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

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

Philippe Bardou – Sarah Maman – Sabrina Legoueix-Rodriguez
Novembre 2015

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.

Quelques statistiques.

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.

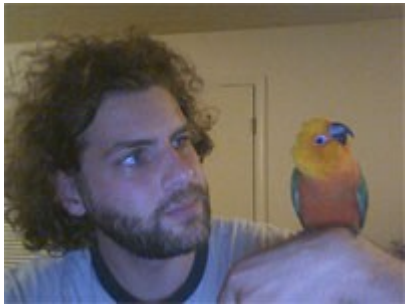
Une communauté active autour de cet outil.

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 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 possible de nouveaux outils / scripts / génomes …

→ **Présentation des particuliarités 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 matin.
- ✓

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.

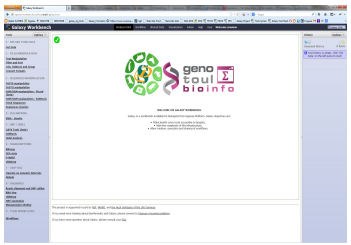
✓Construction de **chaînes de traitement.**

✓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

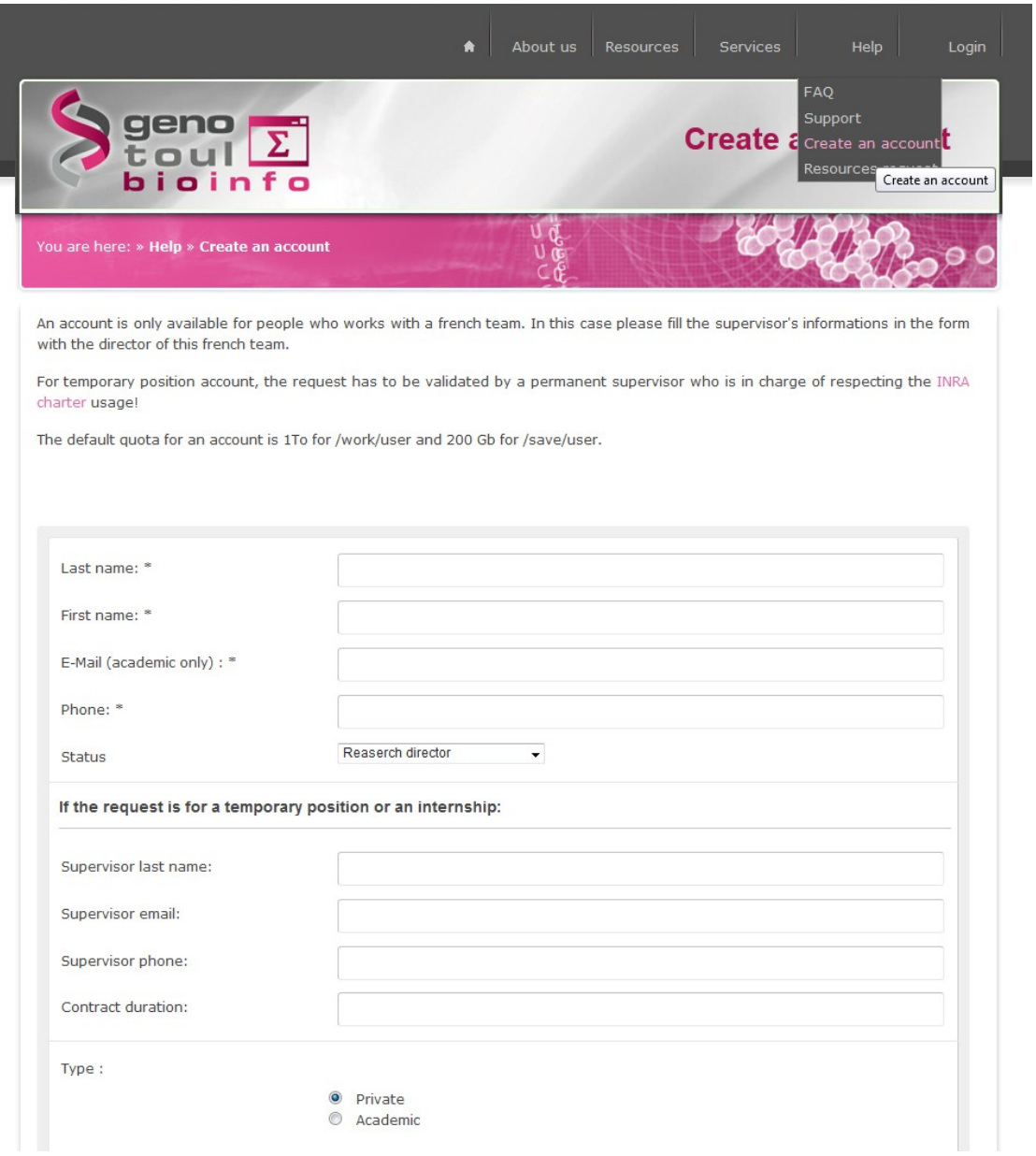
Gère la file d'attente

Gestionnaire de tâches

Exécute

Cluster de calculs

Comment ouvrir un compte sur Genotoul ?



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

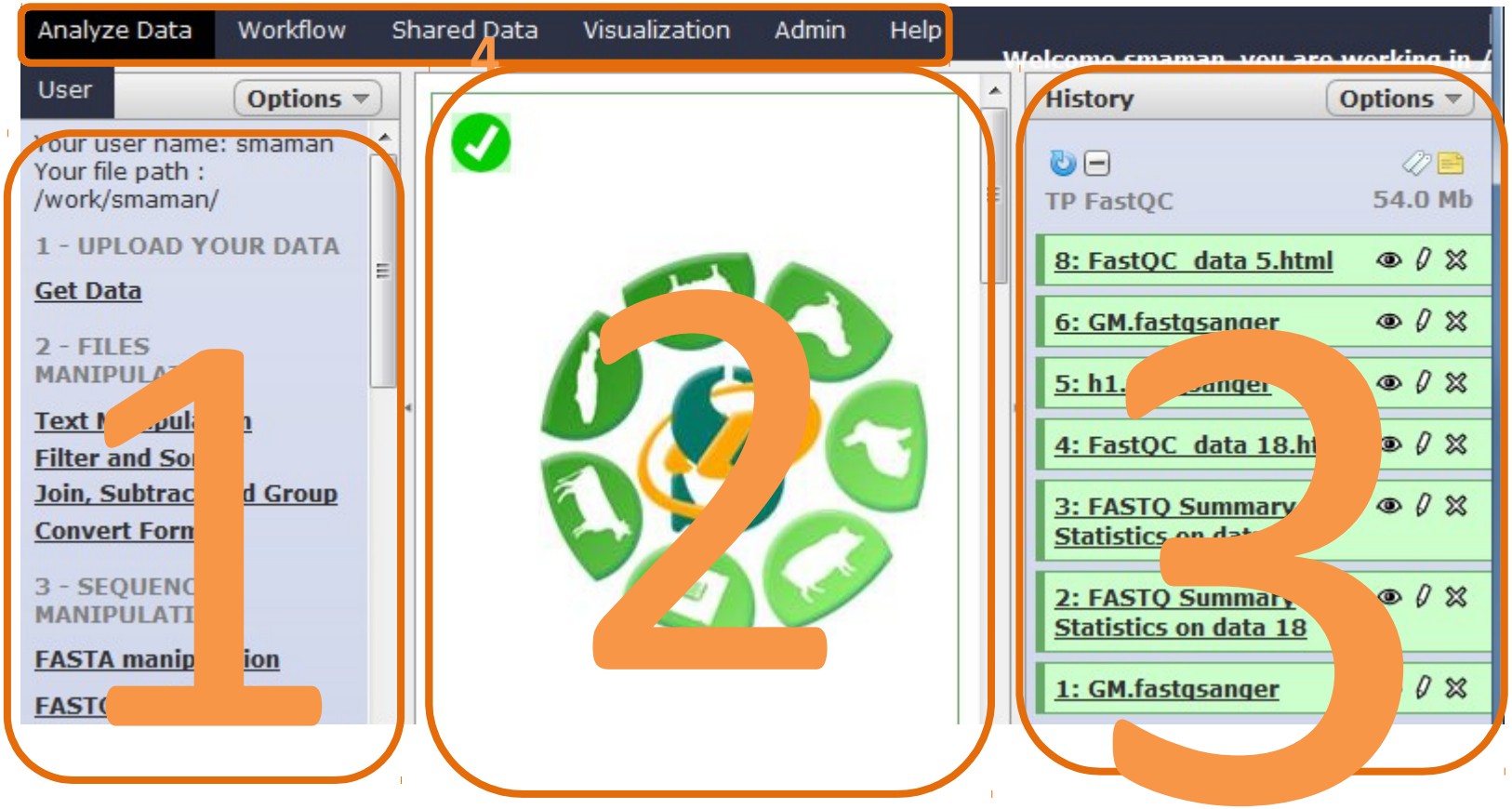
Interface divisée en 4 parties :

1 - Liste des outils disponibles.

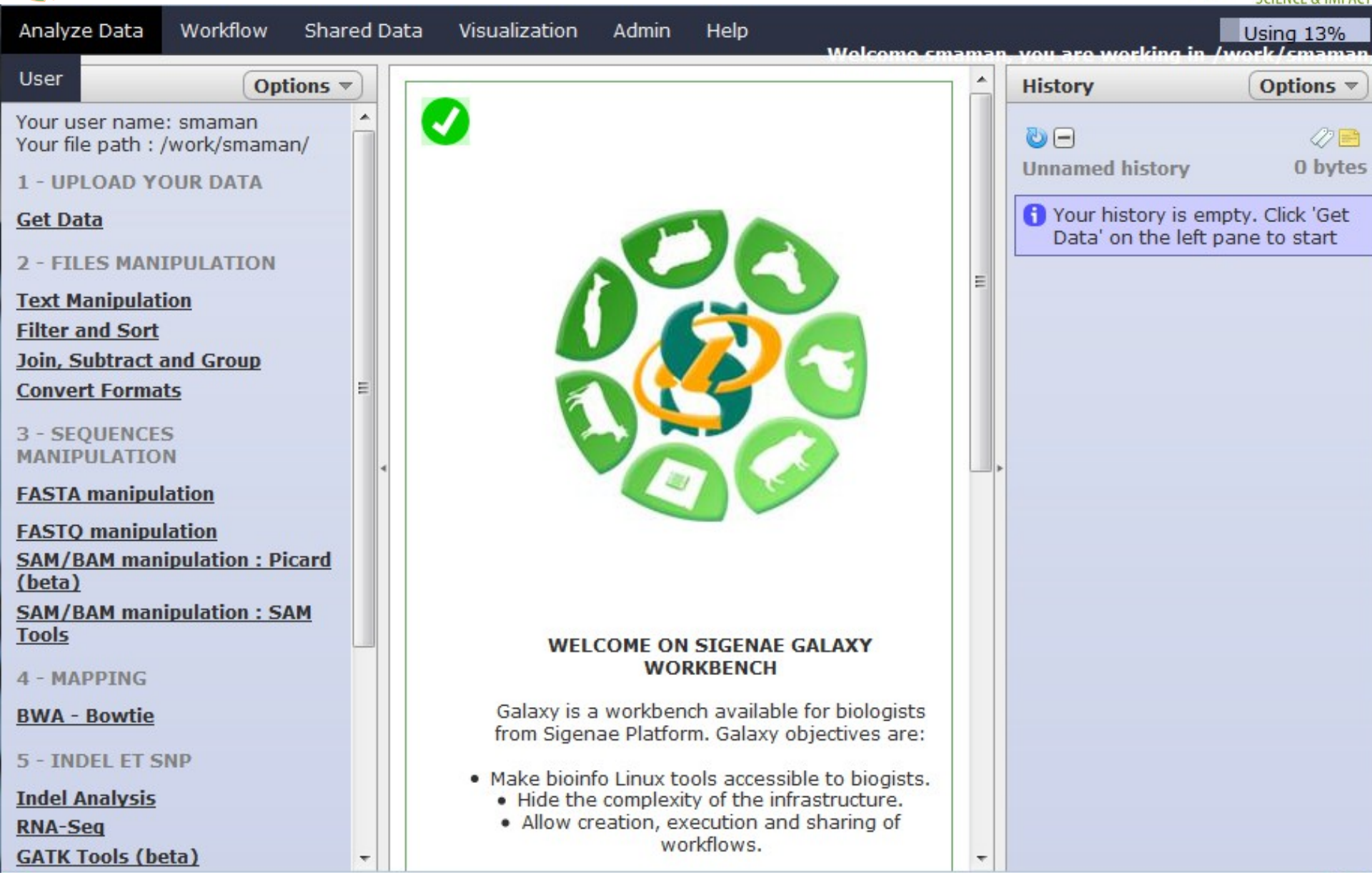
2 - Visualisation de l'outil utilisé, historique, dataset ou workflow.

3 - Historique ou workflow détaillé.

4 - Menu .



