MultiAlign Tutorial 03 – Running an Analysis

BRIAN LAMARCHE

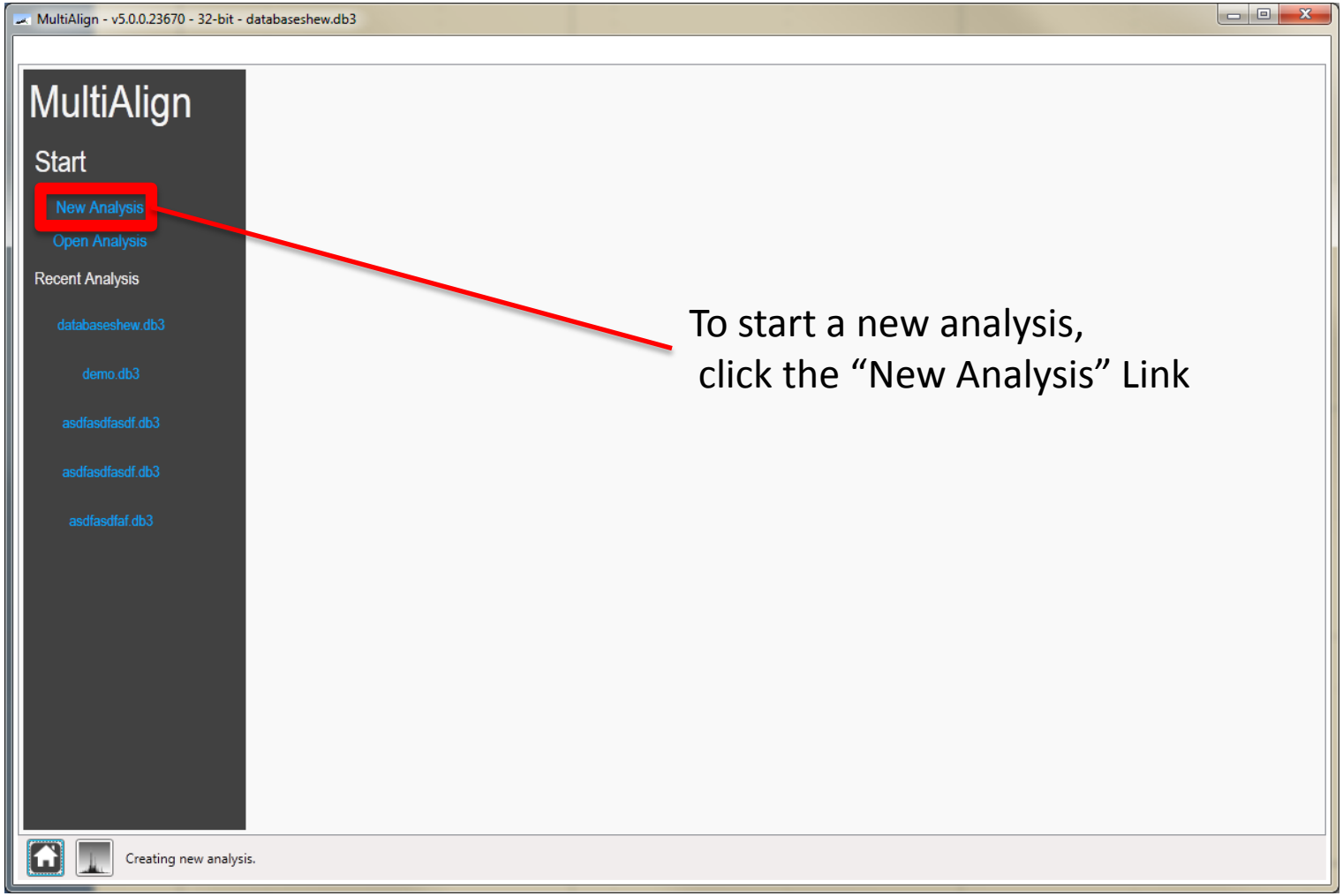
About this tutorial

- 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

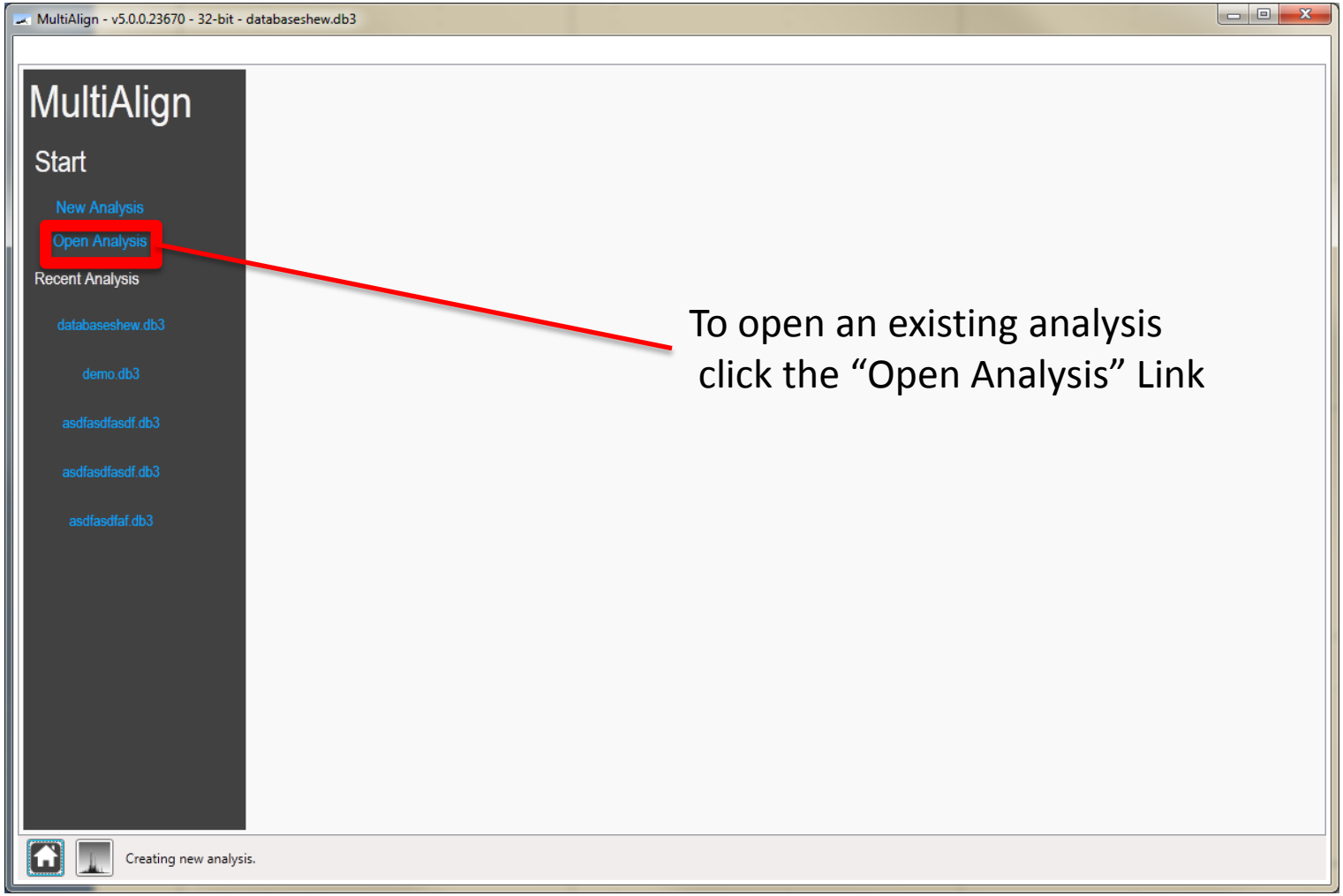
UNDERSTANDING THE SCREENS

GUI Basics



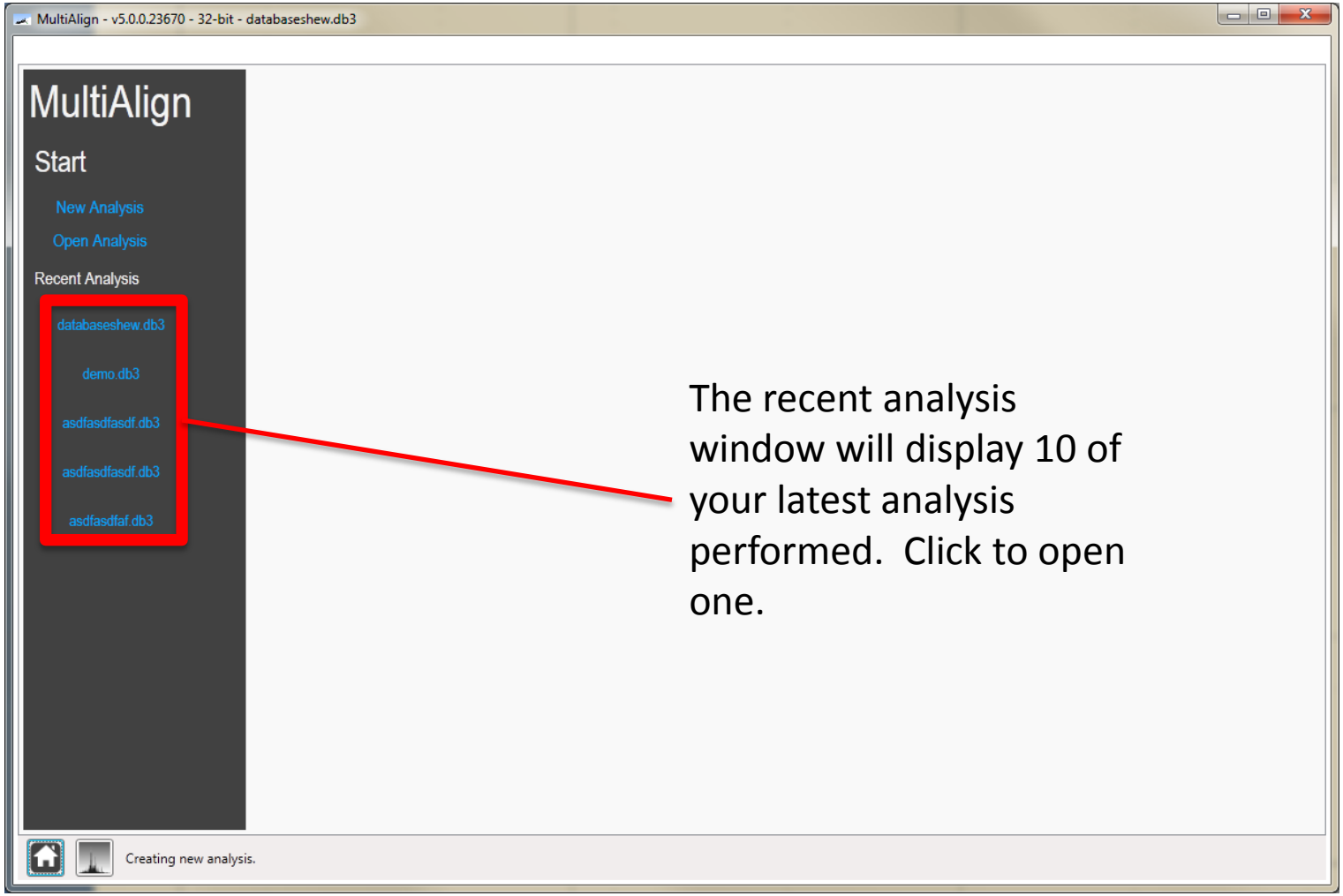
To start a new analysis,
click the “New Analysis” Link

