

Galaxy au Mésocentre

Réunion du comité des utilisateurs du Mésocentre

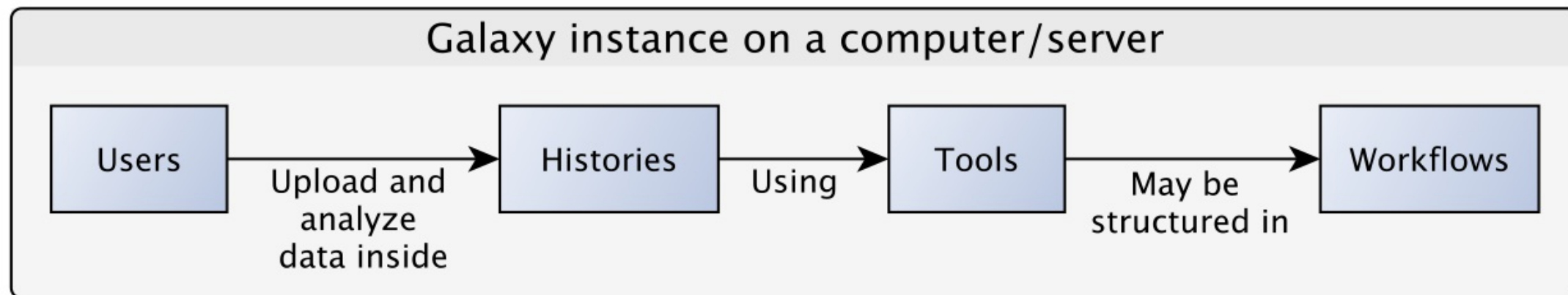
30 Mars 2016

Galaxy

Galaxy

Open source, web-based platform for data intensive
biomedical research

Principe



Galaxy

Galaxy

Tools

search tools

Get Data

Lift-Over

Text Manipulation

Datamash

Convert Formats

Filter and Sort

Join, Subtract and Group

Fetch Alignments/Sequences

NGS: QC and manipulation

NGS: DeepTools

NGS: Mapping

NGS: RNA Analysis

NGS: SAMtools

NGS: BamTools

NGS: Picard

NGS: VCF Manipulation

NGS: Peak Calling

NGS: Variant Analysis

NGS: RNA Structure

NGS: Du Novo

NGS: Gemini

Operate on Genomic Intervals

Statistics

Graph/Display Data

CloudMap

Phenotype Association

BEDTools

Genome Diversity

EMBOSS

Regional Variation

FASTA manipulation

Multiple Alignments

Metagenomic Analysis

Multiple regression

Multivariate Analysis

Motif Tools

STR-FM: Microsatellite Analysis

NCBI SRA Tools

DEPRECATED

NGS: GATK Tools (beta)

Workflows

All workflows

Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy [start here](#) or consult our [help resources](#). You can install your own Galaxy by following the [tutorial](#) and choose from thousands of tools from the [Tool Shed](#).

Tweets by @galaxyproject

Galaxy Project @galaxyproject

In 4 weeks: NGS Bioinformatics Data Analysis Workshop @polyomics @bioteXcel #ngsglasgow bit.ly/22y6Qj7

EmbedView on Twitter

PENNSTATE

The Galaxy Team is a part of the [Center for Comparative Genomics and Bioinformatics](#) at Penn State, and the [Department of Biology](#) and at Johns Hopkins University.

JOHNS HOPKINS UNIVERSITY

TACC

This instance of Galaxy is utilizing infrastructure generously provided by the [CyVerse](#) at the [Texas Advanced Computing Center](#), with support from the [National Science Foundation](#).

