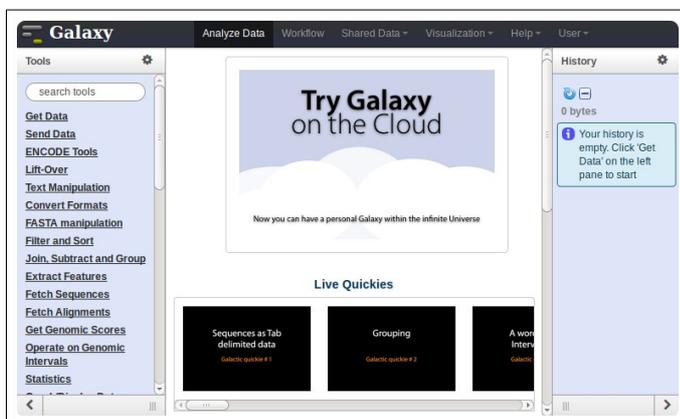




## - Galaxy -

# 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/>



## 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



Gestionnaire de tâches

Gère la file d'attente

Cluster de calculs



Exécute

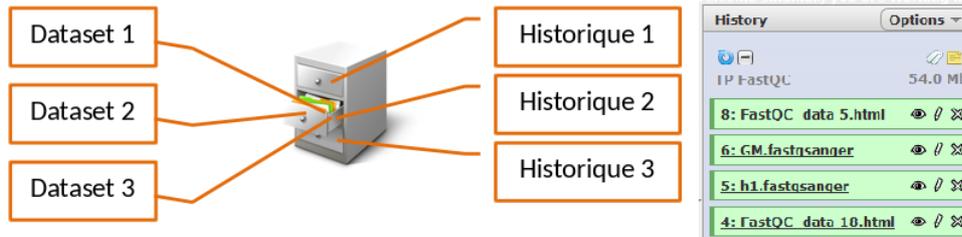


Nous utiliserons un vocabulaire spécifique à Galaxy :

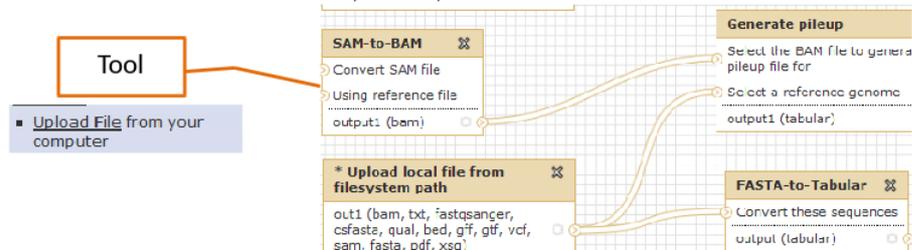
Un **DATASET** est un fichier de données (fichiers d'entrée, fichiers résultats) :



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



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>
- Des fichiers disponibles sur NG6 et des supports disponibles sur : [http://genoweb.toulouse.inra.fr/~formation/1\\_Galaxy\\_Initiation/](http://genoweb.toulouse.inra.fr/~formation/1_Galaxy_Initiation/)