GUI Basics



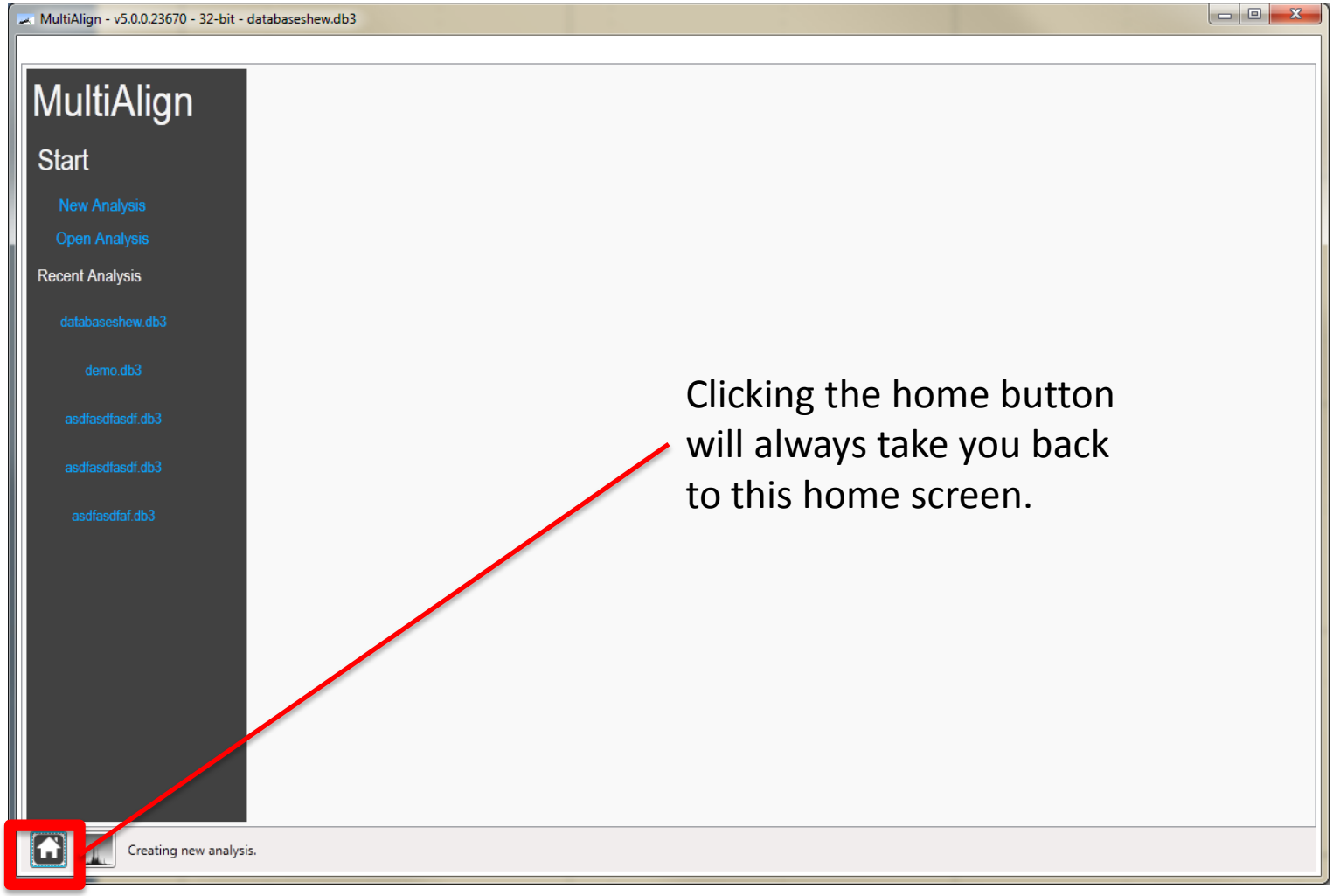
To open an existing analysis click the “Open Analysis” Link

GUI Basics



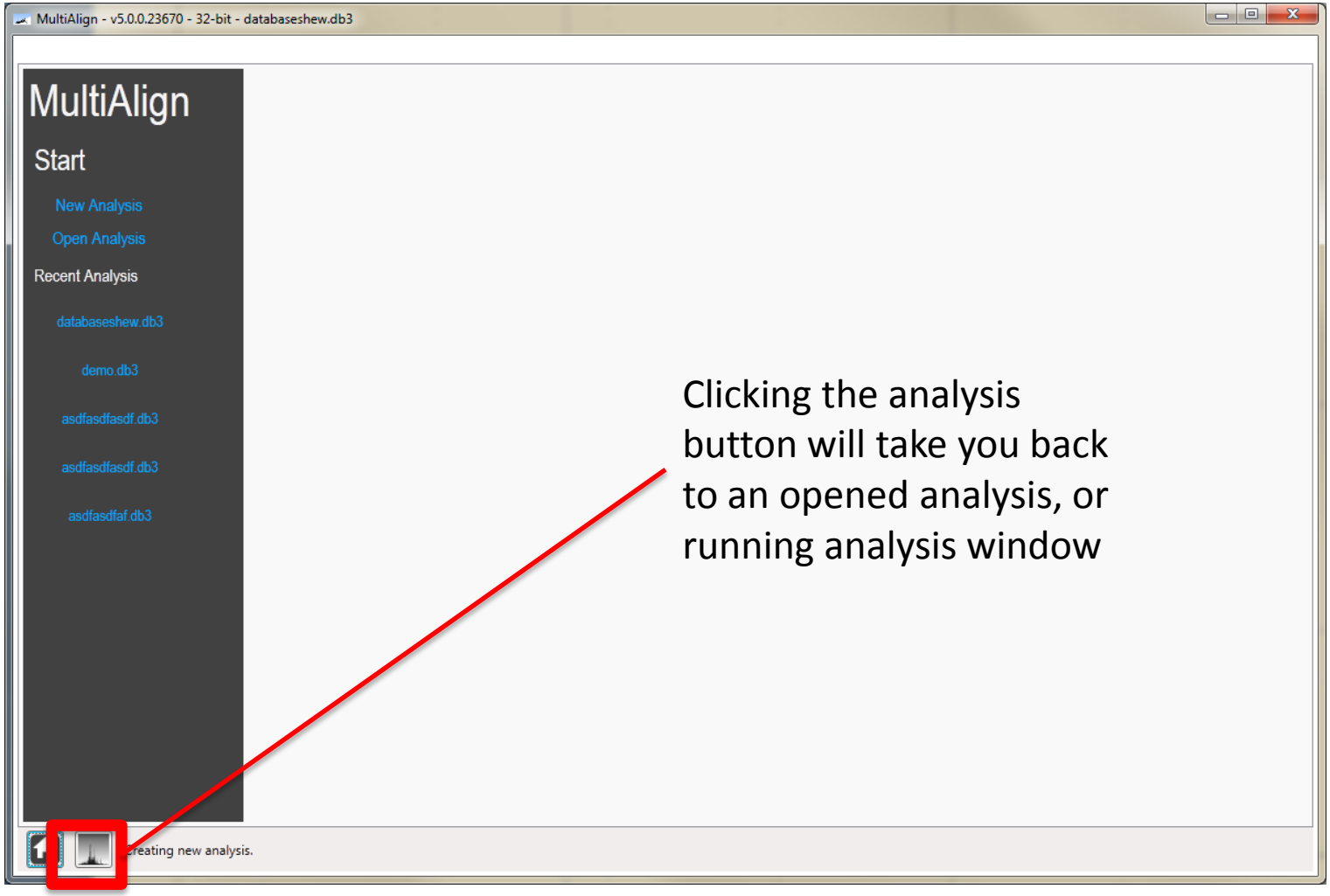
The recent analysis window will display 10 of your latest analysis performed. Click to open one.

