



## - Galaxy -

# Initiation à la plateforme Galaxy et préparation du TP RNAseq

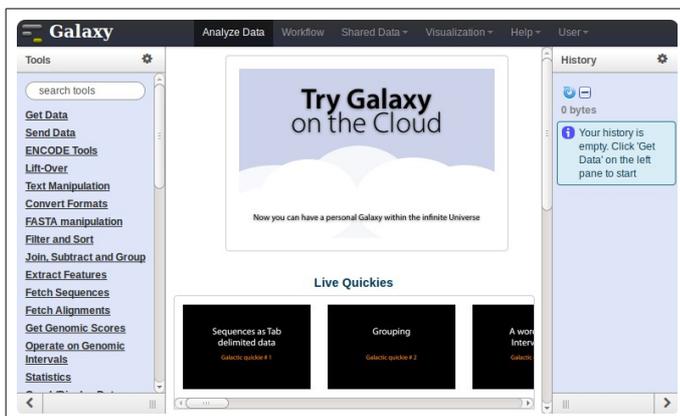
### Remerciements



Fonds Européen  
de Développement Régional



RÉGION  
MIDI-PYRÉNÉES



**Galaxy** plateforme de traitements informatiques et bioinformatiques accessible depuis l'url :

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

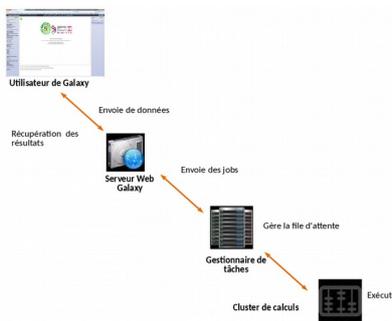


## Déroulement de la matinée « Initiation à Galaxy »

Cette matinée est dédiée :

1. A la prise en main de Galaxy.
2. A la préparation des exercices pour le TP RNAseq.

Clés de lecture de ce document :



Des schémas pour vous expliquer les principes de fonctionnement spécifiques à Galaxy. Les explications des schémas sont notées en dessous, pour vous aider lors de la relecture de ce document, après la formation.



« Lecture facultative »



Vos exercices.



Notes, mémo...



Dans un premier temps, veuillez accéder à votre support au format informatique depuis cette adresse :

[http://genoweb.toulouse.inra.fr/~formation/4\\_Galaxy\\_RNAseq/2018/](http://genoweb.toulouse.inra.fr/~formation/4_Galaxy_RNAseq/2018/)



## Quelques mots sur Galaxy

Galaxy a été créé par l'équipe américaine "Galaxy project" :

- Le Center for Comparative Genomics and Bioinformatics - Penn State,
- Des départements "Biology" et "Mathematics and Computer Science" de l'Université d'Emory.

La communauté autour de cet outil est active. Nous utiliserons l'instance Sigenae/BioInfo Genotoul de Galaxy.



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

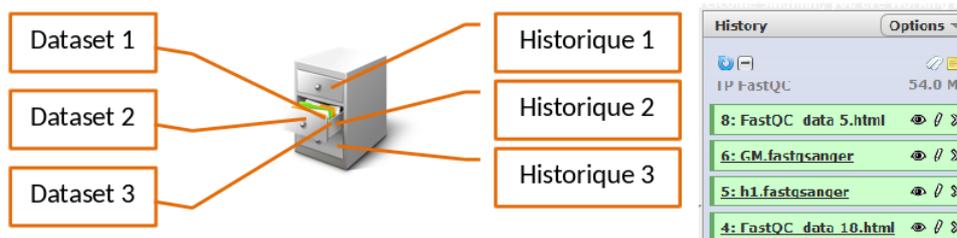


Nous utiliserons un vocabulaire spécifique à Galaxy :

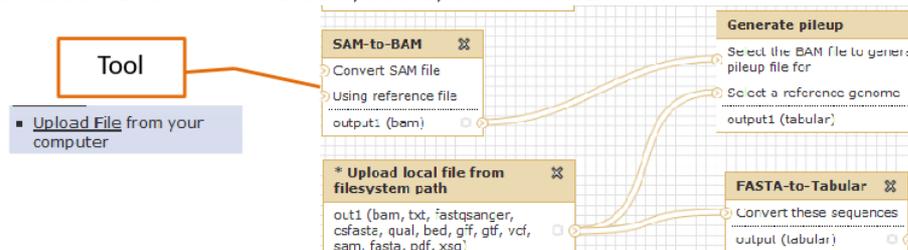
Un **DATASET** est un fichier de données (fichiers d'entrée, fichiers résultats) :



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.



## 1 - Connexion à votre plateforme Galaxy Sigenae / BioInfo Genotoul

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

Cette formation est destinée aux personnes souhaitant traiter des données (bio)informatiques sans connaissances spécifiques en informatique (sans avoir à connaître Linux et la ligne de commande).

Vous découvrirez notamment comment :

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



Veillez vous connecter à Galaxy en tapant cette URL dans votre navigateur :

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



Soit vous avez déjà un compte personnel sur Genotoul, soit vous utilisez un compte de formation :

### J'ai un compte personnel

Si vous utilisez votre compte personnel Genotoul et que vous souhaitez simplifier/modifier votre mot de passe, veuillez suivre le paragraphe « Changer mon mot de passe avec MobaXterm ». Changer son mot de passe est un bon réflexe « sécurité ».

### J'utilise un compte de formation

Logins : anemone arome aster bleuet camelia capucine chardon clematite cobee coquelicot cosmos cyclamen

Password : Demander au formateur.

Ces comptes ne sont valables que le temps de la formation. Nous vous demandons d'utiliser un compte personnel si vous avez besoin de traiter ou stocker des données.

Puis l'identification s'effectue en **deux étapes** :

1- Entrez votre login/mot de passe à l'ouverture de la page internet <http://sigenae-workbench.toulouse.inra.fr/galaxy>

2- Puis entrez de nouveau votre login/mot de passe depuis le menu (barre du haut) « User » puis « Login ».



Bienvenu(e)s sur la page d'accueil de Galaxy :

WELCOME TO 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.

ACCESS TO OLD GALAXY VERSION (2015) : |C|

**Warnings :**

- All jobs running on galaxy are sent to BioInfo Genotoul cluster.
- Your data are stored in work/ directory. Consequently, BioInfo Genotoul platform reserves the right to purge all files not accessed since 120 days on work/ disk space.



