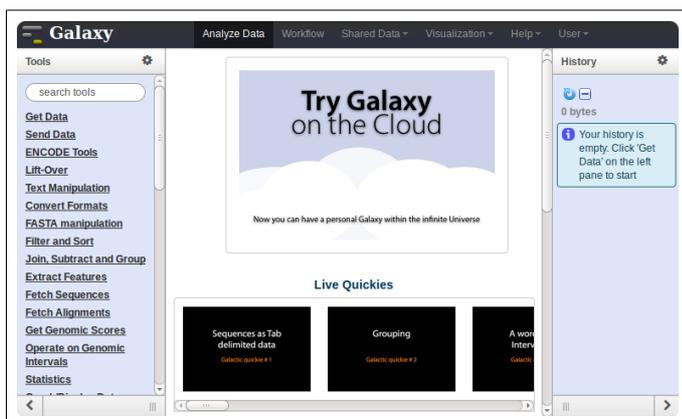


## - *Galaxy* -

# *Initiation à la plateforme Galaxy*

## - *EXERCICES* -



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

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

## Objectifs :

Cette formation a pour objectif de vous familiariser à l'utilisation de votre workbench Galaxy.

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 et présentations disponibles à l'url suivante : [http://genoweb.toulouse.inra.fr/~formation/5\\_ADN16S/](http://genoweb.toulouse.inra.fr/~formation/5_ADN16S/)

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, aster, bleuet, iris, muguet, narcisse, pensee, rose, tulipe, violette, lilas, pervenche, laurier, lavande, lis, capucine, coquelicot, geranium, liseron, arome, chardon
- Password : **f1o2r3!**

Pour répondre à vos questions:



- Mail : [sigenae-support@listes.inra.fr](mailto:sigenae-support@listes.inra.fr)
- 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>

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

## Objectif de l'exercice

Cet exercice expose les différentes méthodes de téléchargement (ou import) de vos données dans l'instance Sigene de Galaxy.

Vos données peuvent être localisées :

- Dans votre PC,
- Sur Internet,
- Sur Genotoul.

L'objectif est de rendre vos données disponibles dans Galaxy quelque soit leur source.

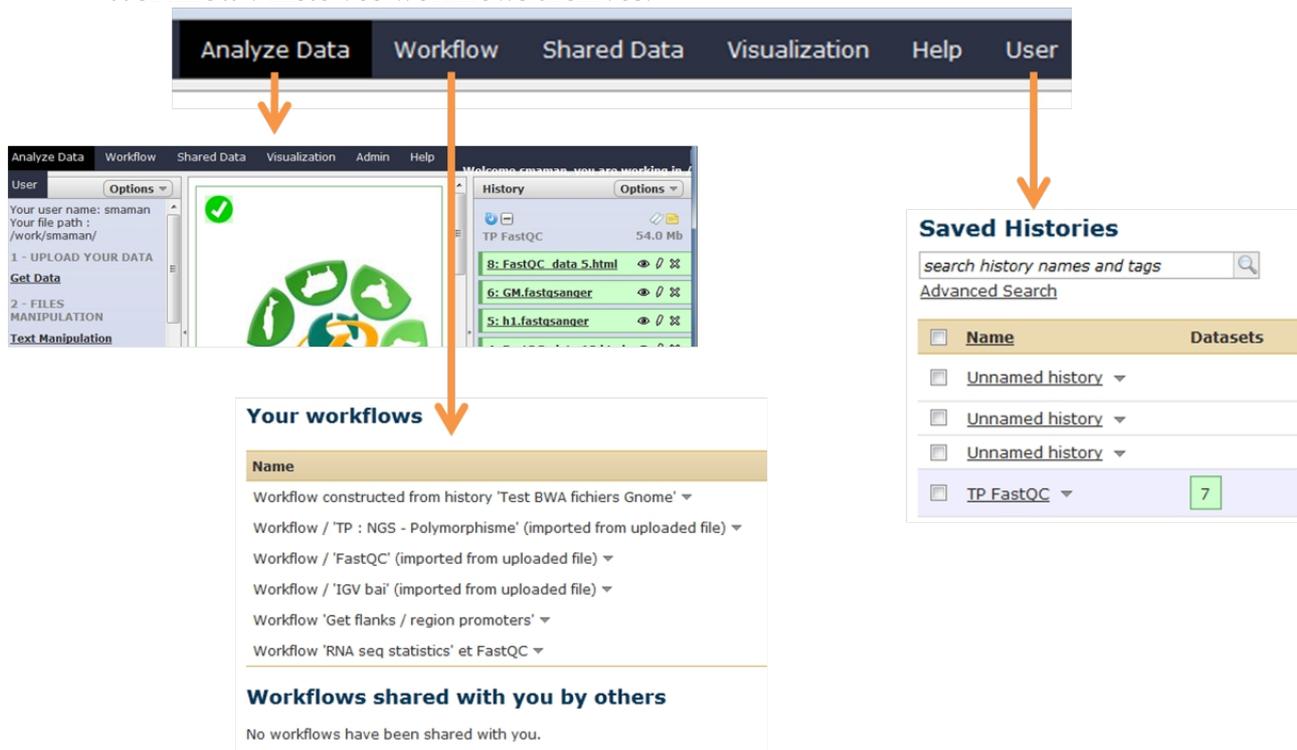
## Connexion à la plateforme Galaxy

Vous pouvez accéder à votre plateforme Galaxy (en précisant votre login et mot de passe LDAP « 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.



