

/TATGTT/ (boite -10)

Parmi les positions trouvées quelles sont celles qui vous paraissent les plus vraisemblables et pourquoi (revoir pour cela la structure des gènes bactériens) ? Pouvez-vous lever l'ambiguïté entre les prédictions de GeneMark et ORF Finder ?

4. Afin de mieux caractériser la structure de gènes trouvés, vous allez maintenant rechercher et positionner les signaux de traduction, à savoir le RBS pour « Ribosome Binding Site » ou séquence de Shine-Dalgarno. Nous allons pour cela utiliser le même site DNA Pattern :

http://www.bioinformatics.org/sms/dna_pattern.html

La séquence à rechercher ici est une séquence consensus dont la syntaxe exacte à copier/coller est :

/AGGA/ (RBS)

Parmi les positions trouvées quelles sont celles qui vous paraissent les plus vraisemblables par rapport à la structure de(s) gène(s) que vous avez supposée être jusqu'ici et pourquoi ? Pouvez-vous lever l'ambiguïté entre les prédictions de GeneMark et ORF Finder ?

5. 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>

Sachant également que 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 cliquez en haut de page sur « Graphics ». Vous observez alors la structure des 5,6Mpb du génome de cette souche d'*E. coli*. Pour zoomer dans la région d'intérêt il vous faut à l'aide du « Find » taper la région d'intérêt (544500-551500). Vous pouvez également glisser cette fenêtre de droite à gauche à l'aide des flèches et donc déplaçant celle-ci de gauche à droite.

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

6. Maintenant que vous avez pu constater de quel(s) gène(s) il s'agit, aidez-vous de la base de données Wikipédia pour savoir le fonctionnement exact de régulation de cet opéron en tapant les mots clés sur votre navigateur internet. Vous pouvez ainsi, pour conclure, rechercher les signaux de régulation qu'il vous manque pour finaliser l'annotation de cette séquence à l'aide de DNA Pattern : http://www.bioinformatics.org/sms/dna_pattern.html.

Les signaux à rechercher doivent avoir la syntaxe suivante :

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

Finir l'annotation de votre séquence et faire un schéma récapitulatif du fonctionnement de cet opéron (*Rq : vous reverrez en partie la régulation de cet opéron en génétique Bi406 !!!*).