

Analyze Data | Workflow | Shared Data | Visualization | Admin | Help | User | Welcome smaman, you are working in / Using 13%

**Tools** 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)

6 - SRNASEQ

Analyse des miRNA  
Annotations  
Alignement sur reference

**History** Options

TP FastQC 54.0 Mb

8: FastQC\_data\_5.html

6: GM.fastqsanger

5: h1.fastqsanger

4: FastQC\_data\_18.html

3: FASTO Summary Statistics on data 18

2: FASTO Summary Statistics on data 18

76 lines, 1 comments  
format: tabular, database: 2  
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
fcolumn	count	min	max	sum	mean
1	99115	2	33	2194703	32.2
2	99115	2	34	2156652	31.8
3	99115	2	34	2145060	31.7
4	99115	2	34	2120431	31.4
5	99115	2	34	2096075	31.2

**WELCOME ON SIGENAE GALAXY WORKBENCH**

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

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

GENO TOULOUSE bioInfo

INRA

# Vos traitements bioinformatiques avec GALAXY

Philippe Bardou – Sarah Maman – Ibouniyamine Nabihoudine – Olivier Rué  
06 Novembre 2013

<http://galaxy-workbench.toulouse.inra.fr>



## Galaxy pour vos traitements (bio)informatiques





Vidéo disponible sur « sig-learning »

- Plateforme
- Vos données
- Historique
- Workflow
- Bioinfo
- Vous

Présentation de la plateforme Galaxy.

Comment récupérer vos données ?

Notions d'outils, d'historique et de workflow.

Lancement de traitements bioinformatiques.

Guide pour les utilisateurs Galaxy.



## Galaxy Project

Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences  
 Jeremy Greenberg<sup>1</sup>, Anton Nekrutenko<sup>2\*</sup>, James Taylor<sup>1,3</sup> and The Galaxy Team



- Plateforme
- Vos données
- Historique
- Workflow
- Bioinfo
- Vous

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



**Une communauté nationale et internationale très active :**

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





**L'instance locale Sigenae de Galaxy :**

- Maintenue par Sigenae.
- Intégration des outils et scripts "locaux".

→ **Présentation des particularités de l'instance Sigenae.**

**Plateforme**

Vos données

Historique

Workflow

Bioinfo

Vous



## Groupe de travail GALAXY IFB <http://www.ifb-galaxy.org/>



- ✓ 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**

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

**Plateforme**

Vos données

Historique

Workflow

Bioinfo

Vous




## Des outils accessibles à tous..

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.

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous




## Galaxy « la bioinformatique pour tous »

Galaxy est :

- « Open source »
- Développé et maintenu par une communauté active.
- Une plateforme proposant un ensemble d'outils bioinformatiques.
- Une "constellation" d'outils (analyser, manipuler, visualiser).

Les biologistes peuvent :

- Lancer des traitements sans Linux.
- Dupliquer des traitements.
- Partager des analyses complètes.

Et ceci de manière très intuitive !




Les bioinformaticiens peuvent :

- Faire ajouter des outils.
- Partager des outils (Tool Shed).
- Partager des traitements.



Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous

**Contexte d'utilisation dans un laboratoire**

✓ 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.

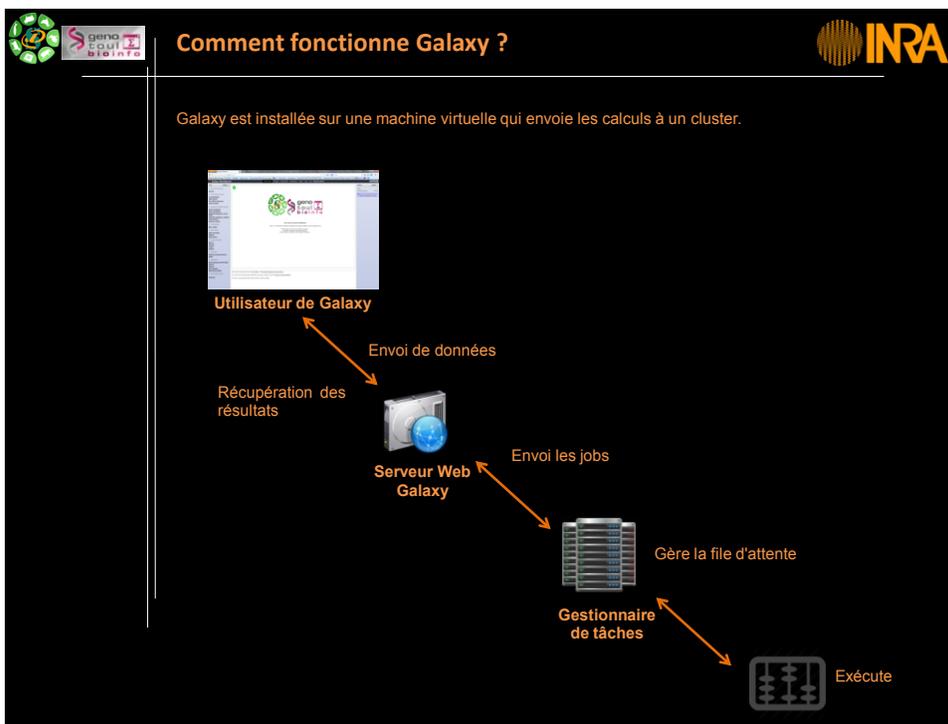
✓ Construction de **workflow** résumant l'ensemble des fonctionnalités utilisées.

✓ Intégration de **nos propres outils** (outils très utiles et fréquemment utilisés)

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

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

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 ouvrir un compte sur Genotoul ?

