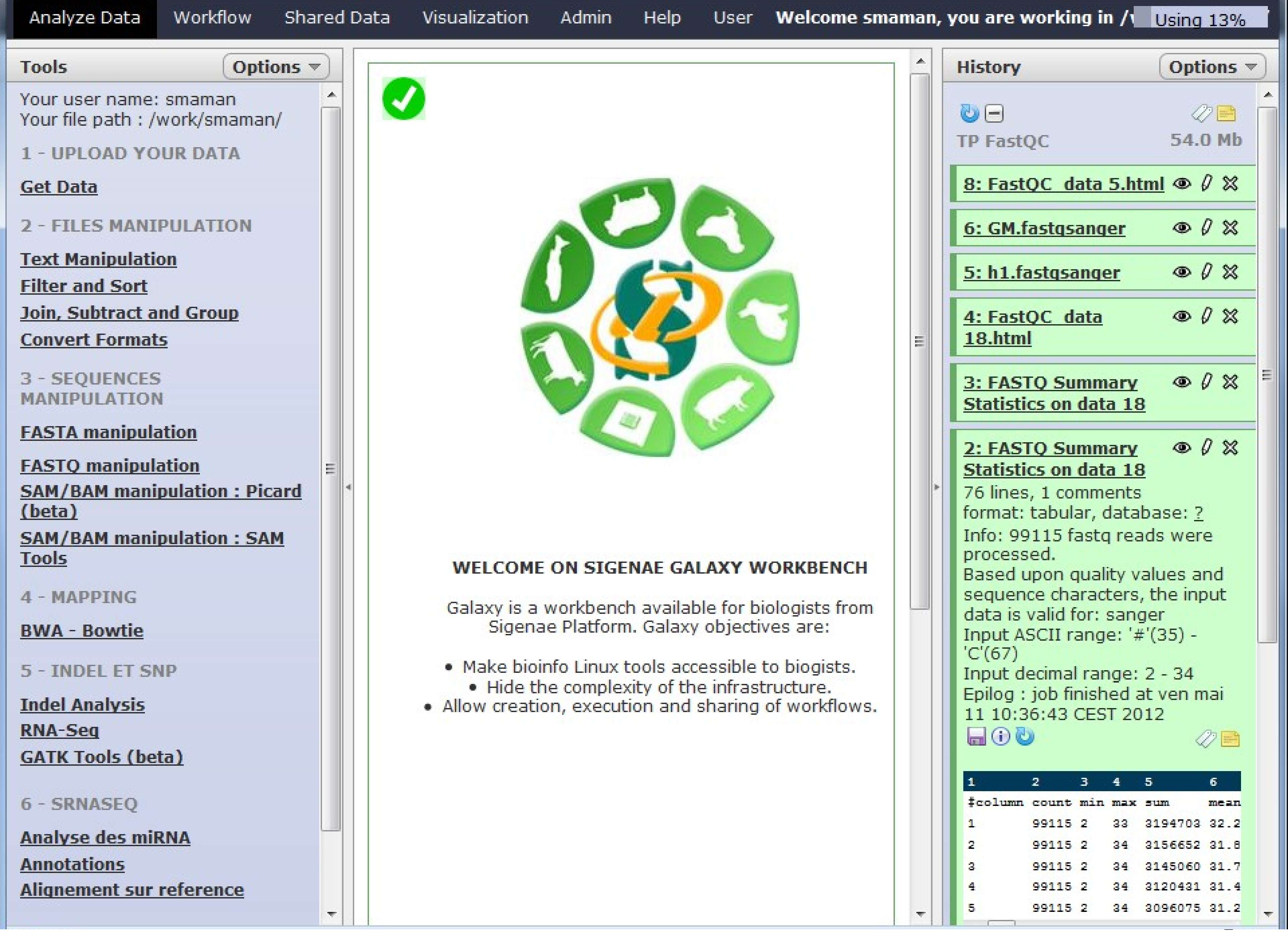


Your user name: smaman
Your file path : /work/smaman/

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

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.

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

Galaxy Training days

Durée / Programme : 2journées ½.

Galaxy : First step.
Galaxy : Reads alignment and SNP calling.

Public : Personnes souhaitant traiter des données (bio)informatiques sans connaissances spécifiques en informatique (sans avoir à connaître Linux et la ligne de commande).

Liste des sessions disponibles : http://bioinfo.genotoul.fr

Les formateurs :

Jour 1 « Initiation » : Philippe Bardou, Sabrina Legoueix-Rodriguez.
Jour 2 + 3 « SNP » : Philippe Bardou, Sabrina Legoueix-Rodriguez.

Vos traitements bioinformatiques avec GALAXY

Sabrina Legoueix-Rodriguez – Philippe Bardou
Oct 2016

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

Vidéo disponible
sur « sig-learning »

Présentation de la plateforme Galaxy.

Premiers pas dans l’instance.

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

Lancement de traitements bioinformatiques.

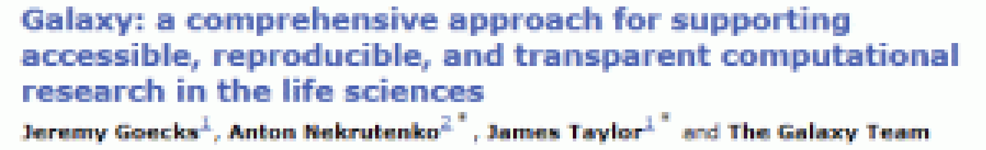
Auto-formations disponibles en ligne.

Galaxy Project

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.

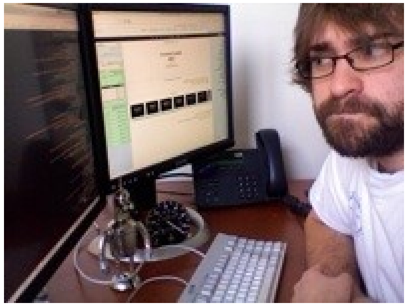
Une communauté active autour de cet outil.



Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences

Jeremy Goecks[1], Anton Nekrutenko[2 *], James Taylor[1 *] and The Galaxy Team

Anton Nekrutenko
Penn State



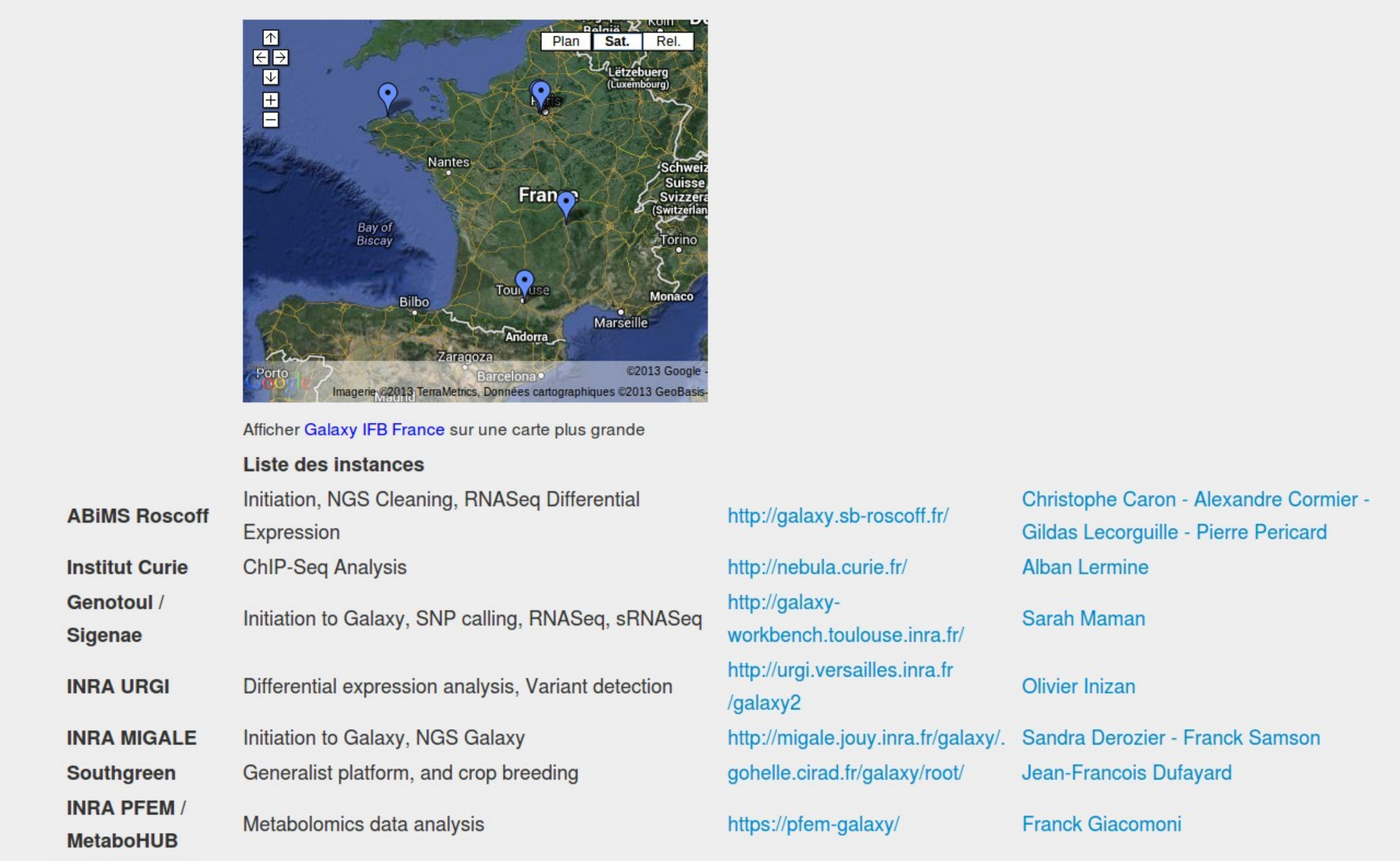
Nate Coraor
Penn State

James Taylor
Emory

Instance intégrée dans une communauté

Groupe de travail Galaxy IFB
- Documentation collaborative (wiki)
- Formations (mise en commun agenda PF)
- Architecture
- Intégration d’outils (Tool Shed)

http://www.ifb-galaxy.org



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

Une « Galaxy » parmi tant d'autres

Serveur public (https://main.g2.bx.psu.edu/ **):**
- Gratuit & "open source",
- Quota limité, petits jeux de donneés,
- Impossibel d'ajouter des banques, génomes, outils.
- Données non protégées.

Une communauté très active :
- Listes de diffusion (US, FR)
- Wiki
- Twitter
- "Galaxy tour de France"

L'instance locale Sigenae de Galaxy :
- Maintenue par Sigenae.
- Intégration possible de nouveaux outils / scripts / génomes …

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

Galaxy « la bioinformatique pour tous »

Inutile de savoir :

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

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:

- Possibilité de lancer plusieurs jobs en parallèle
- Partir prendre un café ..fermer votre navigateur! puis voir les résultats le lendemain.

Vous pouvez :

- Lancer des traitements depuis votre navigateur,
- Dupliquer des traitements,
- Partager des analyses complètes,

et ceci de manière très intuitive !

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.

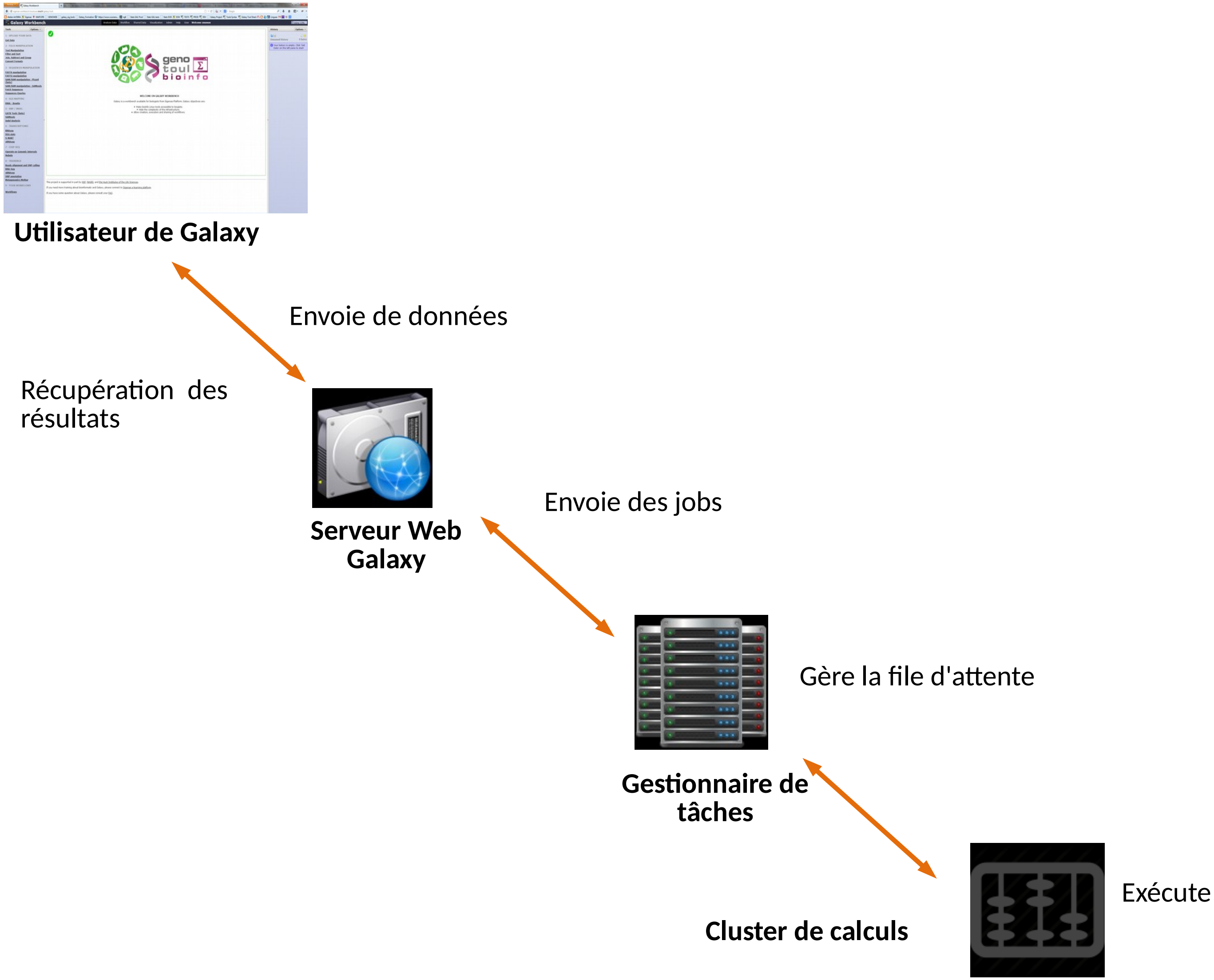
Construction de **chaînes de traitement.**

Intégration de **vos propres outils.**

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

Comment fonctionne Galaxy ?

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



Utilisateur de Galaxy

Envoie de données

Récupération des résultats

Serveur Web Galaxy

Envoie des jobs

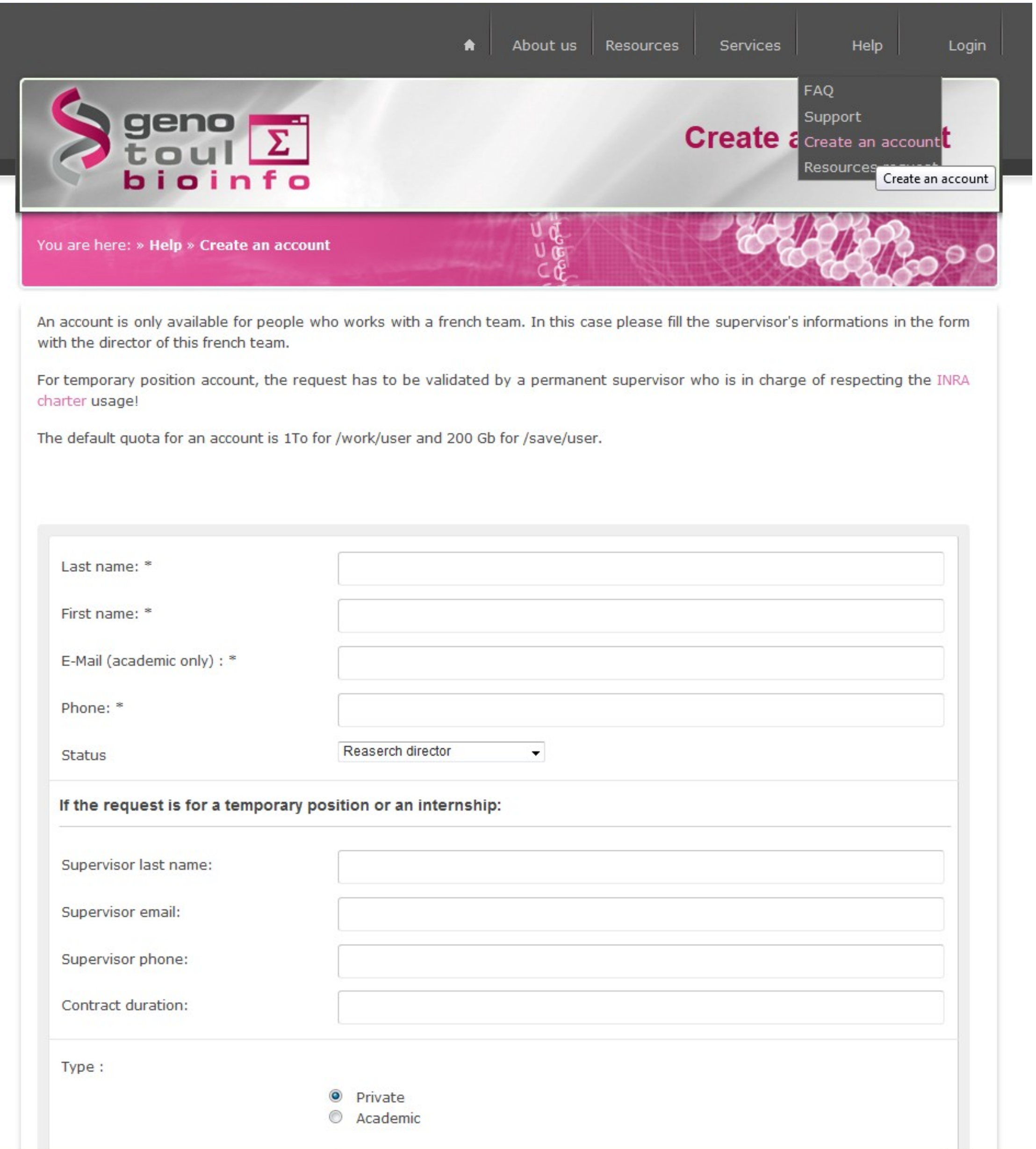
Gestionnaire de tâches

Gère la file d'attente

Cluster de calculs

Exécute

Comment ouvrir un compte sur Genotoul ?