Votre analyse en quelques clics



Votre analyse en quelques clics

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:

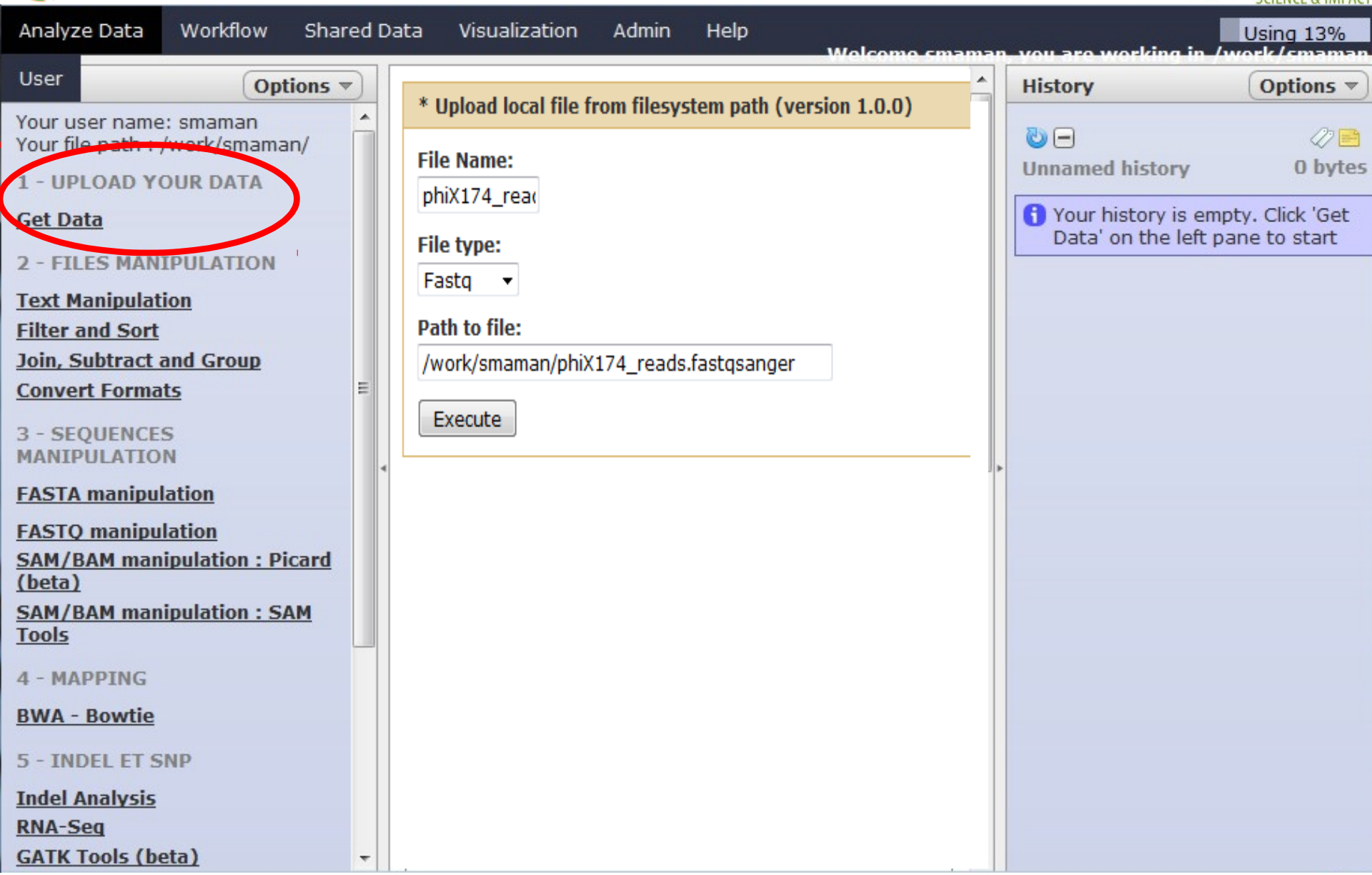
- Make bioinfo Linux tools accessible to biogists.
- Hide the complexity of the infrastructure.
- Allow creation, execution and sharing of workflows.

