

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
        * dotplot : pour generer un dotplot comparant le genome a lui-meme
    [-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
Si aucun algorithme n'est specifie, l'algorithme de Boyer-Moore sera utilise.
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 occurences du mot TATA, de son
reverse, de son complementaire et de son
reverse-complementaire dans le genome du fichier donnees/simple.fasta, en
utilisant les algorithme Karp-Rabin, Brute-Force,
ShiftOr, Boyer-Moore et Knuth-Morris-Pratt
```

Exemple de résultat du programme

```
$ java -jar strand_searching.jar donnees/exemple3.fasta GATA --WITH -comp -rev  
-revComp --USING -bf -so -kr -kmp -bm
```

```
Algorithme naif (BruteForce)
```

```
GATA : [143, 173, 710, 796, 1021]
```

```
CTAT : []
```

```
ATAG : [1022]
```

```
TATC : [557, 1518]
```

```
Temps d'execution : 24081831 nanosecondes.
```

```
Algorithme ShiftOr
```

```
GATA : [143, 173, 710, 796, 1021]
```

```
CTAT : []
```

```
ATAG : [1022]
```

```
TATC : [557, 1518]
```

```
Temps d'execution : 52803673 nanosecondes.
```

```
Algorithme de Karp-Rabin
```

```
GATA : [143, 173, 710, 796, 1021]
```

```
CTAT : []
```

```
ATAG : [1022]
```

```
TATC : [557, 1518]
```

```
Temps d'execution : 56855110 nanosecondes.
```

```
Algorithme de Knuth-Morris-Pratt
```

```
GATA : [143, 173, 710, 796, 1021]
```

```
CTAT : []
```

```
ATAG : [1022]
```

```
TATC : [557, 1518]
```

```
Temps d'execution : 10918704 nanosecondes.
```

```
Algorithme de Boyer-Moore
```

```
GATA : [143, 173, 710, 796, 1021]
```

```
CTAT : []
```

```
ATAG : [1022]
```

```
TATC : [557, 1518]
```

```
Temps d'execution : 8892501 nanosecondes.
```

```
$
```