Technische Fakultät Universität Bielefeld **AG Genominformatik**

Maui and Maltcms

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

Nils Hoffmann^{1,2}, Mathias Wilhelm¹, Kai Stadermann¹, Jens Stoye¹

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

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).

similarity, green colors indicate a medium similarity, whereas blue colors indicate a low similarity.



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).

Preprocessi

BIPACE

Chrom

Import of Peak Features

Binning of Mass Spectra

Pairwise Similarity Calculation

VV vise Anchored DTW Calculat

Center Star Reference Selection

Alignment to Reference

Output of Multiple Alignment



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 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 massspectrometry 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 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.

000	Maui 201104080000							
								Q- Search (X+I)
Projects 🔘 🖗	Chromatogra	am View of mix1-MD.cdf 🛞	Table viev	v 🛞 🛛 succin	atB_pea	kAreas 🛞 succinatA_	peakAreas 🗵 mannitolB_peakAreas 🎗 mannitolA_peakAreas 🎗 glucoseB_peakAreas 🎗 glucoseA_peakAreas 🖇 Table view 🛇	< > -
D01_DefaultVarLoader	Table Chart							
002_MassFilter								
003_DenseArrayProducer	Cell selection							
004_TICPeakFinder	ScanNumber	RetentionTime	Score	RI	MW	Formula	D	SOURCE
005_EIMSDBMetaboliteAssignmer	93	467.5	933	1031.31	167	C8H13NOSi	M000413 A105003-101-xxx NA 1031.31 TRUE VAR5 ALK Pvridine, 2-hvdroxv- (1TMS)	/Users/nilshoffn
EIMSDBMetaboliteAssignment.	93	467.5	876	1154.12	167	C8H13NOSi	M000436_A117003-101-xxx_NA_1154.12_TRUE_VAR5_ALK_Pyridine, 4-hydroxy- (1TMS)	/Users/nilshoffn
glucoseA_peak_assignment.cs	93	467.5	875	1136.97	167	C8H13NOSi	M000419_A114003-101-xxx_NA_1136.97_TRUE_VAR5_ALK_Pyridine, 3-hydroxy- (1TMS)	/Users/nilshoffn
glucoseB peak assignment.cs	302	572.75	736	1104.75	249	C9H27NOSi3	M000416_A110002-101-xxx_NA_1104.75_PRED_VAR5_ALK_Hydroxylamine (3TMS)	/Users/nilshoffn
mannitolA peak assignment.c	454	649.844	912	1208.56	0	null	M000000_A123002-101-xxx_NA_1208.56_PRED_VAR5_ALK_NA	/Users/nilshoffn
mannitolB peak assignment c	454	649.844	795	1929.62	414	C18H34O5Si3	M000803_A194010-101-xxx_NA_1929.62_TRUE_VAR5_ALK_Gallic acid ethyl ester (3TMS)	/Users/nilshoffn
succinatA neak assignment c	454	649.844	723	1945.61	458	C19H38O5Si4	M000801_A196011-101-xxx_NA_1945.61_TRUE_VAR5_ALK_Gallic acid (4TMS)	/Users/nilshoffn
succinate peak assignment of	499	673.25	893	1167.37	256	C7H21O4PSi2	M000507_A119001-101-xxx_NA_1167.37_TRUE_VAR5_ALK_Phosphoric acid monomethyl ester (2TMS)	/Users/nilshoffn
Succinate_peak_assignment.c.	499	673.25	877	1857.73	0	null	M000000_A187004-101-xxx_NA_1857.73_PRED_VAR5_ALK_NA	/Users/nilshoffn
	499	673.25	626	1693.3	0	null	M000000_A171005-101-xxx_NA_1693.3_PRED_VAR5_ALK_NA	/Users/nilshoffn
007_PairwiseDistanceCalculator	585	717.985	735	1416.54	0	null	M000000_A144004-101-xxx_NA_1416.54_PRED_VAR5_ALK_NA	/Users/nilshoffn
MZIDynamicTimeWarp	609	730.25	733	1207.1	261	C11H27NO2Si2	M000030_A122001-101-xxx_NA_1207.1_TRUE_VAR5_ALK_Valine (2TMS)	/Users/nilshoffn
PairwiseAlignment	609	730.25	721	1230.43	261	C11H27N02Si2	M000516_A126001-101-xxx_NA_1230.43_PRED_VAR5_ALK_Norvaline, DL- (2TMS)	/Users/nilshoffn
PairwiseDistances	/22	789.3280000000001	852	1251.31	194	C10H1402Si	M000347_A128003-101-xxx_NA_1251.31_IRUE_VARS_ALK_Benzoic acid, (ITMS)	/Users/nilshoffn
PW_DISTANCE_glucoseA_glucc	/22	789.3280000000001	743	1887.23	325	C15H2/N03Si2	M0000055_A18900b-101-XXX_NA_1887.25_IKUE_VAK5_ALK_TYROSINE (2TMS)	/Users/nilshoffn
PW DISTANCE glucoseA mani	122	789.328000000001	729	1891.29	398	C18H3404SI3	MUU1U2U_A189U23-1U1-XXX_NA_1891.29_IKUE_VAKS_ALK_Lactic acid, 3-(4-hydroxyphenyl)- (31MS)	/Users/nilshoffn

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



(a) Top: Scheme of pipeline definition and execution in (b) Processing Bischeme Maltcms. Bottom: Schematic of remote execution in- PACE/ChromA. frastructure.

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 allagainst-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.



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

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

- N Hoffmann and J Stoye. ChromA: signal-based retention time alignment for chromatography-mass spectrometry data. *Bioinformatics*, 25(16):2080–1, 2009.
- M D Robinson, D P De Souza, W Keen, E C Saunders, M J Mcconville, T P [2] 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