

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 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 ▾

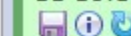


TP FastQC

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

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 » : Sarah Maman, Sabrina Legoueix-Rodriguez.

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

Vos traitements bioinformatiques avec GALAXY

Sabrina Legoueix-Rodriguez - Sarah Maman
Mai 2016

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

Présentation de la plateforme Galaxy.

Premiers pas dans l'instance.

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

Lancement de traitements bioinformatiques.

Quelques statistiques.

Auto-formations disponibles en ligne.



Vidéo disponible
sur « sig-learning »

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 Goecks¹, Anton Nekrutenko^{2, 3}, James Taylor^{1, 3} and The Galaxy Team



EMORY
UNIVERSITY

PENNSTATE



Anton Nekrutenko
Penn State

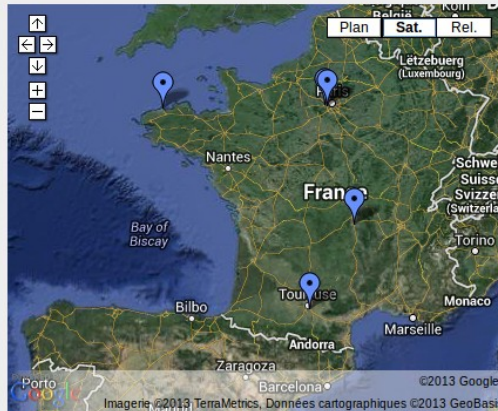


Nate Coraor
Penn State



James Taylor
Emory

- 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



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

- Gratuit & "open source",
- Quota limité, petits jeux de données,
- Impossible d'ajouter des banques, génomes, outils.
- Données non protégées.



Une communauté nationale et internationale très active :

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



L'instance locale Sigeneae de Galaxy :

- Maintenu par Sigeneae.
 - Intégration possible de nouveaux outils / scripts / génomes ...
- **Présentation des particularités de l'instance Sigeneae.**



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 !

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**.

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



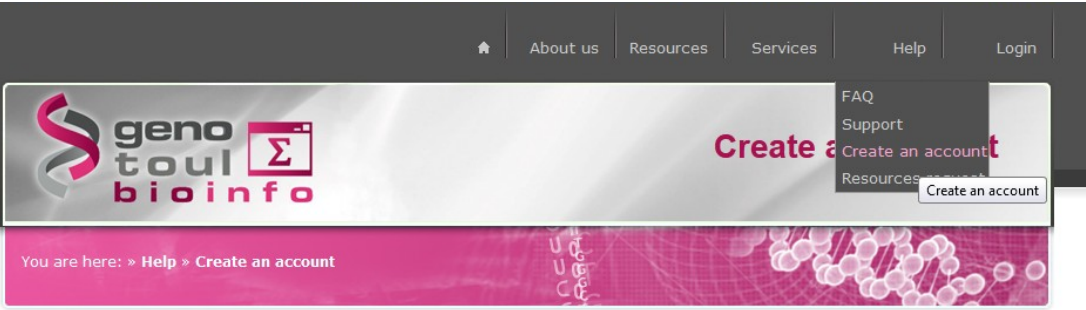
Gestionnaire de tâches

Gère la file d'attente

Cluster de calculs



Exécute



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

1 - Ouvrir un compte sur Genotoul :

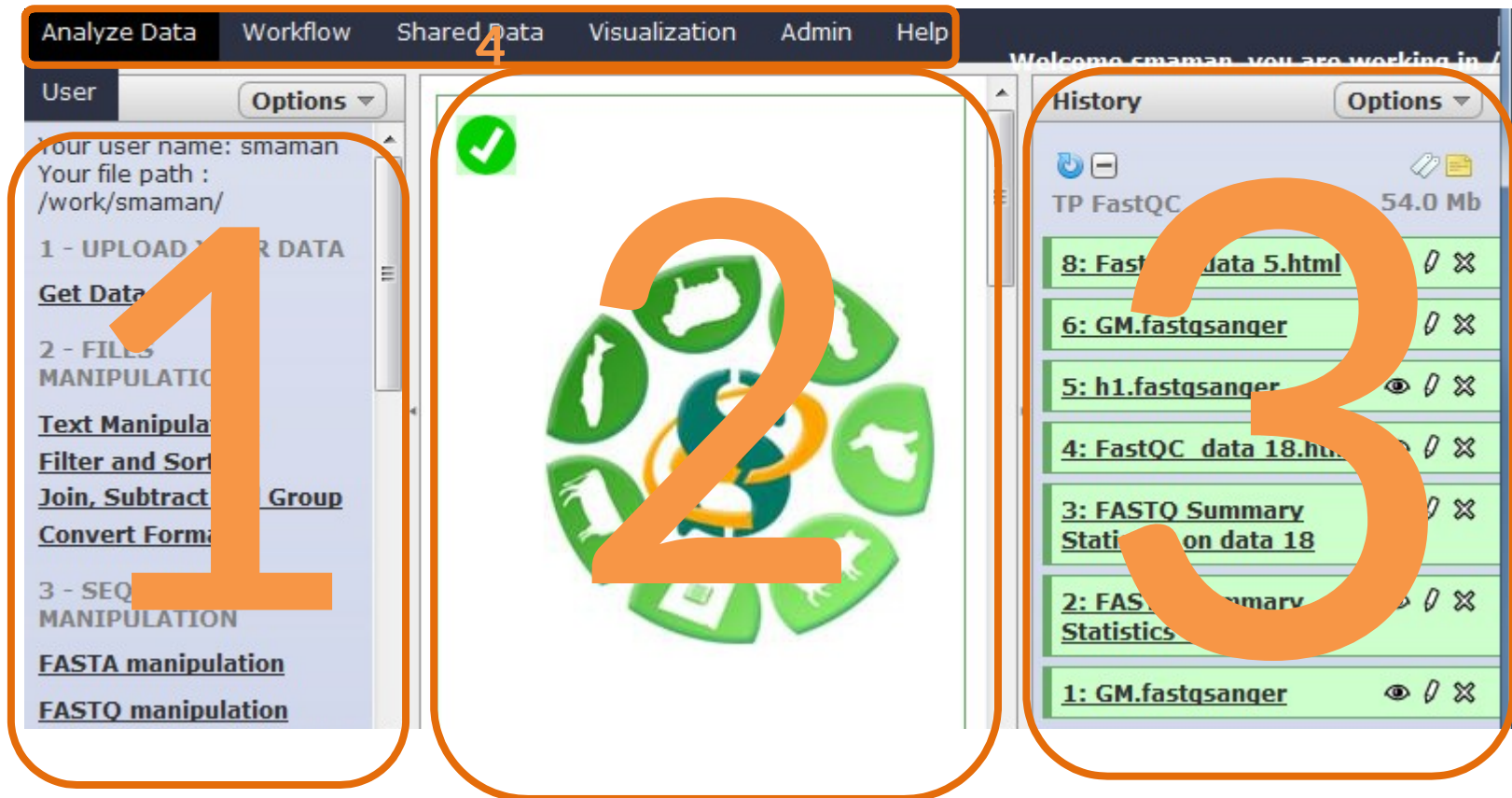
Formulaire de demande de compte:
<http://bioinfo.genotoul.fr>
 (Menu / Help/ Create an account)

