



MultiAlign Tutorial 03 – Running an Analysis

BRIAN LAMARCHE

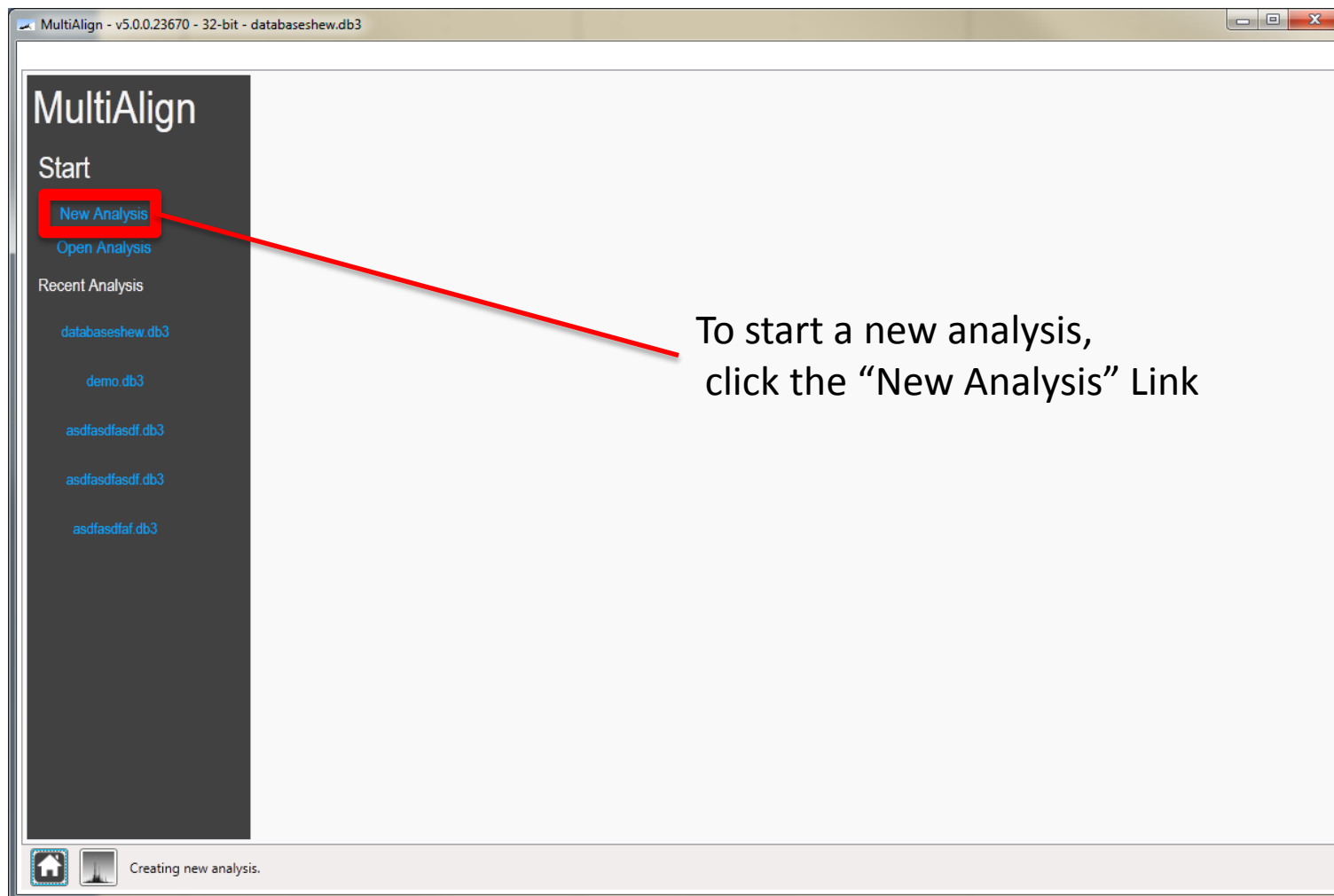
About this tutorial

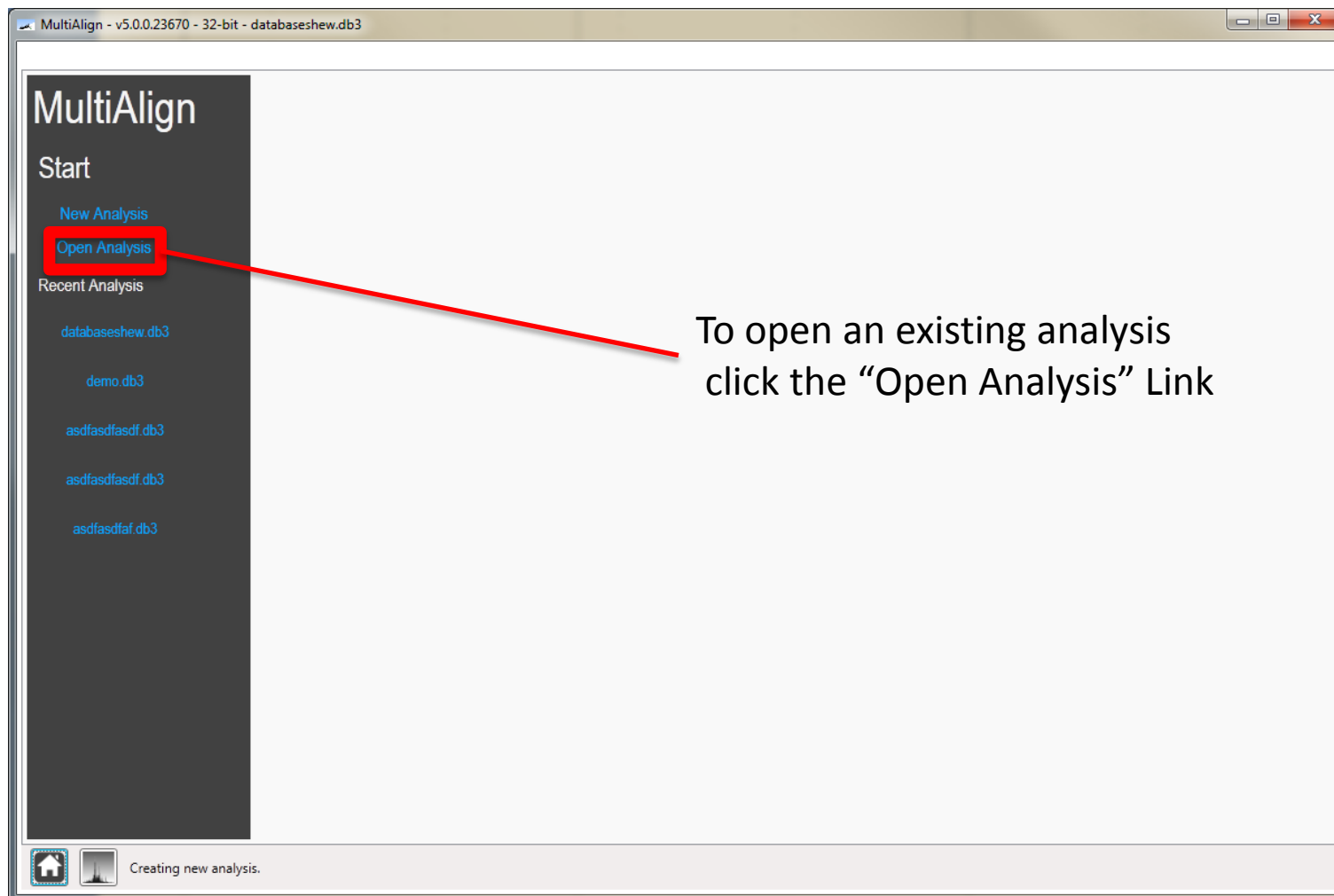
- ▶ This tutorial provides an introduction to the graphical user interface (GUI)
- ▶ This tutorial will walk you through each step of creating a new MultiAlign analysis.