The image shows a composite of three screenshots from the Galaxy web interface. At the top is the main navigation bar with tabs: 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Help', and 'User'. Three orange arrows point from these tabs to the corresponding panels below. The left panel, 'Your workflows', lists several workflows with their names and sources. The middle panel, 'Saved Histories', shows a search bar and a table of history entries. The right panel, 'History', shows a list of recent data files with their names and sizes.

Name	Size
TP FastQC	54.0 Mb
8: FastQC_data 5.html	
6: GM.fastqsanger	
5: h1.fastqsanger	

Name	Datasets
Unnamed history	
Unnamed history	
Unnamed history	
TP FastQC	7

Name
Workflow constructed from history 'Test BWA fichiers Gnome'
Workflow / 'TP : NGS - Polymorphisme' (imported from uploaded file)
Workflow / 'FastQC' (imported from uploaded file)
Workflow / 'IGV bai' (imported from uploaded file)
Workflow 'Get flanks / region promoters'
Workflow 'RNA seq statistics' et FastQC

**Note** : La documentation autour de Galaxy est très aboutie, explorer le site web du « Galaxy Project » (<http://galaxyproject.org/>) 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.

## Import des jeux de données dans Galaxy

### Import du jeu de données 454 Patho\_ID à partir de NG6

Les outils pour importer des fichiers Galaxy sont disponibles dans la section «Upload your data => Get data »

L'outil « **Upload File** » télécharge en copiant votre fichier sur le serveur Galaxy. Cette copie diminue votre quota Galaxy.



Pour obtenir l'adresse de téléchargement, faites un clic droit sur le lien de téléchargement, puis « Copy link location ».

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

Étape 1:Créer un nouvel historique Galaxy et le renommer « PathoID\_454 ». Nous allons ensuite charger les fichiers correspondants à ce jeu de données.

Étape 2:Vos données 454 sont disponibles à partir de la plateforme NG6 (Next Generation Sequencing Information System) :

Contact us About nG6

PROJECTS RUNS DOWNLOAD

### User login

Enter your username and password here in order to log in on the website:

Login

Username:

Password:

### Keep up with news

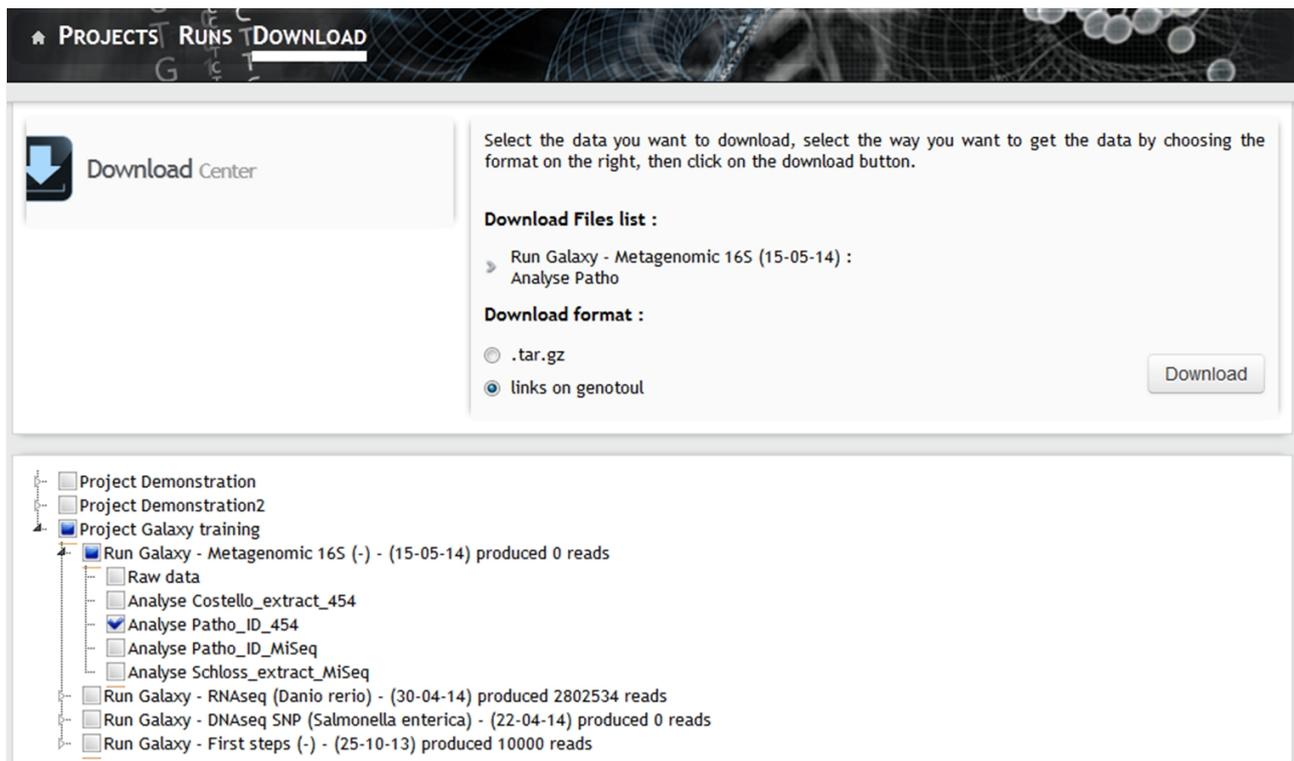
- June 17 2013** NG6 v2.0 is now available. This new version is based upon the jflow workflow engine instead of ergatis. This version is coming with several new features for runs administrators.
- November 27 2012** NG6 is now only available in english.
- September 14 2012** New user management system. 3 rights levels: administrator (in charge to run workflows), manager (in charge to manage project access) and member (browsing projects/runs/analyses).
- September 9 2012** Publication of NG6 in [BMC Genomics](#).
- August 7 2012** NG6 is available in french and in english.
- June 4 2012** Fix a bug in the generation of cigarline graphs produced by the AlignmentStats analysis.
- November 23 2011** Since the 3rd of october 2011 the HiSeq quality encoding is in Sanger format, no longer in illumina format.
- August 2nd 2011** New functionalities are available to browse your projects/runs /analyses. A hierarchical display is now available for analysis. For project administrators new functionalities are also available.

