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

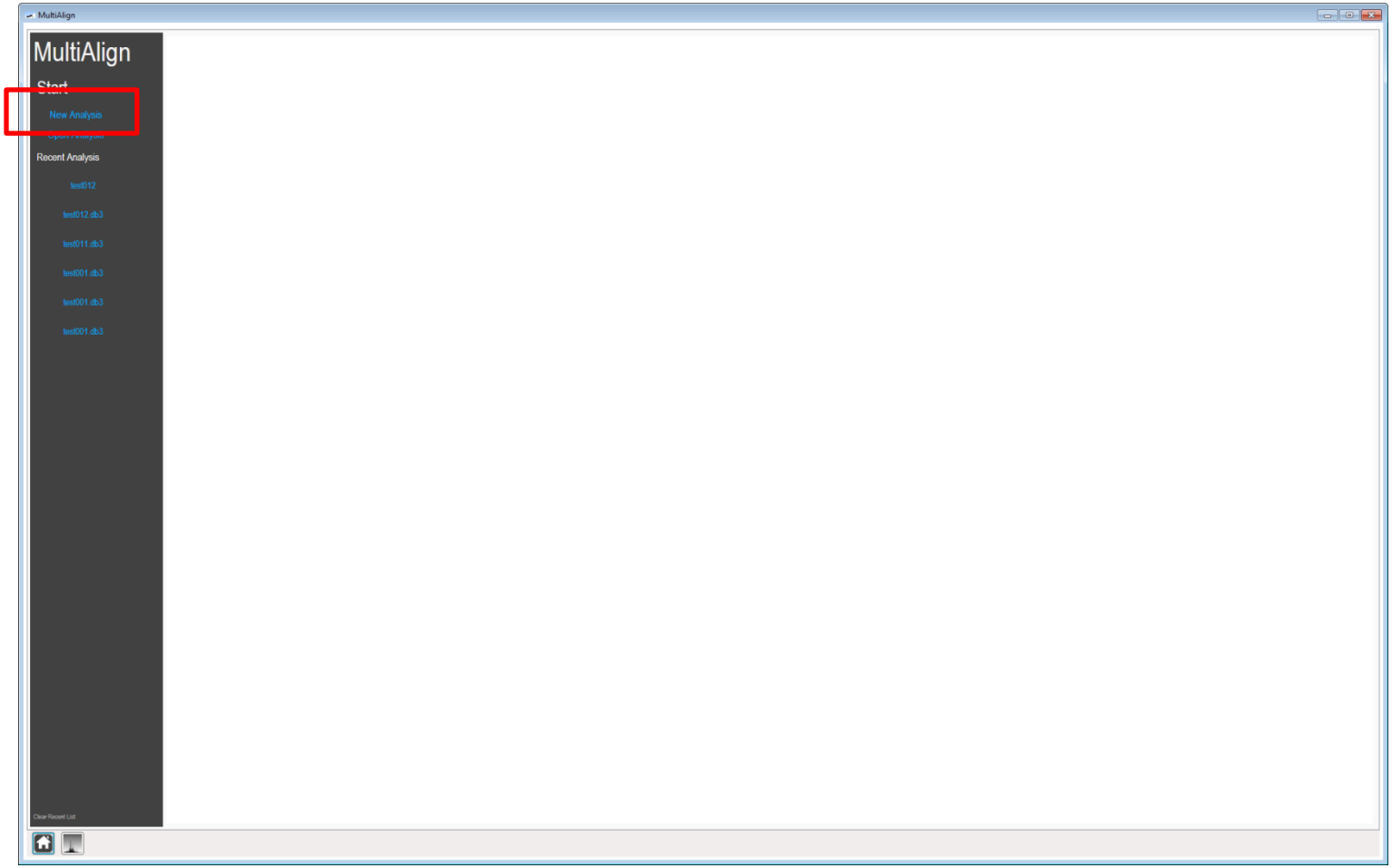
MultiAlign GUI Improvements

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

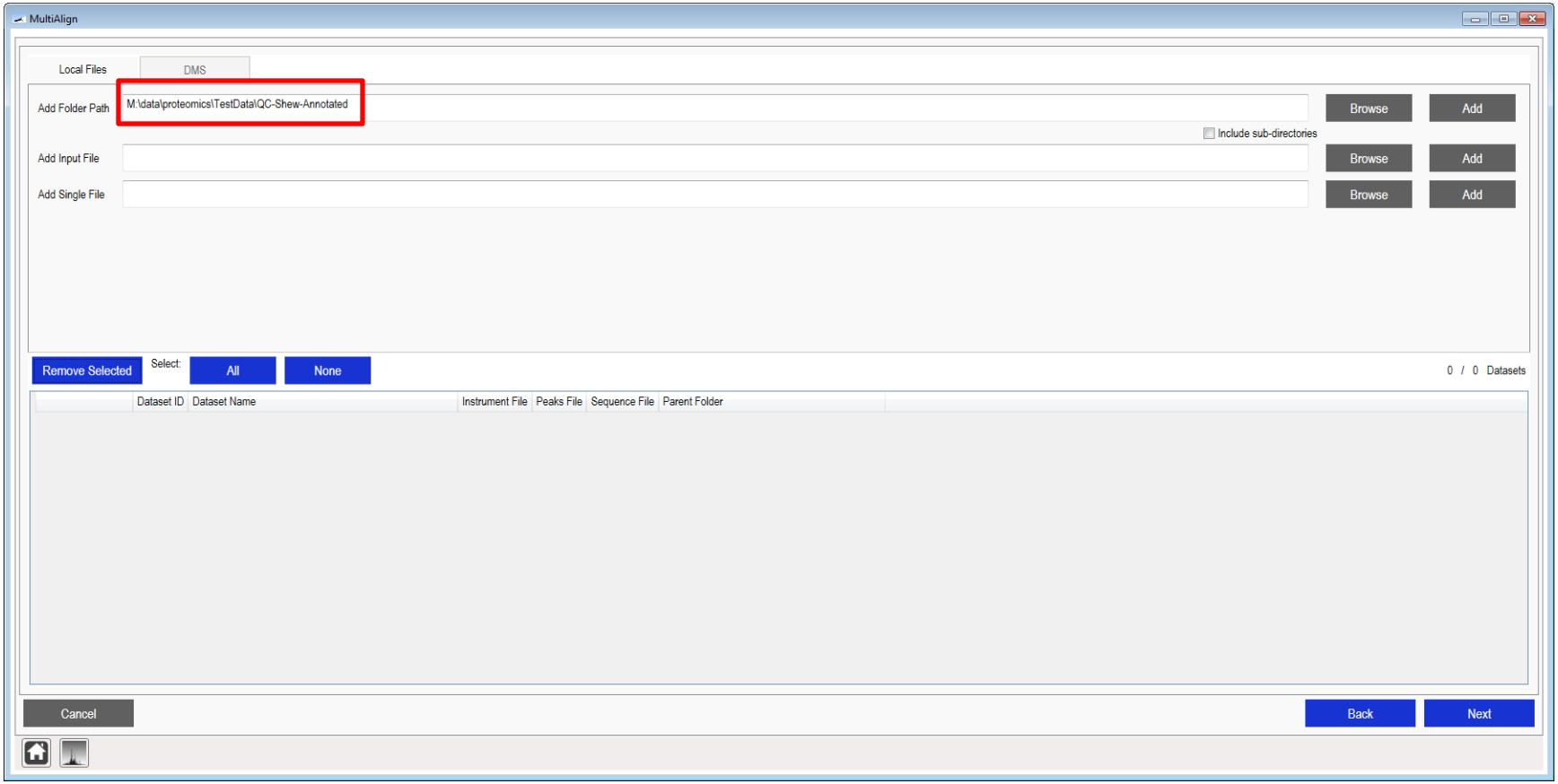
MultiAlign GUI Improvements

- 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

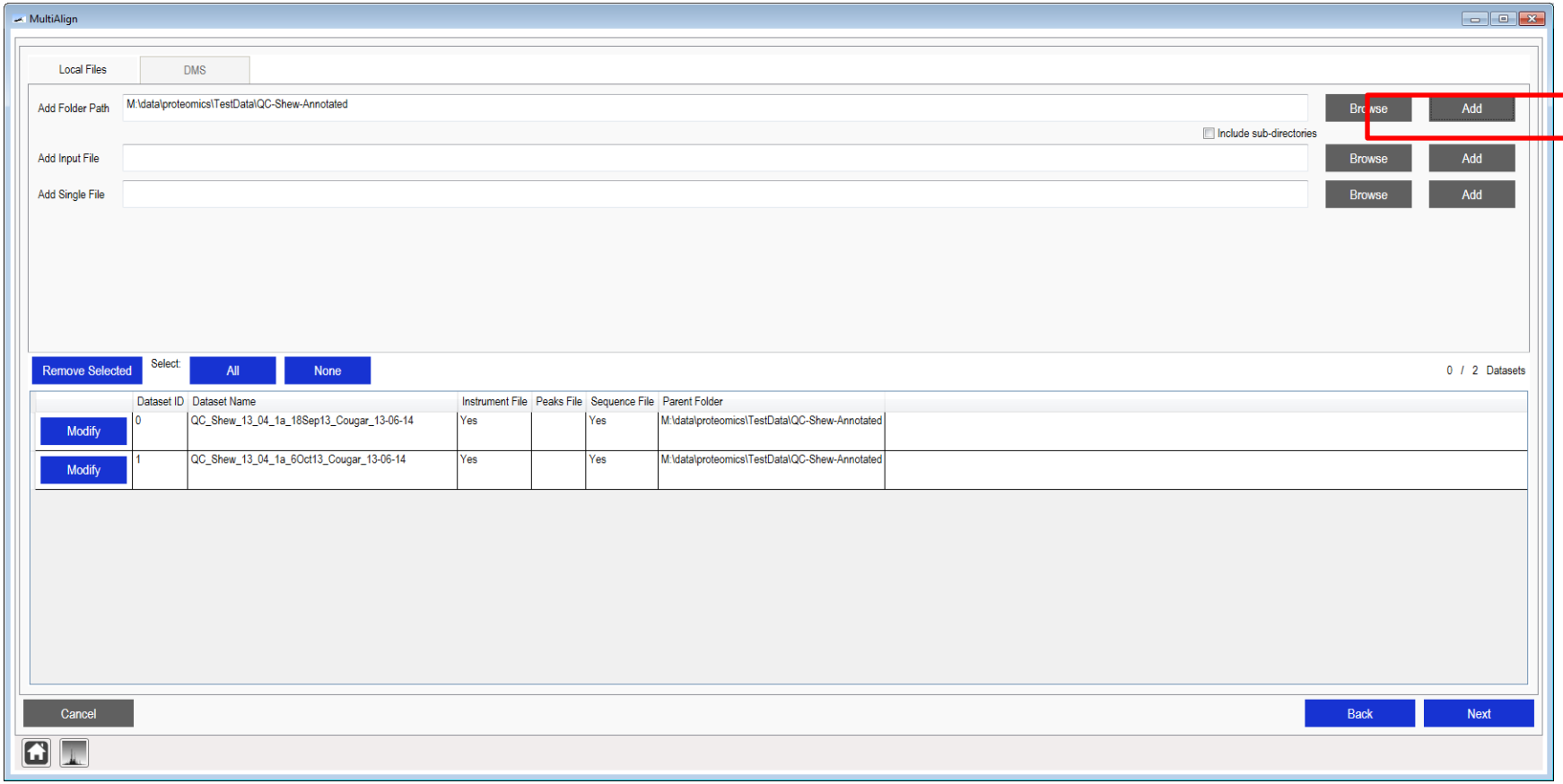
Start An Analysis



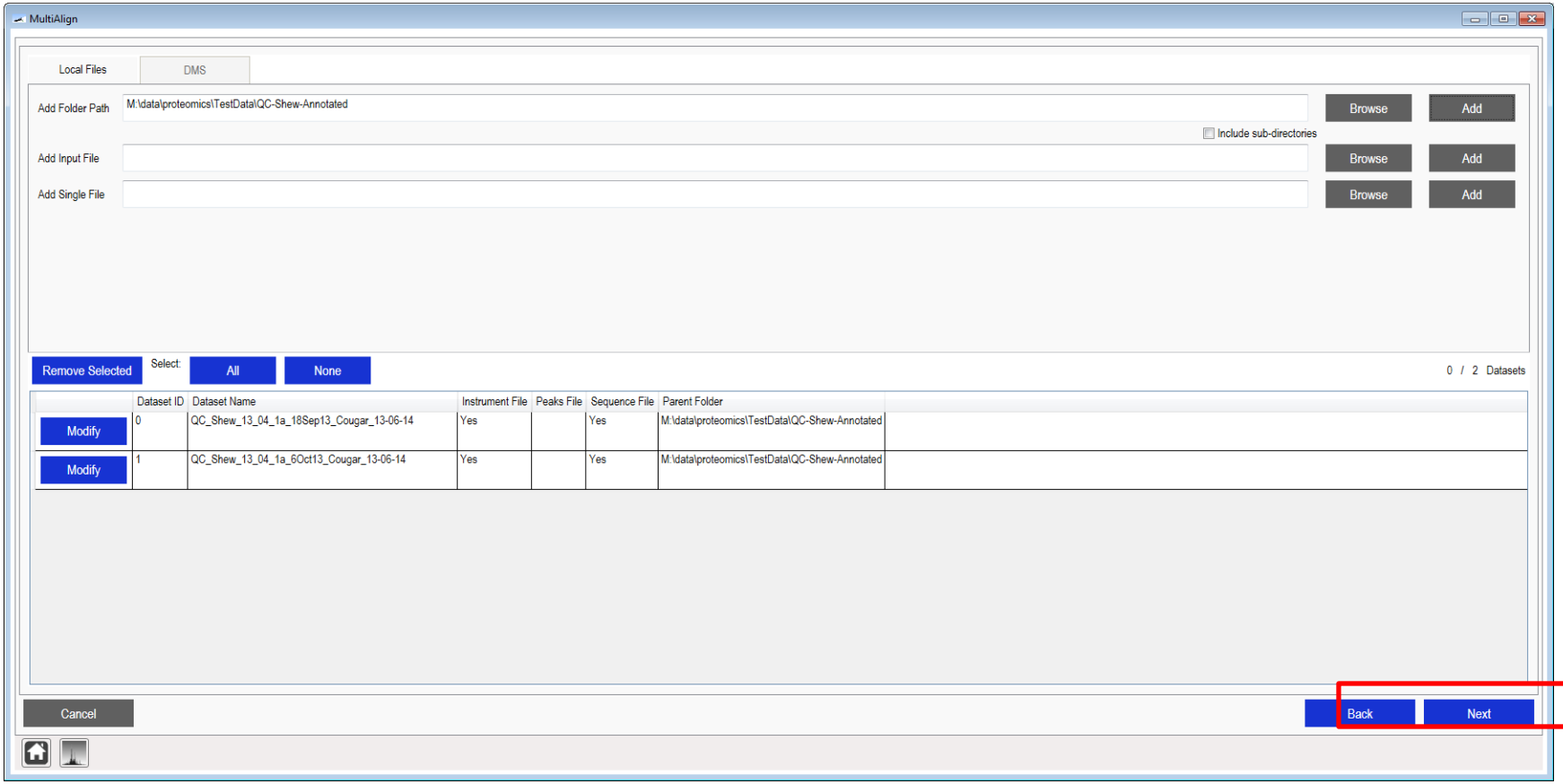
Enter your folder path