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

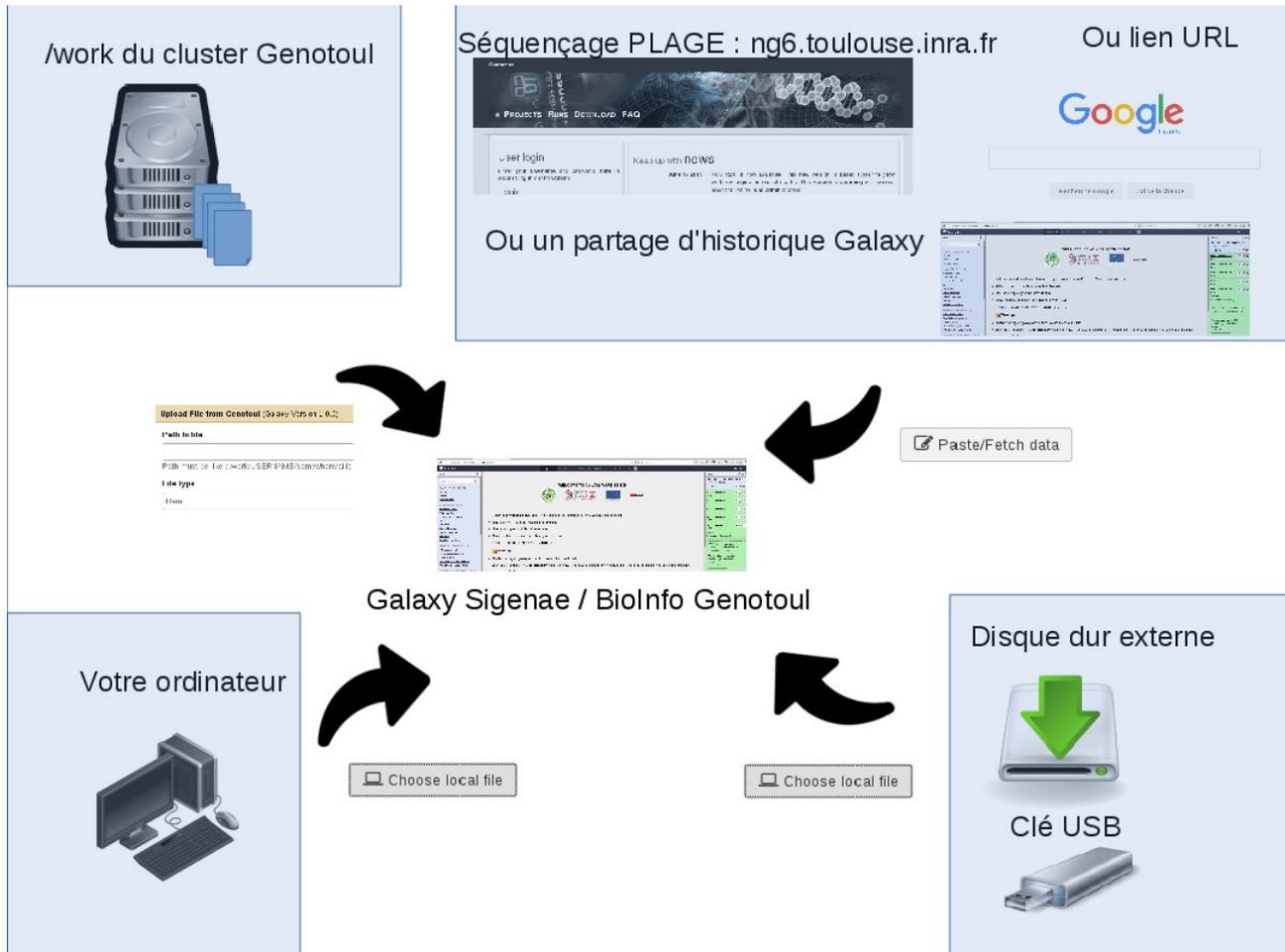
- Logins : anemone arome aster bleuete camelia capucine chardon clematite cobee coquelicot cosmos cyclamen
- Password : Demander au formateur.

**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 reflexe « sécurité ».**

Rappel : 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.



