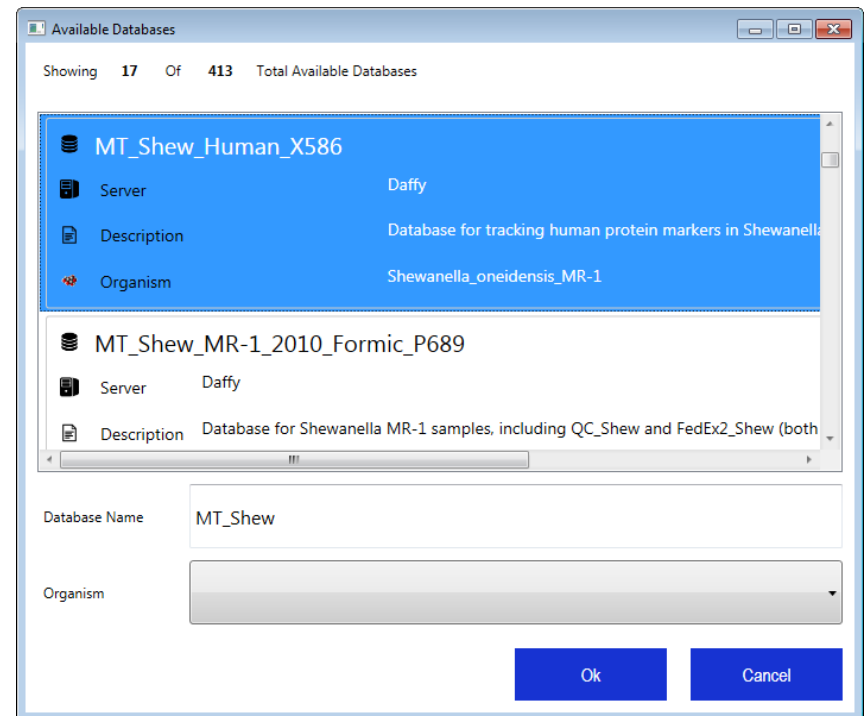
Browse

Find

Cancel Back Next

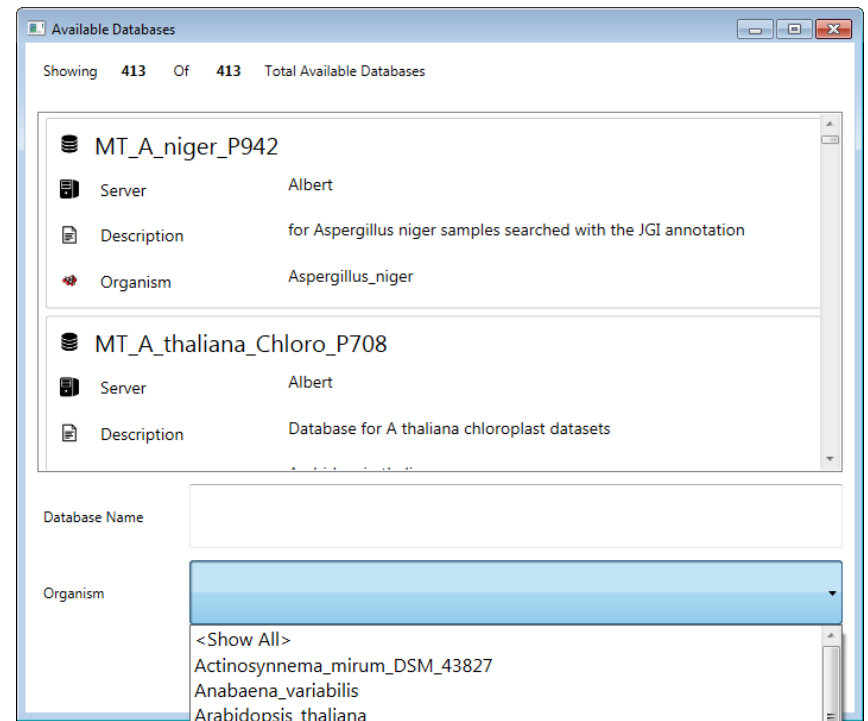
Select your database

- You can type part of the name...and it will filter for you



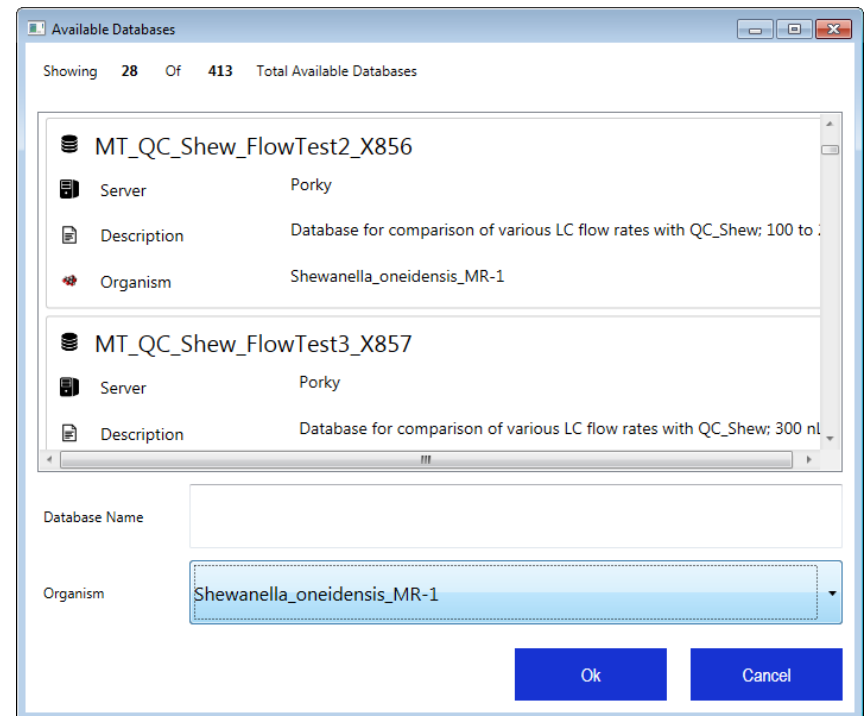
Select your database

- Or find all database for a given organism



Select your database

- Displays all databases for the selected organism



MultiAlign

Alignment Type

☒ Align Dataset(s) to Dataset QC_Shew_13_04_1a_6Oct13_Cougar_13-06-14

☐ Align Dataset(s) to Database

Database for Alignment and/or Identification

☐ Use Local Database

☒ Use Server Database MTShew_Human_X586

☐ Don't use database

Browse

Find

Cancel Back Next

Home

For now, just use a dataset, so press the “don’t use database” button

The screenshot shows the MultiAlign software window. The 'Alignment Type' section has two radio buttons: 'Align Dataset(s) to Dataset' (selected) and 'Align Dataset(s) to Database'. The 'Align Dataset(s) to Dataset' option is highlighted with a blue button. The 'Align Dataset(s) to Database' option is also highlighted with a blue button. The 'Database for Alignment and/or Identification' section has three radio buttons: 'Use Local Database', 'Use Server Database', and 'Don't use database'. The 'Don't use database' option is highlighted with a blue button and a red rectangular box. To the right of these radio buttons are two buttons: 'Browse' and 'Find'. At the bottom of the window are three buttons: 'Cancel', 'Back', and 'Next'. The 'Cancel' button is highlighted with a blue button. The 'Back' and 'Next' buttons are also highlighted with blue buttons. The window title bar shows 'MultiAlign' and standard window controls.