NG6 est :

- \* Un 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.

A partir de la plateforme NG6 (<http://ng6.toulouse.inra.fr>), menu « Projects », « Galaxy training » puis « Run Galaxy – Metagenomic 16S », sélectionner « Analyse Patho\_ID\_454 » puis créer les

liens symboliques dans votre espace de travail sur le serveur genotoul 'Directory :  
/work/VotreUserName)/:



The screenshot shows the 'Download Center' interface. At the top, there are navigation tabs for 'PROJECTS', 'RUNS', and 'DOWNLOAD'. Below the tabs, there is a 'Download Center' section with a blue arrow icon. To the right, there is a text box explaining the download process: 'Select the data you want to download, select the way you want to get the data by choosing the format on the right, then click on the download button.' Below this, there is a 'Download Files list' section showing a single entry: 'Run Galaxy - Metagenomic 16S (15-05-14) : Analyse Patho'. Underneath, there is a 'Download format' section with two radio buttons: '.tar.gz' and 'links on genotoul' (which is selected). A 'Download' button is located to the right of the format options. At the bottom, there is a tree view of projects and runs, including 'Project Demonstration', 'Project Demonstration2', 'Project Galaxy training', and several 'Run Galaxy' entries with their respective dates and read counts.



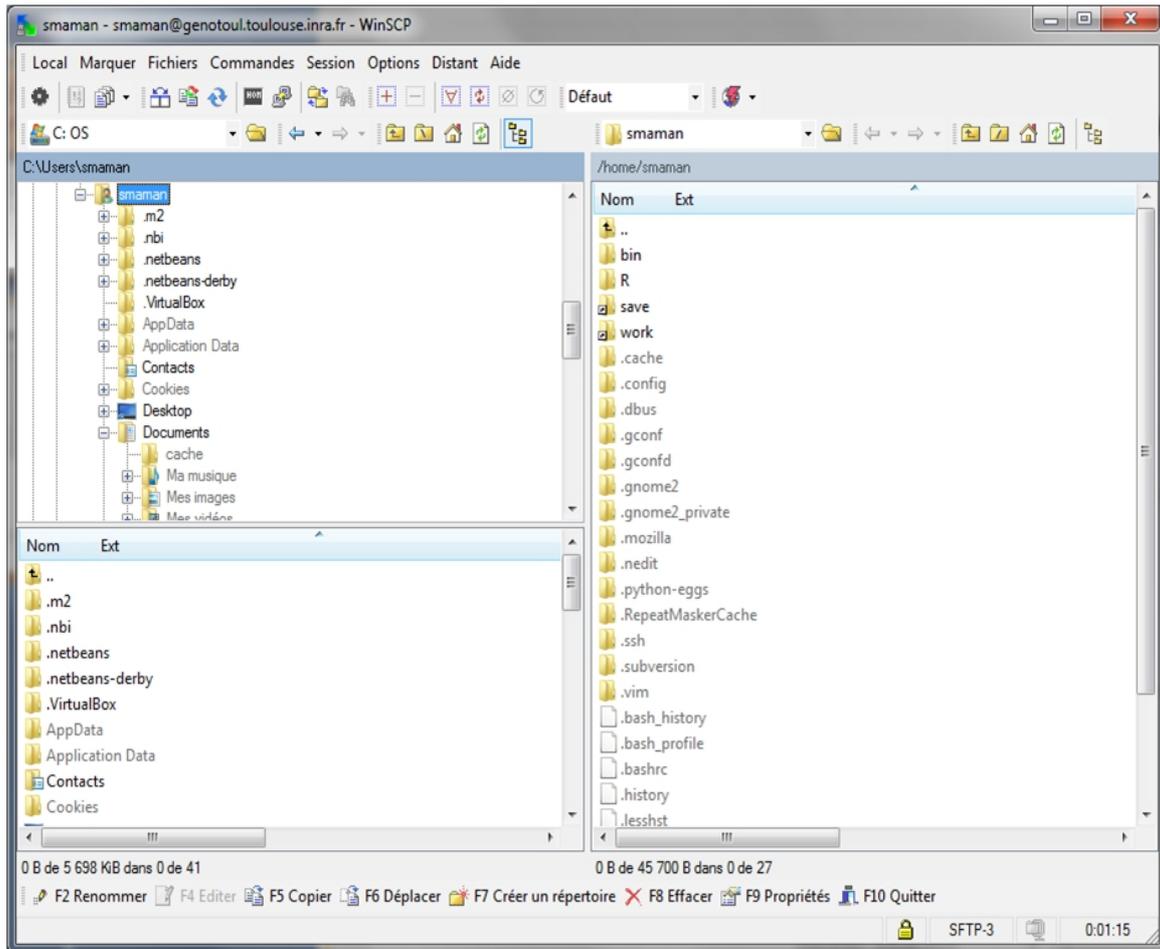
The screenshot shows a dialog box titled 'Create symbolic link in a home directory'. It contains a message: 'This functionality is only available if you have an account on the genotoul server.' Below the message, there is a 'Login' field with the value 'smaman', a 'Password' field with masked characters, and a 'Directory' field with the value '/work/smaman/'. At the bottom right, there are 'Close' and 'Download' buttons.



Étape 3: Ouvrir le logiciel WinSCP (ou FileZilla). Cet outil vous permettra de voir le contenu de vos répertoires sur Genotoul et de gérer les permissions sur ces répertoires et fichiers.

Connectez vous à Genotoul à l'aide des paramètres suivants :

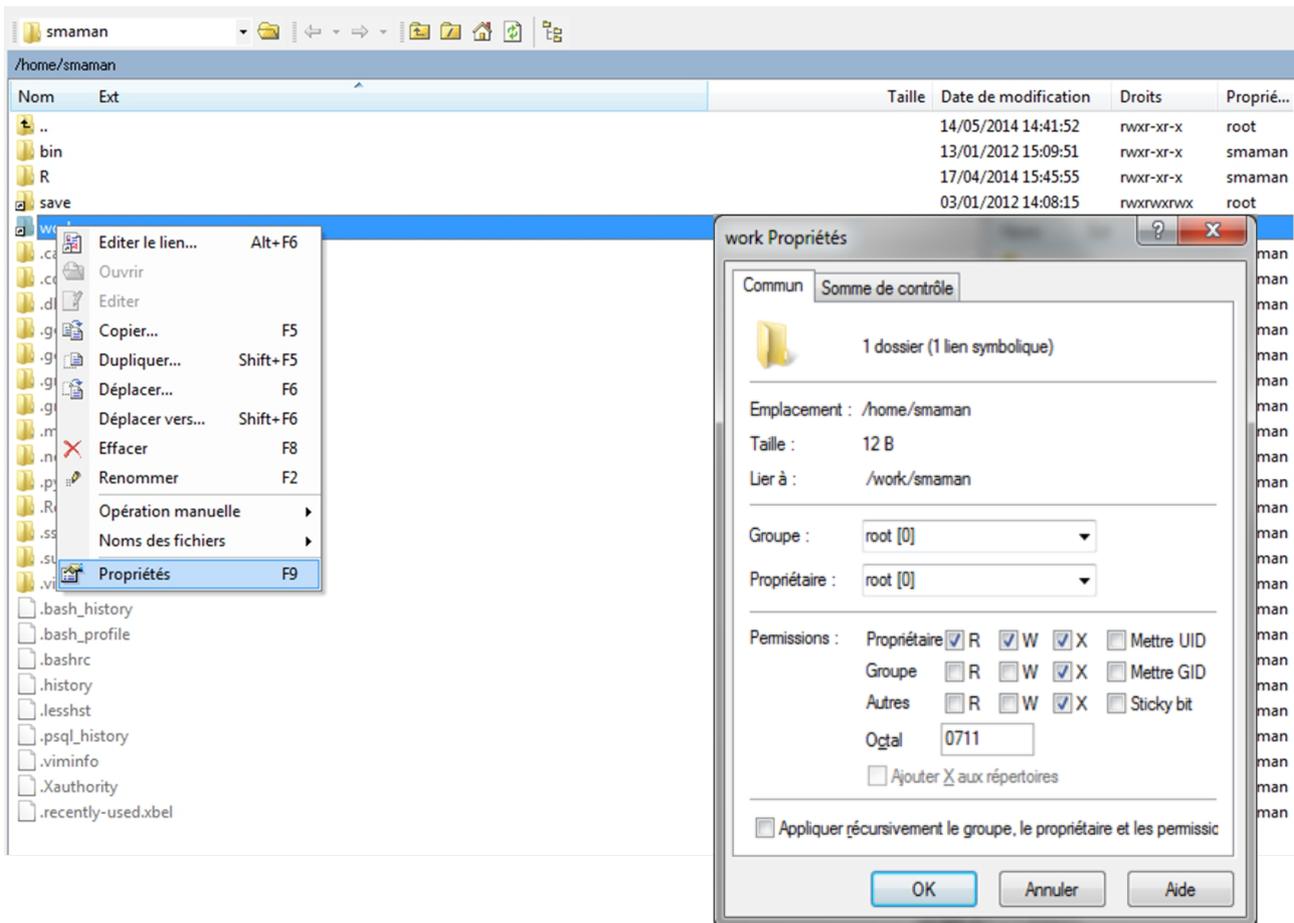
- Hôte : genotoul.toulouse.inra.fr
- Identifiant : Votre login sur Genotoul
- Mot de passe : Votre mot de passe sur Genotoul
- Port : 22