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: Reaserch director

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

Supervisor last name:

Supervisor email:

Supervisor phone:

Contract duration:

Type :

● Private
○ Academic

1 – Ouvrir un compte sur Genotoul :

Formulaire de demande de compte:
http://bioinfo.genotoul.fr
(Menu / Help/ Create an account)

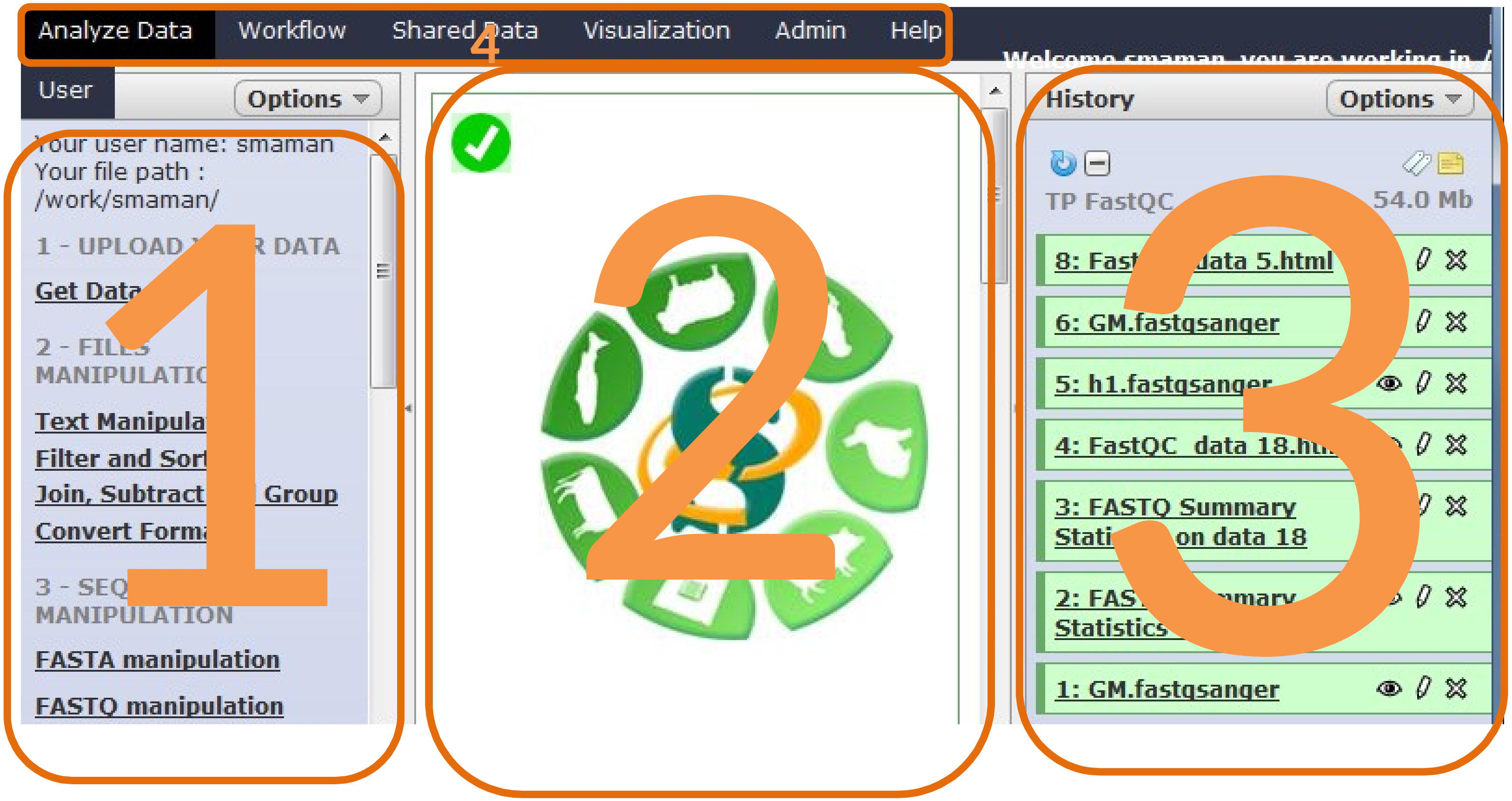
2 – Accéder à Galaxy à l'aide du login/mot de passe obtenus :

http://sigenae-workbench.toulouse.inra.fr/

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 .



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

Votre analyse en quelques clics

Welcome smaman, you are working in /work/smaman

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

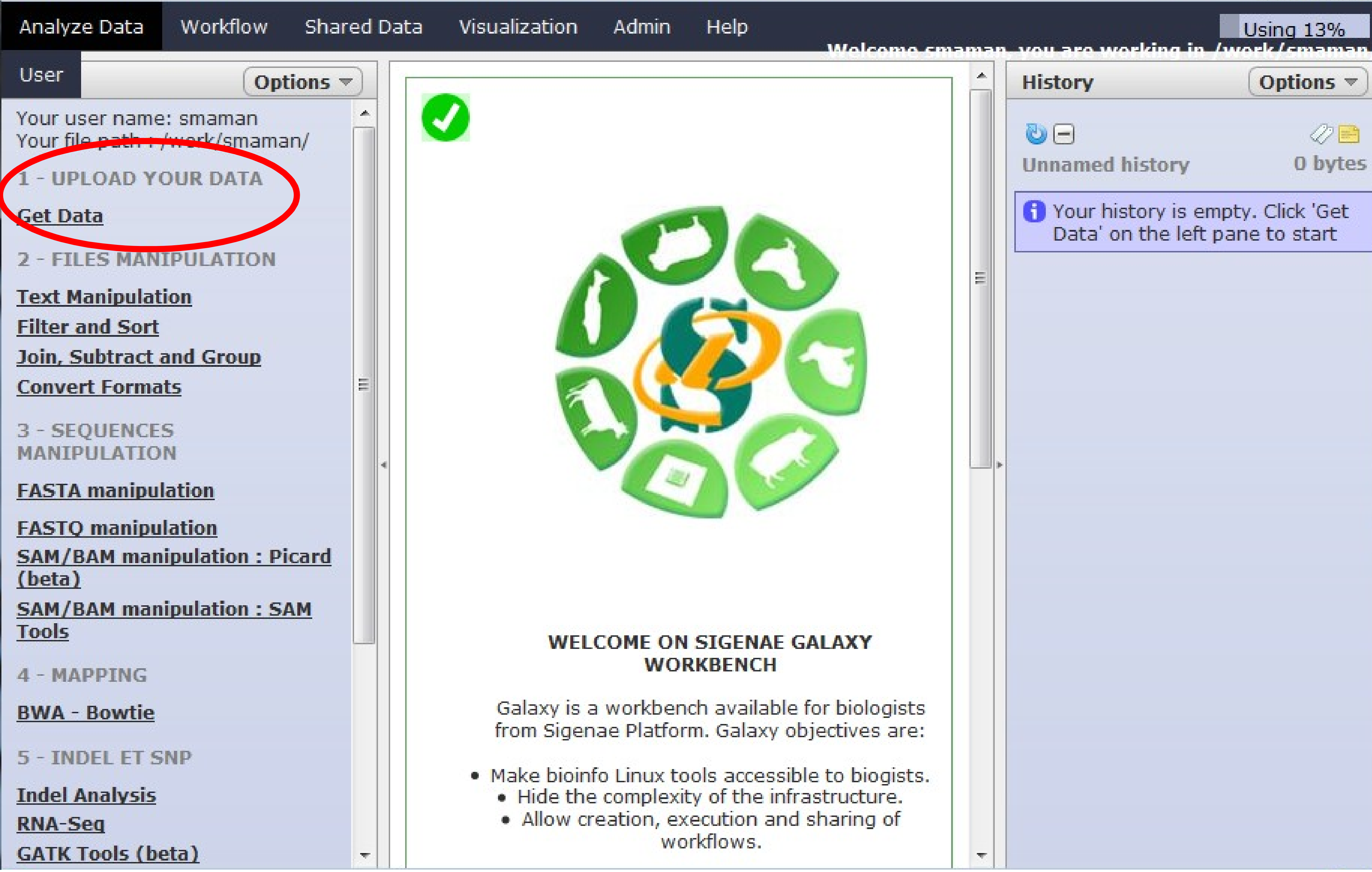
Galaxy is a workbench available for biologists from Sigenae Platform. Galaxy objectives are:

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

Unnamed history 0 bytes

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

Votre analyse en quelques clics



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

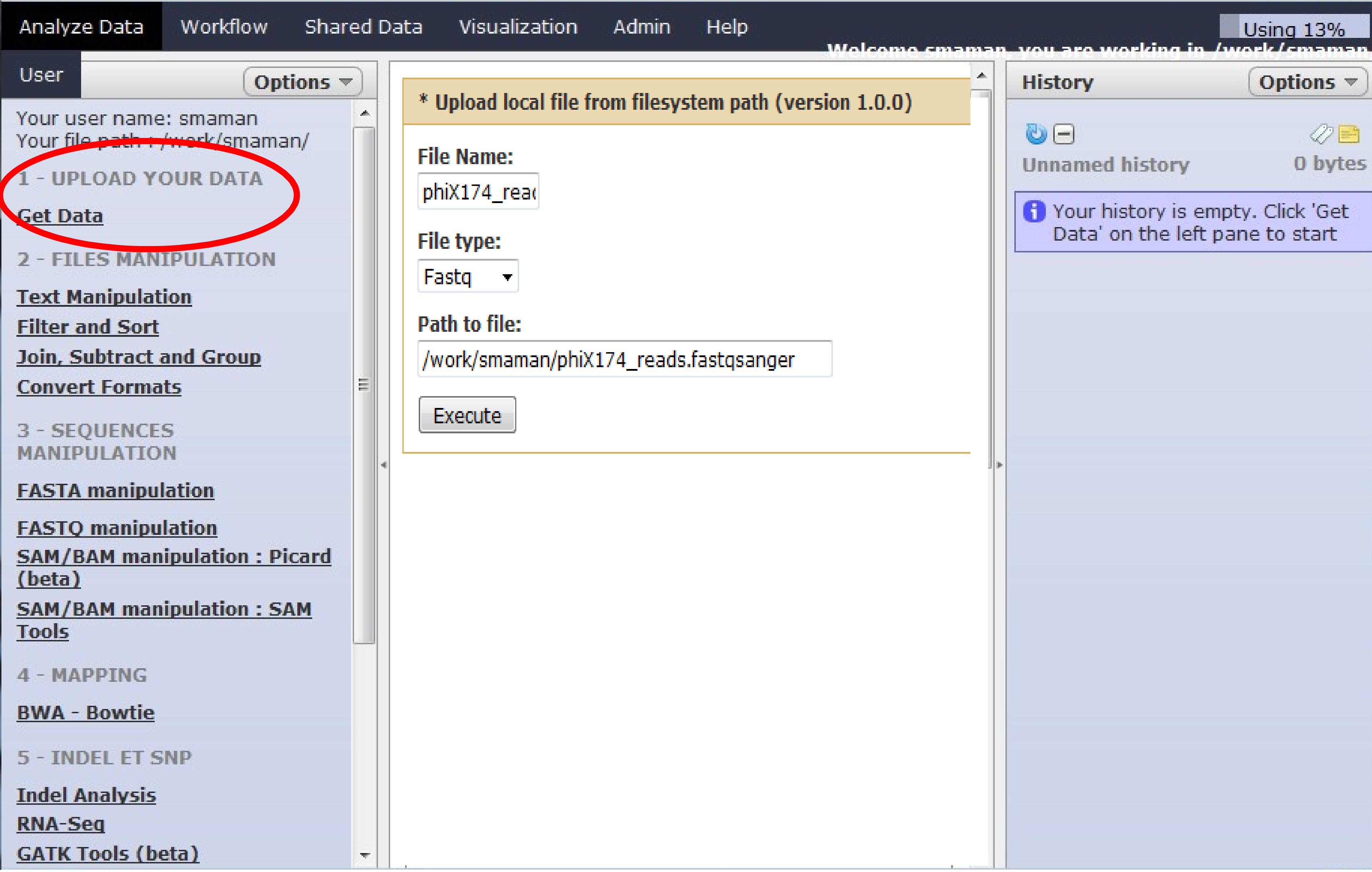
Galaxy is a workbench available for biologists from Sigenae Platform. Galaxy objectives are:

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

Unnamed history
0 bytes

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

Votre analyse en quelques clics



Welcome smaman, you are working in /work/smaman

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_reac

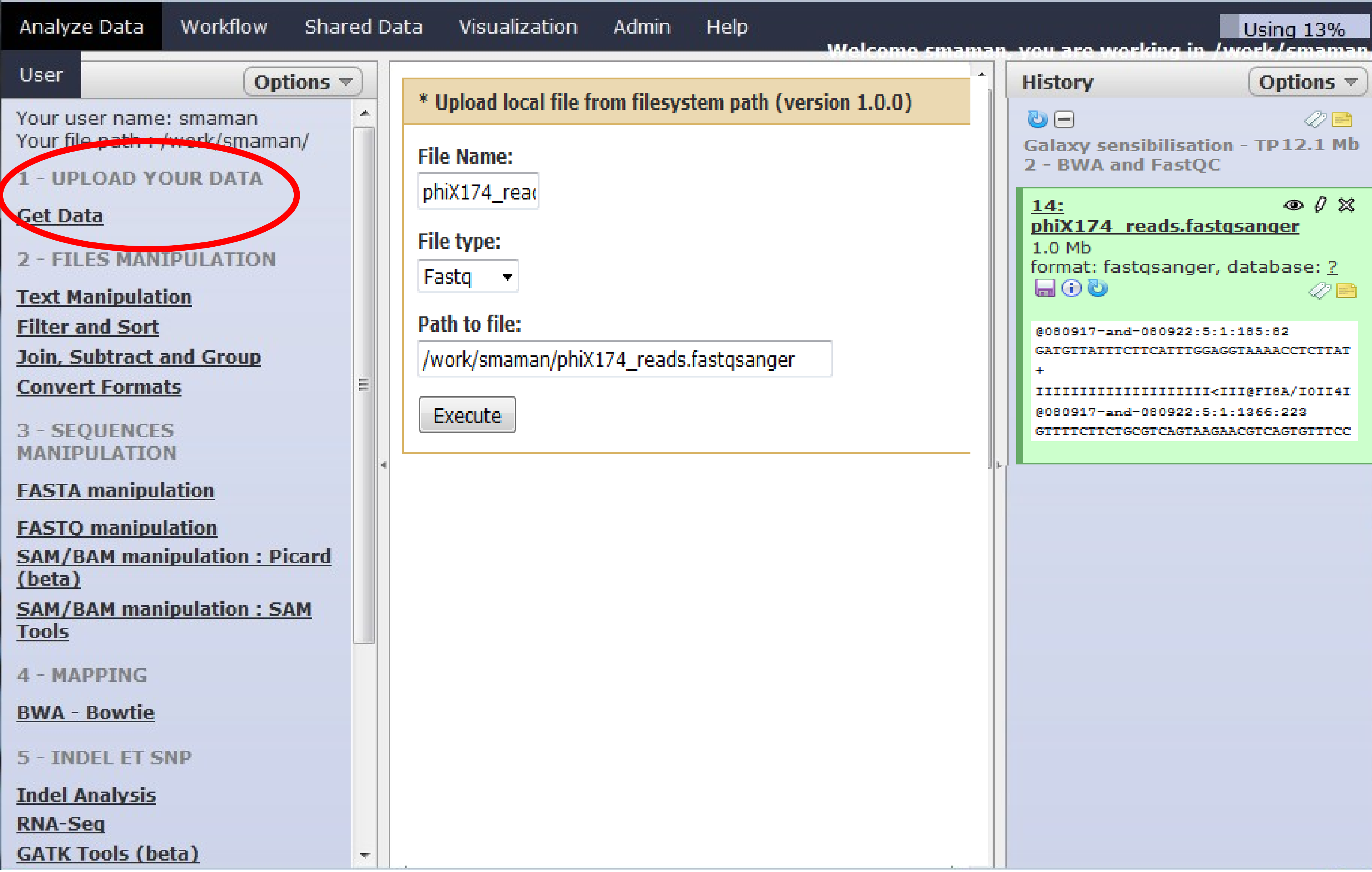
File type: Fastq

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

Unnamed history 0 bytes

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

Votre analyse en quelques clics



Welcome smaman, you are working in /work/smaman

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

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

File Name: phiX174_reac

File type: Fastq

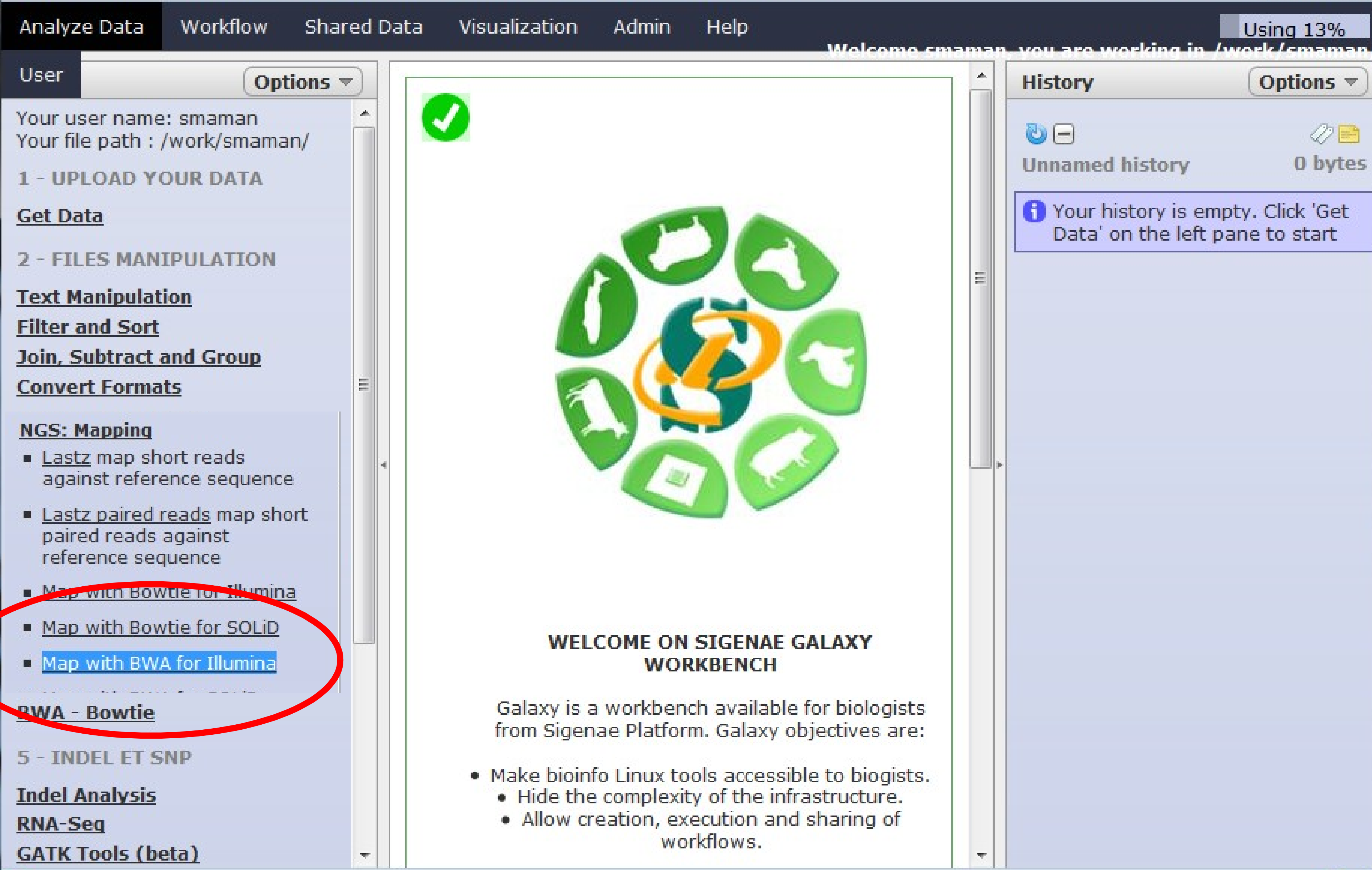
Path to file: /work/smaman/phiX174_reads.fastqsanger

Galaxy sensibilisation - TP 2 - BWA and FastQC 12.1 Mb

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

```
@080917-and-080922:5:1:185:82
GATGTTATTTCTTCATTTGGAGGTAAAACCTCTTAT
+
IIIIIIIIIIIIIIIIIIII<IIII@FI8A/I0II4I
@080917-and-080922:5:1:1366:223
GTTTTCTTCTGCGTCAGTAAGAACGTCAGTGTTTCC
```

Votre analyse en quelques clics



Welcome smaman, you are working in /work/smaman

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

- 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

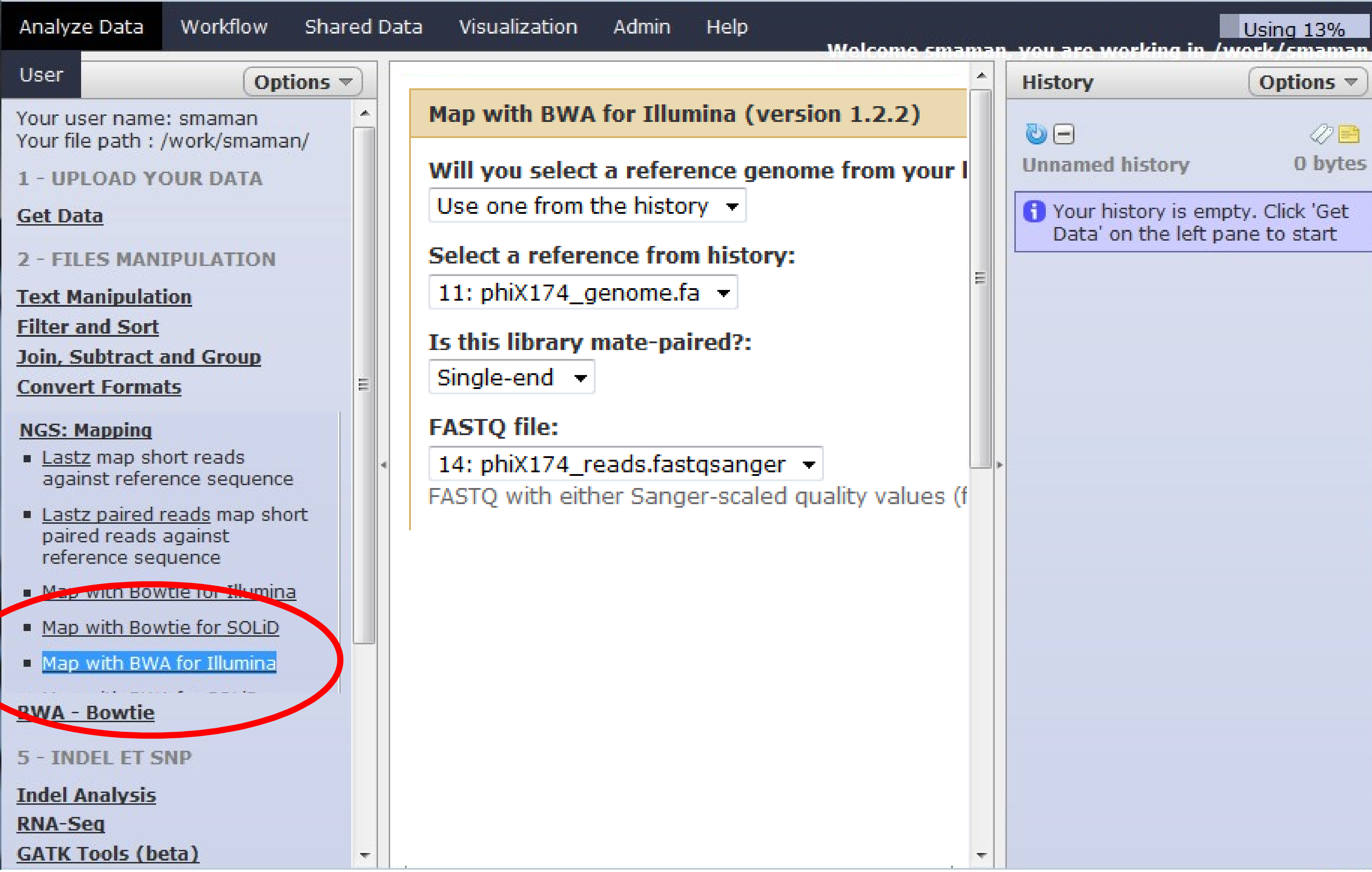
Your history is empty. Click 'Get Data' on the left pane to start

WELCOME ON SIGENAE GALAXY WORKBENCH

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

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

Votre analyse en quelques clics



Welcome smaman, you are working in /work/smaman

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

Map with BWA for Illumina (version 1.2.2)

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

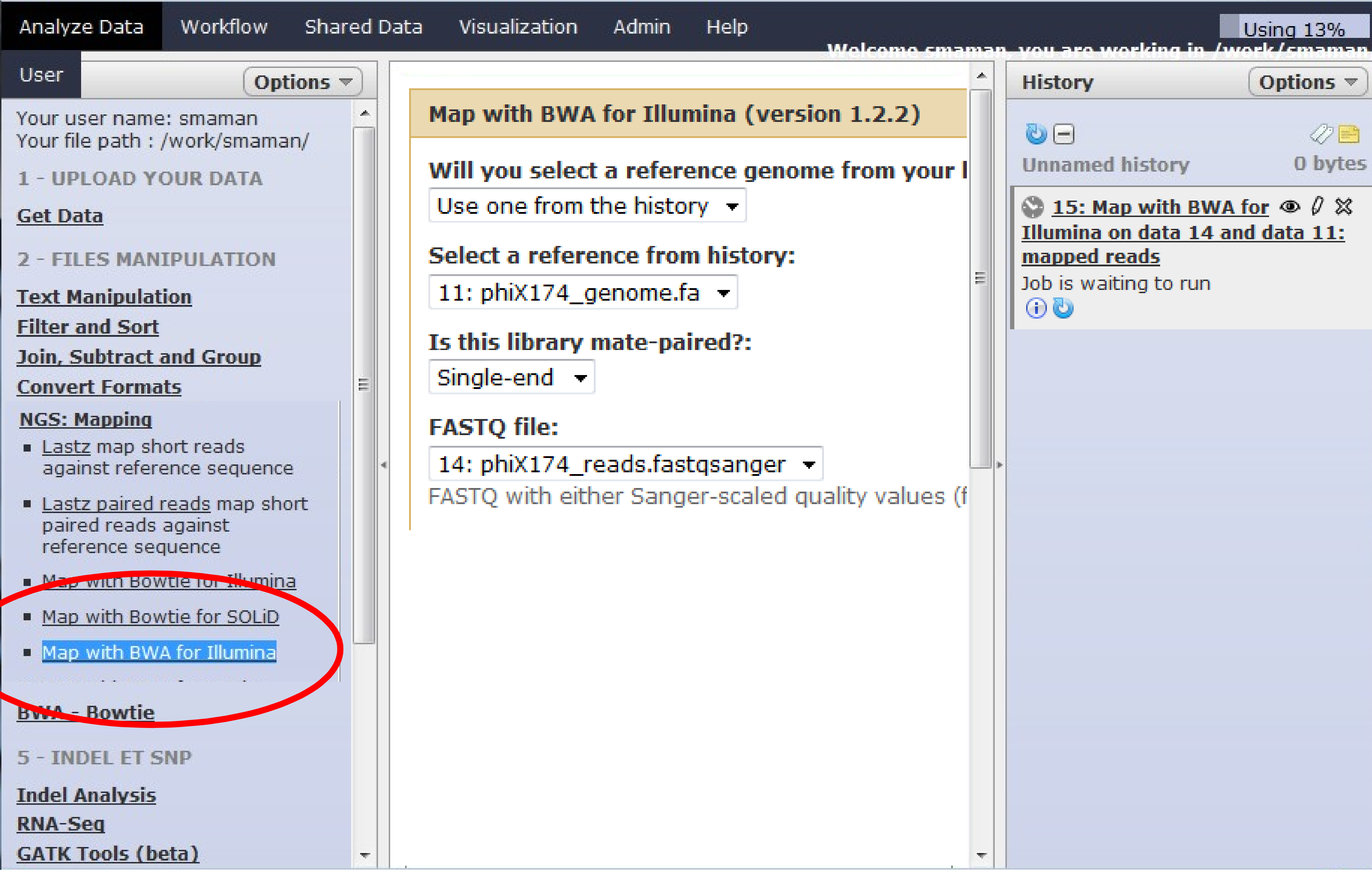
Select a reference from history:
11: phiX174_genome.fa

Is this library mate-paired?:
Single-end

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

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

Votre analyse en quelques clics



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

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

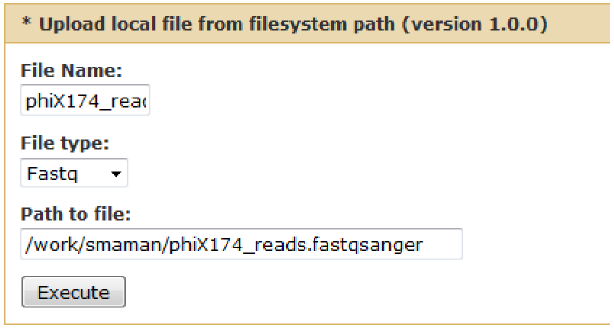
Is this library mate-paired?:
Single-end

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

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

Comment lancer un job sans ligne de commande ?

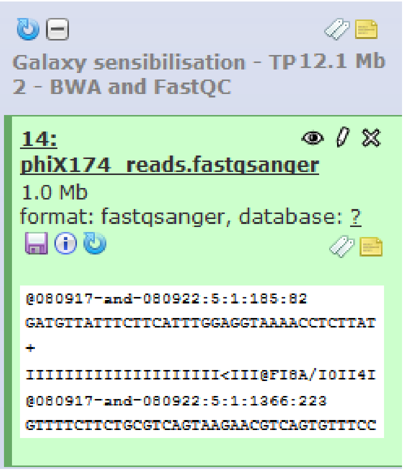
1 – Télécharger vos données :


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

File Name: phiX174_reac

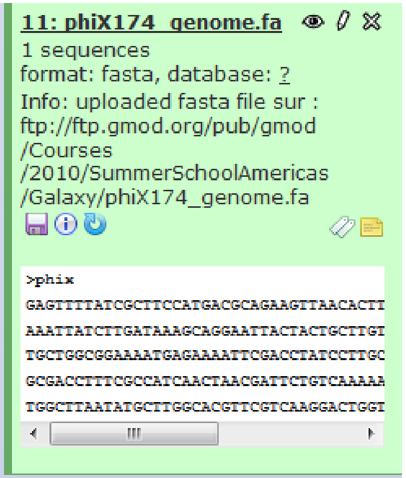
File type: Fastq

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


Galaxy sensibilisation - TP 12.1 Mb
2 - BWA and FastQC

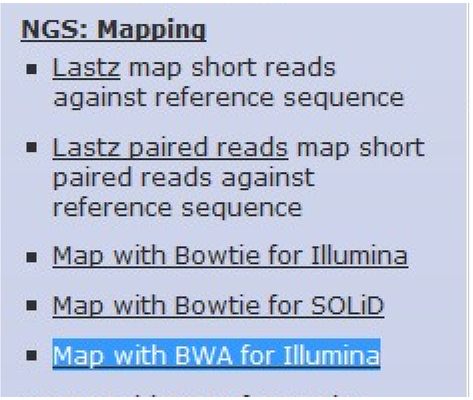
14: phiX174_reads.fastqsanger
1.0 Mb
format: fastqsanger, database: ?

```
@080917-and-080922:5:1:185:82
GATGTTATTTCTTCATTTGGAGGTAAAACCTCTTAT
+
IIIIIIIIIIIIIIIIIII<III@FI8A/I0II4I
@080917-and-080922:5:1:1366:223
GTTTTCTTCTGCGTCAGTAAGAACGTCAGTGTTTCC
```

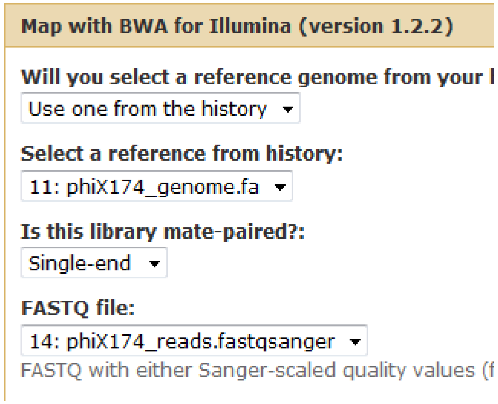

11: phiX174_genome.fa
1 sequences
format: fasta, database: ?
Info: uploaded fasta file sur : ftp://ftp.gmod.org/pub/gmod/Courses/2010/SummerSchoolAmericas/Galaxy/phiX174_genome.fa

```
>phix
GAGTTTTATCGCTTCCATGACGCAGAAGTTAACACTT
AAATTATCTTGATAAAGCAGGAATTACTACTGCTTGT
TGCTGGCGGAAAATGAGAAAATTCGACCTATCCTTGC
GCGACCTTTCGCCATCAACTAACGATTCTGTCAAAAA
TGGCTTAATATGCTTGGCACGTTCGTCAAGGACTGGT
```

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 l Use one from the history

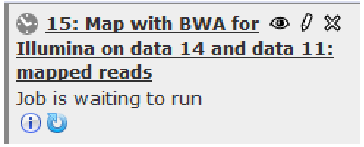
Select a reference from history: 11: phiX174_genome.fa

Is this library mate-paired?: Single-end

FASTQ file: 14: phiX174_reads.fastqsanger

FASTQ with either Sanger-scaled quality values (f

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

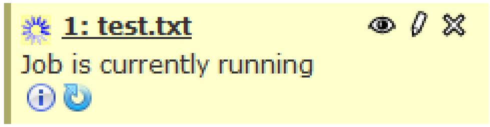
Comment suivre l’execution de mes jobs

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

107: test.txt

Job is waiting to run

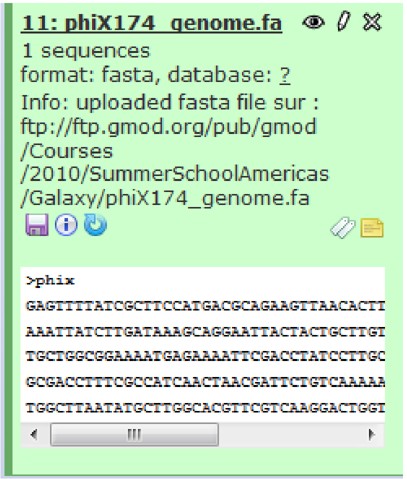
Etat 2 – JAUNE : Votre job est en cours d’execution



1: test.txt

Job is currently running

Etat 3 – VERT : Votre job est fini.