CYVERSE

History

search datasets

Unnamed history

0 b

This history has been purged and deleted

This history is empty. You can [load your own data](#) or [get data from an external source](#)

Galaxy version 16.01, commit 19bf334c7763c2418408083b9a7369386e50ad66

<https://usegalaxy.org/>

Outils dans Galaxy

```
$ vsearch --threads 4 --notrunclabels --derep_fulllength input_file --o
```

VSearch dereplication (Galaxy Tool Version 1.9.7.0) Options

Select your FASTA file

6: Good quality sequences
(--derep_fulllength)

Output just first n sequences

(--topn)

Read abundance annotation from input

(--sizein)

Write cluster abundances to centroid file

(--sizeout)

Strand specific clustering

Both strands
(--strand)

UCLUST-like output

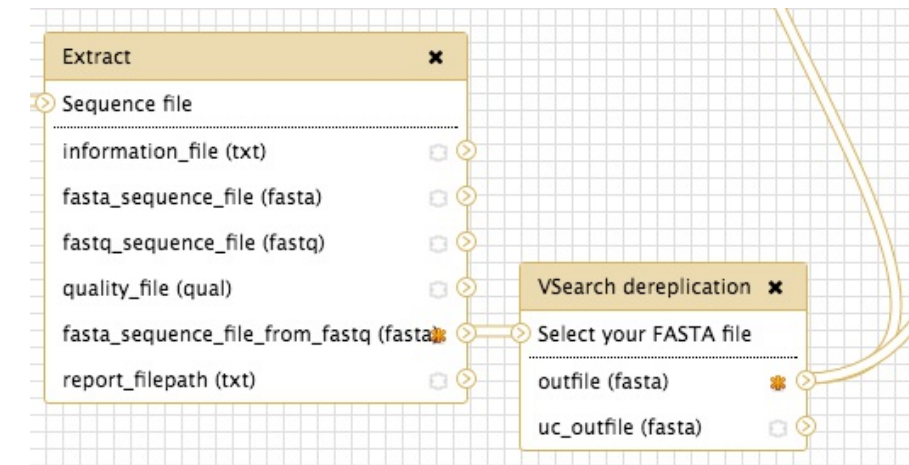
(--uc)

Minimum abundance

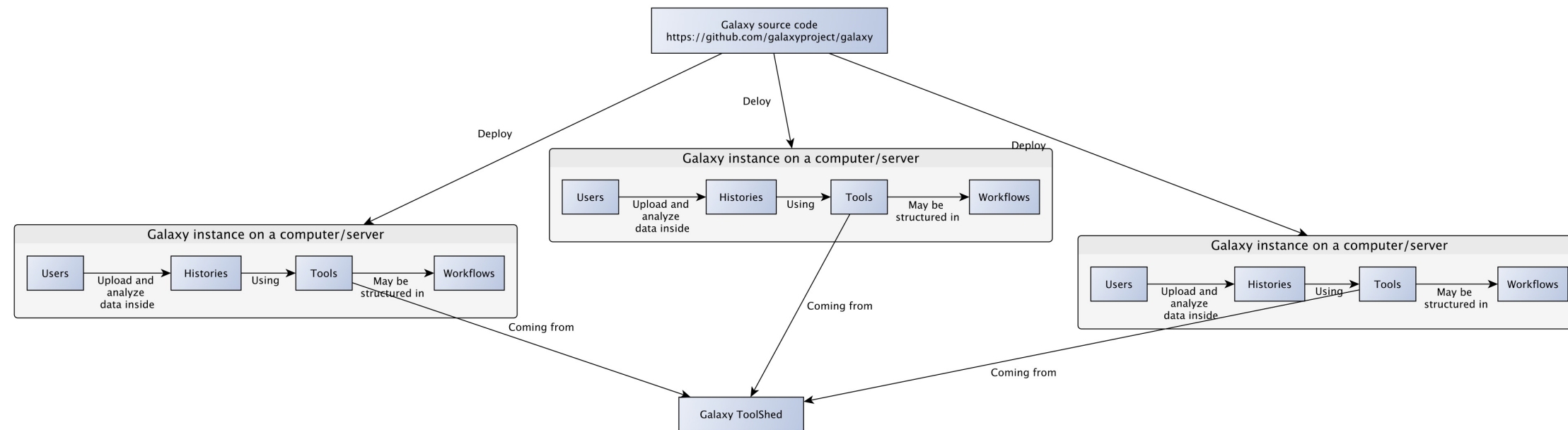
(--minuniquesize)

