

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

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



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.

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.



Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous



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



Plateforme

Une communauté nationale et internationale très active :

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

Vos données



Historique



Workflow

Bioinfo

Vous

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)

Plateforme

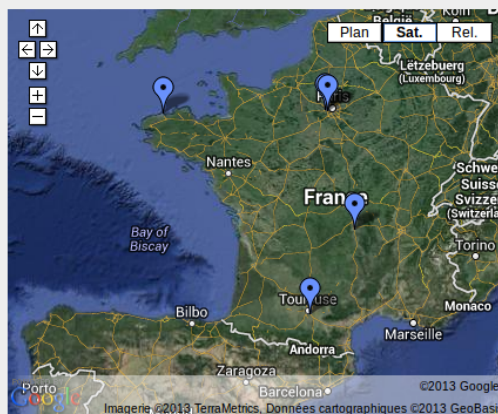
Vos données

Historique

Workflow

Bioinfo

Vous



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 Giacomoni

Inutile de savoir :

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

Plateforme

Inutile de s'inquiéter pour son disque dur:

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

Vos données

Historique

Workflow

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.

Bioinfo

Vous

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

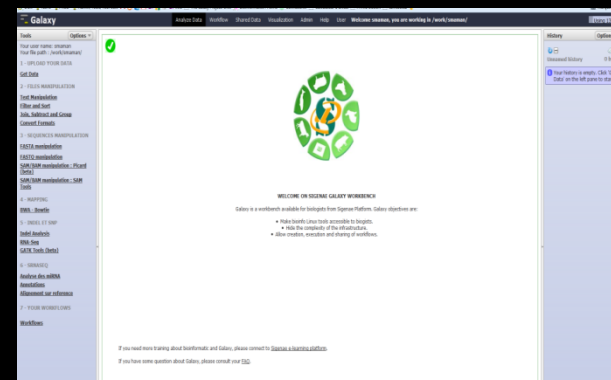
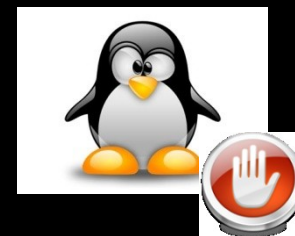
Vos données

Historique

Workflow

Bioinfo

Vous



Plateforme

Vos données

Historique

Workflow

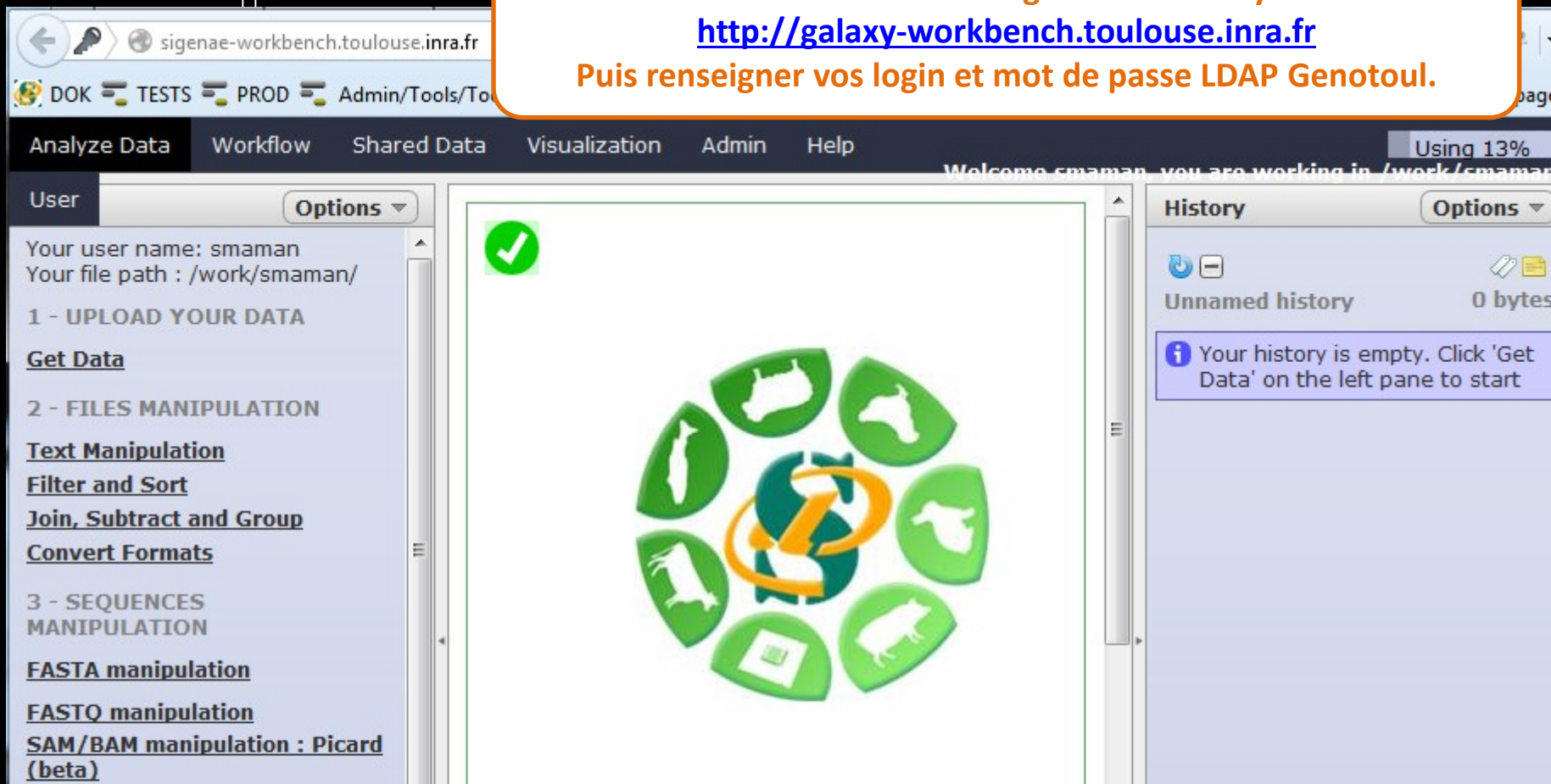
Bioinfo

Vous

- ✓ 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 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



Exécute



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



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)



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.