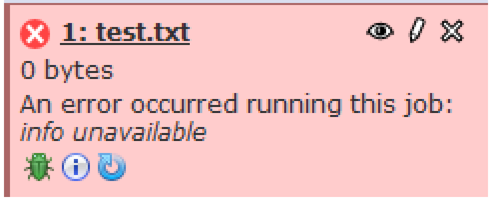


11: phiX174_genome.fa

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

```
>phix
GAGTTTTATCGCTTCCATGACGCAGAAGTTAACACTT
AAATTATCTTGATAAAGCAGGAATTACTACTGCTTGT
TGCTGGCGGAAAATGAGAAAATTCGACCTATCCTTGC
GCGACCTTTCGCCATCAACTAACGATTCTGTCAAAAA
TGGCTTAATATGCTTGGCACGTTCGTCAAGGACTGGT
```

Bug - ROUGE : Votre job est planté !



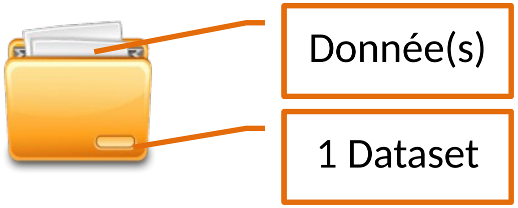
1: test.txt

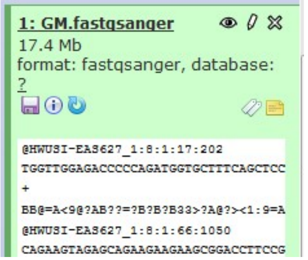
0 bytes

An error occurred running this job: *info unavailable*

Le vocabulaire spécifique à Galaxy

Un **DATASET** est un fichier de données (fichiers d’entrée, fichiers résultats) :

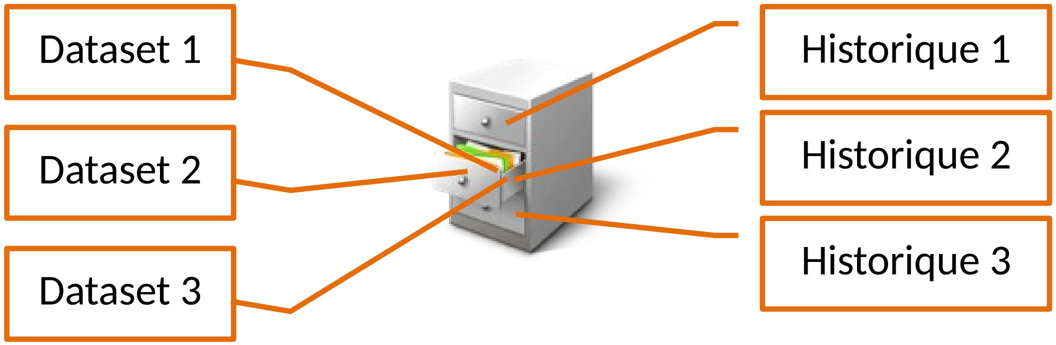


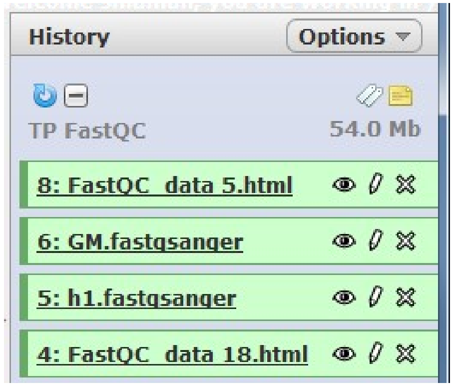


1: GM.fastqsanger
17.4 Mb
format: fastqsanger, database: ?

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

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



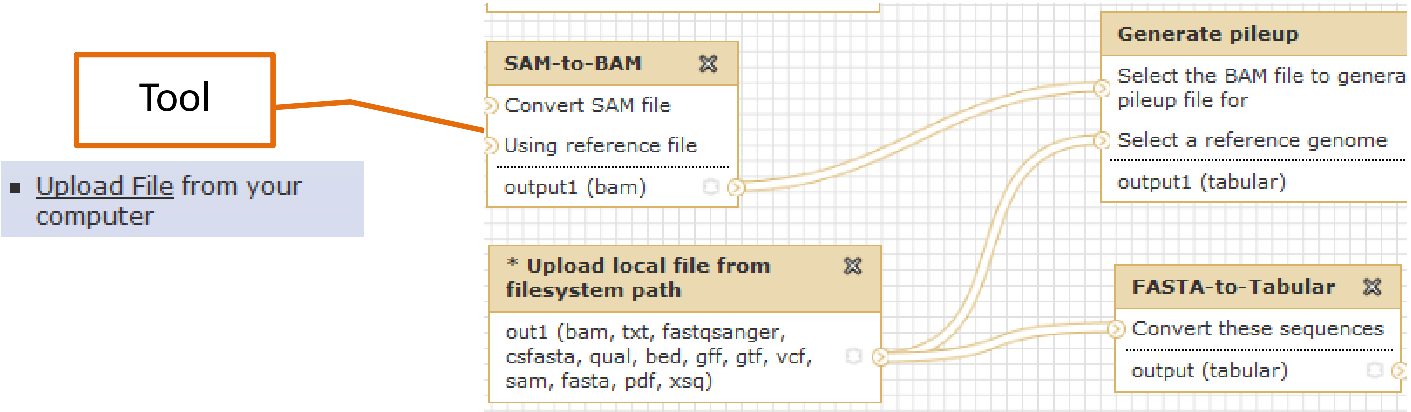


History

TP FastQC 54.0 Mb

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

Votre **WORKFLOW** est un ensemble : fichiers, outils, traitements.



- Upload File from your computer

SAM-to-BAM
Convert SAM file
Using reference file
output1 (bam)

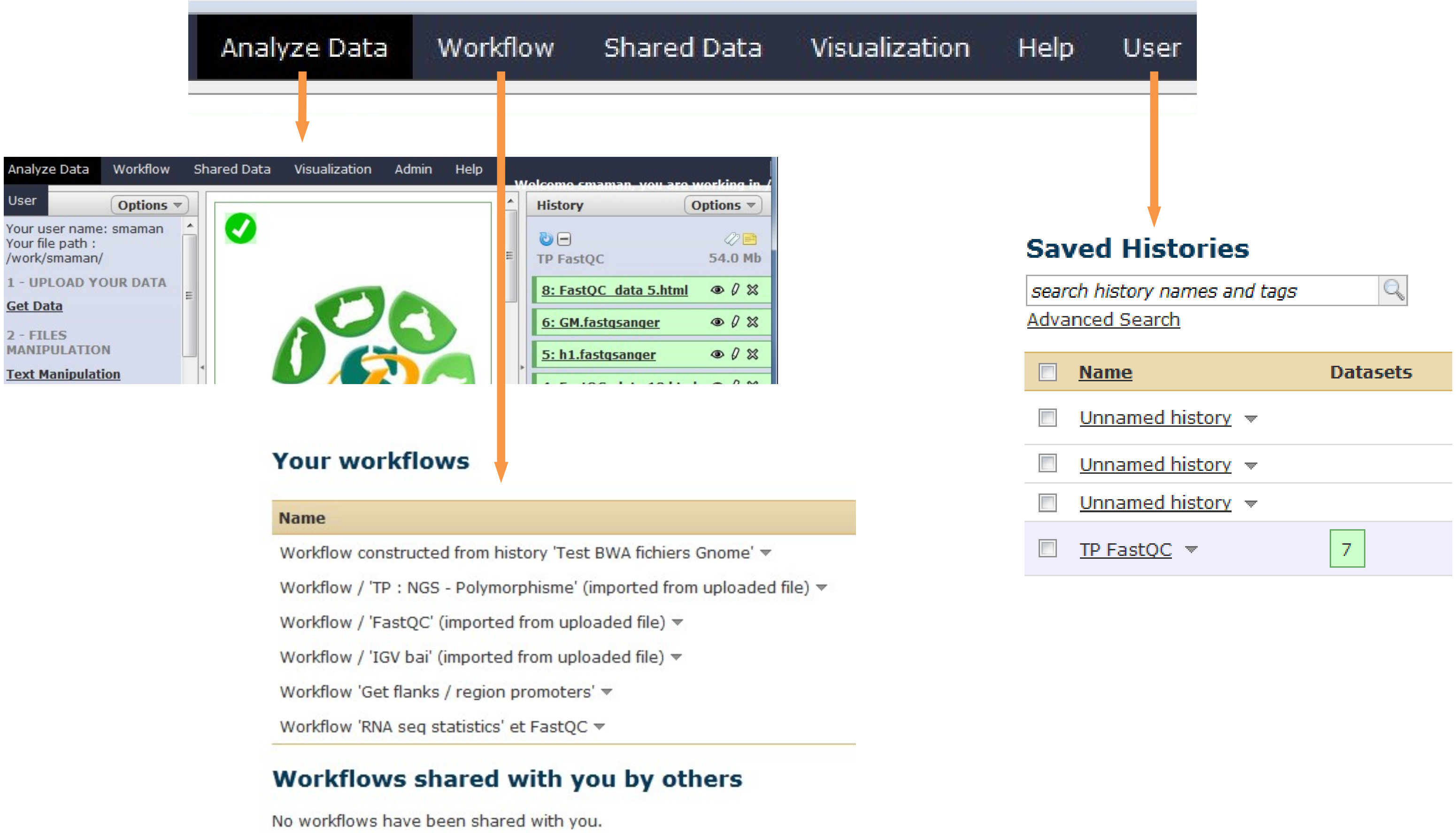
* Upload local file from filesystem path
out1 (bam, txt, fastqsanger, csfasta, qual, bed, gff, gtf, vcf, sam, fasta, pdf, xsq)

Generate pileup
Select the BAM file to genera pileup file for
Select a reference genome
output1 (tabular)

FASTA-to-Tabular
Convert these sequences
output (tabular)

Principaux onglets

- **ANALYSE DATA** : Page d'accueil de Galaxy.
- **WORFLOW** : 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.



Welcome smaman, you are working in /

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

1 - UPLOAD YOUR DATA

Get Data

2 - FILES MANIPULATION

Text Manipulation

TP FastQC 54.0 Mb

8: FastQC_data 5.html

6: GM.fastqsanger

5: h1.fastqsanger

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

☐	Name	Datasets
☐	Unnamed history	
☐	Unnamed history	
☐	Unnamed history	
☐	TP FastQC	7

Présentation de la plateforme Galaxy.

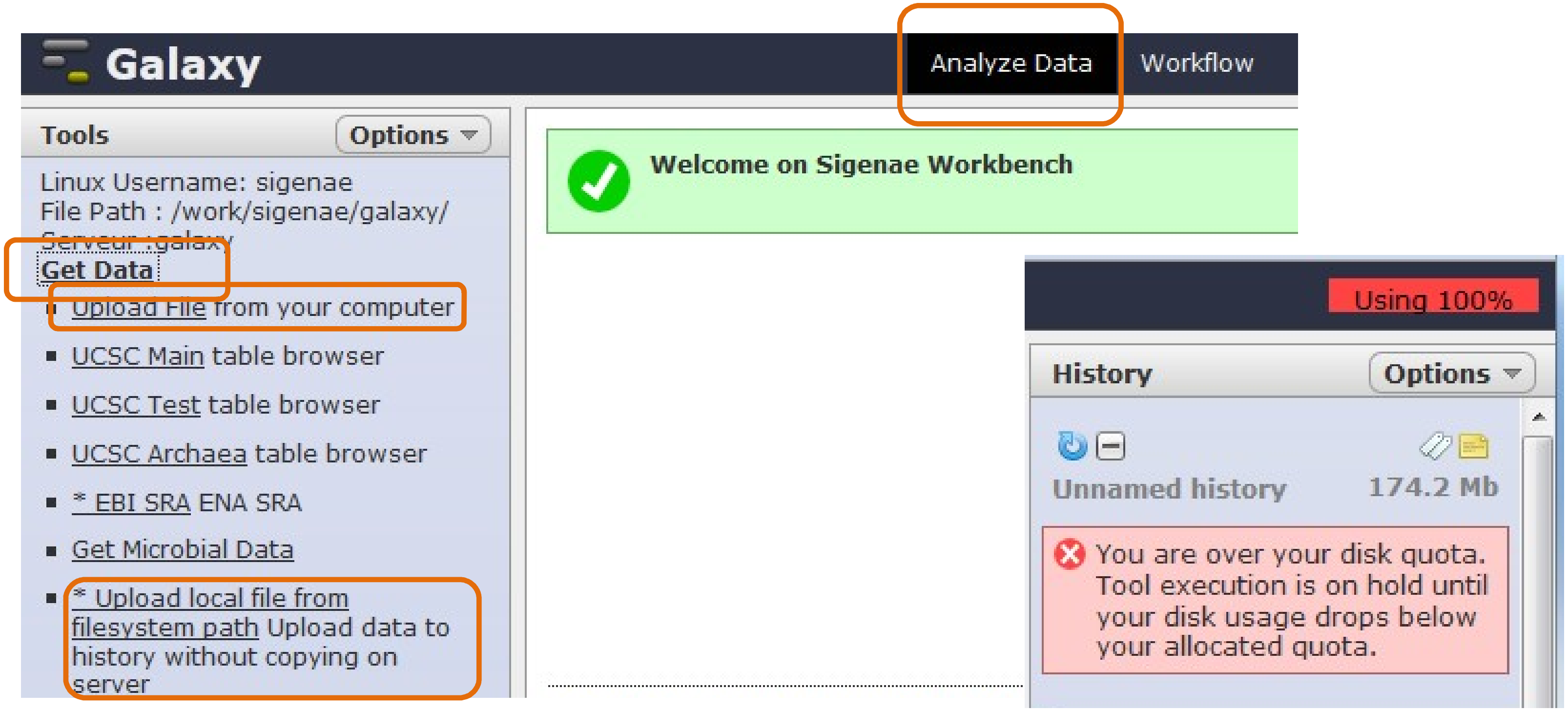
Premiers pas dans l’instance.

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

Lancement de traitements bioinformatiques.

Auto-formations disponibles en ligne.

Deux méthodes de téléchargement de vos données privées



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

Welcome on Sigenae Workbench

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 sur Genotoul) :
« Upload local file from filesystem path ».