## 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 contiged » contenant 1 dataset.

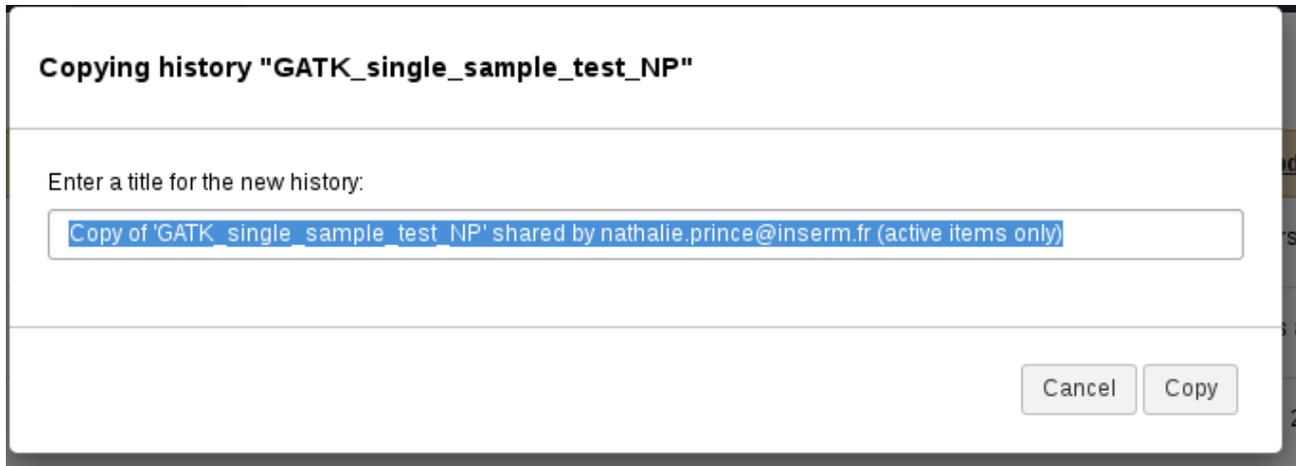
Voici un aperçu de votre plateforme Galaxy en fin de matinée :





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

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.

Pour ouvrir l'historique sur la droite, il suffit simplement de cliquer sur son nom.

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

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

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 contiged ». Pour changer le nom de l'historique, veuillez cliquer une fois sur « Unnamed history », le texte se sélectionne en bleu, taper directement le nouveau nom « Historique contiged » 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.

Dans l'« **historique contiged** », à l'aide de cet outil « [Upload File from your computer](#) », veuillez récupérer l'archive « 100spec\_90000seq\_9samples.tar.gz » disponible depuis cette URL : [http://genoweb.toulouse.inra.fr/~formation/15\\_FROGS/FROGS\\_ini/DATA/100spec\\_90000seq\\_9samples.tar.gz](http://genoweb.toulouse.inra.fr/~formation/15_FROGS/FROGS_ini/DATA/100spec_90000seq_9samples.tar.gz)

Pour récupérer un fichier disponible sur Internet, veuillez utiliser l'outil « Upload File » de Galaxy (outil disponible dans la section « Get Data »).



Fev. 2017

Cliquer sur « Paste/Fetch » data. Puis copier/coller l'URL COMPLETE d'accès au fichier pour le télécharger dans Galaxy. Exemple d'URL complète : [http://genoweb.toulouse.inra.fr/~formation/15\\_FROGS/FROGS\\_ini/DATA/100spec\\_90000seq\\_9samples.tar.gz](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.

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 ».

Chosissez 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 ».

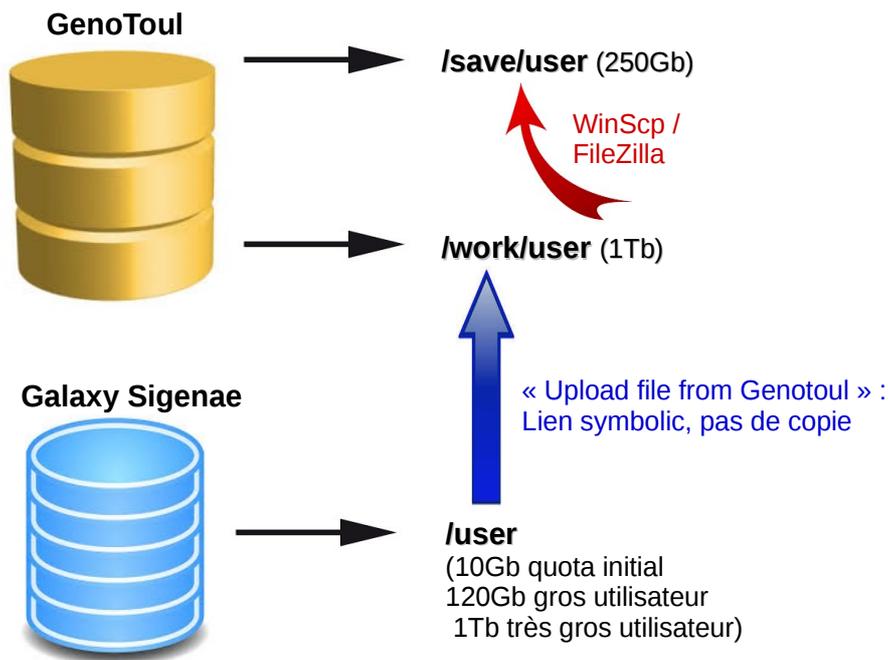
### 3 – 3 – Créer l'historique « 454 » : Télécharger un fichier d'Internet à votre /work puis de votre /work à Galaxy



