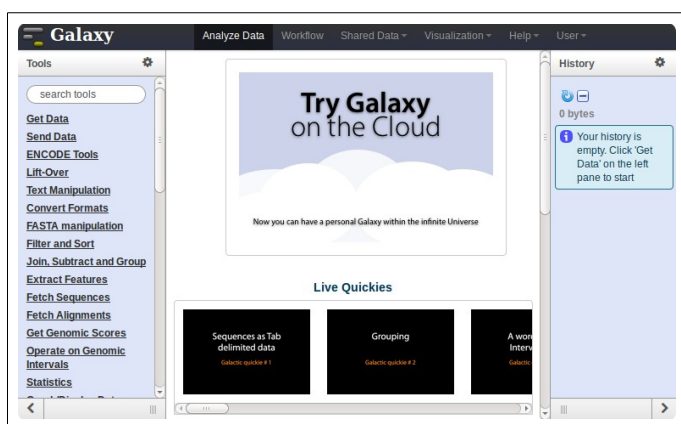




## - Galaxy -

# Initiation à la plateforme Galaxy

## - Utilisation du workflow VIP -



**Galaxy** plateforme de traitements informatiques et bioinformatiques accessible depuis l'url : <http://sigenae-workbench.toulouse.inra.fr/>



## Objectifs :

Cette formation a pour objectif de vous familiariser à l'utilisation de votre workbench Galaxy (<http://galaxy-workbench.toulouse.inra.fr>) et au pipeline de traitement VIP.

Vous découvrirez notamment comment :

- Traiter des fichiers sans utiliser de ligne de commande
- Lancer des traitements bioinformatiques sans Linux



Pour réaliser l'ensemble de ces exercices, vous avez besoin de vous connecter à la plateforme Galaxy en utilisant les login et mot de passe de votre compte « genotoul » : <http://galaxy-workbench.toulouse.inra.fr>

Vous pouvez utiliser vos identifiants et mots de passe de votre compte sur la plateforme bioinfo de Toulouse, ou bien utiliser un des comptes disponibles le temps de la formation :

- Logins : anemone arome aster bleuet camelia capucine chardon clematite cobee coquelicot
- Password : **Demander au formateur.**

Ce TP est disponible depuis cette URL :

[http://genoweb.toulouse.inra.fr/~formation/1\\_Galaxy\\_Initiation/Data/VIP/GALAXY\\_Initiation\\_exercices\\_juin2017.odt](http://genoweb.toulouse.inra.fr/~formation/1_Galaxy_Initiation/Data/VIP/GALAXY_Initiation_exercices_juin2017.odt)

Pour répondre à vos questions:



- Mail : support.siginae@inra.fr
- Une FAQ et un manuel utilisateur sont disponibles depuis la page d'accueil de l'instance Siginae de Galaxy.
- Les formations de la plateforme Bioinfo Genotoul sont disponibles sur <http://sig-learning.toulouse.inra.fr>

En fin de formation, penser à nettoyer votre compte de formation (« Delete permanently ») de l'ensemble des « histories » créés.



Juin 2017

## Exercice n°1 : Connexion à Galaxy, exploration de l'interface, téléchargement de datasets.

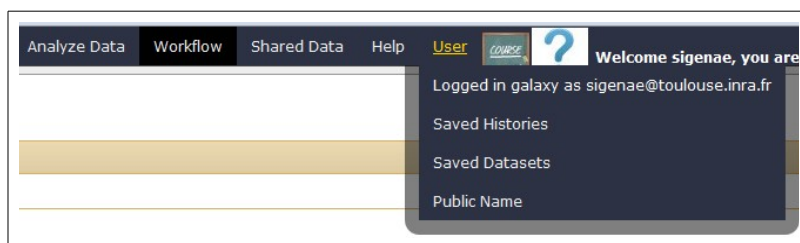
### Connexion à la plateforme Galaxy

Vous pouvez accéder à votre plateforme Galaxy (en précisant votre login et mot de passe « genotoul ») à l'adresse suivante : <http://galaxy-workbench.toulouse.inra.fr>

### Explorer l'interface

Depuis la barre du menu principal, vous avez accès aux onglets suivants :

- **Analyse Data** : Pour télécharger vos fichiers de données privées, et utiliser des modules de traitements.
- **Workflow** : Liste vos workflows archivés.
- **Shared Data** : Accès aux bibliothèques de données, ainsi qu'aux historiques et workflows publiés.
  - Data Libraries
  - Published Histories
  - Published Datasets
- **Help** :
  - Support
  - Galaxy Wiki
  - Video tutorials
  - How to cite Galaxy
- **User** :
  - Logged in galaxy as sigenae@toulouse.inra.fr
  - Saved Histories
  - Saved Datasets
  - Public Name



**Note** : La documentation autour de Galaxy est très aboutie, explorer le menu « help » et notamment la rubrique « Video tutorials »...



