Universität Bielefeld
Technische Fakultät
AG Genominformatik

Maui and Maltcms

a Graphical User Interface and Application Framework for High-Throughput Metabolomics

Nils Hoffmann[1,2], Mathias Wilhelm[1], Kai Stadermann[1], Jens Stoye[1]

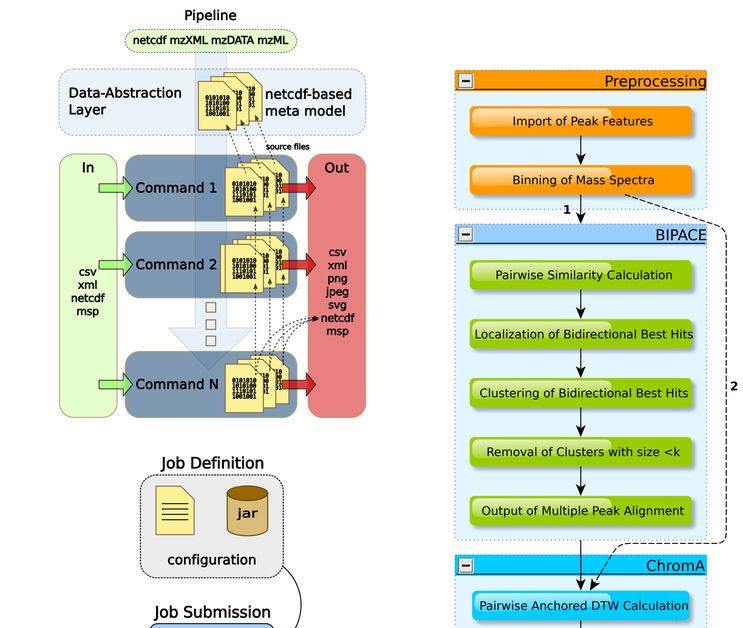
[1] Genome Informatics, Faculty of Technology, Bielefeld University
[2] Graduate School in Bioinformatics and Genome Research, CeBiTec, Bielefeld University

Introduction

Recent advances in analytical technology used for gas chromatography-mass spectrometry based metabolomics produce larger and larger amounts of data in need of processing. Furthermore, metabolomics is being used in clinical diagnostics and studies, generating multiple measurements for hundreds or even thousands of individuals over a period of years, which require subsequent comparison and analysis. File sizes are approaching and exceeding gigabytes per measurement and thus require efficient algorithms and data storage techniques for raw data preprocessing and analysis to be of practical use.

Methods

We present Maui, the Maltcms user interface, an application based on the Netbeans Rich Client Platform for interactive visualization and analysis of large datasets from metabolomics experiments. Maui is built on top of Maltcms, our modular application toolkit for chromatography-mass spectrometry that allows to define and execute typical data preprocessing and analysis workflows in metabolomics. An example of the Maltcms architecture for processing pipelines is given in Figure 1 (a). The framework is focused on processing of data from one-dimensional gas chromatography-mass spectrometry (GC-MS) and comprehensive two-dimensional GCxGC-MS. It provides methods for smoothing, baseline removal, peak detection, peak integration (Figure 5), and peak alignment [1], as well as visualizations of aligned and unaligned data. An example workflow for peak-assisted chromatogram alignment is given in Figure 1 (b).



(a) Top: Scheme of pipeline definition and execution in Maltcms. Bottom: Schematic of remote execution infrastructure.

(b) Processing scheme of BiPACE/ChromA.

Figure 1: Schematics of the Maltcms pipeline and remote processing libraries (left) and of BiPACE/ChromA processing for peak-assisted alignment of GC-MS data (right). Alternative 1 involves calculation of all-against-all pairwise peak similarities and multiple alignment of peaks by successive peak group extension based on bi-directional best hits, using paired peaks as anchor constraints for dynamic time warping alignment (DTW). Preliminary alignment performance results of BiPACE are given in Figure 6. Alternative 2 calculates the DTW alignment directly. Both methods use the calculated DTW scores to select a representative chromatogram as alignment reference.

Maui supports the generation and execution of user-defined workflows in an intuitive way (Figure 3, left) and provides tools in order to create user databases of reference compounds for targeted or comparative analysis. It furthermore provides views for raw and processed GC-MS (Figure 2) and GCxGC-MS data (Figure 3), as well as for other processing results (Figure 4).

