

Contrôle II

BLAST (3 points)

On considère les deux séquences ATTCATTCATTCATTCATTCATTCATTCATTC et ATTGATTGATTGATTGATTGATTGATTGATTG. Quel est, à première vue, leur pourcentage d'identité ? Quand on fait un alignement avec l'algorithme de BLAST, aucune similarité n'est trouvée. Pourquoi ?

Arbre des suffixes (7 points)

Question 1. Construisez l'arbre des suffixes généralisé des trois séquences AAT et TAAA et TAAT.

Question 2. Soient U et V deux séquences. Proposez un algorithme linéaire qui permette de déterminer le plus grand préfixe de U qui soit suffixe de V .

Structures secondaires d'ARN (6 points)

Question 1. On replie la séquence d'ARN GAUUAGC suivant l'algorithme de Nussinov (par minimisation d'énergie), avec les coefficients $\alpha(A, U) = -2$, $\alpha(C, G) = -3$ et $\alpha(G, U) = -1$. Tous les autres types d'appariements sont exclus. Enfin, on autorise l'appariement de la base i avec la base j si, et seulement si, $i < j$. La table de programmation dynamique obtenue est la suivante :

0	0	-2	-3	-4	-4	-7
0	0	-2	-2	-4	-4	-7
0	0	0	0	-2	-3	-5
0	0	0	0	-2	-2	-5
0	0	0	0	0	0	-3
0	0	0	0	0	0	-3
0	0	0	0	0	0	0

Il y a-t-il une structure optimale où la première et la dernière bases sont appariées? Si oui, laquelle ? Vous justifierez à partir de la table de programmation dynamique (vous pouvez annoter la grille sur la feuille, et rendre cette feuille).

Question 2. On peut représenter une structure secondaire sans pseudo-nœuds par un arbre. Donnez l'arbre correspondant à la structure

$$(1, 16), (3, 7), (4, 6), (10, 15)(11, 14)$$

sur la séquence GAACGGUCAGUAUAUC. Comment étendre cette représentation pour autoriser l'appariement (5, 13), qui forme un pseudo-nœud ?

Prédiction de gènes (4 points)

Dans le programme *Ecoparse*, le chevauchement de gènes est géré explicitement quand il s'agit de "petits" chevauchements, qui se produisent entre le codon stop du premier gène et le codon start du gène suivant. Quels sont alors les chevauchements possibles ?

Les chevauchements de taille 4 sont gérés avec l'expression NN[AG]TGANN. Expliquez précisément comment cette expression a été obtenue.