Afin de vous permettre une meilleure prise en main de l'interface Galaxy, nous vous encourageons à rechercher les outils à l'aide du menu « Options » - « Show Tool Search » disponible dans la partie « Tools » tout à gauche de l'interface.



## Notre objectif

Notre objectif est d'importer un fichier d'entrée, puis lancer l'outil VIP :

History

search datasets

**vip**  
3 shown, 93 deleted  
4.85 MB

**96: {/work/edelpuech /data\_test/data VIP /analyses/fastq\_file\_light.fq}-VIP.log**

**95: {/work/edelpuech /data\_test/data VIP /analyses/fastq\_file\_light.fq}-VIP.html**

**94: {/work/edelpuech /data\_test/data VIP /analyses/fastq\_file\_light.fq}**  
100 sequences  
format: **fastqsanger**, database: ?  
Epilog : job finished at lun. juin 26 10:04:38 CEST 2017

```
@SRR1106548.2005967.2005967/1
GTTATCCCAGCAGAAACAGGACAGGAGACGACATACTT
+
CCCCFFFFHHHHJJJJJJJIJIIJJIIHIIJJJJJJJJ
@SRR1106548.1993303.1993303/2
CTTCTTGGGCGGGATCTGTGCGGGGTA AAAATAGGACC
```

L'historique nommé « VIP » contient 3 datasets :

1 – Le fichier FASTA entrant nommé « fastq\_file\_light.fq »  
Il est recommandé de le renommé avec un intitulé plus court, à l'aide de l'icône « stylo » :

**96: {/work/edelpuech /data\_test/data VIP /analyses/fastq\_file\_light.fq}-VIP.log**

Attributes Convert Format Datatype Permissions

Edit Attributes

**Name:**  
fastq\_file\_light.fq

**Info:**  
Epilog : job finished at lun. juin 26 10:04:38 CEST 2017

**Annotation / Notes:**  
Add an annotation or notes to a dataset; annotations are available

**Database/Build:**  
unspecified (?)

Save

2 – Les deux datasets de sortie générées automatiquement par le lancement de l'outil VIP :

- \* Un fichier de log :  
{fastq\_file\_light.fq}-VIP.log
- \* Un report HTML :  
{fastq\_file\_light.fq}-VIP.html



## Import de données

### 1 Téléchargement des fichiers avec copie sur le serveur (non recommandé)

Nous utiliserons l'outil « [Upload File](#) from your computer », disponible dans la section « Get data », pour charger notre FASTA entrant.

#### Get Data

[Upload File](#) from your computer

Cliquez sur « Paste/fetch data » pour copier/coller l'URL d'accès au fichier :

[http://genoweb.toulouse.inra.fr/~formation/1\\_Galaxy\\_Initiation/Data/VIP/fastq\\_file\\_light.fq](http://genoweb.toulouse.inra.fr/~formation/1_Galaxy_Initiation/Data/VIP/fastq_file_light.fq)

Name	Size	Type	Genome	Settings	Status
New File	91 b	Auto-detect	unspecified (?)		None

Puis cliquez sur « Start » puis « Close ». Votre dataset est chargée à droite de votre écran, dans la partie « History ».

Renommer la dataset et votre historique.

Pour visualiser le contenu de votre dataset, veuillez cliquer sur l'icône « oeil » :

96: {/work/edelpuech /data test/data VIP /analyses/fastq file light.fq}-VIP.log



Vos fichiers de données téléchargés apparaîtront dans votre historique courant et seront automatiquement archivés dans « User / Saved Datasets ».



## 2 Téléchargement des fichiers depuis votre work sur Genotoul (recommandé)

L'intérêt d'utiliser l'outil « [Upload File from Genotoul](#) » est de ne pas copier vos fichiers d'entrée en double sur votre work et sur celui de Galaxy. Ainsi, votre quota n'est pas trop impacté. Voici comment renseigner les paramètres de l'outil :

**Upload File from Genotoul** (Galaxy Version 1.0.1)


**Path to file**

Path must be like : /work/USERNAME/somewhere/afile

**File type**

Execute

Si vous avez plusieurs fichiers à télécharger, nous vous conseillons d'utiliser l'outil « [Upload several files from Genotoul](#) without impact too much your Galaxy quota ». En précisant un chemin d'accès à un répertoire contenant plusieurs fichiers d'un même format, Galaxy les charge en un seul clic. Il

est souvent nécessaire d'actualiser votre historique avec l'icône  afin de voir vos datasets s'affichier à droite, dans votre historique.



