New tools and workflows on W4M, the Galaxy metabolomic infrastructure

Marion Landi, Claire Lopez, Marie Tremblay-Franco, Cécile Canlet, Yann Guitton, Daniel Jacob, Gildas Le Corguillé, Jean-Francois Martin, Misharl Monsoor, Nils Paulhe, et al.

NEW TOOLS AND WORKFLOWS ON W4M, THE GALAXY METABOLOMIC INFRASTRUCTURE

Marion LANDI[1‡], Claire LOPEZ[1‡], Marie TREMBLAY-FRANCO[2], Cécile CANLET[2], Yann GUITTON[5-6], Daniel JACOB[3], Gildas LE CORGUILLE[7], Jean-François MARTIN[2], Misharl MONSOOR[7], Nils PAULHE[1], Mélanie PETERA[1], Camille VACQUIE[7], Franck GIACOMONI[1], Estelle PUJOS-GUILLOT[1], Christophe CARON[7], Etienne THEVENOT[4]

1: PFEM, UMR1019, INRA, Centre Auvergne-Rhône-Alpes, 63122 Saint Genès Champanelle, France
2: PF METATOUL-AXIOM, UMR 1331 TOXALIM, INRA, 180 Chemin de Tournefeuille, F-31027 Toulouse, France
3: IBVM, INRA, Metabolome Facility Of Bordeaux Functional Genomics Center, 33140 Villenave d'Ornon, France
4: LIST, Laboratory For Data Analysis And Smart Systems (LADIS), CEA,Saclay Center, F-91191 Gif-sur-Yvette, France
5: IRISA, UMR 6074, CNRS, Campus de Beaulieu, 35042 Rennes, France
6: LINA, UMR 6241, Université de Nantes, 44322 Nantes, France
7: ABIMS, FR2424, CNRS-UPMC, Station Biologique, Place Georges Teissier, 29680 Roscoff, France
‡: MARION LANDI AND CLAIRE LOPEZ CONTRIBUTED EQUALLY TO THIS WORK

Introduction

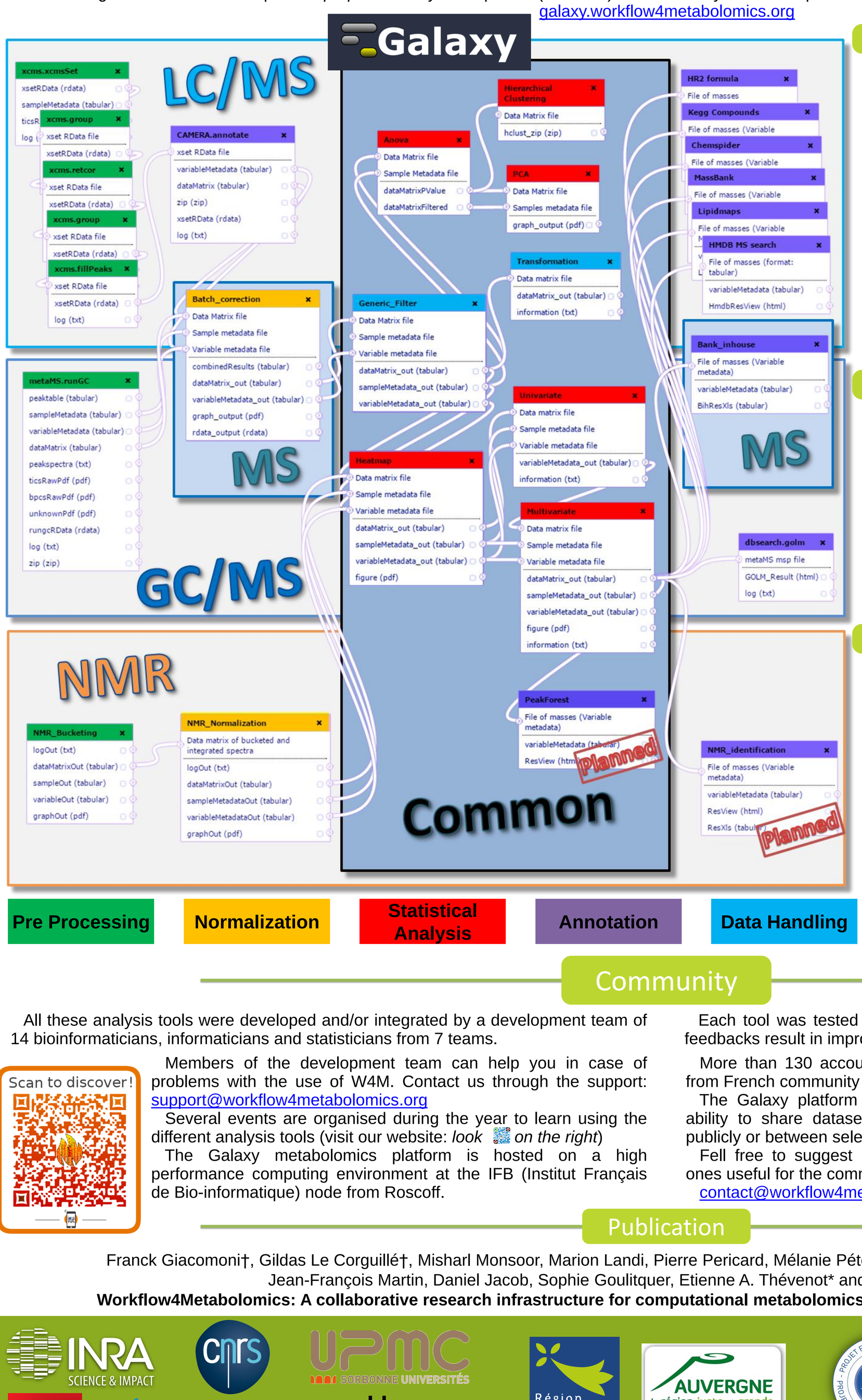
Galaxy is a free and open-source science web-based platform, which provides tools and data processing pipelines (workflow) mainly dedicated to genomics (http://galaxyproject.org). Our platform based on this Galaxy environment http://workflow4metabolomics.org (W4M) is dedicated to metabolomics.

