

AAGB - Devoir de demi-semester

NB. On parle de **matrice de distance** dans le cas des algorithmes UPGMA et NJ car la valeur en ligne i et colonne j d'une telle matrice correspond à la distance entre les espèces i et j . En revanche, on parle de **matrice de coût** dans le cas de l'algorithme de Needleman et Wunsch car cette matrice donne l'information du coût de remplacer le nucléotide A par le nucléotide T par exemple ou l'acide aminé Lysine (K) par l'acide aminé Glutamine (Q).

Exercice 1 : Alignement de séquences

1. Qu'est-ce-qu'une matrice de coût ? Quelles données pour la construire ? Comment la construire ? Donner un exemple.
2. Comment comparer deux séquences ? Quelle est l'utilité de la matrice de coût pour cela ? Citer deux méthodes vues en TD pour comparer deux séquences.
3. Définir orthologie, paralogie, analogie et homologie. Pourquoi chercher des homologies entre séquences ?

Exercice 2 : Reconstruction d'arbres phylogénétiques

1. Définissez l'horloge moléculaire et expliquer pourquoi on emploie ce terme pour l'algorithme UPGMA.
2. Quelle est la partie de l'algorithme Neighbour-Joining qui permet une amélioration par rapport à UPGMA ? Pourquoi ?
3. Dérouler l'algorithme Neighbour-Joining pour la matrice de distance suivante :

	A	B	C	D
A	0	5	4	7
B	5	0	5	6
C	4	5	0	3
D	7	6	3	0