**Ces deux outils ne fonctionnent bien que si et seulement si vos droits sont bien gérés. Plus plus de détails :**

[http://genoweb.toulouse.inra.fr/~formation/1\\_Galaxy\\_Initiation/GALAXY\\_MEMO.pdf](http://genoweb.toulouse.inra.fr/~formation/1_Galaxy_Initiation/GALAXY_MEMO.pdf)



## Etape n°2: Lancement de VIP

### Lancement de VIP depuis l'historique

L'outil « [Detection viral sequences with VIP](#) Detection of viral sequences for the treatment of NGS data from viral metagenomics » est disponible depuis le menu de gauche « Tools » :

[Detection viral sequences with VIP](#)  
Detection of viral sequences for the treatment of NGS data from viral metagenomics

Paramétrez votre outil pour lui indiquer sur quel fichier vous souhaitez travailler, le format de celui-ci, le séquenceur utilisé, et la base de référence à utiliser :

**Detection viral sequences with VIP** Detection of viral sequences for the treatment of NGS data from viral metagenomics (Galaxy Version 1.0.1) Options

**Your NGS data**

**Data format of the NGS data**

**Sequencers used**

**Path to reference database**

Select specific database from the list

Cliquez sur « Execute », l'outil génère automatique un fichier de log et un report HTML.

Le report présente les résultats de VIP.

Le fichier de log indique les étapes du traitement et les éventuels bugs rencontrés. Il est utile au support, si besoin.

**96: {/work/edelpuech /data test/data VIP /analyses/fastq\_file\_light.fq}-VIP.log** 👁️ ✎️ ✕

**95: {/work/edelpuech /data test/data VIP /analyses/fastq\_file\_light.fq}-VIP.html** 👁️ ✎️ ✕

### Reads Distribution

### Top 5 virus distribution

**vip**

3 shown, 96 deleted

4.92 MB

**96: {/work/edelpuech /data test/data VIP /analyses/fastq\_file\_light.fq}-VIP.log** 👁️ ✎️ ✕

**95: {/work/edelpuech /data test/data VIP /analyses/fastq\_file\_light.fq}-VIP.html** 👁️ ✎️ ✕

4.7 KB

format: **html**, database: 2

Debug mode: OK

Job working directory: /galaxydata/galaxy-preprod/my\_workspace/8023

Your linux command line:

```
(cd /galaxydata/galaxy-preprod/my_workspace/8023; /galaxydata/galaxy-preprod/tools/edvip/vip_command_line.sh NGS fastq 454 /galaxydata)
```

HTML file:

**94: {/work/edelpuech /data test/data VIP /analyses/fastq\_file\_light.fq}** 👁️ ✎️ ✕

### Summary Report

Species	Genus	GI	%Coverage	Reads_hit	Reads_num	Average depth of coverage
>AF063223.1 HIV-1 isolate DJ263 from Djibouti, complete genome	Lentivirus	>AF063223.1 HIV-1 isolate DJ263 from Djibouti, complete genome	39.1000	45	53	.7859
>NC_018464.1 Shamonda virus N and NS5 genes, segment 5, genomic RNA, isolate ID An 5550	Orthobunyavirus	>NC_018464.1 Shamonda virus N and NS5 genes, segment 5, genomic RNA, isolate ID An 5550	3.3400	1	1	.0334
>NC_006146.1 Macacine herpesvirus 4, complete genome	Lymphocryptovirus	>NC_006146.1 Macacine herpesvirus 4, complete genome	0	0	1	0



## Lancement de VIP depuis le workflow

Nous vous avons préparé un workflow VIP déjà paramétré. Il est tout à fait possible et recommandé de vérifier et d'ajuster ces paramètres.

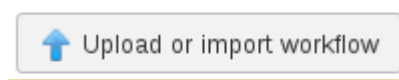
Ce fichier contenant le workflow est disponible depuis cette URL :

[http://genoweb.toulouse.inra.fr/~formation/1\\_Galaxy\\_Initiation/Data/VIP/Galaxy-Workflow-VIP.ga](http://genoweb.toulouse.inra.fr/~formation/1_Galaxy_Initiation/Data/VIP/Galaxy-Workflow-VIP.ga)