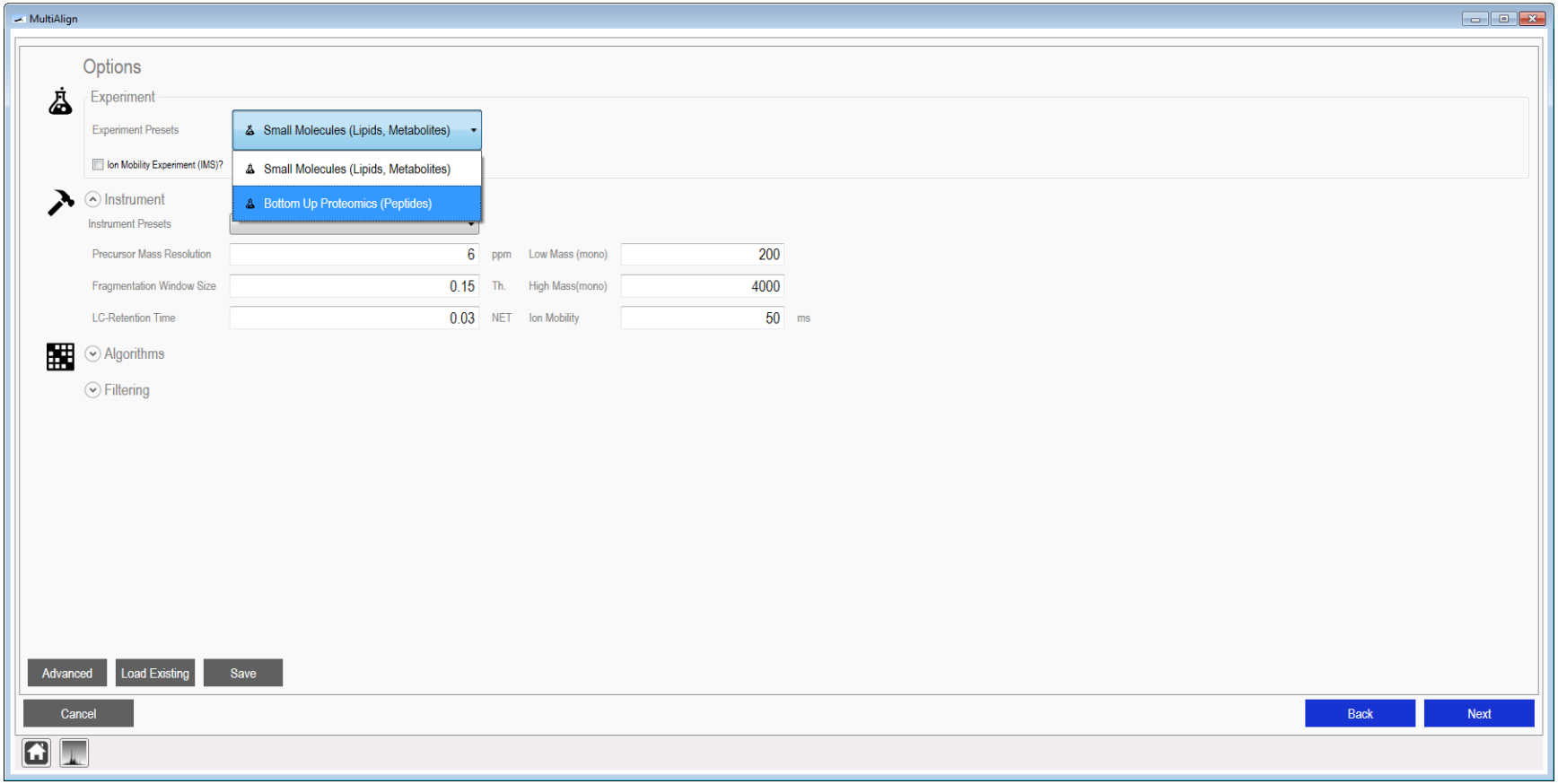
Press Add


MultiAlign
Local Files
DMS
Add Folder Path
M:\data\proteomics\TestData\QC-Shew-Annotated
Include sub-directories
Browse
Add
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

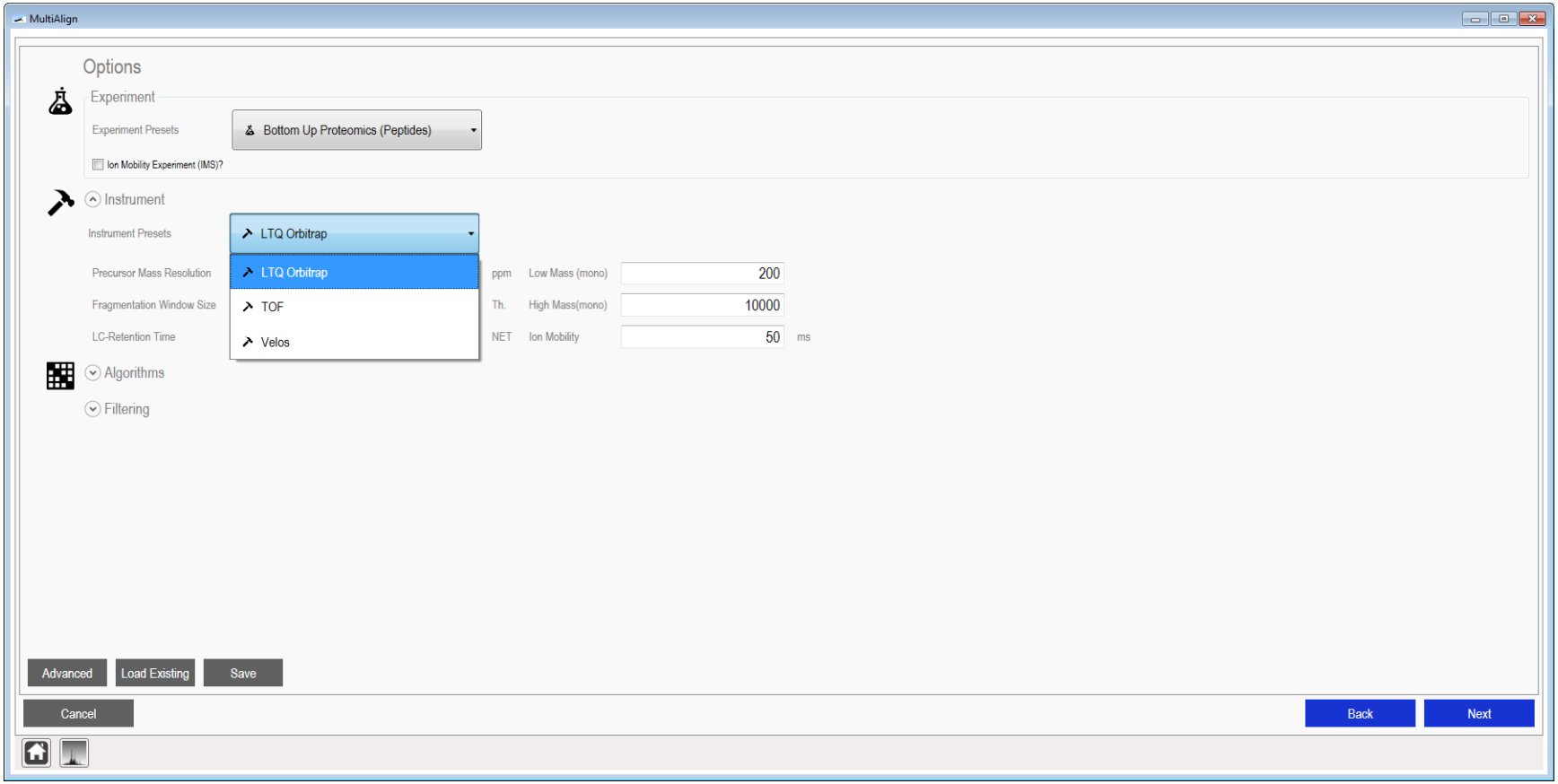
Select Next


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

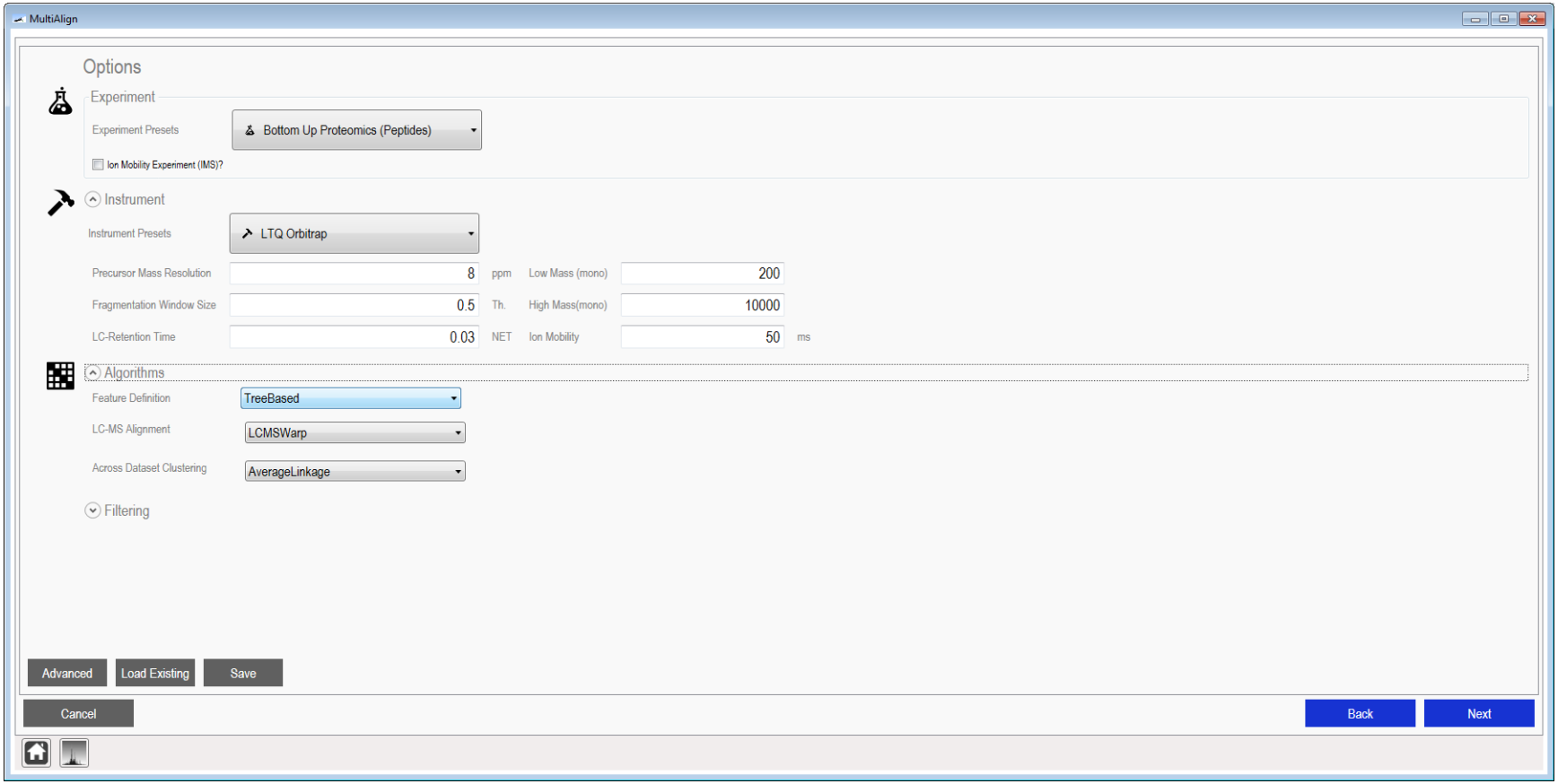
Set your options (presets just modify mass resolution etc)



