Tools

Options

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

1 - UPLOAD YOUR DATA

Get Data

2 - FILES MANIPULATION

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

3 - SEQUENCES MANIPULATION

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

4 - MAPPING

BWA - Bowtie

5 - INDEL ET SNP

Indel Analysis
RNA-Seq
GATK Tools (beta)

6 - SRNASEQ

Analyse des miRNA
Annotations
Alignement sur reference

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

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 pour vos traitements bioinformatiques

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

« sig-learning » pour votre auto-formation continue en ligne

http://sig-learning.toulouse.inra.fr

Vos traitements bioinformatiques avec GALAXY

Galaxy pour vos traitements (bio)informatiques

Vidéo disponible sur « sig-learning »

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

Présentation de la plateforme Galaxy.

Comment récupérer vos données (privées et publiques) ?

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

Lancement de traitements bioinformatiques.

Guide pour les utilisateurs Galaxy.

Galaxy Project

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

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

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.

PENNSTATE 1855

EMORY UNIVERSITY

Anton Nekrutenko
Penn State

Nate Coraor
Penn State

James Taylor
Emory

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

Une « Galaxy » parmi tant d'autres

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

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

Nombreuses autres instances :

- Curie (Nebula)
- URGI

Une communauté nationnale et internationnale très active :

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

L'instance locale Sigenae de Galaxy :

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

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

Des outils accessibles à tous..

Inutile de savoir :

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

Inutile de s'inquiéter pour son disque dur:

- ✓Jobs lancés sur Genotoul.
- ✓Pas d'archivage de fichiers sont votre PC.

Inutile d'attendre la fin d'un traitement:

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

Plateforme
Vos données
Historique
Workflow
Bioinfo
Vous

Galaxy « la bioinformatique pour tous »

Galaxy est :

- « Open source ».
- Développé et maintenu par une communauté active.
- Une plateforme proposant un ensemble d'outils bioinformatiques.
- Accessible : **http://sigenae-workbench.toulouse.inra.fr/**
- Une "constellation" d'outils (analyser, manipuler, visualiser)

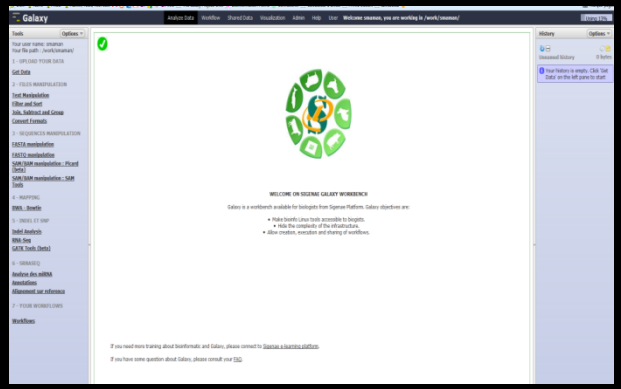
Les biologistes peuvent :

- 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 workflows.



Plateforme

Vos données

Historique

Workflow

Bioinfo

Vous

Contexte d'utilisation dans un laboratoire

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

✓Complémentaire au « **cahier de laboratoire** », avec **archivage** des données de séquençage

→ Retrouver les données, les outils, les références pour la **publication** ….

✓Manipuler les informations contenues dans un fichier, de façon **simple** et **rapide**.

✓Autres fonctionnalités intéressantes :

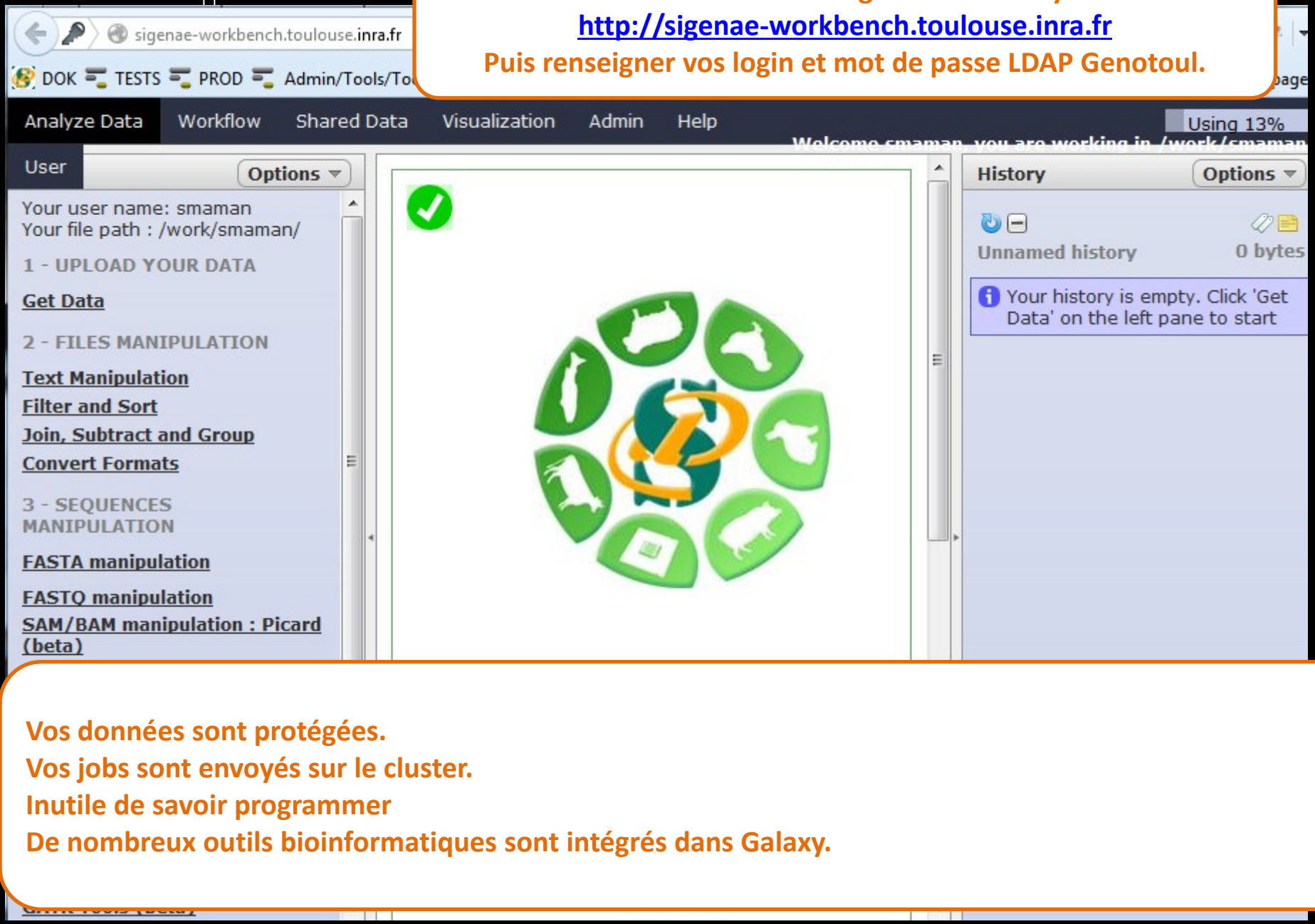
- ✓"**mapping**" des séquences,
- ✓analyse des régions de **variation** ("indel", substitution) ...

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

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

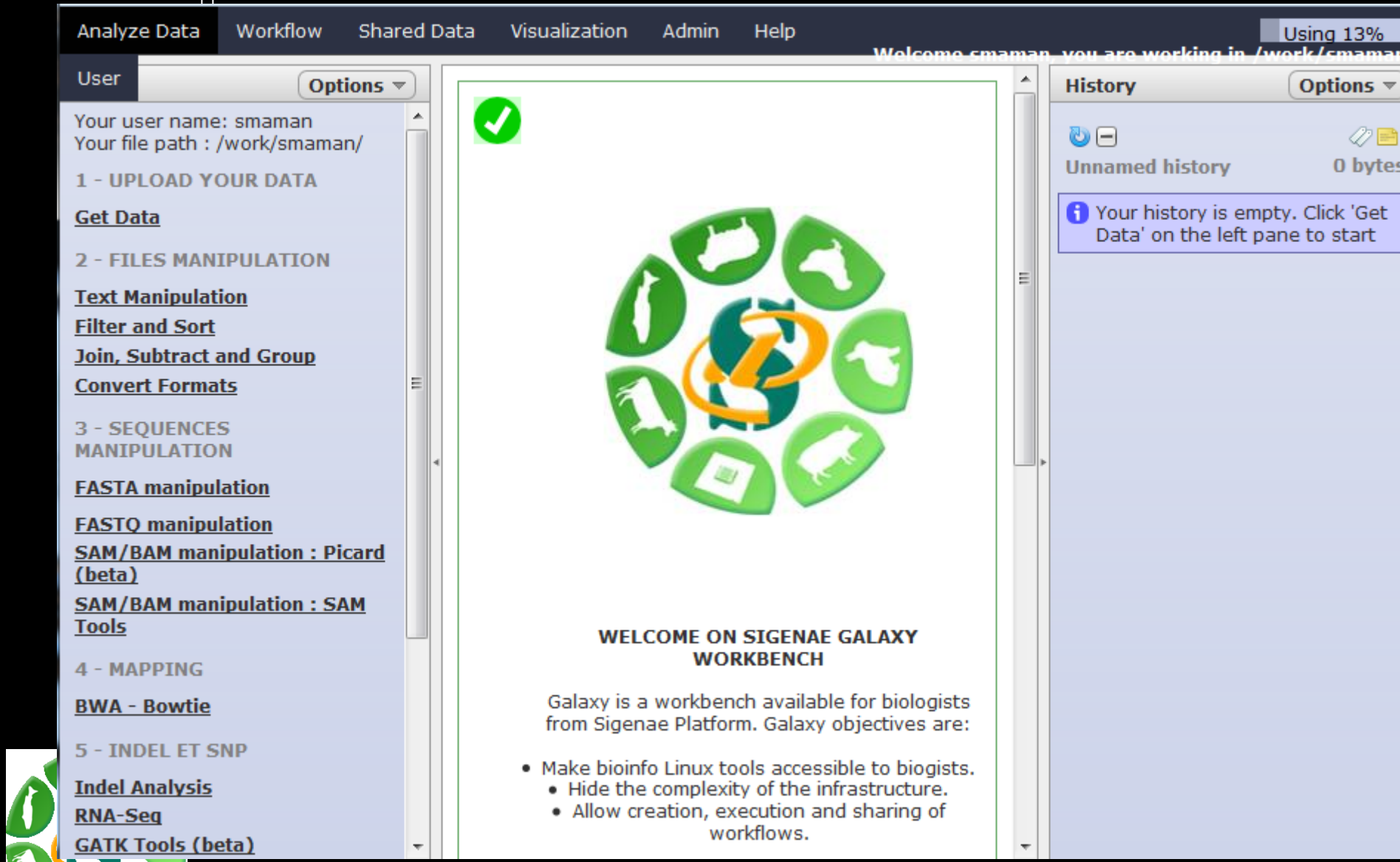
→ Galaxy devient **VOTRE BOITE A OUTILS**.

