## TD6 Annotation d'une séquence bactérienne

## (partie II)

Dans le TD précédent vous avez commencé à annoter une séquence d'ADN. Nous vous avons demandé de faire un schéma récapitulatif pour bien comprendre l'annotation réalisée.

Nous allons aujourd'hui tenter de mieux comprendre la structure du ou des gènes annotés ainsi que leur(s) fonctionnalité(s).

1. Sachant qu'il s'agit d'une séquence d'ADN de la souche O145:H28 str. RM12581 d'*E. coli* nous allons observer la structure de son génome sur le site du « National Center of Biotechnology Information ».

http://www.ncbi.nlm.nih.gov/nuccore/CP007136.1

La région du génome que vous êtes en train d'annoter se situe entre les nucléotides 544500 et 551500, nous allons nous intéresser seulement à cette région. Pour cela, en haut à droite dans « Change region shown » choisir vos début et fin et cliquez en haut de page sur « Graphics ». Vous observez alors la structure des 5,6Mpb du génome de cette souche d'*E. coli*. Sinon vous pouvez zoomer dans la région d'intérêt il vous faut à l'aide de la souris sélectionner la région que vous voulez agrandir, puis dans la fenêtre qui s'ouvre sélectionner « zoom on range ». Refaite plusieurs fois cette opération pour bien arriver dans la zone des nucléotides 544500 à 551500. Vous pouvez également glisser cette fenêtre de droite à gauche pour mieux affiner votre recherche en cliquant avec la souris (maintenir le clic) et déplaçant celle-ci de gauche à droite.

Comparez les annotations que vous venez de faire avec celles trouvées sur ce site notamment concernant les 4 CDS. Que pouvez-vous dire ? De quels gènes s'agit-il ?

2. Vous avez déjà entendu parler dans vos cours de L2 de cet opéron, ainsi expliquez comment s'effectue le contrôle négatif de l'expression de cet opéron ?

3. Il se trouve qu'il existe d'autres types de contrôle de cet opéron avec des protéines régulatrices. Aidez-vous de la base de données Wikipédia (ou autre) pour savoir le fonctionnement exact de régulation de cet opéron en tapant les mots clés sur votre navigateur internet.

Vous pouvez ainsi, dans un premier temps, rechercher les signaux de régulation qu'il vous manque pour finaliser l'annotation de cette séquence à l'aide de DNA Pattern: <a href="http://www.bioinformatics.org/sms/dna">http://www.bioinformatics.org/sms/dna</a> pattern.html.

Les signaux à rechercher doivent avoir la syntaxe suivante :

- pour <u>l'opérateur</u>:
  /AATTGTGAGCGAATAACAATT/ (opérateur)
- pour le <u>site CAP</u> : /AATTAATGTGAGTCAGCTCACTCAT/ (site CAP)

## Ensuite, pour vous aider à comprendre ce que l'on appelle la diauxie répondez à ces questions une à une :

- 1/ Que se passe-t-il en présence de glucose et de lactose ?
- 2/ Que se passe-t-il en présence de glucose mais pas de lactose ?
- 3/ Que se passe-t-il en l'absence de glucose et de lactose?
- 4/ Que se passe-t-il en l'absence de glucose mais en présence de lactose ?
- <u>4. Exercice complémentaire</u> (en fonction du temps): programme python pour réaliser l'annotation.

Vous allez ici à l'aide d'un programme python rechercher les motifs de type boite -35, boite -10, RBS, opérateur et site CAP qui sont contenus dans un fichier appelé MotifsQ7.txt dans la séquence contenue dans le fichier SequenceQ7.txt. Attention la séquence d'ADN à annoter et les motifs sont identiques à ceux du TD ci-dessus.

Trouvez-vous les mêmes positions qu'avec dna pattern.html?