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

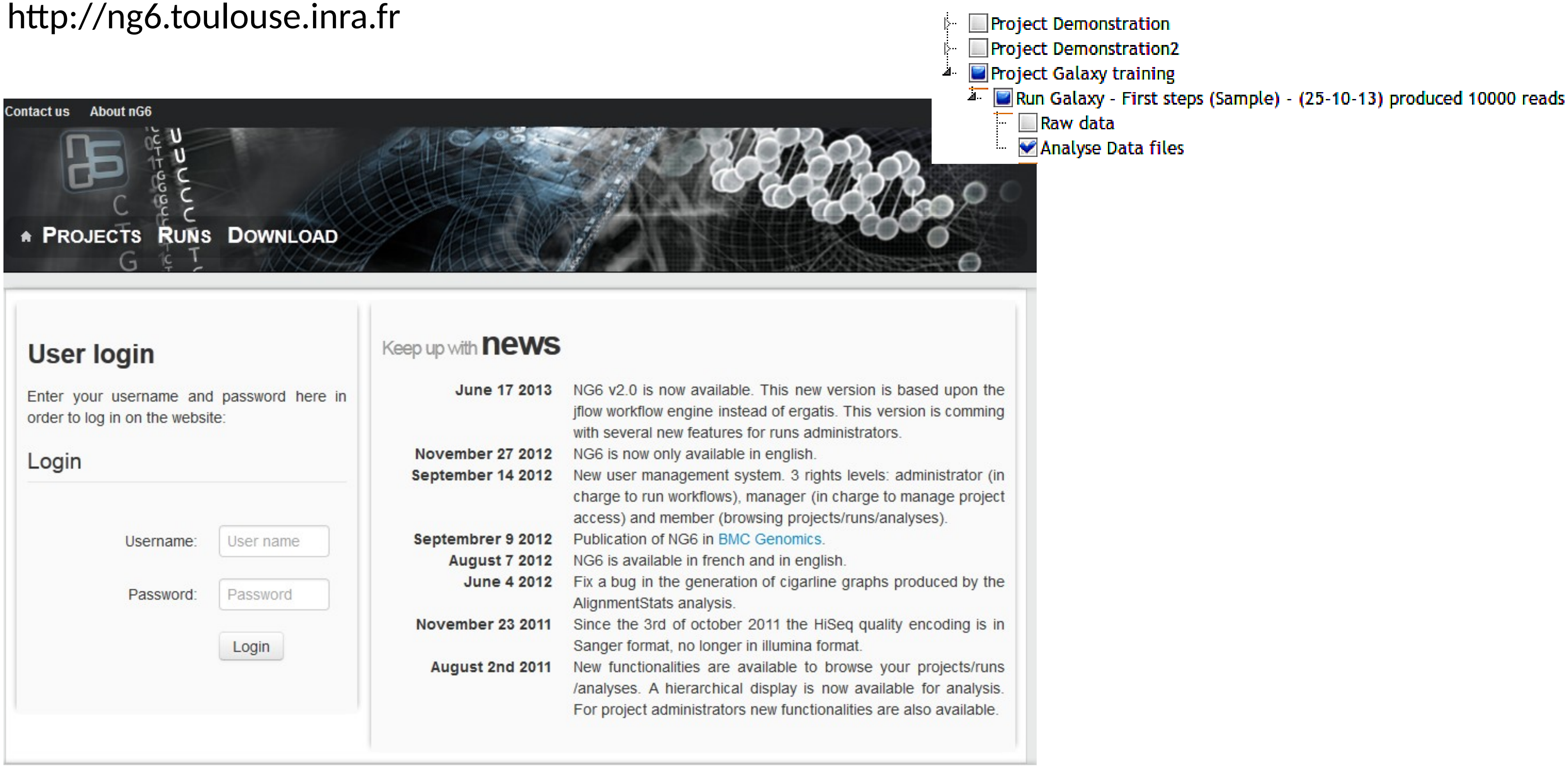
L’espace disque utilisé n’est pas celui de votre PC.
Vos traitements ne sont pas limités par la capacité de votre PC.

Présentation de NG6

NG6 (Next Generation Sequencing Information System) :

- Environnement de stockage et de mise à disposition des données issues des nouvelles technologies de séquençage.
- Organisé autour des notions de projet et de run, permet un accès sécurisé aux données brutes, aux statistiques de traitements ainsi qu'aux assemblages et annotations produites.

http://ng6.toulouse.inra.fr



☐ Project Demonstration
☐ Project Demonstration2
☐ Project Galaxy training
  ☐ Run Galaxy - First steps (Sample) - (25-10-13) produced 10000 reads
    ☐ Raw data
    ☑ Analyse Data files

User login

Enter your username and password here in order to log in on the website:

Login

Username: User name

Password: Password

Keep up with **news**

June 17 2013	NG6 v2.0 is now available. This new version is based upon the jflow workflow engine instead of ergatis. This version is comming with several new features for runs administrators.
November 27 2012	NG6 is now only available in english.
September 14 2012	New user management system. 3 rights levels: administrator (in charge to run workflows), manager (in charge to manage project access) and member (browsing projects/runs/analyses).
Septembrer 9 2012	Publication of NG6 in BMC Genomics.
August 7 2012	NG6 is available in french and in english.
June 4 2012	Fix a bug in the generation of cigarline graphs produced by the AlignmentStats analysis.
November 23 2011	Since the 3rd of october 2011 the HiSeq quality encoding is in Sanger format, no longer in illumina format.
August 2nd 2011	New functionalities are available to browse your projects/runs /analyses. A hierarchical display is now available for analysis. For project administrators new functionalities are also available.

Téléchargement de données publiques

Données UCSC, Ensembl, BIOMART :

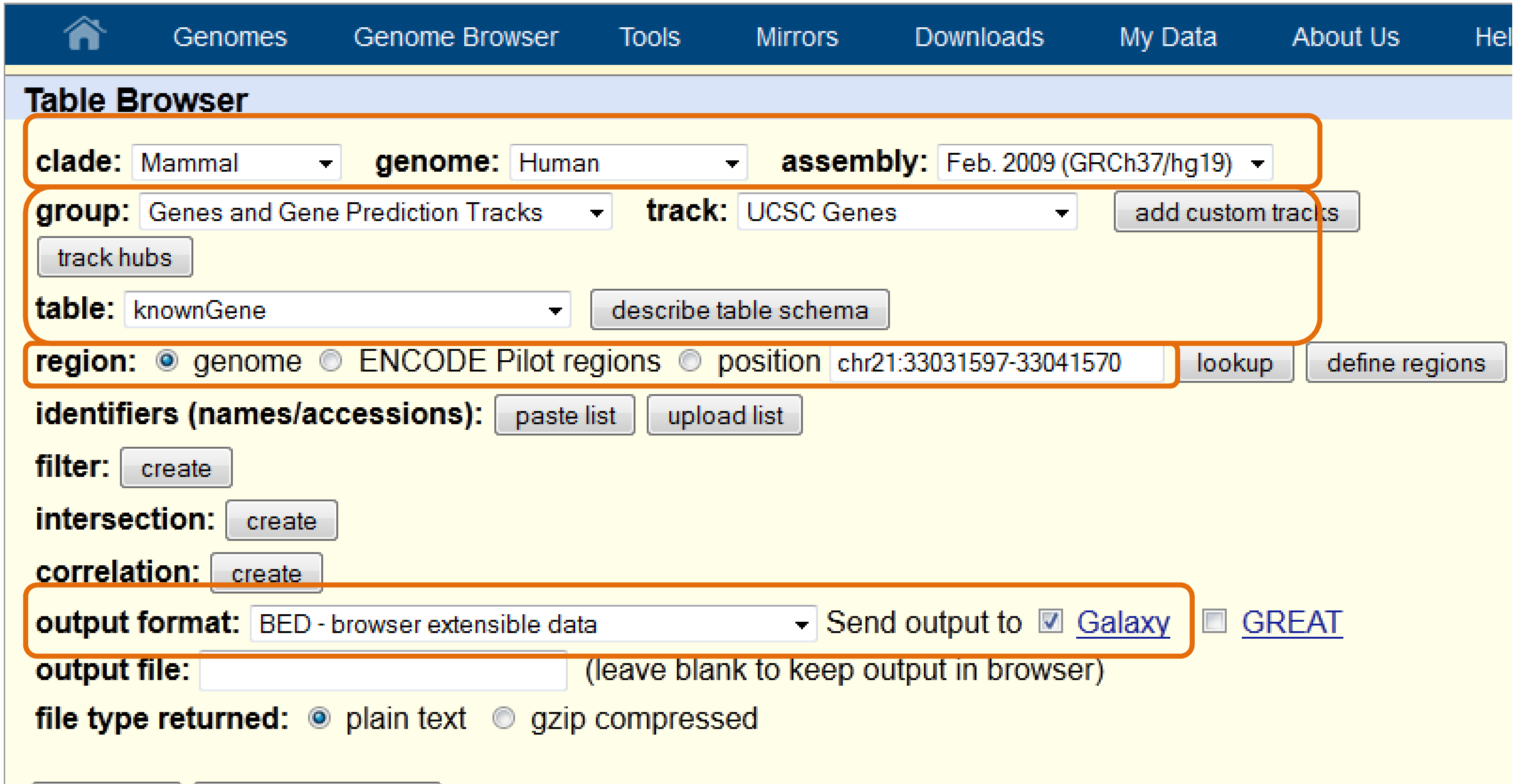


Table Browser

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]

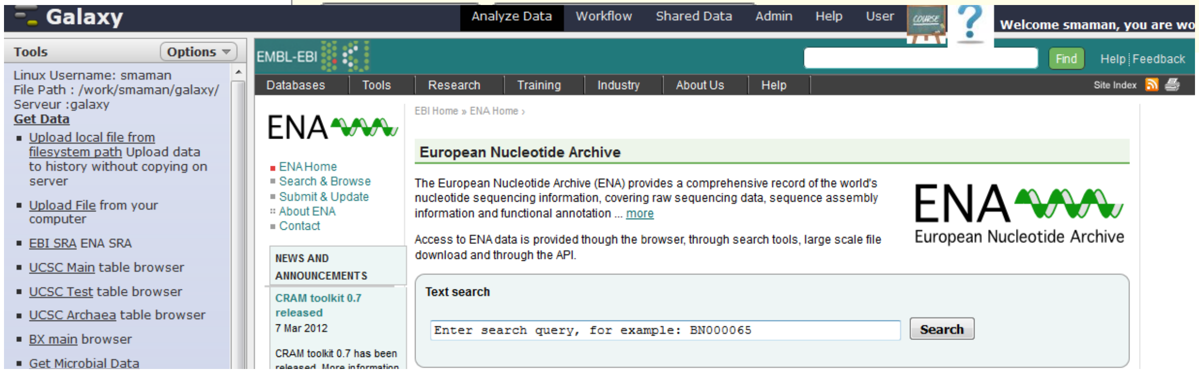
intersection: [create]

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



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

EBI Home » ENA Home ›

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 though 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

Text search

Enter search query, for example: BN000065

Exercice 1 :

Connexion à Galaxy, exploration de l’interface, téléchargement de datasets

Présentation de la plateforme Galaxy.

Premiers pas dans l’instance.

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

Lancement de traitements bioinformatiques.

Auto-formations disponibles en ligne.

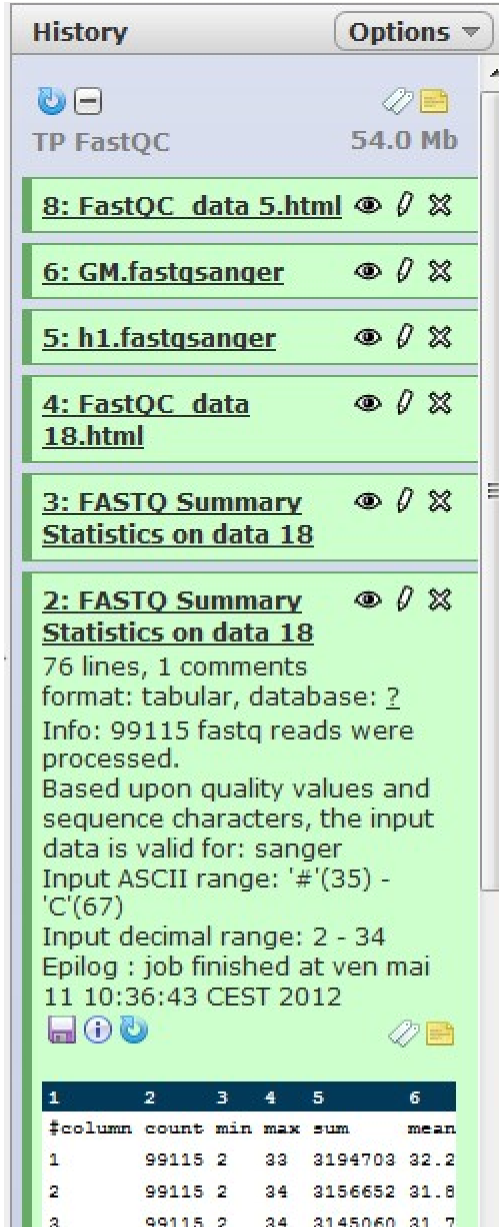
Gestion de vos historiques

Conserver toutes les étapes de vos analyses.

Partager vos analyses.

A chaque (re)run d’un outil, un nouveau dataset est créé. Les données ne sont pas écrasées.

Répéter, autant de fois que nécessaire, une analyse en modifiant vos paramètres pour explorer les différences de résultats.



TP FastQC 54.0 Mb

8: FastQC_data 5.html

6: GM.fastqsanger

5: h1.fastqsanger

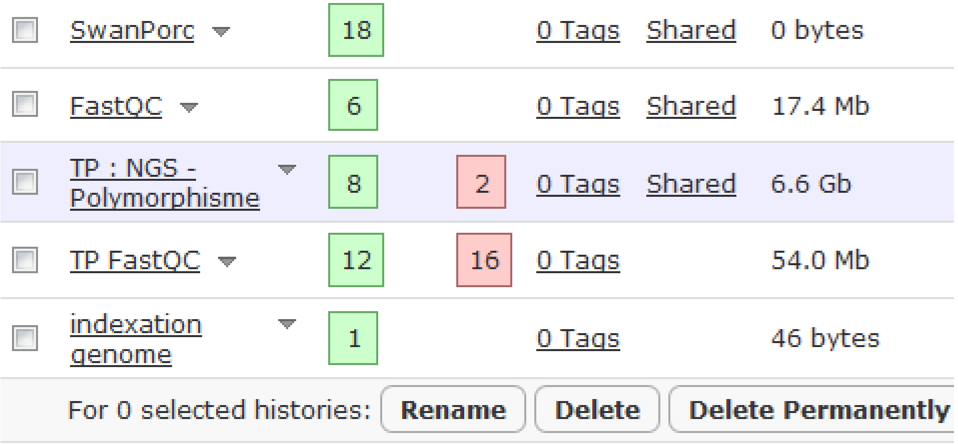
4: FastQC_data 18.html

3: FASTQ Summary Statistics on data 18

2: FASTQ Summary Statistics on data 18

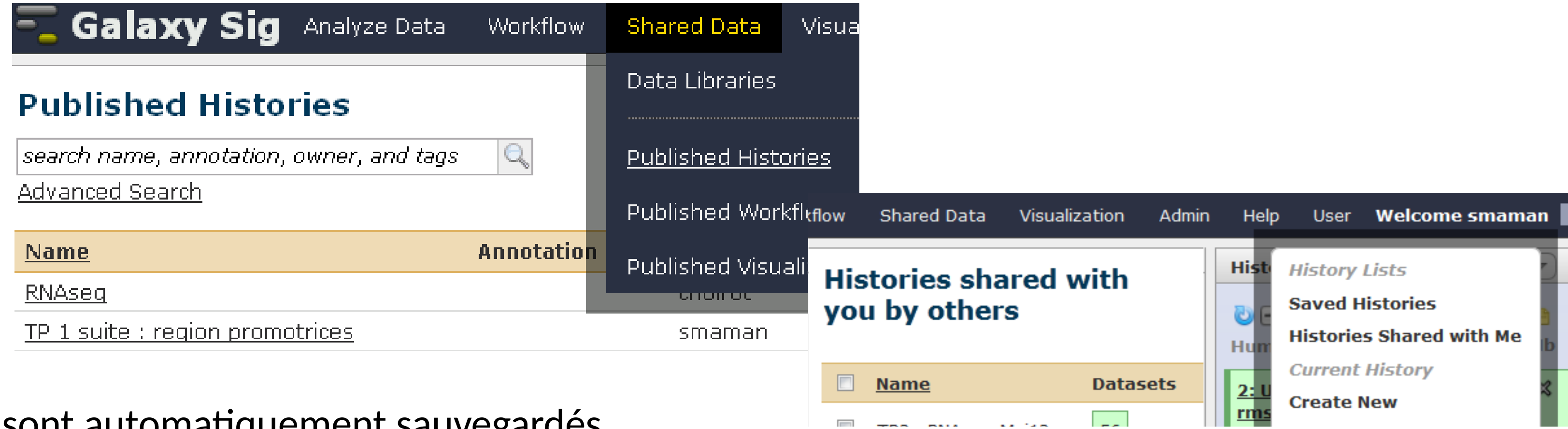
76 lines, 1 comments
format: tabular, database: ?
Info: 99115 fastq reads were processed.
Based upon quality values and sequence characters, the input data is valid for: sanger
Input ASCII range: '#'(35) - 'C'(67)
Input decimal range: 2 - 34
Epilog : job finished at ven mai 11 10:36:43 CEST 2012

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

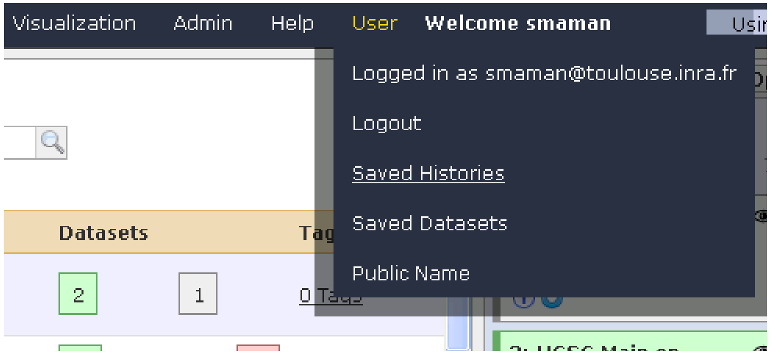


☐	SwanPorc	18		0 Tags	Shared	0 bytes
☐	FastQC	6		0 Tags	Shared	17.4 Mb
☐	TP : NGS - Polymorphisme	8	2	0 Tags	Shared	6.6 Gb
☐	TP FastQC	12	16	0 Tags		54.0 Mb
☐	indexation genome	1		0 Tags		46 bytes

Historiques sauvegardés, publiés et partagés

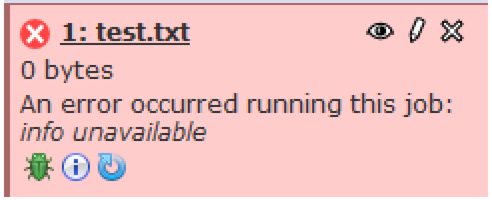


Vos historiques sont automatiquement sauvegardés.
Pour partager ou publier un historique :
User / Saved histories / Cliquer sur le nom de l'historique / Share ou Publish



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 (Hitory / Option / Histories Shared With Me).