History | Options

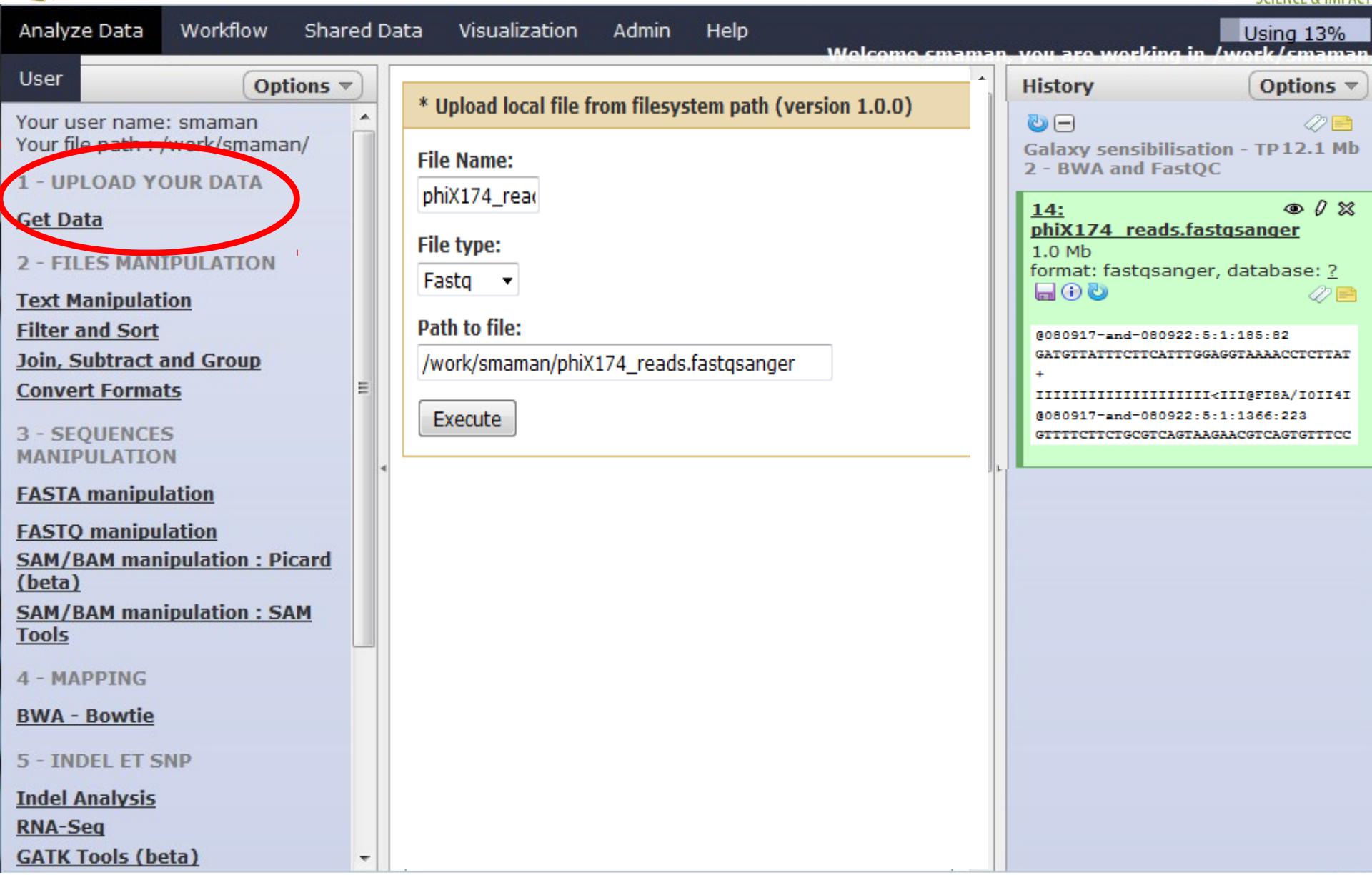
Unnamed history | 0 bytes

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

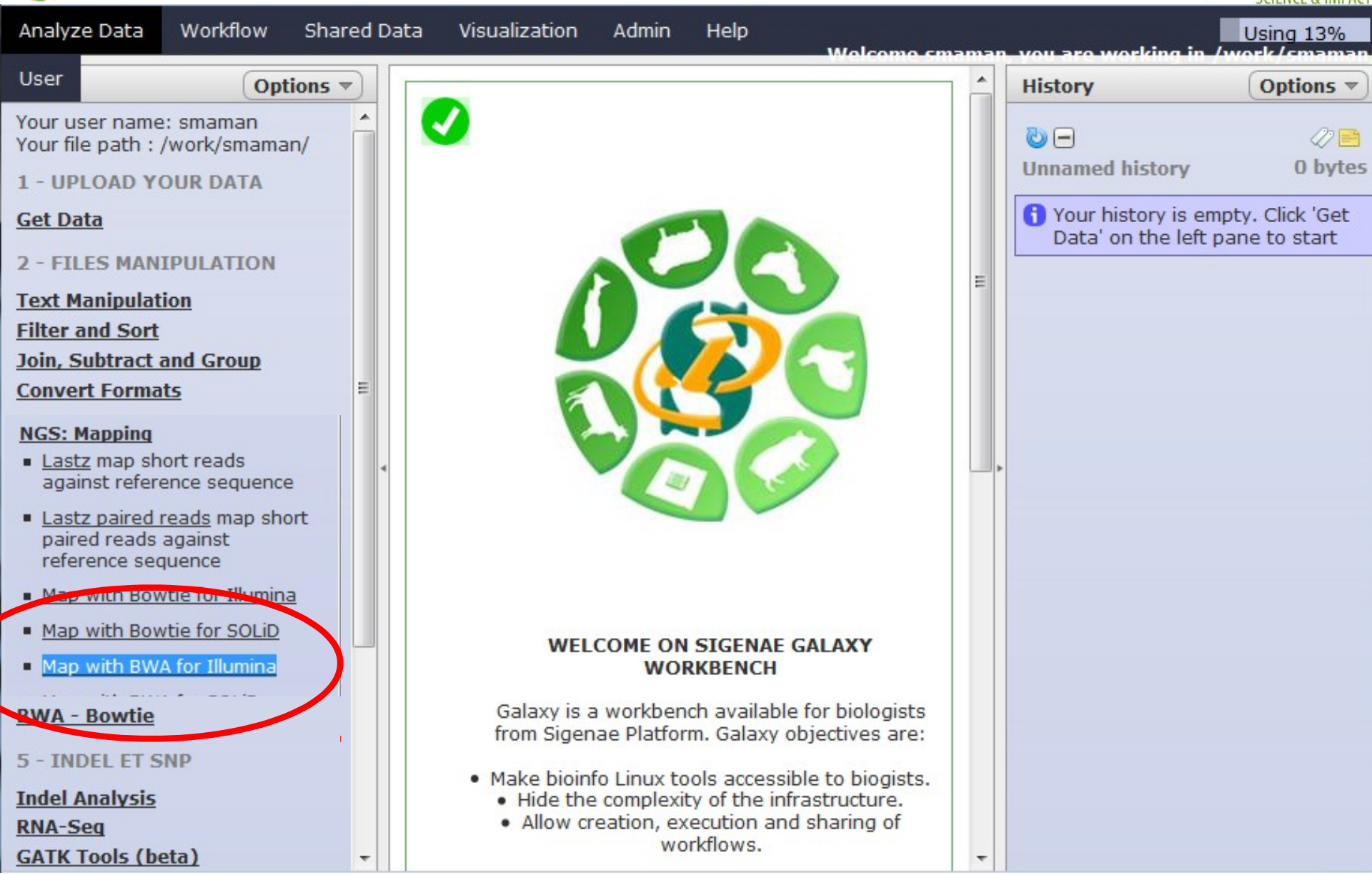
Votre analyse en quelques clics



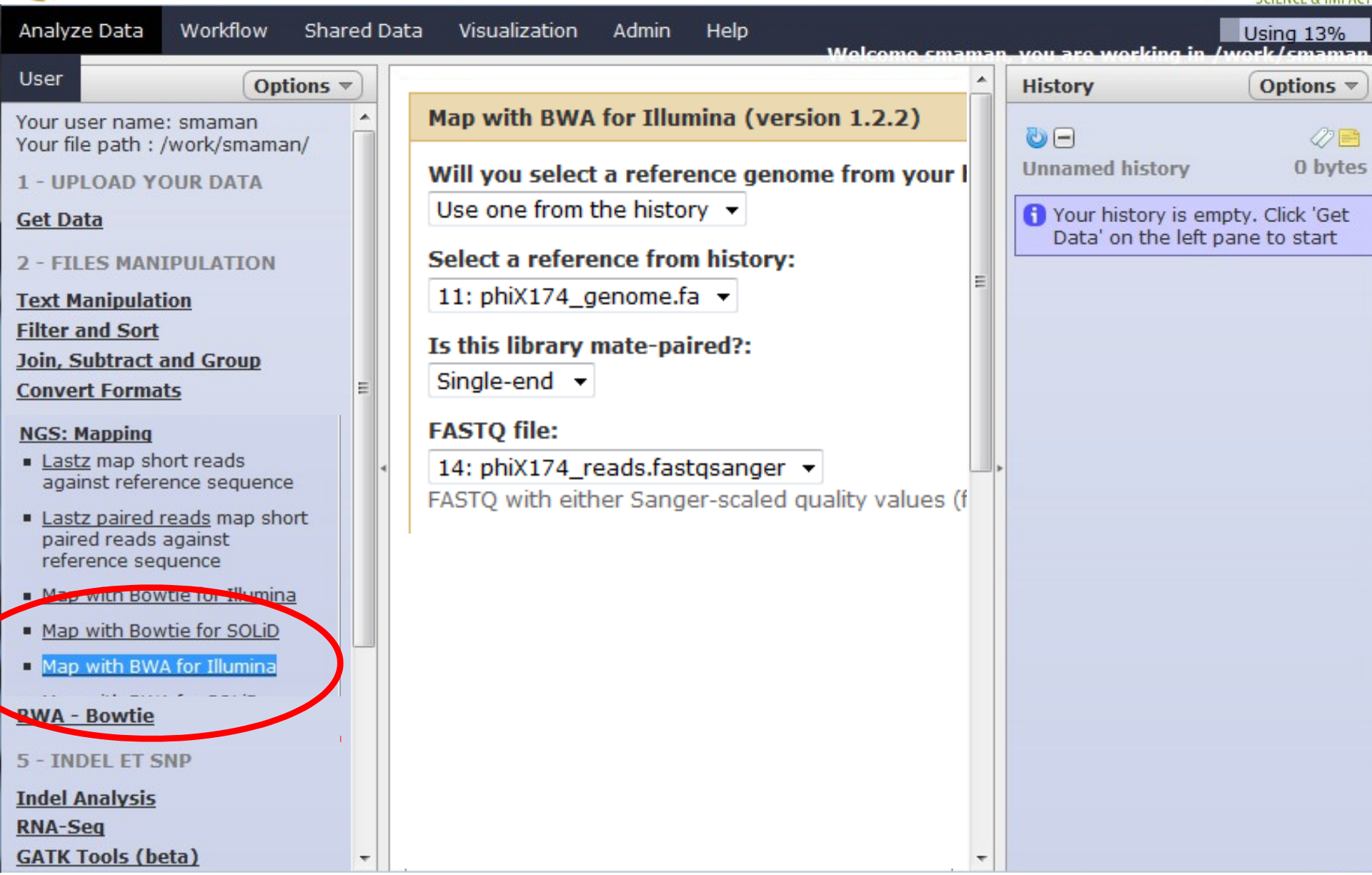
Votre analyse en quelques clics



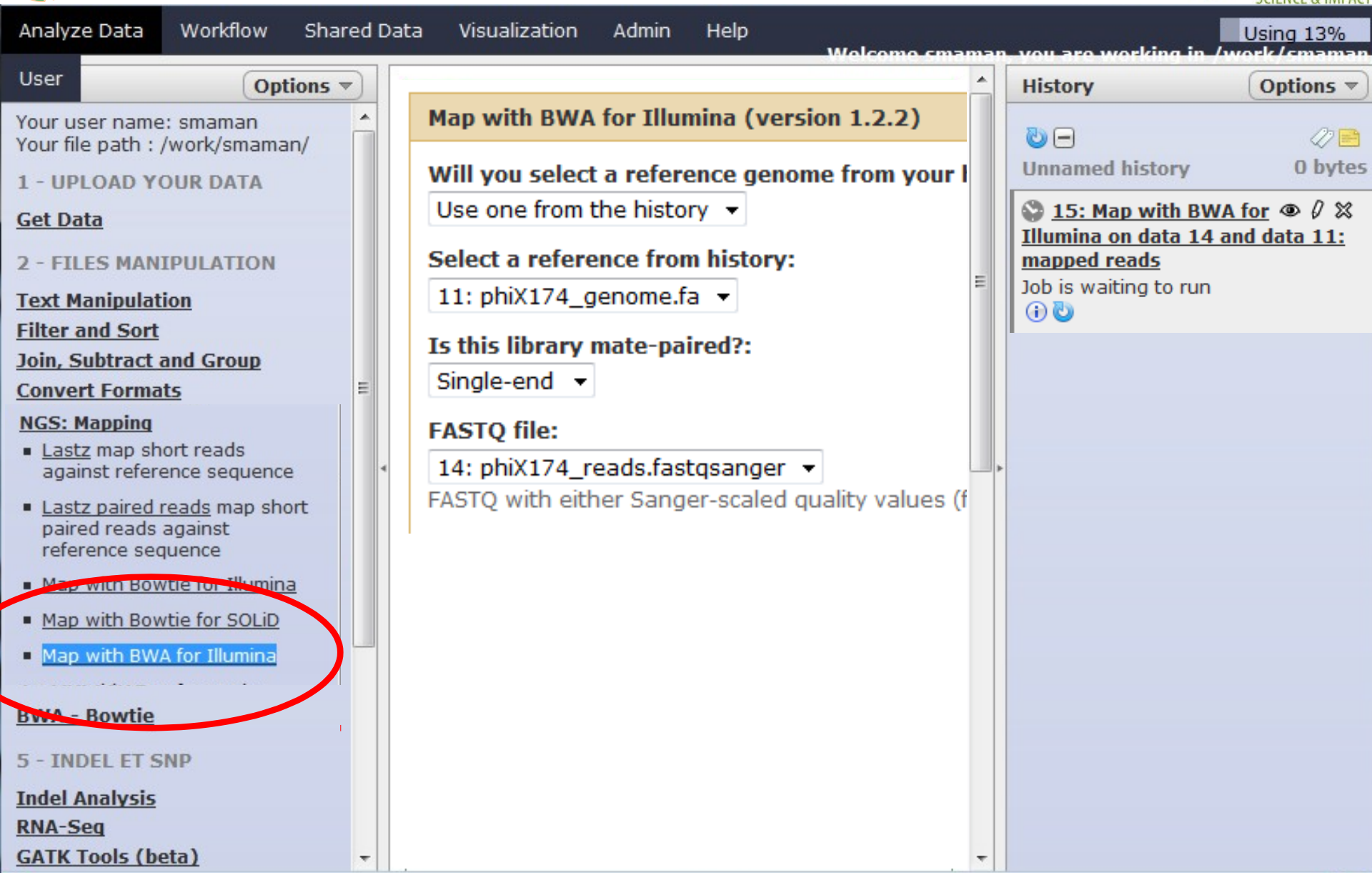
Votre analyse en quelques clics



Votre analyse en quelques clics

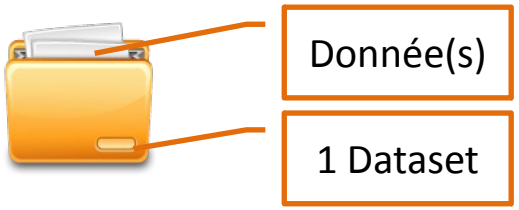


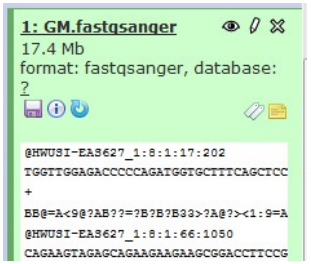
Votre analyse en quelques clics



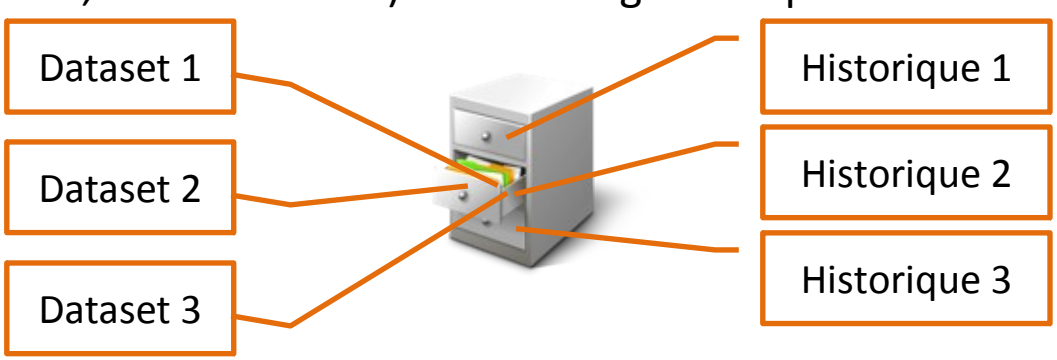
Le vocabulaire spécifique à Galaxy

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

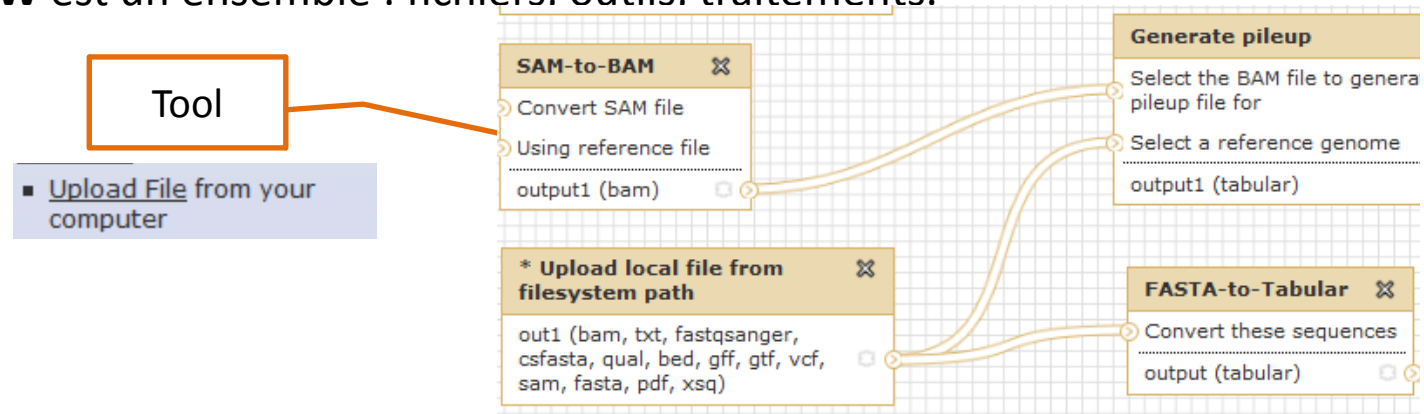




Votre **HISTORIQUE** est un « répertoire » qui « liste » l'ensemble de vos fichiers de données (fichiers d'entrée, fichier résultat) utilisés ou générés par un **TOOL** :

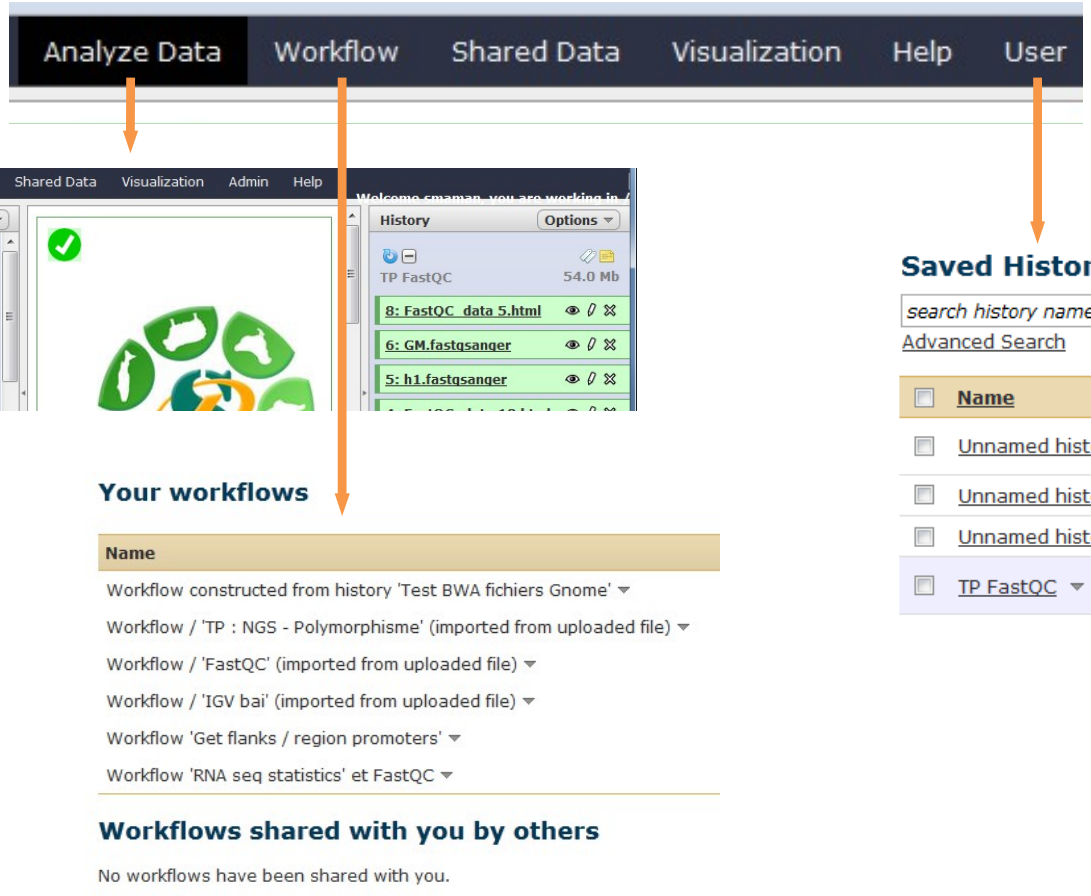


Votre **WORKFLOW** est un ensemble : fichiers. outils. traitements.



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.



Présentation de la plateforme Galaxy.

Premiers pas dans l'instance.

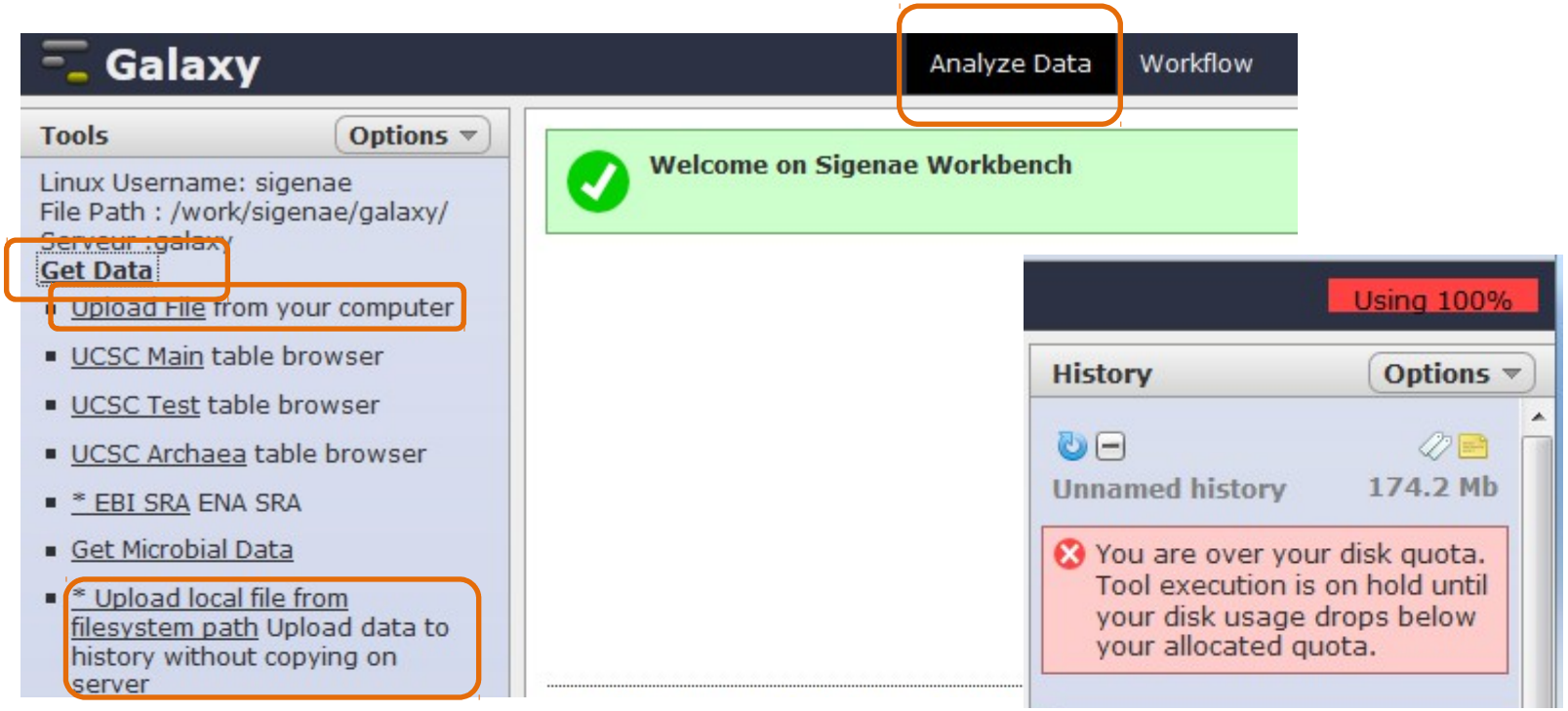
Notions d'outils, d'historique et de workflow.

Lancement de traitements bioinformatiques.

Quelques statistiques.

Auto-formations disponibles en ligne.

Deux méthodes de téléchargement de vos données privées



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.

