

Tools

Options ▾

Your user name: smaman
Your file path : /work/smaman/

1 - UPLOAD YOUR DATA

Get Data

2 - FILES MANIPULATION

Text ManipulationFilter and SortJoin, Subtract and GroupConvert Formats3 - SEQUENCES
MANIPULATIONFASTA manipulationFASTQ manipulationSAM/BAM manipulation : Picard
(beta)SAM/BAM manipulation : SAM
Tools

4 - MAPPING

BWA - Bowtie

5 - INDEL ET SNP

Indel AnalysisRNA-SeqGATK Tools (beta)

6 - SRNASEQ

Analyse des miRNAAnnotationsAlignement sur reference**WELCOME ON SIGENAE GALAXY WORKBENCH**

Galaxy is a workbench available for biologists from Sigena Platform. Galaxy objectives are:

- Make bioinfo Linux tools accessible to biogists.
 - Hide the complexity of the infrastructure.
- Allow creation, execution and sharing of workflows.

History

Options ▾



TP FastQC



54.0 Mb

8: FastQC_data 5.html **6: GM.fastqsanger** **5: h1.fastqsanger** **4: FastQC_data
18.html** **3: FASTQ Summary
Statistics on data 18** **2: FASTQ Summary
Statistics on data 18**

76 lines, 1 comments
format: tabular, database: ?
Info: 99115 fastq reads were processed.
Based upon quality values and sequence characters, the input data is valid for: sanger
Input ASCII range: '#'(35) - 'C'(67)
Input decimal range: 2 - 34
Epilog : job finished at ven mai 11 10:36:43 CEST 2012



1	2	3	4	5	6
#column	count	min	max	sum	mean
1	99115	2	33	3194703	32.2
2	99115	2	34	3156652	31.8
3	99115	2	34	3145060	31.7
4	99115	2	34	3120431	31.4
5	99115	2	34	3096075	31.2



Vos traitements bioinformatiques avec GALAXY

Sarah Maman – Maria Bernard
11 Février 2014



Equipe "Galaxy project" :

- Le Center for Comparative Genomics and Bioinformatics - Penn State,
- Des départements "Biology" et "Mathematics and Computer Science" de l'Université d'Emory.



Anton Nekrutenko
Penn State



Nate Coraor
Penn State



James Taylor
Emory



Une « Galaxy » parmi tant d'autres



Serveur public (<https://main.g2.bx.psu.edu/>):

- Gratuit
- Quota limité : pour se familiariser à l'outil sur des petits jeux de données.
- Données non protégées

Galaxy

Code gratuit et "open source"

Une communauté nationale et internationale très active :

- Listes de diffusion (US, FR)
- Wiki
- Twitter
- "Galaxy tour de France"

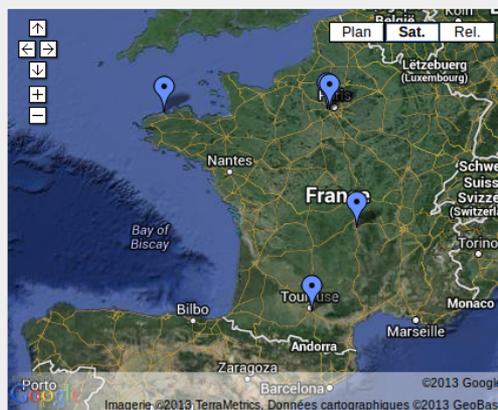


L'instance locale Sigenae de Galaxy :

- Maintenu par Sigenae.
 - Intégration des outils et scripts "locaux".
- **Présentation des particularités de l'instance Sigenae.**



- ✓ Documentation collaborative (wiki)
- ✓ Formations (mise en commun agenda PF)
- ✓ Architecture
- ✓ Intégration d'outils (Tool Shed)



Afficher [Galaxy IFB France](#) sur une carte plus grande

Liste des instances

ABIMS Roscoff	Initiation, NGS Cleaning, RNASeq Differential Expression	http://galaxy.sb-roscoff.fr/	Christophe Caron - Alexandre Cormier - Gildas Lecorguille - Pierre Pericard
Institut Curie	ChIP-Seq Analysis	http://nebula.curie.fr/	Alban Lermine
Genotoul / Sigenae	Initiation to Galaxy, SNP calling, RNASeq, sRNASeq	http://galaxy-workbench.toulouse.inra.fr/	Sarah Maman
INRA URGI	Differential expression analysis, Variant detection	http://urgi.versailles.inra.fr/galaxy2	Olivier Inizan
INRA MIGALE	Initiation to Galaxy, NGS Galaxy	http://migale.jouy.inra.fr/galaxy/	Sandra Derozier - Franck Samson
Southgreen	Generalist platform, and crop breeding	gohelle.cirad.fr/galaxy/root/	Jean-Francois Dufayard
INRA PFEM / MetaboHUB	Metabolomics data analysis	https://pfem-galaxy/	Franck Giacconi