MultiAlign

Alignment Type

☒ Align Dataset(s) to Dataset QC_Shew_13_04_1a_6Oct13_Cougar_13-06-14

☐ Align Dataset(s) to Database

Database for Alignment and/or Identification

☐ Use Local Database

☐ Use Server Database

☒ Don't use database

Browse

Find

Cancel Back Next

Next

The image shows the MultiAlign software interface. The window title is "MultiAlign". The main content area is divided into two sections. The first section, "Alignment Type", has two radio buttons: "Align Dataset(s) to Dataset" (selected) and "Align Dataset(s) to Database". To the right of the selected radio button is a text box containing "QC_Shew_13_04_1a_6Oct13_Cougar_13-06-14". The second section, "Database for Alignment and/or Identification", has three radio buttons: "Use Local Database", "Use Server Database", and "Don't use database". To the right of these radio buttons are two buttons: "Browse" and "Find". At the bottom of the window, there are three buttons: "Cancel", "Back", and "Next". The "Next" button is highlighted with a red rectangle.

MultiAlign

Alignment Type

☒ Align Dataset(s) to Dataset QC_Shew_13_04_1a_6Oct13_Cougar_13-06-14

☐ Align Dataset(s) to Database

Database for Alignment and/or Identification

☐ Use Local Database

☐ Use Server Database

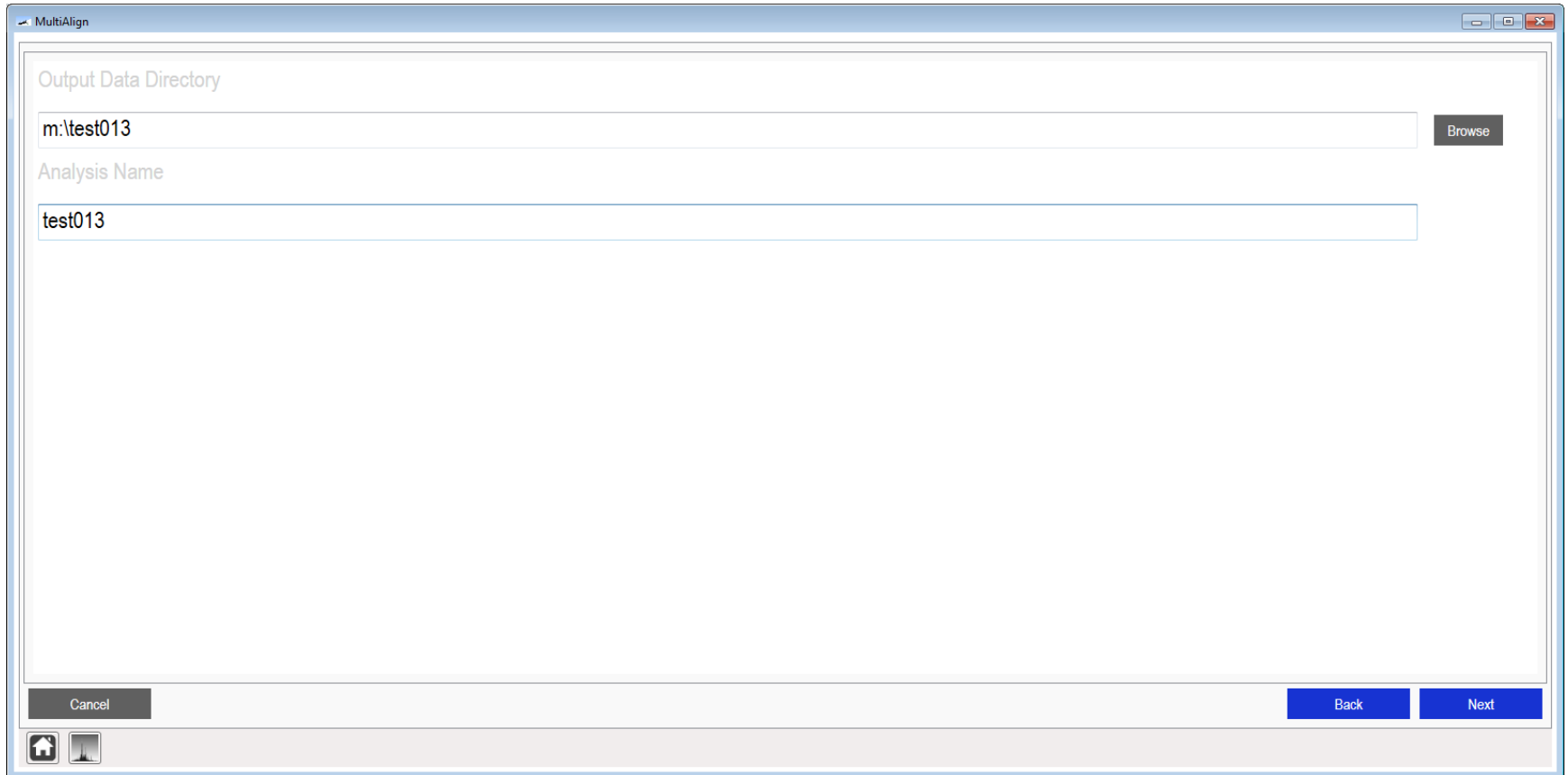
☐ Don't use database

Browse

Find

Cancel Back Next

Enter a name, and output path



The image shows a screenshot of a software window titled "MultiAlign". The window has a standard Windows-style title bar with minimize, maximize, and close buttons. The main content area contains two input fields. The first field is labeled "Output Data Directory" and contains the text "m:\test013". To the right of this field is a "Browse" button. The second field is labeled "Analysis Name" and contains the text "test013". At the bottom of the window, there are three buttons: "Cancel" on the left, and "Back" and "Next" on the right. The "Back" and "Next" buttons are blue, while the "Cancel" button is grey. There are also two small icons in the bottom left corner of the window.

MultiAlign

Output Data Directory

m:\test013

Browse

Analysis Name

test013

Cancel

Back

Next

Next – starts the analysis

The image shows a screenshot of the MultiAlign software window. The window has a title bar with the text "MultiAlign" and standard window controls. The main area contains two input fields: "Output Data Directory" with the text "m:\test013" and a "Browse" button to its right, and "Analysis Name" with the text "test013". At the bottom of the window, there are three buttons: "Cancel", "Back", and "Next". The "Next" button is highlighted with a red rectangular box. The "Next" button is blue with white text, while the "Cancel" and "Back" buttons are grey with black text. There are also two small icons (a home icon and a folder icon) in the bottom left corner of the window.

