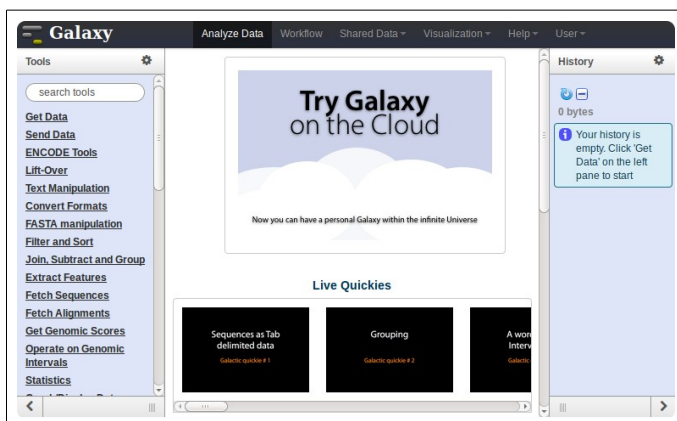




- Galaxy -

Initiation à la plateforme Galaxy et préparation du TP FROGS



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

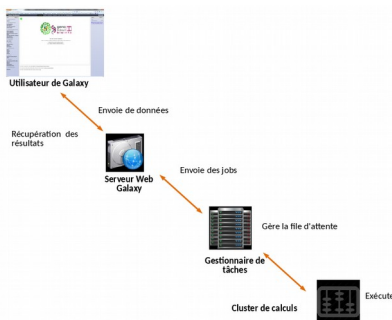


Déroulement de la matinée « Initiation à Galaxy »

Cette matinée est dédiée :

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

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/15_FROGS/FROGS_ini/6-Nov2017/



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 Sigene/BioInfo Genotoul de Galaxy.



Utilisateur de Galaxy

Envoie de données

Récupération des
résultats



Serveur Web
Galaxy

Envoie des jobs



Gère la file d'attente

Gestionnaire de
tâches

Cluster de calculs

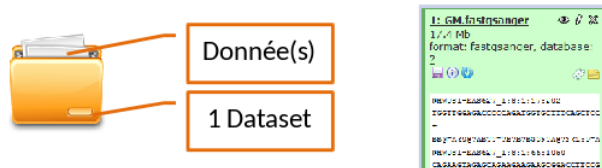


Exécute

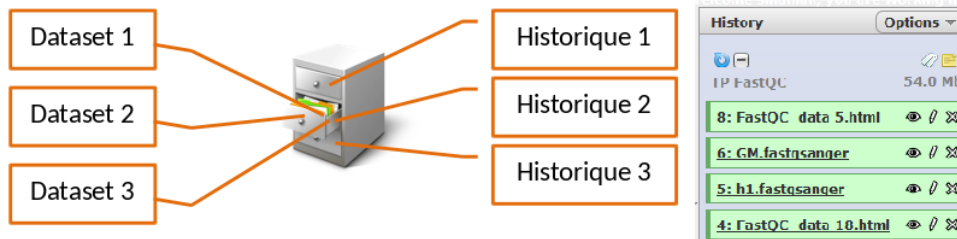


Nous utiliserons un vocabulaire spécifique à Galaxy :

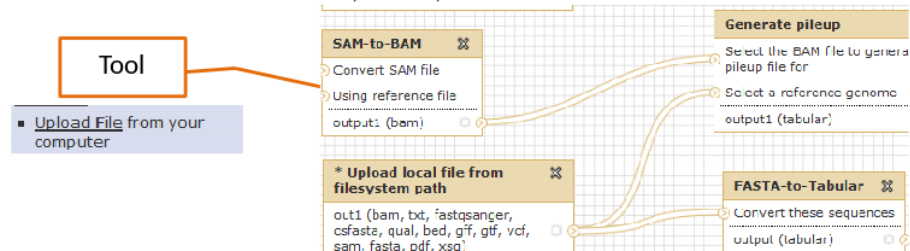
Un **DATASET** est un fichier de données (fichiers d'entrée, fichiers résultats) :



Votre **HISTORIQUE** est un « répertoire » qui « liste » l'ensemble de vos fichiers de donnée (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 Siganae / 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



Veuillez 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, n'hésitez pas à demander à un formateur. 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.