Inutile de savoir :

- ✓ Lancer une ligne de commande
- ✓ Programmer en perl, python, shell ...
- ✓ Lancer un script

Inutile de s'inquiéter pour son disque dur:

- ✓ Jobs lancés sur un cluster de calculs.
- ✓ Pas d'archivage de fichiers sur votre PC.

Inutile d'attendre la fin d'un traitement:

- ✓ Possible de lancer plusieurs jobs en parallèle
- ✓ Partir prendre un café, consulter ses mails, ..fermer Internet !
- ✓ Puis voir les résultats le lendemain matin.



- ✓ Complémentaire au « **cahier de laboratoire** »
→ Retrouver les données, les outils, les références pour la **publication** ...
- ✓ Manipuler **facilement et rapidement** les informations de votre fichier.
- ✓ Utiliser des outils bioinformatiques.
- ✓ Construction de **workflow** résumant l'ensemble des fonctionnalités utilisées.
- ✓ Intégration de **nos propres outils** (outils très utiles et fréquemment utilisés)

→ Galaxy devient **VOTRE BOITE A OUTILS.**



Comment ouvrir un compte sur Genotoul ?



Home About us Resources Services Help Login

genotoul bioinfo Create an account

FAQ
Support
Create an account
Resources
Create an account

You are here: » [Help](#) » [Create an account](#)

An account is only available for people who works with a french team. In this case please fill the supervisor's informations in the form with the director of this french team.

For temporary position account, the request has to be validated by a permanent supervisor who is in charge of respecting the [INRA charter](#) usage!

The default quota for an account is 1To for /work/user and 200 Gb for /save/user.

Last name: *

First name: *

E-Mail (academic only) : *

Phone: *

Status

If the request is for a temporary position or an internship:

Supervisor last name:

Supervisor email:

Supervisor phone:

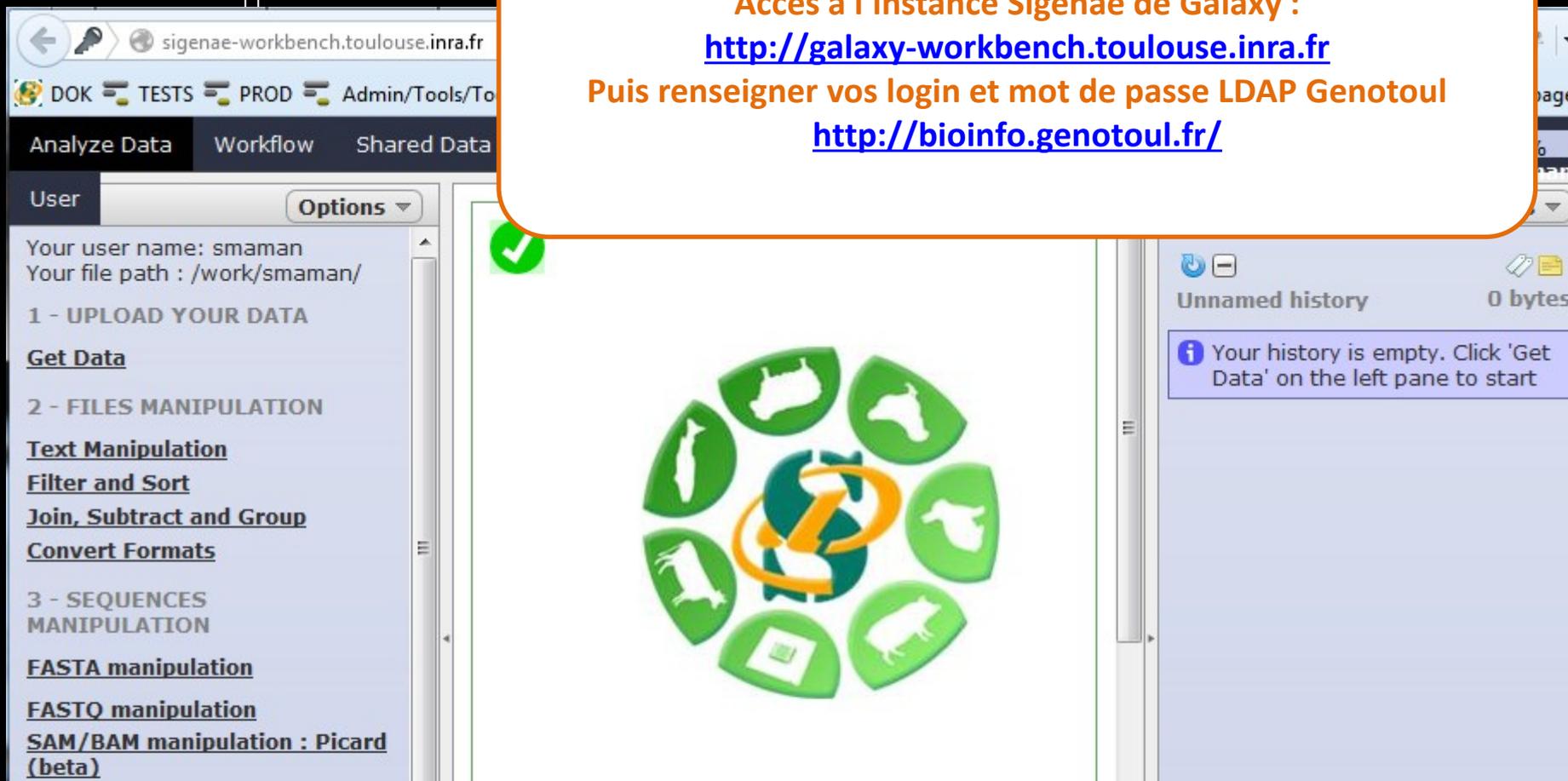
Contract duration:

Type : Private Academic

Formulaire de demande de compte:
<http://bioinfo.genotoul.fr>

! Utiliser un mail académique

Accès à l'instance Sigenae de Galaxy :
<http://galaxy-workbench.toulouse.inra.fr>
Puis renseigner vos login et mot de passe LDAP Genotoul
<http://bioinfo.genotoul.fr/>



Vos données sont protégées.

Vos jobs sont envoyés sur le cluster Genotoul.

Inutile de savoir programmer

De nombreux outils bioinformatiques sont intégrés dans Galaxy.



Comment fonctionne Galaxy ?

Galaxy est installée sur une machine virtuelle qui envoie les calculs à un cluster.



Utilisateur de Galaxy

Envoi de données

Récupération des résultats



Serveur Web Galaxy

Envoi les jobs



Gère la file d'attente

Gestionnaire de tâches

Cluster de calculs



Exécute



- ✓ Les limites et règles d'utilisation sont celles du cluster Genotoul :
 - Temps de calcul: pour l'instant pas limité.
 - Stockage: 1 To

- ✓ Les données sont traités sur la partition /work de Galaxy.
 - Fichiers temporaires de calcul
 - Quota utilisateur 1To
 - Partition purgée si les fichiers ne sont pas utilisés depuis plus de 120 jours.

- ✓ Pour les traitements :
 - Sur la workq : 256 slots et 2048G de mémoire / utilisateur
 - Sur unlimitq : 32 slots et 256G de mémoire / utilisateur



Analyze Data Workflow Shared Data Visualization Admin Help Using 13%

Welcome smaman, you are working in /work/smaman

User Options

Your user name: smaman
Your file path : /work/smaman/

1 - UPLOAD YOUR DATA

Get Data

2 - FILES MANIPULATION

Text Manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats

3 - SEQUENCES
MANIPULATION

FASTA manipulation
FASTQ manipulation
SAM/BAM manipulation : Picard (beta)
SAM/BAM manipulation : SAM Tools

4 - MAPPING

BWA - Bowtie

5 - INDEL ET SNP

Indel Analysis
RNA-Seq
GATK Tools (beta)



**WELCOME ON SIGENAE GALAXY
WORKBENCH**

Galaxy is a workbench available for biologists from Sigenae Platform. Galaxy objectives are:

- Make bioinfo Linux tools accessible to biologists.
 - Hide the complexity of the infrastructure.
 - Allow creation, execution and sharing of workflows.

