

# Maltcms

an Application Framework for Processing of Metabolomics-Data



Mathias Wilhelm<sup>1</sup>, Nils Hoffmann<sup>1,2</sup>, Matthias Keck<sup>3</sup>,  
Karsten Niehaus<sup>3</sup>, Jens Stoye<sup>1</sup>

<sup>1</sup> Genome Informatics Group, Faculty of Technology, Postfach 10 01 31, Bielefeld University, Germany

<sup>2</sup> International NRW Graduate School in Bioinformatics and Genome Research

<sup>3</sup> Proteome and Metabolome Research Group, CeBiTeC, Bielefeld University

## Introduction

We present the current stable release 1.0 of our modular application toolkit for chromatography-mass spectrometry (Maltcms), which is used in our applications ChromA [1] and ChromA4D (Poster 100).

Maltcms provides a flexible, pipeline-based infrastructure for processing data acquired from gas-chromatography (GC) and liquid-chromatography (LC) coupled to one-dimensional (e.g. FID) or mass spectrometry (MS) detectors. Furthermore, it is able to process data from comprehensive two-dimensional gas chromatography (GCxGC) coupled to mass spectrometry. It provides methods for

- signal denoising and background removal
- peak finding, integration and matching
- retention time alignment of peaks and chromatograms by dynamic time warping (DTW)
- clustering of peaks and chromatograms
- annotation of signal peaks against MS database
- visualizations for all methods
- data export in formats compatible for downstream statistical analysis.

The processing steps within the framework center around the pipeline paradigm (see Figure 1), where each step can define dependencies on previous steps to allow for automatic pipeline validation. A pipeline is defined within a simple text file stating the commands to be executed together with the locations of their respective configuration files.

Pipeline commands can be easily implemented in the JAVA programming language or other, compatible languages.

