Universität Bielefeld

Maltcms and Maui — an OpenSource Software Framework and User Interface 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, Germany
[2] Graduate School in Bioinformatics and Genome Research, CeBiTec, Bielefeld University, Germany

▶ http://maltcms.sf.net

Overview

We developed an OpenSource, platform–independent software framework and user interface for chromatography-mass spectrometry based high-throughput metabolomics.

Introduction

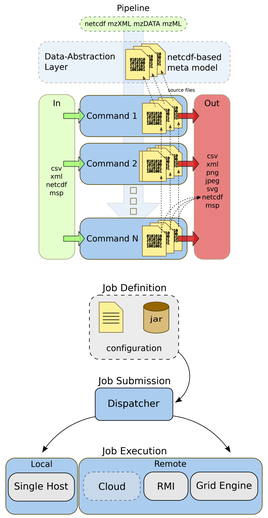
Metabolomics has developed into a mature field of research during the past decade. The establishment of common experimental routines and potent analytical platforms has led to a dramatical increase in throughput and sample size, producing larger and larger amounts of data. High-throughput metabolomics research requires software capable of handling such complex and large samples efficiently and in parallel. Open data exchange formats, such as netCDF and mzML are a key requirement for repeatability, comparability, and validation of data gathered in metabolomics studies. We present a software framework that addresses these challenges and provides the foundation for a user-friendly, rich client application with a focus on one- and comprehensive two-dimensional gas chromatography-mass spectrometry (GC-, GCxGC-MS) data.

Method 1: Maltcms

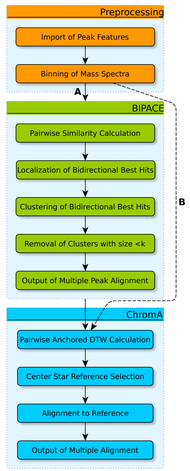
The **Maltcms** framework provides the following features:

- platform–independence
- handling of files of arbitrary size
- support for netCDF (ANDI-MS, ANDI-CHROM), mzML, mzXML and mzData formats
- unified access to 1D- and 2D-chromatograms (GC, GCxGC) with single- or multidimensional detectors (FID, MS)
- pipeline–based data processing abstraction with pipeline validation and result reuse (Figure 1 (a) top)
- putative identification of metabolites on msp-format compatible databases based on mass spectra and retention index
- parallelization of processing steps on local or remote computers (Figure 1 (a) bottom)

Furthermore, it provides methods for smoothing, baseline removal, peak detection, peak integration, and peak alignment [1], as well as visualizations of aligned and unaligned data for GC-MS and GCxGC-MS [2]. 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 using BiPACE and/or ChromA for retention time alignment.

Figure 1: Schematics of the Maltcms pipeline and remote processing libraries (left) and of BiPACE/ChromA processing for automated retention time alignment of GC-MS data. (right).

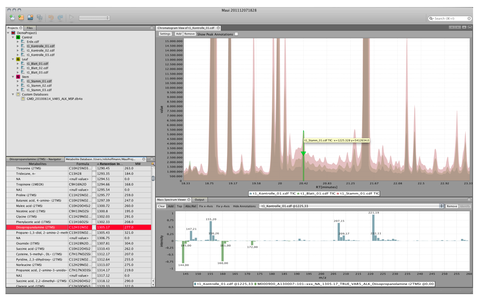


Figure 2: Maui's chromatogram view allows interactive browsing of chromatograms and selection of individual mass spectra for display (top-right). Customization allows to visualize either the total ion count (TIC), summed extracted ion counts (EICs), or co-plotted EICs. Mass spectra can be added or removed to the MS plot (bottom-right) individually and peak annotations can be used for automatic search against the GMD [3] or other AMDIS-format (msp) compatible databases (bottom-left).

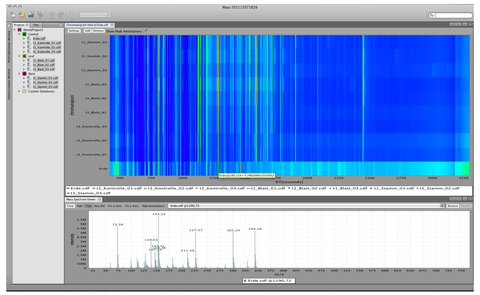


Figure 3: Co-plot view of multiple chromatograms' TICs (top-right). Individual mass spectra can be selected and added to the MS plot (bottom) for manual inspection.