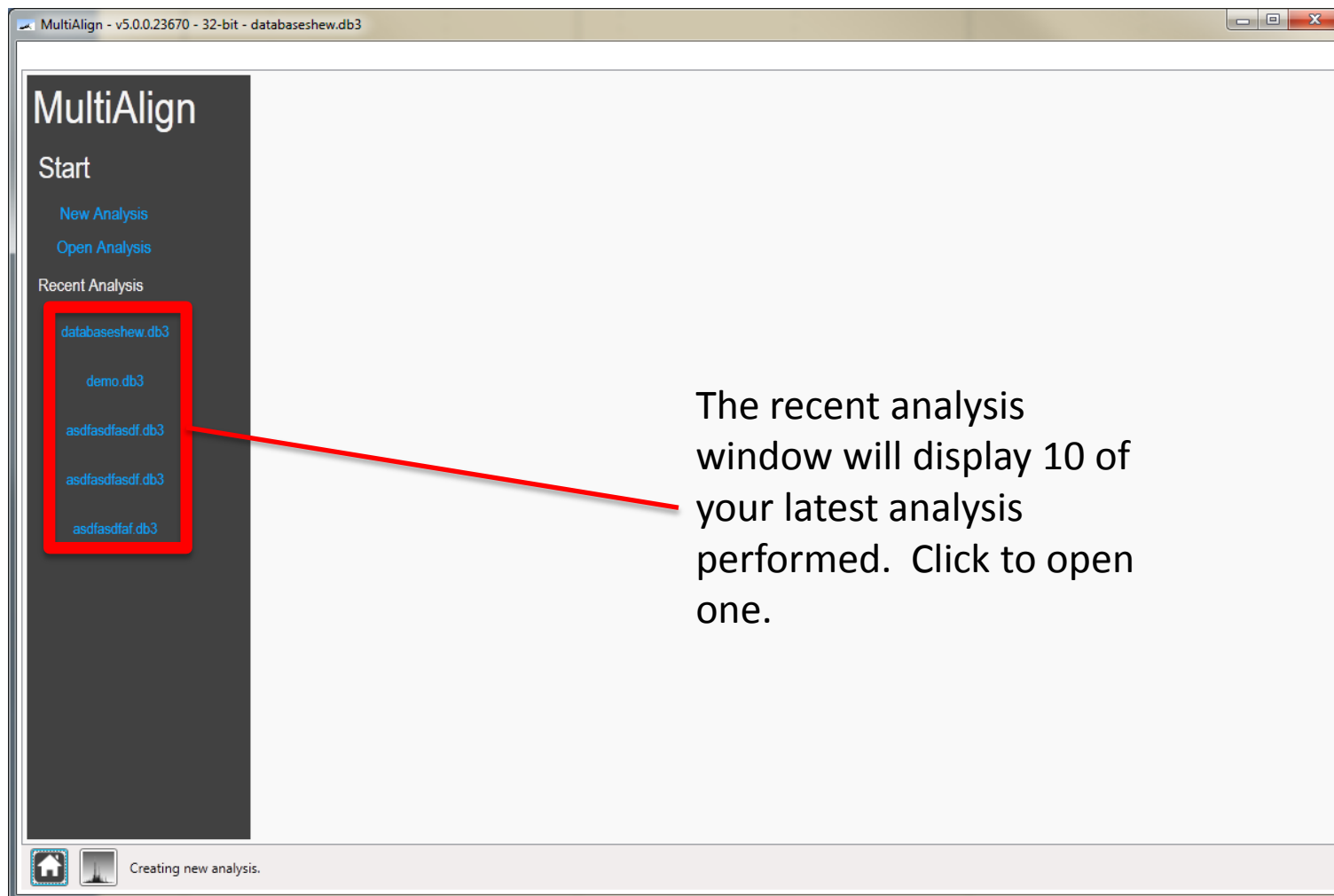
GUI Basics

UNDERSTANDING THE SCREENS

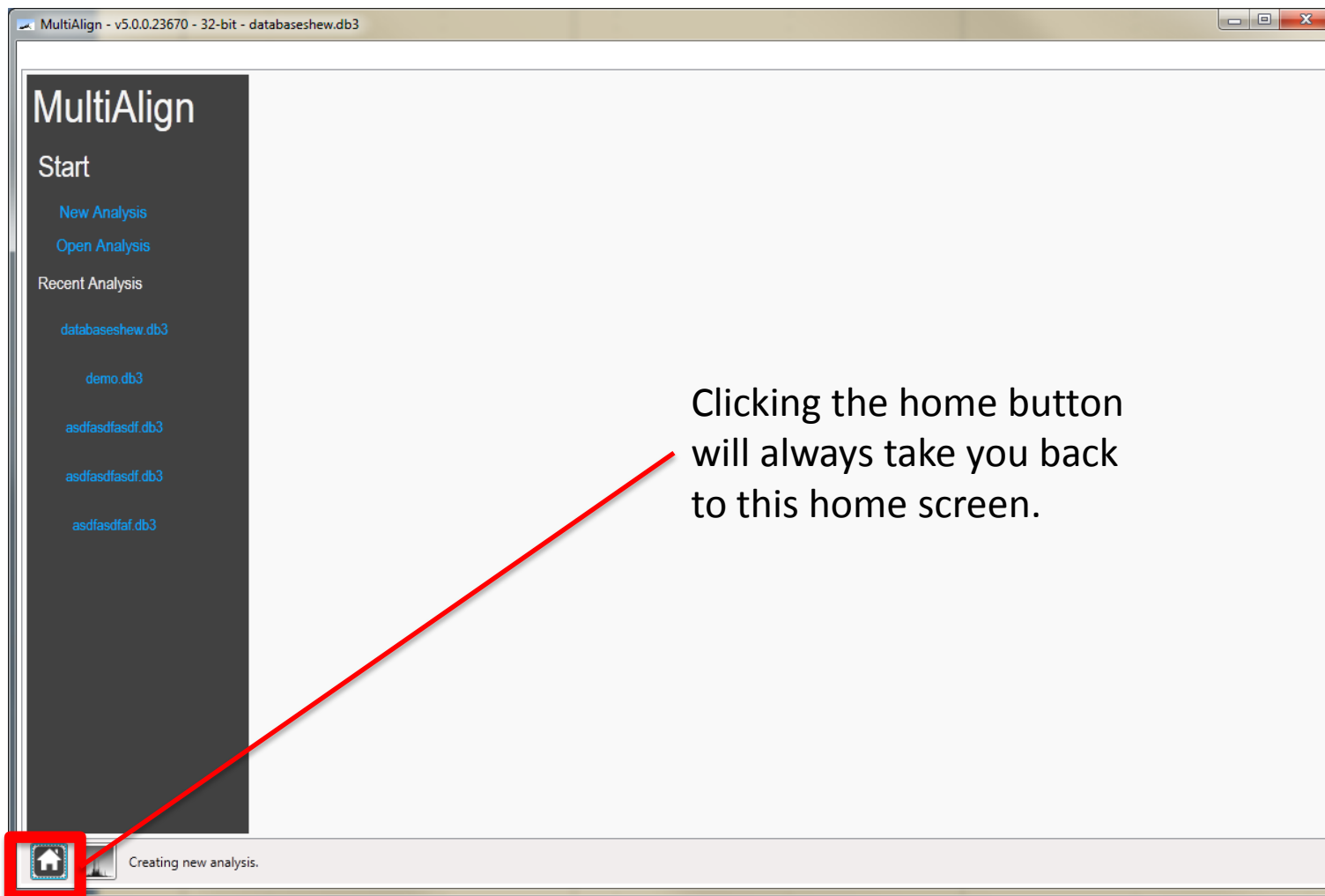




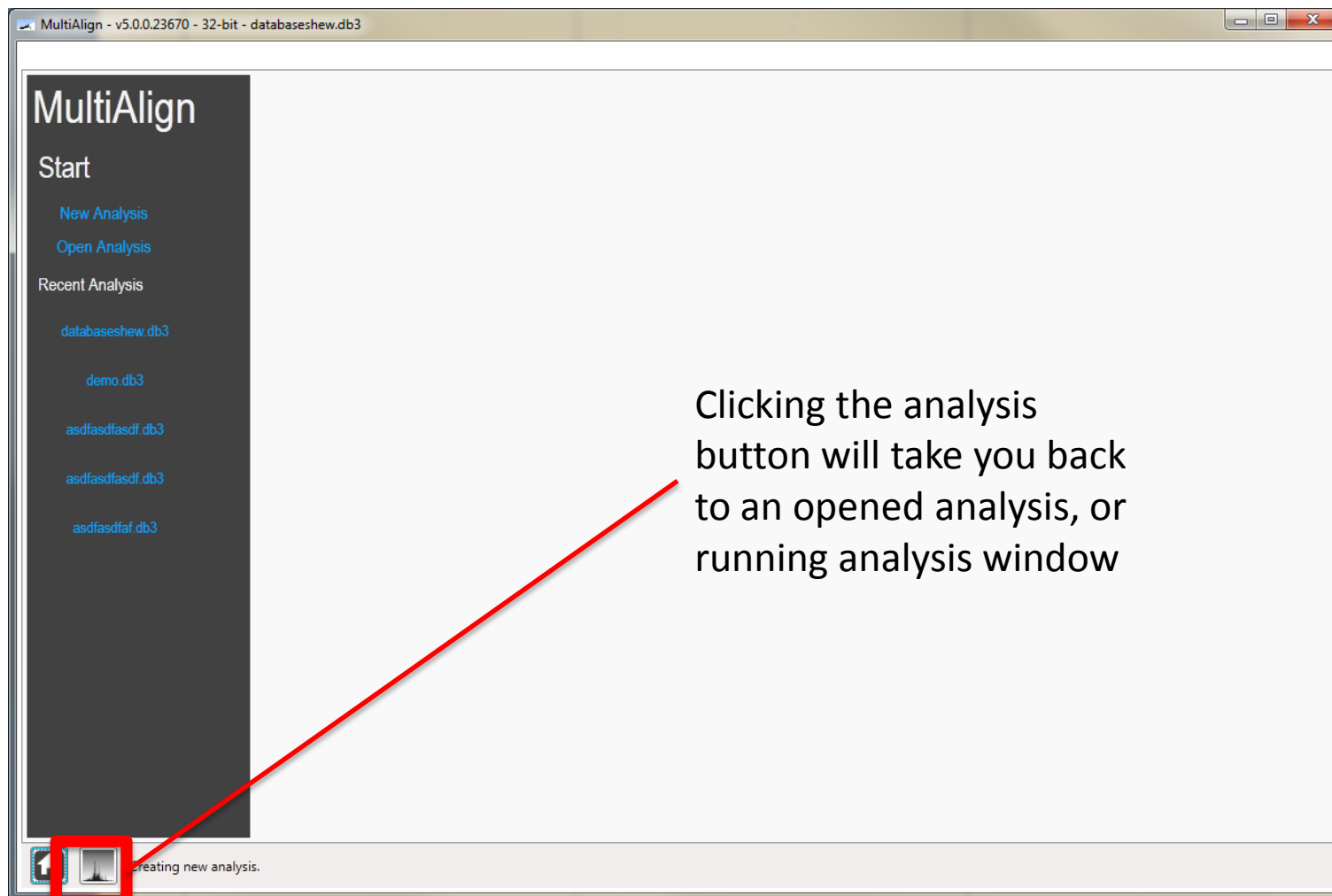
To open an existing analysis
click the "Open Analysis" Link



The recent analysis window will display 10 of your latest analysis performed. Click to open one.



Clicking the home button
will always take you back
to this home screen.

