



- Galaxy -

Initiation à la plateforme Galaxy

- EXERCICES -



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



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 » : <u>http://galaxy-workbench.toulouse.inra.fr</u>
- Des fichiers disponibles à l'url suivante : <u>http://genoweb.toulouse.inra.fr/~formation/1_Galaxy_Initiation/Data/</u>
 Les documents de présentation et d'exercices sont disponibles en PDF à l'url suivante : http://genoweb.toulouse.inra.fr/~formation/1_Galaxy_Initiation_V30min

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
- <u>Password</u>: f1o2r3!

Pour répondre à vos questions:

- Mail : 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 <u>http://sig-learning.toulouse.inra.fr</u>

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

Welcome sigenae, you a

Logged in galaxy as sigenae@toulouse.inra.fr



Exercice n°1 : Connexion à Galaxy, exploration de l'interface, téléchargement de datasets.

Connexion à la plateforme Galaxy

Vous pouvez accéder à votre plateforme Galaxy (en précisant votre login et mot de passe « genotoul ») à l'adresse suivante : <u>http://galaxy-workbench.toulouse.inra.fr</u>

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.
- **Shared Data** : Accès aux bibliothèques de données, ainsi qu'aux historiques et workflows publiés.

Shared Data

Help

User cours

Saved Histories

Saved Datasets

Public Name

Analyze Data Workflow

- Data Libraries
- Published Histories
- Published Datasets
- Help :
 - Support
 - Galaxy Wiki
 - Video tutorials
 - How to cite Galaxy
- User :
 - Logged in galaxy as sigenae@toulouse.inra.fr
 - Saved Histories
 - Saved Datasets
 - Public Name
- **Note** : La documentation autour de Galaxy est très aboutie, explorer le menu « help » 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 de données

Téléchargement des fichiers avec copie sur le serveur (non recommandé)

Télécharger, avec« UploadFile », lesfichiers« reads.fastqsanger »,« NC_012125.1.fasta »,« annotation.txt »et« gene.txt »disponiblesvial'urlhttp://genoweb.toulouse.inra.fr/~formation/1GalaxyInitiation/Data/.

Renommer les datasets.



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 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 ».

Téléchargement sans copie sur le serveur (recommandé)

1 Tutoriel disponible sur sig-learning

Il vous est possible de suivre un tutoriel vidéo pour apprendre à disposer de vos fichiers NG6 dans votre interface Galaxy.

Pour accéder à ce tutoriel :

- 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 »).

2 Mise en pratique

A partir la plateforme NG6, projet public « Galaxy training », créer les liens symboliques dans votre espace de travail sur le server genotoul :

ķ	Project Demonstration	
þ	Project Demonstration2	
.	Project Galaxy training	
	🚈 📃 Run Galaxy - First steps (Sample) - (25-10-13) produced 10000 reads	
	Raw data	
	🔤 Maalyse Data files	

Étape 1 : Parcourer votre espace de travail (logiciel ftp) sur le server genotoul afin de noter le chemin vers le fichier (lien) « linux.txt ».

Étape 2 : 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 cela, vous avez deux méthodes :

* Méthode 1 pour ceux qui préfèrent le Terminal :

chmod +x REPERTOIRE et chmod +r FICHIER

* Méthode 2 pour ceux qui préfèrent les interfaces :

Maria Bernard - Sarah Maman - Ibouniyamine Nabihoudine





Veuillez ouvrir FileZilla (ou WinSCP):

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



Clic droit sur le nom du fichier, puis « Droits d'accès au fichier », puis donner les droits d'exécution sur les répertoires en amont de votre fichier et les droits de lecture sur votre fichier.

Étape 3 : 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.

Avec FileZilla, il est possible de récupérer ce chemin dans le champs « site distant » sur la gauche

Maria Bernard - Sarah Maman - Ibouniyamine Nabihoudine



extenstion.

5 Février 2014 de l'interface.Veillez à ne pas oublier de renseigner aussi le nom du fichier et son

Étape 4 : 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 »).

Étape 5 : Utiliser l'outil « **Upload local file from filesystem path** Upload data to history without copying on server » afin créer le lien dans votre historique galaxy.

Même exercice avec run test.



L'outil « Upload local file from filesystem path Upload data to history without copying on server » 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.

Exercice n°2: Création et partage de datasets, d'historiques et de workflows.

Notions d'historique

Traitements archivés dans un historique

Au fur et à mesure que vous faites appel aux différents outils au sein de votre interface depuis le menu « Analyse Data », l'ensemble des étapes sont enregistrées dans un historique qui est automatiquement archivé dans « User / Saved Histories » et que vous pouvez ensuite, si besoin, partager dans « Shared Data / Published Histories ».



Gérer ses historiques

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

Remarque, 1013	ue voue connexi	on uu wo	INDUICII V	Juluxy,	un «current	1113t01 y // C3t
Tools Options Linux Username: sigenae File Path : /work/sigenae/galaxy/ Serveur :galaxy Get Data	Saved Histories search history names and tags			-7·17		
Send Data	Name	Datasets	Tags Sharing	Size on Disk	Created	Last Updated ↑
Lift-Over	Unnamed history		0 Tags	0 bytes	less than a minute ago	less than a minute ago
Text Manipulation Filter and Sort	RNA seq statistics 🔻	2 4	<u>0 Tags</u>	34.9 Mb	~ 6 hours ago	6 minutes ago
Join, Subtract and Group	Test BWA fichiers Gnome 🔻	4	<u>0 Tags</u>	9.0 Mb	Apr 06, 2012	1 day ago
Extract Features	Test region promoters *	5	<u>0 Tags</u>	23.9 Mb	Mar 08, 2012	Apr 06, 2012
Fetch Sequences Fetch Alignments	Unnamed history -	3	<u>0 Tags</u>	60 bytes	Feb 23, 2012	Mar 09, 2012
Get Genomic Scores Operate on Genomic Intervals	Unnamed history +	1	<u>0 Tags</u>	0 bytes	Mar 07, 2012	Mar 07, 2012
Statistics Wavelet Analysis	Unnamed history +	1	<u>0 Tags</u>	16.0 Kb	Feb 22, 2012	Mar 05, 2012
Graph/Display Data	For 2 selected histories: Rename	elete Delete Permanent	ly Undelete			
Regional Variation Multiple regression	Histories that have been deleted for more that	n a time period specified by	the Galaxy administrator(s) may be permanent	dy deleted.	

5 Février 2014



• Créer un nouvel historique (nommer le en le préfixant par votre login) et ajouter (copie) un ou plusieurs de vos datasets

- Partager ce nouvel historique avec votre voisin
- Importer l'historique de votre voisin
- Modifier votre historique
- Réimporter l'historique modifié de votre voisin
- Supprimer l'historique que vous partagez
- Supprimer les historiques importés

Notions de workflow : convertir un historique en workflow.

Convertir un historique en workflow

A partir d'un de vos historiques, voici les principales étapes pour créer un workflow :

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

Création de workflow :

- 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.

Créer un nouveau workflow puis l'exécuter

Tout d'abord, veuillez choisir un fichier de test par trop lourd pour réaliser cet exercice. Nous vous proposons de travailler sur le fichier « reads.fastqsanger ».

Créer un nouveau workflow à partir des traitements bioinformatiques suivants : FastQC Reprot et Filter Fastq (qualité supérieure à 30 et les reads correctement mappées). Une fois votre workflow créé et chaque outil paramétré, veuillez exécuter votre workflow.



Maria Bernard - Sarah Maman - Ibouniyamine Nabihoudine