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 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 »).

### 3-3-1- D'internet à votre /work

Dans l'« **historique 454** », nous allons télécharger, avec « Upload File », le fichier « 454.fastq » disponibles depuis Internet via l'url [http://genoweb.toulouse.inra.fr/~formation/15\\_FROGS/FROGS\\_ini/DATA/454.fastq](http://genoweb.toulouse.inra.fr/~formation/15_FROGS/FROGS_ini/DATA/454.fastq)

Pour télécharger ce fichier depuis Internet, 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 chaque 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 /work à Galaxy

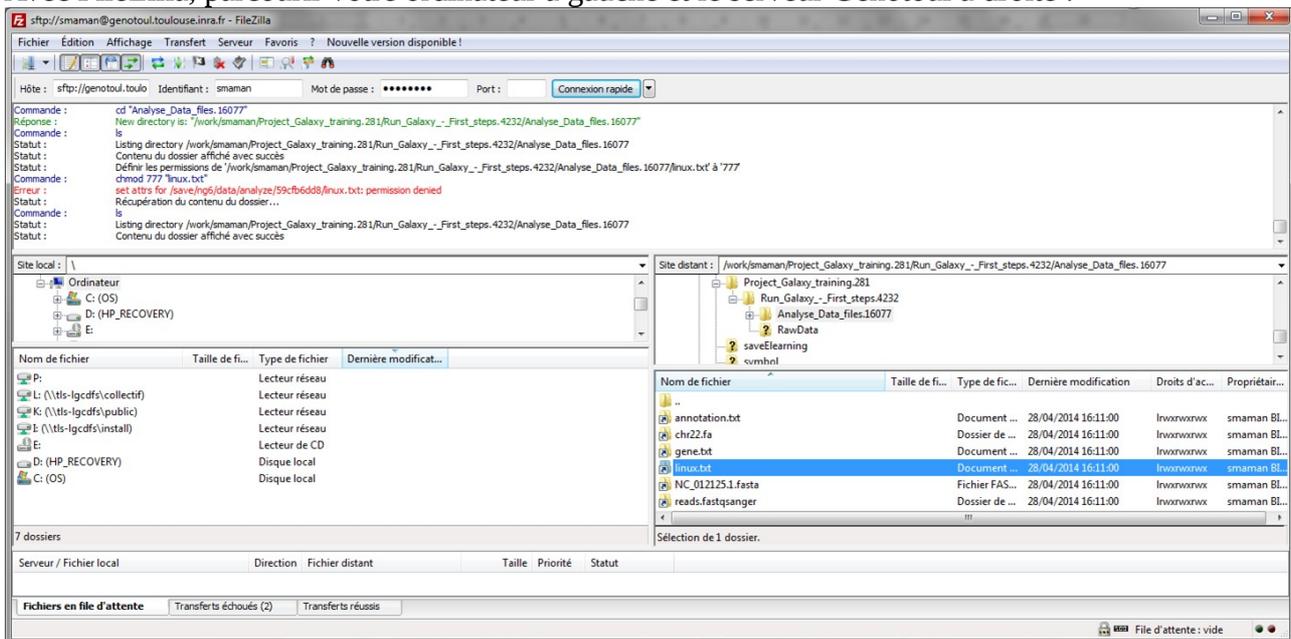
Nous allons ensuite transférer ces fichiers de votre ordinateur à Galaxy. Pour ce faire, veuillez ouvrir le logiciel WinSCP (ou FileZilla). Cet outil vous permet de voir le contenu de vos répertoires sur Genotoul et de gérer les permissions sur ces répertoires et fichiers.