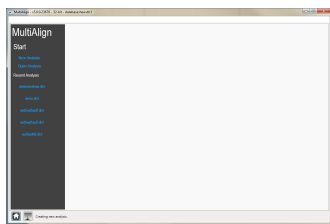


Starting a new analysis

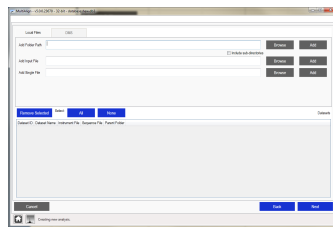
RUNNING A NEW ANALYSIS, LOADING PARAMETER FILES, SETTING PARAMETERS, SELECTING DATA TO ANALYZE, AND SELECTING BASELINE DATASETS OR DATABASES

The Analysis is broken down into several steps

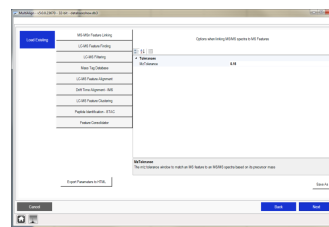
Wizard



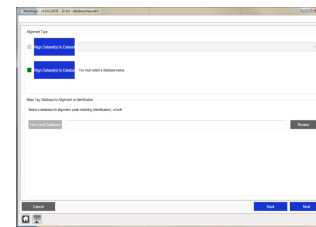
1. Home Screen



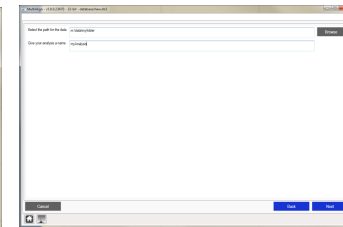
2. Select Data



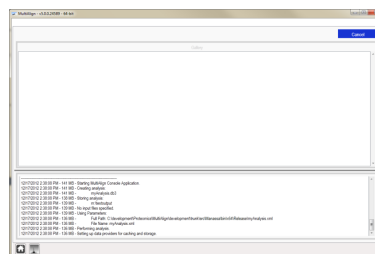
3. Set Parameters



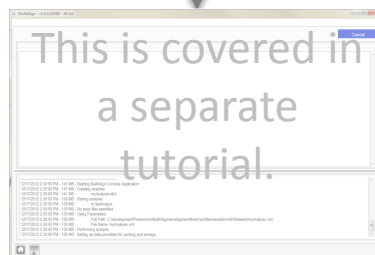
4. Select Baseline
and Mass Tag
Database



5. Set Analysis
Path and Name



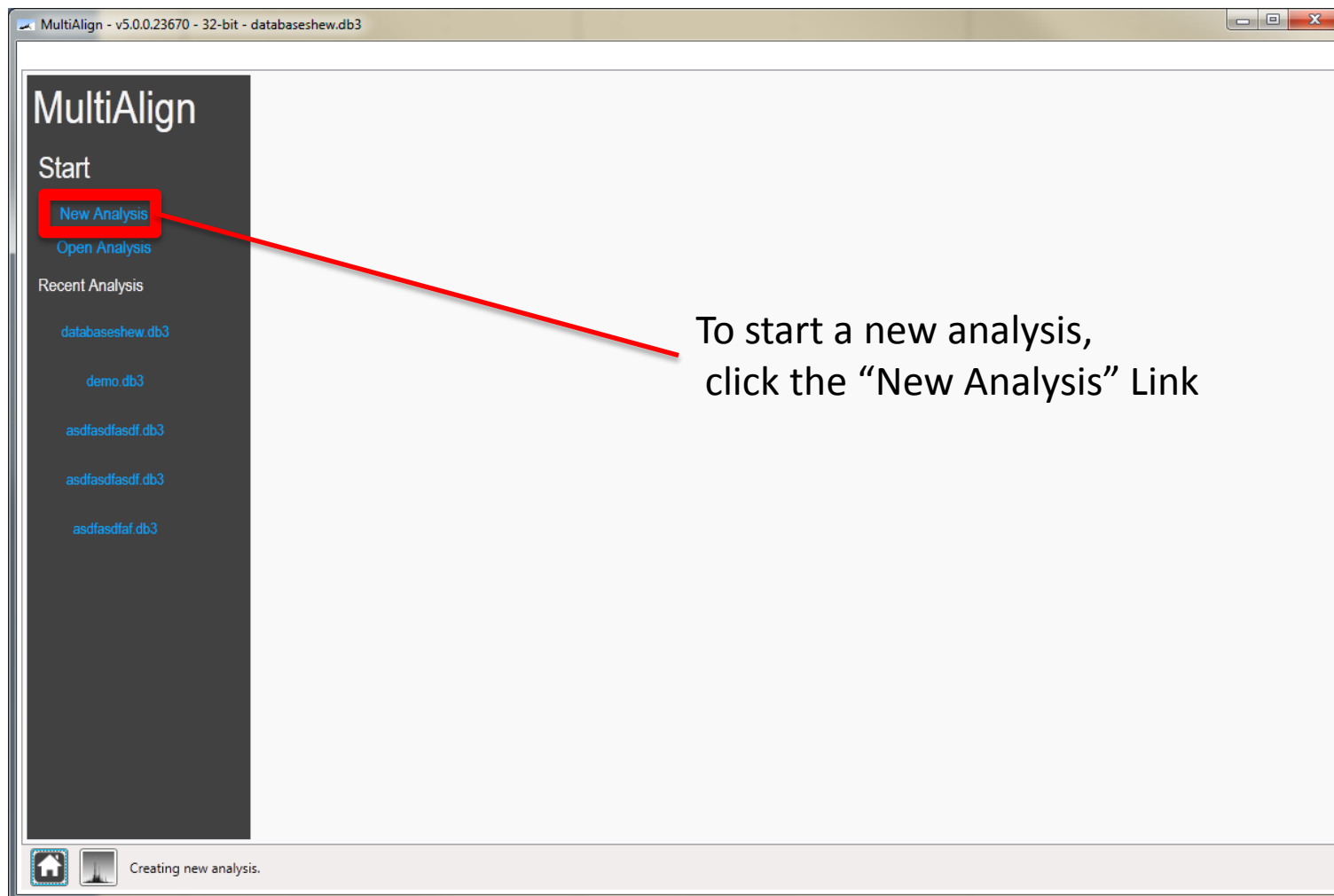
6. Running
Analysis



7. Analysis View
Window



Starting the analysis



To start a new analysis,
click the "New Analysis" Link



Load Data Page

MultiAlign - v5.0.0.23670 - 32-bit - databaseshew.db3

Local Files | DMS

Add Folder Path

☐ Include sub-directories

Browse

Add

Add Input File

Browse

Add

Add Single File



Browse

Add

Remove Selected | Select: All None | Datasets

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

Cancel Back Next

  Creating new analysis.

This page helps
you load
datasets into
the application.



Load data from directory

MultiAlign - v5.0.0.23670 - 32-bit - databaseshew.db3

Local Files | DMS

Add Folder Path

☐ Include sub-directories

Add Input File

Add Single File

Select:

Datasets

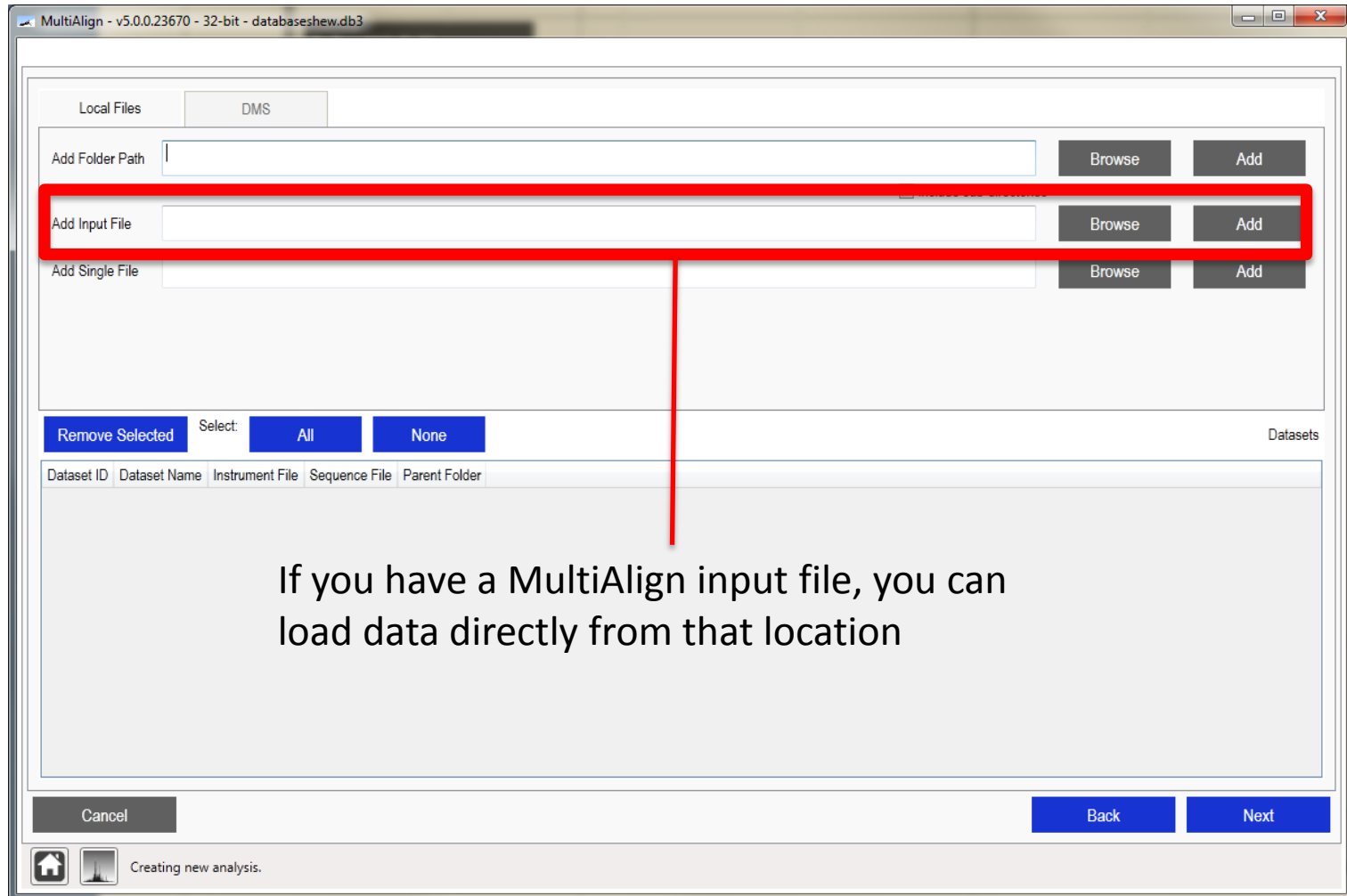
Dataset ID	Dataset Name	Instrument File	Sequence File	Parent Folder
------------	--------------	-----------------	---------------	---------------

Click browse to load files stored in a directory. Select "include sub-directories" if you want to include files in sub-directories.