## 2 - Pour les utilisateurs qui souhaitent changer leur mot de passe ( MobaXterm )

Pour commencer, nous vous conseillons de télécharger MobaXterm. Sur Google, rechercher « MobaXterm ». Cliquer sur le lien : <http://mobaxterm.mobatek.net/> puis sur “Get MobaXterm now”:



GET MOBAXTERM NOW!

Choisir l'édition gratuite en cliquant sur « Download now ».

### Home Edition

#### Free

Full **X server** and **SSH** support  
Remote desktop (RDP, VNC, Xdmcp)  
Remote terminal (SSH, telnet, rlogin, Mosh)  
X11-Forwarding  
Automatic SFTP browser  
Plugins support  
Portable and installer versions  
Full documentation  
Max. **12** sessions  
Max. **2** SSH tunnels  
Max. **4** macros  
Max. **360** seconds for Tftp, Nfs and Cron



Download now

Puis choisir l'édition portable :

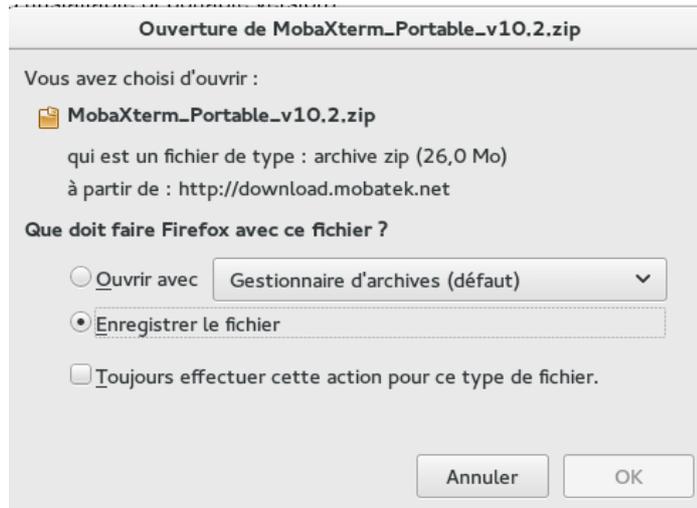
Select the version you want to download (installable or portable version)



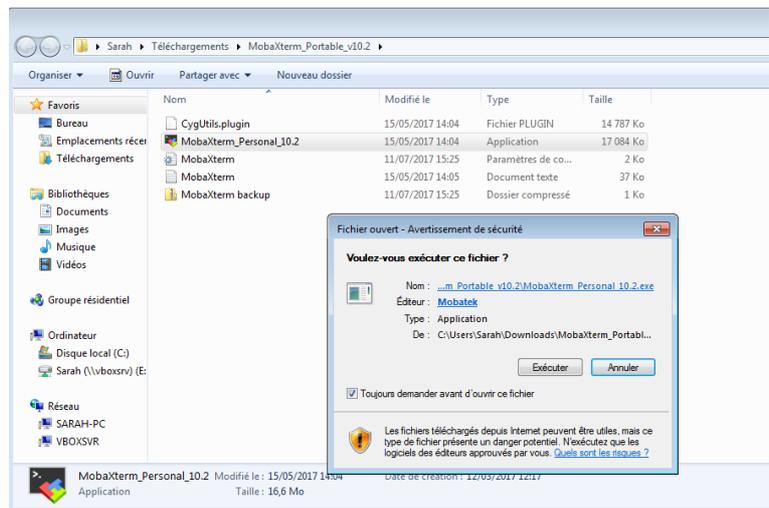
MobaXterm Home Edition v10.2  
(Portable edition)



Enregistrer le zip sur votre ordinateur en cliquant sur OK:



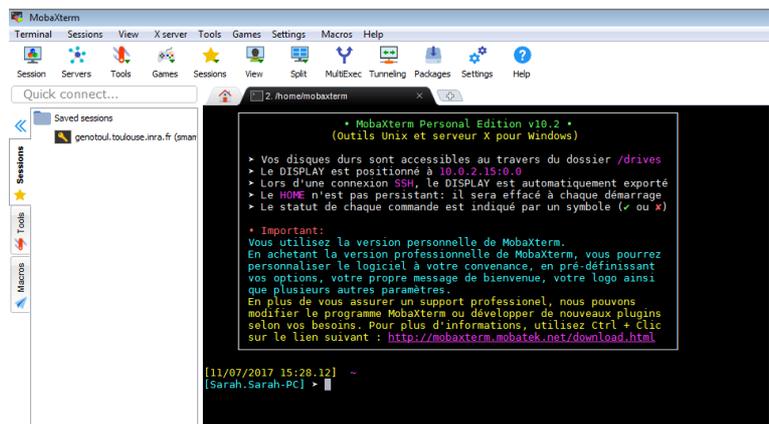
Puis double-cliquer sur le fichier « MobaXterm\_Portable\_v10.2.zip ». Cliquer sur « Extraire » puis sur MobaXterm\_Personal\_10.2.exe



Une fenêtre s'ouvre :



Cliquer sur « Start local terminal ».





Vous êtes sur un terminal. Il vous faut ensuite vous connecter sur Genotoul en tapant la ligne de commande suivante :

