Developments and opportunities with Workflow4Metabolomics

Lain Pavot, Mélanie Pétéra, Nils Paulhe, Yann Guitton, Gildas Le Corguillé, Julien Saint-Vanne, Marie Tremblay-Franco, Cécile Canlet, Jean-François Martin, Ralf Weber, et al.

Developments and opportunities with Workflow4Metabolomics

Lain PAVOT[1], Mélanie PÉTÉRA[1], Nils PAULHE[1], Yann GUITTON[2], Gildas LE CORGUILLÉ[3], Julien SAINT-VANNE[2], Marie TREMBLAY-FRANCO[4,5], Cécile CANLET[4,5], Jean-François MARTIN[4,5], Ralf WEBER[9,10], Cédric DELPORTE[6], Florence SOUARD[7], Céline DALLE[8], W4M Core Team[1,2,3,4,5,6,7,8,9,10], Franck GIACOMONI[1]

[1] *Université Clermont Auvergne, INRAE, UNH, Plateforme d'Exploration du Métabolisme, MetaboHUB Clermont, F-63000 Clermont-Ferrand, France*
[2] *Oniris, INRAE, LABERCA, 44300, Nantes, France*
[3] *ABiMS, FR2424 CNRS-UPMC, Station Biologique, Place Georges Teissier, 29680, Roscoff, France*
[4] *Toxalim – Research Center in Food Toxicology, Toulouse University, INRAE UMR 1331, ENVT, INP-Purpan, Paul Sabatier University, F-31027 Toulouse, France*
[5] *Metatoul-AXIOM platform, National Infrastructure for Metabolomics and Fluxomics, MetaboHUB, Toxalim, INRAE UMR 1331, F-31027 Toulouse, France*
[6] *RD3-Pharmacognosy, Bioanalysis and Drug discovery Unit & Analytical Platform of the Faculty of Pharmacy, Faculty of Pharmacy, Université libre de Bruxelles (ULB), 1050 Brussels, Belgium*
[7] *DPP-Department of Pharmacotherapy and Pharmaceutics, Faculty of Pharmacy, Université libre de Bruxelles (ULB), 1050 Brussels, Belgium*
[8] *French Armed Institute of Biomedical Research (IRBA), Analytical Development and Bioanalysis Unit, F-91220 Brétigny-sur-Orge, France*
[9] *School of Biosciences, University of Birmingham, Birmingham, B15 2TT, UK*
[10] *Phenome Centre Birmingham, University of Birmingham, Birmingham, B15 2TT, UK*

Introduction

Metabolomics data analysis is a complex, multistep process, constantly evolving with the development of new analytical technologies, mathematical methods, bioinformatics tools and databases. **Workflow4Metabolomics**[1] is a collaborative portal dedicated to metabolomics data processing, analysis and annotation for the Metabolomics community. In the latest version of W4M, the core team proposes new upgrades for LC-MS, GC-MS and NMR pipelines, including new preprocessing steps, as well as enhancement of statistical analysis and annotation tools. W4M aims to promote open science in Metabolomics and facilitate knowledge dissemination by providing community resources.

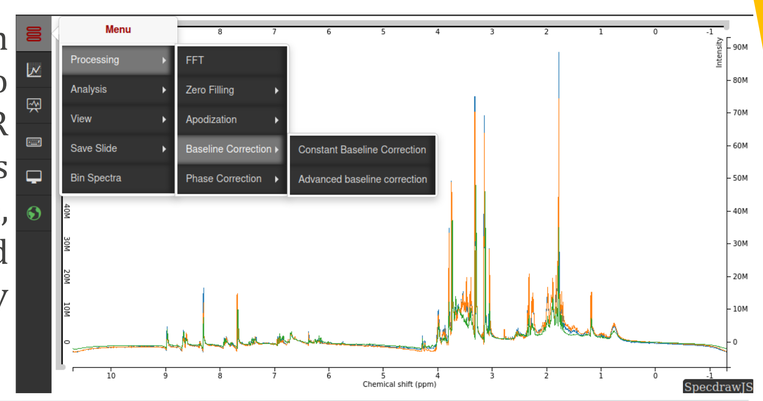
New W4M LC-MS, LC-MSMS, GC-MS and NMR pipelines

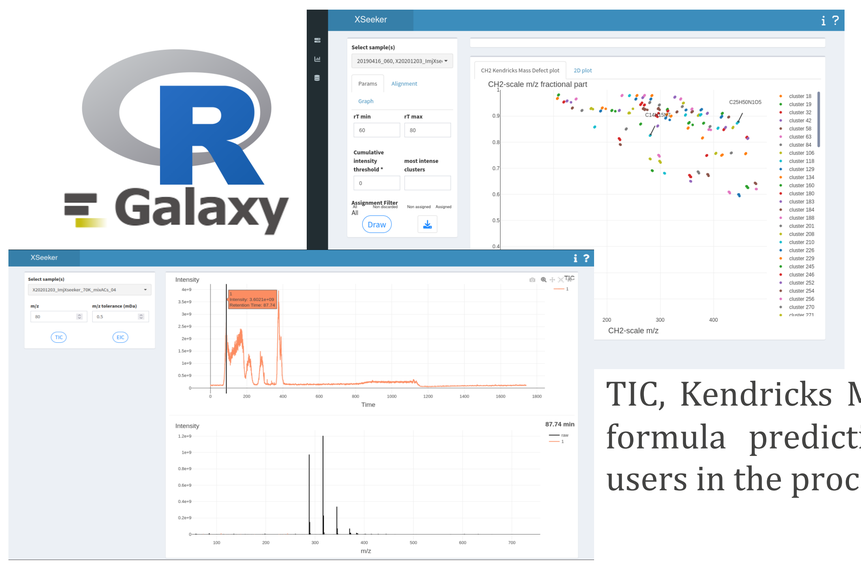
The W4M core team makes continuous upgrades including pipeline enhancement, tool updates and new component development. The latest version of the W4M infrastructure integrates improvements in several parts of the current tool suite, including statistics (*e.g.* new features for the 'mixmodel: ANOVA for repeated measures statistics' tool), annotation (*e.g.* a new version of the '2D NMR Annotation' tool for complex mixture bidimensional NMR spectra)... New tools were also added to the catalogue, as the 'MS2snoop' module to investigate spectra of standards further using the results of the 'MSPurity[2]' MS tool suite. A suite of tools (BankInHouse for GCMS and LCMS, NMR peak-matching) should be proposed to connect with the PeakForest project (peakforest.org[3]), a multi-platform digital infrastructure for interoperable metabolite spectral data and metadata management.