Creating new analysis.

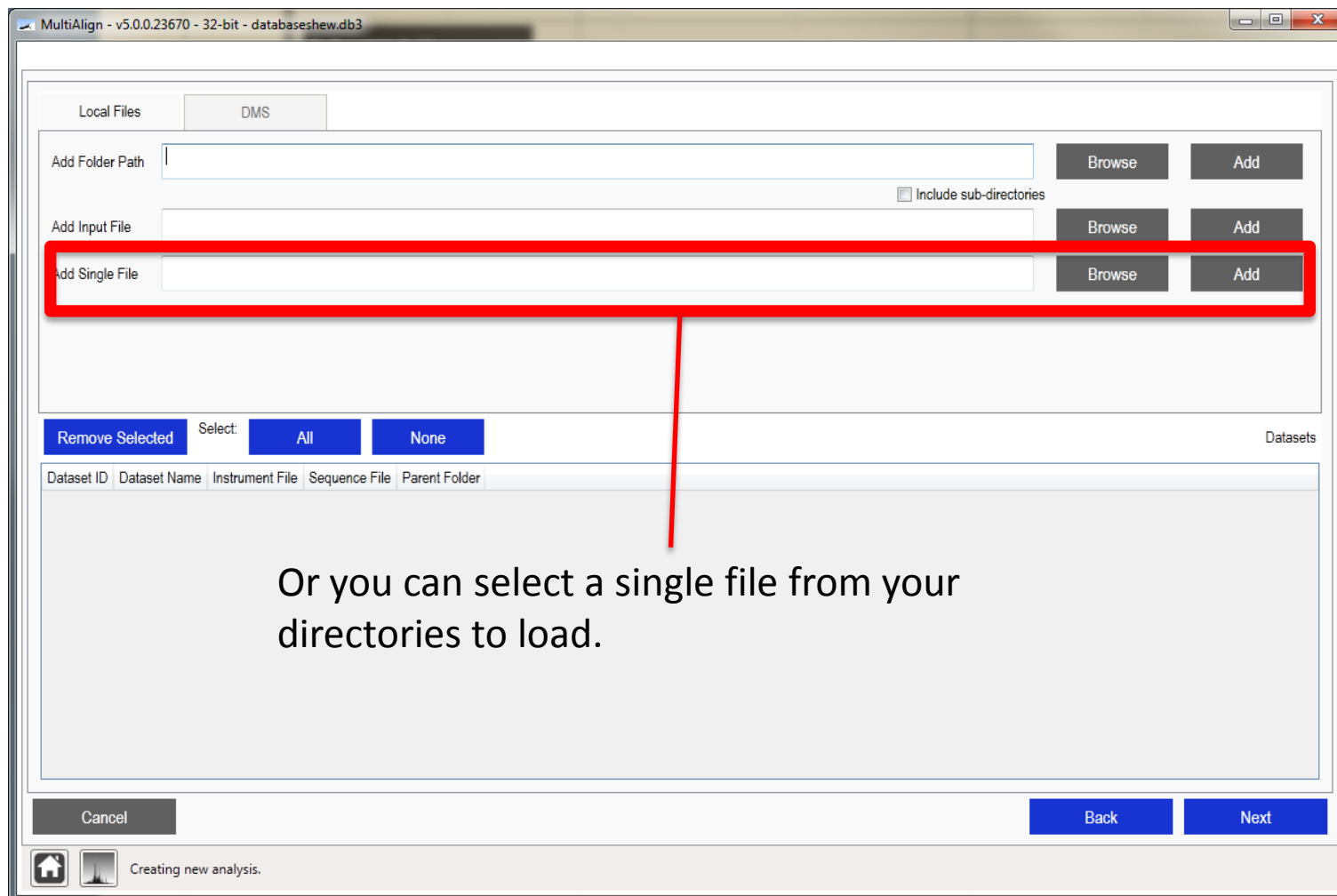


Load from an input file





Load a single dataset file





Load Data Page

MultiAlign - v5.0.0.23670 - 32-bit - databaseshew.db3

Local Files DMS

Analysis files are displayed in the data grid below.

You can remove any files you accidentally loaded or do not want to include by first selecting the items, then pressing the remove selected button

Remove Selected Select: All None

Dataset ID	Dataset Name	Instrument File	Sequence File	Parent Folder
0	bioreactor_rumen1_r1_r_31jul11_jaguar_11-07-17	bioreactor_rumen1_r1_r_31jul11_jaguar_11-07-17		M:\data\proteomics\MAPaper\data\Rumen
1	bioreactor_rumen1_r2_31jul11_jaguar_11-07-19	bioreactor_rumen1_r2_31jul11_jaguar_11-07-19		M:\data\proteomics\MAPaper\data\Rumen
2	bioreactor_rumen1_r3_31jul11_jaguar_11-07-18	bioreactor_rumen1_r3_31jul11_jaguar_11-07-18		M:\data\proteomics\MAPaper\data\Rumen
3	bioreactor_rumen2_r1_31jul11_jaguar_11-07-19	bioreactor_rumen2_r1_31jul11_jaguar_11-07-19		M:\data\proteomics\MAPaper\data\Rumen
4	bioreactor_rumen2_r2_31jul11_jaguar_11-07-17	bioreactor_rumen2_r2_31jul11_jaguar_11-07-17		M:\data\proteomics\MAPaper\data\Rumen
5	bioreactor_rumen2_r3_31jul11_jaguar_11-07-17	bioreactor_rumen2_r3_31jul11_jaguar_11-07-17		M:\data\proteomics\MAPaper\data\Rumen
6	bioreactor_rumen3_r1_31jul11_jaguar_11-07-19	bioreactor_rumen3_r1_31jul11_jaguar_11-07-19		M:\data\proteomics\MAPaper\data\Rumen
7	bioreactor_rumen3_r2_31jul11_jaguar_11-07-19	bioreactor_rumen3_r2_31jul11_jaguar_11-07-19		M:\data\proteomics\MAPaper\data\Rumen
8	bioreactor_rumen3_r3_31jul11_jaguar_11-07-17	bioreactor_rumen3_r3_31jul11_jaguar_11-07-17		M:\data\proteomics\MAPaper\data\Rumen
9	bioreactor_rumen4_r1_31jul11_jaguar_11-07-17	bioreactor_rumen4_r1_31jul11_jaguar_11-07-17		M:\data\proteomics\MAPaper\data\Rumen
10	bioreactor_rumen4_r2_31jul11_jaguar_11-07-19	bioreactor_rumen4_r2_31jul11_jaguar_11-07-19		M:\data\proteomics\MAPaper\data\Rumen
11	bioreactor_rumen4_r3_31jul11_jaguar_11-07-18	bioreactor_rumen4_r3_31jul11_jaguar_11-07-18		M:\data\proteomics\MAPaper\data\Rumen
12	bioreactor_rumen5_r1_31jul11_jaguar_11-07-18	bioreactor_rumen5_r1_31jul11_jaguar_11-07-18		M:\data\proteomics\MAPaper\data\Rumen

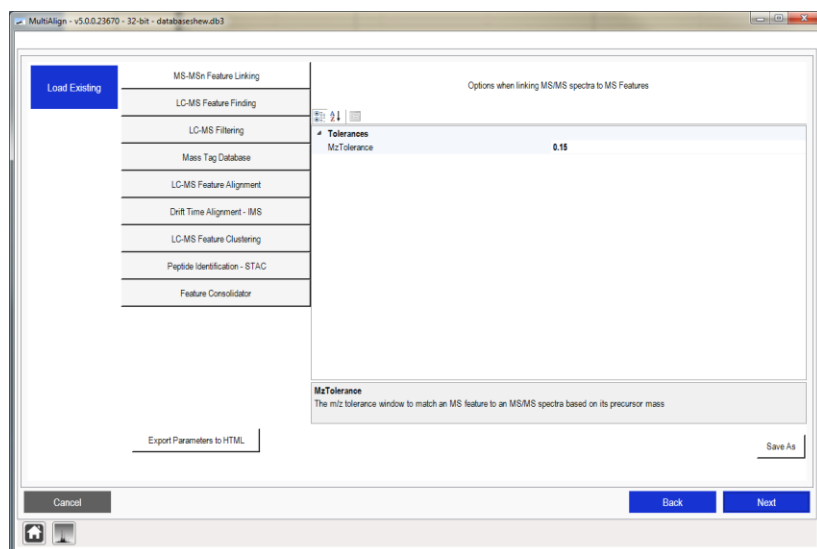
Cancel Back Next

Creating new analysis.