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

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

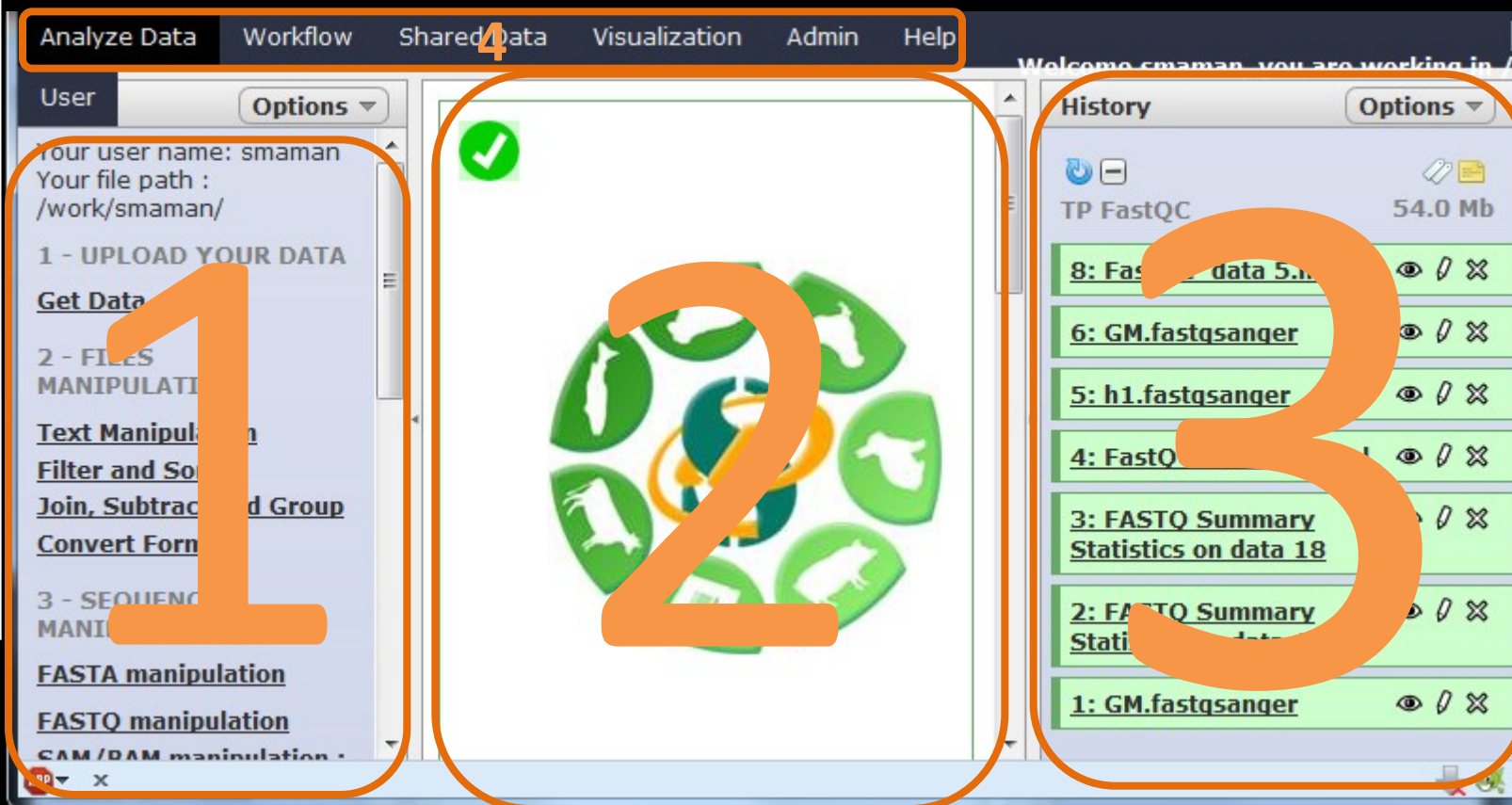
15: Map with BWA for Illumina on data 14 and data 11: mapped reads
Job is waiting to run

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



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.

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

Upload File from your computer

TOOL

génère

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

```
@HWUSI-EAS627_1:8:1:17:202
TGGTTGGAGACCCCGATGGTCTTCAGCTCC
+
BB@=A<9@?AB??=?B?B?B33>?A@?><1:9=A
@HWUSI-EAS627_1:8:1:66:1050
CAGAAGTAGAGCAGAAGAAGACCGGACCTCCG
```

forme

History Options

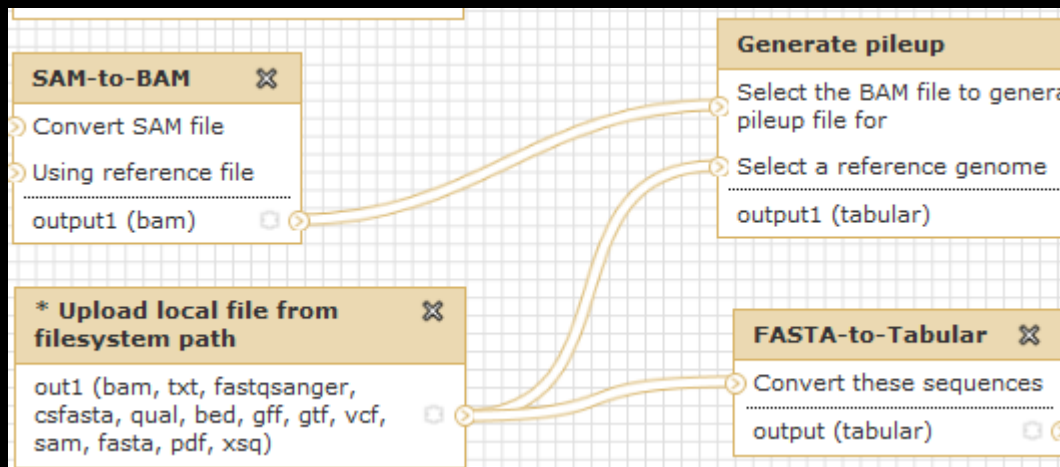
TP FastQC 54.0 Mb

- 8: FastQC_data 5.html
- 6: GM.fastqsanger
- 5: h1.fastqsanger
- 4: FastQC_data 18.html

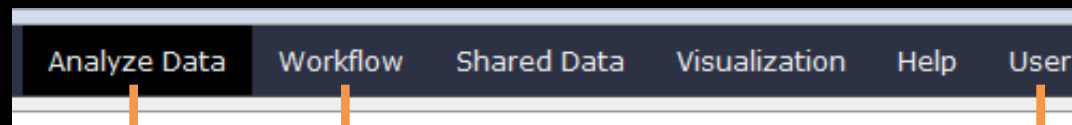
DATASET (S)

HISTORY

WORKFLOW



- **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.



Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous

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.

