Babel Fish: Conversion of MS File Formats and Standards

Paul Gershon[1]; Panagiotis T. Papoulias[2]; Bryan Smith[2]; Philip Andrews[2]
[1]UC-Irvine, Irvine , CA; [2]University of Michigan, Ann Arbor, Michigan

Introduction

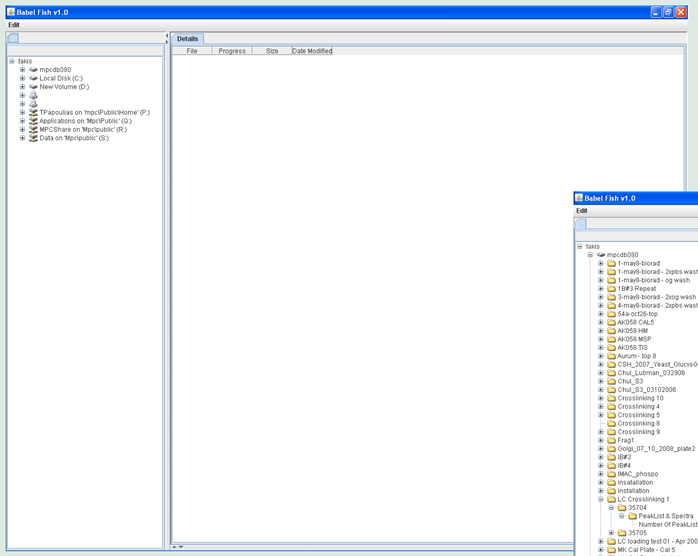
A software tool, Babel Fish, has been developed for the integration of proteomics instruments and software via the speedy conversion of mass spectrometric data to and from existing and emerging standards and data repositories, with the option of efficient data compression. Babel Fish enables the user to quickly navigate to desired datasets that may be stored on local hard drives, network drives, Applied Biosystems 4x00 databases, Tranche, and PRIME. During the compression and recovery of LC-MALDI spectra, distinct MALDI spotting patterns can be interpreted. Babel Fish presents the user with a familiar Windows Explorer-like file system interface in which instruments appear as disk icons on the left pane.

- **Simple, Explorer-like file system interface**
- **Read from hard drives, supported mass spectrometers (AB 4700 and 4800, Virgin 1) and Agilent ChemStation**
- **Export spectra and peak lists, along with associated meta-data) to mzML**
- **Seamless integration with Tranche**
- **Compression of mzML using mzSquash**

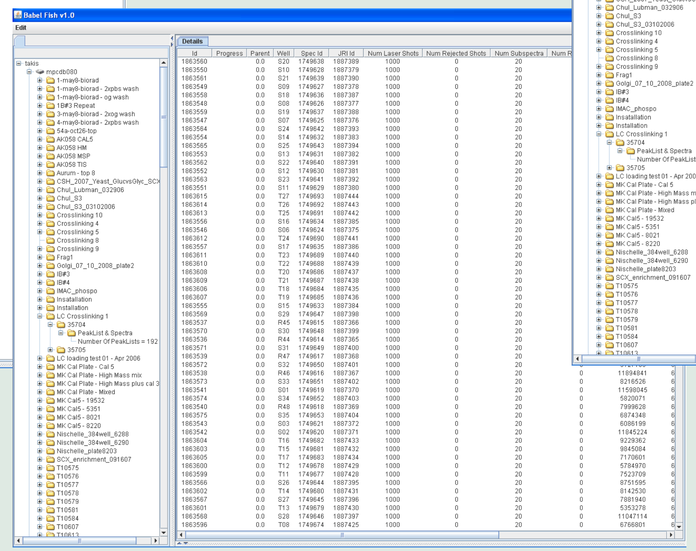
For more information or questions, please feel free to email: panagiot@umich.edu

Methods

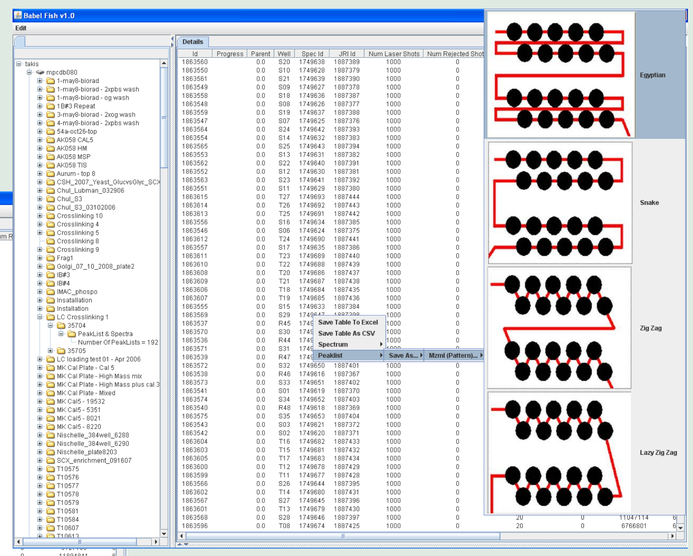
Babel Fish was developed using Sun Java 5.0. Connection to the ProteomeCommons.org Tranche network is established using the Java API (https://proteomecommons.org/tranche/dev/). Data exchange with PRIME was made through an implementation of the Prime-Data-Transfer-Protocol (http://www.prime-sdms.org/). Applied Biosystems proteomics analyzer databases were read directly using the T2Core.jar found on Proteomecommons.org. Spectrum visualization tools were generated by integration with the open source project MSExpedite (https://sourceforge.net/projects/msexpedite/). Compression of mzML files was accomplished using mzSquash (https://sourceforge.net/projects/mzsquash/).



When Babel Fish is launched, you can navigate through any device (e.g., disk, mass spectrometer, etc.) available in the left pane.



After LC MS/MS data is loaded, the data is viewable in the "Details" tab.



While exporting the peak lists to an mzML file, you can specify the LC-MALDI spotting patterns to ensure that the peak lists are in the proper experimental order.

Preliminary Results

Mass spectrometric data sets have been viewed graphically and converted into emerging standards, including mzML. For the Applied Biosystems 4x00 instruments, subfolders are organized hierarchically by job number, then by peaklist, spectra, and spot (well). Experimental metadata are displayed in a tabular format. Menus allow the user to save to various standards and formats (e.g., MS Excel workbook, mzML). Spectra and peak lists can be ordered according to any of four selectable LC-MALDI spotting patterns, thus providing compressed mzML-formatted data that are directly compatible with the "TOF2H" series of programs for HD-exchange by nanoLC-MALDI. mzSquash has an effective compression rate of 30% for mzML containing spectra and 98% for mzML containing peak lists (contrasted to gzip, which averaged 25% and 75%, respectively). A significant benefit of using mzSquash is that the compressed file can be read directly using the mzSquash API; the mzSquash binary file need not be decompressed.