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

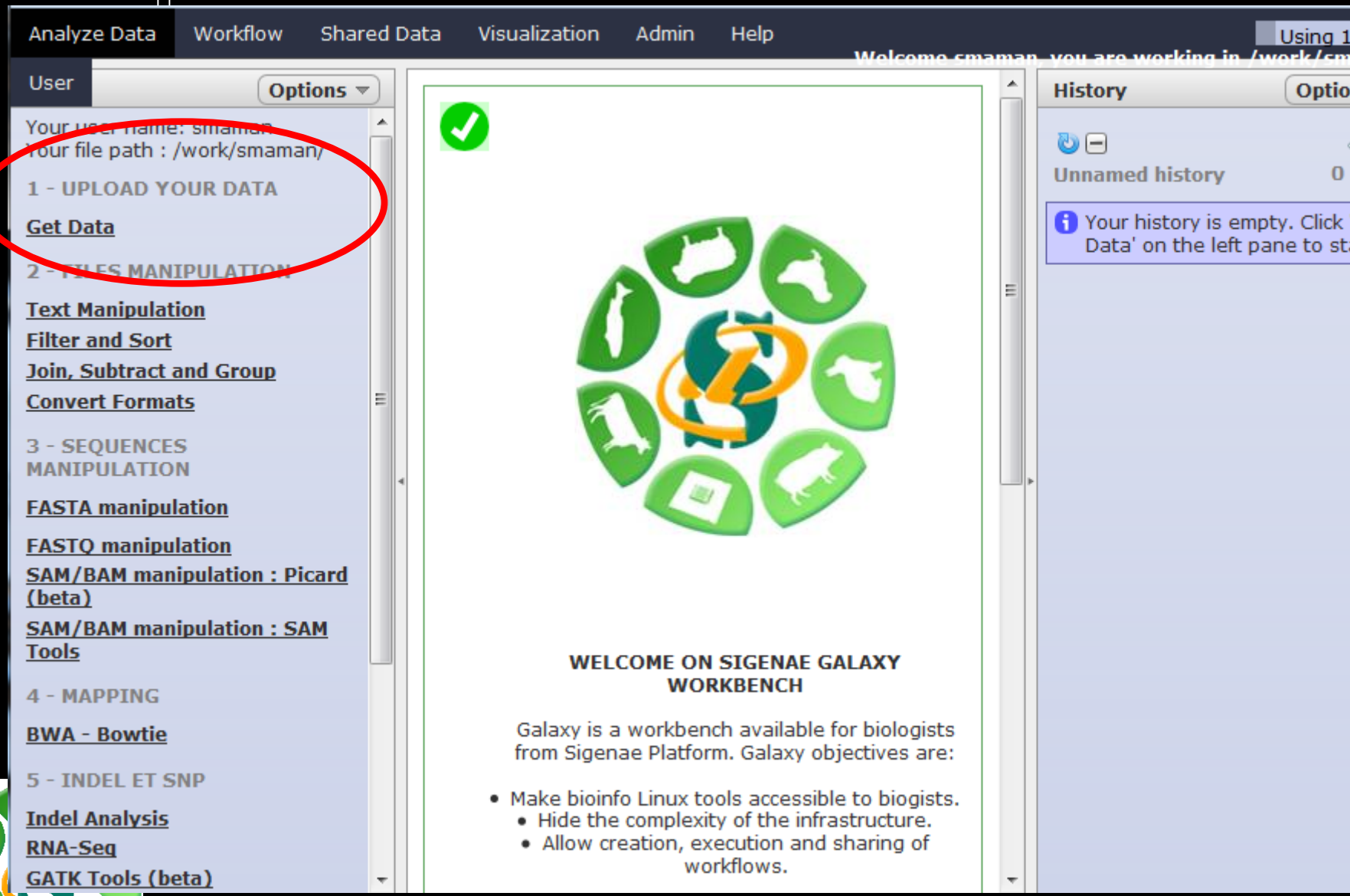


Vos données sont protégées.
Vos jobs sont envoyés sur le cluster.
Inutile de savoir programmer
De nombreux outils bioinformatiques sont intégrés dans Galaxy.

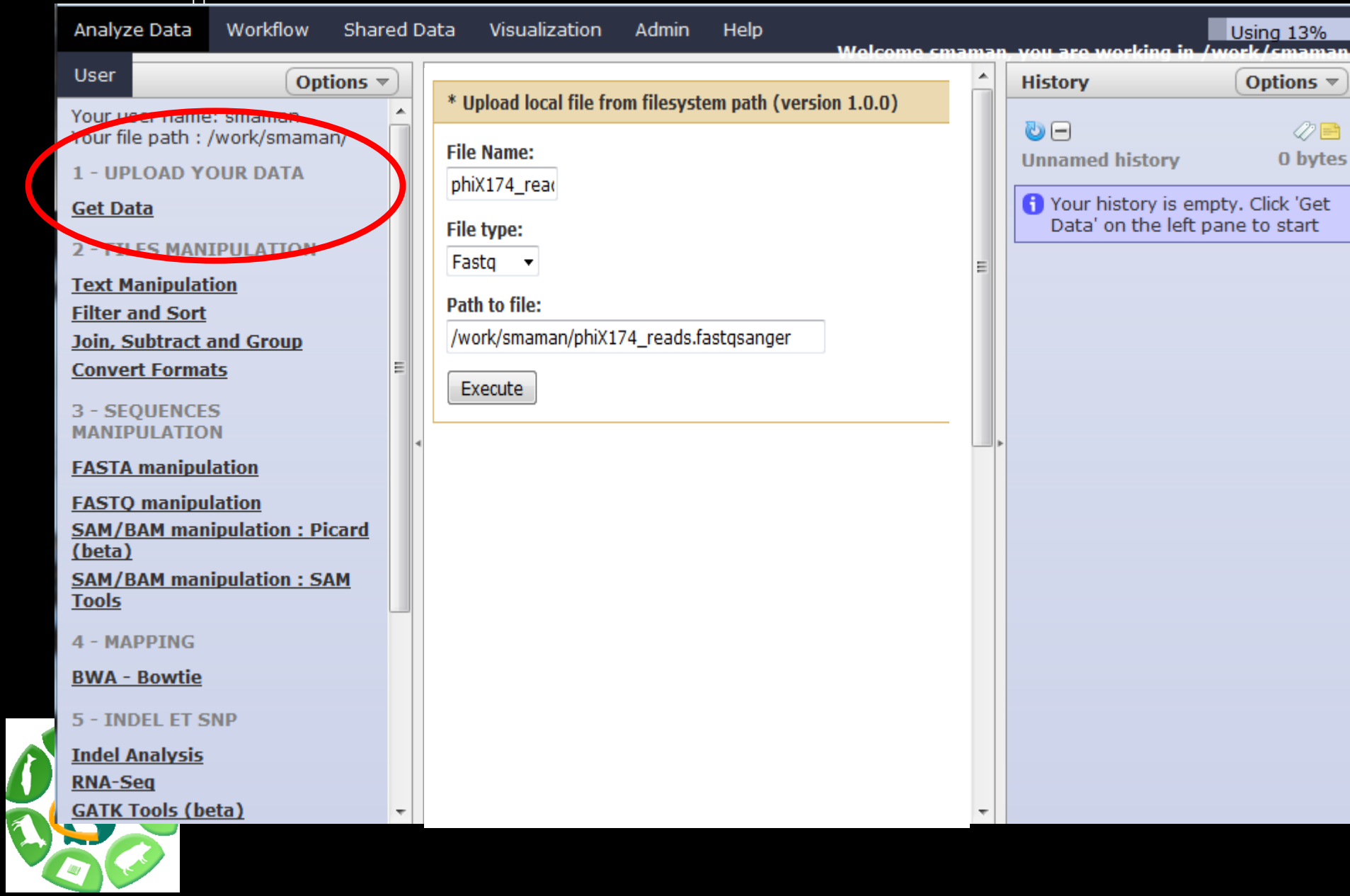
Exemple : Analyse en 3 clics sans utiliser votre disque dur !



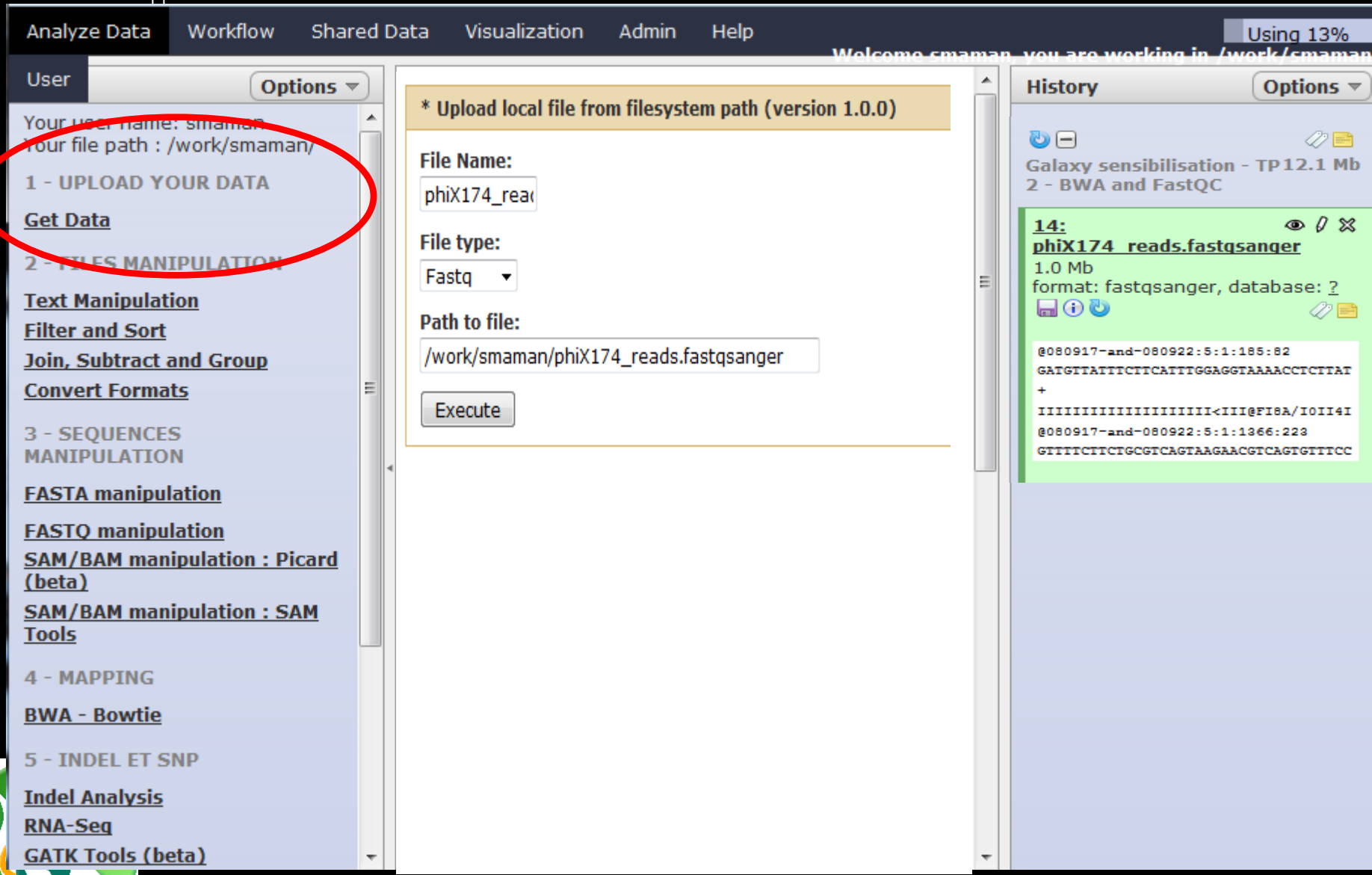
Exemple : Analyse en 3 clics sans utiliser votre disque dur !