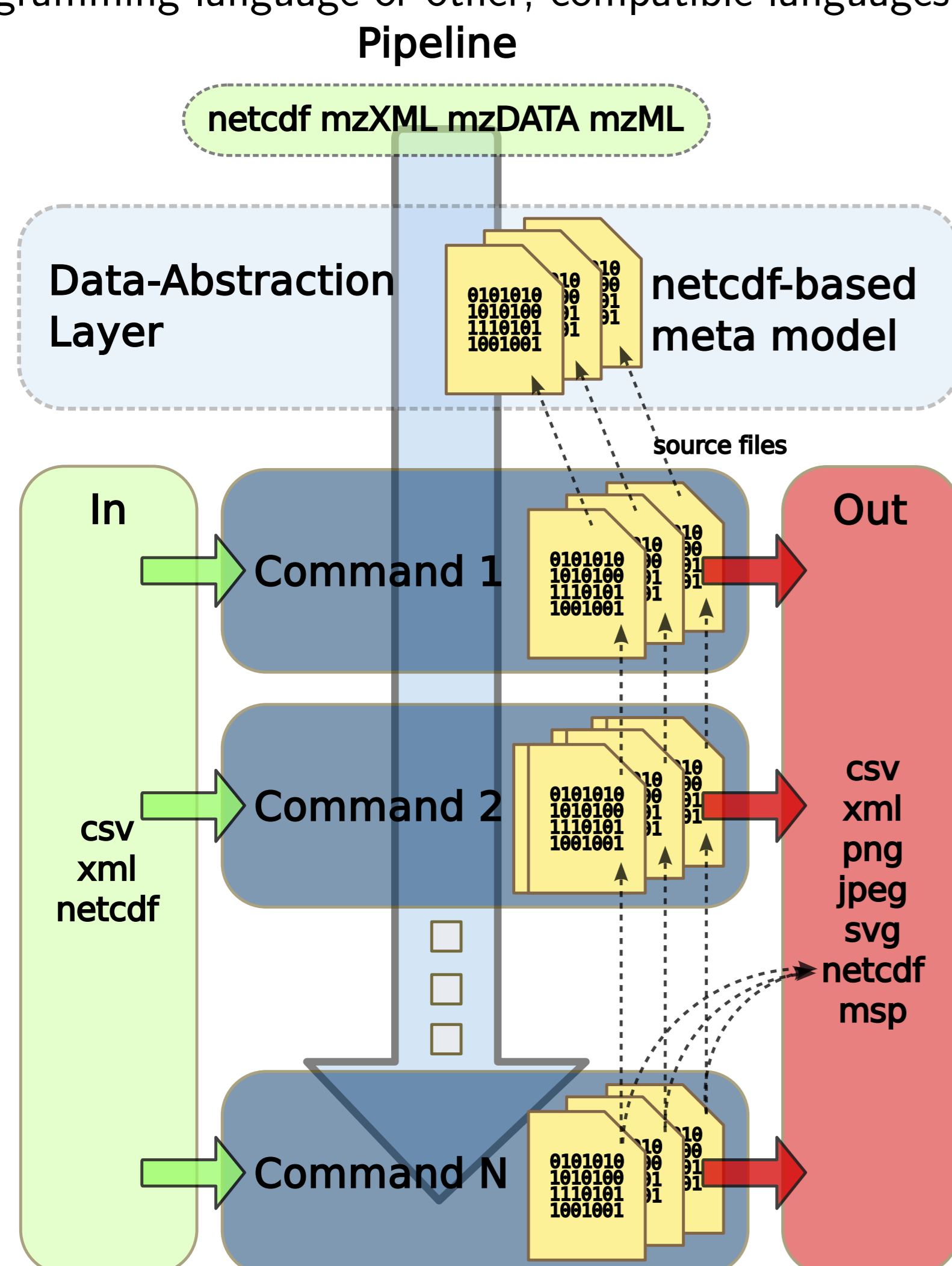


Figure 1: Schematic of pipeline execution with Maltcms. Commands within the pipeline have access to raw data files via the data abstraction layer and to arbitrary data via their source files. This allows incremental saving and aggregation of processing results and the definition of dependencies between commands.

## ChromA

ChromA is our application for processing of GC, LC-MS and GC-MS data, using the algorithms provided by Maltcms. Figure 2 shows the principle behind the bidirectional best hits peak clique assignment algorithm, while Figure 3 shows an example for peak-matching and complete profile alignment on GC-MS data using DTW with different alignment reference chromatograms, using anchors from peak clique assignment to constrain and speed up the alignment calculation.

ment on GC-MS data using DTW with different alignment reference chromatograms, using anchors from peak clique assignment to constrain and speed up the alignment calculation.