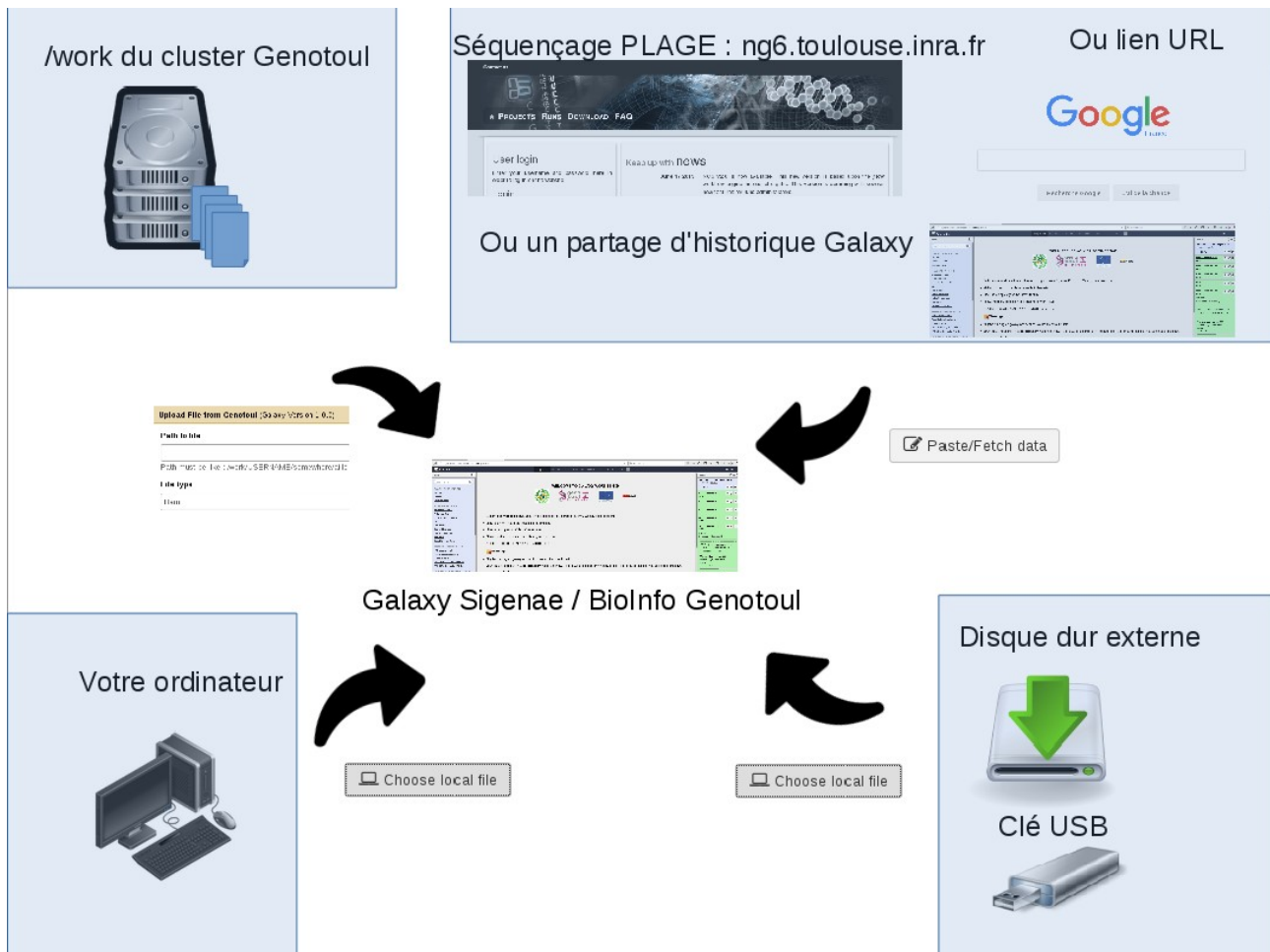
Puis l'identification s'effectue en **deux étapes** :

- 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 :



2 - Les différentes sources de données pour vos datasets



3 - Les 4 historiques FROGS

L'objectif de cette matinée est de préparer 4 historiques FROGS dans lesquels vous allez travailler les jours suivants :

- « historique R1R2 » contenant 2 datasets.
- « historique multiplex » contenant 2 datasets.
- « historique 454 » contenant 1 dataset.
- « historique merged » contenant 1 dataset.



Voici un aperçu de votre plateforme Galaxy en fin de matinée :

Galaxy interface showing a workflow history with the following datasets:

- D-historique merged** (1 shown, 83.8 MB): 1:100spec_90000seq_9samples.tar (data, database: 2, uploaded data file)
- C-historique 454** (1 shown, 26.1 MB): 1:454.fastq (26.1 MB, format: fastq, database: 2, uploaded fastq file)


```

@RR443364.1.2 0/2UCFN1G1C5KH Leng:th=90
AGTCTAGTAGTACGGAGGAGCGAGCTGCGGAATTTTGGCAATGG
+SR443364.1.2 0/2UCFN1G1C5KH Leng:th=90
FFFFFFFFFFFFFF<881:FFFFFFFFDD@@@HHGQIHH
@RR443364.2.2 0/2UCFN1G1A97N Leng:th=446
AGTCTAGTAGTACGGAGGAGCGAGCTGCGGAATCTTGGCAATGG
      
```
- B-historique multiplex** (2 shown, 2.12 MB): 2:barcode_forward.tabular (10 lines, format: tabular, database: 2, uploaded tabular file)


```

1 2 3
MGA468001 ACACGCT TOTAGCT
      
```

