Updates to **xcms**: simplified raw data access and enhanced MS level > 1 capabilities

Johannes Rainer[1], **Laurent Gatto**[2] and **Steffen Neumann**[3,4]

[1]Institute for Biomedicine, Eurac Research, Bolzano, Italy. [2]De Duve Institute, UCLouvain, Brussels, Belgium. [3]Department of Stress and Developmental Biology, Leibniz Institute of Plant Biochemistry, Halle, Germany. [4]German Centre for Integrative Biodiversity Research, Halle-Jena-Leipzig, Germany.

1 Abstract

The **xcms** Bioconductor package is one of the standard toolboxes for the preprocessing of untargeted metabolomics data. Here we present recent updates to **xcms**, which re-use and build upon the support for memory-efficient parallel processing capabilities in the **MSnbase** Bioconductor software package for proteomics and general mass spectrometry data handling.

We have improved large-scale experiment data analysis through memory-efficient parallel processing capabilities and simplify raw spectra data access throughout the whole preprocessing task. This comprises also dedicated functionality to extract ion chromatograms/traces from the original files and to perform chromatographic peak detection directly on such chromatographic data. Besides paving the road for MRM/SRM data analysis with **xcms**, it also allows to evaluate different peak detection settings on selected signals before applying them to the whole data set. Along these lines, we also implemented new visualization capabilities aiding in the definition and evaluation of data set-specific settings for the various preprocessing algorithms. Finally, import of MRM/SRM raw data has been added and a framework for the identification of MS2 spectra for identified chromatographic peaks was implemented.

2 Updates to xcms

- Re-using objects from Bioconductor's MSnbase package ensures native support for MS level > 1 (MSn) data.
- MS peak data is read on-demand from original input files (applying potential data processing steps on-the-fly).
- Simplified extraction of ion chromatograms (EIC, BPC, TIC, ...):

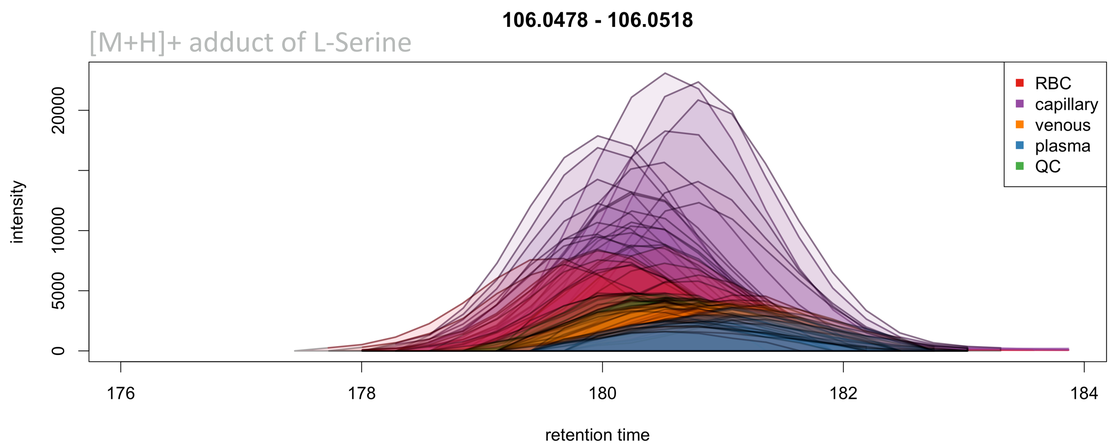
retention time range — m/z range

```
chr <- chromatogram(data, rt = c(176, 184), mz = c(106.047, 106.051))
```

- Returned data (`Chromatograms`) contains chromatographic data for each sample, sample annotation and, if available, identified chromatographic peaks.
- Perform peak detection on extracted ion chromatograms (EIC) to evaluate and finetune settings.

```
cwp <- CentWaveParam(ppm = 40, peakwidth = c(2, 10), snthresh = 5)
xchr <- findChromPeaks(chr, param = cwp)
plot(xchr, col = sample_colors, peakBg = peak_colors)
```

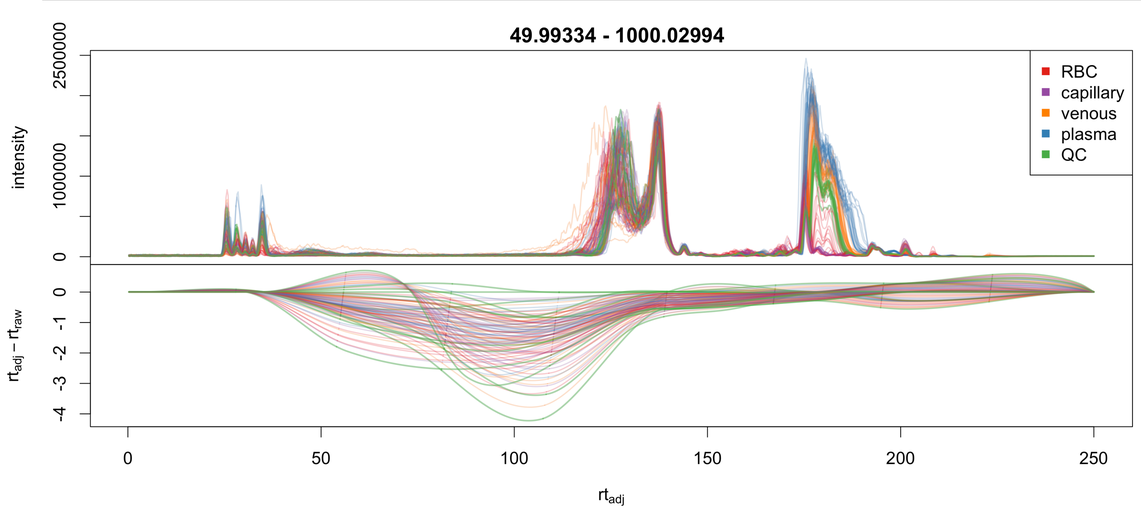
color for each line — background color for each peak



