

- 1 Familiarisation avec l'environnement Linux
- 2 Installation complète d'un environnement Conda RNA-Seq
- 3 Utilisation du cluster de calcul

TD - Introduction à Linux et à l'Environnement Conda

Ghislain Bidaut, Plateforme Cibi, CRCM, Aix-Marseille Université
24/03/2022

Formation
Bioinformatique
Ghislain BIDAUT
Aix-Marseille Université



1 Familiarisation avec l'environnement Linux

1.1 Découverte du terminal

- Entrez les commandes

```
cd  
pwd  
cd /tmp  
cd -  
pwd
```

- Créer le dossier `activite_linux` dans votre répertoire personnel.
- Créer deux fichiers vides `activite_1.txt` et `activite_2.txt` **dans** ce dossier.
- Créer deux fichiers vide en remplaçant les `'_'` des fichiers précédents par des espaces. Est-ce une bonne idée ?
- Afficher le contenu du dossier `activite_linux` en explorant les différentes options de la commande. Enregistrer ce contenu dans un fichier contenant uniquement une ligne par nom de fichier.
- Ajouter les alias suivants de manière permanente dans votre environnement

```
cp = 'cp -i'  
rm = 'rm -i'  
mv = 'mv -i'
```

- Lister les 30 dernières commandes.
- Utiliser la commande `find` pour trouver les fichiers portant l'extension `.txt` dans votre répertoire utilisateur.

1.1.1 Solution

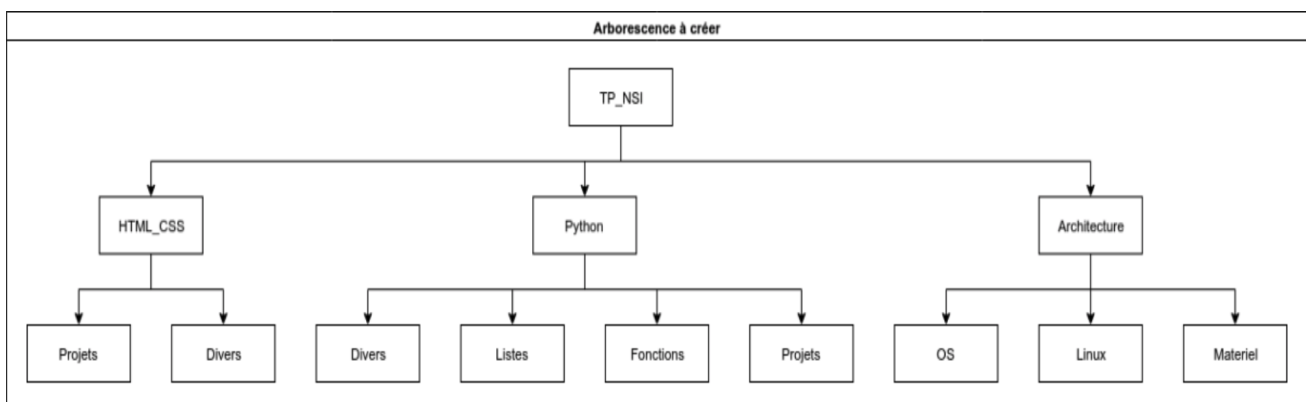
1.2 Manipulation d'un fichier à plusieurs colonnes

- Télécharger le fichier d'annotations de la levure à l'adresse `ftp://ftp.ensemblgenomes.org/pub/release-37/fungi/gtf/saccharomyces_cerevisiae/Saccharomyces_cerevisiae.R64-1-1.37.gtf.gz` (`ftp://ftp.ensemblgenomes.org/pub/release-37/fungi/gtf/saccharomyces_cerevisiae/Saccharomyces_cerevisiae.R64-1-1.37.gtf.gz`)
- Décompresser ce fichier
- Afficher les 5 premières lignes
- Extraire les 4 premières colonnes dans un fichier texte. Parcourir ce fichier avec la commande `less`.
- Extraire uniquement les lignes correspondant aux gènes et les stocker dans un fichier `sc_genes.txt` (indice: utiliser `grep -P`).
- Extraire de `sc_genes.txt` la colonne contenant les identifiants de gènes (exemple `"YDL248W"`). Il faut que les guillemets (") soient supprimés.
- Créer un fichier texte contenant le nom des gènes présents, leur chromosome, leur position chromosomique (début-fin) et le "strand". En créer une version supplémentaire trié par nom de gène.

1.2.1 Solution

1.3 Création d'un script bash

- Ouvrir l'éditeur de fichiers `gedit` (ou votre éditeur préféré).
- Ecrire un script `bash` appelé `creation_arborescence.sh` qui recrée l'arborescence suivante de manière automatisée.
- Commenter ce script en détaillant les opérations.



- Adapter les permissions sur ce script pour le lancer avec la commande

```
./creation_arborescence.sh
```

- Vérifier le résultat final avec `tree`.

1.3.1 Solution

2 Installation complète d'un environnement Conda RNA-Seq

2.1 Installation de l'environnement Conda

- Connectez vous au cluster de l'IFB (IFB-Core).

(Optionnel: créez un **alias** pour cette commande)

```
ssh user@core.cluster.france-bioinformatique.fr
```

- Installer **miniconda** dans votre environnement.
- Installer les dépôts `bioconda` et `conda-forge`.
- Créer un environnement `ngs`.
- Vérifier que cet environnement ait été correctement installé.
- Activer cet environnement.
- Installation des outils suivants: `fastqc`, `multiqc`, `samtools`, `subread`, `star`, `trimmomatic`.
- Tester ces outils et regarder leurs options de ligne de commande. En particulier, comment spécifie-t-on les entrées-sorties ?
- Enregistrer la configuration de cet environnement dans un fichier `RNAseq_env.yml`.
- Désactiver l'environnement `ngs`.
- Effacer l'environnement `ngs`.
- Le recréer sous le nom `rnaseq` à partir du fichier `RNAseq_env.yml`.

2.1.1 Solution

3 Utilisation du cluster de calcul

- Connectez vous au cluster de l'IFB (IFB-Core).
- Créer un script **Bash** permettant de lister les fichiers de votre répertoire en format long.
- Y ajouter les variables `SBATCH` suivantes:
 - Utilisation du nom de job: `lsformation`
 - Envoi d'un mail pour signaler le début et la complétion du job.
 - Utilisation de la partition `fast`
 - Limiter la durée du script à 10 minutes
 - Demander 1Go de mémoire vive
 - Demander 1 CPU (Cœur)
- Soumettre ce script à Slurm (commande `sbatch`) sur le compte `form_2022_09`.
- Le "surveiller" avec `squeue`.
- Examiner le résultat de sortie.



(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Cette œuvre est mise à disposition selon les termes de la Licence Creative Commons:

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International (CC BY-NC-ND 4.0).