



Tutorial

Galaxy environment

Toulouse, 31 mars au 03 avril 2015

Sarah Maman
Plateforme Sigeneae - INRA Toulouse



Galaxy est une plateforme de traitements (bio)informatiques accessible depuis l'URL : <http://galaxy-workbench.toulouse.inra.fr/> (en précisant votre login et mot de passe LDAP « genotoul »).

Quelque soit la source de vos données (PC, Internet, Genotoul), elles peuvent être téléchargées dans Galaxy. Voici les différents cas de figures exposés dans ce mémo :

- Import de jeu de données à partir de fichier present sur les serveurs de la plateforme Bioinformatique (genotoul)
- Import de fichiers depuis Internet
- Import de fichiers depuis votre PC (ou clés Usb, ou disque dur, ou d'un autre serveur).
- Sauver vos fichiers Galaxy dans votre save/ de genotoul
- Sauvegarder une chaîne de traitements

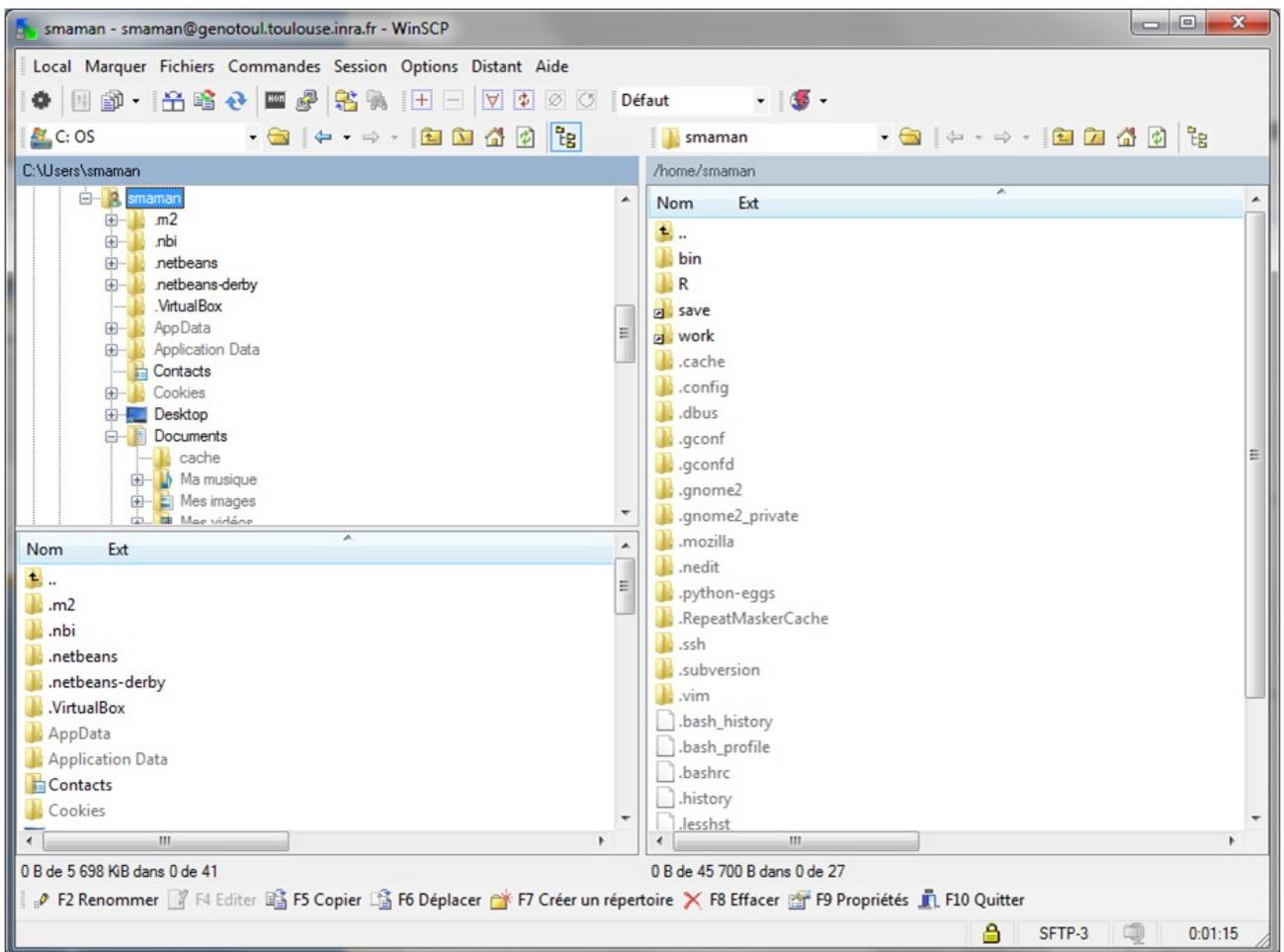
Avant de commencer les téléchargements, il est préférable de renommer son historique.

