

TP UNIX

Duree: 2H

V.2003-2004.2 Daniel Gautheret

Objectifs

A la fin de cette séance, vous devez savoir:

- manipuler sous Unix les fichiers et répertoires
- lancer des programmes, passer le résultat d'un programme à un autre (pipelines et redirections)

Vous pourrez trouver sur [ce site Web](#) une bonne introduction à Unix.

Note : la ligne `export PATH=.:$PATH` peut être nécessaire dans le fichier `.bashrc`

1. Prise en main

- Connectez-vous sur votre station Unix avec votre nom et votre mot de passe (communiqués en TD). S'il s'agit de votre première connexion,
- Les commandes Unix sont entrées dans une fenêtre de type 'xterm'. Sachez créer et supprimer les fenêtres 'xterm'. Voyez bien la différence entre "réduction" et "suppression" des fenêtres.
- Voyez comment rappeler la commande précédente avec la touche "flèche haut", puis "flèche gauche" pour corriger.

2. Fichiers et Répertoires

- Créez un Répertoire TPUNIX à la racine de votre Répertoire personnel (`mkdir`)
- Placez-vous dans le répertoire TPUNIX (`cd`)
- Vérifiez que vous êtes bien dans ce répertoire (`pwd`).
- Copiez dans TPUNIX tous les fichiers se trouvant dans le répertoire `~dgaut/TPUNIX/` (commande `cp`).
- Afficher le contenu de ce répertoire. Quelle est la taille des fichiers? Le propriétaire? La date de création? Quels sont les fichiers executables? (commande `ls` et `ls -l`).
- Le fichier `"test.pl"` est un programme. Exécutez-le.
- Tout en restant dans le répertoire courant, exécutez la commande `"test2.pl"` qui se trouve dans `"~dgaut"`
- Affichez le contenu du fichier `« 16s.seq »` à partir de votre répertoire racine.
- A l'aide de la commande `cat` et des redirections (`>` et `>>`), concatenez le fichier `'16s.seq'` et le fichier `'petiteseq.fasta'` dans le fichier `'allseq'`. Affichez le résultat.

3. Edition

- Lancez l'editeur nedit (commande `nedit <nom de fichier> &`). Vérifier toutes les fonctionnalités de base : Entrée de texte, sauvegarde, recherche, couper et coller via les menus et/ou touches contrôle-xxx.
- Voyez comment effectuer les copier/coller d'une fenêtre à l'autre, à l'aide de la souris. *Les fonctions copier/coller sous Unix-Xwindows sont réalisées à l'aide des boutons de la souris. Bouton de gauche pour copier, bouton central (ou deux boutons en même temps) pour coller.*

4. Commandes diverses

- Familiarisez-vous avec les commandes suivantes:
 - `man` (utilisez-le pour comprendre les commandes ci-dessous)
 - `ps` (quelle option pour voir tous les process?)
 - `rm` (effacez un fichier)
 - `du` (quel espace disque occupez-vous?)
 - `kill` (lancez nedit, voyez son numéro de process avec `ps`, tuez-le a partir d'une autre fenêtre).

5. Analyse d'un fichier Genbank avec egrep

Objectif: Extraire rapidement les informations présentes dans un fichier Genbank.

- Regardez le fichier "mgen.gbk" à l'aide de la commande "more". Repérez les séquences protéiques. Que veut dire "CDS"? Que veut dire "complement" après CDS? Où se trouve la séquence nucléotidique? Quelles informations sont disponibles sur chaque gène?
- Avec la commande egrep, comptez les éléments suivants
 - les gènes protéiques annotés
 - les gènes protéiques présents sur le brin inverse
 - les tRNA
 - Les gènes prédits par similitude de séquence
 - Les gènes prédits par GeneMark .

attention aux majuscules/minuscules: utiliser egrep -i en cas de doute

6) Blast, localement

Blast ne travaille pas sur un fichier Fasta. Il faut pré-traiter les fichiers Fasta avec le programme *formatdb*.

A la racine de votre répertoire, créez un fichier `.ncbirc` contenant:

```
[NCBI]
  Data=/export/homes/personnels/gbma/dgaut/Bin

[BLAST]
BLASTDB=.
```

Ce fichier indique où se trouvent les divers fichiers indispensables à Blast, p. ex. les matrices de substitution

- A l'aide de la commande formatdb préparez le fichier Fasta mgen.fasta au traitement par Blast. Tapez formatdb seul pour voir les arguments. Attention: arguments différents pour les séquences nucléiques et protéiques. Regardez quels fichiers ont été créés.
- Tapez blastall sans argument pour obtenir la liste des arguments de Blast. L'argument -p qui spécifie la version de Blast est indispensable (p. ex: -p blastp pour Blast protéine), ainsi que les arguments -d et -i.
- Récupérez dans le répertoire TPUNIX la séquence 16s.seq. Regardez de quoi il s'agit.
- Avec Blast, recherchez dans le génome de M. genitalia des séquences similaires à 16s.seq.
- Exécutez Blast de façon à n'obtenir que les solutions de E-value inférieure à $10e-4$, en redirigeant la sortie vers un fichier. Conservez ce fichier.