

Sujet DM 1 :

A partir d'un fichier contenant N séquences alignées au format ClustalW, vous allez récupérer les séquences puis les afficher avec dessous les symboles suivants :

- * → identité stricte sur toute la colonne ;
- : → au moins 70 % de caractères identiques ;
- . → au moins 40 % de caractères identiques ;

Dans un deuxième temps, vous allez calculer deux à deux le nombre de différences dans les alignements, et présenter cette matrice des distances sous la forme d'une matrice à deux entrées.

Sujet DM 2 :

A partir de deux segments nucléotidiques de longueur inférieure à 15, vous allez créer un algorithme qui fait un DotPlot des deux séquences avec $w=3$ et un seuil demandé à l'utilisateur mais compris entre [1 et 3]. Vous présenterez le résultat sous la forme d'une matrice à double entrée avec 1 croix par fenêtre au dessus du seuil.

Sujet DM 3 :

A partir d'une séquence nucléotidique, vous recherchez tous les orfs dans les 3 phases directes, ainsi que celles sur le brin reverse complémentaire. Vous donnerez les coordonnées de tous les orfs trouvés ainsi que leur séquence, leur longueur et leur traduction. (inspirez vous de orfinder)

Sujet DM 4 :

Ecrire un programme qui, à partir d'un enzyme de restriction, donne le résultat de la digestion d'une séquence par cette enzyme. C'est à dire que le programme donne le nombre de fragments et leur longueur. Il pourra donner aussi, les fragments triés par longueur (comme dans une migration). Proposer un menu pour ça.

Dans un deuxième temps, la digestion se fera à partir d'une ou plusieurs enzyme que choisira l'utilisateur.

Sujet DM 5 :

Ecrire un programme qui calcule à partir de deux séquences nucléotidiques la matrice des coûts d'alignements (programmation dynamique). L'utilisateur fixe les paramètres (score identité, score mismatch, pénalité ouverture gap). On ne tient compte que d'une pénalité d'ouverture de gap, et pas de l'élongation du gap. Deux gaps adjacents sont considérés comme deux ouvertures de gap.