Saved Histories

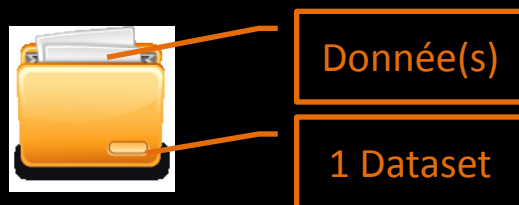
search history names and tags

Advanced Search

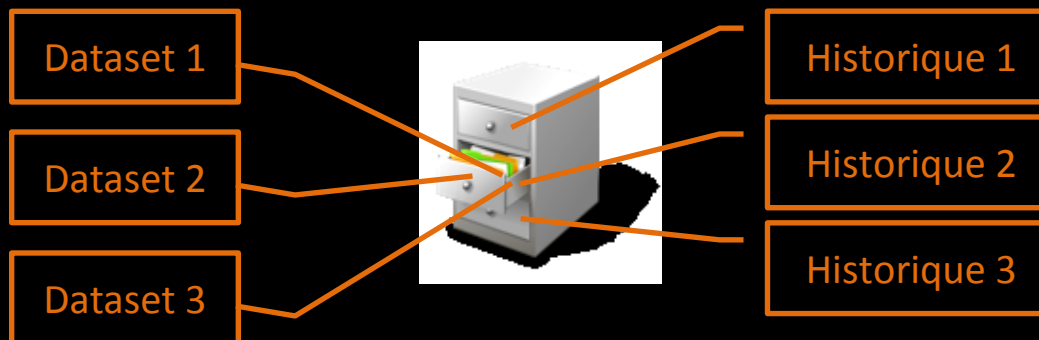
<input type="checkbox"/> Name	Datasets
<input type="checkbox"/> <u>Unnamed history</u> ▾	
<input type="checkbox"/> <u>Unnamed history</u> ▾	
<input type="checkbox"/> <u>Unnamed history</u> ▾	
<input type="checkbox"/> <u>TP FastQC</u> ▾	7

En résumé ...

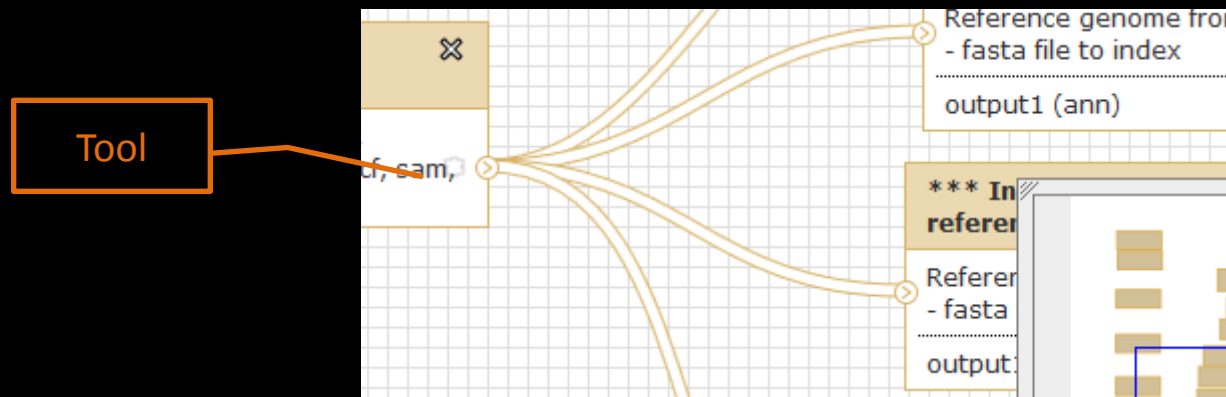
Un **DATASET** est un fichier de données (fichiers d'entrée, fichier résultat) :



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



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.