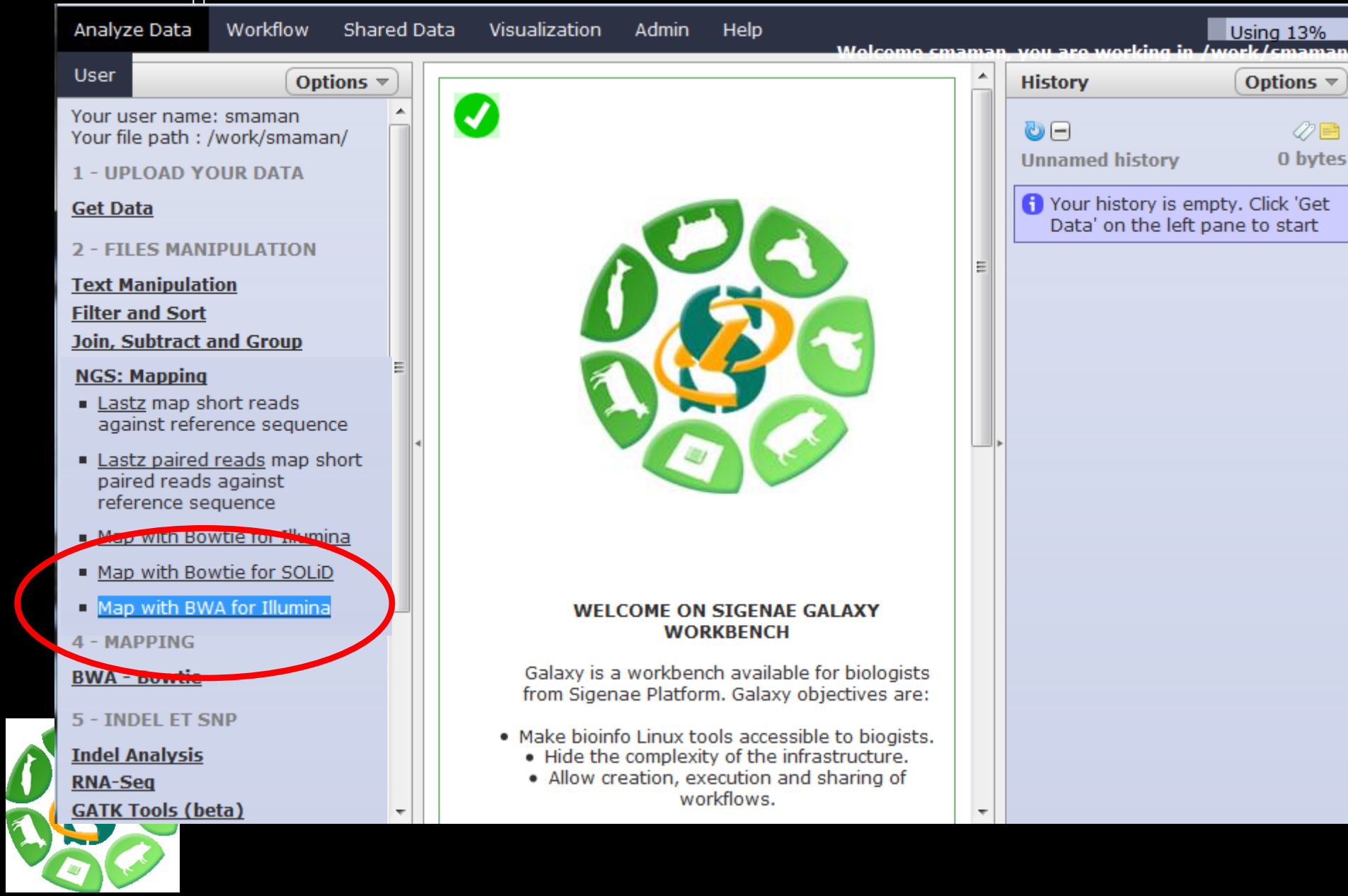
Exemple : Analyse en 3 clics sans utiliser votre disque dur !



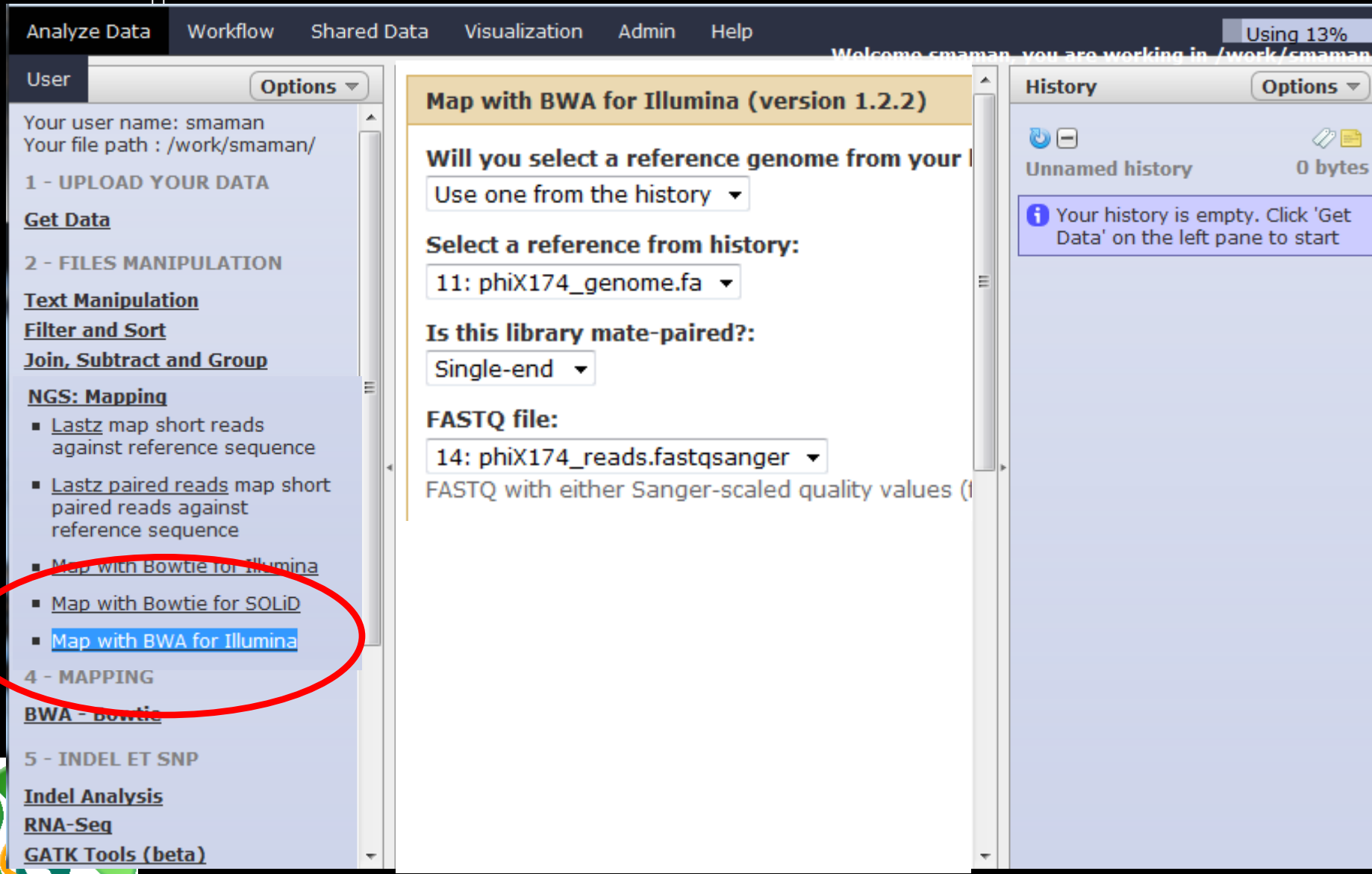
Exemple : Analyse en 3 clics sans utiliser votre disque dur !



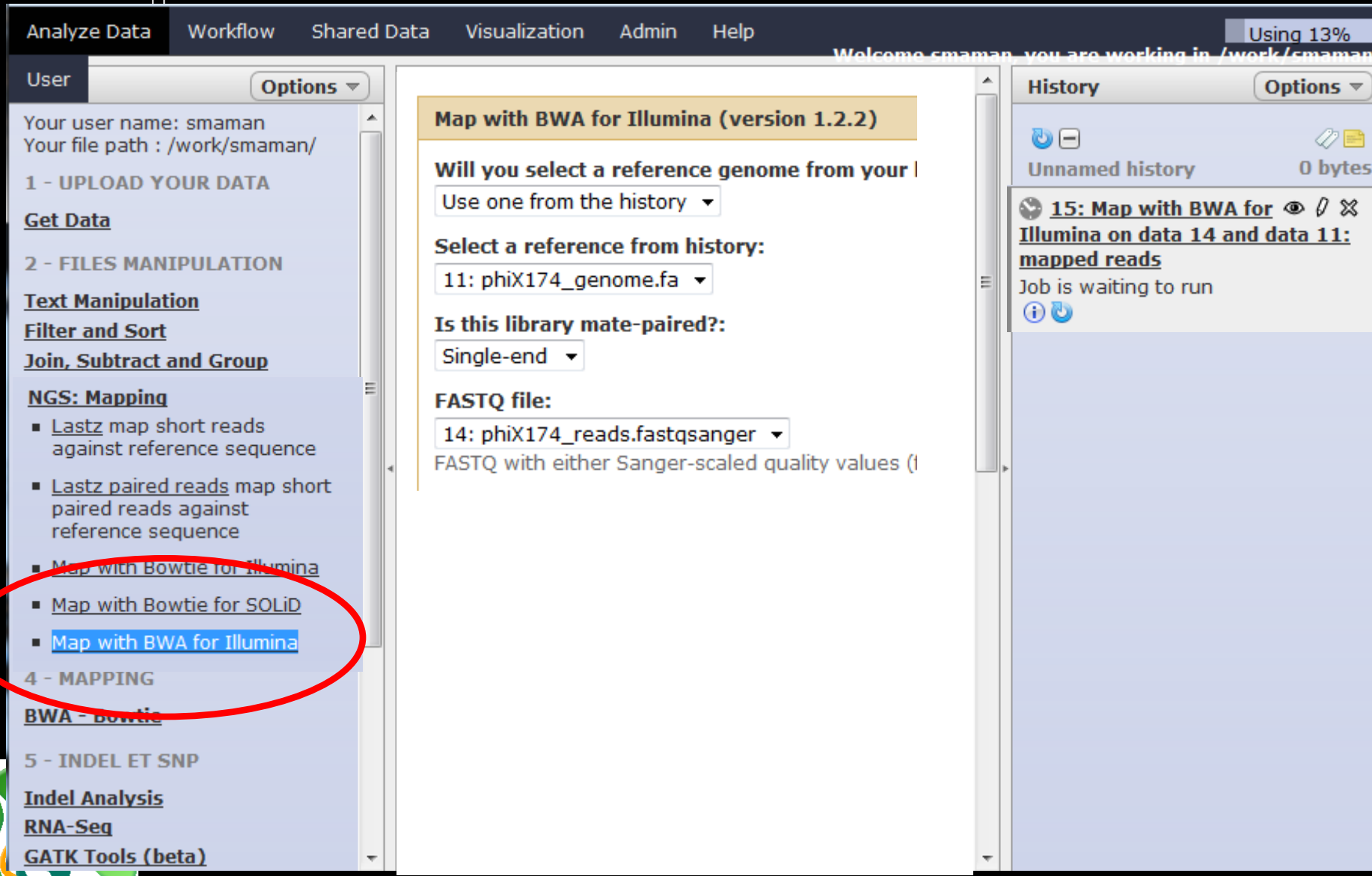
Exemple : Analyse en 3 clics sans utiliser votre disque dur !



Exemple : Analyse en 3 clics sans utiliser votre disque dur !



Exemple : Analyse en 3 clics sans utiliser votre disque dur !