Parameter setup page

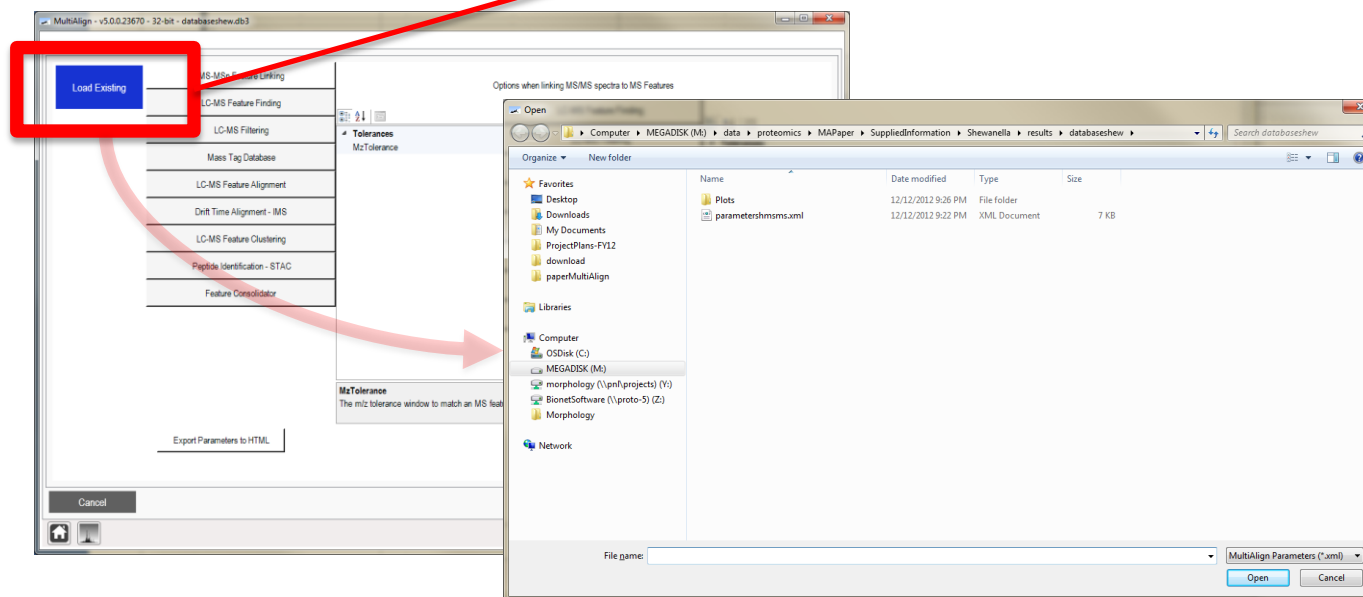
- ▶ The parameter setup page allows you to customize the algorithms par





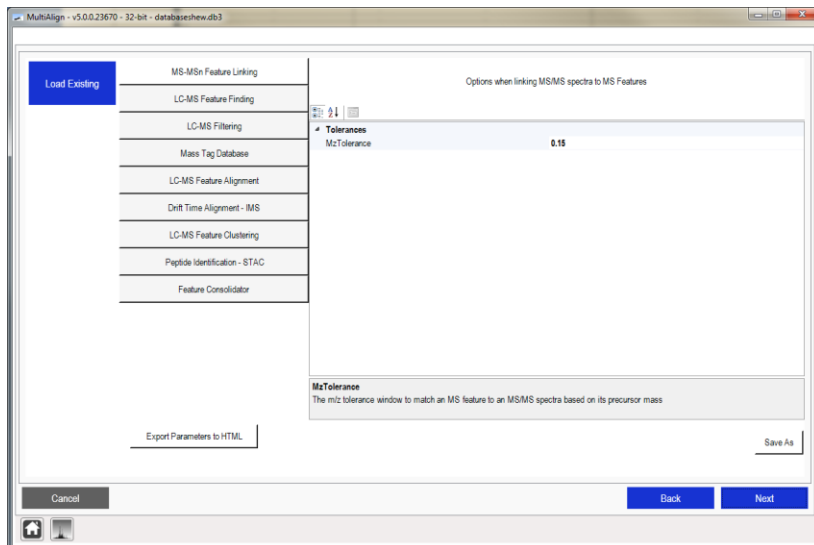
Loading Existing Parameter File

- You can load an existing parameter file by clicking the blue “Load Existing” button.

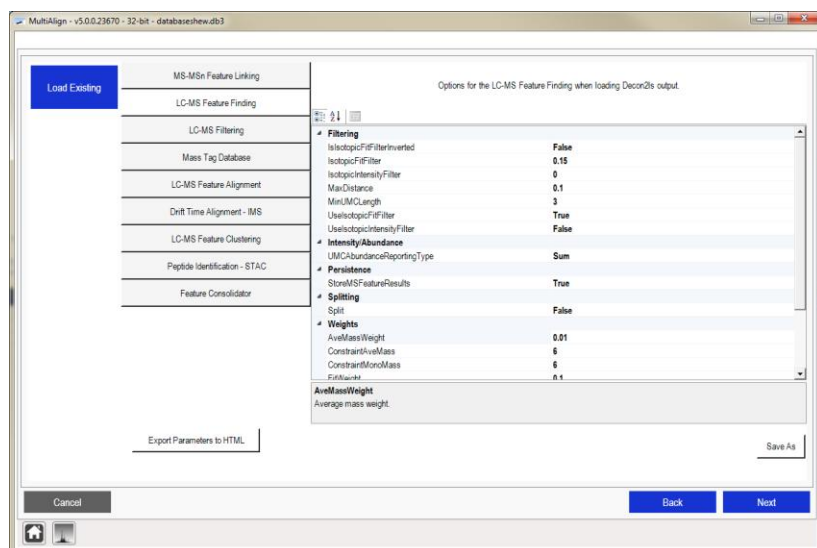




MS/MSn Linking Parameters



- ▶ These parameters specify the tolerance to match a precursor m/z of a deisotoped feature in the parent scan to an MS/MS spectra.
- ▶ Specify the tolerance based on your instrument resolution.
- ▶ Default is
.15 m/z



► These parameters define how to group MS-features, i.e. a feature that is eluting over a number of scans.

► MS-Feature Filtering

■ Isotopic Fit Score

■ Abundance Fit Score

■ Should Invert Isotopic Fit Score

● = True (filter things with low fit scores)

● = False (filter things with low fit score)

► Weights

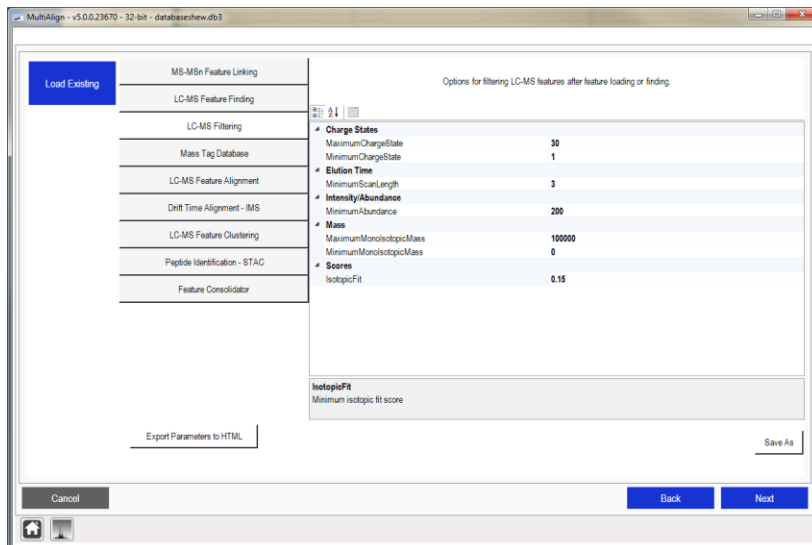
■ Weights used in distance calculation for grouping features



LC-MS Feature Filtering

► These parameters specify how to filter the features.

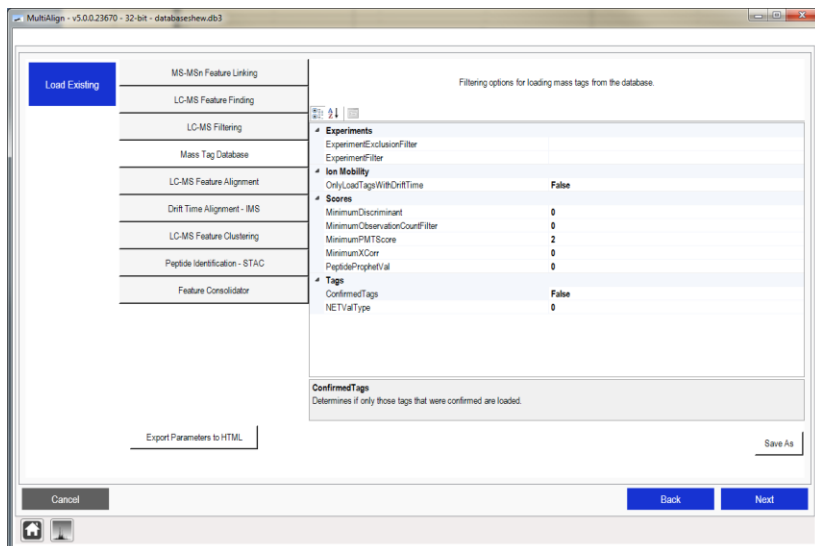
- Minimum LC-Scan Range
- Minimum Abundance cut-offs
- Isotopic Fit Scores





Mass Tag Database Parameters

- ▶ These parameters specify how to filter the mass tag database.
- ▶ Filters map to columns in Mass Tag Database



The screenshot shows the 'Mass Tag Database' section of the MultiAlign software interface. The window title is 'MultiAlign - v5.0.0.23670 - 32-bit - databasenew.db3'. The left sidebar contains a list of options: 'Load Existing' (highlighted in blue), 'MS-MSn Feature Linking', 'LC-MS Feature Finding', 'LC-MS Filtering', 'Mass Tag Database' (selected), 'LC-MS Feature Alignment', 'Drift Time Alignment - RMS', 'LC-MS Feature Clustering', 'Peptide Identification - STAG', and 'Feature Consolidator'. The main area is titled 'Filtering options for loading mass tags from the database.' and contains a table of filters. The table has two columns: the filter name and its value. The filters are grouped into sections: 'Experiments' (ExperimentExclusionFilter, ExperimentFilter), 'Ion Mobility' (OnlyLoadTagsWithDriftTime), 'Scores' (MinimumDiscriminant, MinimumObservationCountFilter, MinimumPMFScore, MinimumXCorr, PeptideProphetVal), and 'Tags' (ConfirmedTags, NETValType). The 'ConfirmedTags' filter is set to 'False' and 'NETValType' is set to '0'. Below the table, there is a section for 'ConfirmedTags' with a description: 'Determines if only those tags that were confirmed are loaded.' At the bottom of the window, there are buttons for 'Cancel', 'Back', and 'Next', and a 'Save As' button.

