Universität Bielefeld

Technische Fakultät
AG Genominformatik

ChromAUI

An integrated OpenSource Application for Metabolomics based on the Netbeans Rich Client Platform

Nils Hoffmann[1,2], Mathias Wilhelm[1], Matthias Keck[3], Anja Döbbe[4], Karsten Niehaus[3], Olaf Kruse[4], Jens Stoye[1]

[1] Genome Informatics Group, Faculty of Technology, Bielefeld University
[2] International Graduate School in Bioinformatics and Genome Research
[3] Proteome and Metabolome Research Group, [4] Algae Biotech and Bioenergy Group, CeBiTeC, Bielefeld University

Introduction

Two-dimensional gas-chromatography mass-spectrometry (GCxGC-MS) has developed into an important technique for the analysis of complex biological samples. It is applied by a growing number of researchers in the field of metabolomics due to its increased peak capacity compared to one-dimensional GC-MS. Yet, this has lead to larger amounts of data, which are even harder to inspect and analyze manually than one-dimensional GC-MS data. Currently, only a few commercial software solutions exist, which cover most parts of the workflow from raw data preprocessing to statistical data analysis, such as ChromaTOF (LECO Corp.) and GC Image (GC Image, LLC).

However, there are no OpenSource solutions available yet, which integrate all steps of processing, analysis and visualization of GCxGC-MS, as well as GC/LC, GC-MS and LC-MS data into a complete solution. We present our software ChromAUI, which captures the typical workflow from data preprocessing to feature detection, grouping and output, as well as visualizations within one comprehensive graphical user interface. All processing steps generate data formats compatible with OpenSource statistics software such as R or spreadsheet programs such as OpenOffice, so that they can be integrated into existing workflows. ChromAUI uses our OpenSource framework Maltcms [1] for data processing and analysis, available at `maltcms.sourceforge.net`.

GC/LC-MS data

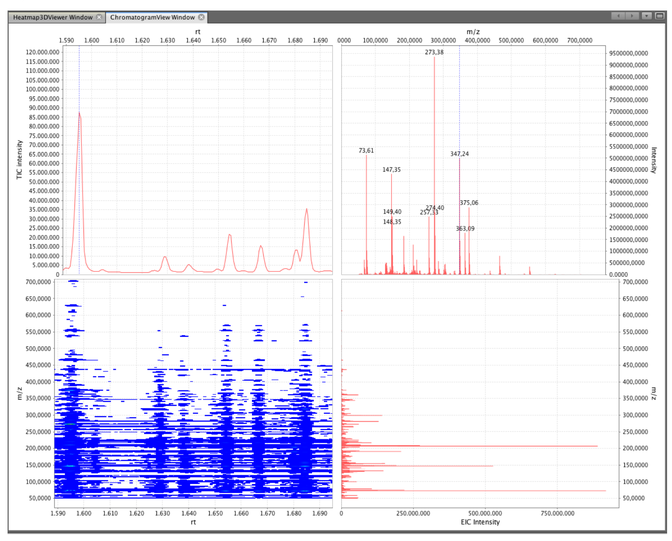


Figure 1: Screenshot of the ChromAUI application, showing the GC/LC-MS heatmap view at the bottom left. Above, the total ion count (TIC) is diplayed. In the upper right hand panel the mass spectra for selected scans are shown. The panel on the lower right hand side displays the cumulative ion channel intensity.

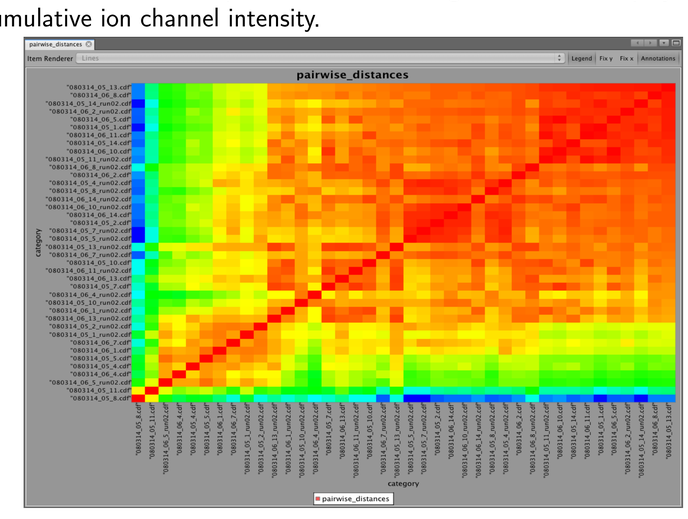
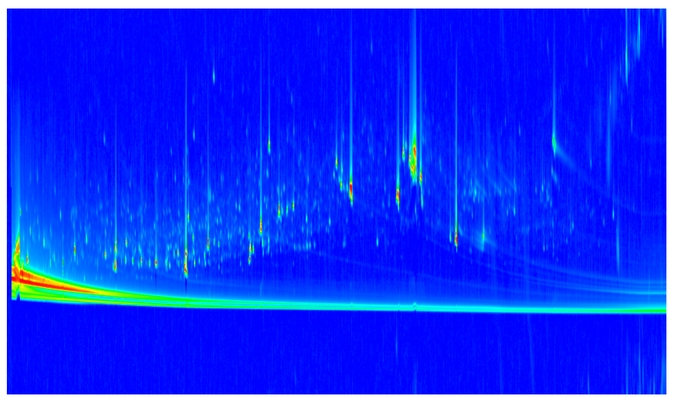