GUI Basics



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

GUI Basics



Clicking the analysis button will take you back to an opened analysis, or running analysis window

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

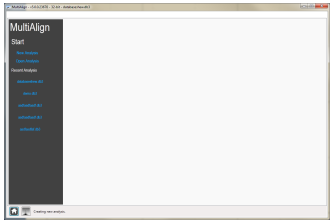
Starting a new analysis

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

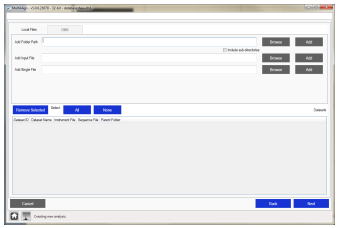
The Analysis is broken down into several steps

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

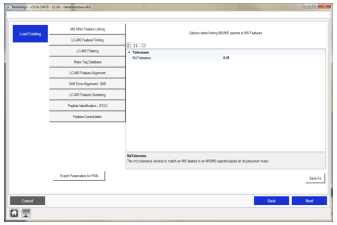
Wizard



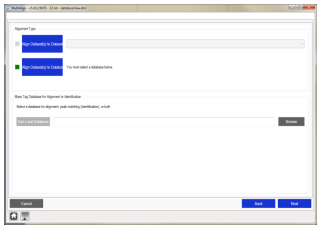
1. Home Screen



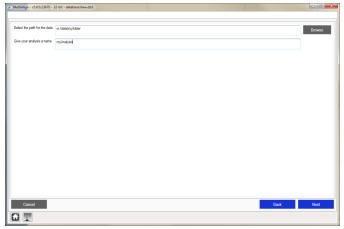
2. Select Data



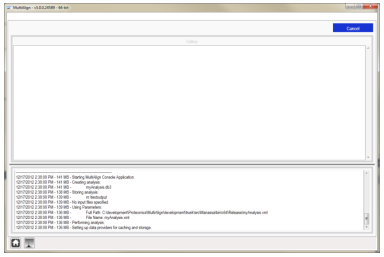
3. Set Parameters



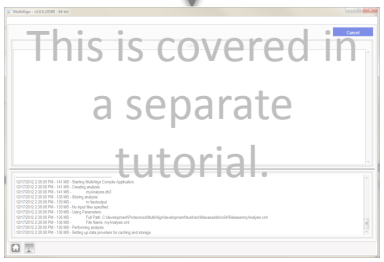
4. Select Baseline and Mass Tag Database



