

# TP 8 - Pratique sur les données de séquençage

## Alignement des lectures contre le génome de référence :

Nous avons des lectures provenant du séquençage et nous voulions l'aligner contre un génome de référence. Pour ce faire, nous allons utiliser l'outil **bwa**.

Sur votre le site de BIN1002 ainsi que le chemin ci-dessous, vous allez trouver les données nécessaires pour faire l'alignement.

```
cp /UDEM/COURS_BAC/BIN1002/dbin1002/genome.fa
```

**chemin/de/votre/repertoire**

**ps** : Attention de bien mettre votre propre chemin en tapant **pwd**

### 1)- Indexage :

On index le génome de référence :

```
bwa index genome.fa
```

### 2)- Alignement :

On aligne les lectures contre la séquence de référence :

```
bwa mem genome.fa reads_total.fastq > output.sam
```

À partir du fichier au format SAM (**S**equences **A**lignment/**M**ap), pouvez-vous indiquer sur quel chromosome les lectures ont été aligné? À première vue, est-ce qu'il a une bonne qualité d'alignement? Justifier votre réponse. Indiquez le score d'alignement pour les lectures suivantes pour un coût de 0 pour un "match", +1 pour un "mismatch" et +1 pour un indel:

**S.cereviceae\_1752**

**C\_elegans\_18S\_523**

**Mus\_musculus\_1791**

Est-ce qu'il a des lectures qui ne sont pas alignés sur le génome? Si oui, citez deux.

Lien utile :

<https://www.france-bioinformatique.fr/sites/default/files/formats.pdf>