ssh -X  
[login@genotoul.toulouse.inra.fr](http://login@genotoul.toulouse.inra.fr)

```
smaman@genotoul.toulouse.inra.fr ~$ ssh -X smaman@genotoul.toulouse.inra.fr
Warning: Permanently added 'genotoul.toulouse.inra.fr' (RSA) to the list of known hosts.
Last login: Mon Jul 10 15:19:32 2017 from Laf31-1-01-56-05-101.fbx.proxad.net

=====
Bienvenue sur les serveurs d'accès de la plate-forme bioinformatique
=====
Les serveurs "genotoul" ne sont PAS destinés à lancer des traitements de données mais uniquement à soumettre des jobs.
Pour cela, utiliser les commandes "qsub" (batch), "qsh" (interactif) ou "qlogin" (redirection graphique).
Tout traitement lancé sur les serveurs genotoul sera immédiatement interrompu par les administrateurs systèmes.
=====
Information concernant les espaces disques
=====
La partition /home est réservée aux fichiers de configuration (quota utilisateur 100Mo)
La partition /save est réservée pour les données à sauvegarder (quota utilisateur 250Go)
La partition /work est réservée pour les fichiers temporaires de calcul (quota utilisateur 1To)
=> ATTENTION CE VOLUME N'EST PAS SAUVEGARDE ET PEUT ÊTRE PURGÉ SI BESOIN
=> IL EST DE VOTRE RESPONSABILITÉ DE GÉRER VOS DONNÉES (organisation, volumétrie, pertinence, ancienneté)
Pour connaître votre consommation d'espace disque, utilisez la commande suivante :
# du -csh /DIR_NAME/USER_NAME/*
=====
Informations concernant l'utilisation de la mémoire
=====
Il existe une limitation de 8Go de RAM par processus utilisateur (sur le cluster).
Pour obtenir plus de mémoire, veuillez consulter la FAQ de notre site web (cf + bas)
=====
Informations concernant le quota de temps de calcul
=====
Il existe un quota de temps de calcul annuel de 100.000 pour les académiques (500h pour les entreprises privées).
Au-delà, il faut renseigner le formulaire de demande de ressources exceptionnelles.
Vous pouvez vérifier votre quota de calcul avec la commande: quota_cpu <login>
=====
Support
=====
Pour plus d'informations, consultez le site web :
http://bioinfo.genotoul.fr/
Pour toute demande de support, adressez-vous à :
support.bioinfo.genotoul@inra.fr
smaman@genotoul2 ~$
```

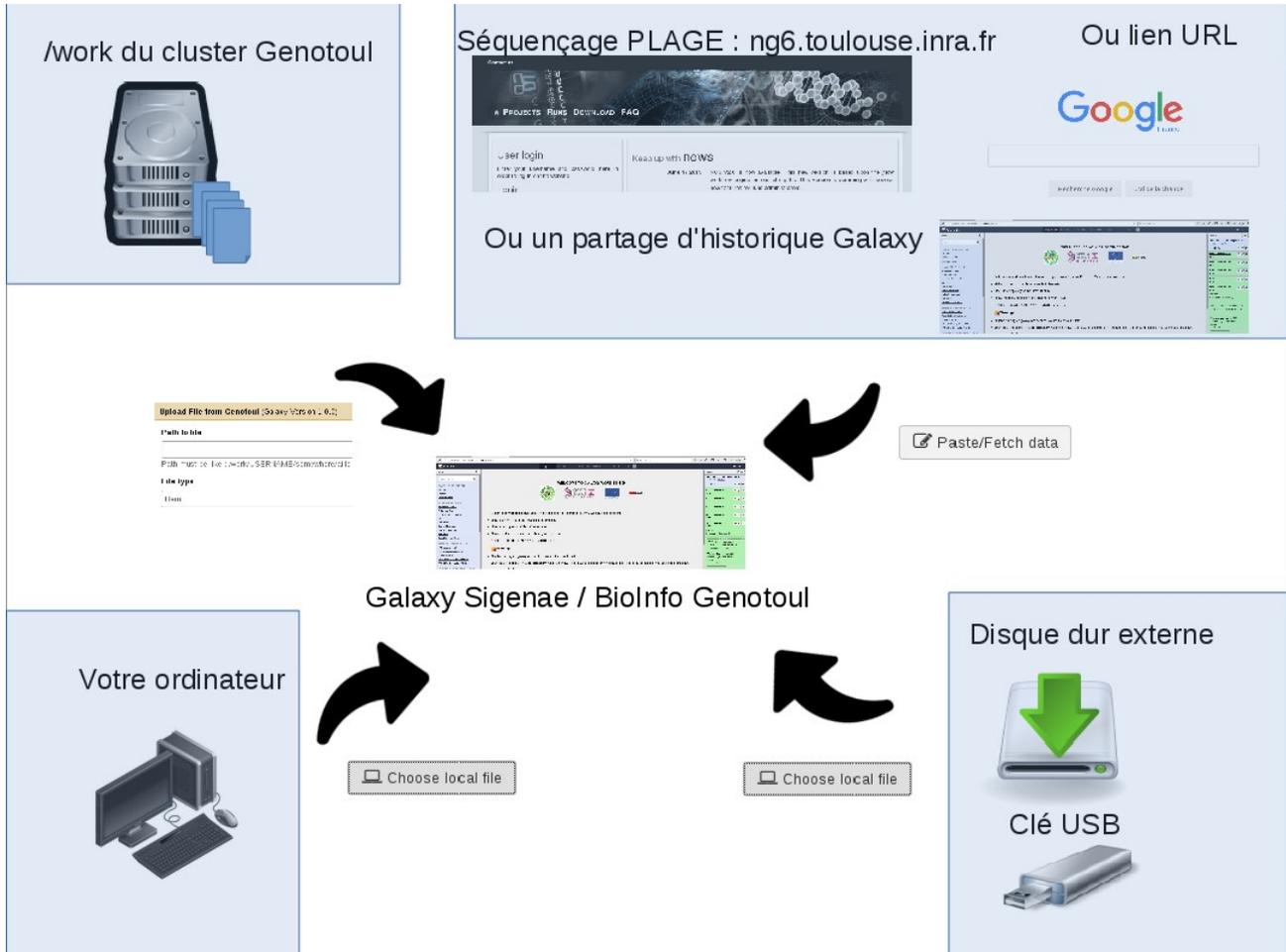
Pour changer votre mot de passe par défaut par un mot de passe plus facilement mémorisable, tapez la commande `passwd`, puis cliquez sur « Entrée ».

```
smaman@genotoul2 ~$ passwd
Changing password for user smaman.
Enter login(LDAP) password:
```

Le terminal vous demande de taper votre ancien mot de passe, puis « Entrée ». Le terminal vous demande de taper votre nouveau mot de passe (simplifié...), deux fois de suite. Votre mot de passe est maintenant mis à jour.



### 3 - Les différentes sources de données pour vos datasets



### 4 - Créer votre espace de travail dans votre interface Galaxy

#### 4-1 - Créer vos deux premiers historiques

L'objectif de cette matinée est de préparer deux historiques nommés « Initiation galaxy » et « TP RNAseq » contenant les données qui seront utilisées en TP les jours suivants.



#### **Nous allons créer un nouvel historique :**

Depuis le menu « History » à droite, cliquez sur « History options » (  petit écrou), choisissez « Create new » et créer ainsi un nouvel historique que vous allez nommé «Initiation Galaxy ».