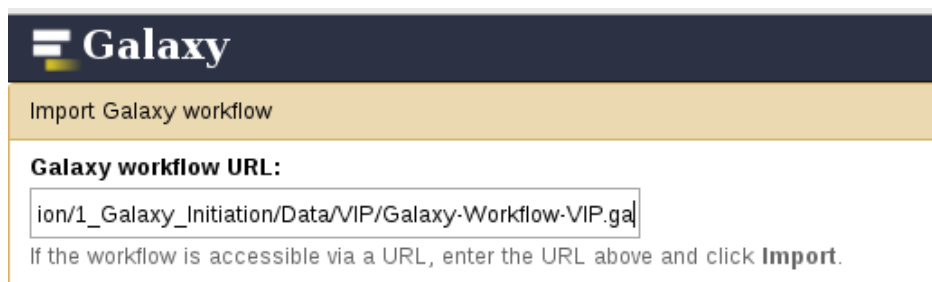
Pour importer ce workflow dans votre instance Galaxy, veuillez cliquer sur le menu « Workflow » :



Puis, à droite de votre écran, importez le fichier dans Galaxy :



Puis renseignez l'URL d'accès au fichier :



Puis cliquez sur «Import». Ce nouveau workflow est maintenant disponible depuis le menu «Workflow»/«Your workflows». Il est possible de le lancer en cliquant sur «Run» (flèche noire).

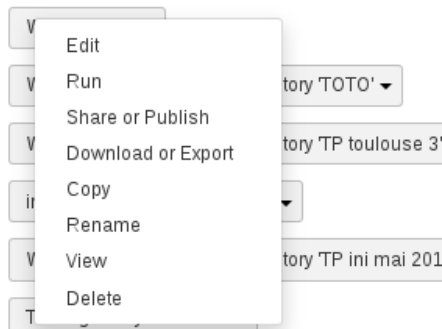
### Your workflows

Name

Workflow VIP (imported from uploaded file) ▼

### Your workflows

Name





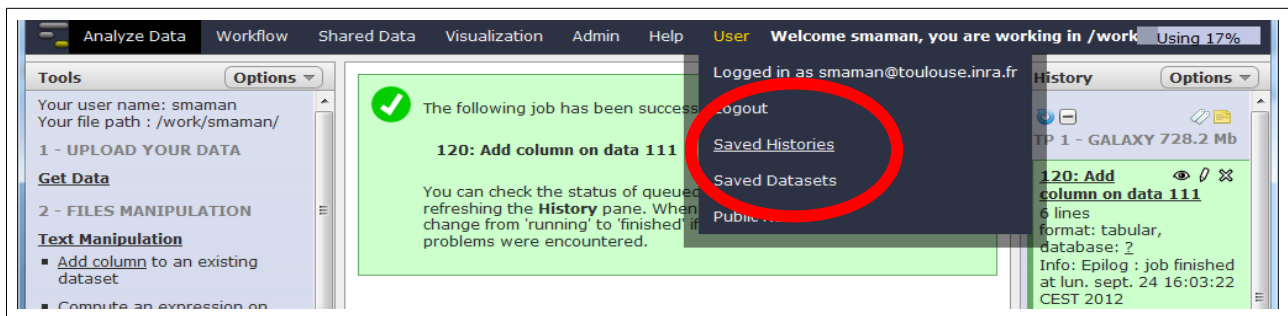


### Etape n°3: Création et partage de datasets, d'historiques et de workflows.

#### Notions d'historique

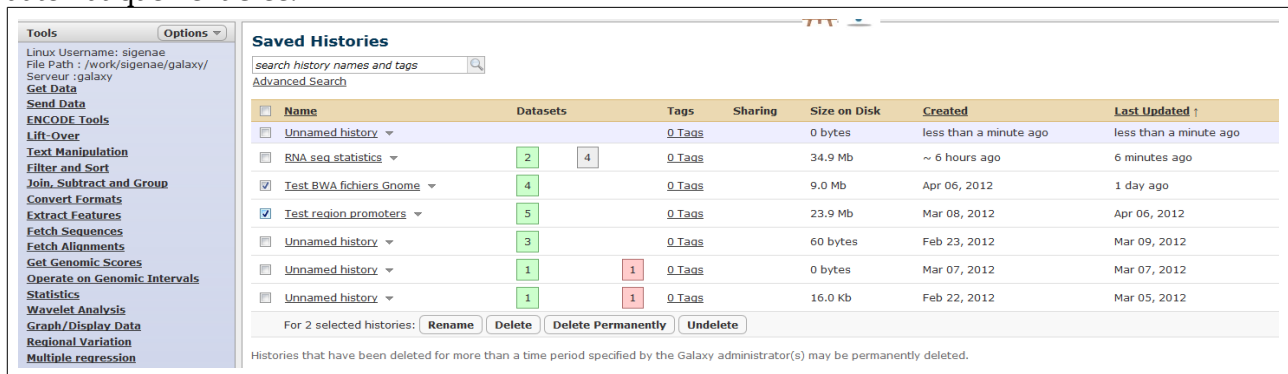
##### Traitements archivés dans un historique

Au fur et à mesure que vous faites appel aux différents outils au sein de votre interface depuis le menu « Analyse Data », l'ensemble des étapes sont enregistrées dans un historique qui est automatiquement archivé dans « User / Saved Histories » et que vous pouvez ensuite, si besoin, partager dans « Shared Data / Published Histories ».