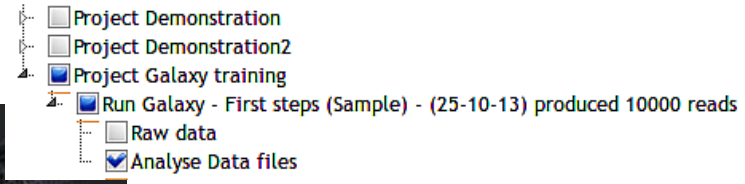
Présentation de NG6

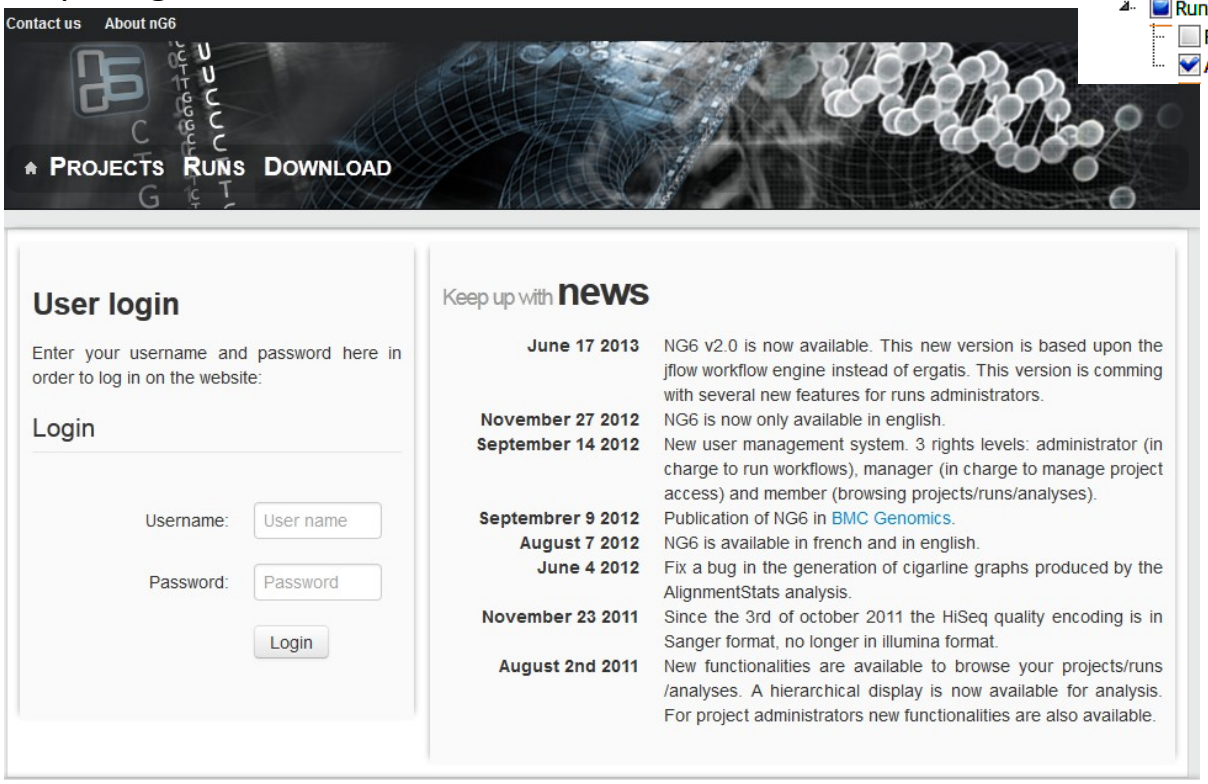
NG6 (Next Generation Sequencing Information System) :

•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.

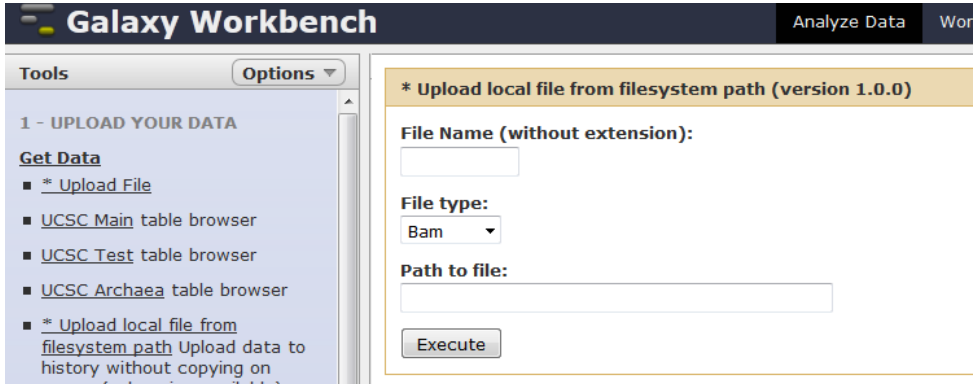
http://ng6.toulouse.inra.fr



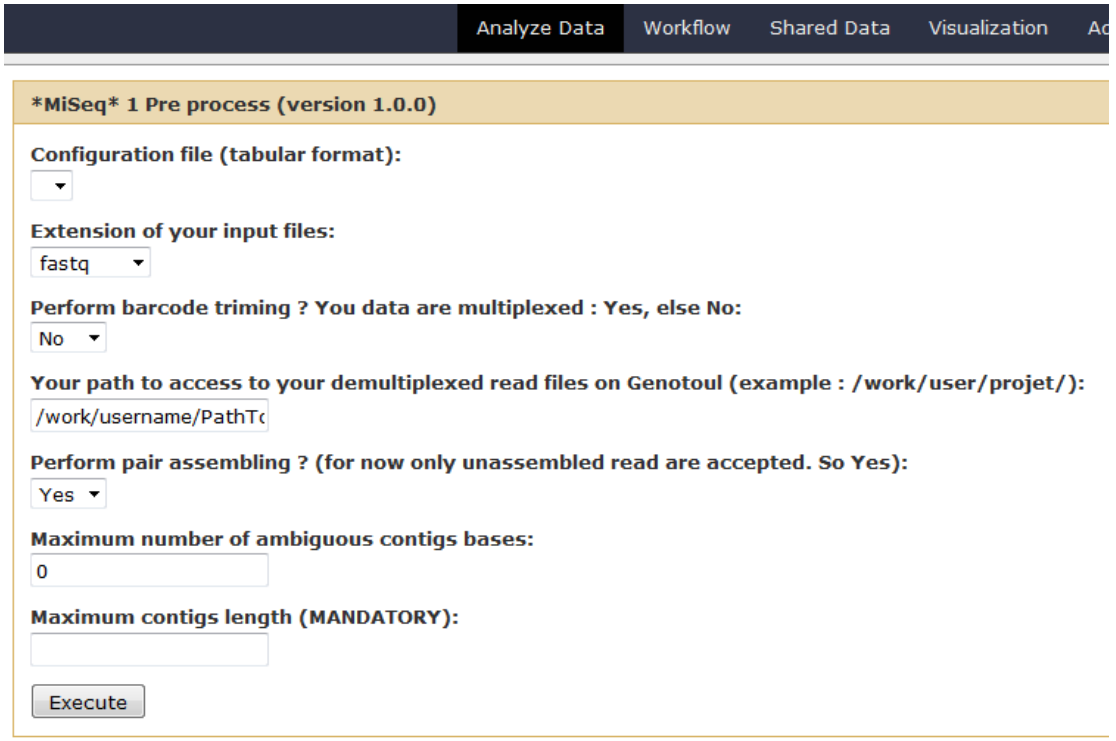


Liens avec NG6

-Soit fichier par fichier (repris en TP) :

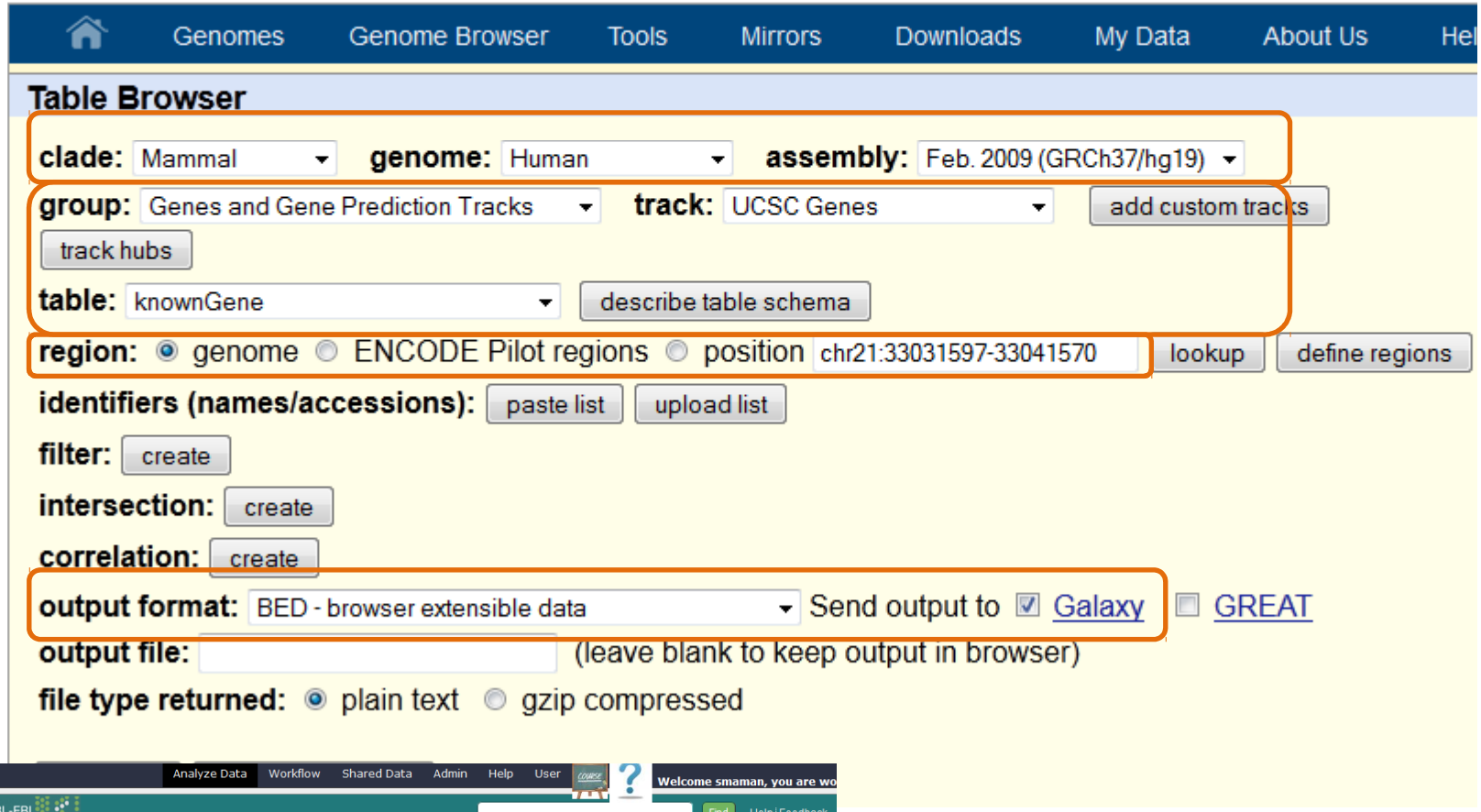


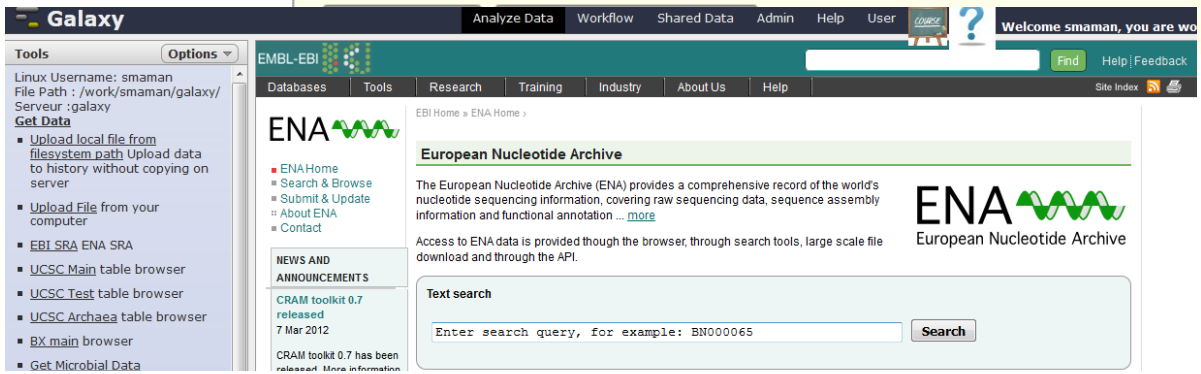
-Soit un répertoire de fichiers :



Téléchargement de données publiques

Données UCSC, Ensembl, BIOMART :





Exercice 1 :

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

Présentation de la plateforme Galaxy.

Premiers pas dans l'instance.

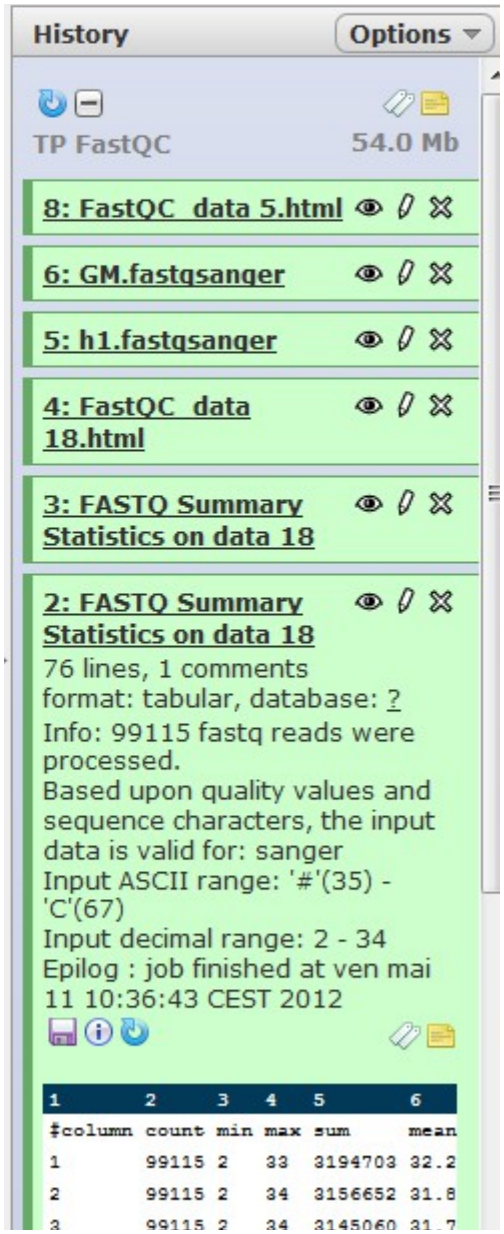
Notions d'outils, d'historique et de workflow.