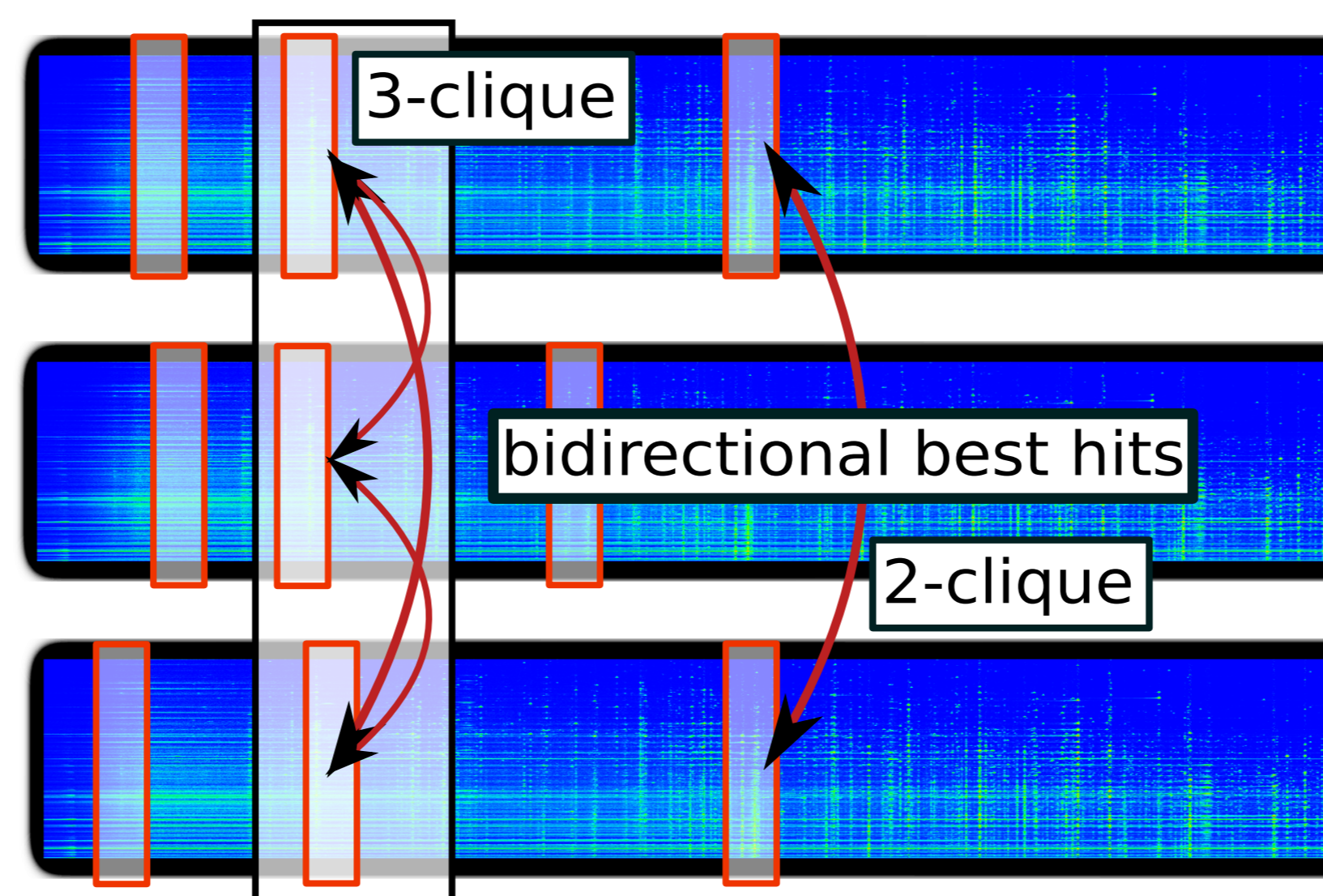


Figure 2: ChromA: Bidirectional best hits-based peak clique selection on GC-MS chromatograms (BPC).

