

Détection de phases ouvertes de lecture et prédiction de gènes

Mathilde Carpentier

29 septembre 2022

Présentation

L'objectif de cette série d'exercice est de vous constituer une bibliothèque de manipulation de séquences biologiques et de détecter des gènes codant pour des protéines (CDS) dans des génomes bactériens. Dans un premier temps nous allons écrire les fonctions de manipulations de séquences, de lecture de fichier puis nous détecterons dans un premier temps toutes les phases ouvertes de lecture ou ORF (*open reading frame*), c'est-à-dire toutes les sous séquences sans codon *stop*, encadrées par deux codons *stop* en phase. Nous chercherons ensuite les start et testerons différentes méthodes pour filtrer les CDS prédits.

Ce projet est découpé en plusieurs parties, chacune correspondant à un fichier .c :

Première partie dans `ManipSeqSimple.c` Fonctions simples de manipulation des séquences ADN sous forme de chaîne de caractères, manipulation de tableaux de caractères. Cette partie correspond au cours 2 (gestion de la mémoire en C). Elle est à réaliser en 2 séances de TP.

Deuxième partie dans `ManipSeqADN.c` Manipulation de la structure pour ces séquences. ,
Lecture des séquences ADN

Troisième partie dans `LectureADN.c`

Quatrième partie dans `ManipORF.c` Manipulation des structures pour les ORFs et implémentation d'une liste

Cinquième partie dans `DetectionORFs.c` Recherche des ORFs puis Comparaison des orfs avec les "vraies" orfs, filtre des orfs, modification du start