Lancement de traitements bioinformatiques.

Quelques statistiques.

Auto-formations disponibles en ligne.

Gestion de vos historiques

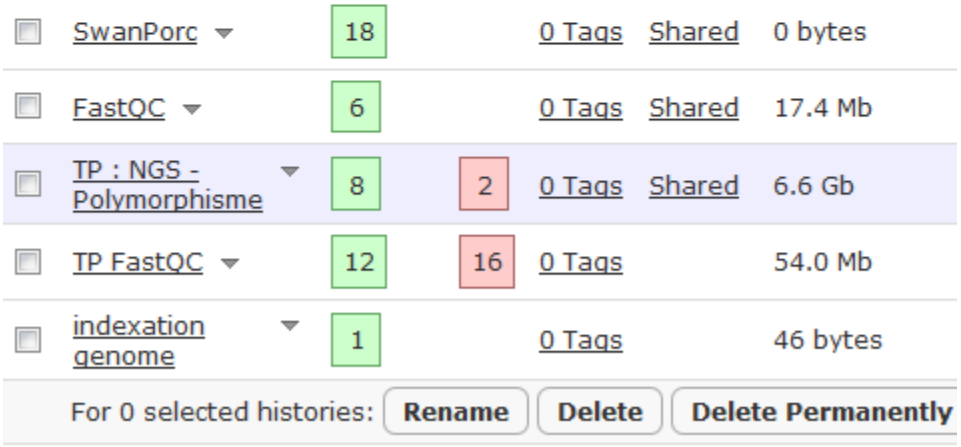


Conserver toutes les étapes de vos analyses.

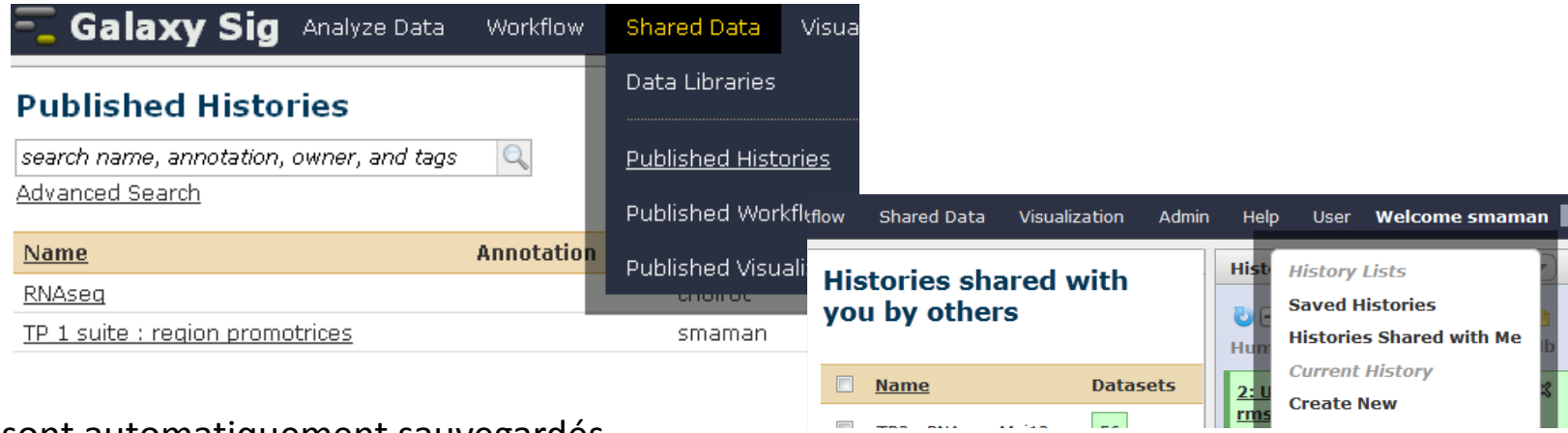
Partager vos analyses.

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

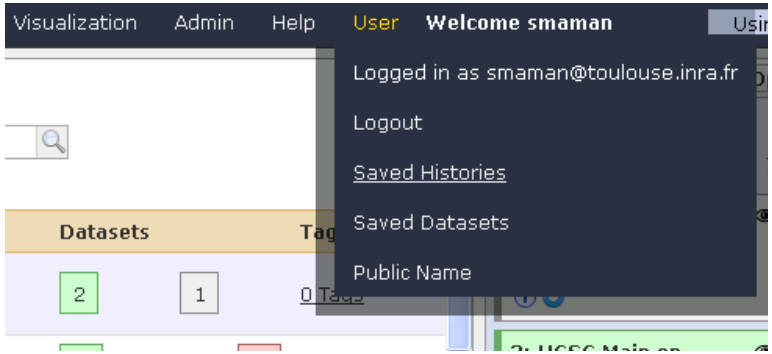
Répéter, autant de fois que nécessaire, une analyse en modifiant vos paramètres pour explorer les différences de résultats.



Historiques sauvegardés, publiés et partagés



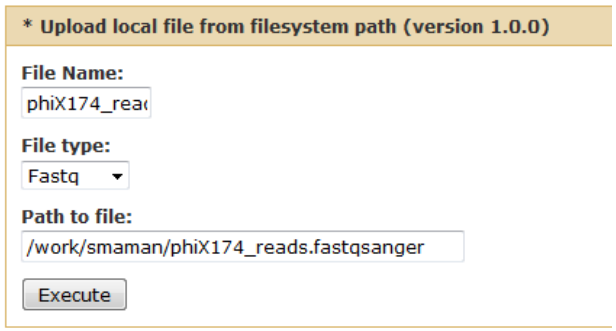
Vos historiques sont automatiquement sauvegardés.
Pour partager ou publier un historique :
User / Saved histories / Cliquer sur le nom de l'historique / Share ou Publish

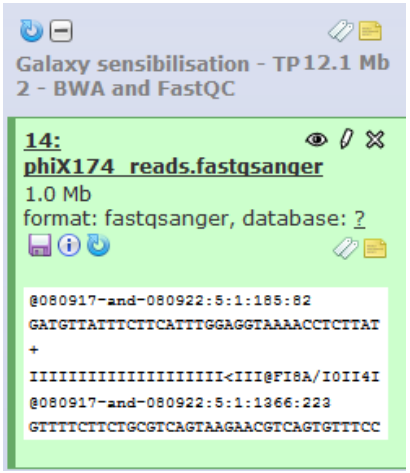


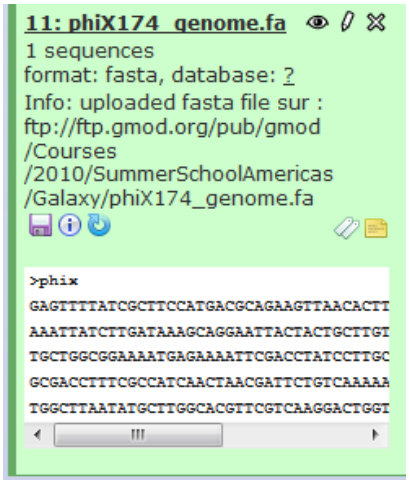
Vos historiques publiés sont accessibles à l'ensemble des utilisateurs loggés sur Galaxy (Shared Data / Published Histories).
Les historiques partagés sont accessibles uniquement à un utilisateur spécifique (Hitory / Option / Histories Shared With Me).

Comment lancer un job sans ligne de commande ?

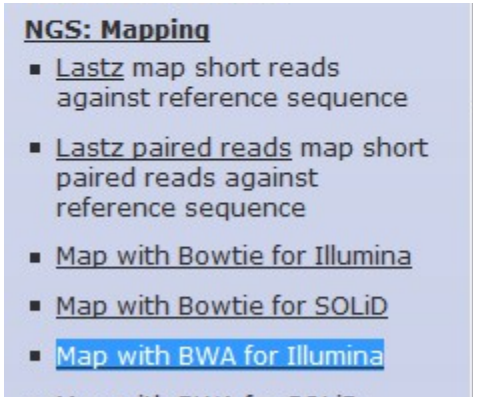
1 – Télécharger vos données :

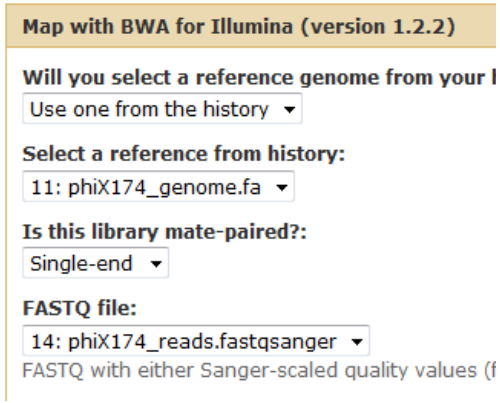






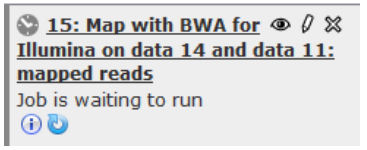
2 – Choisir un outil dans « Tools » :





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

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

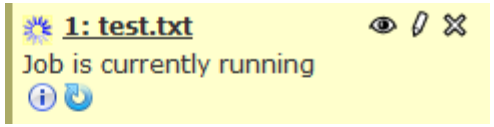


Fini les lignes de commande !

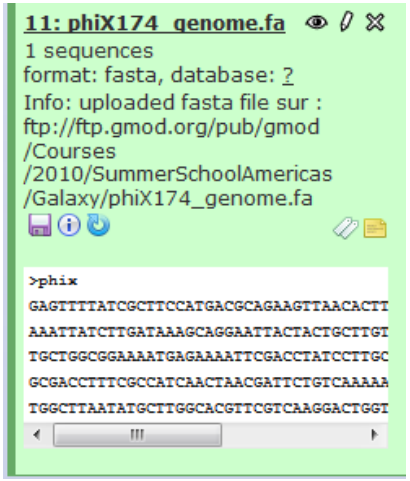
Comment suivre l'execution de mes jobs

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

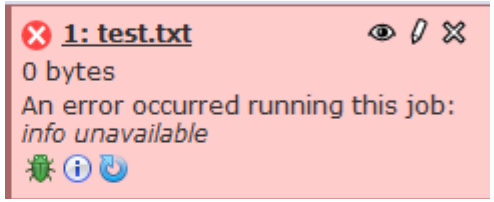
Etat 2 – JAUNE : Votre job est en cours d'execution



Etat 3 – VERT : Votre job est fini.

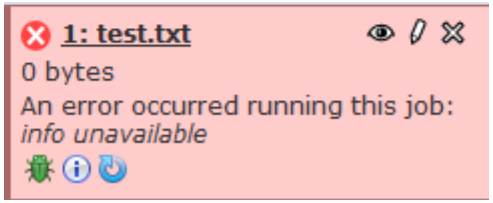


Bug - ROUGE : Votre job est planté !

