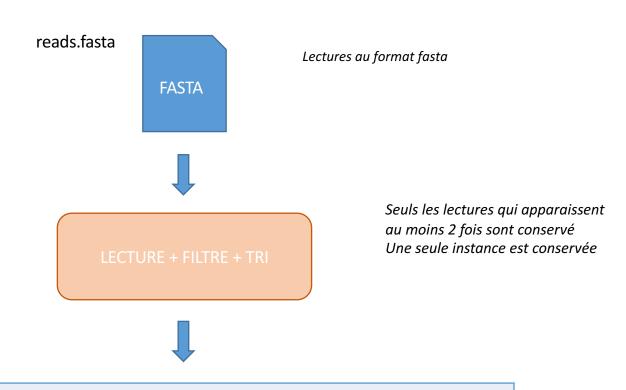
Objectif

Stocker et Trier un jeu de données de séquençage

Données

Format fasta

Taille des lectures : 100 bp



Sort_reads.txt

ATGAGACACAGATAGAGACACAGATATAGAGGACACAGGGATA
CACAGGATATACCACAGAGGATTTAGGAGAGACACAGGAGAA
CTGGAGAGACACAGGAGAAAAGATACACGAGGATTAGGACACGAG
TCTATTATATATAGGCCAGATTAGGACACAGGGAGAGAAAT

Lectures triées par ordre alphabétique

Matériel

- Programme generateFastaFile
 - Usage : generateFastaFile <genome_size> <coverage>

Mbp

• Génère un fichier reads.fasta

TP1: déroulement

1. Charger et compiler le programme **generateFastaReads**

```
gcc -O3 -o generateFastaReads generateFastaReads.c
```

2. Ecrire le programme tp1

```
en entrée : reads.fasta en sortie : sort_reads.txt
Format de sort reads.txt = lectures sans commentaires
```

- 3. Tester le programme sur des petits jeux de données
- 4. Mesurer les performances du programme avec les paramètres suivants
 - Taille génome : 1 Mbp, 3 Mbp, 10 Mbp
 - Couverture: 10, 30, 100

3 situations

A rendre avant le 15/11/2019

- 1. Le code source avec (beaucoup de) commentaires
- 2. Explications sur la stratégie employée
- 3. Rapport sur le comportement du programme en fonction de la taille des données
 - Temps d'exécution
 - Le nombre de lectures sélectionnées

Envoyez à <u>lavenier@irisa.fr</u>:

- Code source (en C)
- Le rapport (format pdf)