Mail à sigeane-support



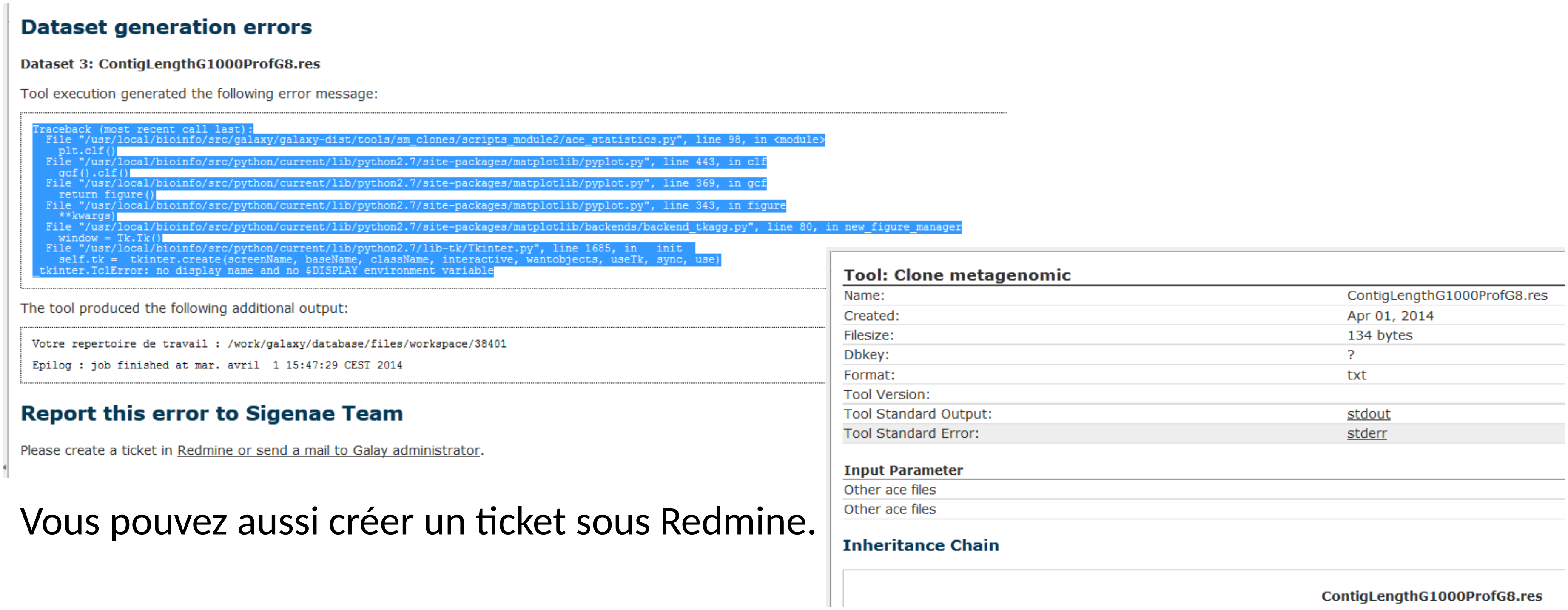
1: test.txt

0 bytes

An error occurred running this job: *info unavailable*

Bug - ROUGE : Votre job est planté !

Voici les informations à transmettre par mail à sigenae-support@listes.inra.fr :
Le contenu (copier/coller) du bug ou/et (i)/stderr
Un share de votre historique.



Dataset generation errors

Dataset 3: ContigLengthG1000ProfG8.res

Tool execution generated the following error message:

```
Traceback (most recent call last):
  File "/usr/local/bioinfo/src/galaxy/galaxy-dist/tools/sm_clones/scripts_module2/ace_statistics.py", line 98, in <module>
    plt.clf()
  File "/usr/local/bioinfo/src/python/current/lib/python2.7/site-packages/matplotlib/pyplot.py", line 443, in clf
    gcf().clf()
  File "/usr/local/bioinfo/src/python/current/lib/python2.7/site-packages/matplotlib/pyplot.py", line 369, in gcf
    return figure()
  File "/usr/local/bioinfo/src/python/current/lib/python2.7/site-packages/matplotlib/pyplot.py", line 343, in figure
    **kwargs)
  File "/usr/local/bioinfo/src/python/current/lib/python2.7/site-packages/matplotlib/backends/backend_tkagg.py", line 80, in new_figure_manager
    window = Tk.Tk()
  File "/usr/local/bioinfo/src/python/current/lib/python2.7/lib-tk/Tkinter.py", line 1685, in __init__
    self.tk =  tkinter.create(screenName, baseName, className, interactive, wantobjects, useTk, sync, use)
_tkinter.TclError: no display name and no $DISPLAY environment variable
```

The tool produced the following additional output:

```
Votre repertoire de travail : /work/galaxy/database/files/workspace/38401
Epilog : job finished at mar. avril  1 15:47:29 CEST 2014
```

Report this error to Sigenae Team

Please create a ticket in Redmine or send a mail to Galay administrator.

Tool: Clone metagenomic

Name:	ContigLengthG1000ProfG8.res
Created:	Apr 01, 2014
Filesize:	134 bytes
Dbkey:	?
Format:	txt
Tool Version:	
Tool Standard Output:	stdout
Tool Standard Error:	stderr

Input Parameter

Other ace files

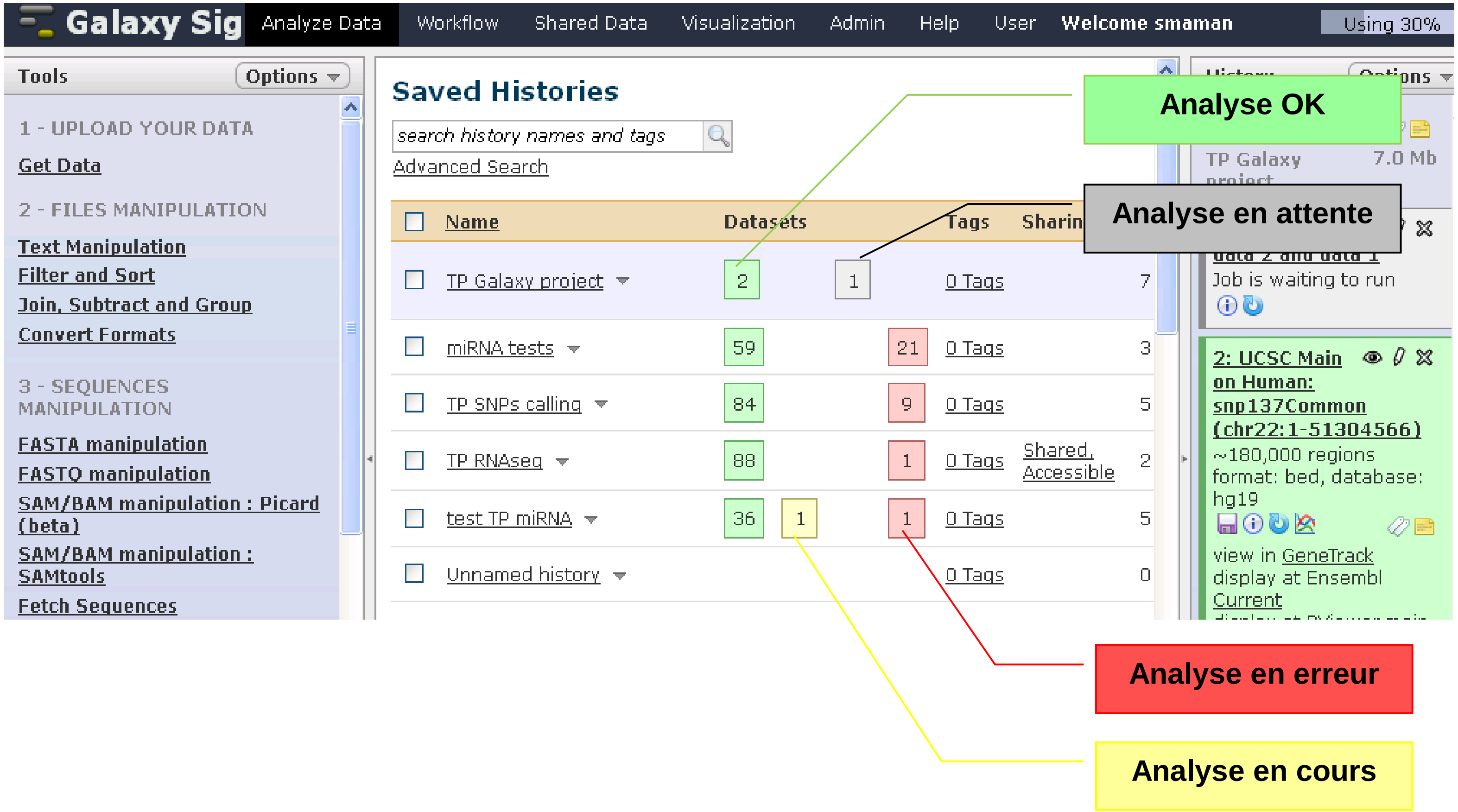
Other ace files

Inheritance Chain

ContigLengthG1000ProfG8.res

Vous pouvez aussi créer un ticket sous Redmine.

Liste des historiques et exécution de mes jobs



Saved Histories

search history names and tags

Advanced Search

☐	Name	Datasets			Tags	Sharin	
☐	TP Galaxy project	2		1	0 Tags		7
☐	miRNA tests	59		21	0 Tags		3
☐	TP SNPs calling	84		9	0 Tags		5
☐	TP RNAseq	88		1	0 Tags	Shared, Accessible	2
☐	test TP miRNA	36	1	1	0 Tags		5
☐	Unnamed history				0 Tags		0

Analyse OK

Analyse en attente

Analyse en erreur

Analyse en cours

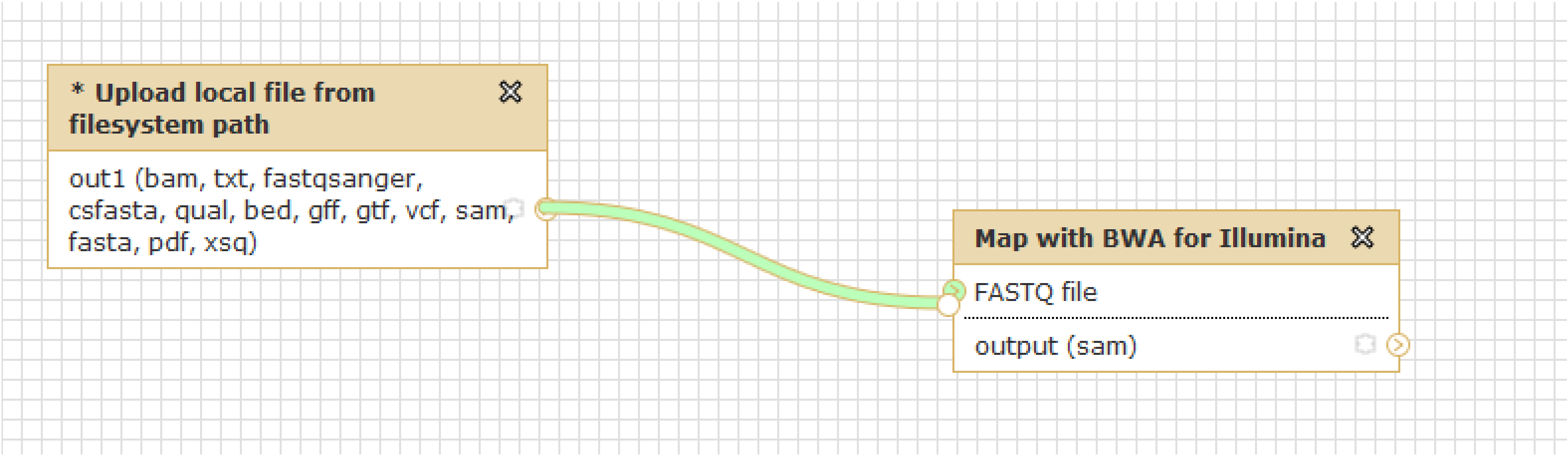
Rapidement, beaucoup de données sont générées

→ D'où l'importance de bien renommer ses historiques / datasets / workflows pour les trier et les organiser au mieux.

Créer un workflow

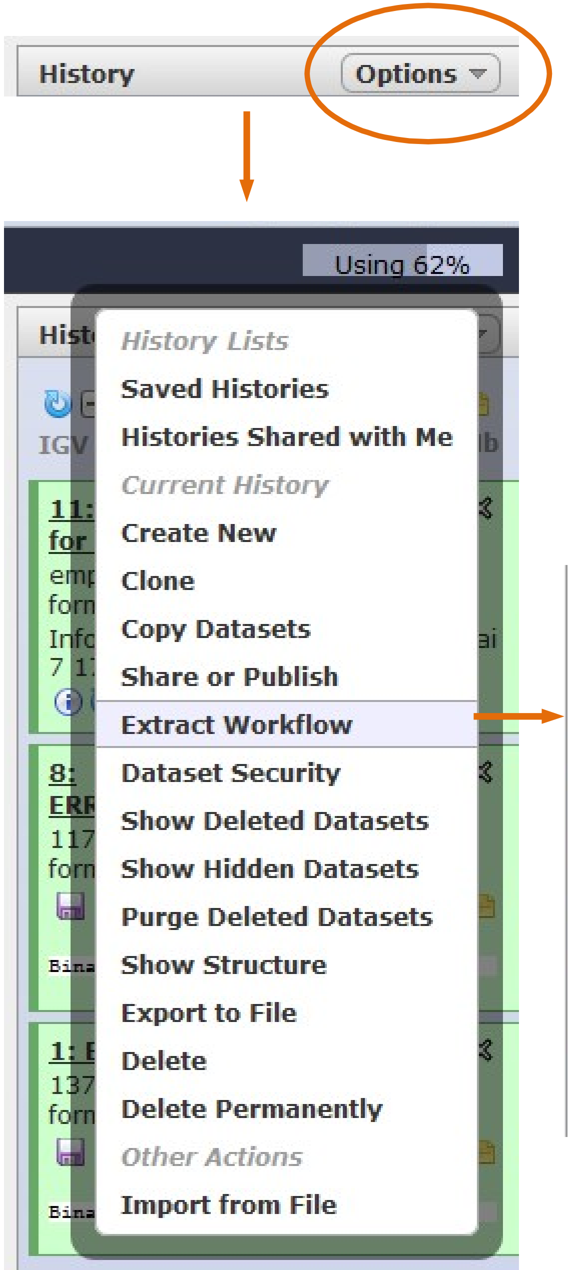
Depuis une page blanche, vous pouvez concevoir un workflow.

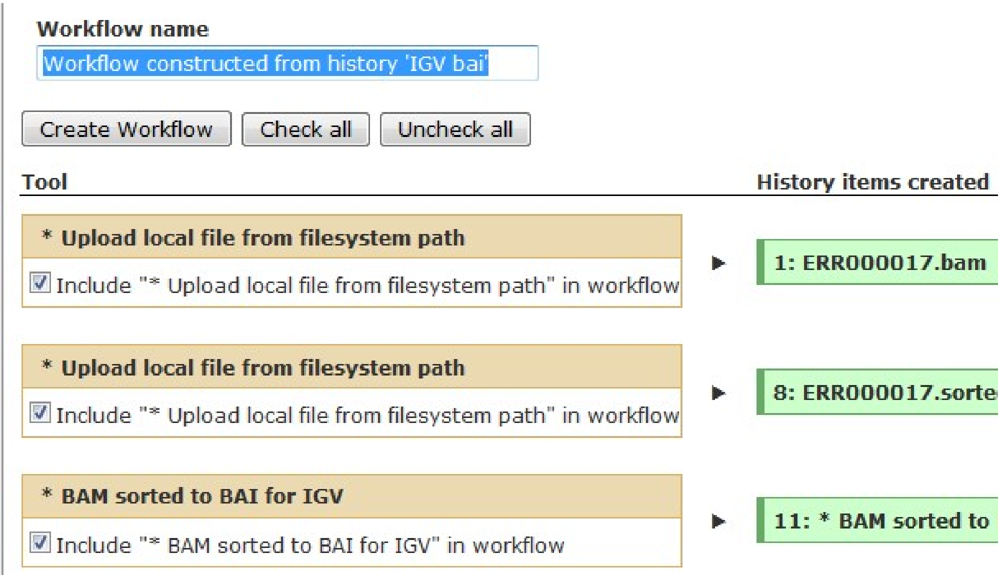
Aide : les résultats produits sont typés, il n’est donc pas possible de brancher un dataset sur un mauvais tool !



Exporter votre historique en workflow.