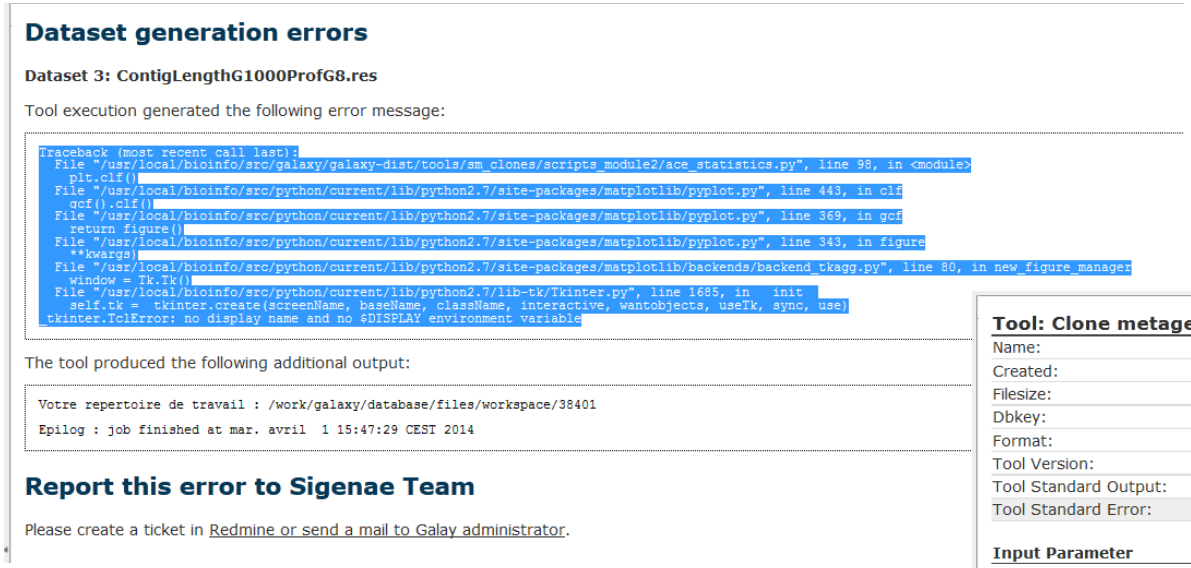


Mail à sigeane-support

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

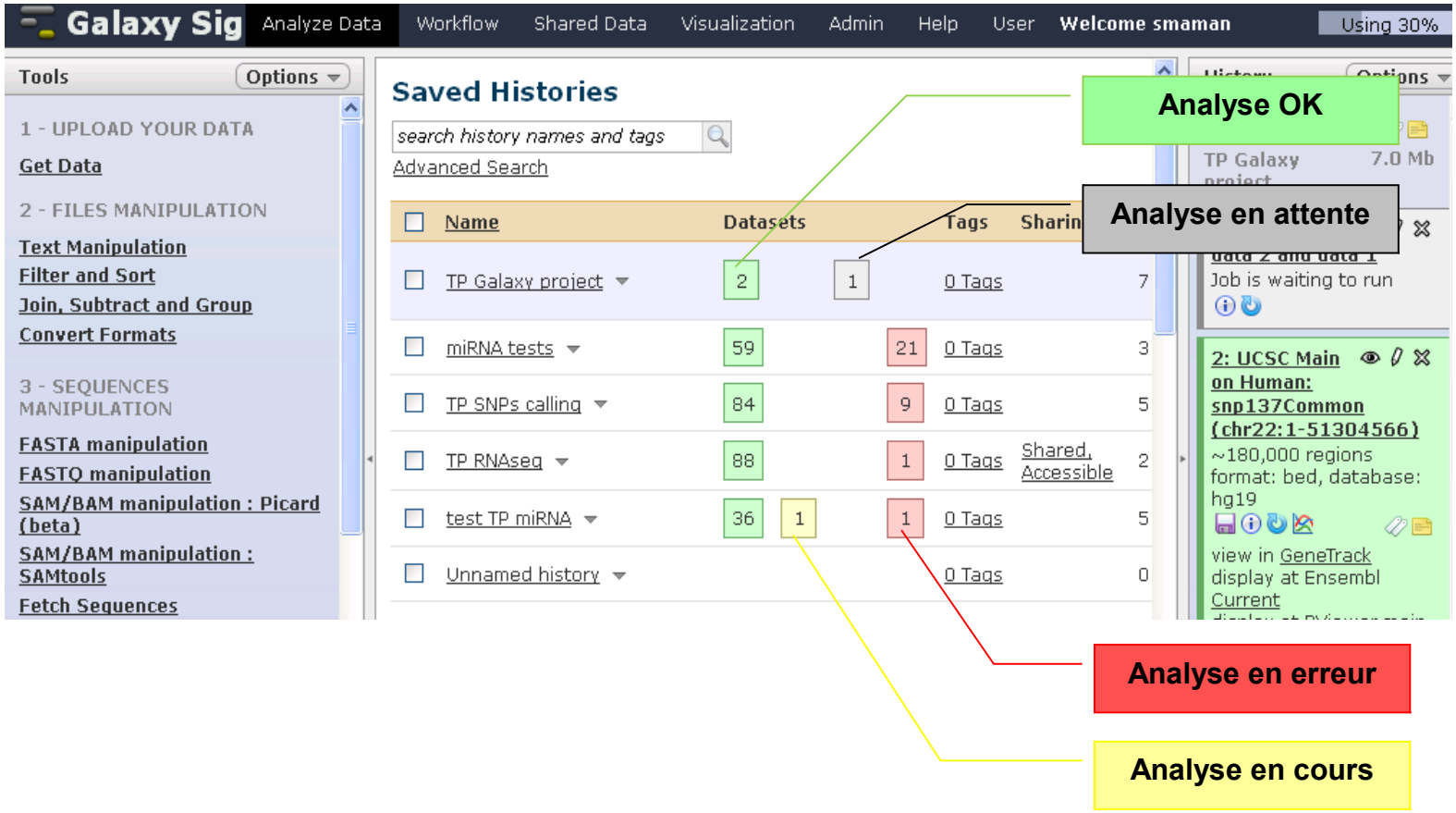
Other ace files	
Other ace files	

Inheritance Chain

ContigLengthG1000ProfG8.res

Vous pouvez aussi créer un ticket sous Redmine.

Liste des historiques et exécution de mes jobs



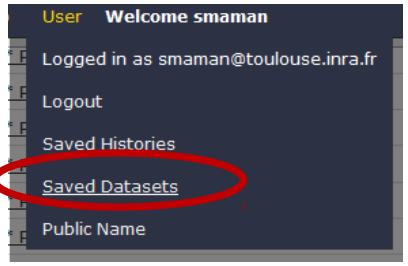
Rapidement, beaucoup de données sont générées

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

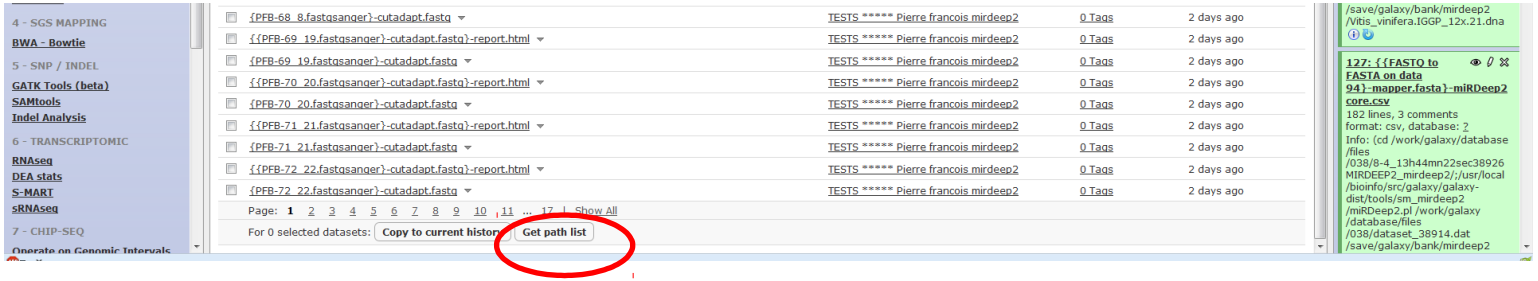
Compléments d'informations sur vos jobs

Pour sauver vos datasets Galaxy dans votre /work

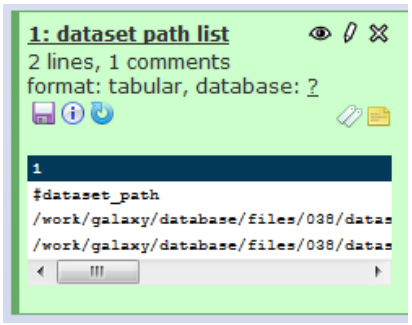
1



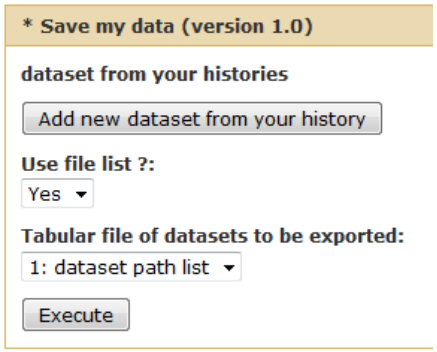
2



3



4

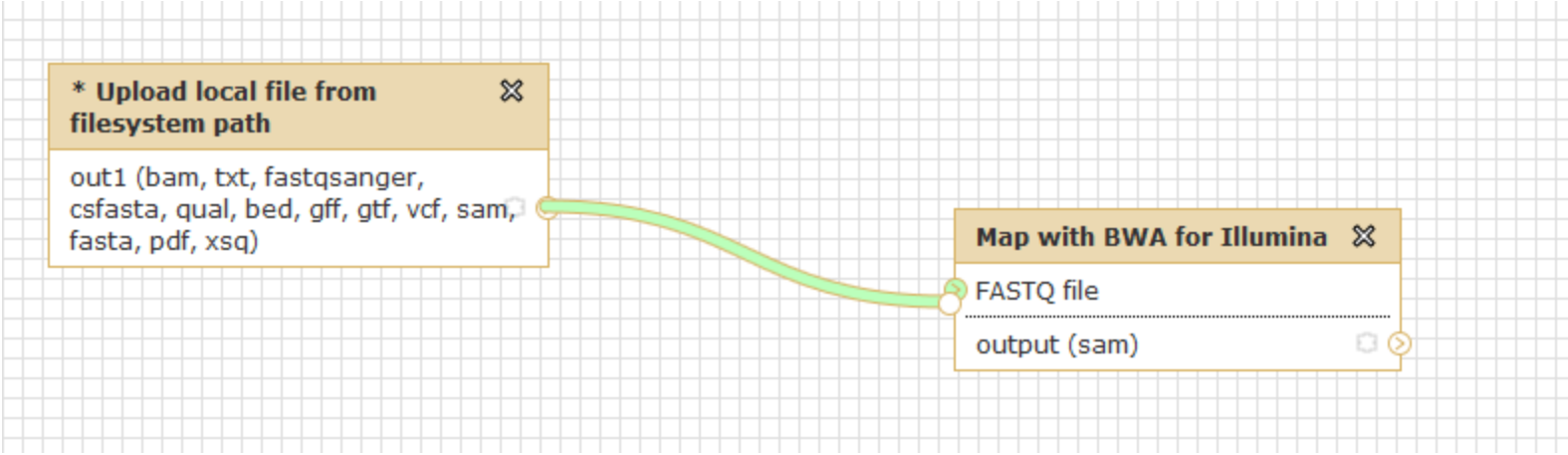


Mise en pratique prévue lors du TP.

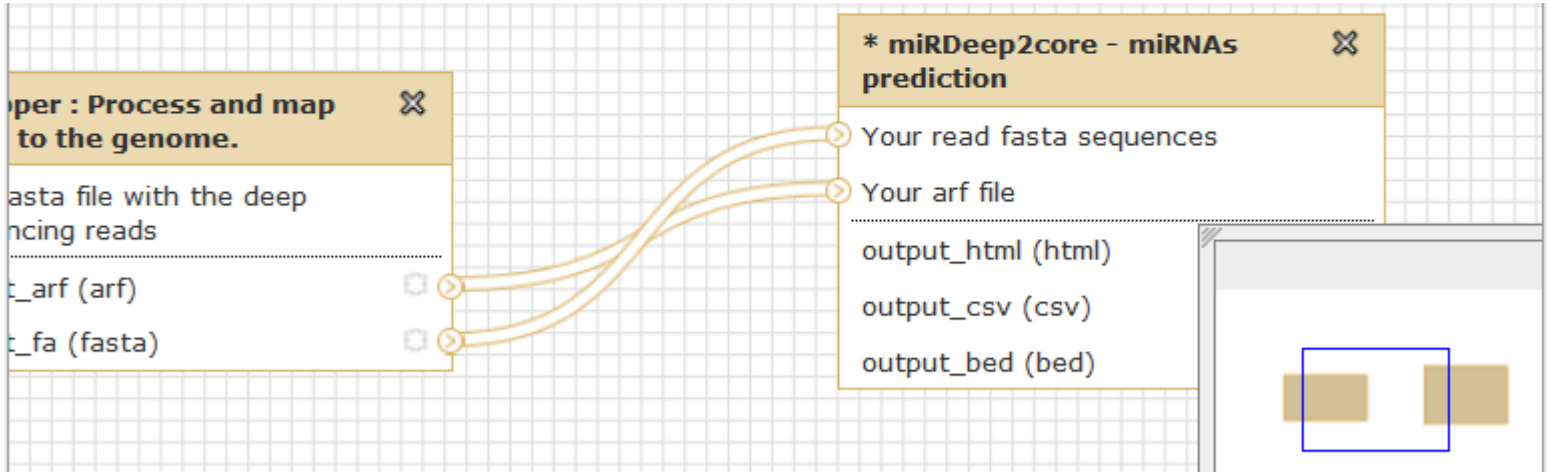
Créer un workflow

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 !



Trop de datasets dans votre workflow ?

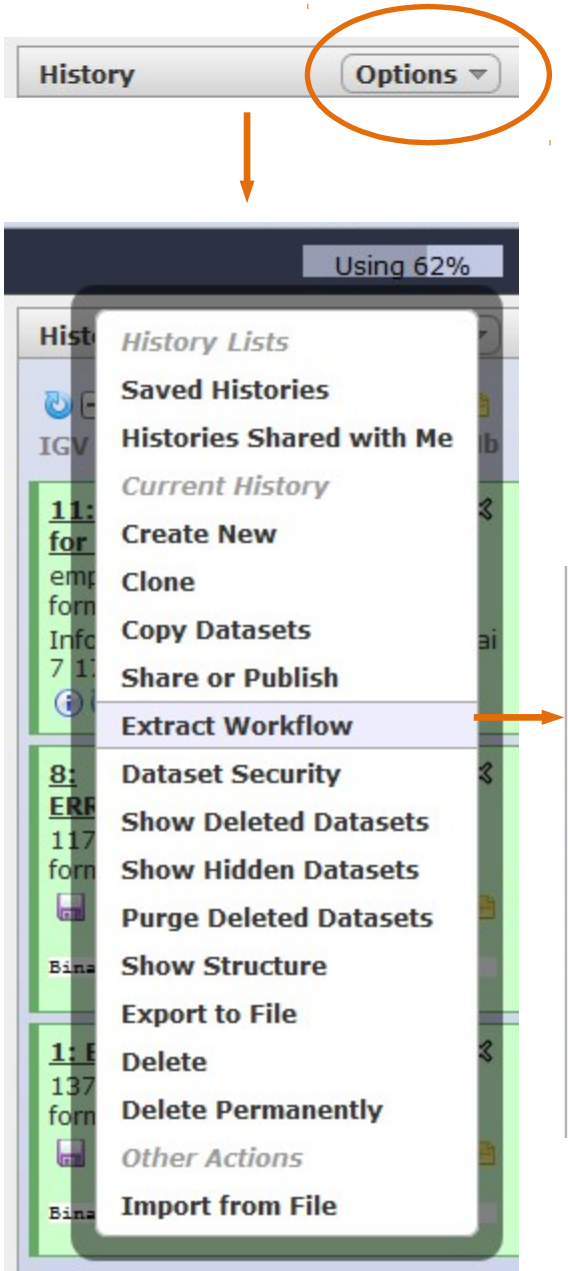


Si tout est coché, alors tout se passe comme si rien est coché.

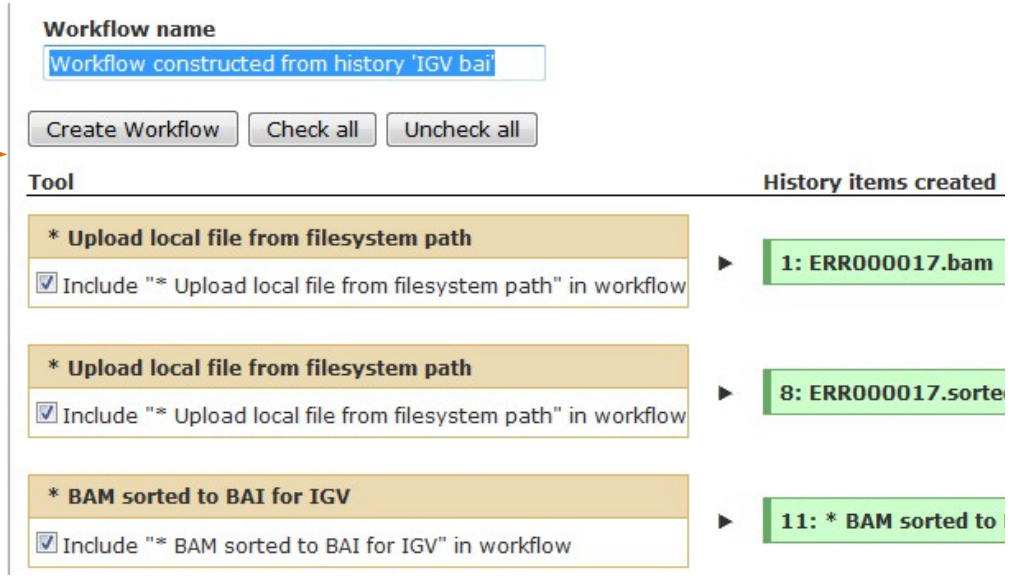
Si le dataset n'est pas coché, alors qu'au moins un autre est coché, alors le dataset non coché ne sera pas visible dans l'historique. Sauf si l'utilisateur choisi d'afficher les fichiers cachés, alors, dans cette config uniquement, il pourra le voir.