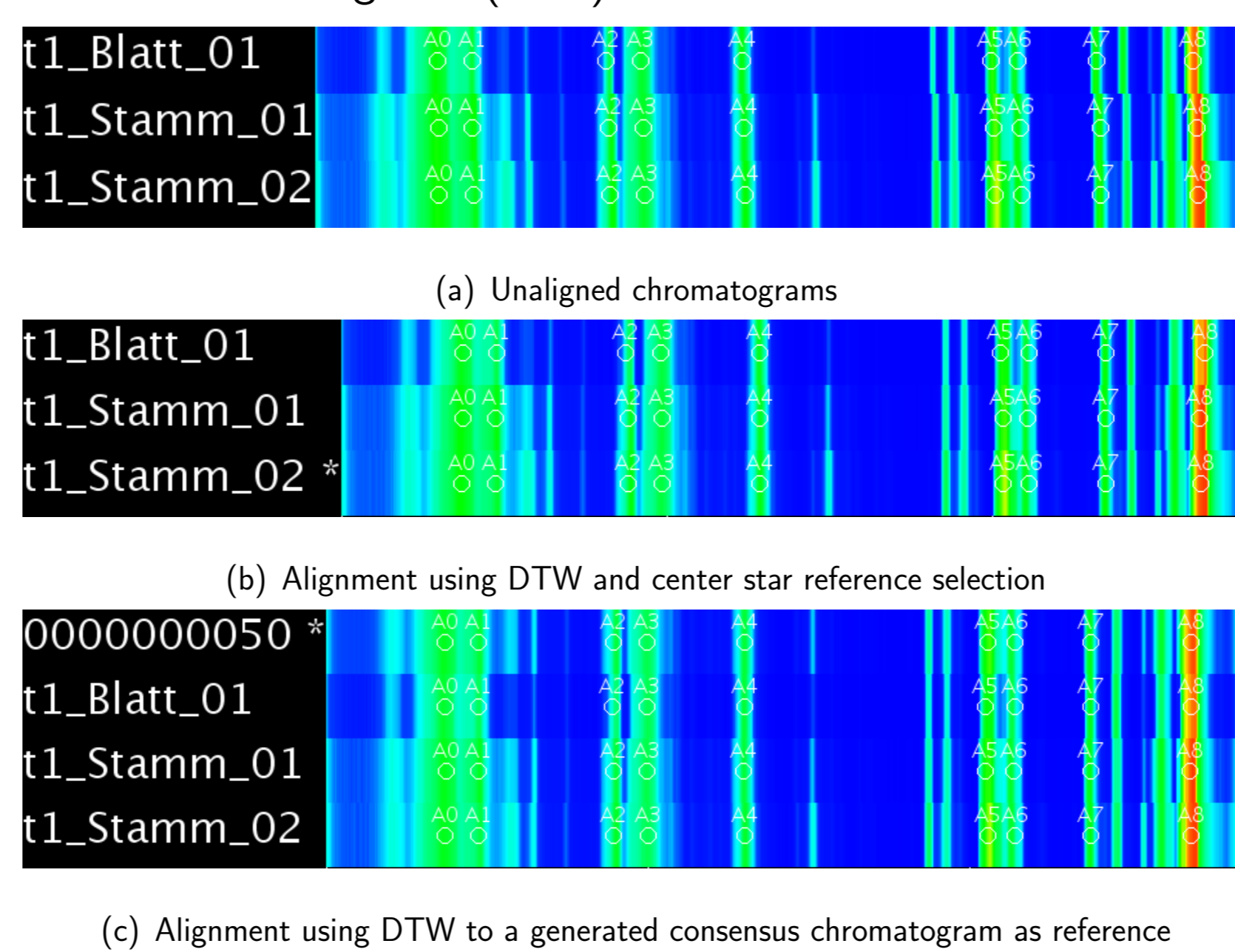


Figure 3: ChromA: Heatmaps of the total ion count (TIC) of sample GC-MS chromatograms from a *Xanthomonas campestris* pv. *campestris* experiment (a) before and (b), (c) after alignment with Dynamic Time Warping (DTW) using a local cosine score. In (b), the chromatogram with the highest sum of pairwise similarities was chosen as alignment reference. (c) shows a consensus chromatogram generated by using a guide tree based on pairwise DTW similarities applying successive alignment, merging and averaging, before aligning all original files to the consensus with DTW. The positions of anchors found by the BPC algorithm are highlighted as white circles.