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

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

Interface 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 .



The screenshot shows a web interface with a dark top navigation bar containing the following menu items: **Analyze Data**, **Workflow**, **Shared Data**, **Visualization**, **Admin**, and **Help**. Below the navigation bar, the interface is divided into three main sections:

- Section 1 (Left):** A sidebar menu titled "User" with an "Options" dropdown. It displays the user's name "smaman" and file path "/work/smaman/". Below this, there are several tool categories: "1 - UPLOAD YOUR DATA", "Get Data", "2 - FILES MANIPULATION", "Text Manipulation", "Filter and Sort", "Join, Subtract", "Group", "Convert Form", "3 - SEQ MANIPULATION", "FASTA manipulation", and "FASTQ manipulation". A large orange number "1" is overlaid on this section.
- Section 2 (Center):** A large workspace area containing a circular icon composed of smaller green icons representing various tools. A green checkmark is visible in the top-left corner of this area. A large orange number "2" is overlaid on this section.
- Section 3 (Right):** A "History" panel with an "Options" dropdown. It shows a list of jobs with their names and sizes. The jobs listed are: "TP FastQC 54.0 Mb", "8: FastQC data 5.html", "6: GM.fastqsanger", "5: h1.fastqsanger", "4: FastQC data 18.h...", "3: FASTQ Summary Statistics on data 18", "2: FASTQ Summary Statistics", and "1: GM.fastqsanger". A large orange number "3" is overlaid on this section.
- Section 4 (Top):** The navigation menu at the top of the page, which is highlighted with an orange border and a large orange number "4" overlaid on it.

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 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
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.

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:

File type:

Path to file:

History Options

Galaxy sensibilisation - TP12.1 Mb
2 - BWA and FastQC

14:
phiX174 reads.fastqsanger
1.0 Mb
format: fastqsanger, database: ?

```
@080917-and-080922:5:1:185:82  
GATGTTATTTCTTCATTTGGAGGTAACCTCTTAT  
+  
IIIIIIIIIIIIIIIIII<III@FI8A/I0II4I  
@080917-and-080922:5:1:1366:223  
GTTTTCTTCTGCGTCAGTAAGAACGTCAGTGTTC
```


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

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](#)



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

Convert Formats

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](#)

BWA - Bowtie

5 - INDEL ET SNP

Indel Analysis

RNA-Seq

GATK Tools (beta)

Map with BWA for Illumina (version 1.2.2)

Will you select a reference genome from your I

Use one from the history

Select a reference from history:

11: phiX174_genome.fa

Is this library mate-paired?:

Single-end

FASTQ file:

14: phiX174_reads.fastqsanger

FASTQ with either Sanger-scaled quality values (f

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

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](#)

BWA - Bowtie

5 - INDEL ET SNP

Indel Analysis

RNA-Seq

GATK Tools (beta)

Map with BWA for Illumina (version 1.2.2)

Will you select a reference genome from your I

Use one from the history ▾

Select a reference from history:

11: phiX174_genome.fa ▾

Is this library mate-paired?:

Single-end ▾

FASTQ file:

14: phiX174_reads.fastqsanger ▾

FASTQ with either Sanger-scaled quality values (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

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
GTTTCTTCTGCGTCAGTAAGAACGTCAGTGTTC
```

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
GAGTTTTATCGCTTCCATGACGCAGAAGTTAACCTT
AAATTATCTTGATAAAGCAGGAATTACTACTGCTTGT
TGCTGGCGGAAAATGAGAAAATTCGACCTATCCTTGC
GCGACCTTTCGCCATCACTAACGATTCTGTCAAAA
TGGCTTAATATGCTTGGCAGTTCGTCAAGGACTGGT
```

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 I

Select a reference from history:

Is this library mate-paired?:

FASTQ file:

FASTQ with either Sanger-scaled quality values (f

3 - Lancer le job en cliquant sur « Executer ».
 L'exécution 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

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'exécution

👁️ ✎️ ✕

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
AAATTATCTTGATAAAGCAGGAATTACTACTGCTTGI
TGCTGGCGGAAAATGAGAAAATTCGACCTATCCTTGC
GCGACCTTTCGCCATCACTAACGATTCTGTCAAAA
TGGCTTAATATGCTTGGCACGTTCTGCAAGGACTGGI
    