#### **Puis changer son nom :**

Cliquez une fois sur « Unnamed history », le texte se sélectionne en bleu, taper directement le nouveau nom « Historique Initiation Galaxy » puis cliquer sur la touche « Entrée ». L'enregistrement est automatique.

**Renouveler ces deux étapes pour créer un autre historique « TP RNAseq ».**



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.

Name	Datasets	Tags	Sharing	Size on Disk	Created	Last Updated
Unnamed history		0 Tags		0 bytes	less than a minute ago	less than a minute ago
RNA seq statistics	2	4	0 Tags	34.9 Mb	~ 6 hours ago	6 minutes ago
Test BWA fichiers Gnome	4	0 Tags		9.0 Mb	Apr 06, 2012	1 day ago
Test region promoters	5	0 Tags		23.9 Mb	Mar 08, 2012	Apr 06, 2012
Unnamed history	3	0 Tags		60 bytes	Feb 23, 2012	Mar 09, 2012
Unnamed history	1	1	0 Tags	0 bytes	Mar 07, 2012	Mar 07, 2012
Unnamed history	1	1	0 Tags	16.0 Kb	Feb 22, 2012	Mar 05, 2012

## 4-2- Télécharger des fichiers de séquences fastq depuis Internet vers Galaxy



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.

### Positionnez vous sur l'historique «TP RNAseq» :

Depuis la liste disponible dans le menu « User »/»Saved histories », il suffit de cliquer sur le nom de l'historique «Historique Initiation Galaxy» pour l'ouvrir à droite dans Galaxy.

### Description des données de TP:

L'ensemble des fichiers fastq contenant les lectures qui seront utilisées pour l'analyse des données sur le chromosome 6 de la Tomate.

Il s'agit des deux jeux de données réduits WT (Wild type, paired) et MT (Wild type, paired).

### Récupérez chaque fichier disponible sur Internet :

Une URL complète contient le chemin d'accès au fichier plus le nom et l'extension du fichier concerné. Pour récupérer une URL complète rapidement et sans faire d'erreur, nous vous conseillons de **faire un clic droit sur le nom du fichier et de choisir « copier l'adresse du lien »**. Puis un « coller » (Ctrl V) permet de récupérer cette URL complète.

Les 4 fichiers « fastq » à récupérer sont dans le répertoire :

[http://genoweb.toulouse.inra.fr/~formation/4\\_Galaxy\\_RNAseq/2018/data/reads/](http://genoweb.toulouse.inra.fr/~formation/4_Galaxy_RNAseq/2018/data/reads/)

### Copier/Coller chaque URL complète dans l'outil « Upload File from your computer » :

Cliquez sur « Paste/Fetch » data. Puis copier/coller l'URL COMPLETE d'accès au fichier.

Il est possible de renseigner plusieurs URL en sautant une ligne entre chaque URL puis exécuter l'outil.





### Lancer le téléchargement :

Puis lancer le téléchargement en cliquant sur « Start » puis sur « Close ». Le fichier sera bientôt disponible dans votre fenêtre « History ».

### Download from web or upload from disk

Regular Composite

You added 1 file(s) to the queue. Add more files or click 'Start' to proceed.

Name	Size	Type	Genome	Settings	Status
New File	371 b	Auto-detect	unspecified (?)		0%

You can tell Galaxy to download data from web by entering URL in this box (one per line). You can also directly paste the contents of a file.

```
http://genoweb.toulouse.inra.fr/~formation/LigneCmd/RNAseq/data/reads/MT_rep1_1_Ch6.fastq.gz
http://genoweb.toulouse.inra.fr/~formation/LigneCmd/RNAseq/data/reads/MT_rep1_2_Ch6.fastq.gz
http://genoweb.toulouse.inra.fr/~formation/LigneCmd/RNAseq/data/reads/MT_rep1_1_Ch6.fastq.gz
```

Type (set all): Auto-detect Genome (set all): unspecified (?)

Choose local file Paste/Fetch data Pause Reset Start Close

### Renommer vos datasets:

Pour gérer vos fichiers de données, nous vous conseillons de renommer chaque dataset. Ainsi, veuillez s'il vous plaît renommer chaque dataset Galaxy avec un nom explicite mais court, sans caractère spécial ni accent. Exemple : «MT\_rep1\_1\_Ch6»

Renommer chaque dataset avec un intitulé plus court et parlant, en cliquant sur l'icône

« stylo »

### Veillez au bon format de fichier:

Choisissez le bon format pour chaque dataset, seulement si ce dernier n'est pas autodetecté.

L'ensemble des outils permettant l'import dans Galaxy est disponible dans la section « 1- Upload your data => Get data »



L'outil « **Upload File** » télécharge en copiant votre fichier sur le serveur Galaxy. Cette méthode d'upload n'est pas recommandée car elle impacte considérablement votre quota.

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



TP RNAseq	
4 shown	
1.38 GB	<input checked="" type="checkbox"/>
<b>4: WT_rep1_2_Ch6.fastq</b>	
<b>3: WT_rep1_1_Ch6.fastq</b>	
<b>2: MT_rep1_2_Ch6.fastq</b>	
<b>1: MT_rep1_1_Ch6.fastq</b>	

*Display*      *Modify datatype*

*delete the file (can be recovered)*

*download the file*

*re-run an analysis with the exact same parameters !!extremely useful!!*

*details about how this file was generated*

Tool: bamCorrelate	
Name:	bamCorrelate on data 5, data 4, and others
Created:	Dec 11, 2013
Filesize:	64.9 KB
Dbkey:	hg19
Format:	png
Galaxy Tool Version:	1.0.1
Tool Version:	
Tool Standard Output:	<code>stdout</code>
Tool Standard Error:	<code>stderr</code>
Tool Exit Code:	0
API ID:	a50c14e4ca28bfa0
Input Parameter	
Bam file	6: IMR90_H3K36me3.bam
Label	