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

Exporter votre historique en workflow.



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

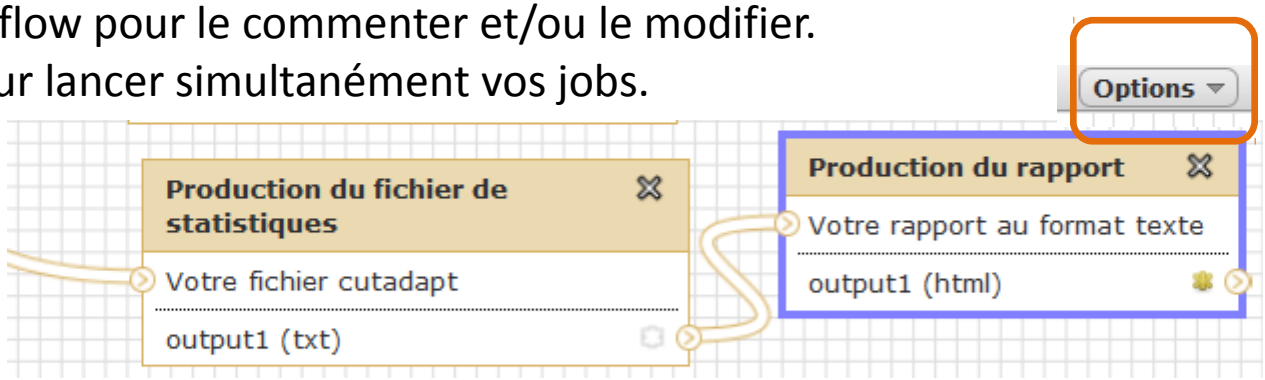


Exécution d'un workflow

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.

Quelques statistiques.

Auto-formations disponibles en ligne.

Les tools Galaxy

Menu évolutif et organisé par thématique

Ajout d'outils sur demande :
- Tool Shed
- Wrapper à façon (scripts maison, outils bioinfos)

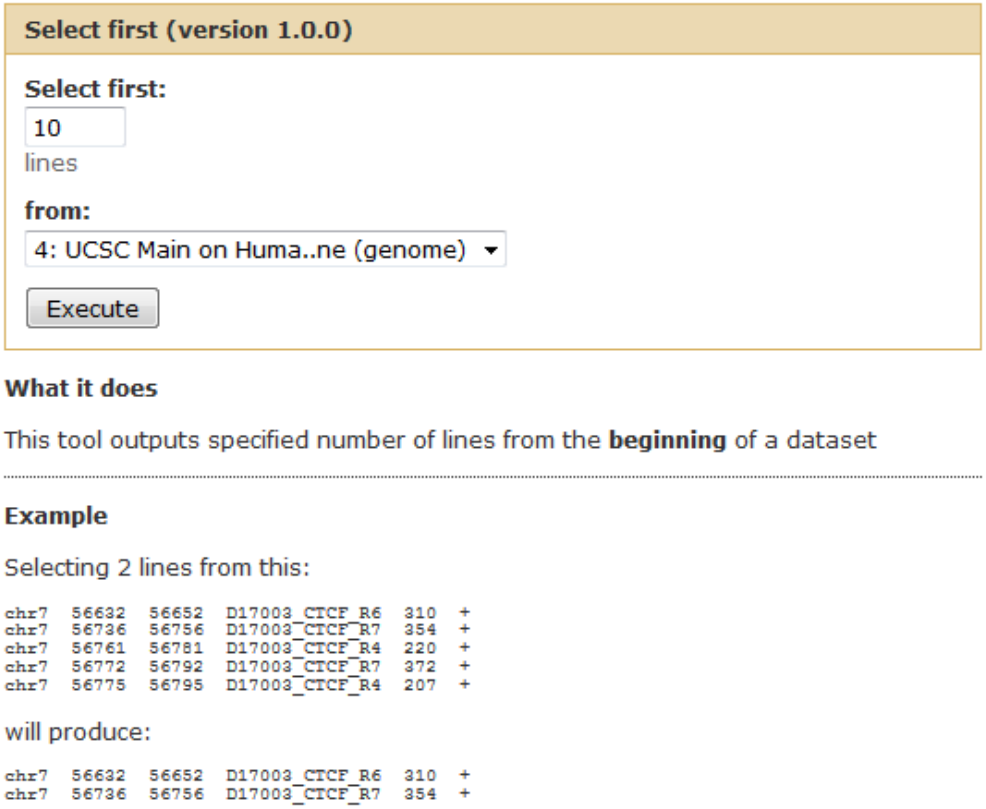
-> Envoyer vos demandes à :
sigenae-support@listes.inra.fr

Outils de traitement de fichiers

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

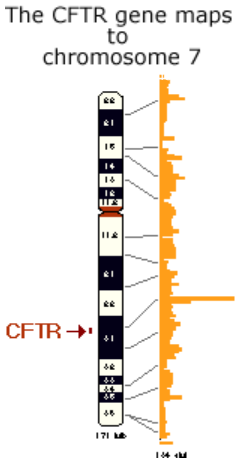
Voici les principaux outils « non bioinfo » proposés :

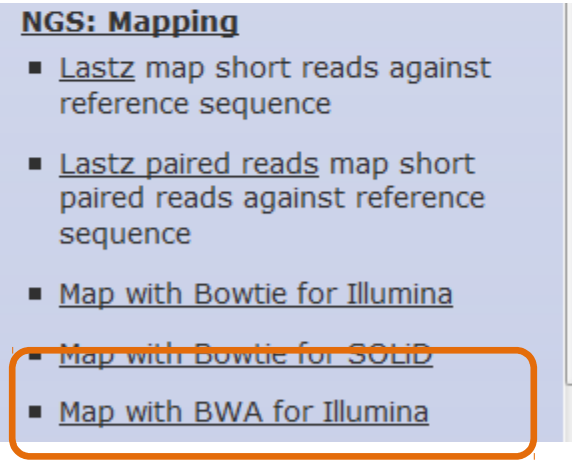
- Join (ex : fichiers lourds), Substract and Group
- Text Manipulation
- Filter and sort
- Convert Formats

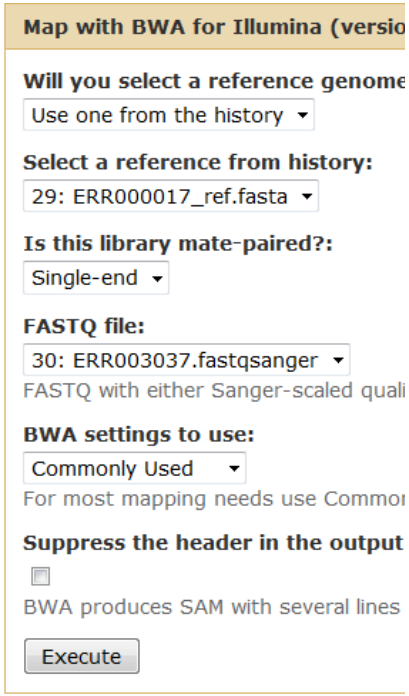


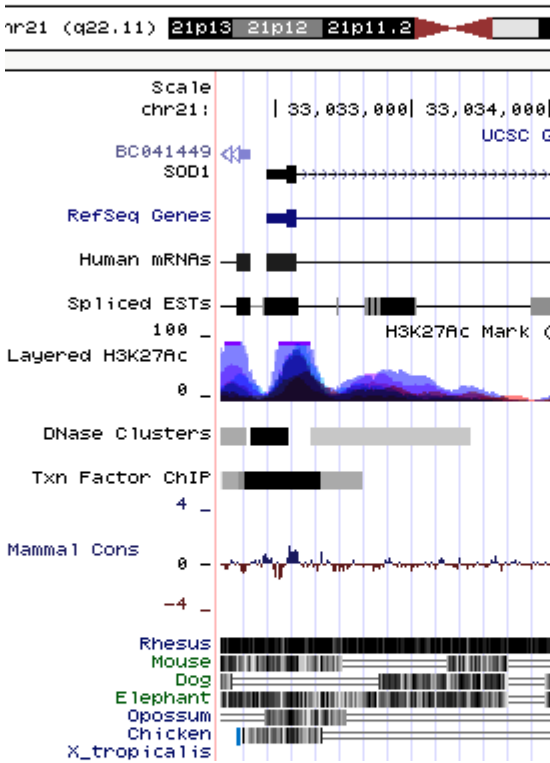
Outils bioinformatiques

Mapper un FASTQ sur une référence avec BWA.



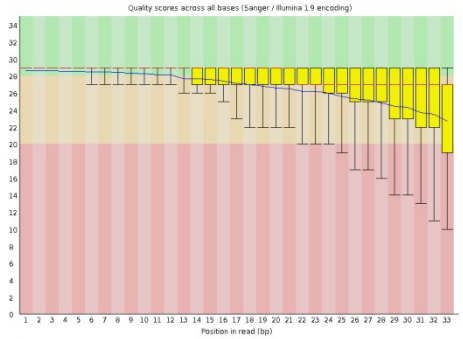






Visualiser la qualité des données avec FASTQC Report.

Visualiser un génome avec UCSC .



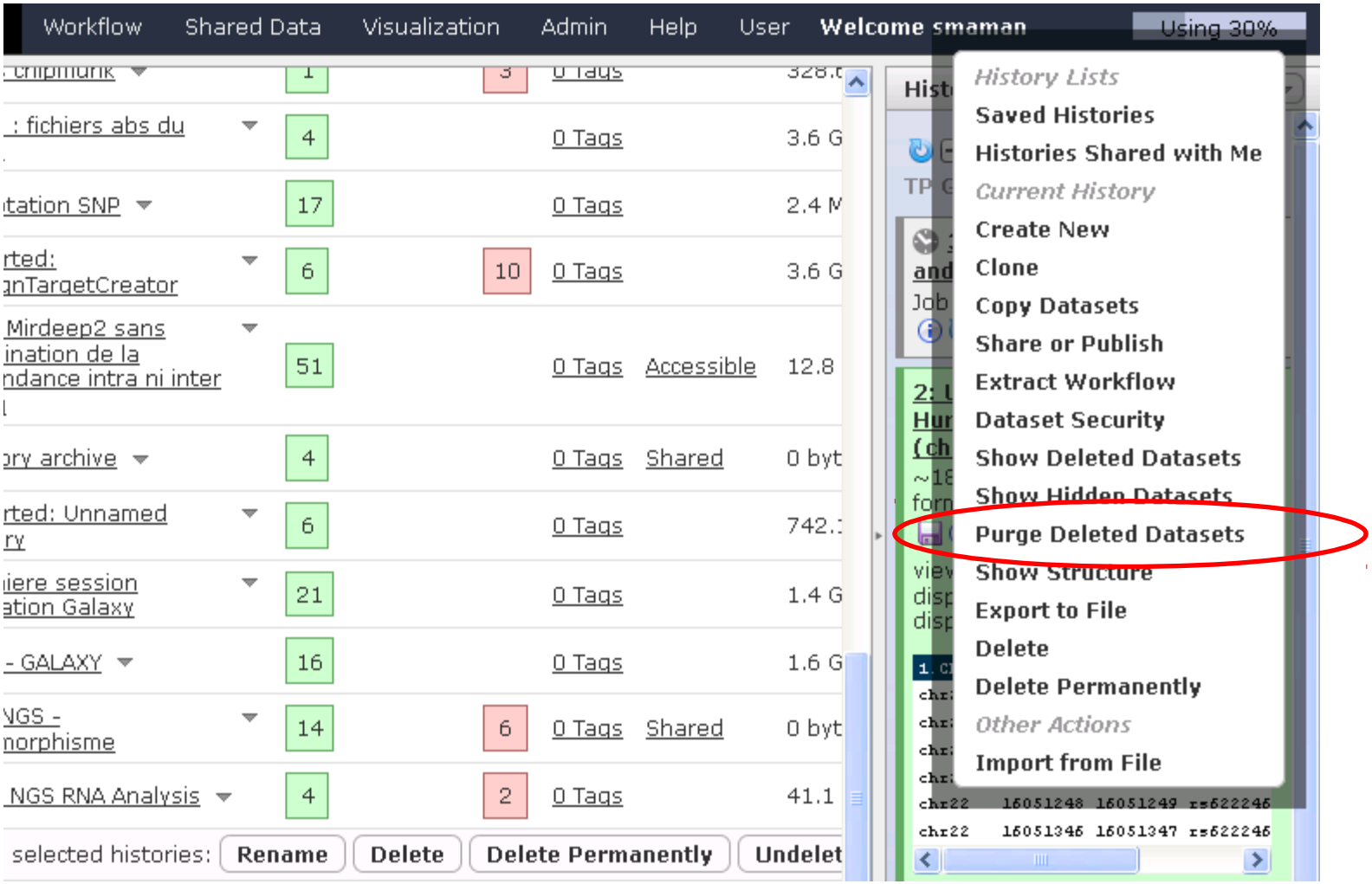
Exercice 2 & 3 :

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

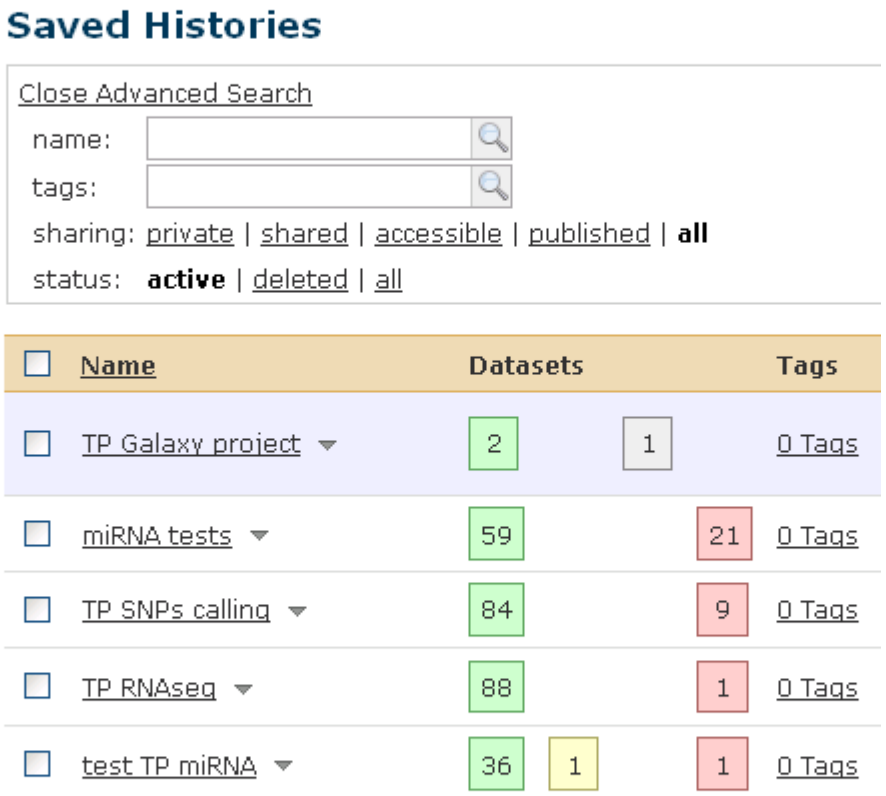
Création et partage d'historiques et de workflows.