Interface simplifiée

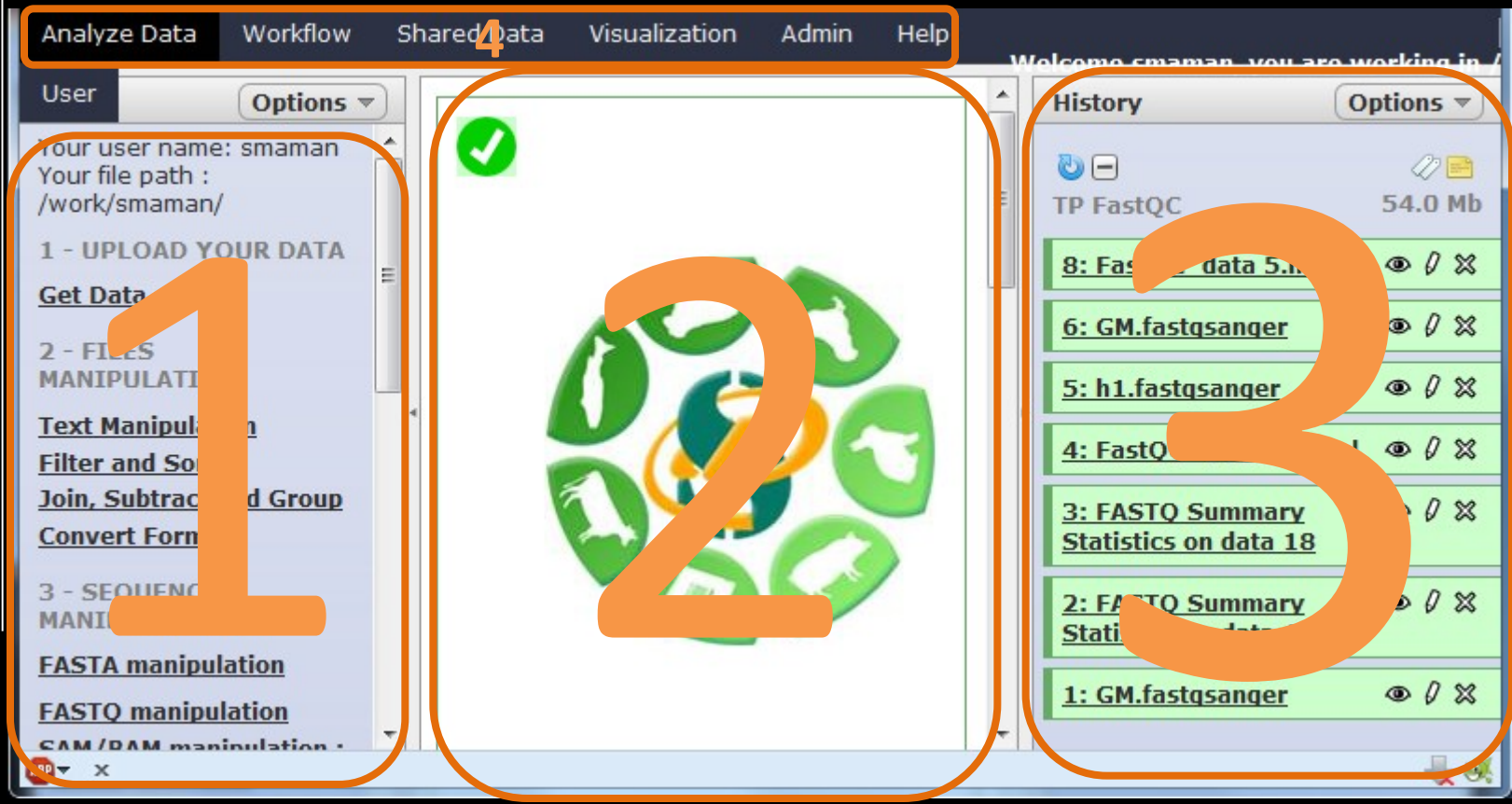
Interface divisée en 4 parties :

1 - Liste des outils disponibles.

2 – Visualisation de l'outil utilisé, historique ou workflow en construction.

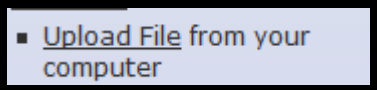
3 - Historique ou workflow détaillé.

4 – Menu .



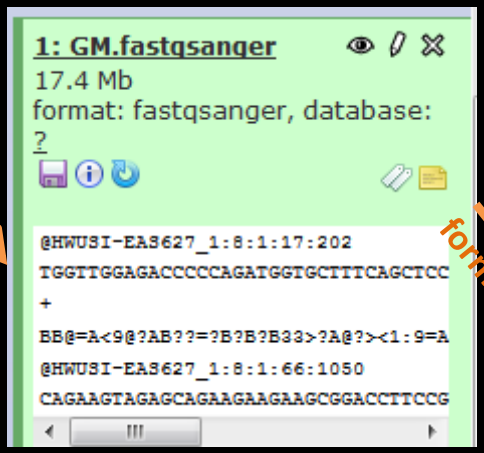
Le vocabulaire spécifique à Galaxy

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.



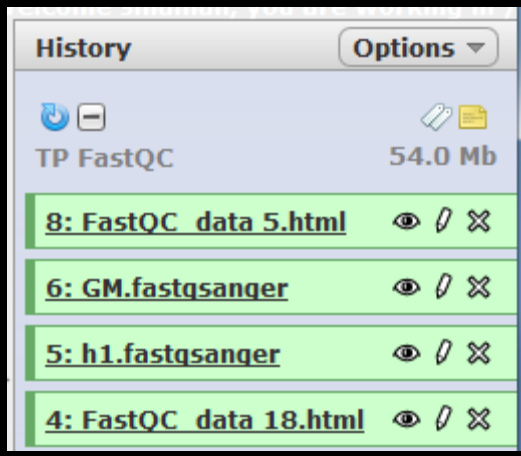
TOOL

génère



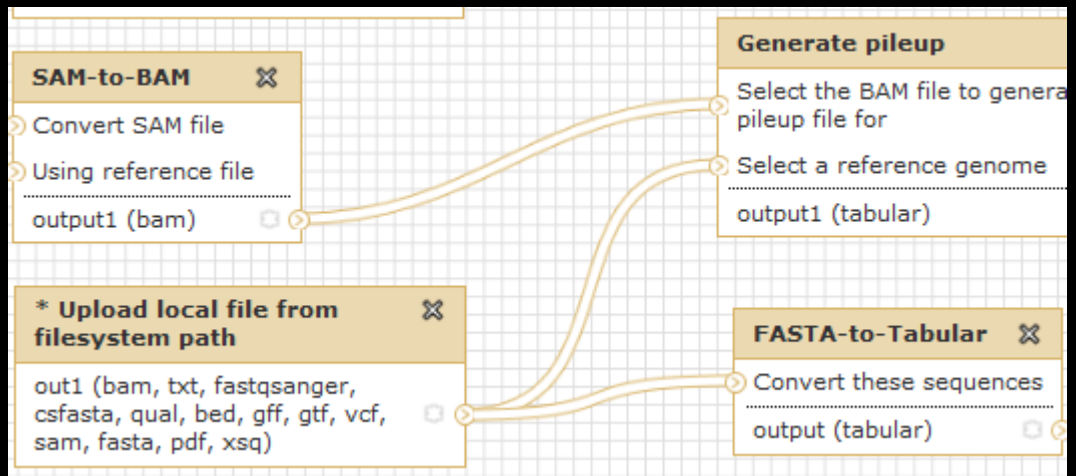
forme

Dont la liste



DATASET (S)

HISTORY



WORKFLOW

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

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.

Plateforme

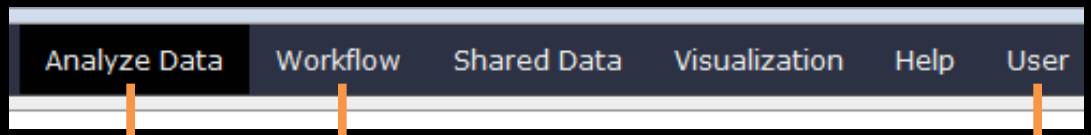
Vos données

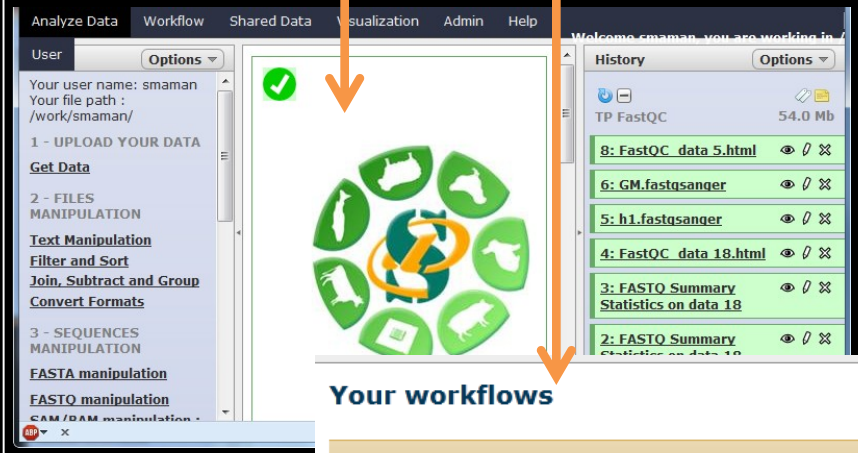
Historique

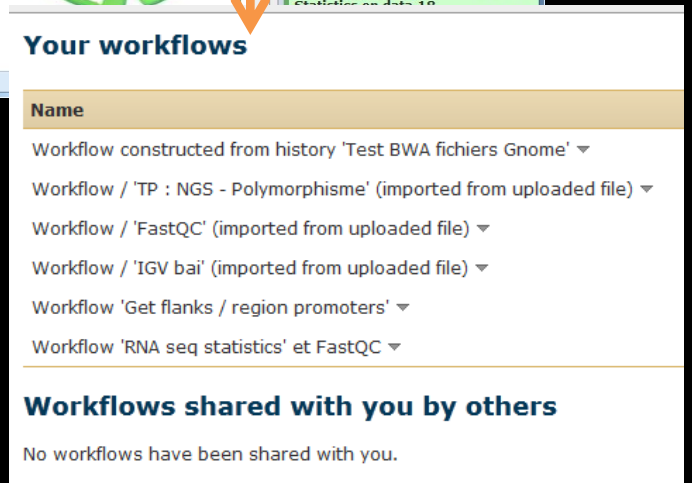
Workflow

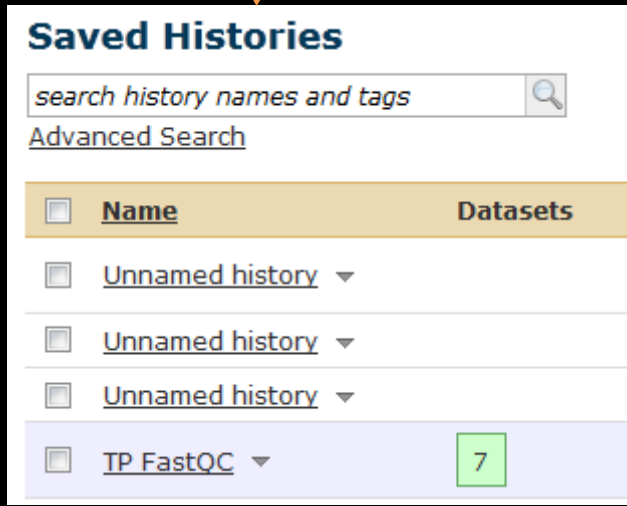
Bioinfo

Vous









En résumé …

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

Données

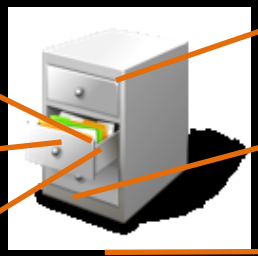
1 Dataset

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

Dataset 1

Dataset 2

Dataset 3



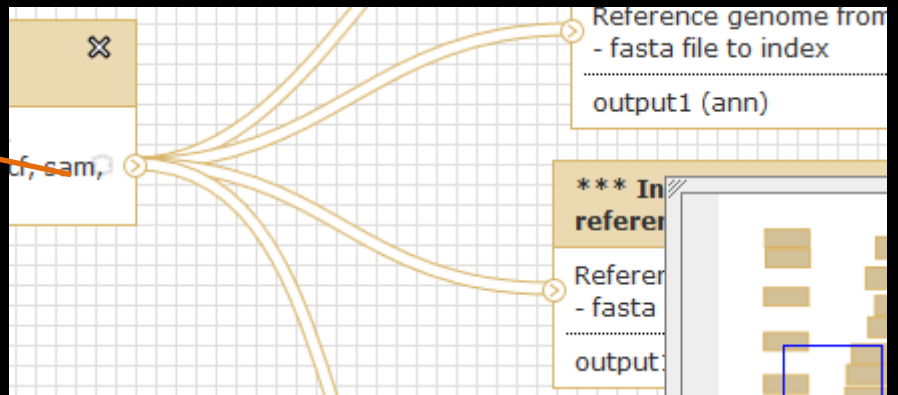
Historique 1

Historique 2

Historique 3

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

TOOL



Présentation de la plateforme Galaxy.