MultiAlign

Output Data Directory

m:\test013

Browse

Analysis Name

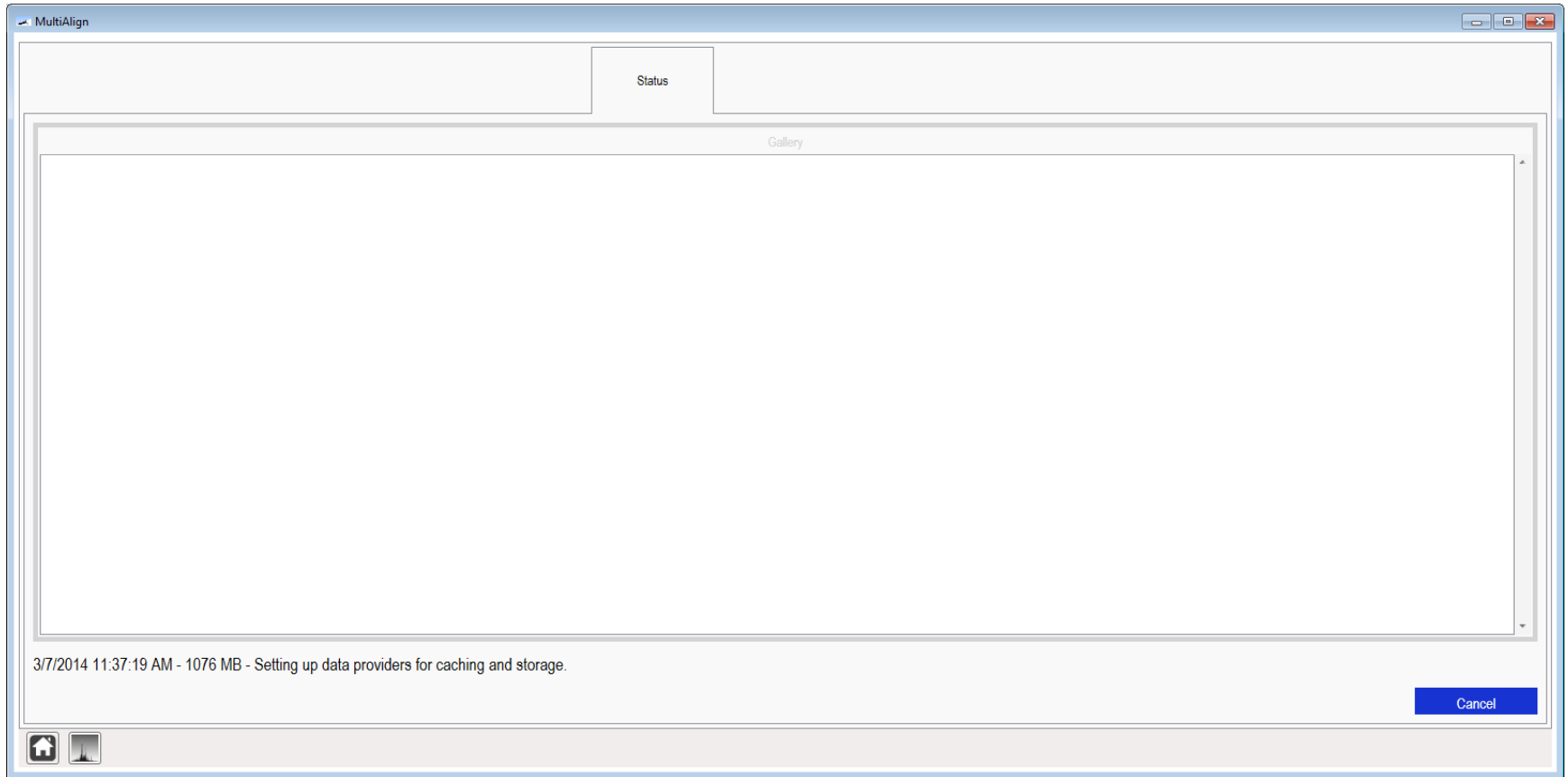
test013

Cancel

Back

Next

Status window will show progress of analysis...

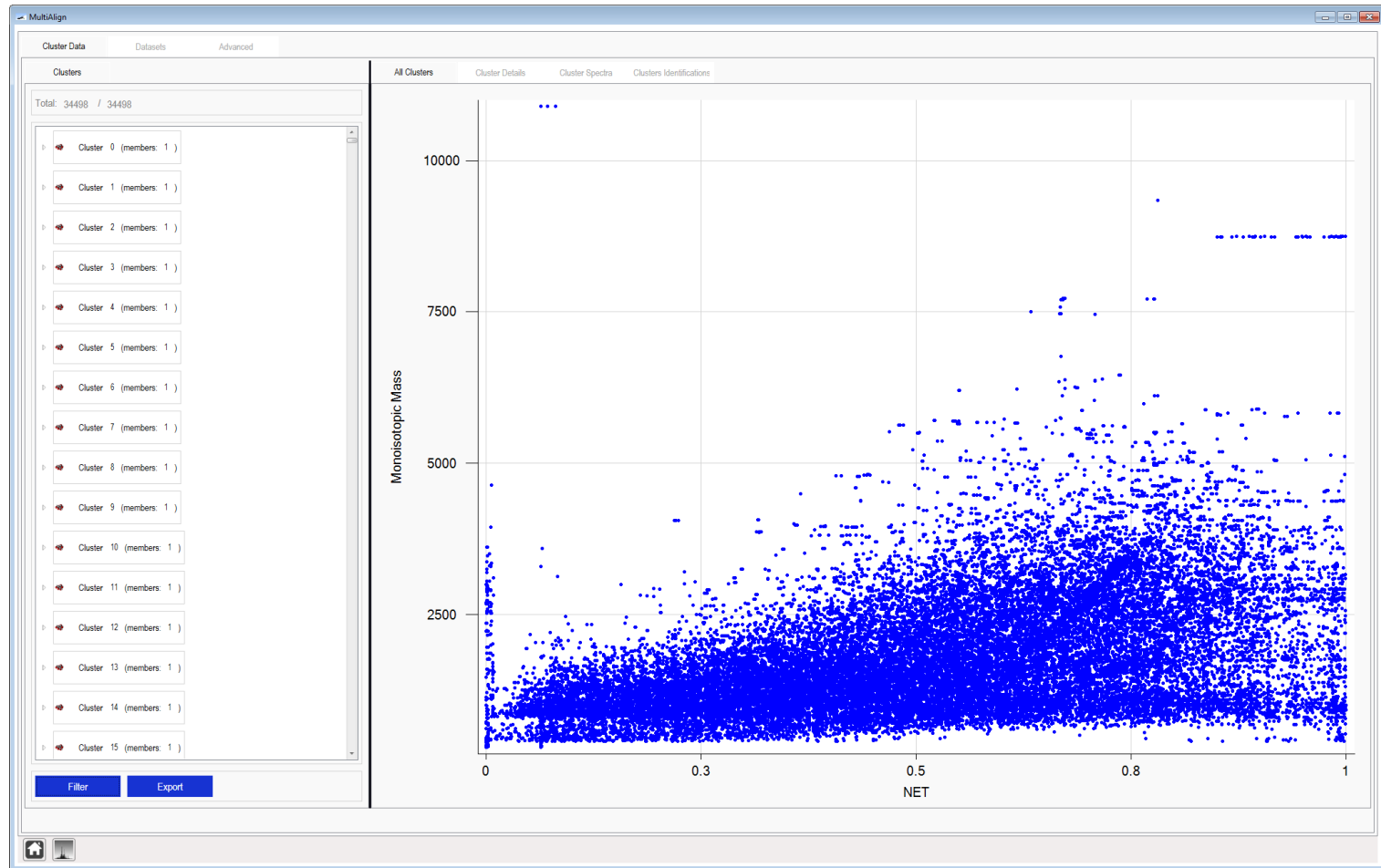


NOTE: this feature is being improved...if it looks like the software is not doing anything it is.