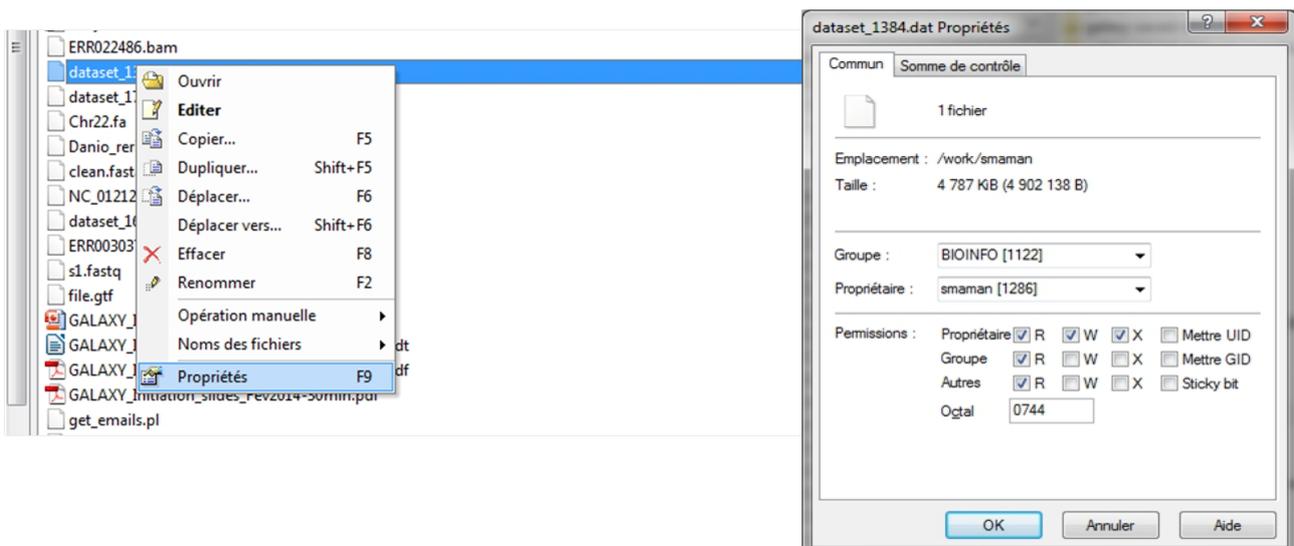
Étape 4 : Parcourez votre espace de travail /work/username/ afin de visualiser les fichiers de données téléchargés depuis NG6.  
 Les droits d'exécution sur le répertoire et de lecture sur les fichiers sont nécessaires pour que vos données puissent être accessibles dans Galaxy.



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.



Étape 5 : Dans Galaxy, utiliser ensuite l'outil « **Upload local file from filesystem path** Upload data to history without copying on server » afin d'ajouter vos jeux de données, un par un, dans votre

historique.



L'outil «**Upload local file from filesystem path**» 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.

