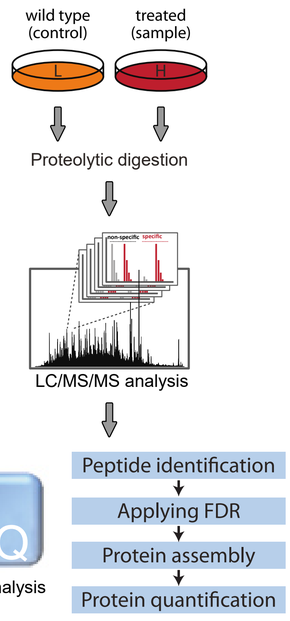
A New Galaxy Wrapper for MaxQuant facilitates Mass Spectrometry-based Proteomics Data Analysis

Damian Glätzer[1], Melanie Föll[1,3], Bettina Warscheid[1,2], Oliver Schilling[2,3], Friedel Drepper[1,3]

[1]Faculty of Biology & [2]CIBSS Centre for Integrative Biological Signalling Studies, University of Freiburg

[3]Institute of Surgical Pathology; University Medical Center, University of Freiburg

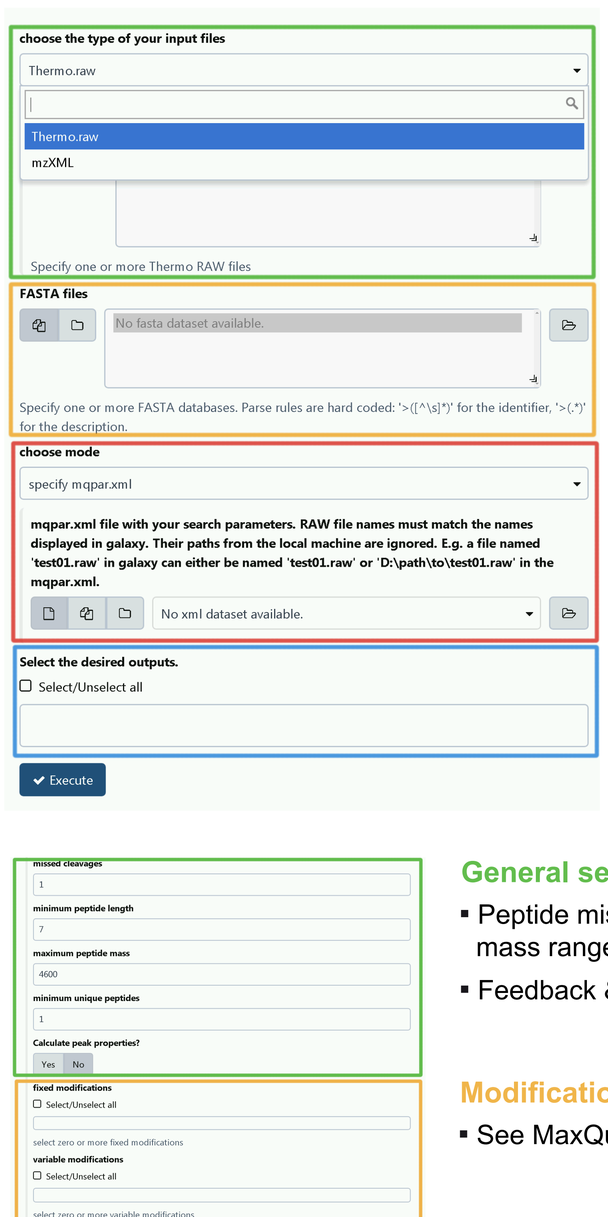
Introduction



MaxQuant is a popular software for the quantitative analysis of large datasets from mass spectrometry based shotgun proteomics experiments [1-3]. Initially restricted to the Microsoft Windows operating system, its recent versions are compatible with Mono, the open source implementation of Microsoft's .NET framework. This allows MaxQuant to be distributed as a Conda package and run on Linux-operated clusters. We present a Galaxy tool that provides a wrapper for MaxQuant.

Workflow of a shotgun proteomics experiment and data analysis. Main steps of data processing in MaxQuant are indicated.

I. Graphical User Interface



Input MS data files

- Accepts Thermo rawfiles or mzXML
- Specify file type (Galaxy detection fails).

