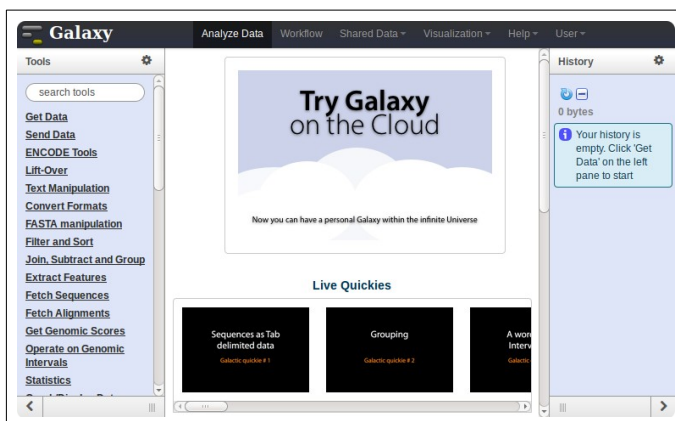




- *Galaxy* -

Initiation à la plateforme Galaxy

- *EXERCICES* -



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



Objectifs :

Cette formation a pour objectif de vous familiariser à l'utilisation de votre workbench Galaxy (<http://galaxy-workbench.toulouse.inra.fr>).

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/~sigenae/sarah/UPS/GALAXY/DATA/>

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.



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.

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



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

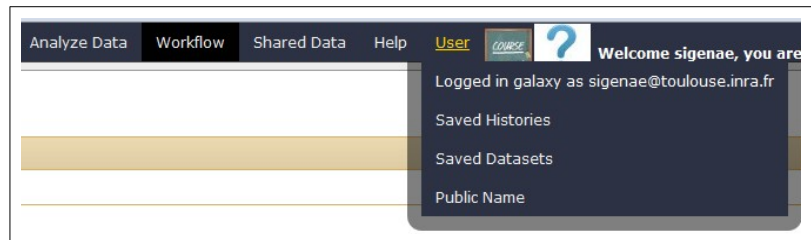
Connexion à la plateforme Galaxy

Vous pouvez accéder à votre plateforme Galaxy (en précisant votre login et mot de passe « genotoul ») à l'adresse suivante : <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.
- **Shared Data** : Accès aux bibliothèques de données, ainsi qu'aux historiques et workflows publiés.
 - 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

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

Télécharger, avec « Upload File », les fichiers « reads.fastqsanger », « NC_012125.1.fasta », « annotation.txt », « linux.txt », « chr22.fa » et « gene.txt » disponibles via l'url <http://genoweb.toulouse.inra.fr/~sigenae/sarah/UPS/GALAXY/DATA/>

Renommer les datasets.



L'ensemble des outils permettant l'import dans Galaxy est disponible dans la section « Upload your 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 ».

2 Télécharger des données de l'UCSC via l'outil « UCSC Main table browser »

Télécharger l'annotation (gènes RefSeq) du chromosome 1 bovin (btau4), paramètres :

- Clade : Mammal
- Genome : Cow
- Assembly : Oct. 2007
- Group : Genes and Genes Prediction Tracks
- Track : RefSeq Genes
- Table : refGene
- Region – position : chr1:1-161106243 (enter « chr1 » puis cliquer sur « lookup »)
- Output format : BED – browser extensible data
- Sélectionner « Send Output to Galaxy » puis cliquer sur « Get Output » et « Send query to Galaxy ».

Visualiser le nouveau dataset, et notamment ses propriétés (« database »). Que remarquez vous ?

Explorer les liens disponibles pour ce dataset.

Relancer le téléchargement en modifiant le Output format : GTF – gene transfert format
Comparer les deux fichiers GTF et BED.

Exercice n°2: Utilisation d'outils de traitement de fichiers (équivalent aux commandes Linux)

Outils de traitement de fichiers

- En utilisant l'outil « Add column to an existing dataset » ajouter une colonne « chr1 » au fichier « linux.txt »
 - Ajouter la colonne
 - Renommer le dataset obtenu en « linux_add »
- Trier numériquement le fichier « linux_add » par ordre descendant sur la première colonne
 - Outil « Sort data in ascending or descending order »
 - Renommer le fichier généré en « linux_add_sort »



