Workflow4Metabolomics: an international computing infrastructure for Metabolomics

Mélanie Pétéra[1], **Romain Dallet**[2], Christophe Duperier[1], Franck Giacomoni[1], Yann Guitton[3], Jean-François Martin[4], Pierrick Roger-Mele[5], Julien Saint-Vanne[3,6], Etienne Thévenot[5], Marie Tremblay-Franco[4], **Gildas Le Corguillé**[6]

[1]INRA, UMR 1019, PFEM, 63122, Saint Genes Champanelle, France

[2]Institut Français de Bioinformatique, CNRS UMS 3601, rue John Von Neumann, 91403 Orsay, France

[3]LUNAM Université, Oniris, Laboratoire d'Etude des Résidus et Contaminants dans les Aliments (LABERCA), Nantes, F-44307, France

[4]Toxalim, Université de Toulouse, INRA, INP-ENVT, INP-EI-Purpan, Univ. Toulouse 3 Paul Sabatier, 31027 Toulouse, France

[5]CEA, LIST, Laboratory for Data Analysis and Systems' Intelligence (LADIS), MetaboHUB, F-91191 Gif-sur-Yvette, France

[6]UPMC, CNRS, FR2424, ABiMS, Station Biologique, 29680, Roscoff, France

Abstract

Metabolomics data analysis is a complex, multistep process, which is constantly evolving with the development of new analytical technologies, mathematical methods, and bioinformatics tools and databases. The Workflow4Metabolomics (W4M) project aim to develop full **LC/MS**, **GC/MS**, **FIA/MS** and **NMR** pipelines using **Galaxy** framework for data analysis including preprocessing, normalization, quality control, statistical analysis and annotation steps.

The W4M Core Team is fully involved building tools for the metabolomics community and its does particular efforts in tool quality and for disseminate their work. Wrappers are openly available on GitHub and automatically tested using Planemo on the TravisCI. The dependencies are managed with Conda packages. Eventually, wrappers are distributed on the ToolShed. Meanwhile, the Workflow4Metabolomics instance provides an on-line, user-friendly and high-performance environment to build, run and share workflows.

Our current developments aim to provide a set of interactive visualization tools in order to make ease the results interpretations. The development of **Shiny** applications, executable through Galaxy Interactive Environments, will allows interactions from graphical features and dataset filters with graphical outputs like chromatograms, RMN spectra, heatmaps or PCA. In parallel, because one of the major issue of the metabolomic approach is the compounds identification. To facilitate this annotation step, tandem mass spectrometry (MS/MS) is able to provide informations about the compounds structure. For that reason, an MS/MS data processing workflow based on msPurity, and two identification tools, metFrag and Sirius-CSI: FingerID, will be available in W4M.

References

1. Giacomoni F., Le Corguillé G. *et al.* (2014), *Workflow4Metabolomics: A collaborative research infrastructure for computational metabolomics*, Bioinformatics
2. 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, 2017, ISSN 1357-2725