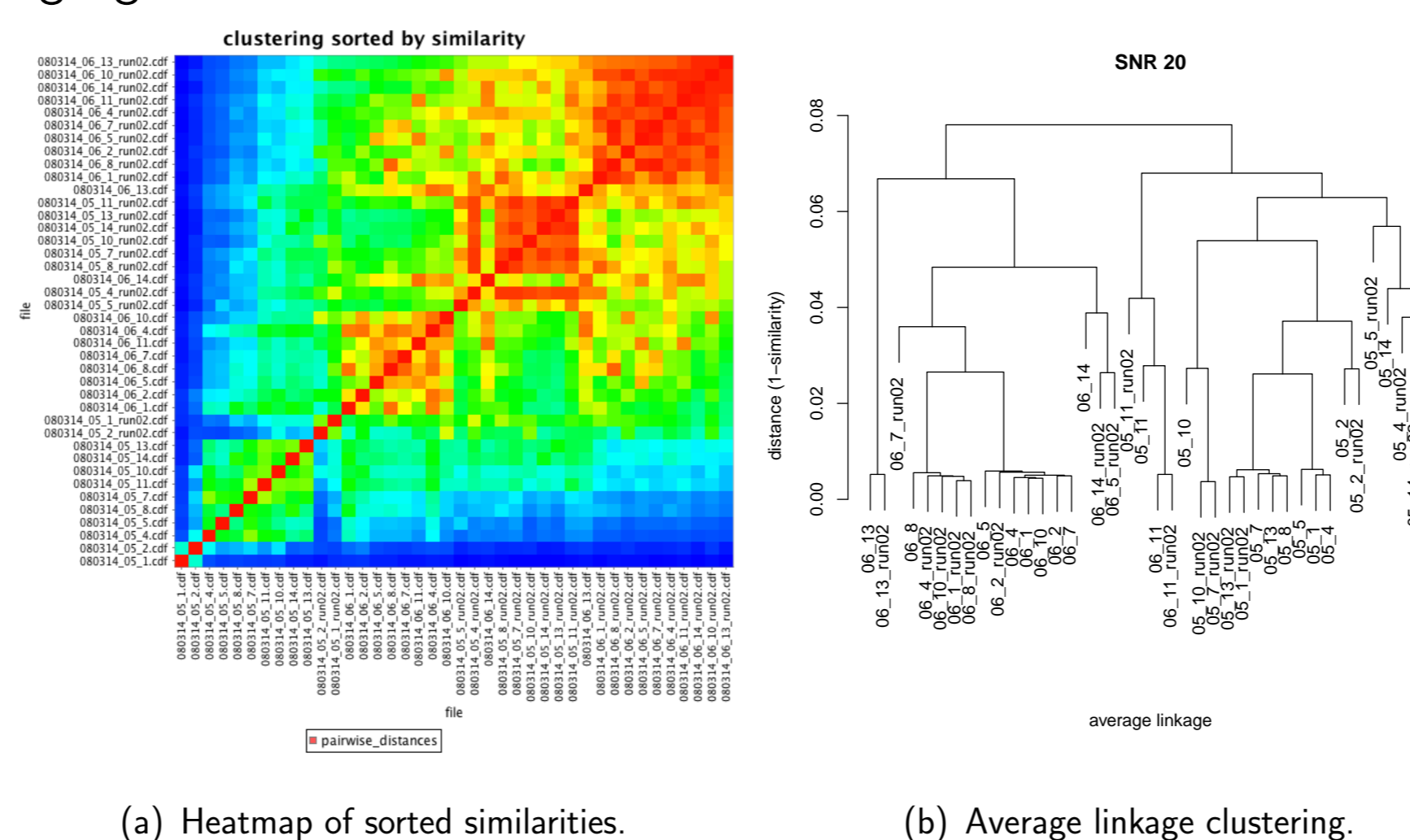