 1:multiplex.fastq (2.1 MB, format: fastq, database: 2, uploaded fastq file)


```

@HHHSHK001ALDDH
ATCTAGTATAAGTTCCTCCTAAGTCCCTAATTAACCAGA
+
FFFFFFFFFFDAS4444889422=>+400044444>;;68898
@HHHSHK001B88SLE
ATAGCTGATTGGTTAAGCGGATAGGGATTAGATACCGCTAGCTG
      
```
- A-historique RL R2** (2 shown, 126.19 MB): 2:sampleA_R2.fastq (63.1 MB, format: fastq, database: 2, uploaded fastq file)


```

@ERR619083.N00794.157.000000000-A5HJE:1.1101.1
NCACGAGCGCGGTAACTACCTAGGCTCGGAGGTTAATCGGAATT
+ERR619083.N00794.157.000000000-A5HJE:1.1101.1
#>AA7CFADBD7EFGGGGOFHGF3FEFEGGOFHMG0G0GH
@ERR619083.N00794.157.000000000-A5HJE:1.1101.1
NCACGAGCTGCGGTAACTAGCTAGGCTCGGAGGTTAATCGGAATT
      
```

 1:sampleA_R1.fastq (63.1 MB, format: fastq, database: 2, uploaded fastq file)


```


@ERR619083.N00794.157.000000000-A5HJE:1.1101.1
TCCCGCTCAATTCCTTTGAGTTTTAATCTGCGAGCGACTCC
+ERR619083.N00794.157.000000000-A5HJE:1.1101.1
CCBCCCCCFGGGGGHHHHHHHHHHHHHHHHHHHGGGHHHH
@ERR619083.N00794.157.000000000-A5HJE:1.1101.1
GOTTCGCTCAATTCCTTTGAGTTTTAATCTGCGAGCGACTCC
      
```
- historique merged** (1 shown, 83.8 MB): 1:100spec_90000seq_9samples.tar (data, database: 2, uploaded data file)

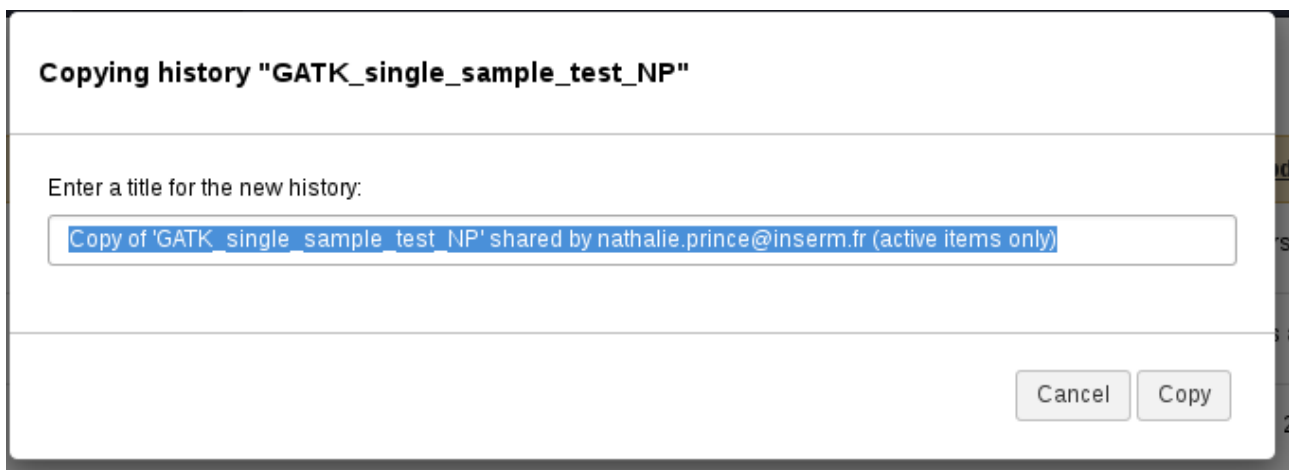


3 – 1 – Récupérer les historiques « R1R2 » et « multiplex » sur votre compte

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 ».



The screenshot shows a dialog box with the title "Copying history 'GATK_single_sample_test_NP'". Below the title, there is a text input field with the placeholder text "Enter a title for the new history:". The input field contains the text "Copy of 'GATK_single_sample_test_NP' shared by nathalie.prince@inserm.fr (active items only)". At the bottom right of the dialog box, there are two buttons: "Cancel" and "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.

Pour ouvrir l'« Historique R1R2 » sur la droite, il suffit simplement de cliquer sur son nom.