5. Set Analysis Path and Name

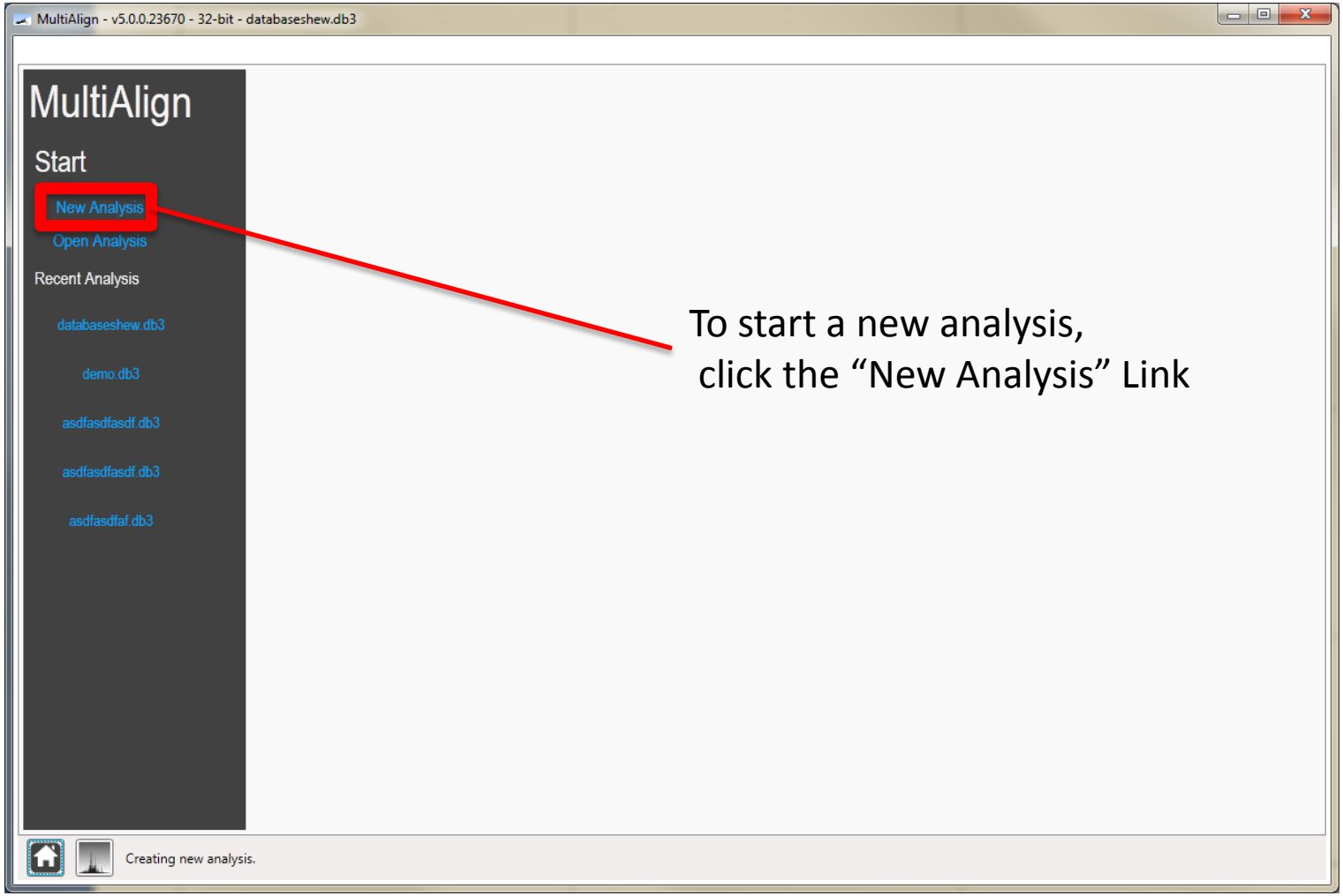


6. Running Analysis



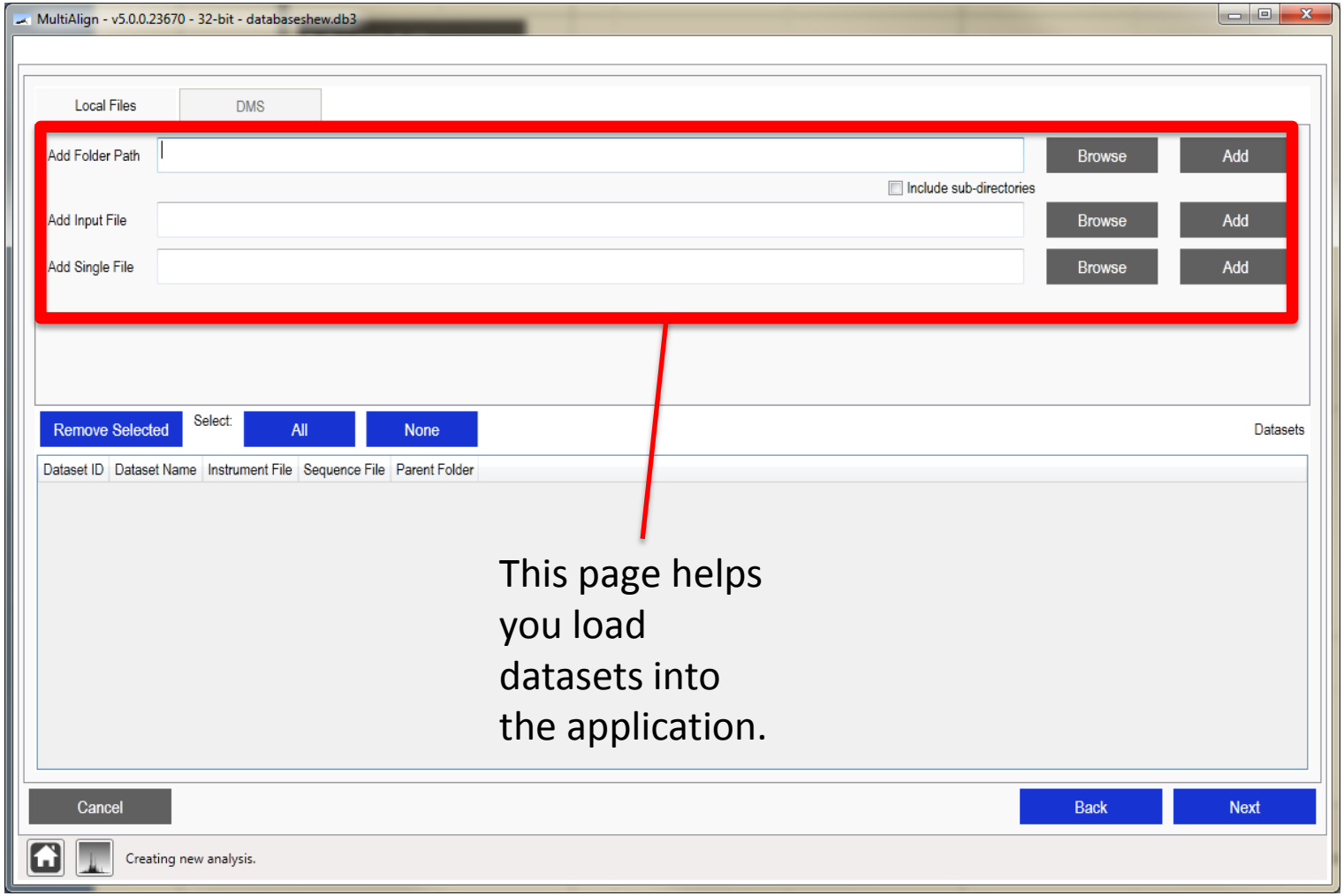
7. Analysis View Window

Starting the analysis



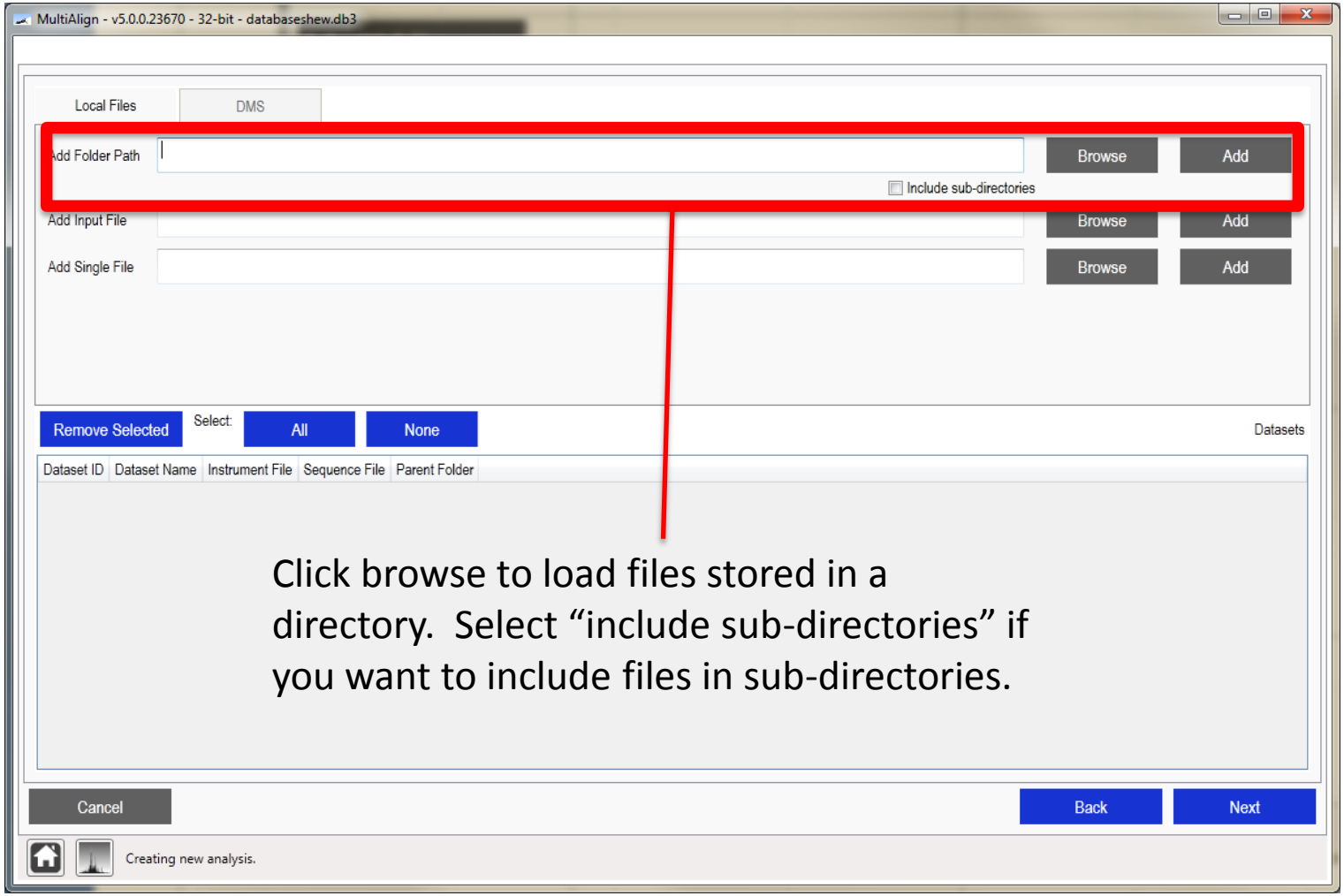
To start a new analysis,
click the “New Analysis” Link

Load Data Page



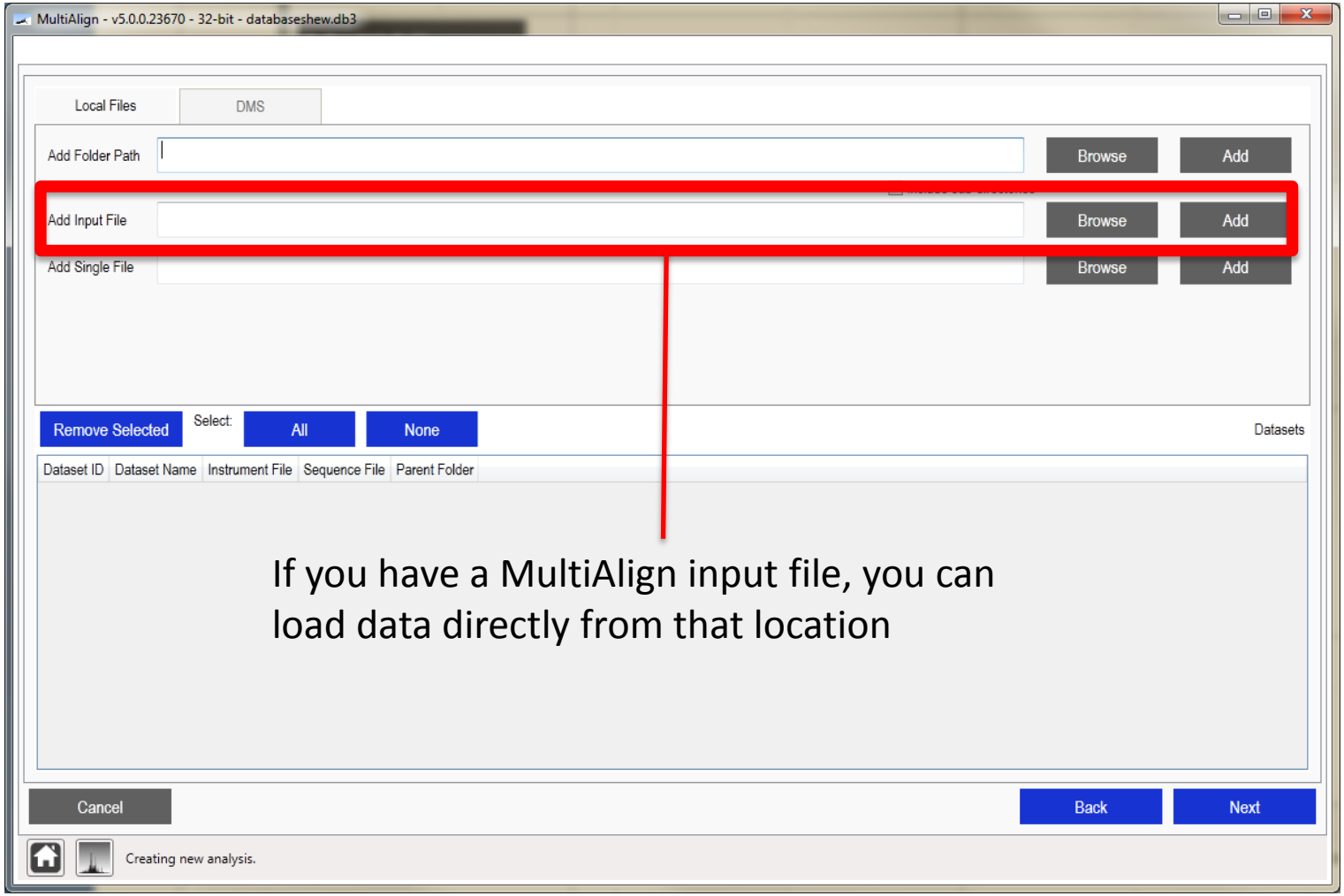
This page helps you load datasets into the application.

Load data from directory



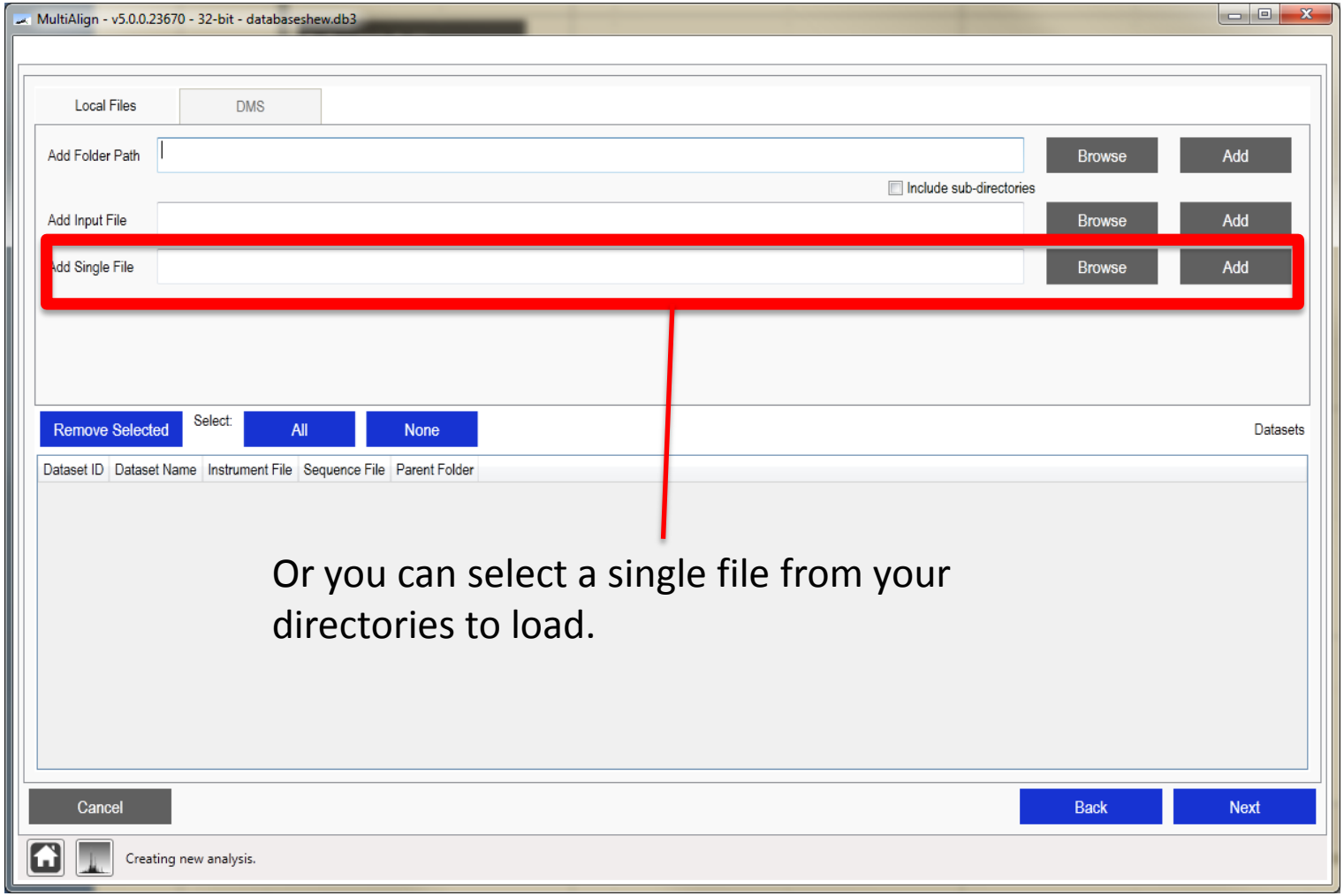
Click browse to load files stored in a directory. Select “include sub-directories” if you want to include files in sub-directories.