Analysis View

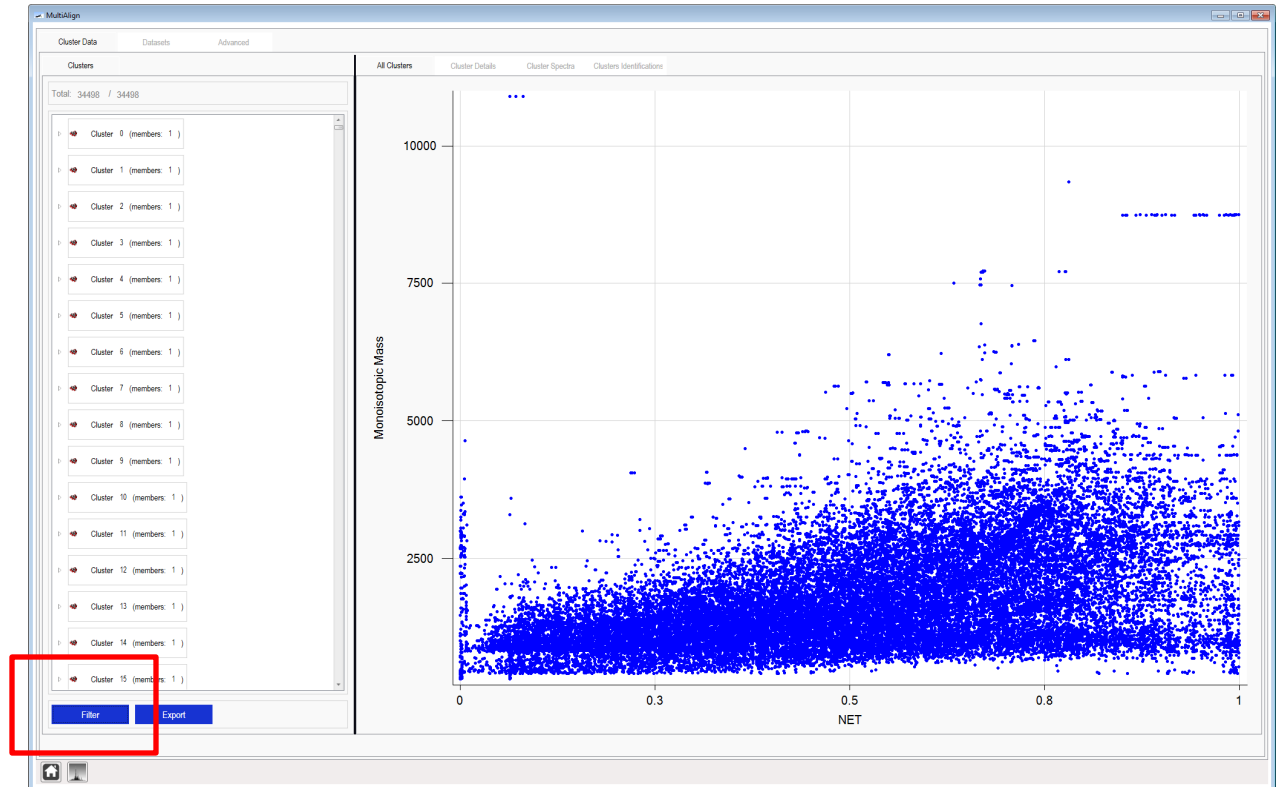
- To demonstrate the GUI
 - 2 Datasets of QC-Shew were aligned and clustered
 - MS/MS spectra were not linked to any MSGF+ results (i.e. the files were not loaded with MultiAlign)

New GUI – View of the analysis



New GUI – View of the analysis

- Click Filter



Filtering

- “Show me clusters that are present in 2 or more datasets”

The screenshot shows a 'Filter Clusters' dialog box with the following settings:

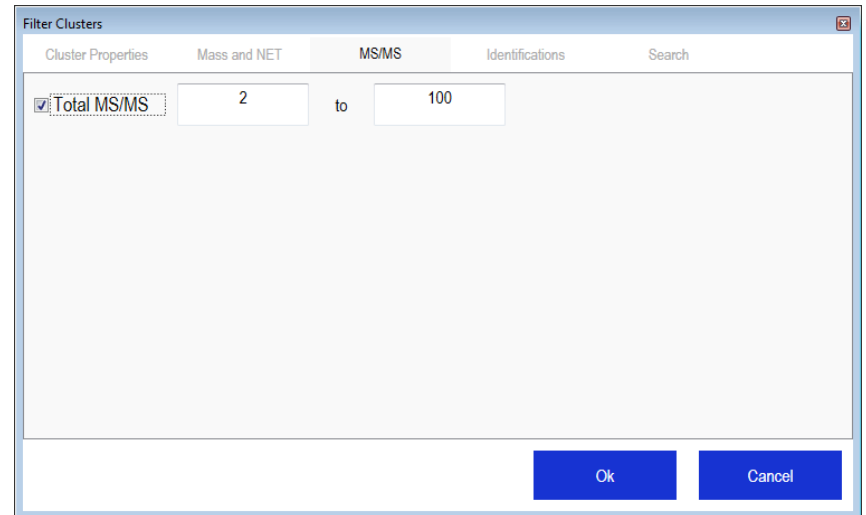
Cluster Properties	Mass and NET	MS/MS	Identifications	Search
<input type="checkbox"/> Total Members	0	to	10000	
<input checked="" type="checkbox"/> Dataset Membe	2	to	10000	
<input type="checkbox"/> Tightness	0	to	10000	
<input type="checkbox"/> Ambiguity	0	to	10000	

Buttons: Ok, Cancel

Note: filtering is an intersection “must have this AND this AND this AND this”

Filtering

- “Show me clusters that have at least 2 MS/MS spectra”

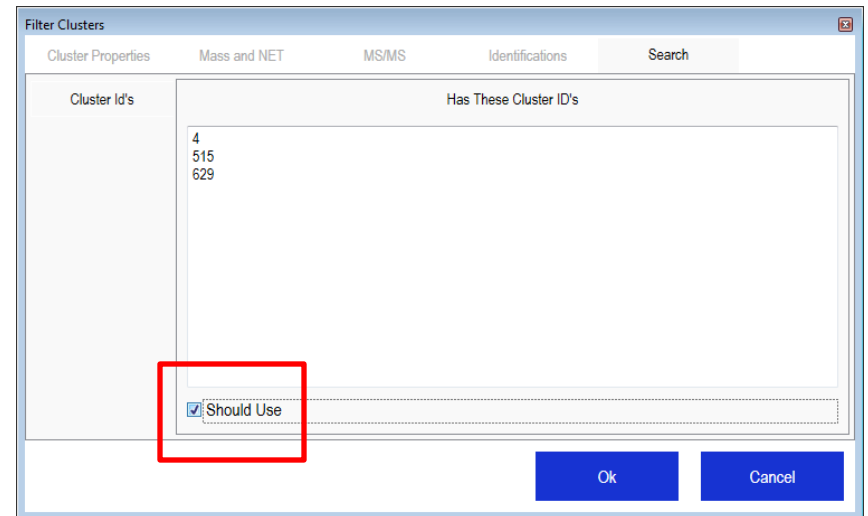


The screenshot shows a software window titled "Filter Clusters". It has five tabs: "Cluster Properties", "Mass and NET", "MS/MS", "Identifications", and "Search". The "MS/MS" tab is currently selected. Inside this tab, there is a checkbox labeled "Total MS/MS" which is checked. To the right of the checkbox is a range selector with two input boxes. The first box contains the number "2" and the second box contains the number "100", with the word "to" positioned between them. At the bottom right of the window, there are two blue buttons labeled "Ok" and "Cancel".

Note: filtering is an intersection “must have this AND this AND this AND this”