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



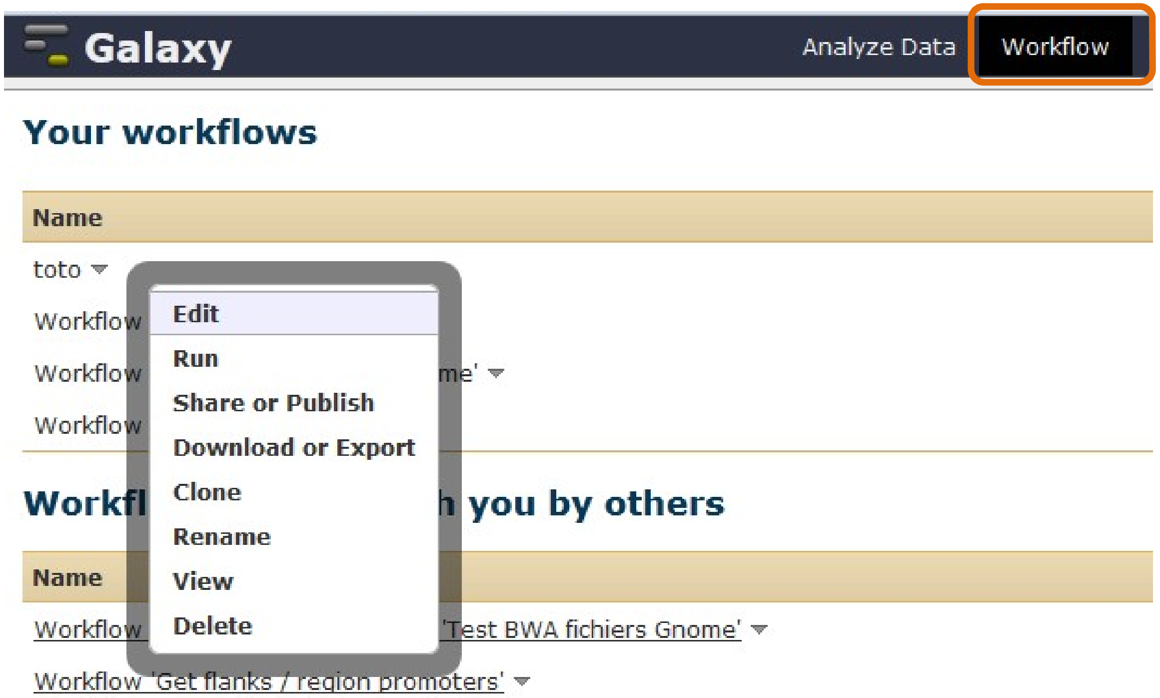


Workflow name
Workflow constructed from history 'IGV bai'

Tool	History items created
*** Upload local file from filesystem path** ☑ Include "* Upload local file from filesystem path" in workflow	1: ERR000017.bam
*** Upload local file from filesystem path** ☑ Include "* Upload local file from filesystem path" in workflow	8: ERR000017.sorte
*** BAM sorted to BAI for IGV** ☑ Include "* BAM sorted to BAI for IGV" in workflow	11: * BAM sorted to

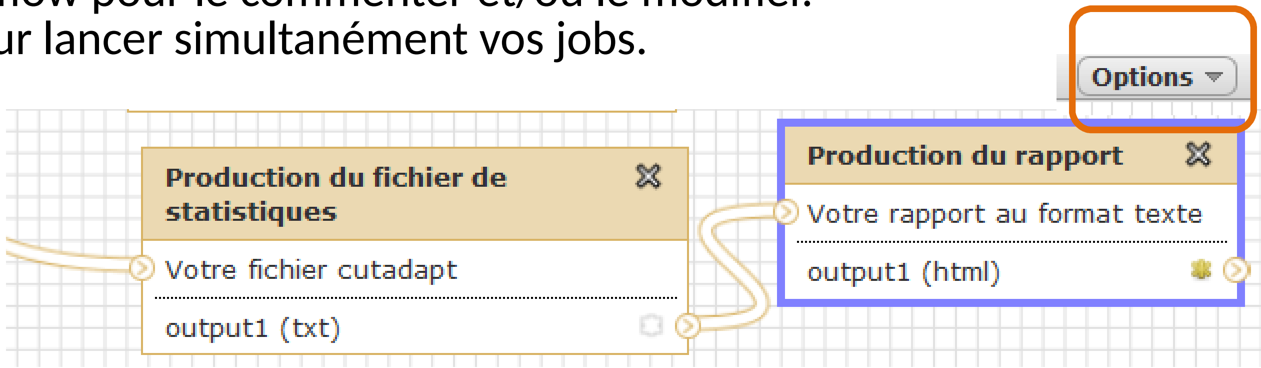
Exécution d’un workflow

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.



Présentation de la plateforme Galaxy.

Premiers pas dans l’instance.

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

Lancement de traitements bioinformatiques.

Auto-formations disponibles en ligne.

Les tools Galaxy

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)

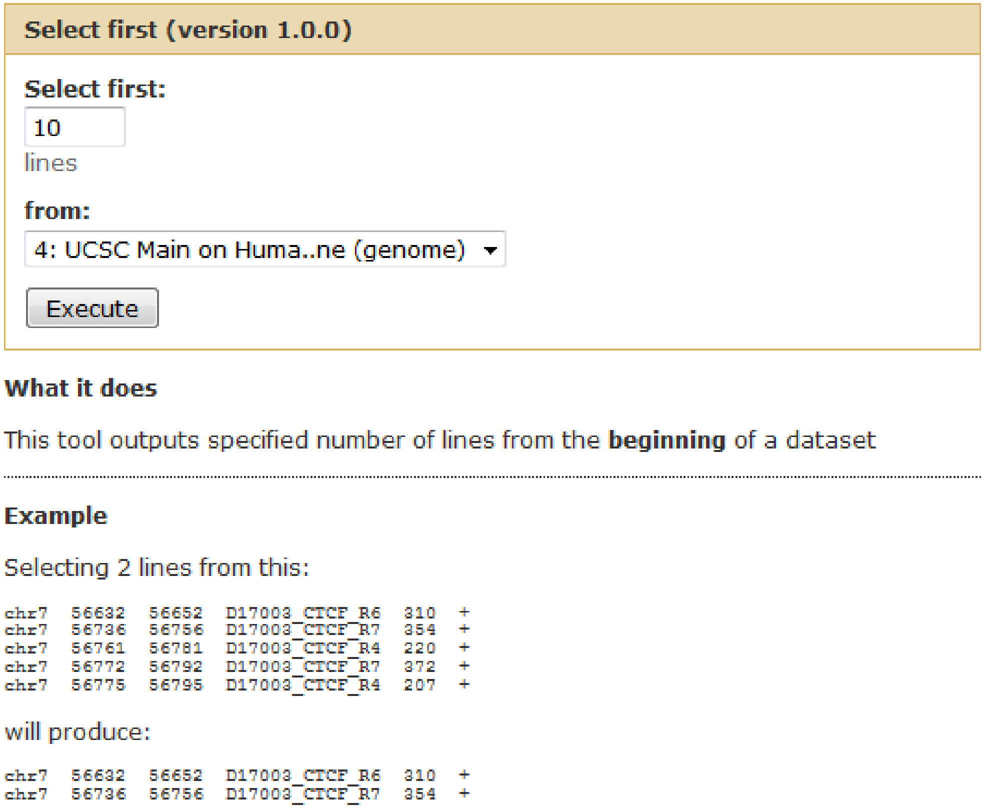
-> Envoyer vos demandes à :
sigenae-support@listes.inra.fr

Outils de traitement de fichiers

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

Voici les principaux outils « non bioinfo » proposés :

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



Select first (version 1.0.0)

