MassChroQ: a complete tool for mass spectrometry-based proteomic quantification

Edlira Nano[1], Benoît Valot[1], Olivier Langella[2], Ludovic Bonhomme[1], Mélisande Blein-Nicolas[1], Michel Zivy[2].

[1]INRA, [2]CNRS, Plateforme d'Analyse Protéomique de Paris Sud-Ouest, UMR de Génétique Végétale, 91190 Gif-sur-Yvette, France.

The context

Recent rapid developments in quantitative mass-spectrometry methods have evidenced the need to efficiently handle large amounts of complex data. Various quantification tools exist, but they often lack flexibility by being specific to data or system types.

Our goal

- To automatically and efficiently quantify the large amounts of data obtained from Liquid Chromatography - Mass Spectrometry experiments.
- To be able to analyse data obtained from:
  - both high and low resolution spectrometers,
  - label-free as well as isotopic labelling experiments,
  - complex sample treatments (e.g. protein or peptide fractionation).

MassChroQ traits

I

MassChroQ (Mass Chromatogram Quantification) can quantify given m/z values or the identified peptides of your data by performing : XIC extraction from mzXML/mzML LC-MS data files, signal denoising, peak detection, retention time alignment and peak matching.

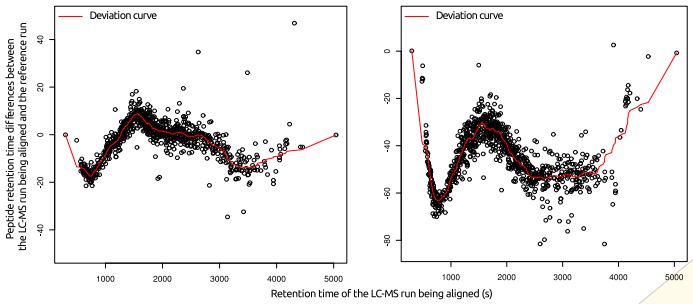
MassChroQ is:

- configurable (every analysis parameter can be user-adjusted);
- traceable (it produces precise alignment and quantification traces);
- modular (integration of/in other tools as OBI-Warp [2] and PROTICdb [3]);
- fast (full analysis of centroid data: < 1min/Go, of profile data: < 3 min/Go);
- light (uses at most 350Mo of RAM, your computer stays fully functional);
- an open-source project (feel free to use it and welcomed to contribute).

II Alignment of LC-MS runs

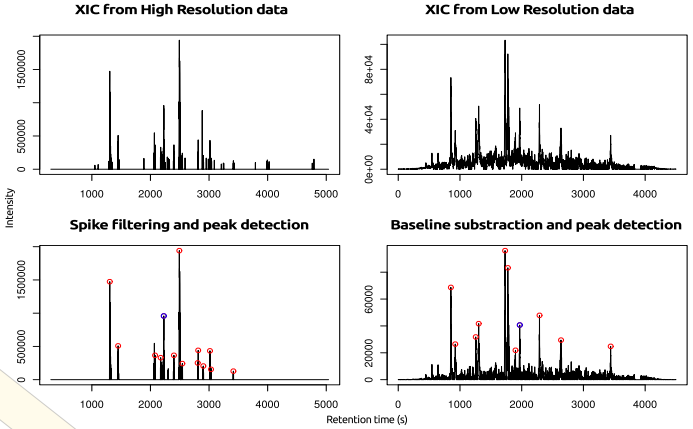
- Alignment is needed to compensate retention time (RT) variations caused by uncontrolled deviations between LCs.
- If a sufficiently large number of peptides is identified, MassChroQ can align two runs by using the RT deviations of the peptides they have in common (figure below). Otherwise, the OBI-Warp alignment method can be used.

Examples of MS2 alignment evaluation of deviation curves



III Quantification

After XIC extraction, various configurable XIC filters can be applied. Peak detection is then performed and peak area (quantification value) is computed.



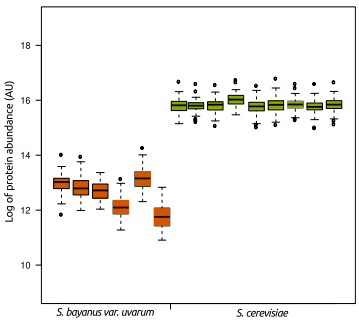
MassChroQ applications

I

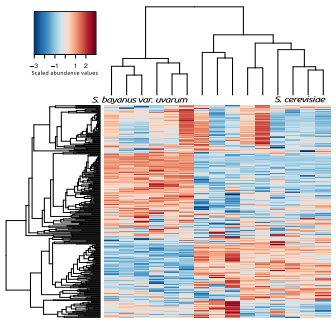
A study of genetic diversity in yeast

(Blein-Nicolas et al, in prep)

- Label-free analysis of:
  9 strains of *Saccharomyces cerivisiae*
  6 strains of *Saccharomyces bayanus var. uvarum*
- 3 independent replicates per strain, resulting in 45 samples, were analysed by LC-MS/MS with an LTQ-Orbitrap (Thermo) high resolution spectrometer.
- MassChroQ quantified, in a total of 250 Go of mzXML data:
  6684 peptides among 622 identified proteins in 25 minutes.
- 360 proteins were shown to vary significantly within and between species.