Plateforme

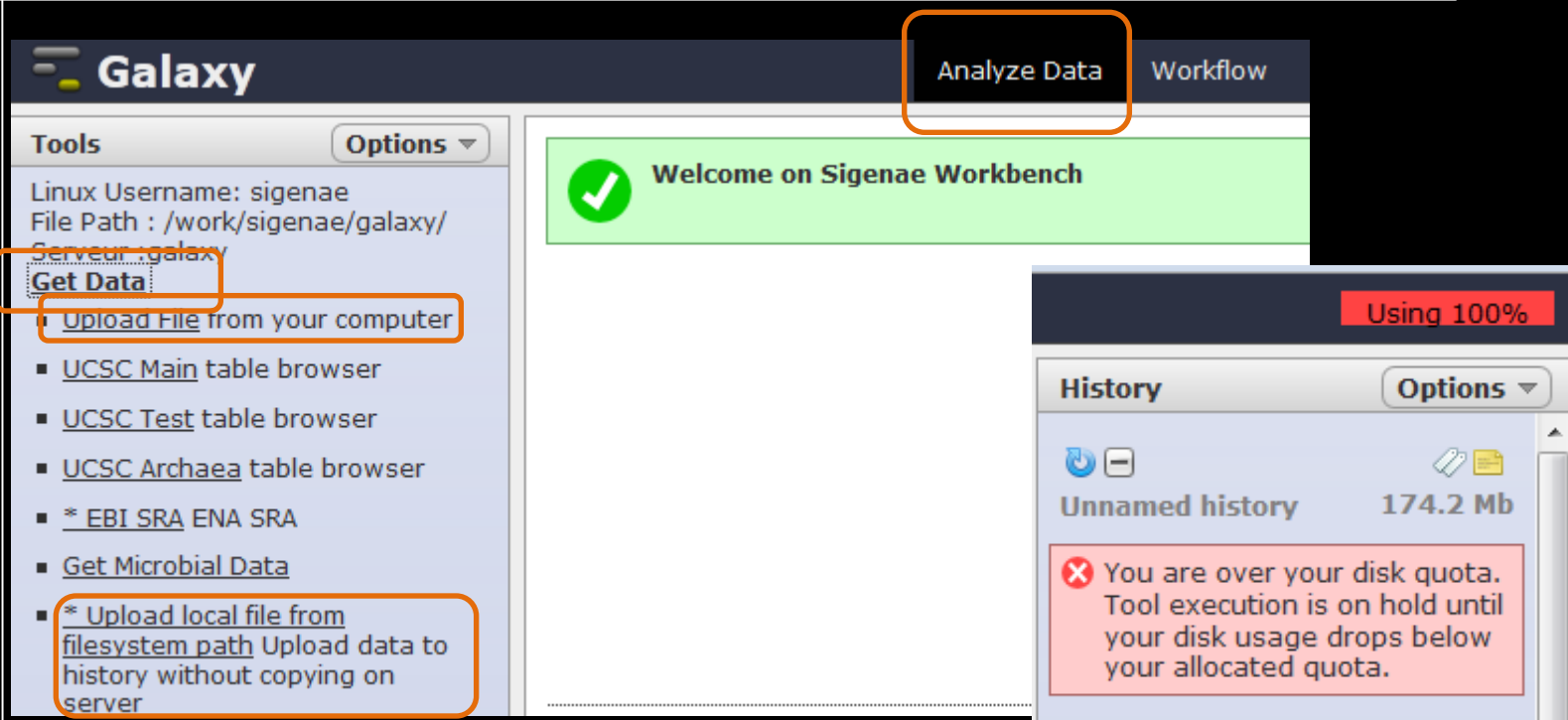
Vos données

Historique

Workflow

Bioinfo

Vous



The screenshot shows the Galaxy web interface. In the top right, the 'Analyze Data' button is highlighted with an orange box. On the left, the 'Tools' menu is open, and 'Get Data' is highlighted with an orange box. Below it, 'Upload File from your computer' and '* Upload local file from filesystem path' are also highlighted with orange boxes. The 'History' panel on the right shows 'Using 100%' in a red box and a red error message: '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.

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous

Données UCSC, Ensembl, NG6, BIOMART :



Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous

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

Galaxy Analyze Data Workflow Shared Data Admin Help User [Welcome smaman, you are wo](#)
Tools Options
 Linux Username: smaman
 File Path: /work/smaman/galaxy/
 Serveur: galaxy
Get Data
 ■ Upload local file from filesystem_path Upload data to history without copying on server
 ■ Upload File from your computer
 ■ EBI_SRA ENA_SRA
 ■ UCSC Main table browser
 ■ UCSC Test table browser
 ■ UCSC Archaea table browser
 ■ BX_main browser
 ■ Get Microbial Data
 ■ BioMart Central server

EMBL-EBI **ENA** **European Nucleotide Archive**
 The European Nucleotide Archive (ENA) provides a comprehensive record of the world's nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation... [more](#)
 Access to ENA data is provided through the browser, through search tools, large scale file download and through the API.
NEWS AND ANNOUNCEMENTS
 CRAM toolkit 0.7 released
 7 Mar 2012
 CRAM toolkit 0.7 has been released. More information with download, installation and usage instructions...

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

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.

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous

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

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

- 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.

<input type="checkbox"/>	SwanPorc ▾	18	0 Tags	Shared	0 bytes	Ju 20
<input type="checkbox"/>	FastQC ▾	6	0 Tags	Shared	17.4 Mb	Ap 20
<input type="checkbox"/>	TP : NGS - Polymorphisme ▾	8	2	0 Tags	Shared	6.6 Gb
<input type="checkbox"/>	TP FastQC ▾	12	16	0 Tags	54.0 Mb	Ma 20
<input type="checkbox"/>	indexation genome ▾	1	0 Tags	46 bytes	Ju 20	

For 0 selected histories: Rename Delete Delete Permanently

The screenshot shows the Galaxy Sig interface. At the top, there are navigation tabs: 'Analyze Data', 'Workflow', 'Shared Data', and 'Visualizations'. The 'Shared Data' tab is active, showing a dropdown menu with options: 'Data Libraries', 'Published Histories', 'Published Workflow', and 'Published Visualizations'. The main content area is titled 'Published Histories' and contains a search bar with the text 'search name, annotation, owner, and tags' and an 'Advanced Search' link. Below the search bar is a table with columns 'Name' and 'Annotation'. One entry is visible: 'RNAseq' with the annotation 'TP 1 suite : region promotrices'. To the right, a 'User' menu is open, showing 'Welcome smaman' and a list of options: 'History Lists', 'Saved Histories', 'Histories Shared with Me', 'Current History', and 'Create New'. Below this, another section titled 'Histories shared with you by others' shows a table with columns 'Name' and 'Datasets'.

- 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).

This screenshot shows the 'User' menu open in the Galaxy Sig interface. The menu options are: 'Logged in as smaman@toulouse.inra.fr', 'Logout', 'Saved Histories', 'Saved Datasets', and 'Public Name'. Below the menu, a table of datasets is visible. The table has columns for 'Datasets' and 'Tags'. The first row shows '2' datasets and '1' tag. The second row shows '59' datasets and '21' tags. The third row shows '0' datasets and '0' tags. A specific dataset is highlighted: '2: UCSC Main on Human: snp137Common (chr22:1-51304566)'.

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

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous

1 – Télécharger vos données :

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

File Name:

File type:

Path to file:

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
GATGTTATTTCTTCATTGGAGGTAACCTCTTAT
+
IIIIIIIIIIIIIIIIIIII<III@FI8A/I0II4I
@080917-and-080922:5:1:1366:223
GTTTCCTTCGCGTCAGTAAGAAGCTCAGTGTTC
```

11: phiX174_genome.fa
 1 sequences
 format: fasta, database: ?
 Info: uploaded fasta file sur :
 ftp://ftp.gmod.org/pub/gmod/Courses/2010/SummerSchoolAmericas/Galaxy/phiX174_genome.fa

```
>phix
GAGTTTTATCGCTTCATGACGCAGAAGTTAACCTT
AAATTATCTTGATAAAGCAGGAATTACTACTGCTTGT
TGCTGGCGGAAAATGAGAAAATTCGACCTATCCTTGC
GCGACCTTTCGCCATCAACTAACGATTCTGTCAAAA
TGGCTTAATATGCTTGGCACGTTCTGTCAAGGACTGGT
```



Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous

2 – Choisir un outil dans « Tools » :

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 1.2.2)

Will you select a reference genome from your history?

Select a reference from history:

Is this library mate-paired?:

FASTQ file:

FASTQ with either Sanger-scaled quality values (

3 – Lancer le job en cliquant sur « Executer ».

L'Execution du job en cours est visible dans votre historique.

Fini les lignes de commande !

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



Plateforme

Vos données


Historique

Workflow

Bioinfo

Vous


Etat 1 – GRIS : Votre job est en file d'attente.

 **107: test.txt** 👁️ 🗑️ ✖️

Job is waiting to run

[i](#) [🔄](#)

Etat 2 – JAUNE : Votre job est en cours d'exécution

 **1: test.txt** 👁️ 🗑️ ✖️

Job is currently running

[i](#) [🔄](#)

Etat 3 – VERT : Votre job est fini.


11: phiX174 genome.fa 👁️ 🗑️ ✖️

1 sequences
format: fasta, database: ?
Info: uploaded fasta file sur :
ftp://ftp.gmod.org/pub/gmod/Courses/2010/SummerSchoolAmericas/Galaxy/phiX174_genome.fa


[📄](#) [i](#) [🔄](#) [📄](#) [📄](#)