History	
48:	heatmapper image
47:	heatmapper sorted/filtered regions
46:	heatmapper matrix of heatmap values
45:	heatmapper image
44:	computeMatrix on data 32 and data 34 regions
43:	computeMatrix on data 32 and data 34 column
42:	computeMatrix on data 32 and data 34
41:	computeMatrix on data 32 and data 34
40:	profiler image
39:	heatmapper image
38:	computeMatrix on data 32 and data 34
37:	Extract reads

source : <http://deeptools.readthedocs.io/>

## 5 - Télécharger la référence GTF d'Internet à Galaxy en passant par votre /work

### 5-1- D'Internet à votre ordinateur

#### Accédez aux fichiers sur Internet:

Connectez vous à l'adresse à laquelle les fichiers « ITAG2.3\_genomic\_Ch6.fasta » et ITAG\_pre2.3\_gene\_models\_Ch6.gtf sont disponibles dans ce répertoire:

[http://genoweb.toulouse.inra.fr/~formation/4\\_Galaxy\\_RNAseq/2018/data/reference/](http://genoweb.toulouse.inra.fr/~formation/4_Galaxy_RNAseq/2018/data/reference/)

#### Téléchargez les fichiers sur votre ordinateur:

Veuillez faire un clic droit sur le nom du fichier pour obtenir l'adresse de téléchargement, puis « enregistrer la cible du lien sous » (« Copy link location »), puis enregistrer le fichier sur votre ordinateur.





Il est conseillé de ne pas ouvrir ou modifier vos fichiers sous Windows car ce dernier ajoute des caractères spéciaux qui ne sont pas pris en charge par le cluster de calculs (linux).

## 5-2- De votre ordinateur à votre /work sur Genotoul

Nous allons ensuite transférer le fichier «ITAG\_pre2.3\_gene\_models\_Ch6.gtf» de votre ordinateur à Galaxy. Pour ce faire, veuillez ouvrir le logiciel FileZilla (ou WinSCP). Cet outil vous permettra de voir le contenu de vos répertoires sur Genotoul et de gérer les permissions sur ces répertoires et fichiers.

### Ouvrir FileZilla:



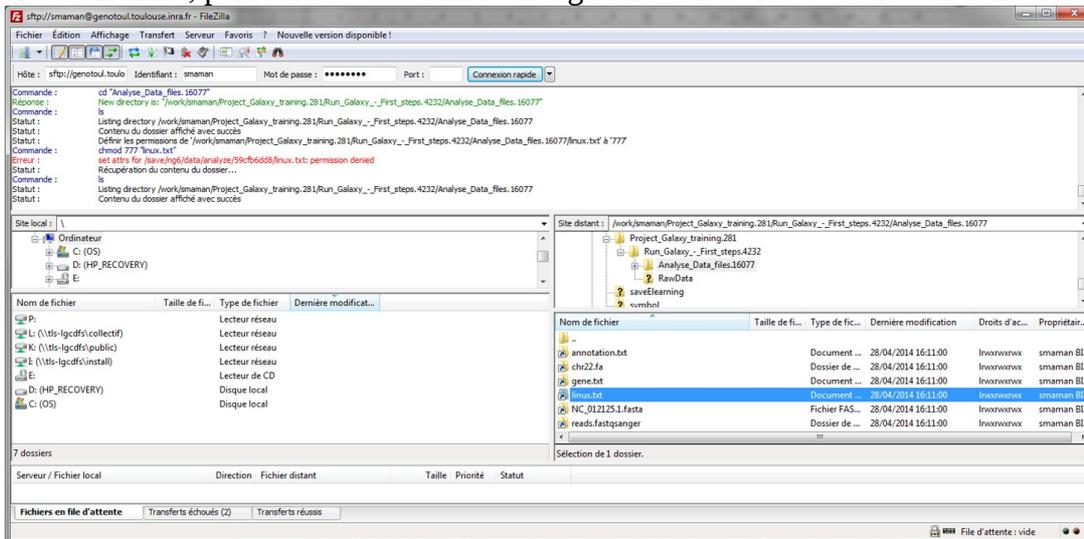
FileZilla vous permettra de transférer votre fichier GTF de votre ordinateur (à gauche de la fenêtre FileZilla) à votre /work sur Genotoul (à droite).

### Connectez vous à votre /work sur Genotoul avec les paramètres suivants:

Hôte :	Identifiant :	Mot de passe :	Port :	Connexion rapide
--------	---------------	----------------	--------	------------------

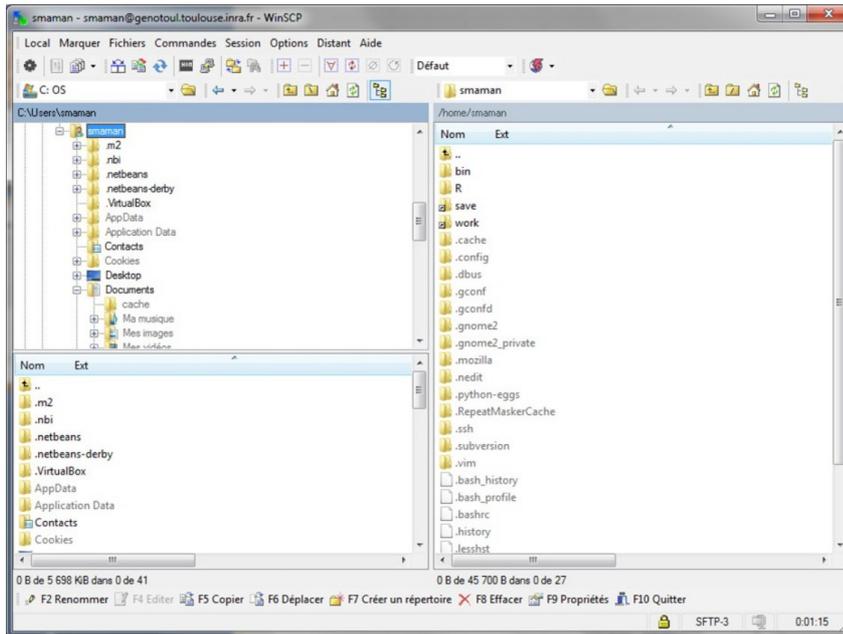
- Hôte : genotoul.toulouse.inra.fr
- Identifiant : Votre login sur Genotoul
- Mot de passe : Votre mot de passe sur Genotoul
- Port : 22

Avec FileZilla, parcourir votre ordinateur à gauche et le serveur Genotoul à droite :





Avec WinSCP :



**Créer un répertoire « galaxy » dans votre /work sur Genotoul:**

A droite de la fenêtre FileZilla, dans Genotoul, créer un répertoire galaxy/ dans votre /work

**Déplacer votre fichier GTF dans le répertoire « galaxy »:**

Déplacer (cliquez sur le nom du fichier et glissez) chaque fichier dans ce répertoire galaxy/ nouvellement créé.