Abundance of the TDH3 protein in the different strains. S. bayanus var.uvarum strains are clearly discriminated from the S. cerevisiae ones



Scaled abundance values heatmap of all the quantified proteins in each strain. Strain and protein classification trees are displayed above and left side, respectively.

II

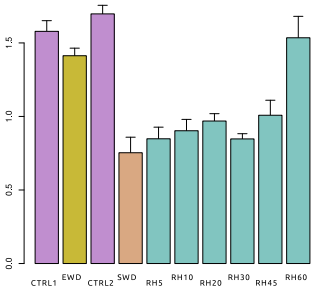
A quantitative phosphoproteomics study

(Bonhomme et al, in prep)

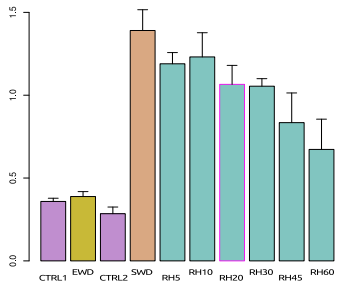
In-vivo changes during drought in the growing zone of maize leaves.

- 40 maize leaf protein lysates treated with:
  dimethyl isotope labelling + SCX-fractionation
  + IMAC phosphopeptide enrichment.
- The 400 resulting fractions were:
  analysed by LC-MS³ with an LTQ (Thermo) spectrometer, resulting in 450 Go of mzXML data.
- 3659 phosphosites accounting for 2462 distinct proteins were identified using the *X!Tandem pipeline**, before being quantified in MassChroQ.
- Significant quantitative variations of phosphopeptide abundances were detected in both deficit and rehydration regimes.

Phosphosite of v-SNARE protein



Phosphosite of 6-phosphofructokinase protein



Histograms of relative abundance of phosphosite in the different water regimes:

EWD - early water deficit *CTRL1 - control for EWD* *RHn - n minutes after rehydration*
SWD - severe water deficit *CTRL2 - control for SWD*

Perspectives

- A full description and evaluation of MassChroQ can be found in [1].
- Further documentation and download possibilities are available at http://pappso.inra.fr/bioinfo/masschroq, community tools at https://sourcesup.cru.fr/projects/masschroq/
- Future work will focus on the development of a graphical-user interface, an MRM analysis mode, automatic parameter detection, and your suggestions.

[1] Valot B., Langella O., Nano E., Zivy M., MassChroQ: A versatile tool for mass spectrometry quantification. Proteomics, 2011, 11: 3572–3577.

[2] Prince J.T., Marcotte E.M., Chromatographic alignment of ESI-LC-MS proteomics data sets by ordered bijective interpolated warping. Anal. Chem 2006, 78, 6140-6152.

[3] Ferry-Dumazet H., Houel G., Montalent P., Moreau L., Langella O., Negroni L., Vincent D., Lalanne C., de Daruvar A., Plomion C., Zivy M., Joets J. PROTICdb: a web-based application to store, track, query and compare plant proteome data. Proteomics, 2005, 8:2069-81.

* The X!Tandem pipeline homepage: http://pappso.inra.fr/bioinfo/xtandempipeline/

This work was partially supported by IBiSA

SourceSup hosts the project

GPL Free Software Free as in Freedom

MassChroQ : A versatile tool for mass spectrometry quantification

B. Valot, O. Langella, E. Nano, M. Zivy, Proteomics, 2011, 11:3572-3577.

Abstract

Recently, many software tools have been developed to perform quantification in LC-MS analyses. However, most of them are specific to either a quantification strategy (e.g. label-free or isotopic labelling) or a mass-spectrometry system (e.g. high or low resolution). In this context, we have developed MassChroQ, a versatile software that performs LC-MS data alignment and peptide quantification by peak area integration on extracted ion chromatograms. MassChroQ is suitable for quantification with or without labelling and is not limited to high resolution systems. Peptides of interest (for example all the identified peptides) can be determined automatically or manually by providing targeted m/z and retention time values. It can handle large experiments that include protein or peptide fractionation (as SDS-PAGE, 2D-LC). It is fully configurable. Every processing step is traceable, the produced data are in open standard format and its modularity allows easy integration into proteomic pipelines. The output results are ready for use in statistical analyses. Evaluation of MassChroQ on complex label-free data obtained from low and high resolution mass spectrometers showed low CVs for technical reproducibility (1.4%) and high coefficients of correlation to protein quantity (0.98). MassChroQ is freely available under the GNU General Public Licence v3.0 at http://pappso.inra.fr/bioinfo/masschroq/.