Load from an input file



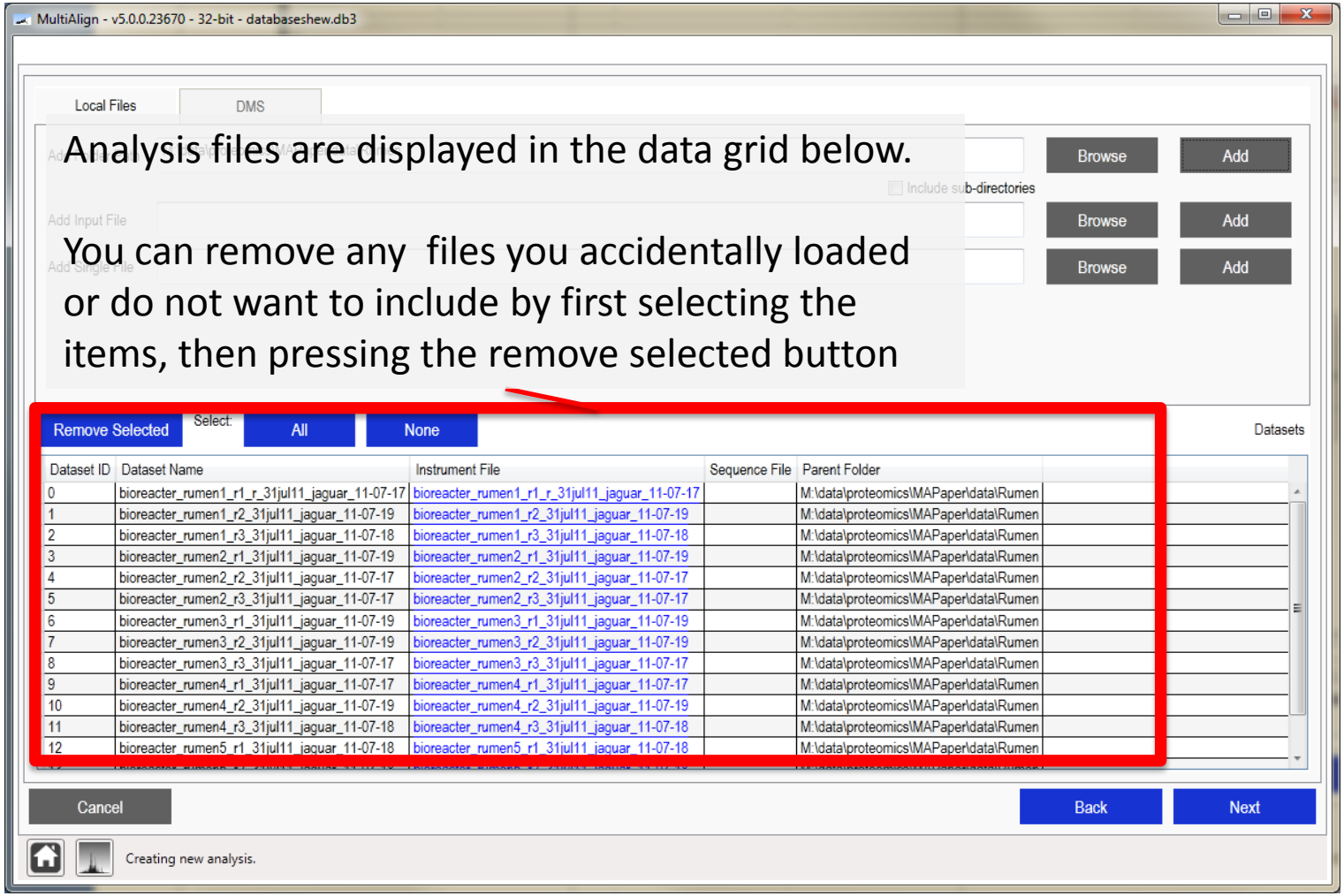
If you have a MultiAlign input file, you can load data directly from that location

Load a single dataset file



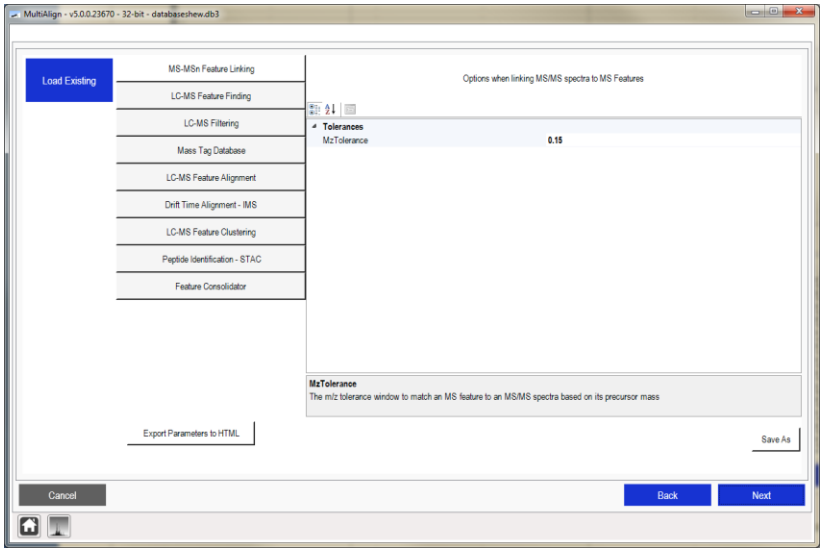
Or you can select a single file from your directories to load.