This year version 2.0 of W4M is available thanks to strong collaboration between software developers and chemists allowing optimization and validation. It is accessible for anyone through a personal account.

Workflows

Until recently, several scattered commercial software were necessary for data treatment in metabolomics. Now, Galaxy environment allows executing tools in a stand-alone mode or easily creating personalized workflows through a graphical interface without programming skills, from preprocessing to annotation. Several settings are available for each analysis step and number of dedicated explanations, examples and tutorials are provided for each one. Demo datasets are ready for use to handle the various analyses tools using the tutorials.

The figure below is an example of a proposed analysis sequence (workflow) for each analytical technique: LC-MS, GC-MS and NMR, published on: galaxy.workflow4metabolomics.org



LC-MS

The LC-MS workflow was the single available one year ago. It already covered all steps of data treatment *via* XCMS pretreatment, batch correction, statistical analyses with ANOVA, (O)PLS(-DA) and annotation through queries in several online database like HMDB, Lipidmaps, MassBank, KEGG…

During the past year, this workflow was enriched and strengthened with the addition of steps for data quality control, improvement of normalization (new pool-less drift modeling) and statistical analyses (N-way anova) as well as the ability to query in a in-house bank.

Currently, 27 tools are accessible for this LC-MS workflow. In the future, they will be further improved and complemented with new tools (*e.g.* MS/MS) especially common tools (*e.g.* PeakForest: spectral database *cf* P1).

GC-MS

A new workflow has emerged very recently. It allows covering all the data processing steps for GC-MS. Many of these analysis steps are common with LC-MS, like normalization and statistical analyses.

But two tools were developed specifically for the GC-MS analysis: "metMS:runGC" for Pre-Processing and "dbsearch.golm" to query spectra into Golm metabolome database.

Like LC-MS, GC-MS workflows will be enriched thanks new tools in the next twelve months.

NMR

New tools were developed on Galaxy for NMR data processing, *e.g.* bucketing and normalization, from Bruker files. NMR spectra preprocessing including Fourier transformation, spectra phase correction, baseline correction and calibration have to be performed first on Topspin software.

For statistical analyses, the same tools as for mass spectrometry are used, *e.g.* univariate and multivariate analyses.

The next stage of NMR workflow corresponding to spectra annotation will be soon developed, with bank queries dedicated to NMR or the use of the MetaboHUB bank: PeakForest.

We planed to enrich these workflows, *e.g.* variable sized bucketing and 2D NMR.

Community

All these analysis tools were developed and/or integrated by a development team of 14 bioinformaticians, informaticians and statisticians from 7 teams.

Each tool was tested and validated by analysts of these teams and many of their feedbacks result in improvements.

Scan to discover!

Members of the development team can help you in case of problems with the use of W4M. Contact us through the support: support@workflow4metabolomics.org

Several events are organised during the year to learn using the different analysis tools (visit our website: *look on the right*)

The Galaxy metabolomics platform is hosted on a high performance computing environment at the IFB (Institut Français de Bio-informatique) node from Roscoff.

More than 130 accounts were opened in the past year, mainly from French community but not only.

The Galaxy platform increases the collaborative work *via* this ability to share datasets and workflows (tools and parameters) publicly or between selected collaborators.

Fell free to suggest improvement of the existing tools or new ones useful for the community contact us at:
contact@workflow4metabolomics.org

Scan to discover!

Publication

Franck Giacomoni†, Gildas Le Corguillé†, Misharl Monsoor, Marion Landi, Pierre Pericard, Mélanie Pétéra, Christophe Duperier, Marie Tremblay-Franco, Jean-François Martin, Daniel Jacob, Sophie Goulitquer, Etienne A. Thévenot* and Christophe Caron* (2014).
Workflow4Metabolomics: A collaborative research infrastructure for computational metabolomics. Bioinformatics doi:10.1093/bioinformatics/btu813

http://www.clermont.inra.fr/plateforme_exploration_metabolisme