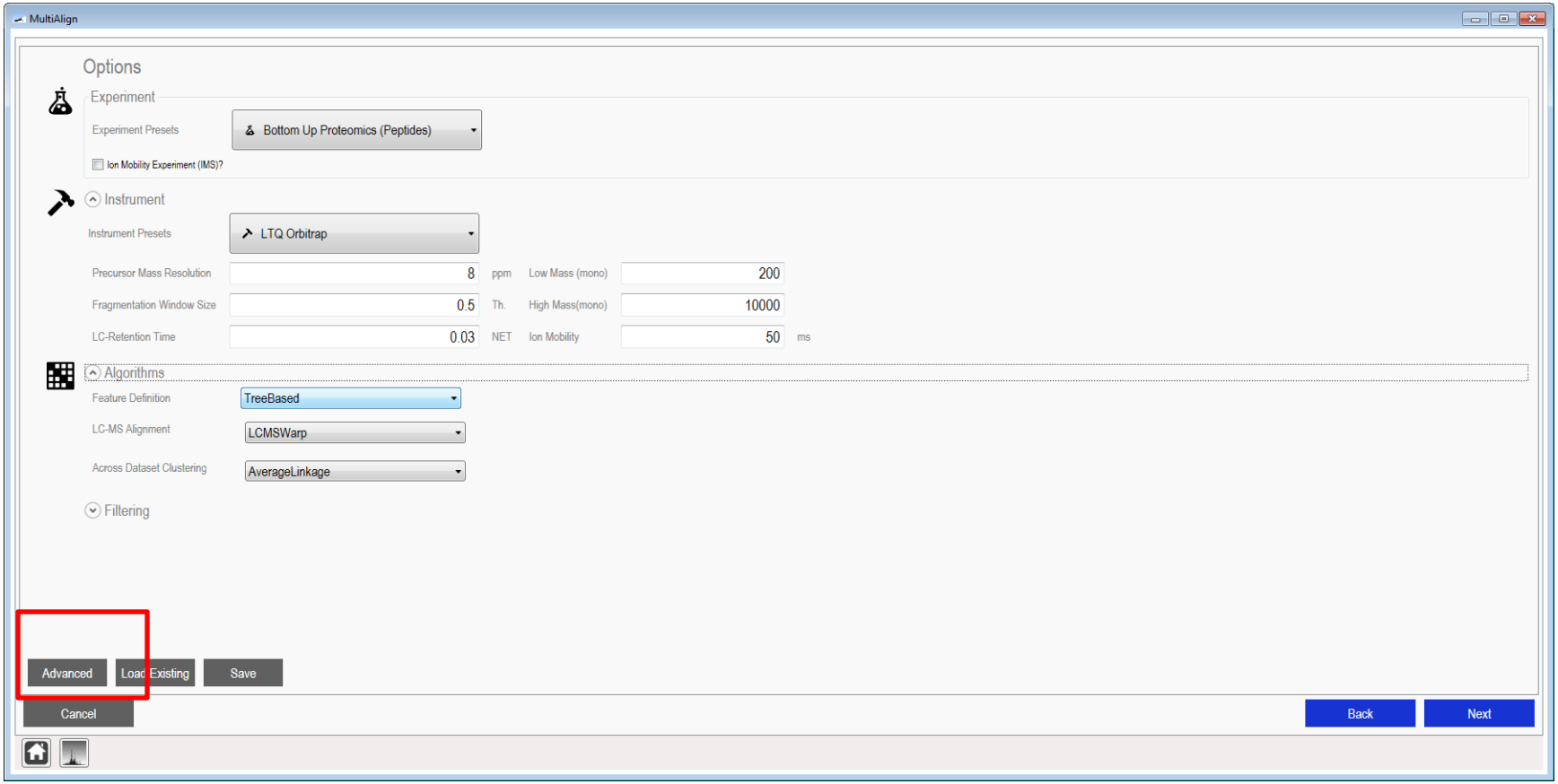
Select Instrument Presets



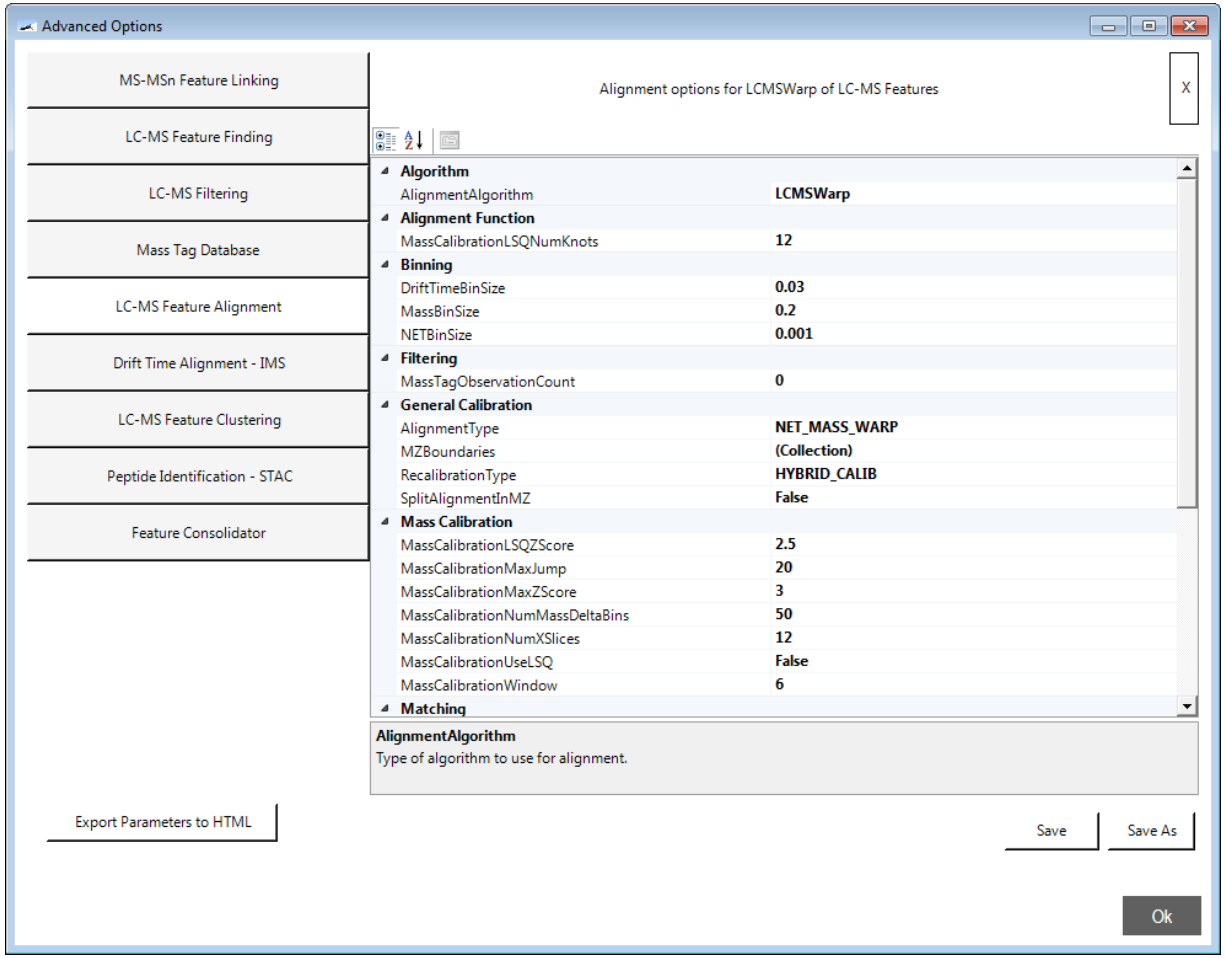
Select Algorithms…if you need to, presets are best.



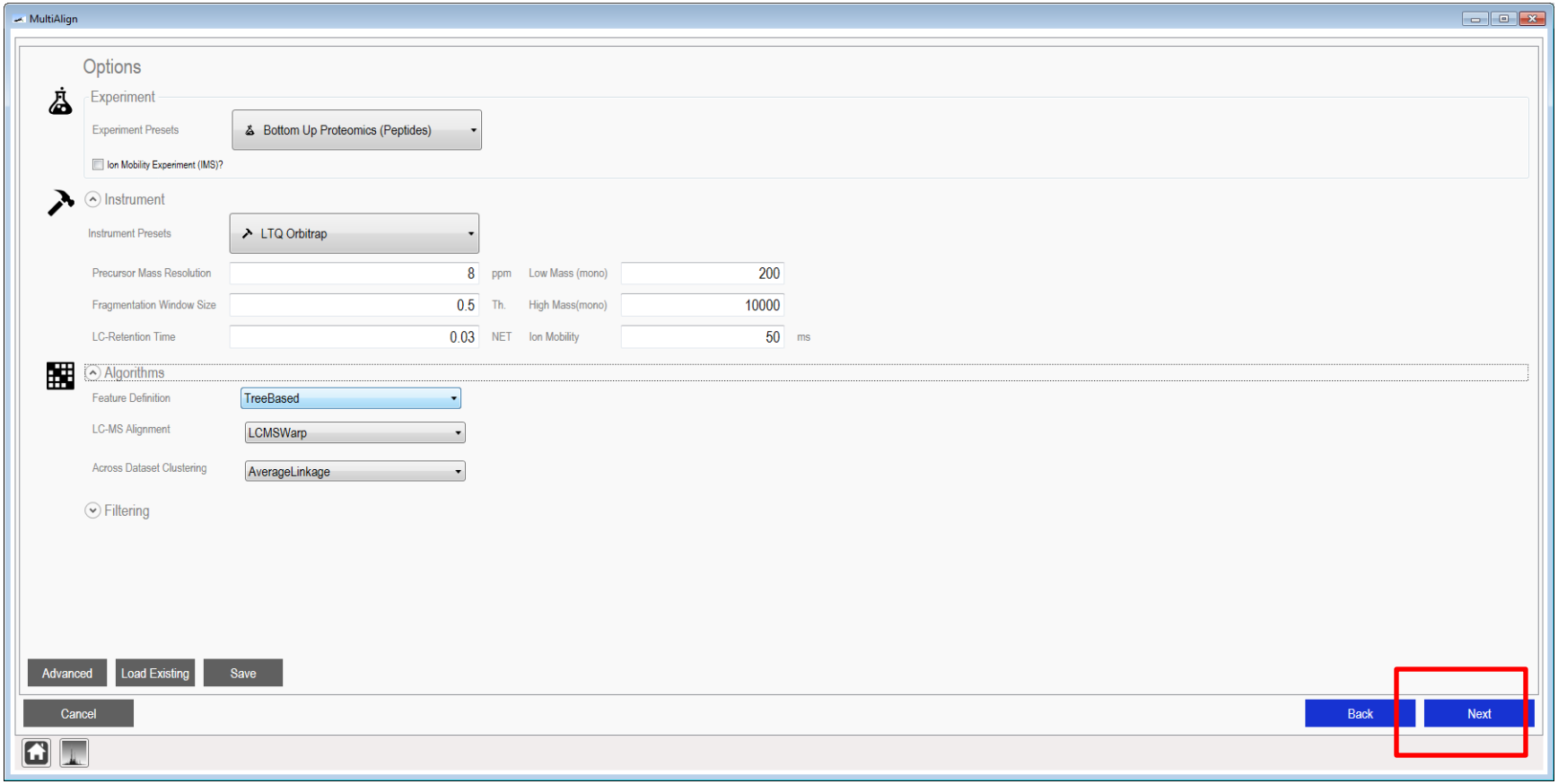
Feeling advanced? All options are still available here…



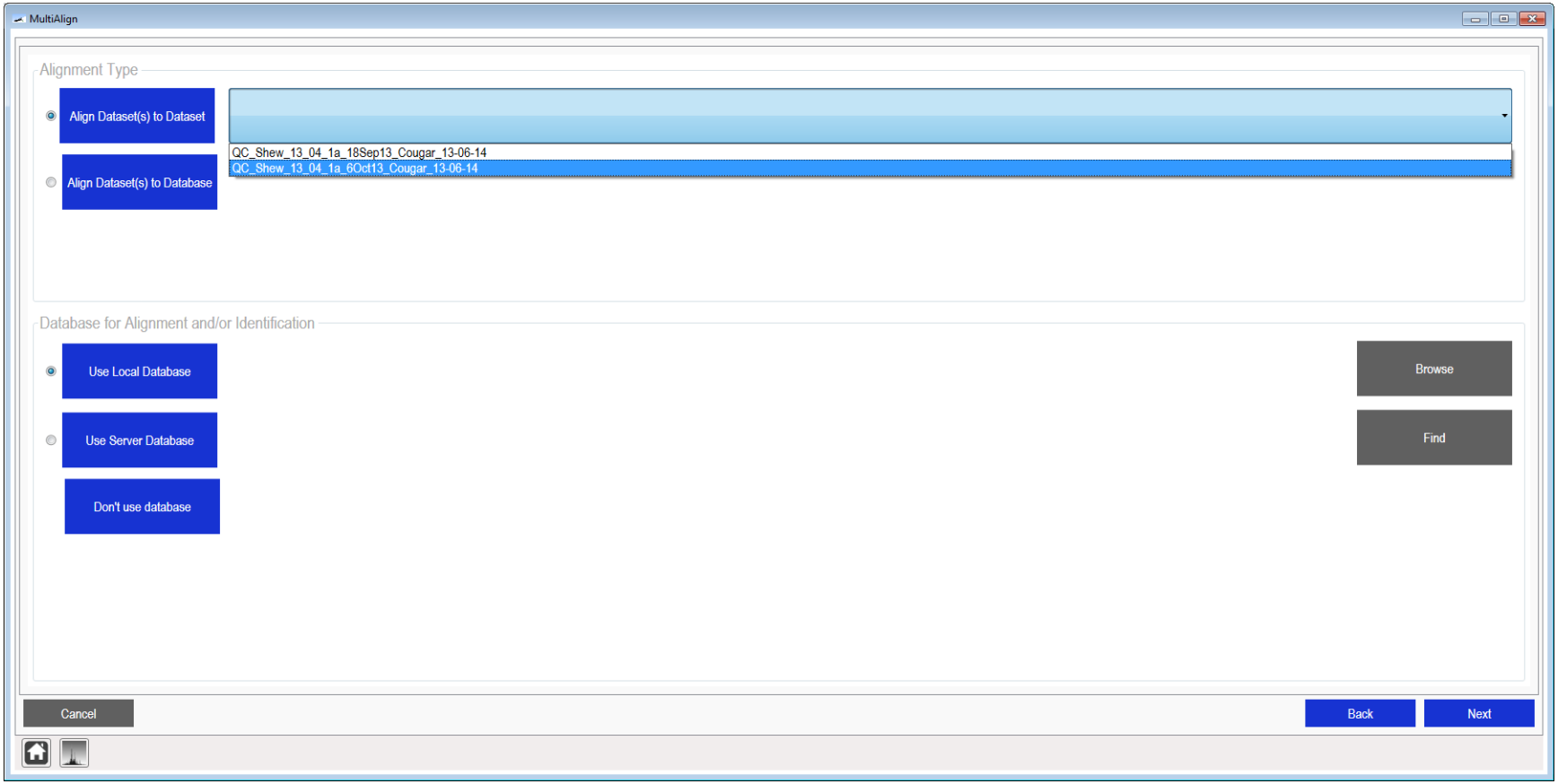
If you need to….