Maximum abundance

(--maxuniquesize)



Principe



Galaxy ToolShed

← → ↺

https://toolshed.g2.bx.psu.edu/repository

☆ 🔥 ⓘ 🖨️ ↕

Galaxy Tool Shed

Repositories Groups Help ▾ User ▾

3792 valid tools on Mar 19, 2016

Search

▪ [Search for valid tools](#)

▪ [Search for workflows](#)

Valid Galaxy Utilities

▪ [Tools](#)

▪ [Custom datatypes](#)

▪ [Repository dependency definitions](#)

▪ [Tool dependency definitions](#)

All Repositories

▪ [Browse by category](#)

Repositories I Can Change

▪ [Repositories I own](#)

▪ [Repositories I can administer](#)

▪ [My writable repositories](#)

▪ [Reset metadata on my repositories](#)

▪ [Latest revision: missing tool tests](#)

▪ [Latest revision: installation errors](#)

▪ [Latest revision: failing tool tests](#)

▪ [Latest revision: skip tool tests](#)

▪ [Latest revision: all tool tests pass](#)

▪ [Latest revision: invalid tools](#)

Available Actions

▪ [Create new repository](#)

▪ [Import repository capsule](#)

Repositories by Category

Name	Description	Repositories
Assembly	Tools for working with assemblies	72
ChIP-seq	Tools for analyzing and manipulating ChIP-seq data.	38
Combinatorial Selections	Tools for combinatorial selection	7
Computational chemistry	Tools for use in computational chemistry	22
Convert Formats	Tools for converting data formats	61
Data Managers	Utilities for Managing Galaxy's built-in data cache	24
Data Source	Tools for retrieving data from external data sources	31
Epigenetics	Tools for analyzing Epigenetic/Epigenomic datasets	
Fasta Manipulation	Tools for manipulating fasta data	70
Fastq Manipulation	Tools for manipulating fastq data	52
Genome-Wide Association Study	Utilities to support Genome-wide association studies	21
Genomic Interval Operations	Tools for operating on genomic intervals	42
Graphics	Tools producing images	41
Imaging	Utilities to support imaging	2
Metabolomics	Tools for use in the study of Metabolomics	26
Metagenomics	Tools enabling the study of metagenomes	51
Micro-array Analysis	Tools for performing micro-array analysis	9
Next Gen Mappers	Tools for the analysis and handling of Next Gen sequencing data	106
Ontology Manipulation	Tools for manipulating ontologies	11
Phylogenetics	Tools for performing phylogenetic analysis	16
Proteomics	Tools enabling the study of proteins	73
RNA	Utilities for RNA	84
SAM	Tools for manipulating alignments in the SAM format	79
Sequence Analysis	Tools for performing Protein and DNA/RNA analysis	394
Statistics	Tools for generating statistics	69

https://toolshed.g2.bx.psu.edu/

Galaxy au Mésocentre

Objectifs

- Déployer une ou plusieurs instances Galaxy au Mésocentre
- Utiliser les ressources du Mésocentre
 - Exécution des outils sur le cluster de calcul (HPC2)
 - Accès au stockage distribué CEPH
- Intégrer des outils nécessaires par les utilisateurs locaux

Groupe de travail

- Première réunion : 15 Janvier 2016
- Participants
 - S. Colin
 - F. Giacomoni
 - N. Guilhot
 - B. Batut
 - P. Leroy
 - O. Guinaldo
 - P. Peyret
 - T. Bellembois
 - E. Peyretailade
 - A. Claude
 - V. Breton
 - A. Mahul
 - C. Vaury
 - P. Pouchin