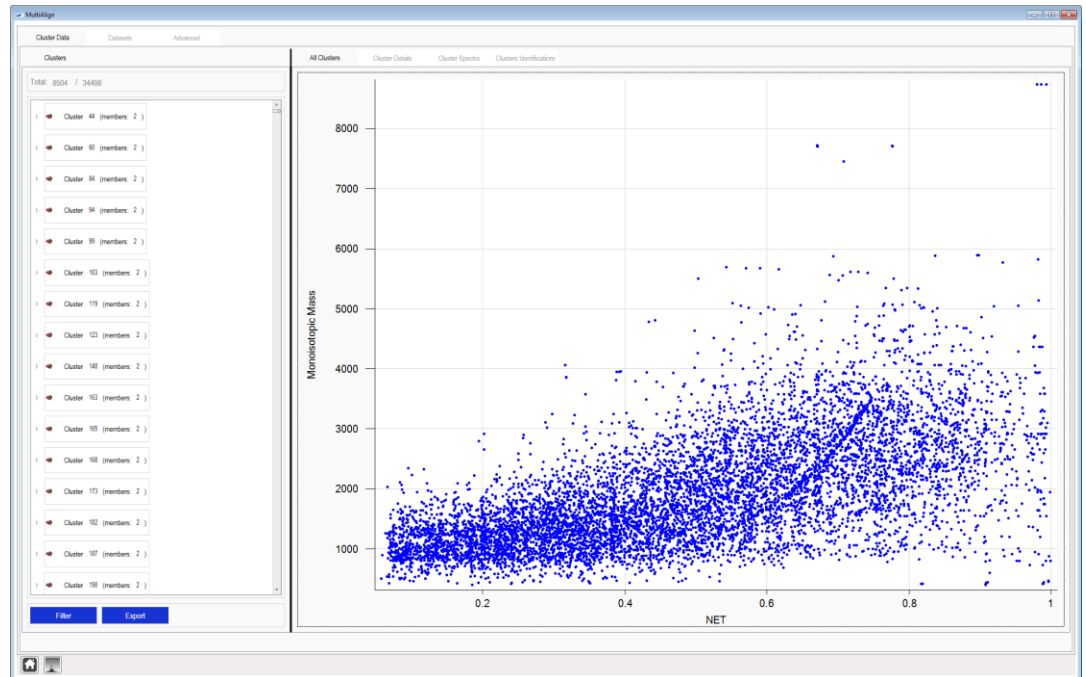
Filtering

- “Show me clusters 4, 515, 629”
- Make sure you click “should use”



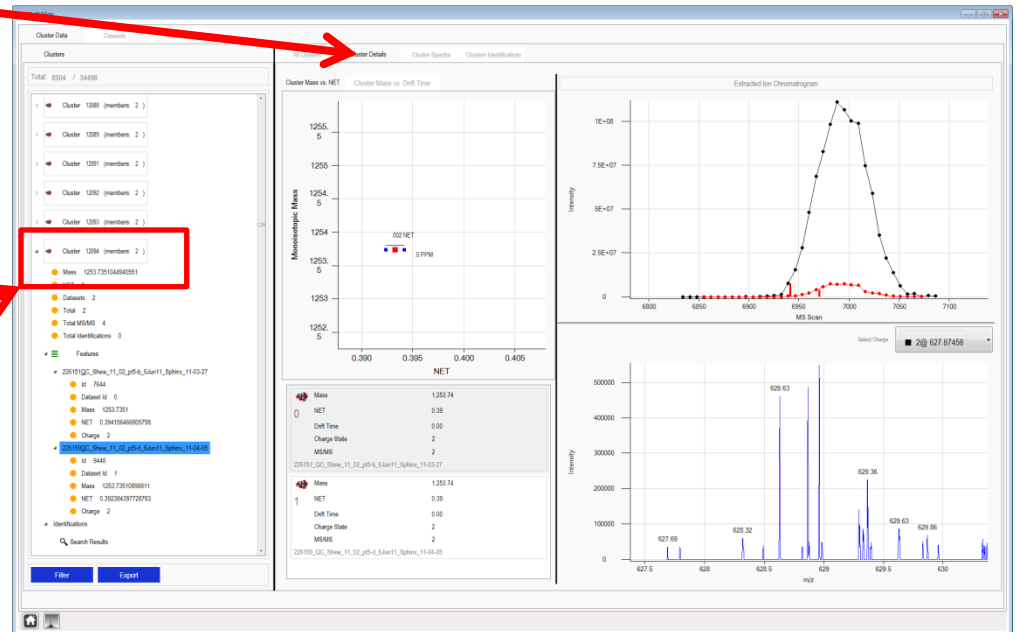
Note: filtering is an intersection “must have this AND this AND this AND this”

- Post – Filtering only shows clusters that met that criteria
 - *NOTE THIS EXAMPLE DID NOT USE THE CLUSTER ID FILTER*



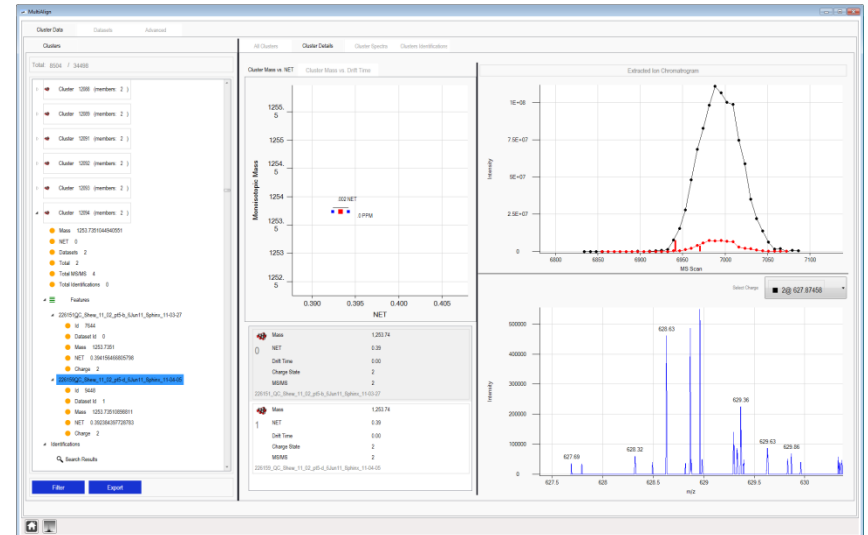
View A Cluster

- Select “Cluster Details”
- Select a cluster
 - Expand the tree view with the arrow



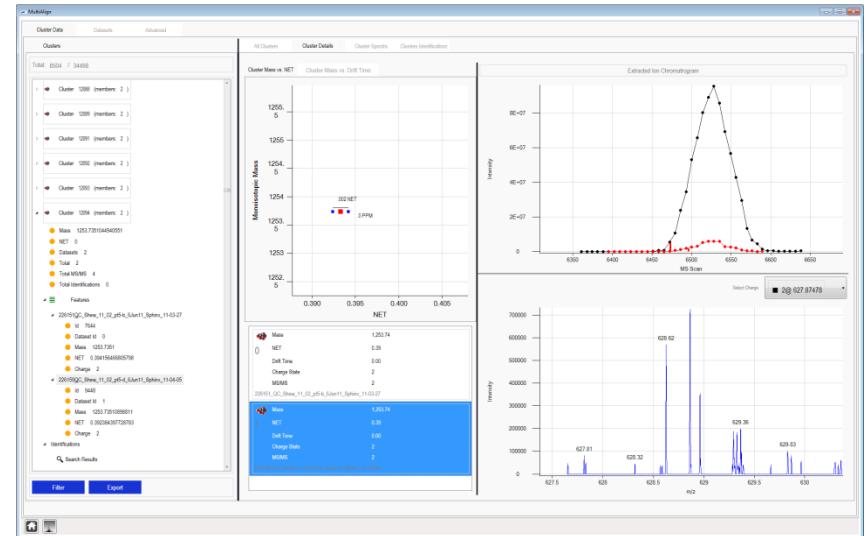
Cluster View

- Dataset 1
- Features shown upper right
 - Both charges states +1 and +2
- Parent Spectra shown lower right
- Cluster Details shown in expandable tree view (left)