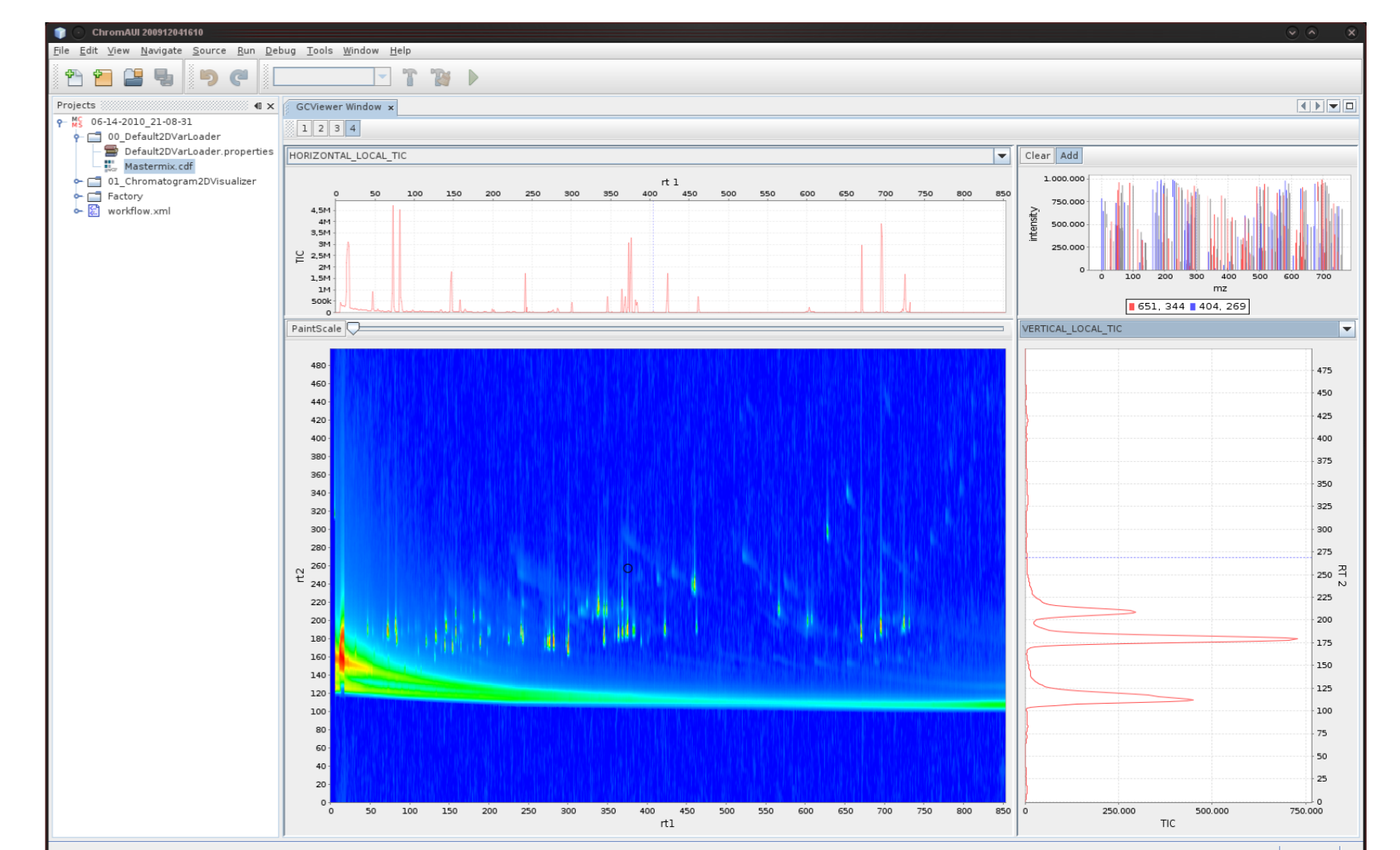