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

! Utiliser un mail académique

genotoul bioinfo

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: Research director

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

Supervisor last name:

Supervisor email:

Supervisor phone:

Contract duration:

Type:  Private  Academic

Analyse en quelques clics sans utiliser votre disque dur ! INRA

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)

History Options

Unnamed history 0 bytes

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 Sigena Platform. Galaxy objectives are:

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

Analyse en quelques clics sans utiliser votre disque dur ! INRA

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)

History Options

Unnamed history 0 bytes

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 Sigena Platform. Galaxy objectives are:

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



Analyse en quelques clics sans utiliser votre disque dur ! INRA

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

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 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.

Analyse en quelques clics sans utiliser votre disque dur ! INRA

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

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

**Map with BWA for Illumina (version 1.2.2)**

Will you select a reference genome from your history?  
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 (

Analyse en quelques clics sans utiliser votre disque dur ! 

Analyze Data Workflow Shared Data Visualization Admin Help

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 I  
Use one from the history

Select a reference from history:  
11: phiX174\_genome.fa

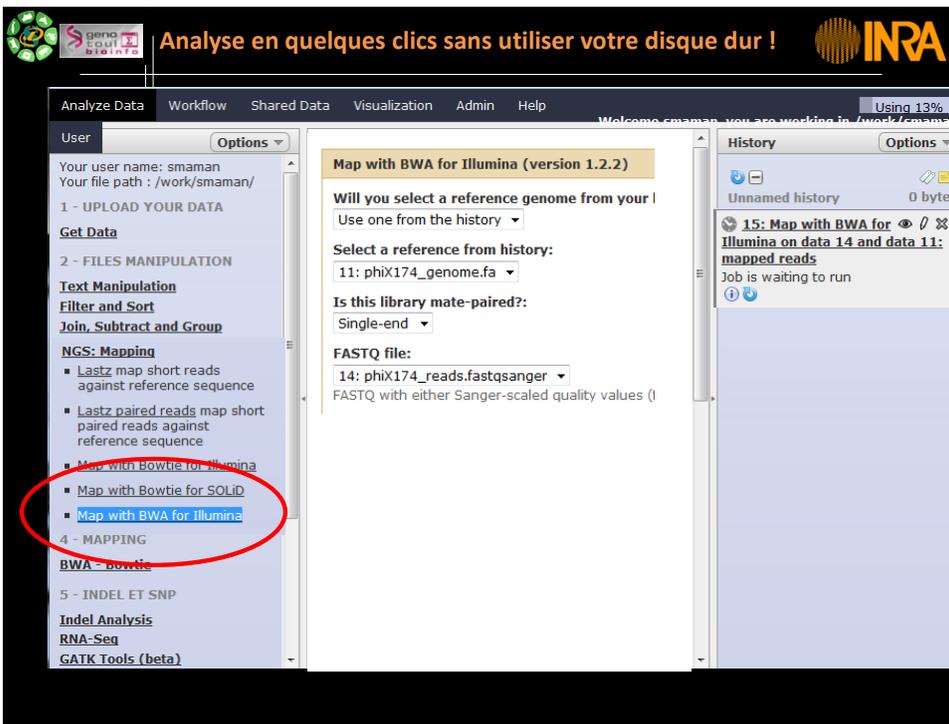
Is this library mate-paired?:  
Single-end

FASTQ file:  
14: phiX174\_reads.fastqsanger  
FASTQ with either Sanger-scaled quality values (I

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



 Interface simplifiée

Interface divisée en 4 parties :

- 1 - Liste des outils disponibles.
- 2 - Visualisation de l'outil utilisé, historique, dataset ou workflow.
- 3 - Historique ou workflow détaillé.
- 4 - Menu .