```
>phix
GAGTTTTATCGCTTCCATGACGCAGAAGTTAACACTT
AAATTATCTTGATAAAGCAGGAATTACTACTGCTTGI
TGCTGGCGGAAAATGAGAAAATTCGACCTATCCTTGC
GCGACCTTTCGCCATCAACTAACGATTCTGTCAAAA
TGGCTTAATATGCTTGGCACGTTTCGTCAAGGACTGGT
```

Bug - ROUGE : Votre job est planté !

 **1: test.txt** 👁️ 🗑️ ✖️

0 bytes
An error occurred running this job:
info unavailable

 [i](#) [🔄](#)



Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous

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	Jobs	Tags	Sharing
<input type="checkbox"/> TP Galaxy project	2	1	0 Tags	7
<input type="checkbox"/> miRNA tests	59	21	0 Tags	3
<input type="checkbox"/> TP SNPs calling	84	9	0 Tags	5
<input type="checkbox"/> TP RNAseq	88	1	0 Tags	Shared, Accessible 2
<input type="checkbox"/> test TP miRNA	36	1	0 Tags	5
<input type="checkbox"/> Unnamed history			0 Tags	0

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

Analyse OK

Analyse en attente

Analyse en erreur

Analyse en cours

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

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 !

Plateforme

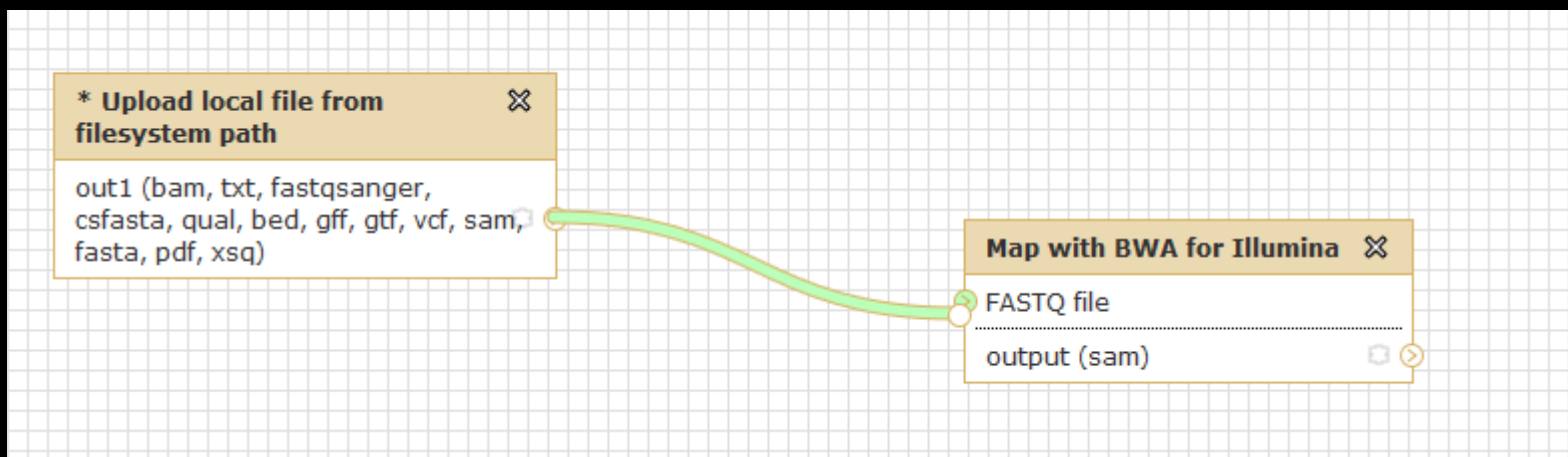
Vos données

Historique

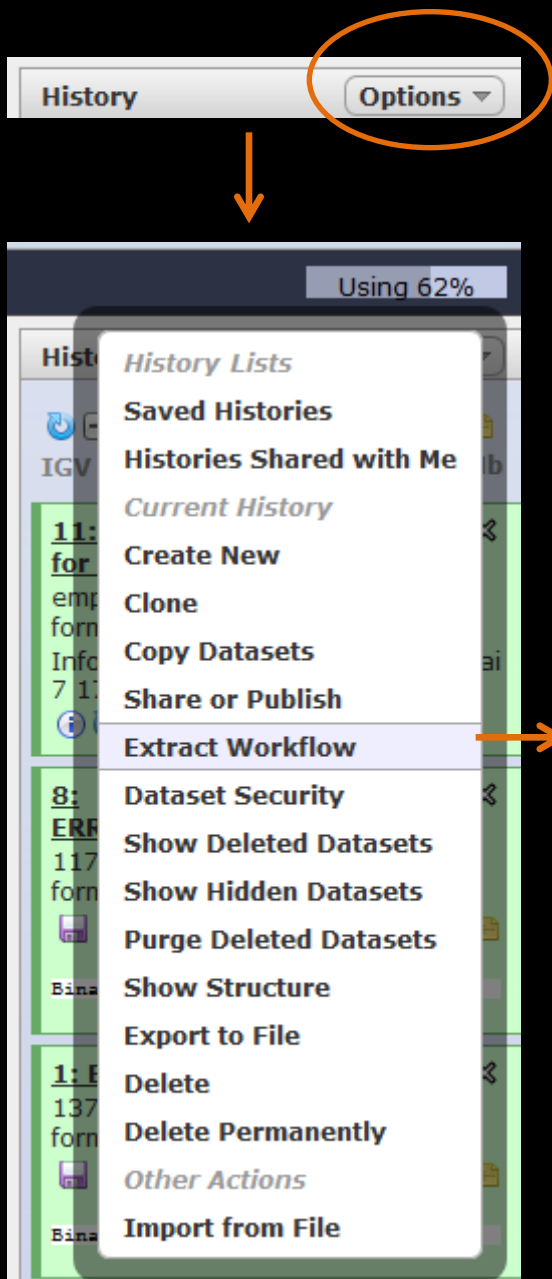
Workflow

Bioinfo

Vous



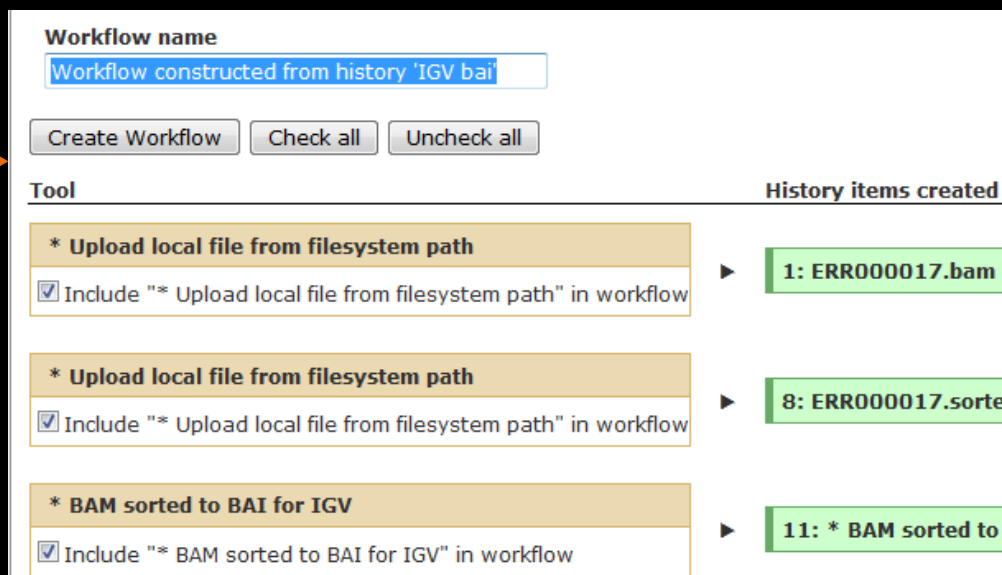
- Plateforme
- Vos données
- Historique
- Workflow
- Bioinfo
- Vous



The screenshot shows the 'History' window with a dropdown menu open. The 'Options' button is circled in orange. The dropdown menu contains the following items:

- History Lists
- Saved Histories
- Histories Shared with Me
- Current History
- Create New
- Clone
- Copy Datasets
- Share or Publish
- Extract Workflow** (highlighted)
- Dataset Security
- Show Deleted Datasets
- Show Hidden Datasets
- Purge Deleted Datasets
- Show Structure
- Export to File
- Delete
- Delete Permanently
- Other Actions
- Import from File

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



The screenshot shows the 'Workflow name' dialog box. The workflow name is 'Workflow constructed from history 'IGV bai''. Below the name are three buttons: 'Create Workflow', 'Check all', and 'Uncheck all'. The dialog also shows a list of tools to be included in the workflow:

Tool	History items created
* Upload local file from filesystem path <input checked="" type="checkbox"/> Include "** Upload local file from filesystem path" in workflow	▶ 1: ERR000017.bam