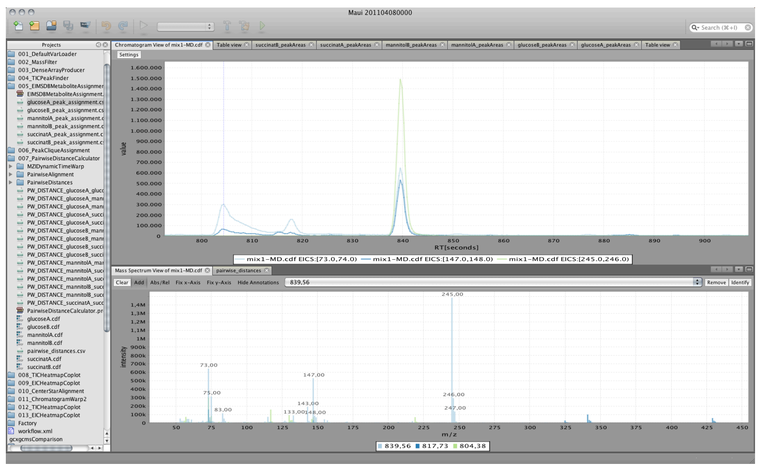


Figure 2: Maui's chromatogram view allows interactive browsing of chromatograms and selection of individual mass spectra for display. Customization allows to visualize either the total ion count (TIC), summed extracted ion counts (EICs), or co-plotted EICs. Mass spectra can be added or removed to the MS plot individually and can be used for manual search against the GMD or other AMDIS-format (msp) compatible databases.

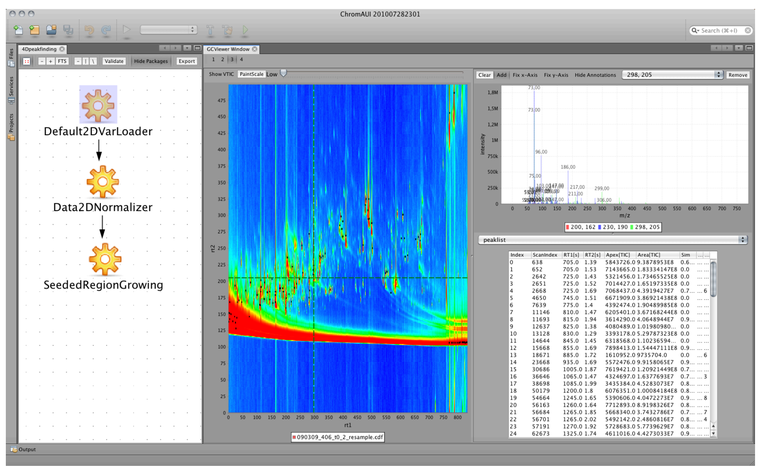


Figure 3: Maui provides a pipeline editor (left) to define Maltcms processing pipelines. Pipelines can be executed and output can be opened immediately to view processing results. Maui's view for GCxGC-MS chromatograms (center) allows interactive browsing and selection of individual mass spectra (top-right) for display along with peaks that were identified using ChromA4D (bottom-right). Different color schemes can be used and adjusted to improve visualization.

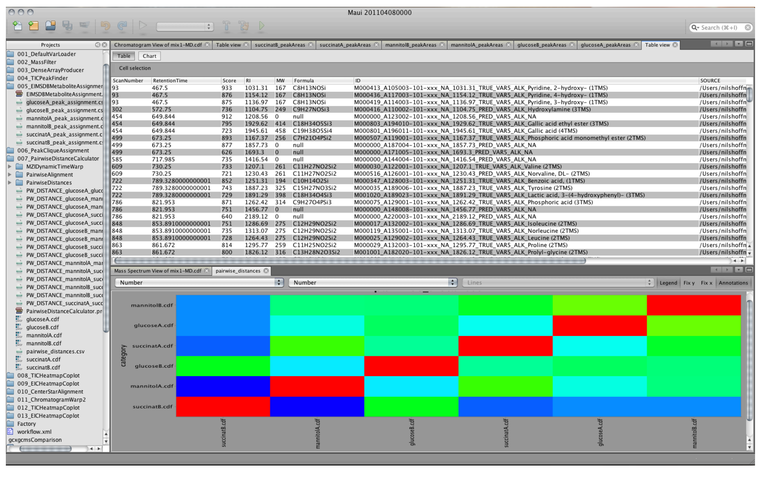


Figure 4: Maui can be used to visualize processing result that were generated with Maltcms. On top, a table view of putative peak identification results for one chromatogram against the Golm Metabolite Database is shown. Below, a pairwise distance matrix obtained from dynamic time warping between chromatograms is shown. Red colors indicate a high similarity, green colors indicate a medium similarity, whereas blue colors indicate a low similarity.

