Towards a cyber Galaxy ?

The IFB working team on the Galaxy platform

Christophe CARON[1], Wilfrid CARRE[1], Alexandre CORMIER[1], Sandra DEROZIER[2], Franck GIACOMONI[3], Olivier INIZAN[4],
Gildas LE CORGUILLE[1], Alban LERMINE[5], Sarah MAMAN[6], Pierre PERICARD[1] and Franck SAMSON[2]

[1] CNRS, UPMC, FR2424, ABiMS, Station Biologique, 29680, Roscoff {christophe.caron, wilfrid.carre, alexandre.cormier, gildas.lecorguille, pierre.pericard}@sb-roscoff.fr –
[2] INRA, UR1077, MIGALE, Centre de Jouy-en-Josas, 78352, Jouy-en-Josas, {sandra.derozier, franck.samson}@jouy.inra.fr – [3] PFEM, UMR1019 INRA, Centre Clermont-Ferrand-Theix, 63122, Saint Genes Champanelle, franck.giacomoni@clermont.inra.fr –
[4] INRA, URGI, Route de St Cyr, Versailles, olivier.inizan@versailles.inra.fr – [5] Institut Curie, Paris 75248, alban.lermine@curie.fr – [6] INRA, UMR444, Laboratoire de Génétique Cellulaire, Centre de Toulouse Auzeville, 24 Chemin de Bordé Rouge, sarah.maman@toulouse.inra.fr

Introduction

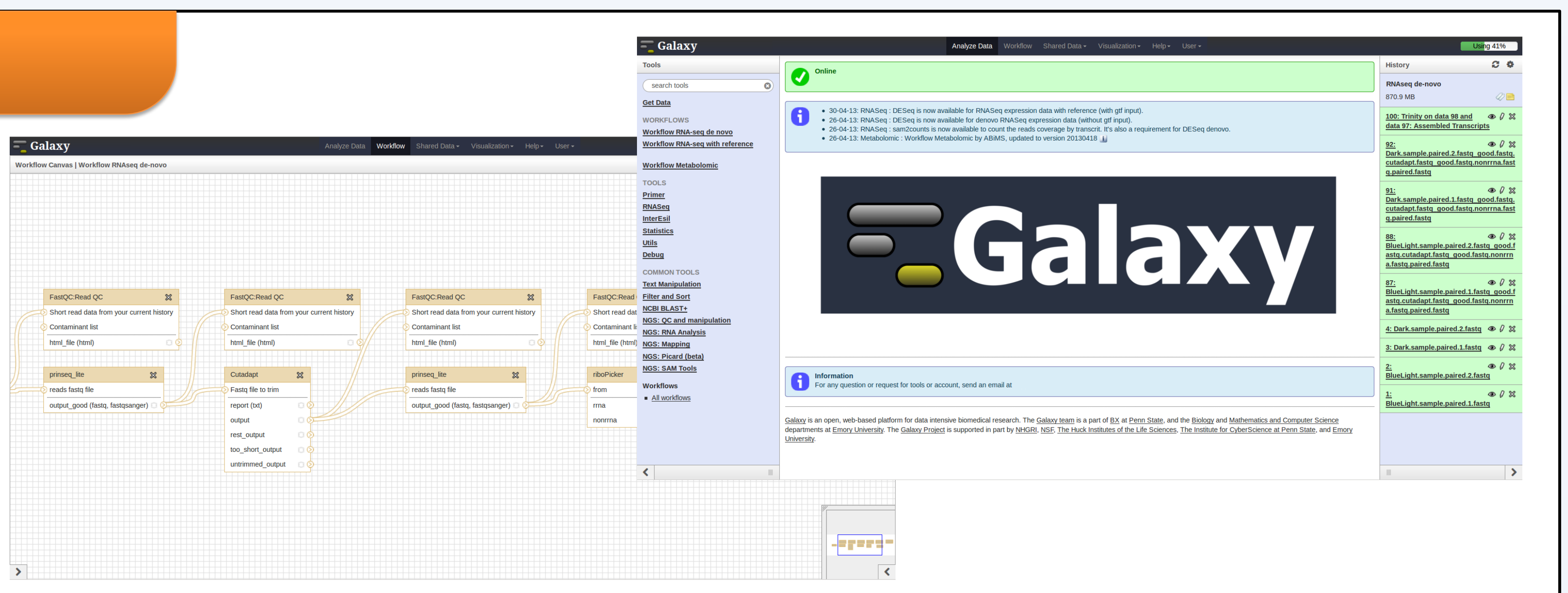
The mission of the French Institute of Bioinformatics (IFB) is to coordinate the French bioinformatics community. Within a context where high-throughput data analysis considerably modifies the way in which analysis is carried out with the use of new infrastructures, new tools and new skills, the IFB has decided to conduct a collaborative action on data analysis workflows and more specifically on the Galaxy platform.