```

Bug - ROUGE : Votre job est planté !

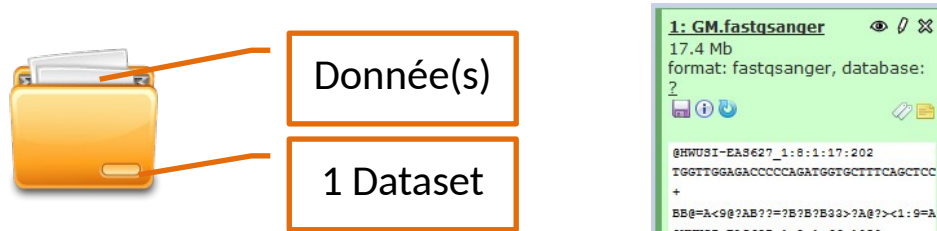
👁️ ✎️ ✕

1: test.txt

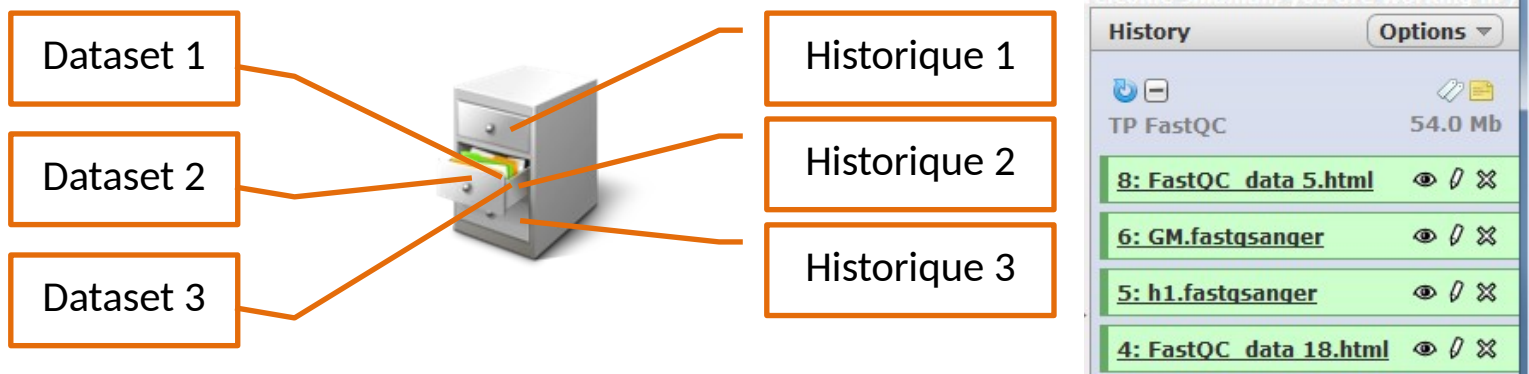
0 bytes
 An error occurred running this job:
info unavailable

🐛 📘 ↻️

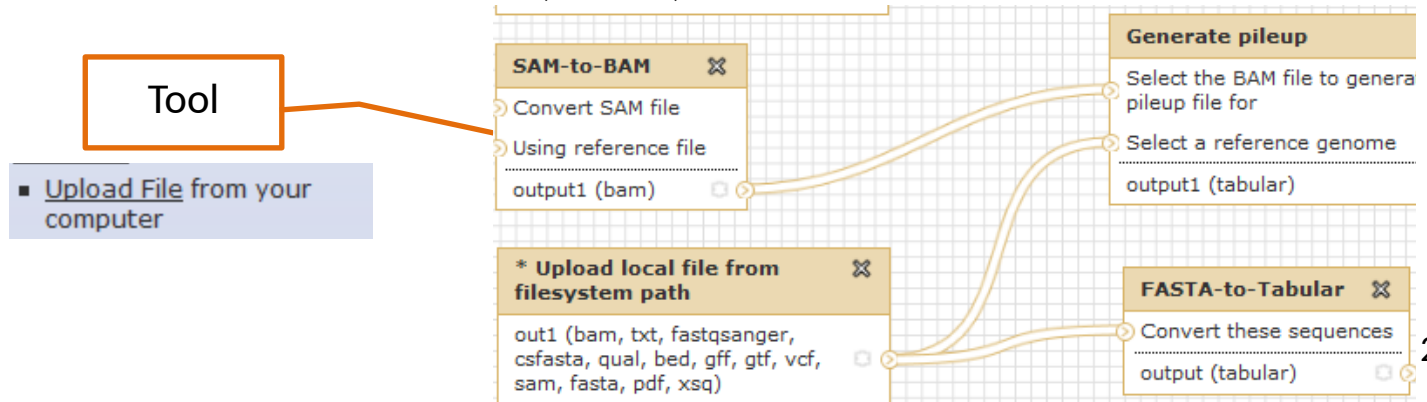
Un **DATASET** est un fichier de données (fichiers d'entrée, fichiers résultats) :



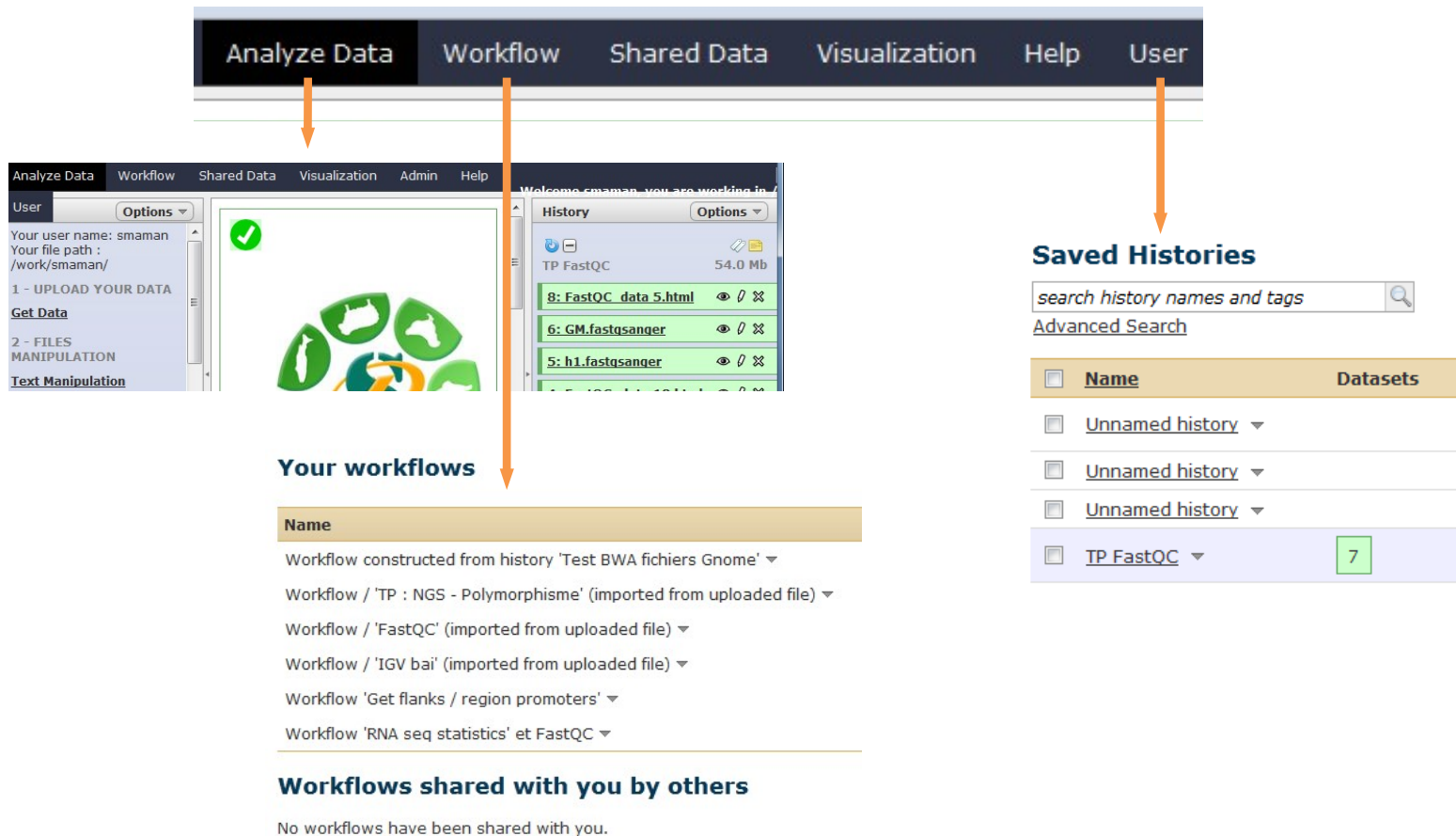
Votre **HISTORIQUE** est un « répertoire » qui « liste » l'ensemble de vos fichiers de données (fichiers d'entrée, fichier résultat) utilisés ou générés par un **TOOL** :



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



- **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.



The screenshot shows the Galaxy web interface with a dark navigation bar at the top containing the following tabs: **Analyze Data**, **Workflow**, **Shared Data**, **Visualization**, **Help**, and **User**. Three orange arrows point from these tabs to their respective content areas below:

- An arrow from **Analyze Data** points to a sidebar menu with options like 'Get Data', '2 - FILES MANIPULATION', and 'Text Manipulation'.
- An arrow from **Workflow** points to a 'Your workflows' section containing a list of workflow names such as 'Workflow constructed from history 'Test BWA fichiers Gnome'', 'Workflow / 'TP : NGS - Polymorphisme'', etc.
- An arrow from **User** points to a 'Saved Histories' section with a search bar and a table of saved history items.

Saved Histories

search history names and tags

Advanced Search

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

Présentation de la plateforme Galaxy.

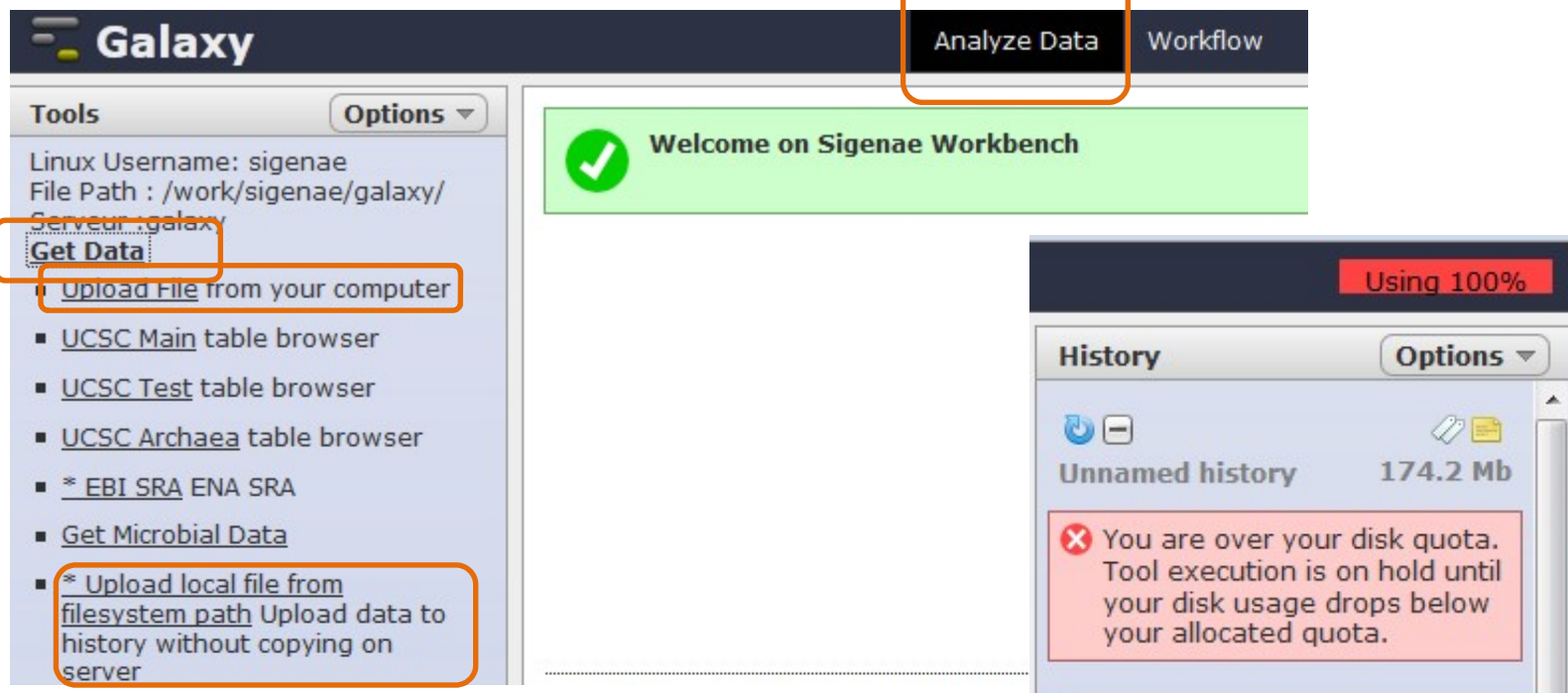
Premiers pas dans l'instance.

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

Lancement de traitements bioinformatiques.

Quelques statistiques.

Auto-formations disponibles en ligne.



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.

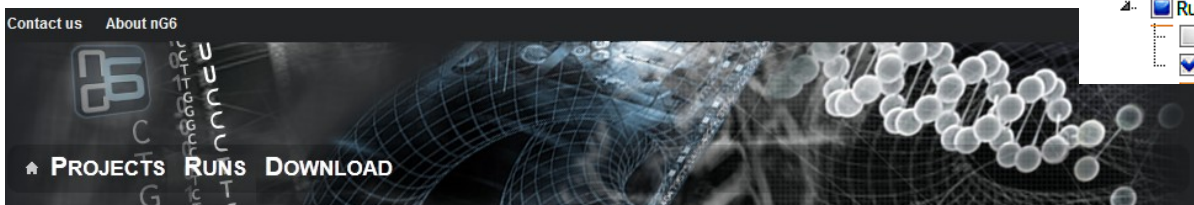
Vos traitements ne sont pas limités par la capacité de votre PC.



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:

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 coming 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).