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



Comment citer Galaxy dans vos publications ?

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.

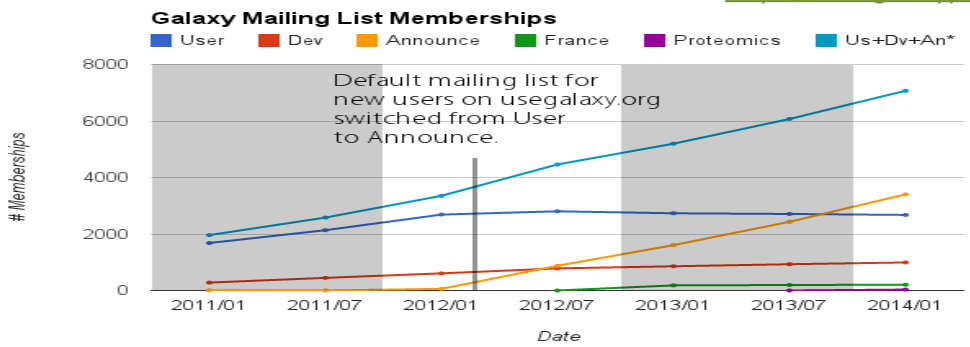
Quelques statistiques.

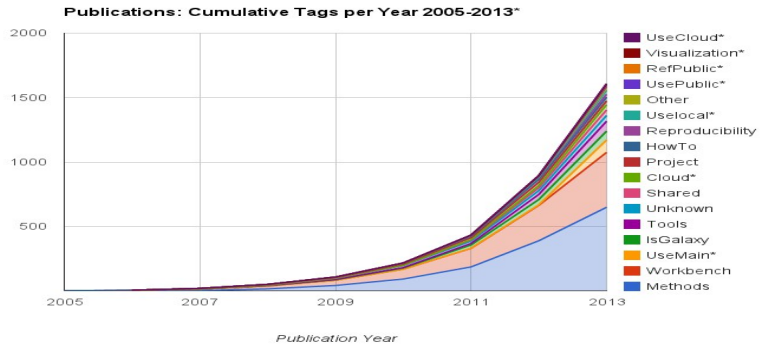
Auto-formations disponibles en ligne.

Statistiques d'utilisation de Galaxy international

Statistiques du Galaxy Project

https://wiki.galaxyproject.org





Une communauté internationale vivante.
Une communauté française grandissante

Belle augmentation des citations Galaxy dans les publications

Statistiques de l'instance Galaxy Sigenae / BioInfo Genotoul

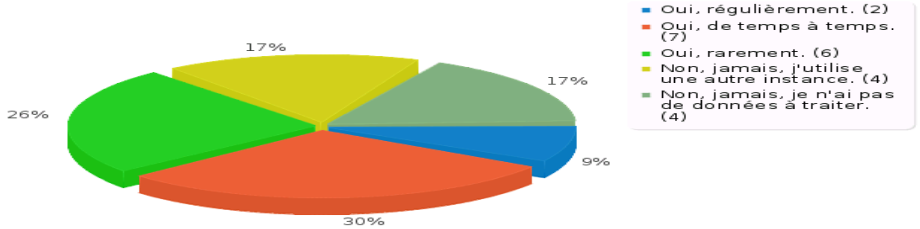
http://galaxy-workbench.toulouse.inra.fr

23 réponses sur ~40 utilisateurs

Historique mensuel des visites

~40 utilisateurs
Les 10 plus gros utilisateurs (hors tests Sigenae) utilisent Galaxy dans le cadre de leur projet.

Utilisez-vous l'instance Sigenae de Galaxy ?



65% des utilsateurs Galaxy ayant répondus à l'enquête utilisent l'instance Sigenae

L'instance Sigenae est intégrée au groupe de travail Galaxy IFB France http://www.ifb-galaxy.org/

Présentation de la plateforme Galaxy.

Premiers pas dans l'instance.

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

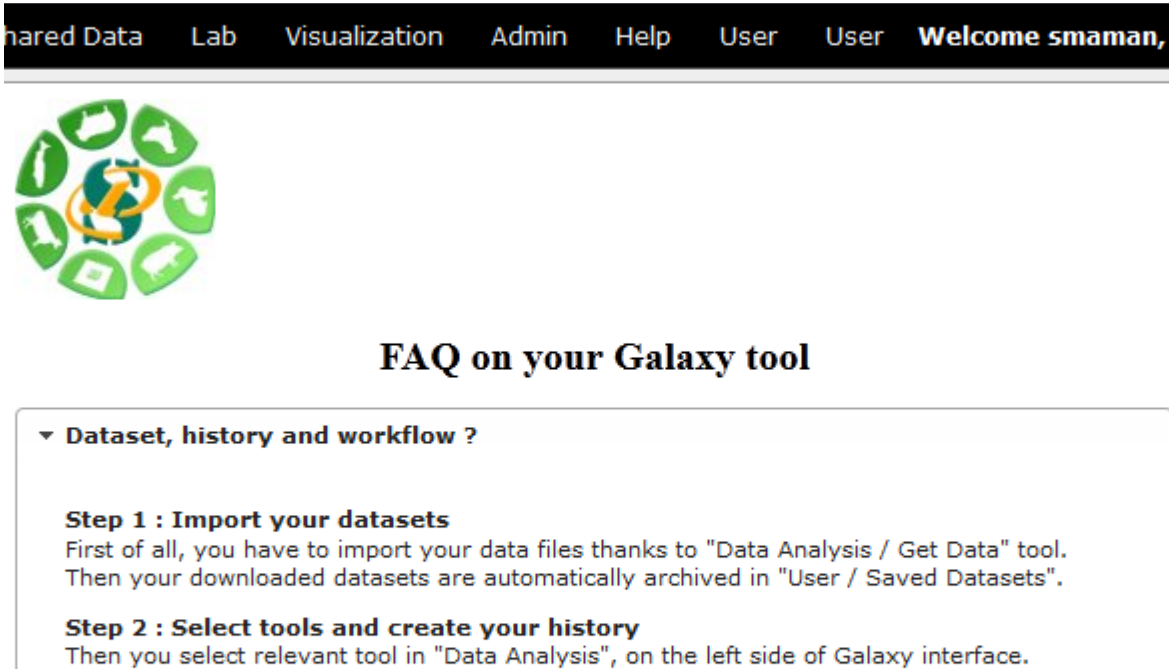
Lancement de traitements bioinformatiques.

Quelques statistiques.

Auto-formations disponibles en ligne.

FAQ et formation en ligne

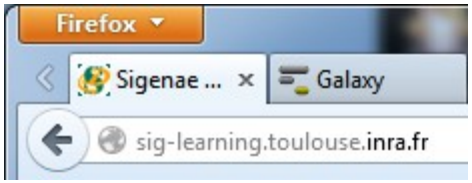
Une FAQ et le lien vers « sig-learning » sont disponibles depuis la page d'accueil.

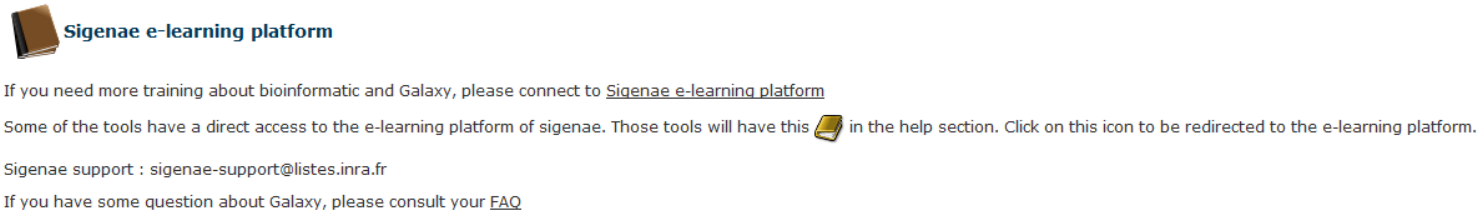


Vos supports sont disponibles depuis : **http://sig-learning.toulouse.inra.fr**

Votre accès à sig-learning

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



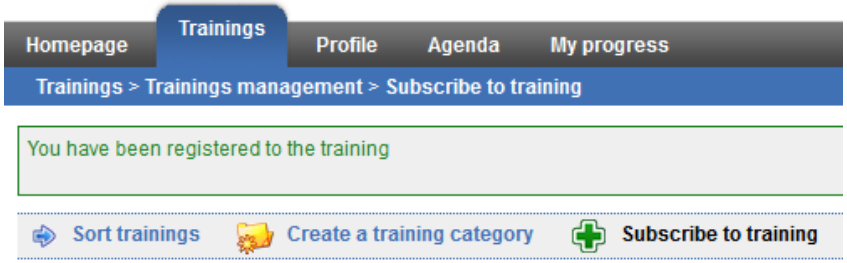


2 Authentification

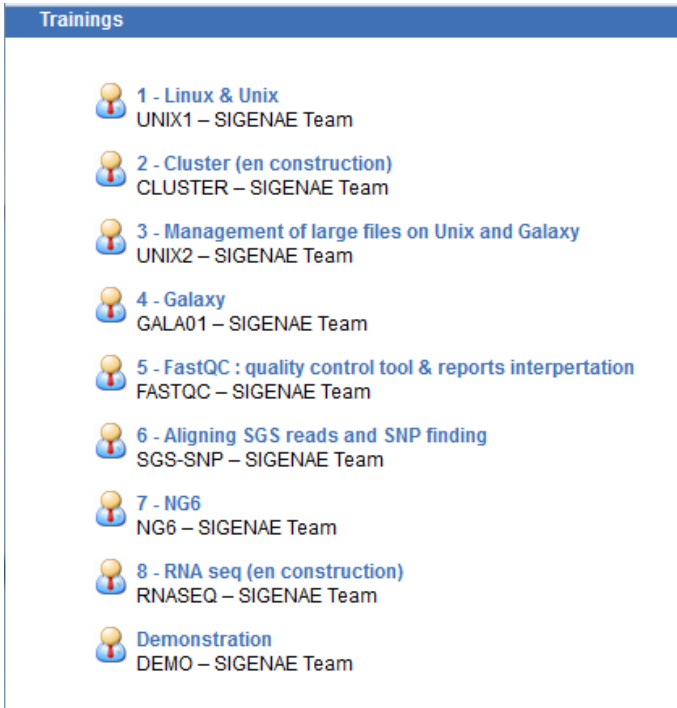
3 Onglet « Trainings » pour accéder à vos e-formations :

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 » :



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



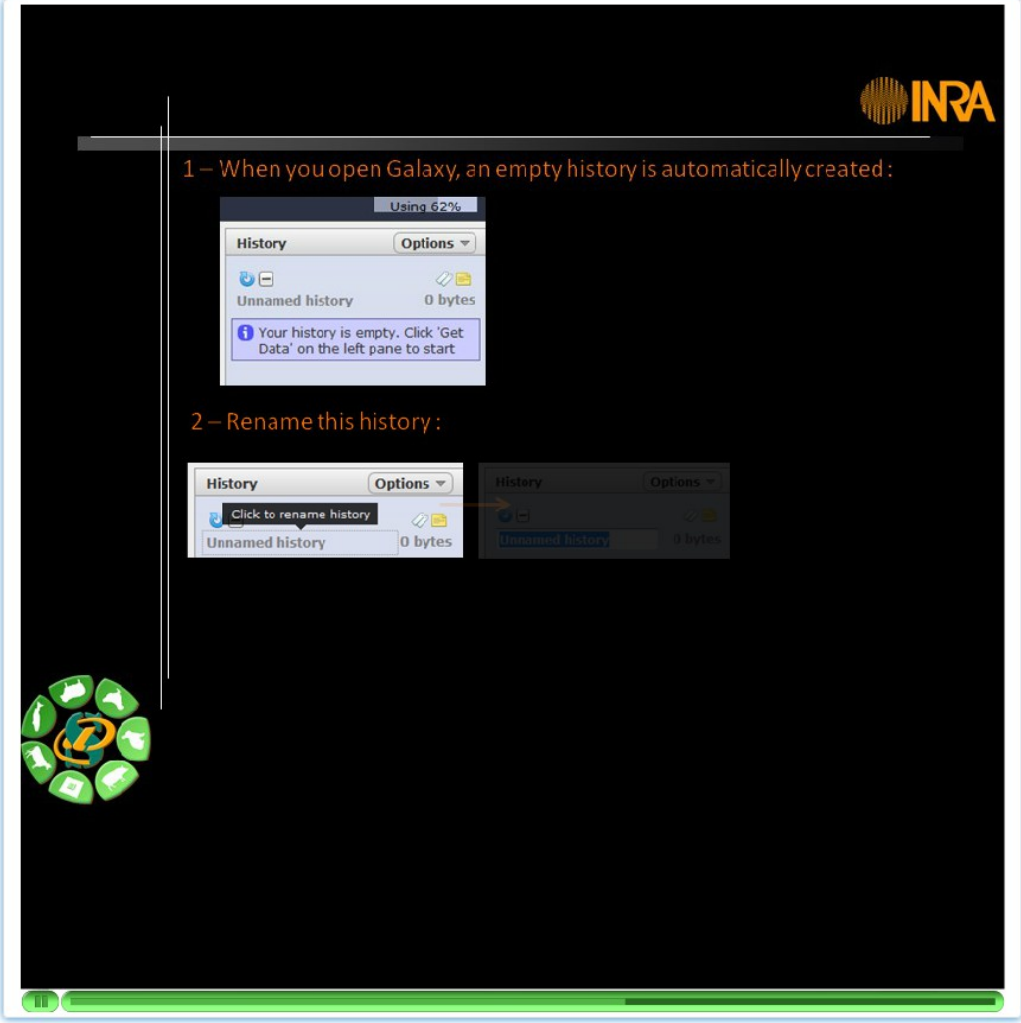
Organisation type d'une formation

Training home

Galaxy

57%

Galaxy > Galaxy User Interface



Commentaires | Comments

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

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

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

Merci pour votre écoute

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

Remerciements