**Donner les bons droits à votre répertoire « galaxy » :**

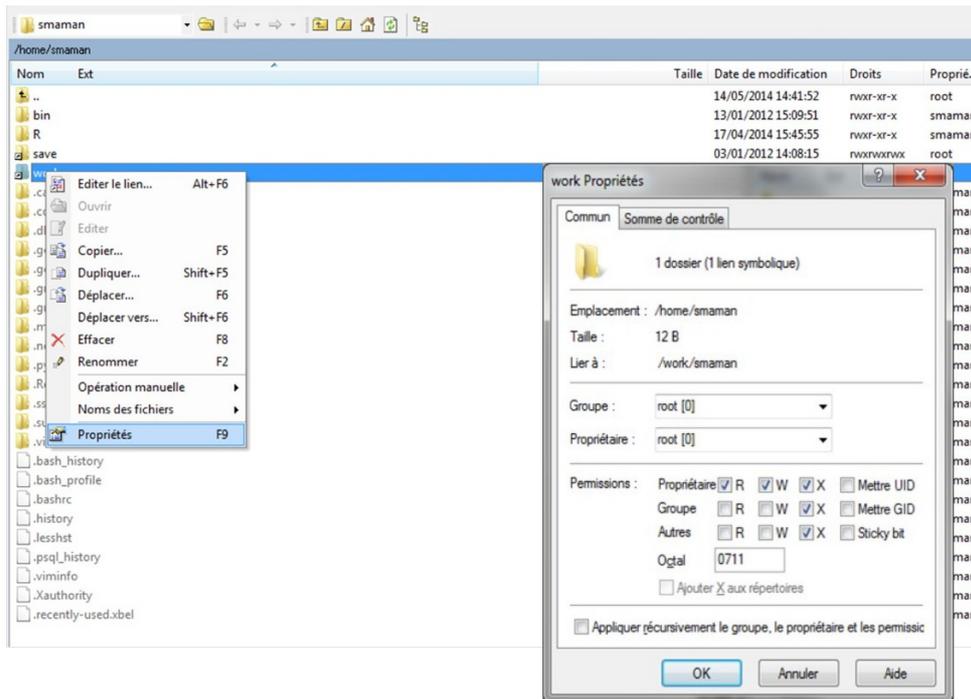
A l'aide d'un clic droit sur le répertoire /galaxy, veillez à ce que le répertoire galaxy/ ai bien les permissions « x » (exécution) pour tous. De même pour votre /work. Puis vérifier, de la même manière, que l'ensemble des fichiers contenus dans galaxy/ (uniquement) aient bien les droits « r »(lecture).

Pour modifier les droits sur votre /work/username/ , clic droit sur votre /work/username/, puis « Droits d'accès au fichier », puis donner les droits d'exécution (X) sur votre /work.

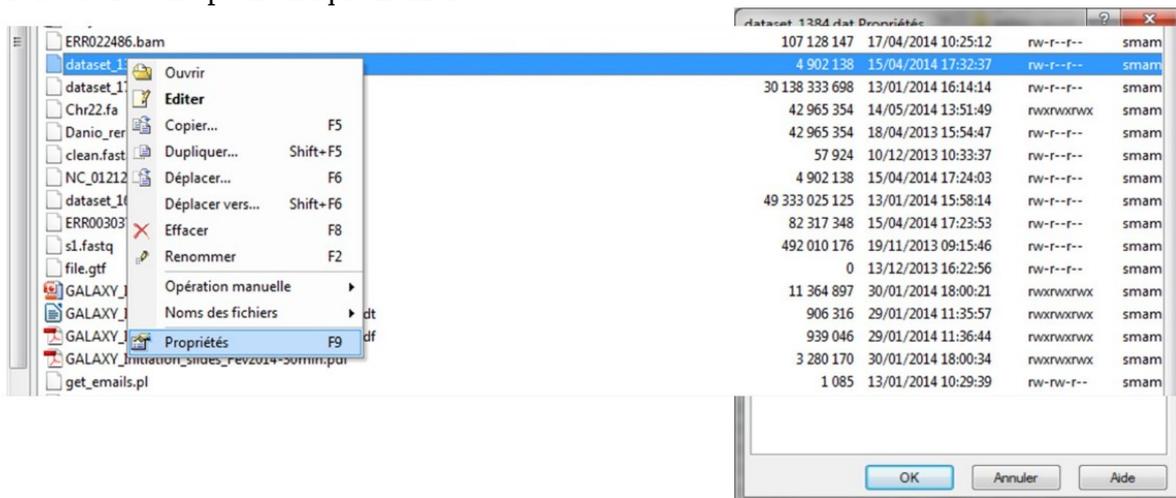
**Donner les bons droits aux fichiers:**

Pour modifier les droits du fichier, clic droit sur le nom du fichier, puis « Droits d'accès au fichier », puis donner les droits de lecture au fichier concerné.





Droits de lecture pour chaque fichier :



Si cela n'est pas fait, l'outil d'upload de Galaxy ne sera pas en mesure d'accéder ni de lire les fichiers que vous souhaitez télécharger et sur lesquels vous allez travailler. Après l'exécution de l'outil « upload » de Galaxy, la dataset sera rouge (donc en erreur).

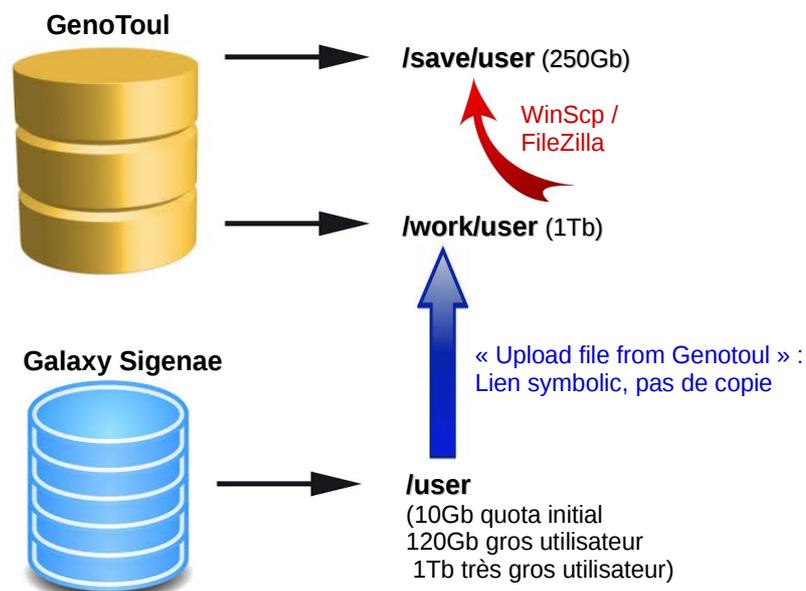


### 5-3- De votre /work à Galaxy

L'outil « Upload file from Genotoul » vous permet de créer un lien symbolique, depuis votre work, sur le serveur Galaxy, sans avoir besoin de copier vos données sur le serveur Galaxy.