Pour vous connecter à Genotoul, veuillez compléter les paramètres suivants :

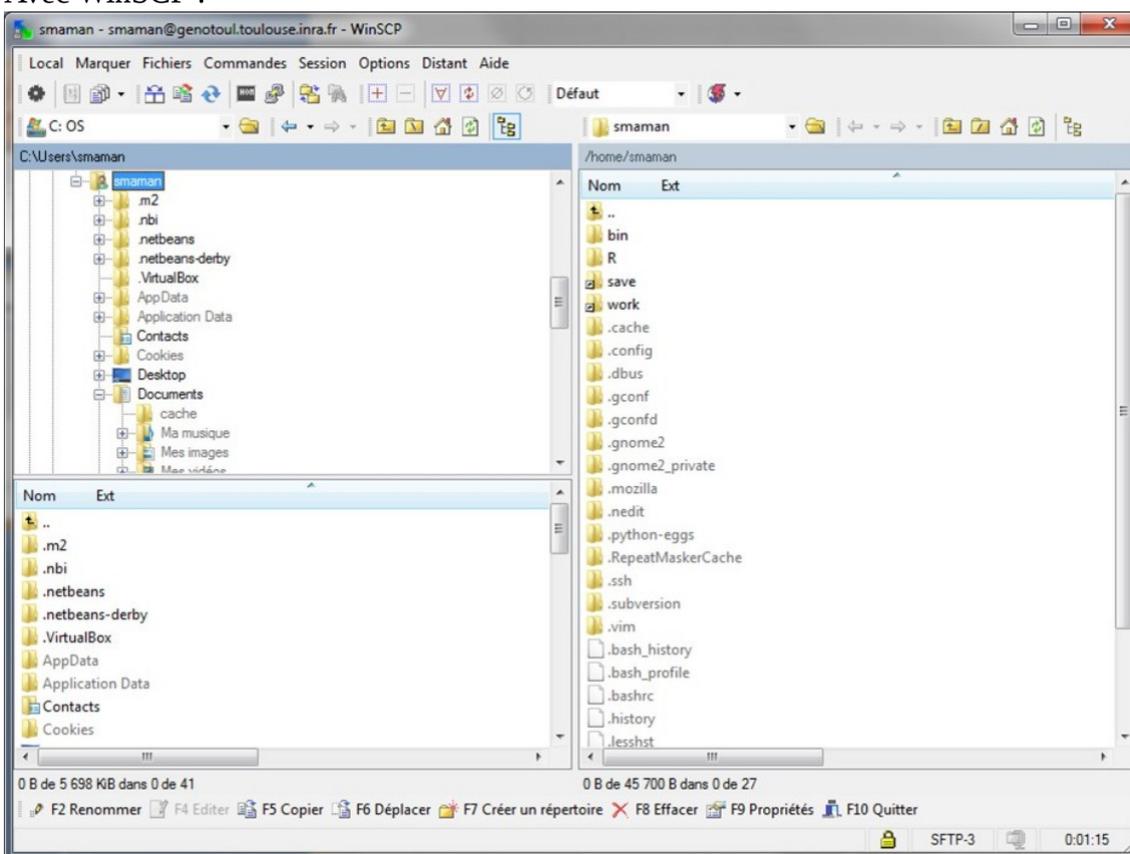
- Hôte : genotoul.toulouse.inra.fr
- Identifiant : Votre login sur Genotoul
- Mot de passe : Votre mot de passe sur Genotoul
- Port : 22



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



Avec WinSCP :