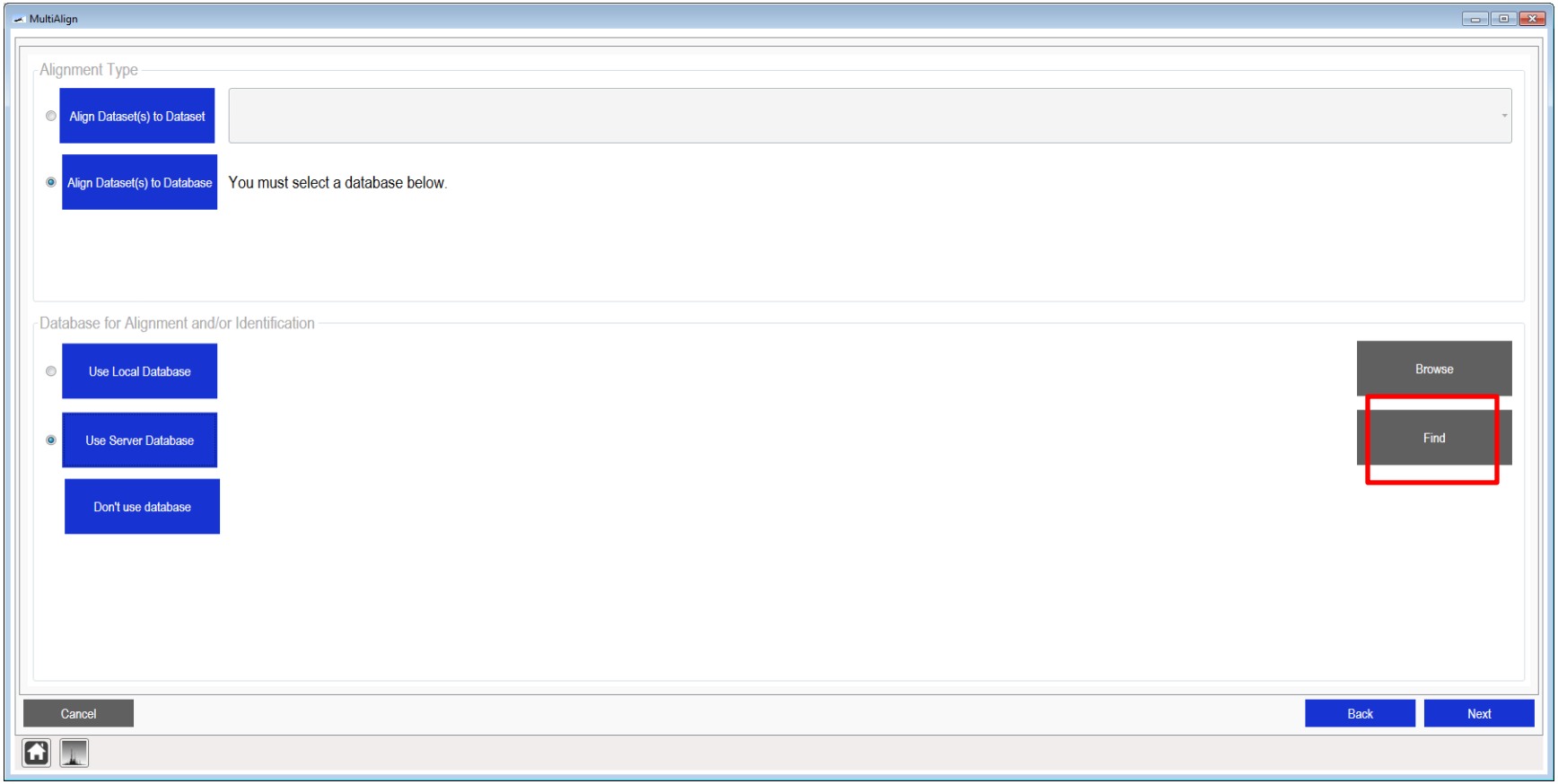
Next…


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
Advanced
Load Existing
Save
Cancel
Back
Next

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

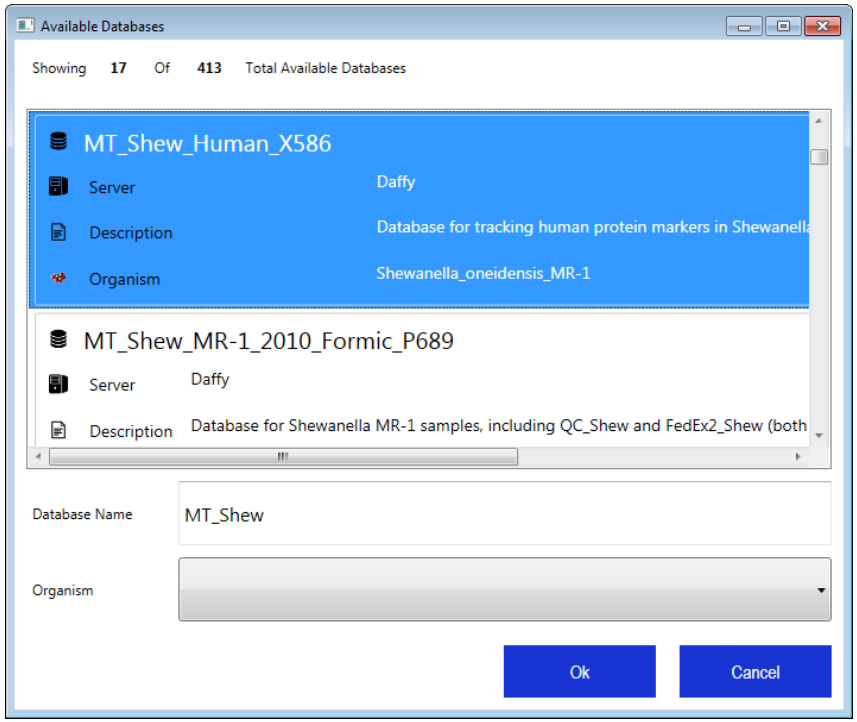


Align to an AMT Tag DB



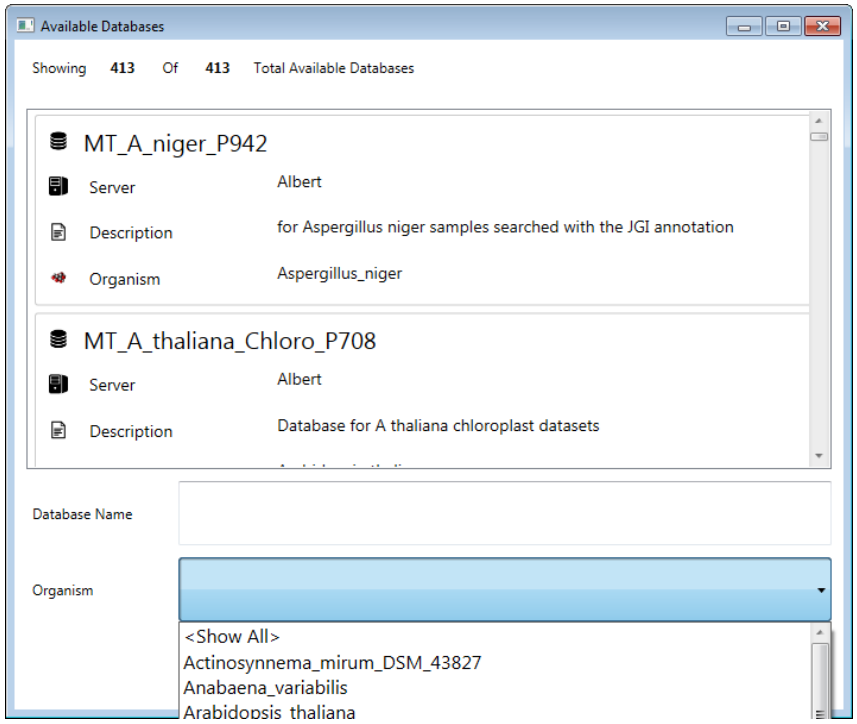
Select your database

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



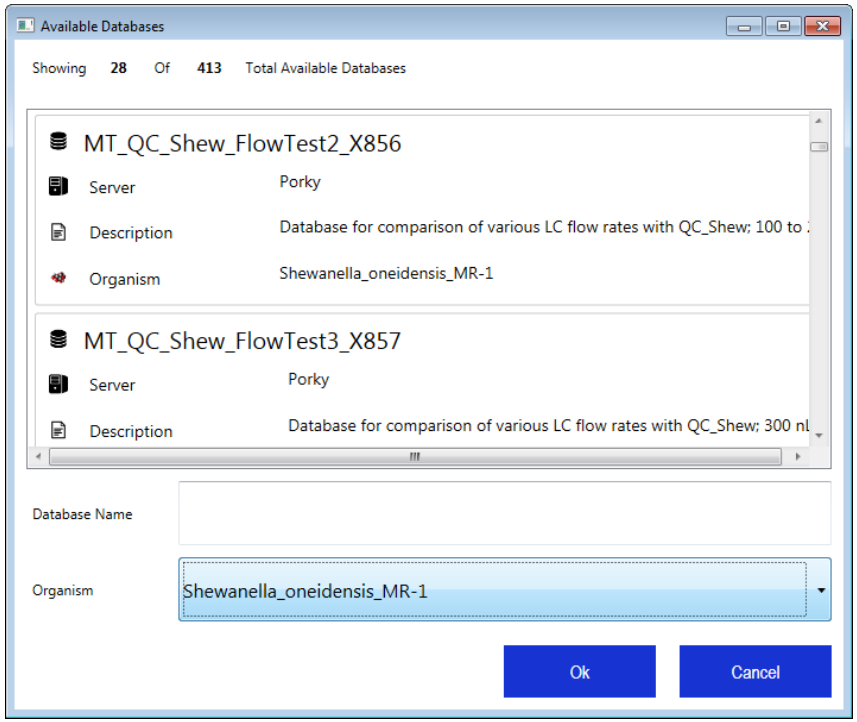
Select your database

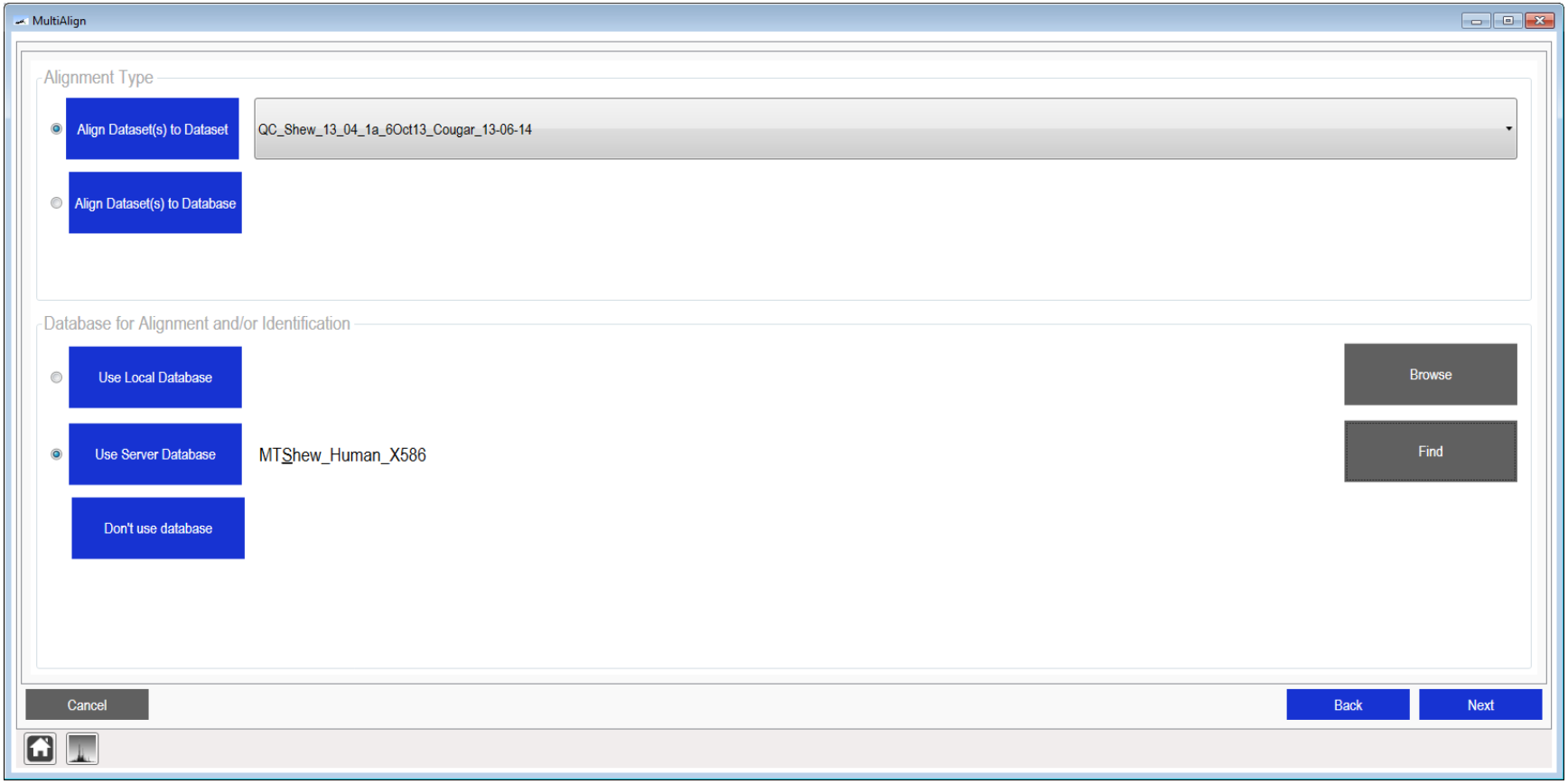
- Or find all database for a given organism