Veuillez renouveler ces mêmes étapes pour l'historique « multiplex ».

3 – 2 – Créer votre premier historique «merged» et télécharger un fichier depuis internet

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 nommer « Historique merged ».

Puis changer son nom :

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



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 « merged » :

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

Récupérez le fichier disponible sur Internet :

Nous allons utiliser l'outil « Upload File from your computer » :

MANAGE YOUR DATA FILES
Get Data
Upload File from your computer

Cliquez sur « Paste/Fetch » data. Puis copier/coller l'URL COMPLETE d'accès au fichier (copiez/collez ce lien):

http://genoweb.toulouse.inra.fr/~formation/15_FROGS/FROGS_ini/DATA/100spec_90000seq_9samples.tar.gz

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. Puis lancer le téléchargement. Le fichier sera bientôt disponible dans votre fenêtre « History ».



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



Organisez votre espace de travail:

Pour gérer vos fichiers de données, nous vous conseillons de renommer chaque dataset. Ainsi, veuillez s'il vous plaît renommer la dataset Galaxy en « 100spec_90000seq_9samples.tar.gz ».

Veillez au bon format de fichier:

Choisissez le format « tar » pour cette dataset.

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 ».

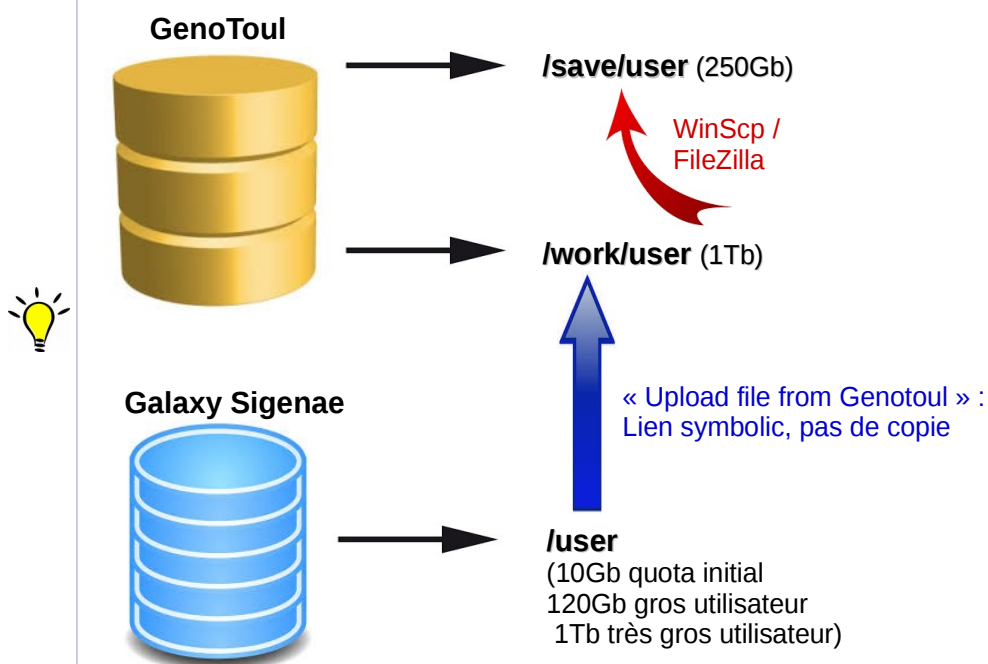


3 – 3 – Créer l'historique « 454 » : Télécharger un fichier d'Internet à votre /work puis de votre /work à Galaxy

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 REPERTOIRE et chmod +r FICHER)



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 »).

3-3-1- D'Internet à votre ordinateur

Accédez au fichier sur Internet:

Connectez vous à l'adresse à laquelle le fichier « 454.fastq » est disponible:

http://genoweb.toulouse.inra.fr/~formation/15_FROGS/FROGS_ini/DATA/

Téléchargez ce fichier « 454.fastq » 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.

3-3-2- De votre ordinateur, à votre /work sur Genotoul puis à Galaxy

Nous allons ensuite transférer ce fichier « 454.fastq » 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:



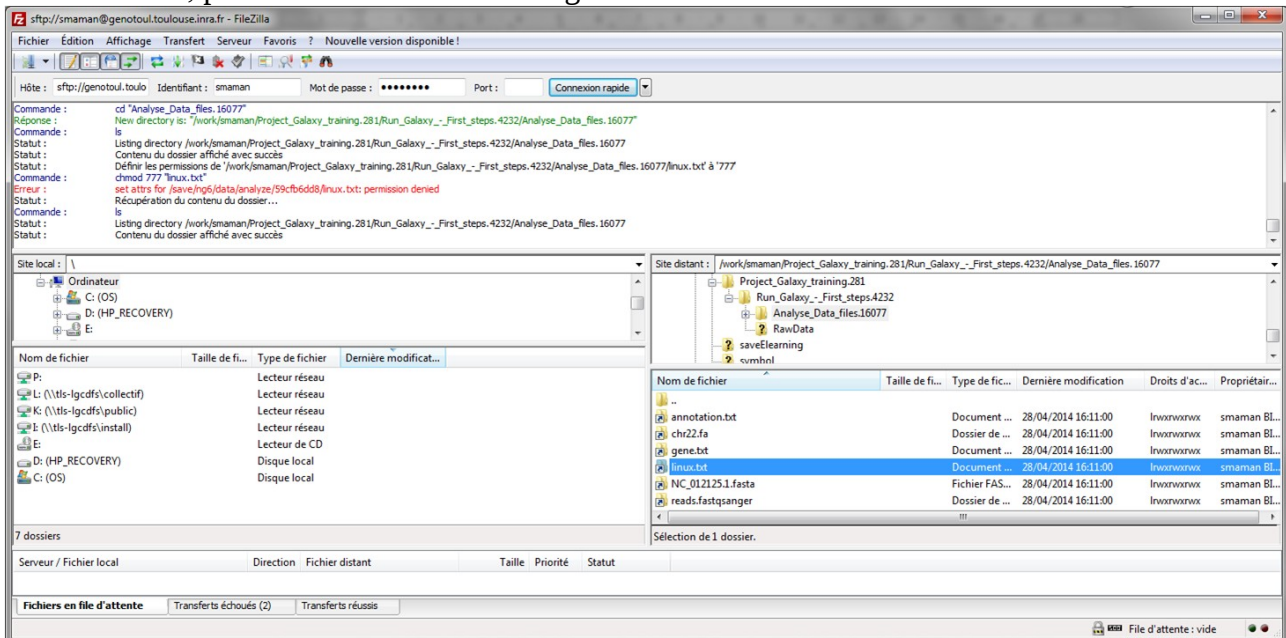
FileZilla vous permettra de transférer votre fichier « 454.fastq » de votre ordinateur (à gauche de la fenêtre FileZilla) à votre /work sur Genotoul (à droite).

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

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

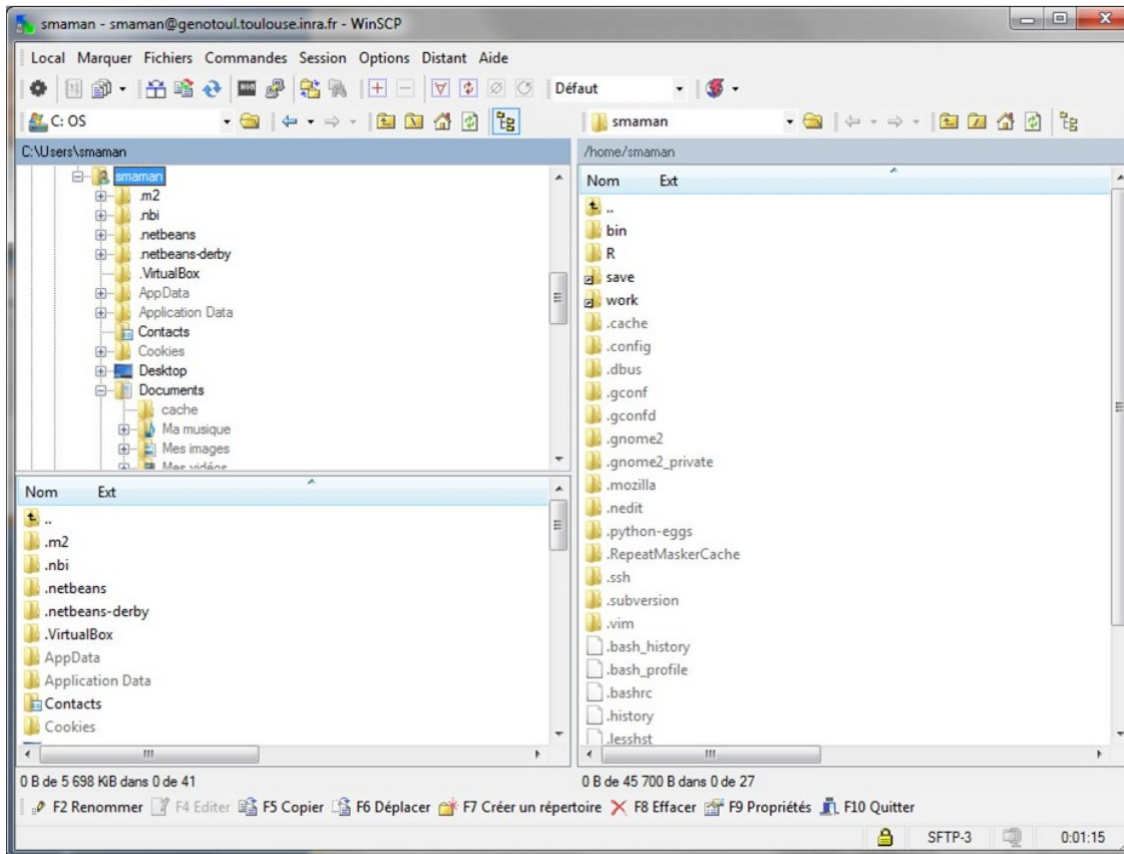
- Hôte : genotoul.toulouse.inra.fr
- Identifiant : Votre login sur Genotoul
- Mot de passe : Votre mot de passe sur Genotoul
- Port : 22