- September 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.



Données UCSC, Ensembl, BIOMART :

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

group: Genes and Gene Prediction Tracks **track:** UCSC Genes

table: knownGene

region: genome chr21:33031597-33041570

output format: BED - browser extensible data

file type returned: plain text

Galaxy

EMBL-EBI

European Nucleotide Archive

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.

Quelques statistiques.

Auto-formations disponibles en ligne.

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 (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.

<input type="checkbox"/>	SwanPorc ▾	18	0	0 Tags	Shared	0 bytes
<input type="checkbox"/>	FastQC ▾	6	0	0 Tags	Shared	17.4 Mb
<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
<input type="checkbox"/>	indexation genome ▾	1	0	0 Tags		46 bytes
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 'Visual'. The 'Shared Data' tab is active. Below the navigation, there is a search bar for 'Published Histories' with the placeholder text 'search name, annotation, owner, and tags'. Below the search bar, there is a table with columns 'Name' and 'Annotation'. One entry is visible: 'RNAseq' with the annotation 'TP 1 suite : region promotrices'. To the right, there is a sidebar menu with options like 'Data Libraries', 'Published Histories', 'Published Workflow', and 'Published Visualizations'. Below the sidebar, there is a section titled 'Histories shared with you by others' with a table showing 'Name' and 'Datasets'. A user profile dropdown is visible in the top right corner, showing 'Welcome smaman' and options like 'History Lists', 'Saved Histories', 'Histories Shared with Me', 'Current History', and 'Create New'.

Vos historiques sont automatiquement sauvegardés.

Pour partager ou publier un historique :

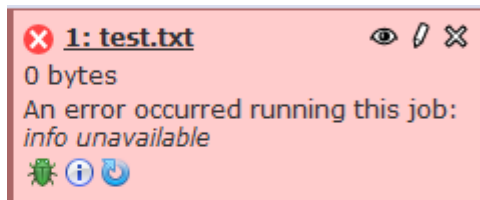
User / Saved histories / Cliquer sur le nom de l'historique / Share ou Publish