History Options

Unnamed history 0 bytes

i Your history is empty. Click 'Get Data' on the left pane to start



Analyze Data Workflow Shared Data Visualization Admin Help Using 13%

Welcome smaman, you are working in /work/smaman

User Options

Your user name: smaman
Your file path : /work/smaman/

1 - UPLOAD YOUR DATA

Get Data

2 - FILES MANIPULATION

Text Manipulation

Filter and Sort

Join, Subtract and Group

Convert Formats

3 - SEQUENCES MANIPULATION

FASTA manipulation

FASTQ manipulation

SAM/BAM manipulation : Picard (beta)

SAM/BAM manipulation : SAM Tools

4 - MAPPING

BWA - Bowtie

5 - INDEL ET SNP

Indel Analysis

RNA-Seq

GATK Tools (beta)





WELCOME ON SIGENAE GALAXY WORKBENCH

Galaxy is a workbench available for biologists from Sigenae Platform. Galaxy objectives are:

- Make bioinfo Linux tools accessible to biologists.
 - Hide the complexity of the infrastructure.
 - Allow creation, execution and sharing of workflows.

History Options

Unnamed history 0 bytes

i Your history is empty. Click 'Get Data' on the left pane to start



Analyze Data Workflow Shared Data Visualization Admin Help Using 13%

Welcome smaman, you are working in /work/smaman

User Options

Your user name: smaman
Your file path : /work/smaman/

1 - UPLOAD YOUR DATA

Get Data

2 - FILES MANIPULATION

Text Manipulation

Filter and Sort

Join, Subtract and Group

Convert Formats

3 - SEQUENCES MANIPULATION

FASTA manipulation

FASTQ manipulation

SAM/BAM manipulation : Picard (beta)

SAM/BAM manipulation : SAM Tools

4 - MAPPING

BWA - Bowtie

5 - INDEL ET SNP

Indel Analysis

RNA-Seq

GATK Tools (beta)

*** Upload local file from filesystem path (version 1.0.0)**

File Name:

File type:

Path to file:

History Options

Unnamed history 0 bytes

Your history is empty. Click 'Get Data' on the left pane to start



Analyze Data Workflow Shared Data Visualization Admin Help

Using 13%

Welcome smaman, you are working in /work/smaman

User Options

Your user name: smaman
Your file path : /work/smaman/

1 - UPLOAD YOUR DATA

Get Data

2 - FILES MANIPULATION

Text Manipulation

Filter and Sort

Join, Subtract and Group

Convert Formats

3 - SEQUENCES MANIPULATION

FASTA manipulation

FASTQ manipulation

SAM/BAM manipulation : Picard (beta)

SAM/BAM manipulation : SAM Tools

4 - MAPPING

BWA - Bowtie

5 - INDEL ET SNP

Indel Analysis

RNA-Seq

GATK Tools (beta)

*** Upload local file from filesystem path (version 1.0.0)**

File Name:
phiX174_read

File type:
Fastq

Path to file:
/work/smaman/phiX174_reads.fastqsanger

Execute

History Options

Galaxy sensibilisation - TP 12.1 Mb

2 - BWA and FastQC

14: phiX174_reads.fastqsanger
1.0 Mb
format: fastqsanger, database: ?

```
@080917-and-080922:5:1:185:82
GATGTTATTTCTTCATTTGGAGGTAAAACCTTTAT
+
IIIIIIIIIIIIIIIIIIII<III@FI8A/I0II4I
@080917-and-080922:5:1:1366:223
GTTTTCTTCTGCGTCAGTAAGAACGTCAGTGTTTC
```



Analyze Data Workflow Shared Data Visualization Admin Help Using 13%

Welcome smaman, you are working in /work/smaman

User Options

Your user name: smaman
Your file path : /work/smaman/

1 - UPLOAD YOUR DATA

Get Data

2 - FILES MANIPULATION

Text Manipulation

Filter and Sort

Join, Subtract and Group

NGS: Mapping

- Lastz map short reads against reference sequence
- Lastz paired reads map short paired reads against reference sequence
- Map with Bowtie for Illumina
- Map with Bowtie for SOLiD
- Map with BWA for Illumina

4 - MAPPING

BWA - Bowtie

5 - INDEL ET SNP

Indel Analysis

RNA-Seq

GATK Tools (beta)

History Options

Unnamed history 0 bytes

i Your history is empty. Click 'Get Data' on the left pane to start

✓



WELCOME ON SIGENAE GALAXY WORKBENCH

Galaxy is a workbench available for biologists from Sigeneae Platform. Galaxy objectives are:

- Make bioinfo Linux tools accessible to biogists.
 - Hide the complexity of the infrastructure.
 - Allow creation, execution and sharing of workflows.