Import du jeu de données existant sur Genotoul

Étape 1 : 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.

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

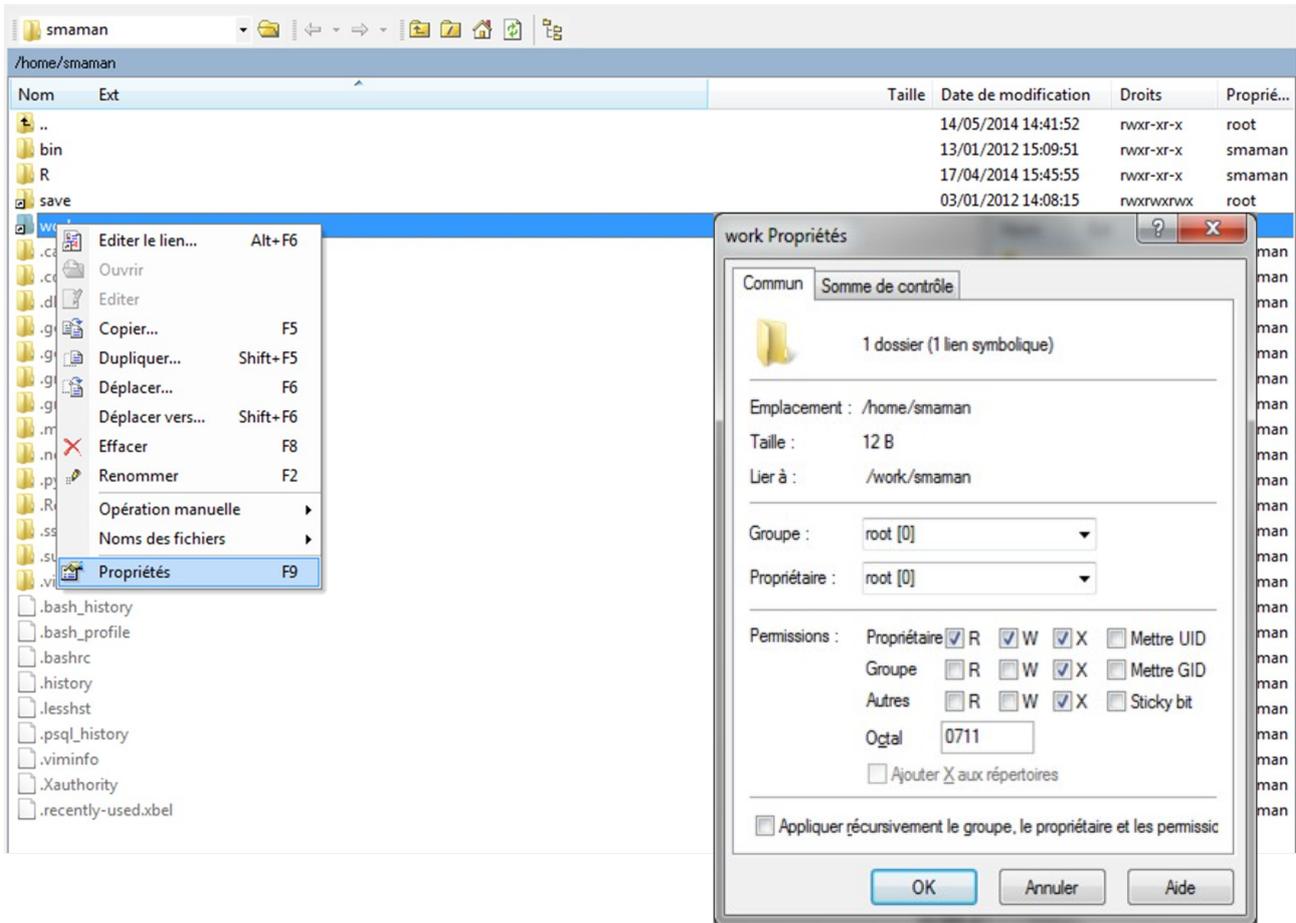


Étape 2 : Parcourez votre espace de travail /work/username/ afin de visualiser les fichiers de données que vous souhaitez importer.