Load Data Page



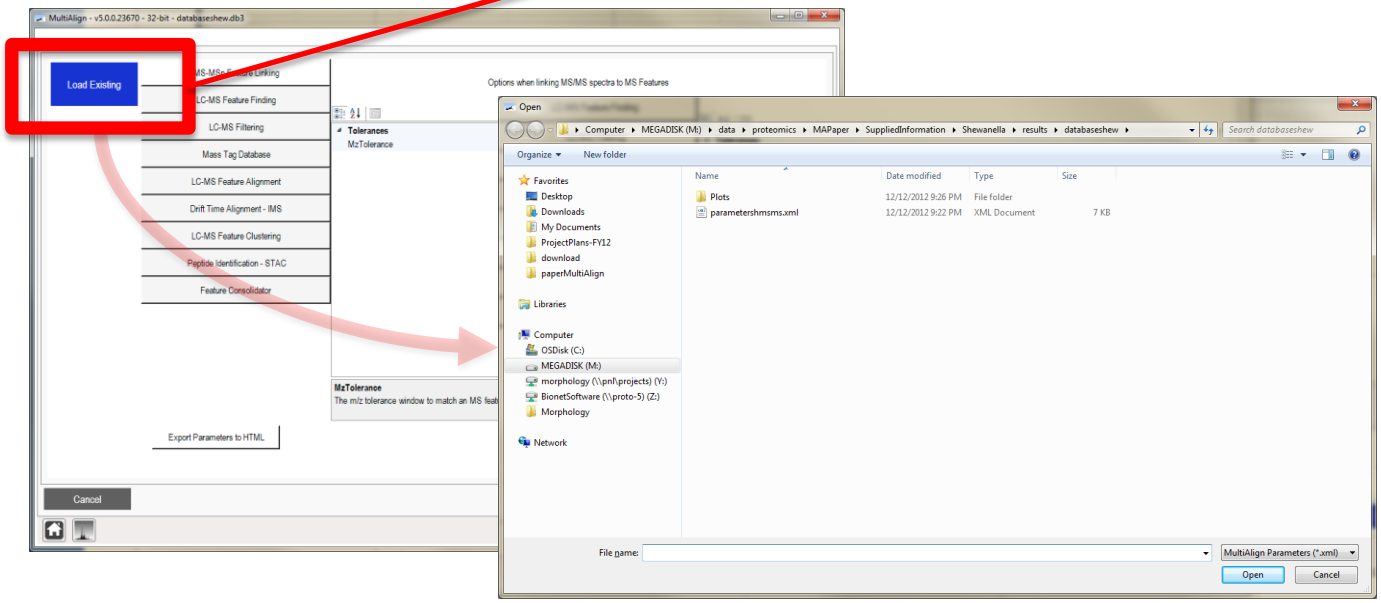
Parameter setup page

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

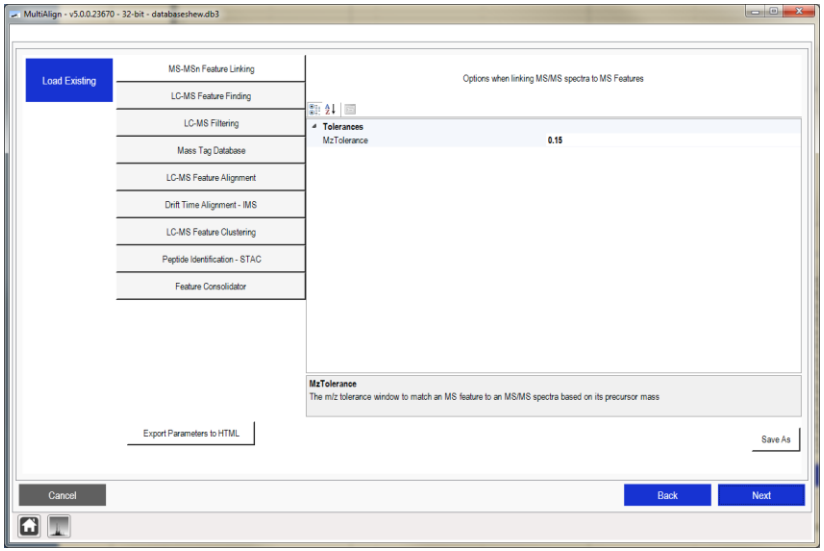


Loading Existing Parameter File

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



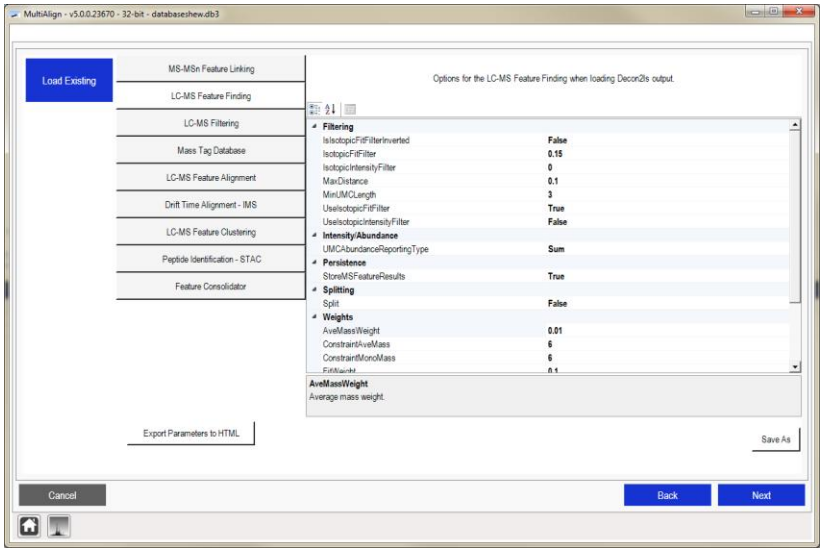
MS/MSn Linking Parameters



- 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

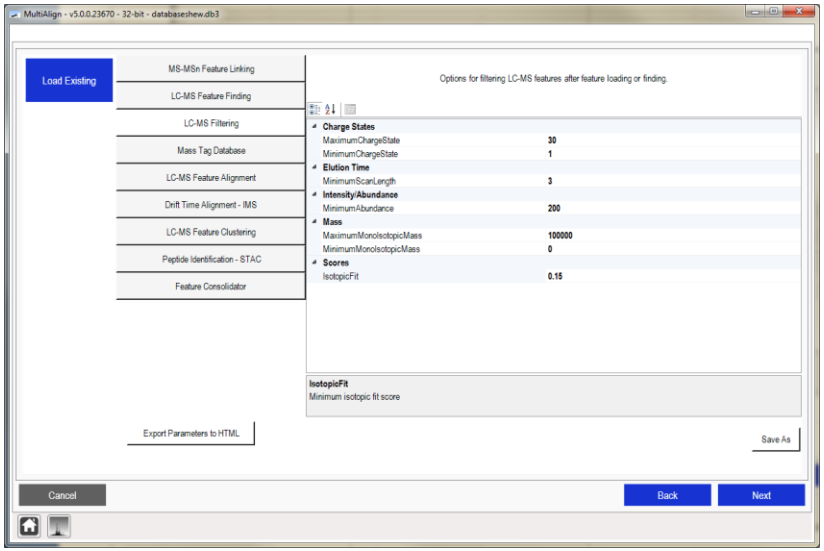
  .15 *m/z*

LC-MS Feature Finding Parameters



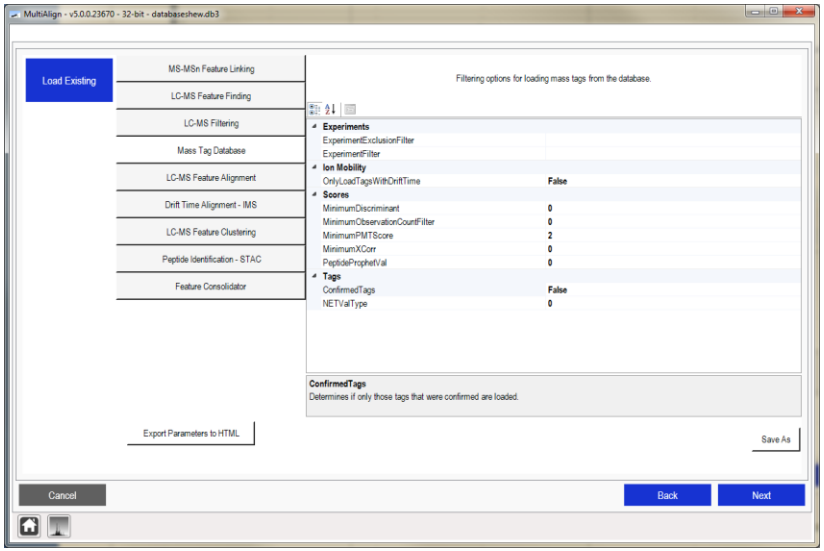
- These parameters define how to group MS-features, i.e. a feature that is eluting over a number of scans.
- MS-Feature Filtering
  - Isotopic Fit Score
  - Abundance Fit Score
  - Should Invert Isotopic Fit Score
    - = True (filter things with low fit scores
    - = False (filter things with low fit score
- Weights
  - Weights used in distance calculation for grouping features

LC-MS Feature Filtering



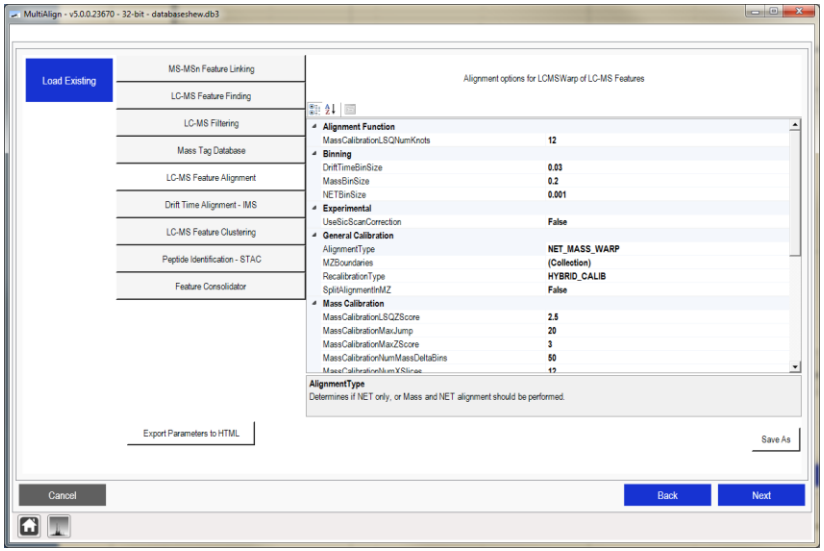
- These parameters specify how to filter the features.
  - Minimum LC-Scan Range
  - Minimum Abundance cut-offs
  - Isotopic Fit Scores

Mass Tag Database Parameters



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

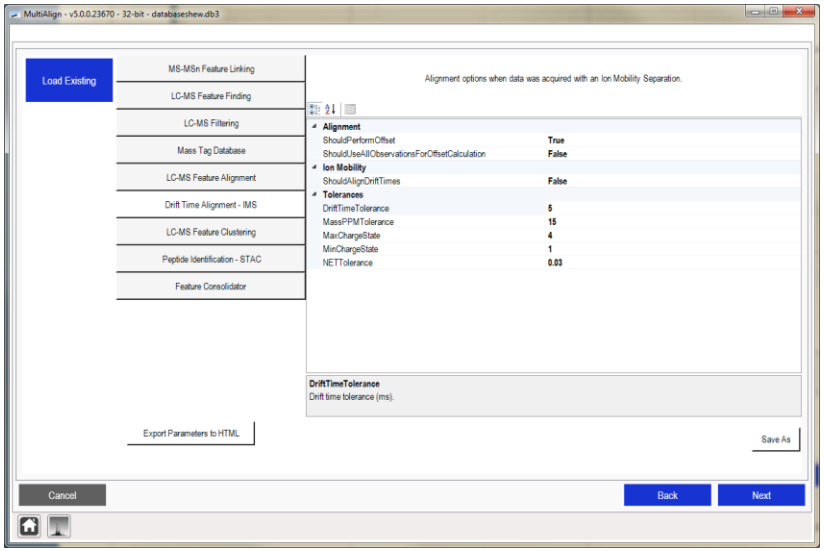
Alignment



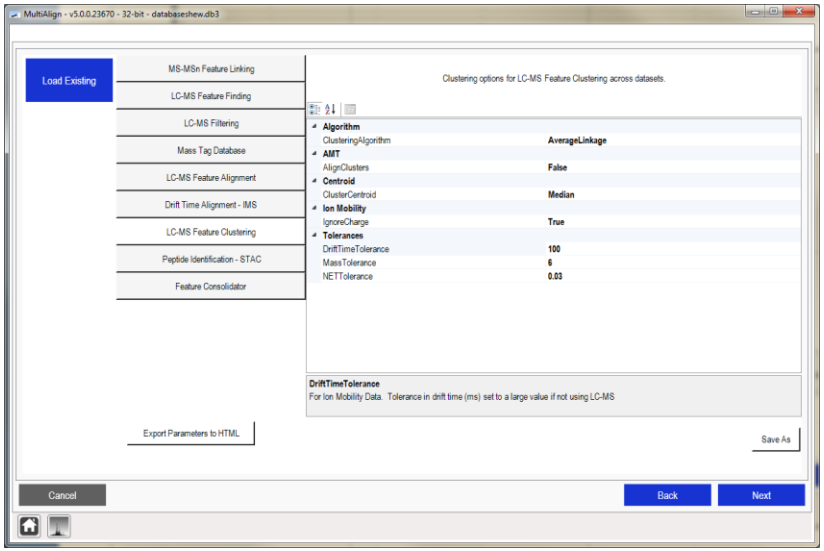
- LC-MS feature finding parameters specify weights, calibration types, and other parameters for the LCMSWarp algorithm.
- These parameters do not need to be changed.