Figure 2: Inverse Visualization of pairwise distances as similarities. Blue color indicates a low similarity between chromatograms, while red indicates a high similarity. Similarities were calculated using the partitioned dynamic time warping (DTW) algorithm with automatically determined anchors on GC-MS data from an experiment that examined the impact of CO_2 concentration on the biomass production of common wheat [2].

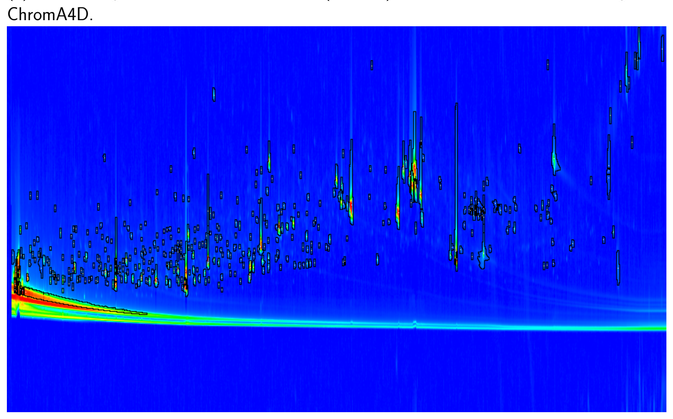
GCxGC-MS data

The GCxGC-MS samples presented on this poster were obtained from the TMS derivatized extract from *Chlamydomonas reinhardtii* wildtpye (406) and mutant (*glc*4) before ($t0$) and during ($t4$) H_2 production [3]. The raw data from these extracts was acquired on a LECO Pegasus® 4D GCxGC-TOFMS machine and was converted to ANDIMS format using the ChromaTOF® software after resampling to 100 Hz scan acquisition rate and baseline removal.

Peakfinding and -integration



(a) Chromatogram heatmap visualization (2D-TIC) of 406_$t0$ after initial filtering with ChromA4D.



(b) Chromatogram heatmap visualization of 406_$t0$ with highlighted peak areas of the 400 most prominent peaks. The other peaks were omitted for an improved visualization.

Figure 3: In order to find candidate peaks, ChromA4D scans for local maxima in the 2D-TIC. Starting from these initial seeds, coherent peak areas are located by using a variant of seeded region growing, based on mass spectral similarities between adjacent 2D-TIC signals.

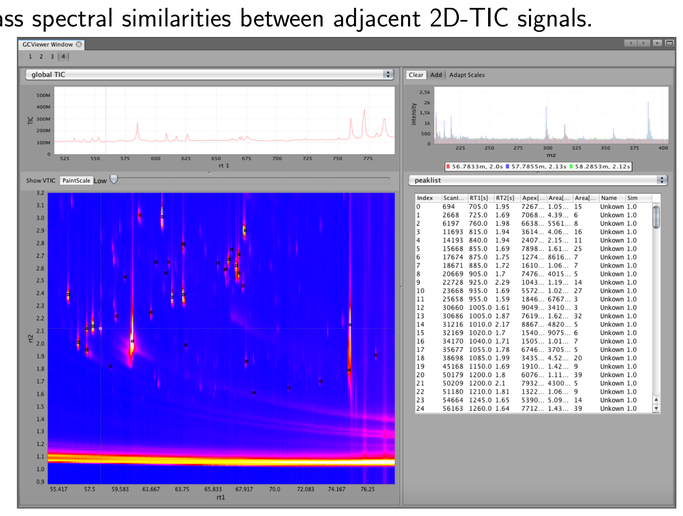
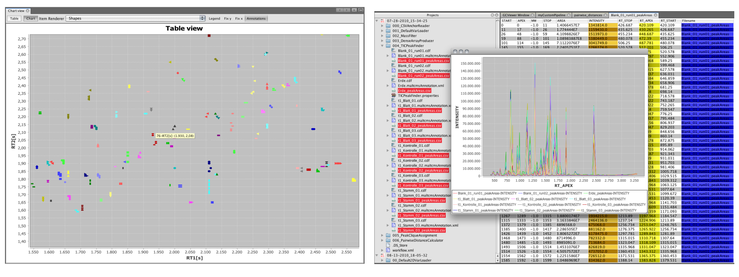


Figure 4: Interactive heatmap view within ChromAUI, clockwise from bottom left, showing a section of a GCxGC-MS chromatogram heatmap in a custom color scheme with peak markers, the total ion count for the modulations, the mass spectrum of selected peaks, and the peak list of the chromatogram.