Figure 4: Visualizations of similarities between chromatograms. (a) shows a heatmap of sorted similarities between pairs of GC-MS chromatograms in ChromAUI. There are four clusters clearly visible corresponding to the different groups within the experiment. (b) shows the average linkage clustering obtained on the same dataset using *R*'s *hclust* method on average similarities calculated from the BPC algorithm. Here, only two groups are immediately visible.

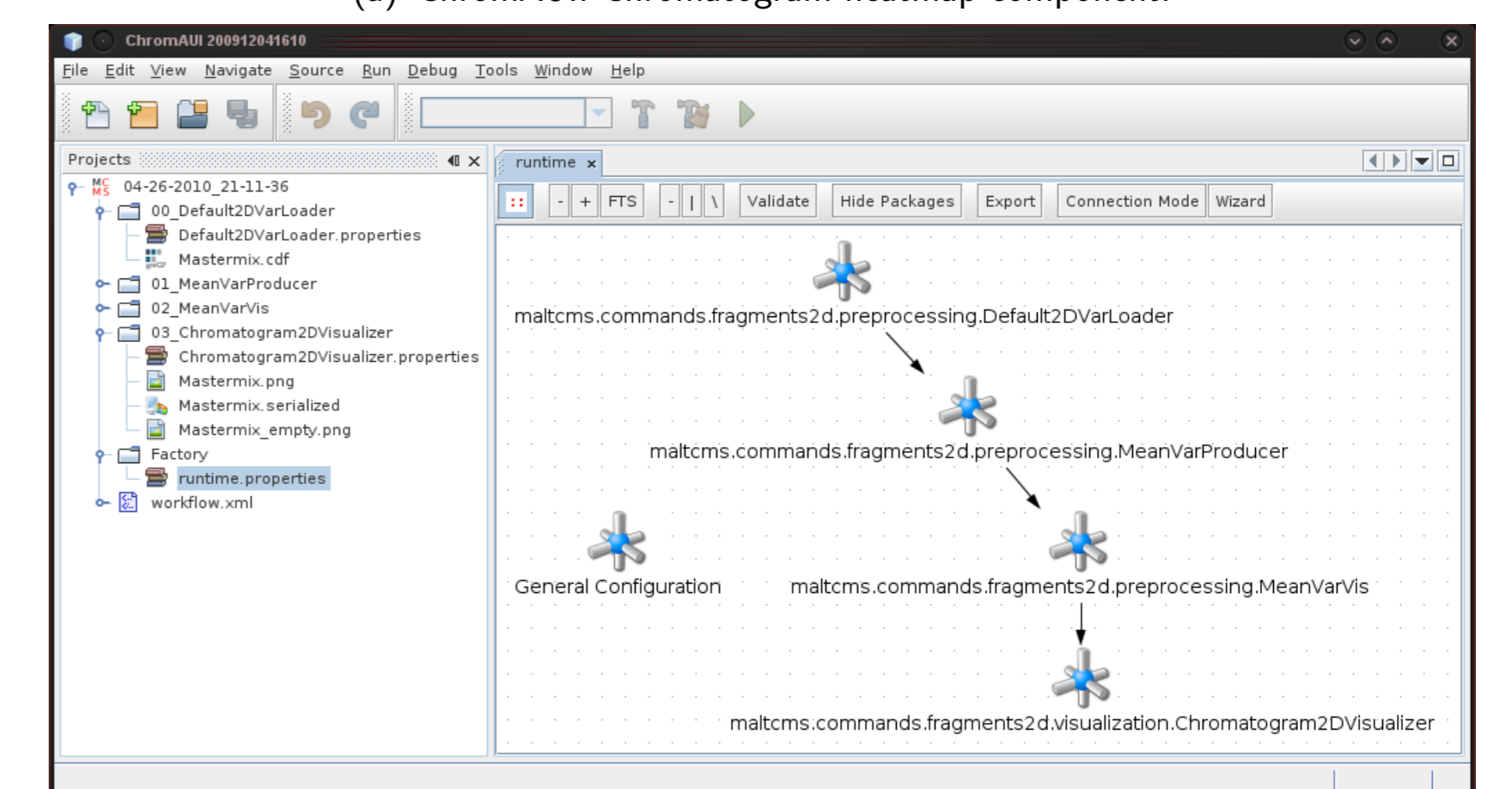
## ChromAUI

We currently focus on the development of a user-friendly, extensible GUI, *ChromAUI*. It provides tools for the creation and configuration of processing pipelines (Figure 5 (b)), as well as for the visual inspection, annotation and exploration of datasets from metabolomics and their processed results (Figure 5 (a)). ChromAUI is based on the Netbeans Rich Client Platform. Currently it is able to visualize all output formats of Maltcms, as well as raw *netcdf*-files. It will be

augmented by plugins for tasks like peak import, e.g. from MeltDB [2] and interactive visualizations.



(a) ChromAUI: Chromatogram heatmap component.



(b) ChromAUI: The pipeline editor component.

Figure 5: Different tools are available within ChromUI. (a) shows the chromatogram heatmap component, allowing user-defined color mappings, peak annotations and interactive retrieval and visualization of mass spectra for GC-MS, LC-MS and GCxGC-MS data. (b) shows a visual processing pipeline configuration in the pipeline editor.

## Outlook

Our future work is aimed at the parallelization of processing tasks on arbitrary grid infrastructures. While Maltcms is already able to execute different jobs on a local, multi-core host, we are currently implementing a system to dispatch tasks to arbitrary computation hosts, which can be either available via JAVA RMI, within a grid infrastructure (e.g. Oracle GridEngine) or within a private cloud infrastructure.