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous

Analyze Data Workflow Shared Data Visualization Admin Help

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 Form**

3 - SEQUENCING MANIPULATION

**FASTA manipulation**

**FASTQ manipulation**

**SAM/BAM manipulation**

1

2

History Options

TP FastQC 54.0 Mb

8: FastQC on data 5

6: GM.fastqsanger

5: h1.fastqsanger

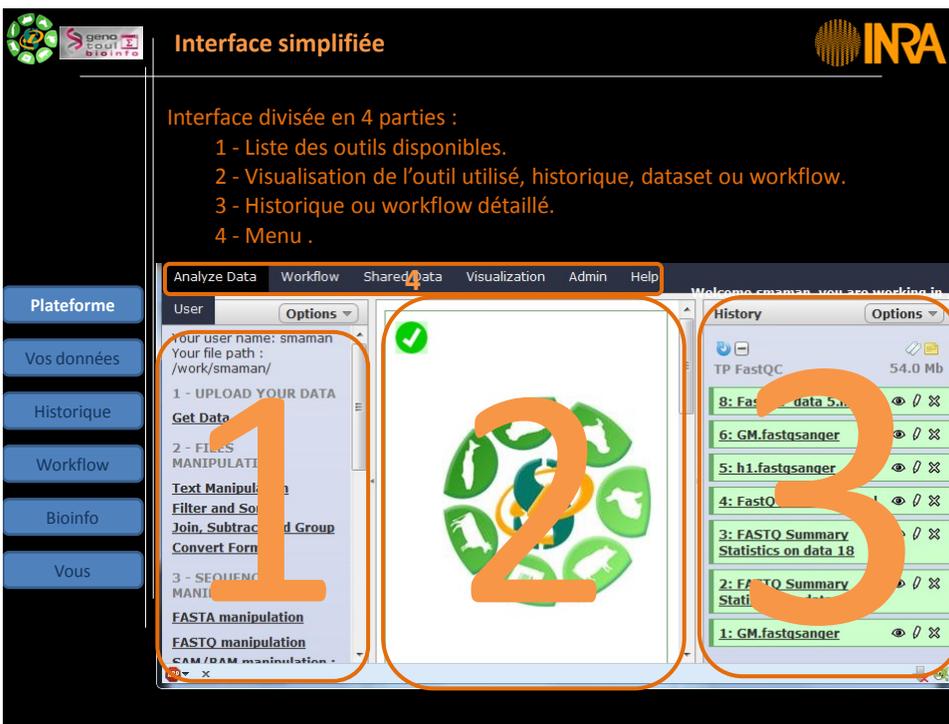
4: FastQC

3: FASTQ Summary Statistics on data 18

2: FASTQ Summary Statistics

1: GM.fastqsanger

3



## Le vocabulaire spécifique à Galaxy




**TOOL** : Outil bioinformatique ou de traitement de fichiers.

**DATASET** : Fichier de données téléchargé dans Galaxy (fichier entrant) ou fichier généré par Galaxy (fichier résultat).

**HISTORY** : Liste des datasets (entrants et résultants) générés par les tools.

**WORKFLOW** : Schématisation des traitements.

Upload File from your computer

**TOOL**

génère

1: GM.fastqsanger 17.4 Mb  
format: fastqsanger, database: 2

```

@HWI-ST-EAS627_1:8:1:17:202
TGTTGGAGACCCCGACATGGTCTTCAGCTCC
+
BB@A<9@?AB??=?B?B?B?>?A@?><1:9=A
@HWI-ST-EAS627_1:8:1:66:1050
CAGAAGTAGACGAGAAAGAGCGGACCTTCCG
                    
```

Dont la liste

**HISTORY**

8: FastQC_data_5.html
6: GM.fastqsanger
5: h1.fastqsanger
4: FastQC_data_18.html

**DATASET (S)**

SAM-to-BAM

- Convert SAM file
- Using reference file

output1 (bam)

Upload local file from filesystem path

out1 (bam, txt, fastqsanger, csfasta, qual, bed, gff, gtf, vcf, sam, fasta, pdf, xsq)

**HISTORY**

Generate pileup

- Select the BAM file to generate pileup file for
- Select a reference genome

output1 (tabular)

FASTA-to-Tabular

- Convert these sequences

output (tabular)

**WORKFLOW**

## Principaux onglets




- ANALYSE DATA** : Page d'accueil de Galaxy.
- WORKFLOW** : Liste des workflows.
- SHARED DATA** : Liste des datasets, historiques et workflows partagés.
- VISUALIZATION** : Outil de visualisation de vos fichiers résultats.
- USER** : Accès à vos historiques et datasets sauvegardés.

Analyze Data Workflow Shared Data Visualization Admin Help

Your workflows

Name
Workflow constructed from history 'Test BWA fichiers Gnome'
Workflow / 'TP : NGS - Polymorphisme' (imported from uploaded file)
Workflow / 'FastQC' (imported from uploaded file)
Workflow / 'IGV bai' (imported from uploaded file)
Workflow 'Get flanks / region promoters'
Workflow 'RNA seq statistics' et FastQC

Workflows shared with you by others

No workflows have been shared with you.

History

8: FastQC_data_5.html
6: GM.fastqsanger
5: h1.fastqsanger
4: FastQC_data_18.html
3: FASTQC Summary Statistics on data_18
2: FASTQC Summary

**Saved Histories**

search history names and tags

Advanced Search

Name	Datasets
Unnamed history	
Unnamed history	
Unnamed history	
TP_FastQC	7

En résumé ... 

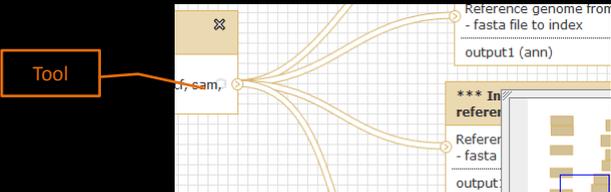
Un **DATASET** est un fichier de données (fichiers d'entrée, fichier résultat) :



Votre **HISTORIQUE** est un « répertoire » qui « liste » l'ensemble de vos fichiers de données (fichiers d'entrée, fichier résultat) utilisés ou générés :



Votre **WORKFLOW** est une représentation de vos traitements : outils utilisés, fichiers



Platforme  
Vos données  
Historique  
Workflow  
Bioinfo  
Vous

Présentation de la plateforme Galaxy.

**Comment récupérer vos données ?**

Notions d'outils, d'historique et de workflow.

Lancement de traitements bioinformatiques.

Guide pour les utilisateurs Galaxy.

Platforme  
**Vos données**  
Historique  
Workflow  
Bioinfo  
Vous



2 méthodes de téléchargement de vos données privées

**Galaxy** Analyze Data Workflow

Tools Options

Linux Username: sigenae  
File Path: /work/sigenae/galaxy/  
Server: galaxy

**Get Data**

- Upload File from your computer
- UCSC Main table browser
- UCSC Test table browser
- UCSC Archaea table browser
- EBI SRA ENA SRA
- Get Microbial Data
- Upload local file from filesystem path Upload data to history without copying on server

Platforme

Vos données

Historique

Workflow

Bioinfo

Vous

Using 100%

History Options

Unnamed history 174.2 Mb

You are over your disk quota. Tool execution is on hold until your disk usage drops below your allocated quota.

**Solution 1 (recommandée si vos fichiers sont dans Genotoul) :**  
« Upload local file from filesystem path ».

**Solution 2 :**  
« Upload file from your computer, with a copy on server ».

⚠ L'espace disque utilisé n'est pas celui de votre PC.  
Vos traitements ne sont pas limités par la capacité de votre PC.