Filter	Value
Experiments	
ExperimentExclusionFilter	
ExperimentFilter	
Ion Mobility	
OnlyLoadTagsWithDriftTime	False
Scores	
MinimumDiscriminant	0
MinimumObservationCountFilter	0
MinimumPMFScore	2
MinimumXCorr	0
PeptideProphetVal	0
Tags	
ConfirmedTags	False
NETValType	0

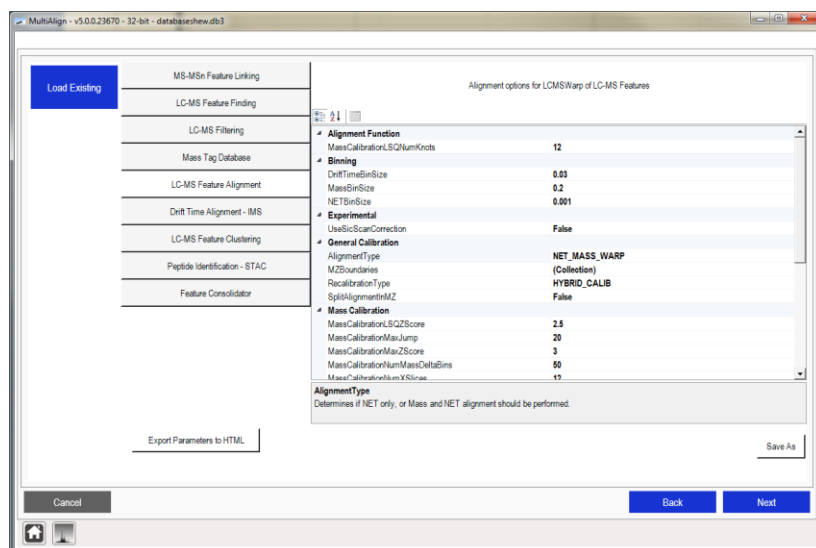
ConfirmedTags
Determines if only those tags that were confirmed are loaded.

Export Parameters to HTML

Cancel Back Next Save As



Alignment

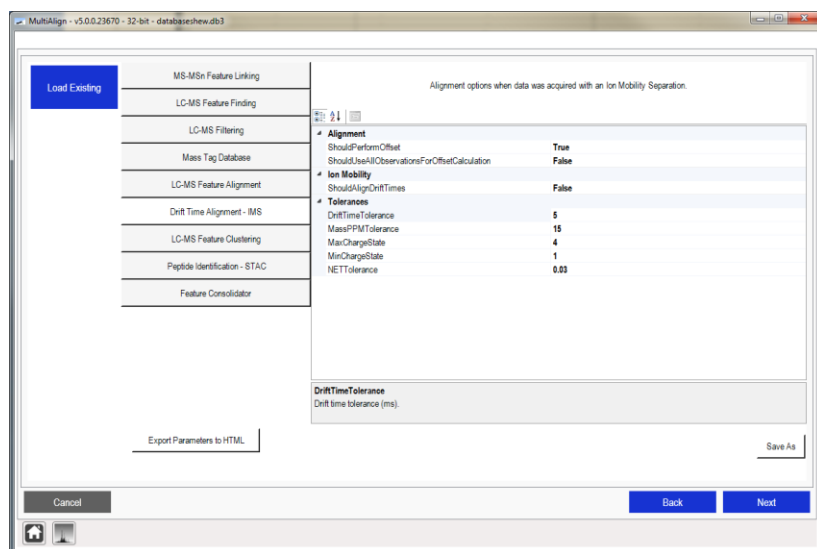


- ▶ LC-MS feature finding parameters specify weights, calibration types, and other parameters for the LCMSWarp algorithm.
- ▶ These parameters do not need to be changed.



Drift Time Alignment Parameters

- ▶ If analyzing Ion Mobility data (LC-IMS-MS) this section could be used to align drift times.





LC-MS Feature Clustering

► These parameters are used to specify how to cluster features across datasets.

■ Clustering Algorithm

■ Cluster Centroid Representation

■ Clustering Tolerances

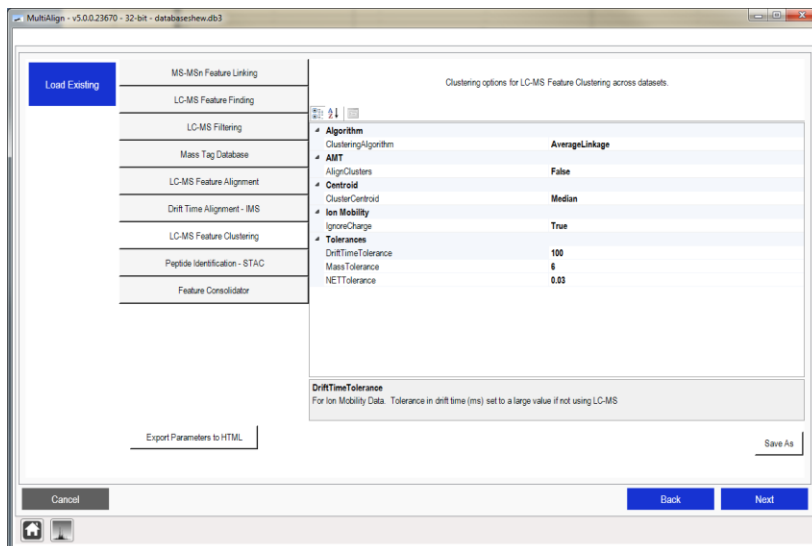
● Monoisotopic Mass

● NET

● Drift Time

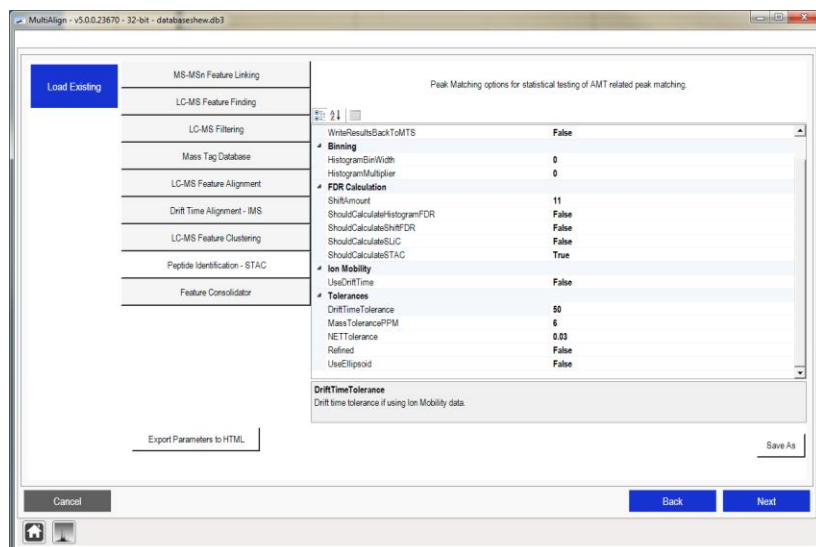
■ Ignore Drift Time

● Set to false if analyzing Ion Mobility data





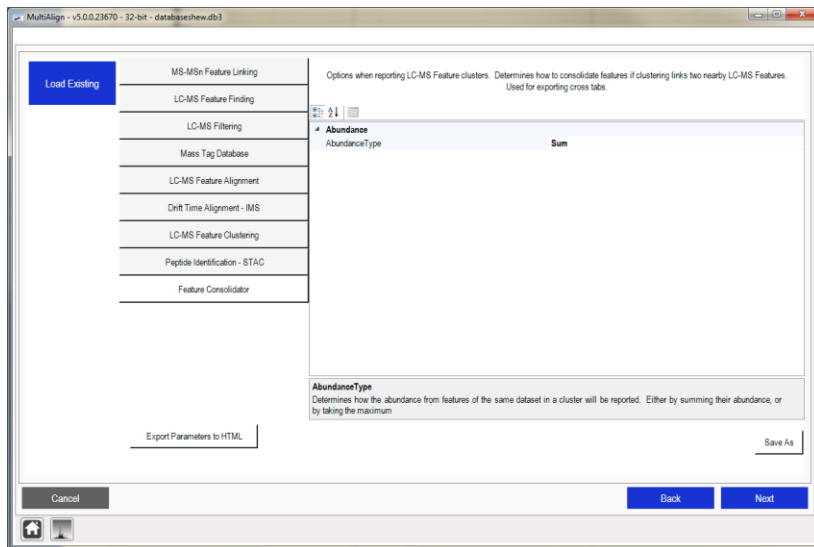
Peptide Identification – STAC parameters