Drift Time Alignment Parameters

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

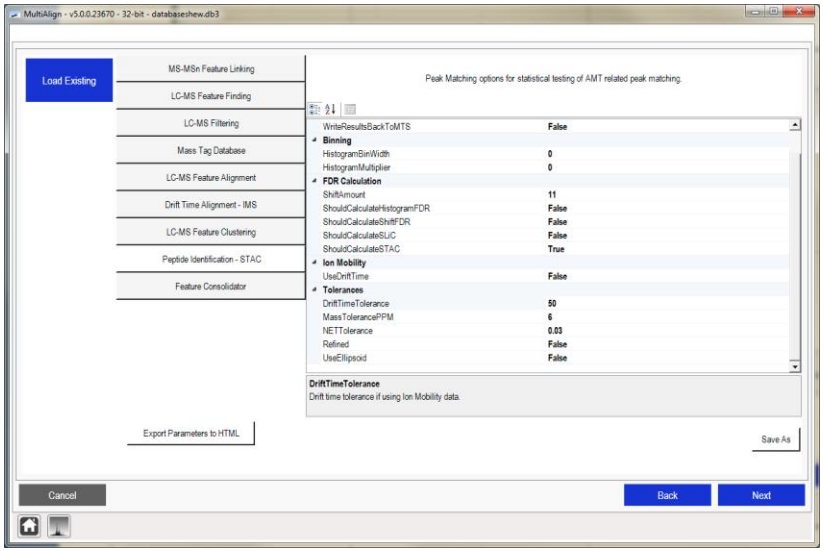


LC-MS Feature Clustering



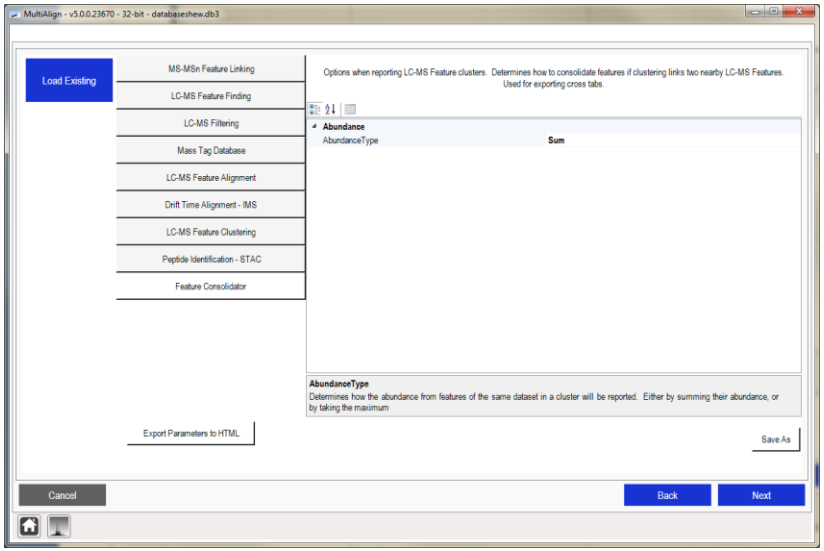
- These parameters are used to specify how to cluster features across datasets.
  - Clustering Algorithm
  - Cluster Centroid Representation
  - Clustering Tolerances
    - Monoisotopic Mass
    - NET
    - Drift Time
  - Ignore Drift Time
    - Set to false if analyzing Ion Mobility data

Peptide Identification – STAC parameters



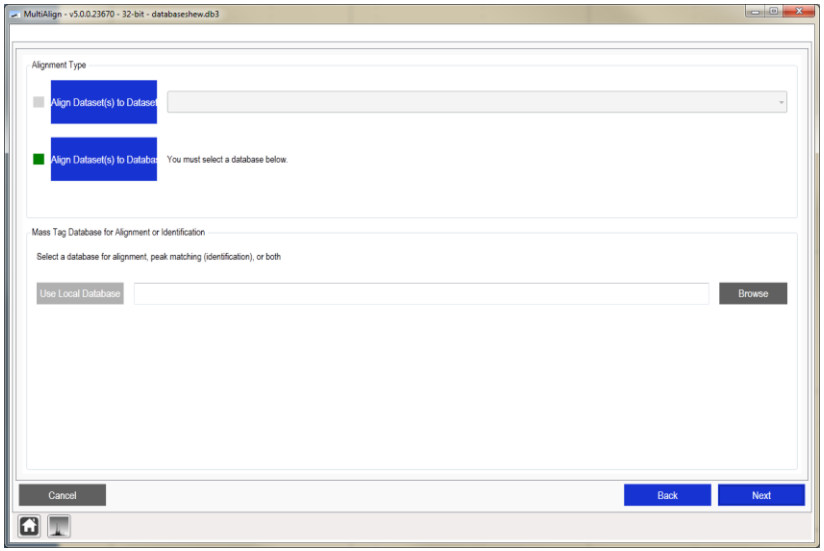
- These parameters are intended for the STAC algorithm.
- Change the tolerances parameters based on your instrument resolution
  - Mass (PPM)
  - Drift Time (Ion Mobility)
    - 3
- You can leave the NET tolerance at .03

Feature Consolidator parameters



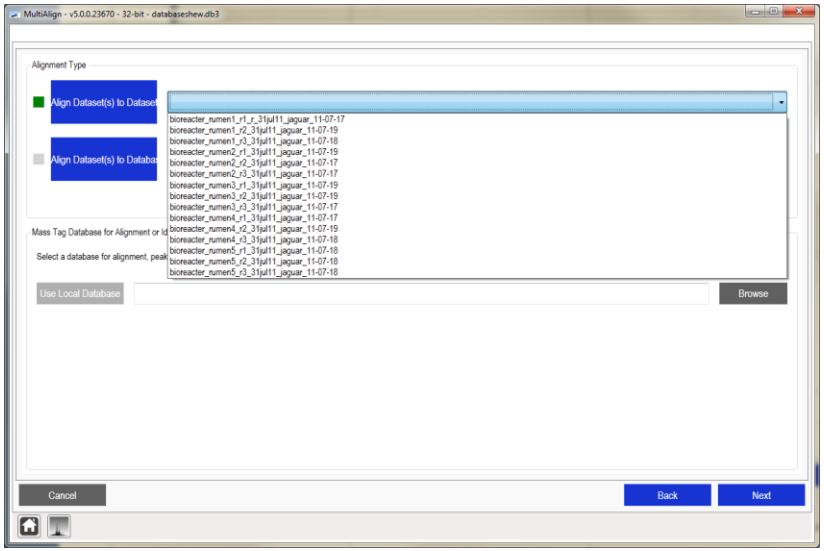
- These parameter define how to consolidate features that may have been improperly split during the LC-MS feature finding stage.
- The Abundance Type parameter specifies whether to sum or use the maximum abundance value if two features from the same dataset are clustered together after alignment.
  -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

Selecting a baseline



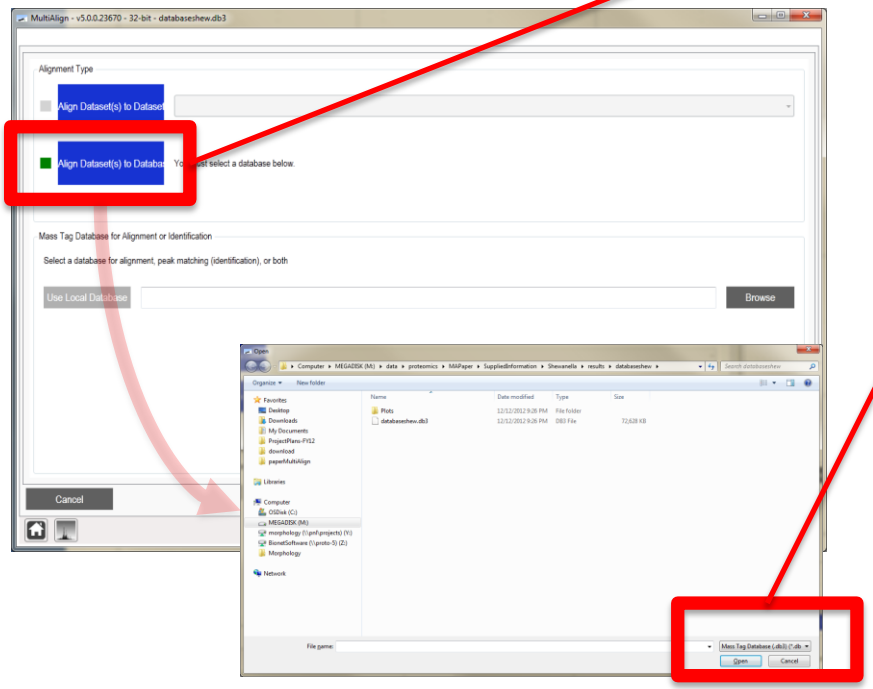
- Once you have setup the parameters, a baseline must be selected.
- Two types of baselines can be used:
  - Mass Tag Database
  - LC-MS dataset
- *NOTE: If you select a mass tag database for alignment, MultiAlign will automatically perform STAC for peptide identification*

Selecting a dataset as a baseline



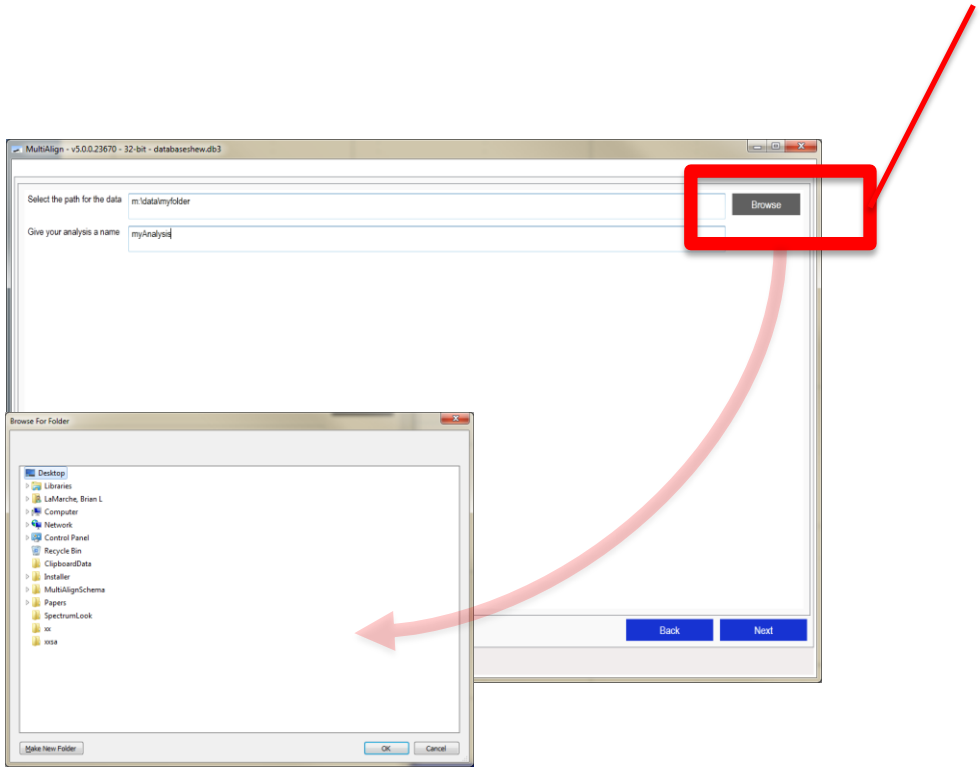
- Click the blue button “Align Dataset(s) to Dataset”
- Select the dataset from the drop down list