Téléchargement de données publiques

Données UCSC, Ensembl, NG6, BIOMART :

Genomes Genome Browser Tools Mirrors Downloads My Data About Us Help

**Table Browser**

clade: Mammal genome: Human assembly: Feb. 2009 (GRCh37/hg19)

group: Genes and Gene Prediction Tracks track: UCSC Genes add custom tracks

track hubs

table: knownGene describe table schema

region:  genome  ENCODE Pilot regions  position chr21:33031597-33041570 lookup define regions

identifiers (names/accessions): paste list upload list

filter: create

intersection: create

correlation: create

output format: BED - browser extensible data Send output to  Galaxy  GREAT

output file: (leave blank to keep output in browser)

file type returned:  plain text  gzip compressed

Platforme

Vos données

Historique

Workflow

Bioinfo

Vous

ENA European Nucleotide Archive

bio::mart

New Count Results

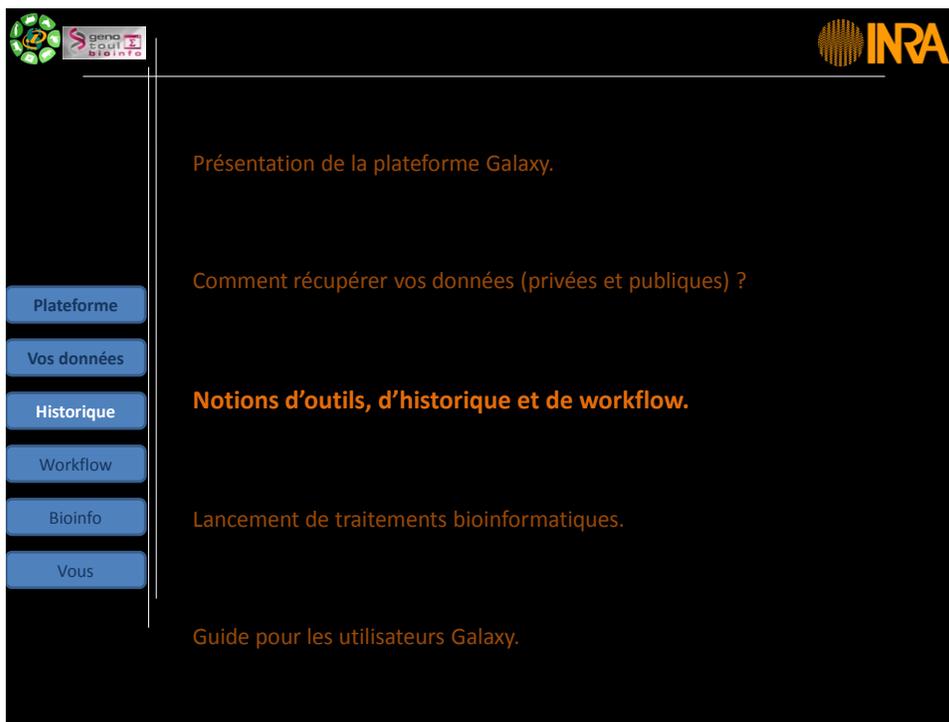
Dataset






Plateforme  
 Vos données  
 Historique  
 Workflow  
 Bioinfo  
 Vous

**Exercice 1 :**  
 Connexion à Galaxy, exploration de l'interface, téléchargement de datasets






Plateforme  
 Vos données  
 Historique  
 Workflow  
 Bioinfo  
 Vous

Présentation de la plateforme Galaxy.  
 Comment récupérer vos données (privées et publiques) ?  
**Notions d'outils, d'historique et de workflow.**  
 Lancement de traitements bioinformatiques.  
 Guide pour les utilisateurs Galaxy.

## Gestion de vos historiques



- Conserver toutes les étapes de vos analyses .
- Partager vos analyses.
- A chaque run d'un outil, un nouveau dataset est créé. Les données ne sont pas écrasées.
- Répéter, autant de fois que nécessaire, une analyse.

Name	Count	Tags	Shared	Size	Date
SwanPorc	18	0 Tags	Shared	0 bytes	Ju 20
FastQC	6	0 Tags	Shared	17.4 Mb	Ag 20
TP : NGS - Polymorphisme	8	2	Shared	6.6 Gb	Ag 20
TP FastQC	12	16	0 Tags	54.0 Mb	Mi 20
indexation genome	1	0 Tags		46 bytes	Ju 20

For 0 selected histories: **Rename** **Delete** **Delete Permanently**

## Historiques sauvegardés, publiés et partagés



- L'ensemble de vos historiques sont automatiquement sauvegardés.
- Vos historiques publiés sont accessibles à l'ensemble des utilisateurs loggés sur Galaxy (Shared Data / Published Histories).
- Les historiques partagés sont accessibles uniquement à un utilisateur spécifique (History / Option / Histories Shared With Me).

Pour partager ou publier un historique :  
 User / Saved histories / Cliquer sur le nom de l'historique / Share ou Publish






## Liste des historiques et exécution de mes jobs

**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 cours  
Analyse en erreur

1 - UPLOADED DATA

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




Présentation de la plateforme Galaxy.

Comment récupérer vos données ?

**Notions d'outils, d'historique et de workflow.**

Lancement de traitements bioinformatiques.

Guide pour les utilisateurs Galaxy.

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous

**Créer un workflow**

Depuis une page blanche, vous pouvez concevoir un workflow.