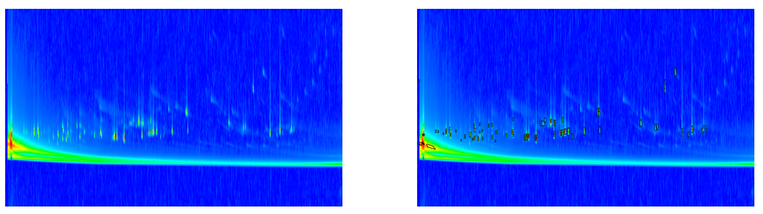


Figure 5: Left: Image of a 2D-TIC of a GCxGC-MS chromatogram from a standard FAME mix. Right: Peak boundary overlay after peak detection and integration with seeded region growing.

Maltcms includes libraries for local and remote parallelization of tasks on a local network of connected computers or within a grid or cloud infrastructure, allowing researchers to scale up their analysis with increasing computing capacities.

It supports the major formats for chromatography mass-spectrometry data as input, like netCDF, mzXML, mzData and mzML. Databases can be imported from msp (AMDIS) compatible text data. The export of user databases to msp is also possible.

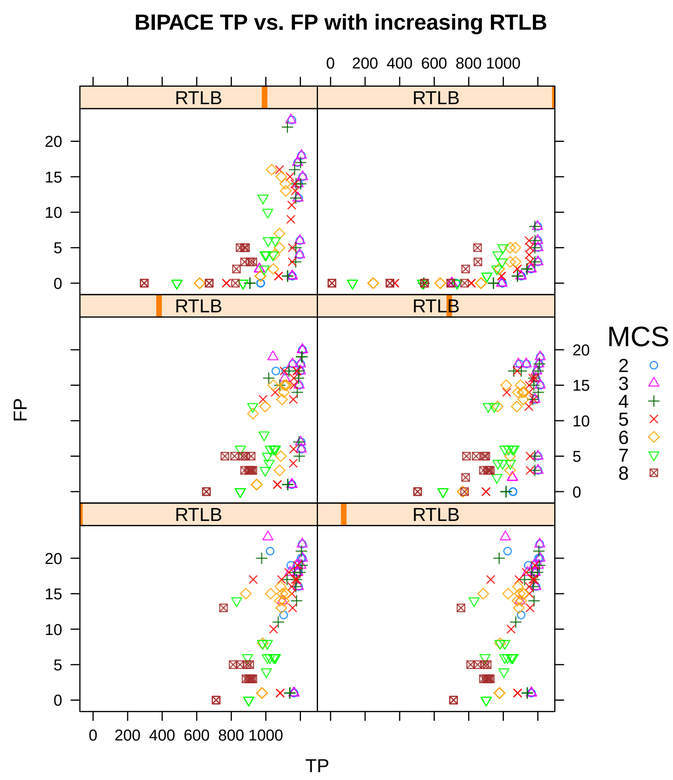


Figure 6: Evaluation results of BiPACE against reference data [2]. TP and FP are given as absolute numbers. RTLB is the minimum retention time deviation-based score required before mass spectral similarities are computed. MCS is the minimum required peak group size.

Results

Maui and Maltcms are implemented using the JAVA programming language allowing them to be extended and adapted easily to custom requirements. Extensions can be developed either in JAVA or in scripting languages like Groovy.

Maltcms is freely available at http://maltcms.sourceforge.net under the L-GPL v3 license. Maui will be officially released early next year and will be available from the same location.

References

[1] N Hoffmann and J Stoye. ChromA: signal-based retention time alignment for chromatography-mass spectrometry data. *Bioinformatics*, 25(16):2080–1, 2009.

[2] M D Robinson, D P De Souza, W Keen, E C Saunders, M J Mcconville, T P Speed, and V A Likić. A dynamic programming approach for the alignment of signal peaks in multiple gas chromatography-mass spectrometry experiments. *BMC Bioinformatics*, 8:419, 2007.

Poster copy available at: `http://www.cebitec.uni-bielefeld.de/~hoffmann`