- ▶ These parameters are intended for the STAC algorithm.
- ▶ Change the tolerances parameters based on your instrument resolution
 - Mass (PPM)
 - Drift Time (Ion Mobility)
 - 3
- ▶ You can leave the NET tolerance at .03



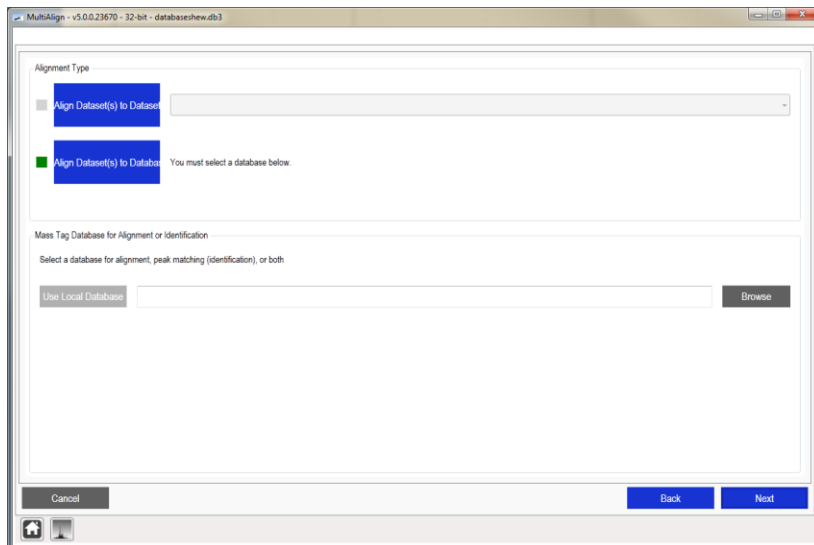
Feature Consolidator parameters



- ▶ These parameters define how to consolidate features that may have been improperly split during the LC-MS feature finding stage.
- ▶ The Abundance Type parameter specifies whether to sum or use the maximum abundance value if two features from the same dataset are clustered together after alignment.
 - Suggested value = Sum



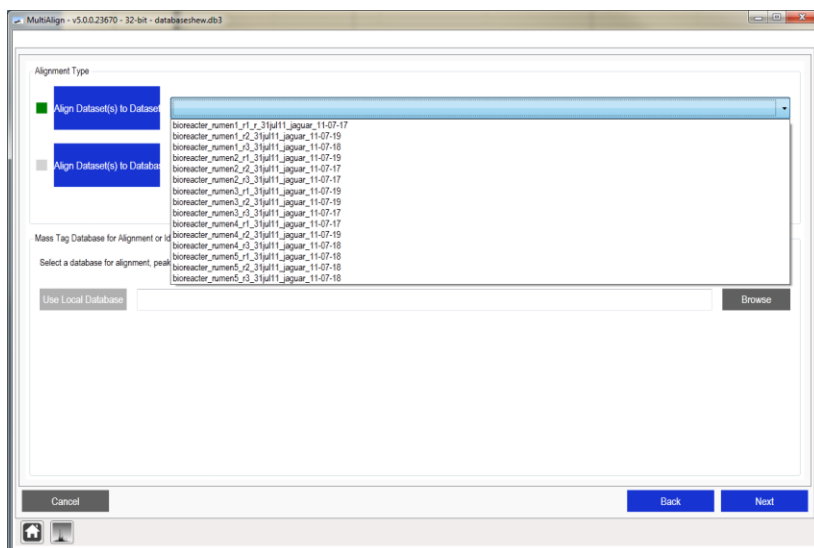
Selecting a baseline



- ▶ Once you have setup the parameters, a baseline must be selected.
- ▶ Two types of baselines can be used:
 - Mass Tag Database
 - LC-MS dataset
- ▶ *NOTE: If you select a mass tag database for alignment, MultiAlign will automatically perform STAC for peptide identification*

Selecting a dataset as a baseline

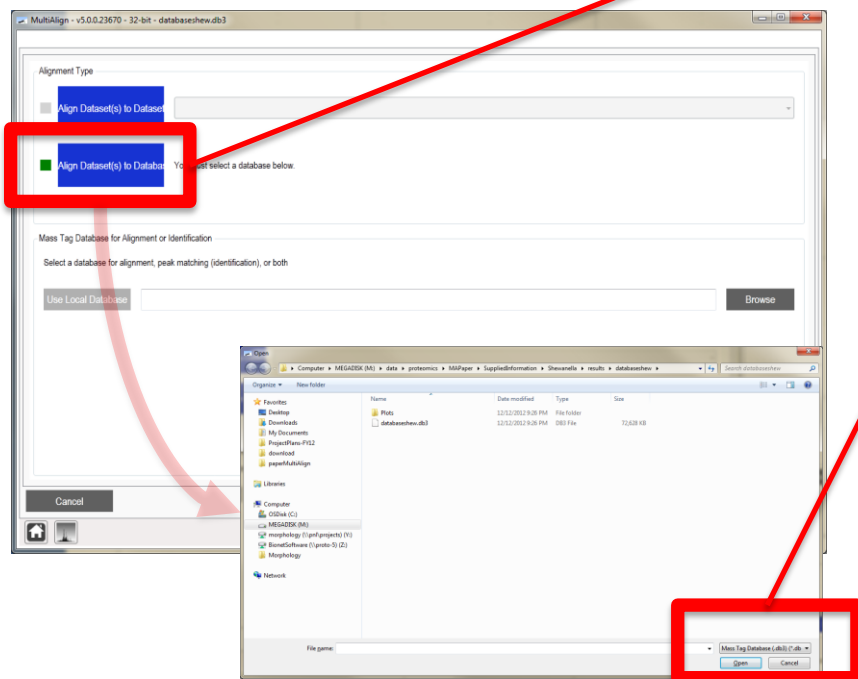
- ▶ Click the blue button “Align Dataset(s) to Dataset”
- ▶ Select the dataset from the drop down list





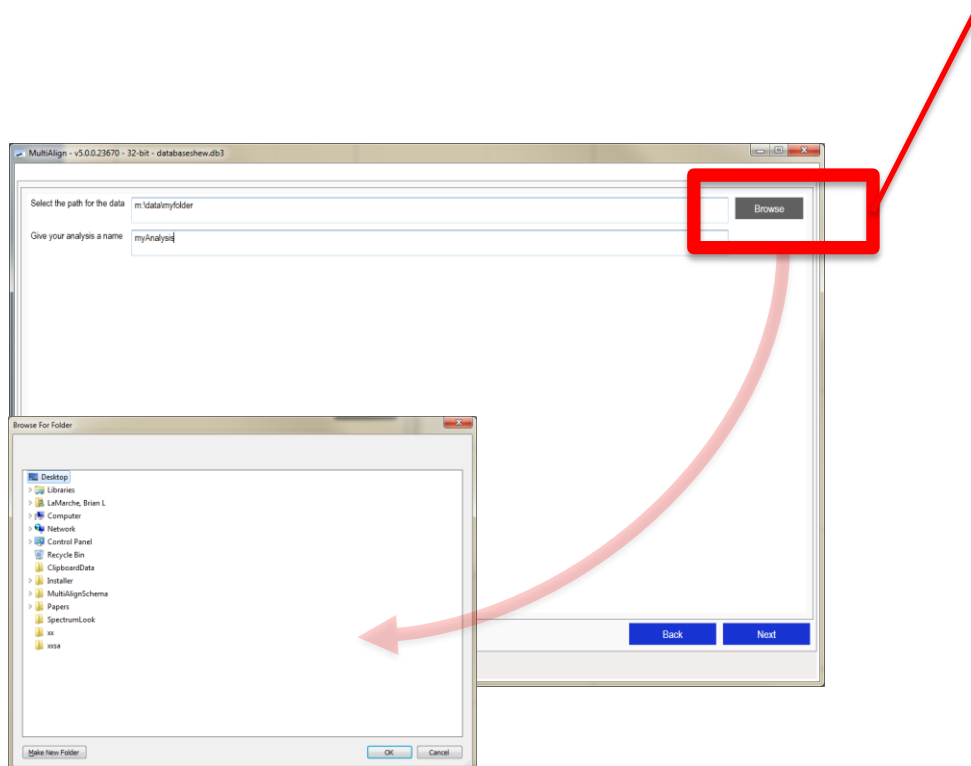
Selecting a database as a baseline

- ▶ Click the blue button “Align Dataset(s) to Database”
- ▶ Click the gray “Browse” button.
- ▶ Find the local mass tag database file on your computer.
- ▶ Choose a Mass Tag database format:
 - APE Cache Database (ape)
 - MTDB Created SQLite database (db3)