Analyze Data Workflow Shared Data Visualization Admin Help Using 13%

Welcome smaman, you are working in /work/smaman

User Options

Your user name: smaman
Your file path : /work/smaman/

1 - UPLOAD YOUR DATA

Get Data

2 - FILES MANIPULATION

Text Manipulation

Filter and Sort

Join, Subtract and Group

NGS: Mapping

- Lastz map short reads against reference sequence
- Lastz paired reads map short paired reads against reference sequence
- Map with Bowtie for Illumina
- Map with Bowtie for SOLiD
- Map with BWA for Illumina

4 - MAPPING

BWA - Bowtie

5 - INDEL ET SNP

Indel Analysis

RNA-Seq

GATK Tools (beta)

Map with BWA for Illumina (version 1.2.2)

Will you select a reference genome from your

Use one from the history ▾

Select a reference from history:

11: phiX174_genome.fa ▾

Is this library mate-paired?:

Single-end ▾

FASTQ file:

14: phiX174_reads.fastqsanger ▾

FASTQ with either Sanger-scaled quality values (f

History Options

Unnamed history 0 bytes

i Your history is empty. Click 'Get Data' on the left pane to start



Analyze Data Workflow Shared Data Visualization Admin Help

Using 13%
Welcome smaman, you are working in /work/smaman

User Options

Your user name: smaman
Your file path : /work/smaman/

1 - UPLOAD YOUR DATA

Get Data

2 - FILES MANIPULATION

Text Manipulation

Filter and Sort

Join, Subtract and Group

NGS: Mapping

- Lastz map short reads against reference sequence
- Lastz paired reads map short paired reads against reference sequence
- Map with Bowtie for Illumina
- Map with Bowtie for SOLiD
- Map with BWA for Illumina

4 - MAPPING

BWA - Bowtie

5 - INDEL ET SNP

Indel Analysis

RNA-Seq

GATK Tools (beta)

Map with BWA for Illumina (version 1.2.2)

Will you select a reference genome from your l
Use one from the history ▾

Select a reference from history:
11: phiX174_genome.fa ▾

Is this library mate-paired?:
Single-end ▾

FASTQ file:
14: phiX174_reads.fastqsanger ▾
FASTQ with either Sanger-scaled quality values (f

History Options

Unnamed history 0 bytes

15: Map with BWA for Illumina on data 14 and data 11: mapped reads
Job is waiting to run

Galaxy Sig Analyze Data Workflow Shared Data Visualization Admin Help User Welcome smaman Using 30%

Tools Options

1 - UPLOAD YOUR DATA
Get Data

2 - FILES MANIPULATION
Text Manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats

3 - SEQUENCES MANIPULATION
FASTA manipulation
FASTQ manipulation
SAM/BAM manipulation : Picard (beta)
SAM/BAM manipulation : SAMtools
Fetch Sequences

Saved Histories

search history names and tags
Advanced Search

Name	Datasets	Tags	Sharing
TP Galaxy project	2	1	0 Tags
miRNA tests	59	21	0 Tags
TP SNPs calling	84	9	0 Tags
TP RNAseq	88	1	0 Tags Shared, Accessible
test TP miRNA	36	1	0 Tags
Unnamed history			0 Tags