Select first: 10 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  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  +
```

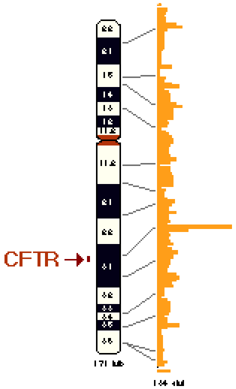
will produce:

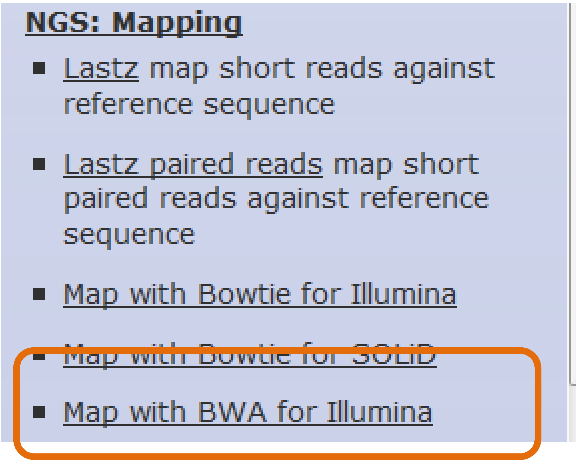
```
chr7  56632  56652  D17003_CTCF_R6  310  +
chr7  56736  56756  D17003_CTCF_R7  354  +
```

Outils bioinformatiques

Mapper un FASTQ sur une référence avec BWA.

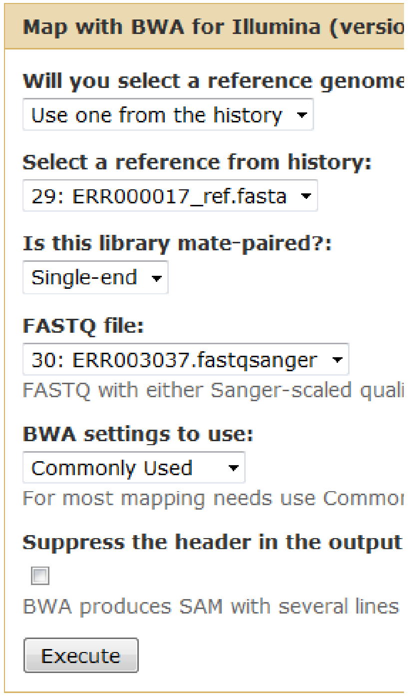
The CFTR gene maps to chromosome 7





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 (versio

Will you select a reference genome
Use one from the history

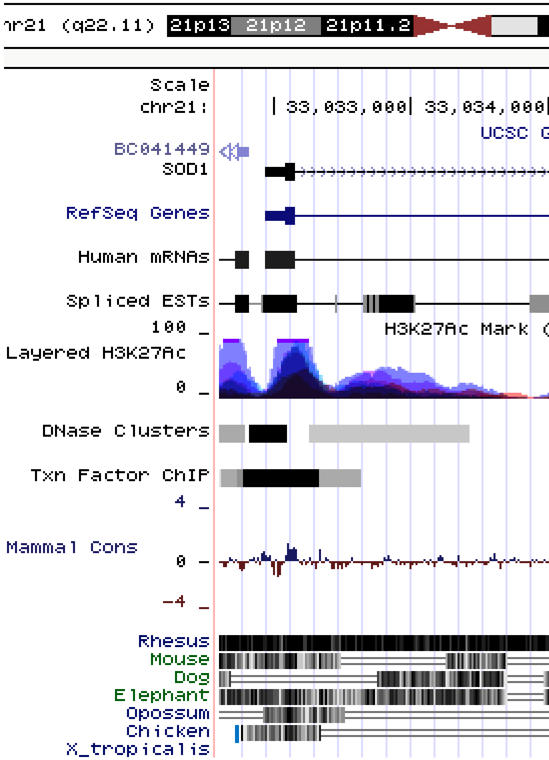
Select a reference from history:
29: ERR000017_ref.fasta

Is this library mate-paired?:
Single-end

FASTQ file:
30: ERR003037.fastqsanger
FASTQ with either Sanger-scaled quali

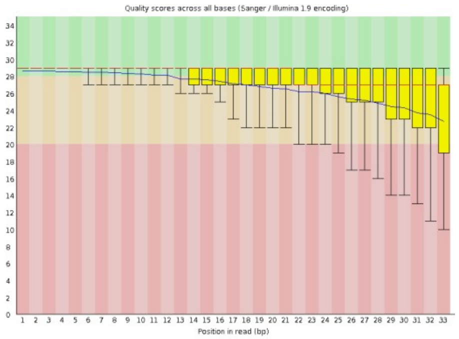
BWA settings to use:
Commonly Used
For most mapping needs use Commo

Suppress the header in the output
☐
BWA produces SAM with several lines



Visualiser la qualité des données avec FASTQC Report.

Visualiser un génome avec UCSC .



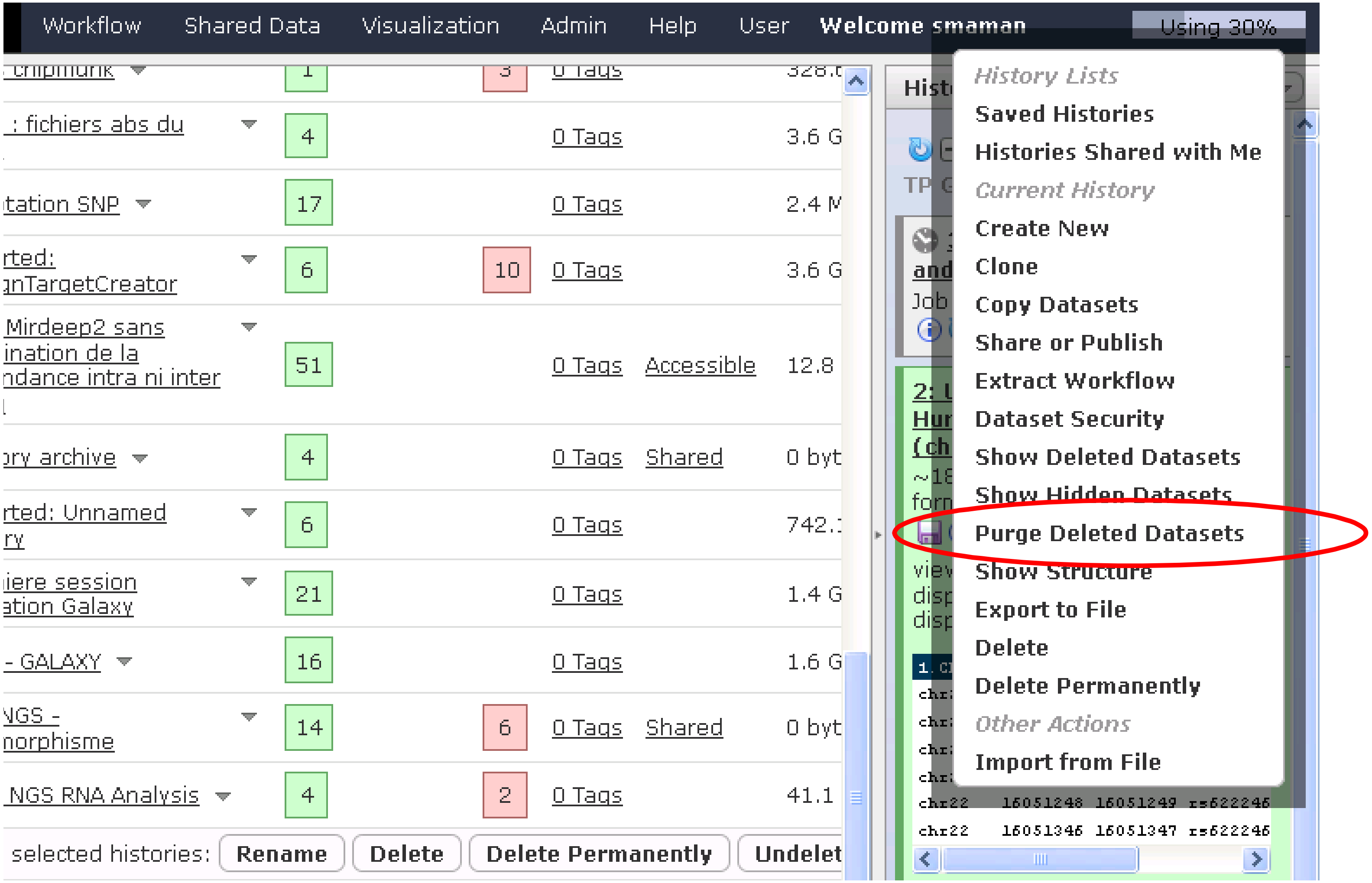
Exercice 2 & 3 :

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

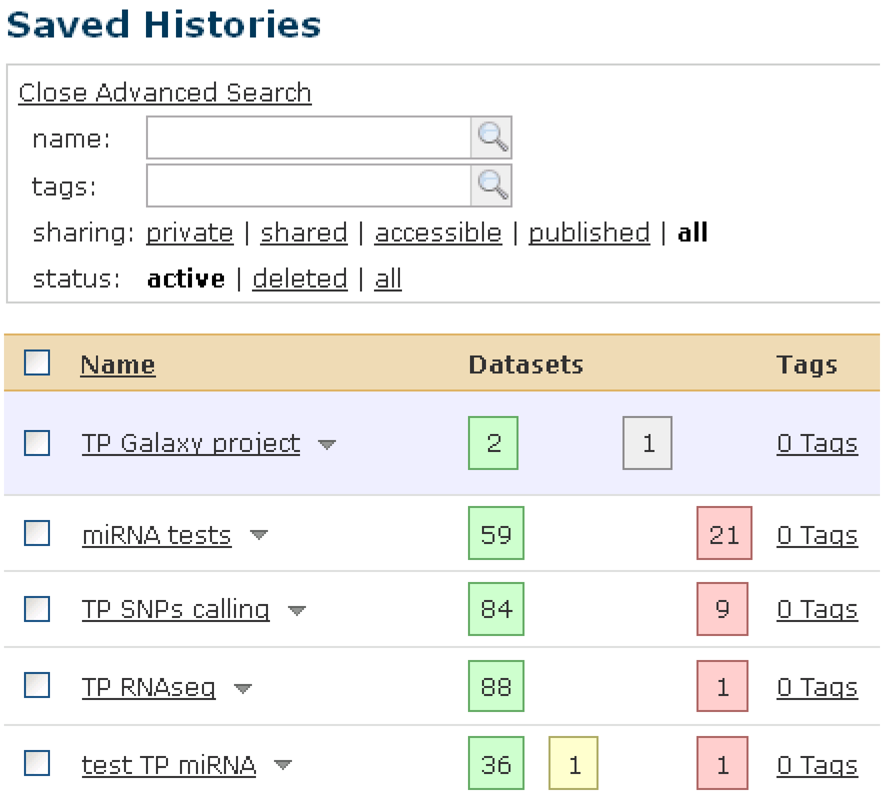
Création et partage d’historiques et de workflows.

Comment gérer son compte ?

Suppression des datasets, historiques et workflows inutiles.
Suppression temporaire ≠ Suppression permanente (purge)
Organiser son espace de travail pour maîtriser son quota.



Revoir les historiques temporairement supprimés



Saved Histories

Close Advanced Search

name:

tags:

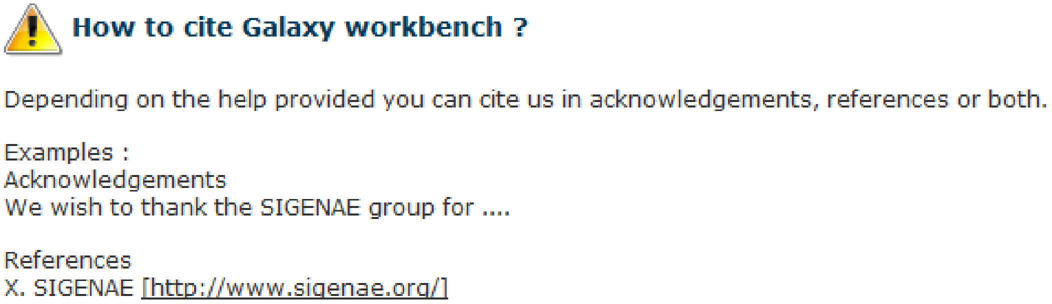
sharing: private | shared | accessible | published | **all**

status: **active** | deleted | all

☐	Name	Datasets				Tags
☐	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
☐	test TP miRNA	36	1		1	0 Tags

Comment citer Galaxy dans vos publications ?

Pour vos publications, citer:
üLes outils utilisés (nom, version).
üLe workflow généré.
üLes références Sigenae et « Galaxy project ».

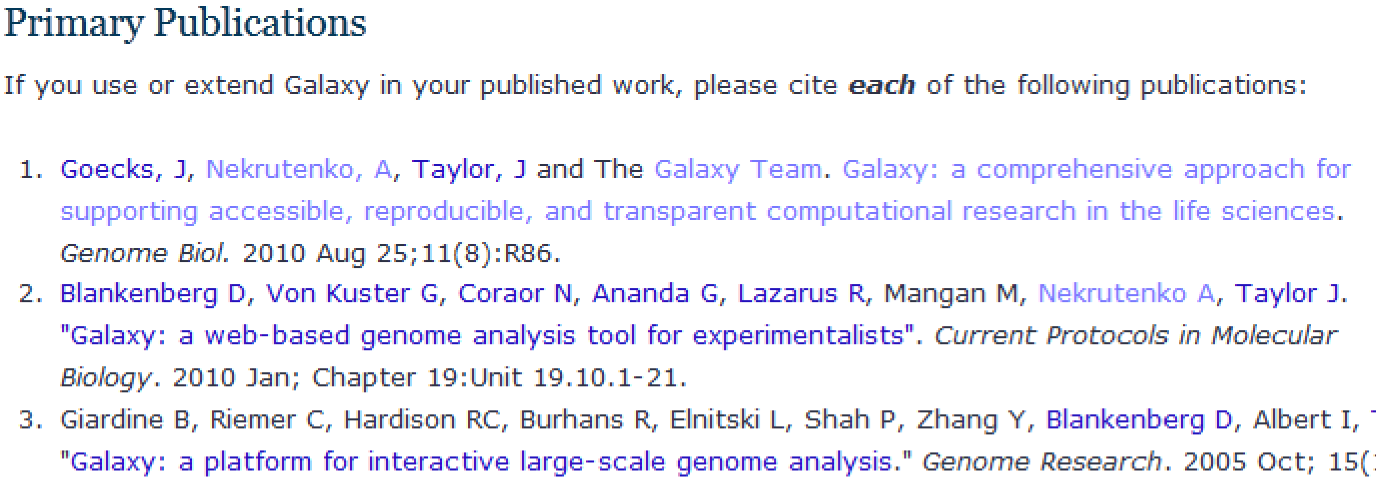


How to cite Galaxy workbench ?

Depending on the help provided you can cite us in acknowledgements, references or both.

Examples :
Acknowledgements
We wish to thank the SIGENAE group for ....

References
X. SIGENAE [http://www.sigenae.org/]



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(

Présentation de la plateforme Galaxy.

Premiers pas dans l’instance.

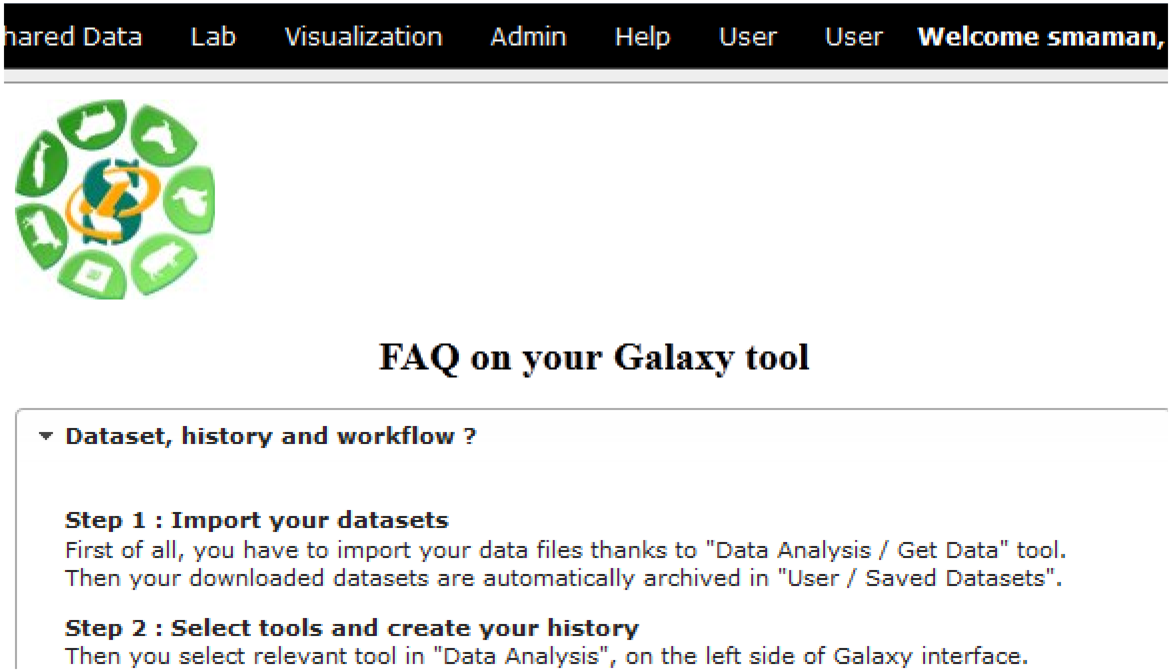
Notions d’outils, d’historique et de workflow.

Lancement de traitements bioinformatiques.

Auto-formations disponibles en ligne.

FAQ et formation en ligne

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



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".

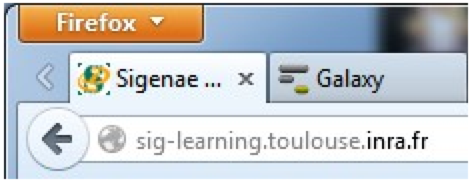
Step 2 : Select tools and create your history
Then you select relevant tool in "Data Analysis", on the left side of Galaxy interface.

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

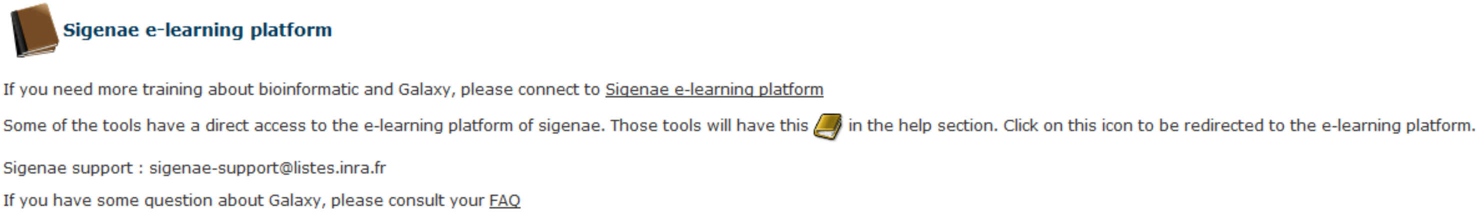
"If you need more training about bioinformatic and Galaxy, please connect to Sigenae e-learning platform.

Votre accès à sig-learning

1 Taper l’adresse de « sig-learning » :
http://sig-learning.toulouse.inra.fr/
Ou directement depuis Galaxy



sig-learning.toulouse.inra.fr



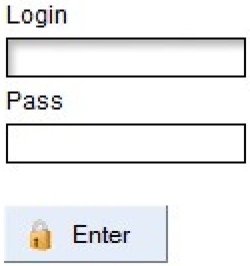
If you need more training about bioinformatic and Galaxy, please connect to Sigenae e-learning platform

Some of the tools have a direct access to the e-learning platform of sigenae. Those tools will have this in the help section. Click on this icon to be redirected to the e-learning platform.

Sigenae support : sigenae-support@listes.inra.fr

If you have some question about Galaxy, please consult your FAQ

2 Authentification



3 Onglet « Trainings » pour accéder à vos e-formations :

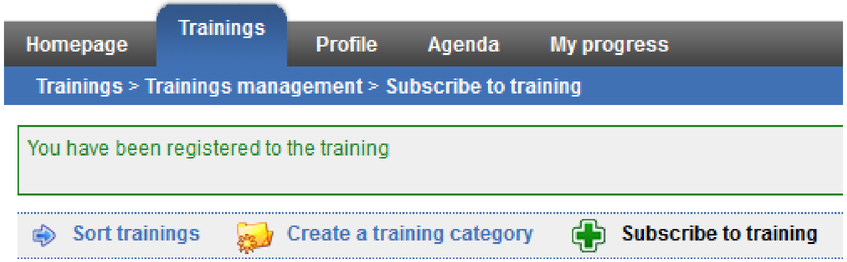
1 - Linux & Unix
UNIX1 – SIGENAE Team

2 - Cluster (en construction)
CLUSTER – SIGENAE Team

3 - Management of large files on Unix and Galaxy
UNIX2 – SIGENAE Team

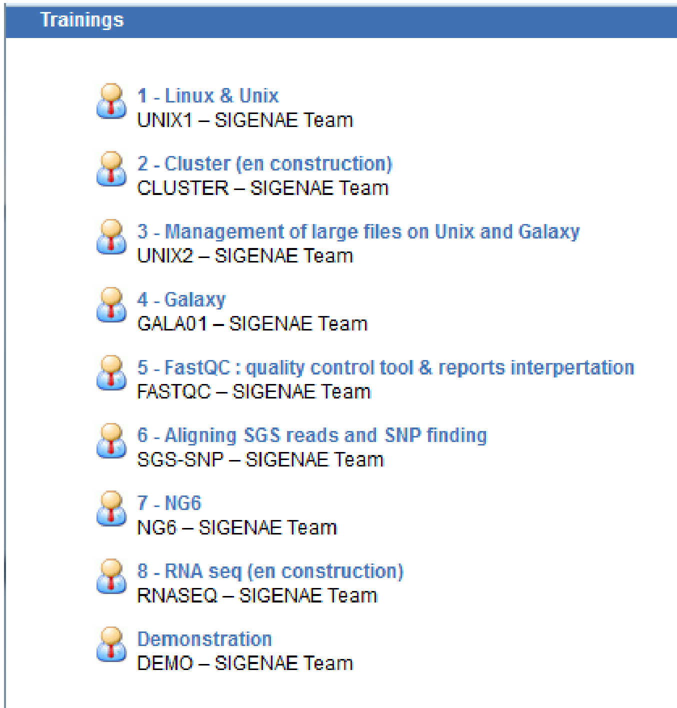
Comment s'inscrire à des e-formations ?

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



You have been registered to the training

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 interpertation
FASTQC – SIGENAE Team

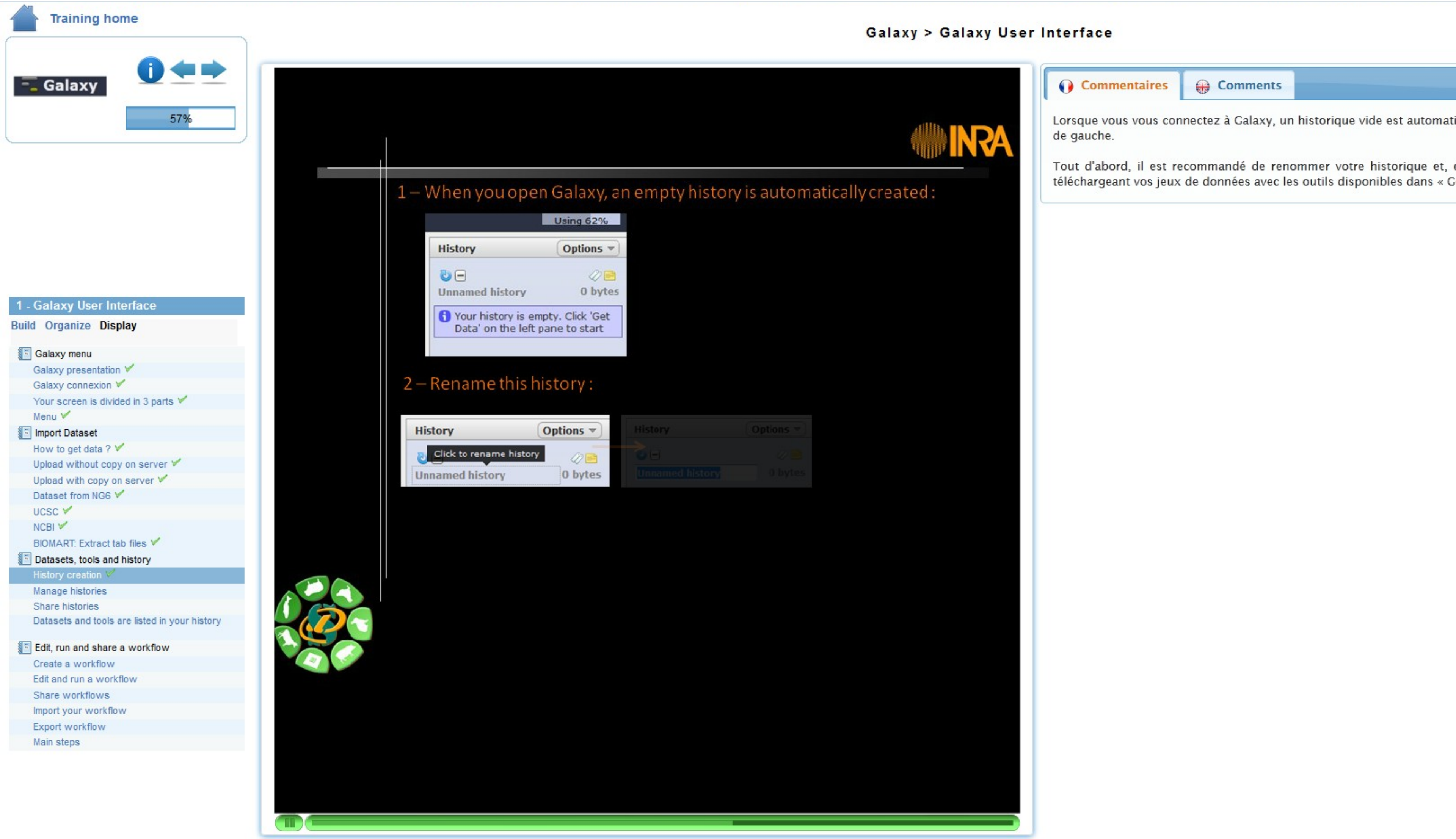
6 - Aligning SGS reads and SNP finding
SGS-SNP – SIGENAE Team

7 - NG6
NG6 – SIGENAE Team

8 - RNA seq (en construction)
RNASEQ – SIGENAE Team

Demonstration
DEMO – SIGENAE Team

Organisation type d'une formation



Galaxy > Galaxy User Interface

1 – When you open Galaxy, an empty history is automatically created :

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

2 – Rename this history :

Lorsque vous vous connectez à Galaxy, un historique vide est automati
de gauche.

Tout d'abord, il est recommandé de renommer votre historique et, e
téléchargeant vos jeux de données avec les outils disponibles dans « G

Merci pour votre écoute

Questionnaire
https://enquetes.inra.fr/index.php?sid=84236&lang=en

Remerciements

Fonds Européen
de Développement Régional

RÉGION
MIDI-PYRÉNÉES