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





Avec WinSCP :



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

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

Déplacez votre fichier « 454.fastq » dans le répertoire « galaxy »:

Déplacez (cliquez sur le nom du fichier et glissez) « 454.fastq » dans ce répertoire galaxy/ nouvellement créé.

Donnez les bons droits à votre répertoire « galaxy » :

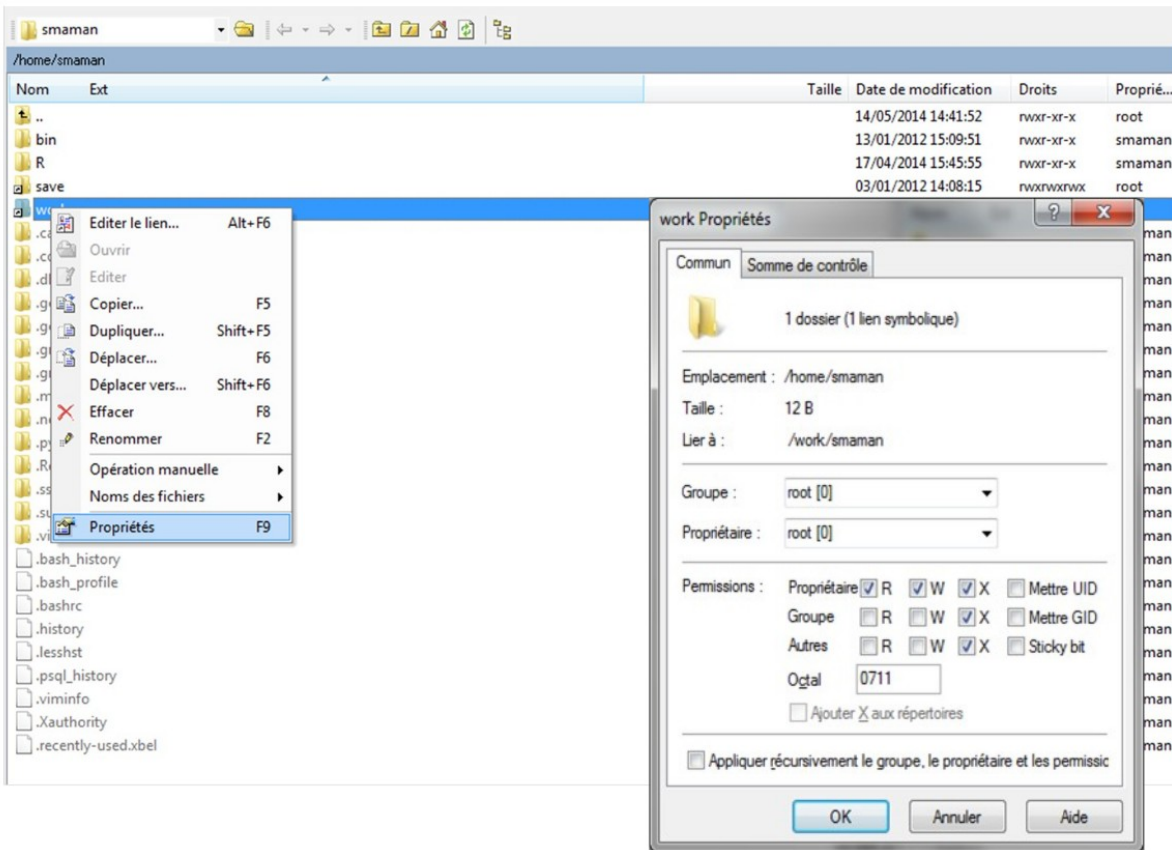
A l'aide d'un clic droit sur le répertoire /galaxy, veillez à ce que le répertoire galaxy/ ai bien les permissions « x » (exécution) pour tous. De même pour votre /work. Puis vérifier, de la même manière, que l'ensemble des fichiers contenus dans galaxy/ (uniquement) aient bien les droits « r »(lecture).

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.