Étape 6 : Voici quelques indications pour paramétrer l'outil Galaxy :

**\* Upload local file from filesystem path (version 1.0.0)**

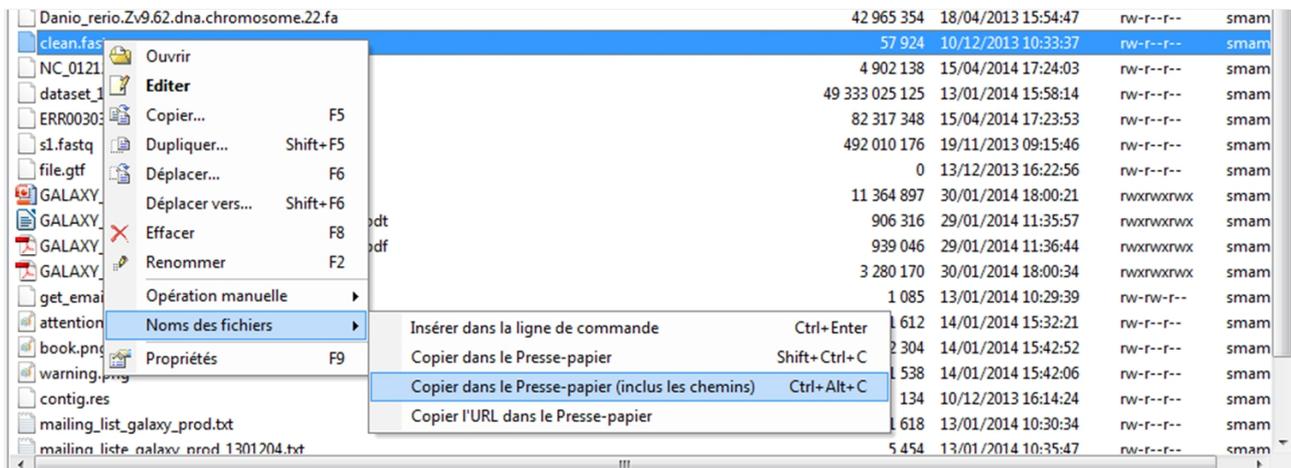
**File Name (without extension):**

**File type:**  
 Bam ▾

**Path to file:**

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) ».



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 »).

## Import de la référence Silva via l'URL du fichier

La référence utilisée dans vos traitements 16S est disponible sur le site Internet de Mothur, à l'adresse suivante : [http://www.mothur.org/wiki/Silva\\_reference\\_files](http://www.mothur.org/wiki/Silva_reference_files)

Cette URL comprends plusieurs fichiers. Il est préférable, pour ce TP, de ne télécharger que les deux fichiers dont nous avons besoin.

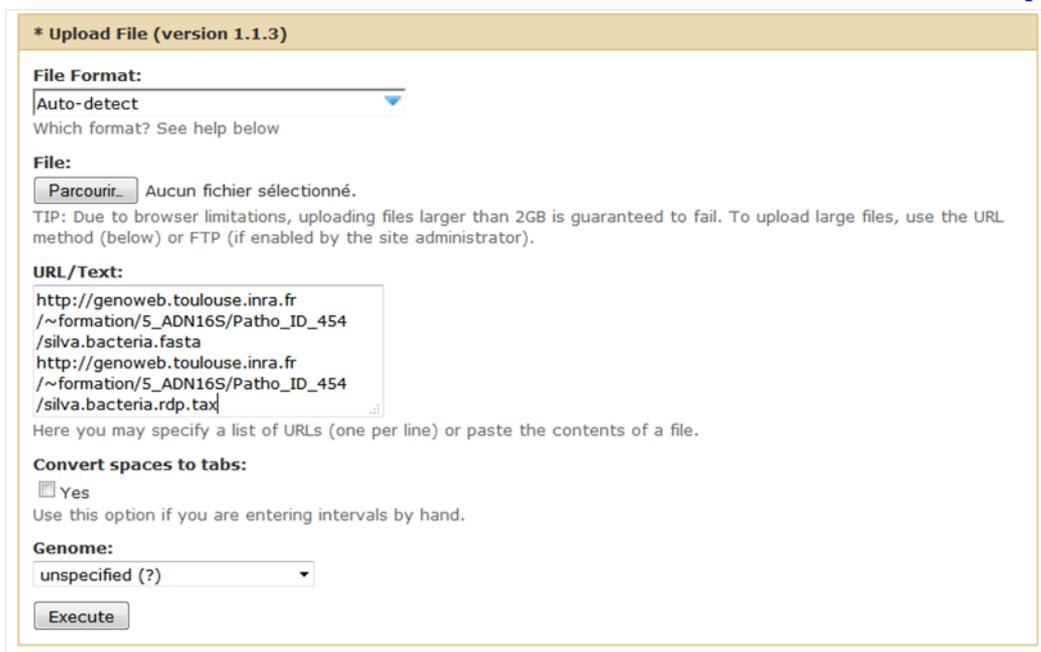
Pour récupérer le fichier fasta et le fichier silva.bacteria.rdp.tax, veuillez utiliser l'outil « Upload File » de Galaxy (outil disponible dans la section « Get Data »).