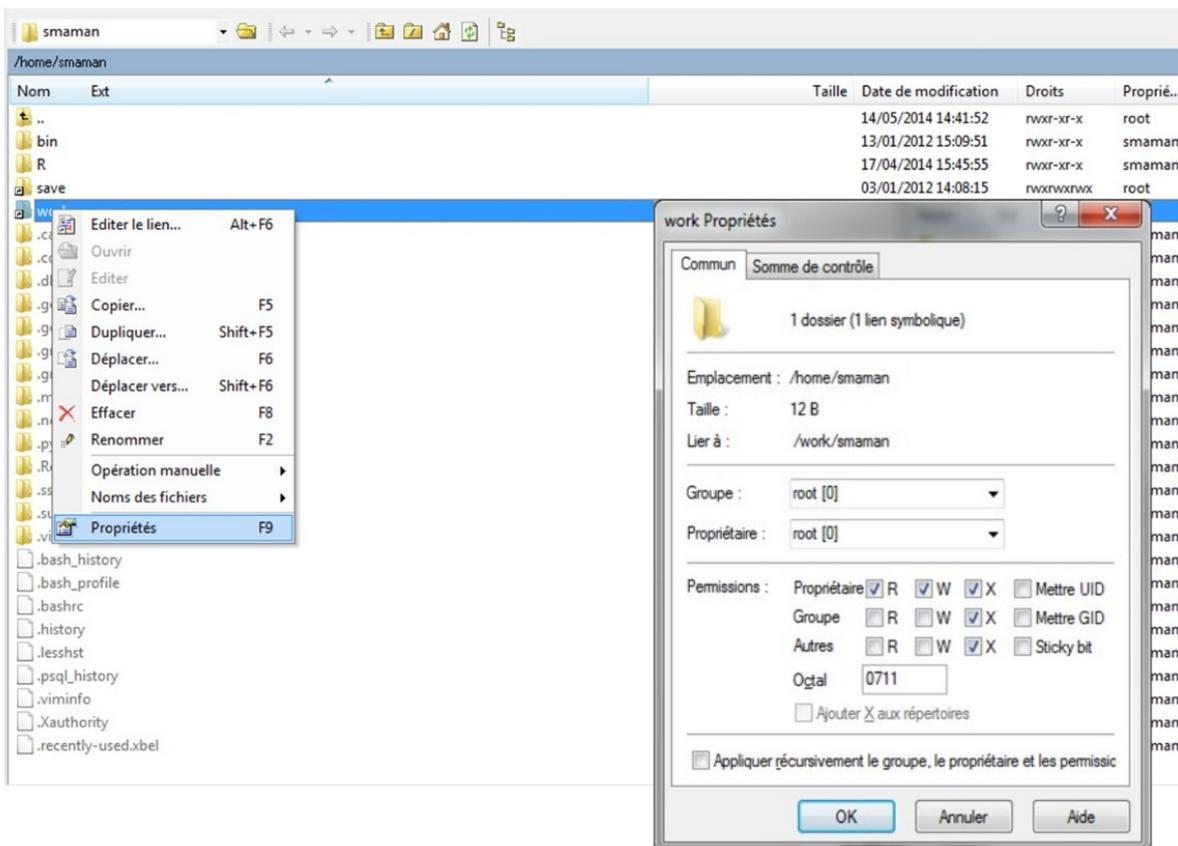
A droite, dans Genotoul, créer un répertoire galaxy/ dans votre /work, puis déplacer l'ensemble des fichiers dans ce répertoire galaxy/ nouvellement créé.