Selecting a database as a baseline

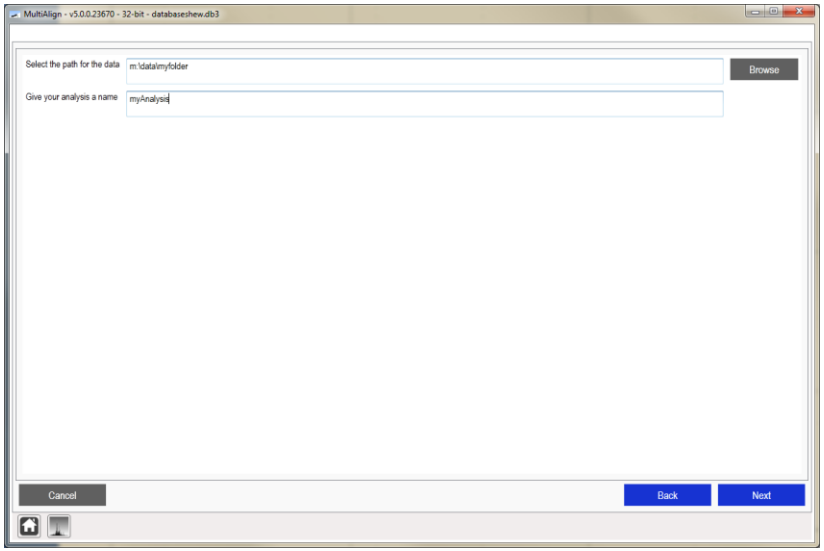


- Click the blue button “Align Dataset(s) to Database”
- Click the gray “Browse” button.
- Find the local mass tag database file on your computer.
- Choose a Mass Tag database format:
  - APE Cache Database (ape)
  - MTDB Created SQLite database (db3)

Select an output path

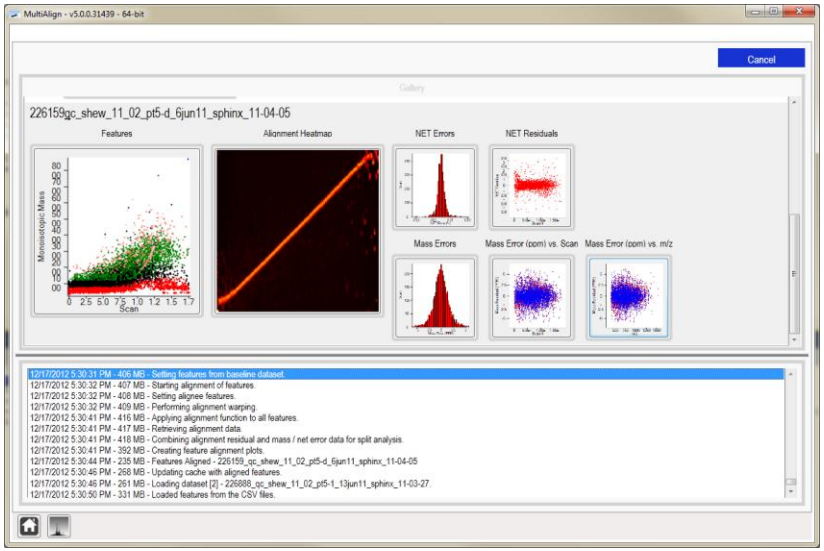


Select an analysis name



- 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

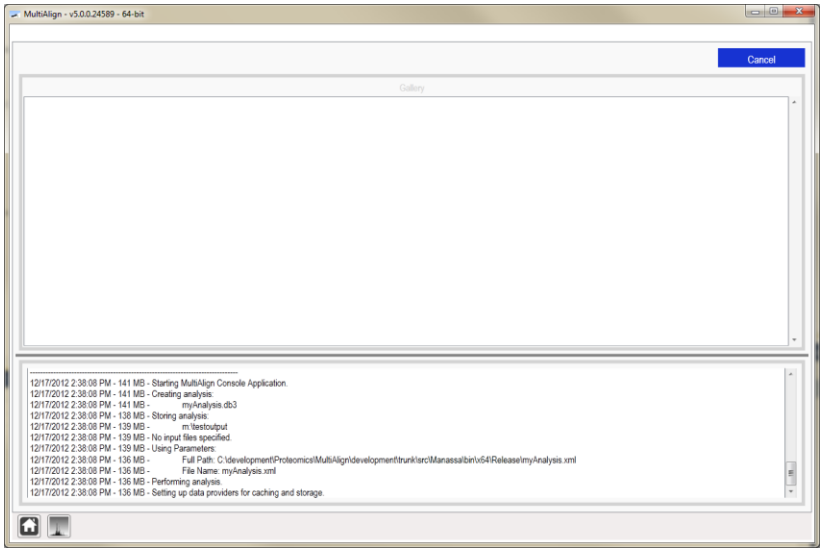
Running Analysis Window



- 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.

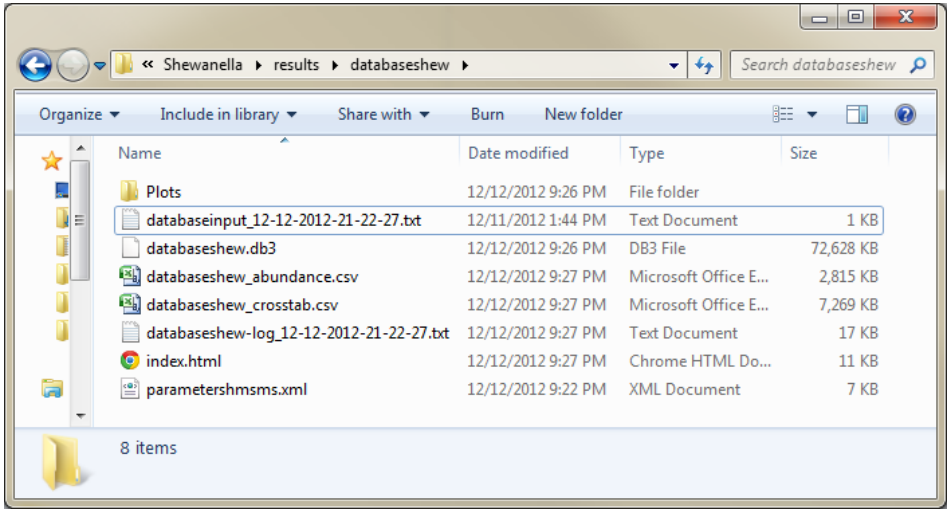
Need to cancel?

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



Output Files

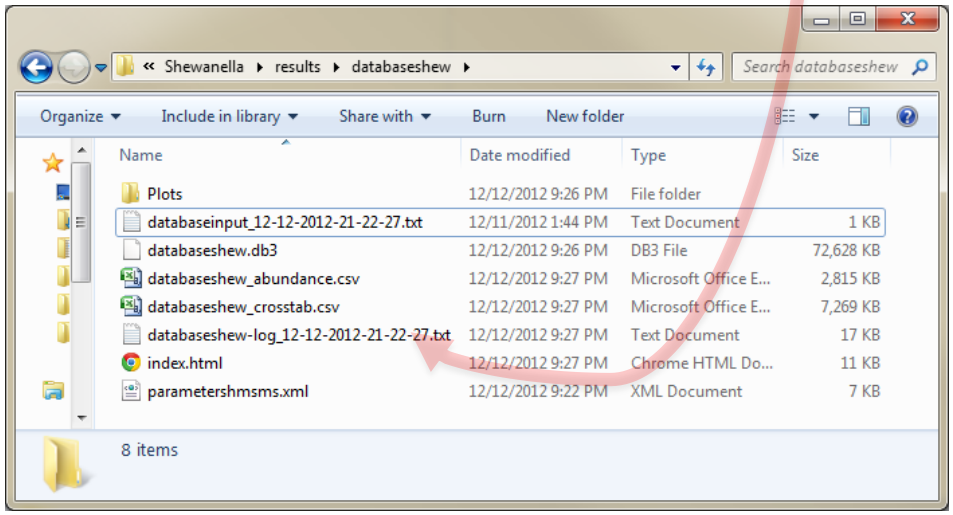
- You can navigate to the folder using Windows Explorer to find the data as the analysis is running.
- This is a list of files that MultiAlign will generate during the analysis.



- 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.

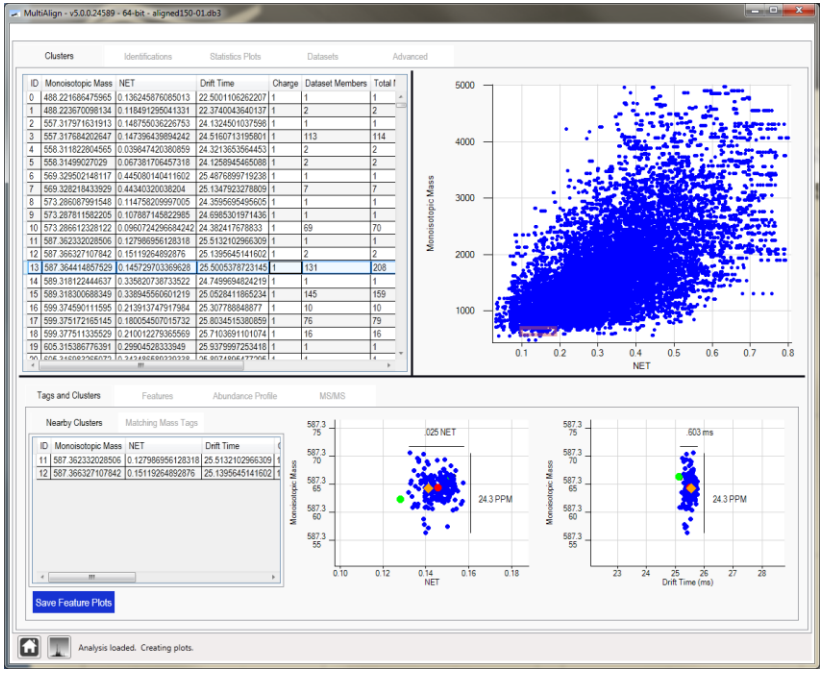
Errors?

- 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.



Analysis View

- 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