Identification d'un projet pilote

- ☐ Déploiement d'une instance
- ☐ Utilisation de l'identification LDAP
- ☐ Execution des outils via des jobs SLURM
- ☐ Accès au stockage distribué CEPH
- ☐ Mise en place de procédures de maintenance automatiques
- ☐ Validation et test de l'instance en conditions réelles par un groupe restreint d'utilisateurs

Groupe de travail restreint

- animateur principal : Nicolas Guilhot
nicolas.guilhot@clermont.inra.fr
- Participants
 - Pierre Pouchin
 - Alexandre Claude
 - Franck Giacomoni
 - Antoine Mahul
 - Alexis Ozwald
 - Bérénice Batut

Etat du projet pilote

1e réunion : 4 Mars 2016

- ☒ Déploiement d'une instance
- ☒ Utilisation de l'identification LDAP
- ☐ Execution des outils via des jobs SLURM
- ☐ Accès au stockage distribué CEPH
- ☐ Mise en place de procédures de maintenance automatiques
- ☐ Validation et test de l'instance en conditions réelles par un groupe restreint d'utilisateurs

Après le projet pilote

- Ouvrir une instance en production
 - Utilisateurs du Mésocentre
 - Outils bioinformatiques de base
- Mise à disposition des outils développés localement
- Développement et installation d'outils spécifiques pour les utilisateurs du mésocentre
- Intégration des outils dans Galaxy ToolShed
- Maintient et évolution de la plateforme

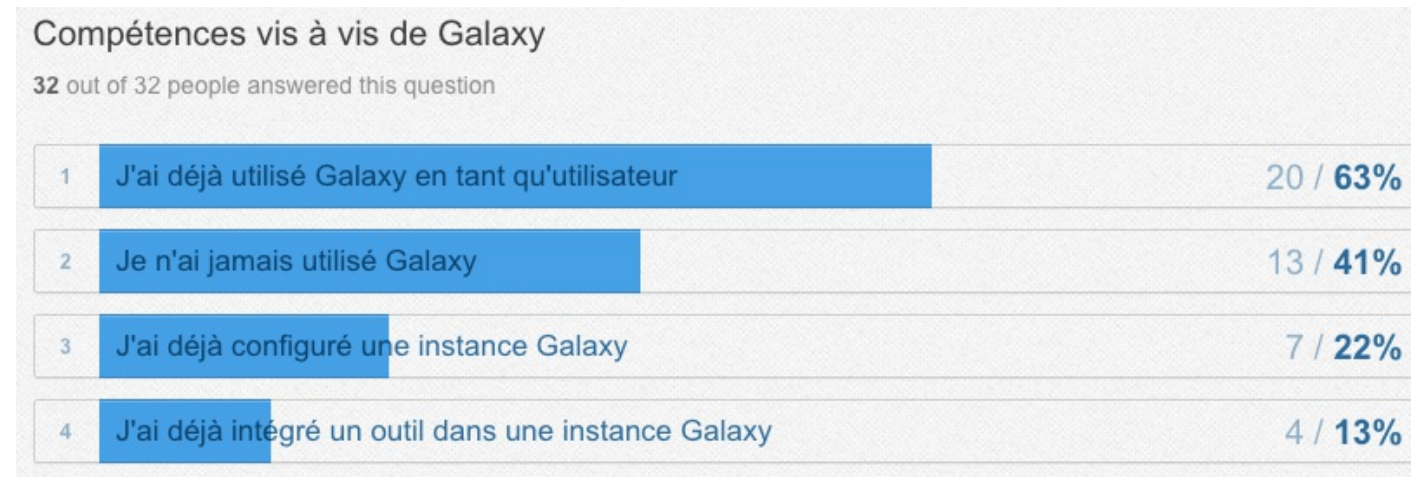
Projet de workshop

Objectifs

- Former à l'intégration d'outils
 - Bases (planemo, ToolShed, gestion des dépendences, ...)
 - Bonnes pratiques
- Initier un groupe local sur l'intégration des outils dans Galaxy

Projet de workshop

Formulaire de pré-enregistrement



A faire : Trouver une date