Comment récupérer vos données (privées et publiques) ?

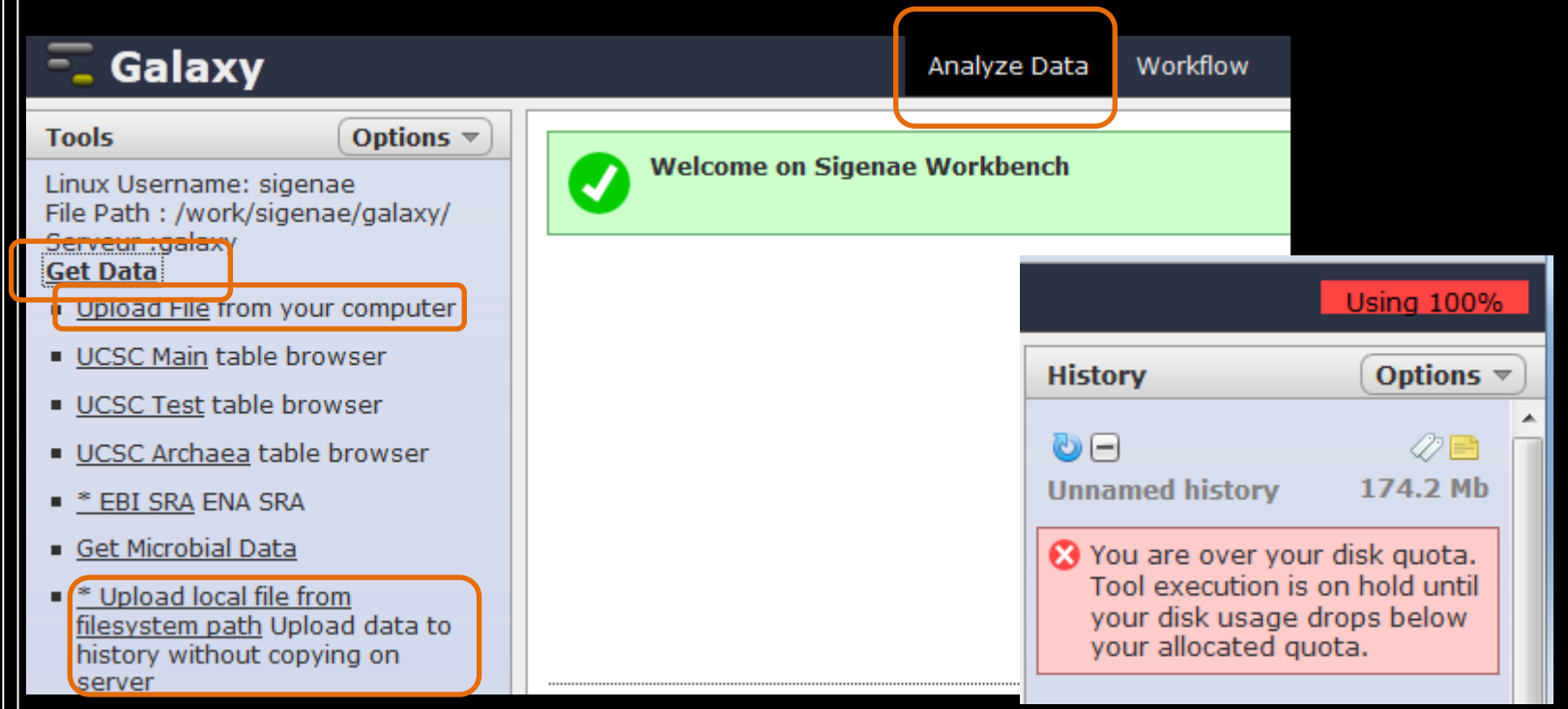
Notions d'outils, d'historique et de workflow.

Lancement de traitements bioinformatiques.

Guide pour les utilisateurs Galaxy.

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

2 méthodes de téléchargement de vos données privées



Solution 1 (recommandée si vos fichiers sont dans Genotoul) :

« Upload local file from filesystem path ».

Solution 2 :

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

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

Téléchargement de données publiques

Données UCSC, Ensembl, NG6, BIOMART :

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

Genomes | Genome Browser | Tools | Mirrors | Downloads | My Data | About Us | Hel

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

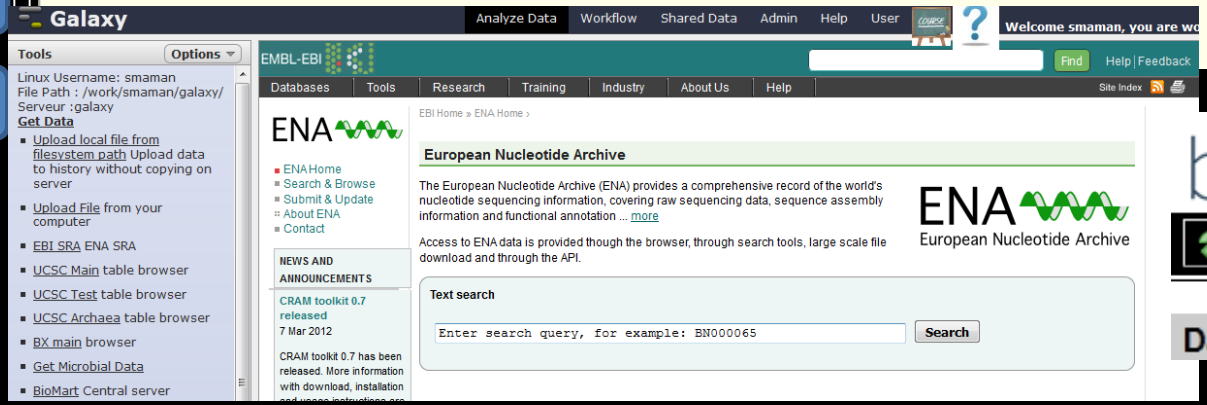
intersection: create

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



Dataset

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

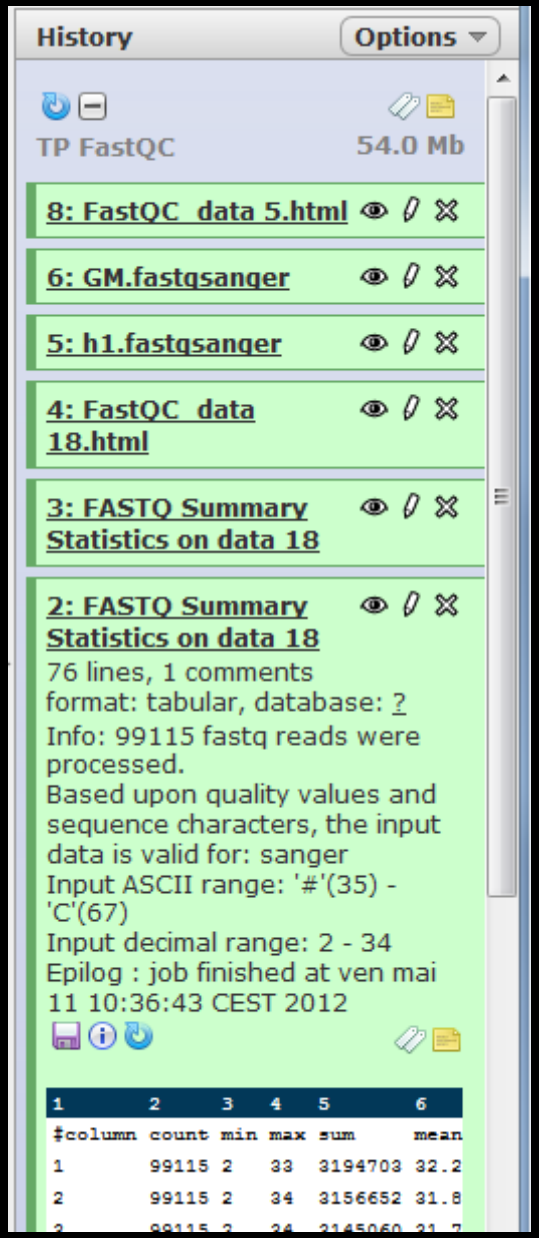
Workflow

Bioinfo

Vous

Gestion de vos historiques

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

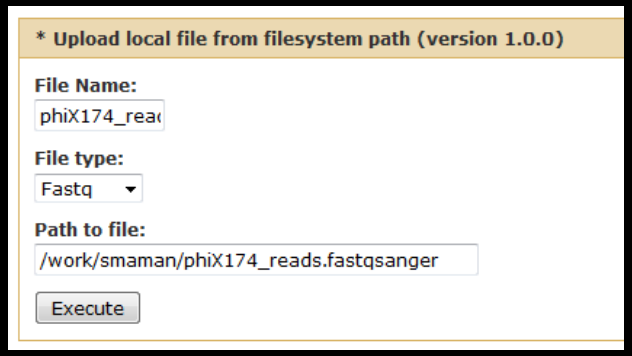


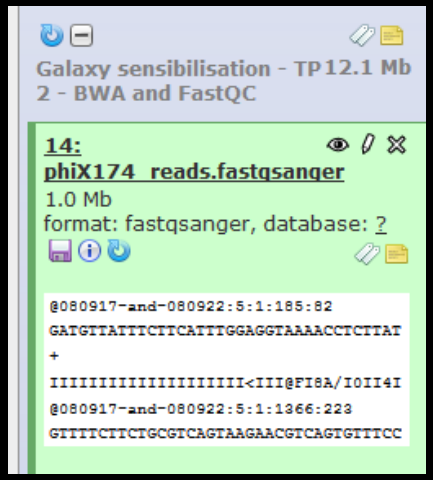
- Conserver toutes les étapes de vos analyses .
- Partager vos analyses.
- A chaque run d'un outil, une nouvelle dataset est créée. Les données ne sont pas écrasées.
- Répéter, autant de fois que nécessaire, une analyse.

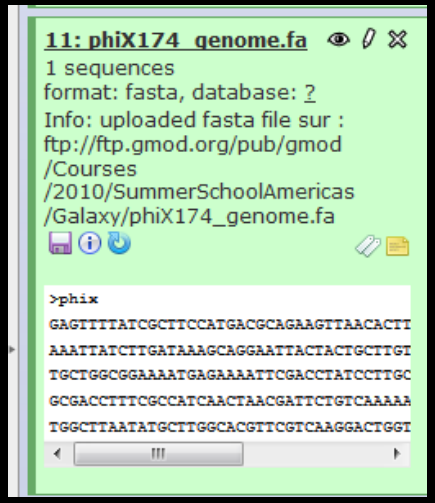
Comment lancer un job sans ligne de commande ?

