Workflow4Metabolomics: A collaborative research infrastructure for computational metabolomics

Mélanie Petera, Gildas Le Corguille, Marion Landi, Misharl Monsoor, Marie Tremblay Franco, Christophe Duperier, Jean-François Martin, Daniel Jacob, Yann Guitton, Marie Lefebvre, et al.

Workflow4Metabolomics: A collaborative research infrastructure for computational metabolomics

Mélanie PÉTÉRA[1], Gildas LE CORGUILLÉ[2], Marion LANDI[1], Misharl MONSOOR[2], Marie TREMBLAY-FRANCO[3], Christophe DUPERIER[1], Jean-François MARTIN[3], Daniel JACOB[4], Yann GUITTON[5,6], Marie LEFEBVRE[4], Estelle PUJOS GUILLOT[1], Franck GIACOMONI[1], Étienne A.THÉVENOT[7] and Christophe CARON[2]

[1] PFEM, UMR1019 INRA, Centre Clermont-Ferrand-Theix, 63122, Saint Genes Champanelle, France
[2] ABiMS, FR2424 CNRS-UPMC, Station Biologique, Place Georges Teissier, 29680, Roscoff, France
[3] PF MetaToul-AXIOM, UMR 1331 Toxalim INRA, 180 chemin de Tournefeuille, F-31027, Toulouse, France
[4] Metabolome Facility of Bordeaux Functional Genomics Center, INRA, IBVM, 33140 Villenave d'Ornon, France
[5] CNRS, IRISA UMR 6074, Campus de Beaulieu, 35042 Rennes, France
[6] Université de Nantes, LINA UMR 6241, 44322 Nantes, France
[7] CEA, LIST, Laboratory for Data Analysis and Smart Systems (LADIS), MetaboHUB Paris, F-91191 Gif-sur-Yvette France

Corresponding Author: melanie.petera@clermont.inra.fr, etienne.thevenot@cea.fr

Paper reference: **Giacomoni et al. (2014) Workflow4Metabolomics: A collaborative research infrastructure for computational metabolomics. Bioinformatics http://dx.doi.org/10.1093/bioinformatics/btu813**

Abstract *In the context of an emergent and fast evolving science, the development of various tools dedicated to metabolomic data processing and data analysis increased. Because metabolomic analyses require a variety of steps involving various disciplines from analytical chemistry to statistics and bioinformatics, it requires many skills and expertise. However, despite this abundance of tools, standardization is lacking in these diversity of programs, as well as infrastructure to handle and link the different steps of metabolomic analyses. We recently implemented Workflow4Metabolomics (W4M), a collaborative online platform hosting and providing a full pipeline for metabolomics from data preprocessing to annotation including statistical analysis. It is not designed to respond to only one specific type of metabolomic analysis, but to cover a maximum range of possible approaches - as metabolomics is a complex science that can be studied through various complementary analytical techniques. Thus, more than just gathering programs, W4M provides relevant combinations of generic and specific tools, a large part of which being developed and sustained by the partners providing this virtual research environment (VRE). Moreover, using Galaxy, a web-based platform technology, W4M provides modules from various sources and of various types. This platform allows hosted tools to be run and linked together via an instinctive and ergonomic interface, which is beneficial for both beginners and experts in metabolomics. W4M gets its strength from the collaboration of complementary teams from bioinformatics and metabolomics environment. Initiated by the collaboration between two platforms, it gathers today six research teams and platforms with a higher diversity in skills and expertise. It allows a continuous enrichment in the service provided, with addition of new modules and new possible workflows dedicated to cover a large scope of the increasing needs of the metabolomic community. Moreover, the 'open-source' aspect of this platform allows to open it to new collaborators bringing specific expertise that can be highlighted and disseminated in the metabolomic community.*

Keywords galaxy environment, computational workflow, metabolomics, mass spectrometry, virtual research environment

1. Introduction and objectives

Pre-processing, statistical analysis and annotation of metabolomic data is a field of intensive research and the number of proposed software tools increases rapidly. Therefore the question of evaluation and integration of these modules within a computational workflow becomes essential. In addition, such a workflow manager has to provide user-friendly functionalities to enable experimenters to select, chain, and tune the tools according to their practices.

We developed Worfklow4Metabolomics (W4M ; http ://workflow4metabolomics.org), a virtual research environment for computational metabolomics. By building it upon the Galaxy environment, W4M enables developers to easily integrate new modules, whatever the programming language and the operating system. The value of the new tools can then be confirmed on available reference datasets. All modules are available on the W4M toolshed and are fully compatible with other Galaxy workflows ; in particular, normalisation and statistical analysis tools, including (orthogonal) partial least-squares modeling, may be of interest for other omics analyses. Finally, a virtual machine is available for local installation of W4M.

2. VRE description

Currently, W4M contains more than 20 modules for pre-processing, analysis and annotation of data from liquid-chromatography coupled to mass spectrometry (LC-MS). The high-performance computing environment of W4M enables to process datasets of more than several hundreds of samples to be analyzed in a few hours while requiring only 1% of the resources. The next release in June 2015 will provide advanced modules for LC-MS analysis (e.g., quality control and outlier filters, as well as network-based annotation functionalities), in addition to the first tools for NMR data processing. Many tutorials, as well as shared datasets and workflows, help users to configure their pipelines and analyze their results. The W4M course 2015 (21-25 sept., IFB/MetaboHub) will help experimenters to learn more about the functionalities and practice with their own data.

We would like W4M to be presented as a highlight in the next JOBIM conference. First, we think that the publication in Bioinformatics is an achievement for the French bioinformatics community, in the highly competitive field of computational metabolomics. Secondly, W4M is a collaborative project : it is currently developed and maintained by a group of 12 bioinformaticians from the French Bioinformatics Institute (IFB) and the French infrastructure for metabolomics and fluxomics (MetaboHUB) and regularly includes new developers from the national and international communities (e.g. in the hackaton organized by the Metabolomics Society in June or during the PhenoMeNal H2020 project lead by EBI and starting next september). Thirdly, both the Galaxy and R communities are very active in all omics data : W4M should therefore be linked in the near future to genomics, transcriptomics, or proteomics modules to provide a systems biology computational environment.

Acknowledgements

This work was supported by Biogenouest®, Lifegrid (Auvergne), and by the IDEALG project [ANR-10-BTBR-04], IFB [ANR-11-INBS-0013] and MetaboHUB [ANR-11-INBS-0010] grants.