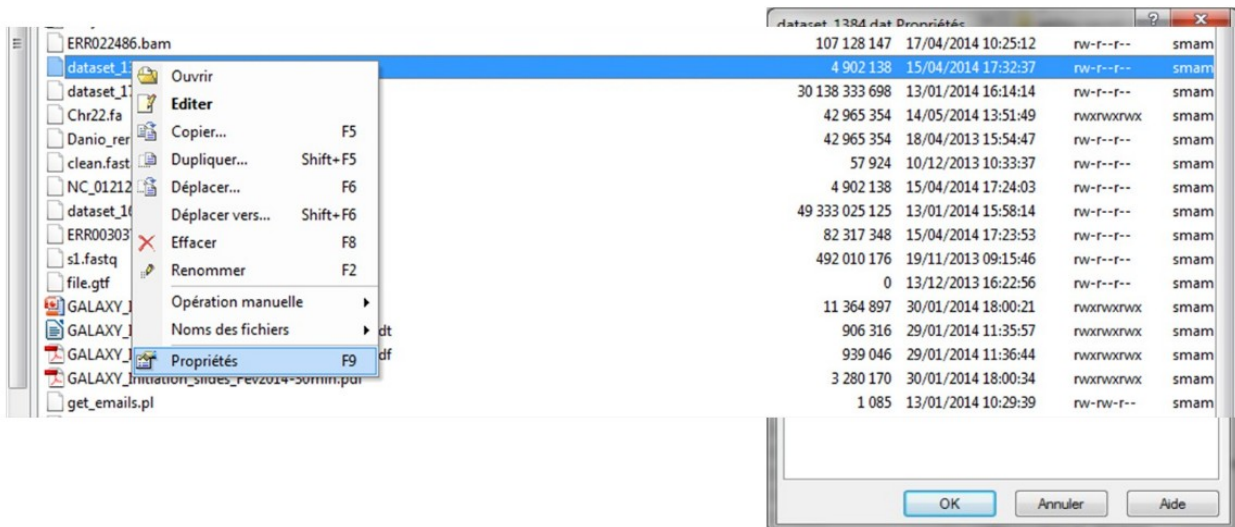
Donnez les bons droits au fichier « 454.fastq » :

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 le fichier « 454.fastq » :

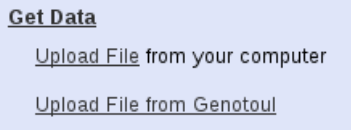


Si cela n'est pas fait, l'outil d'upload de Galaxy ne sera pas en mesure d'accéder ni de lire les fichiers que vous souhaitez télécharger et sur lesquels vous allez travailler. Après l'exécution de l'outil « upload » de Galaxy, la dataset sera rouge (donc en erreur).



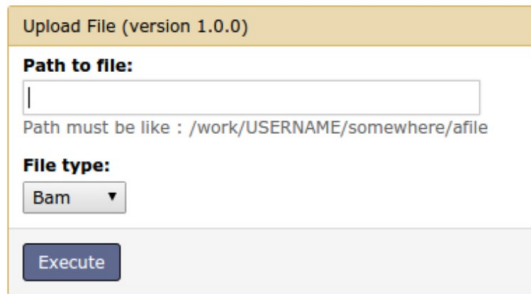
De votre /work sur Genotoul à votre interface Galaxy :

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



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.