Aide : les résultats produits sont typés, il n'est donc pas possible de brancher une dataset sur un mauvais tool !



Platforme

Vos données

Historique

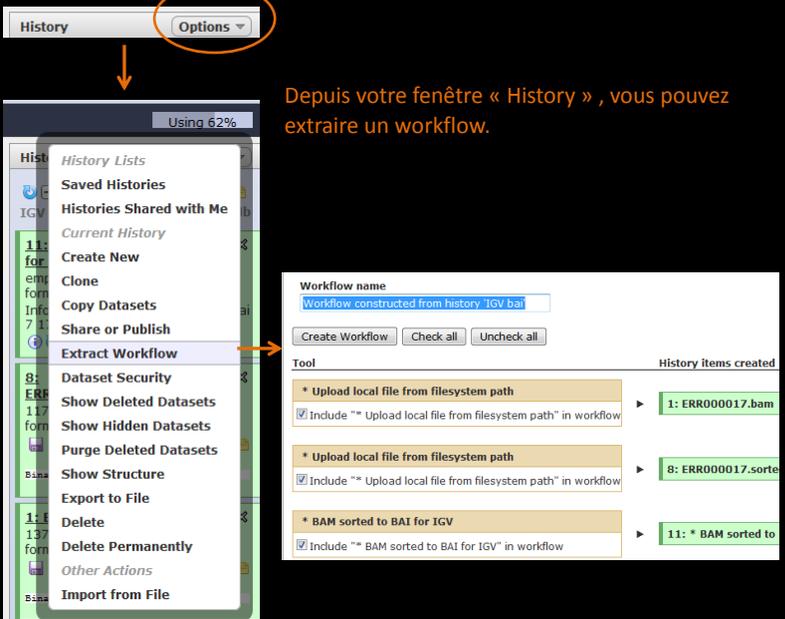
Workflow

Bioinfo

Vous

**Exporter votre historique en workflow.**

Depuis votre fenêtre « History », vous pouvez extraire un workflow.



Platforme

Vos données

Historique

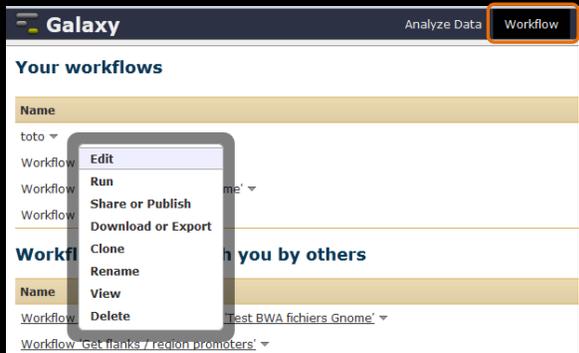
Workflow

Bioinfo

Vous

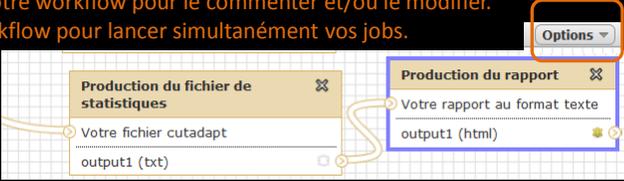
**Lancer plusieurs jobs simultanément**

Cliquer sur le menu « Workflow » pour lister vos workflows :



Vous pouvez ensuite, depuis le menu « Options », soit :

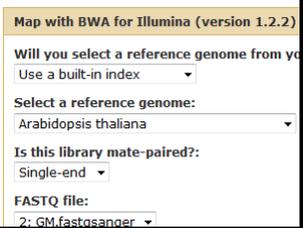
- Editer votre workflow pour le commenter et/ou le modifier.
- Run workflow pour lancer simultanément vos jobs.



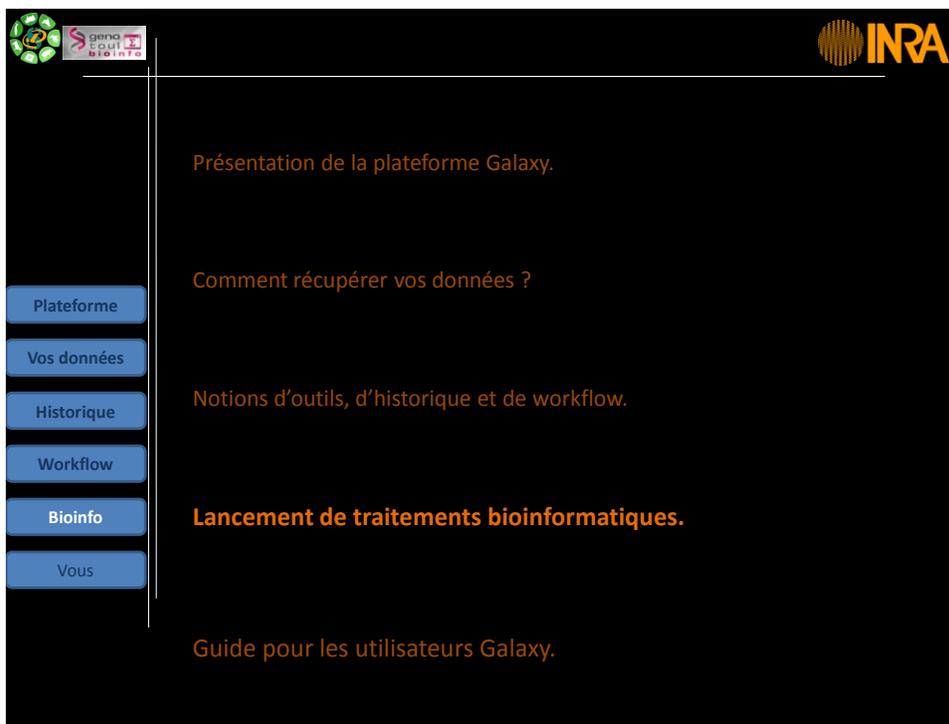
**En résumé ...des datasets aux workflows.**

- 1 – Télécharger vos fichiers de données.
- 2 – Renommer votre historique.
- 3 – Utiliser les outils dont vous avez besoin.
- 4 – Votre historique est sauvegardé automatiquement. Si nécessaire, exporter votre historique en workflow.
- 5 – Editer, partager et lancer vos traitements à volonté (run de votre workflow).