This screenshot shows the user profile dropdown menu in the Galaxy Sig interface. The menu is open, showing the user's name 'smaman' and the email address 'smaman@toulouse.inra.fr'. The menu options include 'Logout', 'Saved Histories', 'Saved Datasets', and 'Public Name'. The background shows a table with columns 'Datasets' and 'Tags', with some data visible.

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).

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>
    pit.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 Sigeneae Team

Please create a ticket in [Redmine](#) or send a mail to [Galaxy administrator](#).

Vous pouvez aussi créer un ticket sous Redmine.

Tool: Clone metagenomic

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

Input Parameter

Other ace files
Other ace files

Inheritance Chain

ContigLengthG1000ProfG8.res



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

Tools Options

1 - UPLOAD YOUR DATA

Get Data

2 - FILES MANIPULATION

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

3 - SEQUENCES MANIPULATION

FASTA manipulation
FASTQ manipulation
SAM/BAM manipulation : Picard (beta)
SAM/BAM manipulation : SAMtools
Fetch Sequences

Saved Histories

search history names and tags

Advanced Search

Name	Datasets	Tags	Sharing
<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 Shared, Accessible
<input type="checkbox"/> test TP miRNA	36	1	0 Tags
<input type="checkbox"/> Unnamed history			0 Tags

Analyse OK

Analyse en attente

Analyse en erreur

Analyse en cours

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

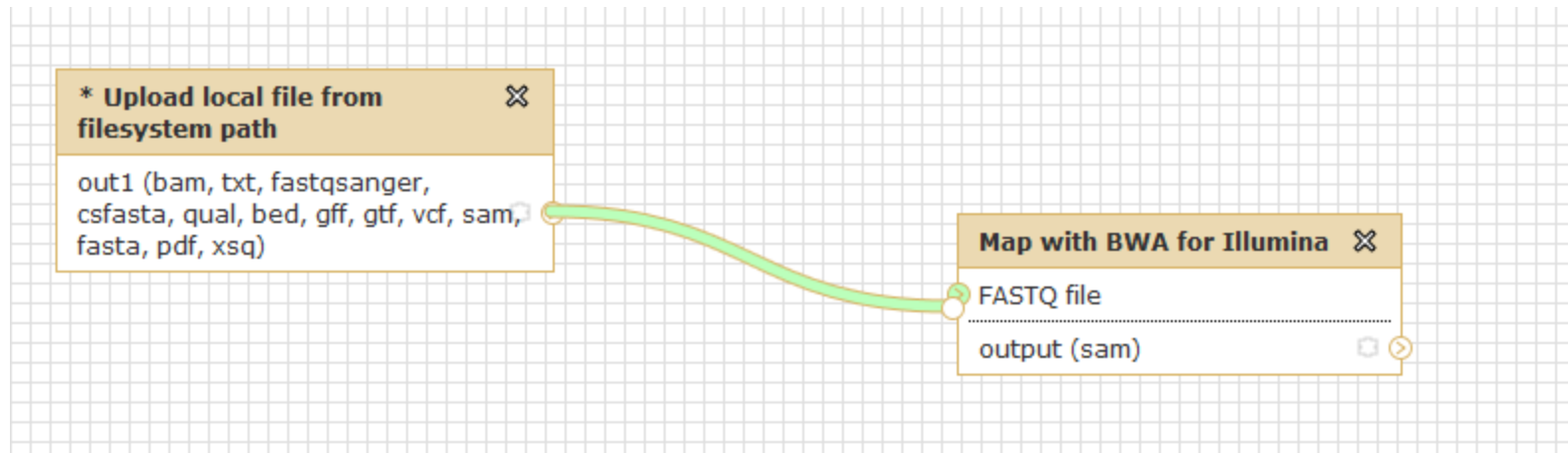


Rapidement, beaucoup de données sont générées

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

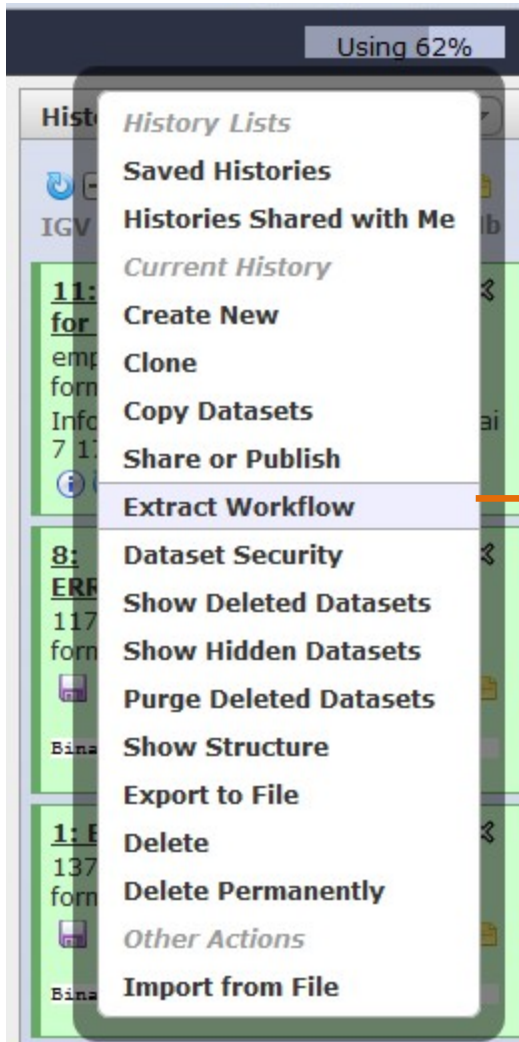
Depuis une page blanche, vous pouvez concevoir un workflow.

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





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 <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, there is a dark navigation bar with the 'Galaxy' logo on the left and 'Analyze Data' and 'Workflow' buttons on the right. The 'Workflow' button is highlighted with an orange border. Below the navigation bar, the page title is 'Your workflows'. A table lists several workflows. A context menu is open over one of the workflows, showing the following options: Edit, Run, Share or Publish, Download or Export, Clone, Rename, View, and Delete. 