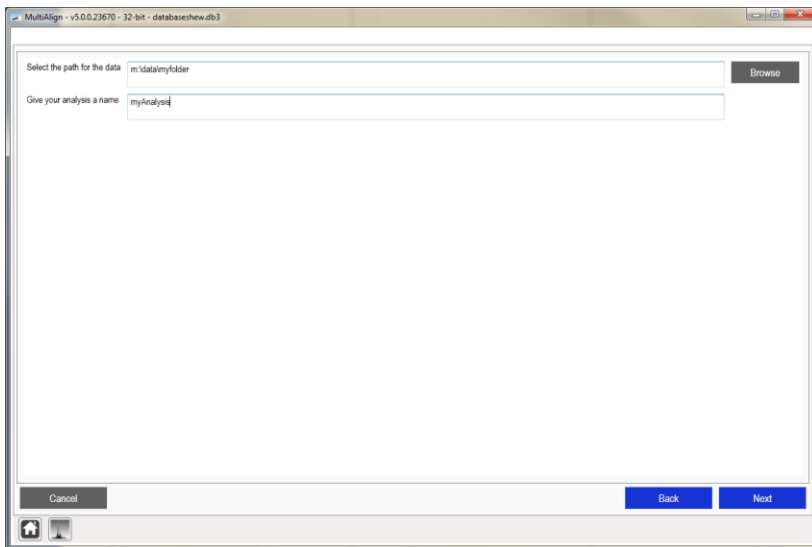
Select an output path





Select an analysis name

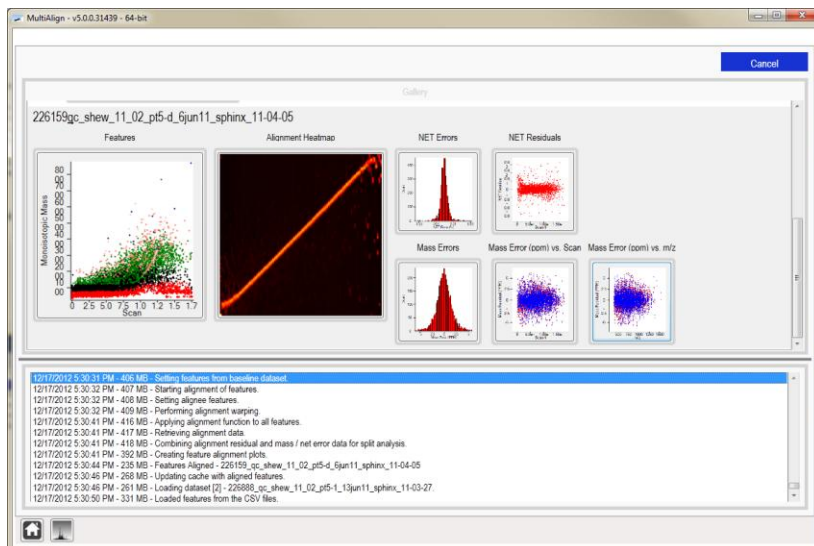
- ▶ Provide a name for the analysis
 - All files that are created (log files, output cross tab files, and the result database will have be prefixed with this name





Running Analysis Window

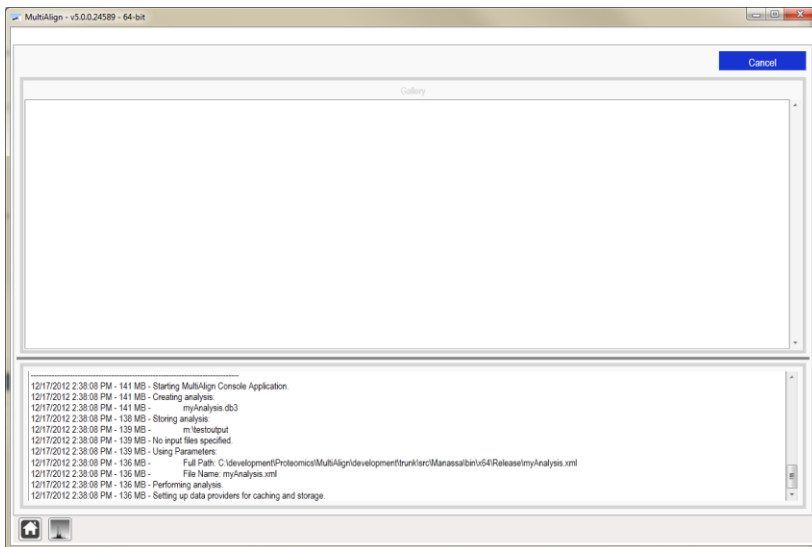
- ▶ The Analysis window will be displayed as shown.
- ▶ The bottom part of the window displays all messages the MultiAlign algorithmic back end puts out.
- ▶ The top part of the window displays images produced during various steps of the analysis.





Need to cancel?

- ▶ To stop the analysis, click the blue “Cancel” button on the top right of the screen.

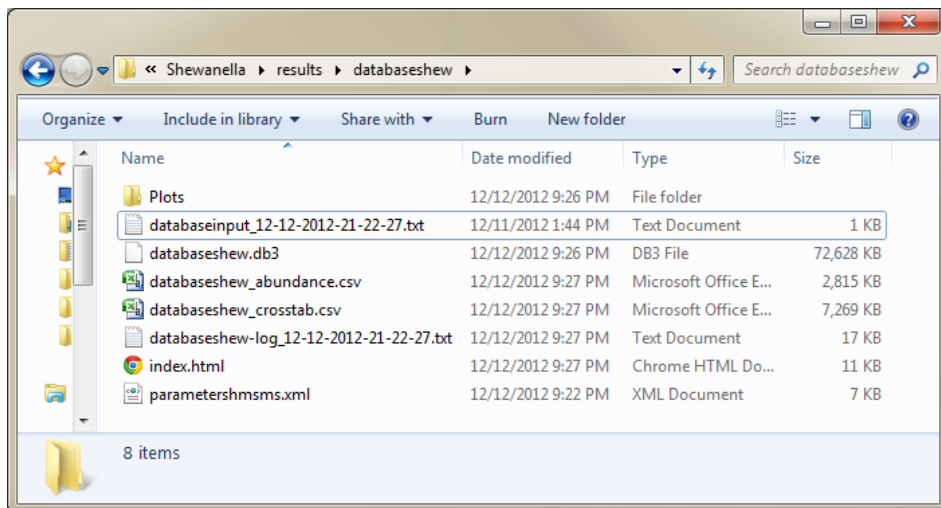




Output Files

- ▶ You can navigate to the folder using Windows Explorer to find the data as the analysis is running.
- ▶ This is a list of files that MultiAlign will generate during the analysis.

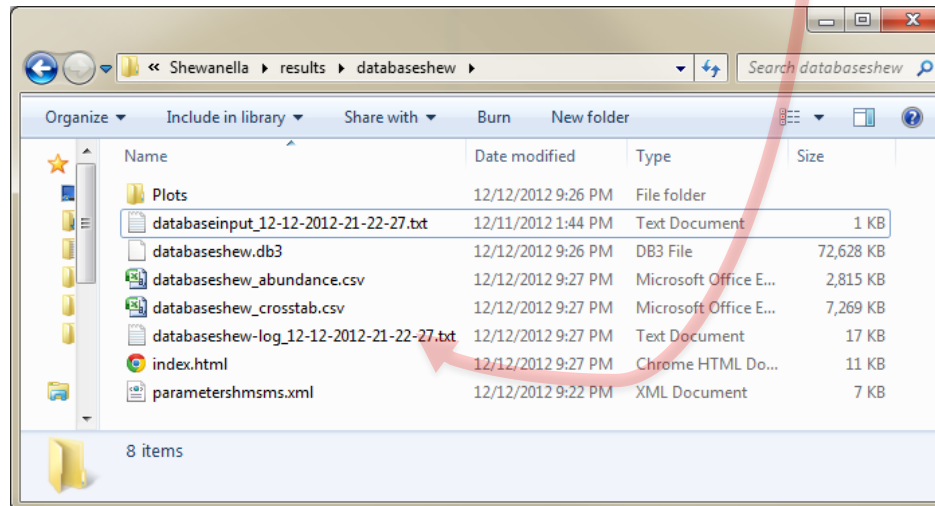
- ▶ Plots
 - Directory containing data
- ▶ Input file
 - A reconstructed input file that could be used to run the console application
- ▶ Database
 - A SQLite formatted database that links all data from raw spectra through clusters of features across datasets, and mass tag identifications (linked to proteins if using a protein MTDB)
- ▶ Cross tabs
 - A set of cross tabs that have data useful for downstream analysis
- ▶ Log
 - A log file with the name of the analysis, date and time the analysis was started.
- ▶ Parameter file
 - Saved parameter file so you can repeat the analysis again
- ▶ HTML summary report
 - Shows all analysis plots and synopsis of data analysis.



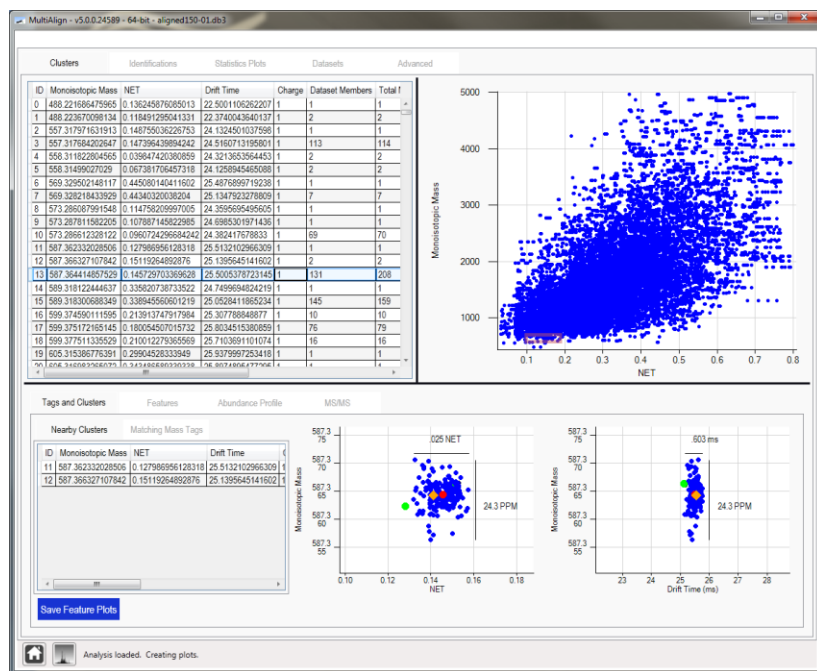


Errors?

- ▶ If you receive an error, and you cannot conclude what the error is, there a log file will give you the specific error, and stack trace that would be useful for our improvement of the tool.



- ▶ After a successful analysis you should see this window



MultiAlign Tutorial Conclusion

- ▶ For more information see the MultiAlign website:

<http://omics.pnl.gov/software/MultiAlign.php>