##### Gérer ses historiques

Depuis le menu « User » / « Saved Histories », vous avez la possibilité de gérer vos historiques (delete, delete permanently, rename, undelete) en cliquant sur l'intitulé de l'historique. Remarque, lors de votre connexion au workbench Galaxy, un « current history » est automatiquement créé.



##### Exercice

- Créer un nouvel historique (nommer le en le préfixant par votre login) et ajouter (copie) un ou plusieurs de vos datasets
- Partager ce nouvel historique avec votre voisin
- Modifier votre historique.
- Supprimer l'historique que vous partagez.



## Notions de workflow : convertir un historique en workflow.

### Convertir un historique en workflow

Les principales étapes :

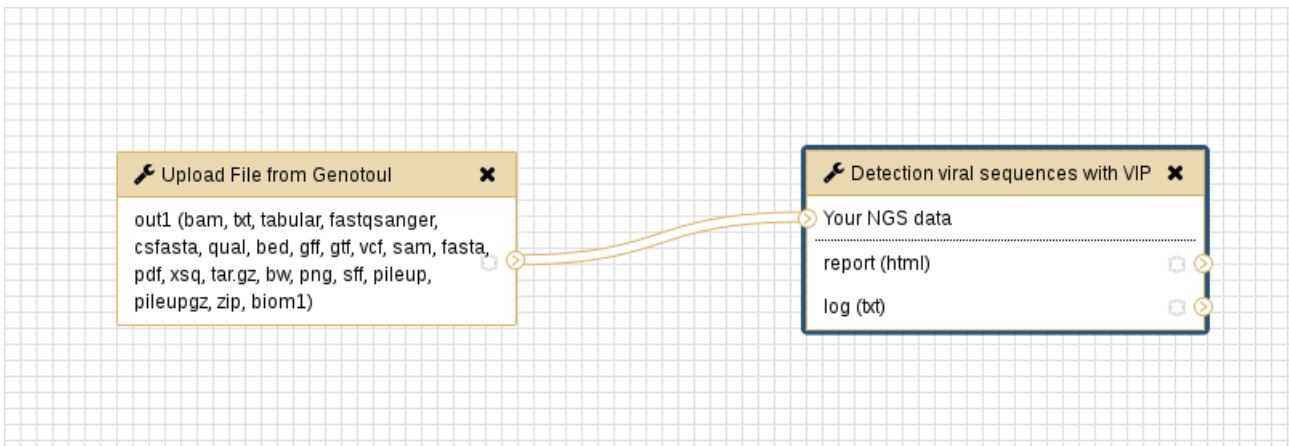
- « History panel » Options → « Extract workflow »
- Sélectionner les bons datasets
- Créer le workflow

Création de workflow :



- A partir de rien : Menu « Workflow » puis « Create a new workflow »
- A partir d'un historique : « History panel » Options → « Extract workflow »

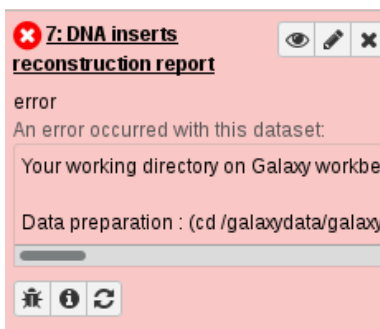
Comme pour les historiques, il est possible de partager des workflows.



### Quelques conseils

Pour vous aider à gérer votre espace de travail, veuillez vous connecter à la plateforme d'auto-formations en ligne <http://sig-learning.toulouse.inra.fr>, vous inscrire à la session « Galaxy », puis lire le chapitre « GOOD PRATICE or How to be a good Galaxy user ? »

Besoin d'aide ? Utilisez de préférence le formulaire web disponible depuis une dataset rouge, sinon, merci d'envoyer un mail à [support.siginae@inra.fr](mailto:support.siginae@inra.fr)



### Report this error to the local Galaxy administrators

Usually the local Galaxy administrators regularly review errors that occur on the se you were trying to do when the error occurred) and a contact e-mail address, we w

Error Report

Your email

Message

Report