Input sequence database files

- Fasta format
- Coming soon: Identifier parse rules support

Mode selection

- Upload your own mqpar.xml - or
- Use GUI for parameter selection (see Figure 3 below)

Output selection

- Most of the tabular outputs of MaxQuant
- gzipped results directory

General search parameters

- Peptide missed cleavages, length and mass ranges
- Feedback & suggestions welcome

Modifications and enzymes

- See MaxQuant distribution for full list

MS1 label based quantification

- SILAC light, medium and heavy labels

Label free quantification

- LFQ [4] parameters and iBAQ [5]
- More tests needed

Conclusions & Outlook

- We provide a Galaxy wrapper based on Mono and the MaxQuant Conda package
- Supports label-based (SILAC) and label-free quantification
- Other quantification strategies, such as MS2 level iTRAQ and TMT, are currently possible via pre-configured mqpar.xml files

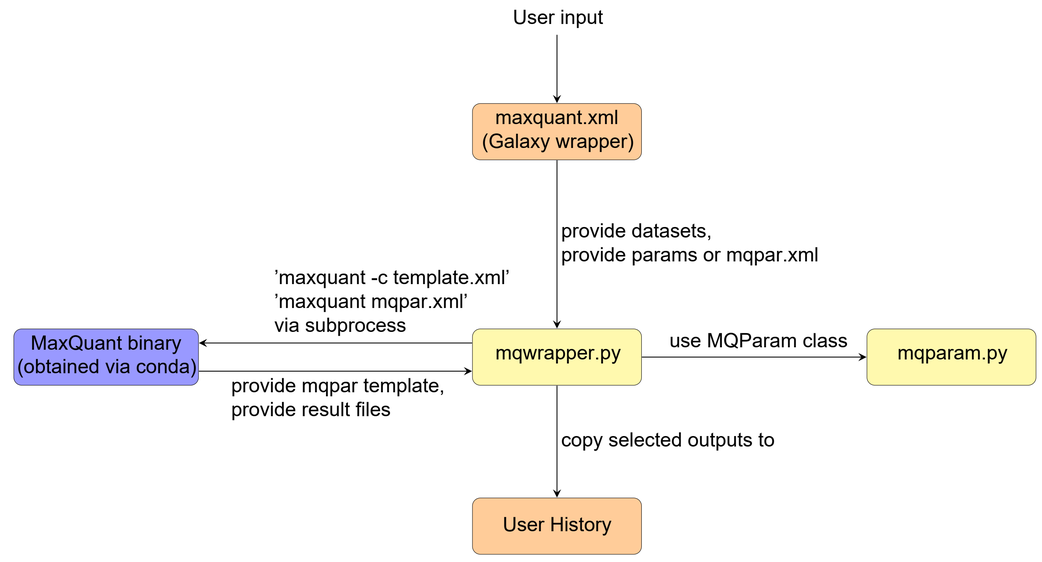
Next steps

- Integration of parameter groups (e.g., multiple proteases, labeled plus label-free)
- Setup a workflow for quality control

Further information

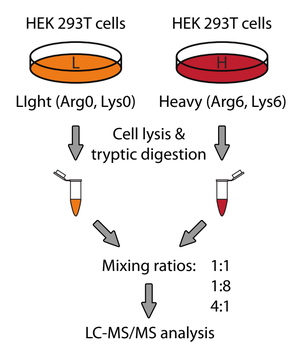
- https://usegalaxy.eu/tool_runner?tool_id=maxquant
- bioconda: MaxQuant version 1.6.3.4
- https://github.com/dglaetzer/tools-galaxyp/tree/master/tools/maxquant