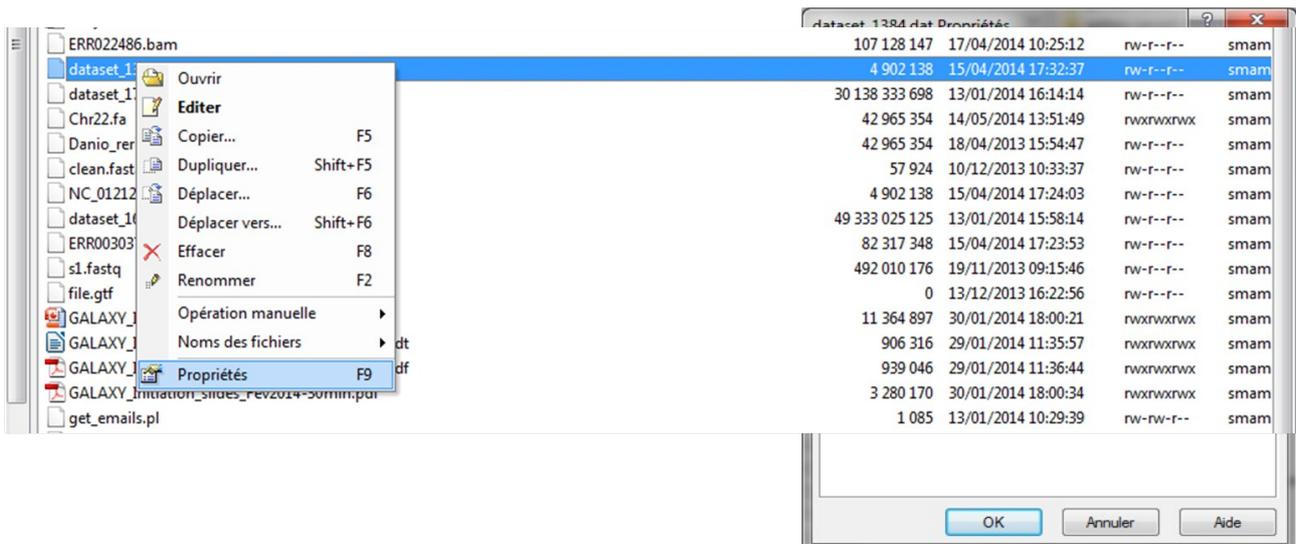
Les droits d'exécution sur le répertoire et de lecture sur les fichiers sont nécessaires pour que vos données puissent être accessibles dans Galaxy.



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 3 : Dans Galaxy, utiliser ensuite l'outil « Upload File from genotoul » 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 4: Voici quelques indications pour paramétrer l'outil Galaxy :

Upload File (version 1.0.0)

Path to file:

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

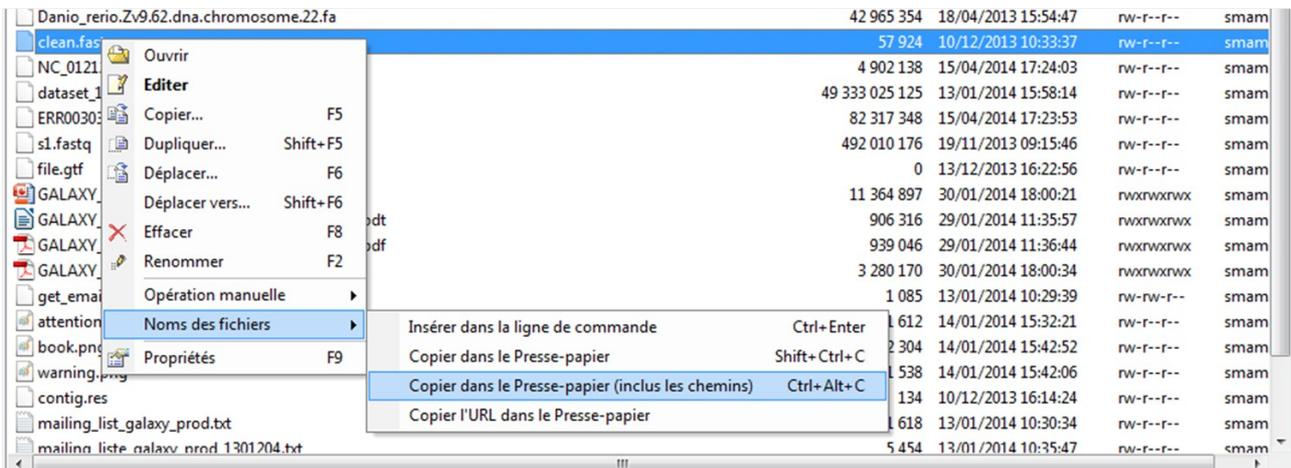
File type:

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) ».