\* Upload local file from filesystem path Upload data to history without copying on server

**Galaxy**  
Your workflows






Présentation de la plateforme Galaxy.

Comment récupérer vos données ?

Notions d'outils, d'historique et de workflow.

**Lancement de traitements bioinformatiques.**

Guide pour les utilisateurs Galaxy.

Plateforme  
 Vos données  
 Historique  
 Workflow  
 Bioinfo  
 Vous






### Les tools Galaxy

Plateforme  
 Vos données  
 Historique  
 Workflow  
 Bioinfo  
 Vous

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

Menu évolutif et organisé par thématique

Ajout d'outils sur demande :

- Tool Shed
- Wrapper à façon (scripts maison, outils bioinfos)

-> Envoyer vos demandes à :  
[sigenae-support@listes.inra.fr](mailto:sigenae-support@listes.inra.fr)



## Outils de traitement de fichiers



Ces outils sont nombreux et constituent une bonne alternative à la ligne de commande.

Les traitements sont automatiquement lancés sur Genotoul (qsub).

Voici les principaux outils « non bioinfo » proposés :

- Join (des fichiers lourds), Substract and Group
- Text Manipulation
- Filter and sort
- Convert Formats

Select first (version 1.0.0)

Select first:  
  
 lines

from:  
 4: UCSC Main on Huma..ne (genome)

**What it does**  
 This tool outputs specified number of lines from the **beginning** of a dataset

**Example**  
 Selecting 2 lines from this:

```
chr7 56632 56652 D17003_CTCF_R6 210 +
chr7 56736 56756 D17003_CTCF_R7 214 +
chr7 56781 56781 D17003_CTCF_R4 220 +
chr7 56792 56792 D17003_CTCF_R7 212 +
chr7 56775 56795 D17003_CTCF_R4 207 +
```

will produce:

```
chr7 56632 56652 D17003_CTCF_R6 210 +
chr7 56736 56756 D17003_CTCF_R7 214 +
```

Plateforme

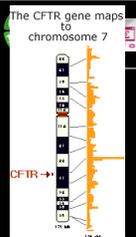
Vos données

Historique

Workflow

Bioinfo

Vous



## Outils bioinformatiques



Mapper un FASTQ sur une référence avec BWA.

**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

Map with BWA for Illumina (version 0.7.3)

Will you select a reference genome?  
 Use one from the history

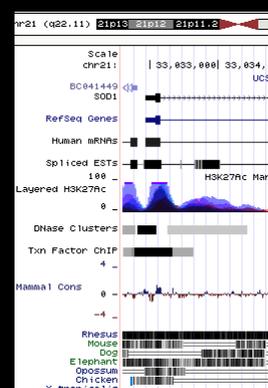
Select a reference from history:  
 29: ERR000017\_ref.fasta

Is this library mate-paired?:  
 Single-end

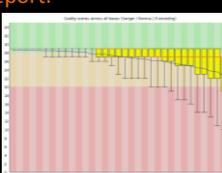
FASTQ file:  
 30: ERR003037.fastqsanger

BWA settings to use:  
 Commonly Used

Suppress the header in the output:



Visualiser la qualité des données avec FASTQC Report.



Visualiser un génome avec UCSC .

Plateforme

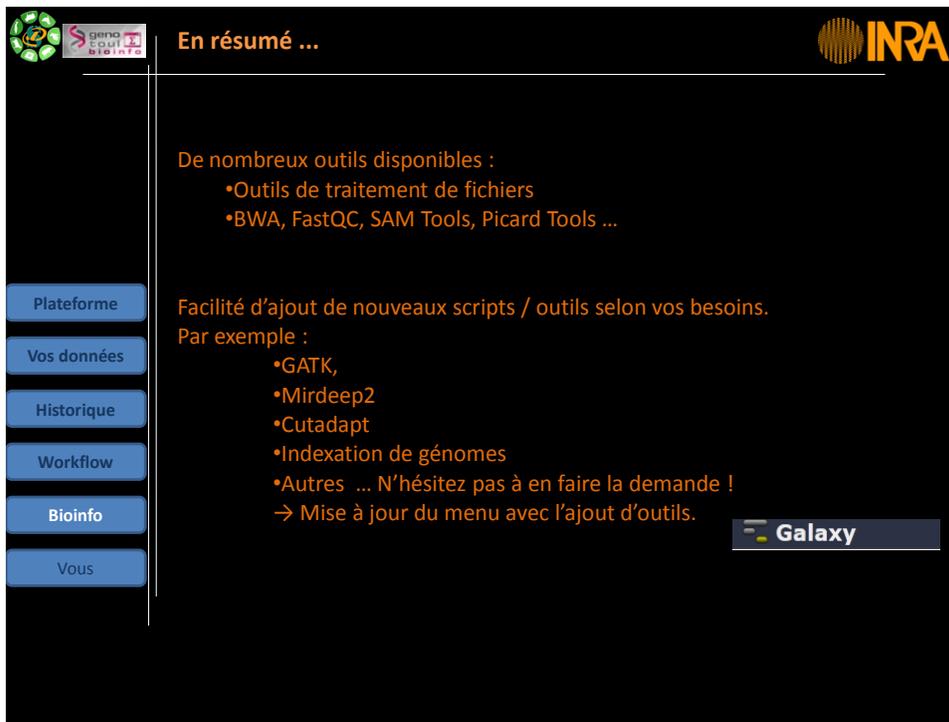
Vos données

Historique

Workflow

Bioinfo

Vous



En résumé ...

