

GalaxEast

An open and powerful Galaxy instance for integrative Omics data analysis

<http://www.galaxeast.fr>

Stephanie Le Gras¹, Serge Uge¹, Matthieu Jung¹, Ludovic Roy, Valerie Cognat², Frederic Plewniak³, Irwin Davidson¹ and Julien Seiler¹

¹IGBMC - CNRS UMR 7104 - Inserm U 964, Illkirch, France

²IBMP - UPR2357, Strasbourg, France

³G.M.G.M - Uds/CNRS UMR7156, Strasbourg, France

Contact : galaxy@igbmc.fr

The exponential growth of high-throughput Omics data has posed a great technical challenge to experimentalists who lack bioinformatics skills and computing power. Moreover, integrative analysis of data from various sources is needed to provide biological insights into biological systems. We present hereby GalaxEast, an open and powerful free web-based platform for integrative analysis of Omics data.

GalaxEast is based upon Galaxy, one of the most popular bioinformatics workflow management systems, which is becoming a standard for sharing bioinformatics tools worldwide. As a Galaxy instance, GalaxEast aims at providing a large range of bioinformatics tools for the analysis of various types of Omics data. It supports reproducible computational research by providing an environment for performing and recording bioinformatics analyses.

The GalaxEast project has the following main objectives:

- Provide the academic scientific community with an open and powerful Galaxy instance with a guaranteed availability. The platform offers access to cutting-edge and up-to-date tools for Omics data analysis with help and support.
- Propose innovative developments and new helpful tools packaged for Galaxy (available in the GalaxEast toolshed)
- Promote the packaging of new developments for Galaxy (through wrappers and/or toolshed packages).

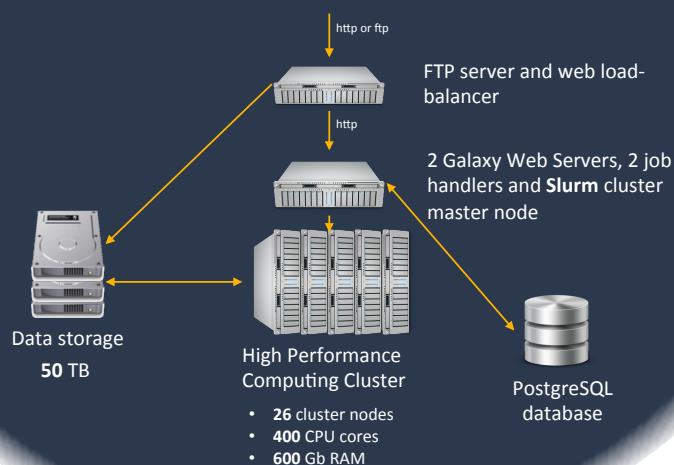
Using GalaxEast

- GalaxEast is freely available for academic users
- Each user is granted 50GB of storage
- More storage can be temporarily granted if needed
- A comprehensive documentation for 26 tools is available (www.galaxeast/wiki)

Key figures

- Launch in June 2013 (for IGBMC users)
- Open to the community since April 2014
- 14,594 jobs
- 188 tools
- 110 users
- 3 training sessions using GalaxEast

An architecture designed for intensive computing



Toolbox & workflows

The platform provides access to up-to-date tools and algorithms such as those devoted to Next Generation Sequencing (NGS) data analysis of:

- Epigenome: MACS, Homer
- Genomic sequence: GATK, Samtools, BEDtools, Picard
- Transcriptome: Cufflinks, HTSeq, TopHat
- Statistical analyses: DeepTools, S-mart, DESeq

GalaxEast provides access to workflows developed for:

- Motif search
- Repetitive element analyses
- ChIP-seq data analysis

Security

- Data transfers are performed through FTPS
- All dataset files are anonymized
- All cluster jobs are anonymized
- All datasets are visible ONLY by the owner
- Database server and compute cluster are accessible from the internet

Innovative developments

GalaxEast implements new bioinformatics developments and tools for:

- Data manipulation
- NGS data analysis for ChIP-seq
- Genomic database querying tool to request the GEO Profiles database
- De-noising methods for spectroscopic data (urQRd)

Available on GalaxEast Toolshed (<http://toolshed.galaxeast.fr/>).

GalaxEast...

From a user point of view

- 50GB of storage per user
- 188 up-to-date tools
- More tools and genomes can be added on request
- A Galaxy platform with a guaranteed availability
- A Responsive GalaxEast workgroup
- A way to facilitate exchanges within the scientific community
- A reliable support for the reproducibility of the results of experiments

From a developer point of view

- Provide a simple way to share bioinformatics tools to the community
- Provide a reliable platform for the publication of research projects
- Exchange best practices and experiences to create Galaxy packages along with the collaborative project initiated by the french Galaxy community from the IFB (French Institute for Bioinformatics).

On-going work

- New server architecture
- Virtualisation of GalaxEast through Vagrant

Future work

- Integrate more tools for imaging, proteomics, statistics,...
- New workflows to come (RNA-seq, variant discovery...).
- Provide access to reviewed toolshed repositories
- Continue analysis tools documentation