Same Cluster – different feature

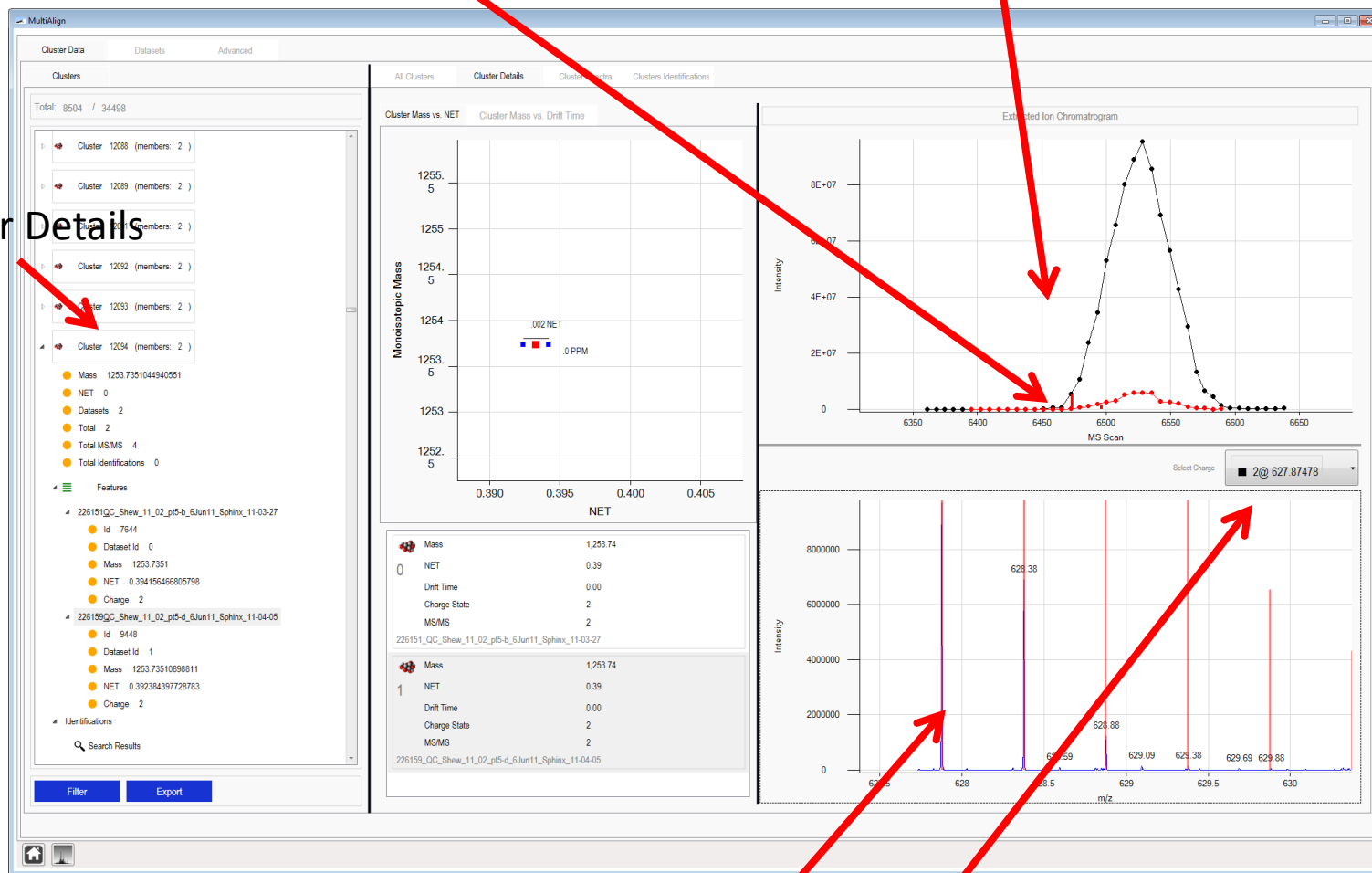
- Same Information displayed as with other dataset



