

<b>université BORDEAUX</b>	ANNÉE UNIVERSITAIRE 2024 - 2025 SESSION 1 DE PRINTEMPS 2025	Collège Sciences et Technologie  Masters
	MENTION : BIOINFORMATIQUE Code UE : 4TBI808 Intitulé de l'épreuve : Nouvelles générations de séquençage  Date : 24/04/2025      Heure : 14h30      Durée : 1h30 Documents : non autorisés Épreuve de Mme Raluca Uricaru.	

## 1 Recherche exacte de motifs (9p)

Question 1.1 Quelle est la différence entre méthodes online et méthodes offline de recherche exacte de motifs ? Donnez plusieurs exemples de méthodes dans chacune des deux catégories. Lesquelles sont les plus pertinentes pour le cas de l'alignement de lectures sur un génome de référence ? Justifiez.

Question 1.2 On donne le texte  $t = \text{AGTTCATCG\$}$ . Construisez la table des suffixes (suffix array) pour ce texte, ainsi que sa transformée de Burrows-Wheeler (BWT). Détaillez les étapes de la construction pour chaque structure d'indexation.

Question 1.3 Pourquoi la BWT rend-elle la recherche d'un motif plus efficace ?

Question 1.4 Déroulez étape par étape la recherche du motif  $p = \text{TC}$  dans le texte  $t$  en utilisant la table de suffixes préalablement construite.

Question 1.5 En supposant que l'on construise un arbre des suffixes pour ce même texte  $t = \text{AGTTCATCG\$}$ , combien de feuilles cet arbre contiendrait-il ? A quoi correspondent ces feuilles ?

## 2 Graphes et assemblage (11p)

On considère les lectures suivantes : GATTACA, TTACAG, ACAGT, CAGTT.

Question 2.1 Construisez le graphe de De Bruijn pour  $k = 4$ . Énumérez les arêtes et les nœuds et dessinez le graphe obtenu. Proposez un chemin dans le graphe pour reconstruire une séquence consensus (résultat de l'assemblage des 4 lectures).

Question 2.2 Dessinez maintenant le graphe de chevauchement (overlap graph) pour ces mêmes lectures avec un chevauchement minimal de 1 entre les lectures. Quels sont les avantages et les inconvénients de cette représentation comparée au graphe de De Bruijn ?

Question 2.3 Écrivez une fonction `deBruijnize(R, k)` qui prend en entrée un ensemble de lectures  $R$  et une valeur  $k$ , et qui construit le graphe de De Bruijn associé. On supposera que chaque lecture est plus longue que  $k$ . Les nœuds sont tous les mots de taille  $k$  extraits des lectures. Les arêtes relient deux mots de taille  $k$  qui se succèdent dans une lecture, c'est-à-dire se chevauchant sur  $k-1$  caractères. La fonction devra renvoyer une liste d'arêtes (paires de nœuds) et un ensemble (set) de nœuds, chaque nœud apparaissant une seule fois.

Exemple d'utilisation de set :

```
nodes = set()
nodes.add("AGC")
```