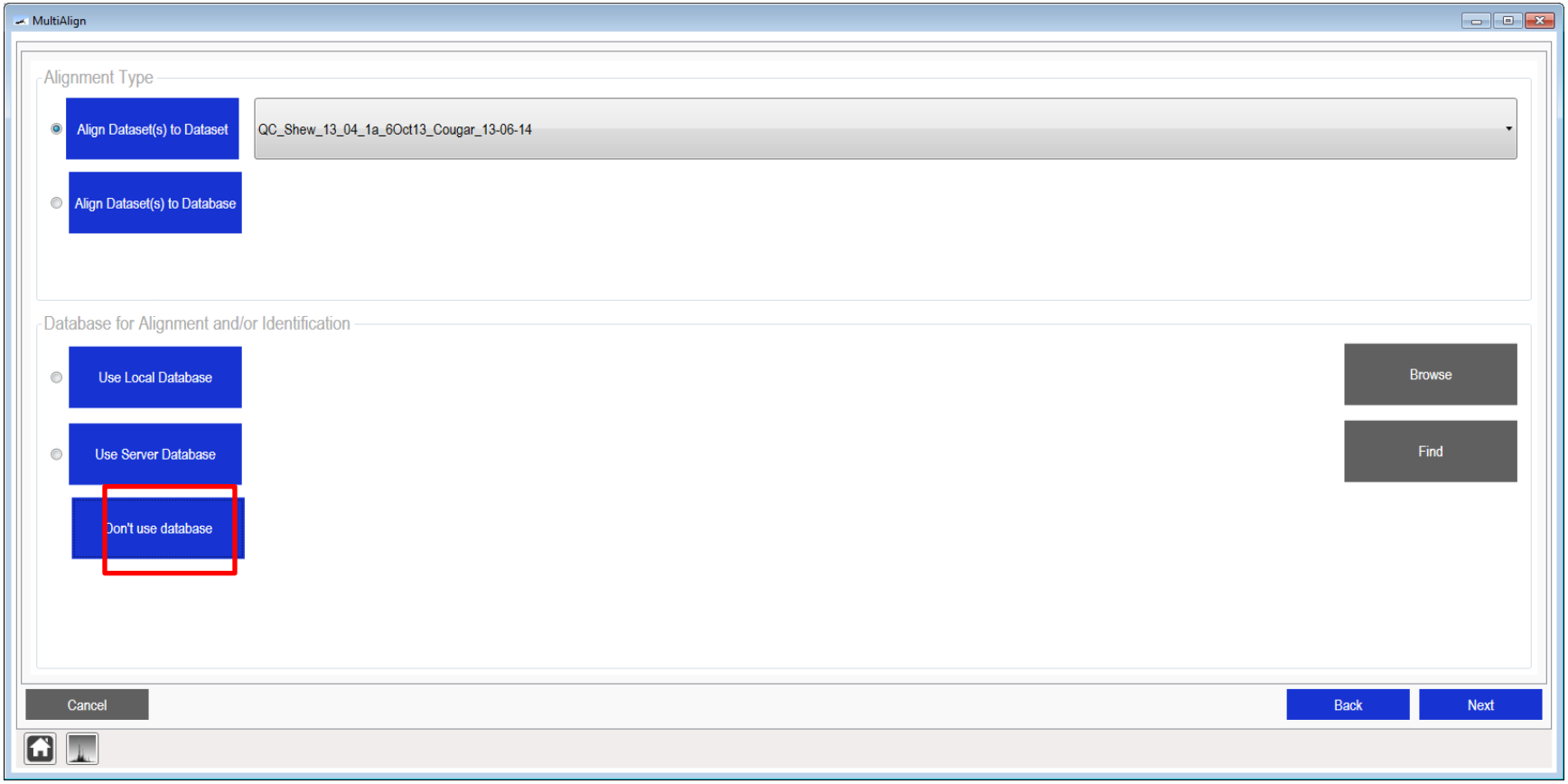
Select your database

- Displays all databases for the selected organism

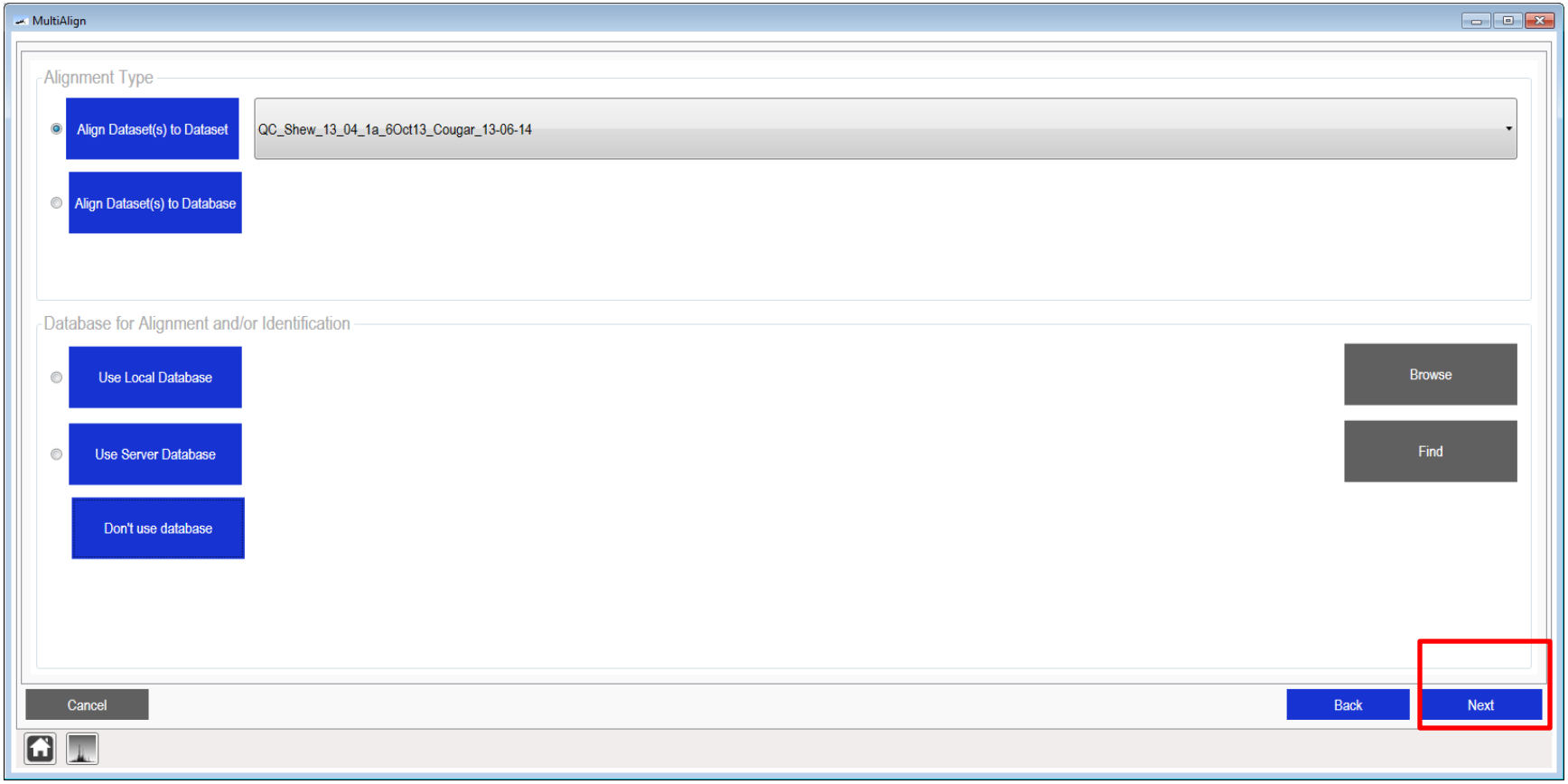



MultiAlign
Alignment Type
Align Dataset(s) to Dataset
QC_Shew_13_04_1a_6Oct13_Cougar_13-06-14
Align Dataset(s) to Database
Database for Alignment and/or Identification
Use Local Database
Browse
Use Server Database
MTShew_Human_X586
Find
Don't use database
Cancel
Back
Next

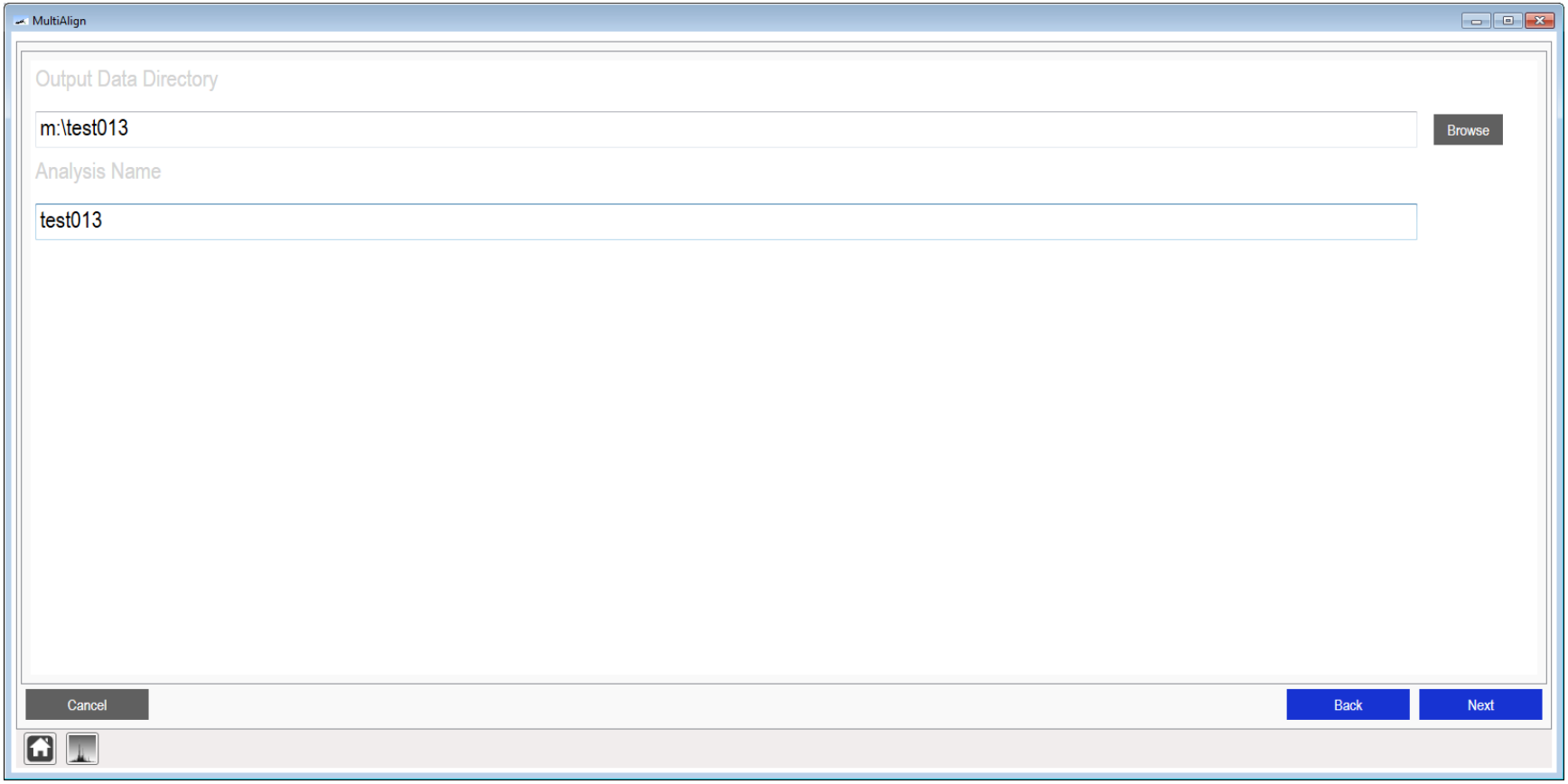
For now, just use a dataset, so press the “don’t use database” button