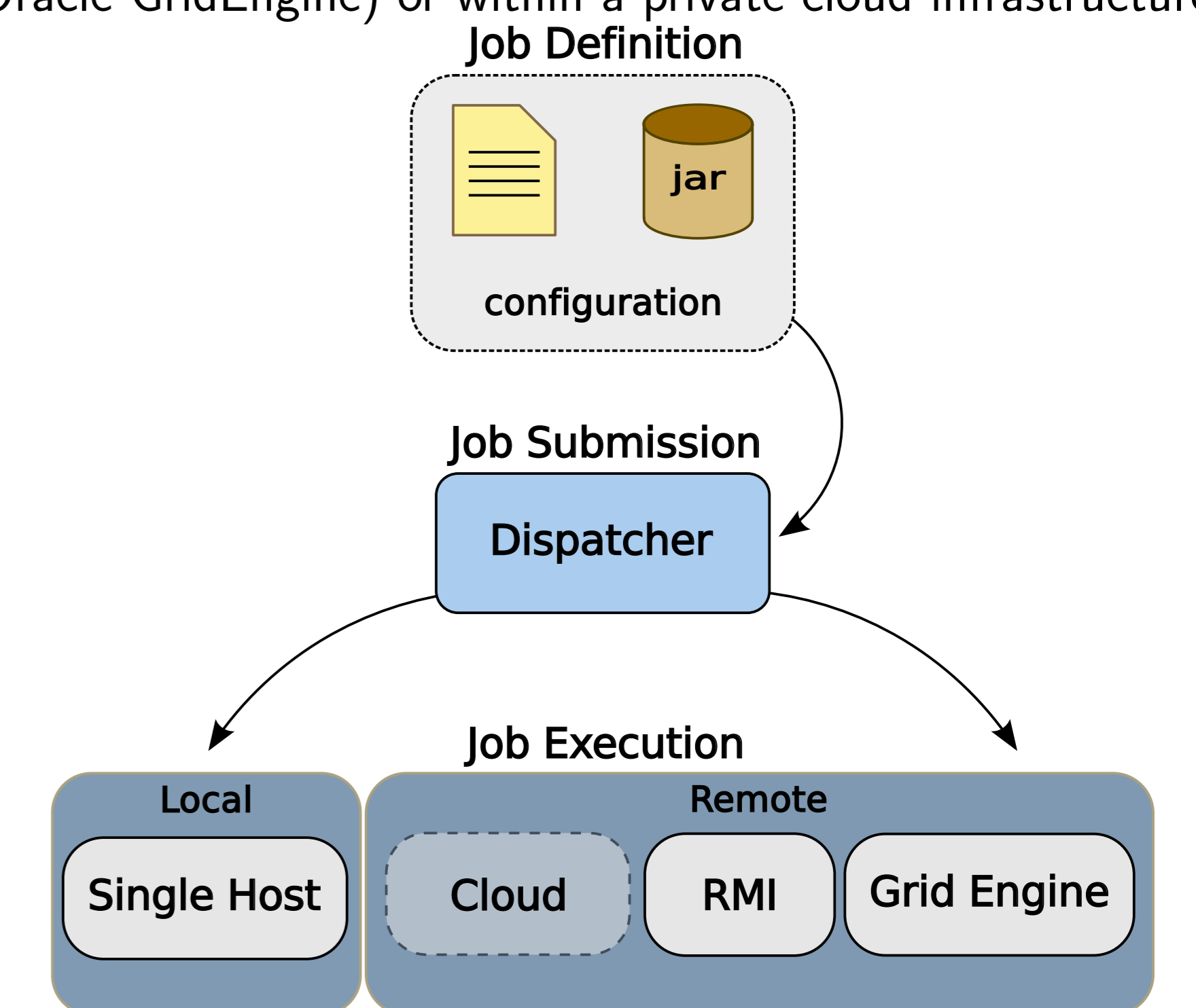


Figure 6: Structure of the future Maltcms job distribution system.

## Availability

Maltcms is freely available under the L-GPL v3 license at <http://maltcms.sourceforge.net>. It runs under all personal computer operating systems for which a JAVA Runtime Environment is available.

## 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] H Neuweger, S Albaum, M Dondrup, and M Persicke. MeltDB: a software platform for the analysis and integration of metabolomics experiment data. *Bioinformatics*, Jan 2008.