MS/MS Fragmentation Event

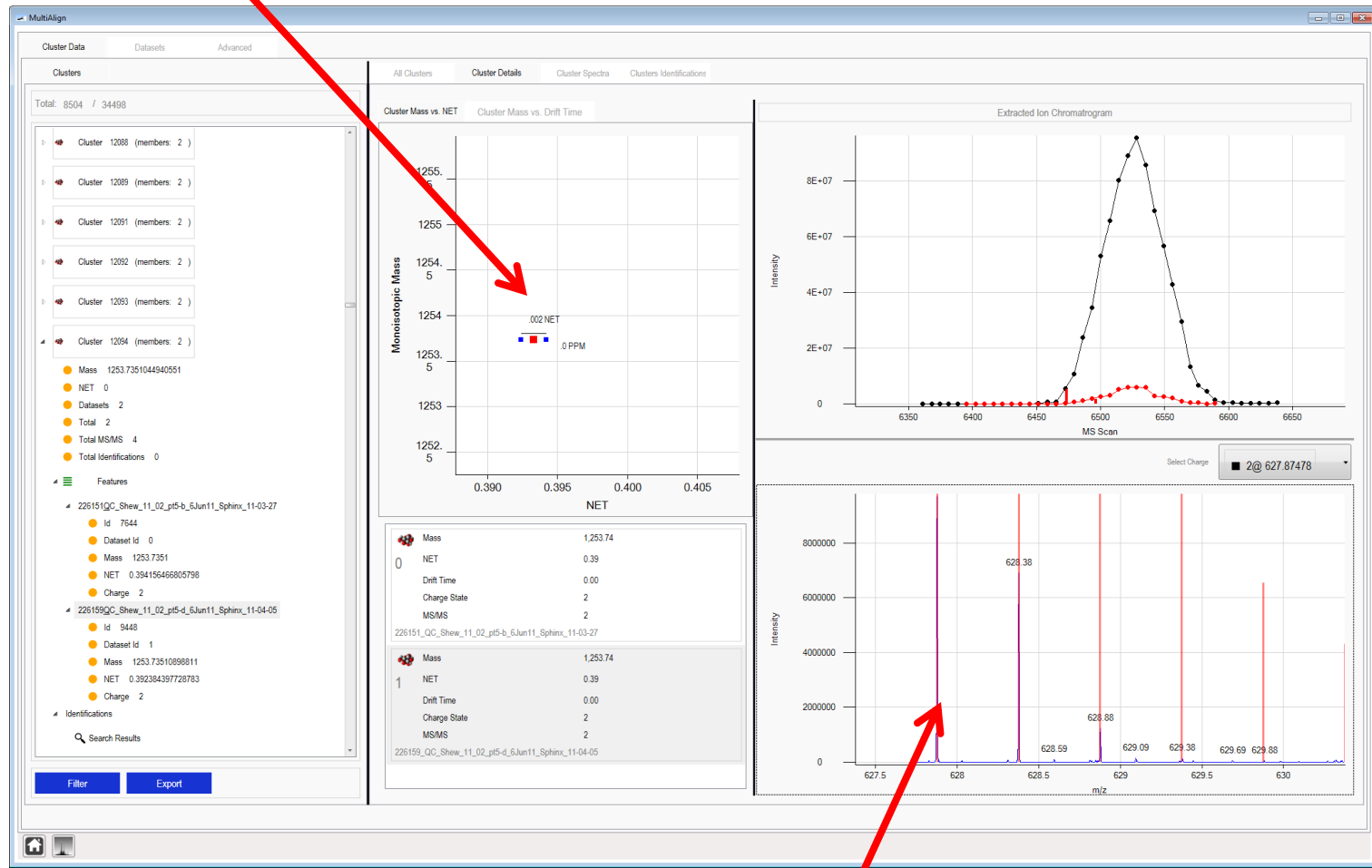
XIC for selected feature

Cluster Details



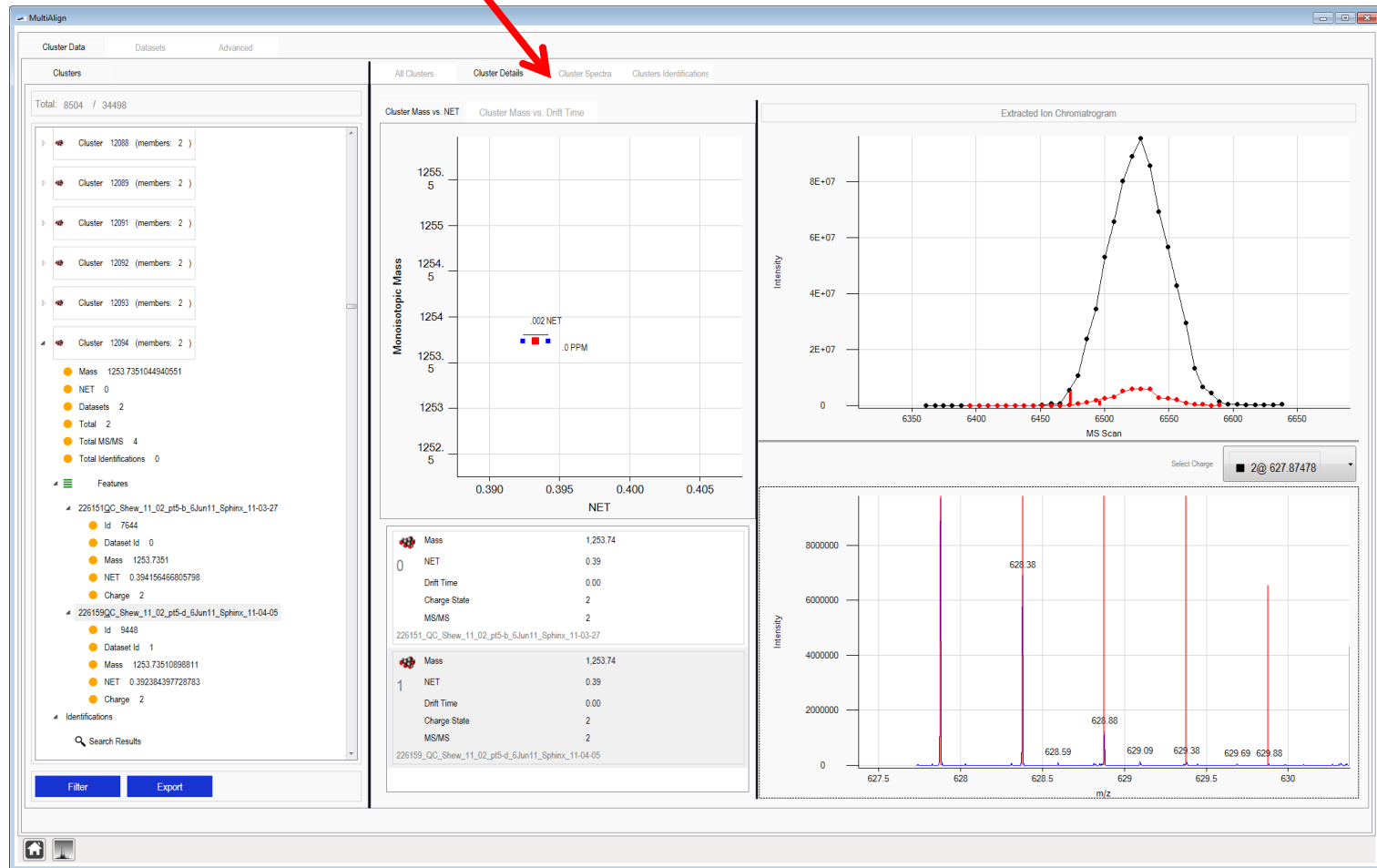
Parent Spectra for selected charge state

Cluster scatter plot



Isotopic distribution (first peak is monoisotopic)

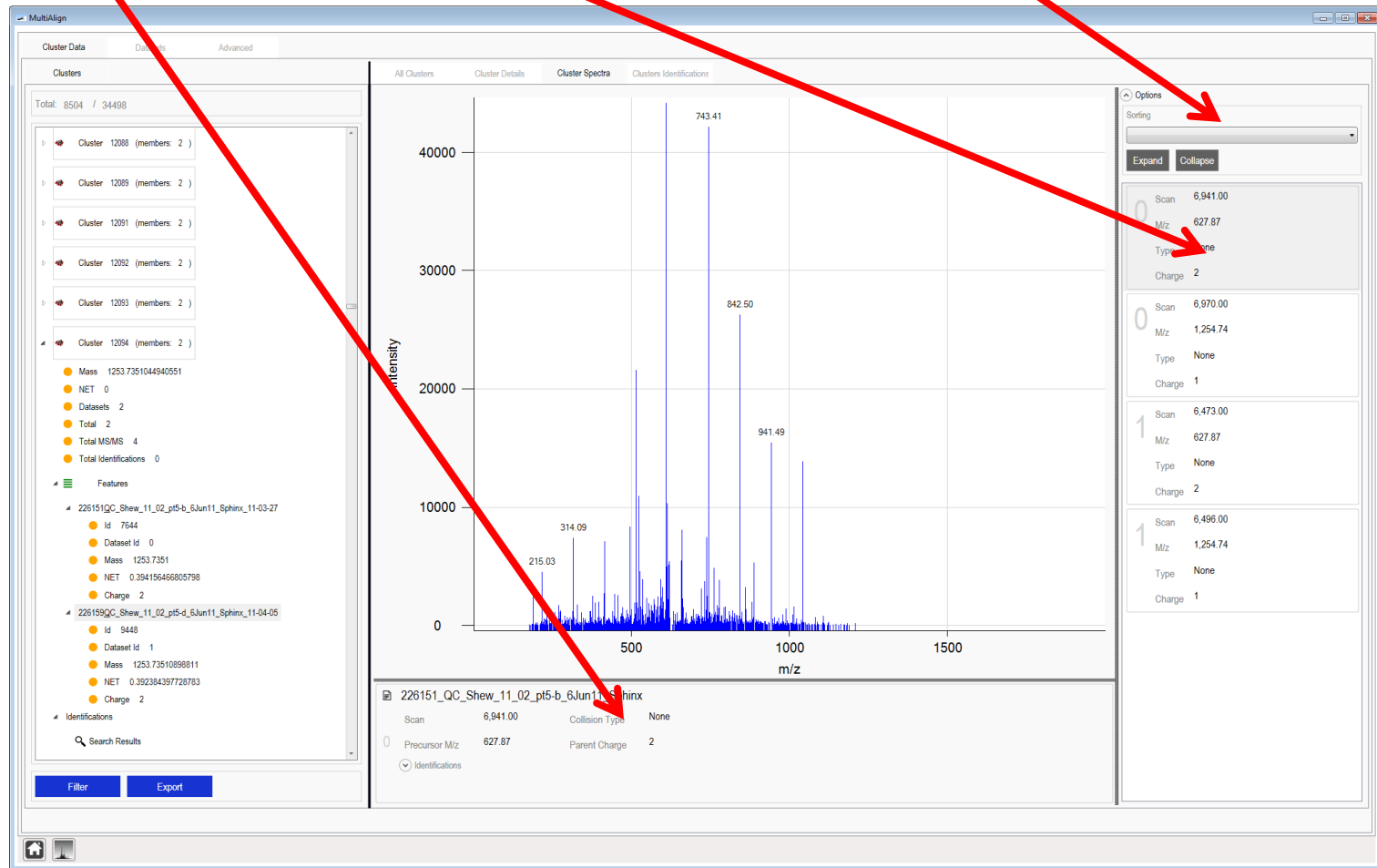
To see MS/MS Spectra for this cluster click here