URL du fichier fasta :

[http://genoweb.toulouse.inra.fr/~formation/5\\_ADN16S/Patho\\_ID\\_454/silva.bacteria.fasta](http://genoweb.toulouse.inra.fr/~formation/5_ADN16S/Patho_ID_454/silva.bacteria.fasta)

URL du fichier tax :

[http://genoweb.toulouse.inra.fr/~formation/5\\_ADN16S/Patho\\_ID\\_454/silva.bacteria.rdp.tax](http://genoweb.toulouse.inra.fr/~formation/5_ADN16S/Patho_ID_454/silva.bacteria.rdp.tax)



Il vous suffit de copier/coller l'URL d'accès au fichier pour le télécharger dans Galaxy. Si vous avez plusieurs URL à indiquer, veuillez sauter une ligne entre chaque URL.

## Import de données MiSeq à partir de votre PC à votre /work/

Avant de commencer les téléchargements, veuillez créer un nouvel historique Galaxy et le renommer « MiSeq ».

### D'Internet à Galaxy

Avec la même démarche et le même outil que pour la référence Silva, veuillez télécharger le fichier « Patho\_ID\_MiSeq.config.txt » depuis cette URL :

[http://genoweb.toulouse.inra.fr/~formation/5\\_ADN16S/Patho\\_ID\\_MiSeq/Patho\\_ID\\_MiSeq.config.txt](http://genoweb.toulouse.inra.fr/~formation/5_ADN16S/Patho_ID_MiSeq/Patho_ID_MiSeq.config.txt)



Le fichier « Patho\_ID\_MiSeq.config.txt » est très léger. Donc le télécharger dans Galaxy avec un upload n'est ni trop long ni trop impactant pour votre quota. Par contre, il ne s'agit pas de la méthode d'upload recommandée.

### De l'historique 454 à l'historique MiSeq

La référence Silva est disponible dans l'historique « PathID\_454 », et nous souhaitons l'importer dans l'historique « MiSeq ».

Tout d'abord, il est nécessaire de vous positionner dans l'historique « MiSeq » que vous venez de créer puis cliquer dans le menu « User », puis « Saved Datasets ». Sélectionner la référence Silva, puis dans le menu déroulant, sélectionner « Copy to current history ».