* Upload local file from filesystem path <input checked="" type="checkbox"/> Include "** Upload local file from filesystem path" in workflow	▶ 8: ERR000017.sorte
* BAM sorted to BAI for IGV <input checked="" type="checkbox"/> Include "** BAM sorted to BAI for IGV" in workflow	▶ 11: * BAM sorted to

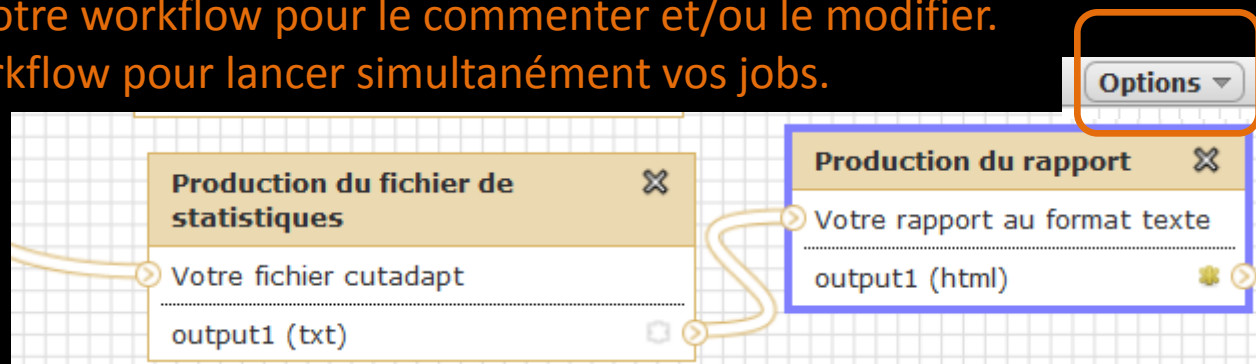
Cliquer sur le menu « Workflow » pour lister vos workflows :



The screenshot shows the Galaxy web interface. At the top right, there are two buttons: 'Analyze Data' and 'Workflow', with the 'Workflow' button highlighted by an orange box. Below the header, the page title is 'Your workflows'. A table lists several workflows, with the first one named 'toto'. A context menu is open over the 'toto' workflow, listing actions: Edit, Run, Share or Publish, Download or Export, Clone, Rename, View, and Delete. The 'Run' option is highlighted. Below the table, there is a section titled 'Workflows shared by you by others' with a similar table structure.

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.



The screenshot shows a workflow diagram on a grid background. Two jobs are connected by a data link. The first job is 'Production du fichier de statistiques' with an output 'output1 (txt)'. The second job is 'Production du rapport' with an input 'output1 (html)'. An 'Options' dropdown menu is highlighted with an orange box in the top right corner of the workflow area.

Plateforme

Vos données

Historique

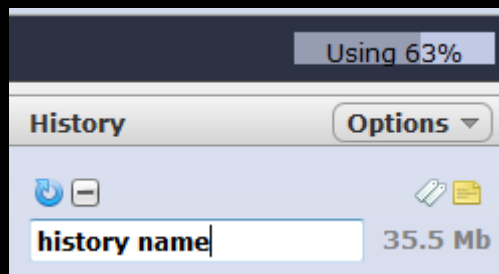
Workflow

Bioinfo

Vous

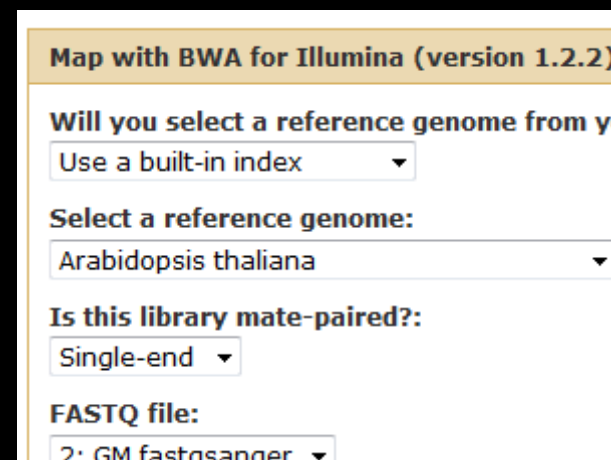
1 – Télécharger vos fichiers de données.

2 – Renommer votre historique.

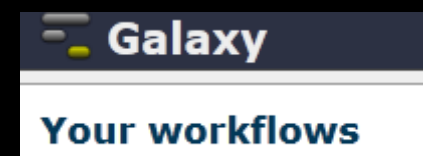


- * Upload local file from [filesystem path](#) Upload data to history without copying on server

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).

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.

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous

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

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



Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous

Select first (version 1.0.0)

Select first:

lines

from:

What it does

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

Example

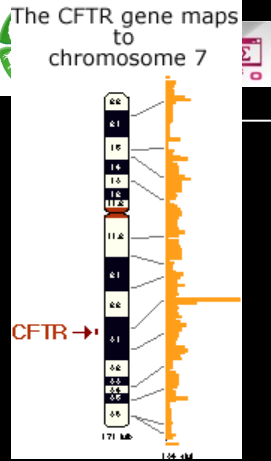
Selecting 2 lines from this:

```
chr7 56632 56652 D17003_CTCF_R6 310 +
chr7 56736 56756 D17003_CTCF_R7 354 +
chr7 56761 56781 D17003_CTCF_R4 220 +
chr7 56772 56792 D17003_CTCF_R7 372 +
chr7 56775 56795 D17003_CTCF_R4 207 +
```

will produce:

```
chr7 56632 56652 D17003_CTCF_R6 310 +
chr7 56736 56756 D17003_CTCF_R7 354 +
```

Outils bioinformatiques



<http://www.genomenewsnetwork.org/>

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

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 0.12.1)

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

Select a reference from history: 29: ERR000017_ref.fasta

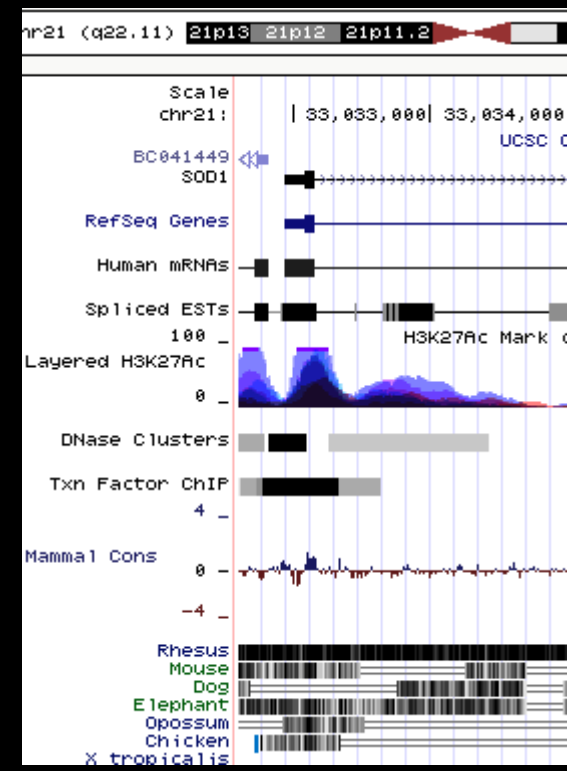
Is this library mate-paired?: Single-end

FASTQ file: 30: ERR003037.fastqsanger

BWA settings to use: Commonly Used

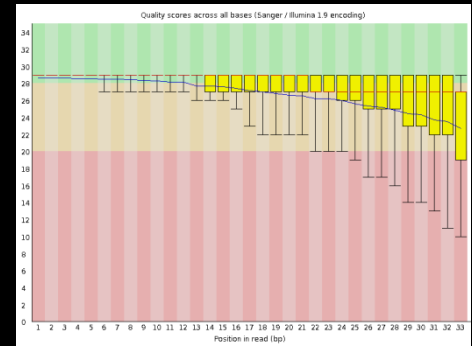
Suppress the header in the output:

Execute



Visualiser la qualité des données avec FASTQC Report.

Visualiser un génome avec UCSC .



En résumé ...

De nombreux outils disponibles :

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

Plateforme

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

Par exemple :

Vos données

• GATK,

• Mirdeep2

• Cutadapt

Historique


• Indexation de génomes

• Autres ... N'hésitez pas à en faire la demande !

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

Workflow

Bioinfo



Galaxy

Vous

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.

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

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

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous

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

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

The screenshot shows the Galaxy interface with a table of datasets and a context menu open over the history list.

Workflow	Shared Data	Visualization	Admin	Help	User	Welcome smaman	Using 30%
scripmonk	1	3	0 Tags			328.0	
: fichiers abs du	4		0 Tags			3.6 G	
otation SNP	17		0 Tags			2.4 M	
orted: gnTargetCreator	6	10	0 Tags			3.6 G	
Mirdeep2 sans ination de la ndance intra ni inter	51		0 Tags	Accessible		12.8	
ory archive	4		0 Tags	Shared		0 byt	
orted: Unnamed	6		0 Tags			742.0	
iere session ation Galaxy	21		0 Tags			1.4 G	
- GALAXY	16		0 Tags			1.6 G	
NGS - morphisme	14	6	0 Tags	Shared		0 byt	