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

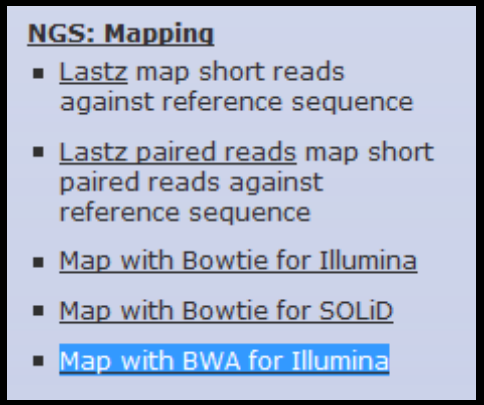
1 – Télécharger vos données :

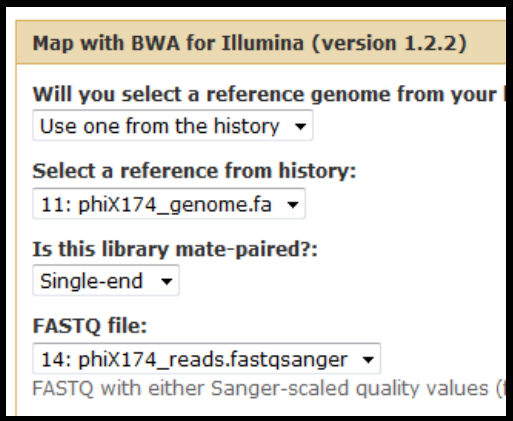






2 – Choisir un outil dans « Tools » :

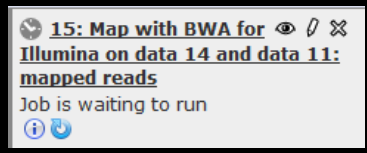




3 – Lancer le job en cliquant sur « Executer ».

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

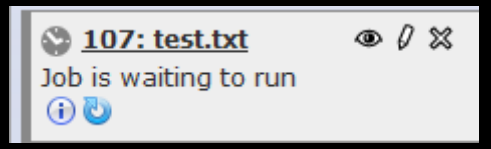
Fini les lignes de commande !



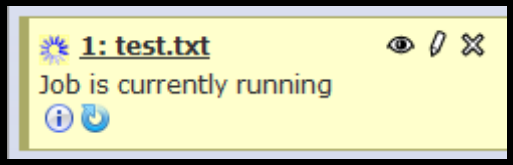
Comment suivre l'execution de mes jobs

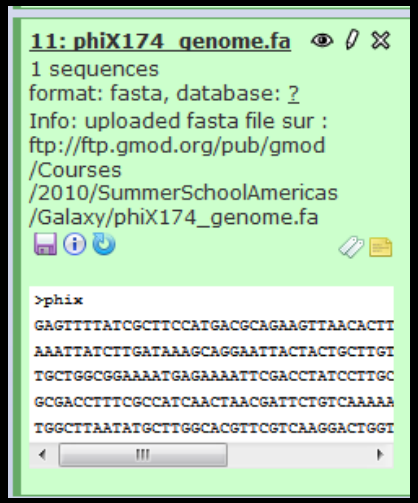
- Plateforme
- Vos données
- Historique
- Workflow
- Bioinfo
- Vous

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

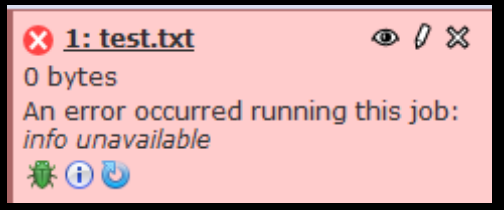


Etat 2 – JAUNE : Votre job est en cours d'execution





Bug - ROUGE : Votre job est planté !



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

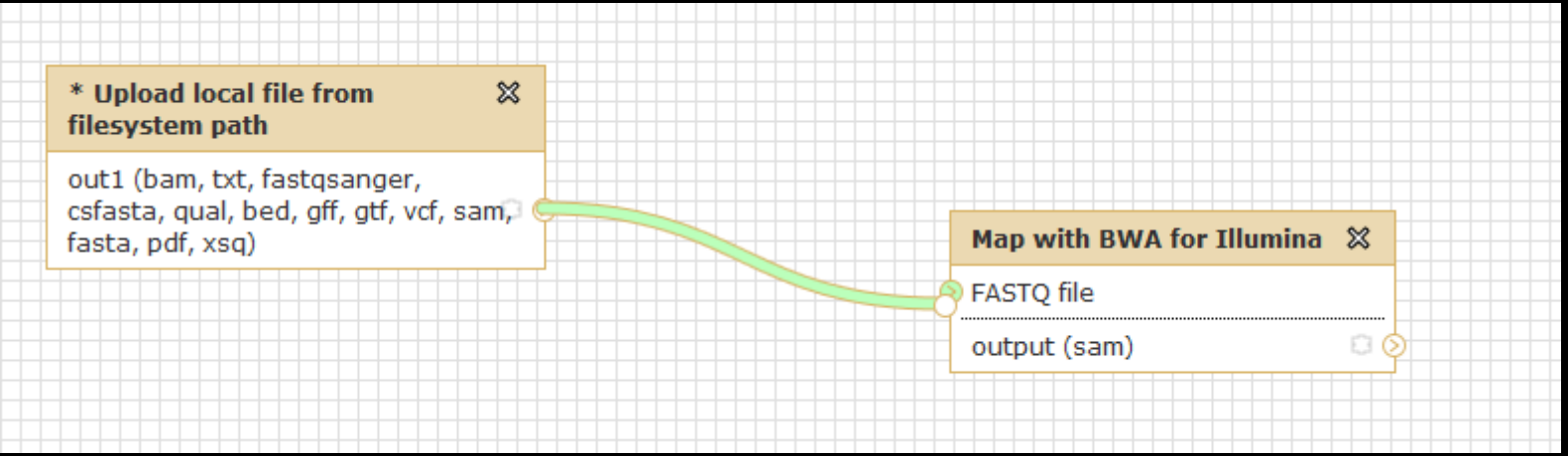
Bioinfo

Vous

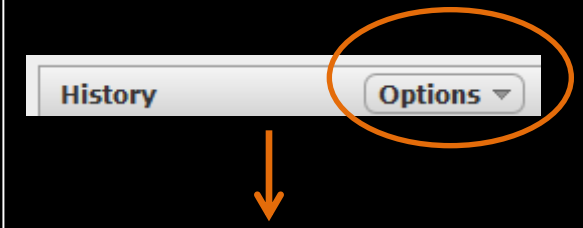
Créer un workflow

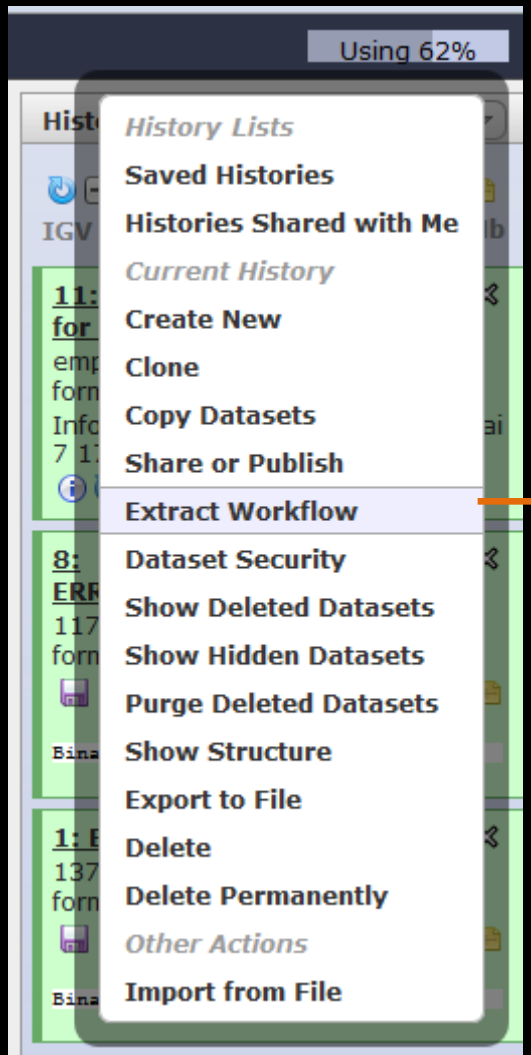
Depuis une page blanche, vous pouvez concevoir un workflow.

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

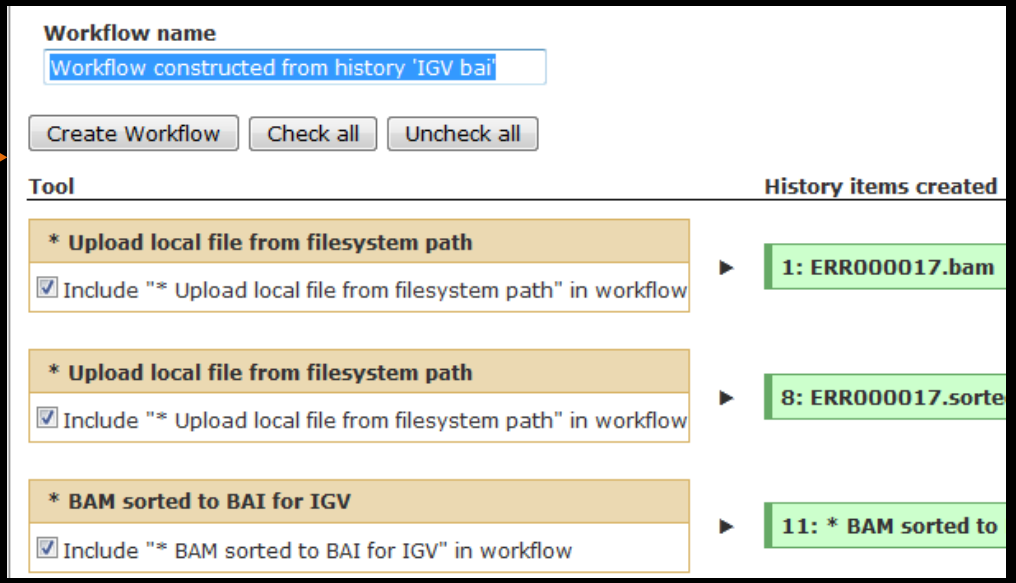


Exporter votre historique en workflow.





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



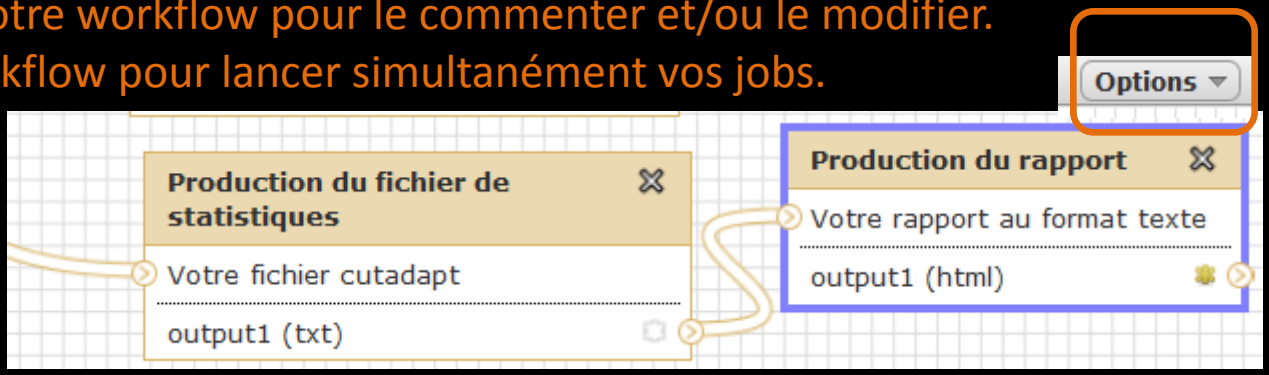
- Plateforme
- Vos données
- Historique
- Workflow
- Bioinfo
- Vous

Lancer plusieurs jobs simultanément

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

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

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

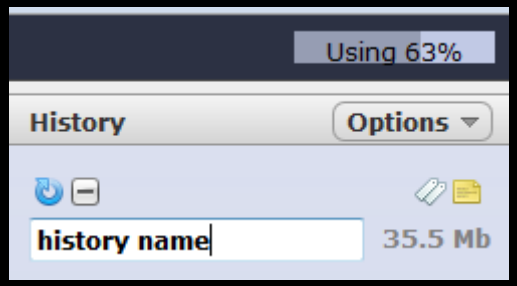


En résumé ….des datasets aux workflows.