W4M recent works also cover the development of Galaxy Interactive Tools, with for example two tools for dynamic data visualisation and annotation:

- XSeeker - a modified version of HaloSeeker[4] to visualize and annotate MS spectra;
- NMRPro[5] - for interactive processing and visualization of NMR spectra.

NMRPro is integrated and can be used in Galaxy [6] to analyse different kind of NMR spectra. Processing such as phase correction, apodization, FFT or peak picking and integration are currently available.





XSeeker is available in Galaxy too, and uses results from the 'XCMS + CAMERA' workflow to help annotating clusters. It provides multiple graphic representations of MS data such as EIC, TIC, Kendricks Mass Defect Plot and uses formula prediction and scoring to help users in the process of clusters annotation.

Open Science - sharing - training: contributing to the community

W4M is fully involved in the Open science movement. In addition to providing Digital Object Identifiers (DOI) for user data history references[7], we invest in training materials based on the Galaxy Training Network (**GTN**)[7] concept and format.

What is the GTN?

Online, open-access resources about data analysis

Interactive learning via hands-on tutorials built around a 'research story'

One website, aggregating training material covering many current research topics

Developed and maintained by the community in a collaborative way using GitHub

What are W4M GTN contributions?

Get to grips with metabolomics data analysis at your own pace using Galaxy Training Materials provided by W4M!

- Galaxy Training! Introduction to Metabolomics
- Galaxy Training! Mass spectrometry: LC-MS analysis
- Galaxy Training! Mass spectrometry: LC-MS preprocessing with XCMS
- Galaxy Training! Mass spectrometry: LC-MS data processing
- Galaxy Training! Mass spectrometry: MSMS analysis with msPurity package (coming soon)

Latest provided materials

W4M contributes to and promotes the use of metabolomics software and open-source software package production.
Contributors share new metabolomics data analysis tools compliant with the galaxy framework on the GitHub **"tools-metabolomics"** repository.

The **W4M** infrastructure is hosted on the **usegalaxy.fr** facility that now provides **t.i.a.s.s** which is a secured space opened on the cluster during your training session

Save the date! W4M keeps up its international school **Workflow4Experimenters**. Next edition will take place in Paris: 20-24th of March, 2023

Special thanks to Jean-François Martin, core team member from the outset, for his long career and all the shared good times!

Join us on workflow4metabolomics.org!

W4M web portal **W4M pipelines** **Events** **Twitter** **Mailing list**

Acknowledgments This work is supported by the French Ministry of Research and National Research Agency as part of the French metabolomics and fluxomics infrastructure (MetaboHUB-ANR-INBS-0010).

Funding: La Région Auvergne-Rhône-Alpes, Région Bretagne, chopin, Investissements d'Avenir

[1] Guitton Y. et al. "Create, run, share, publish, and reference your LC-MS, FIA-MS, GC-MS, and NMR data analysis workflows with the Workflow4Metabolomics 3.0 Galaxy online infrastructure for metabolomics." The international journal of biochemistry & cell biology vol. 93 (2017): 89-101. doi:10.1007/s11306-022-01899-3
[2] Lawson T. N. et al. "msPurity: Automated Evaluation of Precursor Ion Purity for Mass Spectrometry-Based Fragmentation in Metabolomics." Analytical chemistry vol. 89,4 (2017): 2432-2439. doi:10.1021/acs.analchem.6b04358
[3] Paulhe N. et al. "PeakForest: a multi-platform digital infrastructure for interoperable metabolite spectral data and metadata management." Metabolomics : Official journal of the Metabolomic Society vol. 18,6 40, 14 Jun. 2022, doi:10.1007/s11306-022-01899-3
[4] Léon A. et al. "HaloSeeker 1.0: A User-Friendly Software to Highlight Halogenated Chemicals in Nontargeted High-Resolution Mass Spectrometry Data Sets." Analytical chemistry vol. 91,5 (2019): 3500-3507. doi:10.1021/acs.analchem.8b05103
[5] Mohamed A. et al. "NMRPro: an integrated web component for interactive processing and visualization of NMR spectra." Bioinformatics (Oxford, England) vol. 32,13 (2016): 2067-8. doi:10.1093/bioinformatics/btw102
[6] Afgan E. et al. "The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update." Nucleic acids research vol. 46,W1 (2018): W537-W544. doi:10.1093/nar/gky379
[7] Batut B. et al. "Community-Driven Data Analysis Training for Biology." Cell systems vol. 6,6 (2018): 752-758.e1. doi:10.1016/j.cels.2018.05.012