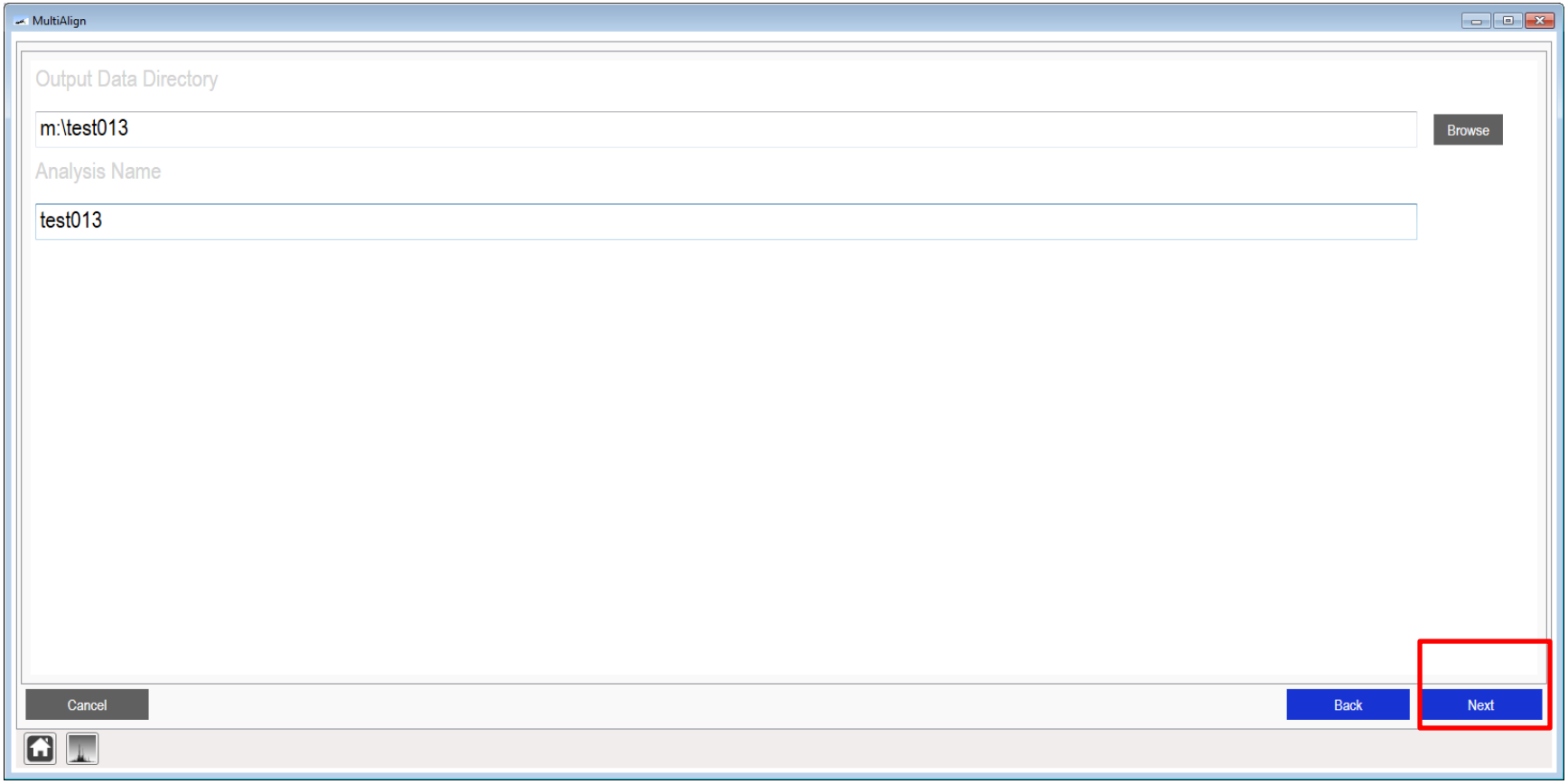
Next


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

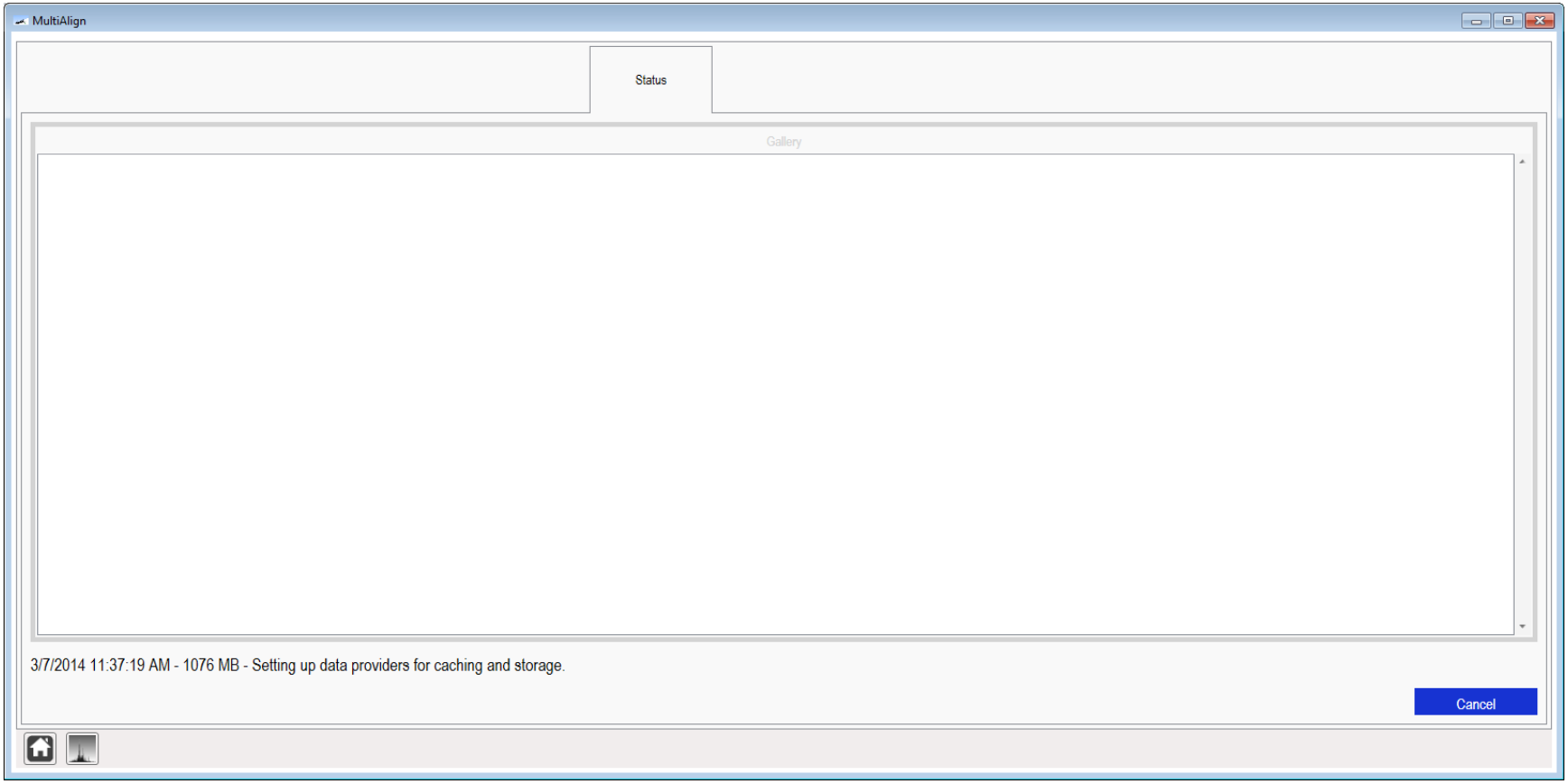
Enter a name, and output path



Next – starts the analysis



Status window will show progress of analysis…

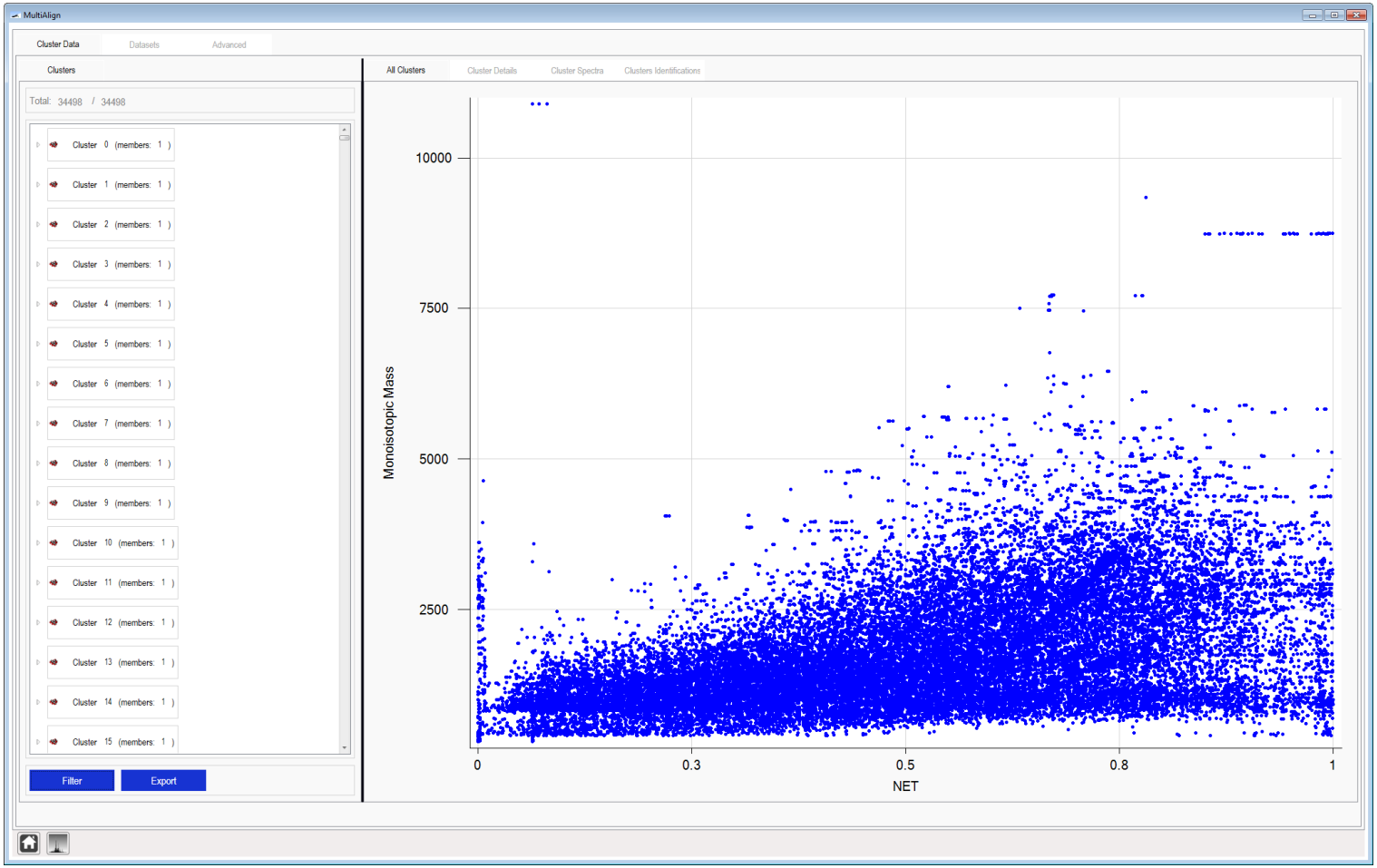


NOTE: this feature is being improved…if it looks like the software is not doing anything it is.

Analysis View

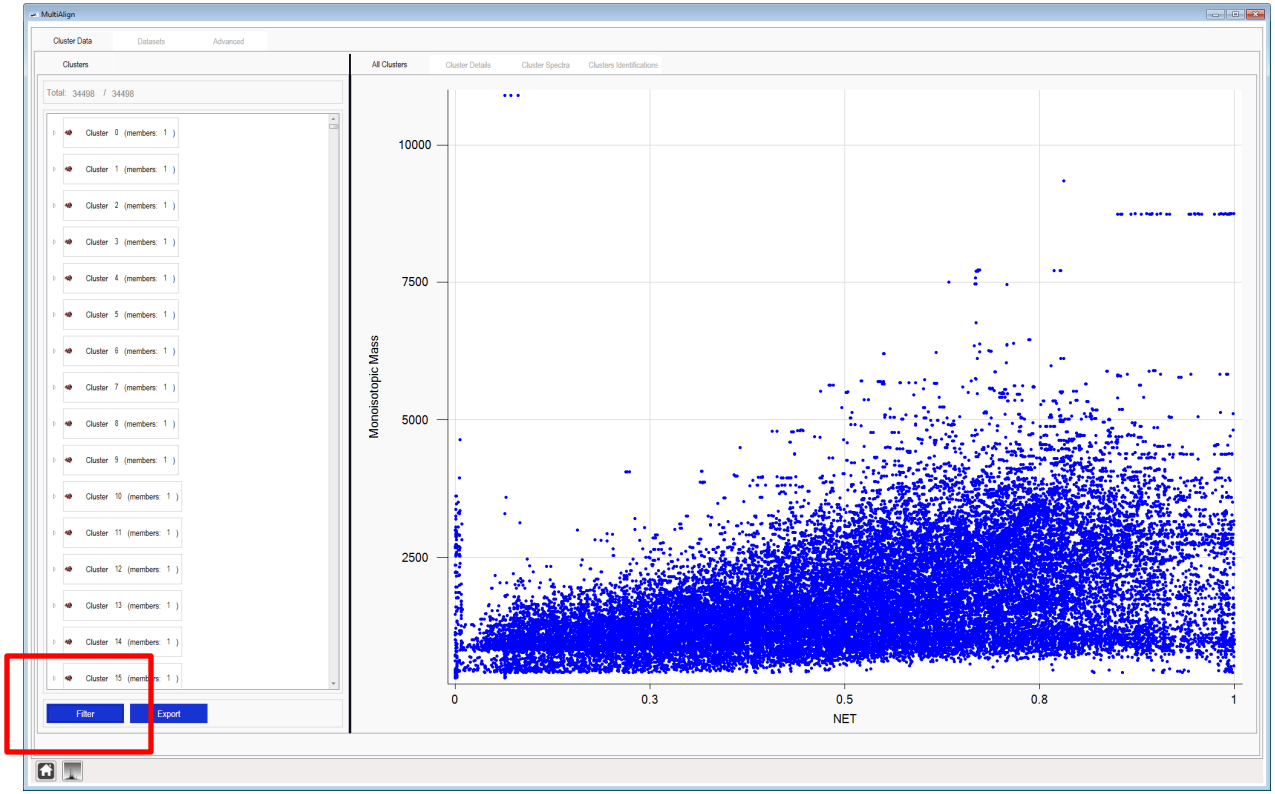
- To demonstrate the GUI
  - 2 Datasets of QC-Shew were aligned and clustered
  - MS/MS spectra were not linked to any MSGF+ results (i.e. the files were not loaded with MultiAlign)

New GUI – View of the analysis



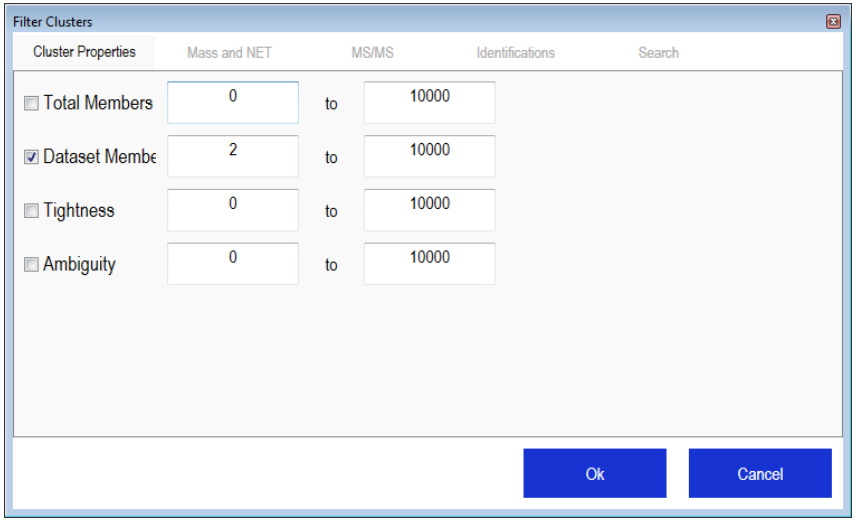
New GUI – View of the analysis

- Click Filter



Filtering

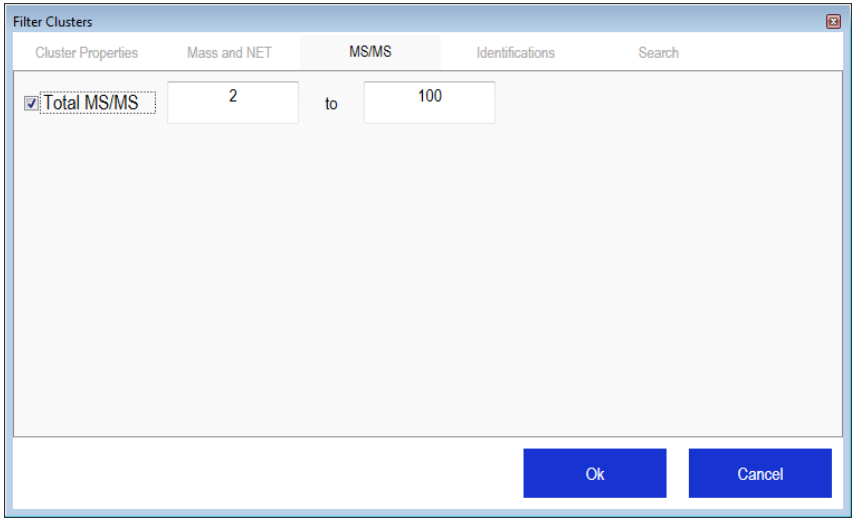
- “Show me clusters that are present in 2 or more datasets”



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

Filtering

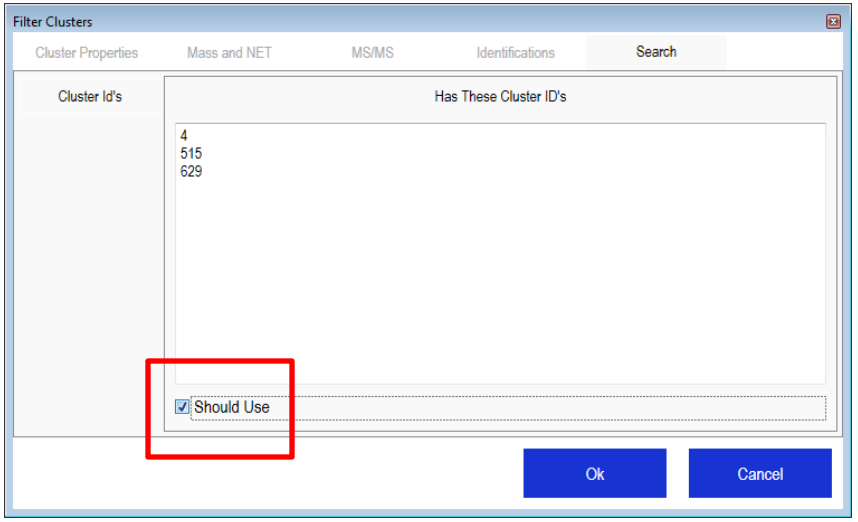
- “Show me clusters that have at least 2 MS/MS spectra”