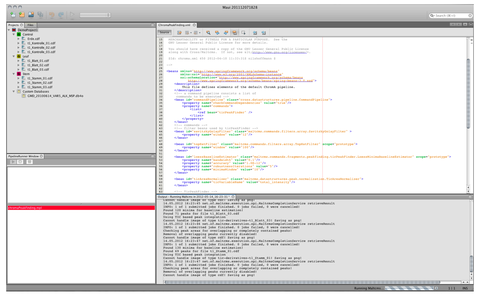


Figure 4: Maui provides a pipeline editor to define Maltcms processing pipelines (top-right). Pipelines can be executed (bottom-left) and output can be opened immediately to view processing results (bottom).

Figure 5: Visualization of peak-based pairwise similarities calculated by BiPACE (top), aligned peak apex retention times (top-right), and raw peak areas (bottom) w.r.t. the automatically determined alignment reference.

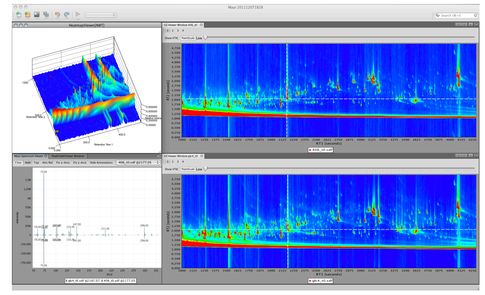


Figure 6: Maui's view for GCxGC-MS chromatograms allows interactive browsing and selection of individual mass spectra. Different color schemes can be used and adjusted to improve visualization.

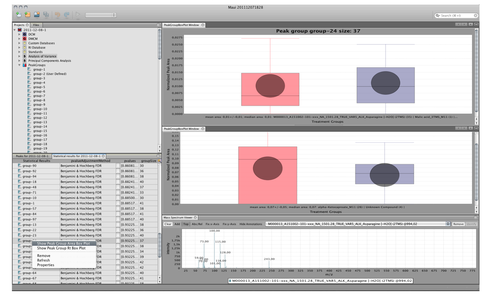


Figure 7: Anova results of peak groups using a correction for multiple testing (bottom-left). Boxplots of two selected peak group areas, each for the two experimental conditions DCM and DMCM (top and center).

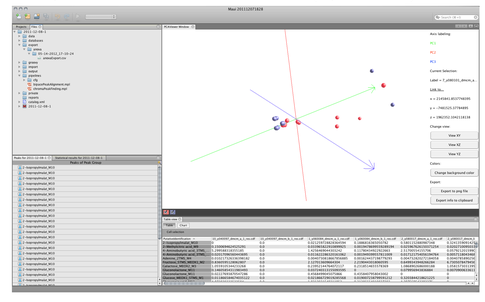


Figure 8: List of peaks contained in a selected peak group (left). Three-dimensional plot of principal components analysis results for experimental conditions DCM and DMCM (center). View of export report of aligned and normalized peak areas (bottom-right).

Method 2: Maui

Maui, the Maltcms user interface, is an application based on the NetBeans Rich Client Platform for interactive visualization and analysis of large datasets from metabolomics experiments with the following features:

- definition and curation of experiments, sample groups and individual normalization factors
- editing and execution of Maltcms processing pipelines (Figure 6)
- import and curation of mass spectral data in msp format (NIST) as custom databases (Figure 2)
- import of ChromaTOF peak lists (LECO Corp.) from GC and GCxGC-MS
- putative batch identification of peak data using mass spectral similarity and retention index
- visualization of raw and processed GC-MS data (Figure 2)
- visualization of raw and processed GCxGC-MS data (Figure 6)
- visualization of Maltcms processing results (Figure 5)
- statistical backend (Rserve) providing Anova and PCA (Figures 7 and 8)
- export of aligned and normalized peak areas with Anova results (Figure 8 bottom)

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. Both projects welcome active contribution.

Maui and Maltcms are freely available under the L-GPL v3 license at `http://maltcms.sf.net`.

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] N Hoffmann and J Stoye. Generic Software Frameworks for GC-MS Based Metabolomics. In Ute Roessner, editor, *Metabolomics*, pages 73–98. InTech / Rijeka, 2012.

[3] J Hummel, J Selbig, D Walther, and J Kopka. The Golm Metabolome Database: a database for GC-MS based metabolite profiling. In Jens Nielsen and Michael Jewett, editors, *Metabolomics*, pages 75–95. Springer Berlin / Heidelberg, 2007.