

Babel Fish: Conversion of MS File Formats and Standards

Paul Gershon¹; Panagiotis T. Papoulias²; Bryan Smith²; Philip Andrews²
¹UC-Irvine, Irvine, CA; ²University 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 MSExpedit (<https://sourceforge.net/projects/msexpedite/>). Compression of mzML files was accomplished using mzSquash (<https://sourceforge.net/projects/mzsquash/>).

ID	Progress	Parent	Vwrt	Spec Id	JRT Id	Num Laser Shots	Num Projected Shots	Num Subspectra	Num P.
1883550	0.0	S20	1749638	1887399	1000	0	0	20	0
1883551	0.0	S10	1749628	1887379	1000	0	0	20	0
1883552	0.0	S21	1749639	1887380	1000	0	0	20	0
1883553	0.0	S09	1749627	1887378	1000	0	0	20	0
1883554	0.0	S18	1749638	1887387	1000	0	0	20	0
1883555	0.0	S08	1749626	1887377	1000	0	0	20	0
1883556	0.0	S19	1749637	1887388	1000	0	0	20	0
1883557	0.0	S07	1749625	1887376	1000	0	0	20	0
1883558	0.0	S24	1749642	1887393	1000	0	0	20	0
1883559	0.0	S14	1749632	1887383	1000	0	0	20	0
1883560	0.0	S25	1749643	1887384	1000	0	0	20	0
1883561	0.0	S13	1749631	1887382	1000	0	0	20	0
1883562	0.0	S12	1749630	1887381	1000	0	0	20	0
1883563	0.0	S23	1749641	1887392	1000	0	0	20	0
1883564	0.0	S11	1749629	1887380	1000	0	0	20	0
1883565	0.0	T27	1749693	1887444	1000	0	0	20	0
1883566	0.0	T26	1749692	1887443	1000	0	0	20	0
1883567	0.0	S17	1749635	1887395	1000	0	0	20	0
1883568	0.0	T24	1749690	1887441	1000	0	0	20	0
1883569	0.0	S16	1749634	1887395	1000	0	0	20	0
1883570	0.0	S06	1749624	1887375	1000	0	0	20	0
1883571	0.0	T24	1749690	1887441	1000	0	0	20	0
1883572	0.0	S17	1749635	1887395	1000	0	0	20	0
1883573	0.0	S17	1749635	1887395	1000	0	0	20	0
1883574	0.0	T22	1749688	1887439	1000	0	0	20	0
1883575	0.0	T22	1749688	1887439	1000	0	0	20	0
1883576	0.0	T20	1749686	1887437	1000	0	0	20	0
1883577	0.0	T18	1749684	1887435	1000	0	0	20	0
1883578	0.0	T19	1749685	1887436	1000	0	0	20	0
1883579	0.0	S15	1749633	1887384	1000	0	0	20	0
1883580	0.0	S29	1749647	1887398	1000	0	0	20	0
1883581	0.0	R45	1749615	1887386	1000	0	0	20	0
1883582	0.0	S30	1749648	1887399	1000	0	0	20	0
1883583	0.0	R44	1749614	1887385	1000	0	0	20	0
1883584	0.0	S31	1749649	1887400	1000	0	0	20	0
1883585	0.0	R47	1749617	1887388	1000	0	0	20	0
1883586	0.0	S32	1749650	1887401	1000	0	0	20	0
1883587	0.0	R46	1749616	1887387	1000	0	0	20	0
1883588	0.0	S33	1749651	1887402	1000	0	0	20	0
1883589	0.0	S01	1749619	1887370	1000	0	0	20	0
1883590	0.0	S34	1749652	1887403	1000	0	0	20	0
1883591	0.0	R48	1749618	1887389	1000	0	0	20	0
1883592	0.0	S35	1749653	1887404	1000	0	0	20	0
1883593	0.0	S03	1749621	1887372	1000	0	0	20	0
1883594	0.0	S02	1749620	1887371	1000	0	0	20	0
1883595	0.0	T16	1749682	1887433	1000	0	0	20	0
1883596	0.0	T15	1749681	1887432	1000	0	0	20	0
1883597	0.0	T17	1749683	1887434	1000	0	0	20	0
1883598	0.0	T12	1749678	1887429	1000	0	0	20	0
1883599	0.0	T11	1749677	1887428	1000	0	0	20	0
1883600	0.0	S26	1749644	1887395	1000	0	0	20	0
1883601	0.0	T14	1749679	1887427	1000	0	0	20	0
1883602	0.0	S27	1749645	1887396	1000	0	0	20	0
1883603	0.0	T13	1749678	1887426	1000	0	0	20	0
1883604	0.0	S28	1749646	1887397	1000	0	0	20	0
1883605	0.0	T08	1749674	1887425	1000	0	0	20	0

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.

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.