II. Code Structure



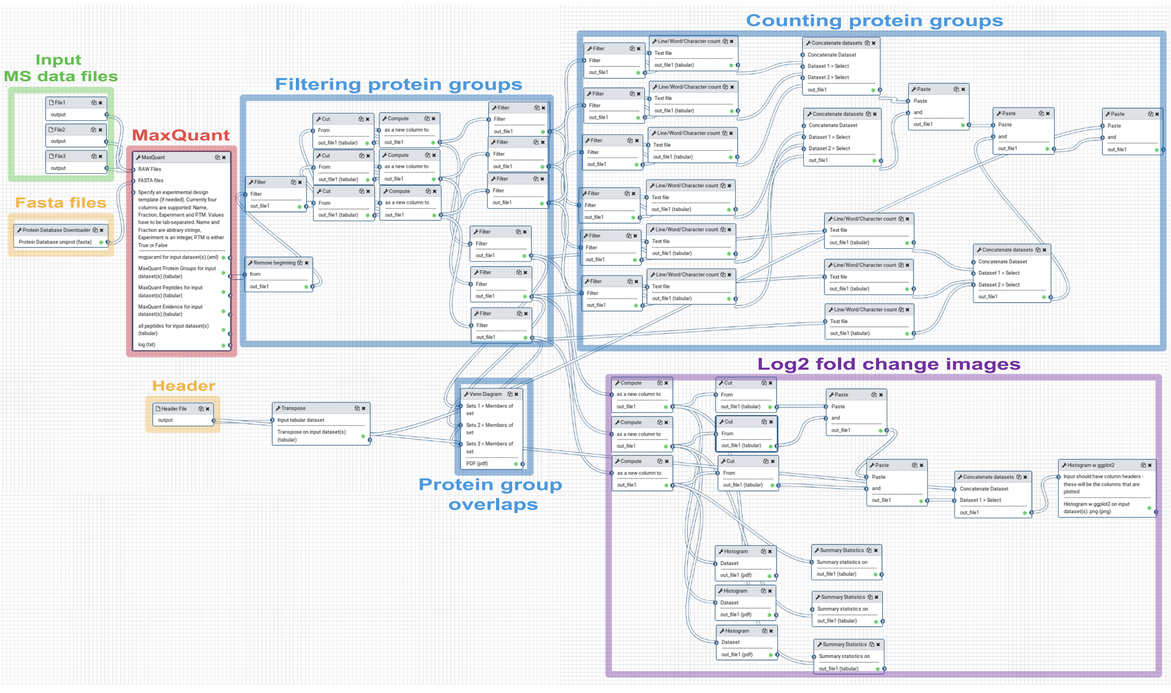
MaxQuant galaxy wrapper schematic structure. Input parameters are transferred into xml parameter file. MaxQuant binary is run as a subprocess. Selected output files are provided in the user history.

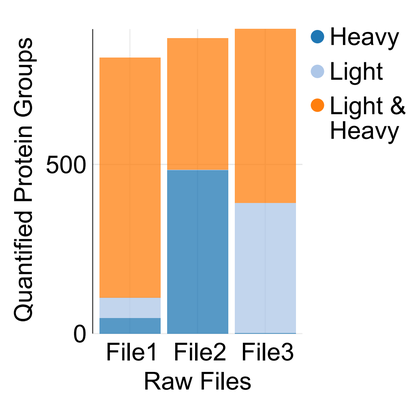
III. Galaxy Workflow for SILAC Analysis



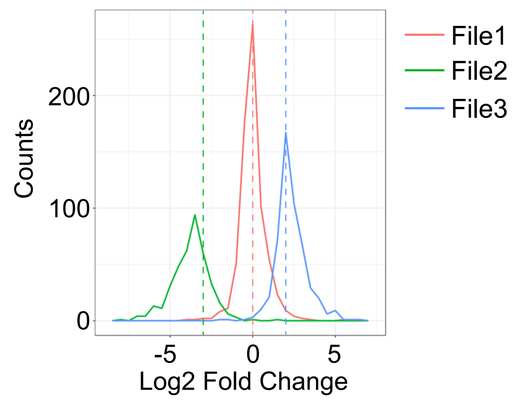
Fixed Ratio SILAC Experiment. HEK293 cells were grown in light or heavy (Arg6, Lys6) SILAC medium. Cells from both conditions were mixed together in different ratios. LC-MS/MS on a Q-Exactive Plus mass spectrometer. Expected log2 fold changes are 0, -3 and 2.

Galaxy Workflow for SILAC ratio analysis.
https://usegalaxy.eu/u/melanie-foell/w/maxquant-silac-ratio-files





Overview of protein identification



Histogram of protein SILAC ratios

References:

[1] Cox & Mann (2008) Nat Biotechnol 26, 1367–72
[2] Cox et al. (2011) J Proteome Res 10, 1794–1805
[3] Tyanova et al. (2016) Nat Protoc 11, 2301-19
[4] Cox et al. (2014) Mol Cell Proteomics 13, 2513-26
[5] Schwanhäusser et al. (2011) Nature 473, 337-42

Acknowledgements:

This project was supported by the Deutsche Forschungsgemeinschaft through the Excellence Initiative (BIOSS - EXC 294) and the Excellence Strategy (CIBSS - EXC 2189). We thank Christian Peikert, Franziska Elsaesser, John Chilton and Björn Grüning for their help especially during the initial phase of this work and Stefan Tholen for providing the MS data files.