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

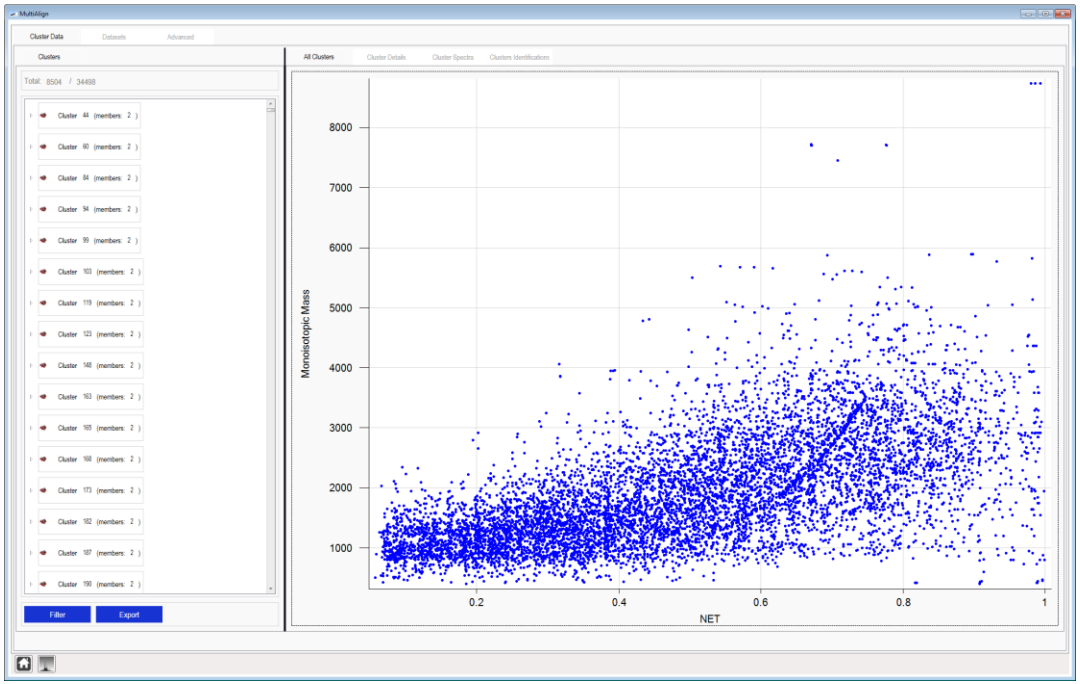
Filtering

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



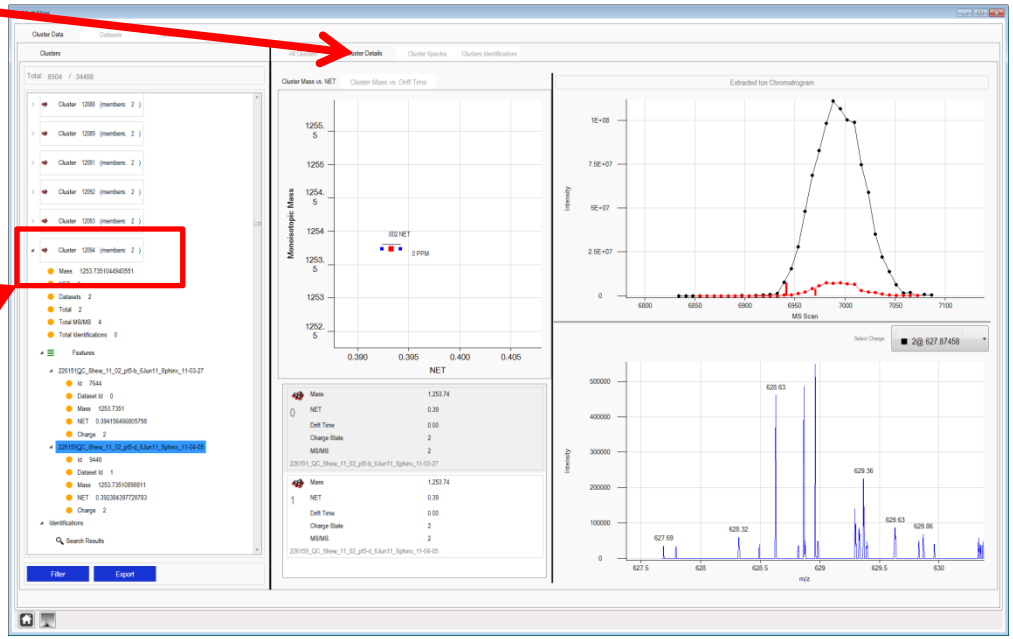
Note: filtering is an intersection “must have this AND this AND this AND this”

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



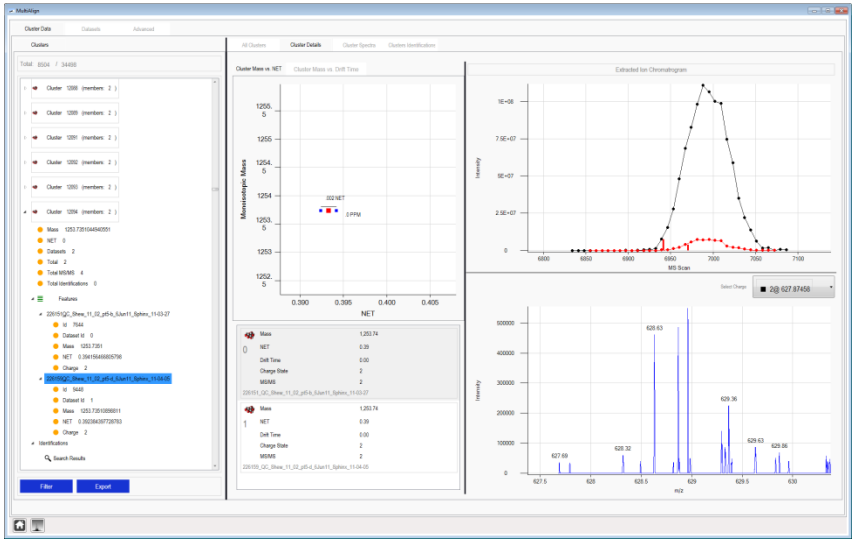
View A Cluster

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



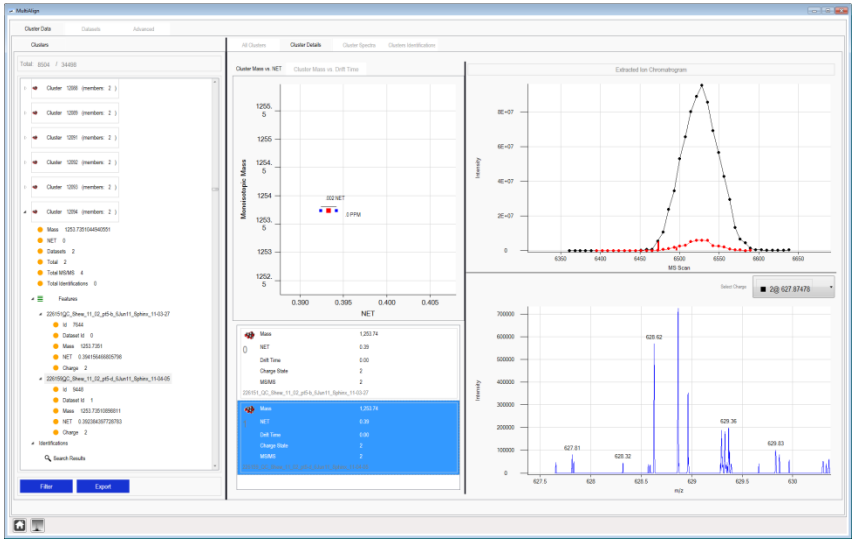
Cluster View

- 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

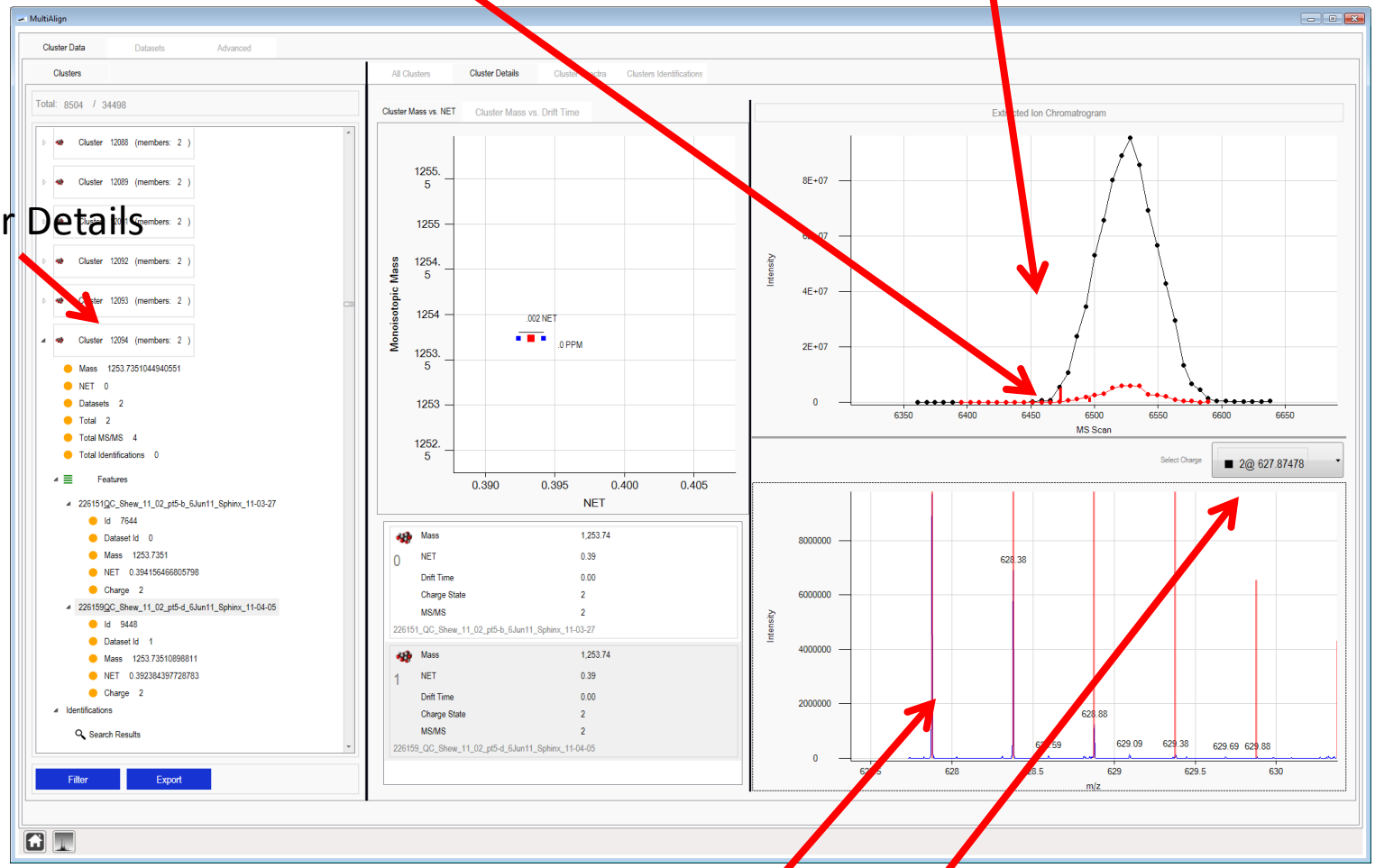
- Same Information displayed as with other dataset