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

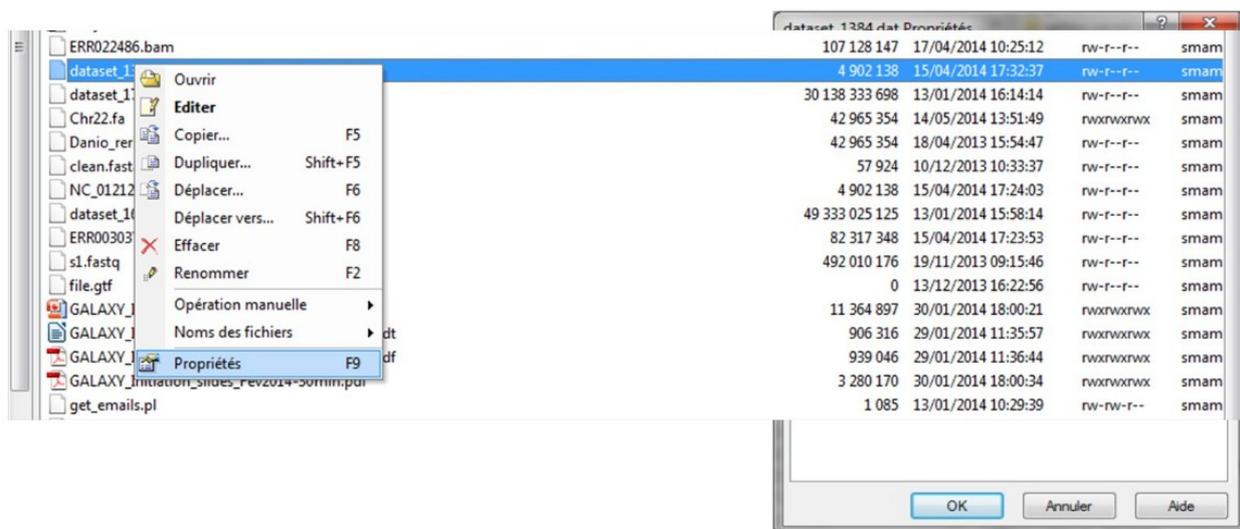


« r »(lecture).

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.



De même pour chacun des fichiers à récupérer, pour modifier leurs droits, clic droit sur le nom du fichier, puis « Droits d'accès au fichier », puis donner les droits de lecture au fichier concerné.



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).

Utiliser l'outil « [Upload File from Genotoul](#) » afin créer le lien dans votre historique galaxy. Cette



méthode de téléchargement de fichiers dans Galaxy entame beaucoup moins votre quota que la méthode exposée précédemment.



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.

Upload File (version 1.0.0)

**Path to file:**

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

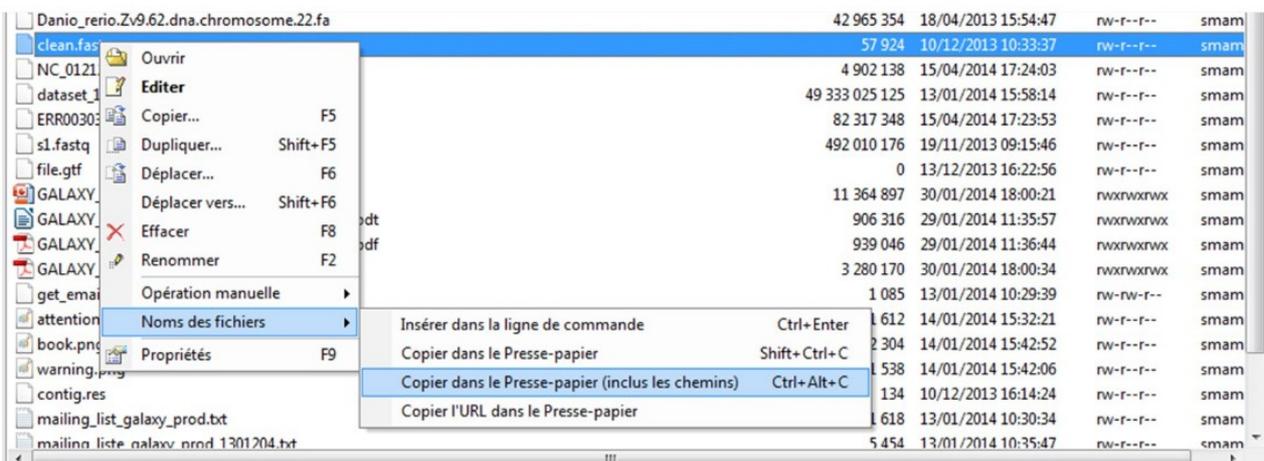
**File type:**

Bam ▼

Execute

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.

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) ».





## 4 – Conseils pour gérer au mieux votre quota

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 ? »

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 : [sigenae-support@listes.inra.fr](mailto:sigenae-support@listes.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 Sigenae de Galaxy.
- Les formations de la plateforme Bioinfo Genotoul sont disponibles sur [http://sig-  
learning.toulouse.inra.fr](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).



Fev. 2017

## Remerciements



Fonds Européen  
de Développement Régional