- Using `chromatogram` after peak detection: `Chromatograms` contains all identified chromatographic peaks.
- `featureChromatograms` to extract ion chromatograms for all features (after correspondence analysis).

- Subset-based alignment: estimate retention time shifts on data subset (e.g. QC samples) and adjust full data using these (interpolating based on injection order of non-subset and subset samples).

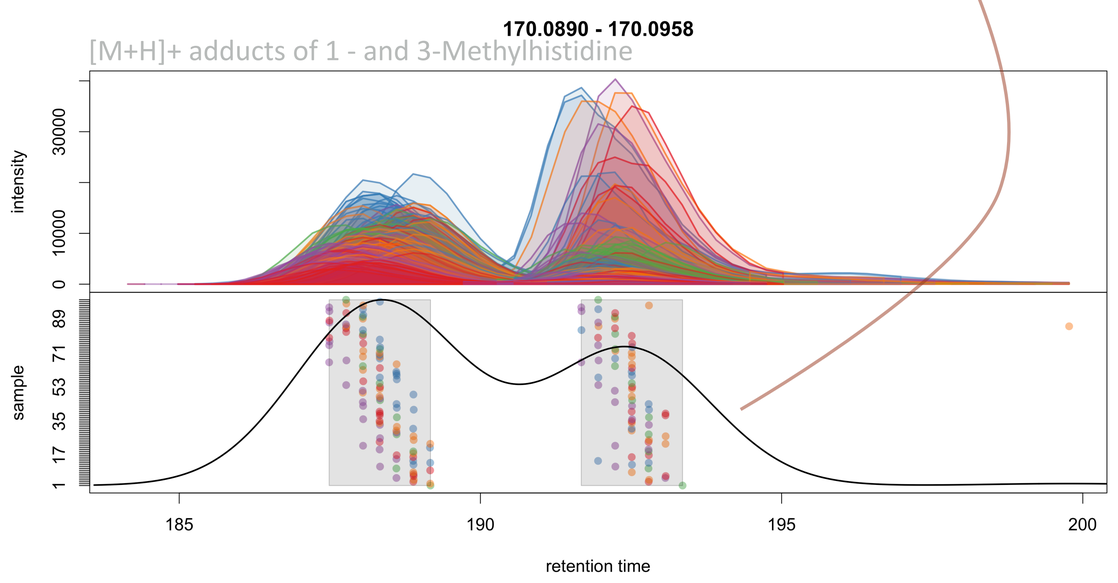
```
owp <- ObiwarpParam(binSize = 0.1, subset = which(data$sample == "QC"))
data <- adjustRtime(data, param = owp)
plot(chromatogram(data, aggregationFun = "max"), peakType = "none",
                  col = sample_colors)
plotAdjustedRtime(data, col = sample_colors)
```



- MSn spectra are always aligned along MS1 spectra.
- Evaluate settings for *peak density*-based correspondence analysis on a m/z slice or EIC:

```
chr <- chromatogram(data, rt = c(183, 200), mz = c(170.089, 170.096))
pdp <- PeakDensityParam(sampleGroups = data$sample_group, bw = 1.3)
plotChromPeakDensity(chr, param = pdp, col = source_colors,
                     peakBg = peak_colors, peakPch = 16)
```

defines *smoothness*



- MSn data: `chromPeakSpectra` and `featureSpectra` allow to extract MS2 spectra with a precursor m/z (and retention time) within the m/z and retention time boundaries of a chromatographic peak. These can be further processed with e.g. the `combineSpectra` function from **MSnbase**.
- Example workflow will be available for GNPS.

3 Conclusion and Outlook

The recent updates made **xcms** an even more useful tool for the preprocessing and analysis of untargeted metabolomics data, specifically for large scale experiments. Next versions will provide support for the analysis of MSn data from data independent acquisition experiments (e.g. SWATH) and will also base on an improved MS data infrastructure in R (https://RforMassSpectrometry.org).

Documentation of the new **xcms** functionality is available at:

- https://bioconductor.org/packages/xcms
- https://github.com/jorainer/metabolomics2018

Acknowledgements:
We thank Michael Witting and Jan Stanstrup for contributions and discussions.

johannes.rainer@eurac.edu

jo_rainer

jorainer