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 fichiers d'Internet à Galaxy

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

*** Upload File (version 1.1.3)**

File Format:

 Which format? See help below

File:
 Aucun fichier sélectionné.
 TIP: Due to browser limitations, uploading files larger than 2GB is guaranteed to fail. To upload large files, use the URL method (below) or FTP (if enabled by the site administrator).

URL/Text:

Here you may specify a list of URLs (one per line) or paste the contents of a file.

Convert spaces to tabs:
 Yes
 Use this option if you are entering intervals by hand.

Genome:

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.



Cette méthode d'upload n'est pas recommandée car elle impacte considérablement votre quota.

Import de données à partir de votre PC

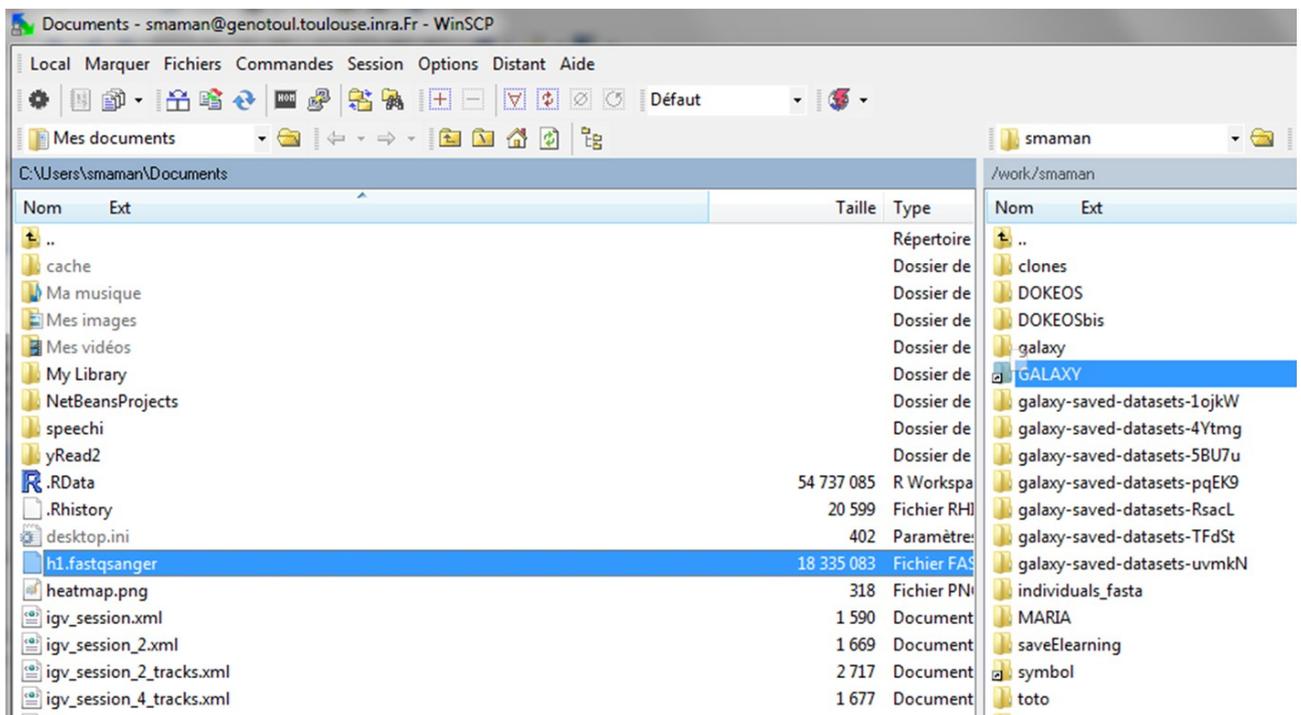
De votre PC à votre /work

Afin d'utiliser un mode de téléchargement peu coûteux en espace disque, il est préférable de travailler à partir du work. Nous allons donc utiliser WinSCP (ou FileZilla) pour déplacer vos fichiers dans votre /work/.

Pour transférer un fichier de votre PC (d'une clé USB, d'un disque dur, d'un autre serveur...) à Galaxy, nous allons de nouveau utiliser WinSCP :

Étape 1: Récupérez le fichier sur votre PC.

Étape 3 : Avec WinSCP, placer une copie de ce fichier de votre PC à votre /work en faisant glisser le fichier de la droite à la gauche de l'interface WinSCP.