Grâce à cet outil, vous économisez de l'espace disque et optimisez votre quota sur Galaxy.

**Important – les droits :** Les droits d'exécution sur le répertoire et de lecture sur les fichiers sont nécessaires pour que vos données puissent être accessibles dans Galaxy. (chmod +x répertoire et chmod +r fichier).



**Chemin d'accès à linux.txt :** Le chemin doit être complet (nom du fichier compris) et pointer sur le work (et non sur le /save ou le /home) afin que le cluster puisse, par la suite, travailler sur ce fichier.

**Important – les formats de fichier :** Les outils Galaxy qui prennent en entrée des fichiers « textes tabulés », ne verront pas vos fichiers textes si le type du fichier n'est pas correctement spécifié (format « tabular »).

#### De votre /work sur Genotoul à votre interface Galaxy :

Sélectionnez l'outil « Upload file from Genotoul » dans la section « Get data ».

##### Get Data

Upload File from your computer

Upload File from Genotoul

#### Complétez le chemin d'accès au fichier:

Le chemin d'accès au fichier (« Path to file ») doit être complet (nom du fichier compris) et pointer sur le work (et non sur le /save ou le /home) afin que le cluster puisse, par la suite, travailler sur ce fichier.





Upload File (version 1.0.0)

**Path to file:**  
  
Path must be like : /work/USERNAME/somewhere/afile

**File type:**  
Bam ▼

Execute

Exemple de **Path to file** : /work/rose/galaxy/ ITAG\_pre2.3\_gene\_models\_Ch6.gtf

**Renseignez le format du fichier:** Complétez le « File type » : gtf

Renommer le fichier Galaxy : « ITAG\_pre2.3\_gene\_models\_Ch6.gtf » à la place de « [http://genoweb.toulouse.inra.fr/~formation/4\\_Galaxy\\_RNAseq/2018/data/reference/ITAG\\_pre2.3\\_gene\\_models\\_Ch6.gtf](http://genoweb.toulouse.inra.fr/~formation/4_Galaxy_RNAseq/2018/data/reference/ITAG_pre2.3_gene_models_Ch6.gtf) »



L'outil «Upload file from Genotoul» vous permet de créer un lien symbolique, depuis votre work, sur le serveur Galaxy, sans avoir besoin de copier vos données sur le serveur Galaxy. Grâce à cet outil, vous économisez de l'espace disque et optimisez votre quota sur Galaxy.



Avec WinSCP, il est possible de récupérer ce chemin complet avec un clic droit sur le nom du fichier, « Noms des fichiers » puis « Copier dans le Presse-papier (inclus les chemins) ».

Fenêtre WinSCP :

File Name	Size	Modified	Permissions	User
Danio_rerio.Zv9.62.dna.chromosome.22.fa	42 965 354	18/04/2013 15:54:47	rw-r--r--	smam
clean.fastq	57 924	10/12/2013 10:33:37	rw-r--r--	smam
NC_0121...	4 902 138	15/04/2014 17:24:03	rw-r--r--	smam
dataset_1	49 333 025 125	13/01/2014 15:58:14	rw-r--r--	smam
ERR0030...	82 317 348	15/04/2014 17:23:53	rw-r--r--	smam
s1.fastq	492 010 176	19/11/2013 09:15:46	rw-r--r--	smam
file.gtf	0	13/12/2013 16:22:56	rw-r--r--	smam
GALAXY...	11 364 897	30/01/2014 18:00:21	rw-rw-rw-	smam
GALAXY...	906 316	29/01/2014 11:35:57	rw-rw-rw-	smam
GALAXY...	939 046	29/01/2014 11:36:44	rw-rw-rw-	smam
GALAXY...	3 280 170	30/01/2014 18:00:34	rw-rw-rw-	smam
get_email...	1 085	13/01/2014 10:29:39	rw-rw-r--	smam
attention	612	14/01/2014 15:32:21	rw-r--r--	smam
book.png	2 304	14/01/2014 15:42:52	rw-r--r--	smam
warning.png	538	14/01/2014 15:42:06	rw-r--r--	smam
contig.res	134	10/12/2013 16:14:24	rw-r--r--	smam
mailing_list_galaxy_prod.txt	618	13/01/2014 10:30:34	rw-r--r--	smam
mailing_liste_galaxy_prod_1301204.txt	5 454	13/01/2014 10:35:47	rw-r--r--	smam



## 6 – Historique « initiation Galaxy » : télécharger des données de l'UCSC vers Galaxy

### 6-1 – Télécharger des données de l'UCSC vers Galaxy



**Télécharger l'annotation (gènes RefSeq) du chromosome 1 bovin (btau4) avec l'outil « UCSC Main table browser » dans l'historique « Intitiation Galaxy » :**

Voici les paramètres :

- Clade : Mammal
- Genome : Cow
- Assembly : Oct. 2007
- Group : Genes and Genes Prediction Tracks
- Track : RefSeq Genes
- Table : refGene
- Region – position : chr1:1-161106243 (enter « chr1 » puis cliquer sur « lookup »)
- Output format : BED – browser extensible data

Sélectionner « Send Output to Galaxy » puis cliquer sur « Get Output » et « Send query to Galaxy ».

**Visualiser ce nouveau dataset:**

Visualiser notamment ses propriétés (« database »). Que remarquez vous ?

Explorer les liens disponibles pour ce dataset.

Relancer le téléchargement en modifiant le Output format : GTF – gene transfert format

Comparer les deux fichiers GTF et BED.