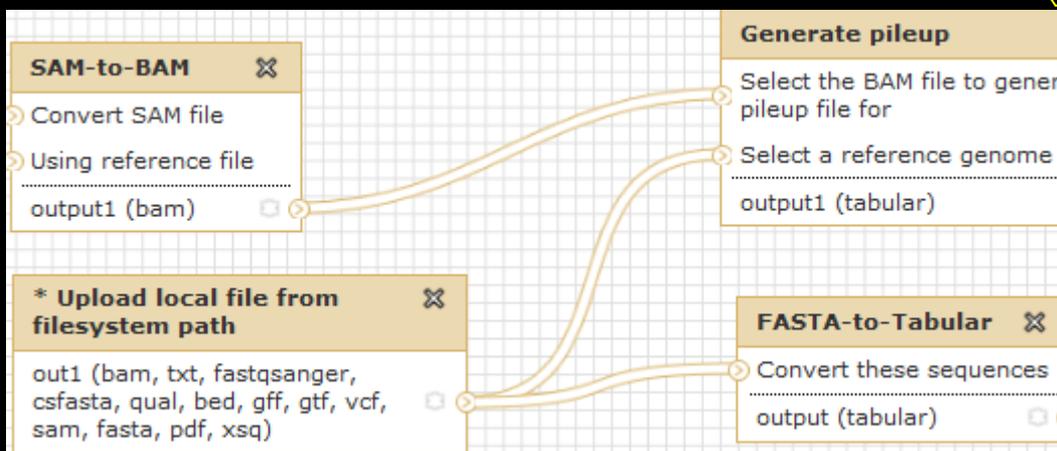
Analyse OK

Analyse en attente

2: UCSC Main on Human: snp137Common (chr22:1-51304566)
~180,000 regions
format: bed, database: hg19
view in GeneTrack
display at Ensembl
Current

Analyse en erreur

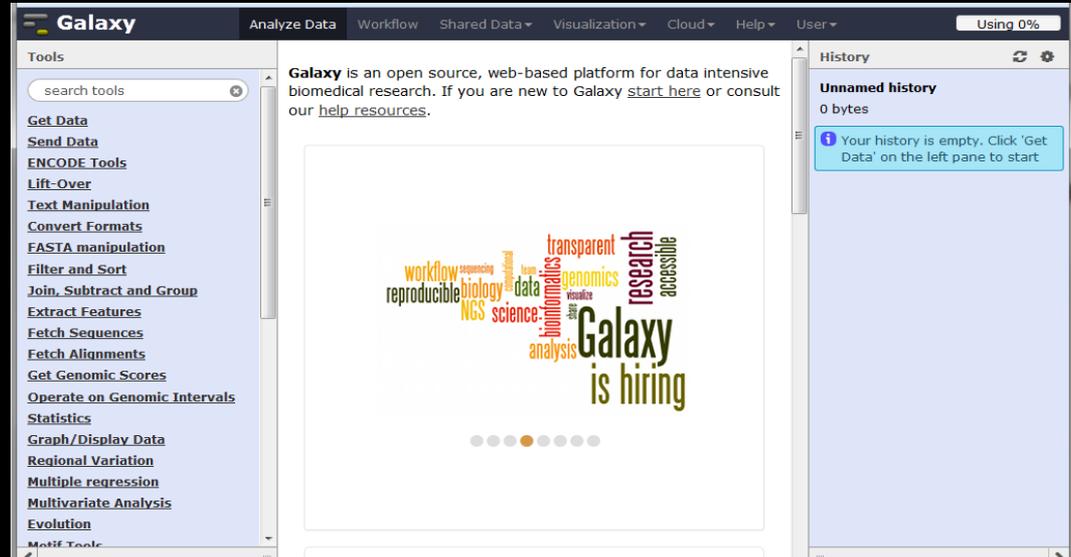
Analyse en cours





Instance publique de Galaxy :

- Sur le serveur public du Galaxy Project : <https://usegalaxy.org/>
- De nombreuses autres instances publiques sont disponibles.



Instance « personnalisée » de Galaxy peut-être disponible :

- Sur un serveur SlipStream : Solution plug-and-play éditée par Galaxy. Aucune installation ni administration nécessaire.
- En local (sur votre PC)
- Sur une machine virtuelle
- Sur le cloud



VirtualBox : logiciel de virtualisation de systèmes d'exploitation.

Principales étapes d'installation de Galaxy sur votre vm

1. Préparer votre espace de travail
2. Télécharger les sources de Galaxy (et Mercurial pour gérer les mises à jour du code)
3. Lancer Galaxy
4. Paramétrer Galaxy : universe ini file

Intérêts d'une vm

1. Faire fonctionner en simultané plusieurs systèmes d'exploitation
2. Possibilité de cloner une vm => Partage de machines Galaxy
3. vm sauvegardée donc restaurable.