De nombreux outils disponibles :

- Outils de traitement de fichiers
- BWA, FastQC, SAM Tools, Picard Tools ...

Facilité d'ajout de nouveaux scripts / outils selon vos besoins.  
Par exemple :

- GATK,
- Mirdeep2
- Cutadapt
- Indexation de génomes
- Autres ... N'hésitez pas à en faire la demande !

→ Mise à jour du menu avec l'ajout d'outils.

Galaxy

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous



Exercice 2 & 3 :

Utilisation d'outils de traitements (bio-)informatiques.

Création et partage d'historiques et de workflows.

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous



Présentation de la plateforme Galaxy.

Comment récupérer vos données ?

Notions d'outils, d'historique et de workflow.

Lancement de traitements bioinformatiques.

**Guide pour les utilisateurs Galaxy.**

Navigation menu items: Plateforme, Vos données, Historique, Workflow, Bioinfo, Vous



**FAQ et formation en ligne**

Une FAQ et le lien vers « sig-learning » sont disponibles depuis la page d'accueil.

Shared Data Lab Visualization Admin Help User User Welcome smaman,

**FAQ on your Galaxy tool**

▼ Dataset, history and workflow ?

**Step 1 : Import your datasets**  
First of all, you have to import your data files thanks to "Data Analysis / Get Data" tool. Then your downloaded datasets are automatically archived in "User / Saved Datasets".

**Step 2 : Select tools and create your history**  
Then you select relevant tool in "Data Analysis", on the left side of Galaxy interface.

Vos supports sont disponibles depuis : <http://sig-learning.toulouse.inra.fr>

"If you need more training about bioinformatic and Galaxy, please connect to [Sigenae e-learning platform](#)"

"How to be a good user ?":

- Maîtrise de son quota
- Organisation de son espace de travail
- Contact: [sigenae-support@listes.inra.fr](mailto:sigenae-support@listes.inra.fr)

Navigation menu items: Plateforme, Vos données, Historique, Workflow, Bioinfo, Vous

**Comment gérer son compte ?**

Suppression des datasets, historiques et workflows inutiles.  
Suppression temporaire / Suppression permanente

The screenshot shows the Galaxy web interface with a sidebar on the left containing navigation buttons: Plateforme, Vos données, Historique, Workflow, Bioinfo, and Vous. The main content area displays a table of workflows with columns for Name, Datasets, Tags, and Size. A context menu is open over a workflow, listing actions such as Delete, Delete Permanently, and Import from File.

Workflow	Datasets	Tags	Size
Workflow	1	0 Tags	320...
...: fichiers_abs_du	4	0 Tags	3.6 G
...ation_SNP	17	0 Tags	2.4 M
...ted: ...nTargetCreator	6	10	3.6 G
...Mirdeep2_sans ...ination de la ...ndance intra ni inter	51	0 Tags	12.8
...ory_archive	4	0 Tags	0 byt
...ted: Unnamed ...ry	6	0 Tags	742..
...iere session ...ation Galaxy	21	0 Tags	1.4 G
...- GALAXY	16	0 Tags	1.6 G
...NGS - ...norphisme	14	6	0 byt
...NGS RNA Analysis	4	2	41.1

selected histories: [Rename](#) [Delete](#) [Delete Permanently](#) [Undelete](#)

**Revoir les historiques temporairement supprimés**

The screenshot shows the Galaxy web interface with the sidebar on the left. The main content area displays the 'Saved Histories' dialog box, which includes search filters and a table of saved histories.

Close Advanced Search

name:

tags:

sharing: [private](#) | [shared](#) | [accessible](#) | [published](#) | [all](#)

status: [active](#) | [deleted](#) | [all](#)

<input type="checkbox"/>	Name	Datasets	Tags
<input type="checkbox"/>	TP Galaxy project	2	1 0 Tags
<input type="checkbox"/>	miRNA tests	59	21 0 Tags
<input type="checkbox"/>	TP SNPs calling	84	9 0 Tags
<input type="checkbox"/>	TP RNAseq	88	1 0 Tags
<input type="checkbox"/>	test TP miRNA	36	1 1 0 Tags



## Comment citer Galaxy dans vos publications ?



Plateforme

Vos données

Historique

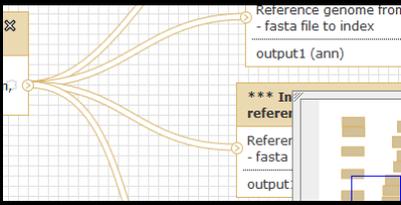
Workflow

Bioinfo

Vous

Pour vos publications, citer:

- ✓ Les outils utilisés (nom, version).
- ✓ Le workflow généré.
- ✓ Les références « Galaxy project ».