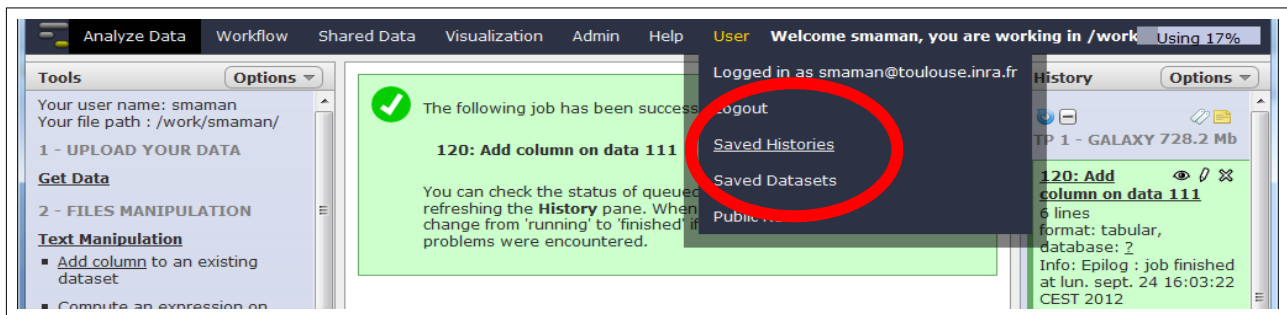
- A partir des deux datasets BED et GTF « UCSC Main on Cow: refGene (chr1:1-161106243) » précédemment importés, extraire du génome la séquence de chacun des gènes.
Utilisation de l'outil « Extract Genomic DNA » de la section « Fetch sequences » (output data type « Interval »)
- Comparer le nombre de lignes dans les nouveaux datasets avec ceux d'origine. Pourquoi cette différence ?
- Convertir le dataset obtenu à partir du BED en multi-fasta.
Utilisation de l'outil « Tabular-to-FASTA converts tabular file to FASTA format »
- Calculer le %GC des gènes (outil "Compute GC content")
Utilisation de l'outil « Compute GC content »
- Calculer la longueur de chaque gène
Utilisation de l'outil « Compute sequence length »

Exercice n°3: 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 automatiquement créé.



Tools Options

Linux Username: sigenae
File Path : /work/sigenae/galaxy/
Serveur : galaxy
Get Data
Send Data
ENCODE Tools
Lift-Over
Text Manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats
Extract Features
Fetch Sequences
Fetch Alignments
Get Genomic Scores
Operate on Genomic Intervals
Statistics
Wavelet Analysis
Graph/Display Data
Regional Variation
Multiple regression

Saved Histories

search_history_names_and_tags

Advanced Search

Name	Datasets	Tags	Sharing	Size on Disk	Created	Last Updated
Unnamed history		0 Tags		0 bytes	less than a minute ago	less than a minute ago
RNA seq_statistics	2 4	0 Tags		34.9 Mb	~ 6 hours ago	6 minutes ago
Test BWA fichiers Gnome	4	0 Tags		9.0 Mb	Apr 06, 2012	1 day ago
Test region promoters	5	0 Tags		23.9 Mb	Mar 08, 2012	Apr 06, 2012
Unnamed history	3	0 Tags		60 bytes	Feb 23, 2012	Mar 09, 2012
Unnamed history	1	1	0 Tags	0 bytes	Mar 07, 2012	Mar 07, 2012
Unnamed history	1	1	0 Tags	16.0 kb	Feb 22, 2012	Mar 05, 2012

For 2 selected histories: **Rename** **Delete** **Delete Permanently** **Undelete**

Histories that have been deleted for more than a time period specified by the Galaxy administrator(s) may be permanently deleted.

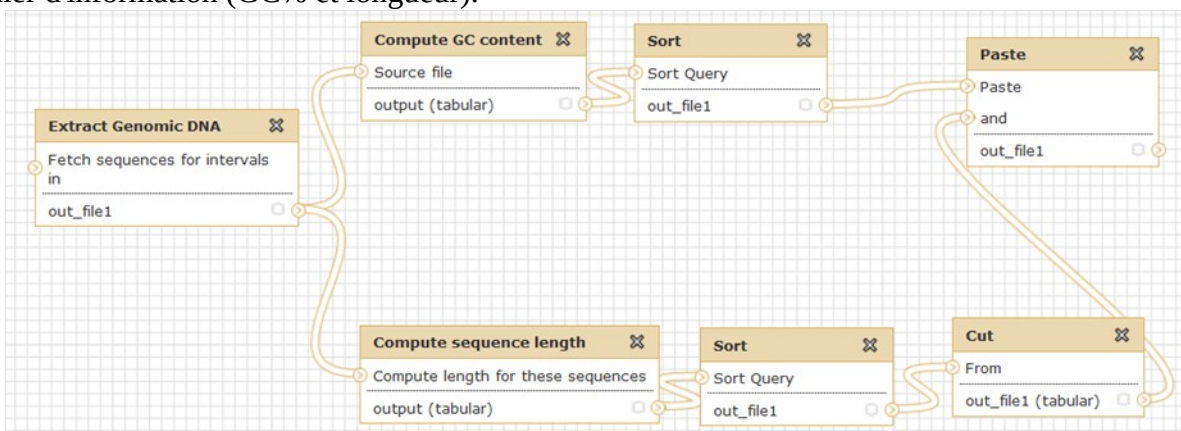
Exercice

- 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 galaxy-prod@toulouse.inra.fr

Notions de workflow : convertir un historique en workflow.

Convertir un historique en workflow

Créer un workflow à partir des traitements bioinformatiques précédemment réalisés. Soit un workflow permettant, à partir d'une annotation (format BED), de générer un multi-fasta ainsi qu'un fichier d'information (GC% et longueur).



Les principales étapes :

- « History panel » Options → « Extract workflow »
- Sélectionner les bons datasets
- Créer le workflow
- Partager le workflow créé avec galaxy-prod@toulouse.inra.fr

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.