Peakmatching

ChromA4D and ChromA use a bidirectional best hits (BBH) peak matching algorithm, based on the similarity of apex mass spectra of peaks, penalized by exponentially weighted time deviation on each separation dimension.

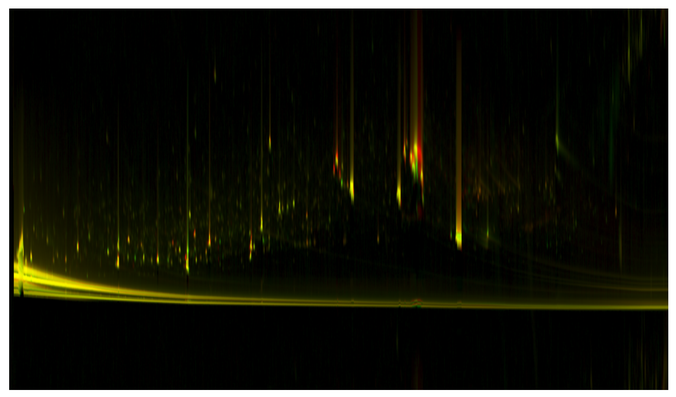


(a) (b)

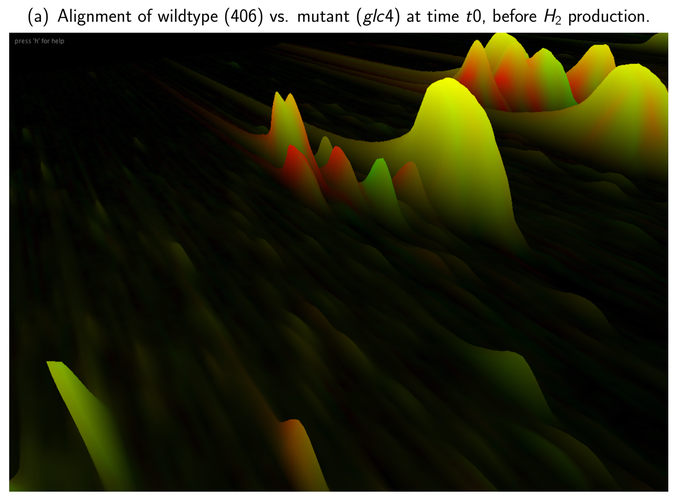
Figure 5: Peak plots in ChromAUI: (a) Plot of retention time (RT1 vs. RT2) of peak groups which were paired between all samples. Identical color indicates association to the same peak group. (b) Plot of column selection from multiple csv files containing one-dimensional GC-MS peak data as overlayed line chart.

Alignment

ChromA4D and ChromA support either an alignment based on the BBH peak matching algorithm or a variant of the dynamic time warping algorithm for alignment of complete profile data. The alignments can be performed either on TIC data or on the mass spectra of the datasets.



(a) Alignment of wildtype (406) vs. mutant (*glc*4) at time $t0$, before H_2 production.



(b) Alignment of wildtype (406) vs. mutant (*glc*4) at time $t0$, before H_2 production visualized as a three dimensional heat map surface.

Figure 6: Result of the application of dynamic time warping using the total ion count of each modulation as a feature vector for pairwise similarity calculation. Yellow peaks are present in both chromatograms at the same intensity level while red ones are only present in the wildtype (406) chromatogram and green ones only in the mutant (*glc*4).

Results and Outlook

ChromAUI together with our Maltcms-based GC-MS workflow ChromA allow to assess the influence of CO_2 concentration on the biomass production of common wheat and the Maltcms-based GCxGC-MS workflow ChromA4D to study wildtype versus mutant samples from *Chlamydomonas reinhardtii* under two different conditions.

We currently work on the interaction of ChromAUI with MeltDB [4] to allow access to experiments which were defined within MeltDB as well as the generation and editing of peak annotations.

ChromAUI will soon be available under the L-GPL v3 license at `http://maltcms.sourceforge.net` together with our framework Maltcms. It runs under all personal computer operating systems for which a JAVA Runtime Environment is available and typically requires $1-2$ GBytes of RAM.

References

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

[2] P Hogy, M Keck, K Niehaus, and J Franzaring. Effects of atmospheric co2 enrichment on biomass, yield and low molecular weight metabolites in wheat grain. *Journal of Cereal Science*, 52(2):215–220, Jan 2010.

[3] A Doebbe, M Keck, M La Russa, Jan H Mussgnug, Ben Hankamer, Ercan Tekce, Karsten Niehaus, and Olaf Kruse. The interplay of proton, electron, and metabolite supply for photosynthetic h2 production in chlamydomonas reinhardtii. *Journal of Biological Chemistry*, 285(39):30247–30260, Jan 2010.

[4] Heiko Neuweger, Stefan P Albaum, Michael Dondrup, Marcus Persicke, Tony Watt, Karsten Niehaus, Jens Stoye, and Alexander Goesmann. Meltdb: a software platform for the analysis and integration of metabolomics experiment data. *Bioinformatics*, 24(23):2726–32, Dec 2008.