The Galaxy platform already benefits from a very active community of users and developers. The Galaxy platform is designed for sharing all the items implied in a biological analysis: tools, data, protocols and results. Consequently, this working environment enhance collaboration between all the actors of a project: algorithmits, scientists, bioanalysts, engineers

Galaxy

- No more need to launch command lines on a terminal
- No more need to have knowledge of programming/scripting
- Jobs submission facilities on computer cluster (HPC)
- Protected data and historical management facilities
- Data and analysis workflows sharing
- Possibility of adding tools as needed: NGS, Metabolomic, Statistic, etc …

[1][2][3]



Training

Pooling, Seminar, Sharing …

Context

- High-throughput data analysis requires new storage and new computing facilities.
- Integrated solutions must be proposed to biologists and bioanalysts to exploit these facilities.

Objectives

- An ideal starting point for biologists and bio-analysts to learn good analysis practices.

Ongoing work

- Training for initiation to Galaxy (Curie, Genotoul, ABiMS, MIGALE, URGI)
- Galaxy as an environment : Transcriptomic, SNP detection, Chip-seq, etc … (all)
- An online self-learning platform (Genotoul/Sigeneae)

Architecture

Optimisation, virtualisation …

Context

- Galaxy components: job manager, web frontal, databases …
- Different usages: training, developmental and exploitation prototype phases
- Methods: instances virtualisation, workstation installation, data sharing, parallelisation, etc…

Objectives

- Address operating problems in the production environment, automation capabilities of deployment tasks and Galaxy servers monitoring solutions.

Ongoing work

- Presentation: "DevOps Ignition to reach Galaxy continuous integration." GCC2013, Oslo. (URGI).

Tool integration

Good practice, Toolshed …

Context

- High-throughput analysis tools integration can be tedious: multiplicity of sources, languages, formats …

Objectives

- Offer solutions facilitating the tools integration within Galaxy: the tool versioning versus the Galaxy releases, tools and workflows unit testing …
- Technological survey
- Reflection on the benefits of a high quality IFB specific Galaxy toolshed

Ongoing work

- Writing a guide to good practice

Community

Users / Platform / Projects / Research institutes

Ongoing initiative

- The Galaxy-France communities, Aviesan bioinformatics school, Galaxy Aplibio group
- A working group has been made from several IFB platforms (ABiMS, Curie, Genotoul/SIGENAE, MIGALE, URGI), and from one platform of the national infrastructure MetaboHub (PFEM).
- Plurality of local instances depending on their specific activities

Conclusion and Perspectives

- The first presented work brought together a community that had already initiated some actions on Galaxy. These actions will be linked with other national research infrastructures financed by "Future Investments" ANR program (MetaboHub, EMBRC-France, etc …).
- This work could also be seen as a first step towards a wider community. Sharing usages and good pratices regarding tools and instances could be seen as first mission for this community.
- Current scientific collaborations illustrate that the Galaxy environment can become an excellent medium and a gateway between bioinformatician and biologist communities.

Schedule

- **Summer 2013**: documentation website creation
- **17-23 November 2013**: new ITMO Aviesan courses in Roscoff (ABiMS, Migale, Institut Curie, URGI and Genotoul/Sigenae)
- **December 2013**: IFB seminar with experience feedback on Galaxy
- **Spring 2014**: 4 days of training: basic and advanced practices (tool integration, e-learning, etc …) in Roscoff by the IFB Galaxy group

Contact

galaxy-france@lists.bx.psu.edu

[1] Goecks, J, Nekrutenko, A, Taylor, J and The Galaxy Team. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. 2010 Aug 25;11(8):R86.
[2] Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J. "Galaxy: a web-based genome analysis tool for experimentalists". Current Protocols in Molecular Biology. 2010 Jan; Chapter 19:Unit 19.10.1-21.
[3] Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, Miller W, Kent WJ, Nekrutenko A. "Galaxy: a platform for interactive large-scale genome analysis." Genome Research. 2005 Oct; 15(10):1451-5.