En fonction du contexte d'utilisation et des ressources IT disponibles :

	NO WAIT TIMES	NO STORAGE QUOTAS	NO JOB SUBMISSION LIMITS	NO DATA TRANSFER BOTTLENECKS	NO IT EXPERIENCE REQUIRED	NO REQUIRED INFRASTRUCTURE
GALAXY MAIN	✗	✗	✗	✗	✓	✓
LOCAL GALAXY	?	?	?	✓	✗	✗
CLOUD GALAXY	✓	✓	✓	✗	✗	✓
SLIPSTREAM GALAXY	✓	✓	✓	✓	✓	✓

Source : <http://bioteam.net/slipstream/galaxy-edition/>



- ✓ Un fichier de configuration de l'instance : `universe.ini`
(paramétrages, configurations, caractéristiques du déploiement).
- ✓ Un fichier de configuration des outils : `tool_conf.xml`
- ✓ Un fichier pour configurer le lancement des jobs : `job_conf.xml`
- ✓ Un fichier pour configurer les types de données traitées : `datatypes_conf`
- ✓ Définir le répertoire de travail dans lequel Galaxy placera les fichiers (tmp, résultats, entrées) : `job_working directory`



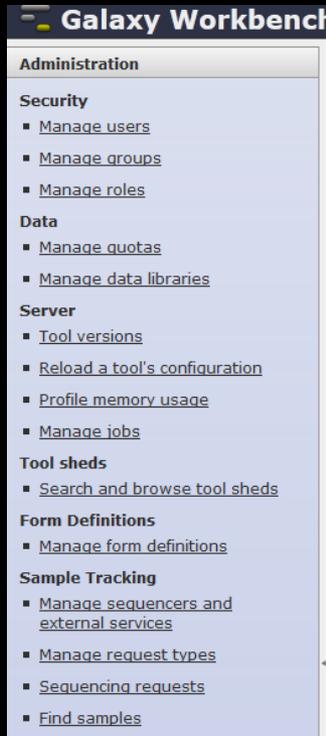
✓ D'où peut provenir un nouvel outil ?

1. Outil de base dans le code de Galaxy
2. Du Galaxy Tool Shed
3. Outil / script maison

✓ Etapes d'installation :

1. Développer ou récupérer un wrapper (1 fichier xml + script(s))
2. Préciser le chemin d'accès à ce wrapper dans le fichier de conf tool_conf.xml
3. Relancer Galaxy pour que ce nouvel outil soit visible depuis l'interface

→ Galaxy devient **VOTRE BOITE A OUTILS.**



Users

Manage users

– Liste des utilisateurs

Manage groups

– Associer un utilisateur à un groupe

Manage roles

– Associer un utilisateur à un rôle

Data

Manage data libraries

– Définir les permissions afin de permettre à certains utilisateurs d'accéder à certaines données

Server

Reload a tool's configuration

– Pour rechargé un outil qui a été commenté dans le tool_conf.xml (préférer le terminal)

Profile memory usage

– Estimer la mémoire utilisée par certaines fonctions de Galaxy

Manage jobs

– Lister les jobs trop longs (en vu d'un kill)



« Problème » :

✓ Tests faciles => Multiplication rapide du nombre de fichiers dans les historiques => Surcharge du /work & impact sur le quota.

Solutions apportées :

✓ Gros avantages à utiliser l'outil d'upload « maison » " Upload local file from filesystem path" plutôt que l'outil "Upload File".

✓ **Classer** vos traitements en créant un nouvel historique par analyse.

✓ **S'organiser** en renommant et taguant vos datasets, historiques & workflows (pour les retrouver facilement et distinguer les tests)

✓ **Supprimer** définitivement (Purge Deleted) et régulièrement les datasets, historiques et workflows inutiles.



Deux méthodes pour gérer les utilisateurs Galaxy :

✓ Gestion des utilisateurs interne à Galaxy

- Autoriser un utilisateur à se créer lui-même un compte :
allow_user_creation : True

✓ Gestion externe des utilisateurs

- Limiter l'accès uniquement aux utilisateurs ayant un compte INRA LDAP :
require_login : True
- URL en https
- Proxy & Virtual host



Avantages :

- ✓ Lancement des jobs sur le cluster eu nom de l'utilisateur logué sur Galaxy et non au nom de « galaxy »
- ✓ Suivi affiné des quotas, statistiques et des priorités.

Inconvénients :

- ✓ Les jobs sont lancés par l'utilisateur connecté avec un sudo.

Solutions apportées :

- ✓ Donc la machine et les répertoires de travail sont isolés du reste du cluster.
- ✓ Réservation d'un nœud et d'une queue spécifique sur le cluster.



Merci pour votre écoute

sigenae-support@listes.inra.fr