Étape 6 : Ensuite, pointer votre fichier sous Galaxy depuis votre /work/username/. Voici quelques indications pour paramétrer l'outil Galaxy :

Upload File (version 1.0.0)

Path to file:

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

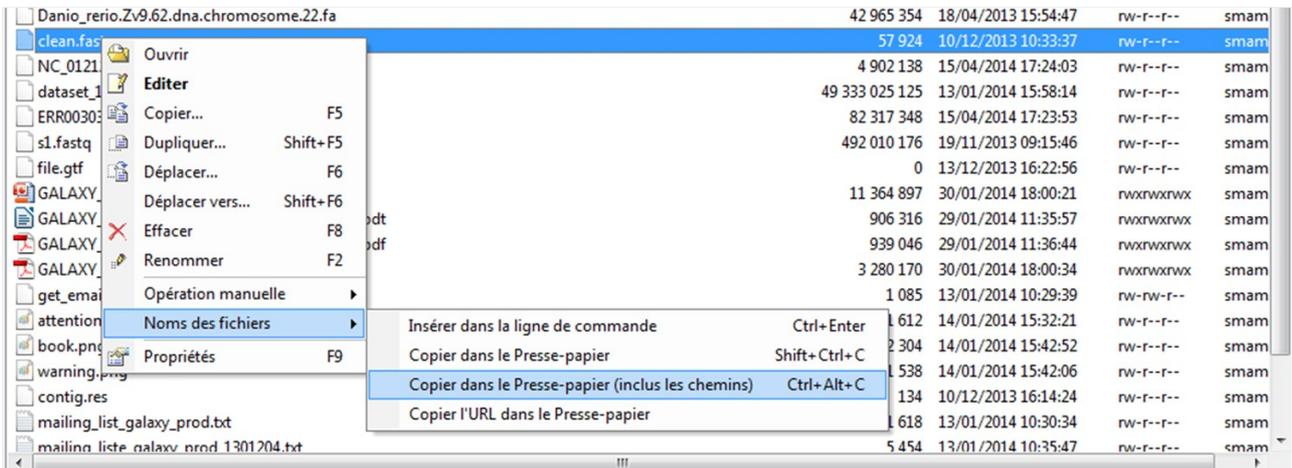
File type:

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) ».



Sauver vos 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 « [Download my Galaxy dataset on Genotoul \(work\)](#) » disponible dans la section « **Download 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.

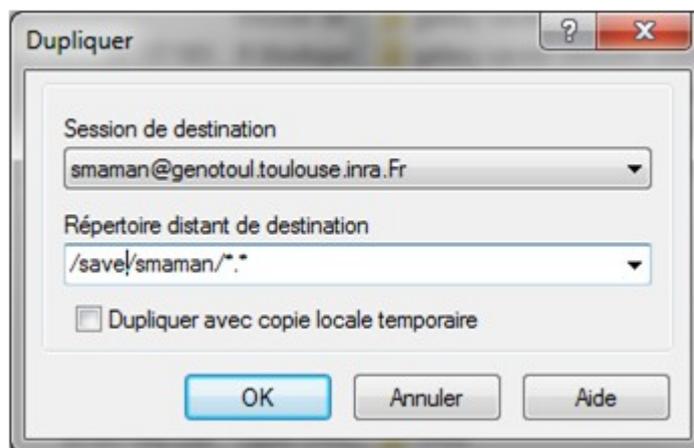
Le /work étant un espace de travail, il est régulièrement purgé par l'administrateur du cluster Genotoul. Nous vous conseillons donc ensuite de déplacer vos fichiers de votre /work vers votre /save.

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.





Sauvegarder une chaîne de traitements

Etape 1 : Depuis le menu « Options » de votre historique en cours, choisir « Extract Workflow ».
Un workflow est ainsi généré automatiquement et disponible depuis le menu « Workflow ».

Etape 2 : Exporter ensuite ce workflow en cliquant sur le flèche noir à côté de l'intitulé du workflow et choisir « Download or Export ».

Etape 3 : Download to File

Veillez cliquer sur « Download workflow to file so that it can be saved or imported into another Galaxy server. »

Puis enregistrer ce fichier sur votre PC

Il vous sera ensuite possible de le ré-importer dans votre instance Galaxy.

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

Tutoriels disponibles sur sig-learning:

Les formations de la plateforme Bioinfo Genotoul sont disponibles sur <http://sig-learning.toulouse.inra.fr>. Pour accéder aux auto-formations en ligne et aux tutoriels vidéo, veuillez suivre cette démarche :



- 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 fomations disponibles. Sinon, veuillez effectuer une recherche par mots clés (exemple : « Galaxy, « Mothur »).



Galaxy gère très mal les caractères spéciaux et les accents.