MS/MS Fragmentation Event

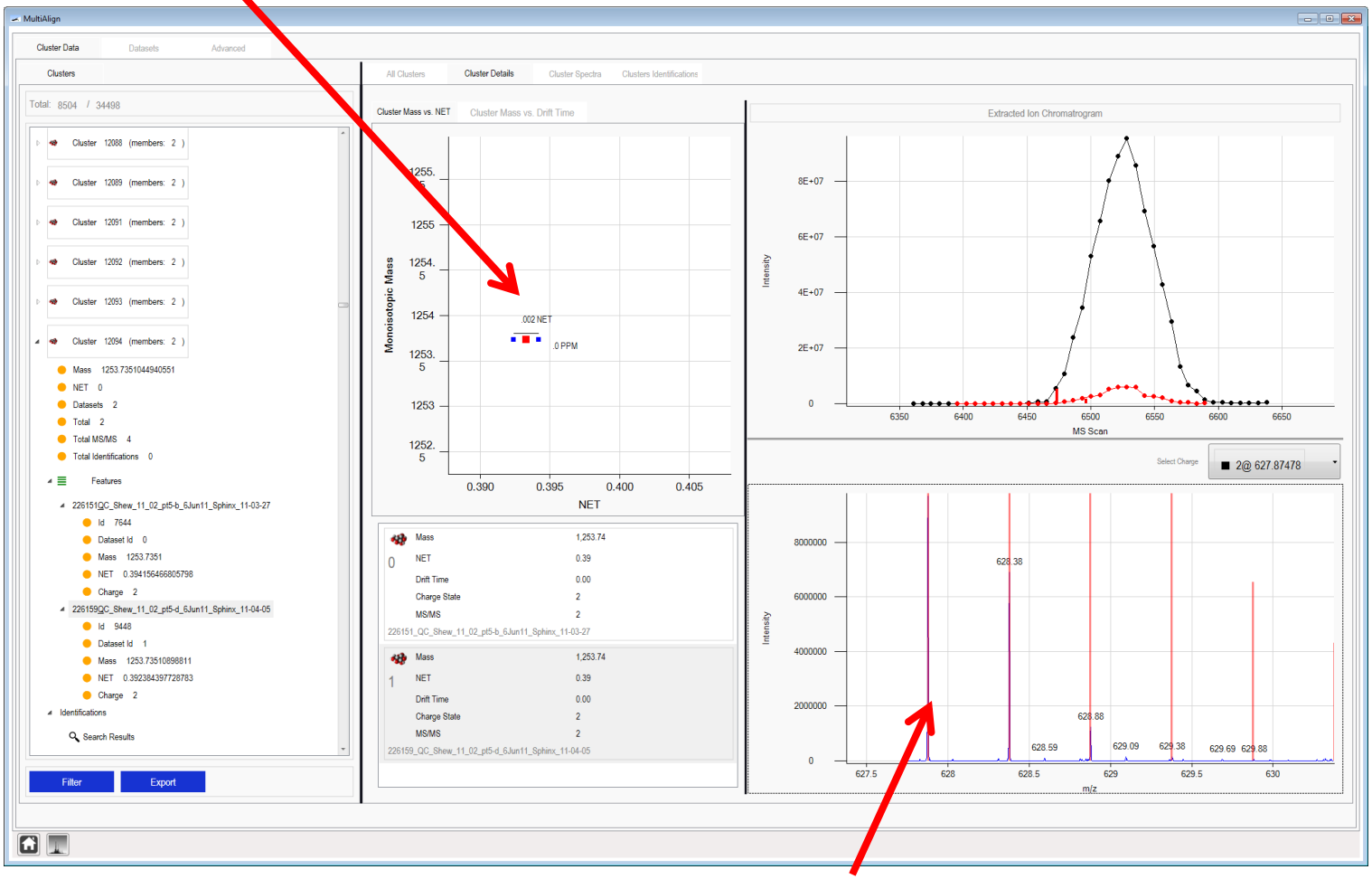
XIC for selected feature

Cluster Details



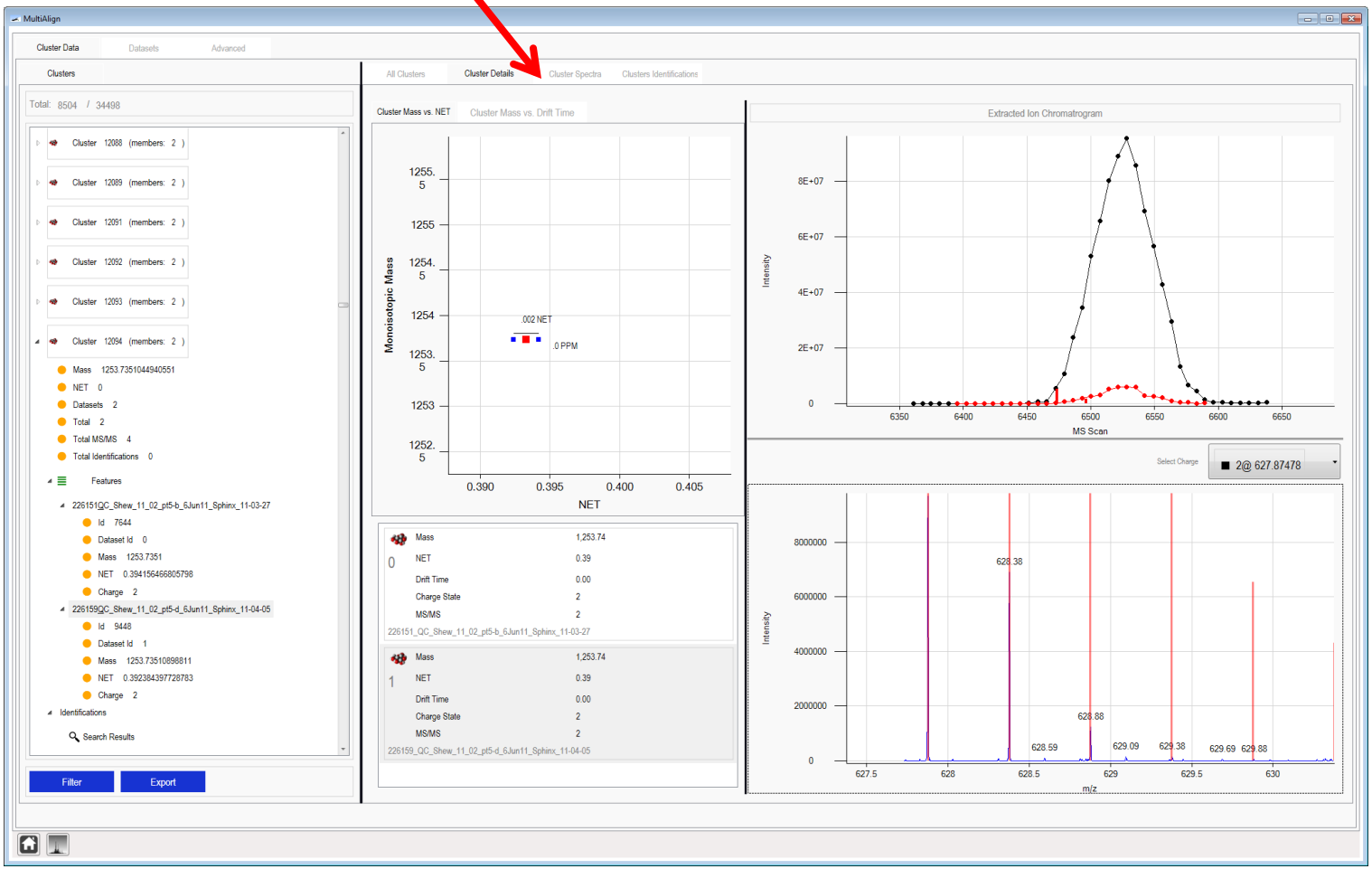
Parent Spectra for selected charge state

Cluster scatter plot



Isotopic distribution (first peak is monoisotopic)

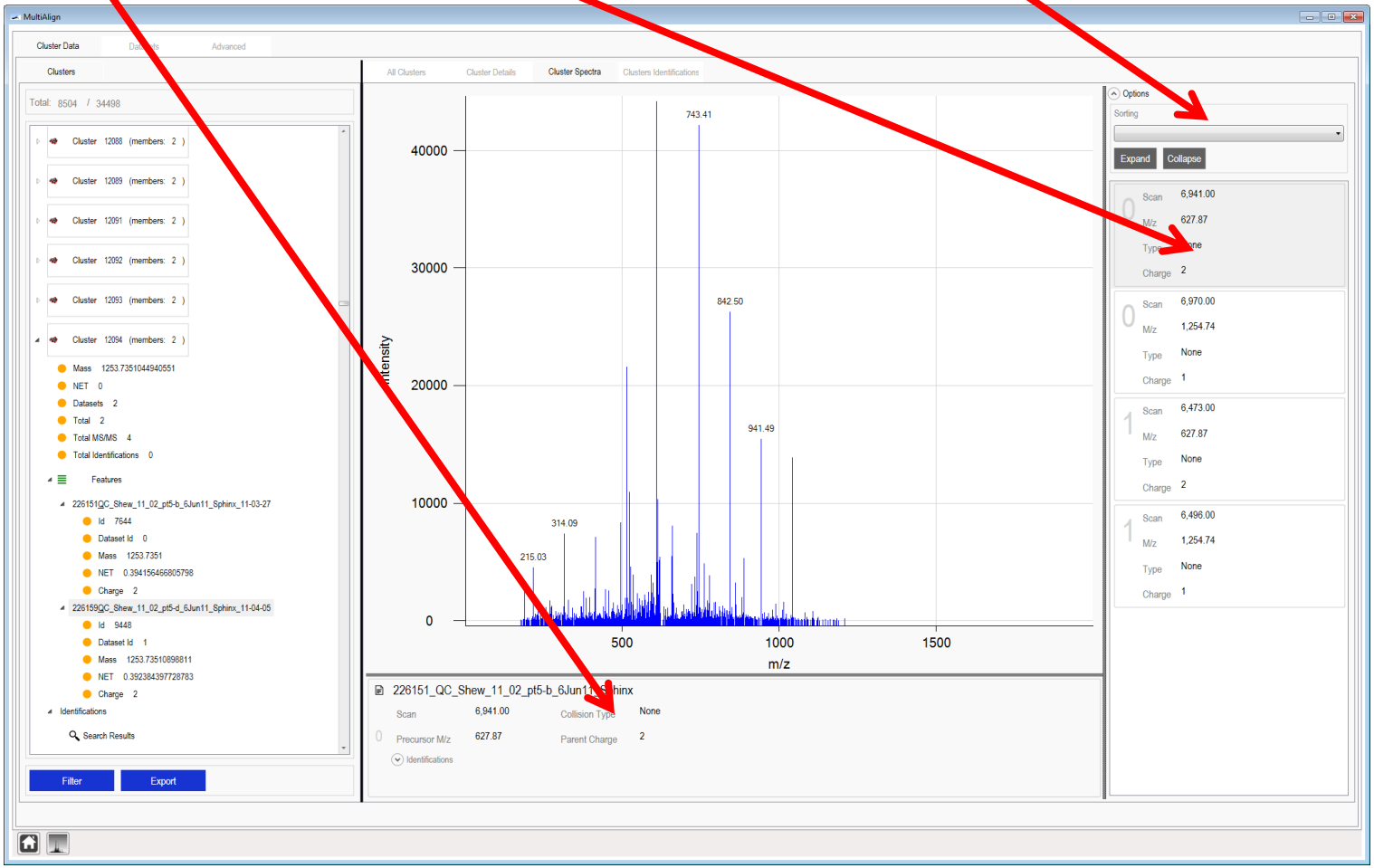
To see MS/MS Spectra for this cluster click here



List of all spectra

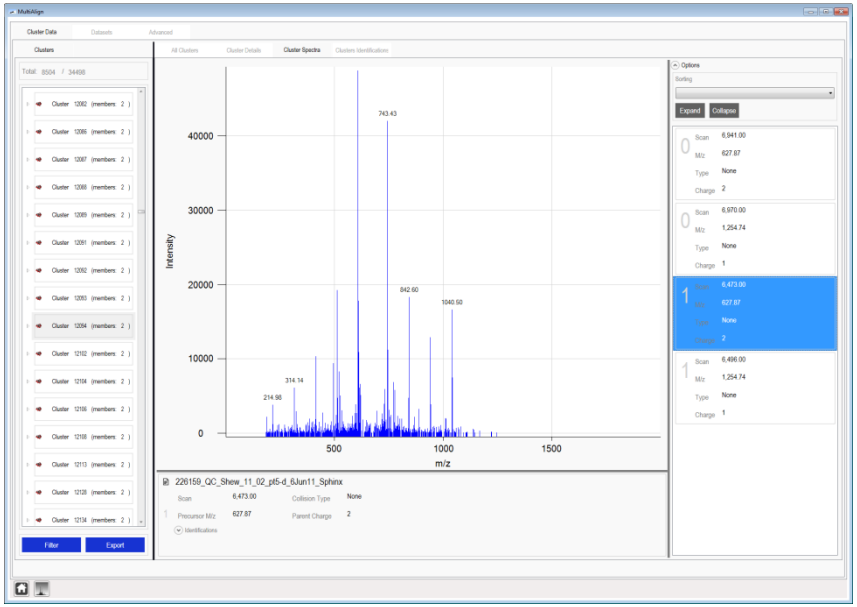
Sorting options

Detailed spectra view

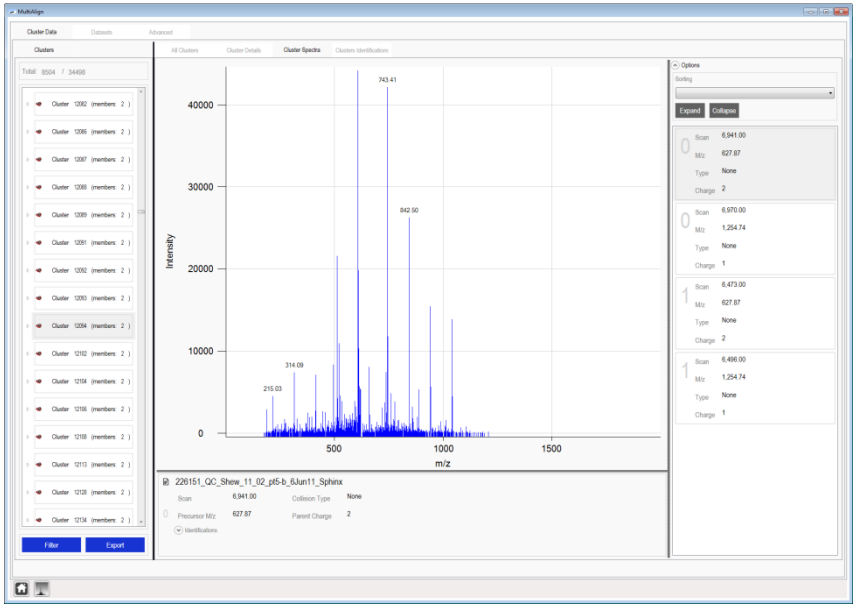


MS / MS Comparisons +2

Dataset 1 – Charge +2

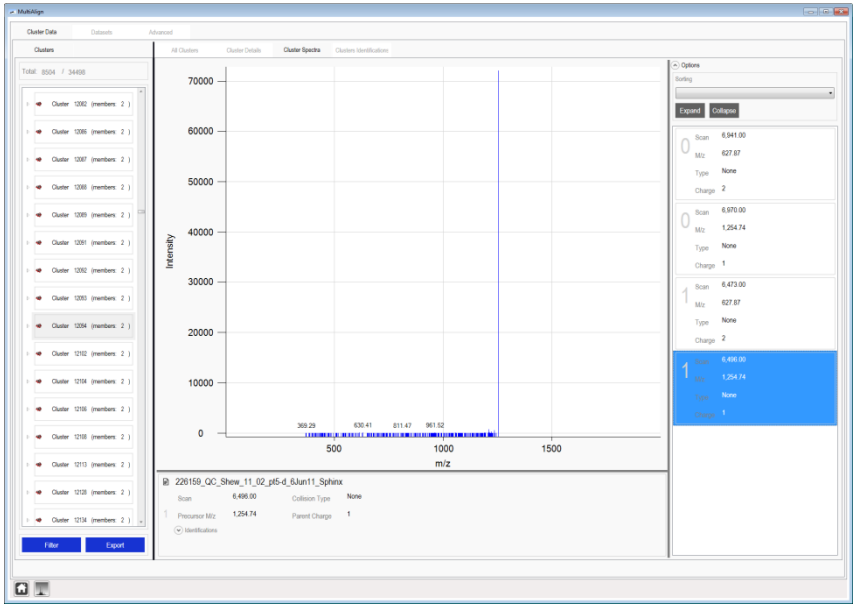


Dataset 0 – Charge +2

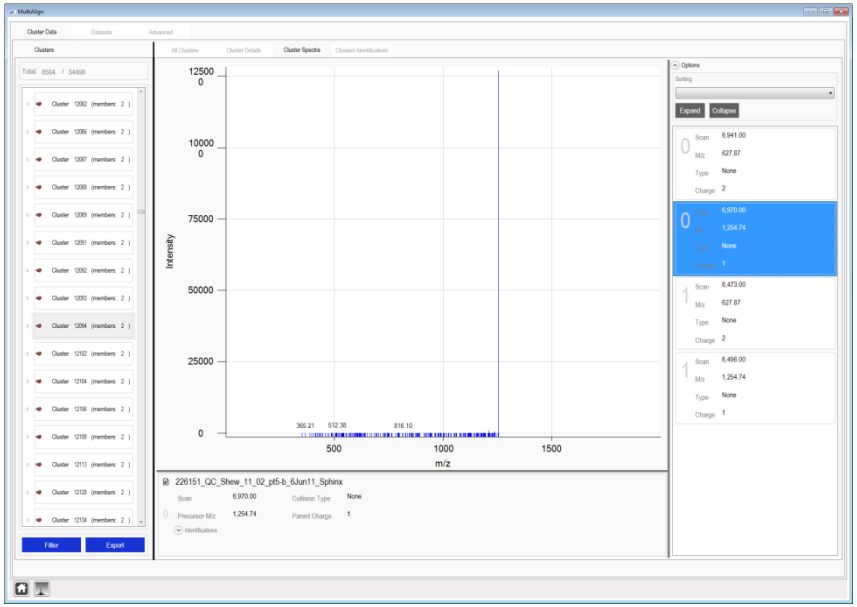


MS/MS Comparisons +1

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