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 editor interface. Two workflow steps are connected by a line. The first step is 'Production du fichier de statistiques' and the second is 'Production du rapport'. The 'Production du rapport' step has an 'Options' dropdown menu highlighted with an orange border. The 'Production du rapport' step has a text input field labeled 'Votre rapport au format texte' and an output field labeled 'output1 (html)'. The 'Production du fichier de statistiques' step has an input field labeled 'Votre fichier cutadapt' and an output field labeled 'output1 (txt)'.

Présentation de la plateforme Galaxy.

Premiers pas dans l'instance.

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

Lancement de traitements bioinformatiques.

Quelques statistiques.

Auto-formations disponibles en ligne.

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.

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:

lines

from:

What it does

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

Example

Selecting 2 lines from this:

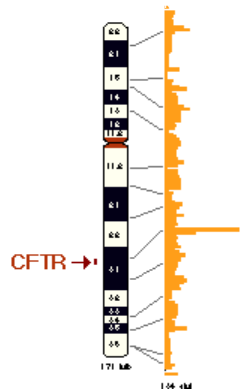
```
chr7 56632 56652 D17003_CTCF_R6 310 +
chr7 56736 56756 D17003_CTCF_R7 354 +
chr7 56761 56781 D17003_CTCF_R4 220 +
chr7 56772 56792 D17003_CTCF_R7 372 +
chr7 56775 56795 D17003_CTCF_R4 207 +
```

will produce:

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

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 (version 0.7.17)

Will you select a reference genome?

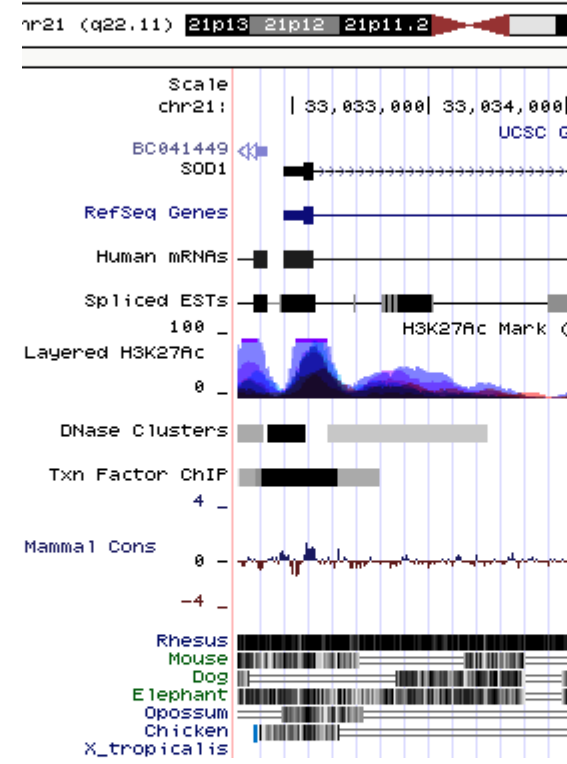
Select a reference from history:

Is this library mate-paired?:

FASTQ file:

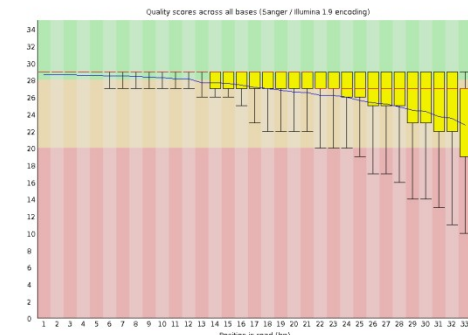
BWA settings to use:

Suppress the header in the output



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.

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

The screenshot shows the Galaxy user interface with a list of workflows and a context menu open over the history list. The workflow list includes columns for name, number of datasets, number of deleted datasets, tags, and size. The context menu, titled 'History Lists', contains several options, with 'Purge Deleted Datasets' circled in red.

