

## TP 1 : Connexion à la plateforme

*Objectifs : S'approprier les notions d'utilisateur et d'arborescence.*

1. Ouvrir une connexion linux sur la plateforme (accès depuis un poste WINDOWS)
  - Lancez Xming
  - Lancez PuTTY
  - Configurez PuTTY de la manière suivante :
    - dans le menu « connection / ssh / X11 » cochez « Enable X11 forwarding »
    - dans le menu « terminal / keyboard » : « the function keys » doit être « Linux »
    - enfin dans le menu « Windows / Translation » sélectionner UTF8
  - Revenir au menu « Session » et saisir le serveur « Host name »  
genologin.toulouse.inra.fr
  - Sauvegarder la configuration GENOLOGIN
  - Cliquer sur « Open » pour ouvrir la session
  - Saisir votre login et votre mot de passe
2. Changer votre mot de passe
3. Dans quel répertoire êtes-vous situé ?
4. Qui est connecté (who) ?
5. Afficher l'aide de la commande 'ls' (utiliser la commande man, taper 'q' pour sortir)
6. Lister les fichiers dans votre 'home directory', combien de fichiers avez vous ?
7. Lister le répertoire des programmes de bioinformatique: /usr/local/bioinfo/src (chemin absolu)
8. Lister les banques blast : /bank/blastdb (chemin absolu)
9. Depuis votre répertoire d'accueil, allez dans le répertoire 'parent' (chemin relatif)  
Lister le contenu de ce répertoire
10. Retourner dans votre répertoire d'accueil
11. Lister les commandes que vous avez exécutées ? Et ré-exécuter une ancienne commande ( !num\_de\_commande)

## TP 2 : Gestion des répertoires, fichiers et droits d'accès.

*Objectifs : Se déplacer dans l'arborescence, sécuriser ses données.*

*État de vos répertoires actuel :*

- /home/user : fichiers de configuration
- /save/user : répertoire sauvegardé mais encore vide
- /work/user : espace de travail, fichiers de résultats d'analyse

*Le but de ce TP est de vous mettre en situation : vous avez lancé des calculs (dans l'espace work) et vous devez organiser et sécuriser les résultats dans votre espace sauvegardé (save).*

*Au final, les fichiers fasta vont se retrouver dans le répertoire « data » de l'espace save. Les fichiers blast vont se retrouver dans le répertoire « blast » de l'espace save. Vous interdirez ensuite l'accès à ces répertoires comme indiqué dans les questions suivantes.*

**Avant de commencer le TP taper la commande suivante :**

```
cp -r /save/formation/public_html/unix/data/* ~/work
```

1. Dans votre répertoire « save », créer un répertoire «tp\_unix», aller dans ce répertoire
2. Créer les sous répertoires « data », « blast\_result ».
3. Lister les fichiers de votre « work directory ».
4. Sauvegarder vos données :
  1. Déplacer les fichiers « \*.fasta » dans data ; vérifier qu'ils n'existent plus dans le work.
  2. Copier les fichiers résultats de blast « \*.blast » dans le repertoire « blast\_result »
5. Depuis votre home directory, rechercher les fichiers nommés « ab\*.fasta » avec la commande find.  
Afficher l'aide pour trouver l'option de recherche dans les répertoires save et work qui sont des liens symboliques.
6. Rechercher les fichiers dont la taille est supérieure à 100Ko (dans le répertoire ~/save/tp\_unix).
7. Aller dans le répertoire « ~/save/tp\_unix », enlever le droit de lecture à tout le monde au dossier « data ». Pouvez vous lister le contenu de data? Rajouter le droit de lecture.
8. Enlever le droit d'exécution à tout le monde au dossier « data ». Pouvez-vous rentrer

dans le répertoire « data » ? ajouter le droit d'exécution au user (vous même).

9. Si vous souhaitez a nouveau faire des calculs sur un fichier fasta (ab073182.fasta), au lieu de copier ce fichier dans l'espace work, vous pouvez créer un lien symbolique (raccourci).

Aller dans répertoire ~/work et créer le lien symbolique vers le fichier  
~/save/tp\_unix/data/ab005233.fasta

## TP 3 : Manipulation d' archives, extraction d'information dans des fichiers.

*Objectifs : Récupérer des données du web pour les analyser, extraire l'information significative dans un résultat d'analyse*

*Transfert d'un fichier de séquences depuis son poste de travail vers le serveur.*

1. Utiliser le navigateur internet de votre poste et récupérer sur **votre poste de travail** le fichier disponible a cette adresse :  
<http://genoweb.toulouse.inra.fr/~formation/unix/tp3/reads.fastq.gz>
2. Transférer ce fichier avec filezilla sur genologin : ~/save/tp\_unix/data
3. Se déplacer dans le répertoire data et décompresser le fichier .gz

*Acquisition d'un fichier de séquences directement sur le serveur genologin.*

4. Toujours dans le répertoire data, télécharger (via la commande wget) une fiche Swissprot :  
<http://www.uniprot.org/uniprot/Q96D37.txt>

*Compression / Décompression*

5. Retourner dans votre répertoire « ~/save/tp\_unix ». Archiver et compresser le répertoire blast\_result. Afficher la taille du répertoire et de son contenu (« du »), afficher la taille de l'archive
6. Supprimer le répertoire blast\_result puis décompresser l'archive précédemment créé

*Nous allons lancer un blast pour travailler sur des résultats au format tabulé (option -m8 ou -m9)*

7. Lancer les commandes suivantes pour lancer le blast:  
module load bioinfo/blast-2.2.26 (charge le logiciel à utiliser)  
blastall -p blastn -i ~/save/ab005233.fasta -d alu.n -m9 -o  
~/work/ab005233\_alu.blast  
Afficher le résultat.

8. Trier le fichier `~/work/ab005233_alu.blast` selon le « %identité » ( 3ieme colonne) .  
*Penser à enlever les 4 premières lignes avec la commande tail*
9. Sur ce même fichier blast, n'afficher à l'écran que les noms des « subject » :  
*Utiliser la commande cut -f.*

#### Concaténation de données

10. Aller dans le répertoire « `~/save/tp_unix/data` », concaténer les fichiers fasta « `ab005*.fasta` » dans un nouveau fichier nommé "mes\_sequences.fasta", compter le nombre de séquences dans le fichier.
11. Ajouter au fichier « `mes_sequences.fasta` » la séquence « `ab017070.fasta` »
12. Afficher le fichier "mes\_sequences.fasta" page par page.  
Rechercher la chaîne de caractères « `AB017070` » pour vérifier que la séquence a bien été ajoutée. *Utiliser le « / ».*
13. Compter le nombre de séquence grâce a la commande grep.
14. Comparer avec la commande diff le fichier "ab106670.fasta" avec le fichier `/save/formation/tp_unix/ab106670_bis.fasta`
15. Rechercher dans les fichiers de séquences fasta celles qui contiennent le motif suivant : "ttatatc" (utiliser la commande grep)

#### Un pas vers la bioinfo ...

16. Taper les commandes commandes suivantes et suivre les instructions :  
module load bioinfo/EMBOSS-6.6.0  
fuzznuc