### 6-2 – Quelques traitements bioinformatiques



**Extraire le génome d'une séquence, avec l'outil « Extract Genomic DNA ».**

A partir des deux datasets BED et GTF « UCSC Main on Cow: refGene (chr1:1-161106243) » précédemment importés, extraire du génome la séquence de chacun des gènes (output data type « Interval »).

Constater et expliquer le nombre de lignes du fichier obtenu.

**Conversion de format:**

Convertir le dataset obtenu à partir du BED en multi-fasta à l'aide de l'outil « Tabular-to-FASTA converts tabular file to FASTA format ».

**Recherche des régions promotrices :**

A partir du fichier d'annotation BED (sans la séquence) utiliser l'outil « Get flanks » pour extraire les régions en amont de chaque gène (longueur 1kb avec un offset de 100pb)

### 6-3 – Partager votre historique

Nous allons copier deux historiques préparés à l'avance :

Depuis le menu « History » à droite, cliquez sur « History options » (  petit écrou),



choisissez « Histories Shared With Me », puis cliquez sur la petite flèche noire à droite du nom de l'historique « Historique R1R2 », choisissez « copy ».

L'historique est en train de se copier chez vous. Pour retrouver cette copie, cliquez dans le menu tout en haut « User », puis « Saved Histories ». Votre historique nouvellement copié est maintenant disponible dans la liste qui s'affiche.

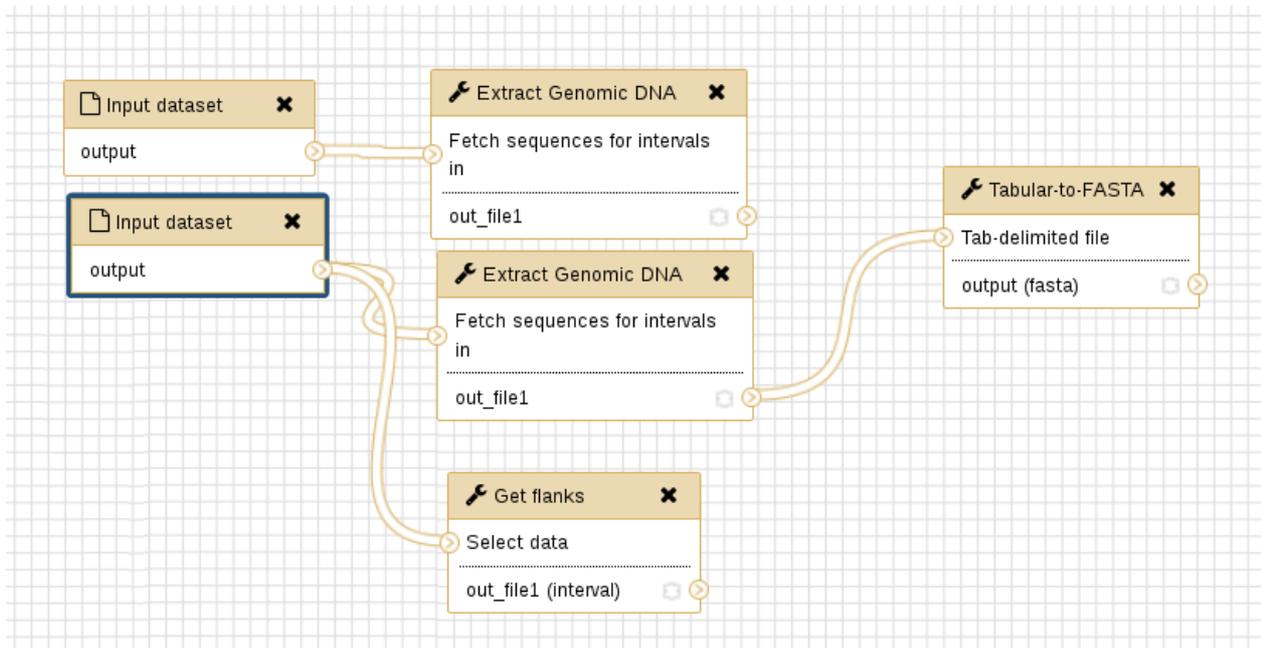
## 6-4 – Convertir votre historique en workflow

Créer un workflow à partir des traitements bioinformatiques précédemment réalisés.



### Créer un workflow depuis l'historique « Initiation Galaxy » existant :

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



### Créer un workflow depuis une page blanche :



- 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.



## 6-5 – Exécuter un workflow



Pour exécuter un workflow, il suffit de se positionner dans un historique contenant vos fichiers entrants, puis menu «Workflow», sélectionner un workflow dans la liste, puis cliquer sur la petite flèche noire, choisir «Run».

## 7 – Conseils pour gérer au mieux votre quota

### **Archiver vos datasets dans votre save/:**

Sauver les datasets générés dans votre compte sur Genotoul, dans votre /work/YourUserName/.

Attention, les données de votre /work sont purgées au bout de 120 jours si les fichiers ne sont pas utilisés.



Puis les déplacer de votre work/ à votre save/ à l'aide de FileZilla.

Quotas Galaxy :

- Première utilisation : 10 Gb
- Gros utilisateur : 120 Gb
- Très gros utilisateur : 1000 Gb



Galaxy gère très mal les caractères spéciaux et les accents.

### Pour répondre à vos questions:

- Mail : [support.siginae@inra.fr](mailto:support.siginae@inra.fr) ou le « bug » de la dataset Galaxy.
- 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 (  petit écrou).

## 8 – Formations en ligne sig-learning

Pour accéder aux formations en e-learning, veuillez suivre la démarche suivante :

- 1- **Tapez l'adresse de « sig-learning »** : <http://sig-learning.toulouse.inra.fr/>
- 2- **Authentification** avec votre compte de formation ou votre compte Genotoul :
- 3- **Ajoutez des formations à votre compte** (si cela n'est pas déjà fait):

Une fois connecté sur la plateforme d'e-learning, pour accéder aux formations proposées, veuillez suivre cette démarche :

- depuis l'onglet "trainings",



- à droite au niveau de "My account", cliquez sur "Trainings management (pour vous inscrire à de nouvelles formations de la plateforme Genotoul)".
- et enfin + "Subscribe to training".

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 :

Dans "Search" , indiquer "%" pour lister l'ensemble des formations disponibles puis cliquer sur le tableau vert des formations qui vous intéressent pour vous y inscrire.