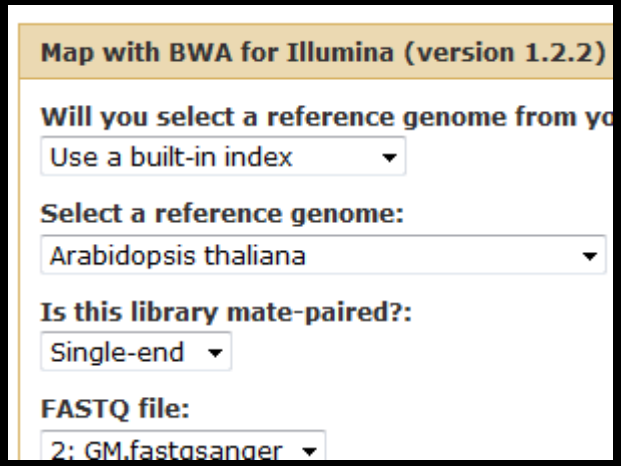
- Plateforme
- Vos données
- Historique
- **Workflow**
- Bioinfo
- Vous

1 – Télécharger vos fichiers de données

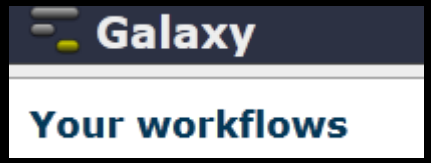
2 – Renommer votre historique :



3 – Utiliser les outils dont vous avez besoin :



4 – Votre historique est sauvegardé automatiquement.
Si nécessaire, exporter votre historique en workflow.



5 – Editer, partager et lancer vos traitements à volonté (run de votre workflow).

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

Les principaux outils Galaxy

Plateforme

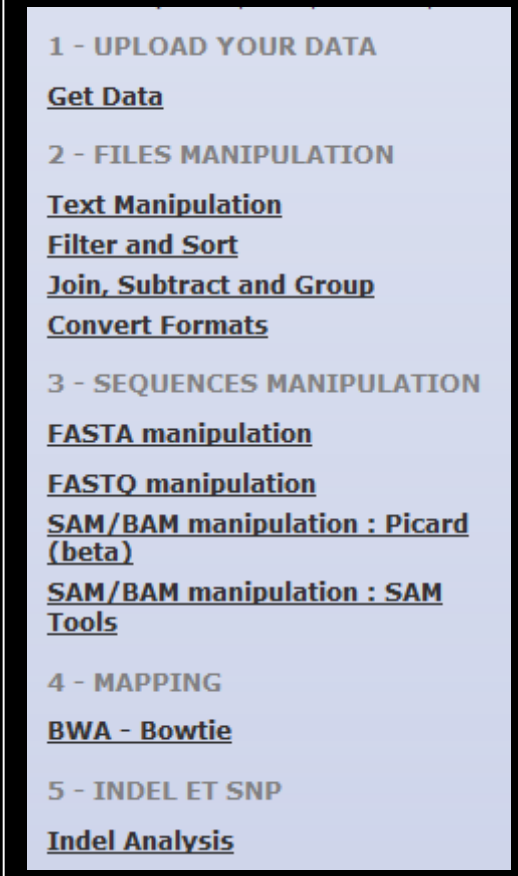
Vos données

Historique

Workflow

Bioinfo

Vous



GET DATA :
Télécharger vos données privées.
Télécharger des données publiques : UCSC, Ensembl, Biomart …

FILES MANIPULATION :
Manipulation de fichiers texte ou autres : Couper, coller, comparer, soustraire, merger, concatener, selectionner, filtrer, trier, convertir, grouper …

SEQUENCE MANIPULATION :
Travailler sur des fichiers FASTA et FastQ, analyse qualité FasQC, Picard tools et samtools.

MAPPING :
BWA, Bowtie, indexation de génome.

Autres :
Recherche d'indel, SNP, RNAseq (TopHat, Cufflinks).

Autres outils en cours de tests et d'ajout .. Selon vos besoins.

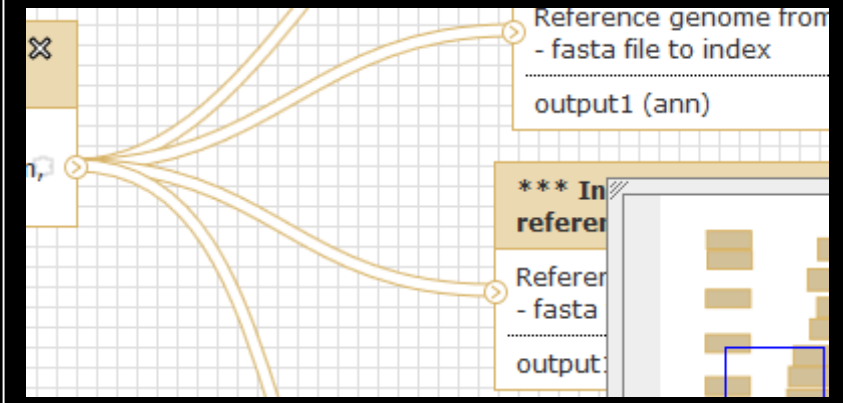
Comment citer Galaxy dans vos publications ?

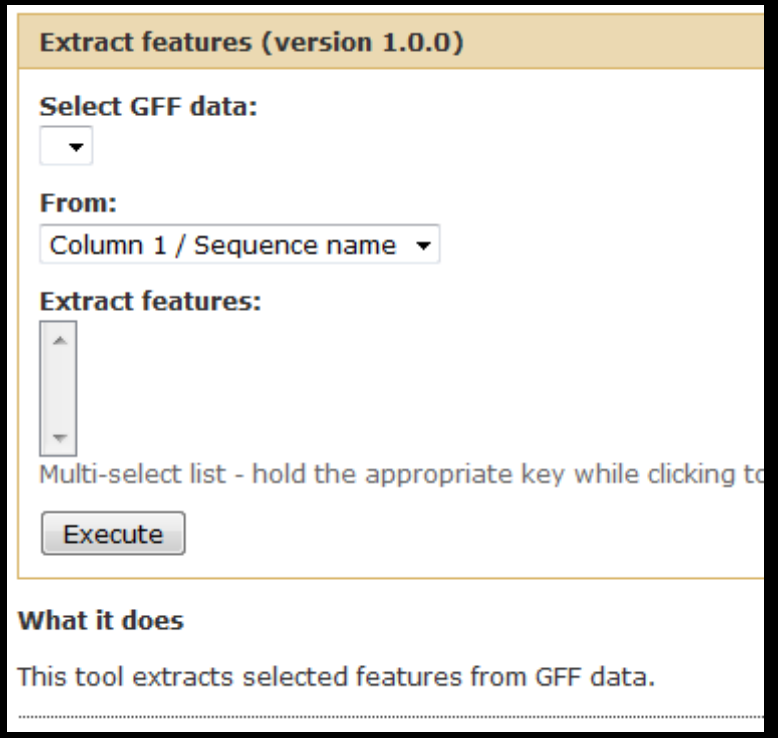
Pour vos publications, citer:

✓Les outils utilisés (nom, version).

✓Le workflow généré.

✓Les références « Galaxy project ».





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

Outils de traitement de fichiers

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

Les traitements sont automatiquement lancés sur GENOTOUL (qsub).

Voici les principaux outils « non bioinfo » proposés :

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

Select first (version 1.0.0)

Select first:

10

lines

from:

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

Execute

What it does

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

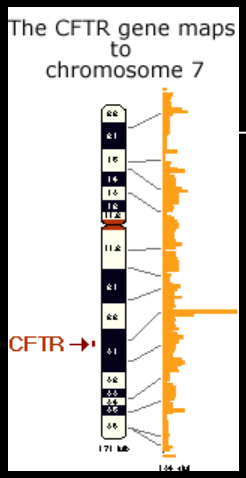
Example

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   +
```



http://www.genomenewsnetwork.org/

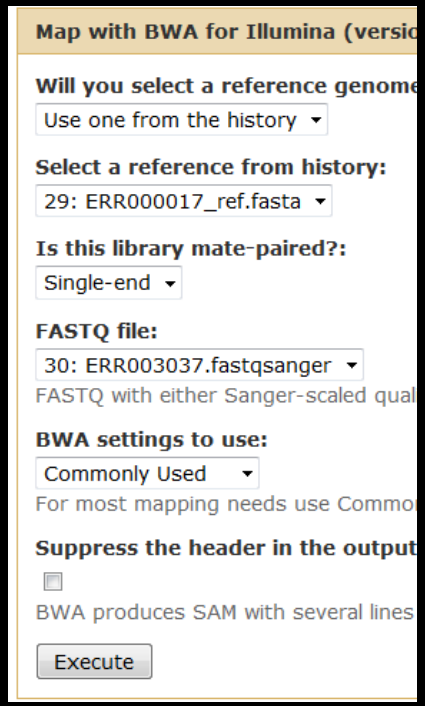
Outils bioinformatiques

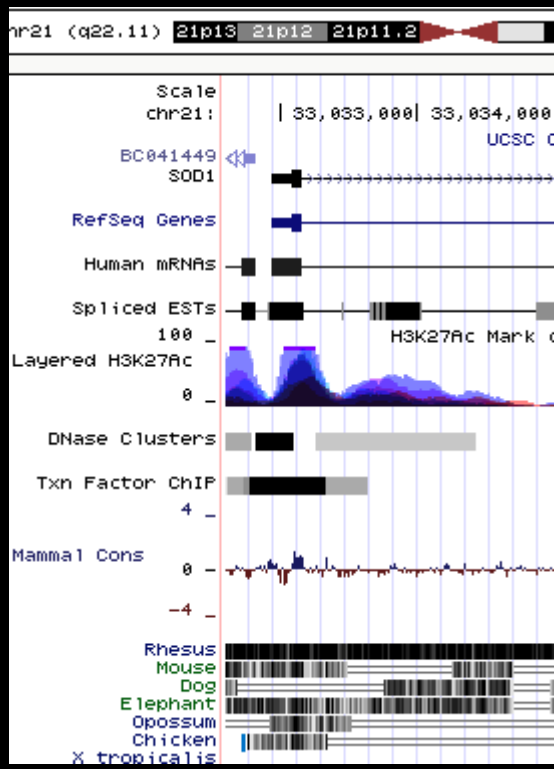
- Plateforme
- Vos données
- Historique
- Workflow
- **Bioinfo**
- Vous

Mapper un FASTQ sur une référence avec BWA.

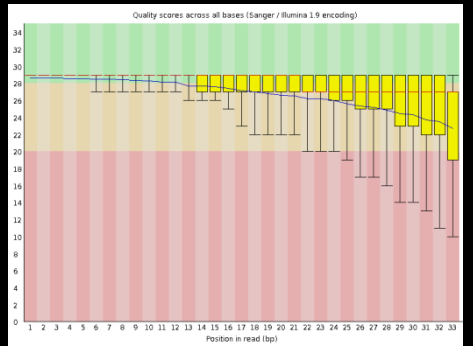
NGS: Mapping

- 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





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 …

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

Par exemple :

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

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

Galaxy

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

FAQ et formation en ligne

Plateforme

Vos données

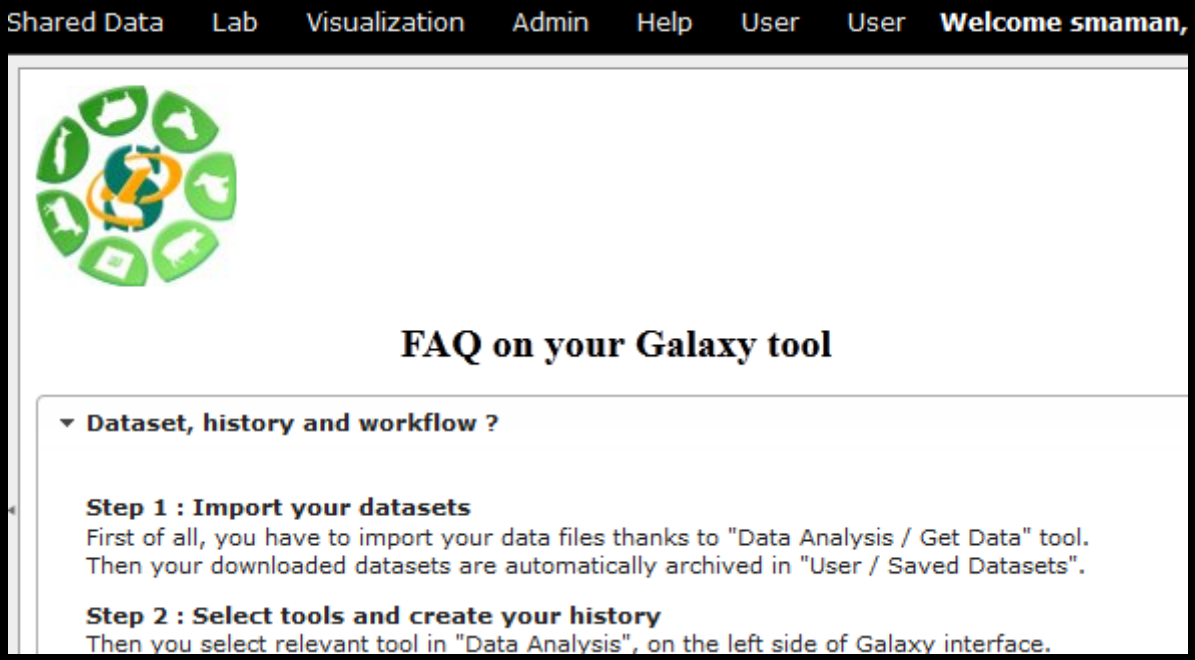
Historique

Workflow

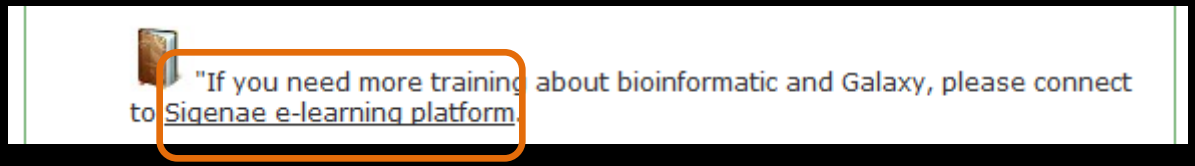
Bioinfo

Vous

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



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



"How to be a good user ?":

- Maîtrise de son quota
- Organisation de son espace de travail
- Demandes via des tickets Redmine ou mail à sigenae-support@listes.inra.fr

En conclusion …

GALAXY

✓Simplicité d'utilisation (sans Linux).

✓Partage de vos datasets, historiques et workflows.

✓Présentation schématique de vos traitements grâce aux workflows.

✓Possibilité d’ajout de nouveaux outils, selon vos besoins.

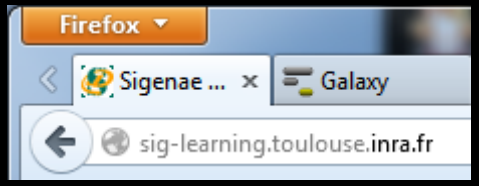
Votre auto-formation continue en ligne avec « sig-learning »

Votre accès à sig-learning

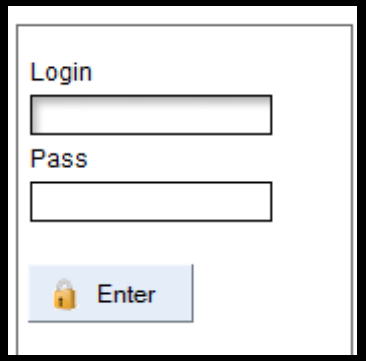
Plateforme

Vos formations

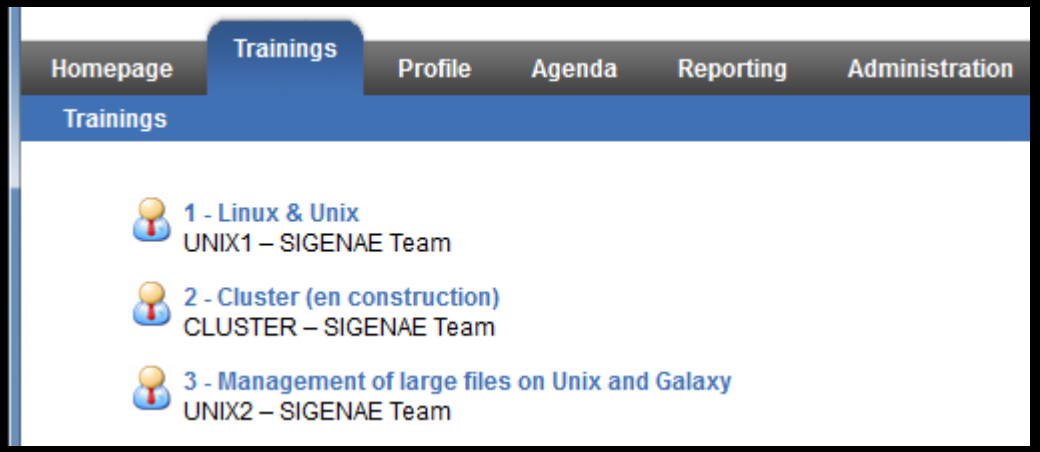
1 Taper l'adresse de « sig-learning » : **http://sig-learning.toulouse.inra.fr/**



2 Authentification

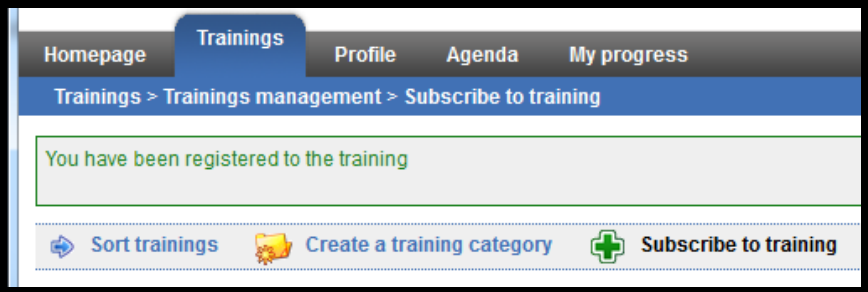


3 Accès à la liste des formations auxquelles l'utilisateur s'est inscrit.



Comment s'inscrire à des formations ?

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

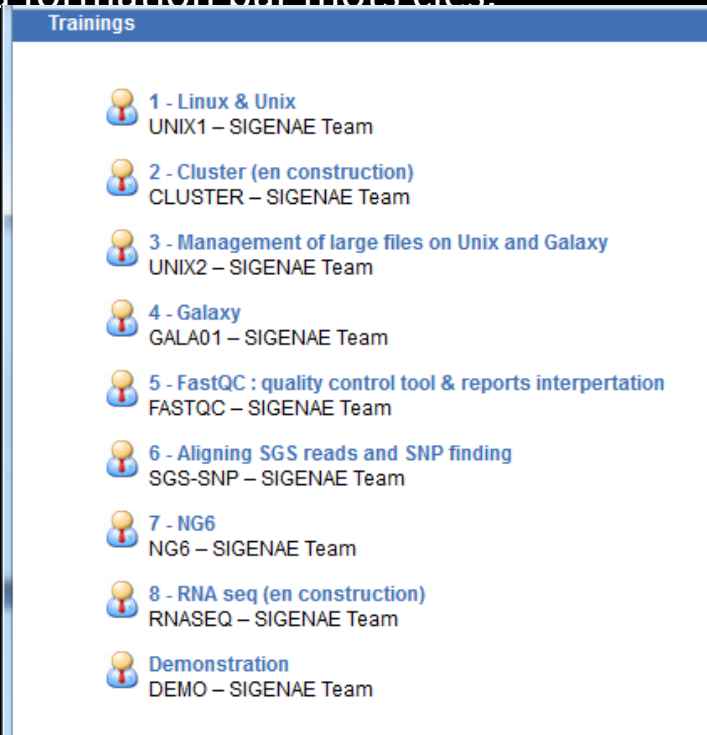


Plateforme

Vos formations

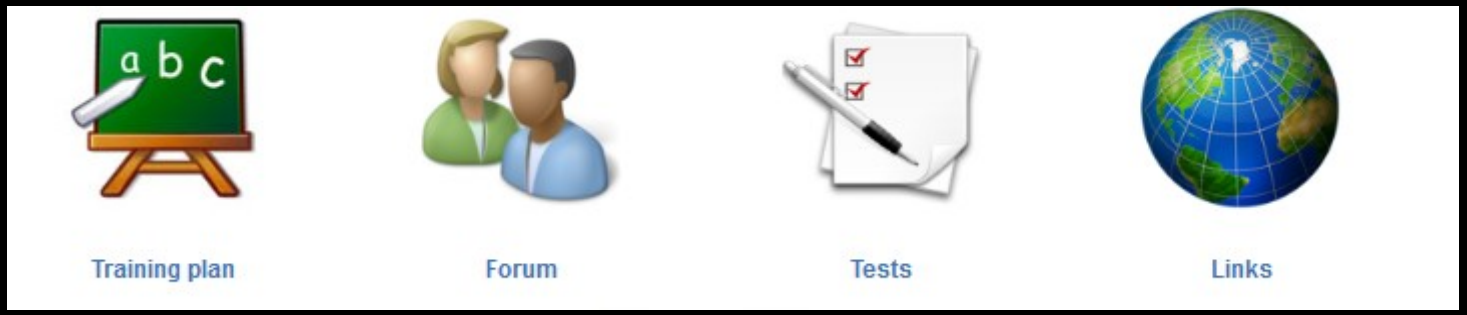
L'inscription s'effectue via une recherche de la formation par mots clés.
Voici donc la liste des formations

(disponibles au 01/2013):



Page d'accueil de votre formation

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



Plateforme

Vos formations

TRAINING PLAN : Parcours pédagogique avec les supports en ligne.

FORUM : support de communication entre stagiaires / formateurs.

TESTS : Tests et exercices.

LINKS : Liens utiles.

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

1 – Demande à compte sur la plateforme BIOINFO GENOTOUL :
http://bioinfo.genotoul.fr/index.php?id=81
Ou :
bioinfo.genotoul.fr puis «menu « Help » , puis « Create an account ».
Vous recevrez un login et mot de passe LDAP Genotoul.

Plateforme

Vos formations

2 – Puis utilisez ce login et mot de passe LDAP Genotoul lorsque vous souhaitez accéder à :

.Instance Sigenae de Galaxy : http://sigenae-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