NGS RNA Analysis	4	2	0 Tags			41.1	

selected histories: **Rename** **Delete** **Delete Permanently** **Undelet**

History Lists

- Saved Histories
- Histories Shared with Me
- Current History
- Create New
- Clone
- Copy Datasets
- Share or Publish
- Extract Workflow
- Dataset Security
- Show Deleted Datasets
- Show Hidden Datasets
- Purge Deleted Datasets
- Show Structure
- Export to File
- Delete
- Delete Permanently
- Other Actions
- Import from File

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous

Saved Histories

[Close Advanced Search](#)

name:

tags:

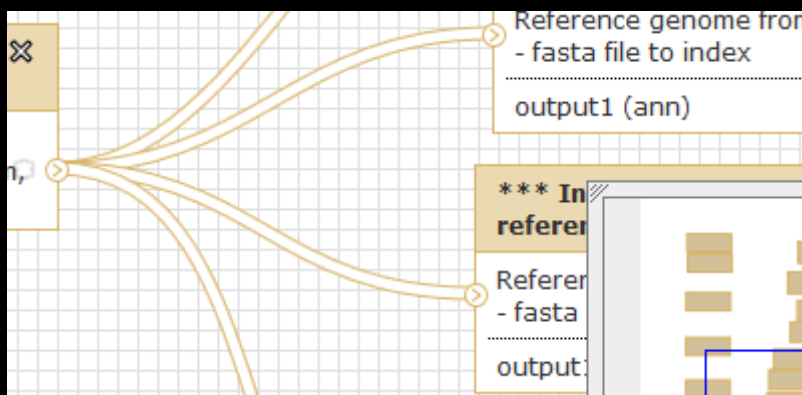
sharing: [private](#) | [shared](#) | [accessible](#) | [published](#) | **all**

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

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:

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(

Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous



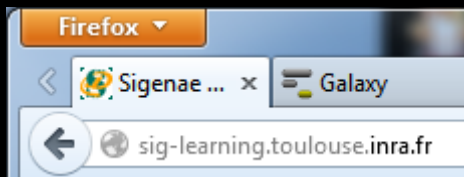
GALAXY

- ✓ **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 Sigenae avec ligne commande)

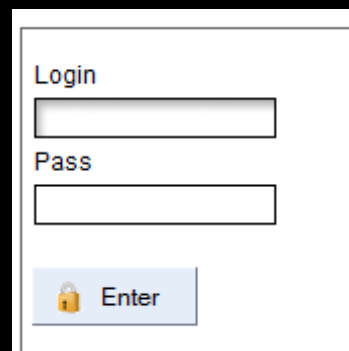


Votre auto-formation
continue en ligne avec
« sig-learning »

- 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.



Homepage **Trainings** Profile Agenda Reporting Administration

Trainings

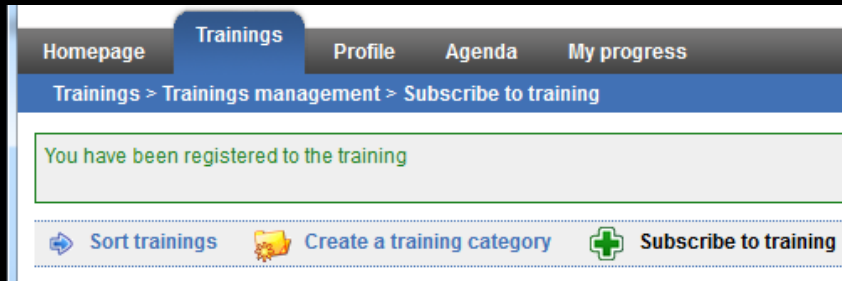
- 1 - Linux & Unix
UNIX1 – SIGENAE Team
- 2 - Cluster (en construction)
CLUSTER – SIGENAE Team
- 3 - Management of large files on Unix and Galaxy
UNIX2 – SIGENAE Team

Plateforme

Vos formations



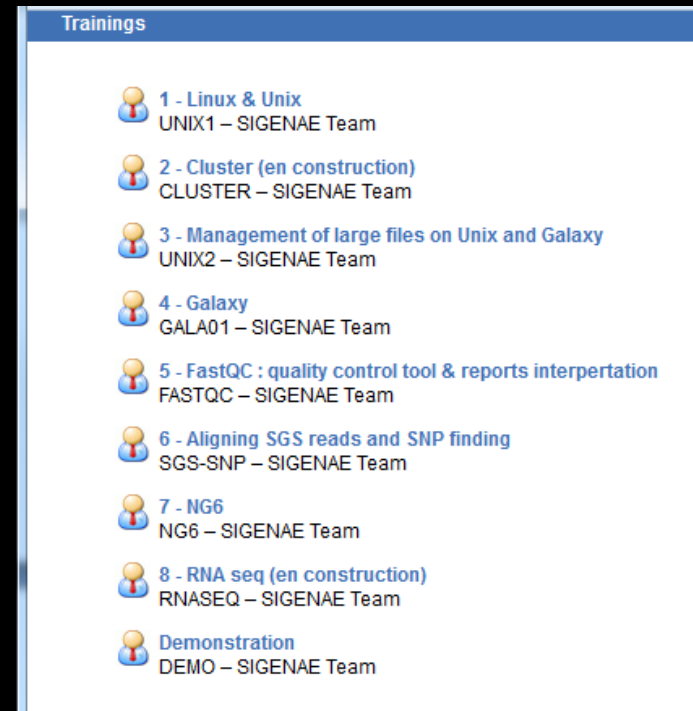
Il vous est possible de vous inscrire directement en ligne à une formation :
« Trainings » « Trainings management » puis « Subscribe to training » :



Plateforme

Vos
formations

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





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



Training plan



Forum



Tests



Links

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.



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 :

Plateforme

Vos
formations

· 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



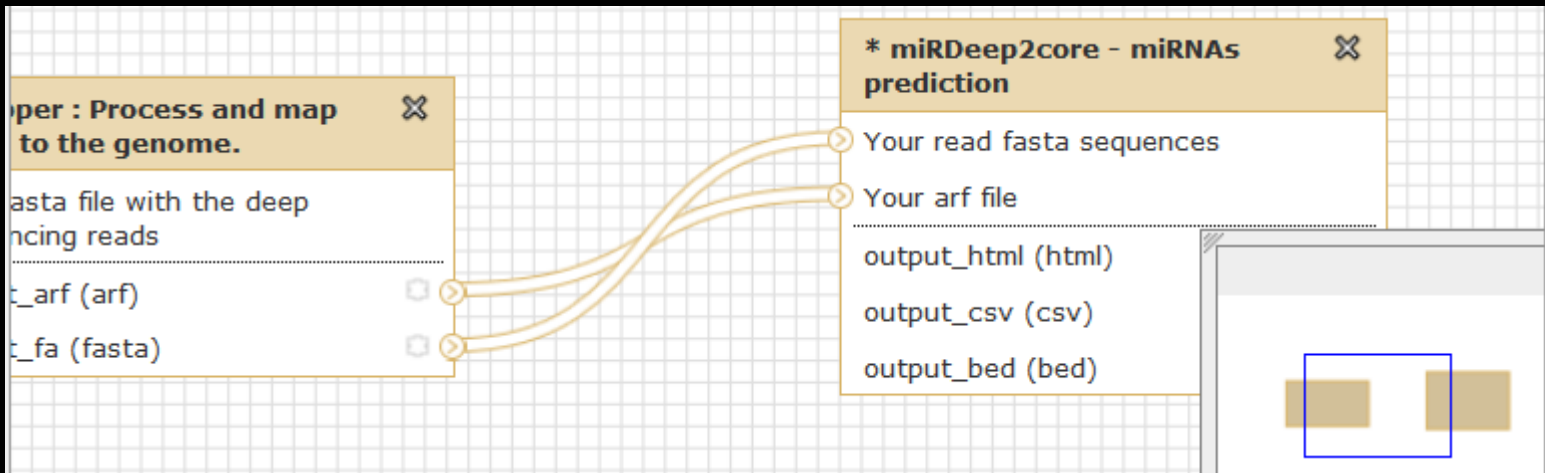
Merci pour votre écoute

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.