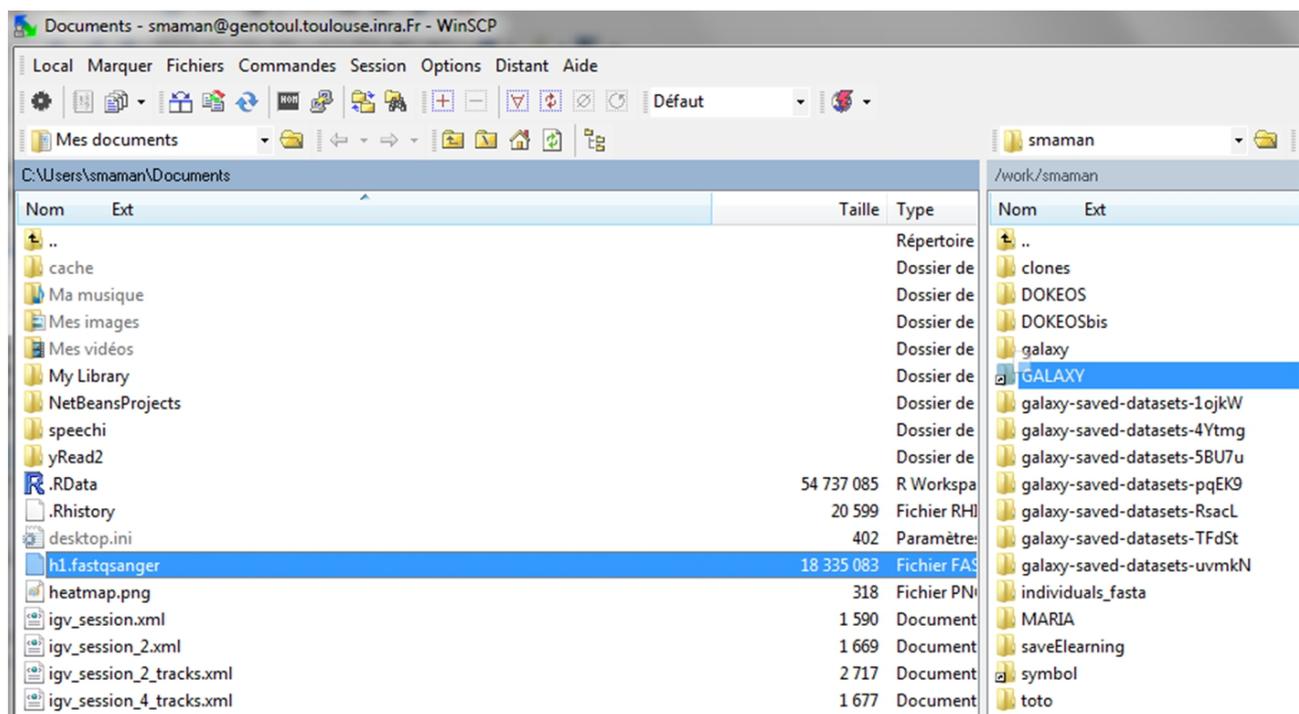
### De votre PC à votre /work

Comme l'outil pre-process de MiSeq nécessite beaucoup de fichiers en entrée, nous n'allons pas les télécharger dans Galaxy un par un mais nous allons fournir un répertoire contenant l'ensemble des fichiers utiles. Ce répertoire « data » est compressé (extension .gz).

Pour transférer cette archive, nous allons de nouveau utiliser WinSCP :

Étape 1: Récupérer le fichier data.tar.gz depuis le lien [http://genoweb.toulouse.inra.fr/~formation/5\\_ADN16S/Patho\\_ID\\_MiSeq/data.tar.gz](http://genoweb.toulouse.inra.fr/~formation/5_ADN16S/Patho_ID_MiSeq/data.tar.gz) sur votre PC.

Étape 2 : Créer un dossier « PathoID\_Miseq » sur votre work via WinSCP, coller le dossier décompressé data dans ce répertoire via WinSCP. en faisant glisser le fichier de la droite à la gauche de l'interface WinSCP.



Étape 3 : Donner les droits d'exécution aux répertoires et de lecture aux fichiers.

### Sauver des fichiers Galaxy dans votre save

#### Récupérer une liste de fichiers de Galaxy vers son /work

Tout d'abord, lister les datasets disponibles depuis le menu « User » puis « Saved Datasets ».

Une liste de fichiers disponibles sont listés. Veuillez cocher ceux que vous souhaitez archiver dans Genotoul , puis tout en bas de cette page, cliquer sur le bouton « Get path list ». Une dataset est

alors automatiquement générée dans votre historique et comprends la liste des fichiers que vous souhaitez archiver.

Ensuite, veuillez utiliser l'outil « [\\* Save my data](#) on Genotoul » disponible dans la section « Get Data » de Galaxy. Indiquer « Yes » dans « Use file list ? » puis pointer vers la dataset que vous venez de générer avec « Get path list ».

Visualisez ensuite les fichiers sauvegardés à l'aide de WinSCP dans votre /work. Un nouveau répertoire du type « galaxy-saved-datasets-NumerosChiffres » vient d'être créé avec vos datasets Galaxy.

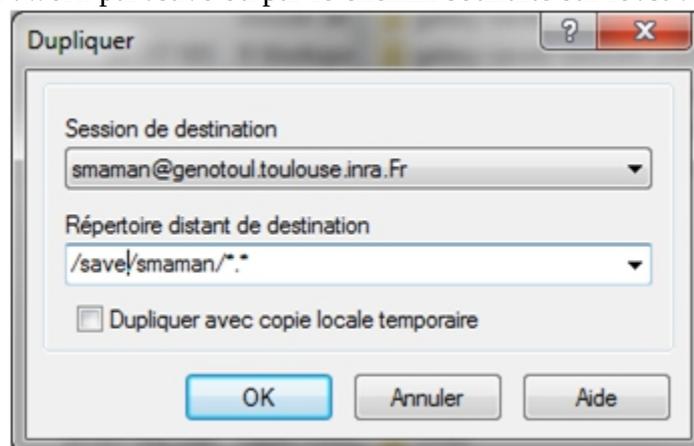
Le /work étant un espace de travail, il est régulièrement purgé par l'administrateur du cluster Genotoul. Nous vous conseillons donc ensuite de déplacer vos fichiers de votre /work vers votre /save.

### Passer ses fichiers du /work au /save avec WinSCP

Etape 1- Depuis WinSCP, cliquer sur l'icône  dans le panneau souhaité. Une troisième fenêtre apparaît alors.

Etape 2 – Clic droit sur le fichier à sauvegarder, puis « Duplicate » (ou menu Fichiers puis Dupliquer).

Dans le path, remplacer /work par /save ou par le chemin souhaité sur le /save.



#### 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 ».

#### Démarche pour accéder aux tutoriels disponibles en e-learning:

- Entrer sur la plateforme d'autoformation sig-learning : <http://sig-learning.toulouse.inra.fr/> (Login et mot de passe de votre LDAP Genotoul)



- A votre première connexion, votre liste de formation est vide. Pour vous inscrire : « Trainings » « Trainings management » puis « Subscribe to training » .

- Astuce : Taper « % » dans le champs de recherche pour lister toutes les formations disponibles. Sinon, veuillez effectuer une recherche par mots clés (exemple : « Galaxy, « Mothur »).