

RECHERCHE DE MOTIFS RÉPÉTÉS DANS UN GÉNOME

Auteurs :

- Leo PERARD
- Salla DIAGNE

Listing des fichiers et répertoires du projet

- *donnees/* : contient des fichiers de tests au format FASTA
- *src/* : contient les fichiers JAVA du projet
- *test/* : contient les tests unitaires concernant le projet
- *strand_searching.jar* : jar contenant le programme principal

Fonctionnement du programme

```
DESCRIPTION : recherche de motifs repetes dans un genome
USAGE : java -jar strand_searching.jar filename [strand|N] [-comp|-rev|-revComp]
* --USING [-bf|-so|-kr|-kmp|-bm]*
    filename : le nom du fichier fasta ou se trouve le genome a etudier
    [strand|N] : permet de rechercher soit :
        * strand : une sequence dont les occurences seront recherchees dans le genome
        * N : rechercher les occurences des mots de taille N
    [-comp|-rev|-revComp] : permettent de rechercher egalement pour le mot entre ou les occurences des mots de taille N :
        * comp : le complementaire
        * rev : le reverse
        * revComp : le reverse-complementaire
    [-bf|-so|-kr|-kmp|-bm] : permet de spécifier le ou les algos a rechercher parmi :
        * bf : Brute-force
        * so : Shift-Or
        * kr : Karp-Rabin
        * kmp : Knutt-Morris-Pratt
        * bm : Boyer-Moore
EXEMPLE : java -jar strand_searching.jar donnees/simple.fasta TATA --WITH -revComp
```

```
-comp -rev --USING -kr -bf -so -bm -kmp
```

Cet exemple affichera sur la sortie standard les occurrences du mot TATA, de son reverse, de son complémentaire et de son

reverse-complémentaire dans le genome du fichier donnees/simple.fasta, en utilisant les algorithmes Karp-Rabin, Brute-Force, ShiftOr, Boyer-Moore et Knuth-Morris-Pratt