Exemple de **Path to file** : /work/rose/galaxy/454.fastq

Renseignez le format du fichier: Complétez le « File type » : Fastq

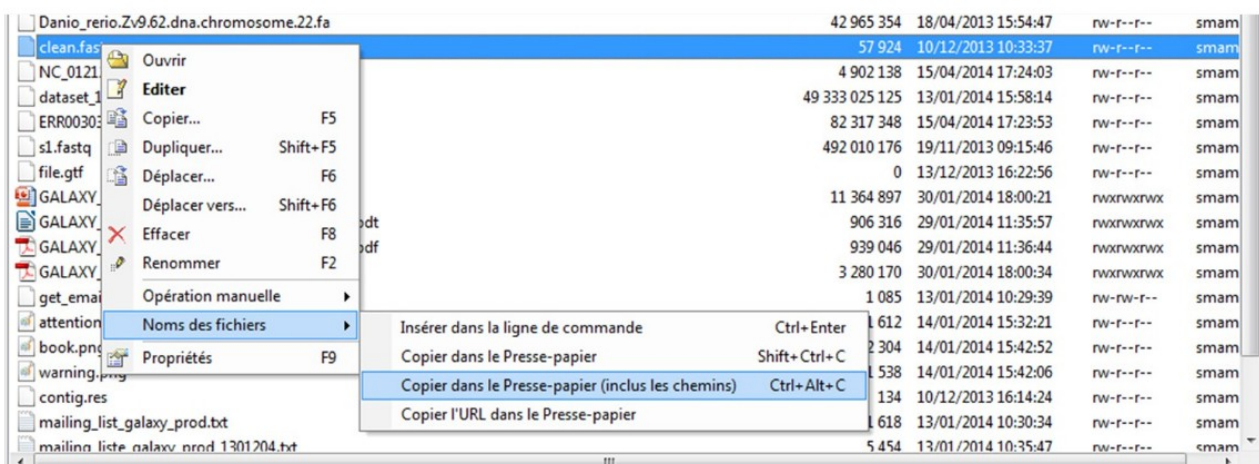


L'outil «Upload file from Genotoul» vous permet de créer un lien symbolique, depuis votre work, sur le serveur Galaxy, sans avoir besoin de copier vos données sur le serveur Galaxy. Grâce à cet outil, vous économisez de l'espace disque et optimisez votre quota sur Galaxy.



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 :





4 - Conseils pour gérer au mieux votre quota

Astuce :



Le /work est purgé régulièrement des fichiers non utilisés de plus de 120 jours mais les liens symboliques ne sont pas purgés.

Vous pouvez donc créer un lien symbolique de votre /save vers votre /work en cochant la case « Dupliquer avec une copie locale temporaire ».

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:



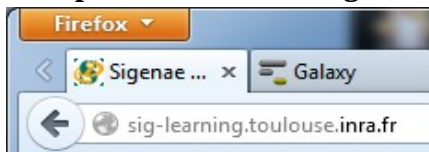
- Mail : support.siginae@inra.fr ou le « bug » de la dataset Galaxy.
- Une FAQ et un manuel utilisateur sont disponibles depuis la page d'accueil de l'instance Siginae de Galaxy.
- Les formations de la plateforme Bioinfo Genotoul sont disponibles sur <http://sig-learning.toulouse.inra.fr>

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

5 - sig-learning (facultatif)

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

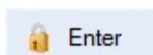
1- Tapez l'adresse de « sig-learning » : <http://sig-learning.toulouse.inra.fr/>



2- **Authentification** avec votre compte de formation ou votre compte Genotoul:

Login

Pass





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 ".

Onglet training pour accéder à vos formations :

The screenshot shows the 'Trainings' menu with options: Homepage, **Trainings**, Profile, Agenda, Progress, and Logout (rodriguez). Below the menu, a list of training courses is displayed:

- 1 - Linux & Unix
UNIX1 – SIGENAE Team
- 3 - Management of large files on Unix and Galaxy
UNIX2 – SIGENAE Team
- 4 - Galaxy
GALA01 – SIGENAE Team
- 5 - FastQC : quality control tool & reports interpretation
FASTQC – SIGENAE Team
- 6 - Aligning SGS reads and SNP finding
SGS-SNP – SIGENAE Team

On the right side, a 'My account' box contains a link: 'Trainings management (pour vous inscrire à de nouvelles formations de la plateforme Genotoul)'.

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

The screenshot shows the 'Trainings' menu with options: Homepage, **Trainings**, Profile, Agenda, and My progress. Below the menu, the breadcrumb trail is: Trainings > Trainings management > Subscribe to training. A green message box states: 'You have been registered to the training'. Below this, there are three buttons: 'Sort trainings', 'Create a training category', and 'Subscribe to training'.

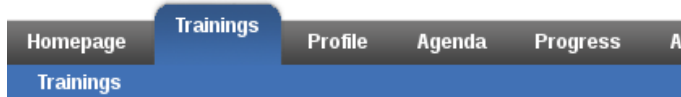
L'inscription s'effectue via une recherche de la formation par mots clés :







The screenshot shows the 'Trainings' menu with options: Homepage, **Trainings**, Profile, Agenda, and My progress. Below the menu, the breadcrumb trail is: Trainings > Trainings management (pour vous inscrire à de nouvelles formations de la plateforme Genotoul) > Su. Below this, there are three buttons: 'Sort trainings', 'Create a training category', and 'Subscribe to training'. Underneath, the 'Training categories' section shows: 'Plateforme GENOTOUL Trainings (5)'. Below this, the 'Trainings in this category' section has a search bar with the text: 'Search trainings (Pour lister l'ensemble des formations disponibles, inscrire « % » dans le champ « Search ».)'. The search bar contains the text '%', and there is a 'Search' button.

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.



-  **1 - Linux & Unix**
UNIX1 – SIGENAE Team
-  **2 - Cluster (en construction)**
CLUSTER – SIGENAE Team
-  **3 - Management of large files on Unix and Galaxy**
UNIX2 – SIGENAE Team
-  **4 - Galaxy**
GALA01 – SIGENAE Team
-  **5 - FastQC : quality control tool & reports interpretation**
FASTQC – SIGENAE Team
-  **6 - Aligning SGS reads and SNP finding**
SGS-SNP – SIGENAE Team

Remerciements



Fonds Européen
de Développement Régional