Workflow	Shared Data	Visualization	Admin	Help	User	Welcome smaman	Using 30%
cripmunk	1	3	0 Tags		328.0		
: fichiers abs du	4		0 Tags		3.6 G		
ation SNP	17		0 Tags		2.4 M		
rted: anTargetCreator	6	10	0 Tags		3.6 G		
Mirdeep2 sans ination de la ndance intra ni inter	51		0 Tags	Accessible	12.8		
ry_archive	4		0 Tags	Shared	0 byt		
rted: Unnamed ry	6		0 Tags		742.0		
iere session ation Galaxy	21		0 Tags		1.4 G		
- GALAXY	16		0 Tags		1.6 G		
VGS - norphisme	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

Saved Histories

[Close Advanced Search](#)

name: 

tags: 

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

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

<input type="checkbox"/> <u>Name</u>	<u>Datasets</u>		<u>Tags</u>
<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 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.

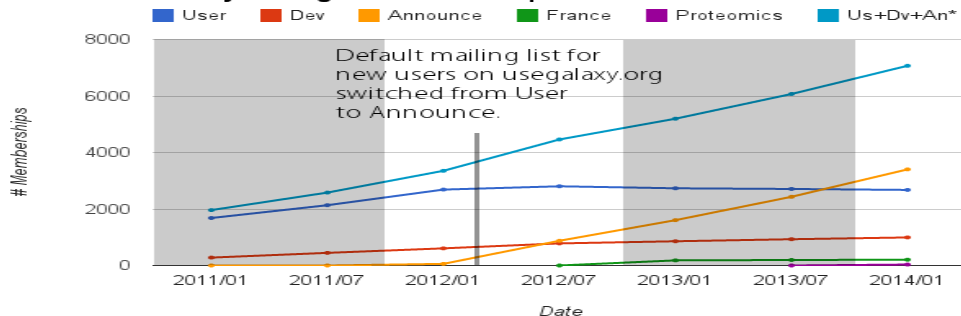
Quelques statistiques.

Auto-formations disponibles en ligne.

Statistiques du Galaxy Project

<https://wiki.galaxyproject.org>

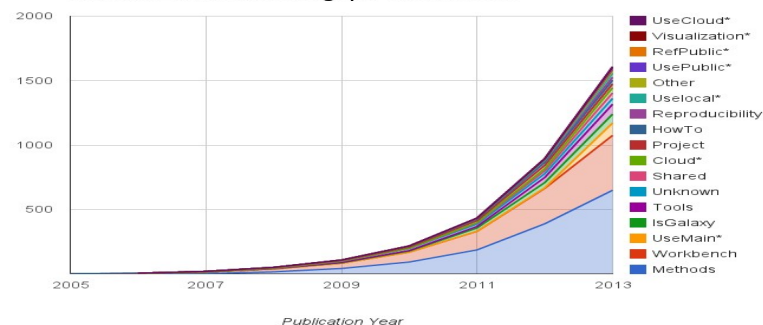
Galaxy Mailing List Memberships



Une communauté internationale vivante.

Une communauté française grandissante.

Publications: Cumulative Tags per Year 2005-2013*



Belle augmentation des citations Galaxy dans les publications

Statistiques de l'instance Galaxy Sigeneae / BioInfo Genotoul

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

Historique mensuel des visites



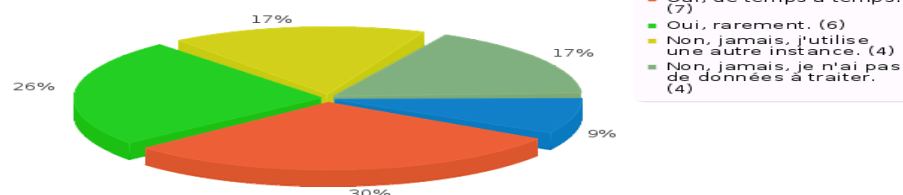
23 réponses sur ~40 utilisateurs

~40 utilisateurs

Les 10 plus gros utilisateurs (hors tests Sigeneae) utilisent Galaxy dans le cadre de leur projet.

65% des utilisateurs Galaxy ayant répondu à l'enquête utilisent l'instance Sigeneae

Utilisez-vous l'instance Sigeneae de Galaxy ?



Présentation de la plateforme Galaxy.

Premiers pas dans l'instance.

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

Lancement de traitements bioinformatiques.

Quelques statistiques.