List of all spectra

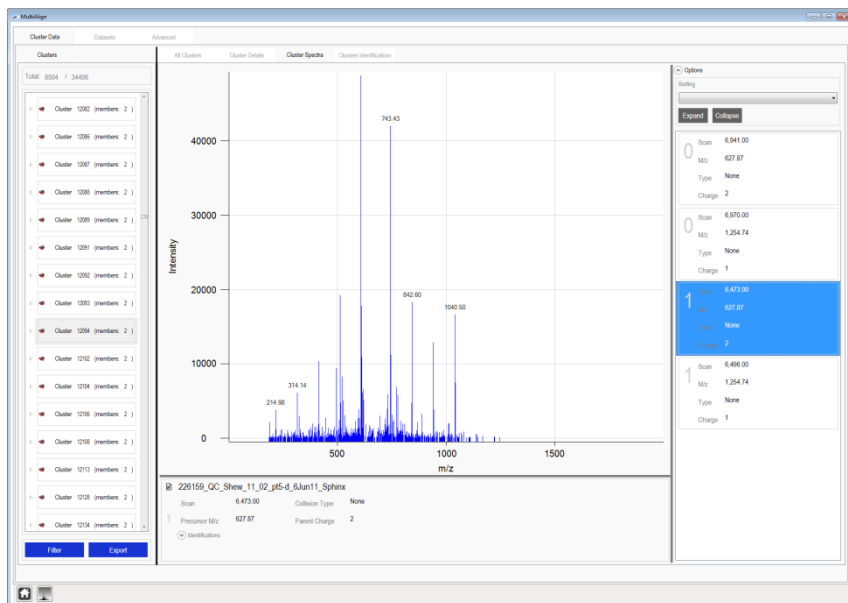
Sorting options

Detailed spectra view

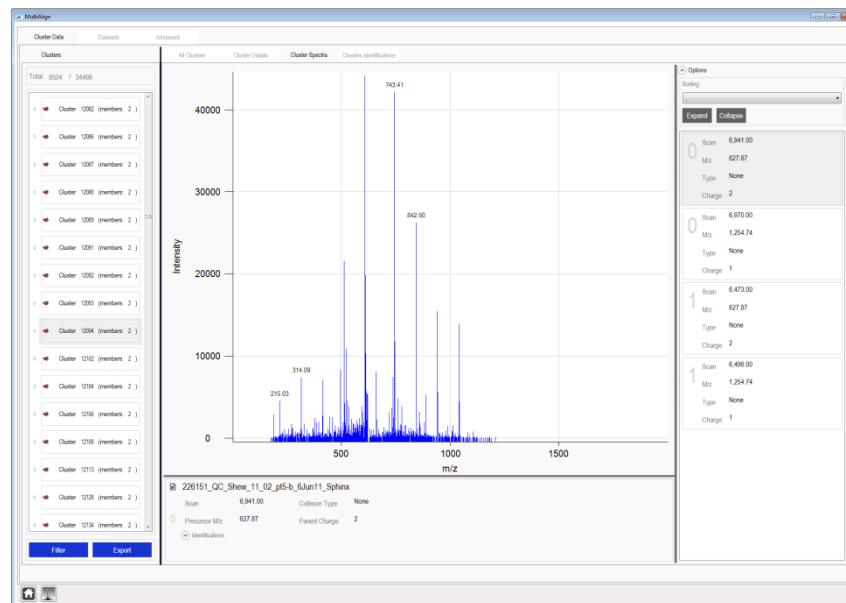


MS / MS Comparisons +2

Dataset 1 – Charge +2

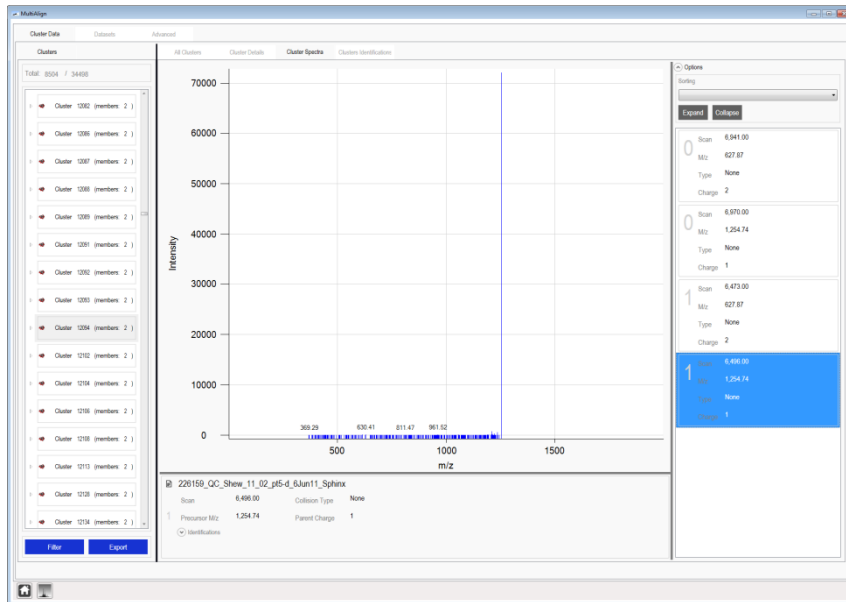


Dataset 0 – Charge +2

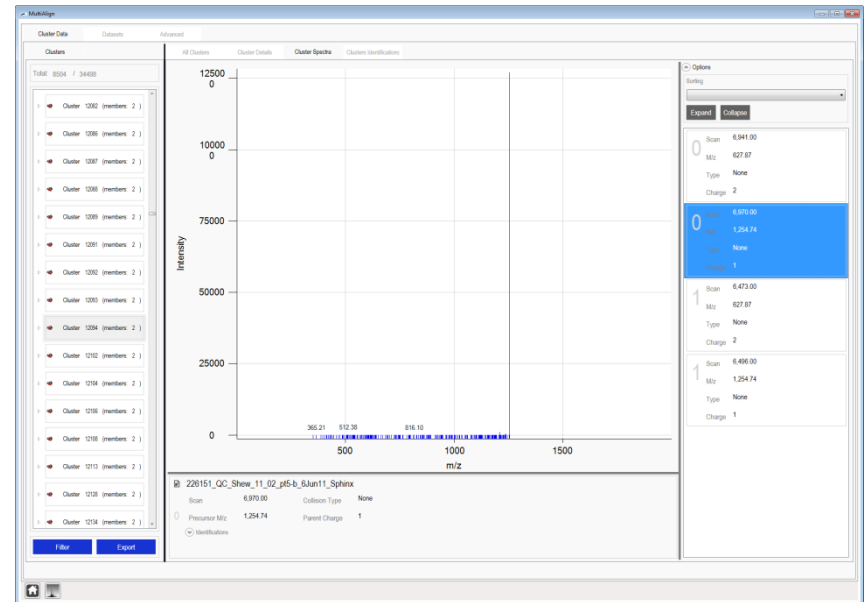


MS/MS Comparisons +1

Dataset 1 – Charge +1



Dataset 0 – Charge +1



Next Improvements

- GUI
 - Better display of Parent MS
 - Better display of MS/MS spectra (for a cluster)
 - Global statistics
 - Alignment Error Histograms (currently just images)
- Result Database Improvements
 - Improve XIC storage in database
- Algorithms
 - Labeled Data
 - O^{16}/O^{18}
 - Cluster Score – based on cluster MS/MS similarity
 - Feature Score & Filtering
 - Based on XIC's
 - Integration of chromatogram scores (XIC correlation)
 - Integration of alignment using MS/MS anchor points (details to come in next group meeting)

Other Improvements

- Integrate with MTDB
 - A recent version of MTDB Creator has been completed
 - Integrating this code into MultiAlign to read AMT tag databases created from SEQUEST/MSGF+ results so external collaborators can build MTDB's and run AMT style workflows
- MultiAlign **will** process IMS data
 - But the GUI traceback (to raw data XIC's or spectra display) is not available with IMS information (i.e. we only show features)
 - Future implementation to do create those features on request

Last Notes

- Speed
 - Since we are going back to the raw data to do a better job of feature finding, running an analysis will take a little more time (~3-5 minutes per dataset)
- You need to include the RAW file with the analysis to do better traceback