### Extract features (version 1.0.0)

Select GFF data:

From:

Extract features:

Multi-select list - hold the appropriate key while clicking to

**What it does**  
 This tool extracts selected features from GFF data.

### Primary Publications

If you use or extend Galaxy in your published work, please cite **each** of the following publications:

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, "Galaxy: a platform for interactive large-scale genome analysis." *Genome Research.* 2005 Oct; 15(



## En conclusion ...



### GALAXY

- ✓ Simplicité d'utilisation (sans Linux) pour lancer des calculs sur le cluster.
- ✓ Gérer les résultats et traitements d'une ou plusieurs analyses.
- ✓ Partage de vos datasets, historiques et workflows (intra et inter Galaxy).
- ✓ Présentation schématique de vos traitements grâce aux workflows.
- ✓ Possibilité d'ajout de nouveaux outils selon vos besoins.  
 (\*Outil Siganae avec ligne commande)




# Votre auto-formation continue en ligne avec « sig-learning »




## Votre accès à sig-learning

- 1 Taper l'adresse de « sig-learning » : <http://sig-learning.toulouse.inra.fr/>
- 2 Authentification
- 3 Accès à la liste des formations auxquelles l'utilisateur s'est inscrit.

Firefox

Sigeneae ... x Galaxy

sig-learning.toulouse.inra.fr

Login

  
 Pass

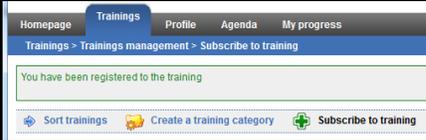
Plateforme

Vos formations

Homepage	Trainings	Profile	Agenda	Reporting	Administration
Trainings					
<ul style="list-style-type: none"> <li> 1 - Linux &amp; Unix UNIX1 – SIGENAE Team</li> <li> 2 - Cluster (en construction) CLUSTER – SIGENAE Team</li> <li> 3 - Management of large files on Unix and Galaxy UNIX2 – SIGENAE Team</li> </ul>					

 **Comment s'inscrire à des formations ?** 

Il vous est possible de vous inscrire directement en ligne à une formation :  
« Trainings » « Trainings management » puis « Subscribe to training » :



**Plateforme**  
**Vos formations**

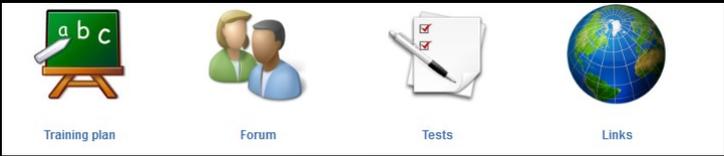
L'inscription s'effectue via une recherche de la formation par mots clés.  
Voici donc la liste des formations :



- 1 - Linux & Unix  
UNIX1 – SIGENAE Team
- 2 - Cluster (en construction)  
CLUSTER – SIGENAE Team
- 3 - Management of large files on Unix and Galaxy  
UNIX2 – SIGENAE Team
- 4 - Galaxy  
GALA01 – SIGENAE Team
- 5 - FastQC : quality control tool & reports interpretation  
FASTQC – SIGENAE Team
- 6 - Aligning SGS reads and SNP finding  
SGS-SNP – SIGENAE Team
- 7 - NGS  
NGS – SIGENAE Team
- 8 - RNA seq (en construction)  
RNASEQ – SIGENAE Team
- Demonstration  
DEMO – SIGENAE Team

 **Page d'accueil de votre formation** 

Outre une introduction et un carrousel permettant d'accéder aux principaux chapitres de la formation, la page d'accueil de la formation donne accès :



**Plateforme**  
**Vos formations**

**TRAINING PLAN** : Parcours pédagogique avec les supports en ligne.

**FORUM** : Support de communication entre stagiaires / formateurs.

**TESTS** : Tests et exercices.

**LINKS** : Liens utiles.




## Comment ouvrir un compte sur GALAXY Sigenae et sig-learning ?

- 1 – Demande à compte sur la plateforme BIOINFO GENOTOUL :  
<http://bioinfo.genotoul.fr>  
 Puis « menu « Help » , puis « Create an account ».  
 Vous recevrez un login et mot de passe LDAP Genotoul.
- 2 – Puis utilisez ce login et mot de passe LDAP Genotoul lorsque vous souhaitez accéder à :
  - Instance Sigenae de Galaxy : <http://galaxy-workbench.toulouse.inra.fr/>
  - « Sig-learning » : <http://sig-learning.toulouse.inra.fr/>
- 3 - Pour demander une augmentation de votre quota utilisateur sur Galaxy, veuillez vous adresser à :  
[sigenae-support@listes.inra.fr](mailto:sigenae-support@listes.inra.fr)

Plateforme

Vos formations

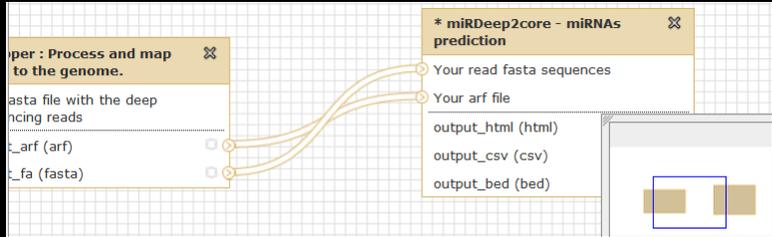



## Merci pour votre écoute

Questionnaire  
<http://bioinfo.genotoul.fr/index.php?id=130>



**Trop de datasets dans votre workflow ?**

\*Si tout est coché, alors tout se passe comme si rien est coché.

\*Si le dataset n'est pas coché, alors qu'au moins un autre est coché, alors le dataset non coché ne sera pas visible dans l'historique. Sauf si l'utilisateur choisi d'afficher les fichiers cachés, alors, dans cette config uniquement, il pourra le voir.

\*Si un 2ieme (ou plus) dataset est coché , alors il sera visible dans l'historique après analyse.