Auto-formations disponibles en ligne.

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

Shared Data Lab Visualization Admin Help User User **Welcome smaman,**



FAQ on your Galaxy tool

▼ Dataset, history and workflow ?

Step 1 : Import your datasets

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

Step 2 : Select tools and create your history

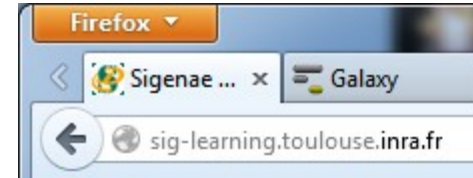
Then you select relevant tool in "Data Analysis", on the left side of Galaxy interface.

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




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

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



Sigenae e-learning platform

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

Login

Pass

 Enter

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

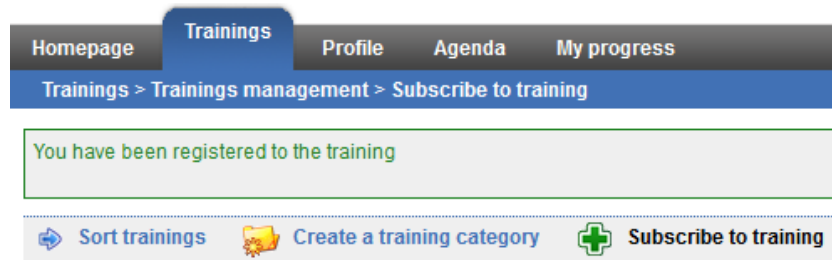


- Homepage
- Trainings**
- Profile
- Agenda
- Reporting
- Administration

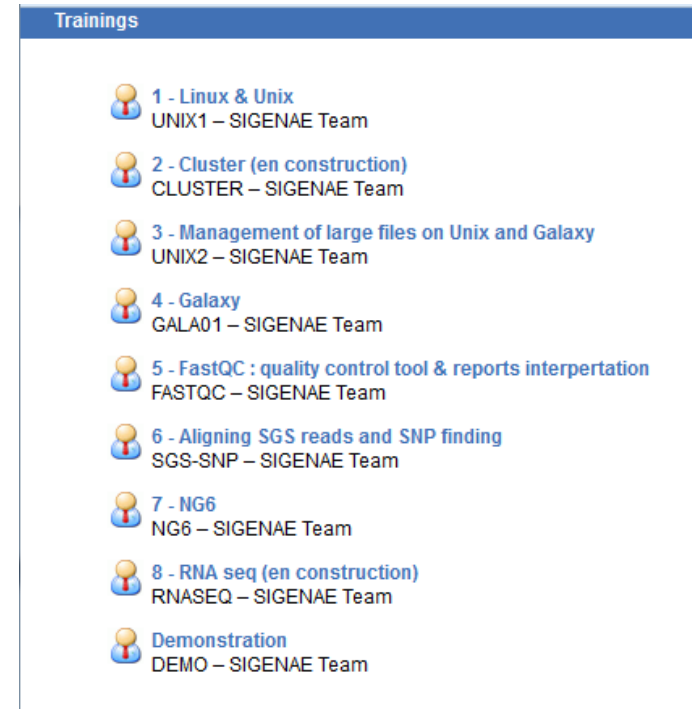
Trainings

-  **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

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



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



Training home

Galaxy 57%

Galaxy > Galaxy User Interface

1 - Galaxy User Interface

Build Organize Display

- Galaxy menu
 - Galaxy presentation ✓
 - Galaxy connexion ✓
 - Your screen is divided in 3 parts ✓
 - Menu ✓
- Import Dataset
 - How to get data ? ✓
 - Upload without copy on server ✓
 - Upload with copy on server ✓
 - Dataset from NG6 ✓
 - UCSC ✓
 - NCBI ✓
 - BIOMART: Extract tab files ✓
- Datasets, tools and history**
 - History creation ✓**
 - Manage histories
 - Share histories
 - Datasets and tools are listed in your history
- Edit, run and share a workflow
 - Create a workflow
 - Edit and run a workflow
 - Share workflows
 - Import your workflow
 - Export workflow
 - Main steps

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

2 – Rename this history :

Commentaires Comments

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

Tout d'abord, il est recommandé de renommer votre historique et de télécharger 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

